

Online Appendix

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

May 13, 2025

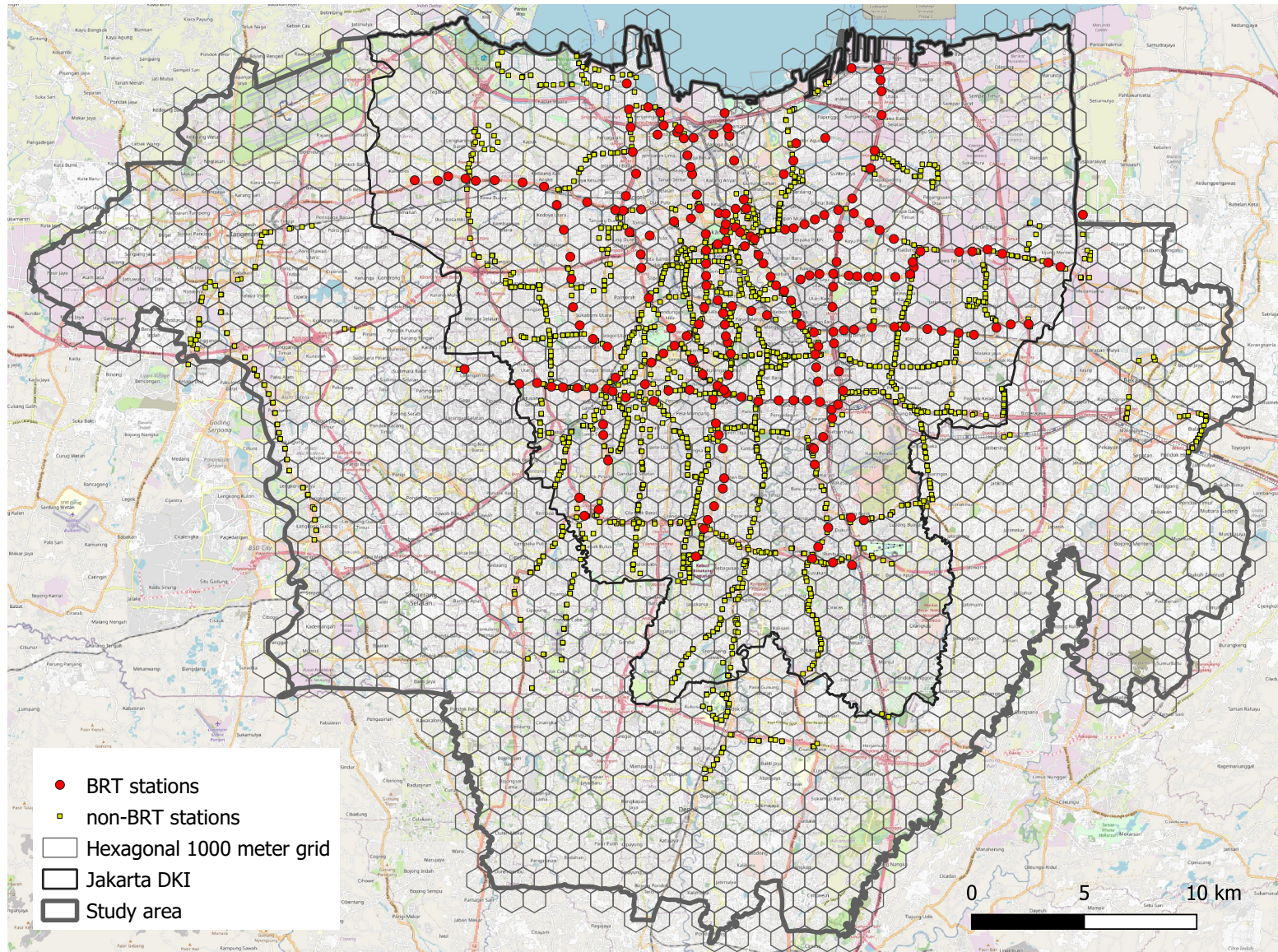
A Online Appendix

Contents

A.1	Appendix Figures	61
A.2	Appendix Tables	87
A.3	Data Processing	99
A.3.1	Bus GPS Data Processing	99
A.3.2	Bus Travel Times	101
A.3.3	Bus Wait Time Distribution	103
A.3.4	TransJakarta Ridership Data Processing	105
A.3.5	Veraset Smartphone Location Data Trip Processing	107
A.4	Reduced Form Estimation	110
A.5	Attention Probabilities	111
A.6	Demand Model Derivations	111
A.7	Optimal Network Design	113
A.7.1	Optimization Environment: Predicted Bus Travel Times	113
A.7.2	Ridership Equilibrium With Bus Capacity Penalties	115
A.7.3	Model Welfare Bound	117
A.7.4	Analytic Results for Optimal Allocations Model	119
A.7.5	The Simulated Annealing (SA) Algorithm Asymptotic Result	120
A.7.6	The Candidate Network Proposal Function Ψ	122
A.7.7	Details on Statistical Test of Mixing	124
A.7.8	A Simple Model of Optimal Wait Times	124
A.8	Impact on Air Pollution and Congestion	126
A.8.1	Impact on Road Traffic Congestion	126
A.8.2	Impact on Air Pollution	127
A.8.3	Monetizing Congestion and Air Pollution Effects	129

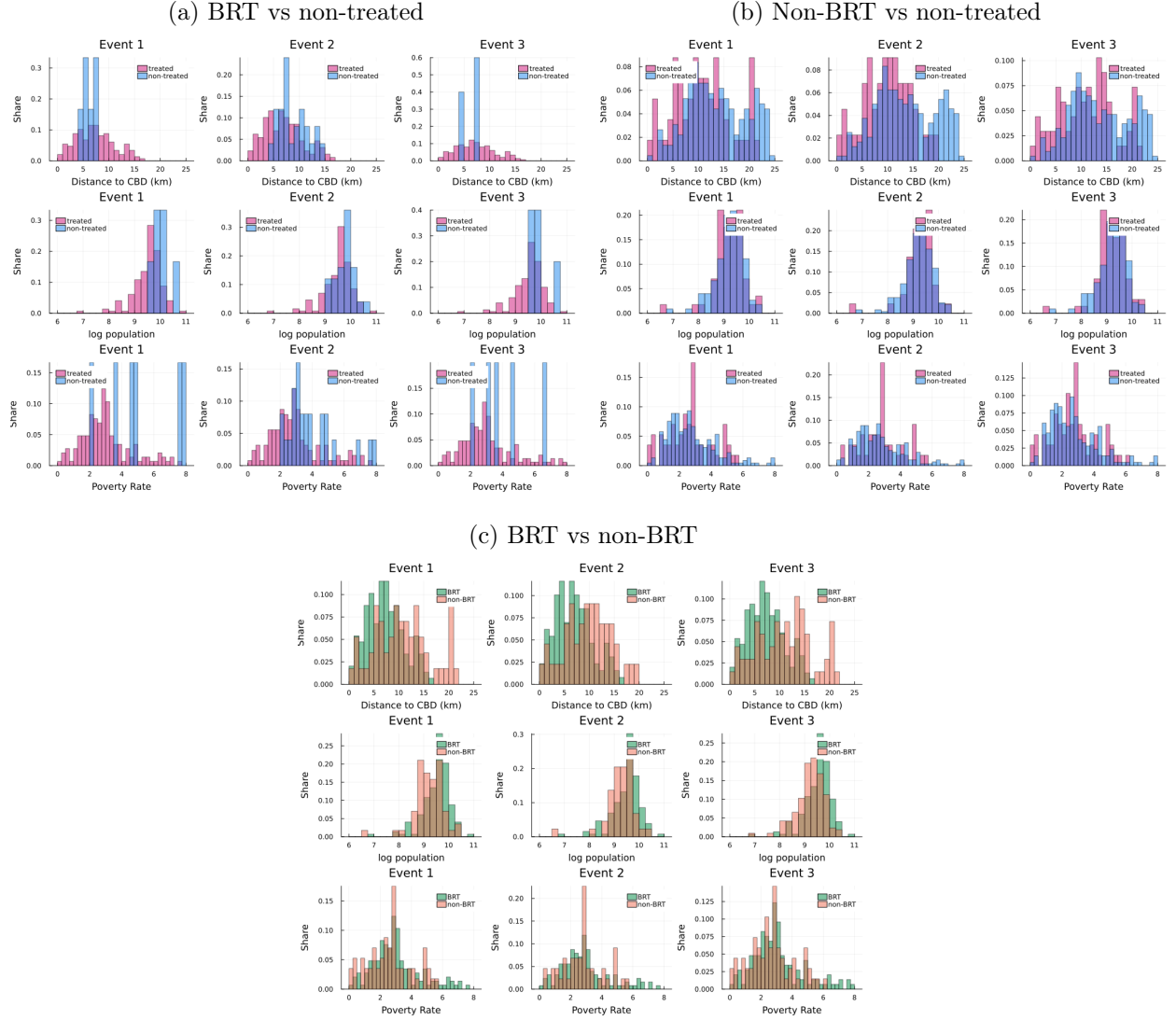
A.1 Appendix Figures

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



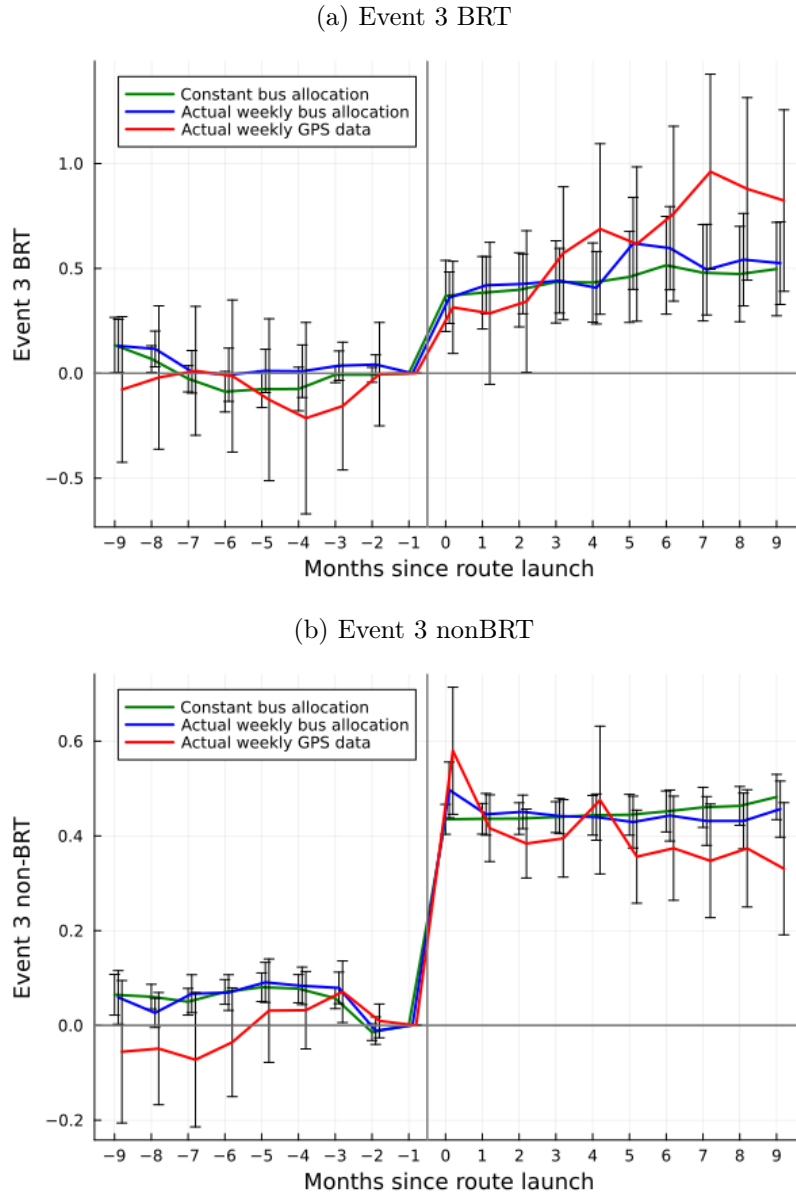
Notes: This figure shows the location of all TransJakarta stations (BRT in red circles, non-BRT in yellow squares) as of March 2020, together with the hexagonal grid with distance 1000 meters between centroids. The thin black link is the boundary of the Special Capital Region of Jakarta (DKI Jakarta),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

Figure A.2: Characteristics of Treated and Non-Treated Origin Grids



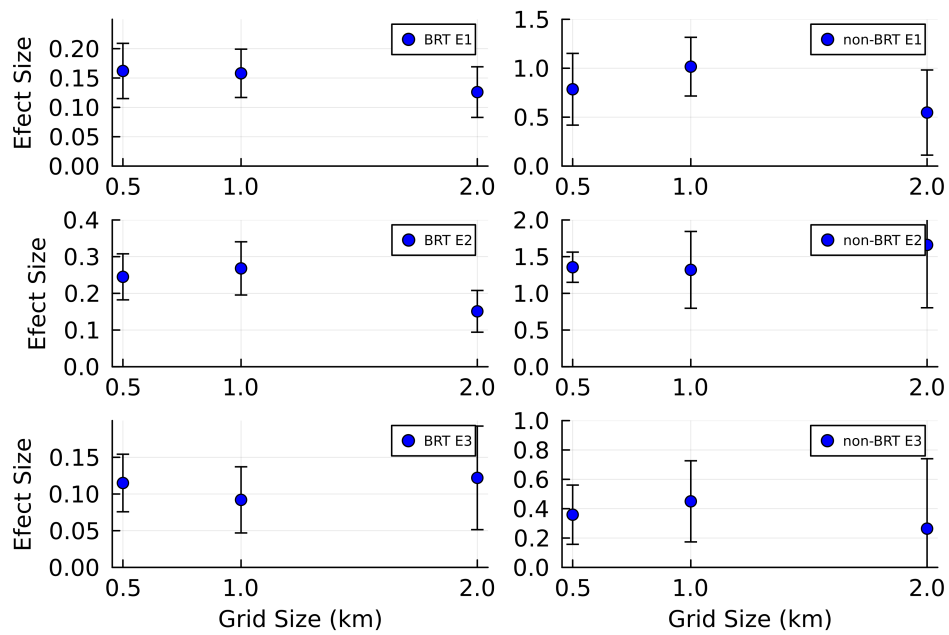
Notes: panels (a) and (b) display histograms of characteristics of treated and non-treated origin grids with bus access. For each event type and for BRT/non-BRT, we define an origin grid as treated if it ever has an event of that type (for some destination). The rest of the origin grids with bus access are the non-treated group. Panel (c) compares treated BRT grids and non-BRT treated grids.

Figure A.3: Event 3 “First-Stage” Robustness with Actual Bus Allocation and GPS Data



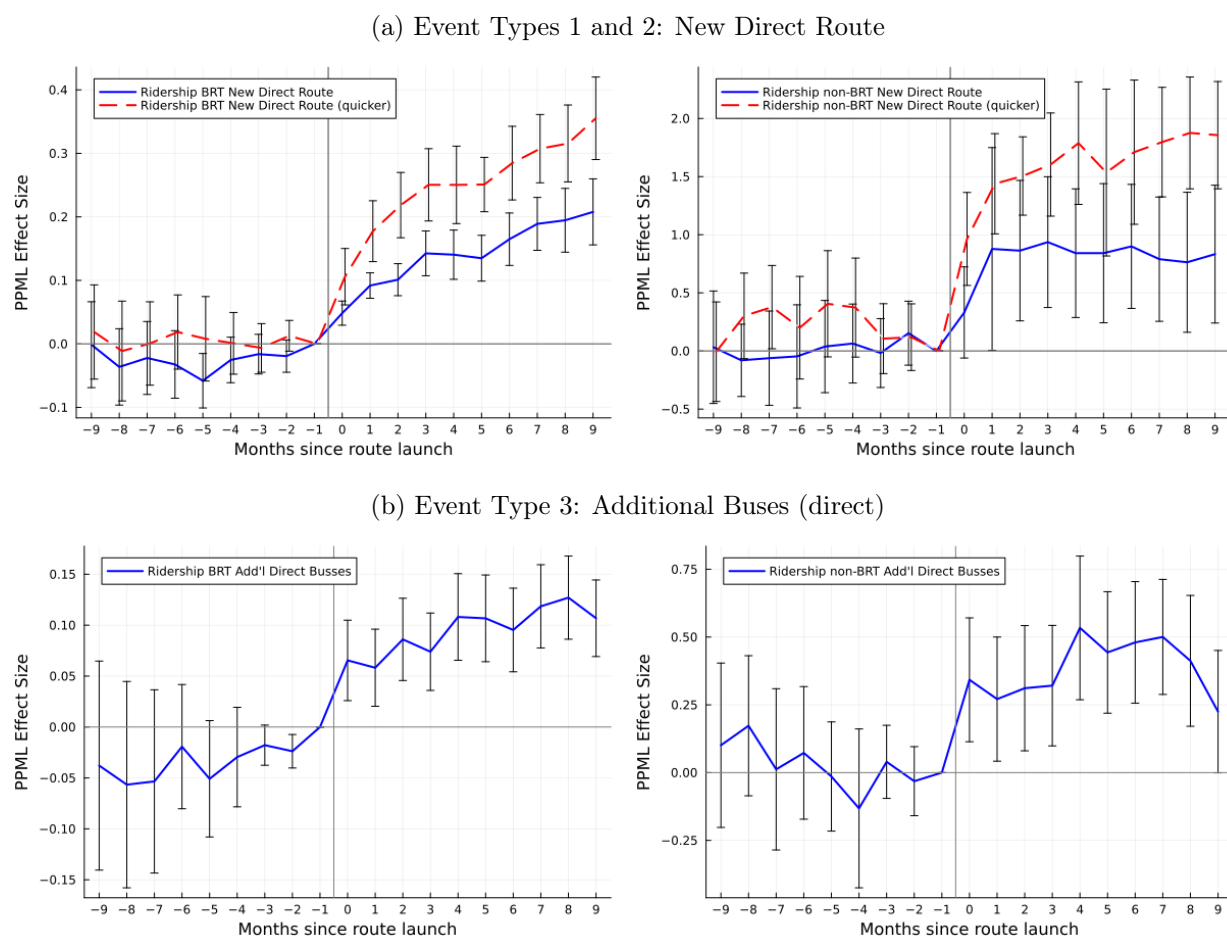
Notes: This graph plots the impact of Event 3 on the log of bus arrival rate. We compute the bus arrival rate in three ways. First, in green, our benchmark method uses for each route a bus allocation that is constant over time. (This avoids endogenous changes in bus allocation in anticipation of changes in demand.) Second, in blue, we use TransJakarta’s actual bus allocation, which varies over time within route. Third, in red, we use the number of trips that appear in TransJakarta’s bus GPS data. In all specifications, the outcome is the log average weekly bus arrival rate per hour between an origin and a destination, in a week, aggregated over all direct routes. We restrict the estimation sample to observations with coverage in the GPS data, which account for around 69% and 83% of all observations for BRT and non-BRT, respectively.

