Impacts of 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

September 5, 2025

Abstract

Roads, buses, parks and other urban public goods aim to improve lives. However, because this infrastructure is built in places not assigned to people, its benefits depend on how mobile residents select into affected areas. We present a framework to clarify how impacts on place relate to impacts on people. We then study labor market impacts of Dar es Salaam's BRT, using novel data that tracks places and people. Those who select in are more likely employed commuters ex-ante, and are not motivated by labor market gains. People effects are smaller than place effects, but both reveal labor market gains.

Keywords: Transport, Urbanization, Gentrification, Tanzania, House Prices
JEL Classification: R40, R31, 018

*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. We thank EDI for assistance with initial fieldwork. We thank Yonas Mchomvu and Sveta Milusheva at the World Bank and Ronald Lwakatare at the Dar es Salaam Rapid Transit Agency for institutional support. We thank Nancy Lozano Garcia for sharing the World Bank LSMS data. We thank audiences at numerous conferences and seminars. We also thank Ed Glaeser, Allan Hsiao, Gabriel Kreindler, and Steve Redding who provided insightful discussions. 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, Betsenat Gebrewold, and Somei Miyashita provided excellent research assistance. Nano Barahona, Chris Becker, Adam Cole, Sally Zhang, and Tom Zohar participated in fieldwork and provided excellent discussion and feedback. 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

Cities are engines of growth and a refuge for the poor (Glaeser, 2012). They reflect a balance between two externalities: agglomeration economies, which raise productivity and amenity and draw people in; and congestion externalities, which reduce health, safety, and effective density, pushing people out. A central role of urban governance is providing local public goods— such as parks, transport, and police stations—that mitigate congestion and enable more people to benefit from agglomeration, a role that is more pressing in the developing world's congested but growing cities (Bryan et al., 2020, 2025).

Local public goods, however, are built in places, not given to people. As a result, their *incidence* (who benefits), and *impact* (how behaviors change) depend on the selection of mobile residents into, and out of, targeted neighborhoods. This selection can reinforce or undermine policy goals, and cause impacts on people and places to differ. For example, a transit line designed to help poor local residents find jobs may be valued most by already employed wealthy commuters, leading to rent increases that price out low-income households (e.g., Balboni et al., 2021). While employment rates in targeted locations may rise, the impact of the infrastructure is to reduce commute times for wealthy newcomers, rather than increase employment for targeted incumbents. Residential mobility determines the importance of these effects (see, e.g., Busso et al. 2013).

While a growing literature studies the impact of infrastructure, we have little direct evidence on how selection shapes incidence and impact. Existing studies either evaluate average impacts on *places* (e.g., Gibbons and Machin, 2005; Donaldson and Hornbeck, 2016), infer distributional effects via structural models (e.g., Tsivanidis, 2024; Couture et al., 2024), or study incidence using sufficient statistics approaches (Busso et al., 2013). More direct approaches are hampered by data scarcity: available urban data sets—especially in developing countries—are typically repeated cross-sections, which preclude tracking of movements in to and out of targeted areas.

We address this data challenge in the context of the first phase of the Dar es Salaam Bus Rapid Transit (BRT) system, a World Bank funded project aiming to provide efficient, green, and inclusive mobility (World Bank, 2022). A specific goal was to improve residents' access to work, and we study how selection affects the ability of the BRT to achieve its labor market objectives.

We collected bespoke data, with four features. First, we surveyed two individuals (one male and one female) in each of 1,748 households living throughout the city before the BRT opened—one of the largest samples of its kind in a developing country city. Second, we tracked this baseline sample and conducted a follow-up survey three years later, including surveying movers. Third, when baseline individuals or households moved out, we surveyed the in-movers, and collected retrospective measures of their pre-BRT outcomes. Finally, the data set is geo-coded and can be linked to pre- and post-BRT travel time data, providing granular measures of exposure to the new infrastructure. This dataset provides an unusually detailed lens into changes in outcomes for places and people in a fast-growing developing country city, where data is typically scarce (Franklin et al. 2024 present the only similar data we know of, which covers Addis Ababa).

We develop a difference-in-differences framework that allows us to to define and estimate relevant selection and causal parameters. Selection can be based on baseline outcomes (e.g., those who already commute select in), or anticipation of endline outcomes (e.g., those for whom the BRT will have the largest impact on commuting select in, what Heckman et al. (2006) call "essential heterogeneity"). Estimating these parameters allows us to understand who demands access to the BRT and why. While a repeated cross section can be used to study selection on demographic characteristics, estimating selection on outcomes at baseline requires our retrospective data, and estimating selection on changes in outcomes requires our tracking data. These selection parameters, combined with measures of population mobility, determine how place- and people-based estimates diverge. We study two policy-relevant groups of people: initial residents (those who stay in a location and those who exit) are those who were targeted; final residents (those who stay in a location and those who arrive) are those who ultimately have access to the BRT.

Impacts on initial residents are identified under a standard parallel trends assumption (e.g., Gibbons and Machin, 2005; Faber, 2014). Our framework shows that initial resident impacts can be decomposed into the causal effect on place, less selection on *endline* outcomes, multiplied by the probability of exit. As such, an increase in employment in a given location can be consistent with no impact on initial residents if those who selected in were more likely to be employed at endline and the population is very mobile. In this case, the endline outcomes of arrivers give the impression of improvements in employment that need not have accrued to initial residents.

Identifying impacts on final residents requires a stronger parallel trends assumption that rules out selection into affected area on pre-existing trends, consistent with assumptions invoked in a wider literature on migration (e.g., Hamory et al., 2021). The impact on final residents is then equal to the place-based effect less selection on *baseline* outcomes, multiplied by the probability of exit. A positive impact on employment for an affected location is in this case consistent with a zero effect for final residents if those who selected in were already working at baseline: a classic selection effect. Under the same parallel trends assumption, the difference between final and initial resident estimates reveals whether those who selected in to affected locations had larger treatment effects than those who selected out and can be used to examine what motivates demand for the BRT.

Applying this framework to our novel data yields four key findings. First, consistent with existing results (Duranton and Turner, 2012; Asher and Novosad, 2020; Tyndall, 2021), the BRT improves labor market outcomes and increases rents in affected *places*. The most strongly treated locations[1] see a 3.7 pp. increase in the employment rate relative to a mean of 68%, a corresponding 5% increase in household income, and a 31% increase in self-reported rent per room.

Second, the BRT is demanded by working people with higher ex-ante incomes. Relative to those who depart, those who arrive in a highly-treated locations have about 0.2 years more education, are 1 percentage point more likely to be employed at baseline, earn 3,700 TSH more (relative to a mean of 27,500 TSH), and are (an insignificant) 1 percentage point more likely to commute. Differences are uniformly larger for homes rented at baseline; for instance, entrants into rented homes are a statistically significant 5 percentage points more likely to commute at baseline.

Third, despite a highly mobile population – 30% of individuals and 50% of renters move between baseline and endline – differences between place and people impacts are small and statistically insignificant. This reflects the fact that, while mobility is high, 70% of individuals are still inframarginal, staying in place and receiving the direct impact of the infrastructure (Busso et al., 2013). People impacts are, however, uniformly smaller than place impacts. While place estimates show a 3.8 pp. increase in employment, this is reduced to 2.9 pp. for final residents, and 3.7 pp. for initial residents. For household income, final resident effects are 36% smaller than place effects and initial resident effects are 24% smaller. Household income results are perhaps most interesting because, while the place effect is statistically significant, neither of the people estimates is: as a result, an analyst using place-based estimates to infer causal effects on people may come to the wrong conclusions. As expected, the differences between place- and people-based effects is larger for renters: for instance, place-based estimates of a 7.6 percentage point increase in employment falls to 5.2 percentage points for final residents and 6.9 percentage points for initial residents.

Fourth, final resident impacts are close to and, if anything smaller than, initial resident effects. Hence, labor market impacts for those who select in to treated locations are no larger than effects for those who move out. This is inconsistent with demand for the BRT being driven by a desire to

[1] To aid interpretation, we define the most strongly treated locations as those that see a greater than 80th percentile reduction in our measure of travel time change induced by the opening of the BRT line.

increase employment, an intuitive result given that arrivers are already more likely to work.

Overall, our results show that BRT access is valued by those who are already employed. These marginal entrants into BRT affected areas are not seeking to improve labor market outcomes, but may benefit from other amenities such as reduced commute times. This type of selection reduces the effectiveness of the new infrastructure in achieving its target of improving labor market outcomes. Despite strong evidence for these selection effects, and high mobility rates, there are still a large number of inframarginal stayers in targeted locations who do see labor market benefits.

We contribute to several literatures. We complement a small but growing set of papers that collect bespoke data to study urban issues in developing countries (Franklin et al., 2024; Harari and Wong, 2021; Gechter and Tsivanidis, 2023; Bryan et al., 2025).

We contribute direct evidence on selection that complements more indirect approaches in a range of settings. A growing literature uses quantitative spatial models to study the impact of transport on aggregates outcomes and inequality (Allen and Arkolakis 2022; Heblich et al. 2020; Zárate 2023; Gaduh et al. 2022; Tsivanidis 2024; Barwick et al. 2024; Couture et al. 2024; Chapelle and Ubeda 2025).[2] A separate literature studies the incidence of local investments (for example, Busso et al. 2013; Gaubert et al. 2025; Lu et al. 2019; Abeberese et al. 2024), and a related literature highlights the importance of accounting for changes in *demographic* composition when estimating causal impacts (Collins and Shester, 2013; González-Pampillón et al., 2020; Weiwu, 2024). A wider literature examines sorting behaviors over space (Albouy and Faberman, 2025).[3] A subset focuses specifically on gentrification (Gechter and Tsivanidis, 2023; Almagro et al., 2024; Almagro and Domínguez-Iino, 2024; Brummet and Reed, 2021; Couture et al., 2024). These papers use structural assumptions, or assume selection on demographics, to deal with selection.

Perhaps most closely related, Autor et al. (2025) study the causal impact of the China shock using tax data that forms a panel. While their focus is different, they have a related emphasis on the difference between impacts on people and impacts on place.

The paper is structured as follows. Section 2 outlines our framework. Section 3 describes the setting and data and shows pre-trends. Section 4 discusses our central results. Section 5 considers the implications of our findings for researchers who wish to account for the selection effects using less detailed data. Section 6 concludes.

[2]Where panel data does exist, it tends to be sourced from administrative datasets which typically contain limited covariates. For example, Warnes (2024) uses electoral registration data in Buenos Aires which captures occupation but not direct employment outcomes.

[3]Neumark and Kawaguchi (2001) consider how bias in longitudinal estimates may arise where panel data sets do not follow movers and therefore exhibit attrition, using the example of the US Current Population Survey.

2 Framework

This section shows how selection links impacts on place and impacts on people, and provides a framework for identification. We start with a simple example to fix ideas and introduce notation, and then develop the framework more fully and discuss parallel trends assumptions. Throughout we consider the impact of the BRT to be its total impact, which would include, for example, a direct impact due to better commute options as well as an indirect impact coming from a change of neighbourhoods induced by the BRT's opening.

2.1 Intuition and notation: place-level versus person-level changes

We begin with a simple framework to clarify how selection links changes in outcomes observed at the level of a *place* to changes in outcomes for the *people* who live there. We also describe how different selection parameters reveal information about the demand for places. This section serves to build intuition and introduce notation; causal interpretation is discussed in the following subsections.

Let y_{hif}^{BL} denote an outcome, measured at baseline, for household h who lived at initial location i, and later moved to final location f. We will think of y as a measure of earnings throughout this subsection for illustrative purposes. Let y_{hif}^{EL} be earnings measured for the same household at endline. Table A summarizes the measured data for the case of a household h who initially lived in place p and moved to f, and a household h' who initially lived in place i and moved to place p. It also highlights the source of data, showing which observations can be taken from a standard repeated cross section consisting of a baseline and endline, and which require our retrospective or tracking data.

Table A: Example of data collection associated with place p

	Baseline			Endline		
	Observation	Place	Data	Observation	Place	Data
Initial resident (exiter)	y_{hpf}^{BL}	p	BL	y_{hpf}^{EL}	f	Tracking
Final resident (arriver)	$y_{h'ip}^{BL}$	i	Retro	$y_{h'ip}^{EL}$	p	EL
Place	y_{hpf}^{BL}	p	BL	$y_{h'ip}^{EL}$	p	EL

The initial resident change $\Delta y_{hpf} = y_{hpf}^{EL} - y_{hpf}^{BL}$ measures the change in earnings for the person originally living in p. The final resident change $\Delta y_{h'ip} = y_{h'ip}^{EL} - y_{h'ip}^{BL}$ measures the change in earnings for the person living in p at endline. The place-level change $\Delta y_{hh'p} = y^{EL}_{h'ip} - y^{BL}_{hpf}$ records the measured change in earnings observed at place p.

The place-level change can be decomposed into two intuitive components: a *person-level change* and a *composition effect* created by selection. We obtain a decomposition via final resident by adding and subtracting $y^{BL}_{h'ip}$:

$$\begin{aligned}
\Delta y_{hh'p} &= y^{EL}_{h'ip} - y^{BL}_{hpf} \\
&= \left(y^{EL}_{h'ip} - y^{BL}_{h'ip}\right) + \left(y^{BL}_{h'ip} - y^{BL}_{hpf}\right) \\
&= \underbrace{\Delta y_{h'ip}}_{\text{Final resident change}} + \underbrace{\Sigma y^{BL}_{hh'p}}_{\text{Baseline composition effect}} .
\end{aligned}$$

A change in earnings at place p results from a change in earnings for the *final resident*, plus the baseline difference in earnings between initial and final residents. The final resident change tells us about how earnings change for a person, while the composition effect tells us about the earnings of the people who value living in p. If p is affected by infrastructure, the composition effect reveals the characteristics of those who demand the infrastructure: for instance, do those who value the BRT have high or low earnings? Table A clarifies that our retrospective data is required to measure either of these objects.

Alternatively, adding and subtracting y^{EL}_{hpf} yields a decomposition via initial resident:

$$\begin{aligned}
\Delta y_{hh'p} &= y^{EL}_{h'ip} - y^{BL}_{hpf} \\
&= \left(y^{EL}_{hpf} - y^{BL}_{hpf}\right) + \left(y^{EL}_{h'ip} - y^{EL}_{hpf}\right) \\
&= \underbrace{\Delta y_{hpf}}_{\text{Initial resident change}} + \underbrace{\Sigma y^{EL}_{hh'p}}_{\text{Endline composition effect}}
\end{aligned}$$

where the endline composition effect is the difference in earnings between the arriver and the exiter, measured at endline. Again the first term tells us about earnings changes for a person, while the composition effect tells us about the endline earnings of those who want to live in p. Both are directly observable only with our tracking data, as shown in Table A.

The endline composition term can also be expressed as:

$$\Sigma y^{EL}_{hh'p} = \Delta y_{h'ip} - \Delta y_{hpf} + \Sigma y^{BL}_{hh'p}.$$

Substituting yields:

$$\Delta y_{hh'p} = \underbrace{\Delta y_{hpf}}_{\text{Initial resident}} + \underbrace{(\Delta y_{h'ip} - \Delta y_{hpf})}_{\text{Trends composition}} + \underbrace{\Sigma y^{BL}_{hh'p}}_{\text{Baseline composition}} .$$

The new term, labeled trends composition, compares the time trends of arrivers and exiters, and reveals a different characteristic of those who value location p. For example, when this term is positive we can infer that those who value the BRT are on a more positive income trend. We provide assumptions below under which this term can be interpreted as the relative causal effect of infrastructure affecting p, in which case the trends composition term reveals whether those who demand infrastructure do so because it will increase earnings differentially.

2.2 Identifying causal and selection effects

This section shows more formally how we identify the causal effect of new transport infrastructure on people—both initial and final residents— and place, as well as the composition effects discussed above.

Different assumptions are required for different estimates, but all identification assumptions take the form of a *conditional parallel trends* assumption. Some notation is helpful to state these assumptions. Assume the city is composed of N locations and let $\Delta T = \{\Delta T_1, \ldots, \Delta T_N\}$ denote the change in travel times to the central business district caused by the new transport infrastructure. Each household h lives in an initial location i at baseline (prior to opening of the new transport infrastructure) and moves to an endogenous final location $f(\Delta T)$ at endline.[4] The potential change,

$$\Delta y_{hif(\Delta T=0)}(\Delta T = 0) = \left(y^{EL}_{hif(\Delta T=0)} - y^{BL}_{hif(\Delta T=0)} \right) (\Delta T = 0),$$

denotes the baseline to endline change in outcome y that would have occurred in the counterfactual scenario in which the new transport infrastructure had not been built (i.e., $\Delta T = 0$).

