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

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

June 9, 2023

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 times, travel times, and transfers. We study these tradeoffs by examining the world's largest bus rapid transit system, in Jakarta, Indonesia, leveraging a large 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, idiosyncratic heterogeneity induced by random wait times, and inattention, matching event-study moments from the route launches. Commuters in Jakarta are 2-4 times more sensitive to wait time compared to time on the bus, and inattentive to long routes. To study the implications for network design, we introduce a new framework to describe the set of optimal networks. Our results suggest that a less concentrated network would increase ridership and commuter welfare.

*We are grateful to TransJakarta for sharing their data. We thank David Atkin, Emily Breza, Dave Donaldson, Allan Hsiao, Myrto Kalouptsidi, Natesh Pillai, Felix Tintelnot, and Aaron Smith for useful conversations, Nick Tsivanidis and Matthew Turner for useful discussions of this paper, and participants at the NBER Summer Institute urban meeting and seminar participants for useful feedback. Lolita Moorena, Nadia Rayhanna, Nadia R Avianti, and Nikhil Kumar provided exceptional research assistance. We thank Paolo Adajar, Aaron Berman, Robert Dulin, Laya Gollapudi, Deivy Houeix, Helen Hu, Drew Johnston, Moulinrouge F Kaspar, Ariba Khan, Van Anh Le, Lauren Li, Anna Lingyi Ma, Kalyan Palepu, Jackson C Phifer, Xiaocun Qiu, Margaret Sands, San Singh, Yulu Tang and Hoi Wai Yu for research assistant work at different stages of the project. We are grateful to Kyle Schindl, Analysia Watley, Aditi Chitkara, and Nikhil Kumar for their help in developing the bus GPS headway engine. The project has IRB approval from MIT COUHES, protocol 1810556772. This material is based upon work supported by the National Science Foundation under Grant No. 2049784, by the Australian Government, the International Growth Centre, the Harvard Asia Center, and the Harvard Data Science Initiative.

†Kreindler: Harvard University (gkreindler@g.harvard.edu), Gaduh: University of Arkansas (agaduh@walton.uark.edu), Graff: Harvard University (tgraff@g.harvard.edu), Hanna: Harvard Kennedy School (rema_hanna@hks.harvard.edu), Olken: MIT (bolken@mit.edu).

1 Introduction

Governments are investing heavily in public transport in the megacities of low- and middle-income countries (LMIC), giving city governments a growing role in planning and operations. For example, the number of LMIC cities with a Bus Rapid Transit (BRT) system went from 14 cities in 2000 to 100 in 2020 (BRT+ Centre of Excellence and EMBARQ, 2023). Bus networks are particularly important because they account for a large share of ridership. For example, Delhi's public bus system carried 3.3 million passengers per day in 2017, much more than its more famous metro system (2.7 million, Hindustan Times 2018). Ensuring that these transport systems are well-designed is key to sustainable and efficient mobility.

Designing bus networks involves tradeoffs. Because the fixed cost of a bus route is relatively small (compared to, say, building a subway or light-rail line), there is great latitude in designing a system. For a given number of buses, one can choose between having a more *direct* network (i.e., more direct routes, rather than a hub-and-spoke system that requires more changes between routes); a more *intensive* route network (i.e., more frequent service to key locations); and a more *expansive* route network (i.e., many routes serving more destinations).

Fundamental to these tradeoffs is an understanding of public transport preferences and demand, which in turn affects bus ridership: how commuters value bus travel time, wait times in stations, and transfers between routes. Different preference parameter configurations may result in very different-looking optimal transportation networks. Estimating these preferences is challenging, however, because it requires plausibly exogenous variation in *system-level* attributes.

In this paper, we study the expansion of the TransJakarta bus system in Jakarta, Indonesia. TransJakarta operates a public Bus Rapid Transit (BRT) network, which is the largest such network in the world (Institute for Transportation and Development Policy, 2019). From 2016-2020, TransJakarta launched a total of 93 routes across the city in a staggered fashion for both BRT routes and non-BRT (feeder) routes. These new routes changed the attractiveness of bus travel in different ways for different potential journeys, which we use to estimate preferences and study the implication for network design.

Our paper proceeds in three steps. First, we estimate the reduced form impact on bus ridership and overall commuting flows from improvements to the TransJakarta network. To do so, we show how the 93 new bus routes create different types of improvements in the availability of direct connections, travel times, and wait times for different origin-destination pairs, and analyze these 'events' using a differences-in-differences framework. Second, we estimate a model of commuter travel behavior, which allows us to back out the underlying

preference parameters that match the reduced-form estimates from step 1. Third, we simulate what optimal transportation networks would look like given the preferences from step 2, and contrast their characteristics with that of the existing network.

For the first step, we define three types of 'events' that correspond to different types of service improvement for origin-destination (o, d) pairs. The first two events focus on new direct routes for (o, d) pairs that were previously served only via transfer connection. The first event type (hereafter, Event 1) considers cases where the total time on the bus is unchanged with the new direct route. This introduces variation in the degree to which the network is *direct*. Event 2 considers new direct routes where the total time on the bus also falls, creating additional variation in service *speed* in addition to becoming more direct. Event 3 considers (o, d) pairs that receive a new direct route that overlaps with the old direct route(s) from o to d. This leads to more frequent service, varying the degree to which service is *intensive*, without substantially changing time spent on the bus or the directness of the route.

We use several new datasets to analyze these events. To measure (o, d) ridership flows on the bus network, we use TransJakarta's detailed administrative data for both BRT routes and non-BRT feeder routes. TransJakarta collects fares using a tap-to-pay smart card system and our data includes every tap — over 500 million taps in the study period — along with an anonymized card identifier.[1] Second, we measure bus travel times and the distribution of wait times that passengers face using GPS data that tracks the position of every TransJakarta bus every 5-10 seconds. Third, we measure overall commuting flows on each (o, d) pair (regardless of transportation mode) using anonymized smartphone location data. We aggregate the outcome data at the weekly level and at the level of 1km-wide hexagonal grids, and show robustness to other grid shapes and sizes.

Using a unified difference-in-differences design, we find that bus ridership is responsive to each of these three types of service quality improvements.[2] For BRT routes, a new direct route between an existing (o, d) pair increases ridership by 0.16 log points. Event 2, which substantially reduced travel times in addition to a new direct route, increased ridership by 0.27 log points. Event 3, which reduced wait times by 0.32 log points (through adding busses via an incidental new route), increased ridership by about 0.09 log points, implying that 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 results for

[1] 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 each trip's likely destination. We validate this approach using the subset of trips for which we directly observe both the origin and destination.

[2] We include a large set of "never-treated" origin-destination pairs, which largely alleviates the class of problems with two-way fixed effects specifications (see, e.g., De Chaisemartin and D'Haultfoeuille 2022).

non-BRT routes are typically larger in magnitude in proportional terms and similar in levels than the BRT routes. Our results are stable when varying grid cell size from 500 meters to 2 kilometers, implying that the events lead to new bus ridership rather than displacing passengers from neighboring TransJakarta stations.

We next look at the impact of the same set of events on the aggregate volume of trips between a pair of locations (regardless of whether people take the bus or an alternative), measured using the smartphone location data. We do not find observable increases in all trips for any of the events in our benchmark specification. In fact, we typically reject moderate positive effects. In short, the results imply substantial increases in ridership when service frequency improves on all three dimensions — directness, speed, and wait times — and with this being driven by substitution towards bus trips, rather than by completely new journeys.[3]

With our reduced form estimates in hand, in the second step of our paper, we develop a model of commuter travel behavior to estimate underlying preference parameters. The model has a new formulation of how commuters choose routes within a geographically realistic transit network, in the presence of stochastic wait times. A commuter traveling from o to d chooses between direct and transfer bus options that differ in terms of type (BRT or non-BRT), total travel time on the bus, and the necessity of a transfer. Buses on each route arrive according to a Poisson process. A commuter who decides to take the bus gets a draw of wait times for all routes in her choice set and selects the best option overall. Because the idiosyncratic term in the model comes from random wait times given by Poisson processes, the model is invariant to aggregating identical routes, a property that allows us to bypass spurious gains from variety when we return to the optimal network design problem.[4]

Importantly, commuters sometimes face large choice sets with dozens of options within TransJakarta's network, many of which are too slow to consider. Therefore, we allow for (and estimate the degree of) partial inattention to bus route options that have high travel time on the bus, relative to the quickest bus option in the choice set (Larcom et al., 2017)

In a higher nest, commuters choose between using the bus network, taking expectations over the wait times they may face, or a private transportation outside option (e.g. motorbike, car), based on a logit specification.[5]

[3]Converting road traffic lanes to dedicated bus lanes can increase road traffic congestion (Gaduh et al., 2022), yet the TransJakarta network expansion we study here did not reduce road traffic lanes. The only major change to BRT infrastructure in this period was the launch of corridor 13, which runs along an elevated busway so did not reduce road traffic lanes.

[4]In typical discrete choice models such as logit, splitting a bus route into two identical routes with half the original bus allocation is the source of an expected utility gain because the new routes have independent idiosyncratic errors; this is the famous 'red-bus, blue-bus' problem discussed by McFadden (1974). We show that the general version of our model, which appears in Daganzo (1979) under the name "negative exponential distribution model," does not have this issue.

[5]While it is possible to further embed the model into a destination-choice nest or in an urban general

We estimate the model for the discretized Jakarta metropolitan area using 1km hexagonal grid cells. We use smartphone data to estimate commuting flows across all possible combinations of these grid cells. We then use the model to compute predicted ridership over each (o, d) pair and examine how it changes over time as the bus network changes, by running the same reduced form analysis as in the actual data. This allows us to estimate the preference and attention parameters that match the actual event-study estimates.

Using this approach, we find a high disutility for waiting time relative to time traveling on the bus. Specifically, we estimate that travelers find wait time 2.4× more costly than travel times for BRT, and 4.2× most costly for non-BRT. Commuters dislike bus options that involve transfers due to the additional wait time and typically longer time on the bus. However, there is no additional transfer penalty, above and beyond the wait time and time on the bus costs. Finally, we find that commuters pay significantly less attention to bus options with bus travel time more than 34%-44% longer than the quickest option in their choice set.

What do these preferences mean for the optimal design of the network? In the third part of the paper, we introduce a theoretical and computational framework that allows us to characterize optimal networks, compare them with the current network, and perform comparative statics with respect to structural model parameters.

We ask how the network of bus routes should be configured across the city and how the existing set of buses should be allocated to these routes to maximize a utilitarian welfare objective informed by our estimated demand model. This involves balancing tradeoffs across reach (having enough bus stops to cover origins and destinations for which people would like to travel), wait times (putting more buses on a given route), and the topology of the network (which stations should be connected directly versus by transfer). We hold fixed TransJakarta's current BRT infrastructure, which allows for faster travel times along the 13 BRT corridors, but we allow routes to enter and exit the BRT corridors, to connect them in new ways, and to travel on existing streets. We study this problem using realistic geography of Jakarta at the levels of 2km-wide square grids.

Our problem displays divergent substitution and complementarity forces at the same time, hence violating the type of complementarity property that enables the characterization of optimal allocations in other settings (Jia, 2008; Arkolakis and Eckert, 2021). This is also a very high-dimensional problem, and the number of possible configurations is extremely large, making it difficult to identify the (generically unique) welfare-maximizing network.[6]

equilibrium model (Tsivanidis, 2022), the null reduced form effect on aggregate trips suggests that over the period we study, the preferences we estimate over wait time, transfers, and travel time are not biased by changes in aggregate trip patterns.

[6]With 418 grid cells and 1,536 possible edges, there are $2^{1,536}$ configurations where each edge is either

Therefore, we set up a slightly modified problem: the social planner chooses a network N that maximizes utilitarian welfare $W(N, \theta)$ from our estimated demand model — average expected utility for all commuters — plus an idiosyncratic shock ϵ_N for each possible network. The shocks capture network-level factors that the social planner cares about, such as cost shocks or benefits from certain networks, outside of our demand model. As such, in our framework, the social planner has a discrete choice problem over the space of all possible networks. A network with high model welfare carries a large probability of being chosen, because small idiosyncratic shocks may tip the planner into preferring it. We parametrize the problem as a multinomial logit with parameter β.

This approach allows us to transform a global optimization problem into a problem of sampling from a large-scale multinomial logit distribution. Several algorithms based on the Metropolis-Hastings algorithm have theoretical guarantees and good practical performance in sampling from such distributions.[7] We derive formulas for properties of optimal networks and comparative statics. We can estimate these objects using their sample analogs. Our framework is general and may be used in other settings where a social planner's (or an agent's) problem is high-dimensional.[8]

Using this approach, we find that optimal networks, based on the preference parameters that we estimated, are substantially more extensive than the actual network. For example, we estimate that a typical optimal network should cover about 66 percent of all grid cells, compared to 42 percent with the actual TransJakarta network. 91% of all trips by Jakarta residents would have access to the bus network, compared to 73% in the current network.[9] Bus frequency in the city center is slightly lower in the optimal network, so commuters in those areas need to wait longer for a bus than in the current network. While our estimated parameters indicate that commuters are highly sensitive to wait times, we find that the current TransJakarta network nevertheless concentrates too many buses in the city center, where wait times are already very short.

In our last exercise, we explore how certain characteristics of the estimated optimal networks would change if preferences differed. For example, doubling the wait time cost leads to more concentrated networks connecting only 23 percent of origin-destination pairs,

connected or not – i.e. about 10^{500} possibilities. This is a lower bound of just the route design problem, and does not even account for the number of buses on each route.

[7] We use a version of the simulated annealing (SA) algorithm, which we run multiple times to obtain a sample of networks drawn (asymptotically) from the planner's distribution.

[8] While the details of the implementation of the sampling algorithm will be application-specific, the theoretical framework can be applied to problems such as optimal spatial policy (Fajgelbaum and Gaubert, 2020), optimal transportation infrastructure Allen and Arkolakis (2022); Fajgelbaum and Schaal (2020), and firms' global sourcing decisions Antràs et al. (2017). The setting is appropriate when the welfare difference (or ratios) can be computed for pairs of counterfactuals, for example, if exact hat methods are applicable.

[9] Our results do not suggest a tension between distributional and efficiency concerns in this setting.

41% less than before. Increasing the transfer penalty by the equivalent of 15 minutes of wait time increases the share of connected location pairs that have a direct connection from 12% to 16%. This type of comparative static exercise connects micro parameters in the model to macro properties of optimal bus networks, which helps uncover which parameters matter for network design and to assess sensitivity of our results to the parameters to estimate.

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 (Mohring, 1972).[10] We build on the literature investigating how commuters value trip attributes, see e.g. Wardman (2004); Currie (2005); Abrantes and Wardman (2011) and Small et al. (2007, page 53). Studies in this literature tend to focus on high-income countries and often infer stated preferences from respondents' choices between hypothetical alternatives. In contrast, we estimate preferences using natural experiments. In that respect, our study is closest to Kreindler (2023), which uses a field experiment to estimate the underlying preference parameters required to calculate optimal congestion pricing in Bangalore, India. Here, by contrast, we use a variety of real-world changes in the attractiveness of the network given by different types of network expansion for identification. Our study is also related to Buchholz et al. (2020) and Goldszmidt et al. (2020), both of which estimate the value of wait time for taxis and ride-hail from choices made on ride-hail platforms in Europe and the United States.[11]

Second, by embedding 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, 2022; Balboni, 2021; Santamaria, 2022; Alder, 2023). This literature has so far focused on road infrastructure (typically inter-city), rather than urban transit.[12] Our detailed microdata allows us to estimate fine parameters that are usually hard to incorporate in these studies, and shows how they are related to features of optimal networks.

Our paper is also related to the literature on the impact of transit systems in LMIC. Gaduh et al. (2022) studies the impact of the initial launch of the TransJakarta system in 2002 to show that at that point in time, the system had no impacts on public transport use but the conversion of road lanes into dedicated lanes for the BRT network worsened road

[10] See also Coulombel and Monchambert (2019) for evidence on dis-economies of scale.

[11] Hörcher et al. (2017) estimate subway crowding costs in Hong Kong with a revealed preferences approach.

[12] Fajgelbaum and Schaal (2020) study a neoclassical trade model where the planner's problem is globally concave. Allen and Arkolakis (2022) study a tractable general equilibrium model with routing and congestion. Balboni (2021) and Santamaria (2022) compare different road investment plans in dynamic settings. Alder (2023) introduces a heuristic algorithm for finding the globally optimal highway network in the context of a quantitative trade model for India.

traffic congestion. Majid et al. (2018) study the impact of a new BRT transit line in Lahore, Pakistan. Tsivanidis (2022), Zárate (2023), and Balboni et al. (2021) measure the impact of the BRT systems in Bogotá, Mexico City, and Dar Es Salaam, respectively, and quantify general equilibrium effects through the lens of quantitative urban models. Tsivanidis, for example, focuses on estimating the welfare and inequality effects of the system and how it affects the organization of the city. 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.

Meta-heuristic algorithms, including the simulated annealing algorithm that we use here, have been used extensively in the transportation and operations research literatures to *approximate* optimal designs in urban transit network design problems (Wei and Machemehl, 2006; Iliopoulou et al., 2019). Our contribution is to link in an asymptotic sense the output of such algorithms to a social planner's discrete choice problem over all possible networks. Instead of approximating an optimal network, we sample multiple networks, which allows us to estimate optimal network characteristics and comparative statics.[13]

The rest of this paper is organized as follows. Section 2 describes the TransJakarta bus network, its expansions, and our data. Section 3 presents the reduced-form event-study results that show how ridership responds to new routes and improved bus frequency. Section 4 introduces the model, describes how we estimate the model using the reduced form moments, and presents estimation results. Section 5 explores the implications for optimal network design.

2 Setting and Data

2.1 The TransJakarta Bus Network

Setting Description. TransJakarta operates an integrated bus system serving the central urban districts in Greater Jakarta — a metropolitan area of about 18 million people. Its routes are concentrated primarily within the city of Jakarta, though some extend to surrounding districts. Established in 2004, TransJakarta serves hundreds of thousands of passengers daily, with daily ridership peaking around 1 million in early 2020.

As of August 2019, TransJakarta has over 139 bus routes, which are a mix of Bus Rapid Transit (BRT) and non-BRT routes. BRT routes operate on dedicated bus lanes along designated stations. Passengers pay at turnstiles to enter the BRT station and board the bus. It has a network length of more than 120 miles of BRT corridors, the longest such

[13] Daganzo (2010), Fielbaum et al. (2016) characterize optimal transit networks in idealized environments.

system in the world (Institute for Transportation and Development Policy, 2019). Non-BRT feeder routes operate partially on normal city streets, and partially along the BRT corridors, where they also stop in BRT stations. These non-BRT routes connect locations that are further away from the BRT system.

TransJakarta costs 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.[14] Free transfers are allowed at BRT stations.

TransJakarta is the primary public rapid transit system in the city. A commuter rail system (KRL CommuterLine) with about 1 million ridership serves outlying areas for longer trips that rarely overlap with TransJakarta trips. In 2019, a single 16 km subway line was opened to serve about 80,000 riders per day. In 2018, privately provided shared transport, using minibuses, accounted for 2.9% of all trips in the larger Jabodetabek region (JUTPI2, 2019). The primary alternative to TransJakarta is private transport, consisting of a mix of motorcycles, private cars, motorcycle taxis (*ojek*), and automobile taxis.

TransJakarta Network Expansions. We study the large expansion of the TransJakarta network, which added more than 93 BRT and non-BRT routes between January 2016 and February 2020. Only one entirely new BRT corridor was launched during this period.[15] The newly launched BRT routes run along existing BRT corridors, sometimes connecting two corridors or running an express route on certain sections. The new non-BRT feeder routes stop both in some BRT stations as well as in roadside non-BRT bus stops, connecting one or multiple BRT corridors to other areas of the city.

The number of operating buses more than doubled during this period, from about 700 to more than 1,600. Figure 1 shows the expansion of the system. These expansions — for both BRT and non-BRT routes — took place at different times throughout the city.

New lines were launched based on a mix of external constraints and discretion. All non-BRT route launches were chosen from a list of routes that were pre-approved in 2016 by the Jakarta transportation department (*Dishub*). For BRT routes, TransJakarta created new routes by connecting existing BRT stations along the existing BRT dedicated lanes. For both BRT and non-BRT, TransJakarta chose routes and launch periods based on new bus fleet delivery dates, bus availability based on bus-operator contracts, inputs from field reports, and other factors. Appendix Table A.2 shows that the order of BRT route launch is balanced with respect to the geographic characteristics of the routes.

[14] Cash payments are allowed on non-BRT busses, accounting for around half of all transactions in 2019.
[15] Corridor 13 is an elevated busway, thereby not affecting the number of lanes available to road traffic.

2.2 Data

We use three different datasets: **TransJakarta ridership** data from administrative data on TransJakarta's smart card transactions; origin-destination **aggregate-trip flows** data from anonymized smartphone location data; and **bus locations** data, which we obtain by processing detailed GPS data on every TransJakarta bus throughout our period.

We focus on the central urban districts of Greater Jakarta.[16] We use consistently defined geographical environments throughout our analysis. We divide space into identical grid cells and aggregate stations at the grid cells level, and trips at the grid cell pair level.[17] Throughout the reduced form and demand estimation analysis, We use a regular hexagonal tiling with 1766 grid cells where adjacent grid-cell centroids are 1,000 meters apart, and then show robustness to using 500-meter and 2000-meter square grid cells. For the optimization section, we use 2,000-meter square grid cells to reduce the problem's dimensionality. We aggregate outcomes (e.g. ridership, trips, bus supply) at the weekday or work-week level, focusing on the 5AM to 10PM interval.

Ridership Data. We use the administrative ridership data that is captured electronically by smart cards to construct a highly granular TransJakarta origin-destination (o, d) ridership matrix at each point in time since 2016 (since 2017 for non-BRT). In BRT stations, passengers tap smart cards at turnstiles to enter the boarding area. In 36 percent of BRT stations, passengers also tap smart cards upon exit. In non-BRT bus stops, passengers board the bus directly from the street and tap to pay for the bus ticket upon boarding. 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.[18]

While we observe each passenger's trip origins, we only observe destinations (d_{it}) for the trips that end at stations where tap-out is enforced. In all other cases, we construct a proxy destination station based on an algorithm that uses the origin station of the next trip for the same smart card, or another frequent origin for that card. We validate this algorithm using the existing tap-out data.[19] We describe the algorithm and ridership data cleaning process and coverage in Appendix A.3.4.

[16] This consists of the Special Capital Region of Jakarta (DKI Jakarta) and all the surrounding urban districts: Tangerang, South Tangerang, Depok, and Bekasi (see Figure A.1).

[17] We collapse the TransJakarta network to the level of grid cells. For transfer options, we consider that origin o and destination d are connected by a transfer option by routes r_1 and r_2 if these two routes intersect in an actual station.

[18] Individuals can purchase new smart cards, but do so infrequently. The median smart card in our data is active for over 4 months, and the median tap belongs to a smart card that is active for over 20 months.

[19] For the set of destination stations where we have actual exit transactions, bivariate regressions of imputed and actual daily ridership shares have $R^2 = 0.85$ (Appendix Figure A.19).

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 or not. We have daily smartphone location data from March 2018 through March 15, 2020 from Veraset, a private data provider. These data cover 35 million weekday trips that belong to 2.3 million unique devices in our study area. (For reference, our study area had 14 million individuals over 15 years old at the 2020 census.) We use these data to: (1) create a panel of trip flows separately for each week in the data as an outcome variable when studying the impact of the network's expansion, and (2) construct a cross-section of typical (o, d) trip flows throughout the period in order to compute choice probabilities for using TransJakarta. See Appendix A.3.5 for details.

The smartphone location data captures rich travel patterns within Jakarta. Figure A.21 shows that the Veraset data is broadly representative across areas with different levels of poverty and density.

Bus Location and Allocation, and Network Expansion. We have data on the planned and realized daily bus allocations for each route, as well as exact GPS locations for every bus every five to ten seconds for the near-universe of routes. These data cover approximately 1,800 busses and 16,000 bus trips per day since January 2017. We process this raw GPS data to compute median bus travel times on all routes, between every origin and destination stations, and to measure the empirical distribution of wait time for each route. See Appendix A.3.2 and A.3.3.

To reconstruct the TransJakarta network at different points in time, we combine data from TransJakarta and data from a mobility planning app available in Jakarta during the study period. We cross-check launch dates inferred from bus allocation data with route- and date-specific TransJakarta ridership data, as well as directly with TransJakarta staff.

3 Reduced Form Results: the Impact of Service Quality

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 the launch of a new bus route changes different dimensions of how attractive it is to take TransJakarta for different (o, d) pairs, based on how it affects the choice set of direct and transfer bus routes between those locations.

We focus on three types of 'events' induced by new route launches. The first two capture the launches of the first *direct* route between two locations that are already connected by *transfer*. In Event 1, the new direct route is not faster than the fastest existing transfer

connection, in terms of total time on the bus. In Event 2, the new direct route is faster. Event 3 captures the addition of busses between two already directly connected locations because the new route overlaps with the existing direct route(s) between o and d. This increases the bus arrival rate and lowers wait times for those traveling from o to d. We examine these event types separately for BRT and non-BRT connections.

3.1 Estimation Framework

For each of the three event types described above, for both BRT ($E \in \{1B, 2B, 3B\}$) and non-BRT network expansions, ($E \in \{1N, 2N, 3N\}$), we estimate the following equation:

$$\ln(\mathbb{E}Y_{odt}) = \alpha^E Post^E_{odt} + \alpha^E_{-10} L^E_{\leq -10,odt} + \alpha^E_{10} L^E_{\geq 10,odt} + \mu^E_{od} + \nu^E_{ot} + \xi^E_{dt} + \varepsilon^E_{odt} \tag{1}$$

Our dataset is at the origin (o) $\times$ destination (d) $\times$ week (t) level, so we include all two-way fixed effects. Specifically, μ^E_{od} are origin$\times$destination fixed effects — i.e. fixed effects for every combination of start and end grid cell — and ν^E_{ot} and ξ^E_{dt} are origin$\times$time and destination$\times$time fixed effects. These fixed effects flexibly capture differences in ridership across origin-destination parts, as well as arbitrary time-related shocks for each origin and destination.

The key variable of interest is $Post^E_{odt}$, a dummy variable for the event having taken place on the od route in the previous 10 months.[20] We control for $L^E_{\leq -10,odt}$ and $L^E_{\geq 10,odt}$, 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, d) in the first 10 months after it occurs, relative to the 9 months prior to the event.

We estimate (1) using robust Poisson Pseudo Maximum Likelihood (PPML). We cluster standard errors two-way by both origin and destination, which allows for arbitrary serial correlation over time for each origin and each destination, as well as arbitrary correlation among destinations for a given origin, and vice-versa (Cameron et al., 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. The sample of pairs od for event type E is restricted to all origins treated at least once ($Post^E_{od't'} = 1$ 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 78.4 percent).[21]

[20] For Event 3, we focus on the first event if there are multiple events for a given o, d pair.

[21] The fact that most of our od pairs are never treated largely alleviates the class of problems with two-way

We use two outcome variables Y_{odt}. First, we examine the total TransJakarta ridership between grid cells o and d during week t, summing up over all station pairs in o and d. Second, we also measure impact on all trips between o and d in week t, computed based on smartphone location data.

We use the full time period of data that we have for each type of event, but since the BRT, non-BRT, and smartphone trip data all begin on different dates, the data start at different times.[22] 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.

In Events 1 and 2, a route launch creates the first direct connection between o and d, a location pair that already has a transfer connection. The two event types differ only in terms of whether the new direct route has faster travel time on the bus compared to the fastest pre-existing (transfer) option.