Figure A.4: Varying the Level of Aggregation: Impact of Network Expansion on Bus Ridership



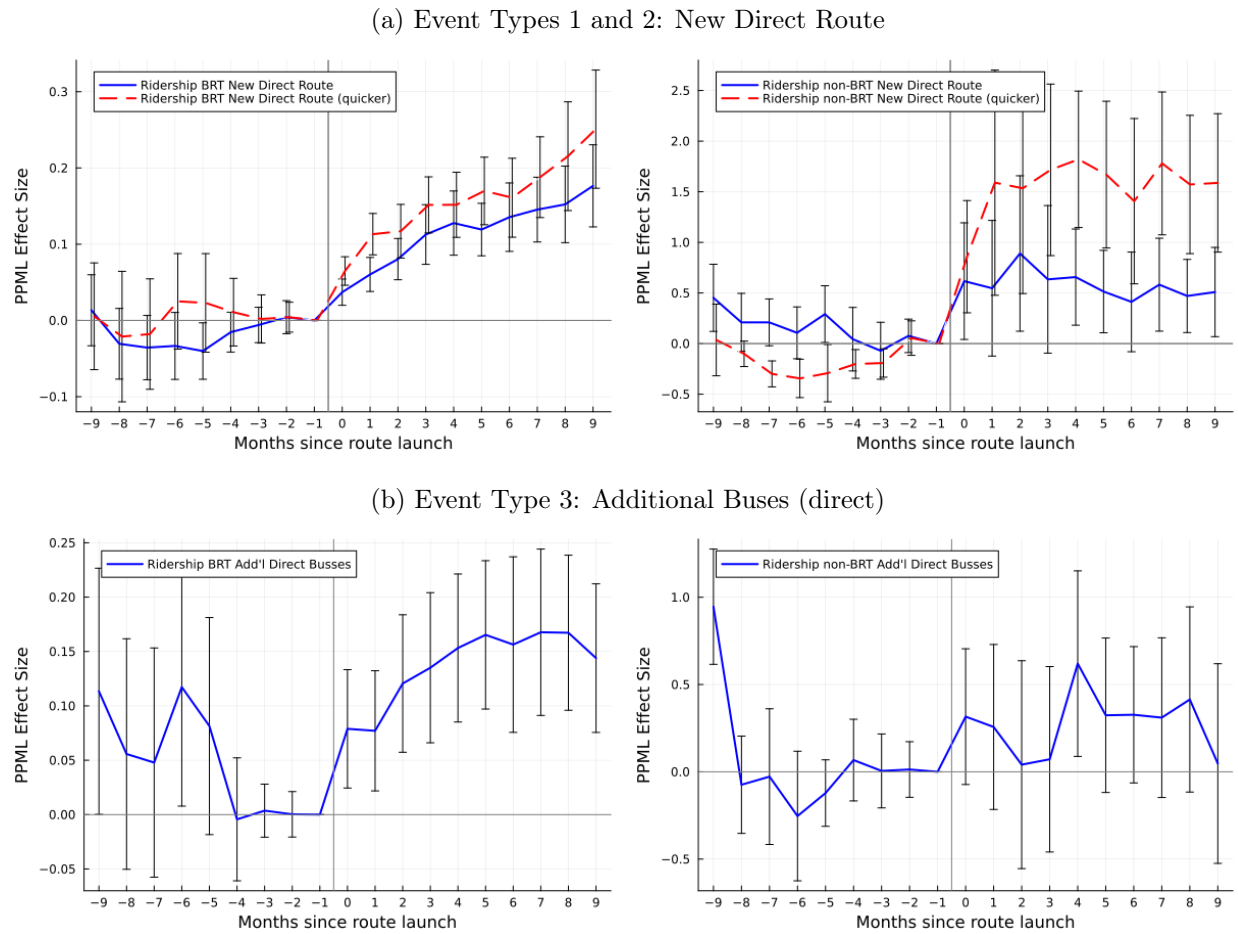
Notes: This graph plots the $Post_{odt}$ coefficients from our main specification with bus ridership as the outcome, for all six events, and for the three aggregation levels we consider: 500-meter square grids, 1000-meter hexagons, and 2000-meter square grids.

Figure A.5: Robustness With 500m Square Grids: **Bus Ridership** Impacts of BRT and Non-BRT Network Expansions (PPML)



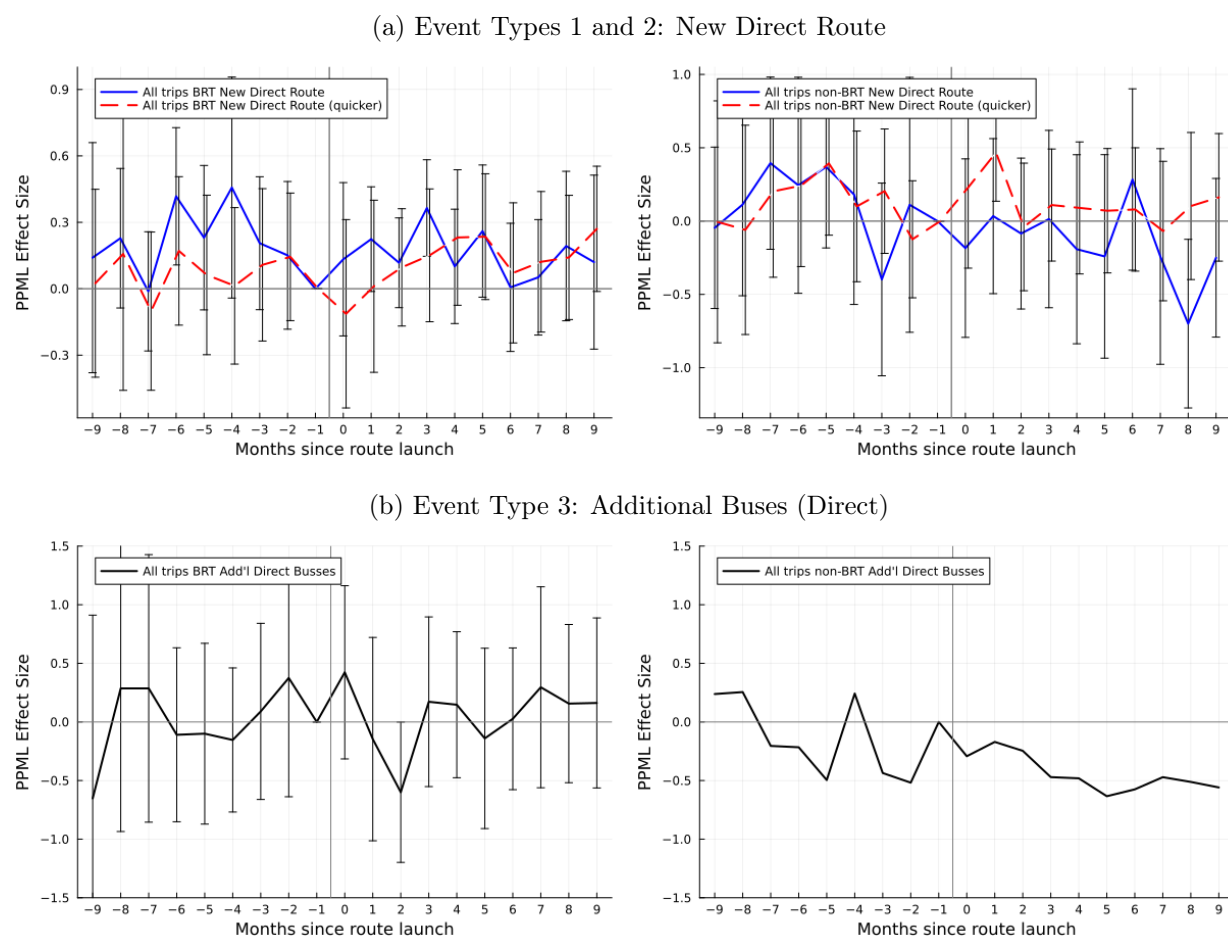
Notes: This graph replicates Figure 2 using 500-meter square grids instead of 1km hexagonal grids.

Figure A.6: Robustness With 2000m Square Grids: **Bus Ridership** Impacts of BRT and Non-BRT Network Expansions (PPML)



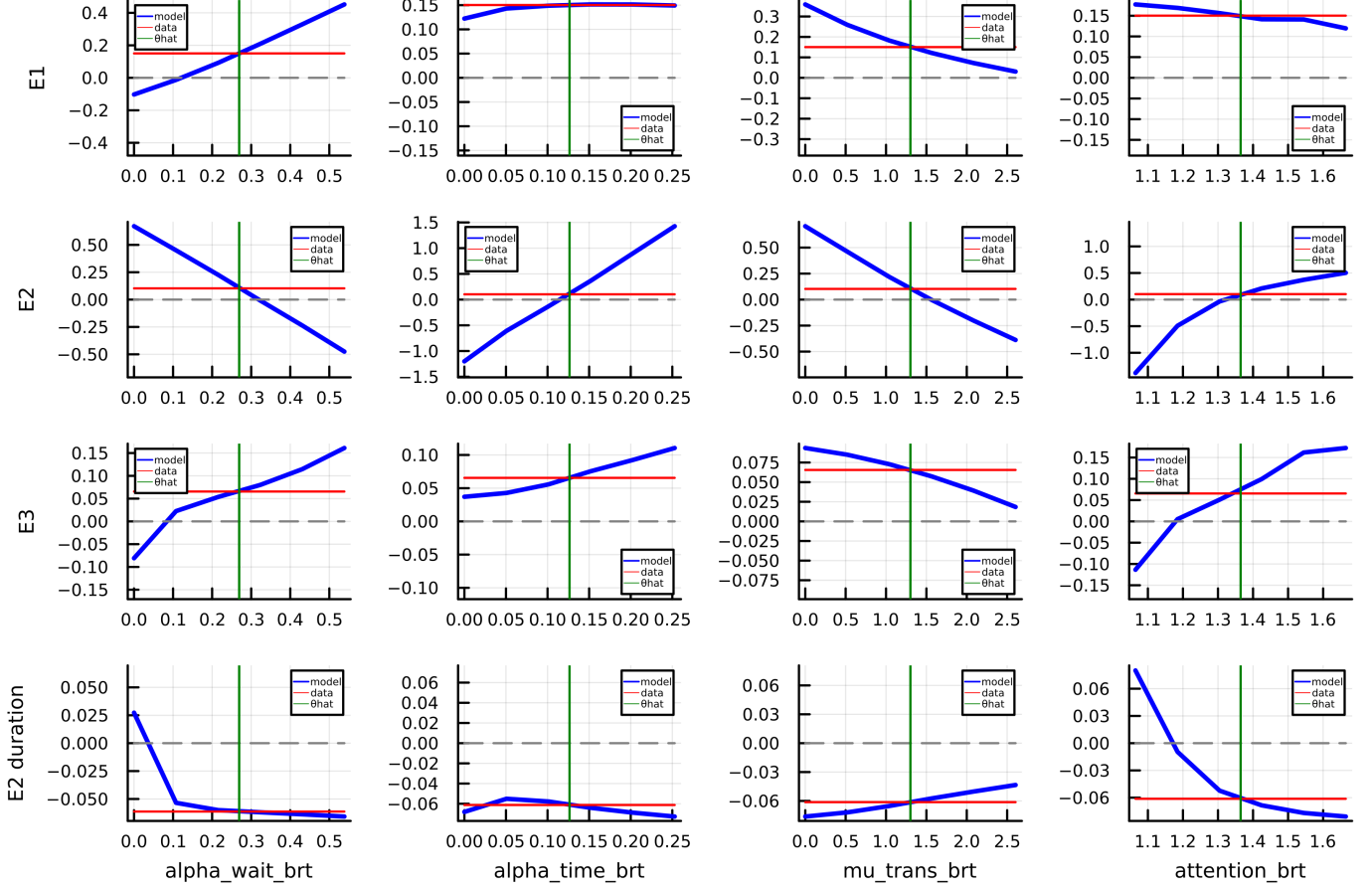
Notes: This graph replicates Figure 2 using 2000-meter square grids instead of 1km hexagonal grids.

Figure A.7: Robustness 500m Square Grids: **Aggregate Trip Volume** Impacts of BRT and Non-BRT Network Expansions



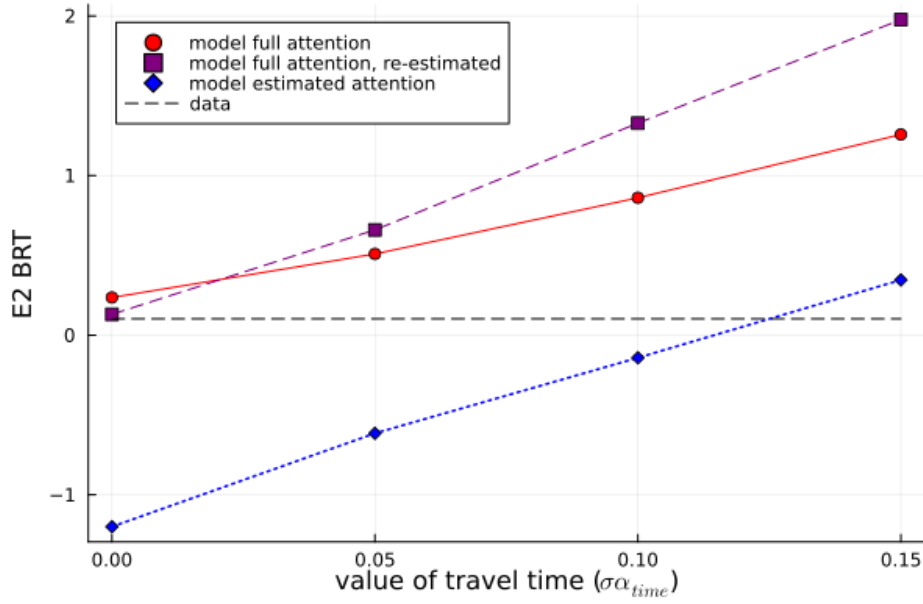
Notes: This graph replicated Figure 3 using 500m square grids. We do not report standard errors for non-BRT Event Type 3 due to a computational issue and refer the reader to the joint effect in Table A.7.

Figure A.8: 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 $\theta_i \in \{\sigma\alpha_{\text{wait}}^{\text{BRT}}, \sigma\alpha_{\text{time}}, \sigma\mu_{\text{transfer}}^{\text{BRT}}, \eta^{\text{BRT}}\}$. For each plot, we use 10 values of θ_i that bracket the estimated value $\hat{\theta}_i$. The thick blue line plots the model's moment value $m_j(\theta_i)$. The green vertical lines indicate the estimated parameter $\hat{\theta}_i$ (column 2 in Table 4). The red horizontal lines indicate m_j^{data} , the value of the moments in the data.

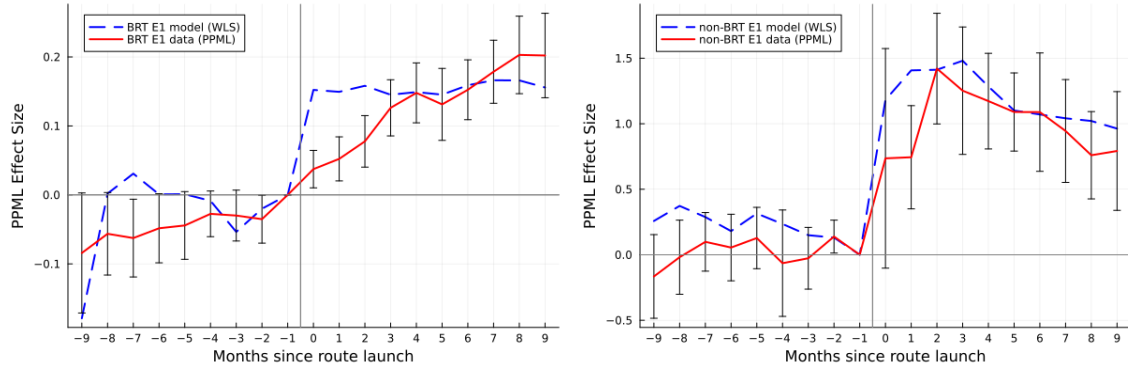
Figure A.9: Model Fit and Full Attention



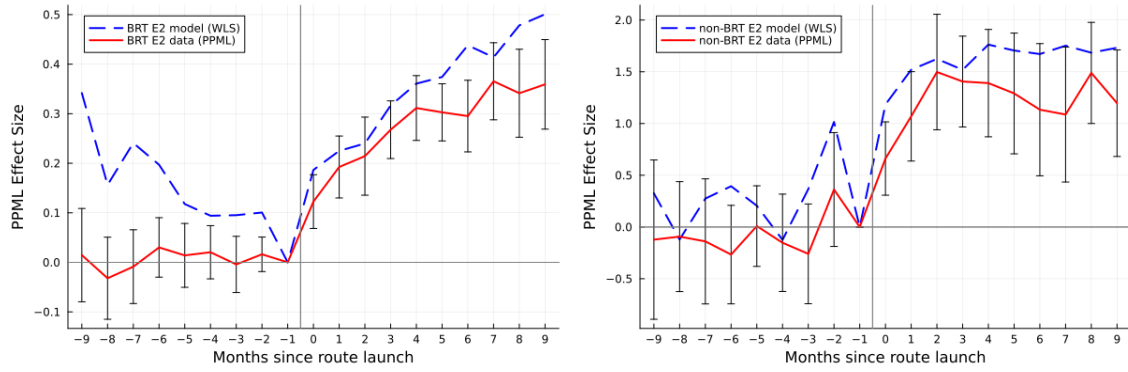
Notes: This figure shows that the model with full attention cannot fit the Event 2 BRT moment. The horizontal dashed line indicates the value of this moment in the data. The blue line (with diamonds) plots the model's predicted value of the event 2 BRT moment using the estimates from Table 4, column 2, where we vary the value of time α_{time} holding the other parameters fixed. (Recall that in the main estimation we normalize $\alpha_{time} = 1$.) The red line (with circles) repeats this exercise in the *full attention* model, i.e. setting $\eta^{BRT} = \infty$. The purple line (with squares) repeats the exercise for the full attention model where the other parameters ($\alpha_{wait}^{BRT}, \mu_{transfer}^{BRT}$) are re-estimated to match the 1st and 3rd events moments, given an assumed value of $\sigma \alpha_{time}$ and the previously estimated σ .

Figure A.10: Demand Estimation Model Fit (Targeted Moments)

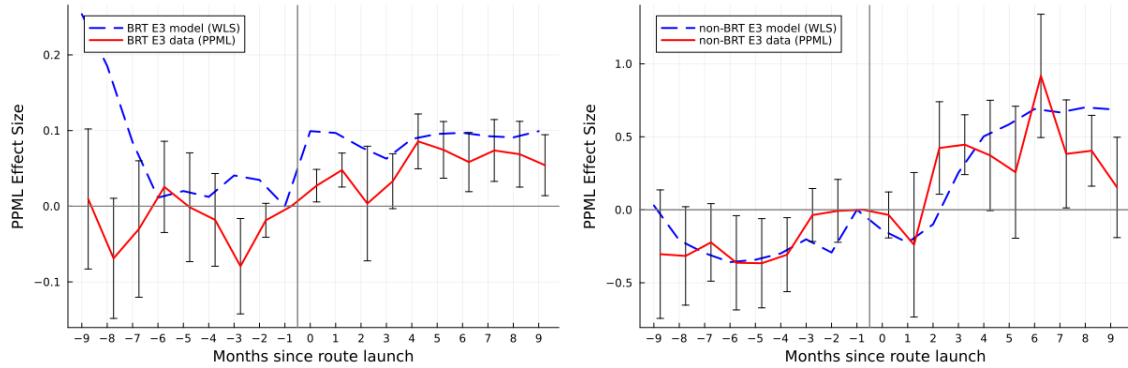
(a) Event Type 1: New Direct Route (Not Quicker)



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

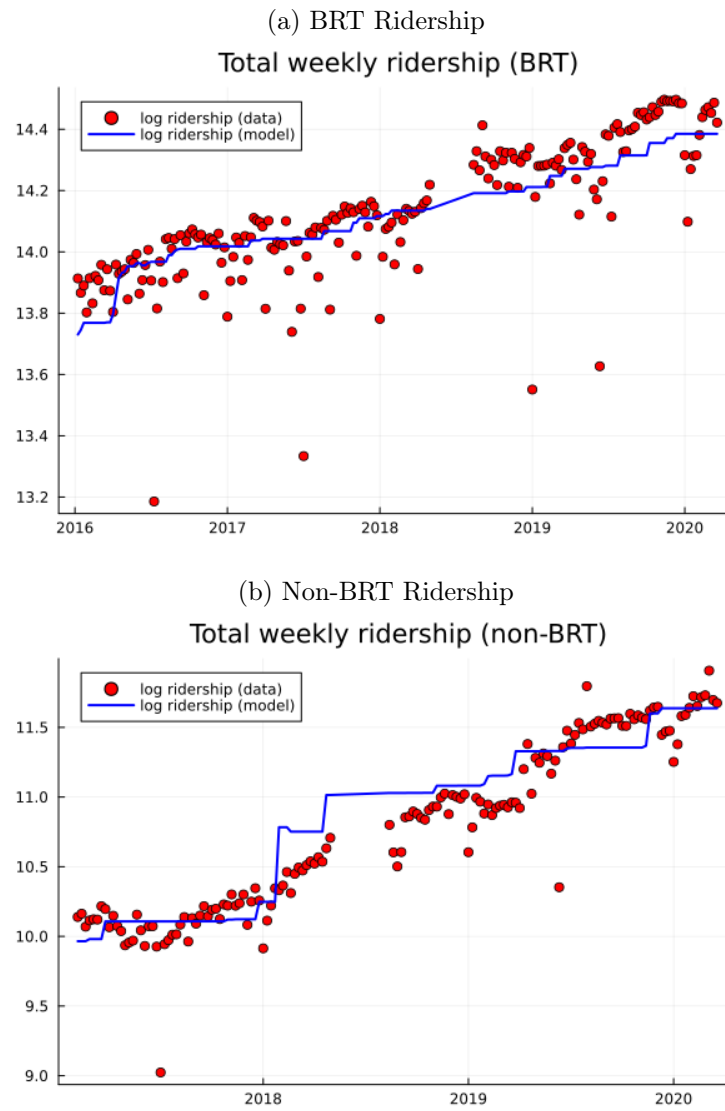


(c) Event Type 3: Additional Buses (Direct)



