Online Appendix

Gaduh, A., Gračner, T. and Rothenberg, A. (2021): "Life in the Slow Lane: Unintended Consequences of Public Transit in Jakarta"

Table of Contents

List of Tables

List of Figures

A Additional Tables and Figures

A.1 Summary Statistics

Table A.1: Summary Statistics on Trip Data for BRT Routes: Pre-BRT (2002)

	Corridor 1 Trips		Other Corridor Trips	
	Mean (sd) (1)	N (2)	Δ Mean (3)	N (4)
Log Number of Commuters (2002)	7.35 (2.88)	237	1.24***	4514
Distance (km)	2.46 (2.27)	237	-2.94***	4514
Delay (min/km)	13.99 (6.32)	235	3.70***	4501
Delay (min/km, residuals)	3.56 (6.16)	235	3.61***	4501
Travel Time (min)	28.23 (15.67)	237	-16.74***	4514
Travel Time (min, residuals)	-5.13 (15.21)	237	-15.35***	4514

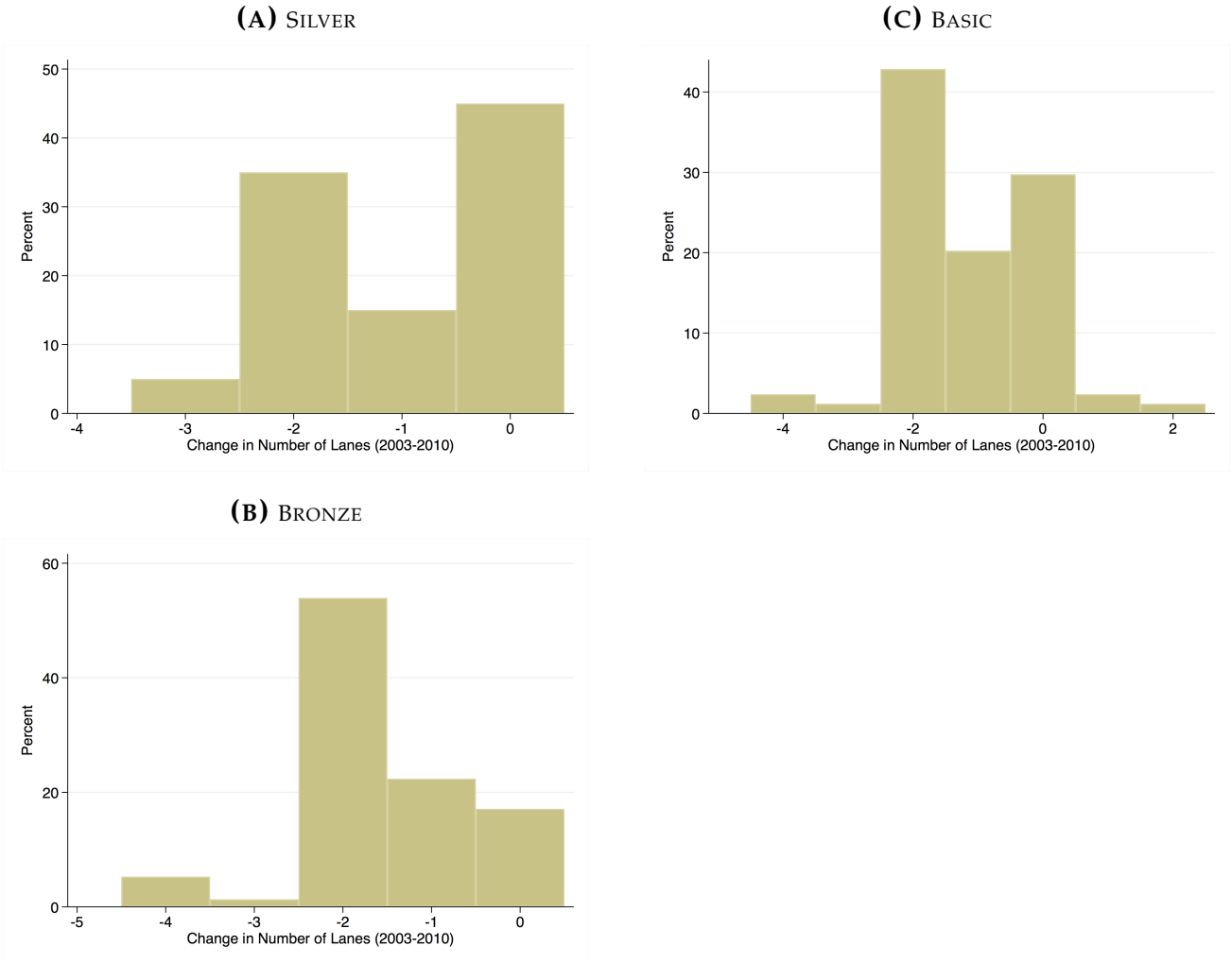
Notes: Authors' calculations. Each observation is an origin-destination pair (at the *kelurahan* level). Column 1 and 2 report the mean, standard deviation (in parentheses), and number of observations of the variable on the left-hand side for routes that originate and terminate within 1 km of a Corridor 1 station in 2010. Columns 3 (4) report the difference in means (number of observations) between routes that originate and terminate within 1 km of another BRT station in 2010. The "residual" variables were constructed after partialling out separate indicators for mode choice, trip purpose, and departure hours. The significance stars in this table are computed by regressing the outcome variable on a treatment indicator. In these regressions, we cluster standard errors at the sub-district (*kecamatan*) level, and significance levels are from the p-values of these treatment indicators. */**/*** denotes significance at the 10% / 5% / 1% levels.

Table A.2: Summary Statistics on BRT Communities: Pre-Treatment Characteristics

	d (CORRIDOR 1) ≤ 1		d (OTHER LINE) ≤ 1	
PANEL A: CENSUS 2000	MEAN (SD) (1)	N (2)	Δ MEAN (3)	N (4)
LOG POPULATION DENSITY	9.75 (0.92)	29	-0.48*	111
AVERAGE YEARS OF SCHOOLING	8.38 (0.78)	29	0.28	111
% OF RECENT MIGRANTS FROM A DIFF. DISTRICT	12.99 (6.28)	29	3.54	111
% OF RECENT MIGRANTS FROM A DIFF. PROVINCE	11.11 (5.56)	29	3.38*	111
PANEL B: JICA 2002				
MONTHLY INCOME < RP 1 MIL	0.44 (0.19)	26	0.05	106
MONTHLY INCOME RP. 1-5 MIL	0.50 (0.14)	26	-0.05	106
MONTHLY INCOME > RP 5 MIL	0.05 (0.09)	26	0.00	106
OWN A CAR (0 1)?	0.25 (0.21)	26	0.00	106
OWN A MOTORCYCLE? (0 1)	0.36 (0.15)	26	-0.02	106
MAIN MODE: TRAIN	0.03 (0.04)	26	0.00	106
MAIN MODE: OTHER PUBLIC TRANSPORT	0.48 (0.16)	26	-0.04	106
MAIN MODE: TAXI / OJEK / BAJAJ	0.07 (0.07)	26	0.01	106
MAIN MODE: CAR	0.21 (0.22)	26	0.03	106
MAIN MODE: MOTORCYCLE	0.22 (0.13)	26	-0.00	106
MAIN MODE: NON-MOTORIZED TRANSIT	0.00 (0.00)	26	-0.00**	106
PANEL C: GIS VARIABLES				
AREA	-0.51 (0.63)	29	-0.38**	111
LOG DIST. TO CITY CENTER	8.17 (1.09)	29	-0.82*	111
ELEVATION	10.05 (7.03)	29	-3.73	111
RUGGEDNESS	0.17 (0.13)	29	0.02	111
PANEL D: PRE-TRENDS (PODES, NIGHT LIGHTS)				
Δ % OF HH W/ A CAR (1996-2000)	-0.15 (0.17)	28	-0.12***	106
Δ % OF HH W/ A MOTORCYLE (1996-2000)	-0.16 (0.26)	28	-0.15*	107
Δ NIGHT LIGHT INTENSITY (1998-2003)	0.00 (0.00)	29	0.31***	111

Notes: Authors' calculations. Each observation is a community (*kelurahan*). Column 1 and 2 report the mean, standard deviation (in parentheses), and number of observations of the variable on the left-hand side for communities that are within 1 km of a Corridor 1 station in 2010. Columns 3 (4) report the difference in means (number of observations) between the communities within 1 km of a Corridor 1 station and communities within 1 km of another BRT station in 2010. The significance stars in this table are computed by regressing the outcome variable on a treatment indicator. In these regressions, we cluster standard errors at the sub-district (*kecamatan*) level, and significance levels are from the p-values of these treatment indicators. */**/*** denotes significance at the 10% / 5% / 1% levels.

Figure A.1: Histograms of Changes in Number of Road Lanes, by BRT Standard (2003-2010)



Notes: This figure provides histograms of changes in the number of lanes of traffic on roads parallel to BRT corridors and stations from 2003 to 2010. To construct this figure, we used Google Earth Pro's satellite imagery, provided by Digital Globe, to visually inspect images of BRT station locations in 2003 (before those stations were constructed) and in 2010 (after those stations had become operational). For each image (see the example in Figure A.2), we counted the number of lanes of traffic in both directions. We then plotted histograms of the change in the number of lanes across BRT corridors of different standards.

Figure A.2: Example of Changes in Number of Road Lanes: Harmoni Sentral Station

(A) 2003

(B) 2010

Notes: This figure provides an example of changes in the number of road lanes from Google Earth Pro's satellite imagery, provided by Digital Globe. The two images depict the area around Harmoni Sentral Station (6.1656°S, 106.8203°E) in 2003 and 2010. Although the resolution of the 2003 image is somewhat coarser than the 2010 image, it shows that there were 10 lanes (5 in each direction) in 2003 and only 8 lanes (4 in each direction) in 2010. We sometimes used nearby imagery along the same roads as the stations to calculate these changes.