Initial resident treatment effects

Consider the regression:

$$\Delta y_{hif} = \alpha^I + \beta^I \Delta T_i + g(x^{BL}_{hif}) + \epsilon_{hif}, \tag{1}$$

[4] It may be that $i = f(\Delta T)$. In this case h is termed a 'stayer'.

where ΔT_i is the change in travel times to the CBD caused by the BRT and x_{hif}^{BL} denotes baseline characteristics of the initial resident. Identification of β^I, the initial resident treatment effect, requires that $\epsilon_{hif} \perp\!\!\!\perp \Delta T_i$. Stated as a parallel trends assumption this requires that

$$\Delta y_{hif(\Delta T=0)}(\Delta T = 0) \perp\!\!\!\perp \Delta T_i \mid x_{hif(\Delta T=0)}^{BL}. \tag{CPT1}$$

Intuitively, this requires that, conditional on the observable baseline characteristics of residents, the BRT was not targeted at locations where residents would have experienced different trends even in the absence of the infrastructure, a common assumption used in the transportation literature (e.g., Gibbons and Machin, 2005; Faber, 2014). The assumption allows for a setting in which those planning the new transport infrastructure had targeted the lines based on characteristics of the population, but did not have access to information on population trends. In line with much of the reduced form literature on transportation impacts, we focus on estimating relative impacts; further stable unit treatment value assumptions would be required to estimate aggregate impacts.

Final resident treatment effects

Consider the related regression:

$$\Delta y_{hif} = \alpha^F + \beta^F \Delta T_f + g(x_{hif}^{BL}) + \epsilon_{hif}, \tag{2}$$

where ΔT_f is the change in travel times at place f caused by the BRT and x_{hif}^{BL} is baseline characteristics of the final resident. Again identification requires that $\epsilon_{hif} \perp\!\!\!\perp \Delta T_f$, or stated as a parallel trends assumption

$$\Delta y_{hif(\Delta T=0)}(\Delta T = 0) \perp\!\!\!\perp \Delta T_{f(\Delta T)} \mid x_{hif(\Delta T=0)}^{BL}. \tag{CPT2}$$

This requires that the BRT was not targeted at places where final residents would have been trending differently even in the absence of the BRT, *and* that final residents who would have trended differently even in the absence of the BRT did not select into BRT-affected locations. While this assumption is stronger than CPT1, it is consistent, for example, with assumptions invoked in a large literature that uses fixed effect estimators to examine migration (e.g., Hamory et al., 2021).

Place-based treatment effects

A place p may have different household residents in the baseline and endline data. To formalize this, let each location $p \in N$ have a potentially different resident at baseline (h) and endline (h'), and let

$$\Delta y_{hh'p}(\Delta T = 0) = \left(y^{EL}_{h'ip(\Delta T=0)} - y^{BL}_{hpf(\Delta T=0)} \right) (\Delta T = 0)$$

be the change in outcomes that would have occurred in place p if the new transport infrastructure had not been built. Authors studying the causal effect of infrastructure on place use a variety of identification assumptions. A common example (e.g., Gibbons and Machin, 2005; Faber, 2014), which we follow here, is that, conditional on the baseline covariates of *initial* residents, the new transport infrastructure was not targeted at places where outcomes would have evolved differently in the absence of the new infrastructure:

$$\Delta y_{hh'p}(\Delta T = 0) \perp\!\!\!\perp \Delta T_p \mid x^{BL}_{hpf}. \tag{CPT3}$$

The implication is similar to (CPT1).

With this assumption, the regression:

$$\Delta y_{hh'p} = \alpha^P + \beta^P \Delta T_p + g(x^{BL}_{hpf}) + \epsilon_{hh'p}. \tag{3}$$

identifies the relative causal effect of the new transport infrastructure on place.

Identifying selection effects

Selection effects determine the difference between people-based and person-based treatment effects, and reveal the nature of demand for the new infrastructure. We derive these results in the notation of our more detailed framework, and then discuss identification.

The set of people living in location p can be decomposed into three groups: stayers $S_p = \{h : i = p, f = p\}$; exiters $E_p = \{h : i = p, f \neq p\}$; and arrivers $A_p = \{h : i \neq p, f = p\}$. Denote $\mathbb{1}^{S_p}_h$ to be an indicator that takes the value one when h is in the stayer set for place p (and similarly for exiters and arrivers), and $\pi_{h \in S_p}$ to the probability that households h is in set S_p.

As discussed in Section 2.1, relating place-based estimates to final resident estimates yields:

$$\begin{aligned}
\Delta y_{hh'p} &= \mathbb{1}_h^{S_p}\left[y^{EL}_{h'\in S_p} - y^{BL}_{h'\in S_p}\right] + \mathbb{1}_h^{E_p}\left[y^{EL}_{h'\in A_p} - y^{BL}_{h\in E}\right] \\
&= \mathbb{1}_h^{S_p}\left[y^{EL}_{h'\in S_p} - y^{BL}_{h'\in S_p}\right] + \mathbb{1}_h^{E_p}\left[y^{EL}_{h'\in A_p} - y^{BL}_{h'\in A_p}\right] + \mathbb{1}_h^{E_p}\left[y^{BL}_{h'\in A_p} - y^{BL}_{h\in E_p}\right] \\
&= \underbrace{\Delta y_{h'p'p}}_{\text{Final resident change}} + \underbrace{\mathbb{1}_h^{E_p}\left[y^{BL}_{h'\in A_p} - y^{BL}_{h\in E_p}\right]}_{\text{baseline composition effect}}.
\end{aligned}$$

Hence, the change in outcomes experienced in a place is equal to the change for final residents, plus a selection term, $y^{BL}_{h'\in A} - y^{BL}_{h\in E}$ that applies only when the initial resident exits. In our difference-in-differences framework, this then implies:

$$E(\Delta y_{hh'p})(\Delta T) = \delta^P + \beta^P \Delta T = \delta^F + \beta^F \Delta T + \pi_{h\in E_p} \cdot \sigma^{BL} \Delta T$$

where $\sigma^{BL} = E\left[y^{BL}_{h'\in A_p} - y^{BL}_{h\in E_p}\right]$. This then implies

$$\beta^P - \beta^F = \underbrace{\pi_{h\in E_p}}_{\text{Mobility}} \cdot \underbrace{\sigma^{BL}}_{\text{BL Selection}} .$$

As a result, the difference between place-based and final resident treatment effects depends on: (i) differences in baseline outcomes between initial and final residents (the term labeled selection); and (ii) the proportion of initial residents that choose to exit (the term labeled mobility). Intuitively, the selection captures which individuals demands the new infrastructure, and the mobility term determines how far this demand alters impacts.

Using a similar approach to relate place-based to initial resident estimates:

$$\begin{aligned}
\Delta y_{hh'p} &= \mathbb{1}_p^{S_p}\left[y^{EL}_{h\in S_p} - y^{BL}_{h\in S_p}\right] + \mathbb{1}_p^{E_p}\left[y^{EL}_{h'\in A_p} - y^{BL}_{h\in E_p}\right] \\
&= \mathbb{1}_p^{S_p}\left[y^{EL}_{h\in S_p} - y^{BL}_{h\in S_p}\right] + \mathbb{1}_p^{E_p}\left[y^{EL}_{h\in E_p} - y^{BL}_{h\in E_p}\right] + \mathbb{1}_p^{E_p}\left[y^{EL}_{h'\in A_p} - y^{EL}_{h\in E_p}\right] \\
&= \underbrace{\Delta y_{hpp'}}_{\text{Initial resident change}} + \underbrace{\mathbb{1}_p^{E_p}\left[y^{EL}_{h'\in A_p} - y^{EL}_{h\in E_p}\right]}_{\text{endline composition effect}}.
\end{aligned} \quad (4)$$

Again, the change in outcomes experienced in a place is equal to the change for initial residents, plus a selection term, $y^{EL}_{h'\in A} - y^{EL}_{h\in E}$ that applies only when the initial resident exits. As above, substituting into our DID framework reveals that

$$\beta^P - \beta^I = \pi_{h \in E_p} \cdot \sigma^{EL},$$

where $\sigma^{EL} = E[y^{EL}_{h' \in A} - y^{EL}_{h \in E}]$. The difference between the place-based and initial resident estimates is again determined by the extent to which those who select in differ from those who select out (here in terms of endline outcomes), and mobility of the population.

As noted in section 2.1, the endline composition term can be further decomposed as follows

$$\begin{aligned} y^{EL}_{h' \in A_p} - y^{EL}_{h \in E_p} &= \left[y^{EL}_{h' \in A} - y^{BL}_{h' \in A}\right] - \left[y^{EL}_{h \in E} - y^{BL}_{h \in E}\right] + \left[y^{BL}_{h' \in A} - y^{BL}_{h \in E}\right] \\ \Rightarrow \pi_{h \in E_p} E\left[y^{EL}_{h' \in A_p} - y^{EL}_{h \in E_p}\right](\Delta T) &= \pi_{h \in E_p}\left(\underbrace{\left[\beta^A - \beta^E\right]}_{\text{Selection on cause}} + \sigma^{BL}\right)(\Delta T), \end{aligned}$$

where β^A is the causal effect of the new transport infrastructure for arrivers, while β^E is the causal effect for exiters and the second line follows from taking expectations and applying the parallel trends assumption (CPT2). The term $(\beta^A - \beta^E)$, which we refer to as "selection on causal effects", is positive if those who select in (arrivers) experience a larger gain because of the new transport infrastructure than those who select out (exiters). This term gives a clear idea of why people value infrastructure. If $\beta^A > \beta^E$ then those who select in have likely done so because they expect to see a gain in outcome y. If $\beta^A \leq \beta^E$ then this gain in outcome y is unlikely to have motivated moving in.

In summary, three selection parameters (baseline and endline composition, and selection on cause), combine with population mobility to determine how place-based and people-based effects differ. The selection terms capturing baseline composition and selection on cause reveal important information about demand for infrastructure and hence who benefits and how. We next discuss how we identify these selection parameters (the mobility parameter can be taken directly from data).

Each parameter can be directly estimated, or computed from a comparison of treatment effects. To directly estimate the baseline composition effect, consider the regression

$$y^{BL}_{h'ip} - y^{BL}_{hpf} = \delta^{BL} + \sigma^{BL}\Delta T_p + g(x^{BL}_{hpf}) + \eta_{h'hp}.$$

Identification requires that $\Delta T_p \perp\!\!\!\perp \eta_{h'hp}$ or, stated as a parallel trends assumption

$$[y^{BL}_{h'ip} - y^{BL}_{hpf}](\Delta T = 0) \perp\!\!\!\perp \Delta T_p | x^{BL}_{hpf}. \tag{5}$$

This requires that, conditional on the characteristics of baseline residents, the BRT was not targeted at places where the composition of the population was changing over time.

If (CPT2) holds, then

$$\sigma^{BL} = \frac{\beta^P - \beta^F}{\pi_{h \in E^p}} \tag{6}$$

and we can identify the baseline composition effect indirectly by comparing the place-based and final resident treatment effects.

The endline composition effect can be directly estimated from the regression

$$y^{EL}_{h'ip} - y^{EL}_{hpf} = \delta^{EL} + \sigma^{EL}\Delta T_p + g(x^{BL}_{hpf}) + \eta_{h'hp}.$$

under the relevant parallel trends assumption requiring that, conditional on the characteristics of baseline residents, the BRT was not targeted at places where the composition of the population was changing over time:

$$[y^{EL}_{h'ip} - y^{EL}_{hpf}](\Delta T = 0) \perp\!\!\!\perp \Delta T_p | x^{BL}_{hpf}.$$

The same parameter can be estimated indirectly by noting that, under (CPT1),

$$\sigma^{EL} = \frac{\beta^P - \beta^I}{\pi_{h \in E^p}}.$$

Finally, under assumption (CPT2), we can identify selection on causal effects by comparing initial and final resident estimates:

$$\begin{aligned} \beta^P &= \beta^I + \pi_{h \in E^p} \cdot \left([\beta^A - \beta^E] + \sigma^{BL} \right) \\ &= \beta^F + \pi_{h \in E^p} \cdot \sigma^{BL} \\ \Rightarrow&\ \pi_{h \in E^p}(\beta^A - \beta^E) = \beta^I - \beta^F. \end{aligned}$$

3 Empirical setting and data collection

This section describes the empirical setting in Dar es Salaam and our data collection.

3.1 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). Traffic congestion is a severe problem (Mpogole and Msangi, 2016), costing 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 being built in the city between 2005 and 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).

3.2 Data collection

We fielded 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 collected a baseline survey of 1,748 structures and interviewed one male and one female respondent from each. We then track along two dimensions during our endline survey three years later. We resurveyed baseline respondents, wherever they were living at endline (creating a panel of households and individuals), and we surveyed any new individuals residing in the original structures surveyed at baseline (creating a panel of structures). For all respondents, we also collected retrospective data on their key outcomes three years prior in order to obtain baseline observations for non-individuals in the panel not surveyed at baseline and to validate the retrospective data against baseline responses for those that were. Appendix Figure A1 outlines the timeline of survey and operational activities.[5]

3.2.1 Baseline survey

Our 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

[5]In addition to the primary baseline and endline survey rounds, we completed a midline attrition survey as well as several rounds of phone surveys, included in Appendix Figure A1. These additional surveys were used primarily to assist with tracking households.

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.[6] At each cluster location, a random walk was used to select 12-14 households to interview, yielding a total of 1,748 households that were available for and consented to 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 3,104 individuals. Throughout the paper we present unweighted statistics that represent the average effect across geography; in Appendix B we construct a household weight that matches the population distribution and home ownership distribution and show robustness of our results to this weighting.

3.2.2 Endline survey

Our endline survey was conducted from February-May 2019. All baseline respondents and structures were physically tracked as part of this survey. Hence, if a baseline household had moved out of its baseline structure, as long as they remained living in Dar es Salaam, 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. 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 A1 summarizes structure- and individual-level recontact rates at endline. Approximately three years after baseline, we locate 92% of our baseline structures; 89% of our initial households; and 85% of our initial individuals. We successfully survey 76% of our baseline structures; 80% of the initial households; and 70% of the initial individual residents. Our successful survey rates between baseline and endline conditional on having successfully located at least one member of the household during the midline attrition survey[7] are 88% for the structure, 92% for

[6] 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.

[7] During the midline attrition survey, we discovered that the initial survey firm had not accurately collected full

the household, and 81% for the individual respondents. We present balance tests for attrition across the three samples in Appendix Table A2. We do not find that distance to the BRT predicts who was successfully surveyed, but, as perhaps expected in locations with high mobility rates, 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.

Our sample frame was to interview one randomly selected male and female respondent from structures we enrolled at baseline. We then followed the male and female respondents if they moved, and attempted to re-enroll a new male or female to replace the mover. Our analysis dataset is therefore at the structure-by-gender level. Appendix Table A3 shows the coverage of our final dataset for measuring initial resident treatment effects (requires the initial resident to be successfully tracked between baseline and endline), final resident treatment effects (requires the initial resident to be surveyed and the replacement resident to be enrolled), and structure (requires the initial resident to be surveyed and the replacement resident to be enrolled). Out of a maximum sample of 3,496 structure-by-gender pairs, we have 322 cases where we did not have a baseline respondent. From this, we are able to estimate all three treatment effects (initial, final, and place) for 60% of baseline structures.[8] This will be our primary analysis sample.[9]

3.2.3 Measuring travel speeds

Our main treatment variable is change in travel time between a structure and the central business district (CBD). We discuss the construction of this variable in the next subsection, but it has two key ingredients: baseline travel speeds, and endline travel speeds along BRT routes. To measure baseline travel speeds we conducted a travel time survey in January 2016. Enumerators traveled between six points on the periphery of Dar es Salaam and the central business district by daladala (minibus), taxi, motorbike, and bajaji (rickshaw), recording GPS locations and time stamps along each journey route. A total of 812 trips were completed, spanning different days of the week and times of the day. These data were used to calculate average baseline travel speeds by public transport, where the latter averaged speeds across daladala (minivan) and bajaji (motobike) travel.

GPS coordinates of all structures, contributing to the 8% of structures that were not located. The other reasons for not completing the survey were household refusal at either midline or endline (10% of structures) and the structure being torn down or empty (5.8% of structures).

[8] In 24% of cases, we cannot compute any of the initial, final, or place estimates. This is primarily due to the structure not being found or the household refusing to be surveyed. For the remainder of the cases we can compute a subset of the three treatment effects.