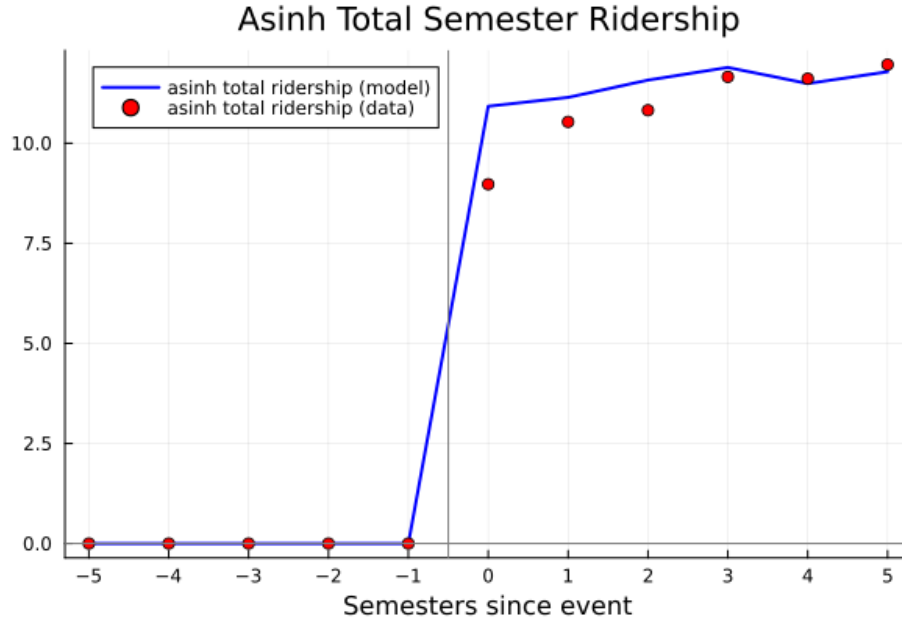
Notes: These figures plot the event impact on ridership data (red solid lines) and the same analysis on model-predicted ridership using the model estimated in Table 4, column 2 (blue dashed lines). For model-predicted ridership, we use weighted least squares on log model ridership.

Figure A.11: Demand Estimation Model Fit Over Time



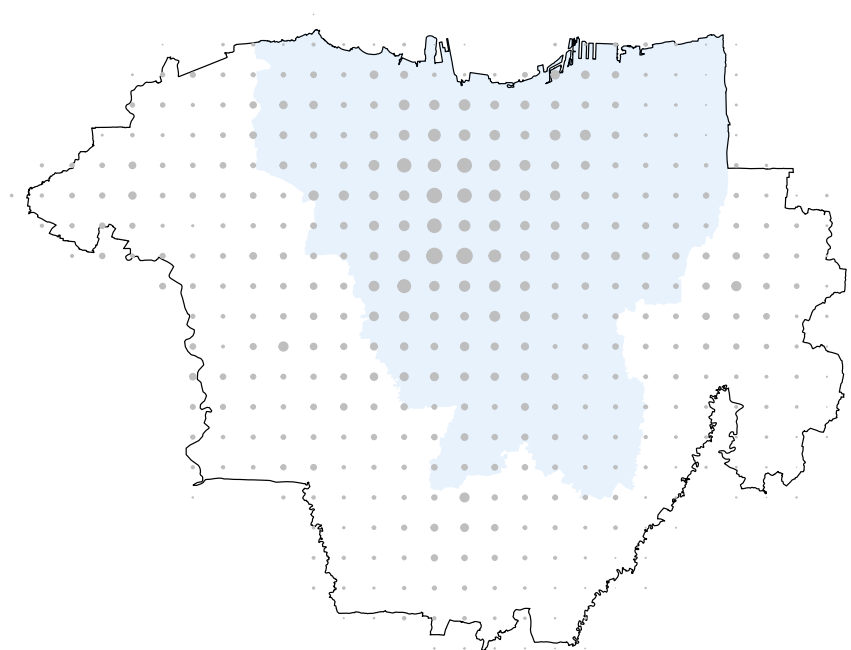
Notes: These figures report actual and model-predicted total ridership over the entire TransJakarta network over time at the weekly level, based on the model estimated in Table 4, column 2.

Figure A.12: Demand Estimation Model Fit for Non-BRT Route Launch



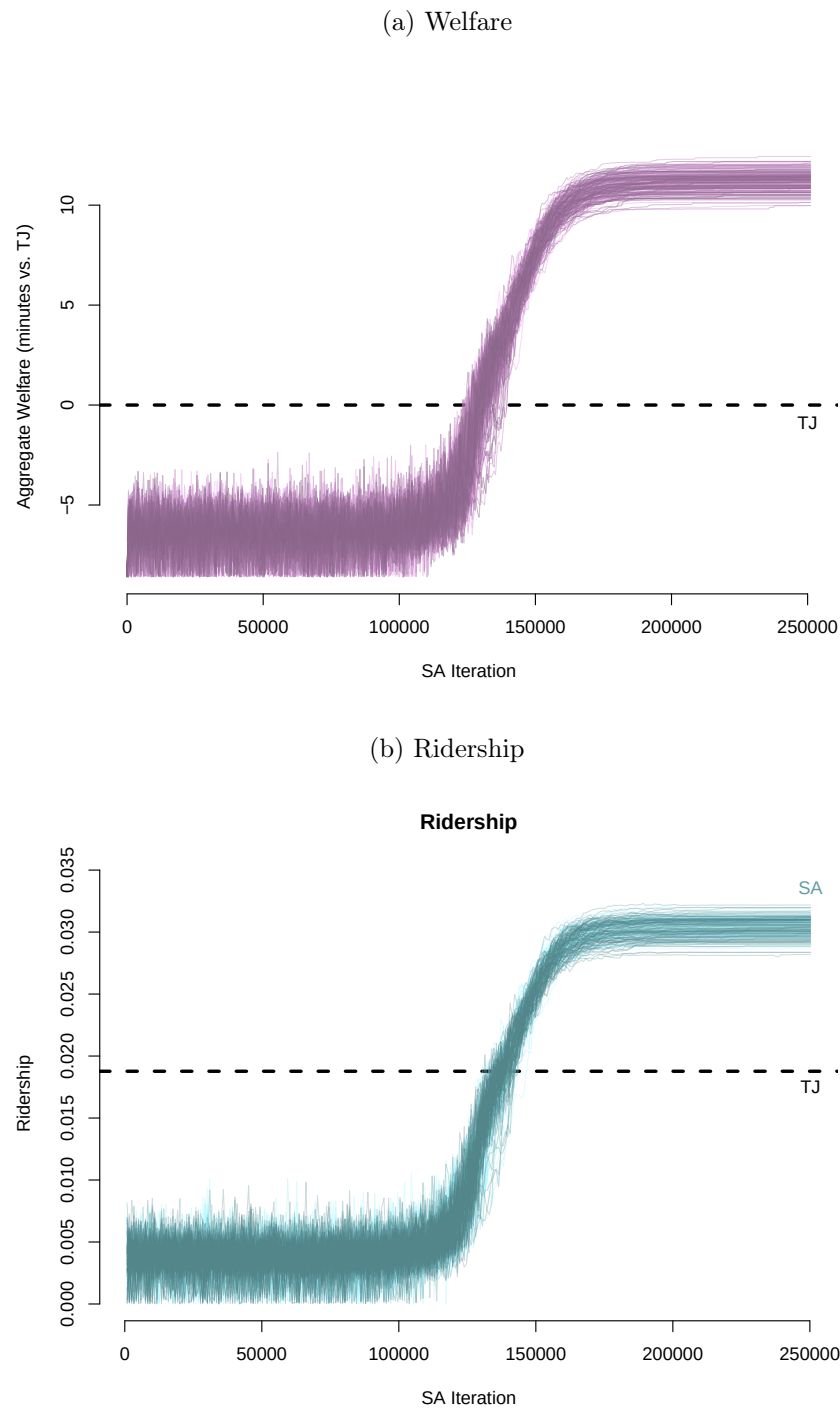
Notes: This figure reports actual and model-predicted ridership after the launch of the first non-BRT connection (direct or transfer). To construct this figure, we consider the 74 origin-destination pairs (using 1km hexagonal grids) that are part of the reduced form analysis and that first become connected 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.13: Jakarta Geography Used for Optimal Networks



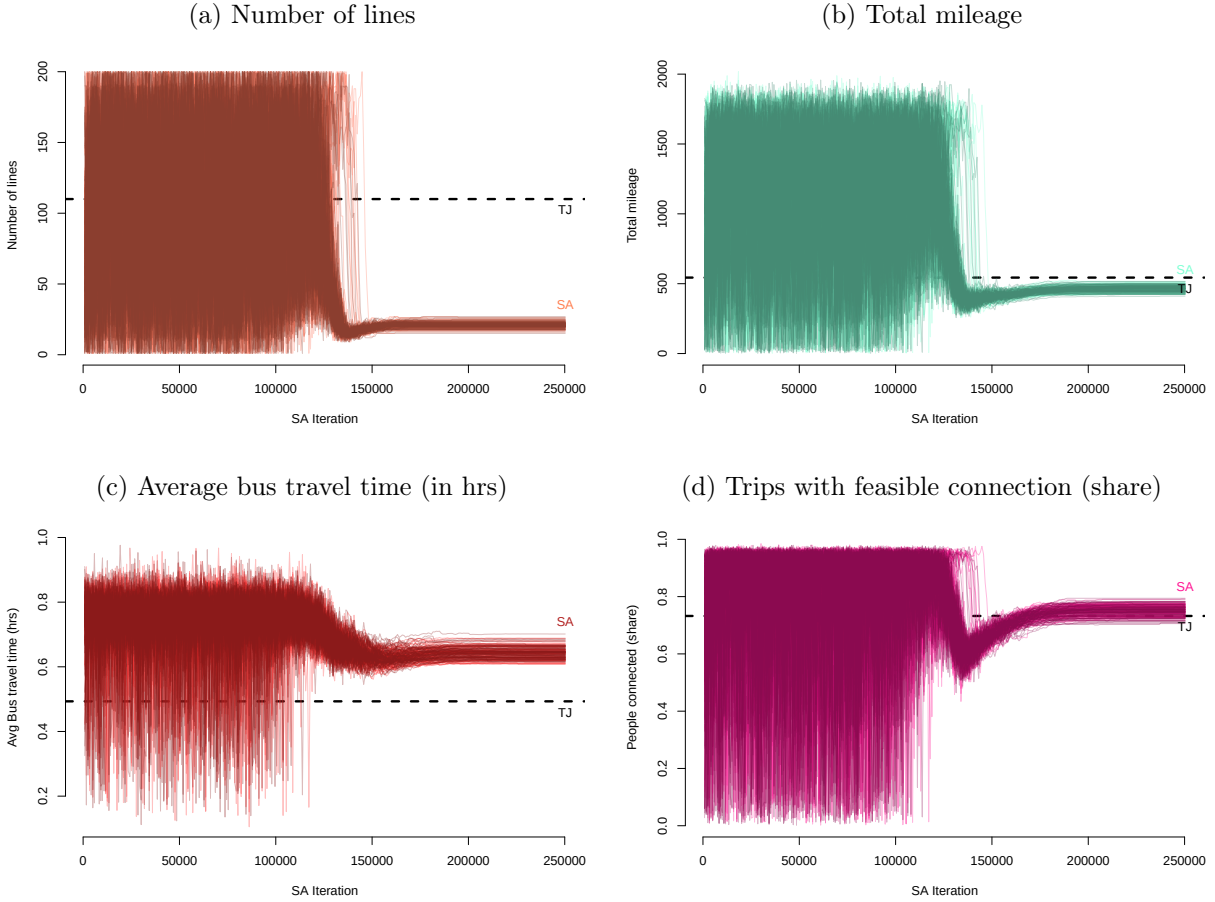
Notes: This map prints the geographic environment with $2\text{km} \times 2\text{km}$ grid cells used to compute optimal networks. The central DKI area is in light blue. The size of each circle is proportional to the number of trips that depart from that grid cell.

Figure A.14: Welfare and Ridership Over the Course of the Simulated Annealing Algorithm



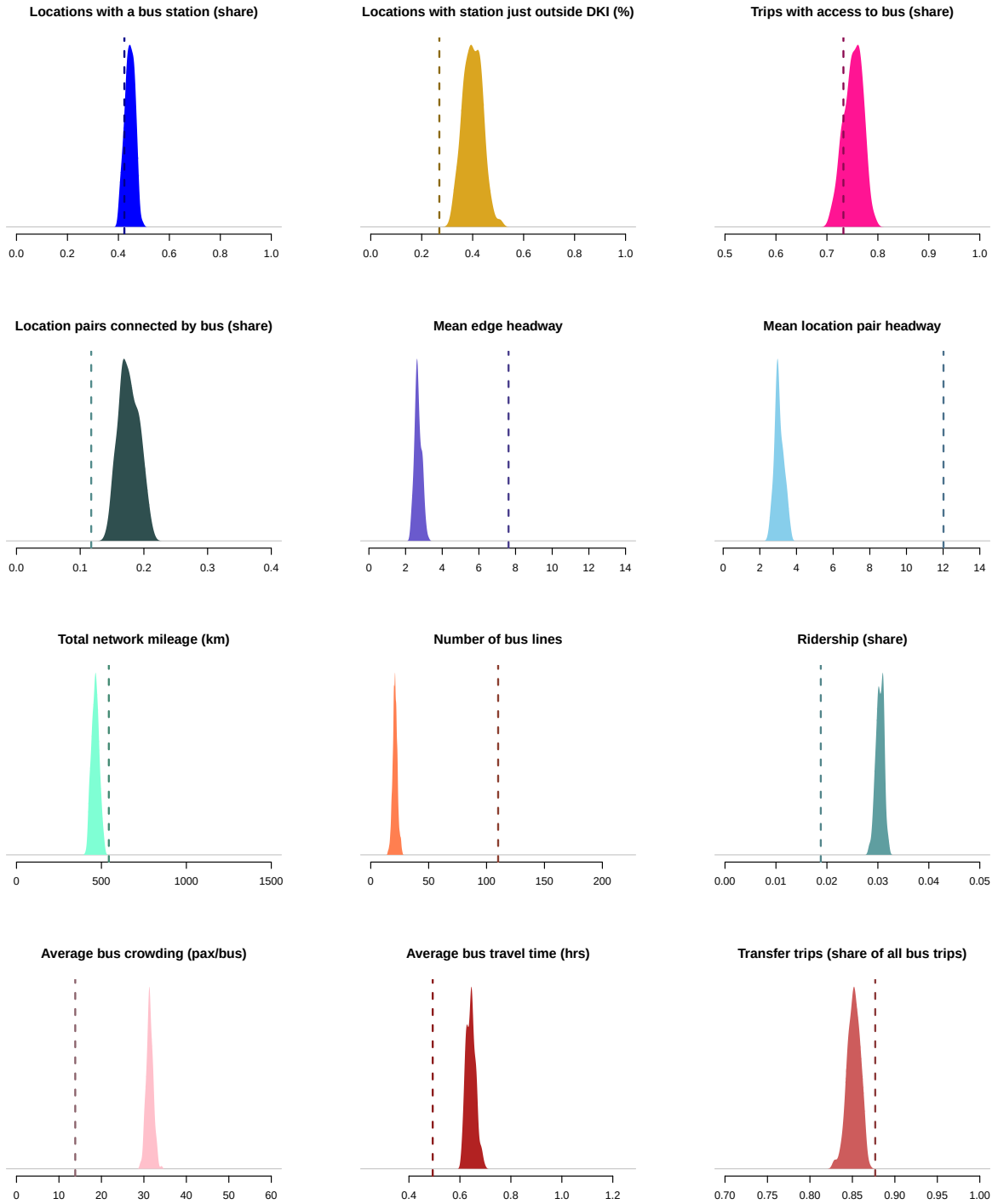
Notes: Each line shows the progression of welfare (panel a, denominated in minutes of travel time equivalent variation relative to the current TJ network) and ridership (panel b) throughout the 200 runs of the simulated annealing algorithm.

Figure A.15: Evolution of Network Characteristics



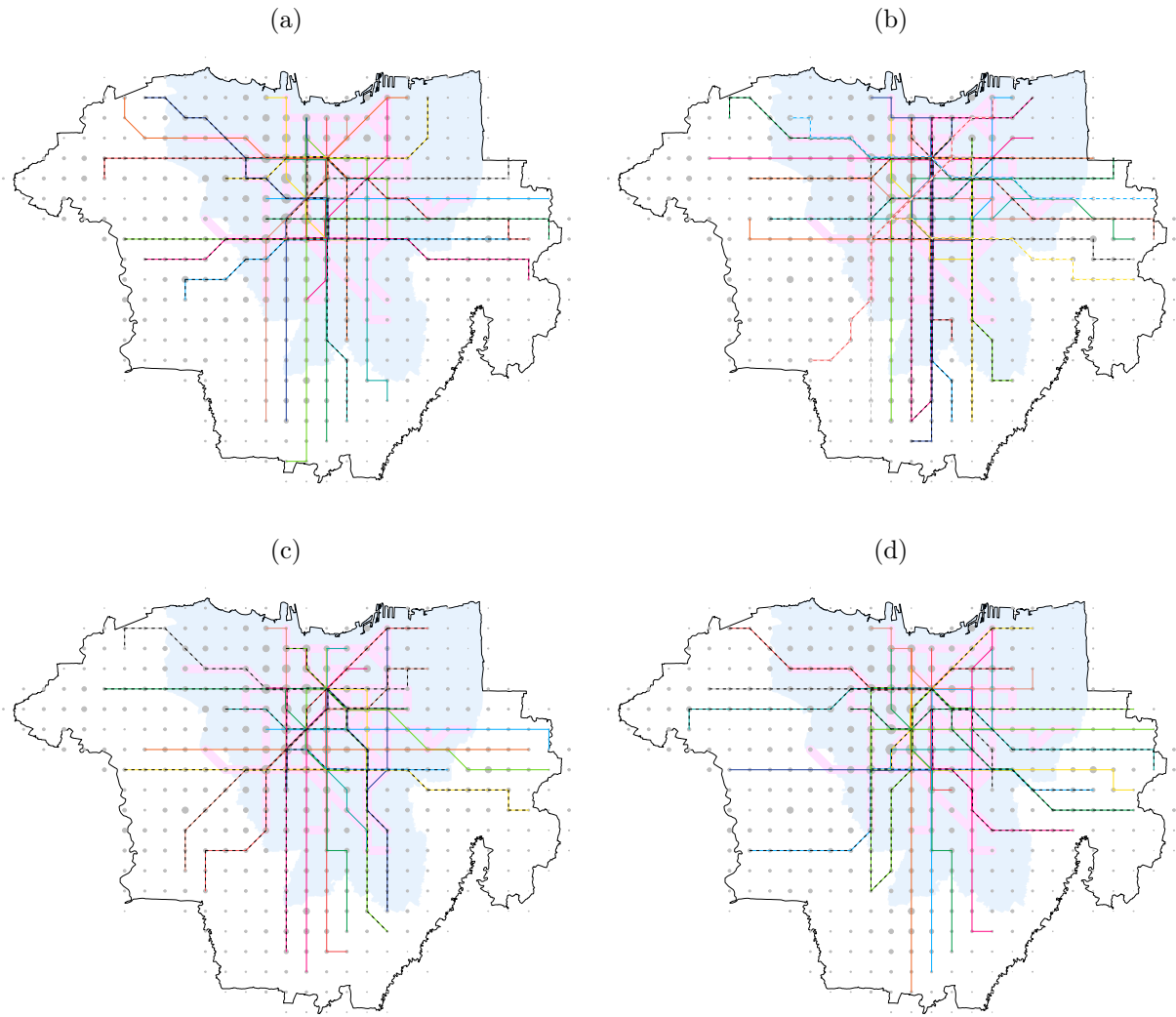
Notes: This figure plots the evolution of selected network characteristics across the 251,089 simulated annealing iterations, for all 200 independent runs. It shows that the algorithm spends roughly the first half of its iterations to explore the full space of available networks, before converging to a fairly stable set of optimal network characteristics.

