

Optimal Public Transportation Networks: Evidence from the World’s Largest Bus Rapid Transit System in Jakarta*

Arya Gaduh, Tilman Graff, Rema Hanna, Gabriel Kreindler, Benjamin A. Olken

December 20, 2022

Please do not circulate.

Abstract

Designing public transport networks involves tradeoffs between extensive geographic coverage, frequent service on each route, and relying on interconnections as opposed to direct service. These choices, in turn, depend on individual preferences for waiting for busses, travel time on the bus, and transfers. We study these tradeoffs by examining the world’s largest bus rapid transit system, in Jakarta, Indonesia, leveraging a large-scale bus network expansion between 2016-2020. Using detailed ridership data and aggregate travel flows from smartphone data, we analyze how new direct connections, changes in bus travel time, and wait time reductions increase ridership and overall trips. We set up and estimate a transit network demand model with multi-dimensional travel costs and idiosyncratic heterogeneity induced by random wait times, matching moments from the route launches. To study the implications for network design, we introduce a new framework to estimate optimal networks and how their characteristics depend on preference parameters. Our results suggest that a less concentrated TransJakarta network would increase ridership and commuter welfare.

*We are grateful to TransJakarta for sharing their data. We thank Allan Hsiao, Felix Tintelnot, Natesh Pillai for useful conversations, and participants at the NBER Summer institute for useful feedback. Lolita Moorena, Nadia Rayhanna, Nadia R Avanti, and Nikhil Kumar provided exceptional research assistance. We thank Deivy Houeix, Aaron Berman, Drew Johnston, Yulu Tang, Van Anh Le, San Singh, Robert Dulin, Kalyan Palepu, Lauren Li, Laya Gollapudi, and Xiaocun Qiu, as well as Margaret Sands, Hoi Wai Yu, Moulinrouge F Kaspar, Ariba Khan, Helen Hu, Paolo Adajar, Jackson C Phifer, and Anna Lingyi Ma, for their research assistant work on different parts of the project. We are grateful to Kyle Schindl, Analysia Watley, Aditi Chitkara, and Nikhil Kumar for their help developing the bus GPS headway engine. This material is based upon work supported by the National Science Foundation under Grant No. 2049784. We acknowledge support from the International Growth Centre, the Harvard Asia Center, and the Harvard Data Science Initiative.

1 Introduction

Megacities of the developing world face severe mobility barriers and traffic congestion, limiting the potential gains from agglomeration forces. As private vehicle ownership rates soar, well-designed public transit, and public bus systems specifically, are key to environmental sustainability and economic inclusiveness for the urban poor with limited access to alternative options. For example, the public bus system of Delhi, India carried 3.3 million passengers per day in 2017, more than its more famous metro system (2.7 million, Hindustan Times 2018).

However, the design of these bus networks involves tradeoffs. For a given number of busses, transportation system designers have a choice between having a more *direct* network (i.e., more direct routes, rather than a hub-and-spoke system which requires more changes between routes); a more *intensive* route network (i.e., more frequent service to a limited number of location); and a more *expansive* route network (i.e., many routes serving more destinations). Because the fixed cost of a bus route is relatively small (compared to, say, building a subway or light-rail line), bus system planners have great latitude in designing a system.

Fundamental to these tradeoffs is an understanding of public transport behavior and demand, and ultimately ridership along the different parts of the bus system: how commuters value bus travel time, wait times in stations for busses, transfers between routes, and walk time to get to the station nearest their origin and destination. Different preference parameter configurations could lead to very different looking optimal transportation networks. Estimating these preferences is challenging, because it requires plausibly exogenous variation in *system-level* attributes of the public transport network (travel time, wait times, availability of direct connections).

In this paper, we examine these questions by studying the expansion of the TransJakarta bus system in Jakarta, Indonesia. TransJakarta is the public bus operator for Jakarta, and operates the largest Bus Rapid Transit (BRT) network in the world. During the period we study, from 2016-2020, TransJakarta launched a total of 93 routes across the city in a staggered fashion. These routes were a combination of feeder routes that operate along regular city streets and then connect to the main BRT trunk lines, and new BRT express routes that created direct connections between TransJakarta's thirteen existing BRT corridors. Moreover, some of these new routes overlap with existing routes for a portion of their journey.

The launch of these 93 new routes creates different types of shocks to the transportation network for different potential journeys throughout the city. We focus on four main types of

‘events.’ For some origin (o) and destination (d) pairs that were previously served only via transfer connection, a new route connecting them would introduce variation in the degree to which the network is *direct*. Second, for other (o, d) pairs that were previously served by a slow transfer connection, a new route may lead to transfer connection options with shorter total travel time, i.e. variation in the degree to which service is *fast*. Third, for some (o, d) pairs that were previously served by direct but infrequent service, a new route that happens to overlap with the old route from o to d leads to more frequent service, i.e. variation in the degree to which service is *intensive*. Fourth, for some (o, d) pairs previously connected by transfer options, a new route leads to more frequent transfer service at the origin.

Our paper proceeds in three steps. We first use the route launch events to estimate the sensitivity of ridership and of overall commuting to changes in expected waiting times for busses, bus travel time, and the necessity of making bus connections. Second, we set up a model of demand for public transportation, introducing a new formulation of network routing with wait times. This framework has an invariance property for route aggregation, and choice probability and expected utility expressions are tractable. We estimate the model’s preference parameters by matching the revealed sensitivity of ridership to the changes in the bus network we observe. Third, we introduce a new framework for estimating the characteristics of optimal networks. We prove that when the planner faces idiosyncratic shocks over networks in addition to the model welfare measure, the problem becomes to sample from a probability distribution over networks. We use a version of the simulated annealing algorithm to achieve this. We use these ‘sampled optimal networks’ to describe the optimal network characteristics (intensity, extent, shape measures), estimate how these network characteristics vary with fundamental model parameters, and compare these results with the characteristics of the network TransJakarta has actually built.

We bring a variety of datasets together to analyze these questions. TransJakarta collects fares using a tap-to-pay electronic money system. To measure (o, d) ridership flows on the TransJakarta network, we use administrative data from TransJakarta that includes every tap into the network (over 500 million taps over the 4 years we study), along with an anonymized card identifier. Since the system enforces tap-in everywhere in the network, but only uses tap-out at a subset of BRT stations, we use an algorithm to infer the likely destination (d_{it}) for each trip it .¹ We validate that this approach works well using the subset of trips that end in a station that uses tap-out and hence for which we directly observe both o_{it} and d_{it} .

To measure overall commuting flows on each (o, d) pair (regardless of whether commuters use TransJakarta or private transportation) we use anonymized smartphone location data,

¹To predict the destination, we use the origin of the next trip made using the same card (i.e., we infer that $d_{it} = o_{it+1}$), or one of the top two stations for that card.

which allows us to compute home and work locations (and, hence, $o - d$ commuting flows) to and from every 500m by 500m grid cell across the city. To measure wait times and travel times on the network, we use GPS data that tracks the position of every TransJakarta bus every 5-10 seconds over the 4 years of our study. Finally, we have a cross-section of typical driving times between every pair of locations across the city, measured using a routing API.

The first step in our analysis is to use the new route launches to estimate how ridership depends on improvements in wait times, travel times, and direct connections. We study four specific types of ‘events’ that affect different (o, d) pairs. The first event type is the launch of the first direct bus line between (o, d) (i.e. which allows a passenger to travel between (o, d) without changing busses), for the sample of (o, d) that were previously connected by transfer.

The second event type is a new route launch that creates a new, faster *transfer* options between o and d that were previously connected via a transfer. This is akin to the first type of event, except that with event 2 we now study a new, faster route that still involves a transfer, whereas with event 1 we study new direct (i.e., no transfer) route.

The third set of events we consider are events that change the frequency of busses, in order to estimate the sensitivity of ridership to wait times. The challenge, of course, is that in general TransJakarta allocates busses to routes in response to demand, so one needs variation in bus frequency that is plausibly exogenous. We study the ‘incidental’ addition of new busses traveling direct between o and d that are caused by the launch of the new route that overlaps, for some portion, with an existing route. That is, many new bus routes serve outlying parts of the city, but then join and overlap with other routes for part of their journey. For this portion (where it overlaps an existing route), the new route launch creates variation in the frequency of service, since a rider traveling only on the overlapping portion can take either one the original route(s) or the new route serving the same stations.

The fourth set of events are a counterpart of the third type, only focusing on transfer routes instead. That is, a new route introduces new transfer options between an origin and a destination, and this increases the arrival rate of busses (and hence lowers wait time) at the origin station.

We analyze the impact of all three types of events in a differences-in-difference setting, using a rich set of fixed effects for origin-destination pairs, origin-time and destination-time.² We run all analysis using PPML, separately for events that affect BRT routes, and events

²We also include a large set of “never-treated” origin-destination pairs, which largely alleviates the class of problems with two-way fixed effects specifications that have been highlighted recently (see, e.g., De Chaisemartin and D’Haultfoeuille (2022) for a discussion).

for non-BRT routes.

For all types of events, we begin by documenting a substantial effect on wait times and travel times, i.e. a substantial first stage. Specifically, using our detailed bus GPS data, we estimate that for BRT routes, the first event type – a new direct route between a given (o, d) pair – reduces travel time on the bus by about 13 percent, in addition to eliminating a transfer. The second event type – the reduction in transfer option travel time due to a new route launch – reduces travel time by 0.54 log points, and increases bus arrivals by 0.47 log points, thus decreasing wait times, but by construction still requires a transfer. The third event type – the ‘incidental’ addition of busses caused by a new route launch – increase bus arrivals by 0.37 log points, but induce only a small 3.1% reduction in travel times. The fourth event type also increases bus arrival rates significantly, with a small effect on travel time.

We then examine how ridership responds to each of these three types of improvement in service quality. We find that ridership is responsive to each of these three types of events. A new direct route between an existing (o, d) pair increases ridership by 20 percent in our preferred specification. The second event type – which leads to substantially faster transfer options – increases ridership by 5-10%. In the third type of event, reducing wait times by 37 percent (through adding busses via an incidental new route) increases ridership by about 6 percent, so ridership also responds substantially to wait times. Event-study graphs show clear jumps in ridership immediately after the launch of new routes, and no pre-trends. The ridership effects take about 1-3 months to fully kick in as riders presumably take some time to adjust their routines to new bus availability. Results for non-BRT routes are qualitatively similar and typically larger in magnitude.

We next look at the impact of the same set of eight events on all trips between a pair of locations (regardless of whether they take TransJakarta or some other means of transportation), measured using the smartphone location data. We do not find any statistically significant and positive effect on all trips for any of the events in our benchmark specification, where we use 1km hexagonal grids. While these estimates are somewhat noisy, we can reject moderate positive effects. For example, for the first type of BRT event, we can reject at 95% level a positive impact on the inverse hyperbolic sine of the number of trips of 0.11, whereas the effect on TransJakarta ridership on the same sample is 0.20. Overall, these results are consistent with mode substitution towards TransJakarta.

In the second step in our analysis, we use a model of commuter travel behavior to interpret the ridership impacts from the reduced-form events, and to estimate underlying preference parameters. The core of the model is a new formulation of how commuters choose public

transport routes within a geographically realistic transit network, in the presence of stochastic wait times. A commuter traveling from an origin station o to a destination station d chooses between direct and transfer bus options that differ in terms of type (BRT or non-BRT), total travel time, and the necessity of a transfer. We assume that commuters are “non-planning,” that is they cannot choose their departure time as a function of bus arrivals, and that busses on each route arrive according to a Poisson process.³ The commuter gets a draw of wait times for all routes in her choice set, and selects the best option overall. This specification allows for realistic choice patterns depending on the vector of wait times that a given commuter draws, such as skipping the first bus that arrives in the station in order to take a better connection. We derive computationally tractable expressions for choice probabilities and expected utility. The model is invariant to aggregating identical routes, a property that allows us to bypass the challenges of defining choices in discrete choice models such as logit.

We embed the network routing model in a wider model of urban mobility decisions. Commuters choose between using the direct routes in the bus network, transfer routes, or a “private transportation” outside option (this captures options such as motorbike, motorcycle taxi, car or private busses) based on a mixed logit specification.⁴

In the model, the commuter’s choice utility depends on 7 unknown parameters. We assume that one parameter, the disutility from bus travel time, is common to BRT and non-BRT options. We estimate the remaining three parameters, two parameters for the disutility of having to change busses, and the disutility from time spent waiting for busses to arrive, separately for BRT and non-BRT routes. We also flexibly estimate origin-destination fixed effects that capture time-invariant attractiveness of the private option between each pair of origin and destination.

To estimate the model, we discretize the Jakarta metropolitan area into 1km hexagonal grid cells. We use the anonymized smartphone data to estimate commuting flows across all possible combinations of these grid cells.⁵ For a given set of preference parameters θ , and the state of the TransJakarta network at time t , the model predicts what share of people traveling from o to d will do so via TransJakarta, so combined with the total number of travelers going from o to d , we have model-predicted ridership. We can then see how predicted ridership

³This implies that wait times in any station are exponentially distributed. We use GPS bus data to show that an exponential wait time distribution is a good approximation for BRT and non-BRT routes. (In the ideal case with equally spaced busses, the wait time distribution would be uniform.)

⁴While it is possible to further embed the model into a destination-choice nest, or in a urban general equilibrium model (Tsivanidis, 2022), given the null reduced form effect on aggregate ridership in our benchmark specification, we do not include this feature in the model.

⁵There about 18 million people who live in the study area within metropolitan Jakarta; so while many (o, d) pairs will have some commuting flows, some will be empty.

$\widehat{R}_{odt}(\theta)$ changes over time as the TransJakarta network changes. To estimate the preference parameters θ , we simulate in the model how much ridership should increase for each of the eight event studies in step 1 (for BRT and non-BRT), and use a classical minimum distance estimator to find the preference values to match the actual event-study estimates.⁶ This allows us to translate the event-study estimates from step 1 into a set of underlying preferences parameters.

Using this approach, we find high disutility for waiting time relative to time traveling on the bus, both for BRT and non-BRT. We also find that most commuters have very large utility costs from taking transfers, but we also estimate a high dispersion in these costs, which implies that for some commuters these terms are positive (they prefer transfers). In general, these results imply that bus ridership is not very responsive to changes in attributes of transfer options, which is what we find in the reduced form results.

What do these parameters mean for the optimal design of the network? In the third part of the paper, we introduce a technique to characterize how certain network statistics of the optimal network depend on structural model parameters.

Specifically, we ask – for a given set of busses (e.g., 1,500 busses, which is approximately the size of TransJakarta’s fleet) – how they should be configured across the city. This involves balancing tradeoffs across reach (having enough bus stops to cover large portions of origins and destinations for which people would like to travel), wait times (putting fewer or more busses on a given route), and the topology of the network (which stations should be connected directly and which to require changes, taking into account the fact that busses travel from station to station). We hold fixed TransJakarta’s current BRT infrastructure, which allows for faster travel times along the 13 BRT corridors.

We study this problem using a realistic geography of Jakarta, and in particular, taking into account actual travel times from each grid cell to an adjacent grid cell. This is this a very high-dimensional problem, and the number of possible configurations is extremely large.⁷ Our problem displays divergent substitution and complementarity forces at the same time, hence violating the type of single-crossing property that enables more efficient characterization of optimal allocations in other settings (Jia, 2008; Arkolakis and Eckert, 2017).

To tackle this complex problem, we proceed in two steps. First, we assume that the

⁶To help estimate the value of travel time, we also include an additional moment that measures for BRT bus trip duration between o and d , measuring from tap-in/tap-out transaction data, depends on the *mean* travel time on the bus among all bus options between o and d , conditional on the *minimum* travel time.

⁷To give a sense of how large this problem is, note that with 418 grid cells and 1,536 possible edges that could be either connected or not, the number of possible networks reaches at least $2^{1,536}$ possibilities – i.e. about 10^{500} possibilities - of just the route design, not even taking into account the number of busses on each route.

planner chooses a network that maximizes welfare as given by our model – average expected utility for all commuters – plus an idiosyncratic shock for each possible network. The latter captures factors that the planner cares about that are not included in our model. We assume shocks are i.i.d. following an extreme value type-1 distribution with standard deviation proportional to β^{-1} , leading to a multinomial logit probability distribution of the network chosen by the planner, which we wish to estimate. In this setting, a network with welfare close to that of the model’s global optimal network carries a large probability of being chosen, because small idiosyncratic shocks may tip the planner into preferring it. As the space of networks is extremely large, we cannot explicitly calculate the planner’s logit probabilities over all networks, but we can calculate the relative likelihoods among possible network choices.

We have transformed a global optimization problem into a problem of sampling from a large-scale logit distribution. This has the advantage that several algorithms have theoretical guarantees and good practical performance in sampling from such distributions.

In the second step, we show that a version of the simulated annealing (SA) algorithm, stopped when the algorithm’s temperature is β (the planner’s logit parameter), allows us to sample from the distribution of optimal networks.⁸ This result holds asymptotically as the duration of the SA algorithm increases. In addition to the formal proof of this result, we report several numerical checks of the output of the SA algorithm. In particular, the distribution of welfare from networks obtained from independent SA runs is highly concentrated. In practice, we run the SA multiple times with random independent initial conditions to obtain networks sampled from the distribution of interest.⁹

We use this framework to compute comparative statics of how optimal networks depend on the structural preference parameters that we have estimated. We do so by defining network statistics by integrating over the distribution of optimal networks calculated using the SA approach for a given preference parameter vector θ , and estimate local and non-local effects of changing structural model parameters by comparing these statistics among optimal networks calculated using different values of θ .

Using this approach, we find that optimal networks, based on the preference parameters we estimate, are substantially more extensive than the actual network. For example, we estimate that a typical optimal network should cover about 88 percent of all grid cells and 99 percent of all potential riders, compared to 42 percent and 73 percent, respectively, with the actual TransJakarta network. It should feature 1,681 km of bus routes, compared to 550

⁸We point out that two other algorithms, Metropolis-Hastings and parallel tempering, also allow sampling from the planner’s distribution over networks.

⁹While computationally intensive, this procedure is analogous to “thinning” the output of a long Metropolis-Hastings process in order to obtain approximately independent samples from a distribution.

km in the actual network, and hence would mean that people would need to wait longer for a bus.

We explore how the distribution of optimal networks changes if preferences differed. For example, halving the wait time cost leads to even more expansive networks covering 1,818 km – 8% more than before. The results suggest that the current operator, which may be used to operating a hub-and-spoke system, may not be fully taking advantage of the flexibility possibilities that bus systems can create.

Our project connects to several literatures. First, we build on classic questions in the transportation economics literature: travel demand estimation and mode choice in particular (McFadden, 1974; Ben-Akiva et al., 1985), and increasing returns in public transport or the “Mohring” effect (Mohring, 1972). In particular, we build on the large literature investigating how commuters value certain trip attributes such as wait time and transfers (see Abrantes and Wardman (2011) for a review and Small et al. (2007, page 53) for a discussion). Studies in this literature tend to focus on rich countries and typically use preference elicitation using stated preference techniques, whereby respondents’ choices between hypothetical alternatives are used to infer valuations. We are focusing on bringing rigorous causal estimation to these questions, estimating them using the natural experiments given by route openings to generate empirical variation in an entire citywide network of routes. In that respect, our study is closest to Kreindler’s 2022 study of congestion pricing, which uses a field experiment with individual drivers to estimate the underlying preference parameters required to calculate optimal congestion pricing in Bangalore, India.

Second, by embedding urban travel demand preferences into a model of optimal bus route network design, our project also contributes to the growing trade-inspired literature that tackles questions of how to design a transport network (Fajgelbaum and Schaal, 2020; Allen and Arkolakis, forthcoming; Balboni, 2021; Santamaria, 2022). This literature has so far focused on road infrastructure (typically inter-city), rather than within-city public transport. Our detailed micro data allows us to estimate fine parameters that are usually challenging to incorporate in these studies, and shows how they are related to features of the (optimal) network.

Finally, our paper is also related to the small-but-growing literature on the impact of transit systems, particularly bus-rapid-transit systems, in the developing world (Tsivanidis, 2022; Balboni et al., 2021). These studies focus on the impact the BRT systems in Bogotá and Dar Es Salaam, respectively. Tsivanidis, for example, focuses on estimating the welfare and inequality effects of the system and how it affects the organization of the city. Gaduh et al. (2022) studies the impact of the initial launch of the TransJakarta system between

2002 and 2010, to show that at that point in time the system had no detectable impacts on public transport use but the conversion of road lanes into dedicated lanes for the BRT network led to worsened road traffic congestion. All these papers take the public transport system design as given. By contrast, we tackle the complementary problem, focusing on estimating underlying preference parameters, and solving for the optimal system design.

The remainder of this paper is organized as follows. Section 2 describes the TransJakarta bus network, the expansions we study, and the data we use. Section 3 presents the reduced-form event-study results that show how ridership responds to new routes and improved bus frequency. Section 4 introduced the model and then describes how we estimate the model using the reduced form moments in the previous section, and presents estimation results. Section 5 explores the implications of the estimated utility parameters for optimal network design.

2 Setting and Data

2.1 The TransJakarta Bus Network

2.1.1 Setting description

We study TransJakarta, the integrated bus system for Greater Jakarta, Indonesia. The city of Jakarta has a population about about 10.5 million people; Greater Jakarta, (known as JaBoDeTaBek, and consisting of the capital city Jakarta plus surrounding districts of Bogor, Depok, Tanggerang, and Bekasi), has a population of about 33 million people, making it by some counts the second most populous metropolitan area in the world.

TransJakarta consists of over 108 bus routes. This consists of a mix of Bus Rapid Transit (BRT) routes, which are bus routes that operate on special reserved bus lanes, as well as non-BRT feeder routes that operate partially on normal city streets as well as on the BRT corridors, which serve to connect locations further away to the BRT system. The BRT system, with a network length of more than 120 miles of BRT corridors, is the longest such system in the world ([Institute for Transportation and Development Policy, 2019](#)). TransJakarta routes are concentrated primarily within the city of Jakarta, though some extend to surrounding districts in the metropolitan area. Established in 2004, TransJakarta has been serving hundreds of thousands of passengers daily: its highest recorded daily ridership was around 998,000 on December 16, 2019.

TransJakarta is the primary public rapid transit system in the city. For most trips, the primary alternative to TransJakarta is private transport, consisting of a mix of motorcycles, private cars, motorcycle taxis, known as *oje*k, and automobile taxis. In addition, there is a

commuter rail system (KRL CommuterLine) that serves outlying areas, which also serves about 1 million people, but these are longer trips that for the most part do not overlap with TransJakarta trips. There is also a single 16km subway line that was opened in 2019 that serves about 80,000 riders per day.

Pricing on TransJakarta is a flat-fare of IDR 3,500 (USD 0.25) per trip regardless of distance. Payment is collected by tap-in smart cards, either at BRT stations or through fare machines on non-BRT busses; tap-out is enforced at a small number of BRT stations as well. Free transfers are allowed at BRT stations.

2.1.2 TransJakarta network expansions

The main source of variation that we use in the paper is the large expansion of the TransJakarta network since 2016. Between January 2016 and February 2020 (the end-point of our study, just before the COVID crisis struck Indonesia), the TransJakarta network added around 90 BRT and non-BRT routes, up from a basis of 23 routes in 2016. The BRT routes consist of one entirely new corridor (i.e. new segment of bus-lanes) as well as new BRT routes that run along existing BRT corridors, sometimes connecting two corridors, or running an express route on certain portions.¹⁰

The new non-BRT feeder routes stop both in some BRT stations as well as in road-side non-BRT bus stops, and they connect one or multiple BRT corridors to other areas of the cities. The number of busses in operation more than doubled, from about 700 at the start of the period we study to more than 1,600 at the end. The expansion of the system is shown graphically in Figure 1, Panel A.

These expansions – for both new BRT and new non-BRT routes – took place at different times throughout the city. Figure 1, Panel B, shows the number of each type of route operating at different times.

This expansion followed soon after the 2014 restructuring of TransJakarta from a government transport department unit to a regional owned enterprise (PT Transportasi Jakarta), a public company with increased autonomy and performance indicators (ridership, headway targets).

The decision process for launching new lines featured a mix of external constraints and discretion. All non-BRT route launches in our sample were chosen from a list of routes that was pre-approved in 2016 by the Jakarta transportation department. For BRT routes, TransJakarta created new routes by connecting existing BRT stations along the existing BRT dedicated lanes. (The only exception are the BRT routes along the new corridor

¹⁰The new corridor 13 is built along an elevated busway, so this did not affect the number of lanes available to road traffic.

13, a primarily elevated route launched in August 2017.) For both BRT and non-BRT, TransJakarta chose routes and launch periods based on new bus fleet delivery dates (which were often delayed), based on bus availability due to expiration of contracts with operators, inputs from field reports, and other factors.¹¹

In our empirical analysis below, we use a differences-in-difference framework to study the impact of new route launches, using a rich set of controls (including origin location by time period and destination location by time period) to account for time-varying trends in ridership that may have influenced the choice of route launches. Appendix Table A.1 shows that the order of BRT route launch was in general balanced with respect to geographic characteristics of the routes.

New routes alter the desirability of using TransJakarta for a given trip differently for different pairs of origin and destination locations throughout the city. TransJakarta routes often overlap, hence a new route can alter the waiting time for certain portions of a pre-existing path due to the additional busses that now travel along the overlapping portion. New routes create new transfer route options and sometimes reduce the travel time necessary to travel between certain origin and destinations. When two locations become connected by a new route, the change in ridership is informative about the transfer penalty (i.e. the utility cost of switching from one route to another), conditional on the other preference parameters.

2.2 Geographical Environments

Study area. Throughout the paper, we focus on the Greater Jakarta area, consisting of the Special Capital Region of Jakarta (DKI Jakarta) and all the surrounding urban districts: Tangerang, South Tangerang, Depok, and Bekasi (see Figure 4). The study area has a population of over 18 million based on the 2020 population census, of which over 14 million individuals are over 15 years old. The TransJakarta network is concentrated in DKI Jakarta, and all its stations are included in the study area.