[9] Sample sizes may differ across variables if a respondent answered "don't know", in which case the response is coded as missing.

To construct endline travel speeds on BRT routes, we obtained recorded departure and arrival times at stops along the BRT route between November 2017 - April 2018 from the Dar Rapid Transit Agency (DART).

3.3 Measuring exposure to the BRT

Exposure to the BRT is measured as the decrease in log travel time from the structure to the 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 the travel speed data described above. 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 public transport 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, since BRT lines are more likely to be constructed along radial routes that connect to 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 0.104. When interpreting coefficient magnitudes, we consider results at this value of the exposure variable (i.e., those households with a demeaned reduction in travel time to the CBD of more than 10.4% due to the BRT) to capture estimated effects for an average respondent who was highly exposed to Phase 1 of the BRT. Regression tables display the average effect for "highly exposed", referring to this definition, to ease interpretation of magnitudes.

Table 1 shows the computation of the demeaned variable. The first row of the table shows the average reduction in log travel time to the CBD, separately for low-exposure and high-exposure households. Low-exposure households have an average reduction in travel time to the CBD of 3% due to Phase 1. However, on average, low-exposure households would have an average reduction of 5% across Phases 1-4. Therefore, on average, the demeaned exposure measure for low-exposed households is an *increase* in demeaned travel time of 2%. Highly-exposed households differ: on average, they face a reduction of 22% in log travel time to the CBD from Phase 1, but would only have an average reduction of 6% across Phases 1-4. Therefore, the demeaned reduction is 16%.

3.4 Sample summary and facts about commuting, residential mobility and BRT usage

Table 2 presents summary statistics during both survey rounds, for both structures and initially-enrolled individual respondents.[10]

The first panel of Table 2 describes structure-level summary statistics from the baseline and endline surveys. At baseline, 68% of surveyed respondents own their house, while 23% are renting. Dwellings comprise 2.01 households on average, with each household occupying an average of 3.02 rooms.

The survey includes several measures of labor force participation and earnings, the key outcomes that are the focus of our analysis. Our primary measure of employment is defined from the question: "What is your main current occupation?", where people who respond that they are unemployed, too old to work, a student, or a housewife are not coded as employed. By this measure, 68% of individual respondents are employed at baseline. We also ask if the respondent engaged in any kind of job or work for payment in the last seven days; employment by this measure is 40% at baseline. The difference in reported employment rates as captured by these two measures reflects both the large amount of labor market informality present in many low-income countries, and the intermittent and sometimes unpaid nature of some types of informal work, as a result of which many respondents who regard themselves as employed may not have carried out paid tasks during the week prior to the survey. Finally, we ask if respondents operated any self-employed business or did any self-employed activity over the last month (including agriculture); 44% of respondents report doing so.

[10]The same table, weighted, is presented in Appendix Table A18. While summary statistics are broadly aligned in the weighted and unweighted samples, with weighting we match census home-ownership rates for Dar es Salaam and so there are some mechanical differences. For instance, a smaller share of households own their house in the weighted sample, with a commensurately larger share of individuals reporting that they moved house between the baseline and endline surveys.

Survey questions on income and earnings are asked in several ways. Our primary measure is earnings in the last 7 days (the answer to the question: "What were your average total weekly earnings from all jobs, after tax, in the last 7 days?"). At baseline, respondents report earning 34,600 TSH (approximately 16 USD) over the last 7 days on average. We also measure self-employed income (the response to: "What is your gross (net) income over the last 1 month (from self-employed business/activity)?"), and household income (the response to: "What is the total gross (net) income for the household over the last 1 month from all sources?").[11]

We measure poverty by household expenditure. As expected for the capital city, poverty levels are relatively low, with 95% of respondents above Tanzania's national poverty line of 49,320 TSH per person per month.

3.4.1 Commuting

Seventy-nine percent of working individuals in the sample commute to work (i.e., work outside the home) at baseline. Commute times are long, averaging 49 minutes across the sample, with especially long commutes among respondents living towards the outskirts of the city. Thirty-seven percent walk to work, while 56% commute by public transport. 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.

Figure 3, panel (a), displays baseline commuting flows at the level of 52 spatial units across Dar es Salaam. These spatial units are based on the 77 wards (which represent third level administrative divisions in Tanzania) within the geographic coverage of our study area, aggregated as necessary to ensure non-zero observations in all units, and are also used for clustering of standard errors in subsequent regression specifications. Commuting flows are strongly concentrated in trips from outlying spatial units towards locations closer to the central business district.

3.4.2 Residential mobility

Rates of household churn in the data are high, suggesting that there may be significant scope for population resorting. Eleven percent of structures surveyed had completely new households with

[11] Table 2 shows a decline in household income between 2016 and 2019. GDP data for Tanzania over this time period show a relatively flat value: based on the World Bank Development Indicator's database, GDP per capita was $953 in 2016 and $1063 in 2019. All analyses will include year fixed effects and so we will focus on how relative reported income changes as a function of exposure to BRT; the level effects between years will not be important for the analysis.

no original members living in the household. Thirty-three percent of structures had at least one of the individual (male or female) respondents move, with especially pronounced churn rates among renters.[12]

Column (2) in Appendix Table A4 shows the location of tracked individuals. Of these individuals, 76% are living in the same structure, 18% have moved elsewhere in Dar es Salaam, 4% have left Dar es Salaam, and 2% have died. Appendix Table A5 shows self-reported reasons for moving house. Thirty-one percent of individuals report family reasons (e.g., moving out of the family home, marriage, needing to look after a family member, or relationship breakup) as the primary reason they moved; 21% report economic reasons, such as being closer to work, school, or public transport; 20% report characteristics of the neighborhood, such as community, safety, or attractiveness of the area; and 14% report moving due to the cost of housing.

Panel (b) of Figure 3 displays mover flows in the sample between the baseline and endline surveys, at the level of the same 52 spatial units as used in the commuting visualization in panel (a). Many of the movers over this period can be seen to have moved from more central areas towards suburban areas further from the city center.

3.4.3 BRT usage

Table 3 reveals that a large share of individuals have ridden the BRT at the time of the endline survey. 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 0.10 (hereafter 'highly exposed'), as described in Section 3.3 – are intuitively more likely to use it.

In high exposure locations, 58% of respondents report having used the BRT in the last seven days, taking an average of 2.25 BRT trips over this period. In these locations, 37% report using the BRT to get to work, 53% report using it for family reasons, 62% for shopping, 38% for leisure and 42% for health purposes.[13] The endline survey asked respondents about their perceptions of the BRT after it had opened. Responses to these questions are summarized in Panels 2 and 3 in Table 3. Highly-exposed respondents intuitively report higher satisfaction with BRT accessibility and usefulness, while reported constraints to use include route usefulness, service reliability and

[12] To benchmark these figures, on average 6% of metropolitan respondents (and 10% of renters) in the 2016-2018 American Community Survey in the United States reported moving house within the last 12 months.

[13] This is consistent with evidence from smartphone app location data in Tanzania that highlights the importance of consumption as a driver of high-frequency mobility Blanchard et al. (2025).

station accessibility.

3.5 Parallel 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.[14] This dataset captures different households to those included in our survey data collection, and includes location identifiers only at the level of enumeration areas. Our analysis of pre-trends is therefore conducted at the level of enumeration areas for a limited number of covariates that we can measure consistently in the Measuring Living Standards Within Cities data. 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 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.

This material data allows us to test the pre-trends required to identify initial resident and place effects (CPT1 and CPT2). Identifying final resident treatment effects, and the difference between exiter and arriver impacts, requires the stronger assumption that selection into BRT affected locations is not based on pre-existing trends (CPT2). Testing this assumption would require (at a minimum) two pre-periods of panel data. Nevertheless, the assumption is common in existing studies of mobility, which assume that people do not select to be mobile based on pre-existing trends (e.g., Hamory et al. 2021).

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. We 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 Equation 3, where the dependent variable $\Delta y_{e,t-1}$ is the change in outcomes at enumeration area e measured between 2015 (LSMS data) and the 2016 baseline survey.

[14] To do this, we accessed anonymized 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.

$$\Delta y_{e,t-1} = \alpha^E + \beta^E \Delta T_e + f(x_{e,t-1}) + \epsilon_{e,t-1}$$

Estimation occurs using the post double selection lasso method to control for pre-period characteristics measured at the enumeration area. Appendix A collates the relevant results. Table A15 indicates no significant differences in pre-existing trends across treatment intensities before the BRT's implementation for any of the outcome variables considered. This is consistent with our conditional parallel trend assumption: areas that receive the first phase of the BRT were not on differential trends in terms of key labor market and housing outcomes before the new line was opened. However, we caution that the limited variables, aggregation, and smaller sample size mean that these are not immediately comparable to our baseline results and limit the power of these tests. Appendix Table A16 shows that each adjustment (choice of variables – illustrated in Panel A – and then aggregating to comparable enumeration areas – illustrated in Panel B) reduces the significance of our treatment effects. Nonetheless, Table A15 offers reassuring support for the identification assumptions underlying the identification strategy above given the available data.

4 Results

This section starts by showing that the BRT had a causal impact on place: rents, average incomes, and employment all increased in locations that received the BRT. Next, we ask who ended up living in the locations that received the BRT. We show that those who selected into affected locations were more-educated, more likely to be employed, and richer. Given this positive selection, we then ask whether the BRT had a positive impact on individuals' labor market outcomes. We find that the gains to people are uniformly smaller than the gains for places, but that there are still positive and statistically significant treatment effects for people.

4.1 Place-based impacts

To estimate place-based effects, we run the following regression, based on Equation 3:

$$\Delta y_{hh'p} = \delta^P + \beta^P \Delta T_p + g(x^{BL}_{hpf}) + \epsilon_{hh'p}$$

where $\Delta y_{hh'p}$ is the change in the outcome of the residents (household h and h') of place p, ΔT_p is the treatment (measured as the demeaned change in log travel time to the CBD) for place

p, β^P measures the casual effect of treatment on the change in outcomes, and $g(x_{hpf}^{BL})$ are baseline covariates of the household living in p at the baseline. We estimate the equation using pdslasso, allowing the estimator to select covariates from a set that includes baseline values of dependent variables and other key demographic and economic attributes. Standard errors are clustered at the aggregate spatial unit.

As discussed in Section 2, the key identification assumption to estimate Equation 3 is the conditional parallel trends assumption which requires that, conditional on baseline characteristics, changes in accessibility are as good as randomly assigned with respect to (counterfactual) outcome trends for the place.

We start by considering how the BRT affects travel times to the CBD, rent, and residential mobility at the location level. Column (1) of Table 4 shows that residents in areas served by the BRT self-report differential reductions in travel times to the CBD, with highly-exposed residents reporting a sizable reduction of nine minutes relative to a mean endline value of 70 minutes. Column (2) reveals that treated locations also experience sharp increases in rental rates: the average highly-exposed structure sees a 31% increase in expected rent per room (accounting for owners' expectations of what the room would rent for).[15] This pattern is consistent with the BRT providing a service that people are willing to pay for. Taken alone, however, this estimate does not reveal which individuals value the BRT, or what aspects of access to the BRT they are willing to pay for. Column (3) shows that there is no differential effect of the BRT on household movement, measured either as a new household living in the structure at endline (Column (3)) or an individual moving out of the household (Column (4)). However, even if there is no change in the probability of a household or individual moving, the composition of who moves in and who moves out may still differ.

The key economic focus of the World Bank and Government of Tanzania in constructing the BRT was to improve labor market outcomes. We turn to these next. The top panel of Table 5 displays estimates of BRT impacts obtained using a place-level specification estimable using standard repeated cross-sectional data. The first two columns of the table consider impacts on employment status (whether the respondent reports having an occupation, and whether they report having worked last week, respectively), while the subsequent two columns consider impacts on income measured at the individual and household level.

[15] While large, the magnitude of these estimates is consistent with sizable estimated impacts of public transport improvements on house prices in other settings such as the construction of the Jubilee Line in London (Gibbons and Machin, 2005).

Consistent with much of the existing literature (for instance, Tsivanidis 2024), these estimates suggest that the BRT has strong positive impacts on place. Column (1) of Table 5 shows that highly-exposed locations experience a 3.8 percentage point increase in the probability that a resident is employed, relative to a mean endline value of 68%. Similarly, the probability that a resident worked in the last seven days increases by 4.9 percentage points, relative to a mean endline value of 39%. The corresponding estimates in columns (3) and (4) reveal that this translates into increases in earnings: point estimates at the individual level are positive but statistically insignificant, while household income increases by 16,400 TSH on average in highly-exposed locations, a 5% increase relative to the mean endline value of 345,400 TSH.

4.2 Selection and demand for the BRT

We found that economic outcomes improved in areas with the BRT, but rents also increased. Did neighborhood composition change near the BRT, reflecting differential demand for the BRT from different types of people?[16] This section reports estimates of the selection parameters discussed above: do those who select in differ from those who select out in terms of baseline characteristics σ^{BL}; or endline characteristics σ^{EL}.

First, using the fact that we asked arrivers about their retrospective outcomes, we estimate σ^{BL} by comparing the baseline characteristics of people who were living near the BRT at endline to the characteristics of people who were living in the same house at baseline by running the following regression

$$\Delta y^{BL}_{hh'p} = \delta^{BL} + \sigma^{BL}\Delta T_p + g(x^{BL}_{hpf}) + \epsilon_{hh'p},$$

where $\Delta y^{BL}_{hh'p} = y^{BL}_{h'ip} - y^{BL}_{hpf}$.

We compare time-invariant characteristics (age and education), which are typically available in repeated cross-sectional data, as well as time-varying outcomes such as labor market participation and commuting status, which are only available if respondents are tracked over time. Results are presented in Table 6. Panel (A) of the table considers all observations, and Panel (B) restricts focus to structures that were rented at baseline. Renters are subject to rental cost increases, and do not get a benefit as owners would, which increases pressure to move. They are also a more mobile population, increasing the likelihood of selection effects.

[16]We note that in the US, demand for public transport is concentrated among the poor (Glaeser et al., 2008). However, in the context of Dar es Salaam, car ownership is much lower, with only 12% of households in the baseline data owning a car. Demand for public transport is therefore plausibly higher for high-income people.

We first consider time-invariant outcomes (age and years of education). As shown in column (1) of Table 6, the evidence here is consistent with positive selection into BRT-affected locations by baseline years of education, though the magnitude of this effect is quantitatively small: on average, those who moved into highly exposed locations (arrivers) have 0.2 more baseline years of education relative to those who move out (exiters). Conversely, there is no statistically significant evidence for selection into treated locations based on age. Our tracking data also allows us to examine evidence for selection in terms of detailed time-varying employment outcomes, as shown in the subsequent columns of Table 6. This reveals that arrivers were 1 percentage point more likely to be employed at baseline, and had higher individual and household income. Positive though statistically insignificant effects are also evident for the likelihood of working in the previous week and commuting at baseline.

Panel (B) of Table 6 demonstrates that the differences between final and initial residents in structures that were rented at baseline do indeed demonstrate stronger evidence for positive selection than in the full sample in Panel (A). Arrivers into treated locations where structures are rented are statistically significantly more educated, more likely to be working and to have higher earnings, relative to exiters from these locations. Rental arrivers were also much more likely to have been commuting (working outside the house) at baseline. The estimated magnitudes of these effects in the sample of structures rented at baseline are also much larger (in some cases even an order of magnitude larger) than those estimated in the full sample.

Appendix Table A7 reports estimates of σ^{EL} the differences between final and endline residents measured at endline, which measures the endline composition effect. The table shows more muted evidence for selection on endline characteristics, implying that those who selected in were either on a lower employment trend than those who selected out, or that the causal impact of the BRT on employment and earnings was smaller for those who selected in. We return to this point below.

These results demonstrate that the BRT is demanded by people who are *already* better off and attached to the labor market, and suggest that positive selection into treated locations may account for at least some of the positive place-based impacts of the BRT described in the previous subsection. Building on this finding, the next section uses the framework described in Section 2 to consider whether these selection effects are sufficiently strong to imply that place based impacts are not a good guide to impacts on people.

4.3 People-based estimates: final and initial residents

Given this evidence suggesting positive selection on baseline characteristics into areas treated with the BRT, we turn next to examining the causal impacts of the BRT on *people*. As outlined in Section 2, our tracking data allows us to distinguish causal from selection effects and isolate true causal effects on final residents (stayers and arrivers) and initial residents (stayers and exiters) in treated locations.