Figure A.16: Distributions of Optimal Network Properties



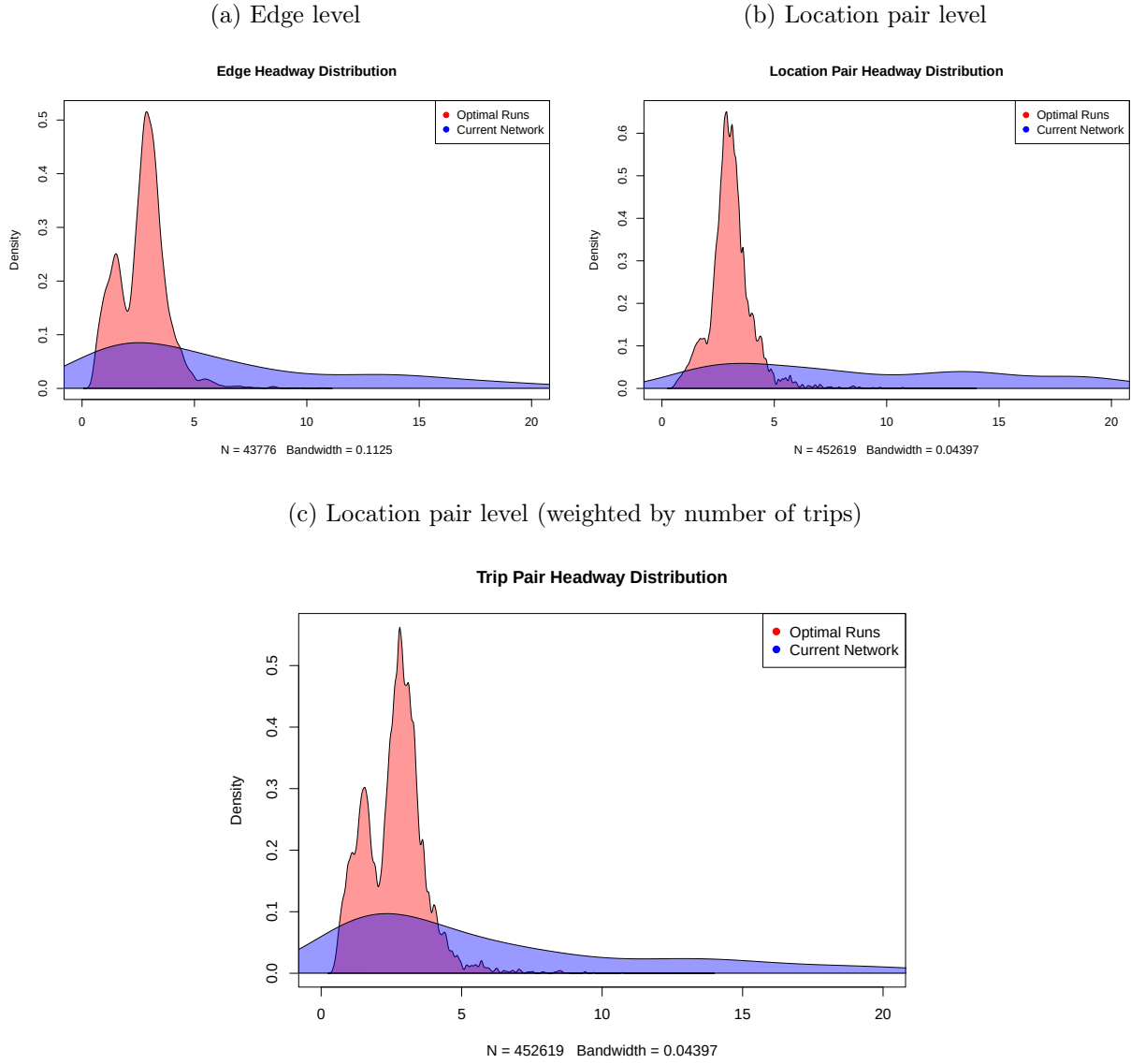
Notes: Each graph plots the kernel density graph of a specific network characteristic, for the final network from the simulated annealing algorithm, over the 200 parallel simulated annealing runs. The vertical line plots the measure for the current TransJakarta network for comparison.

Figure A.17: Optimal Network Examples



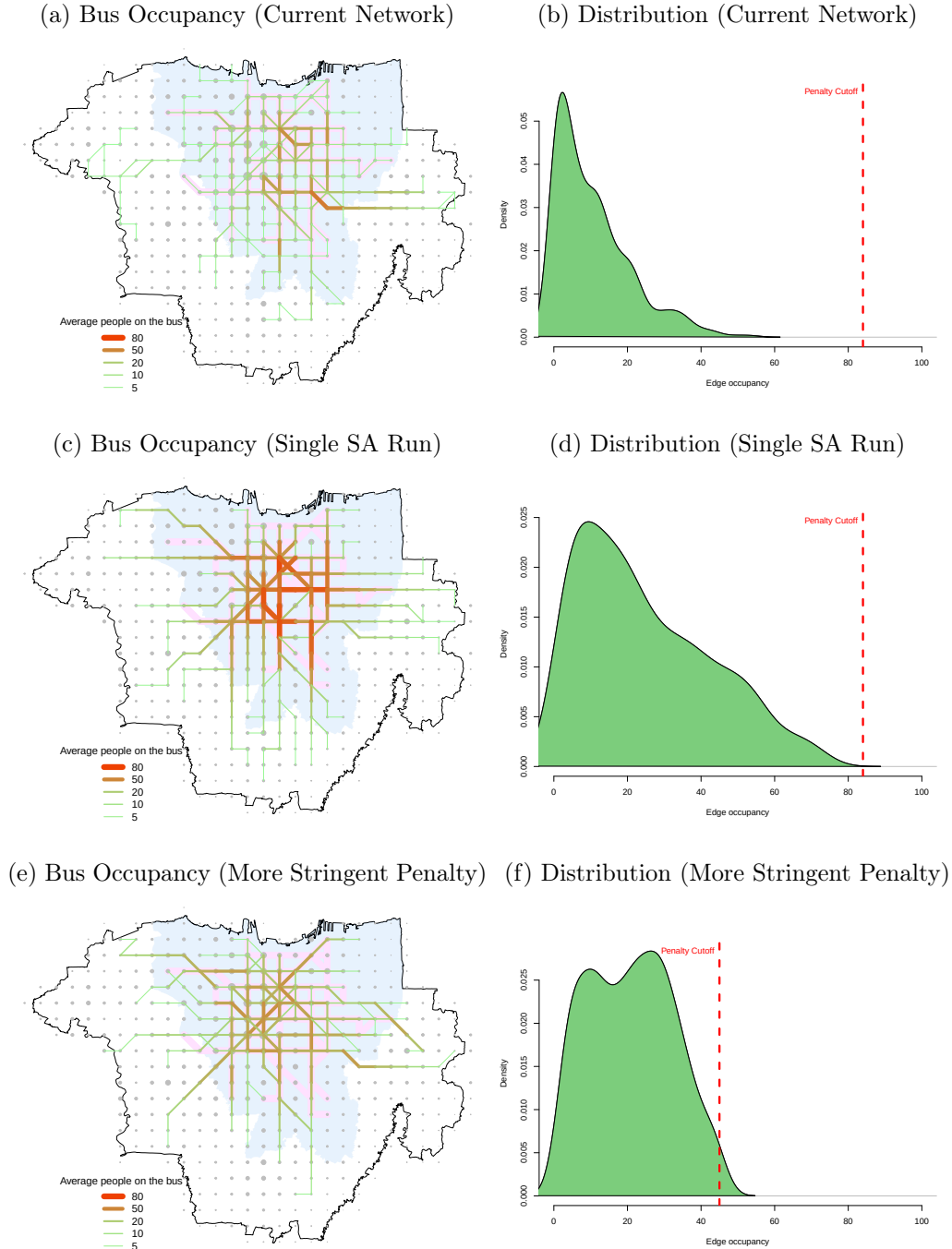
Notes: This figure depicts four additional randomly selected networks from the 200 networks obtained via SA. These represent random draws from the planner's distribution of optimal networks. Each network is obtained from an independent simulated annealing run.

Figure A.18: Distribution of Headways (in minutes)



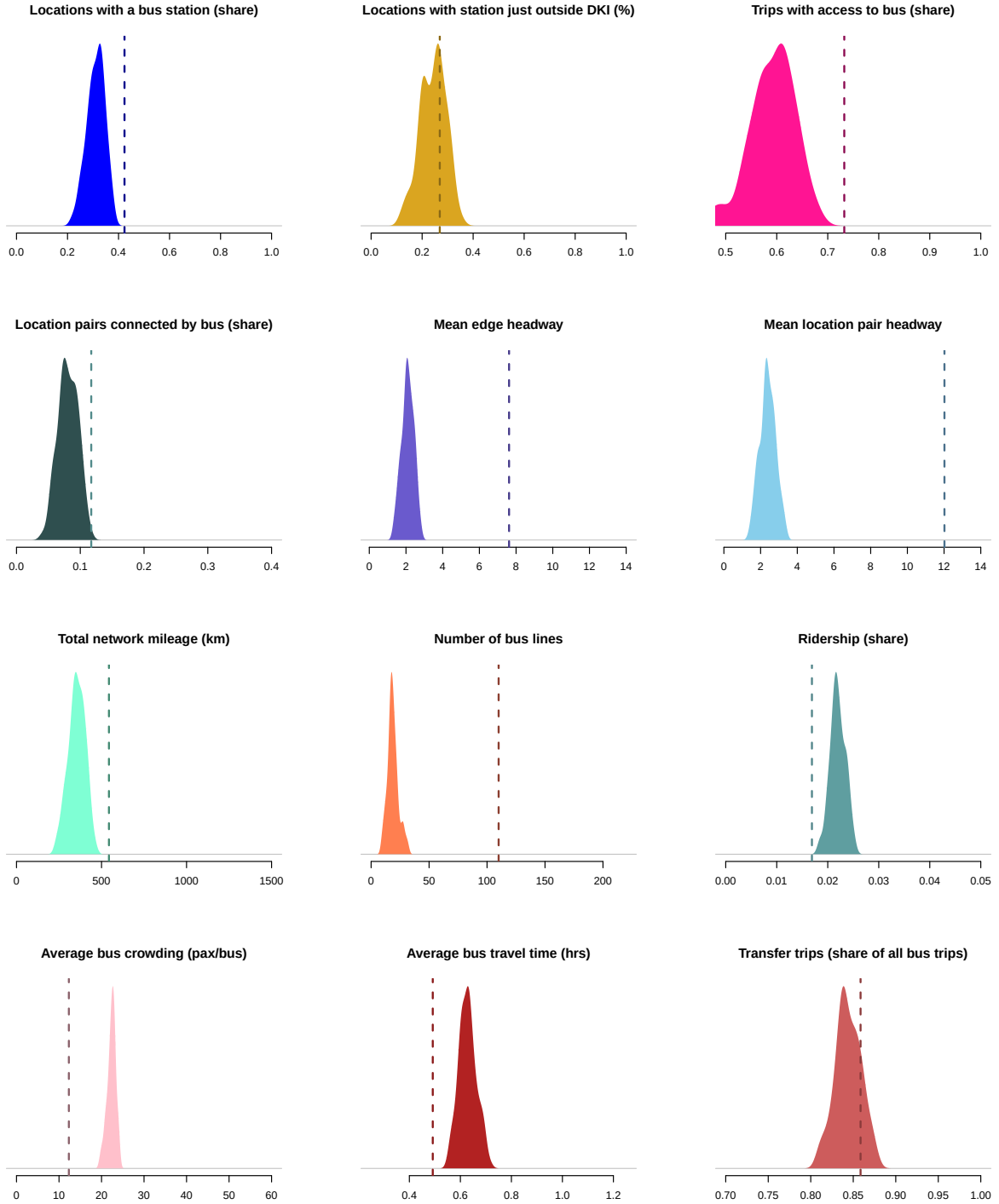
Notes: Distribution of headway times (in minutes) for the 200 optimal runs (in red) and the current TJ network (in blue).

Figure A.19: Average Bus Occupancy in Current and Optimal Networks



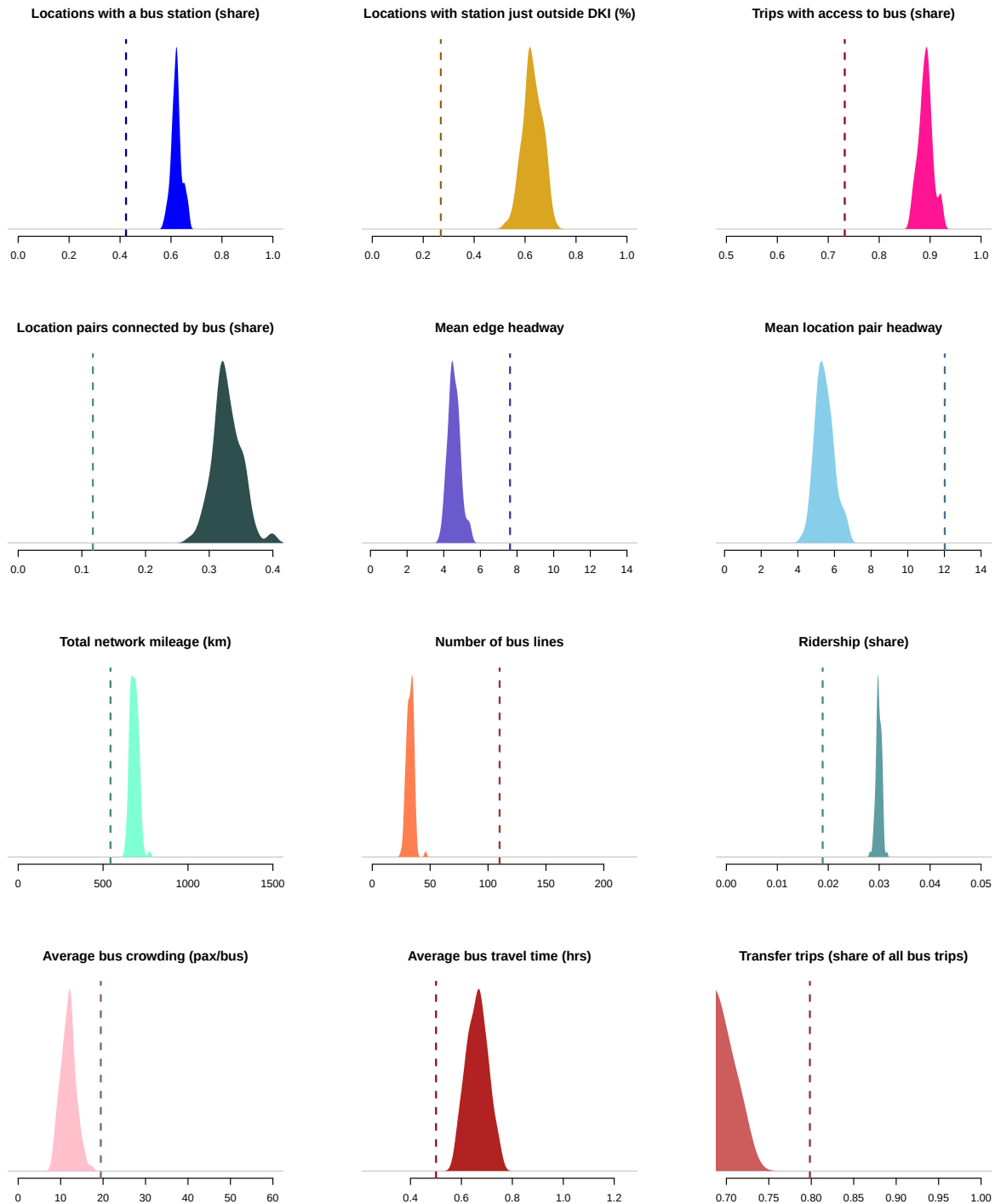
Notes: This figure depicts edge-level bus occupancy, defined as the average number of riders on a typical bus traveling across an edge, for the current TransJakarta network (a), for a randomly selected sampled network (b), and for a random sampled network with stricter (lower) bus capacity. To construct 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 buses run 5AM-10PM each day. Panels (b), (d) and (f) plot the distribution of edge-level occupancy and the assumed bus capacity. See Appendix A.7.2 for details. The counterfactual in panels (e) and (f) uses a bus capacity that is $1.85\times$ lower to account for within-day variation in demand. We obtain this number as follows. We use the aggregate ridership R_h by hour in the current network. Peak ridership (at 7 AM) is 85% larger than the average. We then lower all bus capacities by a factor of 1.85.

Figure A.20: Distributions of Optimal Network Properties: More Stringent Occupancy Penalty



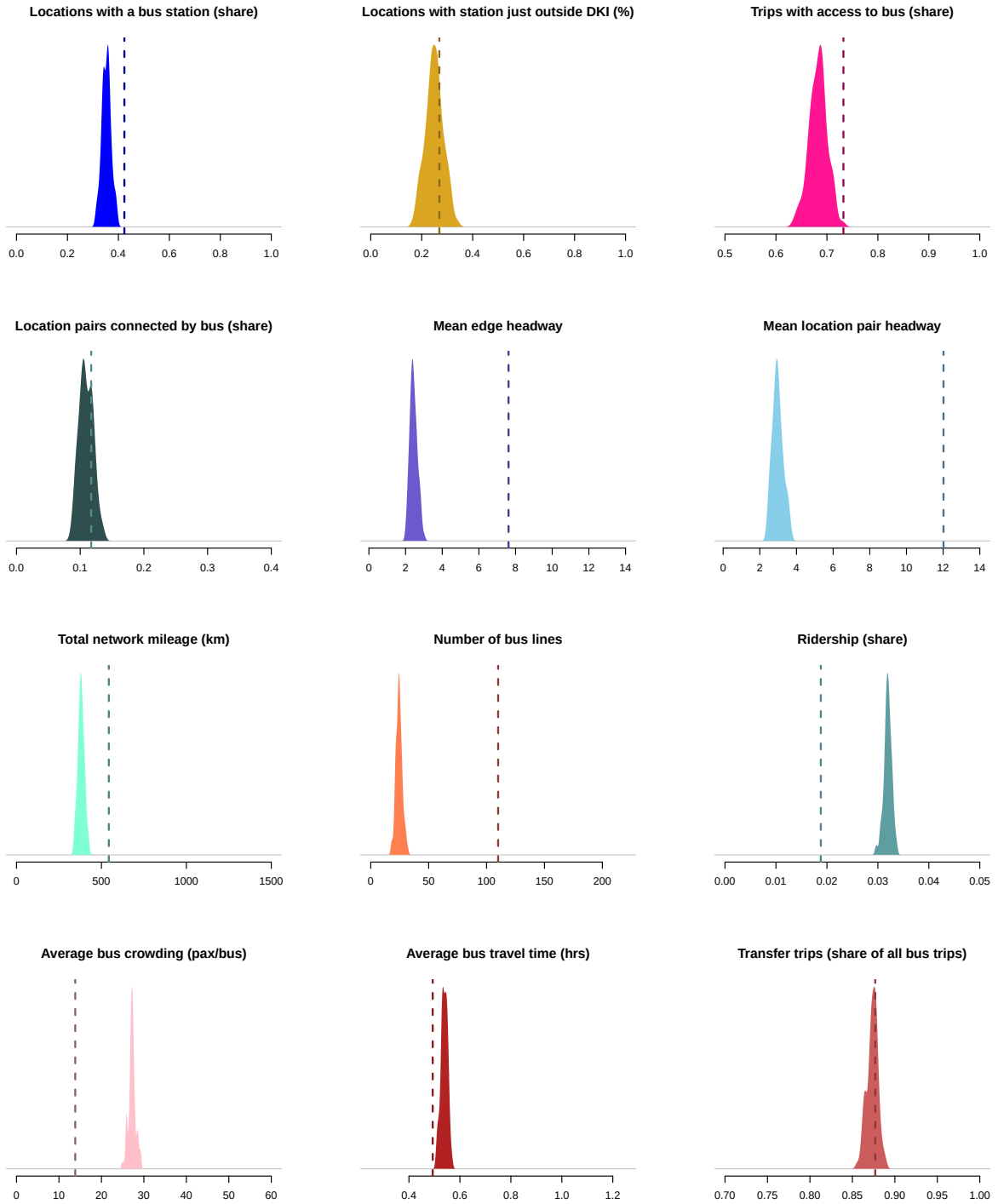
Notes: Each graph plots the kernel density graph of a specific network characteristic, for the final networks from 100 runs of the simulated annealing algorithm where the bus capacity is $1.85\times$ lower than in the baseline runs to account for within-day variation in demand. We obtain this number as follows. We use the aggregate ridership R_h by hour in the current network. Peak ridership (at 7 AM) is 85% larger than the average. We then lower all bus capacities by a factor of 1.85.