Grid cell environments. We use consistently defined geographical environments throughout the analysis in the paper, for the reduced form, demand model estimation, and optimal network simulations. We divide space into identical grid cells and aggregate stations at the grid cells level, and trips at the grid cell pair level.

We focus on a regular hexagonal tiling where adjacent grid cell centroids are 1,000 meters apart throughout the reduced form and demand estimation analysis, and show robustness to using 500 meter square grid cells. For the optimization section, we use 2,000 meter square

¹¹TransJakarta subcontracts operations to several other companies, who own a large share of the TransJakarta busses.

grid cells to reduce the problem’s dimensionality. In future versions of this paper, we plan to use the same definition as in previous sections.

2.3 Data

We use four different datasets for the project – data on **TransJakarta ridership**, which we obtain using administrative data on every entry into the system from TransJakarta’s smart-card tap entry system; data on origin-destination **overall commuting flows**, which we obtain from anonymized smartphone location data; data on **bus locations**, which we obtain by processing detailed GPS data on every TransJakarta bus throughout our period; and cross-sectional data on **driving times**, which we obtain from a different anonymized smartphone-based dataset. We describe these each in turn.

2.3.1 Ridership data

We use the administrative ridership data that is captured electronically by smartcards to construct a highly granular TransJakarta origin-destination (o, d) ridership matrix at each point in time since 2016 (and from 2017 for non-BRT). In BRT stations, passengers tap smartcards to go through turnstiles to enter the boarding area. In non-BRT bus stops, passengers board the bus directly from the street and tap to pay for the bus ticket shortly after boarding. In around 25 percent of BRT stations, passengers also tap smartcards when they exit the station. We observe the time of each tap, as well as an anonymized identifier for the smart-card used, which allows us to link transactions from the same smart-card over time.¹² For taps at BRT stations, we also observe the station identifier. In 2019, these data had around 800,000 transactions per day, with about 1.6 million unique users per month.¹³

To convert this tap data into (o, d) trips, we do two things. First, for non-BRT taps, where people tap into the system on the bus after boarding, we observe only the time and bus identifier, but not a station identifier. We use the detailed bus GPS data (described in Section 2.3.3 below) to identify the most recent station the bus based prior to the tap, and use this to identify the trips origin (o_{it}). Second, note that while we observe each passenger’s entry-point into the system for each trip (origin o_{it}), we only observe destinations (d_{it}) for the trips that end at stations where tap-out is enforced. For the remainder of the trips, we

¹²Note that individuals can, of course, purchase new smart-cards at will, but they do so relatively infrequently. The median smart-card in our data is active over a period of 4 months, and the median tap belongs to a smart-card that is active for over 20 months.

¹³Note that we drop “administrative” cards that are likely used by bus attendants or other TransJakarta employees. We label a card as administrative on a given day if it is used repeatedly throughout the day on the same bus (for non-BRT transactions), or at a BRT shelter. We assume that travel behavior for non-admin cards is representative of other behavior.

construct a proxy destination station for a given trip using an algorithm. If possible, the algorithm picks the next origin station for that smartcard – that is, we infer that $d_{it} = o_{it+1}$. If there is no other trip that day or on the following day, and if the origin station is one of the “top two” stations for that card in a given month, the algorithm infers the destination to be the other “top two” station. The algorithm and ridership data cleaning process more generally are described in detail in Appendix A.3.4.

To validate the algorithm that infers destinations, we use it to construct ridership flows for all trips with and without exit transaction information. For the set of destination stations where we do have actual exit transactions, bivariate regressions of daily ridership flows between stations using the two methods have R^2 of 0.85, suggesting this procedure works well (see Appendix Figure A.11).

2.3.2 Aggregate trip flows

We augment our administrative data with anonymized historical smartphone location data to measure overall trip behavior, regardless of whether a given commuter uses the TransJakarta system. First, we will use this measure of aggregate travel as an outcome variable when studying the impact of the network’s expansion. Second, together with the ridership data, overall trip flows allow us to compute choice probabilities for using TransJakarta. The data processing algorithm and validation results are described in Appendix A.3.5.

We have daily smartphone location data from March 2018 through March 15, 2020 from Veraset, a private provider that aggregates and cleans such smartphone location data. This data covers 35,788,413 weekday trips that belong to 2,370,901 unique devices in our study area. (Recall that the study area had 14 million individuals over 15 years old at the 2020 census.)

We construct two main data sets using the Veraset trips data. First, we construct a cross-section of typical (o, d) trip flows throughout the period. For each device in the data, we reweight all its weekday trips to represent a single typical weekday, and re-weight all devices to make the set of all devices representative of all individuals over 15 years old. Second, we construct a panel of trip flows, where each week we use the same reweighting procedure, separately for each week in the data.

We assess the representativeness of the Veraset smartphone trip data by counting the number of devices that have a home location in a given urban neighborhood (kelurahan) in the study area.

To validate this measure, we classify the home location for each user as the most common recurrent location and assign the home location to an urban neighborhood (kelurahan), the

smallest administrative unit where we have residential population data. We then compare the number of devices with home in a given kelurahan with the population from the 2011 PODES survey, the most recent available population data at fine geographic resolution in Jakarta. Appendix Figure A.12 shows the bivariate scatterplot, which shows that smartphone residential devices and 2011 PODES population are positively related. The veraset commuting data is not a perfect representation of the commuting flows of the city. Appendix Figure A.13 prints the correlation between a location’s smartphone devices and population density (Panel A), as well as an indicator for poverty (Panel B). Poorer, more dense areas are slightly underrepresented in the sample.

The smartphone location data captures rich patterns of travel within Jakarta DKI. Appendix Figure A.14, which shows two maps with origin (residential) and destination totals corresponding to the morning commute. Destinations are more concentrated than origins, especially in Jakarta’s central business district, as well as in other areas around the city.

2.3.3 Bus location and allocation data

We have data on planned and realized daily bus allocation for each route, as well as more detailed bus trip and exact bus location data. Specifically, TransJakarta records the bus location every five to ten seconds, as well as an identifier of the bus route, for the near-universe of routes. These data cover approximately 1,800 BRT and non-BRT (feeder) busses and 16,000 bus trips per day since January 2017 (so we have approximately 70 billion bus-GPS-time location points to work with).

We process this raw GPS data into “headway” information, which allows us to calculate the arrival time of a bus at a given station on a given route. This allows us to calculate both actual and expected wait times for each bus at every station throughout the period we study. We use this bus location data to assign stations to ridership taps for non-BRT busses, as described above.

We also use this data to characterize the *distribution* of wait times separately for each route. Appendix Figure A.9 show that wait times are approximately *exponentially* distributed. This finding supports our later modeling assumption that bus arrivals follow a Poisson process.¹⁴ Moreover, BRT and non-BRT routes have similar levels of variance in wait times for a given average wait time (Appendix Figure A.10).

¹⁴With equally spaced out busses, a traveler who arrives in a station at a random time faces a uniformly distributed wait time. The exponential distribution we see in practice highlights the uncertainty pervasive in real public transport networks.

2.3.4 Geographic Data on Bus Network Expansion

To reconstruct the TransJakarta network at different points in time, we combine data from TransJakarta and from Trafi, a mobility planning app available in Jakarta during the study period.

Using Trafi data, we obtain station locations and route trajectory in each direction (including the sequence of stations). This allows us to construct the network of bus routes, including the BRT stations where transfers are possible. We use the TransJakarta data on planned bus allocation by route and date (described above) to infer the launch date for each route. We cross-check these launch dates with route- and date-specific ridership data, as well as directly with TransJakarta staff.

2.3.5 Driving times

We obtain driving travel times between millions of pairs of locations in Jakarta from a commercial provider of route data derived from smartphones and other GPS-enabled devices. We obtained data for the entire Jakarta region in the year 2020.¹⁵

3 Ridership, overall trips and service quality: Reduced form results

We begin by estimating the ‘reduced form’ impacts of improved service quality on TransJakarta ridership and on all trips (measured using smartphone location data). We use the fact that, for different (o, d) pairs in the network, the launch of a new bus route may have different effects on how attractive it is to take TransJakarta from o to d . A given route may, for example, allow one to travel from o to d without having to change busses, or it may simply reduce travel time by taking a more direct route. In other areas, the new route may overlap with an existing route – for example, many existing non-BRT routes join the BRT for a stretch, and overlap with existing BRT route for a segment. From the perspective of a passenger traveling between these overlap areas, these new routes are simply an increase in service frequency from o to d .

We focus on four types of ‘events’ induced by new route launches, which can be described succinctly as follows. The first type of event captures when two locations that are already connected by *transfer* get a *direct* connection for the first time. The second type of event

¹⁵Through additional partial city coverage from earlier years, we confirmed that it captures pre-COVID traffic congestion patterns and that aggregate congestion appears flat since 2016. Thus, the 2020 data appears sufficient to describe the outside option.

captures when two locations already connected by transfer get a *faster* transfer connection. The third type of event captures when two locations that are already directly connected get additional busses traveling along that direct route (because another route overlaps with the existing (o, d) direct route(s)), thus increasing the bus arrival rate and lowering wait times for those traveling from o to d . The fourth type of event captures when two locations already connected by transfer get one or more additional transfer connections that induce a higher bus arrival rate at the origin location.

In this section, we examine these differential types of service changes, separately for BRT and non-BRT connections, to understand the degree to which ridership and overall trips are sensitive to changes in public transport service quality. We then use these reduced form moments when we fit the model in Section 4 below.

3.1 Estimation framework

We use the following event-study type estimating equation for the reduced form analysis. We will estimate each of the four event types described above both for BRT network expansions ($E \in \{1B, 2B, 3B, 4B\}$) and for non-BRT ($E \in \{1N, 2N, 3N, 4N\}$). In each case, we estimate:

$$Y_{odt} = \alpha^E Post_{odt}^E + \alpha_{-10}^E L_{\leq -10, odt}^E + \alpha_{10}^E L_{\geq 10, odt}^E + \mu_{od}^E + \nu_{ot}^E + \xi_{dt}^E + \varepsilon_{odt}^E \quad (1)$$

Our dataset is at the origin (o) \times destination (d) \times time (t) level, so we include all two-way fixed effects. Specifically, μ_{od}^E are origin \times destination fixed effects – i.e. fixed effects for every combination of start and end grid cell – and ν_{ot}^E and ξ_{dt}^E are origin \times time and destination \times time (week-level) fixed effects.

The key variable of interest is $Post_{odt}^E$. This is just a dummy for the event having taken place on the od route in the past 10 months.¹⁶ We control for $L_{\leq -10, odt}^E$ and $L_{\geq 10, odt}^E$, which are dummies for whether an event between o and d takes place 10 or more months in the future, and in the past, respectively. The coefficient α^E in equation 1 thus captures the overall effect of an event of type E for the pair o and d in the first 10 months after it occurs.

We cluster standard errors two-way at both origin and destination grid cell level, which allows for arbitrary serial correlation over time for each origins and each destination, as well as arbitrary correlation among destinations for a given origin, and vice-versa (Cameron et al.,

¹⁶For event-types 2 and 3, it is possible that there can be multiple events for a given (od) pair that occur at different times. We focus only on the first event. For BRT event type 2, less than 2% of od pairs have multiple events. For BRT event type 3, this number is 39.5%, and the maximum number of events is 5. Our results are robust to defining $Post_{odt}^E$ to measure the number of events for origin-destination pair od in the nine months before week t .

2011).

The sample includes only odt observations that are connected at time t within the TransJakarta network, implying that both o and d contain TransJakarta stations before each type of event.¹⁷ The sample of pairs od for event type E is restricted to all origins treated at least once – $Post_{od't'}^E = 1$ for some d', t' – and to all destinations treated at least once. Most of the origin-destination pairs included in the sample are never treated. (For example, for BRT event type 1 this number is 86.2%.¹⁸

We use two types of outcome variables Y_{odt} . First, we examine the total TransJakarta ridership between o and d during week t , regardless of what route is taken. Second, we also measure impact on all trips between o and d in week t , computed based on smartphone location data.

In our benchmark specification, we estimate (1) using robust Poisson (PPML). In a robustness exercise, we also estimate OLS regressions where the outcome is the inverse hyperbolic sine transformation of the respective outcome variable.

We use the full time period of data we have for each type of event, but since the BRT, non-BRT, and smartphone trip data begin at different dates, the data start at different times.¹⁹ All data end in mid-March 2020, prior to the COVID-pandemic shutdown.

For each of the events described below, we present the overall effects from estimating the corresponding version of equation 1. We also present event study graphs, where we add month-by-month leads and lags of the key explanatory variables.

3.1.1 Event type 1: from transfer route to direct route

The first set of events we examine are events where a new route launch creates a direct connection between o and d where one did not exist before, but where o and d had previously been connected using a transfer.

For BRT events, we define $Post_{odt}^{1B}$ to be a time-varying dummy for the new direct connection between BRT grid cells o and d having been switched on in the past 10 months.²⁰

¹⁷For all non-BRT events E , we also impose that the origin o is never a BRT grid cell, to separate these from BRT events.

¹⁸The fact that most of our od pairs are never treated largely alleviates the class of problems with two-way fixed effects specifications that have been highlighted recently (see, e.g., De Chaisemartin and D'Haultfoeuille (2022) for a discussion), which are primarily a concern when using previously-treated observations as a comparison group.

¹⁹For BRT events, when we estimate impacts on TransJakarta ridership, the sample of time periods t is between January 2016 and mid-March 2020. For non-BRT events, we further restrict to the period after mid-January 2017, when our non-BRT data begins. When we use smartphone trips as an outcome variable, the time periods span March 2018 to March 2020, following the dates we have this data. In all cases, we exclude the period May-July 2018 due to missing BRT ridership data.

²⁰We impose that the the origin and destination grids have BRT stations, while the new route may be a

The coefficient α^{1B} in equation 1 captures the overall effect of a new direct connection being added between o and d in the first 9 months after its launch. The sample of observations is all odt that have a direct or transfer connection between BRT grid cells o and d , and o and d each have at least one treated observation.

For non-BRT events, $Post_{odt}^{1N}$ is a time-varying dummy for the new direct connection between grid cells o and d having been switched on in the past 10 months. We restrict to origin grids o that are never BRT, o and d are connected either directly or by transfer. (In the latter case, the first leg is necessarily non-BRT, while the second leg can be BRT or non-BRT.)

3.1.2 Event type 2: faster transfer

The second type of event is when grid cells o and d are initially connected by transfer, and a new route launch introduces a transfer option that is at least 0.25 log points ($\approx 28\%$) quicker than the fastest transfer option available before.

For BRT events, $Post_{odt}^{2B}$ is a dummy for the first transfer travel time reduction between o and d taking place in the ten months prior to week t . A positive coefficient α^{2B} captures how ridership or all trips between o and d increase after a quicker transfer route becomes available between o and d . The sample of observations is all odt that have a transfer connection between BRT grid o and BRT grid d in week t , grids o and d are never directly connected, and o and d each have at least one treated observation.

For non-BRT events, $Post_{odt}^{2N}$ is defined analogously. The sample is all odt that are connected by transfer at time t , o is never a BRT grid cell, grids o and d are never directly connected, and o and d each have at least one treated observation.

We also consider separate “small” event types $2B$ and $2N$, defined analogously except that the minimum travel time between o and d falls by at most 0.25 log points.

3.1.3 Event type 3: additional busses on direct route

The third type of event we study is when grid-cell pairs o and d that are already directly connected get additional busses due to a new route launch – because the new route overlaps with the existing route for the portion between o and d .

For BRT events, $Post_{odt}^{3B}$ is a dummy for the first event of an additional direct route launched between o and d taking place, in the ten months before week t . A positive coefficient α^{3N} captures the degree to which ridership or all trips between o and d increases after more non-BRT route traveling between these locations. It is often the case that non-BRT route travel along BRT corridors and stop in BRT stations for a portion of the route.

busses are added to the route between o and d due to an additional direct route.²¹ The sample is all odt such that BRT grid cells o and d are directly connected at time t , and o and d each have at least one treated observation.

For non-BRT events, $Post_{odt}^{3N}$ is defined analogously. The sample is all odt that are connected directly at time t , o is never a BRT grid cell, and o and d each have at least one treated observation.

3.1.4 Event type 4: additional busses on transfer route

The fourth type of event we study is a “transfer” version of event 3. It occurs when when grid-cell pairs o and d that are already connected by transfer(s) get additional busses at the origin o due to a new route launch – because the new route creates new transfer options starting from o . Specifically, the bus arrival rate at the origin must go up by at least 0.25 log points ($\approx 28\%$).

For BRT events, $Post_{odt}^{4B}$ is a dummy for the first event of an additional route passing through o that increases the bus arrival rate among all transfer options between o and d by at least 0.25 log points, in the ten months before week t . A positive coefficient α^{4N} captures the degree to which ridership or all trips between o and d increases after more busses are added to transfer connections between o and d due to an additional route. The sample is all odt such that BRT grid cells o and d are connected by transfer at time t , o and d are never directly connected, and o and d each have at least one treated observation.

For non-BRT events, $Post_{odt}^{4N}$ is defined analogously. The sample is all odt that are connected by transfer at time t , o is never a BRT grid cell, o and d are never directly connected, and o and d each have at least one treated observation.

3.2 “First stage” impacts on travel and wait times

We begin by examining how the events we study — new direct lines between o and d , faster transfers between o and d , and additional busses being added to a given segment due to an incidental route launch – affect the attractiveness of riding the TransJakarta network from o to d .

To do so, we estimate equation (1) with two proxy variables that capture different dimensions of the attractiveness of riding TransJakarta – log travel times and log number of busses per hour. The results for BRT and non-BRT network expansion events are presented in Tables 1 and 2.

We first examine how the log minimum travel time from o to d responds to BRT network

²¹As for event type 1, the route itself can be non-BRT as long as it passes through BRT grid cells o and d .

expansions. This captures the pure travel time on the bus – i.e. it does not include any time spent waiting for the bus to arrive, or if a connection to a different bus route is taken, the time spent waiting for the connecting bus. The ‘minimum’ refers to the shortest (i.e. least time on bus) route between o and d . The results show that adding a new direct line between o and d reduces travel time by about 0.13 log points , on average (column 1). Not surprisingly, the second type of BRT event – transfer travel time reduction events that are “large,” with a reduction of at least 0.25 log points – leads to a large average decrease in travel times of 0.54 log points ($\approx 72\%$). While these effects are, of course, somewhat mechanical by the way we define the events, the point here is that the differences are quite large relative to the standard errors, so we have large enough differences that we can pick this up in the data. Adding an additional direct route has a very small effect on travel times (0.03 log points). This is to be expected, as the vast majority of new direct routes between o and d overlap precisely with the existing direct routes and hence there would be no change in travel time. The impact of the fourt type of event on travel time is also relatively small, a 0.05 log points increase.

The second outcome variable we examine is the log number of busses arriving per hour over all direct or transfer options that connect the origin and destination, which captures waiting times. Table 1 shows that the third type of BRT event increases the bus arrival rate by 0.37 log points. Note that two other types of BRT events we consider – faster transfer connections and more frequent transfer connections – also increase the frequency of busses and reduce waiting times. The travel time transfer event (event 2) increases the bus arrival rate by 0.47 log points, while the wait time transfer event (event 4) increases the bus arrival rate by 0.5 log points.

Both “first stage” variables are proxies for the attributes of interest. For example, the second outcome counts bus arrivals even for transfer options between o and d that have very long travel time, which in practice will be valued less by commuters. Similarly, for the first outcome, the minimum travel time between o and d might depend on a very infreency bus line, which is not captured in this definition. The model that we set up in section 4 takes into account how commuters value both these attributes for all bus options in their choice set.

The key point from Table 1, however, is that these different events affect the desirability of using the bus system in different ways – a new direct line reduces travel time *and* reduces waiting times, in addition to eliminating the need to transfer, whereas additional busses essentially *only* reduce waiting times. Faster transfer events lead to substantially larger reductions in travel times and also affect wait times. The fact that these different events have have different types of effects on the desirability of using the TransJakarta system means

that these different events, together, will allow us to identify different preference parameters in Section 4 below.

Results for non-BRT network expansion are broadly similar and typically larger in magnitude (Table 2).

3.3 Reduced form ridership effects

3.3.1 New routes (event types 1 and 2)

The impacts of these events on bus ridership, estimated using equation (1) are presented in column 3 in Table 1 for BRT and in Table 2 for non-BRT. Event-study versions of each of these equations, showing monthly lags and leads from the date of the event, are presented in Figure 2.

Table 1 shows that ridership is highly responsive to improvements in service quality. First, on average, adding an additional direct BRT route between o and d , which column 1 showed led to a 13 percent reduction in travel time, leads to an increase in ridership of 0.20 log points over the 10 months following the introduction of the new route.

The event-study version of equation (1), shown in Figure 2, Panel (a), left column, shows no pre-trends before the event, and a discrete uptick in ridership at the time of the new route launch. The event-study version also shows that riders take some period to adjust to the new route.

Table 1 also shows that introducing a faster transfer route between o and d (event type 2) increases ridership by 0.05 log points on average over the ten months after the event. This is the impact for new transfer routes with a large time reduction (defined as greater than 0.25 log point reduction in time, which column 1 showed a 0.54 log point reduction in time on average. Figure 2, Panel (b), left column, show that this effect appears right after the new route launches.

Table 2 shows the analogous results for non-BRT routes. The results show even larger ridership impacts for all event types, especially from introducing a first direct line in this context.

Non-BRT Results. The non-BRT results, shown on the right side of Figure 2 and in Table 2, paint a qualitatively similar picture of the impact of bus service improvements. However, the impact of the non-BRT events is most frequently larger than for BRT. This is true both for the “first-stage” results in the first two columns of the table and for the impact on bus ridership.

Robustness. To assess the robustness of these results, we conduct two main exercises. First, we use a smaller, more precise, geography to define ridership flows and changes in the TransJakarta network. In Figure A.1 and Tables A.2 and A.3, we use 500-meter by 500-meter square grids and find very similar results across the board.

Second, we use an alternate method to account for zeros in the outcome variable. In our benchmark specifications discussed above, we estimate PPML regressions, using the level of bus ridership as the outcome. In this robustness exercise, we estimate OLS regressions using the inverse hyperbolic sine (asinh) of bus ridership as the outcome variable. We show the results in Figure A.2 and in Tables A.4 and A.5. Using the inverse hyperbolic sine transformation we obtain event study graphs that are broadly similar, and point estimates in a similar range to our benchmark results using PPML. For event type 1, we observe a slight pre-trend in the event study graphs. Nonetheless, there is a clear increase in ridership right after the event of the launch of the direct route. The lack of pre-trends in Figure 2 for our PPML specification suggests that the proportional model the the PPML specification estimates is a better fit for the trends that are captured in our time fixed effects (origin-week and destination-week).

3.3.2 Wait times and the ‘Mohring effect’

We next investigate the addition of busses to a given route from o to d caused by the launch of a new route that happens to intersect that route. Column (3) of Table 1 shows the results. We find that adding additional busses leads to a 0.12 point increase in ridership. Figure 2, Panel (c) shows this in event study form and shows that there is a large, discrete uptick in ridership exactly when the new route is introduced.

Since the first two columns in Table 1 showed that adding additional busses to a given route affects the attractiveness of travel from o to d almost entirely through its effect on wait times (i.e. it doesn’t change whether the trip from o to d is direct or not, and has only a 3 percent appreciable effect on travel times), we can interpret event 3 as being about the effect of wait times on ridership. Combining the estimates from the last two columns of Table 1, we get an elasticity of ridership with respect to wait times of -0.15, i.e. a 10 percent decrease in wait times leads to a 1.5 percent increase in ridership.

These estimates speak to the so-called ‘Mohring effect’ (Mohring, 1972). Mohring argued that, if demand for public transit is responsive to wait times, then there is an externality from riding the bus – more bus ridership allows the bus operator to add more busses to the route, decreasing wait times (and hence improving utility) for other riders. Our estimates show that, indeed, ridership is sensitive to bus frequency, suggesting that this effect operates in this case, and that the optimal public transit subsidy is likely positive for this reason.

The extreme form of the Mohring effect is when the elasticity of ridership with respect to wait times is greater than 1 in absolute value. In this case, the planner may have multiple local optimal levels of service frequency in ridership – a low ridership, high wait time regime and a high ridership, low wait time regime. Our estimate for BRT routes is that the elasticity is -0.15, and we can statistically reject that it is equal to -1 – so while we do find substantial evidence of feedback from wait times to ridership, we can reject multiple local optima in this setting.

For non-BRT routes, however, we find a much larger elasticity. Table 1 shows that adding additional busses increases non-BRT ridership by 0.3 log points. Combined with the “first stage” estimates in column 2, which shows that this was associated with an increase of 0.42 log points in busses per hour, suggests an elasticity of 0.72 – i.e., much closer to 1 and not statistically significantly different from 1. This suggests that adding more non-BRT frequency on some non-BRT routes could increase ridership enough to actually increase *average* ridership per bus.

The difference between the two elasticities likely reflects the fact that baseline bus frequency is much, much lower on non-BRT routes – prior to a new route, the median non-BRT event had busses coming 7 times per hour (i.e. once every 9 minutes), compared to 23 times per hour (i.e. once every 3 minutes) for the median BRT route. Given the very high baseline bus frequency on BRT routes even before the new busses were added, it is not surprising that ridership is somewhat less elastic there.

3.3.3 Bus ridership responsiveness to transfer route improvements

A striking finding from Tables 1 and 2 is that event 2 has larger impacts on travel time and bus arrival rate compared to events 1 and 3, yet it has smaller impacts on ridership. This suggests that commuters in Jakarta are less sensitive to attributes of bus options that involve *transfers*.