A.2 ATT Effects of Station Proximity: Vehicle Ownership and Mode Choice

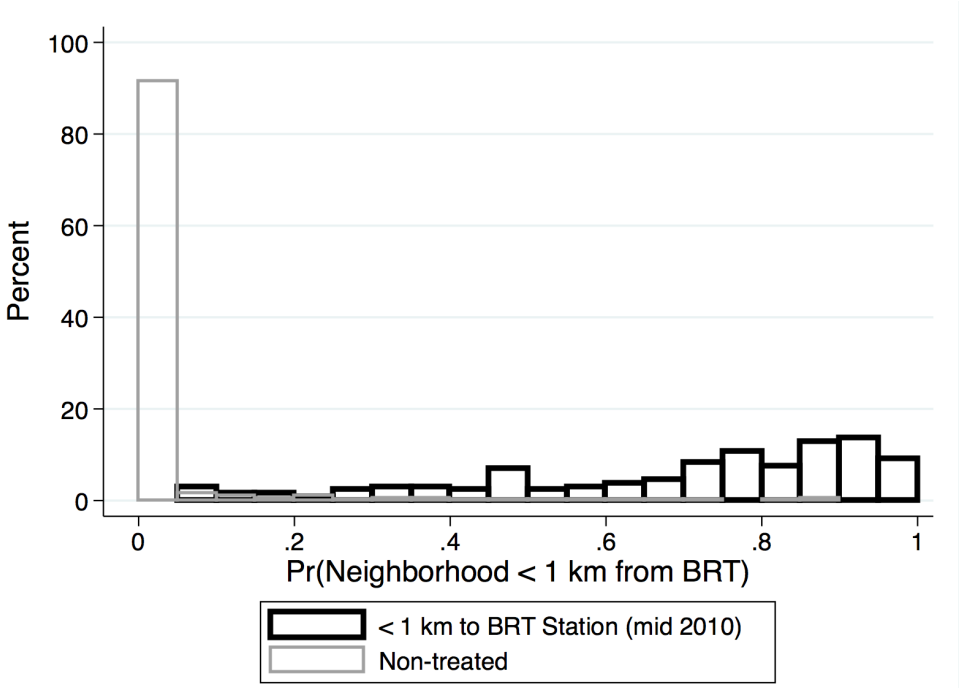
Table A.3: Neighborhood Propensity Score

	Treated vs. All Non-Treated	Treated vs. Planned + Eventual
	(1)	(2)
Population Density (2000)	-0.003	-0.043
	(0.009)	(0.034)
Share of 5-Year District Migrants (2000)	-0.004***	-0.015*
	(0.001)	(0.009)
Monthly Income < Rp 1 mil (%, 2002)	0.083	0.441*
	(0.057)	(0.252)
Monthly Income > Rp 5 mil (%, 2002)	0.041	0.443
	(0.083)	(0.468)
No Primary School Share (2002)	-0.007***	-0.011
	(0.002)	(0.014)
College Completion Share (2002)	0.001	0.006
	(0.002)	(0.010)
Share of Commuting Trips to/from DKI Jakarta	-0.036**	-0.155*
	(0.017)	(0.089)
Log Dist. to City Center	0.105	1.137*
	(0.120)	(0.595)
Log Dist. to City Center (Squared)	-0.007	-0.077**
	(0.007)	(0.038)
Elevation	0.001***	0.011***
	(0.000)	(0.004)
Night Light Intensity (1992)	0.038***	0.092
	(0.009)	(0.061)
Δ Night Light Intensity (1992-2002)	0.029***	0.045
	(0.009)	(0.066)
Ruggedness	0.021	0.302
	(0.028)	(0.202)
Area	-0.016*	-0.035
	(0.008)	(0.044)
N	1487	241
Pseudo R^2	0.718	0.460
Log Likelihood	-125.8	-89.6
LR χ^2	74.6	61.1

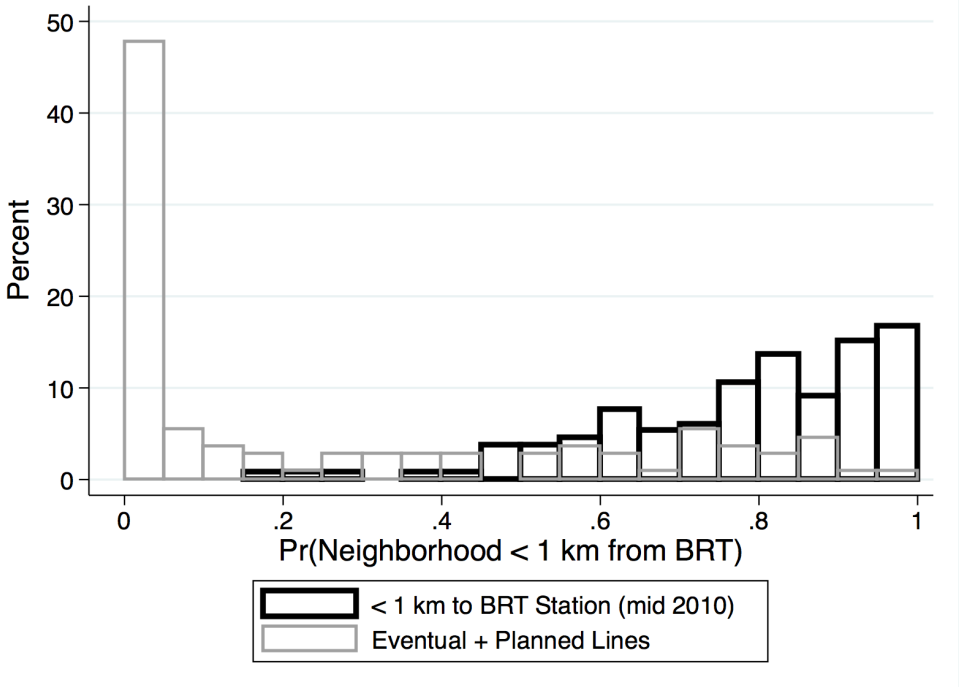
Notes: This table reports marginal effects of our estimated propensity score model. This propensity score is estimated with a logit regression, where the dependent variable is an indicator for whether or not community c is within 1 km of the nearest BRT station, and the independent variables are a vector of pre-treatment variables, x_c. Column 1 includes results comparing treated communities to all non-treated communities, while Column 2 restricts the comparison to only planned or eventually treated neighborhoods. Robust standard errors, clustered at the sub-district (*kecamatan*) level, appear in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Figure A.3: Distribution of Neighborhood Propensity Scores

(A) TREATED VS. NON-TREATED



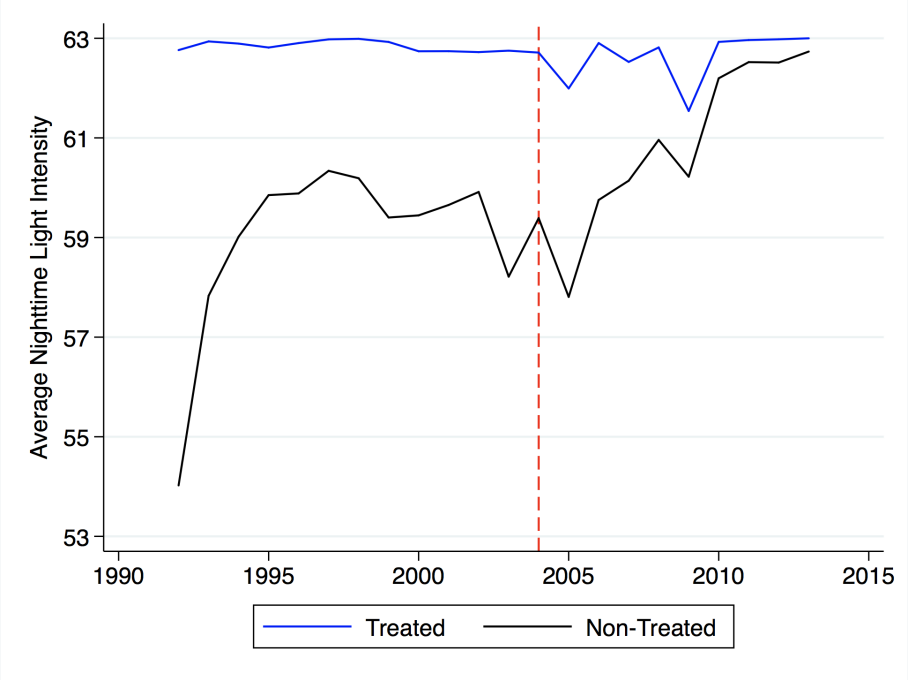
(B) TREATED VS. PLACEBO



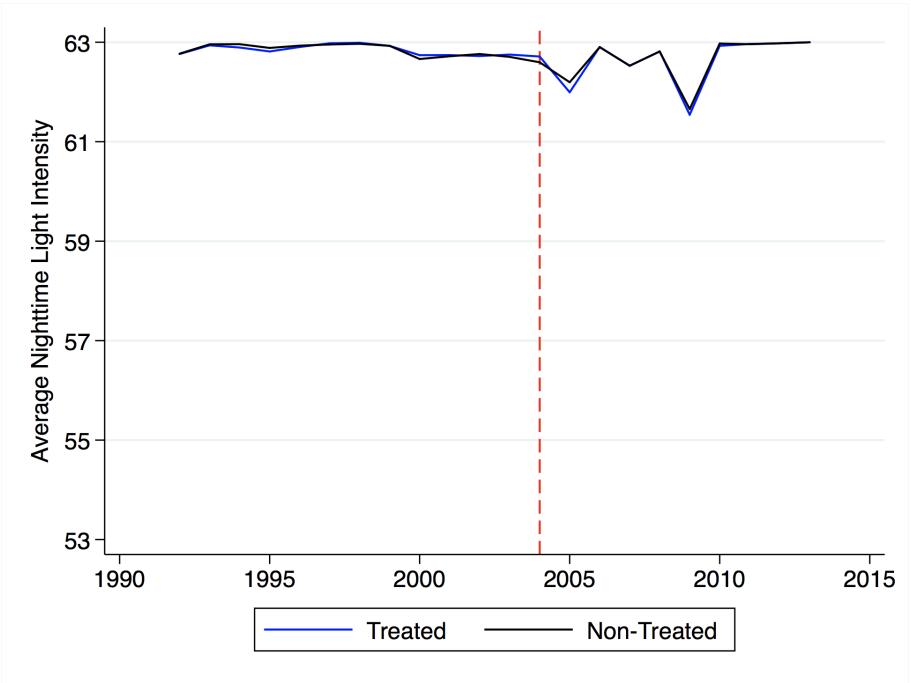
Notes: This figure plots the distribution across communities of the estimated probabilities of being within 1 km of a BRT station, based on the propensity score regressions reported in Appendix Table A.3. Panel A compares propensity scores for kelurahan within 1 km of the closest BRT station to all other communities, while Panel B restricts the comparison to only the planned or eventually treated communities.