In Event 3, grid-cell pairs o and d that are already directly connected get more busses from a new route because it overlaps with the existing route for the portion between o and d. Specifically, $Post^3_{odt}$ 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.

3.2 "First Stage" Impacts on Travel and Wait Times

We first study how the three different events affect the attractiveness of riding the TransJakarta network from o to d (Table 1).

We first estimate the impact on the log minimum travel time on the bus between o and d. We use the bus GPS data to measure bus travel times, and at each time t between 2016 and 2020 we take the minimum over all available bus options (direct or single-transfer). This measure captures the pure travel time on the bus — i.e. it does not include any time spent waiting for a bus to arrive or any time waiting for a connecting bus.

Adding a new direct line between o and d reduces travel time, but only when the new line is faster than the existing transfer connection (Event 2), yielding an average effect of about 0.29 log points with a standard error of 0.027 (column 2). While the fact that travel time reductions are limited to Event 2 is, of course, mechanical to the way we define the events,

fixed effects specifications that have been highlighted recently (see, e.g., De Chaisemartin and D'Haultfoeuille 2022), which are primarily a concern when using previously-treated observations as a comparison group.

[22] For BRT events, the sample comprises January 2016 to mid-March 2020. For non-BRT events, we further restrict the sample 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 from March 2018 to March 2020. We always exclude May-July 2018 due to missing BRT ridership data.

we show that the differences are quantitatively large and statistically quite meaningful.[23] Event 3 has a very small effect on log travel time.

Next, we examine the log number of busses arriving per hour at the origin, over all the direct and transfer options that connect the origin and destination. This is a proxy of waiting times at the origin station. Event 3 increases the bus arrival rate by 0.32 log points (Table 1, column 6). Events 1 and 2 have small effects on the bus arrival rates at the origin. Note that these effects do not capture the reduction in wait time for transfer connections.

Results for non-BRT network expansion are broadly similar and typically larger in magnitude (Table 1, panel B). Importantly, baseline bus arrival rates for treated origin-destination pairs is 2–4 times lower for non-BRT routes compared to BRT (columns 4–6). This means that the absolute reduction in wait time is much larger for non-BRT events compared to BRT events. Results are qualitatively similar when using 500-meter or 2000-meter square grids as the base geography (Tables A.3 and A.4).

The key point from Table 1 is that the three events affect the desirability of using the bus system in different ways: adding a new and faster direct line reduces travel time *and* reduces waiting times, in addition to eliminating the need to transfer, whereas additional busses mainly reduces waiting times.[24] This variation will allow us to identify preference parameters in Section 4.

3.3 The Impact of New Route Events on Bus Ridership

Travel Times and Direct Connections: Impact of New Routes. The impacts of the three events on bus ridership, estimated using equation (1), are presented in columns 7–9 in Table 1 for BRT and non-BRT. Figure 2 shows event-study versions of each of these equations with monthly lags and leads from the date of the event.

Ridership is highly responsive to improvements in service quality. On average, adding an additional direct BRT route between o and d without a travel time improvement (Event 1) leads to an increase in ridership of 0.16 log points over the 10 months following the introduction of the new route (Table 1, column 7). When the new direct route also decreases travel time (Event 2), which column 2 showed led to a 0.29 log point reduction in travel time, this leads to an increase in ridership of 0.27 log points.

[23] The first type of event in fact has a small positive effect on log travel time. This is to some degree mechanical based on how we separate events 1 and 2. New direct routes are categorized under event 2 if the new direct route's travel time is strictly smaller than that of existing transfer routes, and event 1 otherwise.

[24] Both "first stage" variables are imperfect 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. The minimum travel time between o and d shown in Table 1 might depend on a very infrequent bus line. The model in section 4 will lay out how commuters value these attributes jointly for all bus options in their choice set.

The event-study version of equation (1), shown in Figure 2, Panel (a), left column, shows no pre-trends before the first and second events, and a discrete uptick in ridership at the time of the new route launch. The increase is notably higher for Event 2. The event-study graphs also show that riders take some period to adjust to the new route.

The non-BRT results paint a qualitatively similar picture of the impact of bus service improvements (right-side of Figure 2 and Table 1, Panel B). While for all events, the point estimates are larger than those for BRT, the implied *level* effects are much more similar. For Event 1, the BRT effect implies that a treated origin, destination pair has 19.1 ($= 111.3 \cdot (e^{0.158} - 1)$) more riders per week on average in the post period. This compares with an effect of 33.8 additional riders per week for non-BRT Event 1. For Event 2, the BRT and non-BRT level effects are 23.6 and 20.6 additional riders per week, respectively.

Wait Times and the "Mohring Effect." We next investigate the addition of direct busses between o to d caused by the launch of a new route that overlaps with existing direct routes (Event 3). Adding additional busses leads to a 0.09 log point increase in ridership (Table 1, column 9). Figure 2, Panel (b) shows that there is a large, discrete uptick in ridership exactly when the new route is introduced.

Given the first-stage results discussed above, we can interpret Event 3 as being almost entirely about the effect of wait times on ridership. Thus, combining the estimates from columns 6 and 9 of Table 1, we get an implied elasticity of ridership with respect to wait times of -0.29, i.e. a 10 percent decrease in wait times leads to a 2.9 percent increase in ridership.

These estimates speak to the so-called 'Mohring effect.' Mohring (1972) 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 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.

An extreme form of the Mohring effect is when the elasticity of ridership with respect to wait times is greater than 1 in absolute value over some range. 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. For BRT, we can reject an elasticity of -1 at conventional significance levels.

For non-BRT routes, however, we find a much larger implied elasticity of ridership with respect to wait times, of -1.05 ($= 0.450/0.425$, Table 1, panel B). This suggests that adding more non-BRT frequency on some non-BRT routes could increase ridership enough to main-

tain (or even increase) *average* ridership per bus.

Similar to the first two types of events, the *level* impacts of Event 3 are similar for BRT and non-BRT. Specifically, the BRT effect implies that a treated origin, destination pair has 20.3 ($= 210.4 \cdot (e^{0.092} - 1)$) additional riders per week on average in the post period, relative to an effect size of 14.77 for non-BRT.

Alternative Aggregation Levels. One possible concern is that our results are capturing displacement effects within the TransJakarta network. This would be the case if, for example, some commuters would switch from using a certain origin bus station located in a grid cell o' to one of the treated origin grid cells o after an event.

To examine this, we consider different aggregation levels, where we re-estimate all our results using both 500-meter squares (smaller) and 2000-meter squares (larger). If substitution was a central concern, we would find much larger effects when we use smaller units of aggregation than when we use larger units. Figure A.5 compares the magnitude of coefficients for all three events, separately for BRT and non-BRT, placing estimated $Post_{odt}$ coefficients and their 95% confidence intervals side by side.[25] There is no systematic pattern of decreasing coefficients and confidence intervals overlap significantly in all graphs. This implies that our treatment effects are not significantly contaminated by displacement effects.

Heterogeneity Based on Poverty. One important question is whether preference parameters depend on local poverty levels. Table A.6 shows the impact of all six events on bus ridership when we include interactions for whether the origin grid cell has above-median poverty rate in the estimation sample; these regressions also control for interactions with log population. We use poverty and population data at the kelurahan (urban villages) from SMERU (2014) and the PODES 2010 village survey, respectively. The results show no clear patterns based on income levels. It is important to note the lack of heterogeneity here when thinking about optimal network design.

3.4 Impacts on Overall Trips

Having shown how the TransJakarta ridership responds to these events, we next turn to the Veraset smartphone-based trip data to examine whether *aggregate trip volume* — regardless of whether passengers take TransJakarta or not — is responsive to the TransJakarta network expansions.

We re-estimate equation (1) for the total number of trips from o to d at time t from Veraset (Table 2 and Figure 3). Our measure of aggregate trips is derived from individual-

[25] The detailed 500 and 2000 meter results are shown in Tables A.3 and A.4 and Figures A.2 and A.3.

level smartphone traces, and it covers all types of trips, not only commuting trips between home and work. Since our Veraset data only begins in March 2018, for comparability we also re-estimate the effects on TransJakarta ridership for the same time period.

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 1, we find a coefficient of -0.008 with a standard error of 0.051 (column 10). This allows us to reject at the 95% level a positive impact of approximately $+0.091$, compared to the precise 0.11 effect on bus ridership in the same sample (column 7). Thus, we can 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 most event types, and they are robust to analyzing the data at the level of smaller, 500-meter square grids (Table A.5).

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 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.

Overall, the reduced form results suggest that riders are responsive to the three dimensions of service quality we consider: wait times, on-the-bus ride times, and direct connections. We do not find evidence that these changes affect the aggregate volume of trips. Since the events we study affect multiple dimensions of service quality, we next estimate a model of commuter demand, which allows us to jointly infer the underlying preference parameters that best match the responsiveness we observe in the data.

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.[26] We consider the problem of a commuter traveling between an origin grid cell and a destination grid cell who chooses between direct bus routes that link the origin and destination grid cells

[26]While is it 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.

(we assume one bus station per grid cell), single-transfer bus connections, and an outside option. The exact choice set for bus options will depend on the state of the public transport network at a point in time.

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 transit network. The model highlights the importance of wait times, which affect decisions in two key ways. First, overlapping routes between an origin and a destination effectively decrease wait times, because the traveler can take the first bus that arrives among these bus routes. (This is exactly the mechanism that we study empirically in Event 3.) Second, travelers sometimes forego short wait times for other route characteristics. For example, a traveler may decide to not take the first bus that arrives at the station if this option involves a transfer and instead wait longer for a direct (and shorter) route.

In the model, a traveler's utility from a given option is additive in wait times and several deterministic factors: 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 network routing model has four desirable properties. First, the model is invariant to aggregation of identical routes. Second, we confirm using our bus GPS data that the distribution of wait times is approximately exponential, as assumed in our model (Appendix Figure A.18). Third, our model has realistic predictions: overlapping routes lead to shorter wait times, and commuters sometimes choose to wait for a later bus arrival in order to use a better route. Fourth, we obtain tractable expressions for expected utility and choice probabilities.

We add to this baseline formulation the possibility that commuters are partly inattentive to certain options in their choice set. We assume that a route's arrival rate may be attenuated by an attention factor that depends on the travel time on the bus of options using that route, relative to the minimum travel time on the bus in the choice set. The model with full attention is a special case.

In the upper nest, commuters make a logit choice between using the private mode or the TransJakarta network, where the benefit of the latter is given by the risk-neutral expected utility over wait times of choosing from among available TransJakarta options. The private option in the model includes private vehicles, such as motorcycles, taxis, motorcycle taxis, and ride-hail apps, as well as other types of shared transport, such as mini-buses outside the TransJakarta network. We assume that these modes are infinitely elastic to meet demand at a given cost, but their prices and attributes do not vary over time. We also assume that

the TransJakarta network does not affect road traffic congestion.[27]

We estimate the model parameters leveraging the TransJakarta network expansion, focusing on commuters' values for travel time, wait time, and transfers, as well as attention parameters, which we allow to differ for BRT and non-BRT routes. We use a classical minimum distance approach whereby we search for the parameter vector that allows the model-predicted ridership to replicate the reduced form analysis documented above. In addition to the six moments (three events for BRT and non-BRT), we also use a seventh moment, described below, to help pin down the attention parameters.

4.1 Bus Route Choice Model

A Random Utility Model with Waiting Times. The core part of our model is a flexible, static discrete choice over bus options, where idiosyncratic heterogeneity is given by exponentially distributed random wait times. Daganzo (1979) calls this the "negative exponential distribution model." We first describe the general structure and key properties of the model, before delving into the details that are specific to bus network routing.

Consider an agent making a static discrete choice among a finite set of options $k \in K$. Each option has a deterministic utility v_k and a random component $w_k \geq 0$, which measures wait time. The wait time for each option is governed by an independent Poisson process with arrival rate λ_k, which can differ by option. This means that the wait time w_k is drawn independently from an exponential distribution $\Pr(w_k > w) = \exp(-\lambda_r w)$. The agent observes the wait times for all options and picks the option with the highest combined utility $u_k = v_k - \alpha_{\text{wait}} w_k$, where α_{wait} governs the agent's preferences over wait times.

We incorporate partial inattention by assuming that the agent notices each arrival from option k with independent probability p_k. In other words, with probability $1 - p_k$ the agent fails to notice the first arrival for option k. This leads to an "effective" arrival rate for option k of $\widetilde{\lambda}_k = \phi(p_k)\lambda_k$ with $0 \leq \phi(p) \leq 1$ (see Appendix A.5).

This model has a convenient invariance property: it is unchanged when options that have the same indirect utility are combined into a single option with the sum of arrival rates.

Proposition 1. *Assume that the choice set contains two options $k = 1, 2$ with $v_1 = v_2$ and arrival rates λ_1 and λ_2. The model where options 1 and 2 are replaced by a single option 3 with $v_3 = v_1 = v_2$ and $\lambda_3 = \lambda_1 + \lambda_2$ is isomorphic to the original model in terms of choice probabilities and expected utility.*

[27] BRT routes run on dedicated lanes separate from other road traffic, and new BRT routes during our study period were launched along existing BRT corridors, without changes in the supply of lanes for other road traffic. We assume that non-BRT routes did not affect road traffic congestion due to their low frequency and relatively low mode share.

Proof. This follows 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 $\lambda_1 + \lambda_2$. □

This property implies that the demand model is unchanged if a route is split into two identical routes, with the total number of buses (which, as we will see, determine arrival rates) split between the two new routes. In other words, our model does not feature the 'red bus, blue bus' problem discussed by McFadden in the context of the multinomial logit and other random utility models. This will prove to be an attractive invariance property for the planner's optimization problem that we will later set up in section 5.

Choice probabilities and expected utility are as follows:

Proposition 2. *Assume that options are ranked $v_1 < v_2 < \cdots < v_N$. The probability π_k to choose option k is given by*

$$\lambda_k^{-1}\pi_k = \sum_{i=1}^{k} e^{-M_i} \frac{e^{v_i \alpha_{wait}^{-1} \Lambda_i} - e^{v_{i-1} \alpha_{wait}^{-1} \Lambda_i}}{\Lambda_i},$$

where $\Lambda_i = \sum_{\ell=i}^{N} \lambda_\ell$ and $M_i = \sum_{\ell=i}^{N} v_\ell \lambda_\ell$, and $v_0 = -\infty$ by convention. Expected utility is

$$\mathbb{E} \max_k u_k = v_N - \pi_N \frac{\alpha_{wait}}{\lambda_N},$$

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

The proof uses algebraic manipulations of the exponential distribution (Appendix A.6). The first part of this result ensures that computing choice probabilities is computationally tractable. The second part 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 $v_N - \alpha_{\text{wait}}/\lambda_N$. In general, the influence of other options $k \neq N$ on expected utility is summarized by the probability π_N to choose the top option.

Bus Route Choice Model Setup. The travel demand model in the context of the TransJakarta network is as follows. 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 $k \in M_{odt}$, where the choice set depends on the TransJakarta network at calendar date t, which in our application will range between 2016 and 2020. Each option has a utility level $u_k = v_k - \alpha_{\text{wait}} T_k^{\text{wait}}$, where v_k is the deterministic component that depends on characteristics such as travel time and transfer terms (in cases where the bus option k involves a transfer), and T_k^{wait} is the wait time at the origin determined by a Poisson arrival process that we describe in the next section.

The bus route choice set M_{odt} contains all direct and single-transfer bus connections between o and d available at t. The bus connections must start in the origin grid cell o, end in the destination grid cell d, and transfer in an intermediate TransJakarta station m.

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

$$u_k = \underbrace{-\alpha_{\text{time}} T_k^{\text{time}}}_{v_k} - \alpha_{\text{wait}} T_k^{\text{wait}} \tag{2}$$

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

If passengers decide to take a *transfer* option $k \in M_{odt}$ that includes connecting in an intermediate station, we assume that they get the utility from the first leg k_1 of the route up to the intermediate station, and the expected utility from the second leg of the route, taking expectations over the best direct option k_2 connecting the intermediate station to the destination:

$$u_k = \underbrace{-\alpha_{\text{time}} T_{k_1}^{\text{time}} + \mathbb{E} \max_{k_2} \left[-\alpha_{\text{time}} T_{k_2}^{\text{time}} - \alpha_{\text{wait}} T_{k_2}^{\text{wait}} \right] + \mu_{\text{transfer}}}_{v_k} - \alpha_{\text{wait}} T_{k_1}^{\text{wait}}. \tag{3}$$

We 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 over the best route k_2 that they will take from the intermediate station to the destination. μ_{transfer} captures the pure transfer penalty of taking a transfer *above and beyond* the travel time and the (expected) utility for the second leg.

Preference parameters are allowed to differ by BRT/non-BRT. For clarity, we do not track this distinction in notation now and discuss it when we set up the model's estimation.

We assume the following timing for the commuter's decision. Conditional on deciding to use the TransJakarta network, the commuter goes to the (unique) 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 transfer 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 option that maximizes utility.[28]

Poisson Bus Arrival Process and Exponential Wait Time Distributions. 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

[28] During our study period, TransJakarta posted actual arrival times at each BRT station – see Figure A.17 for an example. Users could also look up wait times from home using an app, but these may change between the time the user leaves home and arrives at the station.

route r.[29] We assume that the Poisson processes for different routes are independent.

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 very infrequent service is rare. Under this assumption, the wait time at o for the next bus from a given route r is exponentially distributed. Using bus GPS data, we show that exponentially distributed wait times are a good approximation for small and moderate wait times (Figure A.18, panels A and B). The exponential distribution in the model slightly over-estimates long wait times, suggesting that our estimate of α_{wait} captures steeper-than-linear costs for large wait times.

The mean and variance of wait time for a route follow a tight relationship, which is similar for BRT and non-BRT routes (Figure A.18, panel C). Without independent variation in these moments, we note that our estimate of α_{wait} will capture the combined preference for the wait time distribution.

Large Choice Sets and Partial Inattention. The model outlined so far assumes that commuters consider all possible bus route combinations. However, the TransJakarta bus network is sufficiently interconnected and complex that some choice sets are very large. While the median choice set has four bus options, a quarter of origin-destination pairs have choice sets with at least 14 options, and 10 percent of choice sets have at least 28 bus options.

Many choice set options — those with very long travel times — are ex-ante unattractive. For the choice set at the 75th percentile, the slowest option in the choice set (by total travel time on bus) is 2.8 times slower than the fastest option in the choice set. For 10 percent of choice sets, this ratio is at least 3.6.

We allow commuters in the model to be partially inattentive to options in their choice set that have a long bus travel time. This is a simple heuristic for reducing the size of the consideration set, bearing some similarity to models of rational inattention (Gabaix, 2019).[30]

To capture partial inattention, we consider for each route an effective arrival rate $\widetilde{\lambda}_{rodt} = \lambda_r \cdot \phi_{rodt}$, where $\phi_{rodt} \in [0, 1]$. This is equivalent to assuming that for any bus on route k, the commuter does not observe it with some probability, independently of other buses. Note that under this model, the commuter still draws a waiting time for each route, but this may be longer for some routes. In our parametrization, the ϕ_{rodt} terms will be close to zero or one, so essentially, the commuter will either fully ignore or pay full attention to the route.

We parameterize the attention factor as a function of travel time on the bus. Specifically, let $T^{\min}_{rodt}$ denote the shortest travel time on the bus among bus options that include route r,

[29] Bus allocation and bus speeds are mostly constant throughout the day (Appendix A.3.2).

[30] Larcom et al. (2017) use a short-term disruption due to a strike to show that some commuters in London are using suboptimal routes.

and $T_{odt}^{\min}$ the shortest time in the entire choice set between o and d at time t. We define

$$\phi_{rodt} = \phi\Big(\frac{T_{rodt}^{\min}}{T_{odt}^{\min}}\Big),$$

where $\phi(T_{\text{ratio}}) = \frac{\exp(\gamma(\eta - T_{\text{ratio}}))}{1+\exp(\gamma(\eta - T_{\text{ratio}}))}$ has a decreasing logistic shape in the travel time ratio T_{ratio}. We will estimate the parameter η, which is an attention cutoff. Commuters pay significantly less attention when the travel time ratio is above η, and this function approaches a step function around η as the shape parameter γ grows to infinity.

Partial inattention leads to important differences in how commuters respond to changes in the TransJakarta network. Consider the addition of a faster route on top of a set of existing bus options, as in Event 2. With full attention, the new route can significantly decrease the waiting time at the origin, which increases bus network ridership. This effect is particularly strong when time on the bus is not very costly (low α_{time}) because in that case, all routes are good substitutes for each other. With partial inattention, the commuter will pay less attention to pre-existing bus options after the new faster route is launched, because of the travel time difference. Partial inattention can rationalize low responsiveness to this type of change in the choice set.

4.2 The Logit Choice Between Bus Network and Private Option

At the higher decision nest, a commuter traveling between o and d first decides between using the TransJakarta network or using an outside option, a catch-all for private modes (private motorcycle, for hire motorcycles, car, other private minibusses, etc.). The utilities for the two options for a trip i between o and d at time t are:

$$\begin{aligned} u_{it}^{\text{bus}} &= \underbrace{\Big(\mathbb{E}\max_{k\in M_{odt}} u_k\Big)}_{v_{odt}^{\text{bus}}} + \epsilon_{it}^{\text{bus}} \\ u_{it}^{\text{private}} &= \zeta_{od}^{\text{private}} + \epsilon_{it}^{\text{private}} \end{aligned} \tag{4}$$

where the term in large brackets captures expected utility over different realizations of wait time vectors. The term $\zeta_{od}^{\text{private}}$ captures all time-invariant factors that make the private option more attractive for that specific origin-destination pair.

The terms $\epsilon_{it}^{\text{bus}}$ and $\epsilon_{it}^{\text{private}}$ are Gumbel-distributed error terms with parameter β, giving rise to logit probabilities.

4.3 Model Estimation

Our estimation strategy is based on finding the vector of preference parameters that allows the model to match the reduced form results documented in section 3. We estimate

$$\theta = (\alpha_{\text{time}}, \alpha_{\text{wait}}^{\text{BRT}}, \mu_{\text{transfer}}^{\text{BRT}}, \eta^{\text{BRT}}, \alpha_{\text{wait}}^{\text{non-BRT}}, \mu_{\text{transfer}}^{\text{non-BRT}}, \eta^{\text{non-BRT}}).$$

That is, we estimate a single cost of travel time on the bus, and separate BRT and non-BRT costs of waiting, transfer shifters, and attention cutoffs. 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. However, we impose that travel time on the bus is valued similarly, given that the physical buses used on non-BRT and BRT routes are similar. Because we lack price variation to pin down the level of utility, we normalize the variance of logit shocks in (4).[31] We set the attention shape parameter $\gamma = 30$, which means that attention drops quickly to close to zero above η^{BRT} and $\eta^{\text{non-BRT}}$. We estimate the origin-destination private option attractiveness terms ζ_{od} to match the average bus ridership between o and d over time.

Event Study Moments. We match the following seven moments. First, we match the six "Post" coefficients α^E on bus ridership for all three events, separately for BRT and non-BRT, from Table 1.

To help pin down the attention parameters, we add a seventh moment focused on trip duration. We estimate the impact of BRT Event 2 on the logarithm of the total trip duration from o to d.[32] We first measure BRT trip duration (inclusive of waiting time) in the data using the specific tap-in and tap-out times for the same smart card at the BRT entry and exit station. To do this, we restrict to 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.

Travel times between o and d go down sharply after the new direct and quicker route is launched between o and d (Figure 4 and Table 3). The magnitude is 0.06 log points, which is smaller than the impact on log minimum travel time (-0.28). This reflects a mix of two effects. Travel time on the bus indeed falls to the extent that more commuters substitute to

[31] We normalize the logit parameter $\beta = \frac{\overline{D}}{D_{od}}$ where D_{od} is the straight line distance between grids o and d, and $\overline{D} = 8.5$km is the average distance. Using a constant β would lead logit choices that are nearly random for o, d pairs that are close because the travel time component of utility is generally increasing in straight line distance. Our normalization compensates for this mechanical effect.

[32] We focus on Event 2 because it significantly changes travel time on the bus.

the new direct route. However, realized wait times increase as commuters are more likely to wait for the new direct route. The model estimation will help disentangle the attention and value of travel time parameters that give this result.

Computing Event Study Moments in the Model. For a parameter vector θ, we compute model-predicted ridership for all origin-destination pairs and all time periods. Using the 1km hexagonal grids, we compute ridership for 33,880 o, d pairs (covering a maximum number of 350,493 bus route options), for all versions of the TransJakarta network between January 2016 and March 2020.

Given θ, for any o, d, t we compute choice probabilities

$$\pi_{odt}(\theta) = \frac{\exp(\beta v_{odt}^{\text{bus}})}{\exp(\beta v_{odt}^{\text{bus}}) + \exp(\beta \zeta_{od}^{\text{private}})},$$

which depend on the bus route choice model through the expected utility term v_{odt}^{bus}. Combined with the (cross-sectional) smartphone commuting flow data V_{od}, this yields model-predicted ridership for each o, d, t. We also compute expected trip duration. The choice set between o and d at t includes all TransJakarta routes already launched at t that stop in both o and d. Transfer options consist of a first-leg route r_1 that connects the o to an intermediate station m, such that there is at least one direct route between m and d.[33]

We use GPS data to measure travel times on the bus within the network. (Appendix A.3.2 shows that these are stable within days and across our entire study period.) In the model, we assume that each route has a bus allocation that is fixed over time, equal to the average bus allocation in the data. This determines the route's bus arrival rate.[34]

Given θ, we estimate $\zeta_{od}^{\text{private}}$ to match average bus ridership over time between o and d.[35]

We next run the reduced form analysis on model ridership, which we will then match to the empirical moments from Section 3. Given the high-dimensional fixed effects in equation (1), estimating PPML regressions at each iteration of the model is computationally prohibitive. To address this challenge, we proceed in three steps. First, we approximate the result of PPML estimation of the level of model-predicted ridership $R = R_{odt}^{\text{model}}(\theta)$ with a weighted least-squares regression on $\log(R)$.[36] Second, we obtain a significant computa-

[33] Transfer options require a connection at the level of a real TransJakarta station m, not only a grid cell.

[34] We use a "mechanical" bus allocation to avoid issues of reverse causality in model-predicted ridership.

[35] We compute the bus share as the ratio of bus ridership from TransJakarta data to total commuting estimated using smartphone location data. When this ratio is above 1, we use a Bayesian procedure to estimate the posterior over $(0, 1)$. We then find ζ_{od} to match the mean of the posterior ratio.

[36] In the presence of heterogeneous treatment effects, running PPML with outcome R and OLS on $\log(R)$ will weight treatment effects differently (Tyazhelnikov and Zhou, 2021). That paper shows that using R or a proxy of R as weights leads to a weighting of heterogeneous treatment effects that is similar to PPML. In our

tional speed improvement by pre-computing a vector x^E for each specification E, such that the coefficient of interest α^E from WLS is given by the dot product $\alpha^E = (x^E)'Y_{odt}$ for any outcome vector Y_{odt}.[37] Third, we use a slightly coarser set of fixed effects: origin-destination, origin-quarter-of-the-year, destination-quarter, and week fixed effects. We confirm that the reduced form results are very similar to this specification relative to (1), and after estimation, we report model fit results using the original, finer fixed effects.