4.3.1 Effects on final residents

To estimate the final-resident effects, we run the following regression, based on Equation 2:

$$\Delta y_{hif} = \delta^F + \beta^F \Delta T_f + g(x_{hif}^{BL}) + \epsilon_{hif}$$

Where Δy_{hift} is the change in the outcome of the resident (household h) who was living in i at baseline and f at endline, ΔT_p is the treatment (measured as the demeaned change in log travel time to the CBD) for place f, β^F measures the casual effect of treatment on the change in outcomes, and $g(x_{hif}^{BL})$ are baseline covariates of household h selected by pdslasso. Standard errors are clustered at the aggregate spatial unit.

The results are shown in the second panel of Table 5. Across the labor market outcomes of interest, the point estimates for final resident effects are smaller than the place-based estimates. For example, the place-based treatment effect on employment is 3.8 percentage points and statistically significant at the 1% level. In comparison, the final-resident-based treatment effect on employment is smaller at 2.9 percentage points, and only statistically significant at the 10% level. The coefficient on working last week goes from 4.9 to 4.2 percentage points. Household income, statistically significant in the place-based analysis, is no longer statistically significant for final residents.

The pattern of results is consistent with the positive place-based estimates in the first panel of Table 5 reflecting at least in part the fact that those who had higher education and incomes, previously living elsewhere, moved into areas more strongly treated by the BRT. This is intuitive if, for instance, those already attached to the labor market are more likely to demand convenience in transport, and so move into treated locations, displacing incumbent residents. While the results are somewhat underpowered to distinguish statistically significant differences between structure-level and final resident estimates, the differences between these two sets of estimates are consistent across outcomes and sufficient to yield different statistical conclusions about whether or not the

BRT had positive impacts on labor market outcomes for those living in treated locations.

4.3.2 Effects on initial residents

As outlined in Section 2, our tracking data also allows us to isolate the causal effect of the BRT on initial residents (stayers and exiters) in treated locations. Initial residents may be of particular policy important to policymakers if these households were the target of infrastructure, and public policy concerns about gentrification leading to displacement of initial residents is widespread.

To estimate the initial-resident effects, we run the following regression, based on Equation 1:

$$\Delta y_{hif} = \delta^I + \beta^I \Delta T_i + g(x_{hif}^{BL}) + \epsilon_{hif}$$

Where Δy_{hif} is the change in the outcome of the resident (household h) who was living in i at baseline and f at endline, ΔT_i is the treatment (measured as the demeaned change in log traveltime to the CBD) for the initial location i, β^I measures the casual effect of treatment on the change in outcomes, and $g(x_{hif}^{BL})$ are baseline covariates of household h selected by pdslasso. Standard errors are clustered at the aggregate spatial unit.

The initial resident estimates are included in the third panel of Table 5. For the labor market participation outcomes in columns (1) and (2), initial resident estimates are positive, but slightly attenuated relative to place-based estimates. For the income variables in columns (3) and (4), both the point estimates and statistical significance are reduced in the initial resident relative to the place-based estimates. Across these outcomes, we find no evidence that on average initial residents were harmed from the opening of the BRT line, despite some residents moving away.

Taken together, the findings in Table 5 suggest that, while this is a very mobile population, and there is strong evidence that those who select in are better off to start with, there are sufficient inframarginal households to ensure that people based estimates remain in general positive. The pattern of results also imply that, if anything, final residents gain less than initial residents. As discussed in Section 2, this implies that those who selected in do not have larger treatment effects than those who select out. If we assume that those who select out do not see a positive impact in their labor market attachment and earnings, then this reveals that those who select in do not do so because they have large positive treatment effects. As such, demand for the BRT does not appear to be based on its ability to improve labor market outcomes.

4.4 Results for renters

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). Housing costs increases reflect the fact that some people value the benefits of the BRT, but there may be many who do not. For that population, the BRT brings higher rental rates and little benefit. For owners the implication is unclear, but for renters this is likely to cause movement out of the neighborhood in an attempt to minimize losses.

Table 7 consider results separately restricting attention to the 23% of respondents who were living in rented accommodation at the time of the baseline survey. The table shows that the estimated treatment effects are much larger for renters than for the average resident. Starting with the place-based effects, in Panel A, the employment effect for renters is an increase of 7.6 percentage points, double the effect of 3.8 percentage points estimated for the sample as a whole. A similar result holds for working last week (10.9 compared to 4.9 percentage points). The individual income effect is also statistically significant for the place-based estimate for renters, but not for the average respondent.

However, although the place-based effects are consistently larger for renters than the average sample, the role of selection is also larger. We can proxy the importance of selection by comparing the final resident treatment effect to the place-based treatment effect. Looking at the second panel in the Table 5, for employment, the final resident treatment effect is 23% smaller (2.9 versus 3.8 percentage points) for the overall sample. For renters, the final resident treatment effect is 33% smaller (5.2 versus 7.6 percentage points). The same pattern holds for the other outcome variables. This is consistent with more mobile populations being more readily able to select into treated locations in response to new infrastructure investments, resulting in a higher likelihood that positive place-based estimates may be driven by selection effects rather than positive causal impacts on individuals.

Given the role of selection in explaining the final resident treatment effects, we can then turn to asking whether initial renters were disproportionately harmed by the expansion of the BRT. Based on the employment outcomes in Table 7, initial renters also benefit by around twice as much as the average initial resident. For example, an initial renter experiences an increase in 6.9 percentage points in employment, compared to 3.7 percentage points for the average initial resident. However, the overall impact of the BRT depends both on income earned and increased expenses. An analysis of the expenditure patterns of renters in Appendix Table A11 shows that initial residents end with a larger realized rent change after the BRT than either the place-based or final-resident estimates. This pattern only exists for the renter population: for the average population, the effect of increased house prices for initial residents is smaller than both the place-based and final-resident effects. The fact that initial renters end up more exposed to rent increases than the location itself suggests costs of displacement.

4.5 Robustness tests

In this section, we examine the robustness of the central results to alternative choices in measures of exposure to the BRT, retrospective data, additional outcome variables, and sample weights.

Definition of exposure to BRT

The main results demean calculated exposure to Phase 1 of the BRT by the expected travel time reduction across the first four proposed lines of the BRT system, as described in Section 3.3. The first four lines generally run along radial roads from the CBD, with Phase 2 having a portion that connects Phase 1 and Phase 2. We consider alternative ways of constructing the exposure measure in Appendix Table A8. Columns (1)-(4) consider demeaning by Phases 1-3. Columns (5)-(8) consider phases 1, 3, and 4 (dropping the partially non-radial Phase 2). Columns (9)-(12) use all 6 phases. Results are robust and the same pattern of larger point estimates for place-based than people-based holds.

Next, Columns (13)-(16) of Appendix Table A8 consider a definition of exposure that captures the full market-access change from the BRT instead of just the change in travel time to the CBD. These two measures are closely linked since Phase 1 is primarily connecting along the radial highway to the CBD. Results are robust and the same pattern of large point estimates for place-based rather than people-based effects holds.[17]

Retrospective data

As discussed in Section 3 and Appendix C, we collect retrospective data for all households in order to validate for the initial sample by comparing actual baseline responses to retrospective endline

[17] The market access measure is calculated for each location o by summing the weighted change in travel time (between baseline and endline) to all destinations d multiplied by the destination wage times number of people at baseline. We set $\theta = 3$. We undertake a similar demeaning exercise by demeaning actual exposure to Phase 1 by expected exposure across Phases 1-4. $MA_{ot} = \sum_d \frac{1}{\hat{tt}^{\theta}_{odt}} w_{dt-1} L_{dt-1}$.

responses. Appendix Table A9 shows the robustness of results to using this retrospective data for the full sample, rather than only when required for endline residents. Results are robust.

Additional outcome variables

Next, we consider alternative outcome variables. Appendix Table A10 considers impacts on household expenditure (food, transportation, total spending, and rent). We find statistically significant impacts on food expenditure, a key proxy for welfare in settings where income is noisy. The same pattern holds: place-based estimates are larger than either the final- or initial- based estimates, suggesting an important role for selection. We find noisy increases in transportation expenditure (column (2)), and noisy increases in total expenditure (column (3)). Rent paid increases.

Appendix Table A12 looks at changes in commuting (working outside the home), public transportation, and opinions of public transport. We find noisy increases in commuting, with the individual effects again smaller than the place-based effects. We see no change in using public transportation to commute to work. The results in columns (3) and (4) show strongly statistically significant evidence for improvements in self-reported happiness with public transport and respondent perceptions that where they live is convenient for where they want to go. While these survey questions were not included among the questions asked retrospectively (such that final resident effects cannot be estimated for these outcomes), the effects remain strongly statistically significant for initial residents, suggesting that these impacts are more likely to be causal positive effects of the new infrastructure.

Appendix Table A12 also shows the change in travel time. We find that final residents have a larger reduction in travel time than the average location – consistent with people moving to the BRT from further away – and initial residents have a smaller reduction in travel time than the average location — consistent with those who move out moving further away from the BRT.

Appendix Table A13 considers impacts on other measures of income and employment that are asked independently in the survey: income from self-employment (column (1)), usual working hours at least 4 hours per day (column (2), and whether the respondent was self-employed last month (column (3)). There is no evidence for statistically significant changes in any of these outcomes at the level of either places or people.

Weighting

Our primary results are unweighted to represent the average effect of the BRT across space. In Appendix B we construct a weight accounting for the distribution of the population across neighborhoods in the city as well as home owernship rates matched to the census. Appendix Table A19 presents our key estimation results where we do apply population weighting. The central results are qualitatively robust to applying population weights, with reductions in statistical significance for income effects.

Anticipation effects

Estimation of the BRT's causal effect in our framework relies on the absence of anticipation effects—that is, our baseline data should not reflect respondents' expectations about the system's launch. The most plausible form of anticipation is that people who expect BRT-served areas to become more desirable move in, raising local rents. In this case, our baseline "initial residents" group may exclude some households already displaced and include some who should instead be classified as "final residents". This would bias our estimates of selection downward. Since our main finding is that selection effects matter, accounting for anticipation would, if anything, reinforce rather than weaken our conclusions. Nonetheless, the assumption of no anticipatory effects in this empirical setting may be plausible. As with many infrastructure projects of this type in low income countries, there was significant uncertainty around when and even whether the first phase of the BRT would be completed and 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).[18] This uncertainty is likely to have muted anticipatory behaviors.[19]

[18] 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."

[19] In theory, behavior near later stages of the BRT – which were planned but for which construction had not yet begun during our study period – could be informative about anticipation effects. Empirically this is not possible given that the locations of the different BRT lines are correlated: for instance, the second line of the BRT contains non-radial sections that connect to phase 1, generating a positive correlation, while exposure to the third line is strongly negatively correlated with exposure to phase 1. Appendix Table A14 shows results where the demeaned exposure to Phase 1 is augmented with the demeaned exposure to Phase 2 (columns (1)-(4)) or the demeaned exposure to Phase 3 (columns (5)-(8)). Consistent with these correlations, these tables show positive effects for Phase 2 and negative effects for Phase 3 negative, though these coefficients cannot be interpreted as evidence for positive or negative anticipation effects for the reasons described here.

5 Selection and measurement

Our empirical results suggest that the true casual effect of the BRT on the individuals who end up living near it is smaller than the casual effect on place. Using the place-based estimate would therefore overstate the impacts of the BRT on people. Studies in the existing literature typically take one of two approaches to addressing selection: controlling for changes in demographic composition observable in repeated cross-sectional data, or using structural models. Our unique tracking data allows us to validate these approaches.

5.1 Could selection be controlled for using repeated cross-sectional data?

Could an analyst with access only to repeated cross-sectional data have recovered the true causal effects of the BRT on those who have access to it by controlling for time-invariant demographic characteristics? To answer this, we augment the place-based regression (Equation 3) with additional controls $\widehat{X}^{BL}_{hh'p}$ that represent the difference between the final resident and the initial resident, measured at baseline (i.e., $\widehat{X}^{BL}_{hh'p} = X^{BL}_{h'ip} - X^{BL}_{hpf}$).

$$\Delta y_{hh'p} = \delta^P + \beta^P \Delta T_p + g(x^{BL}_{hpf}) + \widehat{X}^{BL}_{hh'p} + \epsilon_{hh'p}$$

We plot the resulting coefficient β^P in Figure 4. For each of the four outcomes shown, the estimate shown at the bottom of the sub-figure (labeled 'place-based') corresponds to the place-based estimate shown in the first panel of Table 5, while the estimate shown at the top (labeled 'final resident') corresponds to the final resident estimate shown in the second panel of Table 5. We concentrate on understanding whether we can recover the final resident affect as it is more likely to differ from the place effect, and should be recoverable from baseline data if observables predict outcomes.

We first add demographic controls (age and education) that can typically be inferred from repeated cross-sectional data since these are largely time-invariant outcomes. The figure shows that adding such controls barely changes the point estimate relative to the place-based estimate, and for no outcome changes our conclusion about whether there is a statistically significant effect. While those who select in to treated areas are better educated (see Table 6), education alone does not predict labor market outcomes sufficiently to serve as an effective control.

The remaining estimates in each sub-figure of Figure 4 add sequential controls for other characteristics using data that is available in our tracking data, but which would not be available in standard cross-sectional datasets. These include labor market variables such as working during baseline ('labor market'), controls for the original neighborhood of respondents ('neighborhood'), and all of these controls simultaneously ('all'). Similarly to the estimates including only demographic characteristics, there is little change in the point estimate relative to the structure-level estimate in any of these cases, and none lead to changes in statistical significance.

Overall, our results suggest that the strategy of controlling for changes in observed variables is not successful in recovering the causal effects of the BRT on final residents in our sample. This suggests that patterns of selection in this setting are complex and not fully described by the set of observable characteristics captured in our survey. For instance, individuals may sort on motivation, risk tolerance or specific skills that are correlated with valuations of commute time reductions but not adequately captured by observed demographic characteristics and labor market variables. This in turn suggests that the second common strategy in the literature of using structural models may also be insufficient to recover true causal effects, since complex selection patterns may not be consistent with the simple, one-dimensional selection on taste or ability that underpins most existing models.

5.2 Validating retrospective data

In the absence of rich administrative data, researchers need to collect their own data. Our field team tracked individuals over time (leading to estimates of initial resident effects for exiters) and asked retrospective questions of arrivers (leading to the final resident effects for arrivers). Of the two, it is much simpler to ask retrospective questions than track individuals – it relies on a single cross-sectional survey rather than two waves of data with respondent tracking. It may not even be possible to know *whom* to track without an initial baseline survey. With retrospective data in hand, final resident treatment effects can be computed immediately, giving a diagnostic test for selection.

However, a challenge with relying on retrospective data is the validity of responses. In Appendix C, we use our data to assess this by identifying individuals surveyed at both baseline and endline and comparing their baseline responses to their retrospective responses at endline for key outcomes. For binary and categorical variables (e.g., employment status or commuting behavior), we report the proportion of respondents who gave consistent answers across surveys. For continuous variables, we report the share of responses within 10% or 20% of each other, as shown in

Appendix Table A20.

We find that binary variables exhibit moderate consistency: 78% of individuals report the same employment status (correlation = 0.47), and 80% report consistent commuting behavior (correlation = 0.47). We do find high consistency for retrospective location (retrospective location data was only asked for baseline individuals who moved house) – 69% of movers report a previous location that is consistent (correlation = 0.67). However, we find that continuous variables are much less consistent. The correlation for wages in the last 7 days is 0.22, and for gross monthly household income is 0.24. These results align with other studies that find limited reliability in retrospective reporting, such as Fuller et al. (2024). Despite these measurement issues, the retrospective data may still contain useful signal, and indeed when we compute our treatment effects using retrospective data for the entire sample rather than only for those who were not enrolled at baseline, results are very similar (Appendix Table A9). Given this, we cautiously suggest researchers consider asking retrospective data questions of select key outcome variables to compute a simple diagnostic test for selection.

6 Conclusion

Urban infrastructure investments are typically place-based. When the population is mobile impacts on people may not equal the effects on place. This paper develops a framework to show how place, initial resident, and final resident treatment effects are linked through selection. We then estimate these three treatment effects 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 importance of selection, those who demand access to the BRT are substantially different from those who do not, and this demand is not driven by a desire to increase employment. Nevertheless, our results show that positive labor market impacts on place are also reflected in positive impacts for people.

Our results have important implications for both the design and evaluation of place-based policies: failing to account for population mobility either in the policy stage or the research evaluation stage can lead to misleading conclusions about who actually benefits from infrastructure investment.

References

Abeberese, Ama Baafra, Ritam Chaurey, and Radhika Menon, "Place-Based Policies and Migration: Evidence from India," *The World Bank Economic Review*, November 2024, p. lhae052.