To further explore this point, event type 4 looks at a different type of improvement for transfer options. While for event type 2 both the travel time and wait time improve significantly, in event type 4, it is mostly the bus arrival rate that improves (0.5 log points for BRT and 0.32 log points for non-BRT). We find no significant increase in ridership for these two events, which supports the idea that commuters are less responsive to transfer option improvements. (We do see a small positive impact for non-BRT when using the finer geography of 500-meter square grids, see the last row in Table A.3 and the bottom-right panel in Figure A.1.)

Motivated by these results, in the model we will incorporate heterogeneity in the utility

cost of taking a transfer, which will help rationalize these empirical findings.²²

Summary The results in this section suggest that riders are responsive to the three dimensions of service quality we consider – wait times, ride times, and direct connections. Since each of the particular events we study affects multiple dimensions of these, in Section 4 we empirically estimate the model, which allows us to jointly infer the underlying preference parameters that best match the responsiveness we observe in the data.

3.4 Reduced form effects on overall trips (smartphone location data)

We then turn to the Veraset smartphone-based trip data to examine whether *total* trips – regardless of whether the passenger takes TransJakarta or not – are responsive to improvements in the TransJakarta network. To do so, we re-estimate equation (1) with the total number of trips from o to d from Veraset as the dependent variable of interest. Since our Veraset data only begins in March 2018 (as opposed to 2016, when our TransJakarta ridership data begins), for comparability we also re-estimate the effects on TransJakarta ridership for the same time period.

The results for BRT and non-BRT routes are presented in Table 3 and shown in event-study form in Figure 3.

The key result is that we do not see positive and significant impacts of the three types of events on aggregate travel volumes between pairs of 1km hexagon locations. For example, for BRT event type 1, we find a coefficient of 0.008 with a standard error of 0.048. This allows us to reject at the 95% level a positive impact of approximately +0.11, compared to the precise 0.20 effect on bus ridership on the same sample (column 1). Thus, while these effects are somewhat noisy, we can typically rule out moderately positive impacts on all smartphone location trips. Figure 3 shows no clear patterns before and after the events. We find qualitatively similar results for all

These results suggest that our main results on bus ridership reflect an immediate and large mode substitution towards TransJakarta, without an increase in total trips between the affected origin-destination pairs. Of course, we cannot rule out that over a longer time

²²An alternate possible explanation is that commuters pay less attention to transfer options and are thus not aware of such improvements. However, such a model does not necessarily imply lower ridership impacts in logs, as we find here. To see this, imagine that only a share λ of commuters pay attention to transfer options. Then, the level of bus ridership for transfer options will be increasing in λ , but changes in log ridership do not depend on λ . In the model, we have also experimented with a Calvo-(1983)-style updating assumption and found that such a model has a difficult time fitting the low responsiveness to transfer option events.

period compared to the 10 months we focus on here, the pattern of trips will also change. However, based on these results, in our model we will focus on mode choice and hold (origin and) destination trip choices fixed.

Summary Overall, we find that bus ridership responds strongly and quickly to the different types of service quality improvements that we study. We do not find evidence that these changes affect the aggregate volume of trips. The next section sets up a model that will allow us to interpret these findings and parse out the relative strength of commuter preferences for various transit attributes.

4 Model and estimation

We now set up a model of demand for public transportation that describes how commuters choose bus routes within the TransJakarta network, and at a higher level, whether they use the TransJakarta network or an “outside option” that captures private transport modes.²³

How do commuters decide which bus routes to take? The core of our model is a new formulation for how commuters choose routes in a public transport network. Our model highlights the importance of wait times, which affect decisions in two keys ways. First, overlapping routes between an origin and a destination effectively decrease wait times, because the traveller can take the first bus that arrives among these bus routes. (This is exactly the mechanism that we study empirically in the third type of event.) Second, travellers sometimes forego short wait times for other route characteristics. For example, a traveller may decide to not take the first bus that arrives in the station if this option involves a transfer, and instead wait longer for a direct (and shorter) route.

In the transit network routing model that we set up, random wait times play a central role. In the model, a traveler’s utility from a given option is additive in wait times and several deterministic factors: walking time, travel time, and the necessity to make a transfer. For each route and station, bus arrivals follow a Poisson arrival process. Commuters draw a vector of random wait times for the different routes in their choice set. This wait time randomness induces idiosyncratic heterogeneity in route choices. Our model has four desirable properties. First, the model is invariant to aggregation of identical routes. By properties of Poisson processes, “relabeling” a route with $K = K_1 + K_2$ busses as two distinct routes (still identical in terms of trajectory) with K_1 and K_2 busses, respectively, has no effect on

²³While it is possible to further embed these decisions into a model of destination choice, or a general equilibrium urban equilibrium model (Tsivanidis, 2022), given our null results on aggregate trips in this context, we hold these decisions fixed in the model.

the bus arrival process, and hence does not affect expected utility and choice probabilities. This stands in contrast to discrete choice models that use extreme value distributions over modes (McFadden, 1974), which would predict that the combined mode share of the two routes would be higher than that of the single route. Second, we can measure in our data the distribution of wait time, the source of idiosyncratic heterogeneity. We confirm using our GPS data that the distribution of wait times is approximately exponential, as implied in our model. Third, our model is rich enough to capture both phenomena explained above: that overlapping routes lead to shorter wait times, and that commuters sometimes choose to wait for a later bus arrival in order to use a better route. Fourth, despite the new and flexible formulation, we obtain tractable expressions for expected utility and choice probabilities.

We begin in Section 4.1 by introducing the bus network routing model with random wait times, and derive its key properties. We then add the higher nest: a mixed logit choice between direct routes in the bus network, transfer routes, and a private mode (section 4.2). At this time, we introduce heterogeneity in the utility cost of taking a transfer route. We discuss our estimation strategy and estimation results in sections 4.3 and 4.4.

4.1 Bus route choice

We now set up a static discrete choice model over bus options, where idiosyncratic heterogeneity is given by random wait times. In section 4.1.3 we define the Poisson process that determines route wait times, and in section 4.1.2 we derive expressions for choice probabilities, expected utility, and show that the model is invariant to route aggregation.

4.1.1 Model Setup

Consider a commuter i traveling from a given origin grid cell o to a destination grid cell d . They have a choice set with a finite number of bus options $k \in M_{odt}(N_{TJ}(t))$, where the choice set depends on the TransJakarta network $N_{TJ}(t)$ at calendar date t , $2016 \leq t \leq 2020$. Each option has a utility level $u_k = v_k - \alpha_W T_k^{\text{wait}}$ where v_k is the deterministic component that depends on walk time, travel time, and transfer terms (if the bus option k involves a transfer), and T_k^{wait} is a random wait time. We describe how these wait times are determined in the next section. The commuter selects the option k with the highest utility level u_k . Because of random wait times, this model leads to choice probabilities π_k of choosing option k , and expected utility $\mathbb{E}u_k$, with the expectation taken before observing wait times for all options k .

The bus route choice set M_{odt} can contain direct and single-transfer bus connections.^{24,25} The bus connections must start in the origin grid cell o and end in the destination grid cell d .²⁶ Preference parameters other than the value of travel time are allowed to differ by BRT/non-BRT. For clarity, we do not track this distinction in notation now and discuss it at the end of this section.

The utility for a *direct* public transit option k is:

$$u_k = \underbrace{-\alpha_T T_k^{\text{time}}}_{v_k} - \alpha_W T_k^{\text{wait}} \quad (2)$$

where T_k^{time} and T_k^{wait} are travel time and wait time, respectively.

If passengers decide to take a *transfer* public transit option $k \in M_{odt}$ that features a connection, we assume that they get the utility from the first leg k_1 of the route and the expected utility from the second leg k_2 of the route (i.e. taking expectations over wait times).²⁷ This is given by

$$u_k = \underbrace{-\alpha_T T_{k_1}^{\text{time}} + \mu_F + \mathbb{E}[u_{k_2}]}_{v_k} - \alpha_W T_{k_1}^{\text{wait}}, \quad (3)$$

where

$$u_{k_2} = -\alpha_T T_{k_2}^{\text{time}} - \alpha_W T_{k_2}^{\text{wait}} \quad (4)$$

and where μ_F is a fixed transfer utility shifter. Note that μ_F captures the utility for taking a transfer *above and beyond* the travel time and the (expected) wait time for the second leg, which are already included in $\mathbb{E}[u_{k_2}]$.²⁸ For example, if commuters dislike transfers above and beyond the additional travel time and wait time they involve, we expect $\mu_F < 0$.

The attributes we consider (travel time, wait time, the necessity of transfers) depend on

²⁴We only consider bus options where the transfer occurs in a BRT station. In the real TransJakarta network, transfers in BRT stations are free, while commuters who transfer in non-BRT stations must pay again when they board the second leg bus. Overall, total TransJakarta ridership R_{od} where o is a non-BRT station and o and d are only connected by transfer options in non-BRT intermediate stations m accounts for 0.3% of all ridership. By comparison, total ridership where o is a non-BRT station and o and d are only connected by any kind of transfer, accounts for 6.8% of all ridership.

²⁵The model we describe in this section can accommodate choice sets that contain both direct and transfer options. However, when we introduce the mixed logit model in section 4.2 we will assume that commuters optimize separately over direct routes and over transfer routes. This is motivated by computational considerations.

²⁶In estimation, we use hexagonal grid cells 1 kilometer apart. Thus, agents can walk within the grid cell but not to neighboring grid cells.

²⁷Transfer options k where either leg's route already directly connects the origin and destination grids are not allowed.

²⁸We assume that second leg wait times are not known at the time when the commuter makes the initial decision, so the consumer needs to take expectations.

TransJakarta's network and change significantly over the study period due to new routes that launch and change choice sets for various o, d pairs. The corresponding preference parameters will influence the planners' optimal network problem. For example, the transfer penalty μ_F influences the benefits of an interconnected network versus many direct connections. Similarly, the value of wait time influences the returns to bus-allocation intensity on a given route. We will explore how the optimal network varies with respect to these parameters in section 5 below.

BRT and non-BRT parameters. Waiting in non-BRT stations and transfers between non-BRT and BRT bus lines, may differ compared to the costs of using the BRT network. For example, the non-BRT bus stations are on the side of the road and differ significantly from the BRT stations, which are covered and where commuters pay at turnstiles to enter. We allow the parameters α_W and μ_F to differ for BRT and non-BRT routes. However, we impose that travel time on the bus is valued similarly, given that non-BRT and BRT buses themselves are similar.

4.1.2 Poisson bus arrival process and exponential wait times

Before making their decision, the commuter draws a vector of wait times for all the bus options in their choice set. Wait times correspond to the route r (the first leg route in the case of transfers), so if two bus options begin with the same route r , they will have exactly the same wait time.

Busses on a route r arrive at the origin grid o according to a Poisson count process with arrival rate λ_r . The arrival rate is given by $\lambda_r = N_r/RTT_r$, the ratio between the number of busses allocated to that route, and the return travel time needed for one bus to make a full loop on route r .

The wait time at o for the next bus from a given route r is exponentially distributed, namely $\Pr(w_r > w) = \exp(-\lambda_r w)$.²⁹ We assume that the Poisson processes for different routes are independent.

How are wait times distributed in reality? We use the universe of GPS bus data that we process into station bus arrival times to analyze the distribution of wait times at the level of entire routes. Appendix Figure A.9 shows the distribution of wait time for a specific route with high GPS data coverage (1E). If busses were equally spaced, m minutes apart, the wait

²⁹In other words, agents in the model are “non-planning” in that they cannot synchronize their trip departure time with a specific bus given an arrival schedule. This is realistic for TransJakarta, which does not publish a schedule, and where busses are relatively frequent (as described above, the median BRT route has a bus arriving on average once every 3 minutes, and the median non-BRT route has a bus arriving once every 9 minutes).

time distribution would be uniform between 0 and m . Instead, the wait time distribution closely tracks an exponential for most of the support of the distribution. (For large wait times, the distribution has less support than implied by the exponential distribution.) Results are similar for other routes. Appendix Figure A.10 also shows that BRT and non-BRT have similar wait time variance controlling for a route's mean wait time.³⁰

To determine the consumer's decision, we nest the decision-making problem as follows. At the second step, conditional on deciding to take TransJakarta direct routes, or TransJakarta transfer routes, the potential rider goes to the station in her origin grid o , and observes wait times T_k^{wait} for all possible routes for options k in her choice set. For two-leg routes, she observes the wait time for the first leg, and forms expectations about the wait time for the second leg. She then chooses the bus route that maximizes her utility.³¹

4.1.3 Properties of the bus route choice model

In this section, we characterize dominated options, prove that the model is invariant to aggregating identical routes, and characterize the general solution to the commuter's discrete choice problem.

The first result shows that certain options are strictly dominated and hence the commuter never chooses them.

Proposition 1. *If the choice set contains two options $k = 1, 2$ that begin with the same route $r_1 = r_2$ and that satisfy $v_1 < v_2$, then option 1 is dominated and hence $\pi_1 = 0$.*

Proof. The options begin with the same route so they face precisely the same wait time $T_1^{\text{wait}} = T_2^{\text{wait}}$, so $u_1 - u_2 = v_1 - v_2 < 0$. \square

If a commuter has multiple routes that overlap perfectly for the commuter's trip (and hence have identical deterministic utility), the commuter only cares about the total arrival rate. Note, other discrete choice models do not have this property, e.g. a multinomial logit model over direct bus routes.

Proposition 2. *Assume that the choice set contains has two options $k = 1, 2$ with $v_1 = v_2$ and arrival rates λ_1 and λ_2 , the two options start with different routes $r_1 \neq r_2$, and no other options in the choice set begin with r_1 or r_2 . Then, the model is isomorphic to a model where options 1 and 2 are replaced by an option 3 with $v_3 = v_1 = v_2$ and $\lambda_3 = \lambda_1 + \lambda_2$.*

³⁰The exponential distribution has a coefficient of variation equal to 1, so the graphs also show that routes with longer mean wait times are generally more regular.

³¹This is a reasonable modeling assumption because TransJakarta posts actual arrival times at each station – see Figure A.8 for an example. Users could also look up these times from home using an app, but this may change between the time the user leaves home and arrives at the station.

Proof. This follows directly from the fact that the sum of two independent Poisson count processes with arrival rates λ_1 and λ_2 is a Poisson count process with arrival rate $\lambda_1 + \lambda_2$. \square

As a corollary, the demand model is unchanged if two entire identical routes r_1, r_2 with N_1, N_2 busses are replaced by a single route r (identical to the two) with $N_1 + N_2$ busses, so this model does not feature the ‘red bus, blue bus’ paradox discussed by McFadden in the context of the multinomial choice model. This is an attractive invariance property for the planner’s optimization problem that we will later set up in section 5.

Finally, we characterize the general solution.

Proposition 3. *Assume that all options in the choice set are ranked such that $v_1 < v_2 < \dots < v_N$, each option begins with a different route, and the wait time for option k is independently exponentially distributed with parameter λ_k .*

Part 1. *The probability π_k to choose option k satisfies:*

$$\pi_k \frac{\alpha_W}{\lambda_k} = - \left(\sum_{i=1}^{k-1} e^{S_i} \frac{\lambda_i}{\Lambda_i \Lambda_{i+1}} \right) + e^{S_k} \frac{1}{\Lambda_k},$$

where $\Lambda_i = \sum_{\ell=i}^N \lambda_\ell$, $S_N = 0$, and $S_{k-1} = (v_{k-1} - v_k)\Lambda_k + S_k$ for all $1 < k \leq N$.

Part 2. *Expected utility is given by*

$$\mathbb{E}[u_k] = v_N - \pi_N \alpha_W / \lambda_N,$$

where π_N is the choice probability of the option with highest deterministic utility.

The proof is in Appendix A.4. It uses basic algebraic manipulations of the exponential distribution.

Part 1 ensures that computing choice probabilities is computationally tractable. It involves sorting the deterministic components v_k . Computing all π_k has linear complexity in the number of options.³²

Part 2 shows that expected utility has a particularly simple expression. Note that if the commuter only had the option N in their choice set, expected utility would be $v_N - \alpha_W / \lambda_N$. When other options with $v_k < v_N$ exist, their influence on expected utility is summarized by one number, the probability π_N to choose the top option. Consider the two extreme cases.

³²When solving the model in practice, we also consider all options that begin with the same route, drop dominated options, and split probabilities equally among options that begin with the same route and have identical v_k . The latter case arises, for example, when the commuter can transfer in two possible intermediate stations, with identical total travel time, walk time, and expected wait time in the two intermediate stations.

When $\pi_N \approx 1$, the other options are rarely chosen (presumably because v_{N-1} is small), and expected utility is close to the case when only N is available. When $\pi_N \approx 0$, the commuter has a large number of options that closely rival N . This means that many v_k are very close to V_N , and given independent wait time draws, with high probability the commuter obtains a high v_k (close to v_N) and very small wait time.

4.2 The mixed logit choice between direct bus routes, transfer bus routes, and a private option

A commuter traveling between o and d first decides between using the direct or transfer components of the TransJakarta network or using an outside option, a catch-all for private modes (private motorcycle, for hire motorcycles, car, other private minibusses, etc.). This choice takes place before they observe wait times in the TransJakarta network.

The utilities for the three options are:

$$\begin{aligned} u_{it}^{\text{bus direct}} &= \left(E \max_{k \in M_{odt}^{\text{Direct}}} u_k \right) + \epsilon_{it}^{\text{bus direct}} \\ u_{it}^{\text{bus transfer}} &= \left(E \max_{k \in M_{odt}^{\text{Transfer}}} u_k \right) + \mu_i + \epsilon_{it}^{\text{bus transfer}} \\ u_{it}^{\text{private}} &= \zeta_{od}^{\text{private}} + \epsilon_{it}^{\text{private}} \end{aligned}$$

where the terms in large brackets capture expected utility over different realizations of wait time vectors, for all direct route options, and all transfer route options, respectively.

The term μ_i captures the random transfer shifter for individual i . This is drawn iid from a normal distribution $\mathcal{N}(\mu, \sigma_\mu)$ whose parameters we will estimate.

The term $\zeta_{od}^{\text{private}}$ captures all time-invariant factors that make the private option more attractive for that specific origin-destination pair. We will estimate each of these fixed effects in order for the model to match the average bus ridership between o and d .

The term $\epsilon_{it}^{\text{bus direct}}$, $\epsilon_{it}^{\text{bus transfer}}$, and $\epsilon_{it}^{\text{private}}$ are Gumbel-distributed error terms giving rise to multinomial logit probabilities (conditional on μ_i).³³

It is possible to further embed these decisions into a higher nest where commuters decide which destinations to travel to (holding their residences fixed), or in a urban general equi-

³³Many origin destination pairs only have direct options or only have transfer options, so the three-option multinomial logit collapses to a binary choice. In principle, it is possible to include both direct and transfer options in the inner nest in general, leading to a simple binary logit model between the bus network and the private option. In the current paper draft, we are separating direct and transfer options for computational reasons. Specifically, currently we only need to compute the expected utility terms once and then integrate the logit probabilities over the distribution of μ_i . When modeling choice over both direct and transfer options, we would need to recompute the expected utility for every value of μ_i .

librium model where agents choose their home and work locations based on expected travel costs given by the above expression. Given the null result we found on aggregate travel measured using smartphone trips in the previous section, we hold these margins of behavior fixed.

4.3 Model Estimation

We estimate the model by classical minimum distance (CMD), finding the parameter vector θ that best matches the reduced form results from Section 3. We estimate the seven preference parameters $\theta = (\alpha_T, \alpha_W^{\text{BRT}}, \mu^{\text{BRT}}, \sigma_\mu^{\text{BRT}}, \alpha_W^{\text{non-BRT}}, \mu^{\text{non-BRT}}, \sigma_\mu^{\text{non-BRT}},)$. That is, we estimate one cost of travel time and separate BRT and non-BRT costs of waiting, and transfer shifter paramters (the mean and standard deviation of the normal distribution of transfer shifters). We normalize the logit parameter $\beta = \frac{\bar{D}_{od}}{D_{od}}$ where D_{od} is the straight line distance between grids o and d , and $\bar{D}_{od} = 8.5\text{km}$ is the average distance.³⁴

We also estimate the origin-destination private option attractiveness terms ζ_{od} . For each value of θ , and for each (o, d) pair, we find ζ_{od} such that model average bus ridership between o and d over the entire time period matches the empirical counterpart.

Model-predicted ridership. Given a vector of preference parameters and the state of the TransJakarta network at time t , we compute model-predicted ridership combining the smartphone commuting flow data V_{od} and the model choice probabilities

$$\pi_{odt} = \int \frac{\exp(\beta v_{iodt}^{\text{bus direct}}) + \exp(\beta v_{iodt}^{\text{bus transfer}})}{\exp(\beta v_{iodt}^{\text{bus direct}}) + \exp(\beta v_{iodt}^{\text{bus transfer}}) + \exp(\beta v_{iodt}^{\text{private}})} dF(\mu_{it}),$$

which depend on the bus route choice model through the expected utility terms $v_{iodt}^{\text{bus direct}}$ and $v_{iodt}^{\text{bus transfer}}$.

To compute the choice probabilities, we use our data on route launch dates to construct the state of the TransJakarta network at every time period t . To determine wait times, we need to decide how many buses are allocated to each route. We compute the average bus allocation over the entire time period and use it in the model for all t .

Running the reduced form analysis on model-predicted ridership. For a given parameter vector θ , we compute model-predicted ridership for all origin-destination pairs and all time periods. As described above, we first estimate the ζ_{od} terms that match average

³⁴The travel time component of utility is generally increasing in straight line distance. Hence, a constant β would imply “sharper” decisions for more distant (o, d) pairs. The normalization we use compensates for this mechanical effect. Estimation results are similar if we assume $\beta = 1$.

bus ridership over time between o and d . Using the 1km hexagonal grids, we compute ridership for 33,880 o, d pairs (covering a maximum number of 350,493 bus route options), over the different versions of the TransJakarta network, covering 220 weeks between January 2016 and March 2020.

We then run the reduced form analysis given by equation (1) using log model-predicted ridership as outcome variable, for the four types of events and for BRT and non-BRT. As model-predicted is almost always positive, we estimate this equation using OLS. During the estimation of θ , we run exactly the same regressions, changing only the outcome variable. We obtain a significant computational speed improvement by pre-computing a vector x^E for each specification E , such that the coefficient of interest α^E is given by $\alpha^E = (x^E)'Y_{odt}$ for any outcome vector Y_{odt} .³⁵

Event study moments. We match the following nine moments. First, we match the eighth “Post” coefficients α^E , $E \in \{1B, 2B, 3B, 4B, 1N, 2N, 3N, 4N\}$ for the four events for BRT and non-BRT from Tables 1 and 2. That is, we match the average impact on log model-predicted ridership from (a) the introduction of the first direct connection, (b) the introduction of a faster transfer route, and (c) the introduction of additional busses due to an additional direct route, and (d) the introduction of additional busses due to additional transfer routes.

Trip duration moment. We also include an additional moment that captures how the duration of TransJakarta bus trips between o and d depends on the travel time of bus options in the choice set between o and d . This moment helps pin down the value of travel time α_T .

To construct it, we first measure BRT trip duration in the data using the specific tap-in and tap-out times for the same smartcard at the BRT entry and exit station.³⁶ To do this, we restrict to BRT stations that are tap-out compliant, defined as when tap-out transactions are at least 30% of all taps at that station. Thirty-six percent of all stations (92 stations) are tap-out compliant according to this definition. Overall, tap-out transactions cover around 25% of all BRT trips.

Because commuters enter BRT stations through turnstiles, the trip duration computed using this data also includes the wait time for the bus. We compute the same object in our model.

³⁵The vector x^E is a row of the OLS matrix $(X'X)^{-1}X'$ where X is the matrix of covariates, including all fixed effects. We compute this matrix inversion only once before estimation. To further simplify computation, we first select the fixed effect with the largest number of categories, and apply the Frisch–Waugh–Lovell theorem. We project the outcome Y_{odt} and all covariates and other fixed effects onto the chosen fixed effect.

³⁶for non-BRT trips, tap-out time is not recorded.

Our sample is o, d pairs and a time t such that o and d are connected by multiple transfer options at time t , and not connected directly.

We estimate the following regression:

$$\log(Dur_{iodt}) = \gamma MeanTime_{odt} + \beta MinTime_{odt} + \epsilon_{iodt},$$

where Dur_{iodt} is the duration for trip i from o to d at time (week) t , $MeanTime_{odt}$ and $MinTime_{odt}$ are the mean and minimum travel time (excluding wait time) between o and d at t , computed over all transfer options between o and d .

We leverage the variation in mean versus minimum travel time across o, d pairs in our sample. Our ninth moment matches the empirical estimate $\hat{\gamma} = 0.157$ (0.035). (In the data we also estimate $\beta = 0.47$ (0.027).)

Intuitively, the weight on *mean* travel time (holding minimum travel time fixed) is decreasing in α_T , the value of travel time. As commuters value travel time less, they are open to taking transfer options that have longer travel time, and this leads to longer measured trip duration.

Estimation. We use classical minimum distance to match the vector of nine numbers $m = (\alpha^{1B}, \alpha^{2B}, \alpha^{3B}, \alpha^{4B}, \alpha^{1N}, \alpha^{2N}, \alpha^{3N}, \alpha^{4N}, \gamma)$. Specifically, we find the parameter vector θ that minimizes the objective function

$$\min_{\theta} (m(\theta) - m)'W(m(\theta) - m), \quad (5)$$

where $m(\theta)$ is the vector of model moments, and $W = \Omega^{-1}$ is the optimal weighting matrix given by the inverse variance-covariance matrix of the moments m . To compute Ω , we “stack” all reduced form analysis and jointly estimate all regression coefficients in a seemingly unrelated regression framework. We cluster standard errors two-way by origin grid and by destination grid, which introduces dependence between the different regressions.

To reduce the risk of finding a local minimizer θ of (5), we repeat the optimization routine starting from 100 randomly selected initial conditions.