Estimation and Inference. We use classical minimum distance estimation (CMD) to match the vector of seven numbers $\widehat{m} = (\widehat{\alpha}^{1B}, \widehat{\alpha}^{2B}, \widehat{\alpha}^{3B}, \widehat{\alpha}^{1N}, \widehat{\alpha}^{2N}, \widehat{\alpha}^{3N}, \widehat{\alpha}^{2B,\text{duration}})$, where $\{\widehat{\alpha}^{1B}, \widehat{\alpha}^{2B}, \widehat{\alpha}^{3B}, \widehat{\alpha}^{1N}, \widehat{\alpha}^{2N}, \widehat{\alpha}^{3N}\}$ are the empirical estimates for events 1, 2, and 3 for both BRT and non-BRT estimated in Section 3 above and $\widehat{\alpha}^{2B,\text{duration}}$ is the travel time moment discussed above. We find the parameter vector θ that minimizes the objective function

$$\min_{\theta} \; (m(\theta) - \widehat{m})'\widehat{W}(m(\theta) - \widehat{m}), \tag{5}$$

where $m(\theta)$ is the vector of model moments, and $\widehat{W} = \widehat{\Omega}^{-1}$ is the optimal weighting matrix given by the inverse variance-covariance matrix of the moments $\widehat{m}$. The estimate $\widehat{\Omega}$ comes from estimating all reduced form event coefficients "stacked" 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 20 randomly selected initial conditions and confirm that all converge to the same vector.

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

Identification. Our classical minimum distance strategy has two main advantages. First, the reduced form moments that we established in Section 3 correspond to model comparative static exercises of exogenously varying wait time, travel time, and directedness. Second, these moments are also directly relevant for our counterfactuals, where we will be considering

case, we use the measured TransJakarta ridership R^{data}_{odt} as weights, as a proxy for $R^{\text{model}}_{odt}(\theta)$. Unlike model ridership, which depends on θ, actual ridership is pre-determined before estimation, which is an advantage for the procedure discussed in the next paragraph.

[37] The vector x^E is a row of the WLS matrix $(X'WX)^{-1}X'W$ where X is the matrix of covariates, including all fixed effects, and W is the weighting matrix. We compute this matrix inversion only once before estimation, which relies on using pre-determined weights. To further reduce size, we apply the Frisch–Waugh–Lovell theorem to partial out the fixed effect with the largest number of categories.

alternate networks that differ in terms of these attributes.

To shed light on how model parameters and the event study moments used in estimation are linked, we calculate the Jacobian matrix, which measures how, in the model, each moment's value changes when each of the parameters changes. Table A.7 reports the matrix of derivatives $dm_i(\theta)/d\theta_j$, each measuring how moment m_i responds to a marginal change in parameter θ_j, and Figure A.9 plots these relationships for larger parameter changes. For clarity, both exhibits focus on BRT moments and parameters. The value of travel time on the bus α_{time} significantly and positively affects Event 2, with muted impacts on the other events, which is expected given that travel time does not change significantly for Events 1 and 3. The transfer shifter μ_{transfer} affects Events 1 and 2 negatively, with an impact on Event 3 that is almost zero. Again, this is as expected, as treated (o, d) pairs in the third event are already directly connected.

All parameters are estimated jointly, yet given this approximately upper-triangular matrix format, we can describe intuitively how parameters are identified in the following recursive manner. First, Event 3 depends almost only on the value of wait time, so its value pins down the value of wait time. Next, Event 1 depends on wait time and the transfer shifter, so its value pins then down the transfer shifter. Finally, with these two parameters in hand, Event 2 pins down the value of travel time on the bus.

4.4 Estimation Results

Table 4 shows the estimated parameters and 95 percent confidence intervals. We normalize some parameters relative to travel or wait time to ease interpretation. We first estimate the model only for BRT parameters in column 1, using the three BRT events as moments, as well as Event 2 for log trip duration.

Commuters in Jakarta view time spent waiting for the bus as more costly compared to travel time on the bus — the estimated value of wait time $\alpha^{\text{BRT}}_{\text{wait}}$ is 2.4 times larger compared to the value of time on the bus $\alpha^{\text{BRT}}_{\text{time}}$, and we can reject at 95% significance level that the two parameters are equal.[38] The wait time parameter governs the degree to which people are sensitive to frequent service.

There are several reasons why commuters may dislike wait time more than time on the bus. In our model, α_{wait} implicitly captures the importance of the entire wait time distribution, including wait time uncertainty. Waiting conditions may be less comfortable than

[38] By comparison, existing studies of time valuation in public transport contexts focus on Western countries, and most are based on hypothetical choices. The few that use a revealed preferences approach typically use observational variation in attributes, unlike the quasi-random variation we use here. Papers in this literature typically find that the ratio of wait time to in-vehicle time is above one, but with significant dispersion (Wardman, 2004).

riding the bus (for example, bus stations are not air-conditioned). In addition, commuters may mis-perceive wait time as being longer than it actually is.[39]

Turning to the pure transfer penalty, we cannot reject the hypothesis that $\mu^{\text{BRT}}_{\text{transfer}}$ is equal to zero. The 95% confidence interval is relatively tight around zero, allowing us to reject a cost equivalent to 2 minutes of wait time (which is the same as 4.3 minutes of travel time) or a positive benefit of 4.5 minutes of wait time.

Recall that μ_{transfer} captures preferences for transfer routes *above and beyond* the cost of transfer routes in terms of additional wait time (e.g. when making the connection) and any additional time on the bus. Our results suggest that there is no *additional* utility cost of transfers associated with, for example, having to give up one's seat on the first bus and move around. The main factors that matter are time spent waiting and traveling on the bus.[40]

We find that commuters exhibit inattention to the longer bus options in their choice set. Specifically, a commuter with estimated preferences is less attentive to bus options that are more than 34% longer than the quickest option in the choice set. This cutoff parameter is precisely estimated between 1.25 and 1.45.

In the second column, we jointly estimate BRT and non-BRT parameters, adding the moments for non-BRT events 1,2, and 3. The general pattern of results is similar. The ratio of the cost of wait time to travel time is 4.2 for non-BRT, which is statistically significantly larger than the ratio for BRT. This means that waiting for buses in non-BRT stations is even costlier than in BRT stations. The higher wait cost is perhaps not surprising given that non-BRT stations are just stops by the side of the road, not enclosed stations as with the BRT. We cannot reject that the transfer shifter is zero. However, the point estimate is now negative, and we can only reject at 5% level a transfer cost equivalent to 5.3 minutes of wait time, which is the same as 22 minutes of travel time. The attention parameter is $\eta^{\text{non-BRT}} = 1.44$, slightly higher but still broadly similar compared to BRT.

We next explore the model fit. Both models in Table 4 are exactly identified and the model perfectly matches the data moments at the estimated parameters. However, we can learn more about model fit by comparing the time series patterns, which are not explicitly matched. Figure A.6 shows these results, replicating the event study graphs from Figure 2 using model-predicted ridership. In general, for all six moments, the model does a good job of replicating the lack of pre-trends and increase in ridership after the launch of a new

[39]For example, Fan et al. (2016) conducts a survey experiment in Minnesota and finds that people mis-recall wait times, reporting stated wait times that are larger than what they actually experienced.

[40]As in the case of wait time, there is a lack of estimates of the pure transfer penalty based on quasi-random variation, as in this paper. Currie (2005) reviews existing studies, concentrated in Western countries and based mostly on hypothetical choice or, to a lower extent, observational variation, and reports a pure transfer penalty of 22 minutes of bus travel time on average over studies, ranging from 5 to 50 minutes.

route. The model also does a good job to fit aggregate ridership trends, including after an origin-destination pair first becomes connected by non-BRT (Figures A.7 and A.8), with the mention that ridership adjustments happen more gradually in the data.[41] Overall, these results help build confidence in the model predictions for ridership and commuter welfare of counterfactual routes, which we consider in the last section of the paper.

5 Optimal Network Design

We now use the preference parameters estimated in the previous section to study the implications for the bus transit network design problem in Jakarta. We ask how the planner can optimally re-allocate bus routes and the bus allocation to routes, holding the total number of busses and the BRT infrastructure fixed, to improve welfare given the estimated preference parameters. 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.

This problem is discrete and high-dimensional. It does not appear to have an analytical solution and it resembles NP-hard problems such as the traveling salesperson problem. A key challenge is that it exhibits both substitution and complementarity forces, so we cannot use results that rely on complementarity properties (Jia, 2008; Arkolakis and Eckert, 2021).

We introduce a general framework that allows us to characterize optimal allocations and perform comparative static exercises in our setting.[42] The key idea is to allow the planner to care both about model welfare as well as idiosyncratic shocks defined for every possible network. We assume these shocks are unknown to the researcher. The key implication is that for two networks N, N' with similar model welfare $W(N), W(N')$, small shocks may tip the planner into preferring one network over the other. This "smoothes" the planner's problem and leads to a *probability distribution* over networks, where networks with higher model welfare have higher probabilities of being chosen by the social planner as the 'optimal' network. We assume that shocks are type-I generalized extreme value (Gumbel) distributed, giving a multinomial logit distribution.

We can then characterize certain properties of optimal networks by taking expectations over the planner's distribution. We derive analytic formulas for expected optimal properties and their local comparative static with respect to model parameters. These objects can

[41] While these aggregate moments are not directly targeted, the terms $\zeta_{od}^{\text{private}}$ are estimated to match average bus ridership over the estimation sample time period.

[42] The methods we develop may also be applied to quantitative models where optimal policies are high-dimensional, discrete, and analytically intractable, especially when exact hat methods make it possible to compare welfare between two counterfactual policies.

be easily estimated with a sample of networks independently drawn from the planner's distribution.

In this setting, the objective is to sample from the planner's multinomial logit distribution. We show that established algorithms, such as Metropolis-Hastings, parallel tempering, or a modified version of simulated annealing (SA), asymptotically sample from the planner's distribution. In our application, we run multiple independent SA simulations to obtain an i.i.d. sample of optimal networks.

We find that, through the lens of the model, optimal networks are more expansive than the current TransJakarta network. Increasing the cost of wait time leads to less expansive networks, while increasing the transfer penalty makes the network slightly more direct.

5.1 Bus Network Optimization Environment

The planner's problem that we consider here is to design a bus network N that maximizes average commuter expected utility. This involves deciding the number of bus routes, where bus routes run, and how to allocate a fixed total number of busses to these routes. Formally, a network is a tuple $N = (K, (r_1, \ldots, r_K), (b_1, \ldots, b_K))$ where K is the number of routes, r_k is a route, 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. Each grid cell is connected to its eight adjacent grids, including diagonals. A bus route is a non-self-intersecting path on this graph with 1,536 edges. Busses travel in both directions along each route and stop at each grid station. We construct predicted bus travel times for all possible edges using the existing bus travel time data, separately for BRT and non-BRT (see Appendix A.7.1 for details).

The total number of busses and BRT infrastructure are held fixed.[43,44]

We use the travel demand model that we estimate in section 4. A commuter in the model going from o to d chooses among available transportation options given the proposed network N and estimated preferences $\widehat{\theta}$, including inattention parameters. Commuting flows between every (o, d) pair are fixed, computed using the smartphone location data (see section 2.2).[45] The counterfactual networks we consider here comprise intensive and extensive margin changes that resemble the historical changes that our model fits or is estimated to match,

[43] In counterfactual networks, each route has the BRT travel time on edges that have BRT infrastructure at present, and non-BRT travel times on all other edges.

[44] We assume that bus fuel costs only depend on travel time and hence do not change in counterfactuals.

[45] We re-estimate $\zeta_{od} = \zeta(D_{od})$ as a non-parametric function of distance D_{od} such that the model evaluated at the current TransJakarta network matches the total ridership for each distance bin. This gives us an imputed private option attractiveness for all (o, d) pairs. It also accounts for the difference in grid size relative to demand estimation.

which increases our confidence in the model's projections.

To account for potentially higher ridership levels in counterfactual networks, we introduce bus capacity constraints. We assume that when a bus has ridership above its capacity, passengers perceive travel time as flowing more slowly. This time-dilation effect grows steeply as ridership per bus exceeds bus capacity. As ridership in excess of capacity is endogenous, we define a *ridership equilibrium* where commuter optimal choices given crowding are consistent with the assumed crowding levels. Appendix A.7.2 describes the model and our computational approach.

The space we are optimizing over is extremely large. Even the number of unique paths is exponential in the number of grid cells, and a network consists of any combination of routes. The number of edge connection configurations is a very rough lower bound. With 418 grid cells and 1,536 possible edges, there are $2^{1,536}$ configurations where each edge is either connected or not – i.e. about 10^{500} possibilities. The allocation of busses to lines adds even more combinatorial dimensions to the optimisation problem, making it infeasible to derive the global welfare-maximizing network through exhaustive search over all possible networks.

5.2 Characterizing Optimal Policies

The Social Planner's Problem with Idiosyncratic Factors. Consider a social planner who chooses from a finite set of policies (networks, in our setting) $N \in \mathcal{N}$. In a typical application, the set $\mathcal{N}$ is high-dimensional and extremely large. The planner chooses the network N that solves

$$\max_N W(N;\theta) + \epsilon_N,$$

where $W(N;\theta)$ is welfare according to a known, fully specified model, and θ is a vector of parameters. ϵ_N is an idiosyncratic shock for network N, capturing factors not in the model that gives rise to $W(N;\theta)$, such as network construction cost shocks or preferences for specific networks. Assume that the ϵ_N are i.i.d. Gumbel with parameter β.[46]

The probability that network N maximizes social welfare is

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

We next show how to characterize optimal networks by taking expectations over π, and how to estimate these objects using a sample of networks drawn from π. We then introduce

[46] In our application, we choose β to be very large, i.e. we assume that the idiosyncratic shocks are small relative to model welfare.

methods that allow us to asymptotically sample from the π distribution.

Optimal Policies: Properties and Local Comparative Statics. Consider a property defined by a function $f(N, \theta)$. For example, this could measure the number of stations for a bus network N, in which case f only depends on N and not on θ. Alternatively, it could measure model-predicted ridership, in which case it depends on both N and θ.

We define the *expected optimal property* as

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

This is a weighted average of the property f over all possible networks N, with weights given by the probability that N is optimal. Given a sample of networks $N_1, \ldots, N_K$ drawn independently from π, we can estimate $f^*(\theta)$ using the sample counterpart $\widehat{f^*}(\theta) \equiv \sum_{k=1}^{K} f(N_k, \theta)$.

We can compute comparative statics in θ by estimating $\widehat{f^*}(\theta')$ for any other value $\theta' \neq \theta$. To compute *local* comparative statics of the expected optimal property, assume that $W(N; \theta)$ is differentiable in θ for all N. Appendix A.7.3 shows that the gradient of the expected optimal property is

$$D_\theta f^*(\theta) = \sum_N \pi(N; \theta) \Big(\beta D_\theta W(N; \theta) \big(f(N, \theta) - f^*(\theta) \big) + D_\theta f(N; \theta) \Big), \tag{7}$$

where $D_\theta W$ and $D_\theta f$ are the gradients of W and f with respect to θ, respectively. The first term captures the change in f^* due to changes in the probabilities $\pi(N; \theta)$, while the second term captures the direct effect on f of changing θ. When f does not depend directly on θ, as for the network statistics we analyze later, the gradient scales linearly with β and with the scale of welfare W. Hence, we focus on the sign of $D_\theta f^*(\theta)$ and compare the local effects of different parameters in θ. Once again, we can estimate $D_\theta f^*(\theta)$ using a sample of networks from the π distribution.

5.3 The Simulated Annealing Algorithm

Building on the Metropolis-Hastings algorithm, a large family of algorithms has been designed to allow sampling asymptotically from probability distributions that are difficult to compute explicitly. The key idea in our case is to construct a Markov chain over the space of all networks, such that the chain's stationary distribution is exactly (6). A key issue is the speed of convergence (or "mixing") to the stationary distribution. We use a version of the simulated annealing (SA) algorithm (Nikolaev and Jacobson, 2010; Rothlauf, 2011),

which is related to the Metropolis-Hasting algorithm but has a pre-specified number of steps and begins with "noisier" transitions in the initial phase, which helps ensure that the chain explores widely the space of all networks.

Our algorithm has K steps, an "inverse temperature" parameter that increases exponentially from β_1 to β_K, and begins at an arbitrary network N_1. Given network N_k at step k, a probabilistic proposal function $\Psi(N' \mid N_k)$ creates a candidate network N'. This function is user-provided and application-specific. The network N' is accepted with probability:

$$\Pr(N_{k+1} = N' \mid N_k) = \min\left(1, \frac{\exp\left(\beta_k W(N')\right)\Psi(N_k \mid N')}{\exp\left(\beta_k W(N_k)\right)\Psi(N' \mid N_k)}\right). \tag{8}$$

The candidate network is always accepted if it has higher welfare (with an adjustment for proposal probabilities), and otherwise it is accepted with a probability that is increasing in $W(N')$ and decreasing in β_k.

This algorithm asymptotically samples from the planner's distribution π.

Proposition 3. *Assume that the final inverse temperature parameter is set to $\beta_K = \beta$, the logit parameter from the social planner's problem and that the proposal distribution Ψ is irreducible, aperiodic, and satisfies $\Psi(N \mid N') > 0 \iff \Psi(N' \mid N) > 0$. As the number of steps $K \to \infty$, the final state N_K of the simulated annealing algorithm converges in distribution to π, the planner's distribution over allocations.*

The proof first shows that the stationary distribution of MH with parameter β is exactly π. Second, we sketch the argument that as K grows, the last section of SA approximates an MH chain with parameter β. See Appendix A.7.4 for details.

In the initial phase of the SA algorithm, the low inverse temperature implies that new candidate networks are accepted with high probability. This makes the chain approximately a random walk in the space of networks. This helps the algorithm explore widely the space of networks before converging to the π distribution, thus making SA less likely to get stuck near a local optimum close to the initial network.

The proposal function Ψ is important for the practical success of the SA algorithm. We use small changes such as adding or deleting one stop at the end of a route, local route re-routing, and minor changes in bus allocation between routes, and large changes 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. See Appendix A.7.5 for details.

Parameters and Implementation. We set $\beta = 10^8$, so idiosyncratic shocks have a small variance and the planner's choice of network is mostly driven by model welfare. We use

the commuter preference parameters estimated in section 4.3. We use a bus capacity of 84 people per bus based on official TransJakarta data on its bus fleet in early 2020.

We obtain independent asymptotic draws from π using 200 independent SA runs. Each time, we run the SA algorithm for $K = 200,000$ steps.[47]

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 A.11 shows the evolution of the 200 SA runs, and indeed we find that the distribution of welfare for the final network in each run is concentrated. Results for ridership are similar.[48]

5.4 Results

Figure 5a shows the TransJakarta bus network at the end of our study period. The network is relatively dense in the urban core of the city, with most of the 107 bus lines crossing through the downtown areas. 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.

Figure 5c shows one example of a network asymptotically drawn from the planner's distribution. (Figure A.12 shows several other examples.) This network covers more locations compared to the actual TransJakarta network, extending significantly beyond the central Jakarta DKI area. This larger coverage comes at the cost of slightly lower bus frequency in the central area of Jakarta.

Optimal networks are more expansive than the current TransJakarta network. Table 5 reports statistics for the current TransJakarta network in column 1 and for the average over optimal networks in column 2. Panel A reports properties of the network. Optimal networks connect 57% more locations – 66% of all locations in the study area compared to 42% for the current network. The share of location pairs connected by a direct or transfer bus connection also increases considerably, from 12% to 39%. This is achieved with a large network that covers 31% more edges with bus connections, from 544 km to 714 km.[49]

Optimal networks lead to a significant increase in welfare compared to the current net-

[47] The acceptance probability (8) depends on the ratio of proposal kernels Ψ. Calculating this ratio is computationally intensive, so we approximate it by a ratio of 1 when running the simulated annealing algorithm. For most types of network change proposals, this approximation is likely valid. A notable exception is when adding new routes, when this approximation will tend to favor less expansive networks. This effect works against the main result we describe below. See Appendix A.7.5 for details.

[48] Figure A.13 plots the entire distribution of other network statistics.

[49] In our context, efficiency gains do not come with a distributional cost, as coverage also grows for poor areas. 44% of the locations with above-median poverty rate have access to the current TransJakarta network, which grows to 61.6% on average for optimal networks.

work. The average expected utility over all the trips in Jakarta is higher by the equivalent of 21.7 seconds of travel time on the bus, per trip. This may not sound enormous, but recall that this is over all trips, not just bus trips – so given that 1.57% of all trips use the bus with the current network, the equivalent variation improvement is a reduction of 23.1 minutes of bus travel time per initial bus trip. Ridership increases from 1.57% of all trips with the current network to 2.57% on average with optimal networks.[50] As a consequence, the number of passengers per bus also increases. For example, bus occupancy as experienced by the average passenger doubles. Notably, these numbers are well below the average bus capacity of 84 people per bus (see Figure A.14). Bus trips are longer on average in optimal networks, and optimal networks have a slightly higher share of direct bus trips.

Comparative Statics. We next study how the optimal network shape changes when commuter preferences change (Table 6). This type of exercise helps us assess the sensitivity of our results to the parameters to estimate, and reveals how these micro preference parameters affect optimal network design. We focus on changes in the three key preference parameters that we estimated: the value of wait time α_{wait}, the value of travel time on the bus α_{time}, and the transfer shifter μ_{transfer}. We study the impact of large changes in parameters by sampling 40 optimal networks each for these different parameters using the simulated annealing algorithm.[51] We study the impact of these three parameter changes on three classes of bus network shape measures.

First, we find that network coverage shrinks as wait time costs increase (Table 6, panel A). For example, if commuters dislike wait time twice as much as we estimate, the share of location pairs connected by optimal networks goes from 39% with our baseline parameters to 23%. However, even with double wait time costs – which means that wait time is 4.8× and 8.4× as costly as time on the bus for BRT and non-BRT – optimal networks are still substantially more expansive than the current TransJakarta network.

Second, we consider the speed of bus connections (panel B). For each origin-destination pair with a direct bus connection, we consider the quickest bus connection in the network, and divide its duration by the duration of the quickest possible bus route between the two locations, over all possible paths in the geographic environment. In the current TransJakarta network, direct connections are on average 36% slower than the quickest possible bus connection. The average ratio is similar for the baseline optimal networks, but it goes down

[50] These numbers are based on our estimate of 35.1 million weekday trips between different grid cells, which we obtain from the smartphone location data. This implies a weekday average TransJakarta ridership of 551,036 with the current network.

[51] In Table A.9 we report the *local* comparative statics from (7) around the baseline estimated parameters. For most entries, we find the same direction for comparative statics.

significantly, to 26%, when using $2 \times \alpha_{\text{time}}$. This counterfactual also affects the coverage measures in panel A, that is, they lead to more compact networks, although this effect is smaller than that of doubling α_{wait}.

Third, we consider to what extent the network provides direct connections versus offering connections that require a transfer (panel C). In the current TransJakarta network, 21% of all location pairs that are connected by bus also have a direct connection. This number is lower for optimal networks (12%), and it goes up to 16% when we assume that the transfer penalty is higher by the equivalent of 15 minutes of wait time. This counterfactual also increases the (inverse) measure of network speed in panel B, likely because of additional location pairs that are connected by winding direct connections.

None of the parameter vectors we consider leads to network shape parameters that match the current TransJakarta network. However, a higher value of waiting time and a sharper transfer penalty help move the network statistics toward those achieved by the current TransJakarta network.

6 Conclusion

The trend toward centralized urban planning of public transport networks is likely to continue in the coming decades. For example, there are half a dozen BRT systems in African cities, with ten others are under planning or construction (The World Bank, 2022). As the number of large cities increases, this number will grow, impacting traffic congestion and pollution, commuting and work patterns, and more generally the shape of how cities grow.

In this paper, we studied the design of citywide public transport networks. Launching a public transport route affects the attractiveness of using the network throughout the city. Its impact depends on how the new route increases access, and whether it leads to faster, more frequent, or more direct connections, and how potential riders value those attributes. We used the large expansion of the TransJakarta public bus system between 2016-2020 to separate how specific types of network changes affected bus ridership, and then estimated separate commuter preference parameters for these service quality dimensions by matching the reduced form impacts on ridership.

We find that commuters in Jakarta value bus wait time 2-4 times more than time spent on the bus, transfers in BRT stations do not carry a utility penalty above and beyond the additional wait time and time on the bus, and commuters pay little attention to slow options in their bus network choice set. Characterizing the optimal networks in terms of coverage, speed, and directedness metrics, we found that the optimal bus networks in Jakarta would be more expansive — even holding the resources of the bus company fixed — connecting

more people and location pairs, and boosting overall usage. An optimally designed network that is based on these parameters thus has the potential to dramatically expand ridership and consumer welfare compared to the networks in place today.

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.

Alder, S. (2023): "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," *Working Paper*.

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

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 132.

Antràs, P., T. C. Fort, and F. Tintelnot (2017): "The Margins of Global Sourcing: Theory and Evidence from US Firms," *American Economic Review*, 107, 2514–64.

Arkolakis, C. and F. Eckert (2021): "Combinatorial Discrete Choice," *Working Paper*.

Balboni, C. (2021): "In Harm's Way: Infrastructure Investments and the Persistence of Coastal Cities," *Working Paper*.

Balboni, C., G. Bryan, M. Morten, and B. Siddiqi (2021): "Transportation, Gentrification, and Urban Mobility: The Inequality Effects of Place-Based Policies," *Working Paper*.

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.

Buchholz, N., L. Doval, J. Kastl, F. Matějka, and T. Salz (2020): "The Value of Time: Evidence from Auctioned Cab Rides," *NBER Working Paper 27087*.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference with Multiway Clustering," *Journal of Business & Economic Statistics*, 29, 238–249.

Coulombel, N. and G. Monchambert (2019): "Diseconomies of Scale and Subsidies in Urban Public Transportation," *HAL Working Paper, ffhal-02373768v1f*.

Currie, G. (2005): "The Demand Performance of Bus Rapid Transit," *Journal of Public Transportation*, 8, 41–55.

Daganzo, C. (1979): *Multinomial Probit: The Theory and its Application to Demand Forecasting*, Academic Press.

Daganzo, C. F. (2010): "Structure of Competitive Transit Networks," *Transportation Research Part B: Methodological*, 44, 434–446.

De Chaisemartin, C. and X. D'Haultfoeuille (2022): "Two-Way Fixed Effects and Differences-In-Differences with Heterogeneous Treatment Effects: A Survey," *The Econometrics Journal*.

Fajgelbaum, P. D. and C. Gaubert (2020): "Optimal Spatial Policies, Geography, and Sorting," *The Quarterly Journal of Economics*, 135, 959–1036.

Fajgelbaum, P. D. and E. Schaal (2020): "Optimal Transport Networks in Spatial Equilibrium," *Econometrica*, 88, 1411–1452.

Fan, Y., A. Guthrie, and D. Levinson (2016): "Waiting Time Perceptions at Transit Stops and Stations: Effects of Basic Amenities, Gender, and Security," *Transportation Research Part A: Policy and Practice*, 88, 251–264.

Fielbaum, A., S. Jara-Diaz, and A. Gschwender (2016): "Optimal Public Transport Networks for a General Urban Structure," *Transportation Research Part B: Methodological*, 94, 298–313.

Gabaix, X. (2019): "Behavioral inattention," in *Handbook of Behavioral Economics: Applications and Foundations 1*, Elsevier, vol. 2, 261–343.

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.

Goldszmidt, A., J. A. List, R. D. Metcalfe, I. Muir, V. K. Smith, and J. Wang (2020): "The Value of Time in the United States: Estimates from Nationwide Natural Field Experiments," *NBER Working Paper 28208*.

Hörcher, D., D. J. Graham, and R. J. Anderson (2017): "Crowding Cost Estimation with Large Scale Smart Card and Vehicle Location Data," *Transportation Research Part B: Methodological*, 95, 105–125.