Figure A.4: Pre-Trends: Average Nighttime Light Intensity

(A) UNWEIGHTED



(B) WEIGHTED



Notes: This figure reports time-series graphs of average night-time lights intensity for communities that are within 1 km of a BRT station in 2010 ("Treated", in blue) and communities within 1 km of either a planned BRT line that has yet to be constructed or a planned BRT station that was constructed after mid 2010 ("Planned + Eventually Treated", in black). Panel A reports unweighted averages, while Panel B uses weights generated by a first-step propensity score estimation.

Continuous Treatment. Appendix Table A.4 reports results from linear regressions of changes in vehicle ownership and commuting modes on a continuous measure of distance to stations. Column 1 reports OLS estimates, while in Columns 2-4, we use planned and eventual station distances as instrumental variables for actual station distances. All regressions include the same set of pre-determined controls as in (1). These results largely support our main findings that BRT-station proximity neither curbed vehicle ownership nor increased public transit use.

Table A.4: Effects of Continuous BRT Station Distance

	OLS	IV (2SLS)		
	(1)	(2)	(3)	(4)
Δ Share Owning Car	0.000	-0.000	0.381	-0.000
	(0.002)	(0.002)	(4.058)	(0.002)
Δ Share Owning Motorcycle	-0.003	-0.002	0.002	-0.002
	(0.002)	(0.002)	(0.162)	(0.002)
Δ Main Mode Share: BRT	-0.001	-0.001	-0.012	-0.001
	(0.001)	(0.001)	(0.130)	(0.001)
Δ Main or Alternative Mode Share: BRT	-0.001	-0.002	0.038	-0.002
	(0.001)	(0.002)	(0.305)	(0.002)
Δ Main Mode Share: Car	0.002	0.002	0.031	0.002
	(0.001)*	(0.001)*	(0.232)	(0.001)**
Δ Main Mode Share: Motorcycle	-0.000	0.001	0.135	0.001
	(0.003)	(0.003)	(1.125)	(0.003)
Δ Main Mode Share: Train	-0.001	-0.001	0.094	-0.001
	(0.001)	(0.001)	(0.739)	(0.001)
Δ Main Mode Share: Other Public Transport	0.001	0.002	-0.420	0.001
	(0.002)	(0.002)	(3.249)	(0.002)
Δ Main Mode Share: Taxi	0.001	-0.001	0.156	-0.001
	(0.002)	(0.002)	(1.188)	(0.002)
Δ Main Mode Share: Non-Motorized Transit	-0.001	-0.001	0.016	-0.001
	(0.001)	(0.001)	(0.135)	(0.001)
N	696	696	696	696
Controls	X	X	X	X
Planned IV	.	X	.	X
Eventual IV		.	X	X

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column) on a continuous measure of distance to the nearest BRT station. The regression includes all communities within 10 km of the nearest BRT station. Column 1 reports OLS estimates, while in columns 2-4, we use planned and eventual distances to stations as instrumental variables for actual distance to stations. All regressions also include the same set of pre-determined controls as in (1). Robust standard errors, clustered at the sub-district level, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Varying Treatment and Non-Treatment Definitions. The one-kilometer threshold we use to define a community's treatment status is often used in empirical work and consistent with walking behavior of commuters in Jakarta (see Footnote 18). Appendix Table A.5 shows that our mode-choice results are robust to alternative distance thresholds used to define treatment status.

Table A.5: ATT Estimates on Vehicle Ownership and Mode Choice: Varying d

	$d = 0.5$	$d = 1$	$d = 1.5$	$d = 2$	$d = 2.5$	$d = 3$
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Δ Share Owning Car	0.031	-0.035	-0.028	-0.025	-0.052	-0.126
	(0.038)	(0.063)	(0.045)	(0.046)	(0.063)	(0.095)
Δ Share Owning Motorcycle	0.066	0.010	-0.001	0.057	0.187	0.241
	(0.045)	(0.028)	(0.039)	(0.034)*	(0.048)***	(0.065)***
Δ Main Mode Share: BRT	0.029	0.020	-0.017	-0.017	-0.026	-0.008
	(0.037)	(0.019)	(0.025)	(0.031)	(0.037)	(0.035)
Δ Main or Alternative Mode Share: BRT	0.094	0.043	0.019	0.008	-0.082	0.028
	(0.061)	(0.025)*	(0.029)	(0.033)	(0.049)*	(0.052)
Δ Main Mode Share: Car	0.081	-0.011	-0.014	-0.013	-0.128	-0.218
	(0.046)*	(0.027)	(0.037)	(0.035)	(0.074)*	(0.090)**
Δ Main Mode Share: Motorcycle	-0.028	0.027	0.035	0.060	0.093	0.084
	(0.046)	(0.030)	(0.041)	(0.060)	(0.070)	(0.080)
Δ Main Mode Share: Train	0.010	0.004	-0.023	-0.008	0.011	0.057
	(0.032)	(0.018)	(0.029)	(0.021)	(0.041)	(0.049)
Δ Main Mode Share: Other Public Transport	-0.095	-0.041	0.026	-0.007	0.040	0.043
	(0.038)**	(0.029)	(0.028)	(0.030)	(0.048)	(0.066)
Δ Main Mode Share: Taxi	0.000	-0.003	-0.012	-0.022	-0.017	0.017
	(0.013)	(0.007)	(0.008)	(0.010)**	(0.011)	(0.023)
Δ Main Mode Share: Non-Motorized Transit	0.002	0.002	0.004	0.007	0.028	0.025
	(0.007)	(0.004)	(0.007)	(0.008)	(0.011)**	(0.020)
Controls	X	X	X	X	X	X
Oaxaca-Blinder	X	X	X	X	X	X

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column) on an indicator for whether or not the community is within d km of a BRT station, where d is listed in the column header, ranging from $d = 0.5$ in column 1 to $d = 3$ in Column 6. All columns report results of a control function specification based on a Oaxaca-Blinder decomposition, described in Kline (2011), analogous to the estimates in Column 4 of Table 5. Robust standard errors, clustered at the sub-district level, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Dropping Close Non-Treated Neighborhoods. Alternatively, spillovers of the treatment effect between closely located treated and non-treated neighborhoods could be producing our null results. Appendix Table A.6 shows that our main results are mostly unchanged when we drop neighborhoods that are greater than 1 km but less than 2 km away from BRT stations in 2010.

Table A.6: ATT Estimates on Vehicle Ownership and Mode Choice: Dropping $d \in (1, 2]$

	Treated vs Planned + Eventual			
	(1)	(2)	(3)	(4)
Δ Share Owning Car	-0.003	0.009	-0.001	-0.019
	(0.028)	(0.042)	(0.036)	(0.070)
Δ Share Owning Motorcycle	0.014	0.038	-0.012	0.055
	(0.018)	(0.032)	(0.025)	(0.036)
Δ Main Mode Share: BRT	0.033**	-0.011	-0.005	-0.019
	(0.013)	(0.031)	(0.030)	(0.030)
Δ Main or Alternative Mode Share: BRT	0.087***	0.009	0.003	0.031
	(0.019)	(0.036)	(0.039)	(0.035)
Δ Main Mode Share: Car	0.004	0.047	-0.016	-0.025
	(0.022)	(0.033)	(0.022)	(0.044)
Δ Main Mode Share: Motorcycle	-0.060**	0.004	0.043	0.083
	(0.024)	(0.049)	(0.050)	(0.067)
Δ Main Mode Share: Train	0.013	-0.013	-0.012	0.000
	(0.011)	(0.022)	(0.031)	(0.022)
Δ Main Mode Share: Other Public Transport	0.022	-0.024	-0.009	-0.018
	(0.025)	(0.039)	(0.030)	(0.041)
Δ Main Mode Share: Taxi	-0.010	-0.002	-0.003	-0.025**
	(0.007)	(0.009)	(0.009)	(0.012)
Δ Main Mode Share: Non-Motorized Transit	-0.001	-0.000	0.002	0.004
	(0.004)	(0.007)	(0.005)	(0.012)
Controls	.	X	X	X
Logistic Reweighting	.	.	X	.
Oaxaca-Blinder	.	.	.	X

Notes: This table is identical to Table 5 except that we drop all communities that were greater than 1 km but less than 2 km from the closest BRT station. Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column) on an indicator for whether or not the community is within 2 km of a BRT station. Column 2 includes pre-treatment controls, and Column 3 reports a double-robust specification that both includes controls and reweights non-treated communities by $\hat{\kappa} = \hat{P}/(1-\hat{P})$, where $\hat{P}$ is the estimated probability that the community is within 1 km of a BRT station. Columns 4 reports a control function specification based on a Oaxaca-Blinder decomposition, described in Kline (2011). Robust standard errors, clustered at the sub-district level, are reported in parentheses and are estimated using a bootstrap procedure, with 1000 replications, in Column 4 to account for the generated $\hat{\kappa}$ weights. */**/*** denotes significance at the 10% / 5% / 1% levels.

Distance to Destination Stations. Instead of focusing on origin communities as treated, in Appendix Table A.7, we also assessed whether vehicle ownership and mode choice outcomes are impacted by the proximity of BRT stations to commuters' destination communities. To do so, we first assigned individuals in the JICA data to their destination communities based on the trip data, instead of using their origin locations, and we re-estimated the semi-parametric difference-in-differences model using revised treatment status definitions. Overall, we find similar null results as in our main regression estimates.