Inference. To obtain confidence intervals for $\hat{\theta}$, we repeat estimation 100 times where we match the moment vector $m^k = m + \varepsilon^k, k = 1, \dots, 100$, where $\varepsilon^k \sim^{\text{iid}} \mathcal{N}(\mathbf{0}, \Omega)$ are i.i.d. draws from a multivariate normal distribution centered at zero with covariance matrix Ω . We use the resulting $\hat{\theta}^k$ estimates to construct confidence intervals for the elements of θ .

4.4 Estimation Results

Table 4 shows the estimation results for the seven estimated parameters in θ , along with 90 percent confidence intervals.³⁷

For BRT, people appear to view time spent waiting for the bus as more costly compared to travel time on the bus – the estimated parameter α_W^{BRT} is 0.227 compared to $\alpha_T^{\text{BRT}} = 0.037$. This wait time parameter governs the degree to which people are sensitive to frequent service. Note that in our model, this parameter also controls the importance of wait time heterogeneity. In other words, a high α_W means that commuters make more idiosyncratic choices, driven by randomly small wait times for some of the options in their choice set. For non-BRT, the wait time cost is also significantly higher than the time cost for non-BRT.

The distribution of the transfer shifter μ_F has a large negative mean μ , and a large dispersion σ_μ . This means that for most commuters, transfers are very costly, while for a small share of commuters, the shifter is actually positive. Note that μ_F captures costs for transferring busses *above and beyond* the expected utility cost from the second leg of the trip. A positive μ_F may be consistent with commuters that behave as if they underestimate the second leg trip cost. Overall, these transfer cost estimates help match the attenuated bus ridership response to improvements in transfer options that we document in events 2 and 4.

5 Optimal network design

The third step in our analysis is to characterize the planner’s problem of designing an optimal public transportation network. We focus on two key exercises. First, we compare the current network used by TransJakarta to the planner’s solution. Second, we study how the shape of the network solution changes when structural preference parameters change.

The problem of finding the optimal network is discrete and high-dimensional. This poses two challenges. First, our planner’s problem does not appear to have an analytical solution for the global welfare-maximizing network of routes, and it resembles NP-complete problems such as the traveling salesperson problem and the vehicle routing problem. The problem exhibits both substitution and complementarity forces; this precludes us from using formulations based on single-crossing properties from the literature (Jia, 2008; Arkolakis and Eckert, 2017).

Second, it is not clear how to define comparative statics with respect to continuous parameters because the space of networks is discrete. In other words, the global optimal

³⁷ Appendix Table A.7 shows that the nine moments computed using the estimated model are close to the empirical moments.

network is locally constant almost everywhere as a function of a continuous parameter.

We introduce a general framework for characterizing optimal allocation and performing comparative static exercises in discrete, high-dimensional problems. Our theory may be applied to other settings where a planner has discrete and high-dimensional instruments, for example matching, land-use regulation or infrastructure investment in a spatial or network contexts, etc.

The key idea is to smooth the planner’s problem using idiosyncratic shocks. Rather than focusing on the (generically unique) global welfare-maximizing network of routes, we assume that the planner cares both about model welfare as well as idiosyncratic shocks over every possible network, which are extreme value type-1 distributed. These shocks represent factors outside our demand model that the planner cares about.

This leads to a *distribution of optimal networks*, which is simply a multinomial logit distribution over networks. The key implication is that if a network N' has a level of welfare $W(N')$ that is close to the global optimum welfare $W(N^*)$, even small shocks may tip the planner into preferring N' over N^* , and hence the planner’s choice distribution will put non-trivial probability on N' . Of particular interest to us is that while $W(N')$ is close to $W(N^*)$, the shape of the network N' may be very different from N^* .

Adding idiosyncratic shocks to the planner’s problem allows us to solve the two challenges above. First, in this formulation, our objective is to sample from a multinomial logit distribution, rather than to find the global optimal network. We can use established algorithms, such as Metropolis-Hastings, parallel tempering, or a modified version of simulated annealing (SA), to asymptotically sample from the planner’s distribution of optimal networks. In practice, we will run many independent SA simulations to obtain an i.i.d. sample of optimal networks, which we use to characterize the shape of optimal networks.

Second, comparative statics have a natural definition in terms of an expectation over the distribution of optimal networks. We derive formulas for local comparative statics and show how they can be easily computed based on the sample of networks we obtain.

We then introduce our implementation of the simulated annealing algorithm, and the “modifier” functions we choose for our public transport network problem. The key question in practice when applying the theory we have set up is whether the algorithm “mixes,” in the sense of approaching the stationary distribution of the Markov chain that describes it. Our diagnostic checks are consistent with mixing: all parallel runs reach similar levels of welfare and ridership, despite starting from random initial conditions.

We document that the optimal networks we obtain offer large improvements in welfare relative to the current TransJakarta network, and lead to roughly double ridership levels. These networks are more expansive, less concentrated in the urban core of Jakarta.

We then illustrate comparative statics by investigating the impact of higher wait time cost α_W^{BRT} . Both using local and larger changes, we show that a larger wait time cost leads to more concentrated optimal networks.

5.1 Optimization Environment

The planner’s problem we consider is how to design a bus network N . This involves deciding the number of bus routes, where bus routes should run, and how to allocate a fixed number of busses to these routes. Formally, a network is a tuple $N = (K, (r_1, \dots, r_K), (b_1, \dots, b_K))$ where K is the number of routes, r_k indicate routes, and b_k is the number of busses on route r_k .

We divide the greater Jakarta area into 418 2km \times 2km square grid cells. Any grid cell can be connected by a bus route to either of its eight adjacent grids (north, south, east, west, plus four diagonals), leading to 1,536 possible adjacency links. (See section 2.2 for more details on the geographical environment.) A bus route is a sequence of distinct grid cells where every consecutive pair of cells are adjacent. (Self-intersecting routes, including loops, are not allowed.) We assume that busses travel in both directions along the same path.

We use our driving data (section 2.3.5) and the bus travel time data, to construct travel times for driving and bus travel time (separately for BRT and non-BRT) for every adjacency edge. Appendix A.5.1 explains the estimation procedure. We hold fixed the geography of bus rapid transit infrastructure. Only neighboring grid cells that are connected by BRT dedicated lanes in the current network have BRT edge travel times in the optimization model. In future versions of this paper, we plan to endogenize this dimension of the network design as well and optimize the placement of BRT infrastructure.

We consider both where bus lines should run to, how many bus lines there should be, and how many busses should run on each line. We take as constraints that the planner has 1,500 busses, approximately the number of busses that TransJakarta has during the end of the period we study, traveling full time, but they can be arbitrarily allocated across bus lines.

The travel demand model closely follows the model discussed in section 4. We assume that commuting flows between every (o, d) pair are fixed, computed using the smartphone location data (see section 2.3.2). A commuter’s bus network choice set is defined as in section 4.1.1 based on the proposed network N . Driving is the outside option.

The demand model allows us to compute a measure of commuter welfare $W(N; \theta)$ for any network N and vector of preference parameters θ . This is the average over all commuters in the city of the expected utility from the multinomial logit model between direct, transfer

and private options.

The space we are optimizing over is extremely large. Even the number of unique paths is exponential in the size of the grid, and a network consists of any combination of routes. The allocation of busses to lines adds even more combinatorial dimensions to the optimisation problem, making it infeasible to derive the global welfare-maximizing optimal network through exhaustive search over all possible networks.

5.2 Computing Optimal Allocations in High-dimensional Spaces

We now set up a general framework that we will then apply to the planner's network design problem that we just described.

5.2.1 Adding idiosyncratic factors to the planner's problem

Consider a planner who chooses from a finite set of allocations $N \in \mathcal{N}$. In a typical application, the set \mathcal{N} is high-dimensional and extremely large.

A planner chooses the allocation N that maximizes $W(N; \theta) + \epsilon_N$, where $W(N; \theta)$ is welfare according to a fully specified model that the researcher knows, and θ is a vector of structural parameters. ϵ_N is an idiosyncratic preference shock that the planner has for network N , capturing factors that the planner cares about that are not in the model that gives rise to $W(N; \theta)$. Assume that ϵ_N is i.i.d. from a Gumbel distribution with parameter β .³⁸

The probability that network N will be chosen by the planner is

$$\pi(N; \theta) = \frac{\exp(\beta W(N; \theta))}{\sum_{N' \in \mathcal{N}} \exp(\beta W(N'; \theta))}. \quad (6)$$

In a typical case where a researcher wants to solve for the planner's optimum, they are interested in the properties of *the unique* model's global optimum (or global optima if there are ties).

In the setting described here, the researcher is also interested in allocations that have welfare levels that are very close to the global optimum, because small shocks outside the model may tip the planner to preferring one of these. As $\beta \rightarrow \infty$, the distribution π converges to the distribution that puts equal mass on all allocations that maximize welfare.

The goal in this setting is to characterize the probability distribution π . To characterize optimal allocations, properties of optimal allocations can be computed as integrals over π

³⁸We can alternatively assume multiplicative Fréchet shocks. It seems desirable to also consider richer idiosyncratic shock structures, for example where ϵ_N depends on shocks that are defined over some model primitives.

(see section 5.2.2 below).

The key challenge is that in a typical application, the denominator in (6) is prohibitively expensive to calculate explicitly due to the large number of terms. However, it is easy to compute relative probabilities for any pair of networks. This is the key insight for a class of well-developed methods designed to sample from such distributions, such as the Metropolis-Hastings algorithm. We return to these algorithms in section 5.3.

5.2.2 Estimating properties of optimal allocations and comparative statics

We are interested to characterize certain properties of optimal allocations, and how these vary when structural model parameters change.

Consider a property defined by a function $f(N)$. For example, assume f counts the number of stations for a bus network N . Define

$$f^*(\theta) = \sum_{N \in \mathcal{N}} \pi(N; \theta) f(N)$$

We can also consider properties that depend both on the network N and on the demand model, through the parameters θ . For a function $g(N; \theta)$ such as, for example, the level of public transport ridership given N and θ , we define

$$g^*(\theta) = \sum_{N \in \mathcal{N}} \pi(N; \theta) g(N; \theta)$$

We can estimate these expression as long as we can sample from the π distribution. We introduce in the next section the algorithms that allow us to do this.

Comparative Statics. By a slight abuse of notation let θ be a scalar, and assume that $W(N; \theta)$ is differentiable in θ for all N . We can express the change in the property f due to a change in θ as

$$\frac{df^*}{d\theta} = \beta \sum_N \pi(N; \theta) W'(N; \theta) (f(N) - f^*(\theta)) \quad (7)$$

The comparative static for the function g that also directly depends on θ is:

$$\frac{dg^*}{d\theta} = \beta \sum_N \pi(N; \theta) \left(W'(N; \theta) (g(N) - g^*(\theta)) + g'(N; \theta) \right) \quad (8)$$

These expressions gives well-defined local comparative static of the properties given by f and g with respect to model parameter θ . As with f^* and g^* themselves, these expressions can be estimated using a sample of networks from the π distribution.

5.3 The Simulated Annealing Algorithm

Several algorithms have been designed to allow sampling from distributions such as (6). These include Metropolis-Hastings, parallel tempering, and simulated annealing. They all share a key underlying idea. The goal is to construct a Markov chain over networks (or allocations more generally) whose stationary distribution is given by (6). The Markov chain allows transitions between certain pairs of networks, and the transition probability depends on the ratio of probabilities in (6), which is easy to compute.

To sample from the π distribution, we use a modified simulated annealing (SA) algorithm. We define the algorithm and the modifier functions we use formally in Appendix A.5.2.

Intuitively, SA is a probabilistic algorithm based on the idea of controlled experimentation. It starts from a random network and keeps proposing new candidate networks, which are either accepted or rejected. In the initial stages, the algorithm is more likely to accept even candidates which are worse than the current network. As time progresses, the algorithm becomes more conservative and mostly accepts only networks which are better than the current state.

In the later part of the algorithm, it reliably identifies *local* optimal networks. The purpose of the initial phase is to allow the algorithm to explore the entire state space of networks, and escape basins of attraction of any particular local optimum.

The algorithm's tendency to accept new candidate networks that offer lower welfare is controlled by the algorithm's "temperature" at time t . The temperature falls exponentially from an initial value K_0 to a final and lower value K_T . Assuming that the algorithm is at a network N_t at time t , and it proposes candidate new network N_t^{NEW} , this new network will be accepted with the following (easy to compute) probability:

$$\Pr(\text{accept } N_t^{\text{NEW}}) = \begin{cases} 1 & \text{if } W(N_t^{\text{NEW}}) \geq W(N_t) \\ \exp\left(\frac{W(N_t) - W(N_t^{\text{NEW}})}{K_t}\right) & \text{otherwise.} \end{cases} \quad (9)$$

In words, at each step the algorithm proposes a new allocation. Allocations with higher welfare than the current value are always accepted. Allocations that lead to lower welfare are accepted with a probability that is decreasing in the welfare loss, and increasing in temperature. This means that initially in the algorithm, such transitions are more likely to be accepted, and the algorithm tends to travel nearly at random through the set of allocations. Later on during the algorithm, transitions are more heavily biased towards those that increase welfare or that do not decrease welfare very much.

How the algorithm proposes new networks is critical for its success. We propose a nested structure for suggesting new bus networks. First, we apply a random *global* modification to

the current network, such as deleting an entire bus route, adding an entirely new bus route, or re-allocating a large share of buses from one route to other randomly chosen routes. Afterwards, we apply a sequence of randomly chosen *local* modifications, which at each step are accepted only if they improve welfare. Examples of local modifications are adding or deleting one stop at the end of a route, local re-routing, and small changes in bus allocation between routes.

We can prove that our algorithm for suggesting new bus networks has the property that for any two networks N_1 and N_2 , the latter can be reached with positive probability starting from the former network.

5.3.1 Result: Simulated Annealing Identifies π

We now establish that a specific simulated annealing algorithm allows us to sample from the π distribution.

Proposition 4. *Assume the final temperature is set to $K_T = \beta^{-1}$ where β is the logit parameter from the planner’s problem. As the number of steps $T \rightarrow \infty$, the final state N_T of the simulated annealing algorithm converges in distribution to π , the planner’s distribution over allocations.*

This result holds fixed the ending temperature to match the (inverse) parameter of Gumbel shocks in the planner’s problem. The asymptotics are with respect to SA advancing very slowly.

Proof. (Sketch.) The proof uses standard Metropolis-Hastings algorithm arguments. It mirrors a classic proof that as the number of steps increases, the outcome of SA converges to the global optimum (Nikolaev and Jacobson, 2010).

The key observation is that replacing the variable temperature K_t with a fixed temperature $K_t = K = \beta^{-1}$ in the acceptance probability (13) yields the Metropolis-Hastings (MH) algorithm. The above property that any network can be reached from any other network with positive probability ensures that MH induces an irreducible stationary Markov chain. By the classic result for MH, its stationary distribution at state N is proportional to $\exp(K^{-1}W(N))$, thus it must be equal to π .

As the number of steps T grows, the “end” of the SA algorithm becomes close to the Metropolis-Hastings algorithm, in the sense that can be made precise. Hence, SA also converges to π as the number of steps T tends to infinity. \square

Note that the MH algorithm with $K = \beta^{-1}$ also asymptotically samples from the π distribution. The reason why SA is preferable in practice is that the initial high-temperature

period tends to help reduce the mixing time. By initially traversing the state space nearly at random, the SA algorithm is less likely to get stuck near a local optimum close to the initial network.

In practice, we obtain independent draws from π (asymptotically) using independent SA runs. While computationally expensive, the large number of iterations is similar to the use of “burn-in” and “thinning” when using MH to sample from a certain distribution.

5.3.2 Parameters and diagnostic tests

We assume that the planner’s shocks have a parameter $\beta = 1,000,000$, which implies that idiosyncratic shocks have a small variance and most of the planner’s decisions are driven by the model welfare function $W(N)$. We run the simulated annealing algorithm using a set of commuter preference parameters obtained from estimating the demand model in section 4.3. We have also estimated ζ_{od} terms, which capture the overall attractiveness of the private option between o and d , for certain o, d pairs that are connected in the current TransJakarta network. We predict $\hat{\zeta}_{od}$ by projecting them on a constant term and the driving time between o and d , which we have for all o, d pairs.³⁹

We report results from 10 parallel, independent runs of the simulated annealing algorithm. Each run produces one network, so we have a sample of 10 networks that we use to compute network characteristics and comparative statics.

Since we are sampling from a distribution of networks, we don’t expect the algorithm to result in the same network each time. However, given the logit parameter we choose, we expect these networks to be similar in terms of welfare. Figure 6a shows the evolution of the 10 SA runs, each with roughly 30,000 global changes and more than 1.3 million local changes. The SA networks systematically achieve higher welfare than the existing TJ network, and these values are relatively close to each other. Figure 6b repeats this exercise, but plotting the evolution of model-implied ridership across SA runs, again highlighting that resulting networks are very similar in this regard.

Figure A.5 in the appendix expands on this point by plotting the distribution of network statistics for resulting networks from all parallel SA runs. The final networks are very similar in how many locations they cover, how much ridership they achieve, and how many people they connect. However, they differ in their exact design, i.e. how much total network mileage they produce.

³⁹In this draft, all parameters used for the network optimisation exercise are identical to the estimated parameters in 4.3, except the mixed-logit transfer shifter. For computational reasons, we implement a slightly simpler version in which a bus transfer yields a fixed penalty term equivalent to half an hour of on-bus travel time. In future versions of the paper, we aim to harmonise this step between both sections.

5.4 Results

Current TransJakarta network. Figure 4 shows the actual TransJakarta bus network, superimposed on the grid structure (Panel 4a). Different lines are printed in different colors. Edges on which TransJakarta has built infrastructure for BRT travel are denoted by bright green underlay. Note that the network features a relatively dense network in the urban core of the city, with most of the 108 bus lines crossing through the downtown areas. Panel 4b removes the bus lines to reveal the underlying geography of Jakarta, including the central DKI area in light blue as well as the surrounding neighborhoods. Locations' sizes are printed corresponding to their relative outcommuting population weights residing in each grid cell. To improve legibility, Panel C aggregates all bus lines connecting a given edge to compute how often a bus from *any* line arrives and leaves in each direction. Again, while the core of the city seems well connected by frequent bus services, Jakarta's periphery has few busses passing through, and most locations of the city are not connected at all. In the remainder of this section, we try to shed light on whether this network structure is indeed optimal for the economic geography of Jakarta or if there are other networks which do substantially better.

Optimal Networks. Figure 5a shows one example of a network obtained by running the simulated annealing algorithm based on these simulations and the estimated preference parameters from Section 4. Comparing this network to the actual TransJakarta network shown in Figure 4a, it is apparent that the optimal network covers much more of the city, with many more lines than the actual network. Similarly, Figure 5b prints the bus density map analogous to figure 4c, which again reveals a much more spread out network. Many remote areas get connected to the network for the first time under this optimisation scenario, yet locations in the urban core have longer wait times.

To make these types of comparisons more directly, Table 5 presents a number of descriptive statistics about different networks. Column (1) starts by presenting descriptive statistics of the actual TransJakarta network (i.e. the network shown in Figure 4a), and Column (2) presents the same statistics calculated on the sample of optimal networks (achieved through the 10 parallel optimisation runs).

A few differences between the current network and optimal networks are readily apparent. First, the optimal network is much more expansive on average – it covers 88% of all locations with at least one bus route, compared to only 42% of all locations in the actual network. This is also true if we weight locations by population – in the optimal network, 99% of all people live in a network grid with a station, compared to 73% in the actual TransJakarta network. The optimal network also has more than triple the number of kilometers of network than the actual TransJakarta network. This implies that wait times for busses will be substantially

longer in the optimal network than on the actual network. On net, the optimal network is projected to achieve ridership that is more than double the actual TransJakarta network.

Comparative Statics. In Table 6, we re-run simulations to draw from the distribution of an optimal network, yet halving the wait time costs for BRT and non-BRT, α_W^{BRT} and α_W^{nonBRT} . Columns (1) and (2) of Table 6 replicate the first two columns of Table 5 and serve as the baseline estimate of various network characteristics. Column (3) reports our estimate of the derivative for the row statistic with regard to increasing the two waiting cost parameters (α_W^{BRT} and α_W^{nonBRT}), as derived in equations 7 and 8. We see that this derivative is positive for the number of locations with a station and number of pairs with a feasible connection, i.e. a social planner who believes commuters have a lower cost of waiting should construct a more extensive and even slower network with more stations. The change in the parameter and these network changes jointly lead to lower ridership. The effect of a large change in wait time costs, reported in column (4), are qualitatively similar.⁴⁰

References

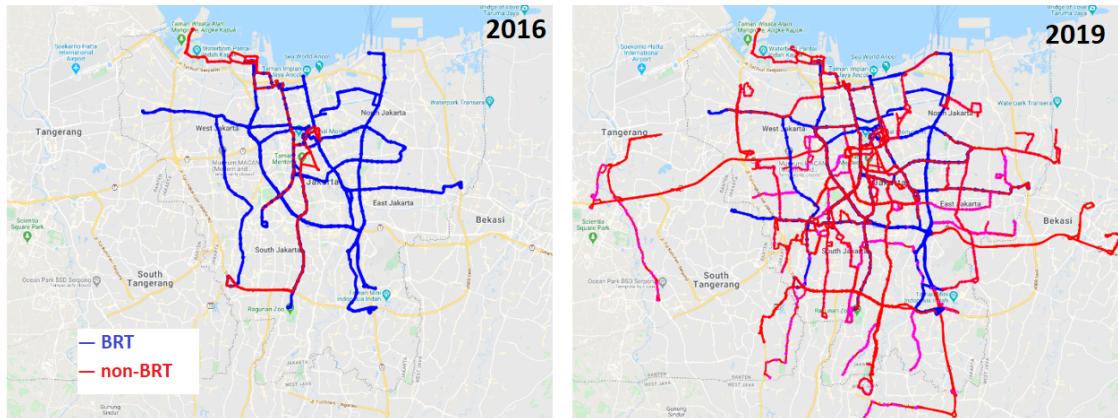
- ABRANTES, P. A. AND M. R. WARDMAN (2011): “Meta-analysis of UK values of travel time: An update,” *Transportation Research Part A: Policy and Practice*, 45, 1–17.
- ALLEN, T. AND C. ARKOLAKIS (forthcoming): “The welfare effects of transportation infrastructure improvements,” *Review of Economic Studies*.
- ARKOLAKIS, C. AND F. ECKERT (2017): “Combinatorial discrete choice,” *Available at SSRN 3455353*.
- BALBONI, C. (2021): “In Harm’s Way: Infrastructure Investments and the Persistence of Coastal Cities.” .
- BALBONI, C., G. BRYAN, M. MORTEN, AND B. SIDDIQI (2021): “Transportation, Gentrification, and Urban Mobility: The Inequality Effects of Place-Based Policies,” .
- BEN-AKIVA, M. E., S. R. LERMAN, S. R. LERMAN, ET AL. (1985): *Discrete choice analysis: theory and application to travel demand*, vol. 9, MIT press.
- BLANCHARD, P., D. GOLLIN, AND M. KIRCHBERGER (2021): “Perpetual Motion: Human Mobility and Spatial Frictions in Three African Countries,” CEPR Discussion Papers 16661, C.E.P.R. Discussion Papers.
- CALVO, G. A. (1983): “Staggered prices in a utility-maximizing framework,” *Journal of Monetary Economics*, 12, 383–398.

⁴⁰We additionally find counter intuitive, negative derivative estimates for the population weighted connectivity measure as well as total mileage. This likely reflects imprecise estimates due to given that in this version of the paper, we only draw from 10 parallel optimised networks.

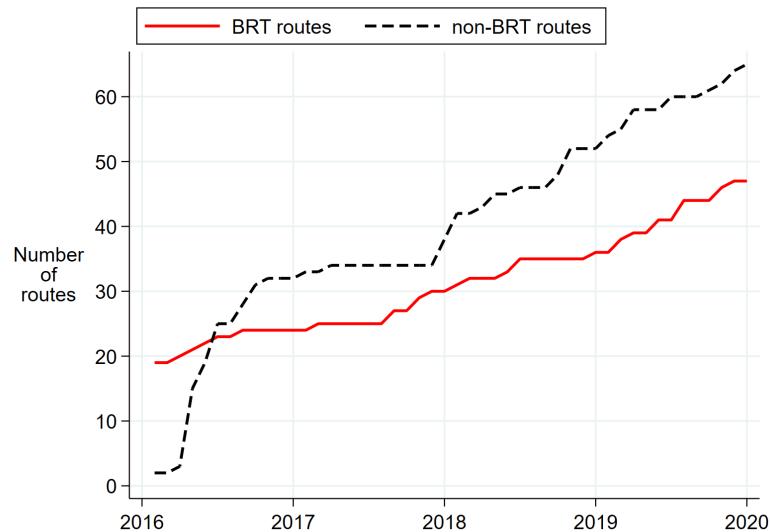
- CAMERON, A. C., J. B. GELBACH, AND D. L. MILLER (2011): “Robust inference with multiway clustering,” *Journal of Business & Economic Statistics*, 29, 238–249.
- DE CHAISEMARTIN, C. AND X. D’HAULTFOUEUILLE (2022): “Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey,” Tech. rep., National Bureau of Economic Research.
- FAJGELBAUM, P. D. AND E. SCHAAL (2020): “Optimal transport networks in spatial equilibrium,” *Econometrica*, 88, 1411–1452.
- GADUH, A., T. GRACNER, AND A. ROTHENBERG (2022): “Life in the Slow Lane: Unintended Consequences of Public Transit in Jakarta,” *Journal of Urban Economics*, 128.
- INSTITUTE FOR TRANSPORTATION AND DEVELOPMENT POLICY (2019): “Transjakarta: A Study in Success,” Tech. rep.
- JIA, P. (2008): “What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry,” *Econometrica*, 76, 1263–1316.
- KREINDLER, G. E. (2022): “Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications,” .
- MCFADDEN, D. (1974): “The measurement of urban travel demand,” *Journal of public economics*, 3, 303–328.
- MOHRING, H. (1972): “Optimization and scale economies in urban bus transportation,” *The American Economic Review*, 62, 591–604.
- NIKOLAEV, A. G. AND S. H. JACOBSON (2010): “Simulated annealing,” in *Handbook of metaheuristics*, Springer, 1–39.
- SANTAMARIA, M. (2022): “Reshaping Infrastructure: Evidence from the division of Germany,” .
- SMALL, K. A., E. T. VERHOEF, AND R. LINDSEY (2007): *The economics of urban transportation*, Routledge.
- TSIVANIDIS, J. N. (2022): “The aggregate and distributional effects of urban transit infrastructure: Evidence from bogotá’s transmilenio,” Tech. rep., University of Chicago.