Albouy, David and R. Jason Faberman, "Skills, Migration, and Urban Amenities over the Life Cycle," 2025.

Allen, Treb and Costas Arkolakis, "The Welfare Effects of Transportation Infrastructure Improvements," *The Review of Economic Studies*, 2022, *89* (6), 2911–2957.

Almagro, Milena and Tomás Domínguez-Iino, "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," *Econometrica (forthcoming)*, 2024.

_ , Eric Chyn, and Bryan A Stuart, "Neighborhood Revitalization and Inequality: Evidence from Chicago's Public Housing Demolitions," 2024.

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," *American Economic Review*, 2020, *110* (3), 797–823.

Autor, David, David Dorn, Gordon Hanson, Maggie R. Jones, and Bradley Setzler, "Places versus People: The Ins and Outs of Labor Market Adjustment to Globalization," *NBER Working Paper 33424*, 2025.

Balboni, Clare, Gharad Bryan, Melanie Morten, and Bilal Siddiqi, "Could Gentrification Stop the Poor from Benefiting from Urban Improvements ?," *AEA Papers and Proceedings*, 2021, *111*, 532–537.

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.

Blanchard, Paul, Doug Gollin, Martina Kirchberger, and Michael Peters, "High-Frequency Human Mobility in Three African Countries," Technical Report, CEPR Discussion Paper 2025.

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

Brummet, Quentin and Davin Reed, "The Effects of Gentrification on Incumbent Residents," 2021.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis, "Cities in the Developing World," *Annual Review of Economics*, August 2020, *12* (1), 273–297.

Bryan, Gharad T, Kyra Frye, and Melanie Morten, "Spatial Economics for Low-and Middle-Income Countries," *NBER Working Paper 33453*, 2025.

Busso, Matias, Jesse Gregory, and Patrick Kline, "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *American Economic Review*, 2013, *103* (2), 897–947.

Chapelle, Guillaume and Morgan Ubeda, "Competing for Opportunity: Transport Infrastructures and Localized Unemployment," Working Paper, CY Cergy Paris University 2025.

Collins, William J and Katharine L Shester, "Slum Clearance and Urban Renewal in the United States," *American Economic Journal: Applied Economics*, 2013, *5* (1), 239–273.

Couture, Victor, Cecile Gaubert, Jessie Handbury, and Erik Hurst, "Income Growth and the Distributional Effects of Urban Spatial Sorting," *Review of Economic Studies*, 2024.

Daily News Reporter, "State Warns against Delays in BRT Projects," *Daily News*, December 2024.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Duranton, Gilles and Matthew A Turner, "Urban Growth and Transportation," *Review of Economic Studies*, 2012, *79* (4), 1407–1440.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Franklin, Simon, Clement Imbert, Girum Abebe, and Carolina Mejia-Mantilla, "Urban Public Works in Spatial Equilibrium: Experimental Evidence from Ethiopia," *American Economic Review*, 2024, *114* (5), 1382–1414.

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.

Gaduh, Arya, Tadeja Gračner, and Alexander D. Rothenberg, "Life in the Slow Lane: Unintended Consequences of Public Transit in Jakarta," *Journal of Urban Economics*, March 2022, *128*, 103411.

Gaubert, Cecile, Patrick Kline, Damian Vergara, and Danny Yagan, "Place-Based Redistribution," 2025.

Gechter, Michael and Nick Tsivanidis, "Spatial Spillovers from High-Rise Developments: Evidence from the Mumbai Mills," 2023.

Gibbons, Stephen and Stephen Machin, "Valuing Rail Access Using Transport Innovations," *Journal of Urban Economics*, 2005, *57* (1), 148–169.

Glaeser, Edward, *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*, Penguin, 2012.

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

González-Pampillón, Nicolás, Jordi Jofre-Monseny, and Elisabet Viladecans-Marsal, "Can Urban Renewal Policies Reverse Neighborhood Ethnic Dynamics?," *Journal of Economic Geography*, 2020, *20* (2), 419–457.

Hamory, Joan, Marieke Kleemans, Nicholas Y Li, and Edward Miguel, "Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata," *Journal of the European Economic Association*, 2021, *19* (3), 1522–1555.

Harari, Mariaflavia and Maisy Wong, "Slum Upgrading and Long-Run Urban Development: Evidence from Indonesia," Technical Report, University of Pennsylvania 2021.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm, "The Making of the Modern Metropolis: Evidence from London," *The Quarterly Journal of Economics*, November 2020, *135* (4), 2059–2133.

Heckman, James J, Sergio Urzua, and Edward Vytlacil, "Understanding Instrumental Variables in Models with Essential Heterogeneity," *The Review of Economics and Statistics*, August 2006, *88* (3), 389–432.

Jacobsen, Malve, "Dar Es Salaam's BRT: Why It Has Been a Long Bumpy Ride," May 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.

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*, January 2021, *13* (3), 1058.

Lu, Yi, Jin Wang, and Lianming Zhu, "Place-Based Policies, Creation, and Agglomeration Economies: Evidence from China's Economic Zone Program," *American Economic Journal: Economic Policy*, August 2019, *11* (3), 325–60.

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

Neumark, David and Daiji Kawaguchi, "Attrition Bias in Economic Relationships Estimated with Matched CPS Files," Working Paper 8663, National Bureau of Economic Research December 2001.

The World Bank, "Untying Dar Es Salaam's 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," *American Economic Review (forthcoming)*, 2024.

Tyndall, Justin, "The Local Labour Market Effects of Light Rail Transit," *Journal of Urban Economics*, 2021, *124*, 103350.

Warnes, Pablo Ernesto, "Transport Infrastructure Improvements and Spatial Sorting: Evidence from Buenos Aires," 2024.

Weiwu, Laura, "Opportunity in Motion: Equilibrium Effects of a Place-based Policy on Economic Mobility," 2024.

World Bank, "With Bus Rapid Transit, African Cities Are Riding toward a Better Future," 2022.

Zárate, Román D., "Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City," *American Economic Review (R&R)*, 2023.

Tables and figures

Table 1: Reduction in travel time

	Low exposure	High exposure
Reduction in log travel time to CBD with Phase 1	0.03	0.22
Average reduction in log travel time to CBD with Phases 1-4	0.05	0.06
Demeaned reduction in log travel to CBD	-0.02	0.16
N	1665	425

Notes: An observation is a structure measured at endline. Table shows the predicted change in log travel to the CBD with Phase 1 compared to without any BRT and with Phases 1-4 compared to without any BRT. High exposure is households above the 80th percentile of demeaned travel time reduction, as explained in the text. Demeaned reduction is the reduction in log travel time due to Phase 1 minus the reduction in log travel time with Phases 1-4. A negative value for the demeaned reduction in log travel time implies that the realized reduction from Phase 1 was less than the expected average reduction across Phases 1-4. Statistics unweighted.

Table 2: Summary stats

	(1) Baseline	(2) Endline
Structure/household-level		
Electricity in house for lighting	0.62	0.73
Street has lights	0.028	0.034
Road is paved	0.098	0.18
HH uses non-latrine toilet	0.52	0.73
Number of households in dwelling	2.01	2.05
Number of members in household	4.50	4.16
Rooms household occupies in dwelling	3.02	3.34
Number of rooms per household member	0.79	1.02
Monthly rent expected per room, Tsh	38164.6	29389.6
Own house	0.68	0.65
Rented	0.23	0.26
Share of consumption on (imputed) rent	0.19	0.18
Share of consumption on food	0.43	0.50
Share of consumption on transportation	0.11	0.11
Above Tanzania national poverty line	0.95	0.89
All initial household members moved out	.	0.11
At least one initial male/female respondent moved out	.	0.33
N	1517	1800
Individuals		
Age	37.8	40.1
Years education	8.25	8.45
Worked for pay last 7 days	0.40	0.40
Employed (as per occupation variable)	0.68	0.66
Operate any self-employed business or activity last month	0.44	0.39
Typical days worked per week	5.85	5.74
Typical hours worked per day	8.38	9.09
Wages last 7 days (1000 TSH)	34.6	29.0
Gross household income (1000 TSH)	544.7	418.2
Net household income (1000 TSH)	387.6	339.5
Gross income from self-employment (1000 TSH)	134.4	107.0
Net income from self-employment (1000 TSH)	89.8	72.9
Commutes (if employed)	0.79	0.70
Commute by walking (if employed)	0.37	0.36
Commute by public transport (if employed)	0.56	0.50
Commute time (mins)	49.2	88.0
Happiness public transport (scale 1-10)	2.85	5.85
Moved house between baseline and endline	.	0.29
Moved house between baseline and endline (renting BL)	.	0.50
N	3104	3824

Notes: The first panel of the table shows summary stats for the structures. The second panel shows summary stats for the male/female respondents (both initial and final residents). Employed is defined from the question: what is your main current occupation, and indicating if someone has an usual occupation other than unemployed, too old to work, student, or housewife. Worked last 7 days is defined from the question: During the last 7 days, were you engaged in any kind of job or work for payment? Self-employment is defined from the question: Did you operate any self-employed business or do any self-employed activity over the last month (including agriculture)? Wages last 7 days is the response to: what were your average total weekly earnings from all jobs, after tax, in the last 7 days? Self-employed income is the response to: what is your gross (net) income over the last 1 month (from self-employed business/activity). Household income is the response to: what is the total gross (net) income for the household over the last 1 month from all sources? All initial household members moved out means that none of the original household (including male/female respondents) remain living in the house. At least one initial male/female respondent moved out means at least one of the male/female respondents moved out, but other household members remained. Commute is a measure that the individual works outside of the home. Modes of public transport are daladala, rickshaw, motorbike, train in baseline, with BRT an additional option in endline. Statistics unweighted.

Table 3: BRT usage and preferences

	Low exposure	High exposure
BRT usage		
Have you ever ridden on the BRT?	0.67	0.94
BRT used in commute (if employed)	0.04	0.20
Number BRT trips last 7 days	0.67	2.25
Rode BRT last 7 days	0.22	0.58
If so, for what reasons?		
... Work	0.36	0.37
... Shopping	0.40	0.62
... Educ	0.06	0.10
... Health	0.28	0.42
... Leisure	0.24	0.38
... Family	0.53	0.53
... Other	0.03	0.01
Out of 10, how much do you agree?		
BRT station is easily accessible	4.12	6.26
BRT driver/vehicle safe	8.73	7.89
BRT is cheap	6.57	5.90
BRT goes to places useful to me	5.62	7.72
BRT makes easier daily jobs done	7.16	7.46
BRT improves public transport options	7.55	7.60
Main constraint from using BRT		
Route not useful	0.36	0.21
Congestion or service not reliable	0.41	0.46
Limited accessibility station	0.18	0.13
Vehicle or driver safety, personal security, harrassment	0.01	0.03
Other	0.05	0.17
N	2179	501

Notes: Table shows BRT usage and preference stats for individual respondents, measured at endline. High exposure is households above the 80th percentile of demeaned travel time reduction, as explained in the text. Number of trips in last 7 days winsorized at the 99th percentile. Reasons for riding BRT in last 7 days are conditional on making at least one trip in the last days. Respondents could respond with multiple trip reasons. Responses measured on a scale of 1-10, 10 is favorable. Respondents could only answer one response to the main constraint from using BRT. Statistics unweighted.

Table 4: Structure-level results

	(1) Self-reported time to CBD (mins)	(2) Log rent per room (expected)	(3) Different household in structure at EL	(4) Different individual in structure at EL
Reduction in log traveltime (demeaned)	-84.243	2.977	-0.074	0.001
	(27.987)***	(0.419)***	(0.094)	(0.174)
N	1180	1125	1215	1215
Mean EL value	70.292	9.882	0.072	0.202
Effect highly-exposed	-8.761	0.310	-0.008	0.000

Notes: Table shows coeffcient with standard errors in parentheses. An observation is a structure that was enrolled at baseline for which we can calculate the structure-level, initial-resident, and final-resident treatment effect for at least one of the initially-enrolled individuals. Demeaned 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. Self-reported time to CBD is the time respondents report it would take to travel to the downtown market. Log rent per room (expected) is the cost to rent the household's section of the house divided by the number of rooms the household occupies. Owners are asked about rent if they were to rent out their section of the house. Different household in structure at EL measures whether the baseline household moved between baseline and endline survey. Different individual in structure at EL measures whether any individual tracked member of the baseline household moved between baseline and endline survey. Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Table 5: Effects of the BRT on place, initial, and final residents

	(1) Employed	(2) Worked last week	(3) Income	(4) Household income
Place-based				
Reduction in log traveltime (demeaned)	0.367	0.473	35.616	157.855
	(0.140)***	(0.247)*	(24.070)	(87.930)*
N	1888	1888	1834	1483
Mean EL value	0.678	0.391	27.445	345.398
Effect highly-exposed	0.038	0.049	3.704	16.417
Initial residents				
Reduction in log traveltime (demeaned)	0.355	0.455	35.156	120.112
	(0.155)**	(0.241)*	(22.789)	(86.223)
N	1888	1888	1832	1502
Mean EL value	0.678	0.405	28.062	345.680
Effect highly-exposed	0.037	0.047	3.656	12.492
Final residents				
Reduction in log traveltime (demeaned)	0.279	0.407	32.909	101.806
	(0.153)*	(0.247)*	(24.015)	(83.273)
N	1888	1885	1826	1470
Mean EL value	0.678	0.391	27.456	345.980
Effect highly-exposed	0.029	0.042	3.422	10.588

Notes: Table shows coeffcient with standard errors in parentheses. An observation is a structure by gender pair. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Demeaned 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. Highly-exposed is the effect for households above the 80th percentile of demeaned travel time reduction, as explained in the text. Employment is measured by whether the household member reports an usual occupation other than unemployed, student, housewife, or too old to work. Worked last 7 days is whether the household member reports working in the last 7 days. Earnings are the earnings (in 1000 TSH) during the last seven days. Household income is the total net income for all household members over the last month (in 1000 TSH). Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Table 6: Demand for public transport: differences between final and initial residents

	(1) Education	(2) Age	(3) Employed	(4) Worked last week	(5) Income	(6) Household income	(7) Commutes
All							
Reduction in log traveltime (demeaned)	1.698	-2.600	0.086	0.058	35.850	125.159	0.074
	(0.511)***	(1.852)	(0.052)*	(0.058)	(16.447)**	(53.703)**	(0.061)
N	1870	1883	1888	1885	1876	1776	1888
Mean EL value	8.187	42.370	0.678	0.391	27.456	345.818	0.321
Effect highly-exposed	0.177	-0.270	0.009	0.006	3.728	13.016	0.008
Structures rented at BL							
Reduction in log traveltime (demeaned)	3.930	-1.407	0.426	0.315	72.124	-22.805	0.523
	(1.979)**	(3.970)	(0.167)**	(0.206)	(26.384)***	(105.784)	(0.193)***
N	296	300	302	301	297	277	302
Mean EL value	8.355	36.803	0.666	0.405	31.145	301.206	0.371
Effect highly-exposed	0.409	-0.146	0.044	0.033	7.501	-2.372	0.054

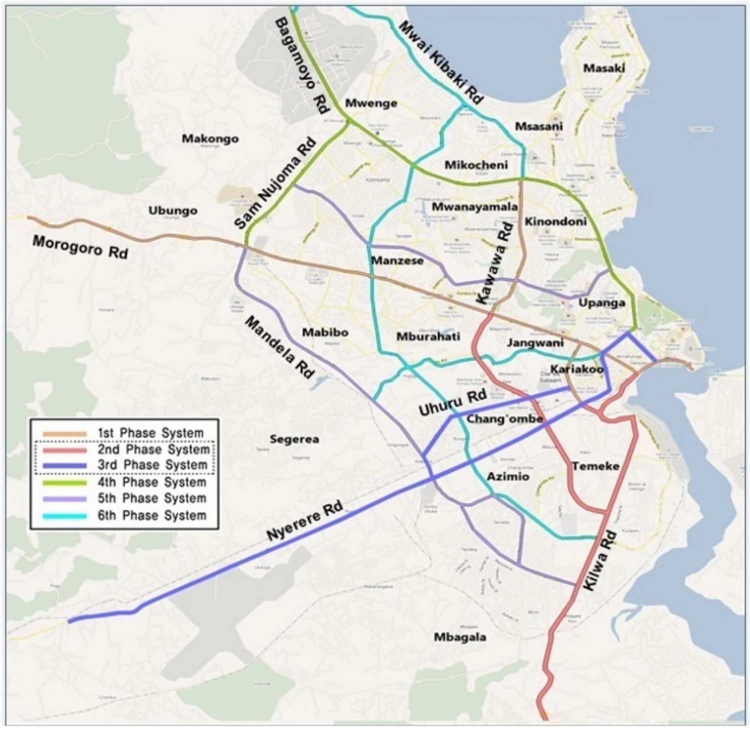
Notes: Table shows coefficient with standard errors in parentheses. An observation is a structure by gender pair. All individuals (movers and non-movers) are included; the different in outcomes for non-movers is zero. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Panel (a) shows all observations. Panel (b) restricts to structures that were rented at baseline. Dependent variable is the difference between the characteristics of the final resident and the initial resident, measured at baseline (through baseline data for a respondent who was enrolled at baseline and retrospective data for a respondent who was not enrolled at baseline) unless noted below. Demeaned 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. Education is measured as years of education, captured at endline for both initial and final resident. Age is age in years, measured at endline. Employment is measured by whether the household member reports an usual occupation other than unemployed, student, housewife, or too old to work. Worked last 7 days is whether the household member reports working in the last 7 days. Earnings are the earnings (in 1000 TSH) during the last seven days. Household income is the total income for all household members over the last month. Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Table 7: Effects of the BRT on place, initial, and final residents (structures rented at BL)