Table A.7: ATT Estimates of BRT Station Proximity to *Destinations*: Vehicle Ownership and Mode Choice

	Treated vs. Planned			
	(1)	(2)	(3)	(4)
Δ Share Owning Car	0.018	0.035*	0.014	0.041*
	(0.016)	(0.020)	(0.016)	(0.022)
Δ Share Owning Motorcycle	0.005	0.005	-0.009	0.003
	(0.014)	(0.014)	(0.014)	(0.017)
Δ Main Mode Share: BRT	0.018**	0.011	0.003	0.019
	(0.007)	(0.009)	(0.015)	(0.012)
Δ Main or Alternative Mode Share: BRT	0.043***	0.016	0.005	0.031*
	(0.011)	(0.012)	(0.016)	(0.016)
Δ Main Mode Share: Car	0.015	0.028**	0.003	0.037**
	(0.012)	(0.012)	(0.014)	(0.017)
Δ Main Mode Share: Motorcycle	-0.036**	-0.001	0.018	0.004
	(0.015)	(0.017)	(0.024)	(0.017)
Δ Main Mode Share: Train	0.006	-0.001	-0.010	0.000
	(0.006)	(0.008)	(0.012)	(0.007)
Δ Main Mode Share: Other Public Transport	0.003	-0.037**	-0.014	-0.061***
	(0.014)	(0.015)	(0.015)	(0.021)
Δ Main Mode Share: Taxi	-0.001	-0.001	-0.001	-0.001
	(0.004)	(0.004)	(0.006)	(0.005)
Δ Main Mode Share: Non-Motorized Transit	-0.004	0.001	0.000	0.002
	(0.003)	(0.003)	(0.005)	(0.004)
N	241	241	241	241
Controls	.	X	X	X
Logistic Reweighting	.	.	X	.
Oaxaca-Blinder	.	.	.	X

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column) on an indicator for whether or not the *destination* community is within 1 km of a BRT station. To construct these outcomes, we first assigned individuals in the JICA data to their destination locations for trips based on the trip data, instead of using their origin locations. Columns 1-4 restrict the non-treated sample to include only almost-treated destination communities. Column 2 includes pre-treatment controls, and Columns 3 reports a double-robust specification that both includes controls and reweights almost-treated destination communities by $\hat{\kappa} = \hat{P}/(1-\hat{P})$, where $\hat{P}$ is the estimated probability that the community is within 1 km of a BRT station. Columns 4 reports a control function specification based on a Oaxaca-Blinder decomposition, described in Kline (2011). Robust standard errors, clustered at the sub-district level, are reported in parentheses and are estimated using a bootstrap procedure, with 1000 replications, in column 3 to account for the generated $\hat{\kappa}$ weights. Sample sizes vary slightly across outcomes but include as many 132 "treated" communities and 109 "almost-treated" communities. */**/*** denotes significance at the 10% / 5% / 1% levels.

Alternative Propensity Score Controls. We next explore the robustness of our estimates to different choices for pre-determined controls used to estimate the propensity score ($\mathbf{x}_c$). Appendix Table A.8 reports how our original Oaxaca-Blinder estimates (Column 4 of Table 5, reported in Column 1) change when we use only geographic controls (Column 2), only demographic controls (Column 3), or the full set of pre-determined covariates (Column 4).[49] Overall, our main results remain robust to these different specifications.

Table A.8: ATT Estimates: Robustness to Propensity Score Specifications

	(1)	(2)	(3)	(4)
Δ Share Owning Car	-0.035	-0.026	-0.016	-0.048
	(0.063)	(0.059)	(0.038)	(0.066)
Δ Share Owning Motorcycle	0.010	-0.003	-0.007	-0.004
	(0.028)	(0.021)	(0.022)	(0.018)
Δ Main Mode Share: BRT	0.020	0.011	0.026	0.026
	(0.019)	(0.016)	(0.019)	(0.020)
Δ Main or Alternative Mode Share: BRT	0.043*	0.038*	0.058**	0.051*
	(0.025)	(0.020)	(0.027)	(0.028)
Δ Main Mode Share: Car	-0.011	-0.014	-0.016	-0.021
	(0.027)	(0.027)	(0.026)	(0.025)
Δ Main Mode Share: Motorcycle	0.027	0.022	-0.021	0.005
	(0.030)	(0.024)	(0.030)	(0.029)
Δ Main Mode Share: Train	0.004	-0.005	0.013	0.014
	(0.018)	(0.014)	(0.017)	(0.018)
Δ Main Mode Share: Other Public Transport	-0.041	-0.017	0.001	-0.026
	(0.029)	(0.024)	(0.025)	(0.025)
Δ Main Mode Share: Taxi	-0.003	0.001	-0.008	0.001
	(0.007)	(0.007)	(0.006)	(0.005)
Δ Main Mode Share: Non-Motorized Transit	0.002	0.002	0.004	0.000
	(0.004)	(0.004)	(0.003)	(0.005)
Oaxaca-Blinder	X	X	X	X
Original Propensity Score $\mathbf{x}$	X	.	.	.
Only Geographic Variables in $\mathbf{x}$	.	X	.	.
Only Demographic Variables in $\mathbf{x}$	.	.	X	.
Full Set of Variables in $\mathbf{x}$	.	.	.	X

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column) on an indicator for whether or not the community is within 1 km of a BRT station. All columns report estimates of a control function specification based on a Oaxaca-Blinder decomposition, described in Kline (2011). Column 1 reports our original estimates (Table 5, Column 4), while Columns 2-4 vary the set of controls used in estimating the propensity score. Robust standard errors, clustered at the sub-district level, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

[49] The original, geographic, and demographic controls are described in Footnote 20. The full set of baseline controls includes baseline measures of vehicle ownership, mode choice shares, incomes, and various light intensity measures.

Machine-Learning Propensity Score Specifications. In Appendix Table A.9, we also implement two different machine learning procedures to select controls for the propensity score and estimate treatment effects. Columns 1 and 2 report our original linear regression with controls and logistic reweighting estimates (Columns 2 and 3 of Table 5). In Column 3, we implement the Belloni et al. (2014) post-double selection estimator, selecting separate controls for the propensity score and the outcome equation by choosing from the large set of controls discussed in Footnote 49 to minimize a Lasso objective function. In Columns 4 and 5, we used generalized boosted regression models to estimate the propensity score, following McCaffrey et al. (2004). Point estimates and confidence intervals under these approaches are very similar to our original results.

Table A.9: ATT Estimates on Vehicle Ownership and Mode Choice: Lasso Robustness

	Original		Lasso	TWANG	
	(1)	(2)	(3)	(4)	(5)
Δ Share Owning Car	0.001	-0.001	0.029	0.035	0.035
	(0.042)	(0.036)	(0.046)	(0.045)	(0.046)
Δ Share Owning Motorcycle	-0.001	-0.012	-0.023	0.003	-0.001
	(0.023)	(0.025)	(0.020)	(0.019)	(0.019)
Δ Main Mode Share: BRT	0.013	-0.005	0.016	0.002	0.007
	(0.020)	(0.029)	(0.019)	(0.027)	(0.025)
Δ Main or Alternative Mode Share: BRT	0.022	0.003	0.029	0.012	0.019
	(0.024)	(0.042)	(0.023)	(0.036)	(0.032)
Δ Main Mode Share: Car	0.006	-0.016	0.004	0.009	0.007
	(0.029)	(0.022)	(0.025)	(0.035)	(0.035)
Δ Main Mode Share: Motorcycle	0.001	0.043	-0.036	0.015	0.003
	(0.034)	(0.051)	(0.029)	(0.039)	(0.036)
Δ Main Mode Share: Train	-0.002	-0.012	0.004	-0.013	-0.006
	(0.016)	(0.033)	(0.015)	(0.028)	(0.024)
Δ Main Mode Share: Other Public Transport	-0.017	-0.009	0.016	-0.010	-0.009
	(0.032)	(0.032)	(0.024)	(0.033)	(0.032)
Δ Main Mode Share: Taxi	-0.004	-0.003	0.001	-0.004	-0.004
	(0.006)	(0.009)	(0.004)	(0.005)	(0.005)
Δ Main Mode Share: Non-Motorized Transit	0.002	0.002	-0.000	0.002	0.001
	(0.003)	(0.005)	(0.004)	(0.005)	(0.005)
Controls	X	X	.	X	X
Logistic Reweighting	.	X	.	.	.
Lasso Double-Selected Controls	.	.	X	.	.
TWANG Reweighting (`es mean`)	.	.	.	X	.
TWANG Reweighting (`ks max`)	.	.	.	.	X

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column) on an indicator for whether or not the community is within 1 km of a BRT station. Columns 1 and 2 report our original linear regression with controls and logistic reweighting estimates (columns 2 and 3 of Table 5). Column 3 provides Belloni et al. (2014) post-double selection treatment effect estimates. We first select separate controls for the propensity score and the outcome equation by choosing from the large set of controls discussed in Footnote 49 to minimize a Lasso objective function. We then regress the outcome on the union of the selected controls. In columns 4 and 5, we used generalized boosted regression models to estimate the propensity score, following McCaffrey et al. (2004). Column 4 uses the average standardized effect size of pre-treatment variables for assessing balance between treated and control groups (`es mean`), while Column 5 uses the maximum Kolmogorov-Smirnov p-values for weighted pre-treatment variables for assessing balance (`ks max`). Robust standard errors, clustered at the sub-district level, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Treatment Effect Heterogeneity. Appendix Table A.10 explores the possibility that the muted effects were driven by competing heterogeneous treatment effects. Separate analyses by gender (Columns 2 and 3), education (Columns 4 and 5), and total monthly household expenditures (Columns 6 and 7) show that our main null results are largely robust to these different splits.[50]

Table A.10: ATT on Vehicle Ownership and Mode Choice: Heterogeneity

		Gender		Education		Monthly HH Expenditure	
	All (1)	Male (2)	Female (3)	Low (4)	High (5)	Low (6)	High (7)
Δ Share Owning Car	-0.035	-0.028	-0.043	0.013	-0.047	-0.040	-0.094
	(0.063)	(0.064)	(0.062)	(0.067)	(0.063)	(0.070)	(0.096)
Δ Share Owning Motorcycle	0.010	0.013	0.007	0.028	0.007	0.041	0.002
	(0.028)	(0.029)	(0.027)	(0.031)	(0.030)	(0.036)	(0.031)
Δ Main Mode Share: BRT	0.020	0.021	0.020	0.024	0.020	0.021	0.020
	(0.019)	(0.019)	(0.019)	(0.018)	(0.021)	(0.022)	(0.019)
Δ Main or Alternative Mode Share: BRT	0.043*	0.048*	0.038	0.020	0.044	0.033	0.057**
	(0.025)	(0.026)	(0.024)	(0.027)	(0.028)	(0.032)	(0.026)
Δ Main Mode Share: Car	-0.011	-0.010	-0.012	0.004	-0.004	-0.003	-0.051
	(0.027)	(0.028)	(0.027)	(0.021)	(0.028)	(0.018)	(0.053)
Δ Main Mode Share: Motorcycle	0.027	0.032	0.024	0.024	0.025	0.037	0.052
	(0.030)	(0.032)	(0.029)	(0.037)	(0.029)	(0.037)	(0.043)
Δ Main Mode Share: Train	0.004	0.005	0.003	-0.002	-0.000	-0.002	0.008
	(0.018)	(0.017)	(0.018)	(0.022)	(0.019)	(0.020)	(0.018)
Δ Main Mode Share: Other Public Transport	-0.041	-0.047	-0.034	-0.053	-0.037	-0.045	-0.026
	(0.029)	(0.029)	(0.029)	(0.042)	(0.029)	(0.038)	(0.027)
Δ Main Mode Share: Taxi	-0.003	-0.002	-0.003	-0.001	-0.005	-0.014	0.002
	(0.007)	(0.006)	(0.009)	(0.009)	(0.007)	(0.023)	(0.018)
Δ Main Mode Share: Non-Motorized Transit	0.002	0.003	0.002	0.003	0.001	0.006	-0.004
	(0.004)	(0.005)	(0.004)	(0.006)	(0.005)	(0.006)	(0.006)
Oaxaca-Blinder	X	X	X	X	X	X	X