Iliopoulou, C., K. Kepaptsoglou, and E. Vlahogianni (2019): "Metaheuristics for the Transit Route Network Design Problem: A Review and Comparative Analysis," *Public Transport*, 11, 487–521.

Institute for Transportation and Development Policy (2019): "Transjakarta: A Study in Success," https://www.itdp.org/2019/07/15/transjakarta-study-success/, [Online; accessed 9-June-2023].

Jia, P. (2008): "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 76, 1263–1316.

JUTPI2 (2019): "Jabodetabek Urban Transportation Policy Integration Project Phase 2 – Annex 2: JABODETABEK Urban Transportation Master Plan," https://openjicareport.jica.go.jp/pdf/12356366.pdf, [Online; accessed 9-June-2023].

Kreindler, G. E. (2023): "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications," *NBER Working Paper 30903*.

Larcom, S., F. Rauch, and T. Willems (2017): "The Benefits of Forced Experimentation: Striking Evidence from the London Underground Network*," *The Quarterly Journal of Economics*, 132, 2019–2055.

Majid, H., A. Malik, and K. Vyborny (2018): "Infrastructure Investments, Public Transport Use and Sustainability: Evidence from Lahore, Pakistan," *Working Paper*.

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.

Rothlauf, F. (2011): *Design of Modern Heuristics: Principles and Application*, vol. 8, Springer.

Santamaria, M. (2022): "Reshaping Infrastructure: Evidence from the Division of Germany," *Working Paper*.

Small, K. A., E. T. Verhoef, and R. Lindsey (2007): *The Economics of Urban Transportation*, Routledge.

SMERU (2014): "Poverty and Livelihood Map of Indonesia 2015," https://www.indonesiapovertymap.org/, [Online; accessed 9-June-2023].

The World Bank (2022): "With Bus Rapid Transit, African Cities Are Riding Toward a Better Future," *Feature Story, Online [9-June-2023]*, [Online; accessed 9-June-2023].

Tsivanidis, N. (2022): "The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," *Working Paper*.

Tyazhelnikov, V. and X. Zhou (2021): "PPML, Gravity, and Heterogeneous Trade Elasticities," *Working Paper*.

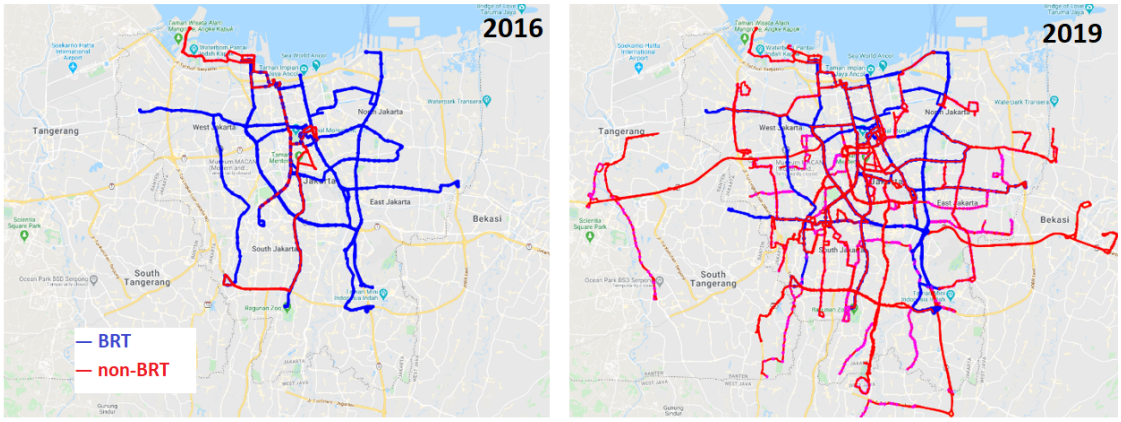
Wardman, M. (2004): "Public Transport Values of Time," *Transport Policy*, 11, 363–377.

Wei, F. and R. B. Machemehl (2006): "Using a Simulated Annealing Algorithm to Solve the Transit Route Network Design Problem." *Journal of Transportation Engineering*, 132, 122 – 132.

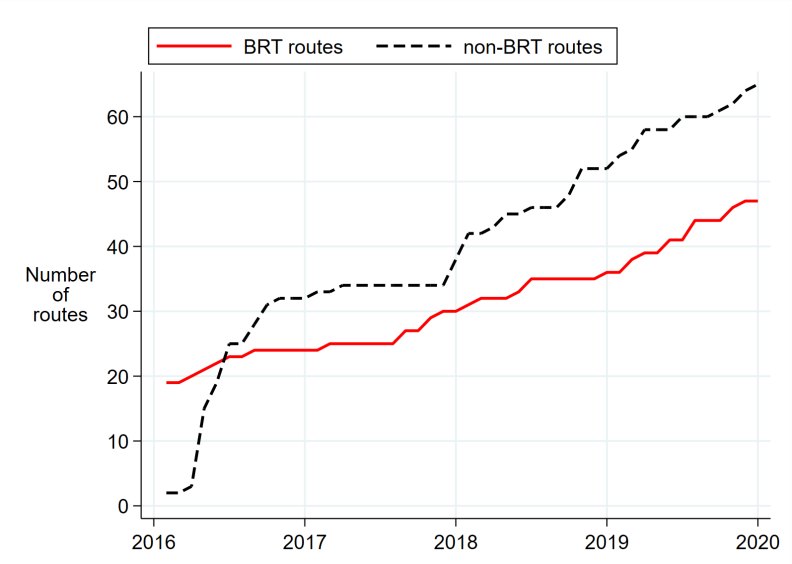
Zárate, R. D. (2023): "Spatial Misallocation, Informality and Transit Improvements: Evidence from Mexico City," *Working Paper*.

Figure 1: TransJakarta Network Expansion Since 2016

(a) Expansion of TransJakarta Route Network



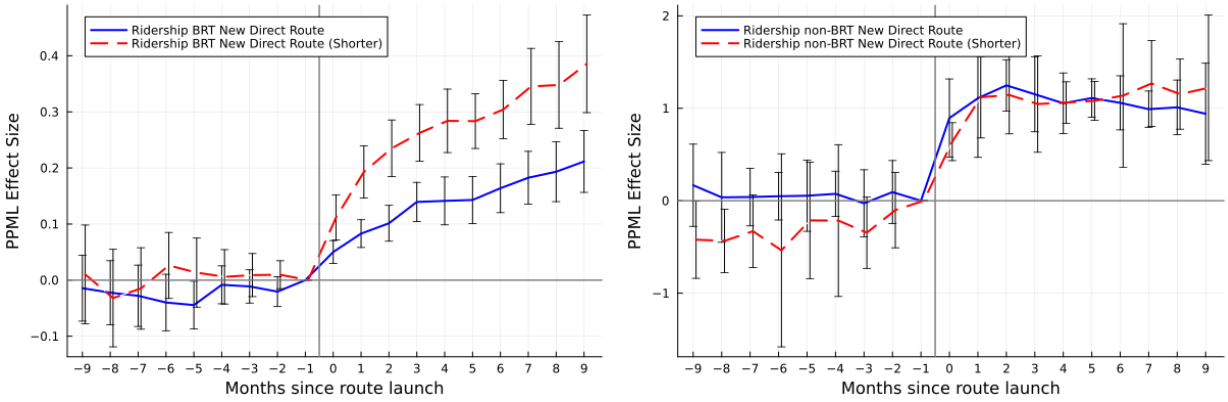
(b) Number of BRT and Non-BRT Routes Over Time



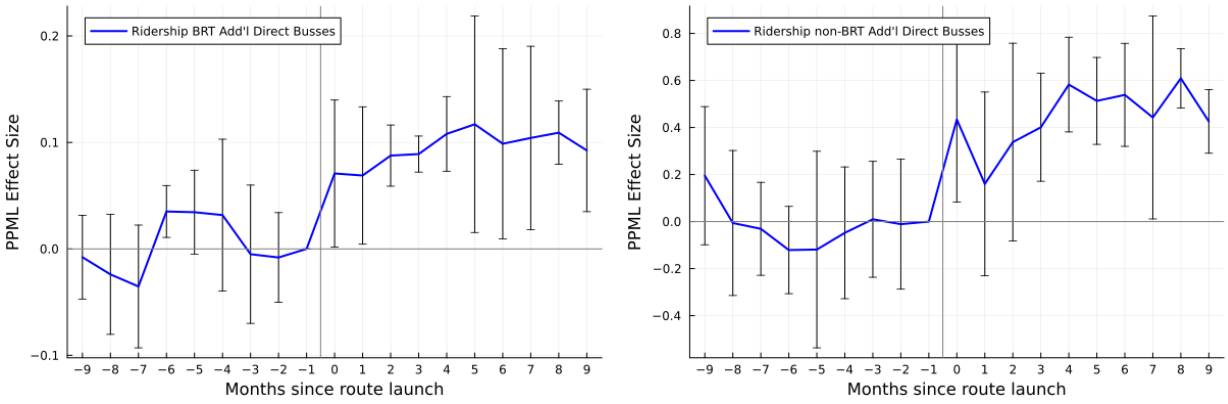
Notes: The top panel shows two snapshots of the TransJakarta network in 2016 and 2019. The bottom panel shows the number of active routes over time.

Figure 2: **Bus Ridership** Impacts of BRT and Non-BRT Network Expansions

(a) Event Types 1 and 2: New Direct Route



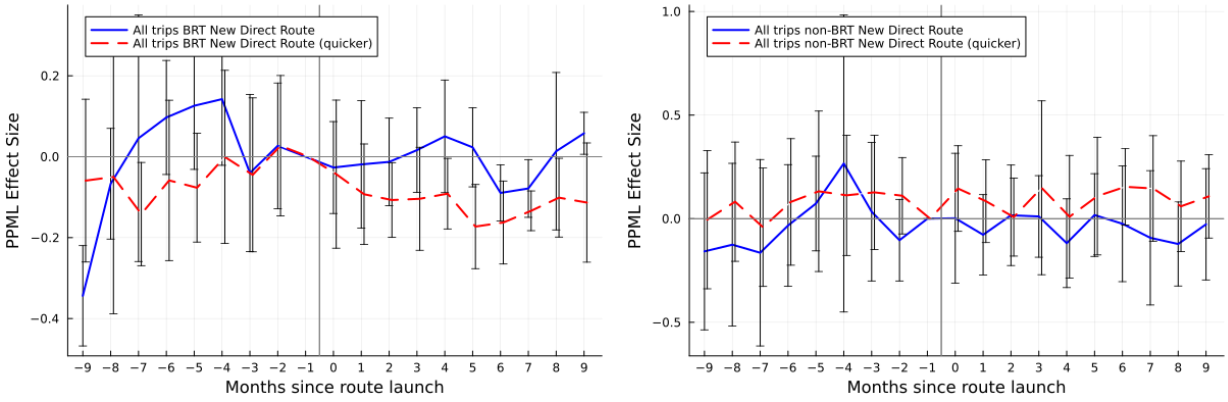
(b) Event Type 3: Additional Busses (Direct)



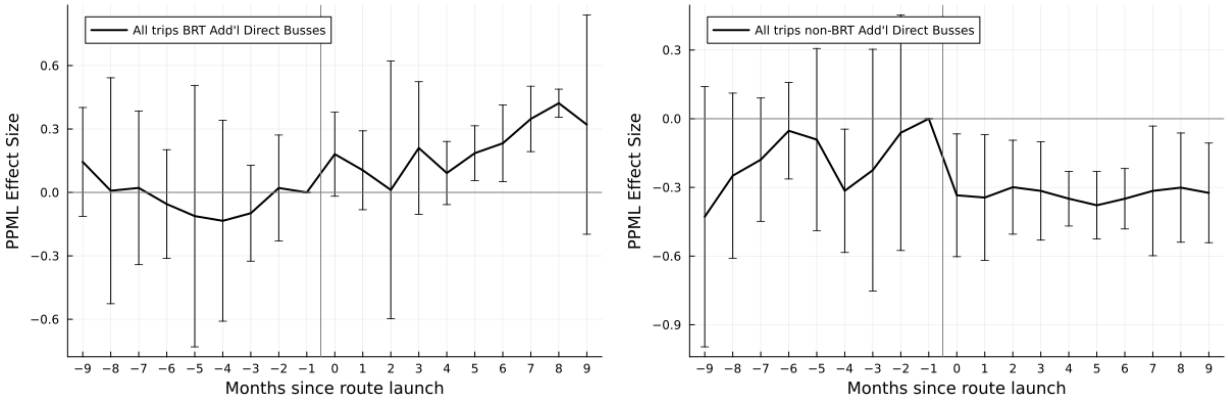
Notes: BRT graphs on the left, and non-BRT graph on the right. In the graphs on the first row, the blue, solid lines report monthly coefficients for the first event type (direct route with no change in travel time on the bus). The red, dashed lines report monthly coefficients for the second event type (quicker direct route). The graphs in the second row show the impact of the third event type. All specifications are the event study version of equation (1). 95% confidence intervals are based on standard errors clustered two-way at the origin and destination grid level.

Figure 3: **Aggregate Trip Volume** (Lack of) Impacts of BRT and non-BRT Network Expansions

(a) Event Types 1 and 2: New Direct Route

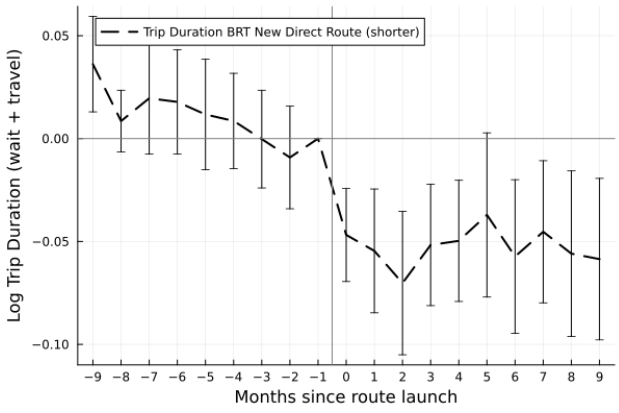


(b) Event Type 3: Additional Busses (direct)



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

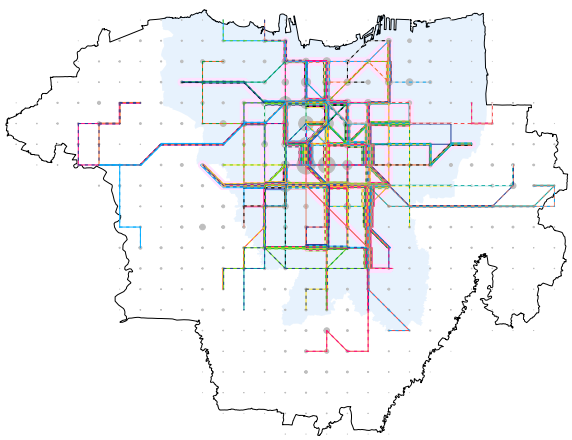
Figure 4: Trip Duration Impact of BRT Event 2 (New and Quicker Direct Route)



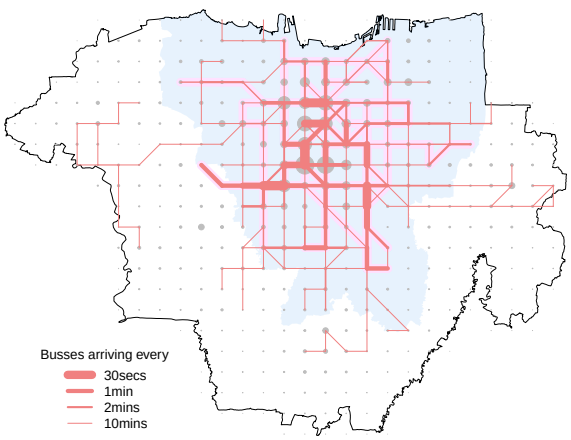
Notes: The outcome is the log of trip duration (wait time plus time on the bus), measured from linked tap-in and tap-out observations for the same smart card. The sample is restricted to BRT destination stations that are tap-out compliant, defined as stations where tap-out transactions account for at least 30% of all transactions. The specification is the same as for Event 2 in Figure 2. 95% confidence intervals are based on standard errors clustered two-way at the origin and destination grid level.

Figure 5: Current TransJakarta Network and Optimal Network Example

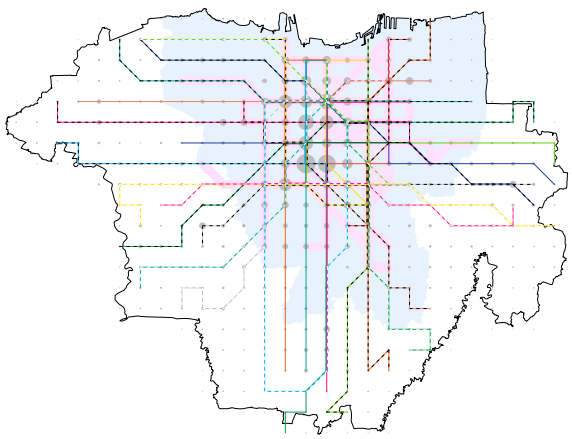
(a) The Current TransJakarta Network



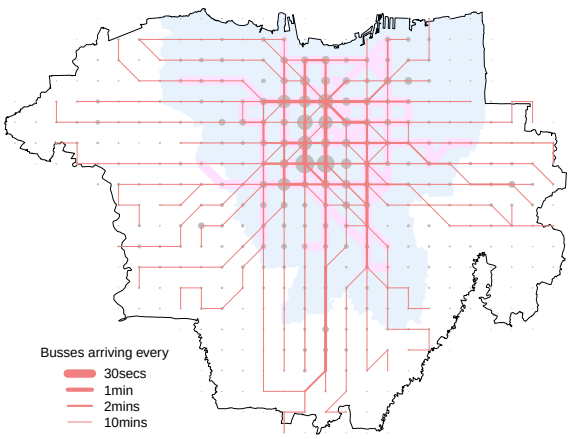
(b) Edge-level Bus Frequency (Current)



(c) Optimal Network $N \sim \pi$ (Single SA Run)



(d) Edge-level Bus Frequency (Single SA Run)



Notes: This figure depicts the current TransJakarta network and one of the networks N obtained using the simulated annealing (SA) algorithm, asymptotically sampled from the planner's distribution π. The study area (black outline) includes the Jakarta DKI area (light blue) as well as the urban districts of Tangerang, South Tangerang, Depok, and Bekasi. Different bus lines are printed in different colors. Edges connected through BRT dedicated bus lanes are denoted by bright pink underlay. Panel (a) plots the current TransJakarta network (107 bus lines). Panel (b) shows bus frequency on each edge in the current TransJakarta network. To construct this graph, for each directed edge we compute the total bus arrival over all routes that share that edge. For each edge, we then take the average over the two directions. Panels (c) and (d) repeat the exercise for one of the 200 SA networks, chosen randomly.

Table 1: Impact of Network Expansion on Travel Time, Wait Times, and **Bus Ridership**

(a) BRT

	log Min Travel Time			log Bus/hr (origin)			Bus Ridership		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
E1: New Direct Line	0.036***			-0.051*			0.158***		
	(0.007)			(0.026)			(0.021)		
E2: New Direct Line (quicker)		-0.294***			0.070			0.268***	
		(0.027)			(0.036)			(0.037)	
E3: Additional Busses			-0.027*			0.320***			0.092***
			(0.011)			(0.035)			(0.023)
Orig × Dest FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	PPML	PPML	PPML
Mean outcome	33.2	44.2	12.9	35.2	27.0	20.0	111.3	76.9	210.4
Median outcome	31.0	42.4	10.6	28.7	20.7	16.9	55.9	39.1	94.6
Unique origin x destination pairs	18,433	18,449	5,319	18,433	18,449	5,319	18,433	18,449	5,319
Unique origins	148	148	148	148	148	148	148	148	148
N	3,154,672	3,143,019	793,965	3,154,672	3,143,019	793,965	3,154,672	3,143,019	793,965
R^2	0.981	0.980	0.998	0.923	0.921	0.997			

(b) Non-BRT

	log Min Travel Time			log Bus/hr (origin)			Bus Ridership		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
E1: New Direct Line	0.030**			0.292***			1.016***		
	(0.009)			(0.044)			(0.153)		
E2: New Direct Line (quicker)		-0.716***			0.205***			1.320***	
		(0.072)			(0.030)			(0.267)	
E3: Additional Busses			-0.089***			0.425***			0.450**
			(0.021)			(0.015)			(0.141)
Orig × Dest FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	PPML	PPML	PPML
Mean outcome	34.8	49.3	13.0	11.6	11.2	7.1	19.2	7.5	26.0
Median outcome	24.8	47.2	11.0	7.8	8.5	6.3	0.0	0.0	0.0
Unique origin x destination pairs	2,588	2,652	1,263	2,588	2,652	1,263	2,588	2,652	1,263
Unique origins	57	57	64	57	57	64	57	57	64
N	306,722	313,994	144,763	306,722	313,994	144,763	306,722	313,994	144,763
R^2	0.992	0.987	0.998	0.967	0.967	0.992			

Notes: The two panels report results for BRT and non-BRT, respectively. Columns 1-6 capture the "first stage" impacts of the relevant events and are estimated using OLS. The bus ridership effects (columns 7-9) are estimated by PPML. "Min Travel Time" captures the time on the bus (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 the routes connecting the origin and destination. (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. The origin and destination data is aggregated at the level of 1km hexagonal grids. Standard errors clustered two-way at the origin and destination grid level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Impact of BRT and Non-BRT Network Expansions on **Aggregate Trip Volume**

(a) BRT

	log Min Travel Time			log Bus/hr (origin)			Bus Ridership			All trips		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
E1: New Direct Line	0.057***			-0.158***			0.108***			-0.008		
	(0.008)			(0.031)			(0.016)			(0.051)		
E2: New Direct Line (quicker)		-0.214***			-0.108***			0.208***			-0.073	
		(0.026)			(0.030)			(0.032)			(0.049)	
E3: Additional Busses			-0.006			0.221***			0.046			0.166
			(0.036)			(0.058)			(0.041)			(0.117)
Orig × Dest FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	PPML	PPML	PPML	PPML	PPML	PPML
Mean outcome	32.6	42.0	13.4	37.1	32.9	13.6	122.9	93.7	195.5	1895.5	1729.5	10117.3
Median outcome	30.6	42.2	10.9	32.5	27.7	11.6	61.9	54.8	153.7	0.0	0.0	6412.4
Unique origin x destination pairs	13,802	13,675	2,497	13,802	13,675	2,497	13,802	13,675	2,497	13,802	13,675	2,497
Unique origins	124	124	89	124	124	89	124	124	89	124	124	89
N	1,184,781	1,172,746	197,039	1,184,781	1,172,746	197,039	1,184,781	1,172,746	197,039	1,184,781	1,172,746	197,039
R^2	0.986	0.985	0.999	0.948	0.948	0.999						

(b) Non-BRT

	log Min Travel Time			log Bus/hr (origin)			Bus Ridership			All trips		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
E1: New Direct Line	0.027*			0.253***			0.999***			-0.042		
	(0.011)			(0.048)			(0.199)			(0.060)		
E2: New Direct Line (quicker)		-0.736***			0.200***			1.337***			0.031	
		(0.094)			(0.036)			(0.322)			(0.060)	
E3: Additional Busses			-0.097**			0.365***			0.425*			-0.177*
			(0.031)			(0.023)			(0.176)			(0.073)
Orig × Dest FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	PPML	PPML	PPML	PPML	PPML	PPML
Mean outcome	35.9	44.3	12.4	12.0	12.0	7.4	13.0	7.6	40.7	3948.6	6060.7	6880.5
Median outcome	24.5	43.6	10.8	7.8	8.5	6.9	0.0	0.0	0.0	635.5	956.9	1017.5
Unique origin x destination pairs	1,906	1,950	860	1,906	1,950	860	1,906	1,950	860	1,906	1,950	860
Unique origins	49	49	53	49	49	53	49	49	53	49	49	53
N	156,656	160,409	61,377	156,656	160,409	61,377	156,656	160,409	61,377	156,656	160,409	61,377
R^2	0.995	0.989	0.999	0.976	0.976	0.997						

Notes: This table reports the impact on the aggregate trip volume as measured from smartphone data (columns 10-12). The outcome is aggregate trip volume from smartphone data. The time sample is restricted to after March 2018. For comparison, columns 7-9 report the impact on bus ridership in the same sample. 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 Event Type 2 on Total Trip Duration

	log Min Travel Time	log Bus/hr (origin)	Bus Ridership	log Trip Duration
	(1)	(2)	(3)	(4)
E1: New Direct Line	0.055***	-0.212***	0.110***	-0.003
	(0.012)	(0.038)	(0.025)	(0.005)
E2: New Direct Line (quicker)	-0.284***	0.165***	0.071*	-0.057***
	(0.038)	(0.045)	(0.036)	(0.012)
Orig × Dest FE	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	PPML	OLS
Median outcome (E1)	30.1	33.7	68.4	54.2
Median outcome (E2)	40.1	27.8	61.2	57.6
Unique origin x destination pairs	13,455	13,455	13,455	13,455
Unique origins	148	148	148	148
N	595,685	595,685	595,685	595,685
R^2	0.991	0.970		0.881

Notes: Column 4 in this table reports the impact on log trip duration (wait time plus time on the bus) of BRT Event 2. The outcome is trip duration measured from linked tap-in and tap-out observations for the same smart card. The sample is restricted to BRT destination stations that are tap-out compliant, defined as stations here tap-out transactions account for at least 30% of all transactions. We jointly estimate Events 1 and 2, so the "E2: New Direct Line (quicker)" coefficient captures the *additional* effect for Event 2. First-stage and bus ridership impacts are reported in columns 1-3. Standard errors clustered two-way at the origin and destination grid level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Estimated Demand Model Parameters

	(1)	(2)	
	BRT	BRT	non-BRT
Wait time $\alpha_{\text{wait}}/\alpha_{\text{time}}$	2.4	2.4	4.2
	[1.8, 4.3]	[1.9, 4.3]	[3.2, 6.6]
Travel time α_{time}	1	1	1
Transfer shifter $\mu_{\text{transfer}}/\alpha_{\text{wait}}$ (minutes)	2.3	2.3	-1.1
	[-2.0, 4.5]	[-2.0, 4.4]	[-5.3, 0.9]
Attention Cutoff η	1.34	1.34	1.44
	[1.25, 1.45]	[1.25, 1.45]	[1.43, 1.52]
Logit parameter σ	0.060	0.060	
	[0.023, 0.113]	[0.023, 0.105]	
Moments:			
BRT events (1-3)	Yes	Yes	
BRT event 2 trip duration	Yes	Yes	
non-BRT events (1-3)		Yes	

Notes: We use a classical minimum distance with the optimal weighting matrix, and 20 random initial conditions. To construct the 95% confidence intervals, we re-estimate the model 100 times. Each time, we target a data moment vector $\widehat{m}^k = \widehat{m} + \varepsilon^k$, where ε^k is randomly drawn from the multivariate normal distribution $\mathcal{N}(\mathbf{0}, \widehat{\Omega})$. $\widehat{\Omega}$ is the estimated variance-covariance matrix of the reduced form analysis, jointly estimated in a seemingly unrelated regression framework. During this procedure, we use 4 random initial conditions for each estimation.