	(1) Employed	(2) Worked last week	(3) Income	(4) Household income
Place-based				
Reduction in log traveltime (demeaned)	0.735	1.047	77.198	111.621
	(0.238)***	(0.353)***	(44.971)*	(184.781)
N	302	302	295	256
Mean EL value	0.666	0.407	31.058	299.855
Effect highly-exposed	0.076	0.109	8.029	11.609
Initial residents				
Reduction in log traveltime (demeaned)	0.662	0.868	45.374	53.849
	(0.251)***	(0.349)**	(31.319)	(174.118)
N	302	302	294	260
Mean EL value	0.689	0.447	32.591	316.446
Effect highly-exposed	0.069	0.090	4.719	5.600
Final residents				
Reduction in log traveltime (demeaned)	0.496	0.952	61.729	148.766
	(0.226)**	(0.341)***	(47.412)	(185.963)
N	302	301	290	248
Mean EL value	0.666	0.405	31.145	301.201
Effect highly-exposed	0.052	0.099	6.420	15.472

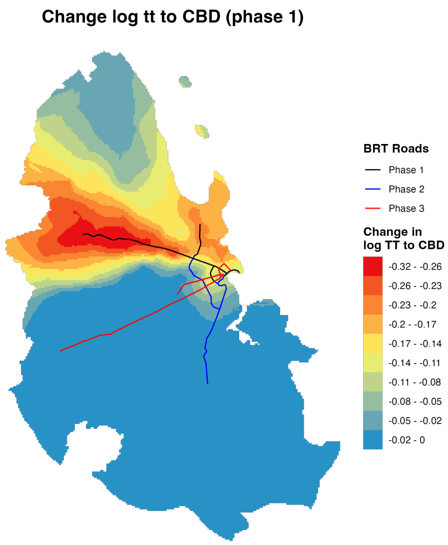
Notes: Table shows coeffcient with standard errors in parentheses. An observation is a structure by gender pair. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Demeaned variable is equal to predicted decrease in travel time to downtown along BRT Phase 1 minus average predicted decrease in travel time to downtown along planned BRT Phases 1-4. Highly-exposed is the effect for households above the 80th percentile of demeaned travel time reduction, as explained in the text. Employment is measured by whether the household member reports an usual occupation other than unemployed, student, housewife, or too old to work. Worked last 7 days is whether the household member reports working in the last 7 days. Earnings are the earnings (in 1000 TSH) during the last seven days. Household income is the total net income for all household members over the last month (in 1000 TSH). Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Figure 1: BRT network in Dar es Salaam

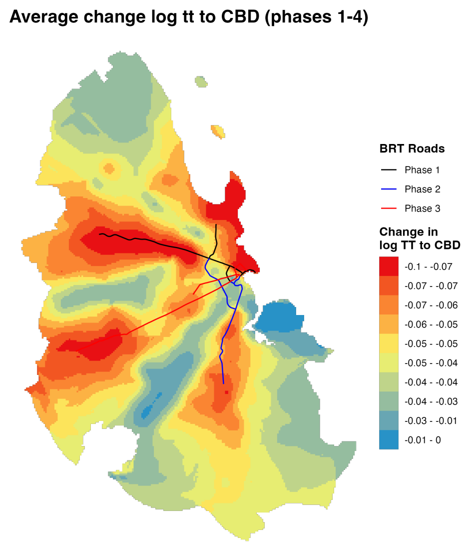


Notes: The figure shows the planned 6-phase BRT network in Dar es Salaam. We examine the impact of Phase 1 that runs northwest along Morogoro Road. Image source: Dar Rapid Transit (DART).

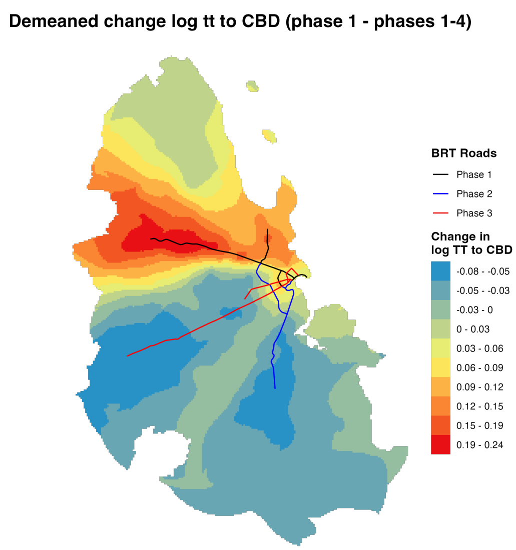
Figure 2: Reduction in log travel time to CBD due to BRT



(a) Change (Phase 1)



(b) Change (average across Phases 1-4)

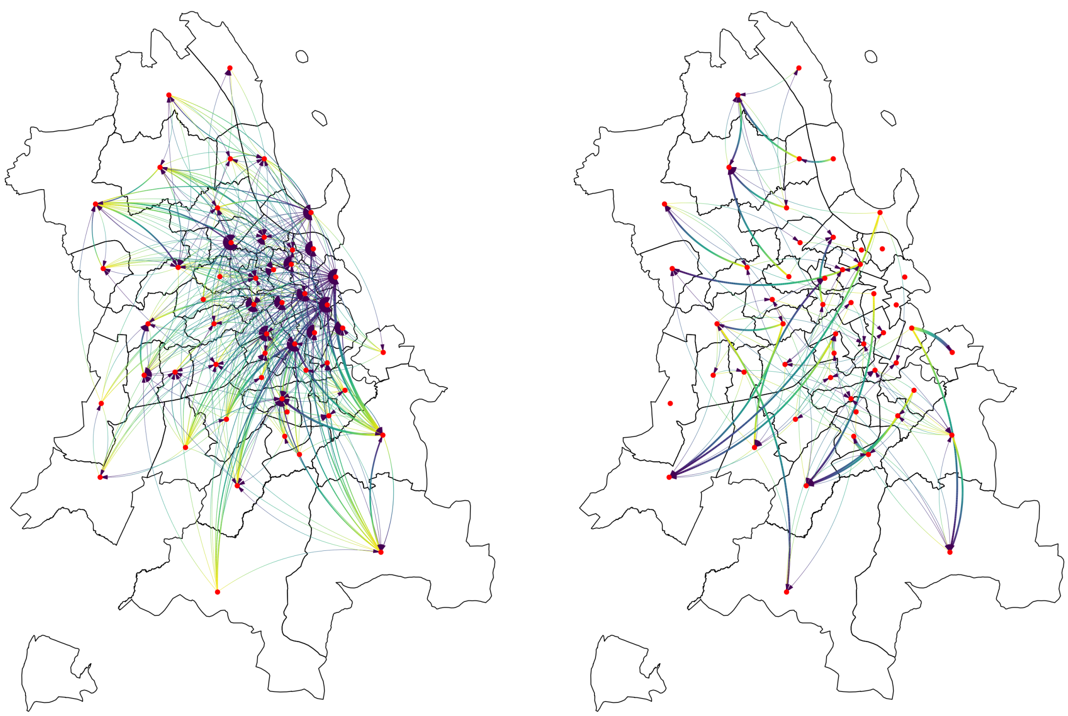


(c) Demeaned travel time

Notes: The figure shows the calculated reduction in log travel time to the CBD from phases of the BRT. Panel (a) shows the value comparing Phase 1 to no BRT. Panel (b) shows the average reduction in travel time to CBD from Phases 1-4. Panel (c) shows the value after demeaning log travel time by the average decrease for Phases 1 to 4.

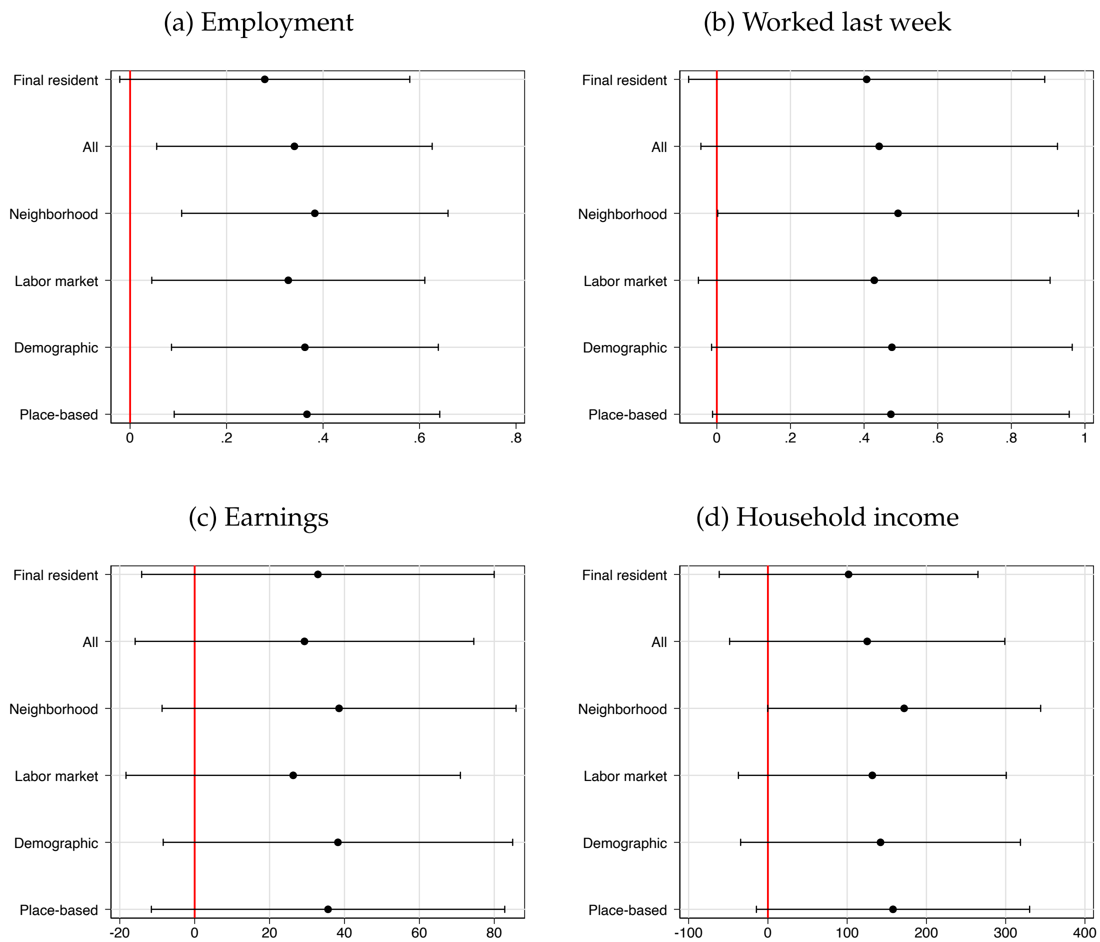
Figure 3: Commuting and migration flows

(a) Baseline commuting flows (b) Migration flows from baseline to endline



Notes: The figure shows directional baseline commuting flows (left hand panel) and migration flows of baseline respondents between baseline and endline (right hand panel). Direction of travel is shown from yellow towards purple. Arrow thickness indicates magnitude of commuting or migrant flows.

Figure 4: Decomposing the selection effect



Notes: The figure plots the point estimate and 95th percentile confidence interval of the demeaned coefficient from the place-based regression. An observation is a structure x gender pair. Place-based effect is the place-based coefficient. Selection is accounted for by allowing Post Double Selection Lasso to select the difference in baseline characteristics of the baseline and endline respondents within the structure. Demographic adds the change in demographic outcomes (age and years of education) between the final and initial resident as a potential control. Labor market is the change between being employed and weekly earnings (both measured at baseline) between final and initial resident. Neighborhood is the change in location fixed effects at baseline. All includes demographic, labor market, and neighborhood controls. The final resident treatment effect is included for comparison. Standard errors clustered at aggregate spatial unit. Regressions unweighted.

APPENDIX: FOR ONLINE PUBLICATION ONLY

Figure A1: Timeline of research and operational activities

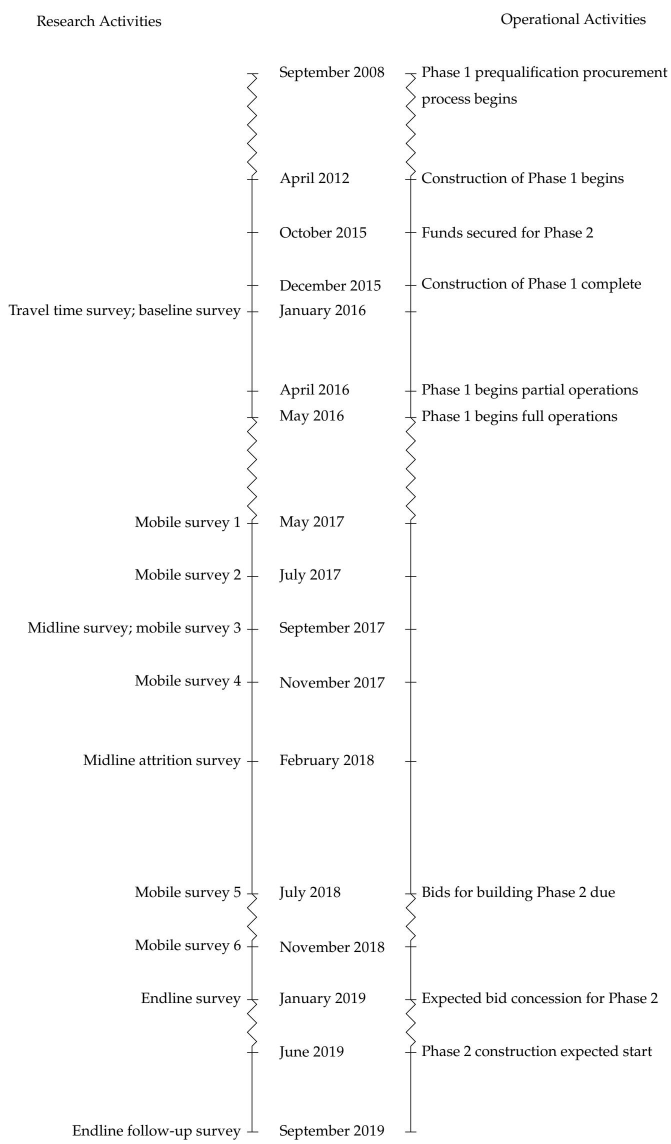
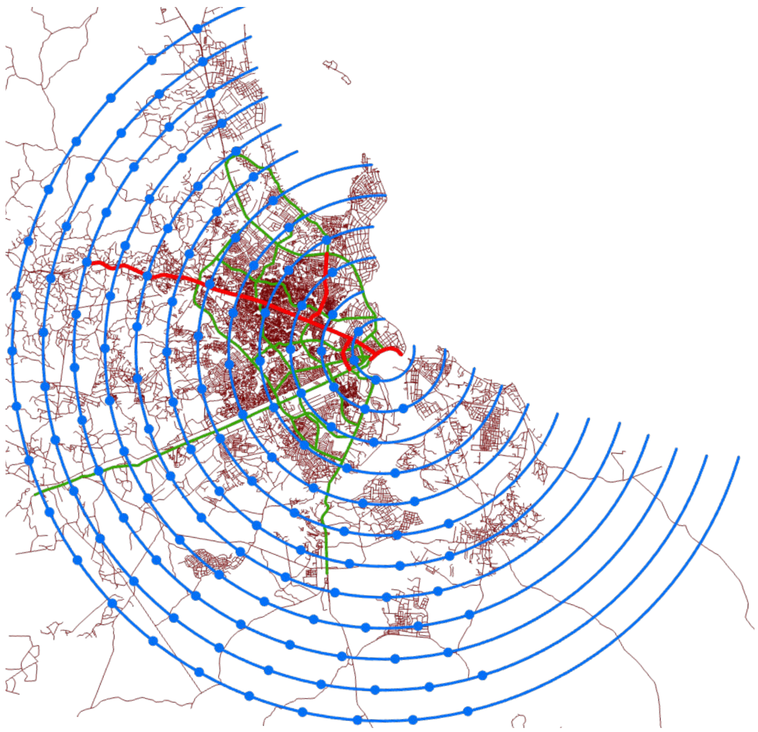
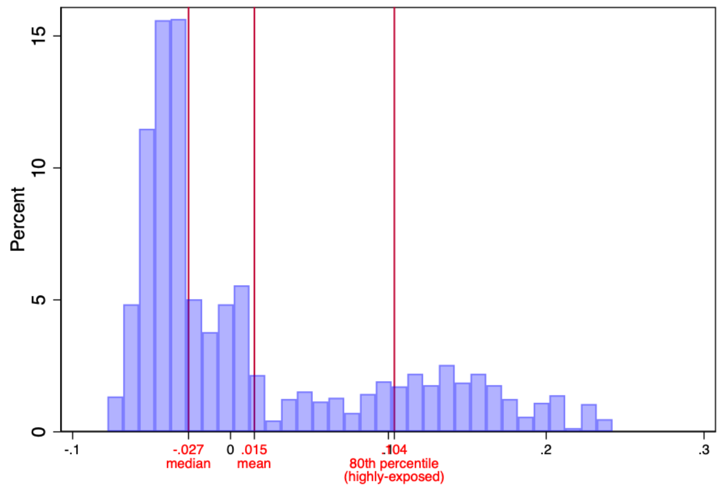


Figure A2: Spatial sampling frame



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 blue arcs. We enrolled households in 125 of the 141 clusters shown as blue dots along these arcs; 12-14 households were enrolled in each cluster, yielding a sample of 1748 households. The full BRT network is shown in green, with the Phase 1 route highlighted in red.