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column) on an indicator for whether or not the community is within 1 km of a BRT station. All columns report results of a control function specification based on a Oaxaca-Blinder decomposition, described in Kline (2011). Column 1 reproduces our original results (Column 4 of Table 5). In the remaining columns, the dependent variable is the community-level average of the outcome variable for males (Column 2), females (Column 3), below median years of schooling (Column 4), above median years of schooling (Column 5), below median monthly household expenditure (Column 6), and above median monthly household expenditure (Column 7). Robust standard errors, clustered at the sub-district level, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

[50] We find some suggestive evidence, however, that the main or alternative BRT effect may be coming from males. Witoelar et al. (2017) show that many females are concerned about safety while riding the BRT. TransJakarta has enacted several policies trying to increase female ridership, including providing female-only BRT cars.

Sorting. Sorting may drive our results if, for example, areas with BRT stations attracted migrants who were less likely to demand public transportation. Appendix Table A.11 increasingly adds a series of (bad) demographic controls to our original Oaxaca-Blinder specification (Table 5, Column 4). Column 2 controls for changes in density and migration shares, Column 3 adds changes in education shares, and Column 4 adds changes in monthly household expenditure shares. Overall, the inclusion of these additional controls do not change our main findings that close-proximity BRT communities had little changes in vehicle ownership and positive, though small, changes in BRT mode shares. This is suggestive evidence that associated changes in neighborhood composition cannot explain the lion's share of TransJakarta's muted impacts.[51]

Table A.11: ATT Estimates on Vehicle Ownership and Mode Choice: (Bad) Controls

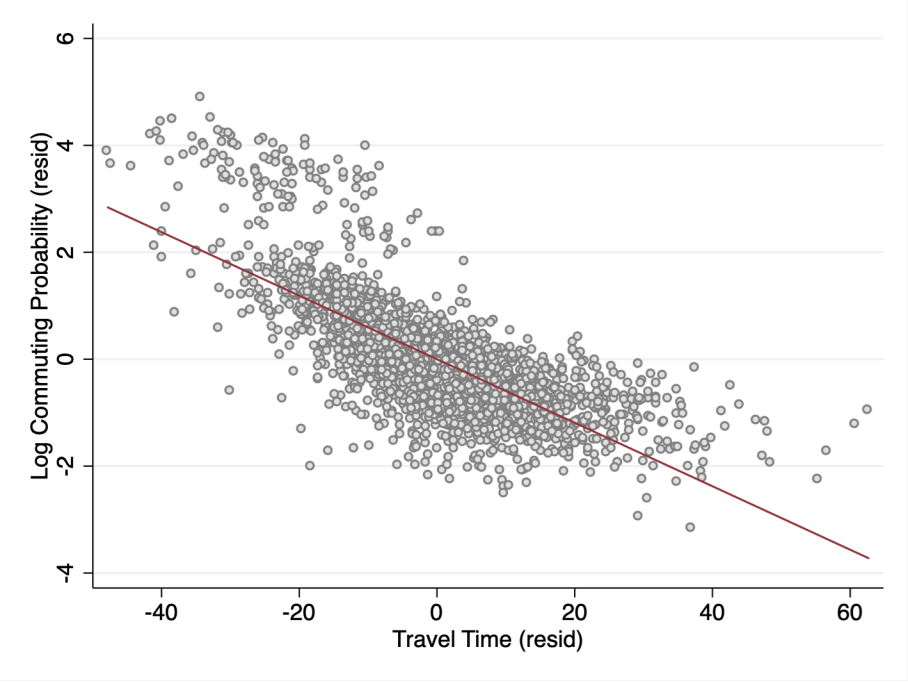
	Treated vs. Planned + Eventual			
	(1)	(2)	(3)	(4)
Δ Share Owning Car	-0.035	-0.032	-0.020	-0.011
	(0.063)	(0.067)	(0.064)	(0.050)
Δ Share Owning Motorcycle	0.010	0.009	0.013	0.023
	(0.028)	(0.028)	(0.027)	(0.025)
Δ Main Mode Share: BRT	0.020	0.019	0.009	0.012
	(0.019)	(0.019)	(0.021)	(0.022)
Δ Main or Alternative Mode Share: BRT	0.043*	0.040	0.024	0.028
	(0.025)	(0.026)	(0.027)	(0.027)
Δ Main Mode Share: Car	-0.011	-0.011	-0.008	-0.002
	(0.027)	(0.029)	(0.029)	(0.023)
Δ Main Mode Share: Motorcycle	0.027	0.028	0.048	0.042
	(0.030)	(0.030)	(0.037)	(0.038)
Δ Main Mode Share: Train	0.004	0.003	-0.005	-0.006
	(0.018)	(0.019)	(0.022)	(0.021)
Δ Main Mode Share: Other Public Transport	-0.041	-0.040	-0.047	-0.045
	(0.029)	(0.029)	(0.029)	(0.030)
Δ Main Mode Share: Taxi	-0.003	-0.002	0.001	-0.005
	(0.007)	(0.007)	(0.007)	(0.006)
Δ Main Mode Share: Non-Motorized Transit	0.002	0.003	0.002	0.003
	(0.004)	(0.005)	(0.006)	(0.006)
Oaxaca-Blinder	X	X	X	X
Controls for Δ Density		X	X	X
Controls for Δ Migrant Share		X	X	X
Controls for Δ Education Shares			X	X
Controls for Δ Monthly HH Expenditure Shares				X

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column) on an indicator for whether or not the community is within 1 km of a BRT station. All columns report results of a control function specification based on a Oaxaca-Blinder decomposition, described in Kline (2011). Column 1 reproduces our original results (Column 4 of Table 5). In Column 2, we add controls for changes in community-level population density and the share of recent province-level and district-level migrants. In Column 3, we add 7 controls for changes in different levels of educational attainment. In Column 4, we add 7 additional controls for changes in the share of households with different monthly household expenditures. Robust standard errors, clustered at the sub-district level, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

[51] If people with strong tastes for public transportation moved into treated areas, this would cause us to overestimate the average impacts of the program in the absence of sorting. The absence of strong program impacts reduces concerns over this type of sorting bias.

A.3 Gravity Commuting Regressions

Figure A.5: Linear Gravity Regression: Residual-on-Residual Plot



Notes: This figure presents a residual-on-residual plot of the regression in Table 6, Panel A, Column 3. The y-axis depicts the log commuting probability residual, after partialling out the origin and destination FE, and the x-axis depicts the travel time residual, after partialling out the origin and destination FE. Each point on the graph represents a bilateral pair of origin and destination sub-districts.

Table A.12: Gravity Commuting Regressions (All Observations)

	Cross-Section				Panel	
	2002		2010			
Panel A: Fixed Effects Least Squares	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Travel Time	-0.036***		-0.037***			
	(0.001)		(0.001)			
BRT Route (OD)		1.685***		1.587***	0.002	0.002
		(0.127)		(0.113)	(0.064)	(0.064)
N	6746	6746	5687	5687	10906	10906
Adjusted R^2	0.608	0.217	0.523	0.179	0.841	0.841
Adjusted R^2 (within)	0.533	0.068	0.451	0.056	0.004	0.003
Panel B: Poisson Pseudo ML w/ FE	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Travel Time	-0.095***		-0.121***			
	(0.003)		(0.004)			
BRT Route (OD)		3.351***		3.332***	0.027	0.010
		(0.162)		(0.145)	(0.065)	(0.061)
N	6746	6746	5687	5687	10906	10906
Wald χ^2	1306.436	429.312	1020.740	525.693	2.863	14.043
Pseudo R^2	0.199	0.038	0.204	0.048	0.230	0.230
Origin FE	Yes	Yes	Yes	Yes	.	.
Destination FE	Yes	Yes	Yes	Yes	.	.
Origin × Destination FE	.	.	.	.	Yes	Yes
Year FE	.	.	.	.	Yes	Yes
Orig. District × Year FE	.	.	.	.	Yes	Yes
Dest. District × Year FE	.	.	.	.	Yes	Yes
O-D Density Controls	.	.	.	.	.	Yes

Notes: This table reports estimates of the gravity commuting regression equation, (3). This table is similar to Table 6, except that it uses all of the data, instead of dropping observations with fewer than 10 commuters. Columns 1-4 report cross-sectional regressions and only include separate fixed effects for each origin and destination location. Columns 5-7 use panel data, with a full set of origin × destination fixed effects and year fixed effects. Column 7 adds origin district × year fixed effects and destination district × year fixed effects. Robust standard errors, two-way clustered at the origin and destination sub-district level, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Table A.13: Gravity Commuting Regressions (Actual vs. Eventually Treated or Planned)