Figure 1: TransJakarta Network Expansion Since 2016

(a) Expansion of TransJakarta Route Network

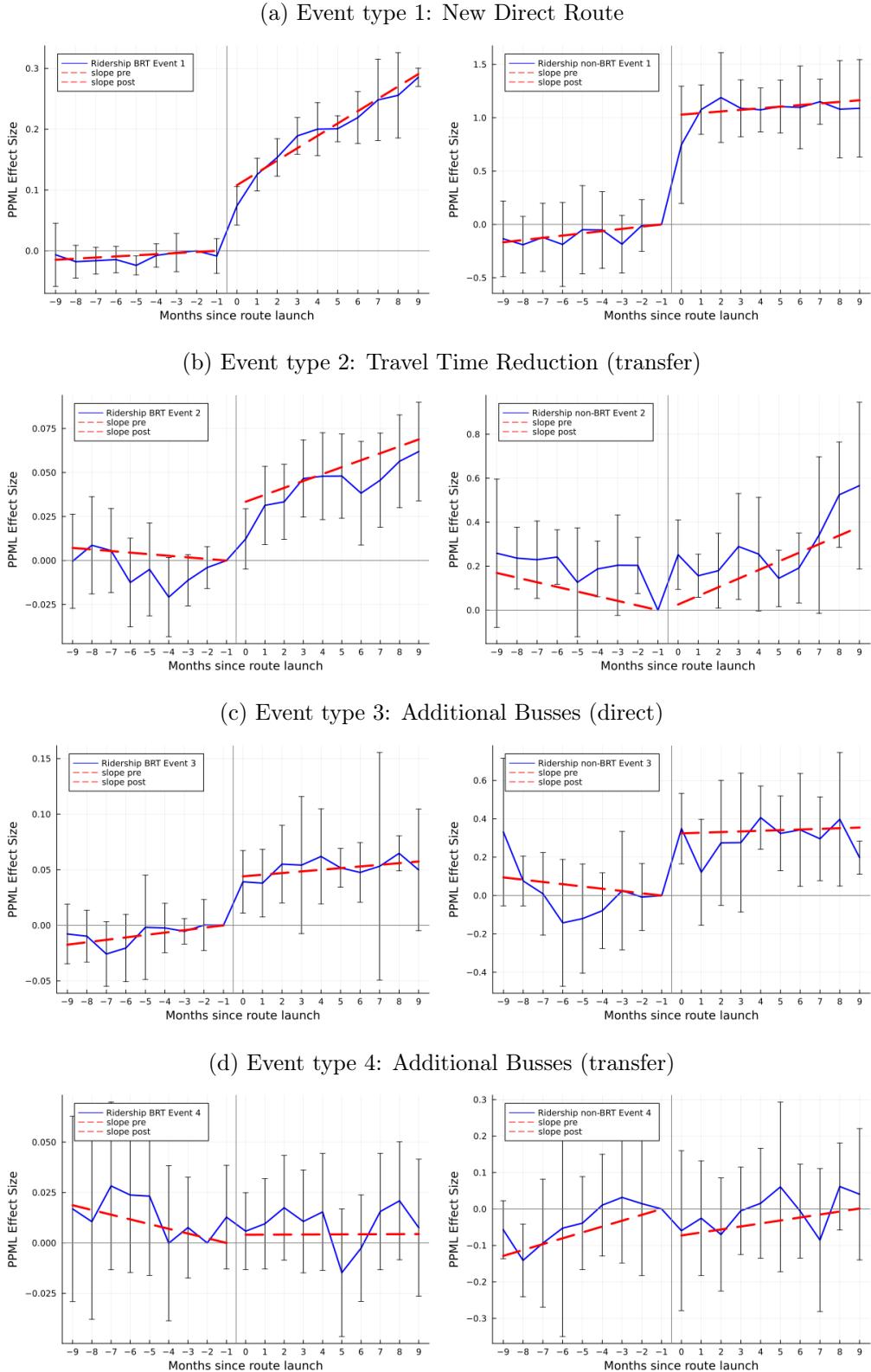


(b) Number of BRT and non-BRT routes over time



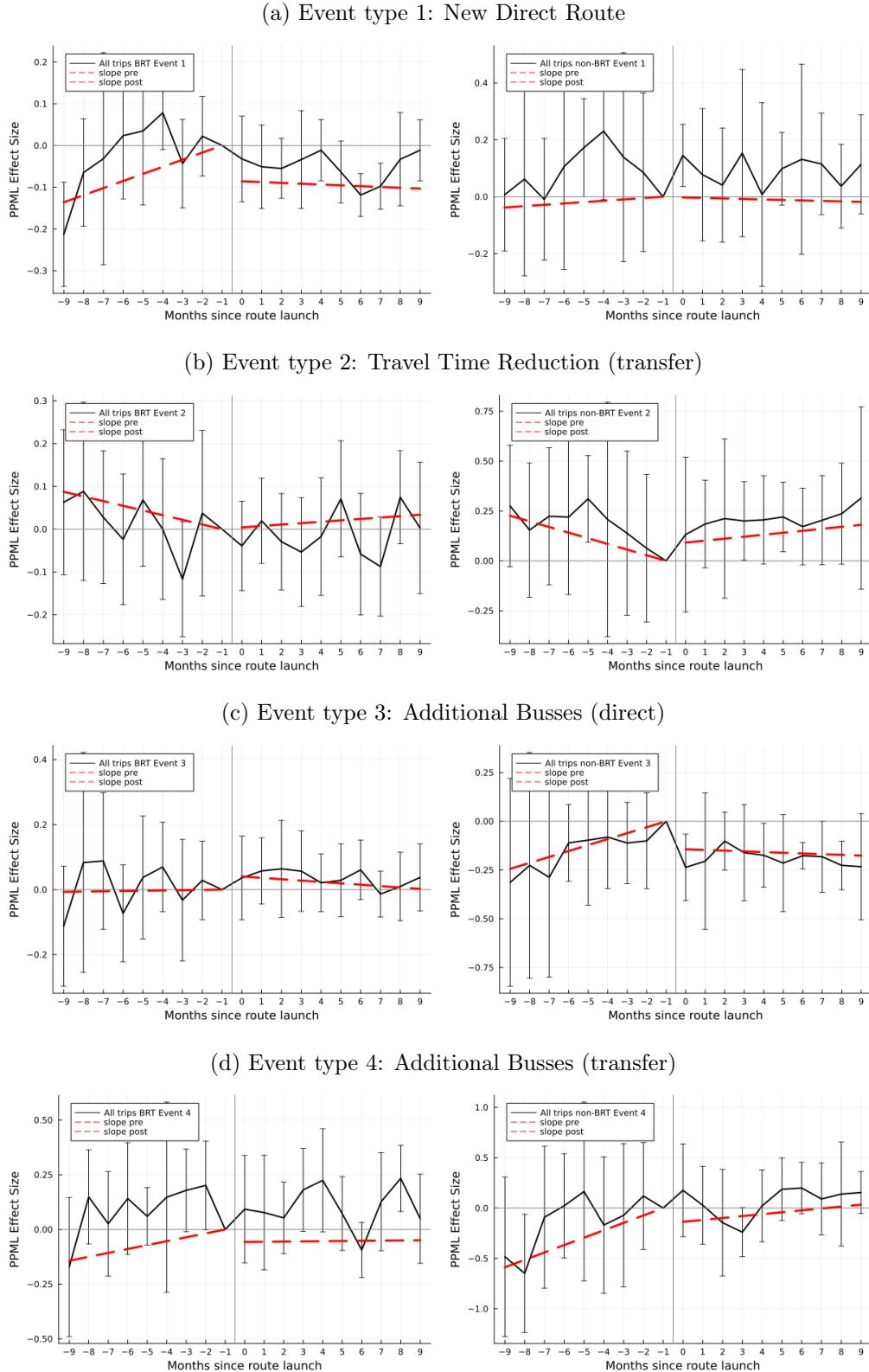
Notes: TO ADD

Figure 2: **Bus ridership** impacts of BRT and non-BRT network expansions



Notes: The blue, solid lines report monthly coefficients from the event study version of equation (1). The red, dashed lines represent linear pre- and post-trends

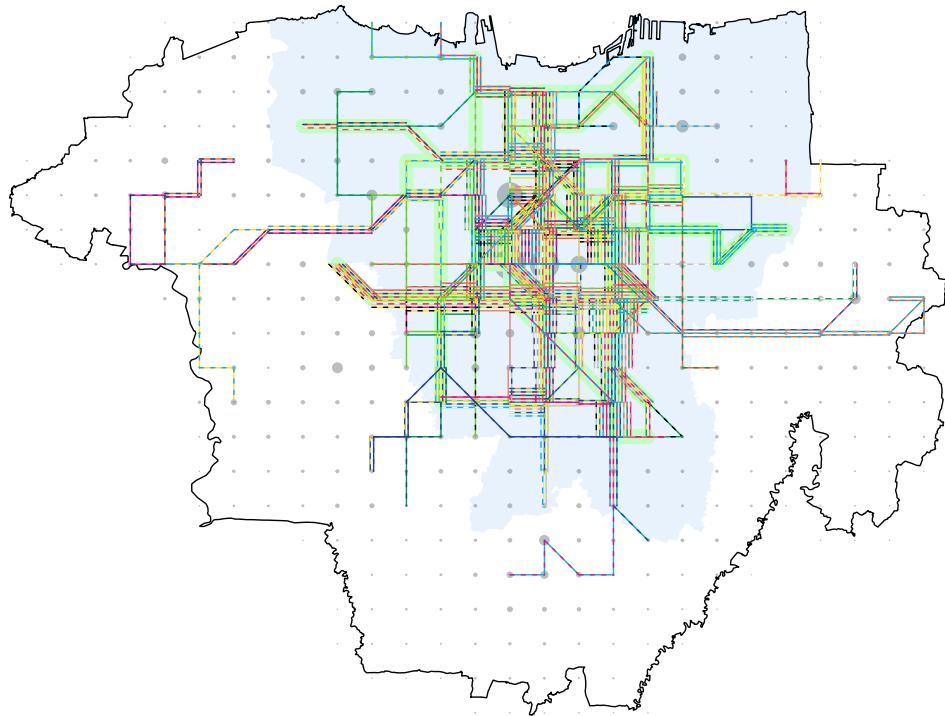
Figure 3: All Trips Impacts of BRT and non-BRT Network Expansions



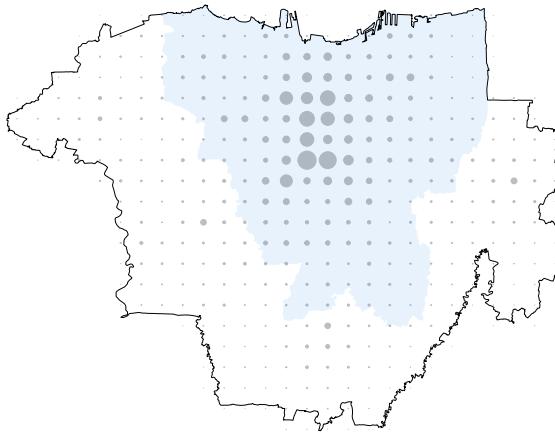
Notes: The outcome is all trips from smartphone the data. The time sample is restricted to after March 2018. The specifications are otherwise the same as in Figure 2.

Figure 4: Current TransJakarta Network

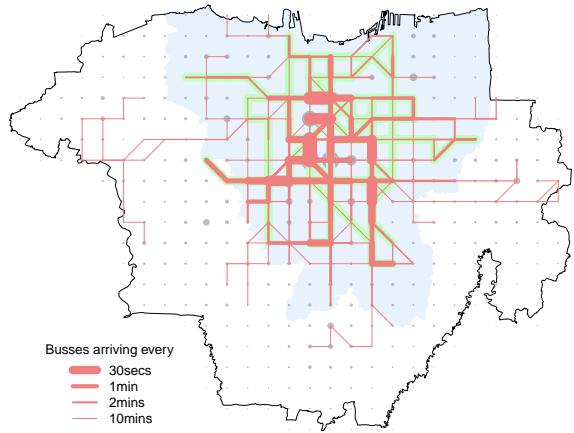
(a) The Bus Line Network



(b) Grid Cells with Trip Weight (origin totals)



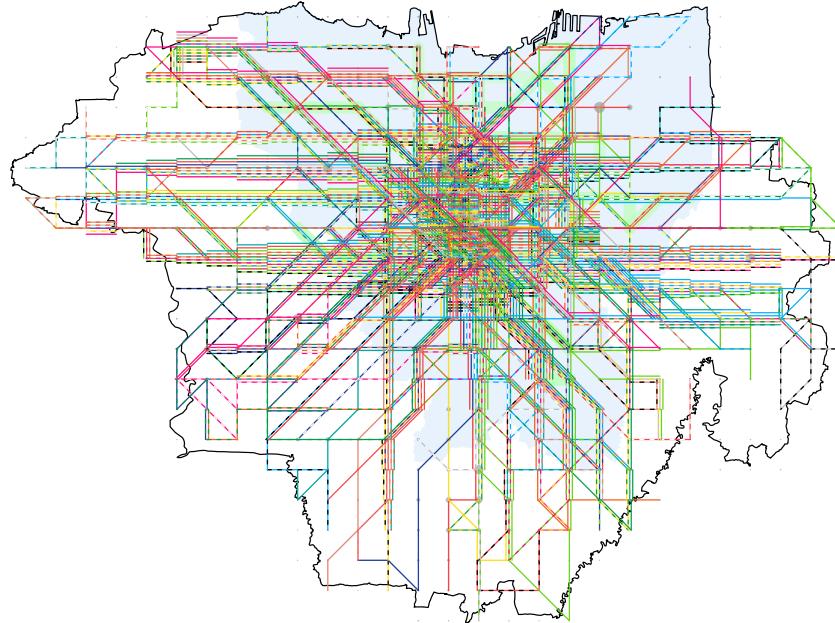
(c) Edge-level Bus Frequency of Service



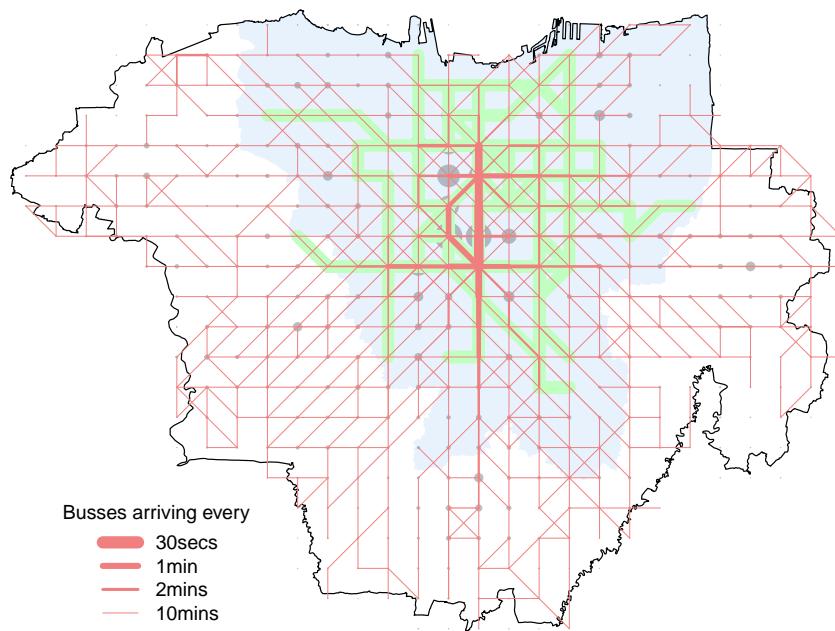
Notes: This figure depicts the current TransJakarta network. The study area is marked by the black outline. It includes the Jakarta DKI area (light blue) as well as the urban districts of Tangerang, South Tangerang, Depok, and Bekasi. Panel (a) plots all 109 bus lines (BRT and non-BRT). In panels (a) and (c), the edges with existing BRT infrastructure are highlighted in light green. Panel (b) shows the 2km square grid with the size of each cell proportional to the total number of trips starting there. Panel (c) shows bus frequency on each edge. To construct this graph, for each route we compute the number of busses per hour (based on the route length and number of busses allocated to that route), and for each (directed) edge we compute the total bus arrival from all routes that share that edge. We then take the mean in both directions and plot the bus arrival rate.

Figure 5: A network draw from the distribution of optimal networks

(a) Resulting bus map from a simulated annealing run



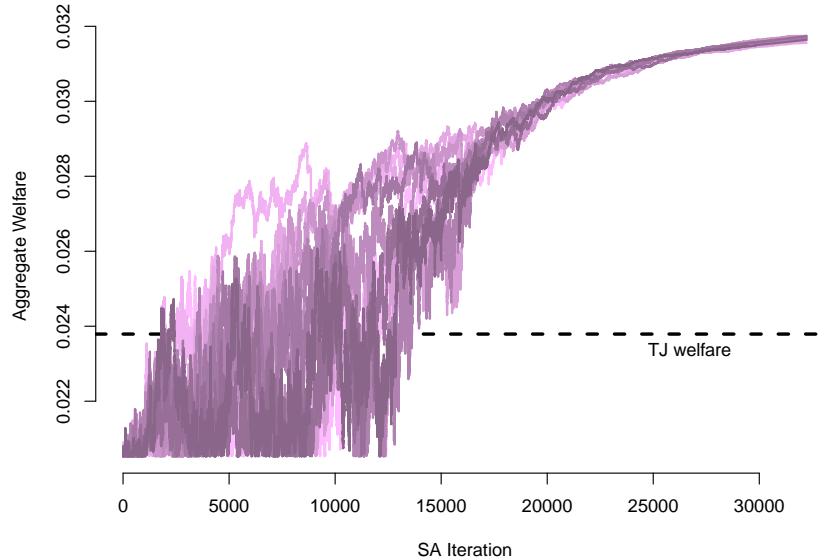
(b) Edge-level Bus Frequency of Service



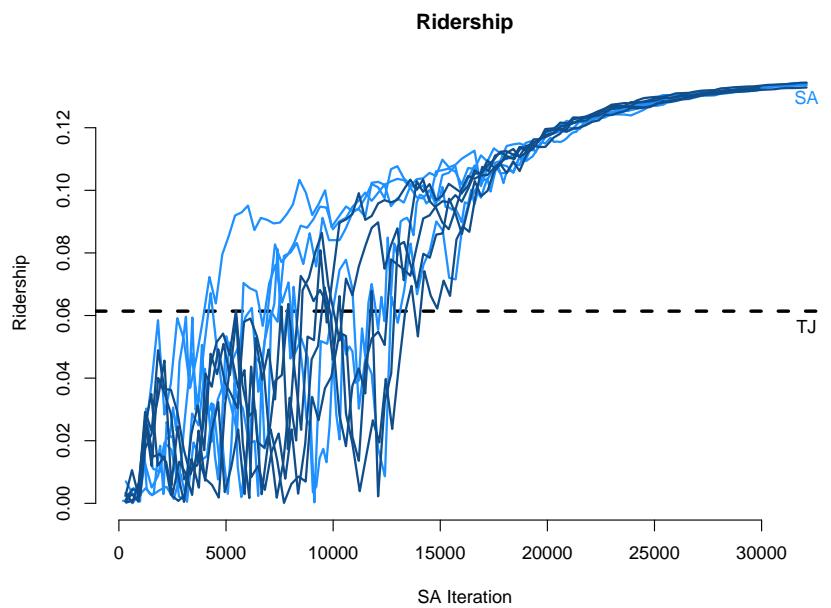
Notes: One example outcome of the simulated annealing algorithm.

Figure 6

(a) Welfare progression over the parallel runs



(b) Ridership progression over the parallel runs



Notes: Each line in these graphs shows the progression of welfare (panel a) and ridership (panel b) for one run of the simulated annealing algorithm.

Table 1: Impact of BRT network expansion on on travel time, wait times, and **bus ridership**

	log Min Travel Time	log Bus/hr (origin)	Bus Ridership
	(1)	(2)	(3)
E1: New Direct Line	-0.130*** (0.016)	0.012 (0.026)	0.203*** (0.020)
E2: Travel Time Large Reduction (transfer)	-0.542*** (0.015)	0.470*** (0.038)	0.047*** (0.010)
E3: Additional Busses (direct)	-0.031*** (0.007)	0.367*** (0.020)	0.056*** (0.017)
E4: Additional Busses (transfer)	0.053*** (0.011)	0.500*** (0.055)	-0.004 (0.012)

Table 2: Impact of non-BRT network expansion on on travel time, wait times, and **bus ridership**

	log Min Travel Time	log Bus/hr (origin)	Bus Ridership
	(1)	(2)	(3)
E1: New Direct Line	-0.471*** (0.071)	0.235*** (0.030)	1.163*** (0.169)
E2: Travel Time Large Reduction (transfer)	-0.571*** (0.035)	0.321*** (0.036)	0.113* (0.050)
E3: Additional Busses (direct)	-0.090*** (0.020)	0.421*** (0.017)	0.302* (0.146)
E4: Additional Busses (transfer)	0.023* (0.009)	0.323*** (0.044)	0.022 (0.046)

Notes: Each entry reports the coefficient on *Post* from a separate regression corresponding to equation (1). Columns 1-2 capture the “first stage” impacts of the relevant events and are estimated using OLS. The bus ridership effects (column 3) are estimated by PPML. “Min Travel Time” captures the duration (excluding wait time) of the quickest route between an origin and a destination, given the routes available in the TransJakarta network at that time. “Buses per Hour” measures the total number of buses arriving at the origin grid over all routes connecting the origin and destination at that point. (The set of routes is restricted to direct routes for event 3.) The coefficients capture the average effect in the first 10 months after the respective event. Regressions performed on a 1km hexagonal grid. Standard errors clustered two-way at origin and destination grid level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Impact of BRT and non-BRT network expansions on **all trips**

	Bus Ridership		All Trips	
	(1)	(2)	(3)	(4)
	BRT	non-BRT	BRT	non-BRT
E1. New Direct Line	0.203*** (0.020)	1.163*** (0.169)	-0.032 (0.035)	0.008 (0.048)
E2. Travel Time Large Reduction (transfer)	0.047*** (0.010)	0.113* (0.050)	-0.020 (0.044)	0.035 (0.079)
E3. Additional Busses (direct)	0.056*** (0.017)	0.302* (0.146)	0.027 (0.035)	-0.061 (0.074)
E4. Additional Busses (transfer)	-0.004 (0.012)	0.022 (0.046)	0.011 (0.067)	0.238 (0.180)
Orig \times Dest FE	Yes	Yes	Yes	Yes
Orig \times Week FE	Yes	Yes	Yes	Yes
Dest \times Week FE	Yes	Yes	Yes	Yes

Notes: This table reports the impact of the eight types of event study on all trips as measured from smartphone data (columns 3 and 4). The outcome is all trips from smartphone data. For comparison, columns 1 and 2 report the impact on bus ridership on the same sample. The specification is as in Table 1, column 3. Specifically, reported coefficients from equation (1). Each coefficient captures the average effect in the first 10 months after the respective event. The time sample is restricted to after March 2018. Standard errors clustered two-way at origin and destination grid level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Estimated demand model parameters

	Wait Time (minutes)	Travel Time (minutes)	Transfer Shifter μ	Transfer Std. Dev. σ_μ
	α_W	α_T		
BRT:	0.227 [0.13, 2.13]	0.037 [0.015, 0.31]	-45.2 [-83.9, -17.6]	37.3 [16.0, 62.6]
non-BRT:	0.157 [0.074, 0.87]		-72.6 [-78.9, -1.5]	32.8 [11.9, 40.7]

Notes: We use a classical minimum distance with the optimal weighting matrix, and 100 random initial conditions. To construct the 90% confidence intervals, we re-estimate the model 100 times. Each time, we target a data moment vector that is randomly drawn from a multivariate normal distribution $\mathcal{N}(0, \Omega)$ where Ω is the variance-covariance matrix of the reduced form analysis, jointly estimated in a seemingly unrelated regression framework. During this procedure, we use 10 random initial conditions for each estimation.

Table 5: Characteristics of the current and optimal networks

Statistic	TJ	Mean	SD	Min	Max
Locations with a station (share)	0.421	0.875	0.008	0.866	0.890
Ridership (share)	0.061	0.151	0.006	0.134	0.153
People living in locations with a station (share)	0.729	0.992	0.001	0.990	0.994
Location pairs with feasible connection (share)	0.113	0.761	0.014	0.744	0.787
Total network mileage (in km)	549.611	1,681.502	51.840	1,616.513	1,769.184
Number of lines	107	150	0	150	150

Notes: This table reports characteristics of the current TransJakarta network (column 1 labelled “TJ”) and of the planner’s distribution of optimal networks in columns 2-5. The statistics in columns 2-5 are computed over 20 independent runs of the simulated annealing algorithm described in Section 5.3.