Figure A.21: Distributions of Optimal Network Properties: Heterogeneous Preferences by Poverty



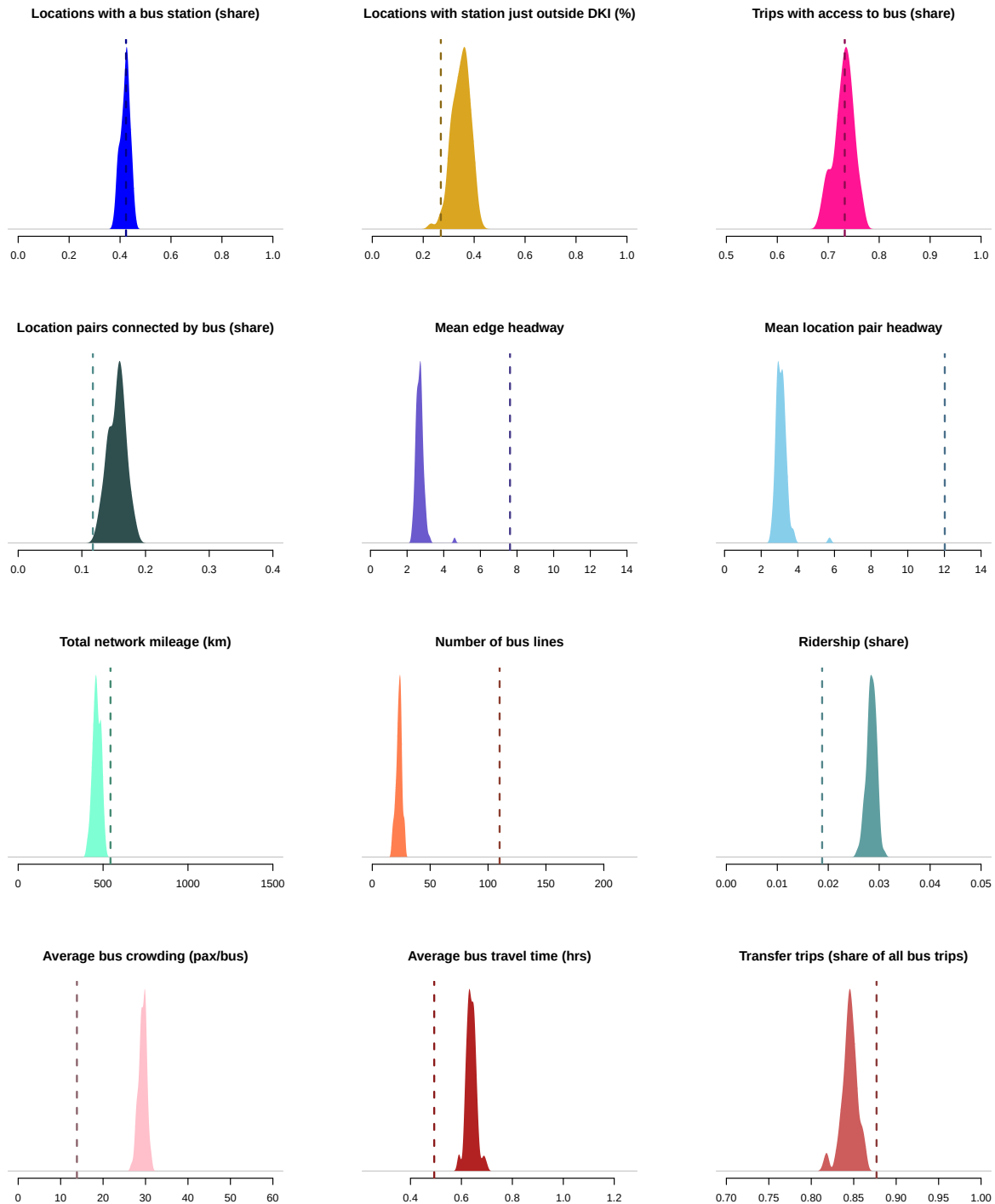
Notes: Each graph plots the kernel density graph of a specific network characteristic, for the final networks from 100 runs of the simulated annealing algorithm where preference parameters are allowed to vary for locations above and below median poverty. The planner puts equal weight on equivalent variation welfare expressed in travel time on the bus for the two groups. Bus crowding is computed based on joint preferences to ease computational burden.

Figure A.22: Distributions of Optimal Network Properties: Ridership as Objective



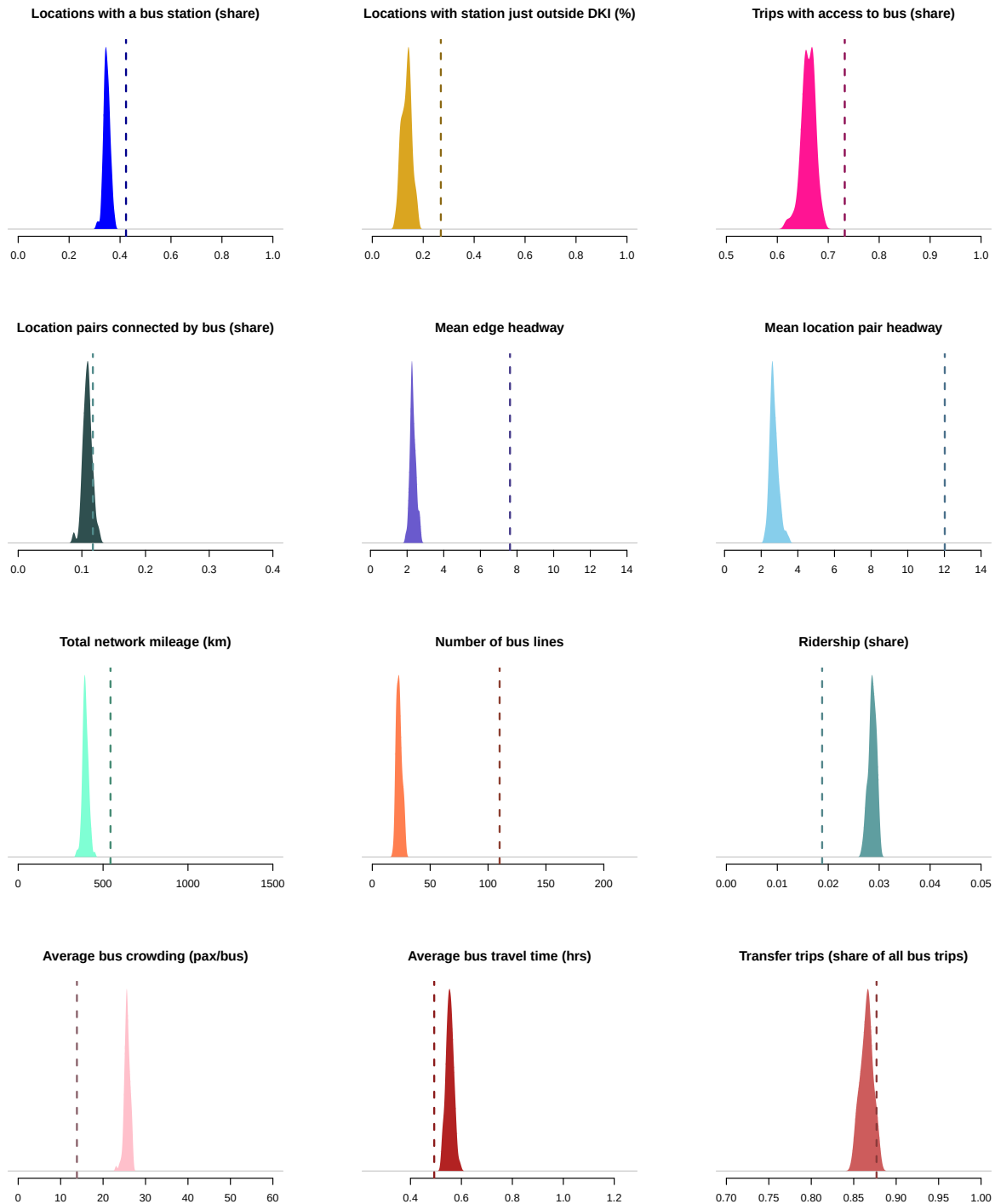
Notes: Each graph plots the kernel density graph of a specific network characteristic, for the final networks from 100 runs of the simulated annealing algorithm where the objective is total bus ridership instead of commuter welfare. The vertical line plots the measure for the current TransJakarta network for comparison.

Figure A.23: Distributions of Optimal Network Properties: Welfare Bias Towards Low Poverty Regions



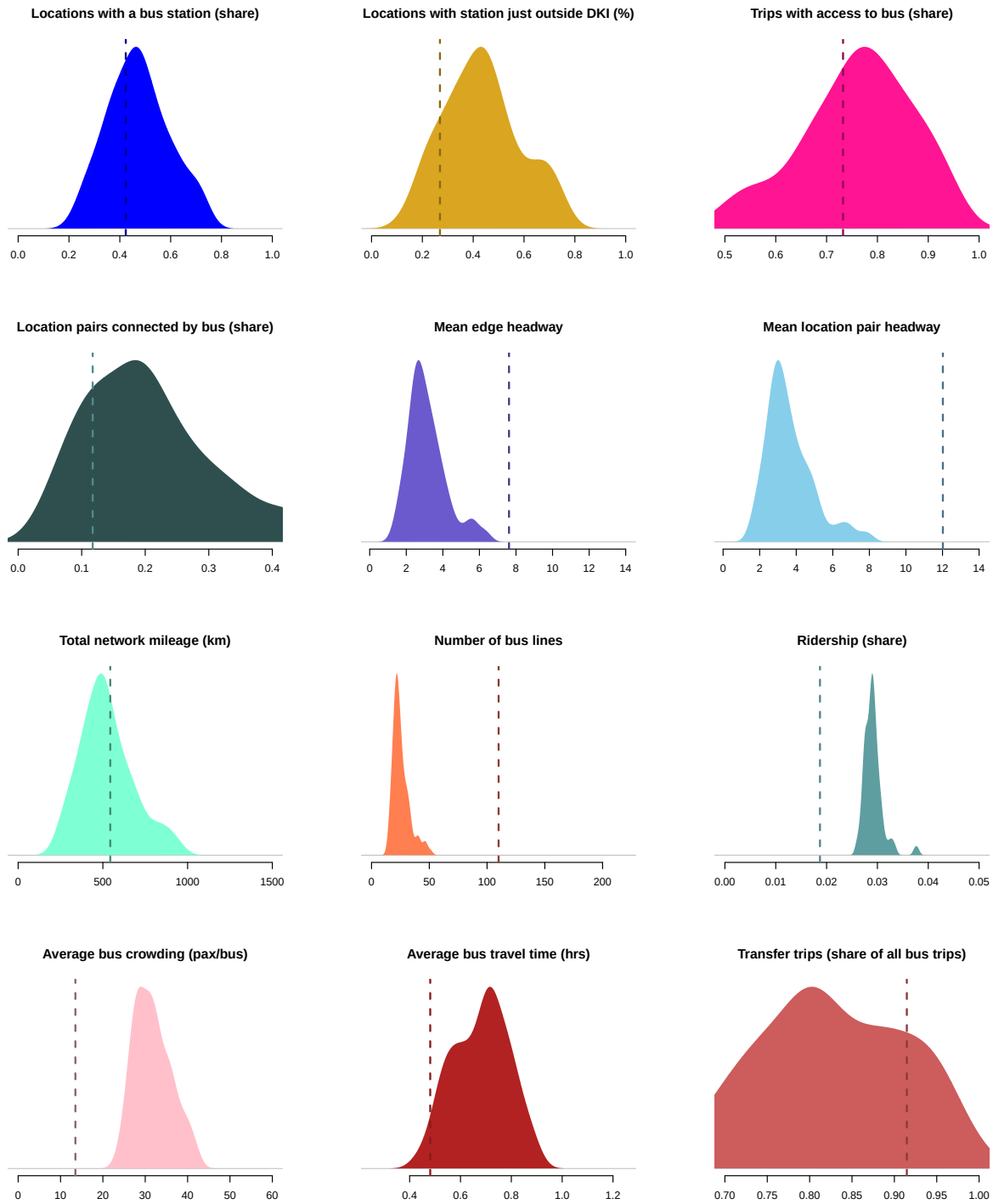
Notes: Each graph plots the kernel density graph of a specific network characteristic, for the final networks from 100 runs of the simulated annealing algorithm, in which the social planner is assumed to care more about low poverty regions. Grid cells with below-median poverty receive a welfare weight of $1.5\times$, while all grid cells above get a weight of $0.5\times$. The vertical line plots the measure for the current TransJakarta network for comparison.

Figure A.24: Distributions of Optimal Network Properties: Zero Welfare Weights Outside Central Jakarta (DKI Jakarta)



Notes: Each graph plots the kernel density graph of a specific network characteristic, for the final networks from 100 runs of the simulated annealing algorithm where we use a modified welfare function that puts zero welfare weights on commuters who live outside DKI Jakarta (the light blue region in maps). The vertical line plots the measure for the current TransJakarta network for comparison.

Figure A.25: Distributions of Optimal Network Properties with Bootstrapped Preference Parameters



Notes: Each graph plots the kernel density graph of a specific network characteristic, for the final networks from the simulated annealing algorithm with preference parameters drawn from the joint bootstrapped distribution. The vertical line plots the measure for the current TransJakarta network for comparison.

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: 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*** (0.008)			-0.039 (0.023)			0.162*** (0.024)		
E2: New Direct Line (quicker)		-0.273*** (0.025)			0.069* (0.034)			0.245*** (0.032)	
E3: Additional Busses			-0.018* (0.008)			0.278*** (0.023)			0.115*** (0.020)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson
Obs.	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	0.927	0.925	0.998			
Mean outcome	31.9	43.3	11.9	34.8	28.1	21.1	69.4	48.3	122.2
Median outcome	29.4	41.2	9.9	27.6	22.7	18.3	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

(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	0.024* (0.012)			0.296*** (0.040)			0.785*** (0.187)		
E2: New Direct Line (quicker)		-0.711*** (0.072)			0.225*** (0.033)			1.356*** (0.105)	
E3: Additional Busses			-0.086*** (0.015)			0.425*** (0.019)			0.359*** (0.103)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson
Obs.	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			
Mean outcome	27.8	50.7	12.2	10.1	9.8	7.8	12.0	2.9	6.4
Median outcome	21.4	49.4	10.5	8.7	8.4	7.0	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

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.3: 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 (0.011)			-0.042 (0.034)			0.126*** (0.022)		
E2: New Direct Line (quicker)		-0.298*** (0.036)			0.064 (0.049)			0.151*** (0.029)	
E3: Additional Busses			-0.092** (0.028)			0.328*** (0.068)			0.122*** (0.036)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson
Obs.	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			
Mean outcome	37.5	46.3	16.2	48.5	38.7	21.4	465.3	384.7	841.3
Median outcome	36.7	45.1	12.6	39.9	29.2	18.3	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

(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* (0.017)			0.281*** (0.057)			0.547* (0.222)		
E2: New Direct Line (quicker)		-0.655*** (0.119)			0.163* (0.060)			1.660*** (0.437)	
E3: Additional Busses			-0.120** (0.039)			0.389*** (0.024)			0.264 (0.243)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson
Obs.	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	0.938	0.940	0.991			
Mean outcome	45.6	66.2	20.1	10.2	9.5	7.1	117.7	27.6	55.7
Median outcome	32.8	66.2	19.3	8.7	8.4	7.0	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

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.4: Impact of Network Expansion on Travel Time, Wait Times, and **Bus Ridership**. Spatially Clustered Standard Errors.

(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*** (0.008)			-0.051 (0.031)			0.158*** (0.025)		
E2: New Direct Line (quicker)		-0.294*** (0.030)			0.071 (0.045)			0.268*** (0.036)	
E3: Additional Busses			-0.027* (0.012)			0.309*** (0.039)			0.092*** (0.024)
Orig $\times$ Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig $\times$ Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest $\times$ Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson
Obs.	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	0.922	0.921	0.997			
Mean outcome	33.2	44.2	12.9	36.6	28.0	21.1	111.3	76.9	210.4
Median outcome	31.0	42.4	10.6	29.4	22.0	18.3	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

(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*** (0.008)			0.294*** (0.051)			1.016*** (0.162)		
E2: New Direct Line (quicker)		-0.716*** (0.076)			0.208*** (0.033)			1.320*** (0.268)	
E3: Additional Busses			-0.089** (0.027)			0.417*** (0.022)			0.450** (0.138)
Orig $\times$ Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig $\times$ Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest $\times$ Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson
Obs.	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	0.964	0.964	0.992			
Mean outcome	34.8	49.3	13.0	10.9	10.8	7.7	19.2	7.5	26.0
Median outcome	24.8	47.2	11.0	7.0	8.9	7.0	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

Notes: This table replicates Table 1, with standard errors clustered two-way at the level of 2km hexagonal grids at the origin and destination location. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5: Robustness: Impact of Network Expansion on Travel Time, Wait Times, and Bus Ridership, **Excluding Observations after MRT Opening**

(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	0.033*** (0.009)			0.022 (0.032)			0.189*** (0.038)		
E2: New Direct Line (quicker)		-0.352*** (0.034)			0.185*** (0.051)			0.362*** (0.064)	
E3: Additional Busses			-0.035** (0.012)			0.315*** (0.034)			0.086*** (0.020)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson
Obs.	2,212,872	2,200,389	530,810	2,212,872	2,200,389	530,810	2,212,872	2,200,389	530,810
R^2	0.988	0.987	0.998	0.941	0.940	0.997			
Mean outcome	33.8	45.1	12.9	36.4	25.8	21.5	113.1	75.7	211.9
Median outcome	31.4	42.9	10.6	29.3	19.1	18.3	55.5	35.8	92.0
Unique origin x destination pairs	16,820	16,834	4,360	16,820	16,834	4,360	16,820	16,834	4,360
Unique origins	148	148	148	148	148	148	148	148	148

(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	0.065** (0.020)			0.299*** (0.046)			1.042*** (0.144)		
E2: New Direct Line (quicker)		-0.617*** (0.067)			0.204*** (0.051)			1.362*** (0.324)	
E3: Additional Busses			-0.088** (0.026)			0.416*** (0.024)			0.093 (0.124)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson
Obs.	175,779	179,716	82,418	175,779	179,716	82,418	175,779	179,716	82,418
R^2	0.994	0.993	0.999	0.970	0.970	0.993			
Mean outcome	41.1	52.9	13.3	12.8	11.3	7.7	23.4	6.5	9.3
Median outcome	24.9	51.8	11.5	7.0	8.9	7.0	0.0	0.0	0.0
Unique origin x destination pairs	2,283	2,344	923	2,283	2,344	923	2,283	2,344	923
Unique origins	57	57	63	57	57	63	57	57	63

Notes: Version of Table 1 excluding observations after March 12, 2019 – the day the sole Jakarta metro line (MRT) opened operation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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*** (0.030)	0.326*** (0.057)		0.972* (0.383)	1.400** (0.467)	
New Direct Line x High Poverty Origin	-0.096** (0.032)	-0.053 (0.061)		-0.079 (0.366)	0.489 (0.417)	
E3: Additional Busses			0.074* (0.036)			0.281 (0.242)
E3: Add'l Busses x High Poverty Origin			-0.048 (0.034)			0.591** (0.229)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	3,154,672	3,143,019	793,965	306,722	313,994	144,763
Event type	BRT 1	BRT 2	BRT 3	non-BRT 1	non-BRT 2	non-BRT 3
Median outcome x Low Poverty	49.5	39.4	102.7	0.0	0.0	0.0
Median outcome x High Poverty	69.2	38.6	86.6	0.0	0.0	0.0

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 an indicator for within-sample above median poverty rate. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.7: 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*** (0.006)			-0.134*** (0.027)			0.106*** (0.018)			-0.057 (0.119)		
E2: New Direct Line (quicker)		-0.212*** (0.027)			-0.122*** (0.028)			0.199*** (0.027)			0.046 (0.084)	
E3: Additional Busses			0.012 (0.032)			0.197** (0.060)			0.094* (0.040)			-0.007 (0.314)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Obs.	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						
Mean outcome	31.6	39.7	11.8	35.3	34.1	15.9	83.4	53.8	98.3	237.6	176.5	572.1
Median outcome	28.7	38.6	10.1	30.0	29.4	17.4	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

(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 (0.015)			0.278*** (0.039)			0.736*** (0.216)			-0.239 (0.177)		
E2: New Direct Line (quicker)		-0.738*** (0.084)			0.233*** (0.035)			1.276*** (0.147)			0.071 (0.129)	
E3: Additional Busses			-0.087*** (0.023)			0.353*** (0.018)			0.394** (0.128)			-0.223 (0.138)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Obs.	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	0.979	0.996						
Mean outcome	28.5	46.5	11.5	10.1	10.2	8.2	10.3	3.3	9.4	446.4	704.3	1238.0
Median outcome	22.1	46.7	10.1	8.7	8.4	8.4	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

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.8: Impact of BRT and Non-BRT Network Expansions on **Aggregate Trip Volume**. Spatially Clustered Standard Errors.