	Cross-Section				Panel	
	2002		2010			
Panel A: Fixed Effects Least Squares	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Travel Time	-0.034***		-0.034***			
	(0.001)		(0.001)			
BRT Route (od)		1.044***		1.369***	0.063	0.063
		(0.152)		(0.169)	(0.064)	(0.064)
N	3316	3316	3053	3053	6108	6108
Adjusted R^2	0.716	0.530	0.596	0.434	0.840	0.840
Adjusted R^2 (within)	0.422	0.043	0.328	0.059	0.003	0.003
Panel B: Poisson Pseudo ML w/ FE	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Travel Time	-0.084***		-0.098***			
	(0.002)		(0.004)			
BRT Route (od)		3.744***		4.278***	0.024	0.021
		(0.421)		(0.450)	(0.064)	(0.063)
N	3316	3316	3053	3053	6108	6108
Wald χ^2	1160.453	79.174	583.371	90.565	1.288	10.955
Pseudo R^2	0.177	0.087	0.180	0.105	0.200	0.200
Origin FE	Yes	Yes	Yes	Yes	.	.
Destination FE	Yes	Yes	Yes	Yes	.	.
Origin × Destination FE	.	.	.	.	Yes	Yes
Year FE	.	.	.	.	Yes	Yes
Orig. District × Year FE	.	.	.	.	Yes	Yes
Dest. District × Year FE	.	.	.	.	Yes	Yes
O-D Density Controls	.	.	.	.	.	Yes

Notes: This table reports estimates of the gravity commuting regression equation, (3). This table is similar to Table 6, except that it only uses bilateral subdistrict location pairs that begin and end within 1 km of actual, eventually constructed, or planned stations. Columns 1-4 report cross-sectional regressions and only include separate fixed effects for each origin and destination location. Columns 5-7 use panel data, with a full set of origin × destination fixed effects and year fixed effects. Column 7 adds origin district × year fixed effects and destination district × year fixed effects. Robust standard errors, two-way clustered at the origin and destination sub-district level, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Table A.14: Gravity Commuting Regressions (Corridor 1 vs. Other Corridors)

	Cross-Section				Panel	
	2002		2010			
Panel A: Fixed Effects Least Squares	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Travel Time	-0.058***		-0.059***			
	(0.002)		(0.002)			
Corridor 1 (od)		2.398***		2.575***	0.009	0.006
		(0.652)		(0.360)	(0.305)	(0.318)
Other Lines (od)		0.614**		0.616**	-0.124	-0.124
		(0.263)		(0.248)	(0.119)	(0.120)
N	2210	2210	2210	2210	4402	4402
Adjusted R^2	0.620	0.030	0.550	0.064	0.880	0.881
Adjusted R^2 (within)	0.610	0.006	0.522	0.006	0.010	0.012
Panel B: Poisson Pseudo ML w/ FE	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Travel Time	-0.107***		-0.133***			
	(0.004)		(0.005)			
Corridor 1 (od)		4.985***		5.684***	0.241	0.188
		(1.236)		(1.521)	(0.285)	(0.246)
Other Lines (od)		1.406***		1.732***	-0.158	-0.108
		(0.407)		(0.406)	(0.122)	(0.113)
N	2210	2210	2210	2210	4402	4402
Wald χ^2	847.069	18.993	672.641	23.990	20.698	33.137
Pseudo R^2	0.148	0.030	0.164	0.047	0.177	0.177
Origin FE	Yes	Yes	Yes	Yes	.	.
Destination FE	Yes	Yes	Yes	Yes	.	.
Origin × Destination FE	.	.	.	.	Yes	Yes
Year FE	.	.	.	.	Yes	Yes
Orig. District × Year FE	.	.	.	.	Yes	Yes
Dest. District × Year FE	.	.	.	.	Yes	Yes
O-D Density Controls	.	.	.	.	.	Yes

Notes: This table reports estimates of the gravity commuting regression equation, (3). This table is similar to Table 6, except that it only uses bilateral subdistrict location pairs that begin and end within 1 km of corridor 1 or non-corridor 1 stations, and it looks for separate effects for the different line quality. Columns 1-4 report cross-sectional regressions and only include separate fixed effects for each origin and destination location. Columns 5-7 use panel data, with a full set of origin × destination fixed effects and year fixed effects. Column 7 adds origin district × year fixed effects and destination district × year fixed effects. Robust standard errors, two-way clustered at the origin and destination sub-district level, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

A.4 Effects on Travel Times and Congestion

Robustness to Different Fixed Effects and Controls. In Appendix Table A.15, we show that our estimates of β are robust to different controls and fixed effects. Panel A reports the overall BRT results, while Panel B reports separate effects for different corridors. In column 1, we just control for origin-by-destination fixed effects. Column 2 adds separate controls for purpose-by-year, mode-by-year, and departure-hour-by-year effects. Column 3 adds the control for the number of trips taken each year for each origin-destination pair. Column 4 additionally controls for time-varying origin population density and destination population density. Column 5 is the specification we report in columns 1 and 2 of Table 7. Overall, the effects remain robust to these different specifications.

Table A.15: Negative Spillovers: Robustness to Different Fixed Effects

Panel A: Overall Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
BRT Route	0.109***	0.129***	0.125***	0.120***	0.099***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.019)
N	1139216	1137900	1137900	1119916	1137898
Adjusted R^2	0.432	0.449	0.449	0.447	0.464
Adjusted R^2 (within)	0.001	0.001	0.001	0.001	0.001
Panel B: Corridor 1 vs. Other Corridor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Corridor 1	-0.049	-0.017	-0.026	-0.033	0.003
	(0.071)	(0.074)	(0.074)	(0.074)	(0.061)
Other Corridors	0.121***	0.141***	0.137***	0.131***	0.104***
	(0.016)	(0.016)	(0.016)	(0.017)	(0.019)
N	1139216	1137900	1137900	1119916	1137898
Adjusted R^2	0.432	0.449	0.449	0.447	0.464
Adjusted R^2 (within)	0.001	0.001	0.001	0.001	0.001
Year FE	Yes				
Origin × Destination FE	Yes	Yes	Yes	Yes	Yes
Purpose × Year FE		Yes	Yes	Yes	Yes
Departure Hour × Year FE		Yes	Yes	Yes	Yes
Mode × Year FE		Yes	Yes	Yes	Yes
Number of Trips Control			Yes	Yes	Yes
Origin Density Control				Yes	
Destination Density Control				Yes	
Origin × Year FE					Yes
Destination × Year FE					Yes

Notes: Each cell in this regression corresponds to a separate estimate of β from the specification (4) to assess the differential impact on travel times for trips originating and terminating within 1 km of a BRT station. The dependent variable is log travel times, and the parameters are estimated from the pooled 2002 HVS / 2010 CS sample. In Panel A, we use all trips, while Panel B restricts the sample to train trips. In column 1, we only control for year fixed effects and origin-by-destination community FE (effectively controlling for distance). In column 2, we add controls for purpose × year, mode × year, and departure hour × year. In column 3, we add a control for changes in the total number of trips made for each origin-by-destination pair over time. In column 4, we add controls for origin and destination population density. Column 5 adds separate origin-by-year and destination-by-year fixed effects. Robust standard errors, two-way clustered by origin and destination community, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Robustness to Survey Weights. The results in Table 7 use survey weights to adjust for differences in population origin and destination locations. We report the unweighted version of this table in online Appendix Table A.16, which shows qualitatively similar results, although there are some differences in the magnitude of the point estimates.

Table A.16: Negative Spillovers: Unweighted Results

Panel A: All Trips	**(1)**	**(2)**	**(3)**	**(4)**
BRT Route	0.092***		0.026	
	(0.020)		(0.028)	
Corridor 1		-0.009		-0.176***
		(0.056)		(0.062)
Other Corridors		0.097***		0.037
		(0.020)		(0.028)
N	1137898	1137898	696308	696308
Adjusted R^2	0.444	0.444	0.407	0.407
Adjusted R^2 (within)	0.001	0.001	0.001	0.001
Panel B: Train Trips	**(1)**	**(2)**	**(3)**	**(4)**
BRT Route	-0.040		0.217	
	(0.173)		(0.462)	
Corridor 1		-0.083		0.376
		(0.297)		(0.531)
Other Corridors		-0.063		0.200
		(0.171)		(0.443)
N	35743	35743	22224	22224
Adjusted R^2	0.431	0.431	0.383	0.383
Adjusted R^2 (within)	0.001	0.001	0.000	0.000
Origin × Destination FE	Yes	Yes	Yes	Yes
Origin × Year FE	Yes	Yes	Yes	Yes
Destination × Year FE	Yes	Yes	Yes	Yes
Number of Trips Control	Yes	Yes	Yes	Yes
Only Non Peak-Time Trips			Yes	Yes

Notes: Each cell in this regression corresponds to a separate estimate of β from the specification (4) to assess the differential impact on travel times for trips originating and terminating within 1 km of a BRT station. The dependent variable is log travel times, and the parameters are estimated from the pooled 2002 HVS / 2010 CS sample. In Panel A, we use all trips, while Panel B restricts the sample to train trips. All columns include separate origin-by-year and destination-by-year fixed effects, and all columns additionally include separate purpose-by-year effects, mode-by-year effects, and departure-hour-by-year indicators. Columns 3 and 4 restrict the sample to only include non-peak time trips. Robust standard errors, two-way clustered by origin and destination community, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Variation by Mode. In Appendix Table A.17, we report the impact of the BRT system on different transport modes. We find effects that are large and significant for many road-based modes, including private cars (Panel A), private motorcycles (Panel B), and traditional public buses (Panel C). As expected, the impact on private cars is the largest, at 19.6 percent, relative to smaller effects on motorcycles and public buses. Our effects are also similar if we look for the effect of the BRT system on all bus trips, including both BRT and non-BRT buses (Panel D). This is explained by the slow BRT speeds along BRT routes, and the fact that non-BRT buses were using fewer traffic lanes.