Table 5: Properties of Current and Optimal Networks

Statistic	Current Network	Baseline Optimal	Difference Baseline – Current
Panel A: Network and access properties			
Locations with a bus station (share)	0.42	0.66	0.24
			(0.0015)
Location pairs connected by bus (share)	0.12	0.39	0.27
			(0.0018)
Trips with access to bus (share)	0.73	0.91	0.18
			(0.00094)
Total network mileage (km)	543.9	713.9	169.9
			(2.0)
Number of bus lines	107	27.76	-79.24
			(0.2)
Panel B: Ridership equilibrium properties			
Welfare	-0.0031	-0.0027	0.00036
			(0.00000065)
Ridership (%)	1.57	2.57	1.0063
			(0.0028)
Average bus occupancy (pax/bus)	13.62	33.00	19.39
			(0.034)
Average experienced bus occupancy (pax/bus)	23.82	46.1	22.27
			(0.056)
Average bus trip duration (hrs)	0.59	0.81	0.23
			(0.00073)
Transfer trips (share of all bus trips)	0.76	0.71	-0.049
			(0.00054)

Notes: This table reports characteristics of the TransJakarta network (column 1) and of optimal networks, on average over 200 independent simulated annealing (SA) runs (column 2). Panel A reports network properties, while Panel B reports properties of the ridership equilibrium. "Trips with access to bus" denotes the share of all trips (measured using smartphone data and assumed fixed in counterfactuals) that can be completed using (have access to) the bus network. "Total network mileage" computes the total length of all edges traversed by the network. The difference in welfare in Panel B in equivalent variation terms is a reduction of 23.1 minutes of bus travel time for each bus trip in the current network. "Ridership" corresponds to the share of total trips across Jakarta using the bus network. "Average bus occupancy" is the occupancy level of the average bus, assuming busses are active for 17 hours each day (5AM-10PM) and using the model to predict total bus ridership for each route and pair of consecutive stations. "Average experienced occupancy" weights this number by ridership itself to get at the average amount of other passengers on a bus for a typical passenger. In column 3, in parentheses, bootstrapped standard errors are obtained using the 200 independent runs.

Table 6: Optimal Networks Comparative Statics

Statistic	Current Network	Baseline Optimal	Comparative Statics (Large Changes)		
			Wait time $2 \times \alpha_{\text{wait}}$	Time on Bus $2 \times \alpha_{\text{time}}$	Transfer $\mu_{\text{transfer}} - 15\text{mins}$
Panel A: Coverage measures					
Locations with a bus station (share)	0.42	0.66	0.5	0.59	0.65
		(0.0015)	(0.0032)	(0.0024)	(0.0031)
Location pairs connected by bus (share)	0.12	0.39	0.23	0.3	0.41
		(0.0018)	(0.0028)	(0.0025)	(0.004)
Total network mileage (in km)	543.95	713.87	498.9	664.89	744.068
		(2.032)	(3.15)	(3.24)	(3.73)
Panel B: Speed measures					
Bus time relative to quickest possible bus route	1.36	1.35	1.3	1.26	1.48
		(0.01)	(0.024)	(0.021)	(0.029)
Panel C: Directness measures					
Connected directly (share of all connected pairs)	0.21	0.12	0.14	0.13	0.16
		(0.0006)	(0.0016)	(0.0011)	(0.0018)

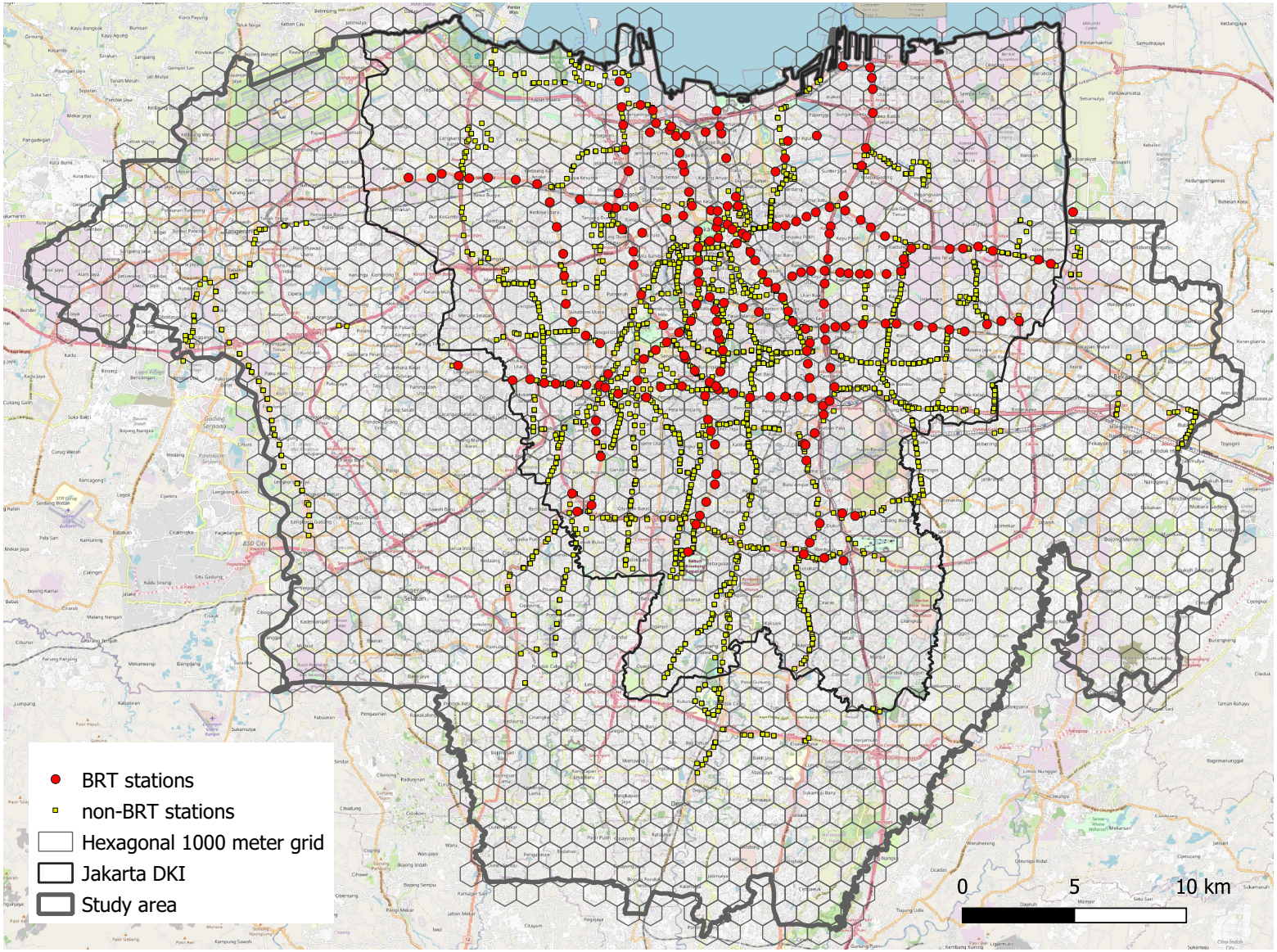
Notes: This table reports how network shape measures (rows) change as a function of parameter changes (columns). Columns 3-5 report characteristics of optimal networks achieved with different sets of preference parameters, each obtained from 40 independent simulated annealing runs with altered parameter vectors: double value of wait time ($\alpha_{\text{wait}}^{\text{BRT}}$ and $\alpha_{\text{wait}}^{\text{non-BRT}}$) relative to our estimate, double value of time on bus (α_{time}) relative to our estimate, and an additional penalty of taking a transfer ($\mu_{\text{transfer}}^{\text{BRT}}$ and $\mu_{\text{transfer}}^{\text{non-BRT}}$) equivalent to 15 minutes of wait time. See Table 5 for variable definitions. To construct "bus time relative to quickest possible bus route," for each origin-destination pair that is directly connected, we consider the quickest direct bus connection in the network, and divide its duration by the duration of the quickest possible bus route between the two locations in the entire geographic environment. We then average this ratio over all directly connected origin-destination pairs. "Connected directly (share of all connected pairs)" computes the share of all location pairs that are connected by bus, for which a direct connection also exists. Bootstrapped standard errors are in parentheses. Table A.9 reports the *local* comparative statics version of this table.

A Appendix - For Online Publication

Contents

A.1 Appendix Figures

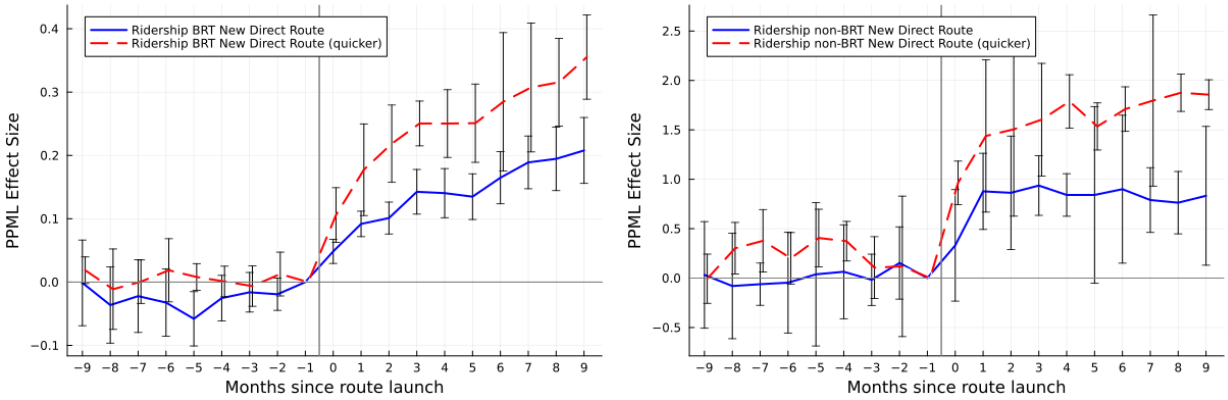
Figure A.1: Map of Study Area With 1km Hexagonal Grids and TransJakarta Bus Stations



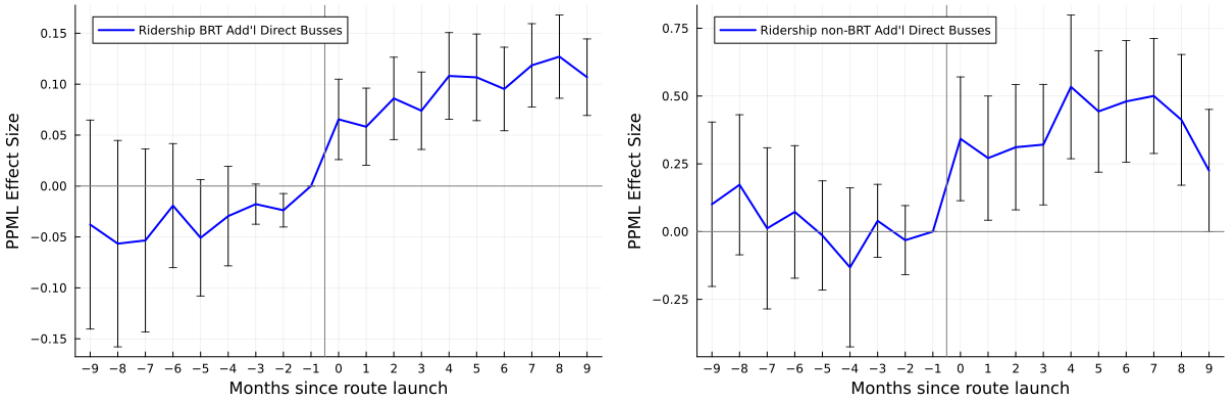
Notes: This figure shows the location of all TransJakarta stations (BRT in red circles, non-BRT in yellow squares) as of March 2020, together with the hexagonal grid with distance 1000 meters between centroids. The thin black link is the boundary of the Special Capital Region of Jakarta (Jakarta DKI), whereas the thicker gray boundary also includes the adjacent cities of (in counterclockwise order) Tangerang, South Tangerang, Depok, and Bekasi.

Figure A.2: Robustness With 500m Square Grids: **Bus Ridership** Impacts of BRT and Non-BRT Network Expansions (PPML)

(a) Event Types 1 and 2: New Direct Route



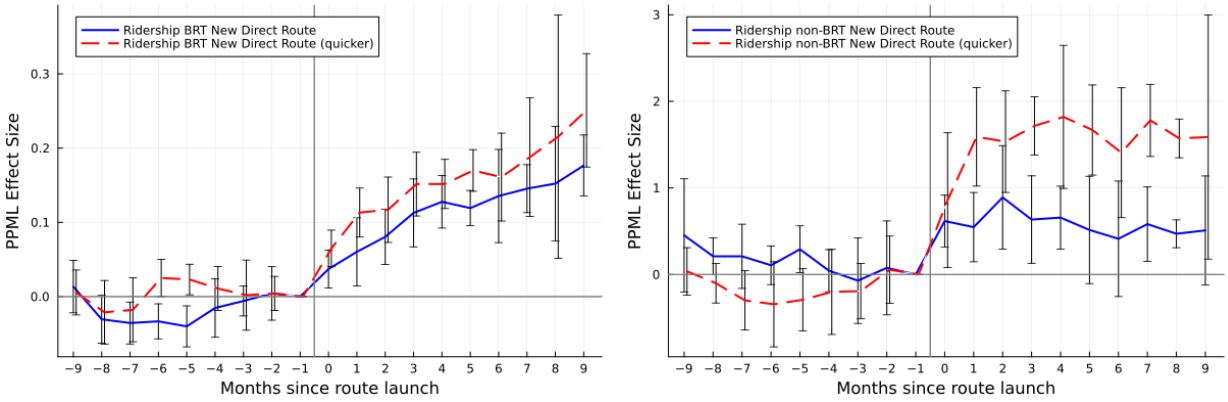
(b) Event Type 3: Additional Busses (direct)



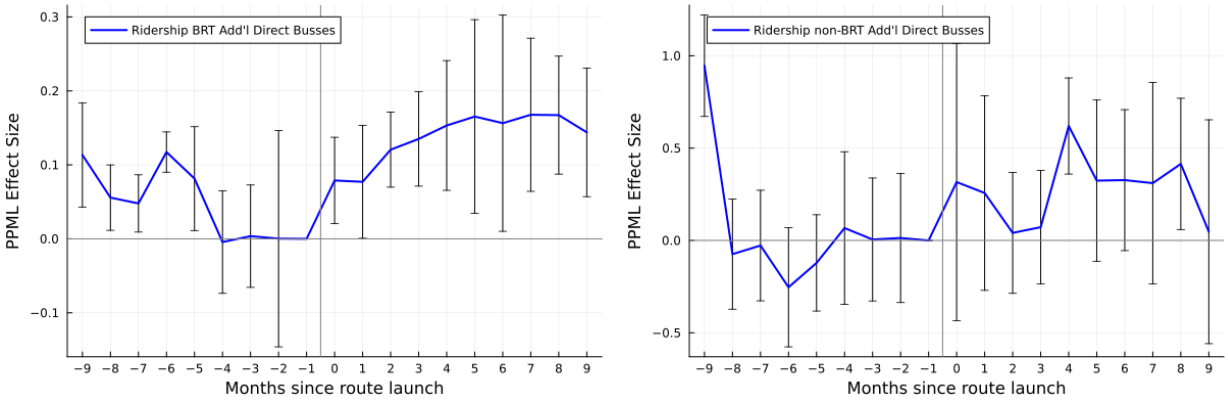
Notes: This graph replicates Figure 2 using 500-meter square grids instead of 1km hexagonal grids.

Figure A.3: Robustness With 2000m Square Grids: **Bus Ridership** Impacts of BRT and Non-BRT Network Expansions (PPML)

(a) Event Types 1 and 2: New Direct Route



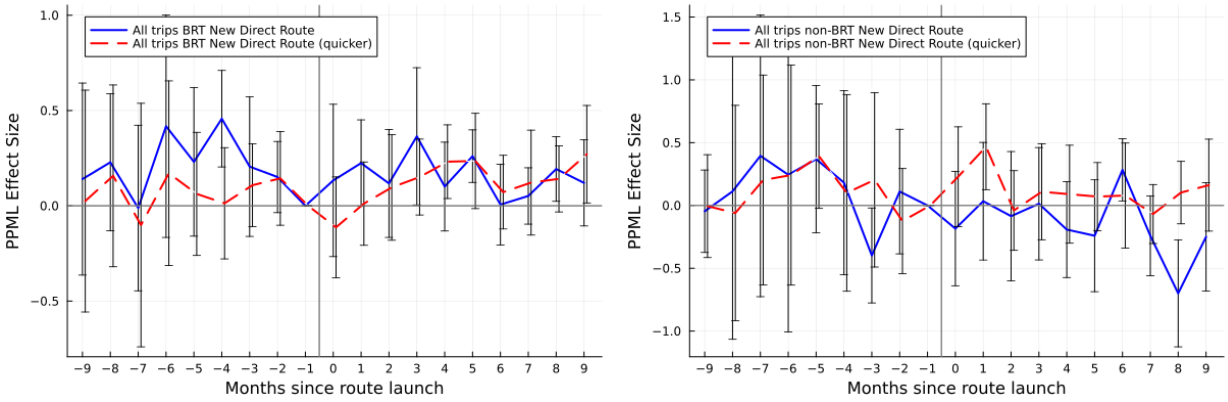
(b) Event Type 3: Additional Busses (direct)



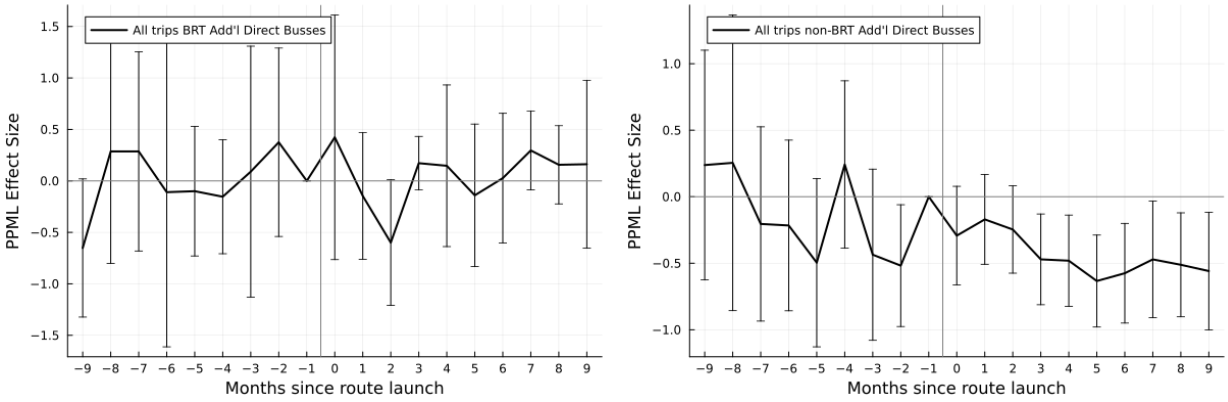
Notes: This graph replicates Figure 2 using 2000-meter square grids instead of 1km hexagonal grids.

Figure A.4: Robustness 500m Square Grids: **Aggregate Trip Volume** Impacts of BRT and Non-BRT Network Expansions

(a) Event Types 1 and 2: New Direct Route

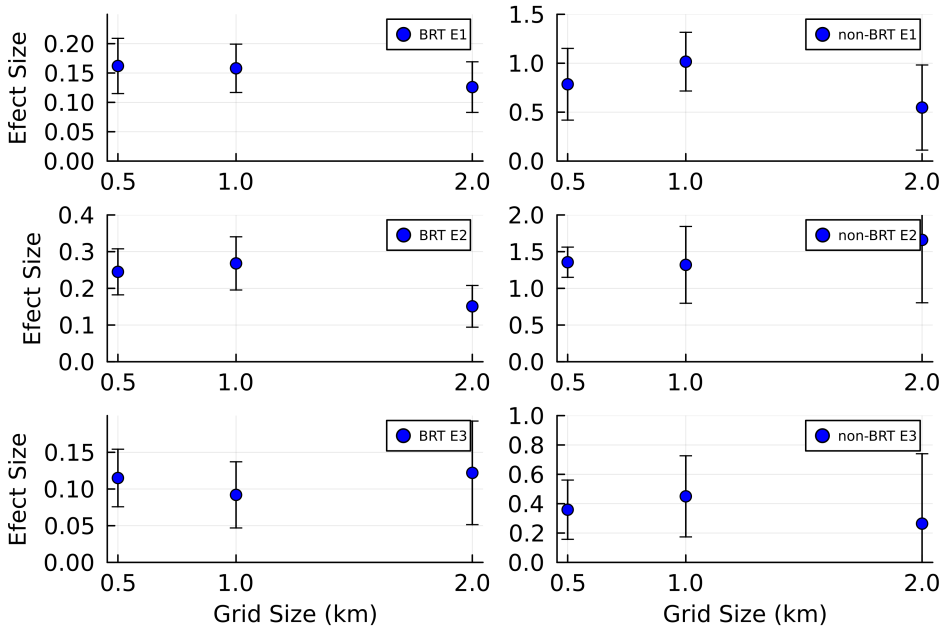


(b) Event Type 3: Additional Busses (Direct)



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

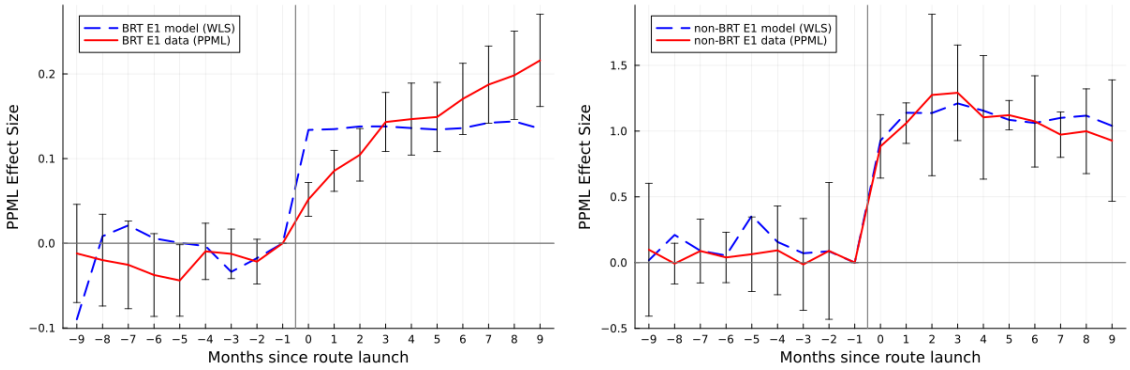
Figure A.5: Varying the Level of Aggregation: Impact of Network Expansion on Bus Ridership



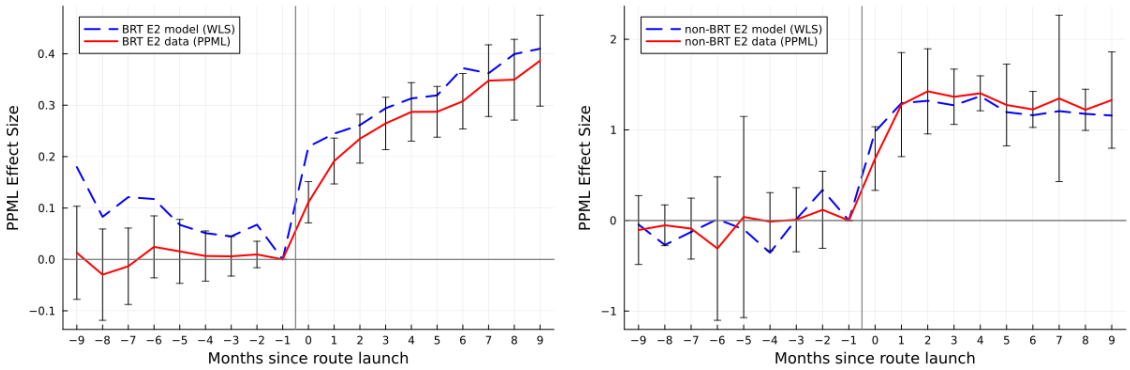
Notes: This graph plots the $Post_{odt}$ coefficients from our main specification with bus ridership as the outcome, for all six events, and for the three aggregation levels we consider: 500-meter square grids, 1000-meter hexagons, and 2000-meter square grids.

Figure A.6: Demand Estimation Model Fit (Targeted Moments)

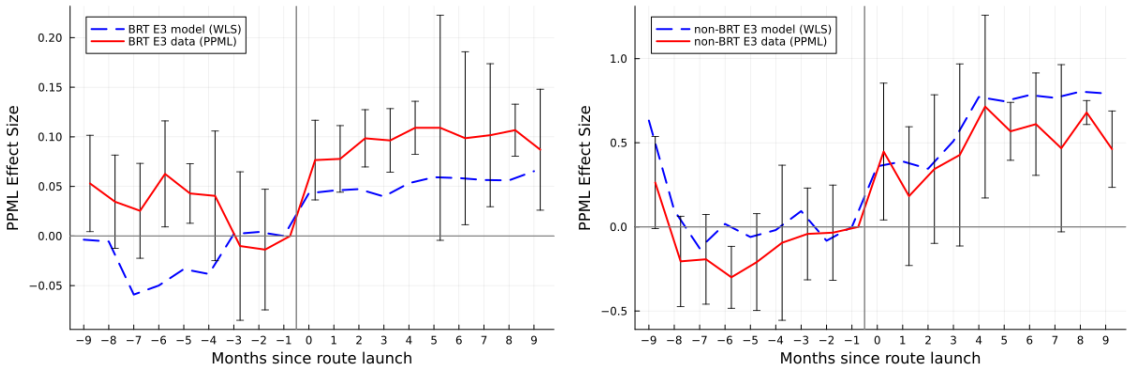
(a) Event Type 1: New Direct Route (Not Quicker)



(b) Event Type 2: New Direct Route (Quicker)



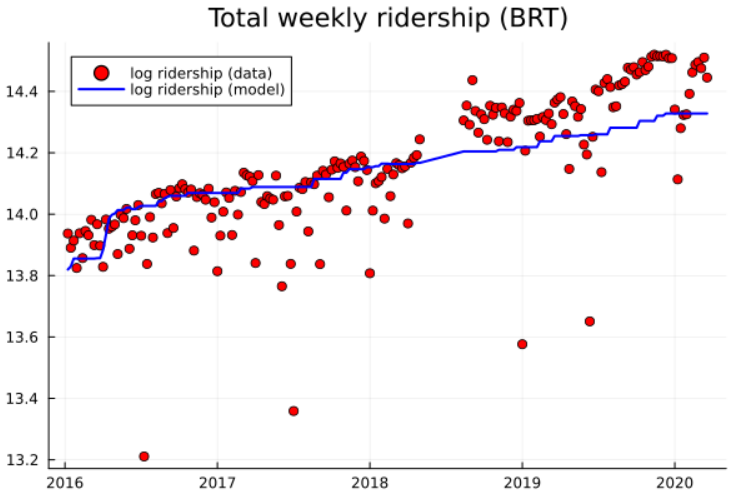
(c) Event Type 3: Additional Busses (Direct)



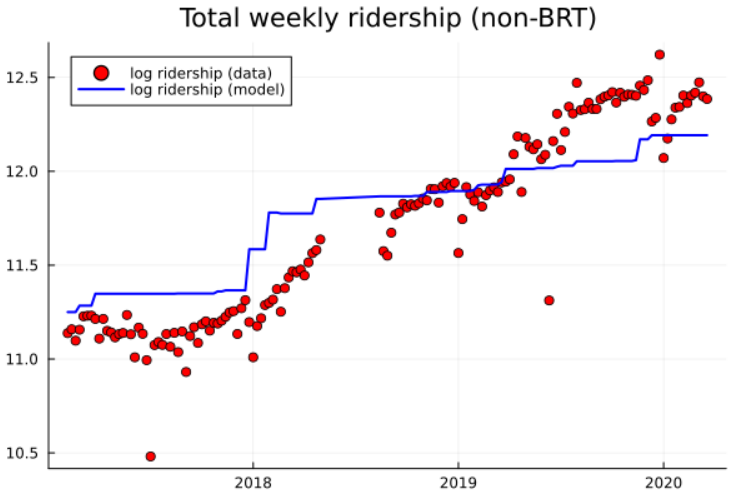
Notes: These figures overlay the event study analysis from Figure 2 with the same analysis on model-predicted ridership using the model estimated in Table 4, column 2. For model-predicted ridership, we use weighted least squares on log model ridership. All regressions have origin-week and destination-week fixed effects as in equation (1).