Table 6: Comparative statics with respect to the cost of wait time

Statistic	(1)	(2)	(3)	(4)
	TJ Network	Baseline Optimal	Local derivative $dF^*/d\alpha_W$	Large change $0.5 \times \alpha_W$
Locations with a station (share)	0.42	0.88	0.45	0.92
Ridership (share)	0.06	0.15	-0.012	0.18
People living in locations with a station (share)	0.73	0.99	-0.60	0.99
Location pairs with feasible connection (share)	0.41	0.76	0.94	0.83
Total network mileage (in km)	549	1,681	-15,692	1,818
Number of lines	107	150	0	150

Notes: The third column reports the local derivative of the characteristic indicated in a given row, with respect to the same proportional change in the cost of wait time for BRT (α_W^{BRT}) and for non-BRT (α_W^{nonBRT}). This corresponds to equation (8) in the 2nd row, and equation (7) for all other rows. In the last column we report the impact of a larger change. It shows the average characteristics from 20 runs where we double the value of α_W^{BRT} . All statistics are computed over 20 independent runs of the simulated annealing algorithm described in Section 5.3.

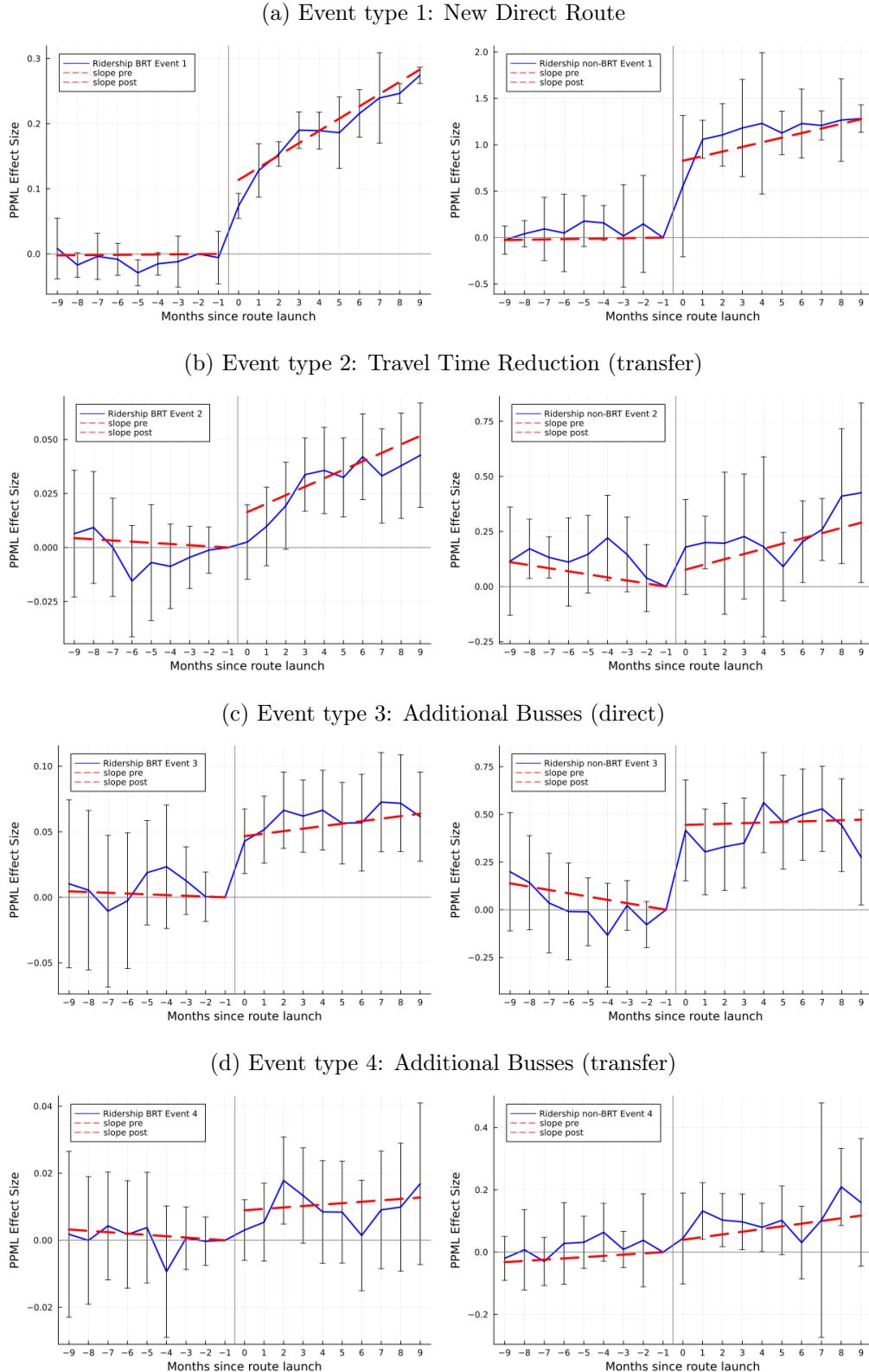
A Online Appendix

Contents

A.1	Appendix Figures	57
A.2	Appendix Tables	63
A.3	Data Processing	67
A.3.1	Bus GPS Data Processing	67
A.3.2	Bus Travel Times	68
A.3.3	Bus Wait Time Distribution	71
A.3.4	TransJakarta Ridership Data Processing	75
A.3.5	Veraset smartphone location data trip processing	77
A.4	Model Derivations and Computation	81
A.4.1	Choice probabilities	81
A.4.2	Expected Utility	84
A.4.3	Expected Travel Time and Expected Wait Time	84
A.5	Optimal network design	85
A.5.1	Optimization environment: driving and bus travel time	85
A.5.2	The Simulated Annealing (SA) Algorithm	85
A.5.3	The candidate network proposal algorithms	86
A.5.4	Obtaining the derivative in equation 7	87

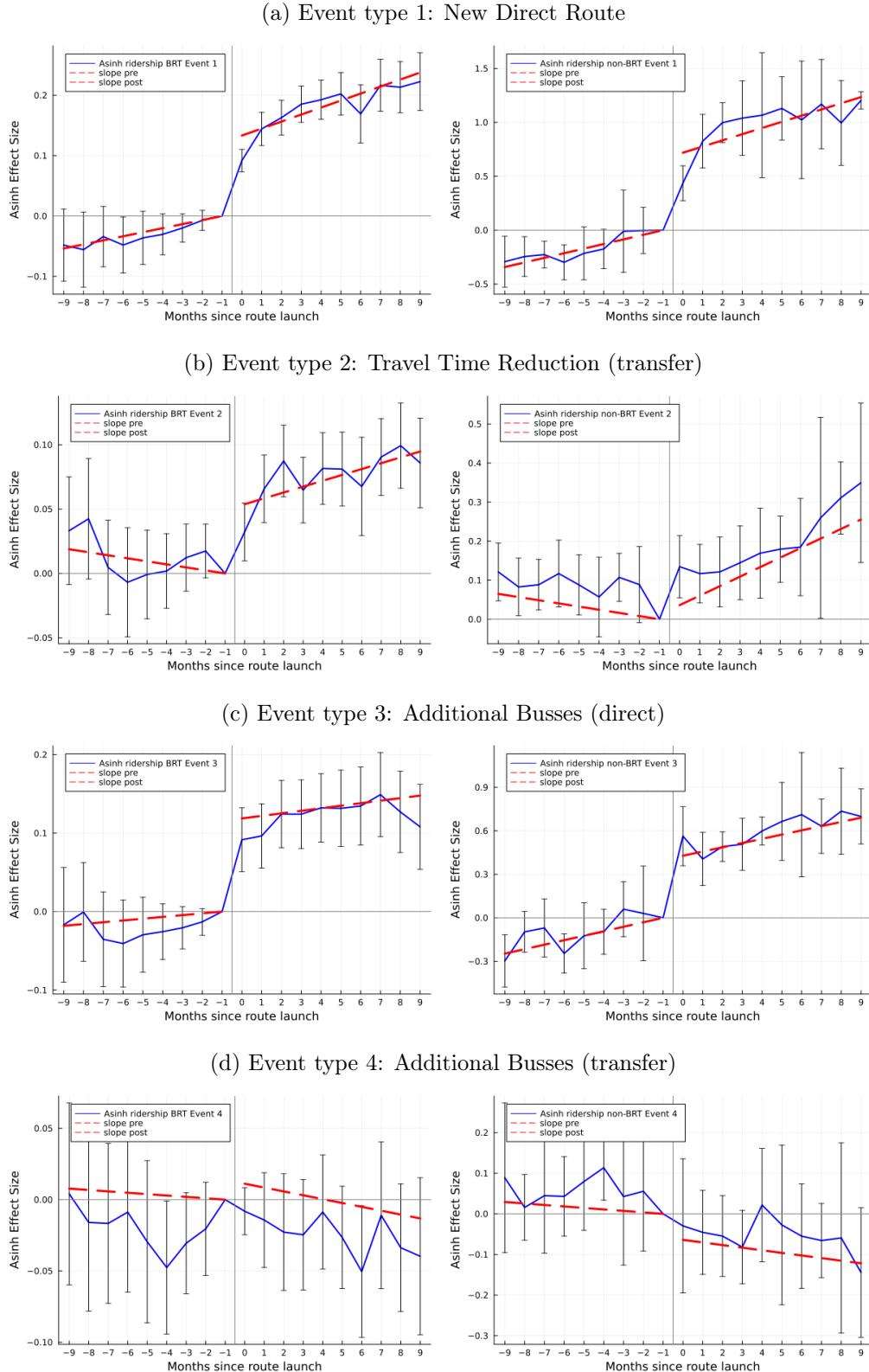
A.1 Appendix Figures

Figure A.1: Robustness with 500m square grids: **bus ridership** impacts of BRT and non-BRT network expansions (PPML)



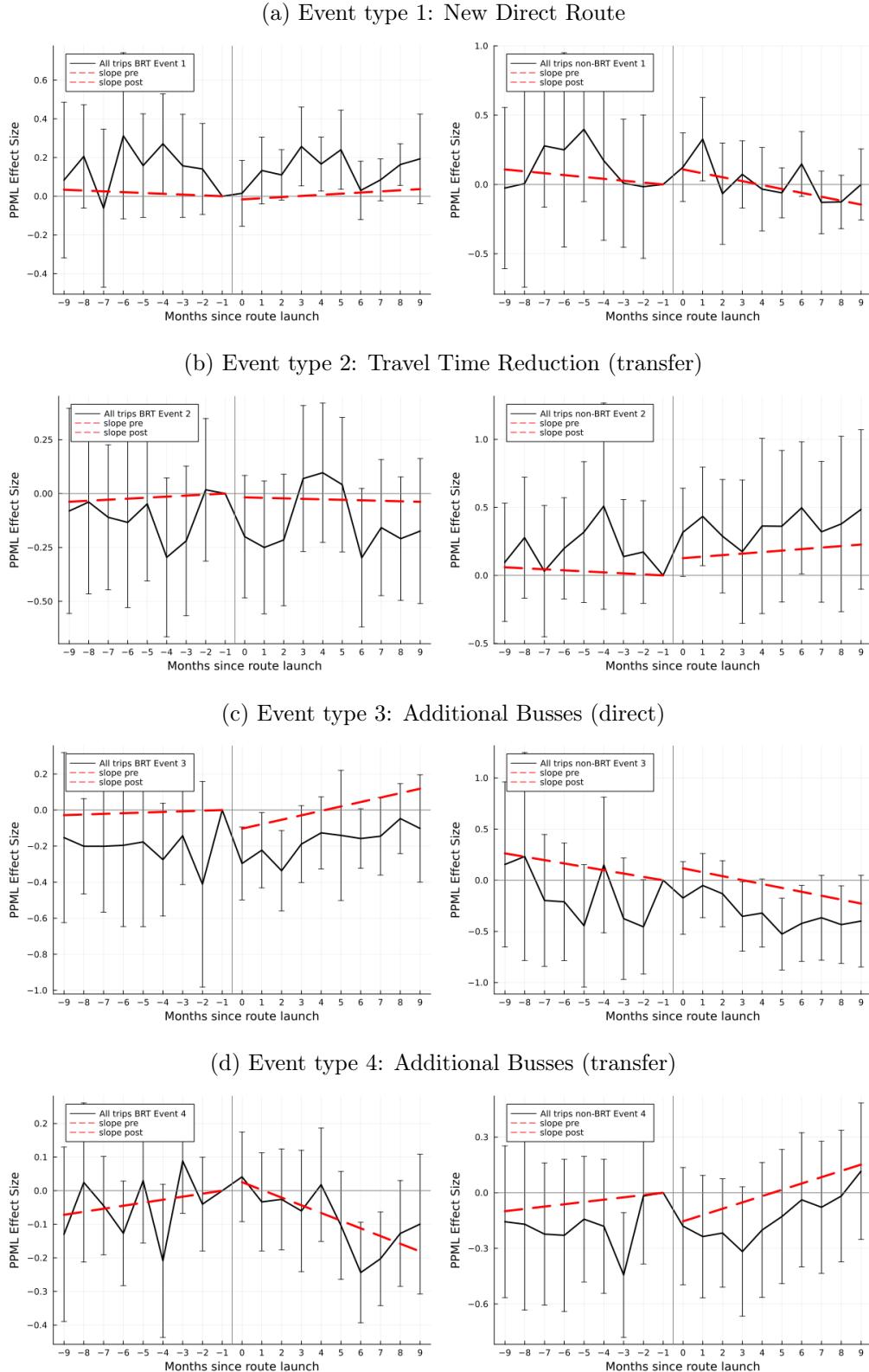
Notes: This graph replicates Figure 2 using 500-metres square grids instead of 1km hexagonal grids.

Figure A.2: Robustness with OLS: inverse hyperbolic sine of **bus ridership** impacts of BRT and non-BRT network expansion



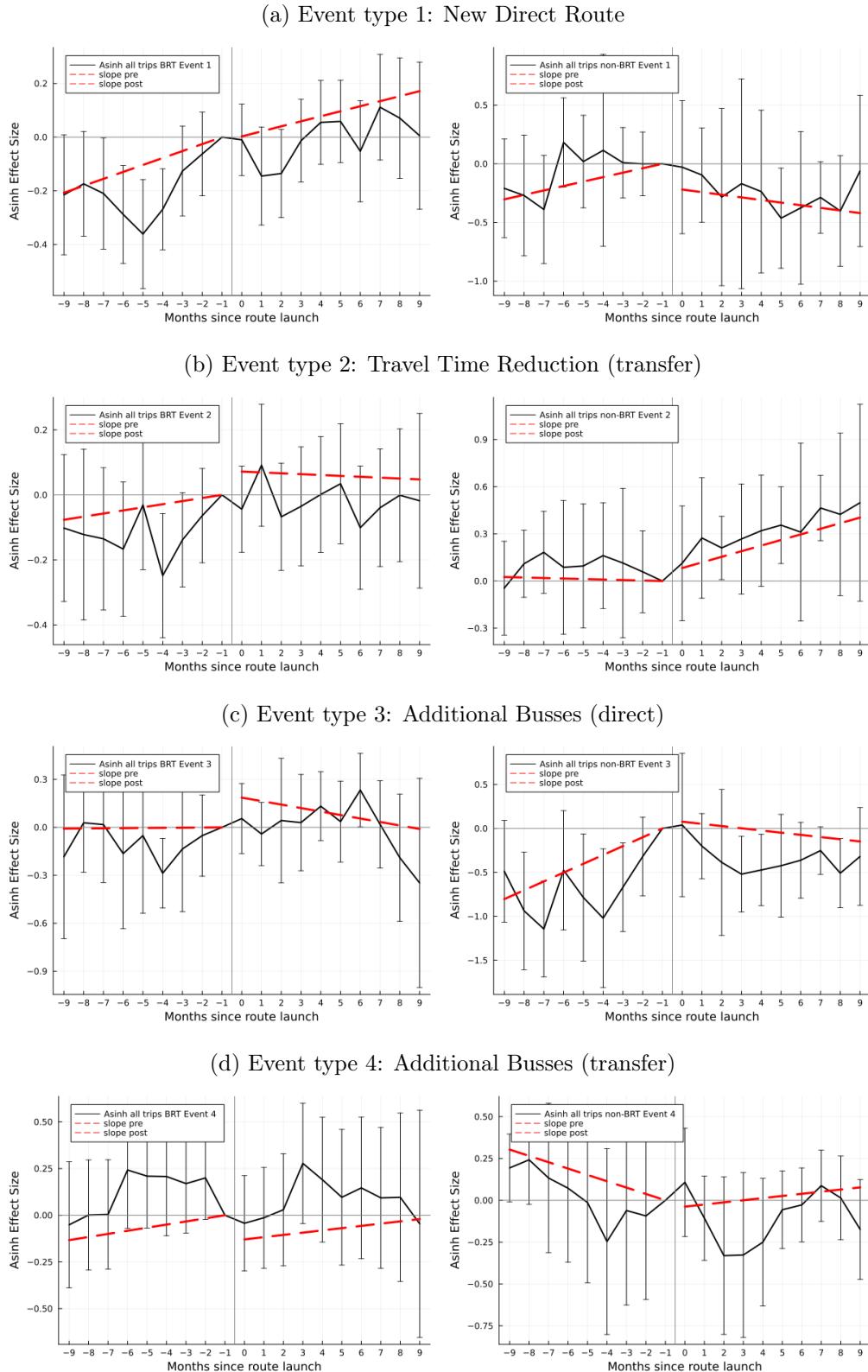
Notes: These graphs replicate Figure 2 using an OLS specification using the inverse hyperbolic sine of ridership as the outcome variable, instead of the PPML specification in our benchmark results.

Figure A.3: Robustness 500m square grids: **all trips** impacts of BRT and non-BRT network expansions



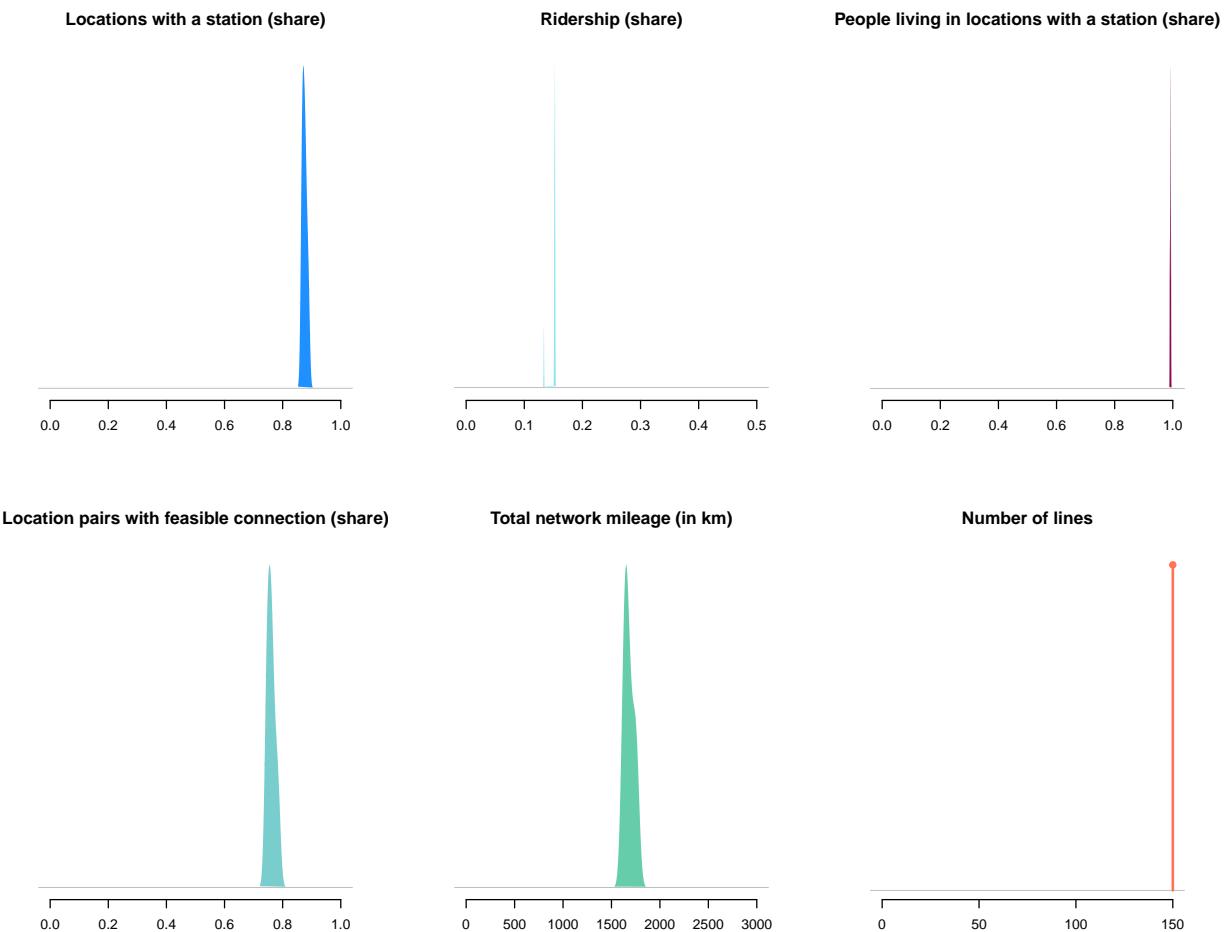
Notes: This graph replicated Figure 3 using 500m square grids.

Figure A.4: Robustness with OLS (asinh, hex 1000m): **all trips** impacts of BRT and non-BRT network expansions



Notes: This graph replicated Figure 3 using OLS estimation and asinh transformation of all trips. 61

Figure A.5: Distribution of network characteristics over parallel optimisation runs



Notes: Each graph plots the kernel density graph of a specific network characteristic, for the final network from the simulated annealing algorithm, over the 10 parallel simulated annealing runs.

A.2 Appendix Tables

Table A.1: BRT route launch order: balance on baseline geographical variables

	β	(1) BRT RI p-value
Outcome Variables		
Total length along the path in meters	-420.38	0.62
# of stations on new route	-1.42*	0.09
Distance from the nearest station on the new route to Sudirman CBD (in km)	0.35 (0.01)	0.38
Average distance across stations on the new route to Sudirman CBD (in km)	0.33 (0.01)	0.28
# of existing baseline stations on the new route	-1.41 (0.02)	0.19
Share of existing baseline stations on the new route	0.01 (0.00)	0.82
# of connecting baseline routes on existing stations on the new route (avg)	0.03 (0.00)	0.75
# of connecting baseline routes on the new route in total	-0.07 (0.01)	0.89
Avg daily ridership across existing baseline stations first 7 days of Jan2016	-81.75 (1.14)	0.19
Median Initial Planned Buses Allocation (at launch)	-1.11*** (0.01)	0.01
N		28

Notes: The following OLS regression equation is performed $Outcome_i = \beta WeekLaunch_i + \epsilon_i$ and the coefficient β and its robust SE are displayed. The coefficients β are multiplied by 52 to be interpreted as “a route launched one year later has...”. The randomization inference (RI) p-value of the t-test is computed using `ritest` in Stata, with 1,000 permutations without replacement.

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

Table A.2: Robustness with 500-meter square grids: Impact of **BRT network** expansion on travel time, wait times, and bus ridership

	log Min Travel Time	log Bus/hr (origin)	Bus Ridership
	(1)	(2)	(3)
E1: New Direct Line	-0.129*** (0.015)	0.020 (0.023)	0.197*** (0.019)
E2: Travel Time Large Reduction (transfer)	-0.548*** (0.014)	0.514*** (0.036)	0.032** (0.010)
E3: Additional Busses (direct)	-0.019*** (0.006)	0.355*** (0.021)	0.053** (0.019)
E4: Additional Busses (transfer)	-0.171*** (0.009)	0.449*** (0.027)	0.009 (0.007)

Table A.3: Robustness with 500-meter square grids: Impact of **non-BRT network** expansion on travel time, wait times, and bus ridership

	log Min Travel Time	log Bus/hr (origin)	Bus Ridership
	(1)	(2)	(3)
E1: New Direct Line	-0.435*** (0.065)	0.251*** (0.027)	1.025*** (0.131)
E2: Travel Time Large Reduction (transfer)	-0.545*** (0.026)	0.348*** (0.033)	0.118** (0.043)
E3: Additional Busses (direct)	-0.087*** (0.014)	0.434*** (0.015)	0.404*** (0.106)
E4: Additional Busses (transfer)	-0.267*** (0.026)	0.424*** (0.028)	0.089* (0.039)

Notes: versions of Tables 1 and 2 using 500-meters square grids.

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

Table A.4: Robustness with OLS: Impact of **BRT network** expansion on travel time, wait times, and the inverse hyperbolic sine of **bus ridership**

	log Min Travel Time	log Bus/hr (origin)	Asinh Bus Ridership
	(1)	(2)	(3)
E1: New Direct Line	-0.130*** (0.016)	0.012 (0.026)	0.205*** (0.022)
E2: Travel Time Large Reduction (transfer)	-0.542*** (0.015)	0.470*** (0.038)	0.065*** (0.014)
E3: Additional Busses (direct)	-0.031*** (0.007)	0.367*** (0.020)	0.139*** (0.026)
E4: Additional Busses (transfer)	0.053*** (0.011)	0.500*** (0.055)	-0.003 (0.015)

Table A.5: Robustness with OLS: Impact of **non-BRT network** expansion on travel time, wait times, and the inverse hyperbolic sine of **bus ridership**

	log Min Travel Time	log Bus/hr (origin)	Asinh Bus Ridership
	(1)	(2)	(3)
E1: New Direct Line	-0.471*** (0.071)	0.235*** (0.030)	1.100*** (0.152)
E2: Travel Time Large Reduction (transfer)	-0.571*** (0.035)	0.321*** (0.036)	0.104* (0.044)
E3: Additional Busses (direct)	-0.090*** (0.020)	0.421*** (0.017)	0.671*** (0.112)
E4: Additional Busses (transfer)	0.023* (0.009)	0.323*** (0.044)	-0.106* (0.048)

These tables repeat the exercise of Tables 1 and 2 using OLS and asinh bus ridership as outcome variable.