Figure A3: Distribution of demeaned log travel time



Notes: This figure shows the distribution of demeaned log travel time to the CBD. A unit of observation is a baseline structure. The 80th percentile of the demeaned distribution is shown on the figure. The 80th percentile value is used in the tables and text to indicate an "highly exposed" household.

Table A1: Contact rates of sample at endline

	(1) All	(2) Located ML
Structures		
Found and survey complete	76.0	87.5
Found and refused /incomplete	10.0	3.6
Not found	8.2	2.1
Torn down / empty	5.8	6.7
N	1748	1517
Households		
Found and survey complete	79.6	92.2
Found and refused/incomplete	9.2	2.3
Not found	11.2	5.6
N	1748	1510
Male/female respondents		
Found and survey complete	69.9	81.3
Found and refused /incomplete	13.1	6.0
Not found	15.1	10.6
Died	1.8	2.1
N	3104	2668

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. Column (2) shows contact rates conditional on locating at least one member of the household in the midline attrition survey. We stopped tracking structures after the midline attrition survey 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. Statistics are unweighted.

Table A2: Correlates of attrition at the individual and structure level

	(1) Structure	(2) Individual
Log BL HH monthly gross income per capita	-0.032***	-0.030**
	(0.010)	(0.012)
Log BL monthly rent per room	-0.035***	-0.020
	(0.013)	(0.013)
BL number of years lived in the structure	0.002	-0.001
	(0.001)	(0.001)
BL number of HH in the structure	0.006	0.009
	(0.006)	(0.006)
Dwelling rented by HH	-0.145**	-0.136***
	(0.055)	(0.036)
Dwelling owned by HH	0.100**	0.009
	(0.046)	(0.030)
BL distance to phase 1	0.004	0.002
	(0.003)	(0.003)
BL distance to phase 2	0.005	-0.001
	(0.004)	(0.003)
BL Age		-0.000
		(0.001)
Male dummy		-0.057***
		(0.014)
N	1748	2646
Mean dependent variable	0.760	0.699

Notes: Dependent variable is completing the endline survey. A unit of analysis is either a structure that was enrolled at baseline (Column (1)) or an individual that was enrolled at baseline (Column (2)). Standard errors, clustered by aggregate spatial units, are reported in parentheses. Missing values for the independent variables are dummied out.

Table A3: Tabulation of which treatment effects can be computed for baseline structures

	(1) Male	(2) Female	(3) All
Structure, initial, final	859	1029	1888
Structure, final only	72	104	176
Initial only	139	143	282
Final only	47	27	74
None (BL respondent exists)	409	345	754
None (no BL respondent)	222	100	322
N	1748	1748	3496

Notes: An observation is a structure x gender. Sample is restricted to structures present at BL. The table shows the categorization of each baseline structure by the types of treatment effects that can be calculated. Structure, initial, final means that the initial resident was surveyed originally and was successfully tracked and resurveyed (initial), that the new respondent in the household was also surveyed (structure and final). No BL respondent exists refers to whether or not there was an eligible adult to be a baseline respondent. This could be if the baseline respondent did not complete the baseline (too busy or refused) or cases where there was no eligible person on the roster (e.g., if the household was a single male, there could not be a female respondent). If no treatment effects can be calculated this is due to the individual or the structure not being tracked or the household refusing to participate in the endline.

Table A4: Location of tracked individuals between baseline and endline

	(1) Overall share	(2) Share if found
Same structure	57	76
Moved within Dar	13	18
Left Dar	3	4
Died	2	2
Unknown	25	
Total	100	100
N	3104	2331

Notes: Table shows the location of individuals enrolled at baseline at endline. Column (1) includes the whole sample. Column (2) drops indivduals where location is unknown. Some location information is derived from household member responses and tracking even if respondent did not complete the survey. Statistics unweighted.

Table A5: Why did individuals move

	Share
Family reasons	31
Proximity to public transport, work, education, or social services	21
Community/safety/attractiveness of area	20
Cheaper rent	14
Other	13
Total	100
N	504

Notes: Table shows answers to the question: What is the main reason you moved house? Statistics unweighted.

Table A6: Time use

	Low exposure	High exposure
Work	4.24	4.57
Travel to/from work	0.65	0.66
Travel to/from other activities	1.79	1.77
Education	0.17	0.31
Sleep	7.89	7.52
Rest of day	9.33	9.38
N	2021	461

Notes: Table shows time use at endline, measured in hours, across activities for individual respondents. High exposure is households above the 80th percentile of demeaned travel time reduction, as explained in the text. Rest of day is computed from 24 hours minus the sum of other activities.

Table A7: Demand for public transport: differences between final and initial residents (differences measured at EL)

	(1) Education	(2) Age	(3) Employed	(4) Worked last week	(5) Income	(6) Household income	(7) Commutes
All							
Reduction in log traveltime (demeaned)	1.698	-2.600	0.024	0.007	-3.949	43.930	0.101
	(0.511)***	(1.852)	(0.058)	(0.097)	(12.891)	(56.743)	(0.073)
N	1870	1883	1888	1888	1832	1480	1888
Mean EL value	8.187	42.370	0.678	0.391	27.292	346.304	0.321
Effect highly-exposed	0.177	-0.270	0.002	0.001	-0.411	4.569	0.010
Structures rented at BL							
Reduction in log traveltime (demeaned)	3.930	-1.407	0.062	0.112	32.417	41.347	0.339
	(1.979)**	(3.970)	(0.306)	(0.364)	(34.644)	(141.960)	(0.306)
N	296	300	302	302	292	248	302
Mean EL value	8.355	36.803	0.666	0.407	30.949	301.121	0.371
Effect highly-exposed	0.409	-0.146	0.006	0.012	3.371	4.300	0.035

Notes: Table shows coefficient with standard errors in parentheses. An observation is a structure by gender pair. All individuals (movers and non-movers) are included; the different in outcomes for non-movers is zero. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Panel (a) shows all observations. Panel (b) restricts to structures that were rented at baseline. Dependent variable is the difference between the characteristics of the final resident and the initial resident, measured at endline. Demeaned 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. Education is measured as years of education, captured at endline for both initial and final resident. Age is age in years, measured at endline. Employment is measured by whether the household member reports an usual occupation other than unemployed, student, housewife, or too old to work. Worked last 7 days is whether the household member reports working in the last 7 days. Earnings are the earnings (in 1000 TSH) during the last seven days. Household income is the total income for all household members over the last month. Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Table A8: Effects of the BRT on place, initial, and final residents (alternative measures of exposure to BRT)

	Phases 1-3				Phases 1,3,4				Phases 1-6				Market access			
	(1) Employed	(2) Worked last week	(3) Income	(4) Household income	(5) Employed	(6) Worked last week	(7) Income	(8) Household income	(9) Employed	(10) Worked last week	(11) Income	(12) Household income	(13) Employed	(14) Worked last week	(15) Income	(16) Household income
Place-based																
Reduction in log traveltime (demeaned)	0.353	0.485	38.801	170.785	0.445	0.850	63.374	229.911	0.352	0.454	34.112	123.925	0.345	0.875	57.679	469.816
	(0.140)**	(0.240)**	(23.572)*	(85.560)**	(0.161)***	(0.293)***	(26.616)**	(104.586)**	(0.140)**	(0.248)*	(23.781)	(83.952)	(0.229)	(0.263)***	(26.513)**	(149.939)***
N	1888	1888	1834	1483	1888	1888	1834	1483	1888	1888	1834	1483	1888	1888	1834	1483
Mean EL value	0.678	0.391	27.445	345.398	0.678	0.391	27.445	345.398	0.678	0.391	27.445	345.398	0.678	0.391	27.445	345.398
Effect highly-exposed	0.037	0.050	4.035	17.762	0.046	0.088	6.591	23.911	0.037	0.047	3.548	12.888	0.036	0.091	5.999	48.861
Initial residents																
Reduction in log traveltime (demeaned)	0.342	0.474	34.312	131.407	0.389	0.793	58.614	176.605	0.336	0.424	31.001	91.359	0.144	0.785	48.817	318.039
	(0.146)**	(0.237)**	(23.087)	(81.981)	(0.171)**	(0.277)***	(25.223)**	(105.934)*	(0.156)**	(0.243)*	(22.946)	(81.526)	(0.229)	(0.264)***	(27.856)*	(180.181)*
N	1888	1888	1832	1502	1888	1888	1832	1502	1888	1888	1832	1502	1888	1888	1832	1502
Mean EL value	0.678	0.405	28.062	345.680	0.678	0.405	28.062	345.680	0.678	0.405	28.062	345.680	0.678	0.405	28.062	345.680
Effect highly-exposed	0.036	0.049	3.568	13.666	0.040	0.082	6.096	18.367	0.035	0.044	3.224	9.501	0.015	0.082	5.077	33.076
Final residents																
Reduction in log traveltime (demeaned)	0.294	0.428	31.391	117.091	0.329	0.793	59.484	160.154	0.264	0.392	30.741	66.212	0.121	0.754	55.179	286.495
	(0.148)**	(0.242)*	(23.635)	(81.087)	(0.173)*	(0.288)***	(28.059)**	(108.683)	(0.154)*	(0.248)	(23.888)	(78.686)	(0.233)	(0.281)***	(32.169)*	(177.900)
N	1888	1885	1826	1470	1888	1885	1826	1470	1888	1885	1826	1470	1888	1885	1826	1470
Mean EL value	0.678	0.391	27.456	345.980	0.678	0.391	27.456	345.980	0.678	0.391	27.456	345.980	0.678	0.391	27.456	345.980
Effect highly-exposed	0.031	0.045	3.265	12.177	0.034	0.082	6.186	16.656	0.027	0.041	3.197	6.886	0.013	0.078	5.739	29.796

Notes: Table shows robustness to definition of exposure to BRT. Columns (1)-(4) demean by using phases 1-3. Columns (5)-(8) demean by using phases 1,3,4. Columns (9)-(12) demean by using phases 1-6. Columns (13)-(16) define exposure by demeaned market access instead of demeaned traveltime to CDB. Table shows coeffcient with standard errors in parentheses. An observation is a structure by gender pair. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Demeaned 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. Highly-exposed is the effect for households above the 80th percentile of demeaned travel time reduction, as explained in the text. Employment is measured by whether the household member reports an usual occupation other than unemployed, student, housewife, or too old to work. Worked last 7 days is whether the household member reports working in the last 7 days. Earnings are the earnings (in 1000 TSH) during the last seven days. Household income is the total net income for all household members over the last month (in 1000 TSH). Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Table A9: Effects of the BRT on place, initial, and final residents (using only retro responses)

	(1) Employed	(2) Worked last week	(3) Income	(4) Household income
Place-based				
Reduction in log traveltime (demeaned)	0.375	0.517	28.934	163.227
	(0.145)***	(0.248)**	(24.596)	(87.482)*
N	1888	1888	1834	1483
Mean EL value	0.678	0.391	27.445	345.398
Effect highly-exposed	0.039	0.054	3.009	16.976
Initial residents				
Reduction in log traveltime (demeaned)	0.302	0.464	30.847	116.933
	(0.163)*	(0.233)**	(22.838)	(87.115)
N	1888	1888	1832	1502
Mean EL value	0.678	0.405	28.062	345.680
Effect highly-exposed	0.031	0.048	3.208	12.161
Final residents				
Reduction in log traveltime (demeaned)	0.302	0.427	25.517	138.996
	(0.153)**	(0.243)*	(23.761)	(82.164)*
N	1888	1885	1826	1470
Mean EL value	0.678	0.391	27.456	345.980
Effect highly-exposed	0.031	0.044	2.654	14.456

Notes: Table shows coeffcient with standard errors in parentheses. An observation is a structure by gender pair. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Demeaned 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. Highly-exposed is the effect for households above the 80th percentile of demeaned travel time reduction, as explained in the text. Employment is measured by whether the household member reports an usual occupation other than unemployed, student, housewife, or too old to work. Worked last 7 days is whether the household member reports working in the last 7 days. Earnings are the earnings (in 1000 TSH) during the last seven days. Household income is the total net income for all household members over the last month (in 1000 TSH). Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Table A10: Treatment effects: expenditure

	(1) Food exp.	(2) Transport exp.	(3) Total exp.	(4) Rent
Place-based				
Reduction in log traveltime (demeaned)	104.538	46.732	171.534	3.163
	41.723**	30.541	104.757	0.431***
N	1789	1589	1888	1751
Mean EL value	243.975	59.466	398.091	9.860
Effect highly-exposed	10.872	4.860	17.840	0.329
Initial residents				
Reduction in log traveltime (demeaned)	79.595	33.807	145.832	3.025
	44.020*	25.122	82.464*	0.470***
N	1793	1600	1888	1754
Mean EL value	241.057	59.038	395.909	9.864
Effect highly-exposed	8.278	3.516	15.167	0.315
Final residents				
Reduction in log traveltime (demeaned)	120.030	45.441	197.170	3.218
	43.387***	28.437	104.513*	0.520***
N	1774	1586	1888	1554
Mean EL value	244.268	59.250	398.091	9.833
Effect highly-exposed	12.483	4.726	20.506	0.335

Notes: Table shows coeffcient with standard errors in parentheses. An observation is a structure by gender pair. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Demeaned 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. Food spending is spending on food. Transport spending is spending on transport. Total is total spending. Expenditure data is measured at the household level and duplicated for each gender respondent. Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Table A11: Treatment effects: expenditure (structures rented at BL)

	(1) Food exp.	(2) Transport exp.	(3) Total exp.	(4) Rent
Place-based				
Reduction in log traveltime (demeaned)	165.239	62.720	239.347	2.092
	94.906*	33.621*	106.850**	0.672***
N	292	271	302	290
Mean EL value	218.521	49.702	381.532	10.026
Effect highly-exposed	17.185	6.523	24.892	0.218
Initial residents				
Reduction in log traveltime (demeaned)	97.989	66.569	143.018	2.111
	81.318	48.004	90.384	0.641***
N	293	274	302	287
Mean EL value	221.184	49.053	383.233	10.011
Effect highly-exposed	10.191	6.923	14.874	0.220
Final residents				
Reduction in log traveltime (demeaned)	136.070	49.774	219.203	1.452
	91.409	32.904	114.552*	0.709**
N	286	270	302	201
Mean EL value	219.748	49.331	381.532	9.954
Effect highly-exposed	14.151	5.176	22.797	0.151