Figure A.7: Demand Estimation Model Fit Over Time

(a) BRT Ridership

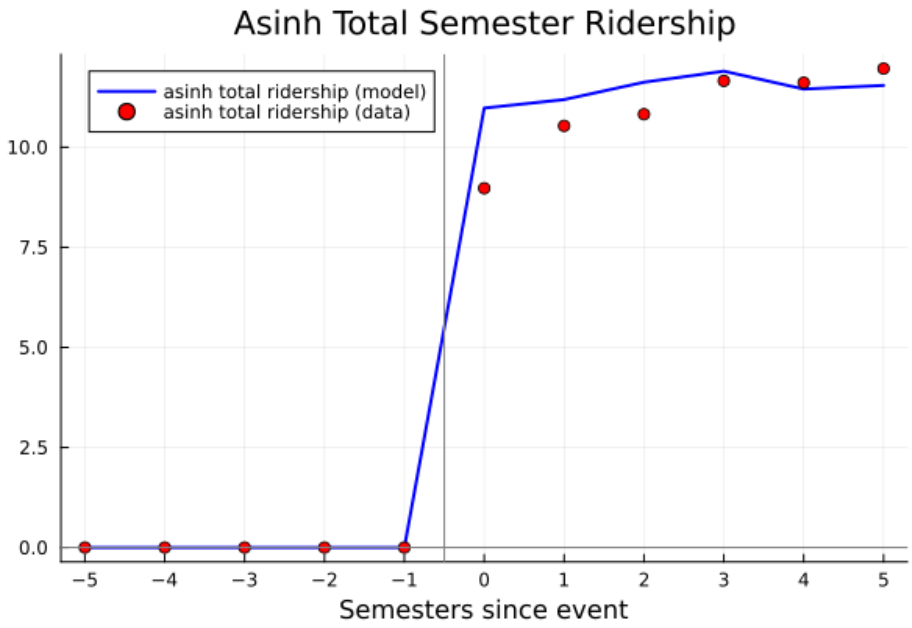


(b) Non-BRT Ridership



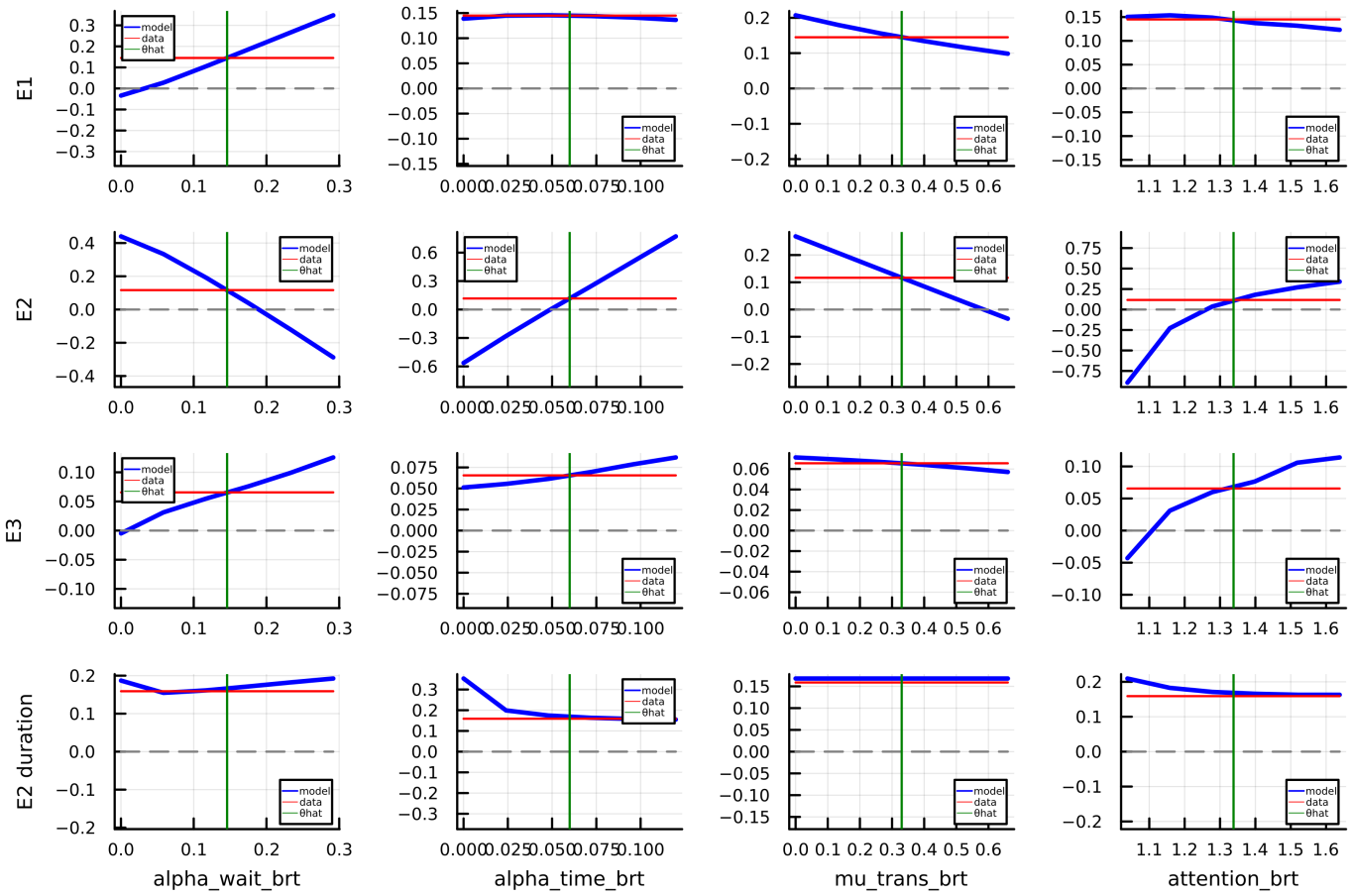
Notes: These figures report actual and model-predicted total ridership over the entire TransJakarta network over time at the weekly level, based on the model estimated in Table 4, column 2.

Figure A.8: Demand Estimation Model Fit for Non-BRT Route Launch



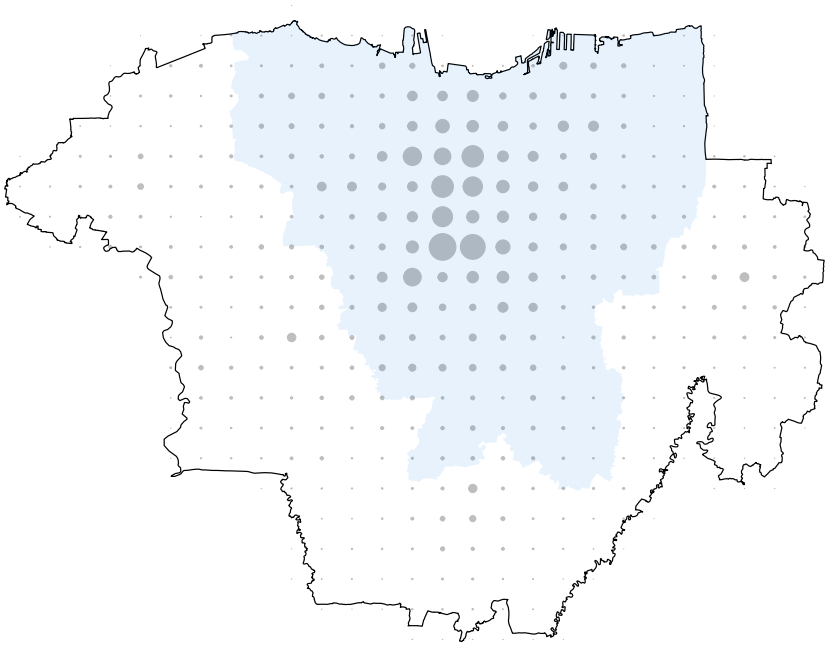
Notes: This figure reports actual and model-predicted ridership after the launch of the first non-BRT connection (direct or transfer). To construct this figure, we consider the 74 origin-destination pairs (using 1km hexagonal grids) that are part of the reduced form analysis and that first become connected through non-BRT after 2017, when our non-BRT ridership data begin. We then aggregate total bus ridership by semester after the route launch. (Note, the sample of origin destinations pairs changes by semester. However, 86% of the origin destination pairs appear in semester 3 after the route launch.). We then plot the inverse hyperbolic sine of total semester ridership in the data and in the model, based on the model estimated in Table 4 column 2.

Figure A.9: Moment Dependence on Parameters (Jacobian)



Notes: This figure shows how, in the model, the moments we use in estimation (rows) vary as a function of parameters (columns). Each graph shows how the value of one of the moments $m_j \in \{\alpha^{1B}, \alpha^{2B}, \alpha^{3B}, \alpha^{2B,\text{duration}}\}$ depends on one of the parameters $\theta_i \in \{\alpha^{\text{BRT}}_{\text{wait}}, \alpha_{\text{time}}, \mu^{\text{BRT}}_{\text{transfer}}, \eta^{\text{BRT}}\}$. For each plot, we use 10 values of θ_i that bracket the estimated value $\widehat{\theta}_i$. The thick blue line plots the model's moment value $m_j(\theta_i)$. The green vertical lines indicate the estimated parameter $\widehat{\theta}_i$ (column 2 in Table 4). The red horizontal lines indicate m_j^{data}, the value of the moments in the data.

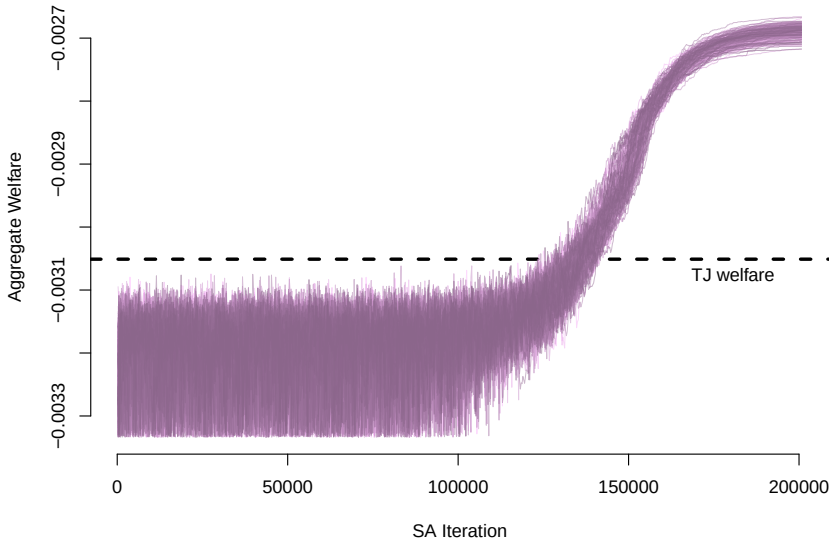
Figure A.10: Jakarta Geography Used for Optimal Networks



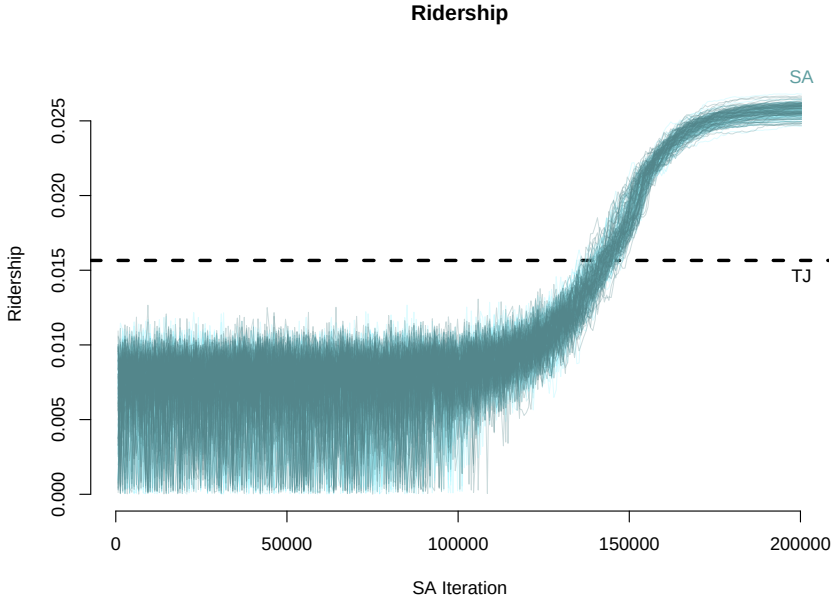
Notes: This map prints the geographic environment with 2km×2km grid cells used to compute optimal networks. The central DKI area is in light blue. The size of each circle is proportional to the number of trips that depart from that grid cell.

Figure A.11: Welfare and Ridership Over the Course of the Simulated Annealing Algorithm

(a) Welfare

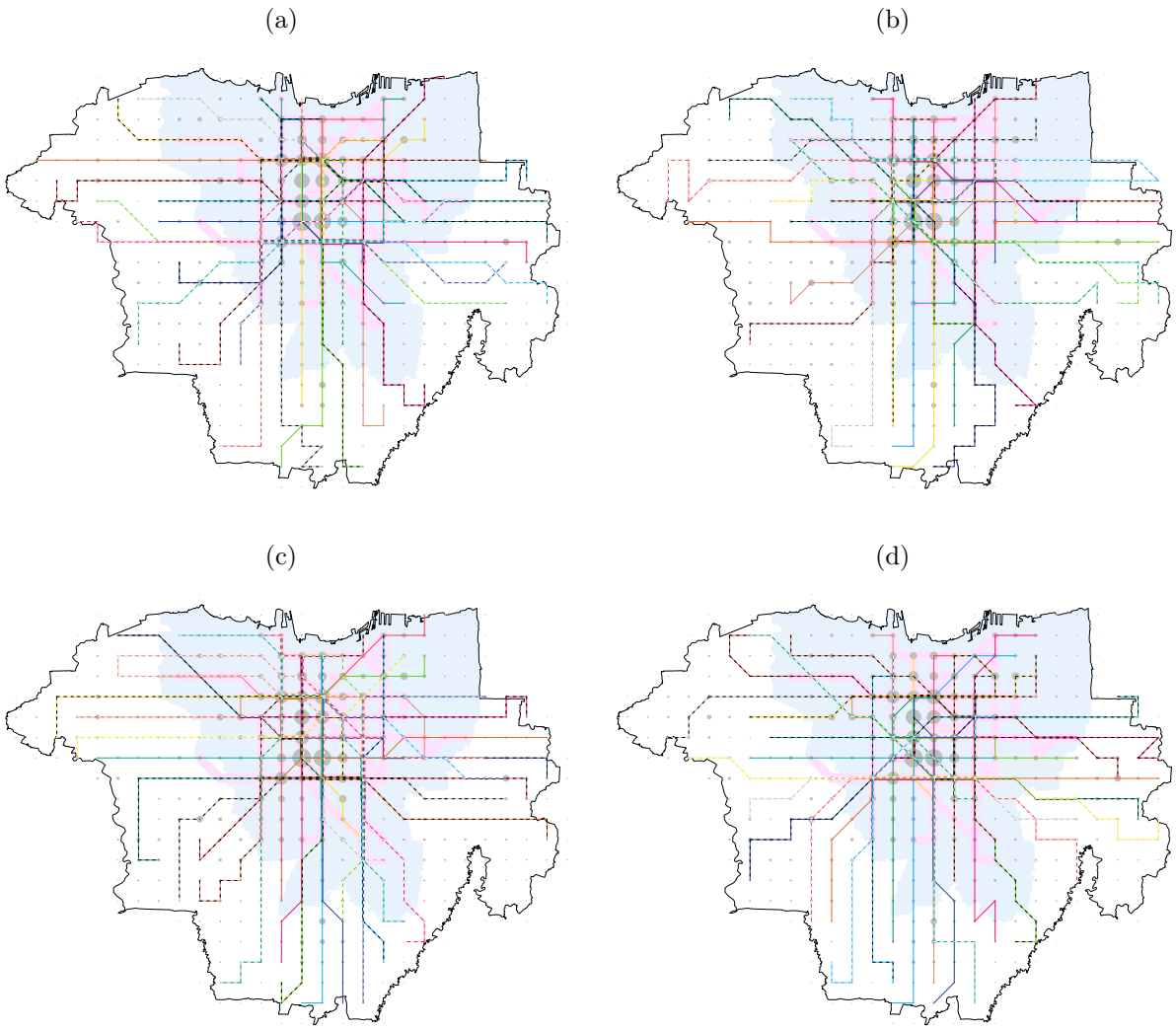


(b) Ridership



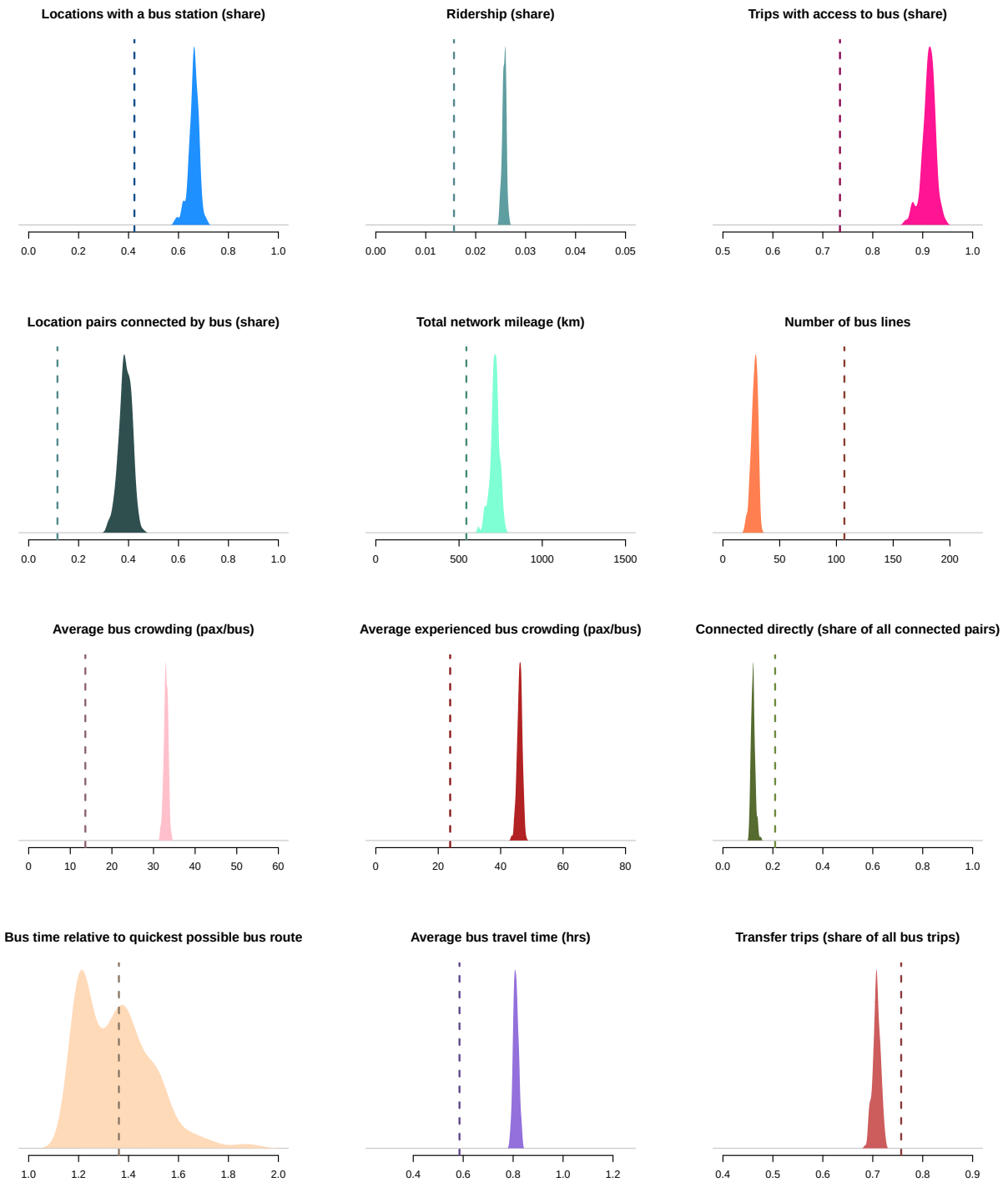
Notes: Each line shows the progression of welfare (panel a) and ridership (panel b) throughout one of the 200 runs of the simulated annealing algorithm.

Figure A.12: Optimal Network Examples



Notes: This figure depicts four additional examples of draws from the planner's distribution of optimal networks. Each is obtained from an independent simulated annealing run.

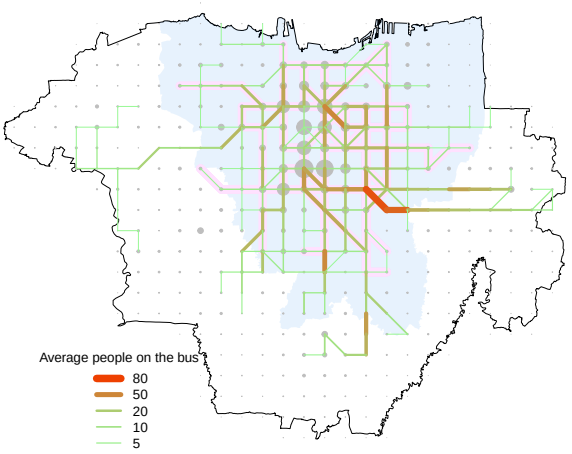
Figure A.13: Distributions of Optimal Network Properties



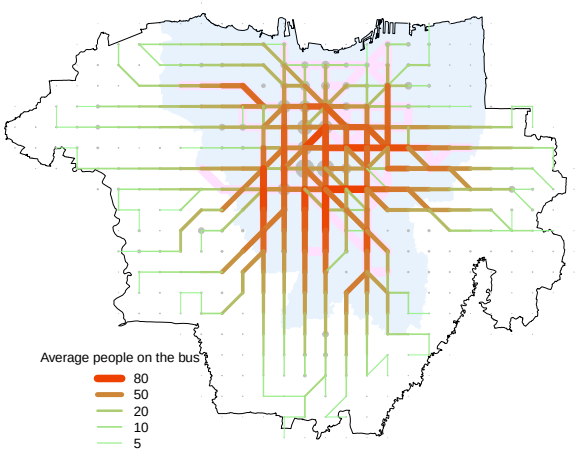
Notes: Each graph plots the kernel density graph of a specific network characteristic, for the final network from the simulated annealing algorithm, over the 200 parallel simulated annealing runs. The vertical line plots the measure for the current TransJakarta network for comparison.

Figure A.14: Average Bus Occupancy in Current and Optimal Networks

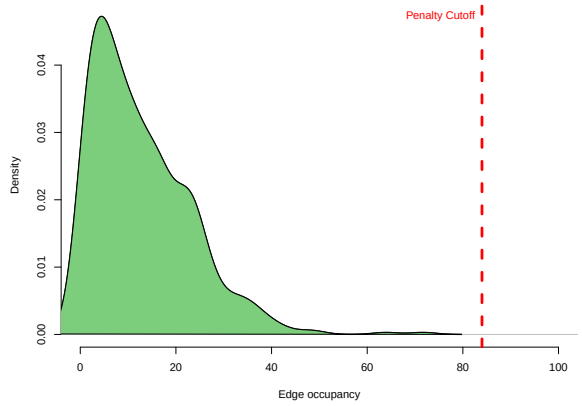
(a) Edge-level Bus Occupancy (Current Network)



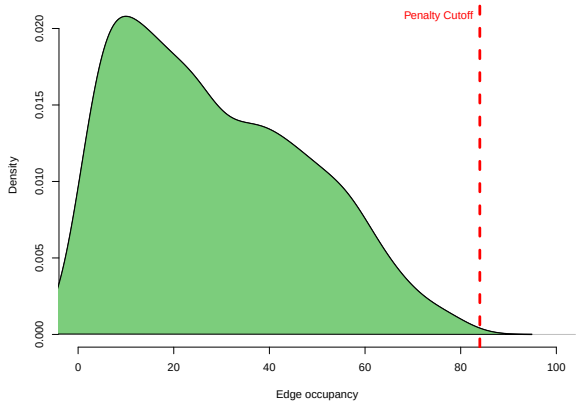
(b) Edge-level Bus Occupancy (Single SA Run)



(c) Distribution (Current Network)



(d) Distribution (Single SA Run)



Notes: This figure depicts edge-level bus occupancy, defined as the average number of riders on a typical bus traveling across an edge, for the current TransJakarta network (a) and one of the sampled networks (b). Edges on which TransJakarta has built infrastructure for BRT travel are denoted by bright pink underlay. To construct this graph, we use model-implied ridership percentages to compute how many people take each bus line and where they get on and off the line. We assume busses run for 17 hours each day (5AM-10PM). For each edge, the graph displays the average occupancy over all the lines using that edge. Figures (c) and (d) plot the distribution of edge-level occupancies, as well as the cutoff $\overline{B}$ of 84 people per bus. See Appendix A.7.2 for details.

A.2 Appendix Tables

Table A.1: Number of Transjakarta Routes by Type

	(1) No. of routes
BRT	43
Non-BRT (inner city)	84
Non-BRT (outer city)	12
Royaltrans	13
Mikrotrans	45
Others	17
Total	214

Notes: A snapshot of the TransJakarta network in August 2019.