(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*** (0.011)			-0.157*** (0.040)			0.108*** (0.016)			-0.008 (0.052)		
E2: New Direct Line (quicker)		-0.214*** (0.027)			-0.107** (0.035)			0.208*** (0.034)			-0.073 (0.052)	
E3: Additional Busses			-0.006 (0.033)			0.174*** (0.047)			0.046 (0.040)			0.166 (0.112)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Obs.	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						
Mean outcome	32.6	42.0	13.4	38.5	33.9	14.4	122.9	93.7	195.5	1895.5	1729.5	10117.3
Median outcome	30.6	42.2	10.9	33.5	29.3	12.5	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

(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	0.028** (0.010)			0.260*** (0.055)			0.999*** (0.198)			-0.042 (0.056)		
E2: New Direct Line (quicker)		-0.737*** (0.101)			0.205*** (0.038)			1.337*** (0.321)			0.031 (0.056)	
E3: Additional Busses			-0.097** (0.035)			0.375*** (0.028)			0.425* (0.177)			-0.177* (0.076)
Orig × Dest FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dest × Week FE Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Obs.	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	0.973	0.973	0.996						
Mean outcome	35.9	44.3	12.4	11.2	11.6	7.9	13.0	7.6	40.7	3948.6	6060.7	6880.5
Median outcome	24.5	43.6	10.8	7.0	8.9	7.1	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

Notes: This table replicates Table 2, with standard errors clustered two-way at the level of 2km hexagonal grids at the origin and destination location. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.9: Demand Estimation Robustness and Extensions

	(1)	(2)	(3)	(4)	(5)	(6)
	Benchmark	Poverty: Low	Poverty: High	Constant σ_{od}	γ : Slope = 15	γ : Slope = 45
BRT: Wait time $\alpha_{\text{wait}}/\alpha_{\text{time}}$	2.1 [1.5, 3.4]	1.7 [1.3, 2.5]	3.6 [1.9, 6.5]	2.4 [2.1, 4.0]	2.2 [1.6, 3.7]	2.1 [1.5, 3.4]
BRT: Travel time α_{time}	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]
BRT: Transfer shifter μ/α_{wait} (minutes)	4.8 [0.8, 7.4]	5.6 [1.8, 7.5]	-0.2 [-6.6, 3.3]	6.2 [1.3, 9.6]	4.3 [0.4, 7.2]	4.9 [0.9, 7.2]
BRT: Attention cutoff η	1.36 [1.25, 1.46]	1.26 [1.18, 1.38]	1.49 [1.34, 1.86]	1.43 [1.34, 1.65]	1.32 [1.23, 1.43]	1.39 [1.27, 1.47]
Non-BRT: Wait time $\alpha_{\text{wait}}/\alpha_{\text{time}}$	3.41 [2.72, 5.01]	2.51 [1.56, 4.67]	5.08 [3.15, 7.96]	4.52 [3.29, 8.58]	3.53 [2.74, 5.40]	3.40 [2.75, 5.02]
Non-BRT: Transfer shifter μ/α_{wait} (minutes)	2.22 [-1.35, 4.57]	3.84 [-2.09, 10.34]	-1.91 [-11.91, 1.12]	-0.02 [-5.89, 4.12]	1.69 [-1.87, 4.08]	2.33 [-1.18, 4.24]
Non-BRT: Attention cutoff η	1.43 [1.42, 1.44]	1.44 [1.44, 1.44]	1.27 [1.20, 2.56]	1.42 [1.41, 1.44]	1.40 [1.39, 1.42]	1.44 [1.43, 1.45]
Logit parameter σ	0.13 [0.04, 0.32]	0.19 [0.07, 0.38]	0.03 [0.01, 0.07]	0.20 [0.05, 0.78]	0.11 [0.04, 0.30]	0.13 [0.04, 0.28]

Note: This table replicates the estimation from Table 4, column 2, under several different assumptions. The first column repeats the benchmark demand estimation. The second and third columns report estimates from the model restricted to above- and below-median poverty grid cells, respectively. The fourth column uses a constant $\sigma_{od} = \sigma$ instead of the normalization $\sigma_{od} = \sigma \frac{\bar{D}}{D_{od}}$. The fifth and sixth columns use the inattention slope in equation (4) set to 15 and 45, respectively, rather than 30 as in the main estimation. 95% confidence intervals in column 2-6 are based on re-estimating the model 200 times to match perturbed moments, see the notes for Table 4.

Table A.10: Moment Dependence on Parameters (Jacobian)

	Event 1 (α^{1B})	Event 2 (α^{2B})	Event 3 (α^{3B})
α_{Wait}	1.1	-2.0	0.2
α_{Time}	0.0	9.2	0.5
μ_{Transfer}	-0.1	-0.4	-0.0

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 $\sigma \partial m_j(\theta) / \partial \theta_i$, for moments $m_j \in \{\alpha^{1B}, \alpha^{2B}, \alpha^{3B}\}$ and parameters $\theta_i \in \{\alpha_{\text{wait}}^{\text{BRT}}, \alpha_{\text{time}}, \mu_{\text{transfer}}^{\text{BRT}}\}$. The partial derivative is scaled by the logit parameter σ .

Table A.11: Sensitivity Measure: Estimated Parameter Dependence on Moment Values

	Event 1 (α^{1B})	Event 2 (α^{2B})	Event 3 (α^{3B})
α_{Wait}	4.4	-0.5	25.0
α_{Time}	0.7	1.4	16.7
μ_{Transfer}	-3.5	-0.3	23.5

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

Table A.12: Optimal Network Characteristics: Robustness to Simulated Annealing Run Time

Statistic	Current Network	Baseline Optimal	One Day Run Optimal
Locations with a bus station (share)	0.42	0.45	0.45
Locations with station just outside DKI (%)	0.27	0.4	0.4
Trips with access to bus (share)	0.73	0.75	0.75
Location pairs connected by bus (share)	0.12	0.18	0.18
Mean edge headway	7.62	2.69	2.84
Mean location pair headway	12.02	3.056	3.26
Total network mileage (km)	543.6	464.01	485.13
Number of bus lines	110	21	23.44
Welfare (minutes travel time)	0	10.15	8.56
Ridership (%)	1.88	3.039	2.85
Average bus crowding (pax/bus)	13.84	31.37	30.081
Average bus travel time (hrs)	0.49	0.64	0.65
Transfer trips (share of all bus trips)	0.88	0.85	0.85

Notes: Comparison of characteristics of optimal networks to different run times of the simulated annealing algorithm. In our baseline results, we run the SA algorithm for 251,089 iterations (taking roughly 7 days of computing time). In some robustness runs, we run the SA algorithm for approximately 46,041 iterations (taking about 1 day). This table shows that networks achieved from the two approaches are generally similar, yet the shorter runs achieve slightly lower welfare and ridership.

Table A.13: Optimal Network Characteristics: Robustness to Planner's Parameter β

Statistic	Current Network	$\beta = 4 \times 10^2$	$\beta = 4 \times 10^3$	$\beta = 4 \times 10^4$	$\beta = 4 \times 10^5$
Locations with a bus station (share)	0.423	0.430	0.443	0.445	0.445
Locations with station just outside DKI (%)	0.269	0.380	0.398	0.400	0.400
Trips with access to bus (share)	0.732	0.739	0.750	0.751	0.752
Location pairs connected by bus (share)	0.117	0.163	0.175	0.176	0.176
Mean edge headway	7.618	2.840	2.920	2.926	2.926
Mean location pair headway	12.020	3.222	3.326	3.336	3.337
Total network mileage (km)	543.604	450.953	462.873	463.834	463.773
Number of bus lines	110	20.965	20.965	20.975	20.975
Ridership (share)	0.019	0.030	0.030	0.030	0.030
Average bus crowding (pax/bus)	13.836	30.817	31.296	31.357	31.376
Average bus travel time (hrs)	0.493	0.638	0.643	0.643	0.643
Transfer trips (share of all bus trips)	0.877	0.852	0.852	0.851	0.852

Notes: This table reports optimal networks characteristics for different values of β , the inverse standard deviation of planner shocks. Higher levels of β correspond to idiosyncratic shocks with lower variance.

Table A.14: Properties of Current and Optimal Networks with Bootstrapped Preference Parameters

Statistic	Network in 2016	Current Network	Optimal Networks	Difference Optimal – Current
Locations with a bus station (share)	0.17	0.42	0.47	0.049 [0.12]
Locations with station just outside DKI (%)	0.0077	0.27	0.43	0.16 [0.15]
Trips with access to bus (share)	0.43	0.73	0.76	0.027 [0.11]
Location pairs connected by bus (share)	0.022	0.12	0.2	0.087 [0.099]
Mean edge headway (minutes)	3.03	7.62	3.14	-4.48 [1.047]
Mean location pair headway (minutes)	3.86	12.02	3.65	-8.37 [1.31]
Total network mileage (km)	180.093	543.6	523.75	-19.85 [157.42]
Number of bus lines	19	110	25.51	-84.49 [7.36]
Ridership (%)	0.83	1.87	2.91	1.044 [0.19]
Average bus crowding (pax/bus)	14.74	13.51	31.74	18.23 [4.33]
Average bus travel time (hrs)	0.48	0.48	0.68	0.2 [0.11]
Transfer trips (share of all bus trips)	0.87	0.91	0.83	-0.088 [0.086]

Notes: This table replicates Table 5 using 100 bootstrapped runs of the optimisation algorithm, each with an independent draw from the joint distribution of estimated preference parameters from column 2, Table 4. In column 4 we report the mean and standard deviation of the difference between optimal and current networks, computed over all optimal network draws. See Figure A.25 for the distributions of properties of optimal networks.

Table A.15: Optimal Networks Local Comparative Statics

Statistic	Current Network	Baseline Optimal	Local derivative $dF^*/d\alpha_W$	Local derivative $dF^*/d\alpha_T$	Local derivative $dF^*/d\mu_T$
Panel A: Coverage measures					
Locations with a bus station (share)	0.42	0.45 (0.0013)	0.0042 [-0.022; 0.032]	-0.001 [-0.039; 0.039]	-0.0017 [-0.017; 0.013]
Location pairs connected by bus (share)	0.12	0.18 (0.0011)	-0.0048 [-0.025; 0.018]	-0.016 [-0.047; 0.017]	-0.0004 [-0.012; 0.011]
Total network mileage (km)	543.6	464.064 (1.66)	5.096 [-25.36; 38.58]	10.9 [-35.0041; 60.29]	-3.56 [-22.79; 14.18]
Panel B: Speed measures					
Bus time relative to quickest possible bus route	1.33	1.24 (0.0085)	-0.17 [-0.33; 0.0088]	-0.36 [-0.63; -0.11]	-0.029 [-0.12; 0.05]
Panel C: Directness measures					
Connected directly (share of all connected pairs)	0.21	0.15 (0.00064)	-0.00019 [-0.01; 0.0098]	-0.0014 [-0.019; 0.015]	-0.0034 [-0.01; 0.0025]

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 (19). 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 $\mu_{\text{transfer}}^{\text{non-BRT}}$. Derivatives are scaled down by a factor of 10^8 to aid legibility. 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 buses 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 buses 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.

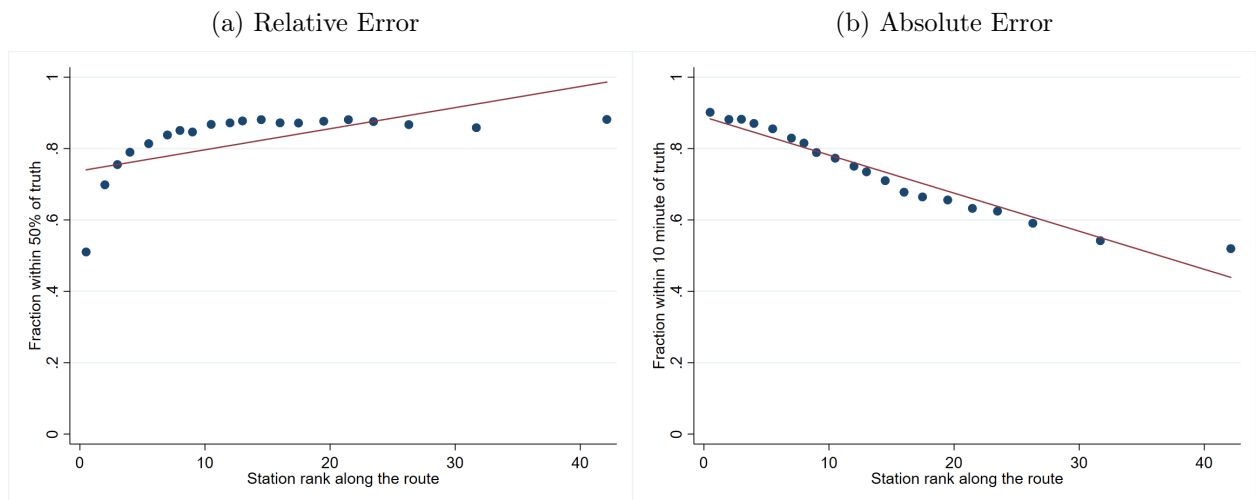
We developed three algorithms to identify when a bus arrives at a given bus station. When both GPS data and trip logs are available, we map match the bus GPS locations to the path of the bus route (from separate data), starting from the trip start time recorded in the trip log, and find arrival times 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.26 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.26: Accuracy of Predicting Station Arrival Time Based Only on Trip Start Time



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

A.3.2 Bus Travel Times

In this section, we describe how we compute travel times between an origin station o , a destination station d , along a route r . These travel times are used in the reduced form analysis to compute bus route travel times 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.27 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.16).

Figure A.27: Bus Travel Delay by Departure Time and Year

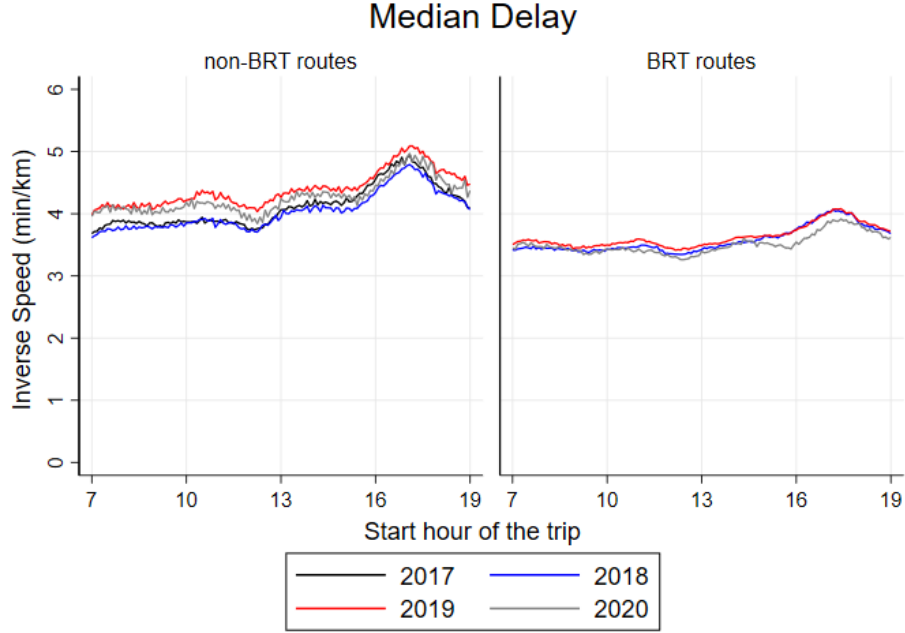


Figure Notes: This figure reports median delay (inverse speed, in minutes per km) by departure time, separately for BRT routes and for non-BRT routes using data from January 2017 - March 2020 for routes that were active throughout this period. To construct this figure, we assign each trip to 5-minute 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.16: Bus Travel Time Correlation Across Time

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

Table Notes: We organize the data at the route-origin-destination level. The outcomes variables are the log of median delay (inverse speed) 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).