Notes: Table shows coeffcient with standard errors in parentheses. An observation is a structure by gender pair. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Demeaned 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. Food spending is spending on food. Transport spending is spending on transport. Total is total spending. Expenditure data is measured at the household level and duplicated for each gender respondent. Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Table A12: Treatment effects: commmuting behavior and preferences

	(1) Commutes	(2) Public transport to work	(3) Time Kariakoo	(4) Happy public transport	(5) Live convenient
Place-based					
Reduction in log traveltime (demeaned)	0.151	0.016	-69.062	5.802	6.700
	0.213	0.142	28.232**	1.749***	1.778***
N	1888	1888	1793	1882	1799
Mean EL value	0.321	0.345	71.298	5.764	6.118
Effect highly-exposed	0.016	0.002	-7.182	0.603	0.697
Initial residents					
Reduction in log traveltime (demeaned)	0.078	-0.107	-67.057	5.197	5.662
	0.199	0.137	30.559**	1.600***	1.764***
N	1888	1888	1771	1882	1796
Mean EL value	0.329	0.346	71.385	5.764	6.119
Effect highly-exposed	0.008	-0.011	-6.974	0.540	0.589
Final residents					
Reduction in log traveltime (demeaned)	0.097	-0.099	-80.386		
	0.207	0.130	30.015***		
N	1888	1888	1780		
Mean EL value	0.321	0.345	71.111		
Effect highly-exposed	0.010	-0.010	-8.360		

Notes: Table shows coeffcient with standard errors in parentheses. An observation is a structure by gender pair. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Demeaned 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. Commute is a measure that the individual works outside of the home.Public transport to work is a measure of taking public transport to work. Happy public transport is the response, out of 10, to the question: how happy are you with the public transport options from the house. Live convenient is the response, out of 10, to agreeing with the statement: where I live is convenient for where I want to go to? Happy public transport and live convenient were not asked retrospectively so the final resident treatment effect cannot be computed. Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Table A13: Treatment effects: additional income and employment variables

	(1) Income from self-employment	(2) Works at least 4 hours a day	(3) Self employed last month
Place-based			
Reduction in log traveltime (demeaned)	-12.067	-0.110	-0.062
	38.859	0.155	0.167
N	1815	1882	1882
Mean EL value	74.256	0.327	0.412
Effect highly-exposed	-1.255	-0.011	-0.006
Initial residents			
Reduction in log traveltime (demeaned)	9.143	0.005	0.068
	39.854	0.128	0.135
N	1812	1882	1882
Mean EL value	74.211	0.327	0.417
Effect highly-exposed	0.951	0.001	0.007
Final residents			
Reduction in log traveltime (demeaned)	-49.210	-0.096	-0.039
	38.251	0.159	0.181
N	1811	1674	1674
Mean EL value	73.437	0.331	0.422
Effect highly-exposed	-5.118	-0.010	-0.004

Notes: Table shows coeffcient with standard errors in parentheses. An observation is a structure by gender pair. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Demeaned 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. Self employment is the individual net income from self employment in the last month, if the individual operates a self-employment business. It is measures in 1000 TSH. Household income is the total net household income from all sources in the last month (measured in 1000 TSH). We only provide the household average as it is a household variable. Works at least four hours a day is equal to one if the respondent reports their usual daily hours are at least four. It is set to zero if the respondent does not work. Worked last 7 days is equal to one if the respondent reports earning income in the last 7 days. It is zero otherwise. Self employed last month is equal to one if the respondent reports earning income from self employment in the last month. It is zero otherwise. Standard errors clustered at aggregate spatial unit. Unweighted regressions.

Table A14: Effects of the BRT on place, initial, and final residents (include effects from Phases 2 and 3)

	Include Phase 2				Include Phase 3			
	(1) Employed	(2) Worked last week	(3) Income	(4) Household income	(5) Employed	(6) Worked last week	(7) Income	(8) Household income
Place-based								
Reduction in log traveltime from Phase 1 (demeaned)	0.454	0.974	70.464	224.209	0.260	0.007	6.750	75.278
	0.164***	0.248***	24.411***	105.974**	0.155*	0.213	23.559	96.616
Reduction in log traveltime from Phase 2 (demeaned)	0.159	0.911	63.787	118.380				
	0.146	0.195***	21.328***	97.576				
Reduction in log traveltime from Phase 3 (demeaned)					-0.237	-1.033	-63.147	-179.124
					0.156	0.151***	19.591***	98.104*
N	1888	1888	1834	1483	1888	1888	1834	1483
Mean EL value	0.678	0.391	27.445	345.398	0.678	0.391	27.445	345.398
Initial residents								
Reduction in log traveltime from Phase 1 (demeaned)	0.384	0.910	68.393	186.526	0.286	0.004	6.978	42.431
	0.170**	0.244***	23.569***	106.914*	0.178	0.216	21.866	90.873
Reduction in log traveltime from Phase 2 (demeaned)	0.051	0.827	60.937	118.437				
	0.142	0.180***	18.882***	83.658				
Reduction in log traveltime from Phase 3 (demeaned)					-0.155	-0.999	-61.944	-170.949
					0.148	0.130***	18.010***	92.034*
N	1888	1888	1832	1502	1888	1888	1832	1502
Mean EL value	0.678	0.405	28.062	345.680	0.678	0.405	28.062	345.680
Final residents								
Reduction in log traveltime from Phase 1 (demeaned)	0.323	0.900	66.220	170.697	0.213	-0.028	7.187	18.333
	0.172*	0.245***	26.036**	107.889	0.173	0.215	24.499	83.916
Reduction in log traveltime from Phase 2 (demeaned)	0.078	0.873	59.320	120.499				
	0.149	0.190***	22.798***	86.805				
Reduction in log traveltime from Phase 3 (demeaned)					-0.151	-0.992	-58.017	-183.949
					0.157	0.150***	21.411***	91.131**
N	1888	1885	1826	1470	1888	1885	1826	1470
Mean EL value	0.678	0.391	27.456	345.980	0.678	0.391	27.456	345.980

Notes: Table shows robustness to including additional phases. Columns (1)-(4) included demeaned time to CBD arising from Phase 2 as well as Phase 1. Columns (5)-(8) included demeaned time to CBD arising from Phase 3 as well as Phase 1. Table shows coeffcient with standard errors in parentheses. An observation is a structure by gender pair. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Demeaned 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. Highly-exposed is the effect for households above the 80th percentile of demeaned travel time reduction, as explained in the text. Employment is measured by whether the household member reports an usual occupation other than unemployed, student, housewife, or too old to work. Worked last 7 days is whether the household member reports working in the last 7 days. Earnings are the earnings (in 1000 TSH) during the last seven days. Household income is the total net income for all household members over the last month (in 1000 TSH). Standard errors clustered at aggregate spatial unit. Unweighted regressions.

A Pre-trends

Table A15: Pre-trends analysis

	(1) Log rent per room (expected)	(2) Lived in house < 3 yrs	(3) Typical number of hours worked per week	(4) Wealth index (PCA)	(5) Earnings in last 7 days (1000 TSH)
De-meaned pred. decrease TT CBD Ph1	1.129	-0.027	0.049	0.257	61.692
	0.740	0.397	16.911	1.162	40.435
N	83	83	82	83	82
Mean baseline value	10.261	0.325	51.323	-0.075	33.481

Notes: Unit of observation is enumeration area (EA). Individual-level data is weighted using constructed survey weight when collapsed at the EA level. Standard errors are 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; however, only post-period structures that lie within 2 km of an EA are included in the sample. Pre-period data come from WB Measuring Living Standards in Cities 2014-2015 survey, and post-period data come from Dar BRT baseline survey.

Table A16: Comparing treatment effects at different aggregations (variables used in pretrend analysis)

	(1) Log rent per room (expected)	(2) Hours worked per week	(3) Wealth index	(4) Income	(5) Different household in structure at EL
Structure-gender level					
Reduction in log traveltime (demeaned)	3.163	-1.022	0.281	35.616	0.037
	(0.431)***	(9.258)	(0.707)	(24.070)	(0.084)
N	1751	1882	1888	1834	1888
Mean EL value	9.860	18.081	0.100	27.445	0.059
Effect highly-exposed	0.329	-0.106	0.029	3.704	0.004
Collapsed to enumeration area					
Reduction in log traveltime (demeaned)	1.423	-10.484	-0.968	27.347	-0.132
	(0.652)**	(20.160)	(0.918)	(28.305)	(0.246)
N	80	80	80	80	83
Mean EL value	10.030	17.119	0.200	28.828	0.125
Effect highly-exposed	0.148	-1.090	-0.101	2.844	-0.014

Notes: Table shows coeffcient with standard errors in parentheses. An observation is a structure that was enrolled at baseline. Demeaned 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. Self-reported time to CBD is the time respondents report it would take to travel to the downtown market. Log rent per room (expected) is the cost to rent the household's section of the house divided by the number of rooms the household occupies. Owners are asked about rent if they were to rent out their section of the house. Different household instructure at EL measures whether the baseline household moved between baseline and endline survey. Standard errors clustered at aggregate spatial unit. Unweighted regressions.

B Construction of household weights

We construct a household weight that reweights our sample to match the 2012 Tanzanian Census on ownership population and the distribution of population across wards. This section describes the procedure.[20]

Population data for enumeration areas in Dar es Salaam were obtained from the country's most recent population census, the 2012 Population and Housing Census. This gives aggregated total population counts by sex for each enumeration area. The surveyed area covers 87% of the population of Dar es Salaam.

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 $\Omega_{bct} = N\_hh_{st}$.

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 $\tilde{\pi}_{bc} = \tilde{\pi}_{bc,t=1}$.

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).[21]

Appendix Table A17 shows the distribution of structures across ownership status. Column (1) shows the raw data. Column (2) shows the data after applying the constructed weights. Column (3) shows the distribution arising from the 2012 Census. Column (4) compares the data from the World Bank MLSC data that we use for pretrends.

[20] 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.

[21] Accessed via https://www.statcompiler.com/en/

Table A17: Distribution of households by ownership status of dwelling units

	Endline data		2012 Census data	World Bank MLSC data
	(1) Unweighted	(2) Weighted	(3)	(4)
Owned	65.0	44.7	36.9	36.8
Lived in without paying rent	8.8	8.4	5.5	6.4
Rented	26.2	46.9	57.6	56.8
N	1795	1795	1083381	1964

Notes: Endline data sample consists of all households that answered StatDwel at endline. Constructed weight accounts for number of households in the structure, distribution of owners and renters, and population across city wards.

Table A18: Summary stats (weighted)

	(1) Baseline	(2) 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
Number of members in household	4.02	3.85
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	32299.8
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 food	0.45	0.50
Share of consumption on transportation	0.11	0.11
Above Tanzania national poverty line	0.96	0.90
All initial household members moved out	.	0.19
At least one initial male/female respondent moved out	.	0.40
N	1517	1800
Individuals		
Age	36.8	39.1
Years education	8.39	8.53
Worked for pay last 7 days	0.43	0.42
Employed (as per occupation variable)	0.68	0.65
Operate any self-employed business or activity last month	0.43	0.38
Typical days worked per week	5.87	5.69
Typical hours worked per day	8.56	9.03
Wages last 7 days (1000 TSH)	35.1	30.5
Gross household income (1000 TSH)	566.2	378.0
Net household income (1000 TSH)	419.9	309.9
Gross income from self-employment (1000 TSH)	133.4	103.5
Net income from self-employment (1000 TSH)	91.5	70.0
Commutes (if employed)	0.84	0.73
Commute by walking (if employed)	0.36	0.39
Commute by public transport (if employed)	0.58	0.50
Commute time (mins)	46.7	69.2
Happiness public transport (scale 1-10)	3.21	6.23
Moved house between baseline and endline	.	0.40
Moved house between baseline and endline (renting BL)	.	0.51
N	3104	3824

Notes: The first panel of the table shows summary stats for the structures. The second panel shows summary stats for the male/female respondents (both initial and final residents). Employed is defined from the question: what is your main current occupation, and indicating if someone has an usual occupation other than unemployed, too old to work, student, or housewife. Worked last 7 days is defined from the question: During the last 7 days, were you engaged in any kind of job or work for payment? Self-employment is defined from the question: Did you operate any self-employed business or do any self-employed activity over the last month (including agriculture)? Wages last 7 days is the response to: what were your average total weekly earnings from all jobs, after tax, in the last 7 days? Self-employed income is the response to: what is your gross (net) income over the last 1 month (from self-employed business/activity). Household income is the response to: what is the total gross (net) income for the household over the last 1 month from all sources? All initial household members moved out means that none of the original household (including male/female respondents) remain living in the house. At least one initial male/female respondent moved out means at least one of the male/female respondents moved out, but other household members remained. Commute is a measure that the individual works outside of the home. Modes of public transport are daladala, rickshaw, motorbike, train in baseline, with BRT an additional option in endline. Statistics weighted by constructed survey weight.

Table A19: Effects of the BRT on place, initial, and final residents (weighted estimates)

	(1) Employed	(2) Worked last week	(3) Income	(4) Household income
Place-based				
Reduction in log traveltime (demeaned)	0.479	0.688	33.573	-4.055
	(0.185)***	(0.295)**	(25.639)	(116.411)
N	1888	1888	1834	1483
Mean EL value	0.678	0.391	27.445	345.398
Effect highly-exposed	0.050	0.072	3.492	-0.422
Initial residents				
Reduction in log traveltime (demeaned)	0.306	0.624	36.530	1.785
	(0.202)	(0.254)**	(22.637)	(111.108)
N	1888	1888	1832	1502
Mean EL value	0.678	0.405	28.062	345.680
Effect highly-exposed	0.032	0.065	3.799	0.186
Final residents				
Reduction in log traveltime (demeaned)	0.174	0.631	19.675	-15.849
	(0.160)	(0.249)**	(27.585)	(117.232)
N	1888	1885	1826	1470
Mean EL value	0.678	0.391	27.456	345.980
Effect highly-exposed	0.018	0.066	2.046	-1.648

Notes: Table shows coeffcient with standard errors in parentheses. An observation is a structure by gender pair. Sample is restricted to structures where we can calculate the structure-level, initial-resident, and final-resident treatment effect. Demeaned 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. Highly-exposed is the effect for households above the 80th percentile of demeaned travel time reduction, as explained in the text. Employment is measured by whether the household member reports an usual occupation other than unemployed, student, housewife, or too old to work. Worked last 7 days is whether the household member reports working in the last 7 days. Earnings are the earnings (in 1000 TSH) during the last seven days. Household income is the total net income for all household members over the last month (in 1000 TSH). Standard errors clustered at aggregate spatial unit. Sample is weighted by constructed sample weight.

C Validation of retrospective data

To assess the reliability of the retrospective data we collected at endline, we identify individuals who were surveyed at both baseline and endline and compare their baseline responses with their endline retrospective responses for key outcome variables of interest.

For binary or categorical outcomes, such as employment status or the aggregate spatial unit one resides in, we calculate the proportion of respondents who gave consistent answers at baseline and retrospectively at endline. For continuous variables, we report the share of responses that fall within 10% or 20% of each other across the two surveys.

Appendix Table A20 presents the results of this exercise. Binary and categorical variables show higher consistency when reported retrospectively, whereas continuous variables are less likely to be reported with similar values at baseline and endline.

Home aggregate spatial unit has a lower sample size as it is only asked retrospectively to individuals who have moved to a new structure at endline.

Table A20: Comparison between individuals' baseline and endline retrospective responses

	(1) Employed (main job)	(2) Commutes (distance to work > 0)	(3) Home aggregate spatial unit	(4) Wages in last 7 days	(5) Gross monthly HH income	(6) Total monthly HH consumption
Prop. consistent	0.775	0.797	0.690	0.296	0.066	0.013
Prop. within 10%				0.306	0.097	0.096
Prop. within 20%				0.333	0.195	0.194
Correlation coef.	0.467	0.471	0.666	0.221	0.237	0.179
N	2164	862	525	2032	1556	2864

Notes: Employment status is determined by the respondent's reported main occupation, and is a binary variable. Commute status is defined as one if a respondent reports working and the distance to their work is greater than 0 km, and is defined as zero if the respondent reports not working or reports working but the distance to their work is 0 km. Home aggregate spatial unit is the aggregated geographical neighborhood the respondent reported living in at baseline. This question was only asked to people who moved, so we can only validate it on a smaller sample. Prop. consistent shows the proportion of individuals whose EL retrospective response matches their BL response – this is primarily informative for the binary and categorical variables. Prop. within 10% [20%] shows the proportion of individuals whose EL retrospective response was within 10% [20%] of their BL response – this is only informative for the continuous variables.