Table A.2: BRT Route Launch Order: Balance on Baseline Geographical Variables

	(1) BRT	
	β	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	0.38
	(0.01)	
Average distance across stations on the new route to Sudirman CBD (in km)	0.33	0.28
	(0.01)	
# of existing baseline stations on the new route	-1.41	0.19
	(0.02)	
Share of existing baseline stations on the new route	0.01	0.82
	(0.00)	
# of connecting baseline routes on existing stations on the new route (avg)	0.03	0.75
	(0.00)	
# of connecting baseline routes on the new route in total	-0.07	0.89
	(0.01)	
Avg daily ridership across existing baseline stations first 7 days of Jan2016	-81.75	0.19
	(1.14)	
Median Initial Planned Buses Allocation (at launch)	-1.11***	0.01
	(0.01)	
N	28	

Notes: This table reports results from the OLS regression $Outcome_i = \beta WeekLaunch_i + \epsilon_i$ 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.3: Robustness with 500m Square Grids: Impact of Network Expansion on Travel Time, Wait Times, and Bus Ridership

(a) BRT

	log Min Travel Time			log Bus/hr (origin)			Bus Ridership		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
E1: New Direct Line	0.028***			-0.039			0.162***		
	(0.008)			(0.023)			(0.024)		
E2: New Direct Line (quicker)		-0.272***			0.069*			0.245***	
		(0.025)			(0.034)			(0.032)	
E3: Additional Busses			-0.018*			0.286***			0.115***
			(0.008)			(0.023)			(0.020)
Orig × Dest FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	PPML	PPML	PPML
Mean outcome	31.9	43.3	11.9	33.6	27.1	20.0	69.4	48.3	122.2
Median outcome	29.4	41.2	9.9	26.8	21.5	16.9	32.4	21.2	48.1
Unique origin x destination pairs	29,518	29,657	8,027	29,518	29,657	8,027	29,518	29,657	8,027
Unique origins	185	185	190	185	185	190	185	185	190
N	4,962,030	4,964,566	1,202,193	4,962,030	4,964,566	1,202,193	4,962,030	4,964,566	1,202,193
R^2	0.982	0.980	0.998	0.928	0.926	0.998			

(b) Non-BRT

	log Min Travel Time			log Bus/hr (origin)			Bus Ridership		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
E1: New Direct Line	0.023			0.283***			0.785***		
	(0.012)			(0.036)			(0.187)		
E2: New Direct Line (quicker)		-0.715***			0.226***			1.356***	
		(0.072)			(0.032)			(0.105)	
E3: Additional Busses			-0.085***			0.427***			0.359***
			(0.015)			(0.015)			(0.103)
Orig × Dest FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	PPML	PPML	PPML
Mean outcome	27.8	50.7	12.2	10.3	9.9	7.3	12.0	2.9	6.4
Median outcome	21.4	49.4	10.5	8.5	7.1	6.4	0.0	0.0	0.0
Unique origin x destination pairs	6,656	6,749	3,890	6,656	6,749	3,890	6,656	6,749	3,890
Unique origins	110	110	139	110	110	139	110	110	139
N	755,706	765,671	435,225	755,706	765,671	435,225	755,706	765,671	435,225
R^2	0.993	0.990	0.998	0.967	0.968	0.992			

Notes: Version of Table 1 using 500-meters square grids. The two panels report results for BRT and non-BRT events, respectively. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.4: Robustness With 2000m Square Grids: Impact of Network Expansion on Travel Time, Wait Times, and Bus Ridership

(a) BRT

	log Min Travel Time			log Bus/hr (origin)			Bus Ridership		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
E1: New Direct Line	0.017			-0.042			0.126***		
	(0.011)			(0.034)			(0.022)		
E2: New Direct Line (quicker)		-0.298***			0.063			0.151***	
		(0.036)			(0.048)			(0.029)	
E3: Additional Busses			-0.092**			0.348***			0.122***
			(0.028)			(0.069)			(0.036)
Orig × Dest FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	PPML	PPML	PPML
Mean outcome	37.5	46.3	16.2	47.0	37.5	20.4	465.3	384.7	841.3
Median outcome	36.7	45.1	12.6	38.9	27.6	16.9	315.3	246.5	459.1
Unique origin x destination pairs	3,868	3,869	1,405	3,868	3,869	1,405	3,868	3,869	1,405
Unique origins	68	68	70	68	68	70	68	68	70
N	676,039	672,111	205,889	676,039	672,111	205,889	676,039	672,111	205,889
R^2	0.984	0.979	0.996	0.940	0.938	0.996			

(b) Non-BRT

	log Min Travel Time			log Bus/hr (origin)			Bus Ridership		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
E1: New Direct Line	0.038*			0.270***			0.547*		
	(0.017)			(0.053)			(0.222)		
E2: New Direct Line (quicker)		-0.655***			0.152*			1.660***	
		(0.119)			(0.059)			(0.437)	
E3: Additional Busses			-0.121**			0.388***			0.264
			(0.039)			(0.022)			(0.243)
Orig × Dest FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	PPML	PPML	PPML
Mean outcome	45.6	66.2	20.1	10.4	9.2	6.6	117.7	27.6	55.7
Median outcome	32.8	66.2	19.3	10.6	7.8	6.3	25.8	0.0	14.1
Unique origin x destination pairs	422	435	156	422	435	156	422	435	156
Unique origins	16	16	16	16	16	16	16	16	16
N	53,315	54,412	19,078	53,315	54,412	19,078	53,315	54,412	19,078
R^2	0.989	0.983	0.998	0.942	0.944	0.990			

Notes: Version of Table 1 using 2000-meters square grids. The two panels report results for BRT and non-BRT events, respectively. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5: Robustness With 500m Square Grids: Impact of Network Expansion on Aggregate Trip Volume

(a) BRT

	log Min Travel Time			log Bus/hr (origin)			Bus Ridership			All trips		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
E1: New Direct Line	0.047***			-0.134***			0.106***			-0.057		
	(0.006)			(0.027)			(0.018)			(0.119)		
E2: New Direct Line (quicker)		-0.212***			-0.123***			0.199***			0.046	
		(0.027)			(0.028)			(0.027)			(0.084)	
E3: Additional Busses			0.012			0.235***			0.094*			-0.007
			(0.032)			(0.067)			(0.040)			(0.314)
Orig × Dest FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	PPML	PPML	PPML	PPML	PPML	PPML
Mean outcome	31.6	39.7	11.8	34.1	33.0	15.2	83.4	53.8	98.3	237.6	176.5	572.1
Median outcome	28.7	38.6	10.1	29.4	28.7	16.9	38.6	27.8	63.0	0.0	0.0	0.0
Unique origin x destination pairs	21,813	21,852	3,960	21,813	21,852	3,960	21,813	21,852	3,960	21,813	21,852	3,960
Unique origins	156	156	123	156	156	123	156	156	123	156	156	123
N	1,850,929	1,854,359	310,496	1,850,929	1,854,359	310,496	1,850,929	1,854,359	310,496	1,850,929	1,854,359	310,496
R^2	0.988	0.986	0.999	0.949	0.949	1.000						

(b) Non-BRT

	log Min Travel Time			log Bus/hr (origin)			Bus Ridership			All trips		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
E1: New Direct Line	0.023			0.268***			0.736***			-0.239		
	(0.015)			(0.035)			(0.216)			(0.177)		
E2: New Direct Line (quicker)		-0.737***			0.235***			1.276***			0.071	
		(0.084)			(0.033)			(0.147)			(0.129)	
E3: Additional Busses			-0.088***			0.342***			0.394**			-0.223
			(0.023)			(0.018)			(0.128)			(0.138)
Orig × Dest FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	PPML	PPML	PPML	PPML	PPML	PPML
Mean outcome	28.5	46.5	11.6	10.3	10.2	7.6	10.3	3.3	9.4	446.4	704.3	1238.0
Median outcome	22.1	46.7	10.1	8.5	7.8	7.8	0.0	0.0	0.0	0.0	0.0	0.0
Unique origin x destination pairs	4,728	4,770	2,481	4,728	4,770	2,481	4,728	4,770	2,481	4,728	4,770	2,481
Unique origins	90	90	110	90	90	110	90	90	110	90	90	110
N	371,173	374,650	173,907	371,173	374,650	173,907	371,173	374,650	173,907	371,173	374,650	173,907
R^2	0.996	0.993	0.998	0.979	0.980	0.996						

Notes: This table replicates Table 2 using 500-meter square grids instead of 1000 hexagonal grids. The two panels report results for BRT and non-BRT events, respectively. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.6: Heterogeneity by Poverty Level: Impact of Network Expansion on Bus Ridership

	Bus Ridership					
	(1)	(2)	(3)	(4)	(5)	(6)
New Direct Line	0.215***	0.326***		0.957***	1.416**	
	(0.030)	(0.057)		(0.250)	(0.453)	
New Direct Line x High Poverty Origin	-0.095**	-0.052		-0.078	0.500	
	(0.032)	(0.061)		(0.147)	(0.429)	
E3: Additional Busses			0.074*			0.329
			(0.036)			(0.236)
E3: Add'l Busses x High Poverty Origin			-0.049			0.658**
			(0.034)			(0.216)
Estimator	PPML	PPML	PPML	PPML	PPML	PPML
Event Type	BRT 1	BRT 2	BRT 3	non-BRT 1	non-BRT 2	non-BRT 3
Median outcome pre x High Poverty	67.8	38.4	84.3	0.0	0.0	0.0
Median outcome pre x Low Poverty	50.0	39.6	104.6	0.0	0.0	0.0
N	3,154,672	3,143,019	793,965	306,722	313,994	144,763
R^2						

Notes: This table reports heterogeneity by whether the origin grid is high- or low-poverty for the bus ridership impacts in Table 1. We use poverty and population data at the kelurahan level from (SMERU, 2014) and PODES 2010, which we assign at the grid cell level based on surface area intersection. The interaction variable is a dummy for within-sample above median poverty rate. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.7: Moment Dependence on Parameters (Jacobian)

	Event 1 (α^{1B})	Event 2 (α^{2B})	Event 3 (α^{3B})
$\alpha^{\text{BRT}}_{\text{wait}}$	1.37	-2.62	0.27
α_{time}	0.01	10.72	0.52
$\mu^{\text{BRT}}_{\text{transfer}}$	-0.16	-0.46	-0.02

Notes: The (i, j) entry in this matrix reports how the j-th moment varies locally in the model when the i-th parameter changes, namely $\partial m_j(\theta)/\partial\theta_i$, for moments $m_j \in \{\alpha^{1B}, \alpha^{2B}, \alpha^{3B}\}$ and parameters $\theta_i \in \{\alpha^{\text{BRT}}_{\text{wait}}, \alpha_{\text{time}}, \mu^{\text{BRT}}_{\text{transfer}}\}$.

Table A.8: Sensitivity Measure: Estimated Parameter Dependence on Moment Values

	Event 1 (α^{1B})	Event 2 (α^{2B})	Event 3 (α^{3B})
$\alpha_{\text{wait}}^{\text{BRT}}$	4.9	-2.2	49.3
α_{time}	-4.1	2.4	38.0
$\mu_{\text{transfer}}^{\text{BRT}}$	-15.6	-2.3	52.5

Notes: The (i, j) entry in this matrix reports the (Andrews et al., 2017) sensitivity measure $(\widehat{SE}_i)^{-1}\partial\widehat{\theta}_i/\partial m_j^{\text{data}}$ where the scaling factor $\widehat{SE}_i$ is the estimated standard error of $\widehat{\theta}_i$.

Table A.9: Optimal Networks Local Comparative Statics

			Comparative Statics (Local Changes)		
Statistic	Current Network	Baseline Optimal	Wait Time $dF^*/d\alpha_{\text{wait}}$	Time on Bus $dF^*/d\alpha_{\text{time}}$	Transfer $dF^*/d\mu_{\text{transfer}}$
Panel A: Coverage measures					
Locations with a station (share)	0.42	0.66	-0.08 [-0.14; -0.028]	-0.065 [-0.2; 0.062]	0.0061 [-0.0065; 0.02]
Location pairs connected by bus (share)	0.12	0.39	-0.075 [-0.14; -0.013]	-0.014 [-0.19; 0.18]	-0.0013 [-0.017; 0.016]
Total network mileage (in km)	543.95	713.87	-78.45 [-157.11; -6.095]	-99.87 [-301.061; 84.056]	-0.62 [-16.77; 16.24]
Panel B: Speed measures					
Bus time relative to quickest possible bus route	1.36	1.35	0.16 [-0.25; 0.54]	0.76 [-0.053; 1.58]	-0.14 [-0.25; -0.054]
Panel C: Directness measures					
Connected directly (share of all connected pairs)	0.21	0.12	0.019 [-0.0035; 0.042]	0.021 [-0.024; 0.062]	-0.01 [-0.021; -0.002]

Notes: This table reports how network shape measures (rows) change locally as a function of parameter changes (columns). It replicates Table 6 for local parameter changes, using the sample analogue of equation (7). Column 3 reports local changes to both $\alpha_{\text{wait}}^{\text{BRT}}$ and $\alpha_{\text{wait}}^{\text{non-BRT}}$, column 5 reports local changes to both $\mu_{\text{transfer}}^{\text{BRT}}$ and $\mu_{\text{transfer}}^{\text{non-BRT}}$. Note that a positive change in the transfer shifter corresponds to a *lower* transfer penalty. 90% bootstrapped confidence intervals (columns 3-5) in parentheses.

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. 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.

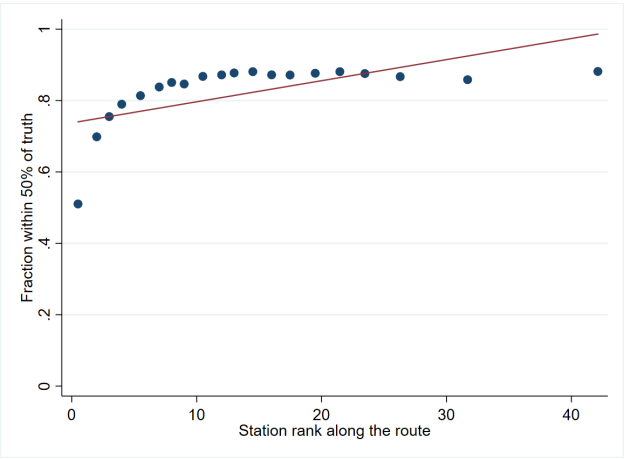
We developed three algorithms to identify when a bus arrives at a given bus station. When both GPS data and trip logs are available, we map match the bus GPS locations to the path of the bus route (from separate data), starting from the trip start time recorded in the trip log, and find arrival times 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.

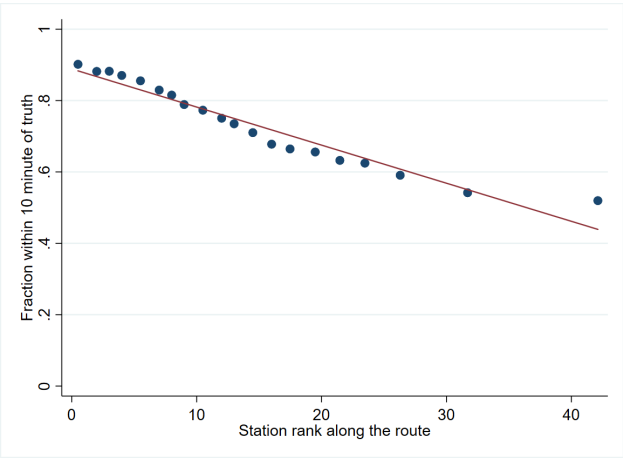
When only trip log data is available, the algorithm proceeds in two steps. First, we predict *station* arrival time as a function of trip start time, bus route, and time of the day. We estimate this model using the output of the first algorithm for the same route, on bus-days when both GPS and trip log data is available. Figure A.15 shows that we achieve high accuracy using this model. Second, we apply these predicted times using the trip start time from the log.

Figure A.15: Accuracy of Predicting Station Arrival Time Based Only on Trip Start Time

(a) Relative Error



(b) Absolute Error



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 for defining Events 1 and 2, 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.16 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). Travel time and delay in different years are strongly correlated after including route, origin and destination fixed effects (Table A.10).

Figure A.16: Bus Travel Delay by Departure Time and Year

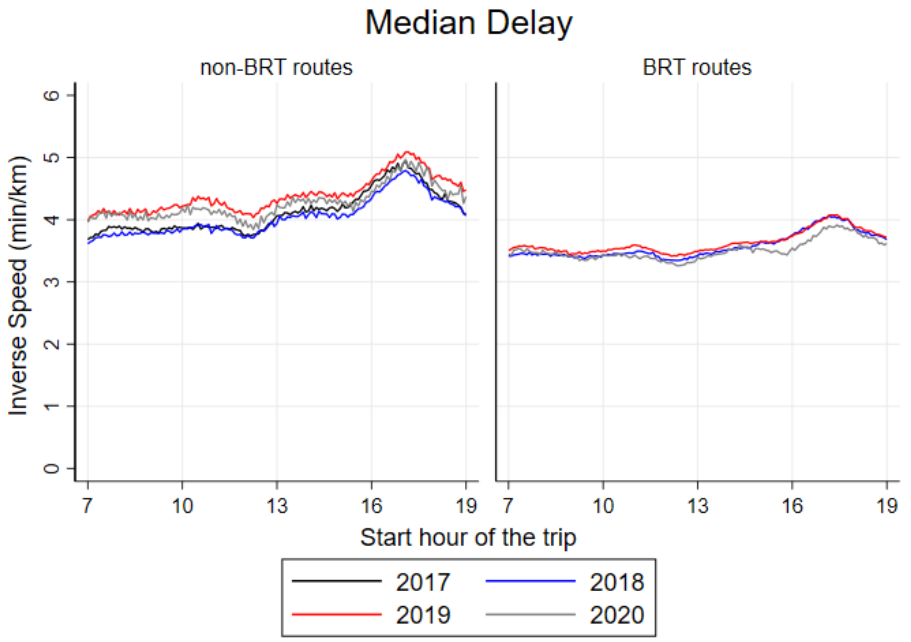


Figure Notes: This figure reports median delay (inverse speed, in minutes per km) by departure time, separately for BRT routes and for non-BRT routes using data from January 2017 - March 2020 for routes that were active throughout this period. To construct this figure, we assign each trip to 5-minute 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 traveled 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.10: **Bus Travel Time Correlation Across Time**

	Log(Median Delay) 2017	Log(Median Delay) 2018	Log(Median Travel Time) 2017	Log(Median Travel Time) 2018
	(1)	(2)	(3)	(4)
Log(Median Delay) 2019	0.716***	0.794***		
	(0.048)	(0.038)		
Log(Median Travel Time) 2019			1.031***	1.009***
			(0.007)	(0.002)
Constant	0.368***	0.222***	-0.193***	-0.083***
	(0.062)	(0.047)	(0.043)	(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) and log of median travel time. 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 from 2019.

We first calculate bus headways (difference in minutes between two consecutive buses) for every station, route, and direction. Then, we calculate the frequency of headway occurrence by route and minute. Assuming that TransJakarta passengers are "non-planning" and thus arrive at bus stations at a (locally) constant rate, we calculate the implied histogram of wait times by taking the reverse cumulative total of headway frequency. We compute the wait time distribution for 45 routes (18 BRT and 27 non-BRT) that have only two main route versions (i.e. where the top two trip variants account for $> 95\%$ of all trips on that route).

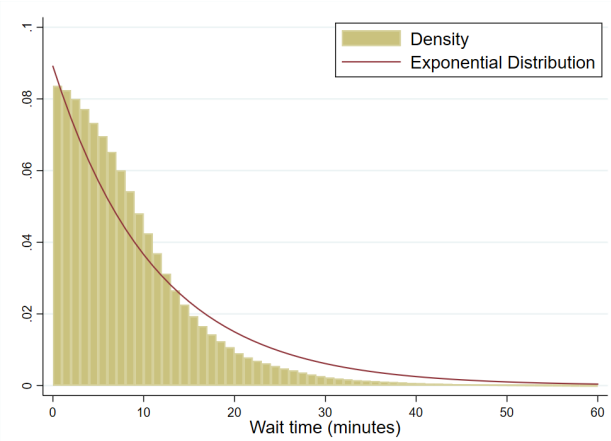
Panels A and B in Figure A.18 plot the wait time distribution for non-BRT route 1E. The wait time distribution is approximately exponential, except for wait times over 30 minutes, where the empirical distribution puts lower weight than the exponential. Panel B plots results for all the routes in our sample, showing that there is a tight linear relationship between a route's mean wait times and the coefficient of variance of wait time. BRT and non-BRT routes share the same relationship. In other words, BRT routes are not less (or more) uncertain than non-BRT routes, conditional on average wait time.

Figure A.17: Sample Arrival Time Monitor at a TransJakarta Station

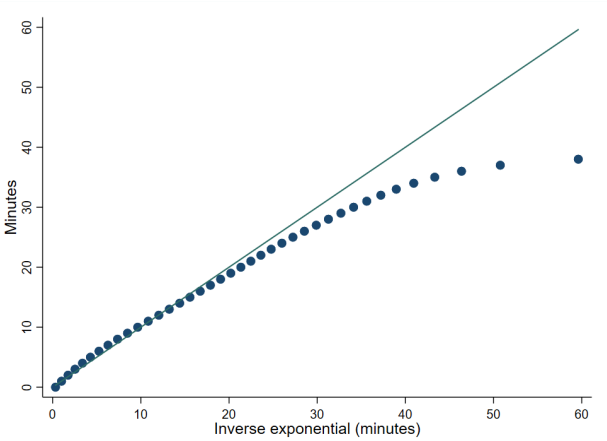
Notes: During the study period, all BRT bus stops were equipped with screens that displayed real-time estimated bus wait times. (The estimates were based on real-time bus GPS location data.)

Figure A.18: Wait Time Distribution Within and Across Routes

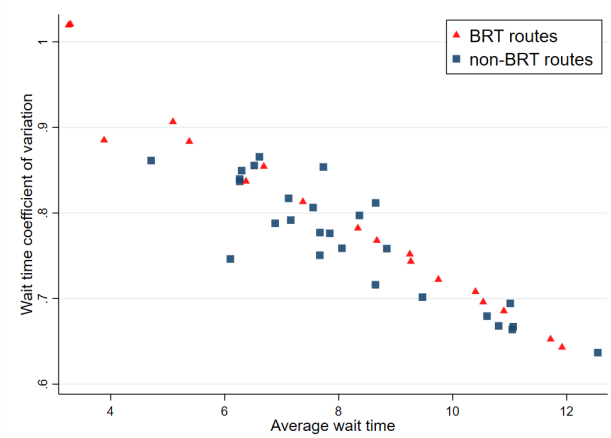
(a) Wait Time Distribution (route 1E)



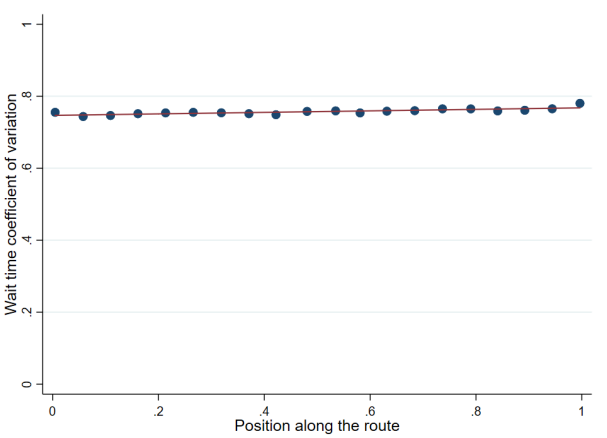
(b) Wait Time Q-Q Plot (route 1E)



(c) Wait Time Reliability on BRT and Non-BRT Routes



(d) Within-Route Reliability



Notes: Panel (A) and (B) report the wait time distribution for non-BRT route 1E. To create these graphs, we calculate bus headways (defined as the duration in minutes between two consecutive buses of the same route and direction at a particular station) 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 (A), we fit an exponential distribution to the empirical wait time distribution. Panel (C) 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. Panel (D) reports the coefficient of variation of wait time over position along the route (0 for first station, 1 for last station). For all panels, we use data from January to October 2019 and restrict the sample to only include arrival times during morning peak hours (8 AM-12 PM). Each Transjakarta route may have several route variants that differ slightly in length and stations reached. Therefore, for panels (C) and (D), we restrict the sample to routes where two trip variants account for above 95% of all trips on that, consisting of 45 routes (18 BRT and 27 non-BRT routes).

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. In 2019, 57% of transactions were cash. 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 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 for **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 all trips in the system.

Second, we process "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.

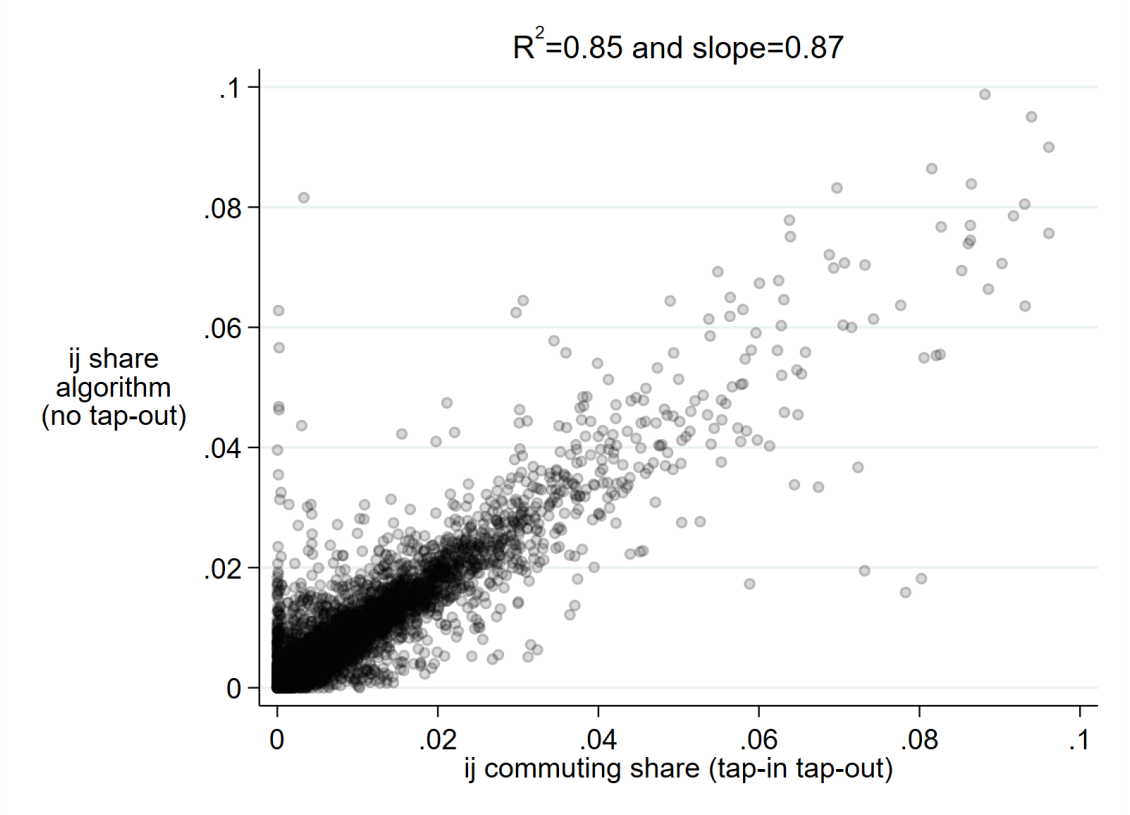
We exclude 15 non-BRT routes from ridership coding for reasons related to the way payments were made. Each non-BRT bus accepts payments with smart cards provided by a specific bank. We lack detailed micro-transaction data corresponding to one of the banks. Overall, tap data from dropped routes accounts for 14.96% of total tap data in 2017-2020. Coverage is 44.11% of total routes in 2017 and 72.14% in 2020.

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. Commuters were not required to tap-out on non-BRT routes, and tap-out was enforced only at some BRT bus stops. Given these limitations, we developed an algorithm

to infer a commuter's destination for each ridership tap-in. The algorithm consists of three main rules for assigning a trip's destination. The first rule is to use the actual tap-out location as the destination, when the tap-out occurs within 4 hours after the tap-in. For the second rule, we proxy for the destination using the next trip's tap-in, if that next trip takes place on the same day or the next day. For the third rule, for each smart card, each month, we computed the top 2 stations with the most taps. If the trip origin is one of the top 2 stations, we proxy for the trip destination using the other top 2 stations. This method is only applied for frequent commuters, defined as those with more than 10 taps in a month and more than 75% of total taps located within its top-2 stations.

We validate the algorithm by comparing ridership flow destination shares originating from station i towards station j (destination) using data with only tap-in entry (algorithm) and data with both tap-in and tap-out entry (actual). Figure shows that the two measures are highly correlated.

Figure A.19: Comparison of Actual and Inferred (o, d) Ridership Shares



A.3.5 Veraset Smartphone Location Data Trip Processing

We use the smartphone trip processing algorithm from Kreindler (2023) to convert raw GPS data into individual trips and common location for each device in the data.

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 labeled as "home."