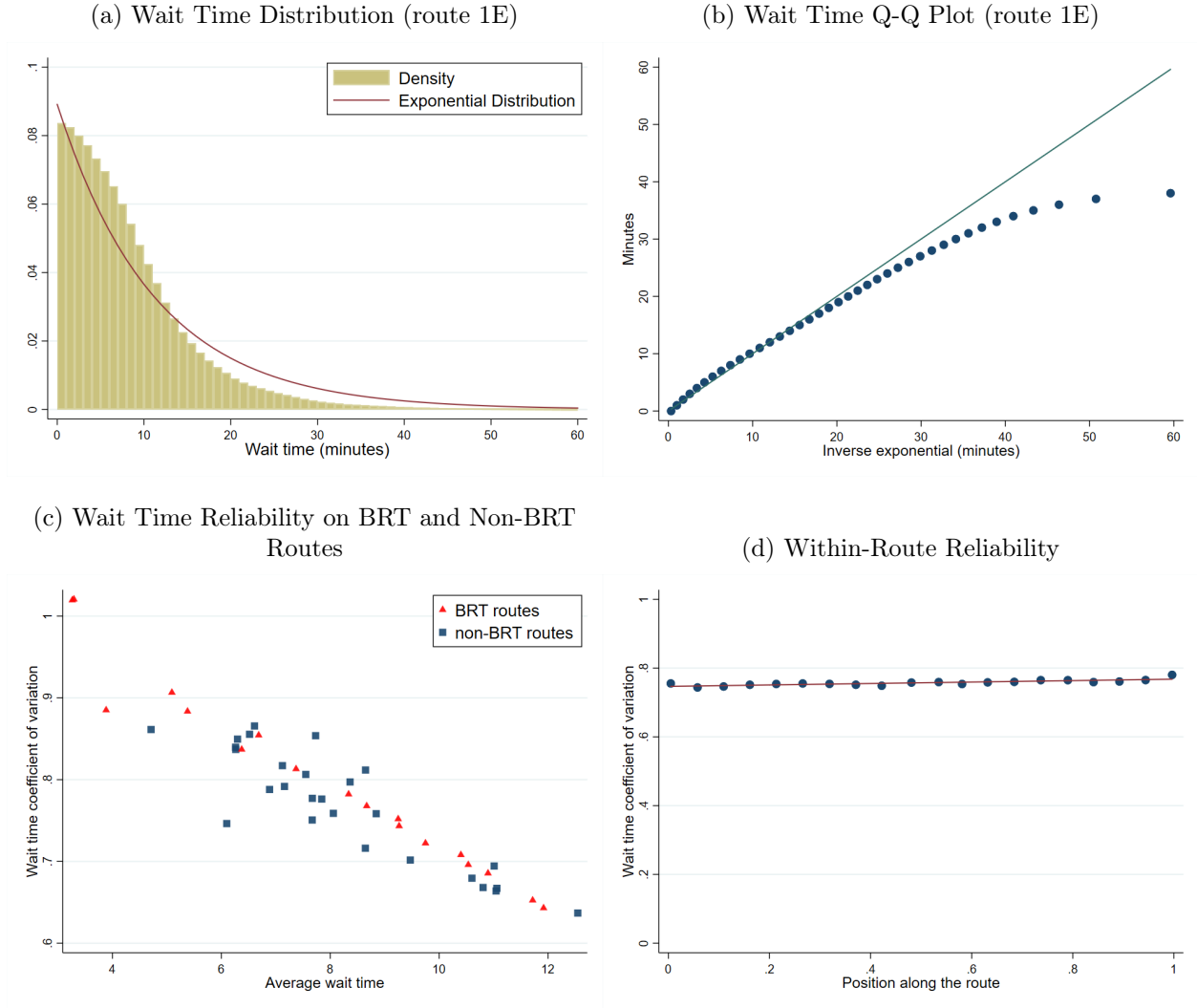
Panels A and B in Figure A.29 plot the wait time distribution for non-BRT route 1E. The wait time distribution is approximately exponential, except for wait times over 30 minutes, where the empirical distribution puts lower weight than the exponential. Panel B plots results for all the routes in our 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.28: 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.29: Wait Time Distribution Within and Across Routes



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 buses. 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.

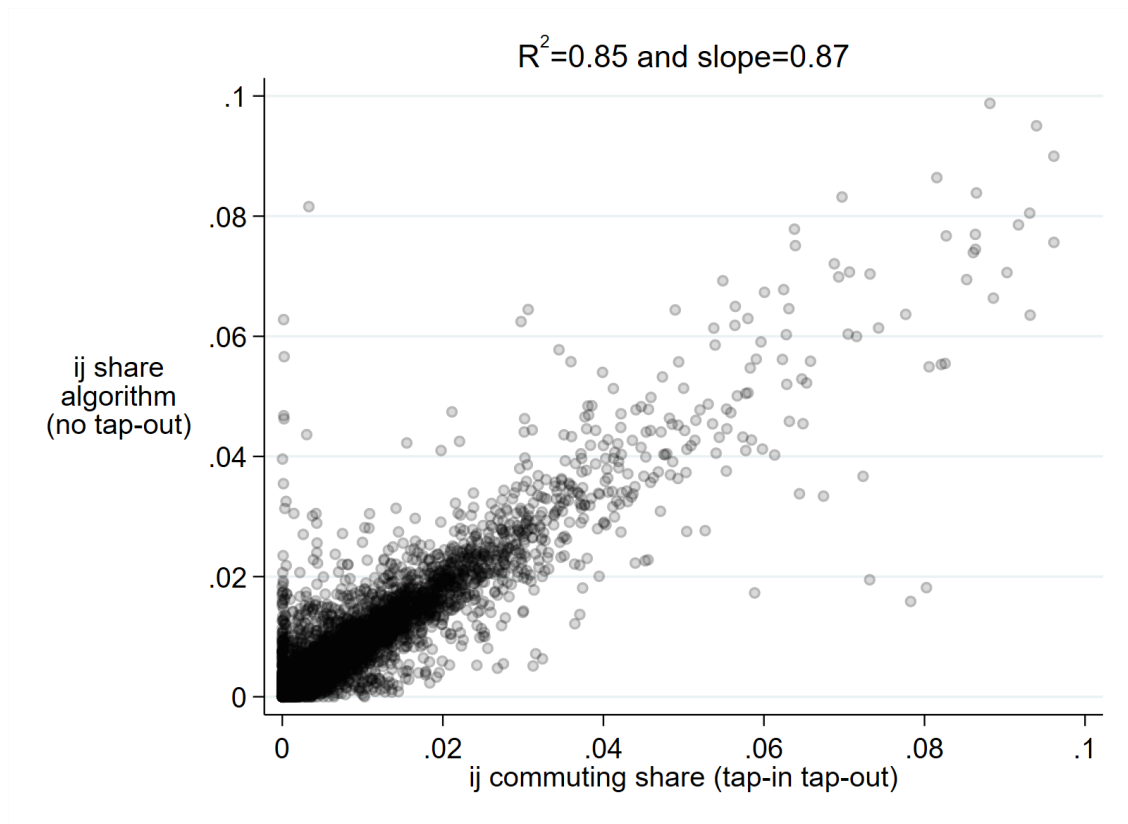
We exclude 15 non-BRT routes from ridership coding for reasons related to the way payments were made. Each non-BRT bus accepts payments with smart cards provided by a specific bank. We lack detailed micro-transaction data corresponding to one of the banks. Overall, tap data from dropped routes accounts 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.30: 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 \(2024\)](#) 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.”

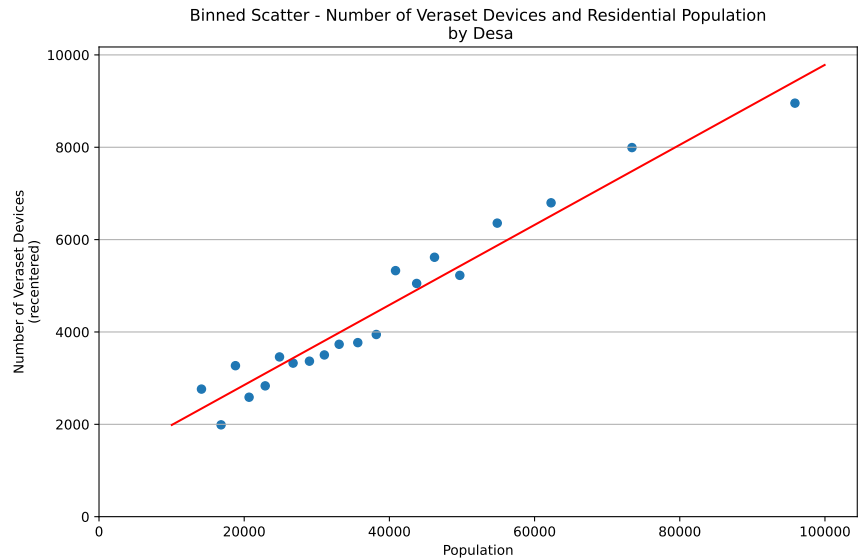
The sample of trips used in the analysis is all weekday trips starting and ending at a known locations (i.e. locations classified by DBSCAN as belonging to a cluster), excluding trips starting before 5AM or after 11PM. We also drop trips that are unusually long or short (in both duration and distance), “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.31](#)). 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 device, 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 \(2024\)](#).

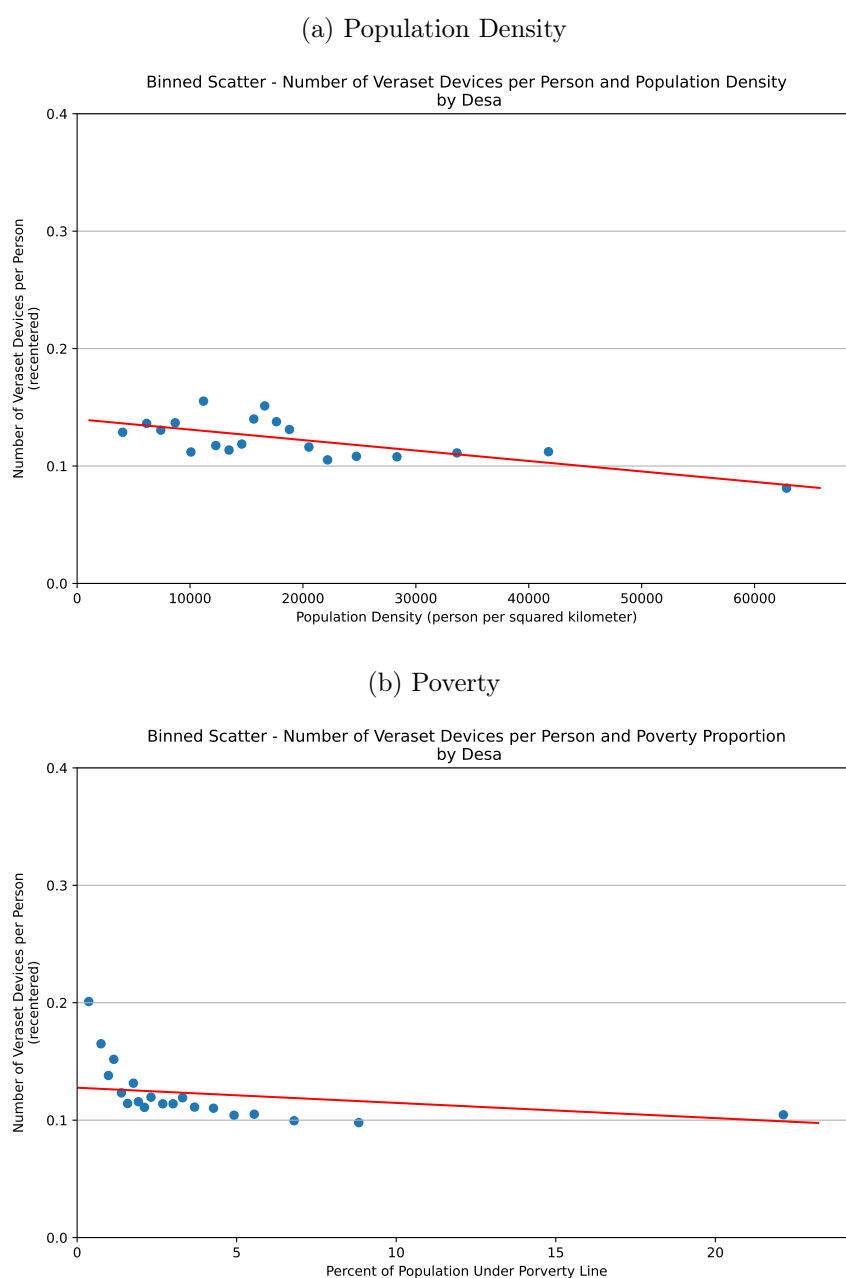
Second, we construct a cross-section of typical (o, d) trip flows throughout the period. We apply the same procedure as for the panel, except that we first pool all data for the entire two year period.

Figure A.31: 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.32: Representativeness of the Veraset Trip Data



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 $Post_{odt}^{iB}$, $i = 1, 2$, to be a time-varying indicator for the new direct connection between BRT grid cells o and d having been switched on in the past 10 months.⁶⁹ 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.⁷⁰

For non-BRT events, $Post_{odt}^{iN}$, $i = 1, 2$, is a time-varying indicator 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 buses from a new route because it overlaps with the existing route for the portion between o and d . Specifically, $Post_{odt}^3$ is an indicator 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 buses are added to the route between o and d due to an additional direct route.⁷¹ 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.⁷²

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

⁶⁹We 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.

⁷⁰This 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.

⁷¹As 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 .

⁷²We 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 $\tilde{\lambda}_k = \phi(p_k)\lambda_k$ 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:

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

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 $X_1, X_2, \dots, X_N$, then

$$\Pr(k \in \arg \max_j X_j) = \int_{-\infty}^{\infty} f_k(x) \prod_{i \neq k} F_i(x) dx, \quad (12)$$

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

$$\mathbb{E} \max_k X_k = \int_{-\infty}^{\infty} x \sum_k f_k(x) \prod_{i \neq k} F_i(x) dx. \quad (13)$$

The probability density that x is the maximum is the sum over k that the k -th options is $X_k = x$ and that all other options are strictly lower than x .

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 $v_1 < v_2 < \dots < v_N$. Replacing the pdf and cdf in (12) and separating the integral by intervals delimited by the v_k 's, the probability that option k is optimal is:

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

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

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

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

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

Expected Utility. Plugging the exponential pdf and cdf formulae in (13) 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:

$$\mathbb{E} T_k^{\text{time}} \equiv \mathbb{E} (T_k^{\text{time}} \mid k \in \arg \max) = \sum_k \pi_k T_k^{\text{time}}. \quad (14)$$

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 2020.⁷³ 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 $\hat{T}_{od}^{\text{bus}}$ 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 primitives 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 $\hat{T}_{od}^{\text{bus}}$, 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 $\tilde{T}_p = T_p + \epsilon_p$, 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

⁷³Through 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

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

Allen and Arkolakis (2022) prove that the matrix $B = (\exp(-\nu \tilde{T}_{od}))_{od}$ 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 $\hat{T}_{od}^{\text{bus}}$. We minimize the following objective function

$$\min_{\nu, (T_{ij})_{ij}} \sum_{o,d} \psi_{od} \left(\log(\hat{T}_{od}^{\text{bus}}) - \log(\tilde{T}_{od}) \right)^2$$

using weights ψ_{od} equal to the inverse squared distance between grid cells o and d . We estimate $\hat{\nu} = 118$ for BRT and $\hat{\nu} = 116$ 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 $\hat{T}_{ij}$ for BRT and non-BRT as bus travel time in all our counterfactual network exercises.

Figure A.33 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'

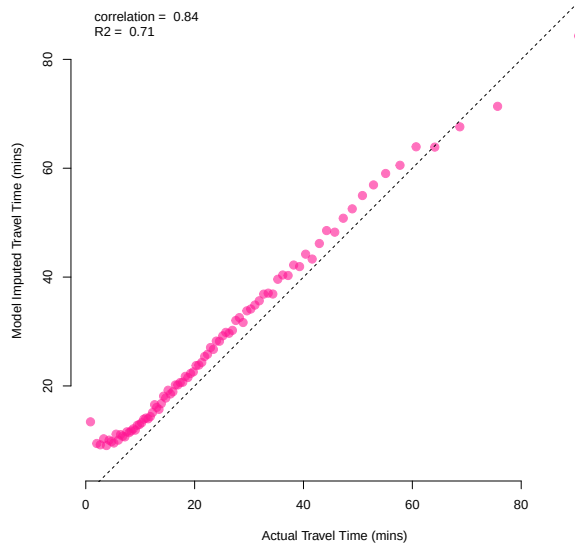
$$B_{ss'r} = \frac{1}{D_r} \sum_{o \leq s < s' \leq d} R_{odr}, \quad (15)$$

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

$$T_{ss'r}^p = T_{ss'r} \cdot \kappa(B_{ss'r}), \quad (16)$$

Figure A.33: 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 $\hat{T}_{od}$ 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, $B \leq \bar{B}$, and convex and increasing for $B > \bar{B}$. We parameterize κ as

$$\kappa(B) = \max(1, \exp(0.2 \cdot (B - 84))). \quad (17)$$

We assume there is no penalty up to $\bar{B} = 84$ 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 (15) based on route-specific ridership flows R_{odr} , and ridership patterns R_{odr} are given by choice probabilities based on equations (2), (3) and (5), where we replace travel time T_{odr} with perceived travel time T_{odr}^p given by (16), 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 $\bar{B}$. 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 Model Welfare Bound

Here we prove a slightly stronger version of Proposition 3. Without loss of generality, we assume $W^{\max} = 0$, so all other welfare levels are non-positive. Denote $\mathcal{N}^{\neg\max}$ the set of networks excluding the welfare maximizing network (picking one arbitrarily if there are ties). We rewrite expected welfare as

$$\mathbb{E}_\pi W = \sum_{N \in \mathcal{N}^{\neg\max}} x_N \pi(N) = \sum_{N \in \mathcal{N}^{\neg\max}} x_N \frac{\exp(\beta x_N)}{1 + \sum_{M \in \mathcal{N}^{\neg\max}} \exp(\beta x_M)}. \quad (18)$$

We focus on minimizing $\mathbb{E}_\pi W$ with respect to welfare levels $x_N \leq 0$ for all $N \in \mathcal{N}^{\neg\max}$.

Lemma 1. *To minimize $\mathbb{E}_\pi W$, all welfare levels are interior, meaning $-\infty < x_N < 0$ for all $N \in \mathcal{N}^{\neg\max}$.*

Proof. Write $\mathbb{E}_\pi W$ as a function the welfare of one of the networks $N \in \mathcal{N}^{\neg\max}$, fixing the other:

$$\mathbb{E}_\pi W = f(x) = \frac{x \exp(\beta x) + \Sigma_2}{1 + \exp(\beta x) + \Sigma_1},$$

where

$$\begin{aligned} \Sigma_1 &\equiv \sum_{M \in \mathcal{N}^{\neg\max} \setminus \{N\}} \exp(\beta x_M) \geq 0 \\ \Sigma_2 &\equiv \sum_{M \in \mathcal{N}^{\neg\max} \setminus \{N\}} x_M \exp(\beta x_M) \leq 0 \end{aligned}$$

The function f is continuously differentiable with endpoints $f(-\infty) = \frac{\Sigma_2}{1+\Sigma_1} < f(0) = \frac{\Sigma_2}{2+\Sigma_1} \leq 0$. This means $x = 0$ cannot minimize f . We next rule out that $x = -\infty$ can minimize f . It is easy to check that $f'(x) \propto 1 + \Sigma_1 - \beta\Sigma_2 + \exp(\beta x) + \beta x(1 + \Sigma_1)$, which is negative for sufficiently negative x . This means that f is minimized for $-\infty < x < 0$. $\square$

Lemma 2. *To minimize $\mathbb{E}_\pi W$, all welfare levels are equal to the solution of*

$$(n - 1) \exp(\beta x) = -(1 + \beta x),$$

where $n = |\mathcal{N}|$ is the total number of networks.

Proof. Using standard results for logit probabilities, the partial derivative of (18) is

$$\begin{aligned} \frac{d}{dx_N} \mathbb{E}_\pi W &= \frac{d}{dx_N} x_N \pi(N) + \frac{d}{dx_N} \left(\sum_{M \neq N} x_M \pi(M) \right) \\ &= \left(\pi(N) + x_N \pi(N) (1 - \pi(N)) \beta \right) + \left(\sum_{M \neq N} -x_M \pi(N) \pi(M) \beta \right) \\ &= \pi(N) \left(1 + \beta x_N - \sum_M \beta x_M \pi(M) \right) \end{aligned}$$

where the sums in the first two rows are over $M \in \mathcal{N}^{\neg \max} \setminus \{N\}$.

Setting this to zero, and noting that $\pi(N) > 0$ for all N and that the last summation term in the last row does not depend on N , we find that all optimal x_N are equal. Setting all $x_N = x$ and using the logit expression for π_M we get

$$1 + \beta x = (n - 1) \beta x \frac{\exp(\beta x)}{1 + (n - 1) \exp(\beta x)}.$$

This is equivalent to the expression in the lemma. $\square$

To finish up the proof of Proposition 3, note that the function $g(x) = (n - 1) \exp(\beta x) + 1 + \beta x$ is increasing in $x \leq 0$, that $g(0) > 0$, and it is easy to check that plugging in $x < 0$ that satisfies $1 + \beta x = -\log(n)$ gives $g(x) = \frac{n-1}{n} e^{-1} - \log(n - 1) < 0$.

This implies that at the common welfare level $X_N = x^*$ for all $N \in \mathcal{N}^{\neg \max}$ we have $1 + \beta x^* > -\log(n)$. Minimized expected welfare then satisfies

$$\begin{aligned} \min \mathbb{E}_\pi W &= \frac{x^*(n - 1) \exp(\beta x^*)}{1 + (n - 1) \exp(\beta x^*)} \\ &= \frac{-x^*(1 + \beta x^*)}{1 - 1 - \beta x^*} = \frac{1 + \beta x^*}{\beta} = \\ &> -\frac{\log(n)}{\beta} \quad \square \end{aligned}$$

A.7.4 Analytic Results for Optimal Allocations Model

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

$$D_\theta f^*(\theta) = \sum_N \pi(N; \theta) \left(\beta D_\theta W(N; \theta) (f(N, \theta) - f^*(\theta)) + D_\theta f(N; \theta) \right), \quad (19)$$

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.