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

Table A.6: Robustness on **all trips**

	500m square PPML		1000m hex OLS	
	(1) BRT	(2) non-BRT	(3) BRT	(4) non-BRT
E1. New Direct Line	-0.007 (0.075)	-0.050 (0.109)	0.169** (0.059)	-0.171 (0.155)
E2. Travel Time Large Reduction (transfer)	-0.011 (0.079)	0.140 (0.152)	0.094 (0.061)	0.203 (0.111)
E3. Additional Busses (direct)	0.006 (0.074)	-0.117 (0.138)	0.101 (0.089)	0.303 (0.217)
E4. Additional Busses (transfer)	-0.029 (0.047)	0.018 (0.082)	-0.025 (0.093)	-0.114 (0.122)
Orig × Dest FE	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes

Notes: This table reports robustness exercises of the analysis in Table 3. The first two columns repeat the analysis on a 500m square grid. Columns three and four repeat the analysis on a 1000m hexagonal grid using OLS instead of PPML.

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

Table A.7: Demand Estimation Model Fit (9 Moments)

	(1)	(2)	(3) BRT	(4)	(5)	(6)	(7) non-BRT	(8)	(9)
	Event 1	Event 2	Event 3	Event 4	Event 1	Event 2	Event 3	Event 4	Trip Duration
Data	0.192	0.043	0.056	-0.003	1.163	0.113	0.293	0.022	0.159
Model	0.219	0.072	0.032	0.026	1.134	0.135	0.318	-0.017	0.255

Notes: This table reports the estimated demand model fit for the nine moments, the eight event study coefficients, and the regression coefficient of log trip duration on mean trip among all transfer options.

A.3 Data Processing

A.3.1 Bus GPS Data Processing

This section explains how we process bus GPS data to find bus station arrival times, which we use when describing the wait time distribution and when assigning bus transactions to bus stations (appendix A.3.4). We use GPS data every 5 or 10 seconds available for most TransJakarta busses between January 2017 and March 2020. We also use bus trip logs entered manually by bus dispatchers. For each trip, this data contains the bus code, the bus route code (with direction) and the trip start time. **2,798** TransJakarta busses appear in the GPS data.

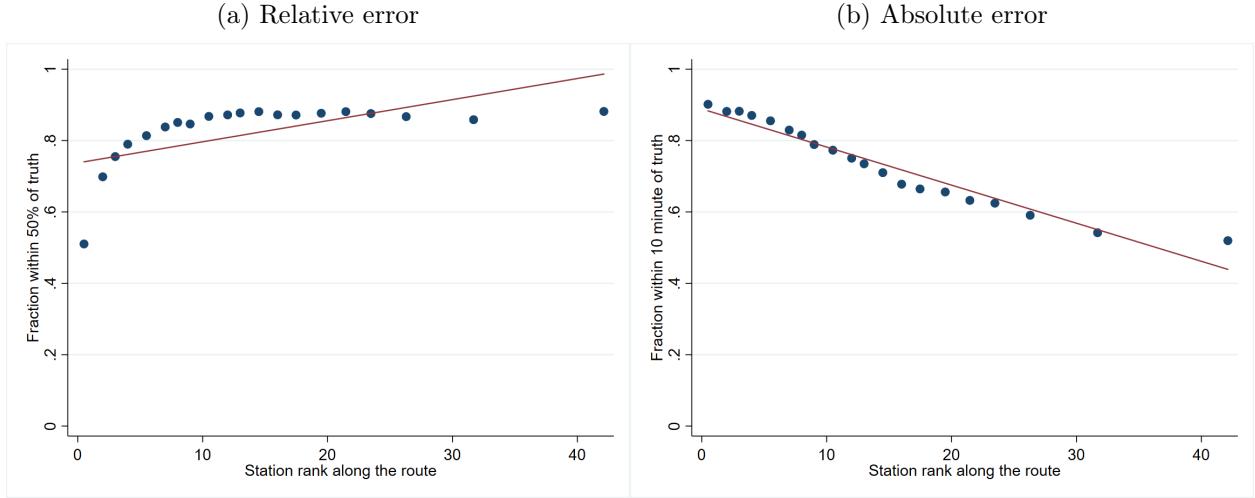
Combined data on GPS and trip logs is significantly better starting in 2018. In 2017, **22.2%** of bus days contain both GPS data and trip logs, **76.2%** contain only GPS data but no logs, and the rest contain only trip logs without GPS data. However, in 2018 - March 2020, **70.7%** of bus days contain both GPS data and trip logs, **22.8%** contain only GPS data but no logs, and the rest contain only trip logs without GPS data.

We developed three algorithms to identify when a bus arrives at a given bus station. When both GPS data and trip logs are available, we map match the bus GPS locations to the path of the bus route (from separate data), starting from the trip start time recorded in the trip log, and find arrival times for all bus stations along that route. (The algorithm automatically identifies bus trips where the “return” trip log is missing, which happens **15.4%** of the time.)

When only GPS data is available, the algorithm proceeds in two steps. First, given a bus and a date, it ranks bus routes in decreasing order of overlap with the GPS traces for that bus day and generates a set of candidate routes where the traces overlap at least **30%** with the route. Second, the algorithm map matches the GPS traces to the best-fit bus route, trying multiple candidates, starting from the first GPS point that is near to the first station of a candidate bus route.

When only trip log data is available, the algorithm proceeds in two steps. First, we predict station arrival time as a function of trip start time, bus route, and time of the day. We estimate this model using the output of the first algorithm for the same route, on bus-days when both GPS and trip log data is available. Figure A.6 shows that we achieve high accuracy using this model.

Figure A.6: Accuracy of predicting station arrival time based only on trip start time



We end up with a total of **7,315,854** trips out of which **5,568,890 (76.1%)** are identified when both GPS data and trip logs are available, **859,776 (11.8%)** are identified when only GPS data is available and the remaining **12.1%** are identified when only trip log data is available.

A.3.2 Bus Travel Times

In this section, we describe how we compute travel times between an origin station o , a destination station d , along a route r . These travel times are used in the reduced form analysis to compute bus route travel times and changes (especially for defining the second type of event), and in the model to characterize the bus network choice set of any given traveler.

For every triplet (r, o, d) we consider all trips along r and the time they take to go from o and d . We then take the median travel time within this set, over all trips between 7 am and 7 pm in our study period.

Figure A.7 shows that there is only a very small amount of variation in “delay” (median travel time per kilometer, or inverse speed) for trips starting at different times of the day between 7 am - 7 pm. The variation is even smaller for BRT routes. Moreover, we can also observe that delay is mostly stable over the years in our sample. This supports our choice of using medians of travel times of trips starting between 7 am - 7 pm, over the entire data period (January 2017 - March 2020). Moreover, columns 3 and 4 of table A.8 also show that log median travel times across years have a slope close to 1, after controlling for routes,

origins and destinations.

Figure A.7: Delay by departure time and year

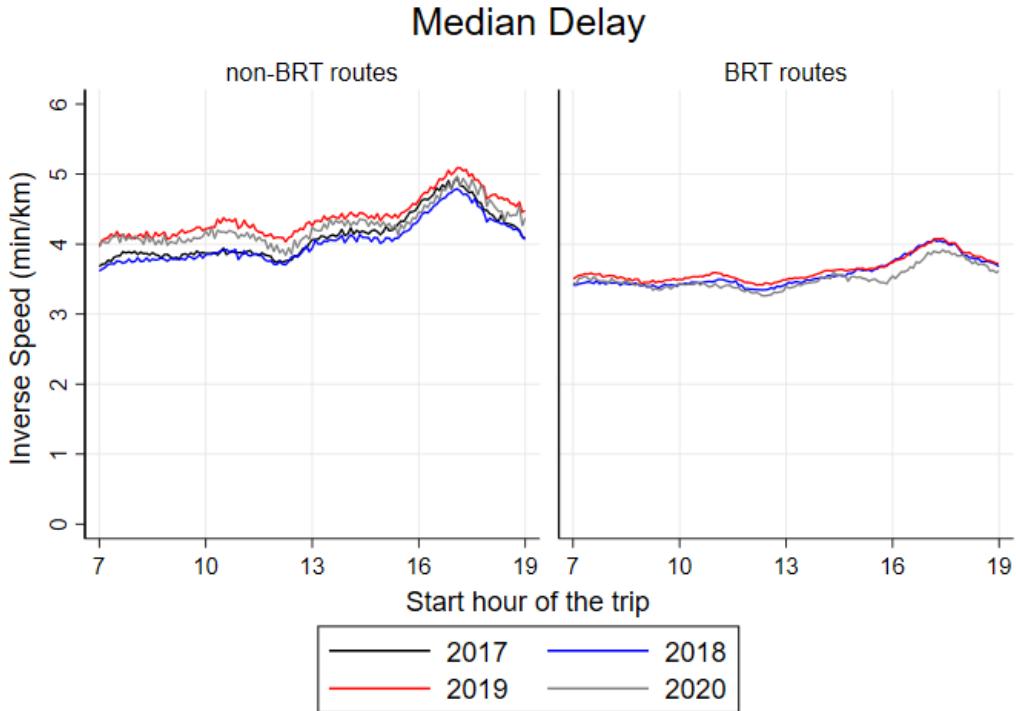


Figure Notes: This figure reports median delay (inverse speed, in minutes per km) by departure time, separately for BRT routes and for non-BRT routes using data from January 2017 - March 2020 for routes that were active throughout this period. To construct this figure, we assign each trip to 5-minute bins by their departure time at first station on the route. We then find the time taken by the bus to complete each trip and the distance travelled for the entire route (start to end). Using this, we calculate the inverse speed for each trip (time/distance). Within each 5-minute bin, we finally plot the median of inverse speed of trips throughout the day, grouping the trips by year.

Table A.8: **Travel Time Correlation at Different Points in Time**

	Log(Median Delay) 2017 (1)	Log(Median Delay) 2018 (2)	Log(Median Travel Time) 2017 (3)	Log(Median Travel Time) 2018 (4)
Log(Median Delay) 2019	0.716*** (0.048)	0.794*** (0.038)		
Log(Median Travel Time) 2019			1.031*** (0.007)	1.009*** (0.002)
Constant	0.368*** (0.062)	0.222*** (0.047)	-0.193*** (0.043)	-0.083*** (0.018)
R ²	0.476	0.685	0.961	0.983
N	26990	38624	26990	38624

Table Notes: We organize the data at the route-origin-destination level. The outcomes variables are the log of median delay (inverse speed, in minutes per km) and log of median travel time (in minutes). Standard errors are clustered three-way by route, origin and destination, and reported in parentheses: * p < 0.05, ** p < 0.01, *** p < 0.001

A.3.3 Bus Wait Time Distribution

In the model, we assume that wait times are exponentially distributed. In this section we analyze how wait times for different TransJakarta routes are distributed in reality, using the GPS data to measure bus headways and to compute the implied wait times.

To compute the wait time distribution, we calculate bus headways (time difference in minute between two consecutive buses) for every station, route, and direction. Then, we calculate the frequency of headway occurrence by route and minute. Assuming that people arrive at bus stations at a constant rate, we calculate how many people wait for every minute from 1 to 30 (we restrict to 30 to exclude overestimation of headway due to unavailability of bus GPS data) by taking the reverse cumulative total of headway frequency. Since each Transjakarta route might have route variants that differ in length and stations reached, we compute the wait time distribution for 45 routes (18 BRT and 27 non-BRT) with low trip variations (the top two trip variants account for above 95% of all trips on that route).

Taking a sample of non-BRT route 1E, Figure A.9, we see that wait time distribution is approximately exponential except for larger wait times. Moreover, Figure A.10 shows BRT and non-BRT routes share a similar pattern in which shorter wait times fit exponential distribution better than larger wait times.

Figure A.8: Sample Arrival Time Monitor at a TransJakarta Station



Figure A.9: Headway and wait time distribution of route 1E

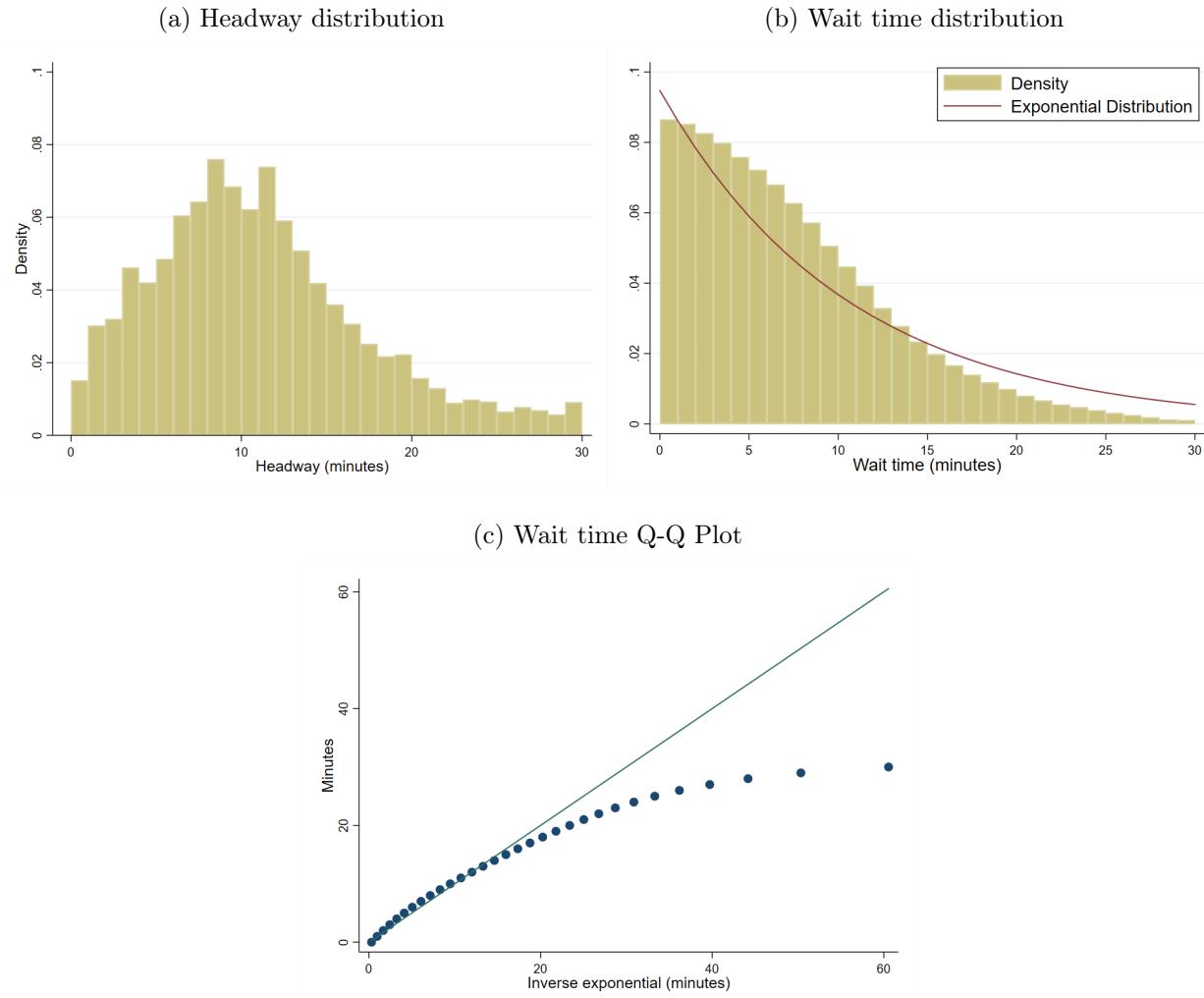


Figure Notes: This figure reports the headway (defined as the duration in minutes between two consecutive buses) and the wait time distribution for non-BRT route 1E, using data from January to October 2019. We restrict the sample to only include arrival times during morning peak hours (8 AM-12 PM). To create this graph, we calculate bus headway using the bus GPS data. To obtain the wait time distribution from the headway distribution, we assume that passengers arrive at the station at a constant rate. In panel (C), we fit an exponential distribution to the empirical wait time distribution.

Figure A.10: Correlation between average wait time and wait time coefficient of variation

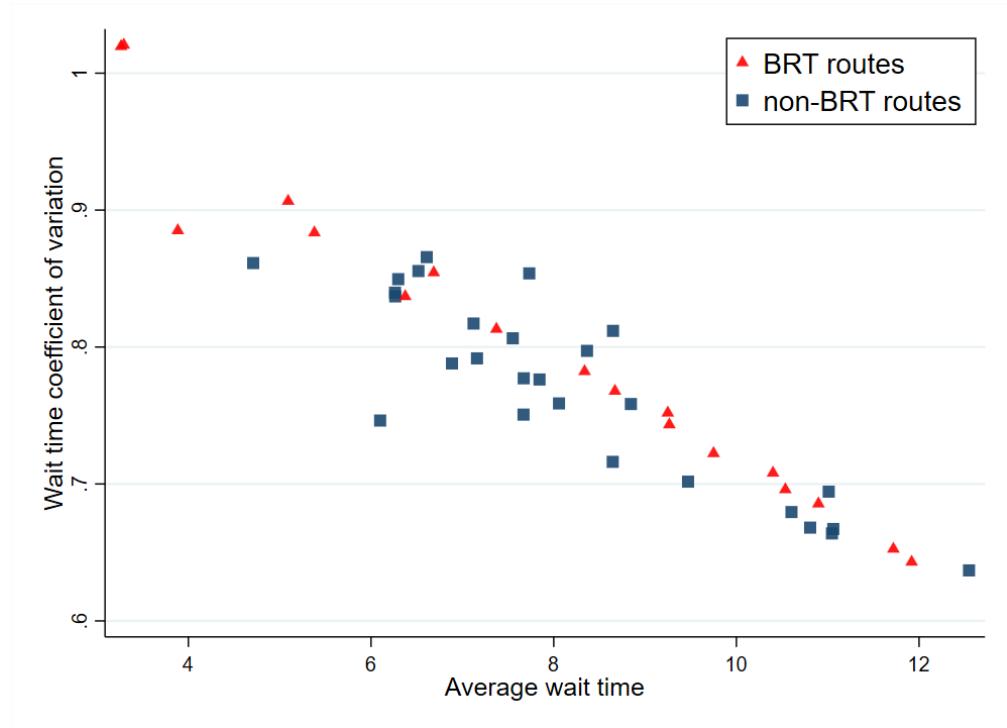


Figure Notes: This figure reports the mean of wait time (in minutes) and the coefficient of variation of wait time, separately for BRT routes and for non-BRT routes, using data from January to October 2019. The sample is restricted to 45 routes (18 BRT and 27 non-BRT routes) where headway calculation is straightforward. Each Transjakarta route may have several route variants that differ slightly in length and stations reached. We restrict the sample to routes where two trip variants account for above 95% of all trips on that route.

A.3.4 TransJakarta Ridership Data Processing

We use two main TransJakarta ridership data sources. First, we use transactions (“taps”) in BRT bus shelters, where passengers pass through turnstiles to enter the bus shelter. In theory, passengers also need to “tap out” when they exit a bus shelter. This is only enforced in **35.9%** of bus shelters (accounting for **34.0%** of shelter ridership).

Second, we use transaction data from non-BRT busses. When a passenger gets on the bus from a (non-BRT) bus station on the side of the road, they pay inside the bus using their card or cash. When a passenger uses cash, the bus attendant uses their own card on the card reading machine. To assign these transactions to bus station, we use the bus station arrival time from the GPS or trip log data (section A.3.1).

Origin-destination ridership flows. To construct a measure of origin-destination ridership flows at each point in time, we proceed in several steps. We focus on a sample of cards for which we observe ridership behavior over time. At each step, we construct weights assuming that the sample of cards and transactions that we use is representative.

First, we drop “administrative” cards that are likely used by bus attendants or other TransJakarta employees. We label a card as administrative on a given day if it is used repeatedly throughout the day on the same bus (for non-BRT transactions), or at a BRT shelter. Administrative cards account of **2.1%** of all BRT shelter transactions, and **63.2%** of all non-BRT bus transactions. We assume that travel behavior for non-admin cards is representative of

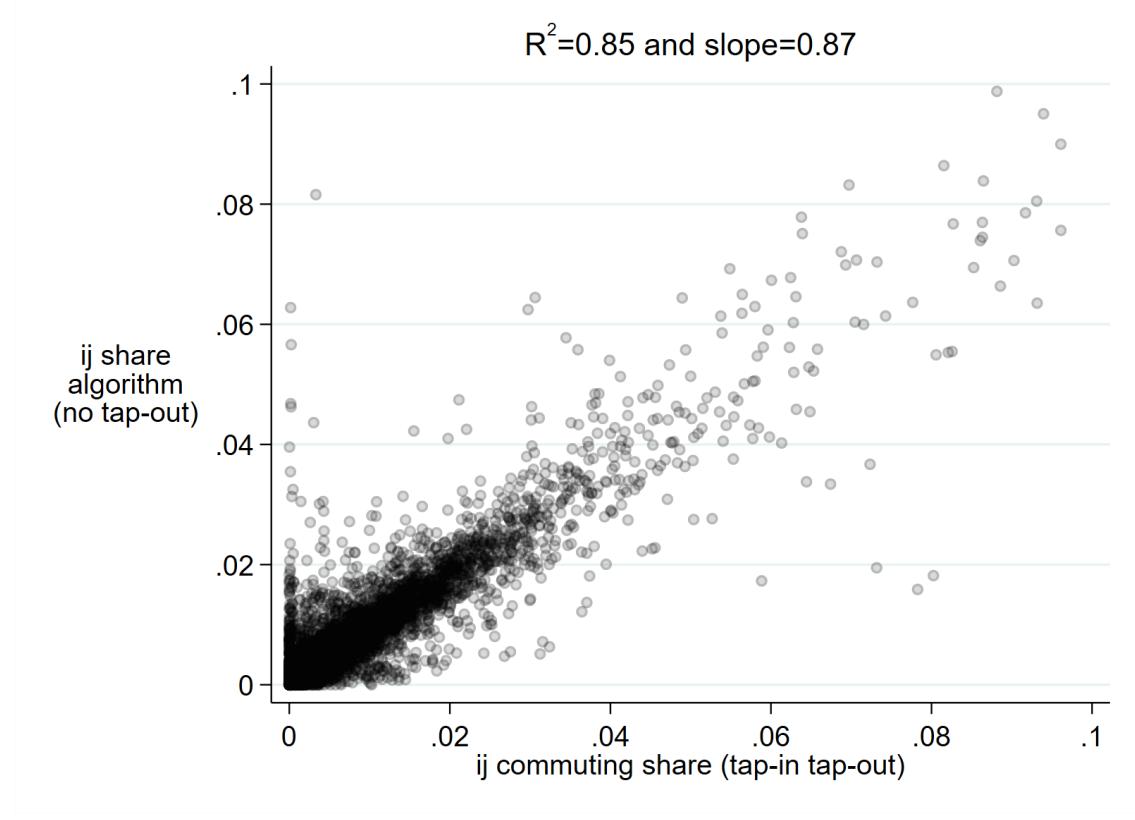
Out of 67 routes that were initially in our sample, there were 15 routes with ridership data that we decided to exclude from our study for reasons which stem from the way payments were made for boarding a non-BRT bus. For non-BRT trips, passengers could pay using cash or cards (card balance deducted using bank EDC machines and/or Transjakarta tap-on-bus machines, depending on their availability). A route might be excluded from the sample if the majority of its trips were 1) paid with cash, hence no digital record of smartcard taps and/or 2) paid using EDC machines from one of the banks. Note that EDC machines could only be used with their corresponding bank cards and were usually assigned to certain routes for a period of time. The payment data provided by this particular bank contained micro data of individual taps, but did not have enough information of card identifiers and precise time of the taps which are necessary to construct the origin-destination. Tap data from these two sources accounts for 14.96% of total tap data in 2017-2020. Additionally, data availability varies over time and we have relatively lower data coverage during the earlier period of the study with 44.11% coverage of total routes operating in 2017 compared to 72.14% coverage of total routes operating in 2020.

Serial taps. We combine consecutive taps using the same card into a single transaction, likely capturing groups traveling together and using a single card. 8.1% of transactions have two or more consecutive taps.

Algorithm to Infer Trip Destination and Validation. Ideally, we want to determine a commuter’s origin-destination based on the tap-in and tap-out locations of their trip. However, tap-out information is not always available across all routes and stations in TransJakarta, especially because on non-BRT routes, commuters are not obligated to tap out when they get off the bus. Even on BRT routes, tap-outs on TransJakarta stations are not always enforced. Given these limitations, we developed an algorithm to infer a commuter’s destination for each ridership tap, in which we create a set of probable trips, and determine a certain order of priority to assign the trip destination. There are three main methods of assigning destinations that account for XXX% of monthly trips. The first method is straightforward using actual tap-out location as the destination of trips with tap-out data within 4 hours after its tap-in. For the second method, we predetermined a pair of top-2 stations where most taps of a card located at and we assign one of the these two stations as destination of the trip. We only use the second method if the tap is located at the other station in the pair and the card is defined as frequent commuter (have more than 10 taps in a month and more than 75% of total taps located within its top-2 stations). The third method is using the next tap-in entry as destination of the previous tap-in entry for trips with consecutive tap-ins at different stations.

To assess our algorithm of determining destination of Transjakarta trips, we compare the probability of assigning trips originating from station i towards station j (destination) using data with only tap-in entry (inferred o,d) and data with both tap-in and tap-out entry (actual o,d). Figure shows that the inferred o,d corresponds to actual o,d.