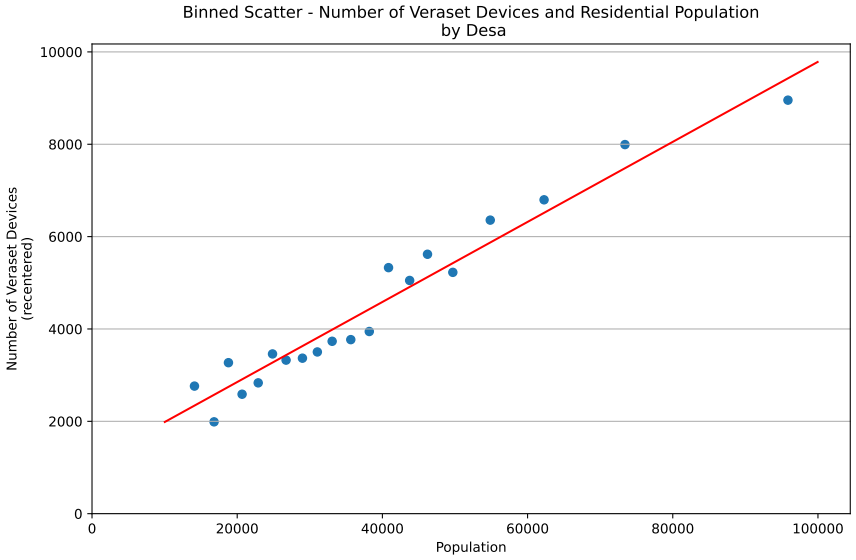
The sample of trips used in the analysis is all weekday trips starting and ending at a known locations (i.e. locations classified by DBSCAN as belonging to a cluster), excluding trips starting before 5AM or after 11PM. We also drop trips that are unusually long or short (in both duration and distance), "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 A.20). 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.

We construct two main data sets using the Veraset trips data. First, we construct a panel of trip flows at the origin grid by destination grid by week level. Each week, for each device in the data, we re-weight all its weekday trips that week to represent a single typical weekday. (This is essentially using inverse weights given by the number of weekdays when we observe the decide, except that we also account for partial days.) We also use a single overall adjustment factor such that the sum of all device weights equals approximately 14 million, the number of individuals over 15 years old in the study area. To construct trips, we consider any individual-day with sufficient GPS data frequency and include all the trips that can be identified in the data on that day, and use the trip detection algorithm from Kreindler (2023).

Second, we construct a cross-section of typical (o, d) trip flows throughout the period. We apply the same procedure as for the panel, except that we first pool all data for the entire two year period.

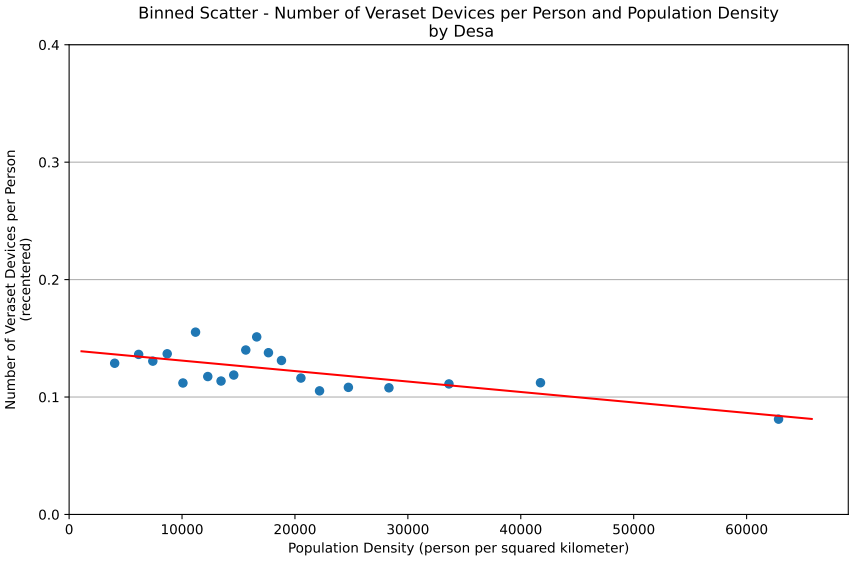
Figure A.20: Binned Correlation Between Veraset Devices and Residential Population



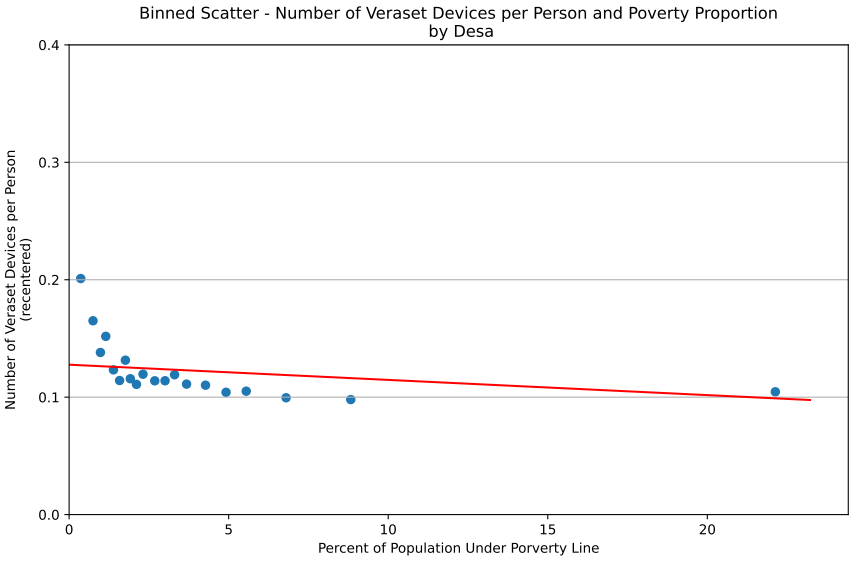
Notes: Each observation is an urban neighborhood (kelurahan or desa), $N = 538$. Population data is from the 2011 PODES survey, the most recent source for population at this level of geographical detail. For each desa, we count all devices which have their "home" location assigned inside the desa.

Figure A.21: Representativeness of the Veraset Trip Data

(a) Population Density



(b) Poverty



Notes: Binscatter plots analyzing the representativeness of the Veraset trip data at the kelurahan level. The sample excludes kelurahans below the 10th percentile of the distribution of PODES population. These locations tend to represent commercial or leisure areas. In these locations, the ratio of Veraset to PODES population is high. Panel A compares the ratio of Veraset home locations in a kelurahan to the kelurahan's population density. Panel B uses a poverty indicator from SMERU (2014).

A.4 Reduced Form Estimation

This section describes the differences-in-difference specifications.

For BRT events, we define $Post^{iB}_{odt}$, $i = 1, 2$, 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.[52] The coefficient α^{iB} in equation 1 captures the overall effect of a new direct connection being added between o and d in the first 10 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.[53]

For non-BRT events, $Post^{iN}_{odt}$, $i = 1, 2$, 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, and o and d each have at least one treated observation. (In the latter case, the first leg is necessarily non-BRT, while the second leg can be BRT or non-BRT.)

In Event 3, grid-cell pairs o and d that are already directly connected get more busses from a new route because it overlaps with the existing route for the portion between o and d. Specifically, $Post^{3}_{odt}$ 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 busses are added to the route between o and d due to an additional direct route.[54] The sample is all odt such that BRT grid cells o and d are directly connected at time t, o and d each have at least one treated observation, and o and d are only connected by direct connections before the event.[55]

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

[52] We impose that the origin and destination grids have BRT stations, while the new route may be a non-BRT route traveling between these locations. It is often the case that non-BRT route travel along BRT corridors and stop at BRT stations for a portion of the route.

[53] This sample includes o, d pairs that are treated at some point, as well as o, d pairs that are never treated, yet there exist o', d' such that o, d' and o', d are treated.

[54] As for event types 1 and 2, the route itself can be non-BRT as long as it passes through BRT grid cells o and d.

[55] We make this last restriction to focus on the case where the change in the choice set is very simple. Results where we include o, d pairs that have both direct and transfer connections before the event are similar.

A.5 Attention Probabilities

We incorporate partial inattention by assuming that the agent notices each arrival from option k with independent probability p_k. In other words, with probability $1 - p_k$ the agent fails to notice the first arrival for option k. This leads to an "effective" arrival rate for option k of $\widetilde{\lambda}_k = \phi(p_k)\lambda_k$ with $0 \leq \phi(p) \leq 1$. $\phi(p)$ is given by $-\frac{p\log(p)}{1-p}$, which is increasing and concave in p, $\phi(0) = 0$, and $\phi(1) = 1$. To see this, note that the effective arrival rate is a weighted sum of the first arrival rates, the second arrival, etc. Dropping the k subscript, we can write:

$$\begin{aligned}\widetilde{\lambda} &= p\lambda + (1-p)p\frac{\lambda}{2} + (1-p)^2 p\frac{\lambda}{3} + ... \\ &= p\lambda(\sum_{k=0}^{\infty} k^{-1}(1-p)^k) \\ &= -\frac{p\log(p)}{1-p}\lambda.\end{aligned}$$

The expression on the second line is the Mercator series for $p - 1$.

A.6 Demand Model Derivations

Proof of Proposition 2. Here we derive expressions for the probability π_k to choose option 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 $X_1, X_2, \ldots, X_N$, then

$$\Pr(k \in \arg\max_j X_j) = \int_{-\infty}^{\infty} f_k(x) \prod_{i \neq k} F_i(x) dx, \tag{9}$$

where f denotes a pdf function, and F denotes a cdf function. Expected utility is

$$\mathbb{E}\max_k X_k = \int_{-\infty}^{\infty} x \sum_k f_k(x) \prod_{i \neq k} F_i(x) dx. \tag{10}$$

In our model, have $u_k = v_k - \alpha_{\text{wait}} w_k$ where w_k is exponentially distributed with parameter λ_k. Assume that $\alpha_{\text{wait}} = 1$, if necessary replacing all λ_k by $\lambda_k/\alpha_{\text{wait}}$. 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.

Choice Probabilities. Assume that options are ranked such that $v_1 < v_2 < \cdots < v_N$. Replacing the pdf and cdf in (9) and separating the integral by intervals delimited by the v_k's, the probability that option k is optimal is:

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

(For any $u > v_k$, the probability that k is optimal is zero.)

We use the notation: $\Lambda_i = \sum_{\ell \geq i} \lambda_\ell$ and $M_i = \sum_{\ell \geq i} \lambda_\ell v_\ell$. We have

$$\begin{aligned}
\lambda_k^{-1} \pi_k &= e^{-M_1} \int_{-\infty}^{v_1} e^{u \Lambda_1} du + \cdots + e^{-M_k} \int_{v_{k-1}}^{v_k} e^{u \Lambda_k} du \\
&= \sum_{i=1}^{k} e^{-M_i} \int_{v_{i-1}}^{v_i} e^{u \Lambda_i} du \\
&= \sum_{i=1}^{k} e^{-M_i} \frac{e^{v_i \Lambda_i} - e^{v_{i-1} \Lambda_i}}{\Lambda_i},
\end{aligned}$$

were we use the convention $v_0 = -\infty$.

Expected Utility. Plugging the exponential pdf and cdf formulae in (10) we get

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

The $\exp(\cdot)$ term can be factored out of the sum as it does not depend on k, so we get

$$\begin{aligned}
\mathbb{E}\max_k u_k &= \sum_{i=1}^{N} \int_{v_{i-1}}^{v_i} u\Lambda_i e^{u\Lambda_i - M_i} du \\
&= \sum_{i=1}^{N} \Lambda_i^{-1} e^{-M_i} \int_{v_{i-1}}^{v_i} u\Lambda_i e^{u\Lambda_i} d(u\Lambda_i) \\
&= \sum_{i=1}^{N} \Lambda_i^{-1} e^{-M_i} \left[e^{\Lambda_i v_i}(\Lambda_i v_i - 1) - e^{\Lambda_i v_{i-1}}(\Lambda_i v_{i-1} - 1) \right] \\
&= \underbrace{\sum_{i=1}^{N} e^{-M_i} \left[e^{\Lambda_i v_i} v_i - e^{\Lambda_i v_{i-1}} v_{i-1} \right]}_{v_N} - \underbrace{\sum_{i=1}^{N} \Lambda_i^{-1} e^{-M_i} \left[e^{\Lambda_i v_i} - e^{\Lambda_i v_{i-1}} \right]}_{\lambda_N^{-1}\pi_N}
\end{aligned}$$

The first sum is telescopic and evaluates to $e^{-M_N + \Lambda_N v_N} v_N = v_N$. This concludes the proof of part 2 in Proposition 2.

Expected Travel Time and Expected Wait Time. Travel time is non-random, so expected travel time is given by:

$$\mathbb{E}T_k^{\text{time}} \equiv \mathbb{E}\left(T_k^{\text{time}} \mid k \in \arg\max\right) = \sum_k \pi_k T_k^{\text{time}}. \tag{11}$$

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

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

A.7 Optimal Network Design

A.7.1 Optimization Environment: Predicted Bus Travel Times

In this section, we describe how we estimate bus travel times for every edge in the grid cell environment, in a manner that makes these edge costs consistent with our entire GPS data on bus travel times. The key challenge is that certain origin-destination pairs are not connected by bus in the current network.

We proceed in two steps. First, we use data that we collected on driving times between

every pair of grid cells, and we project bus log travel time – for those origin-destination-route combinations available – onto log driving time for that origin-destination pair. Second, we estimate edge-specific bus travel times that "micro-found" these predicted bus travel times. We estimate edge times in the routing model from Allen and Arkolakis (2022), which allows us to express the predicted bus travel time between any origin and destination as the noisy shortest route in our grid cell environment network. We repeat this exercise separately for BRT and non-BRT.

Step 1. Predicted bus travel time for all origin-destination pairs. We obtain driving travel times between tens of thousands 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.[56] We then spatially interpolate this data to construct driving time T_{od}^{drive} between every origin and every destination.

We then predict bus travel times between all (o, d) pairs. We start with our data on median bus travel times T_{odr}^{bus} between o and d on route r (only for o, d pairs where this data is available). We then estimate the following linear regression

$$\log(T_{odr}^{\text{bus}}) = \alpha_0 + \alpha_1 \log(T_{od}^{\text{drive}}) + \epsilon_{odr}.$$

We then construct the prediction $\widehat{T}_{od}^{\text{bus}}$ for all origin-destination pairs, separately for BRT and non-BRT bus travel times.

Step 2. Estimate edge-level bus travel times. In our counterfactual simulations, we need to predict bus travel time for routes following any path. The model primites are *edge-specific* bus travel times, which we sum up over any possible route. We now describe how we estimate edge-specific bus travel times using our data on predicted bus travel times for all origin-destination pairs.

Given that we do not have specific bus routes for $\widehat{T}_{od}^{\text{bus}}$, we model these travel times as noisy shortest routes over the underlying network (Allen and Arkolakis, 2022).

For every edge ij in the grid cell network – this includes horizontal, vertical and diagonal neighboring grids – denote by T_{ij} the bus travel time parameter in the model. We denote by p a path between o and d, and by T_p the sum of edge-wise travel times along the path. Assume that a bus commuter traveling from o to d can choose any finite path between o and d and faces realized travel time $\widetilde{T}_p = T_p + \epsilon_p$, where ϵ_p is an iid Gumbel-distributed idiosyncratic shock with parameter ν. These shocks capture in a stylized way the fact that bus routes

[56] Through additional partial city coverage from earlier years, we confirmed that this data captures pre-COVID traffic congestion patterns and that aggregate congestion appears flat since 2016.

between o and d do not always follow the shortest path, with small ν corresponding to larger deviations. The commuter selects the path with the shortest realized travel time. By properties of multinomial logit, the expected bus travel time between o and d is

$$\widetilde{T}_{od} \equiv \mathbb{E} \max_p \widetilde{T}_p = -\nu^{-1} \log \Big(\sum_p \exp(-\nu T_p) \Big)$$

Allen and Arkolakis (2022) prove that the matrix $B = (\exp(-\nu \widetilde{T}_{od}))_{od}$ is given by $B = (I - A)^{-1}$ where $A = (\exp(-\nu T_{ij}))_{ij}$ is the travel time adjacency matrix, which has entries equal to zero whenever ij is not an edge.

We next estimate T_{ij} for all ij edges, as well as ν, using data on $\widehat{T}^{\text{bus}}_{od}$. We minimize the following objective function

$$\min_{\nu, (T_{ij})_{ij}} \sum_{o,d} \psi_{od} \Big(\log(\widehat{T}^{\text{bus}}_{od}) - \log(\widetilde{T}_{od}) \Big)^2$$

using weights ψ_{od} equal to the inverse squared distance between grid cells o and d. We estimate $\widehat{\nu} = 118$ for BRT and $\widehat{\nu} = 116$ for non-BRT, suggesting that bus routes are best explained as nearly shortest-route paths. (After this point, we no longer use ν.) We use the resulting $\widehat{T}_{ij}$ for BRT and non-BRT as bus travel time in all our counterfactual network exercises.

Figure A.22 prints the empirical fit of this exercise.

A.7.2 Ridership Equilibrium With Bus Capacity Penalties

Here we set up the formal model where commuters incur time penalties when bus occupancy exceeds bus capacity. Consider a route r and two consecutive stations s and s' on r. The route-edge occupancy level $B_{ss'r}$ is given by the expected number of passengers on a bus on route r between s and s'

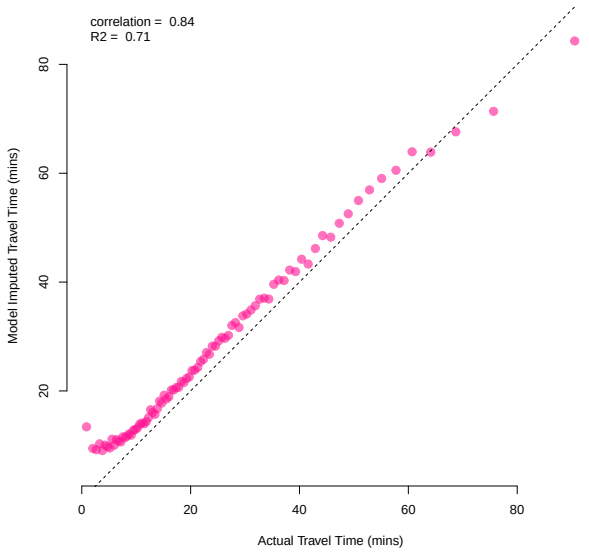
$$B_{ss'r} = \frac{1}{D_r} \sum_{o \leq s < s' \leq d} R_{odr}, \tag{12}$$

where R_{odr} is the daily number of commuters traveling on route r between any pair of stations o and d, and D_r is the daily number of bus departures on route r. The sum is over all stations o on route r that precedes s (in the direction from s to s'), and all d that come after s'.

We assume that the perceived travel time on route r and edge ss' is given by

$$T^p_{ss'r} = T_{ss'r} \cdot \kappa(B_{ss'r}), \tag{13}$$

Figure A.22: Bus Travel Time Model Fit



Notes: This binscatter plot assesses GMM model fit. It correlates our data on bus travel times T_{odr}^{bus} at the origin-destination-route level, restricting to origin-destination pairs for which this data is available, on the x-axis, with model-predicted expected bus travel time $\widetilde{T}_{od}$ on the y-axis.

where $T_{ss'r}$ is the baseline (objective) travel time, and $\kappa(B)$ is a function equal to 1 when bus occupancy is less than bus capacity, $B \leq \overline{B}$, and convex and increasing for $B > \overline{B}$. We parameterize κ as

$$\kappa(B) = \max\left(1, \quad \exp\left(0.2 \cdot (B - 84)\right)\right). \tag{14}$$

We assume there is no penalty up to $\overline{B} = 84$ persons per bus (the average bus capacity in TransJakarta's fleet at the end of our study period), and the penalty increases rapidly afterward. Perceived travel time with 90 people on the bus is already more than three times as high as objective travel time.

We define the perceived travel time between any two stations o, d on route r as the sum of edge-wise perceived travel times along that route, i.e. $T_{odr}^{p} = \sum_{k=1}^{n} T_{s_k s_{k+1} r}^{p}$ where $s_1 = o$ and $s_n = d$. As commuters get on and off, crowding on a given route fluctuates at different points on the route. Our model assumes that commuter disutility is additive over route segments, based on the crowding level at each point.

A ridership equilibrium is defined by a set of crowding levels (for each route r and each pair of consecutive stations s, s') such that optimal commuter choices lead to ridership patterns and a set of crowding levels that is exactly the assumed one. That is, it involves a

fixed point, whereby crowding $B_{ss'r}$ is determined by equation (12) based on route-specific ridership flows R_{odr}, and ridership patterns R_{odr} are given by choice probabilities based on equations (2), (3) and (4), where we replace travel time T_{odr} with perceived travel time T^p_{odr} given by (13), which depends on $B_{ss'r}$.

We compute a ridership equilibrium by iterating until we reach a fixed point. For computational reasons, in our simulated annealing algorithm, we approximate the equilibrium by a single one-step updating rule. We compute route choices and ridership assuming no penalties, then compute occupancy levels and re-compute ridership once more. This allocation is very close to the equilibrium. For the 200 bus networks we obtain from the simulated annealing algorithm, we compare the one-step allocation with the fixed-point equilibrium. 170 of the 200 networks are already at equilibrium because all edge-occupancy levels are below $\overline{B}$. Even when they differ, the maximum deviation is negligible, equivalent to less than a second of wait time for every original TransJakarta passenger. The maximum ridership difference is 0.03% of daily bus ridership.

A.7.3 Analytic Results for Optimal Allocations Model

Formula for Local Comparative Statics. Here we derive the expression for the derivative of the expected optimal property $f^*(\theta)$ with respect to a scalar θ, defined in equation (7).

Using the logit formula we can show that

$$D_\theta \pi(N,\theta) = \beta \pi(N,\theta)\Big(D_\theta W(N,\theta) - \mathbb{E}_{N'} D_\theta W(N',\theta)\Big) \tag{15}$$

The term in parentheses says that when θ changes, the probability of a given network N increases if and only if the derivative of welfare W with respect to θ at N is higher than its expectation over all possible networks. Using this expression and rearranging sums several

times we get

$$
\begin{aligned}
D_\theta f^*(\theta) &= \sum_N D_\theta \pi(N,\theta) f(N,\theta) + \pi(N,\theta) D_\theta f(N,\theta) \\
&= \sum_N \beta \pi(N,\theta) \Big(D_\theta W(N,\theta) - \mathbb{E}_{N'} D_\theta W(N',\theta) \Big) f(N,\theta) + \pi(N,\theta) D_\theta f(N,\theta) \\
&= \Big[\sum_N \beta \pi(N,\theta) D_\theta W(N,\theta) f(N,\theta) \Big] - \beta \Big[\mathbb{E}_{N'} D_\theta W(N',\theta) \Big] \Big[\underbrace{\sum_N \beta \pi(N,\theta) f(N,\theta)}_{f^*(\theta)} \Big] + \\
&\qquad + \Big[\sum_N \pi(N,\theta) D_\theta f(N,\theta) \Big] \\
&= \sum_N \beta \pi(N,\theta) D_\theta W(N,\theta) f(N,\theta) - \beta \Big[\sum_{N'} \pi(N',\theta) D_\theta W(N',\theta) \Big] f^*(\theta) + \sum_N \pi(N,\theta) D_\theta f(N,\theta) \\
&= \sum_N \pi(N,\theta) \Big[\beta D_\theta W(N,\theta) \Big[f(N,\theta) - f^*(\theta) \Big] + D_\theta f(N,\theta) \Big].
\end{aligned}
$$

A.7.4 The Simulated Annealing (SA) Algorithm Asymptotic Result

This section reports the sketch of the proof of Proposition 3. The proof uses standard Metropolist-Hastings algorithm arguments. It mirrors a classic proof that as the number of steps increases, the outcome of SA converges to the global optimum.

The key observation is that replacing the variable inverse temperature β_k with a fixed temperature β in the acceptance probability (8) yields the Metropolis-Hastings (MH) algorithm. Because Ψ is irreducible and aperiodic, MH is an irreducible and aperiodic stationary Markov chain. Its stationary distribution at N is proportional to $\exp(\beta W(N))$ and hence is exactly π. This follows from checking that the *detailed balance* condition holds for any N, N' such that $\Psi(N' \mid N) > 0$:

$$
\pi(N) \Pr(N' \mid N) = \pi(N') \Pr(N \mid N').
$$

Returning to the SA algorithm, as the number of steps grows, the algorithm can be approximated by a sequence of MH algorithms with increasing inverse temperatures. In particular, the number of steps with an inverse temperature nearly equal to β grows to infinity, and hence the endpoint of the SA algorithm is asymptotically distributed according to π. (These arguments can be made precise, see Nikolaev and Jacobson (2010).)

A.7.5 The Candidate Network Proposal Function Ψ

How to produce new candidate networks is critical for the success of the SA algorithm. Starting from a network N_k, we obtain a proposed network N' by applying one of the "modifier" operations described below. We first select one of the four categories with equal probability, then a type of modifier within the category with equal probability, and finally, we generate a network N' according to the selected modifier, uniformly randomly.

Local Bus Modifiers

1. Exchange one bus between two randomly drawn bus routes.
2. Exchange a randomly drawn 10% of busses between two randomly drawn bus routes.

Global Bus 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.

Local Route Modifiers

1. Draw a bus route at random at add one random new adjacent stop to one end of the route.
2. Draw a bus route at random and take away one random stop at one end of the route.
3. 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.
4. 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.

Global Route Modifiers

1. 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.
2. Delete a randomly drawn route and replace it with 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, where routes with less ridership are more likely to be chosen. Redistribute the busses among other, randomly chosen, bus routes.
4. Delete a randomly chosen set of 10 bus routes at once. Redistribute the busses among other, randomly chosen, bus routes.
5. Add 10 random bus lines at once. Assign a random number of busses to the new routes, redistributed from randomly chosen other bus routes.

Approximating Proposal Probabilities Ratio In our implementation of the simulated annealing algorithm, we use the following approximation when using equation (8)

$$\frac{\Psi(N_k \mid N')}{\Psi(N' \mid N_k)} = 1.$$

We make this approximation because computing the proposal function ratio term precisely is computationally intensive. $\Psi(N' \mid N_k)$ is the probability that network N' is proposed when starting from network N_k. To compute this, we need to enumerate all possible modifications starting from N_k. (Given that we stratify, we only need to do this within each category and type of modification.)

For most modifiers described above, this ratio is likely to be close to 1. For example, this ratio is approximately equal to 1 for the proposal function that exchanges a bus between two randomly drawn bus routes. There are two types of exceptions. First, certain modifiers that we use are not reversible. Second, for the modifiers that add or delete routes, the ratio is systematically far from 1. We argue that approximating the ratio with 1 will lead our algorithm to converge to a stationary distribution that puts more weight on expansive networks. This would only strengthen our main results that optimal networks are more expansive than the current TransJakarta network.

To see this more clearly, denote by R the (very large) number of possible routes in our square grid cell environment, and by $r(N)$ the number of routes in network N. For ease of exposition, assume that only modifiers that add or remove one route are allowed, and assume that N' is obtained from N_k by adding a randomly chosen route. Then the proposal function ratio is given by

$$\frac{\Psi(N_k \mid N')}{\Psi(N' \mid N_k)} = \frac{\frac{1}{r(N_k)+1}}{\frac{1}{R-r(N_k)}} = \frac{R - r(N_k)}{r(N_k)+1}.$$

This is significantly larger than 1, meaning that transitions towards networks with many routes is favored in equation (8). It easy to prove that if we use the ratio equal to 1, the Markov chain will converge to a modified stationary distribution given by

$$\widetilde{\pi}(N) \propto \pi(N) \cdot \binom{R}{r(N)+1}^{-1}$$

In other words, the modified stationary distribution puts less weight on expansive networks. Note that this adjustment effectively compensates for the very large number of expansive networks, which is of the same magnitude as the adjustment factor.