Table A.17: Negative Spillovers: Impact of BRT on Other Modes

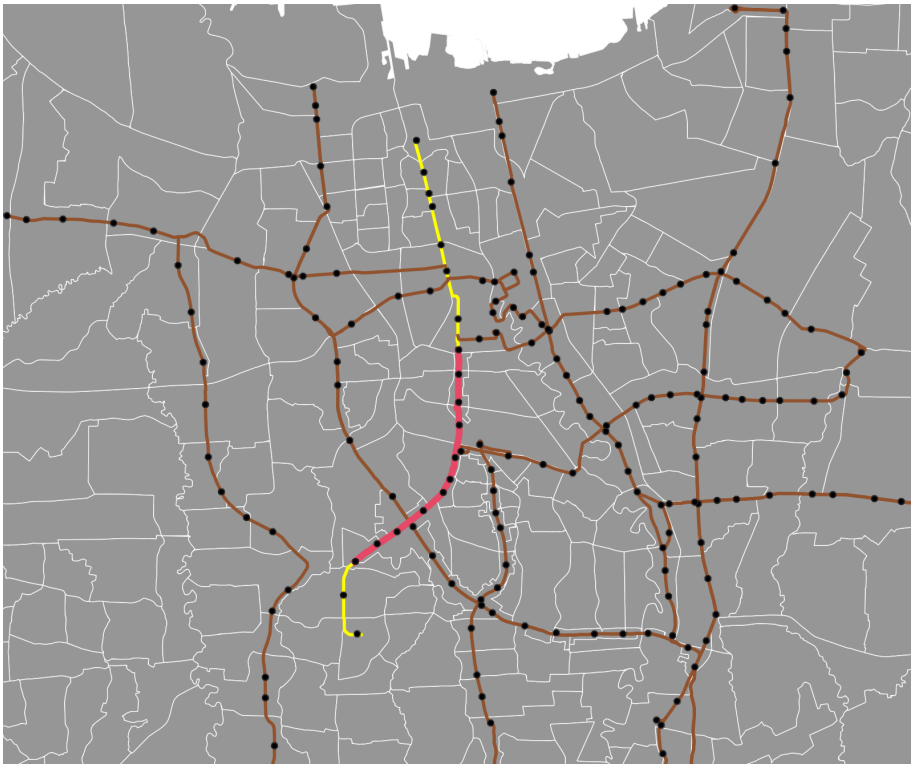
Panel A: Private Car	**(1)**	**(2)**	**(3)**	**(4)**
BRT Route	0.196***		0.134	
	(0.060)		(0.126)	
Corridor 1		-0.301		-1.279***
		(0.284)		(0.424)
Other Corridors		0.207***		0.163
		(0.060)		(0.127)
Panel B: Private Motorcycle	**(1)**	**(2)**	**(3)**	**(4)**
BRT Route	0.105***		0.042	
	(0.027)		(0.044)	
Corridor 1		0.008		-0.218***
		(0.066)		(0.080)
Other Corridors		0.108***		0.055
		(0.027)		(0.045)
Panel C: Other Public Bus	**(1)**	**(2)**	**(3)**	**(4)**
BRT Route	0.090**		0.021	
	(0.040)		(0.048)	
Corridor 1		0.009		-0.171
		(0.096)		(0.110)
Other Corridors		0.103***		0.041
		(0.040)		(0.049)
Panel D: Any Bus	**(1)**	**(2)**	**(3)**	**(4)**
BRT Route	0.104***		0.041	
	(0.034)		(0.045)	
Corridor 1		0.015		-0.171*
		(0.090)		(0.099)
Other Corridors		0.117***		0.063
		(0.034)		(0.045)
Origin × Destination FE	Yes	Yes	Yes	Yes
Origin × Year FE	Yes	Yes	Yes	Yes
Destination × Year FE	Yes	Yes	Yes	Yes
Number of Trips Control	Yes	Yes	Yes	Yes
Only Non Peak-Time Trips			Yes	Yes

Notes: Each cell in this regression corresponds to a separate estimate of β from the specification (4) to assess the differential impact on travel times for trips originating and terminating within 1 km of a BRT station. The dependent variable is log travel times, and the parameters are estimated from the pooled 2002 HVS / 2010 CS sample. Across the panels, different samples of trips are used based on different trip modes: private cars (Panel A); private motorcycles (Panel B); other public bus (Panel C); and any bus (including the traditional public bus and BRT, Panel D). All columns include separate purpose-by-year effects, mode-by-year effects, and departure-hour-by-year indicators, as well as separate fixed effects for each origin × year and destination × year. Columns 3 and 4 restrict the sample to only include non peak-time trips. Robust standard errors, two-way clustered by origin and destination community, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

HOV Expansion and Corridor 1 Results. We explore whether our findings for Corridor 1 were driven by an expansion of the high-occupancy vehicle (HOV) policy (Hanna et al., 2017). The HOV policy in Jakarta was first implemented in 1992 and until the end of 2003, only covered Sudirman and Thamrin Streets — a 6.4 km stretch of road along BRT's Corridor 1 route — and was operational only during the morning peak hours (7-10 am). In December 2003, the governor extended the policy's coverage to the whole 12.9 km of Corridor 1 route; furthermore, its hours of operation were also expanded to include evening peak hours (4-7 pm in 2003 but changed to 4.30-7 pm in September 2004).

We use the policy adjustments to examine its impact on travel time along Corridor 1. The spatial adjustment introduced two sets of road segments along Corridor 1: the "original HOV" segments that had already been covered at the time of our baseline in 2002; and the "new HOV" segments, which were included just before the launch of the BRT system and after our baseline. These different segments of Corridor 1 are depicted in Appendix Figure A.6.

Figure A.6: Corridor 1: Original HOV vs. Expanded HOV



Notes: Jakarta's HOV policy in Jakarta was first implemented in 1992 and until the end of 2003, only covered Sudirman and Thamrin Streets — a 6.4 km stretch of road along BRT's Corridor 1 route — and was operational only during the morning peak hours (7-10 am). In December 2003, the governor extended the policy's coverage to the whole 12.9 km of Corridor 1 route; furthermore, its hours of operation were also expanded to include evening peak hours (4-7 pm in 2003 but changed to 4.30-7 pm in September 2004). The red portion of Corridor 1 (depicted above) refers to the "Original HOV" section of Corridor 1, while the yellow portions refer to the expanded or "new HOV" segments.

Appendix Table A.18 reports the results of the travel time regression where we replace the "Corridor 1" indicator from Table 7 with its interaction with the "Original HOV" indicator. If the HOV policy drove our Corridor 1 results, we would expect to see a weaker negative spillover in the latter segment. We did not find any heterogeneous impact by segment both overall (Column 3) and for the original morning HOV hours of operation (Column 4). Overall, these results suggest that the HOV policy likely had little role in explaining the lack of negative spillover along Corridor 1.

Table A.18: Negative Spillovers: Historical HOV Policy

	(1)	(2)	(3)	(4)
BRT Route	0.099***			
	(0.019)			
Corridor 1		0.003	-0.012	0.021
		(0.061)	(0.073)	(0.099)
Corridor 1 × Original HOV			0.057	-0.070
			(0.086)	(0.148)
Other Corridors		0.104***	0.104***	0.095***
		(0.019)	(0.019)	(0.035)
N	1137898	1137898	1137898	239409
Adjusted R^2	0.464	0.464	0.464	0.437
Adjusted R^2 (within)	0.001	0.001	0.001	0.001
Origin × Destination FE	Yes	Yes	Yes	Yes
Origin × Year FE	Yes	Yes	Yes	Yes
Destination × Year FE	Yes	Yes	Yes	Yes
Number of Trips Control	Yes	Yes	Yes	Yes
Only Early Morning Trips (7-10 AM)	.	.	.	Yes

Notes: This table is identical to the specification from Table 7, Panel A, and columns 1-2 repeat the corresponding estimate from this table. In columns 3 and 4, we add an indicator for whether the Corridor 1 trip was between stations affected by the historical HOV policy; the corresponding sections of Corridor 1 appear in Appendix Figure A.6. All columns include separate origin-by-year and destination-by-year fixed effects, in addition to separate purpose-by-year effects, mode-by-year effects, and departure-hour-by-year indicators. Column 3 uses all of the data, while column 4 restricts the sample to only include early morning trips (7-10 AM). Robust standard errors, two-way clustered by origin and destination community, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Closer Route Comparisons. One concern with the analysis in Table 7 is that trips along BRT corridors may not be comparable to trips made further away. Appendix Table A.19 reports estimates of β by only using comparison routes that begin and end within 10 km of a BRT station (Panel A), within 5 km of a BRT station (Panel B), and within 3 km of a BRT station (Panel C). Although the magnitude of the effects vary somewhat, the point estimates are still positive and significant. In Panel D, we compare trips along BRT corridors to trips along planned or eventual BRT corridors, and again we find similar qualitative results, although our estimates are smaller and less precise.

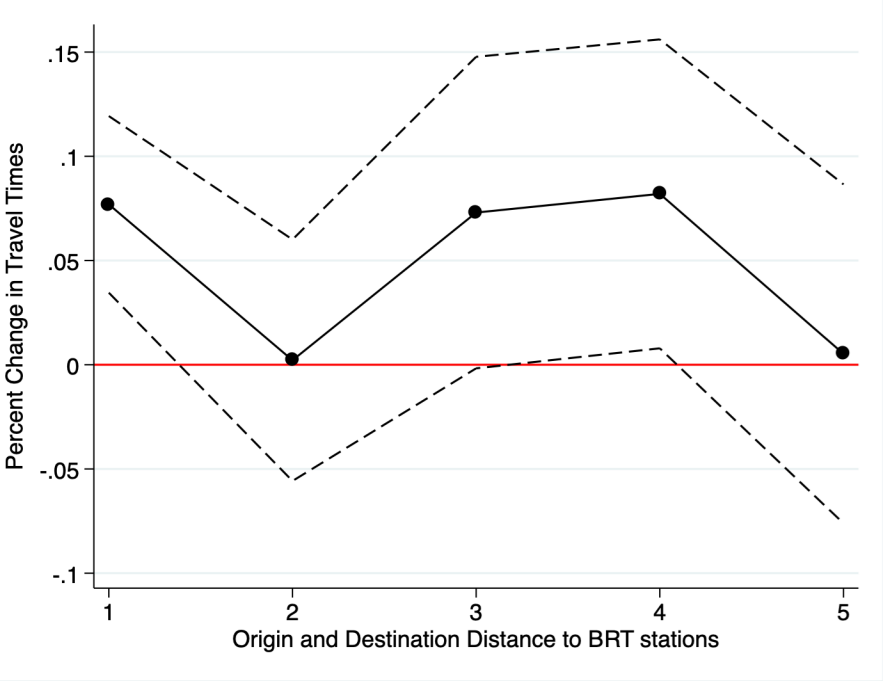
Table A.19: Negative Spillovers: Varying Treatment Comparisons