Figure A.11: Comparison of actual and inferred (o, d) ridership



A.3.5 Veraset smartphone location data trip processing

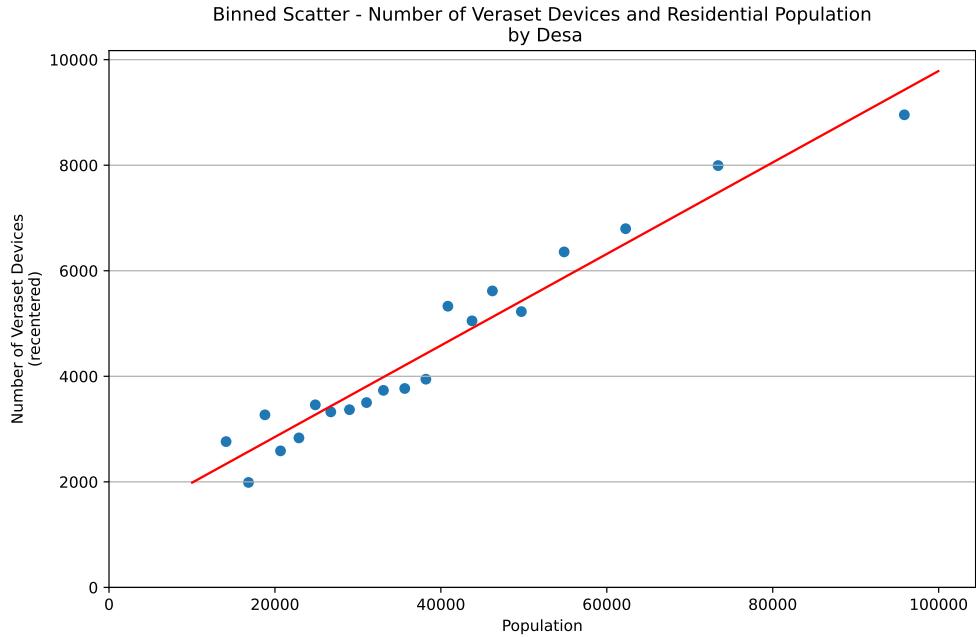
We use a smartphone trip processing algorithm to convert raw GPS data into individual trips and common location for each device in the data [Kreindler \(2022\)](#).

For each individual i , the algorithm first classifies individual trips and “stays,” periods of time when the device is observed to be stationary in a given area. Then, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is used to cluster all locations for any given devices. The most common cluster is labelled as “home.”

The sample of trips used in the analysis is all weekday trips starting and ending at a known locations (i.e. locations classified by DBSCAN as belonging to a cluster), excluding trips starting before 5AM or after 11PM. We also drop trips that are unusually long or short (in both duration and distance), “swiggly” trips (where ratio of the largest distance between any two points on the trip to path length is less than 0.3), and “short loop” trips (trips that are less than 2km where the ratio of origin-destination distance to path length is less than 0.3).

Given the selection concerns when using smartphone location data [Blanchard et al. \(2021\)](#), we examine how representative the users in our data are of the general Jakarta population. First, we compare the distribution of users' home locations obtained from the data to that of residential population from the PODES survey (Figure). The number of devices in our data with home locations in each desa is correlated with populations in the desa, suggesting that distribution of Veraset devices is consistent with population distribution. However, the coverage of our data (number of devices per total population) is slightly lower in areas with higher population density and proportion of population under poverty.

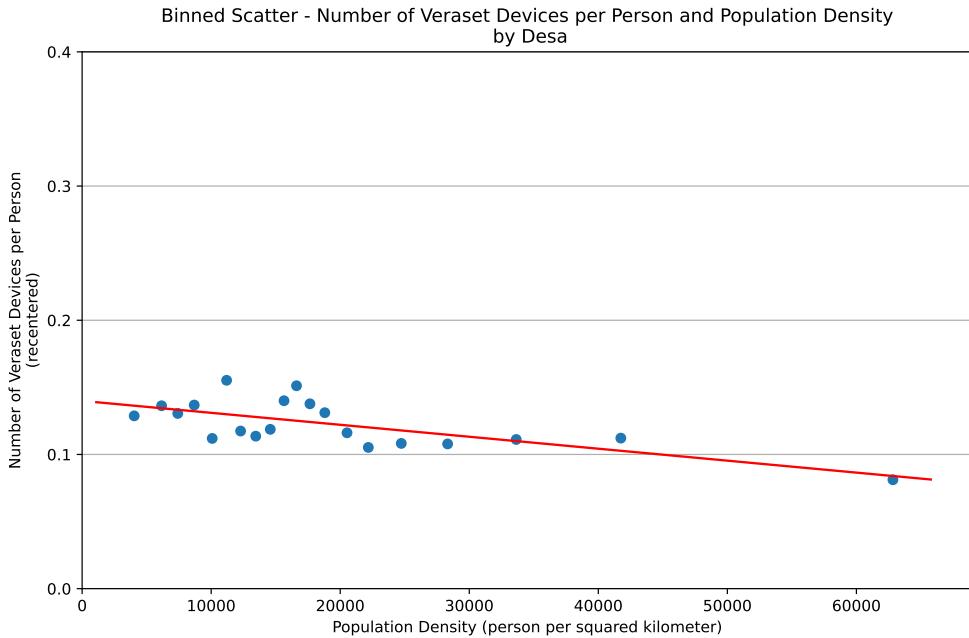
Figure A.12: Correlation between Veraset devices and residential population



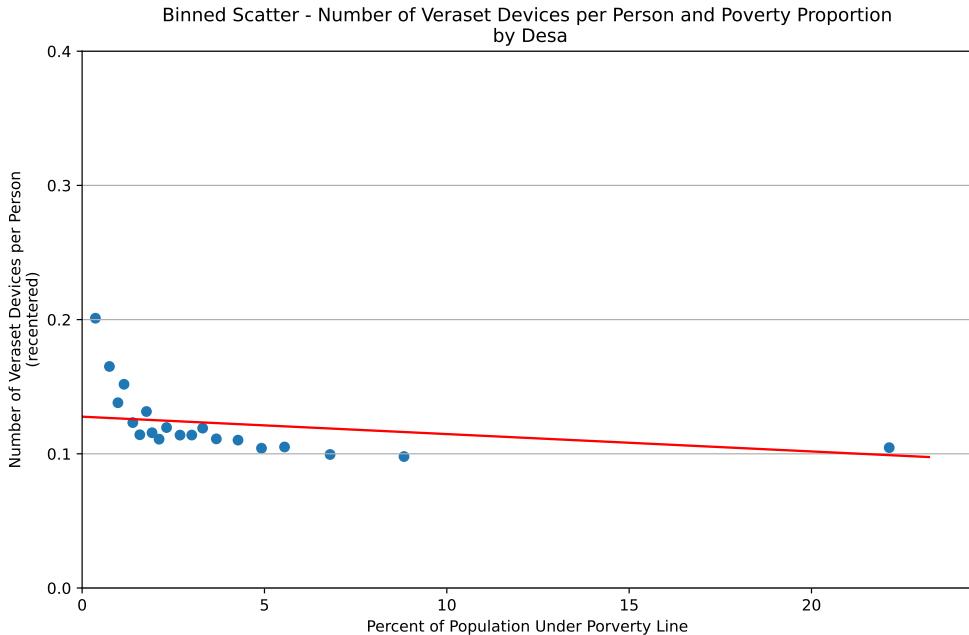
Each observation is an urban neighborhood (kelurahan or desa), $N = 538$. Population data is from the 2011 PODES survey, the most recent source for population at this level of geographical detail. For each desa, we count all devices which have their "home" location assigned inside the desa.

Figure A.13: Representativeness of the Veraset trip data

(a) Population density

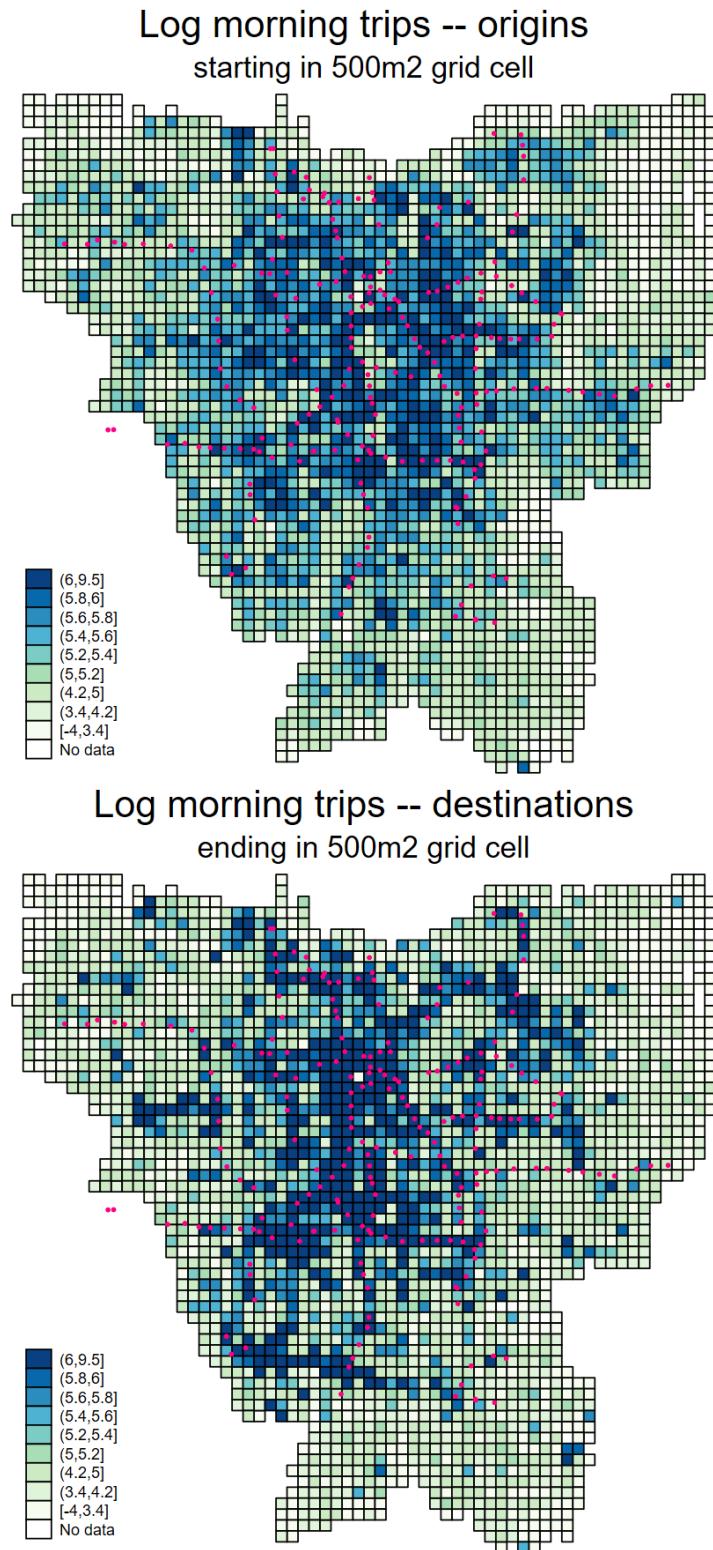


(b) Poverty



Notes: Binscatter plots highlighting the representativeness of the veraset trip data locations. Panel A shows that more dense areas of Jakarta are less well represented in the dataset. Panel B shows that poorer areas are also less well represented.

Figure A.14: Map of origin and destination locations in Jakarta from smartphone data



A.4 Model Derivations and Computation

Proof of Proposition 3. Here we derive expressions for the probability to choose option k , $\Pr(u_k > u_i, i \neq k)$, and for expected utility $\mathbb{E} \max_k u_k$. We also derive expressions for expected travel time and expected wait time.

In general, assume that we have independent random variables X_1, X_2, \dots, X_N , then

$$\Pr(k \in \arg \max_j X_j) = \int_{-\infty}^{\infty} f_k(x) \prod_{i \neq k} F_i(x) dx \quad (10)$$

where f is pdf and F is cdf. The probability that $X_k = x$ is the max is given by the pdf, and the probability that all other variables are lower, is given by the product of CDFs (note independence). Expected utility is

$$\mathbb{E} \max_k X_k = \int_{-\infty}^{\infty} x \sum_k f_k(x) \prod_{i \neq k} F_i(x) dx \quad (11)$$

The probability that x is the maximum is the sum over k that $X_k = x$ is the max, which is the same as the integrand above.

In our model, have $u_k = v_k - \alpha_W w_k$ where w_k is exponentially distributed with parameter λ_k . We assume without loss of generality that $\alpha_W = 1$. (For example, we can replace all λ_k by λ_k/α_W .)

The pdf and cdf for exponential variables are given by

- $F_k(u) = \exp(\lambda_k(u - v_k))$ for $u \leq v_k$ and 1 above, and
- $f_k(u) = \lambda_k \exp(\lambda_k(u - v_k))$ for $u \leq v_k$ and 0 above.

A.4.1 Choice probabilities

Assume that options are ranked such that $v_1 < v_2 < \dots < v_N$. Replacing the pdf and cdf in (10) and separating the integral by intervals delimited by the v_k 's, the probability that option k is optimal is:

$$\begin{aligned}
\Pr(k \in \arg \max) &= \int_{-\infty}^{\infty} f_k(u) \prod_{i \neq k} F_i(u) du \\
&= \lambda_k \int_{-\infty}^{v_1} e^{\lambda_k(u-v_k)} \times \prod_{i \geq 1, i \neq k} e^{\lambda_i(u-v_i)} du \\
&\quad + \lambda_k \int_{v_1}^{v_2} e^{\lambda_k(u-v_k)} \times \prod_{i \geq 2, i \neq k} e^{\lambda_i(u-v_i)} du \\
&\quad \dots \\
&\quad + \lambda_k \int_{v_{k-1}}^{v_k} e^{\lambda_k(u-v_k)} \times \prod_{i > k} e^{\lambda_i(u-v_i)} du
\end{aligned}$$

(For any $u > v_k$, the probability that k is optimal is zero.) The following notation will be useful: $\Lambda_i = \sum_{\ell \geq i} \lambda_\ell$ and $M_i = \sum_{\ell \geq i} \lambda_\ell v_\ell$.

We can re-write the above as

$$\begin{aligned}
\Pr(k \in \arg \max) &= \lambda_k e^{-M_1} \int_{-\infty}^{v_1} e^{u\Lambda_1} du \\
&\quad + \lambda_k e^{-M_2} \int_{v_1}^{v_2} e^{u\Lambda_2} du \\
&\quad \dots \\
&\quad + \lambda_k e^{-M_k} \int_{v_{k-1}}^{v_k} e^{u\Lambda_k} du
\end{aligned}$$

Evaluating the integrals we get:

$$\begin{aligned}
\lambda_k^{-1} \cdot \Pr(k \in \arg \max) &= e^{-M_1} e^{v_1 \Lambda_1} / \Lambda_1 \\
&\quad + e^{-M_2} (e^{v_2 \Lambda_2} - e^{v_1 \Lambda_2}) / \Lambda_2 \\
&\quad \dots \\
&\quad + e^{-M_k} (e^{v_k \Lambda_k} - e^{v_{k-1} \Lambda_k}) / \Lambda_k
\end{aligned}$$

Note that alternate terms from consecutive lines are the same. This is because

$$\begin{aligned}
-M_{i-1} + v_{i-1}\Lambda_{i-1} &= -\sum_{\ell \geq i-1} \lambda_\ell v_\ell + \sum_{\ell \geq i-1} v_{i-1} \lambda_\ell \\
&= -\sum_{\ell \geq i} \lambda_\ell v_\ell + \sum_{\ell \geq i} v_{i-1} \lambda_\ell \\
&= -M_i + v_{i-1}\Lambda_i.
\end{aligned}$$

Hence, we can rewrite the sum as:

$$\begin{aligned}
\lambda_k^{-1} \cdot \Pr(k \in \arg \max) &= e^{-M_1+v_1\Lambda_1} \left(\frac{1}{\Lambda_1} - \frac{1}{\Lambda_2} \right) \\
&\quad + e^{-M_2+v_2\Lambda_2} \left(\frac{1}{\Lambda_2} - \frac{1}{\Lambda_3} \right) \\
&\quad \dots \\
&\quad + e^{-M_{k-1}+v_{k-1}\Lambda_{k-1}} \left(\frac{1}{\Lambda_{k-1}} - \frac{1}{\Lambda_k} \right) \\
&\quad + e^{-M_k+v_k\Lambda_k} \frac{1}{\Lambda_k} \\
&= -\sum_{i=1}^{k-1} e^{S_i} \frac{\lambda_i}{\Lambda_i \Lambda_{i+1}} + e^{S_k} \frac{1}{\Lambda_k},
\end{aligned}$$

where we used the notation

$$S_k = -M_k + v_k \Lambda_k = \sum_{i \geq k} \lambda_i (v_k - v_i).$$

It is easy to see that $S_N = 0$ and S_k satisfies the (inverse) recursion

$$S_{k-1} = (v_{k-1} - v_k) \Lambda_k + S_k.$$

A.4.2 Expected Utility

Plugging the exponential pdf and cdf formulae in (11) we get

$$\begin{aligned}
\mathbb{E} \max_k u_k &= \sum_{i=1}^N \int_{v_{i-1}}^{v_i} u \sum_{k \geq i} f_k(u) \prod_{j \geq i, j \neq k} F_j(u) du \\
&= \sum_{i=1}^N \int_{v_{i-1}}^{v_i} u \sum_{k \geq i} \lambda_k \exp \left(\sum_{j \geq i} \lambda_j (u - v_j) \right) du \\
&= \sum_{i=1}^N \int_{v_{i-1}}^{v_i} u \Lambda_i e^{u \Lambda_i - M_i} du \\
&= \sum_{i=1}^N \Lambda_i^{-1} e^{-M_i} \int_{v_{i-1}}^{v_i} u \Lambda_i e^{u \Lambda_i} d(u \Lambda_i) \\
&= \sum_{i=1}^N \Lambda_i^{-1} e^{-M_i} \left[e^{\Lambda_i v_i} (\Lambda_i v_i - 1) - e^{\Lambda_i v_{i-1}} (\Lambda_i v_{i-1} - 1) \right]
\end{aligned}$$

Again noting that $-M_{i-1} + v_{i-1} \Lambda_{i-1} = -M_i + v_{i-1} \Lambda_i$ we can group terms together and write

$$\begin{aligned}
\mathbb{E} \max_k u_k &= \sum_{k=1}^{N-1} e^{S_k} \left(\frac{\Lambda_k v_k - 1}{\Lambda_k} - \frac{\Lambda_{k+1} v_k - 1}{\Lambda_{k+1}} \right) + e^{S_N} \frac{\Lambda_N v_N - 1}{\Lambda_N} \\
&= \sum_{k=1}^{N-1} e^{S_k} \frac{\lambda_k}{\Lambda_k \Lambda_{k+1}} - e^{S_N} \frac{1}{\Lambda_N} + v_N \\
&= v_N - \pi_N \alpha_W / \lambda_N,
\end{aligned}$$

where we again used the notation $S_k = -M_k + v_k \Lambda_k$, and the expression for $\pi_N \equiv \Pr(N \in \arg \max)$. This concludes the proof of part 2 in Proposition 3.

A.4.3 Expected Travel Time and Expected Wait Time

Travel time is non-random, so we can compute expected travel time simply using choice probabilities:

$$\mathbb{E} T_k^{\text{time}} \equiv \mathbb{E} (T_k^{\text{time}} \mid k \in \arg \max) = \sum_k \pi_k T_k^{\text{time}}. \quad (12)$$

By a similar argument, $\mathbb{E}v_k = \sum_k \pi_k v_k$. We use this result to derive expected wait time:

$$\begin{aligned}\mathbb{E}u_k &= \mathbb{E}v_k - \alpha_W \mathbb{E}T_k^{\text{wait}} \\ v_N - \pi_N \alpha_W / \lambda_N &= \sum_k \pi_k v_k - \alpha_W \mathbb{E}T_k^{\text{wait}} \\ \Rightarrow \mathbb{E}T_k^{\text{wait}} &= \pi_N \lambda_N^{-1} - \alpha_W^{-1} \left(\sum_k \pi_k (v_N - v_k) \right)\end{aligned}$$

A.5 Optimal network design

A.5.1 Optimization environment: driving and bus travel time

In this section, we describe how we estimate driving times and bus travel times (for BRT and non-BRT) at the level of each edge in the grid cell environment, so that these edge costs are consistent with our entire data on driving times and bus travel times.

Edge driving times are estimated based on our data on (o, d) driving times (section 2.3.5). We estimate the vector of edge travel times that best rationalize all these (o, d) driving times, assuming that the travel time between o and d is determined by a noisy shortest cost route between o and d , using the noisy routing model with Fréchet shocks from [Allen and Arkolakis \(forthcoming\)](#).

To compute edge bus times, we proceed in two steps. First, we seek to predict bus travel times between all (o, d) pairs. We start with our (o, d, r) data on median bus travel times between o and d on route r (only for o, d pairs where this data is available), and we use OLS to predict these travel times based on driving time data and distance between o and d . Second, we use the prediction for all (o, d) pairs and a similar procedure as for driving times (above). We repeat this exercise separately for BRT and non-BRT.

A.5.2 The Simulated Annealing (SA) Algorithm

Simulated annealing is defined by an objective function $W(N)$ over a finite set \mathcal{N} of allocations, the number of steps, T , initial and final “temperatures” $K_0 > K_T$, and a probabilistic neighbor suggestion function $nbs(N)$. Denote the algorithm by $SA(\mathcal{N}, W, T, K_0, K_T, nbs(\cdot))$.

The algorithm works as follows:

- Initialize with state N_0 at time $t = 0$.
- Temperature falls according to a preset schedule with T steps, between K_0 and K_T . We assume it falls exponentially, with the temperature at time t given by $K_t = K_0 \left(\frac{K_T}{K_0} \right)^{\frac{t}{T}}$.
- At time t , starting from state N_t , the algorithm proposes a modification to state $N_{t+1} = nbs(N_t)$. (Recall $nbs(N)$ is probabilistic, so there are usually multiple possible

values for N_{t+1} .)

- This modification is accepted with the following probability

$$\Pr(\text{accept } N_{t+1}) = \begin{cases} 1 & \text{if } W(N_{t+1}) \geq W(N_t) \\ \exp\left(\frac{W(N_{t+1}) - W(N_t)}{K_t}\right) & \text{otherwise.} \end{cases} \quad (13)$$

In words, at each step the algorithm proposes a new allocation based on the $nbs(\cdot)$ function. Allocations with higher welfare than the current value are always accepted. Allocations that lead to lower welfare are accepted with a probability that is decreasing in the welfare loss, and increasing in temperature. This means that initially in the algorithm, such transitions are more likely to be accepted, and the algorithm tends to travel nearly at random through the set of allocations. Later on during the algorithm, transitions are more heavily biased towards those that increase welfare or that do not decrease welfare very much.

A.5.3 The candidate network proposal algorithms

How to produce new candidate networks is critical for the success of the SA algorithm. We propose a nested structure for suggesting neighboring bus networks: first, the current network experiences a random modification drawn from a set of *global* modifiers, which we define below. Afterwards, we apply a sequence of randomly chosen *local* modifications, which are accepted only if they improve welfare.

The combination of one global change plus a subsequent series of local changes jointly produces our candidate neighboring network.

We use the following set of *global* modifiers:

1. Give or take away busses from a randomly selected bus route. Redistribute busses among randomly chosen other lines to stay at the constraint of 1,500 total busses.
2. Create a random new bus route. Pick two locations at random and create a bus route on the shortest path between these locations. Assign a random number of busses to the new route, redistributed from randomly chosen other bus routes.
3. Delete a randomly drawn existing bus route. Redistribute the busses among other, randomly chosen, bus routes.

and the following set of *local* modifiers

1. Exchange one bus between two randomly drawn bus routes.
2. Exchange a randomly drawn small fraction of busses between two randomly drawn bus routes.
3. Draw a bus route at random and add one random new adjacent stop to one end of the route.
4. Draw a bus route at random and take away one random stop at one end of the route.

5. Draw a bus route at random, pick two locations A and B on the route and “straighten” the route by replacing the intermediate stops by the shortest path between A and B .
6. Draw a bus route at random and add a detour to it. Pick two stops on the bus route, pick one new location on the map at random and let the route go between the two stops through the new location.

A.5.4 Obtaining the derivative in equation 7

$$\begin{aligned}
f^*(\theta) &= \sum_n \pi(n, \theta) f(n) \\
&= \sum_n \frac{\exp(\beta W(n, \theta))}{\sum_{n'} \exp(\beta W(n', \theta))} f(n) \\
\frac{\partial f^*(\theta)}{\partial \theta} &= \sum_n \frac{\partial \pi(n, \theta)}{\partial \theta} f(n) \\
&= \sum_n \frac{\sum_{n'} \exp(\beta W(n', \theta)) \beta W'(n, \theta) \exp(\beta W(n, \theta)) - \exp(\beta W(n, \theta)) [\sum_{n'} \beta W'(n', \theta) \exp(\beta W(n', \theta))]}{(\sum_{n'} \exp(\beta W(n', \theta)))^2} \\
&= \sum_n \frac{\beta W'(n, \theta) \exp(\beta W(n, \theta))}{\sum_{n'} \exp(\beta W(n', \theta))} f(n) - \sum_n \frac{\exp(\beta W(n, \theta))}{\sum_{n'} \exp(\beta W(n', \theta))} f(n) \frac{\beta \sum_{n'} W'(n', \theta) \exp(\beta W(n', \theta))}{\sum_{n'} \exp(\beta W(n', \theta))} \\
&= \sum_n \beta W'(n, \theta) \pi(n, \theta) f(n) - \sum_n \pi(n, \theta) f(n) \left[\frac{\beta \sum_{n'} W'(n', \theta) \exp(\beta W(n', \theta))}{\sum_{n'} \exp(\beta W(n', \theta))} \right] \\
&= \sum_n \beta W'(n, \theta) \pi(n, \theta) f(n) - f^*(n) \left[\frac{\beta \sum_{n'} W'(n', \theta) \exp(\beta W(n', \theta))}{\sum_{n'} \exp(\beta W(n', \theta))} \right] \\
&= \sum_n \beta W'(n, \theta) \pi(n, \theta) f(n) - f^*(n) \left[\beta \sum_{n'} \pi(n', \theta) W'(n', \theta) \right] \\
&= \sum_n \beta W'(n, \theta) \pi(n, \theta) f(n) - f^*(n) \left[\beta \sum_n \pi(n, \theta) W'(n, \theta) \right]
\end{aligned}$$