To prove this, we first use the logit formula to show that

$$D_\theta \pi(N, \theta) = \beta \pi(N, \theta) \left(D_\theta W(N, \theta) - \mathbb{E}_{N'} D_\theta W(N', \theta) \right) \quad (20)$$

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

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

A.7.5 The Simulated Annealing (SA) Algorithm Asymptotic Result

In this section, we define the Metropolis-Hastings (MH) and (modified) Simulated Annealing (SA) algorithms formally, and include a brief discussion of mixing times. This material draws on (Levin and Peres, 2017).

The MH algorithm asymptotically samples from the planner’s distribution π in the following sense:

Proposition 4. *Assume that the proposal distribution Ψ is irreducible, aperiodic, and satisfies $\Psi(N | N') > 0 \iff \Psi(N' | N) > 0$. The Metropolis-Hastings algorithm with transition probabilities given by (11) is an ergodic Markov chain, meaning it has a unique stationary distribution, and that unique stationary distribution is the planner’s probability distribution π . As $k \rightarrow \infty$, the state N_k converges in distribution to π .*

This is a classic result, see for example (Levin and Peres, 2017), section 3.2.

Mixing Times for MH. The mixing time of a Markov chain formalizes how many steps it takes to get “close” to the stationary distribution. For the k -th step of MH, we define the worst-case distance to the stationary distribution π as

$$d(k) = \max_{N_1 \in \mathcal{N}} \|\nu_k(N_1) - \pi\|_{TV},$$

where $\nu_k(N_1)$ is the distribution of N_k when MH starts at N_1 and $\|\alpha - \beta\|_{TV} = \max_{A \subseteq \mathcal{N}} |\alpha(A) - \beta(A)|$ is the total variation distance between two probability distributions α and β over $\mathcal{N}$.

The convergence theorem for ergodic Markov chains states that $d(k)$ falls exponentially in k . Specifically, there exist constants $C > 0$ and $\gamma \in (0, 1)$ such that $d(k) \leq C\gamma^k$.

The mixing time is typically defined as $t_{\text{mix}} = \min\{k : d(k) \leq \frac{1}{4}\}$.⁷⁴

The convergence theorem implies that the constant γ is the crucial determinant of mixing times. The closer γ is to 1, the slower the chain is guaranteed to mix.

The concepts of slow and fast mixing are defined with respect to a family of problems underlying the MH algorithm (or any Markov chain) that depend on a growing parameter n that captures the problem size. In our setting, n could be, for example, the number of grids in the geographic environment. The object of interest is then how the expectation of the mixing time depends on n . When this dependence is itself exponential, we say that MH mixes slowly. Mixing times that are polynomial in n are considered fast (see footnote 58 for an example).

⁷⁴Neither the distance measure nor the value of the $\frac{1}{4}$ constant (which is used by convention), matter for the substantive mixing time properties of a chain.

The problem with the MH algorithm is that it can exhibit slow mixing when, loosely speaking, the welfare function W has multiple local maxima (defined with respect to the proposal function Ψ).

We next define the SA algorithm, which is one of several algorithms that build on MH with the aim to improve its convergence properties.

The Modified Simulated Annealing Algorithm SA runs for a fixed number of steps K . SA begins at an arbitrary network N_1 . Given network N_k at step k , it draws a candidate network N' from a probabilistic proposal distribution $\Psi(N' | N_k)$. SA accepts the network N' with probability:

$$\Pr(N_{k+1} = N' | N_k) = \min \left(1, \frac{\exp(\beta_k W(N')) \Psi(N_k | N')}{\exp(\beta_k W(N_k)) \Psi(N' | N_k)} \right), \quad (21)$$

and keeps $N_{k+1} = N_k$ otherwise. That is, 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 . The term β_k is called the “inverse temperature” and it follows a deterministic schedule. It grows exponentially from some initial value β_1 to $\beta_K = \beta$. (Hence, at the end of the SA algorithm the inverse temperature coincides with the planner’s parameter β). That is, $\beta_k = \beta_1 \rho^k$ for an appropriate $\rho > 1$.

This algorithm also asymptotically samples from the planner’s distribution π .

Proposition 5. *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 $\Psi(N | N') > 0 \iff \Psi(N' | N) > 0$. As the number of steps $K \rightarrow \infty$, the final state N_K of the simulated annealing algorithm converges in distribution to π , the planner’s distribution over allocations.*

The proof uses standard Metropolis-Hastings algorithm arguments. It mirrors a classic proof that as the number of steps increases, the outcome of SA converges to the global optimum.

As the number of steps K grows, the SA 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.7.6 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 buses between two randomly drawn bus routes.

Global Bus Modifiers

1. Give or take away buses from a randomly selected bus route. Redistribute buses among randomly chosen other lines to stay at the constraint of 1,500 total buses.

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 buses 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 buses 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 buses among other, randomly chosen, bus routes.
4. Delete a randomly chosen set of 10 bus routes at once. Redistribute the buses among other, randomly chosen, bus routes.
5. Add 10 random bus lines at once. Assign a random number of buses 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 (21)

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

We make this approximation because computing the proposal function ratio term precisely is computationally intensive. $\Psi(N' | N_k)$ 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

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

This is significantly larger than 1, meaning that transitions towards networks with many routes is favored in equation (21). 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

$$\tilde{\pi}(N) \propto \pi(N) \cdot \left(\frac{R}{r(N) + 1} \right)^{-1}$$

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.

A.7.7 Details on Statistical Test of Mixing

To conduct the test of statistical mixing outlined in the main test, we need an upper bound on the number of hypothetically feasible networks $\bar{n}$. This number is extremely large, but can nevertheless be characterised and employed for our purposes.

We assume that each hypothetical network consists of at most N lines, each of length ℓ , arranged over 418 grid cells, each with 8 neighbors. B total buses are distributed across these lines. The total number of potential networks $\bar{n}$ is then at most equal to

$$\underbrace{(418 * 8^{\ell-1})^N}_{\text{line designs}} \times \underbrace{\frac{(B + N - 1)!}{B!(N - 1)!}}_{\text{bus allocations}} \quad (22)$$

where the second term is just the number of combinatorial ways to allocate B indistinguishable balls into N distinguishable urns.

We are interested in

$$\begin{aligned} \log \bar{n} &= N \log 418 + N(\ell - 1) \log 8 + \log((B + N - 1)!) - \log(B!) - \log((N - 1)!) \\ &\approx N \log 418 + N(\ell - 1) \log 8 + (B + N - 1) \log(B + N - 1) \\ &\quad - B \log(B) - (N - 1) \log(N - 1) \end{aligned}$$

where we have used Ramanujan's formula that $\log(x!) \approx x \log x$.

In our simulations, we assume that there can be at most $N = 200$ networks of maximum length of $\ell = 60$ stops and a total of $B = 1500$ buses. Plugging these in, we arrive at an upper bound of $\log \bar{n} \approx 26,258$.

Overall, the upper bound from Proposition 3 is:

$$\frac{\log \bar{n}}{\beta} \approx \frac{26,258}{4.0 \times 10^5} \approx 0.0656.$$

A.7.8 A Simple Model of Optimal Wait Times

In this section, we describe a simple model of optimal wait times when the planner allocates buses between two partly overlapping routes to different areas of the city.

The city center (location 0) is connected by two bus routes to two locations $i = 1$ (near center) and $i = 2$ (periphery), with bus travel times $T_1 < T_2$ and number of commuters $V_1 > V_2$. (We assume no commuters travel between 1 and 2.)

The bus route to location 2 passes through location 1, so both bus routes connect the city center to location 1.

The planner allocates B_i buses to route i to maximize aggregate ridership, subject to a budget constraint $B_1 + B_2 = B$. Bus arrival rates between the city center and the two locations are

$$\lambda_1 = \frac{B_1}{T_1} + \frac{B_2}{T_2},$$

$$\lambda_2 = \frac{B_2}{T_2}.$$

The expected wait time to location i is $W_i = 1/\lambda_i$, which is convex in B_i .

We assume that the share of commuter who take the bus is linear in wait time, $\pi(W_i) = \pi_0 - \pi_1 W_i$ with $\pi_0 < 1$ and $\pi_1 > 0$. The key assumption is that the share is bounded even when wait times are zero, which is also a feature of our full demand model. This means that in the limit, very small wait times are not useful. Their benefit is bounded, and they are “expensive” to achieve because wait time is inversely proportional to B_i .

The planner’s problem is hence

$$\max_{B_1, B_2} V_1 \pi(W_1) + V_2 \pi(W_2) \quad \text{s.t. } B_1 + B_2 = B$$

With a linear demand function π , the first order conditions imply that at an interior optimum, the ratio of wait times between the city center and the two destinations is

$$\frac{W_2^*}{W_1^*} = \sqrt{\frac{V_1 \cdot \left(\frac{1}{T_1} - \frac{1}{T_2}\right)}{V_2 \cdot \frac{1}{T_2}}} = \sqrt{\frac{V_1}{V_2} \left(\frac{T_2}{T_1} - 1\right)}.$$

Wait times to reach the periphery are relatively higher when the ratio of commuters V_1/V_2 is higher. The ratio of travel times may attenuate this effect when $T_2 < 2T_1$. Intuitively, buses on both routes decrease wait times to get to location 1, although buses allocated to route $i = 2$ are less efficient because they spend more time traveling beyond location 1. Hence, the relative marginal benefit of adding a bus to route 1 is proportional to $\frac{1}{T_1} - \frac{1}{T_2}$, which can be smaller than $\frac{1}{T_2}$ when the time to the periphery is not much higher than to location 1. The square root moderates both of these forces, reflecting the fact that it is increasingly “expensive” to reduce wait times that are already small.

This simple model shows how several core features of demand and geography interact to determine the optimal degree of dispersion in wait times in the bus network.

A.8 Impact on Air Pollution and Congestion

In this section we estimate the impact of TransJakarta network improvements on air pollution and road traffic congestion using back-of-the-envelope calculations.

A.8.1 Impact on Road Traffic Congestion

Consider the impact of 1% of trips diverted from other travel modes to TransJakarta buses. We will estimate the reduction in road traffic volume, and then use a congestion elasticity to convert the change in volume to a change in travel time. For context, recall that our main simulations, optimal networks increase ridership by 1.16 percentage points (Table 5).

Reduction in Road Traffic Volume. The change in traffic volume depends on what modes the new TransJakarta passengers are substituting away from, and the relative volumes of those modes.⁷⁵

In our benchmark specification, we assume that a 1% shift toward TransJakarta from other modes lower road traffic volumes by 1%. We obtain this number by measuring initial modal shares and assuming that substitution is proportional to initial shares. The 2018 JUTPI household survey measured modal shares in the larger JaBoDeTaBek area, home to around 30 million people (JUTPI2, 2019). The modal shares were 14.5% car, 78.9% motorcycle or motorcycle taxi, 1.35% TransJakarta, 2.9% other buses, and 1.7% train. We assume a volume of 1 for motorcycles, 2.5 for cars, and zero for other buses and trains, which means that we are assuming that there is no reduction in road traffic volume from train and other bus switchers. This leads to a reduction of road traffic volume of

$$1.15\% = 14.5 \times 2.5\% + 78.9 \times 1\% + (2.9 + 1.7) \times 0\%$$

We obtain a 1% reduction in road traffic volume by rounding down to obtain a more conservative number.

Congestion Elasticity. (Gaduh et al., 2022) estimate the elasticity of road travel time to road lanes, and obtain a value of 0.29.⁷⁶ We set the elasticity of road travel time to traffic volume to be the same. Akbar and Duranton (2017) and Kreindler (2024) estimate this elasticity for Bogota, Columbia, and Bangalore, India, and find maximum elasticities of

⁷⁵For this calculation, we ignore the equilibrium aspect of road traffic congestion. Due to lighter congestion, some commuters would switch away from TransJakarta.

⁷⁶This is the parameter δ_1 in their notation, given by $\delta_1 = \lambda\theta = 0.055 \times 5.36$.

around 0.20.

Using the number for Jakarta, this means that a 1% bus modal share increase increases travel times by 0.3%. We assume that an average trip takes 30 minutes at baseline and that all non-train modes benefit from the lower congestion. The total daily travel time savings due to lower congestion are

$$58,820 \text{ hours} = (39,891,951 \text{ trips}) \times (0.983 \text{ share of trips using roads}) \times \\ (0.5 \text{ hours trips duration}) \times (0.003 \text{ travel time reduction})$$

One way to benchmark these results is using the equivalent variation benefits from optimal networks cited in section 5 of 11.2 minutes of bus travel time per initial bus trip. This adds up to

$$139,994 \text{ hours} = (39,891,951 \text{ trips}) \times (0.019 \text{ initial TransJakarta share}) \times (11.2/60 \text{ EV gain})$$

The benefit due to lower traffic congestion (58,820 hours of time saved) is approximately 42% of the (direct) equivalent variation benefit from re-organizing the bus network (equivalent to 139,994 hours of time on the bus saved). This calculation assumes that time on the bus and time in traffic are valued similarly.

At the end of this section, we return to these estimates and monetize them using a back-of-the-envelope calibration of the value of time at 50% of the wage in Jakarta.

Limitations. Our simple estimates aim to give a rough sense of possible impacts on congestion but have several limitations. First, we do not have the data to estimate the substitution between buses and other modes. We have also assumed, based on the null effect of TransJakarta expansion on total trips, that all new bus trips replace other modes. If, in fact, some of the bus trips are new, this would diminish the congestion effects. We used an overall elasticity of travel times with respect to congestion, thus ignoring the spatial heterogeneity in bus ridership and traffic congestion. It is possible that bus users select from routes with high marginal congestion as in [Anderson \(2014\)](#), or vice-versa. Some commuters might value time on congested conditions higher than the value of time on the bus, leading to higher values than we find here.

A.8.2 Impact on Air Pollution

Our main approach is to use estimates from [Gendron-Carrier et al. \(2022\)](#), who quantify the impact of the opening of subway systems on air pollution. We begin with a short

description of air pollution and emissions patterns in Jakarta.

Vehicle emissions in Jakarta. (Mahalana et al., 2022) measured real-world emissions for over 93,000 vehicles in Jakarta at 18 sites using devices installed at the side of the road, and matching readings with license plate information and vehicle registration information.

Table A.17 below lists median emissions by vehicle type for three different pollutants, expressed in grams of pollutant per kilogram of fuel. TransJakarta buses, which are certified to the Euro II standard, are less polluting than other buses for all pollutant types. Private vehicles emit fewer nitrogen oxides (NO_x) than TransJakarta buses per kilogram of fuel, but more than twice the carbon monoxide (CO) and three times the hydrocarbons (HC). While motorcycles account for more than the majority of vehicles in Jakarta, this study has a small sample of motorcycle because the measurement devices were located near toll booths, where motorcycles are prohibited. Based on 135 motorcycle readings in the data, the report finds suggestive evidence that motorcycles pollute significantly more than private vehicles and than buses per kg of fuel. These figures do not adjust for the smaller fuel consumption per kilometer for private vehicles and motorcycles, nor for the higher occupancy of buses.

Overall, these results suggest that shifting commuters in Jakarta towards using TransJakarta buses is likely to reduce emissions and pollution.

Table A.17: Median emissions by vehicle category (in g/kg fuel).

	Nitrogen Oxides (NO _x)	Carbon Monoxide (CO)	Hydrocarbon (HC)	Sample Size
TransJakarta buses	24	16	3	789
Other buses	56	24	6	1,504
Private vehicles	5	34	9	77,989
Motorcycles	45	156	67	135

Note: The source is Mahalana et al. (2022). The numbers in this table are estimated based on the bar charts from Figures 7, 14, and 15.

The impact of public transport on air pollution. Gendron-Carrier et al. (2022) measure air pollution using satellite measurements of Aerosol Optical Depth (AOD). They show that a new subway system lowers AOD by -0.0182 in a sample of 42 cities where they have ridership data (Table 8). The same event increases ridership per 1,000 people in the city by 36.8857. The implied effect of public transport ridership is that each subway passenger per 1,000 population lowers AOD by 4.934×10^{-4} .

In Jakarta, the population of our study area is 18,384,186, and the weekday total number of trips is 39,891,951. Consider a 1% increase in the mode share of TransJakarta. The increase in ridership per 1,000 inhabitants is

$$21.70 = \frac{1\% \text{ bus mode share} \times 39,891,951 \text{ trips}}{18,384,186 \text{ inhabitants}} \times 1,000$$

The implied effect on AOD is

$$21.70 \times -4.934 \times 10^{-4} = -0.0107$$

Gendron-Carrier et al. (2022) estimate that one unit of AOD is equivalent to $114.6\mu\text{g}/\text{m}^3$ of PM10 and to $60.6\mu\text{g}/\text{m}^3$ of PM2.5 (columns 2 and 8, Appendix Table A.2). This means that the -0.0107 AOD effect above is equivalent to a reduction in PM10 of $1.227\mu\text{g}/\text{m}^3$ and a reduction of $0.648\mu\text{g}/\text{m}^3$ in PM2.5. The mean annual concentration of PM2.5 in the Jakarta area in 2021 was around $30\mu\text{g}/\text{m}^3$, so this represents a 2.16% decrease in air pollution.

Limitations. The ridership impact in Gendron-Carrier et al. (2022) is measured in an 18 month period after the opening of the system, while the level of ridership after 5 years is roughly double. If we interpret the ridership effects implied by our model as the long-run effects, we should use this longer-term ridership numbers and hence divide all our estimates by a factor of approximately 2. By using the estimates from this paper, we are also assuming that subway and bus systems have similar effects on private mode usage and hence air pollution. Finally, we are using the estimates from a global sample of cities to the case of Jakarta.⁷⁷

A.8.3 Monetizing Congestion and Air Pollution Effects

Direct effects. As derived above, the direct effects of implementing an optimal TransJakarta network is equivalent to saving around 139,994 hours per day (in bus travel time) in the entire study area, or 36,398,440 hours per year assuming 260 workdays per year.

We use an average wage of 3.5 USD per hour in Jakarta and assume that the value of time on the bus is 50% of the wage, leading to 63.7 million USD per year in direct gains.

Congestion effects. The reduction in congestion calculated above (58,820 hours of time in traffic per day), valued the same way as time on the bus, leads to estimated gains of 26.7

⁷⁷Gendron-Carrier et al. (2022) show that the air pollution effects are larger in cities with high pollution levels at baseline, such as Jakarta.

million USD per year.

Air pollution effects. The monetary value of health savings from air pollution reductions are more difficult to estimate. We follow [Gendron-Carrier et al. \(2022\)](#), who use two methodologies. The first focuses only on benefits from lower infant mortality, while the second uses the WHO Global Burden of Disease methodology to capture reductions in mortality and morbidity throughout the life-cycle.

We replicate here the first methodology, and note that the gains based on the second methodology are around 20 times larger. We use the estimate in [Gendron-Carrier et al. \(2022\)](#) that $1\mu\text{g}/\text{m}^3$ PM10 leads to 10 infant deaths averted per 100,000 births, based on their literature review of the causal impact of air pollution on infant mortality. The yearly birth rate for Indonesia is 17 per 1,000 people, which implies 312,531 births per year in our study area. This means that the number of infant deaths averted by an optimal TransJakarta network is

$$38.34 = (1.227\mu\text{g}/\text{m}^3 \text{ of } PM10) \times \frac{10}{100,000} \times (312,531 \text{ births per year})$$

We use a Value of Statistical Life (VSL) of 6 million USD in the United States with a 0.6 elasticity of VSL with respect to country GDP. This leads to a value of 2.57 million USD per life saved in Jakarta.⁷⁸ In total, this means gains of $98.41 = 38.34 \times 2.57$ million USD per year.

⁷⁸This is based on a GDP per capita of 18k USD for the Jakarta area in 2022, relative to 76k USD for the United States.