Panel A: vs. Other Trips w/ O-D $\leq$ 10 km of BRTs	**(1)**	**(2)**	**(3)**	**(4)**
BRT Route	0.101***		0.037	
	(0.019)		(0.028)	
Corridor 1		-0.003		-0.166***
		(0.054)		(0.061)
Other Corridors		0.105***		0.048*
		(0.019)		(0.028)
Panel B: vs. Other Trips w/ O-D $\leq$ 5 km of BRTs	**(1)**	**(2)**	**(3)**	**(4)**
BRT Route	0.085***		0.018	
	(0.020)		(0.029)	
Corridor 1		-0.021		-0.190***
		(0.054)		(0.062)
Other Corridors		0.090***		0.031
		(0.020)		(0.029)
Panel C: vs. Other Trips w/ O-D $\leq$ 3 km of BRTs	**(1)**	**(2)**	**(3)**	**(4)**
BRT Route	0.091***		0.028	
	(0.022)		(0.033)	
Corridor 1		0.008		-0.154**
		(0.049)		(0.063)
Other Corridors		0.094***		0.039
		(0.022)		(0.033)
Panel D: vs. Eventual and Planned Lines	**(1)**	**(2)**	**(3)**	**(4)**
BRT Route	0.044*		0.004	
	(0.024)		(0.035)	
Corridor 1		-0.062		-0.190***
		(0.056)		(0.066)
Other Corridors		0.048*		0.014
		(0.024)		(0.036)
Origin $\times$ Destination FE	Yes	Yes	Yes	Yes
Origin $\times$ Year FE	Yes	Yes	Yes	Yes
Destination $\times$ Year FE	Yes	Yes	Yes	Yes
Number of Trips Control	Yes	Yes	Yes	Yes
Only Non Peak-Time Trips			Yes	Yes

Notes: This table is identical to Table 7, except that we restrict our estimate of β from the specification (4) to only use comparison routes listed in the panel titles. In Panel A, we only use comparison routes that begin and end within 10 km of a BRT station. In Panel B, we restrict the sample to routes that originate or terminate within 5 km of a BRT station, while in Panel C, we restrict the sample to routes within 3 km of a BRT station. In Panel D, we restrict the sample to routes within 1 km of planned or eventual corridors. The dependent variable is log travel times, and the parameters are estimated from the pooled 2002 HVS / 2010 CS sample. All columns include origin-by-year and destination-by-year fixed effects, as well as separate purpose-by-year effects, mode-by-year effects, and departure-hour-by-year indicators. Columns 3 and 4 restrict the sample to only include non peak-time trips. Robust standard errors, two-way clustered by origin and destination community, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Varying Treatment Distances. In Appendix Figure A.7, we re-estimate the overall BRT effect from equation (4) by including several separate BRT distance indicators, where d ranges from 1 to 5 km, and their interaction terms. After estimating this regression, we plot the coefficient estimates on these different indicators (i.e. $\beta(d)$) as a function of distance d. This figure demonstrates that the negative externality impacts of the BRT system are highly localized. The travel time impacts of the system are positive for trips beginning and ending very close to BRT stations, but dissipate at larger levels of distance.

Figure A.7: Negative Spillovers: Impact of BRT on Travel Times by Distance



Notes: This figure reports estimates of β from the specification (4) to assess the differential impact on travel times for trips originating and terminating within d km of a BRT station. The dependent variable is the log travel times, and the parameters are estimated from the pooled 2002 HVS / 2010 CS sample. In this specification, we include indicators for whether or not a trip originates or terminates within d km of a BRT station, with $d \in \{1, 2, 3, 4, 5\}$ and we plot separate effects of the different distance-specific interaction terms. The regression includes separate purpose-by-year effects and separate departure-hour-by-year indicators. Robust standard errors, two-way clustered by origin and destination community, are represented by the dashed lines.

Variation by Initial Levels of Traffic. In Appendix Table A.20, we explore the extent to which the BRT system's infrastructure impacted travel times differently on different routes based on their initial traffic levels. Intuitively, routes that were already quite congested should have grown even worse as a consequence of the BRT taking away lanes, while other routes should be less impacted. We first measure traffic on a given route from o to d at departure hour h by constructing a delay measure (minutes per km, or the inverse of speed) using the 2002 trip data. We then construct a high delay indicator equal to 1 if those trips had above median delay in 2002. Finally, we included separate interaction terms of this high delay indicator with all of the BRT variables in (4). We find that the BRT effects seem to be more pronounced for the high delay routes, as expected.

Table A.20: Negative Spillovers: Heterogeneity by Initial Traffic

	(1)	(2)
BRT Route	0.099***	-0.059
	(0.019)	(0.046)
BRT Route × High Delay		0.201***
		(0.047)
N	1137898	1137898
Adjusted R^2	0.464	0.471
Adjusted R^2 (within)	0.001	0.014
Origin × Destination FE	Yes	Yes
Origin × Year FE	Yes	Yes
Destination × Year FE	Yes	Yes
Number of Trips Control	Yes	Yes

Notes: This table is identical to the specification from Table 7, Panel A, and indeed, column 1 repeats the estimate from this table, column 1. In column 2, we add an indicator for high delay and interact it with the BRT route indicator. This indicator was constructed by first constructing a delay measure (minutes per km, or the inverse of speed) for a given route from o to d at departure hour h using the 2002 trip data. The high delay indicator equal to 1 if those trips had above median delay in 2002. Both columns include separate origin-by-year and destination-by-year fixed effects, in addition to separate purpose-by-year effects, mode-by-year effects, and departure-hour-by-year indicators. Robust standard errors, two-way clustered by origin and destination community, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

A.5 Model and Counterfactual Results

Table A.21: Estimates of $\widetilde{\gamma_1}$ and $\widetilde{\gamma_2}$

	(1)	(2)
γ_1 (BRT)	-3.675***	-3.464***
	(0.044)	(0.054)
γ_2 (TRAIN)	-3.650***	-3.497***
	(0.044)	(0.049)
LOG TRAVEL TIMES^{-1} (SURFACE) - LOG TRAVEL TIMES^{-1} (m)		0.352***
		(0.050)
N	9723	9723
N OF CLUSTERS	3241	3241
R^2	0.348	0.352
R^2 WITHIN	0.490	0.493
F STATISTIC	9240.844	6293.124
ORIGIN × DESTINATION FE	X	X

Notes: This table reports estimates of $\widetilde{\gamma_1}$ and $\widetilde{\gamma_2}$ from equation (15). We use ordinary least squares to obtain these estimates, with separate origin × destination fixed effects as controls. Robust standard errors, clustered at the origin × destination level, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Table A.22: Different Sets of Model Parameters

PANEL A: PREFERRED PARAMETERS	VALUE	SOURCE
θ	5.364	LINEAR GRAVITY $+\kappa = 0.012$
α	-0.120	AHLFELDT ET AL. (2015, BERLIN)
β	-0.100	AHLFELDT ET AL. (2015, BERLIN)
λ	0.055	IMPLIED, BASED ON Δ LANE EST. AND LINEAR GRAVITY θ
γ_1	0.646	IMPLIED, BASED ON $\widetilde{\gamma}$ REGRESSIONS AND θ
γ_2	0.652	IMPLIED, BASED ON $\widetilde{\gamma}$ REGRESSIONS AND θ
STABILITY CONDITION, α	YES	
STABILITY CONDITION, β	YES	
PANEL B: TSIVANIDIS (2019)	**VALUE**	**SOURCE**
θ	3.020	TSIVANIDIS (2019, AVG. FOR HIGH AND LOW SKILL WORKERS)
α	-0.120	AHLFELDT ET AL. (2015, BERLIN)
β	-0.100	AHLFELDT ET AL. (2015, BERLIN)
λ	0.098	IMPLIED, BASED ON Δ LANE EST. AND AVG. θ (TSIVANIDIS, 2019)
γ_1	1.147	IMPLIED, BASED ON $\widetilde{\gamma}$ REGRESSIONS AND θ
γ_2	1.158	IMPLIED, BASED ON $\widetilde{\gamma}$ REGRESSIONS AND θ
STABILITY CONDITION, α	YES	
STABILITY CONDITION, β	YES	
PANEL C: ALLEN AND ARKOLAKIS (2020)	**VALUE**	**SOURCE**
θ	6.830	AHLFELDT ET AL. (2015, BERLIN)
α	-0.120	AHLFELDT ET AL. (2015, BERLIN)
β	-0.100	AHLFELDT ET AL. (2015, BERLIN)
λ	0.071	ALLEN AND ARKOLAKIS (2020, SEATTLE)
γ_1	0.507	IMPLIED, BASED ON $\widetilde{\gamma}$ REGRESSIONS AND θ
γ_2	0.512	IMPLIED, BASED ON $\widetilde{\gamma}$ REGRESSIONS AND θ
STABILITY CONDITION, α	YES	
STABILITY CONDITION, β	YES	
PANEL D: PPML GRAVITY	**VALUE**	**SOURCE**
θ	12.091	PPML GRAVITY $+\kappa = 0.012$
α	-0.120	AHLFELDT ET AL. (2015, BERLIN)
β	-0.100	AHLFELDT ET AL. (2015, BERLIN)
λ	0.025	IMPLIED, BASED ON Δ LANE EST. AND PPML θ
γ_1	0.286	IMPLIED, BASED ON $\widetilde{\gamma}$ REGRESSIONS AND θ
γ_2	0.289	IMPLIED, BASED ON $\widetilde{\gamma}$ REGRESSIONS AND θ
STABILITY CONDITION, α	YES	
STABILITY CONDITION, β	YES	

Notes: This table reports the parameter values and sources of different parameters used to assess the robustness of our counterfactual welfare impacts, as reported in Table 8.

Table A.23: Counterfactuals: Relative Delay

		BRT			Train	Surface		
		Overall	Corr 1	Oth. Corr	Overall	Overall	Corr 1	Oth. Corr
	Description	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Baseline	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	Removing the BRT				0.9999	0.9838	1.0001	0.9080
3	Keeping BRT but Removing Spillovers	1.0036	0.9814	1.0037	0.9965	0.9806	0.9957	0.9104
4	Improving the BRT, Corridor 1 Standards	1.0037	1.0029	1.0038	1.0010	0.9839	0.9996	0.9109
5	Improving the BRT, TransMilenio Standards	1.0485	1.0052	1.0486	1.0018	0.9840	0.9994	0.9130
6	Building Planned Lines, Low Quality	1.0000	1.0000	1.0000	1.0000	1.0069	1.0083	1.0000
7	Building Planned Lines, High Quality	1.0464	1.0022	1.0466	1.0011	0.9803	1.0000	0.8896

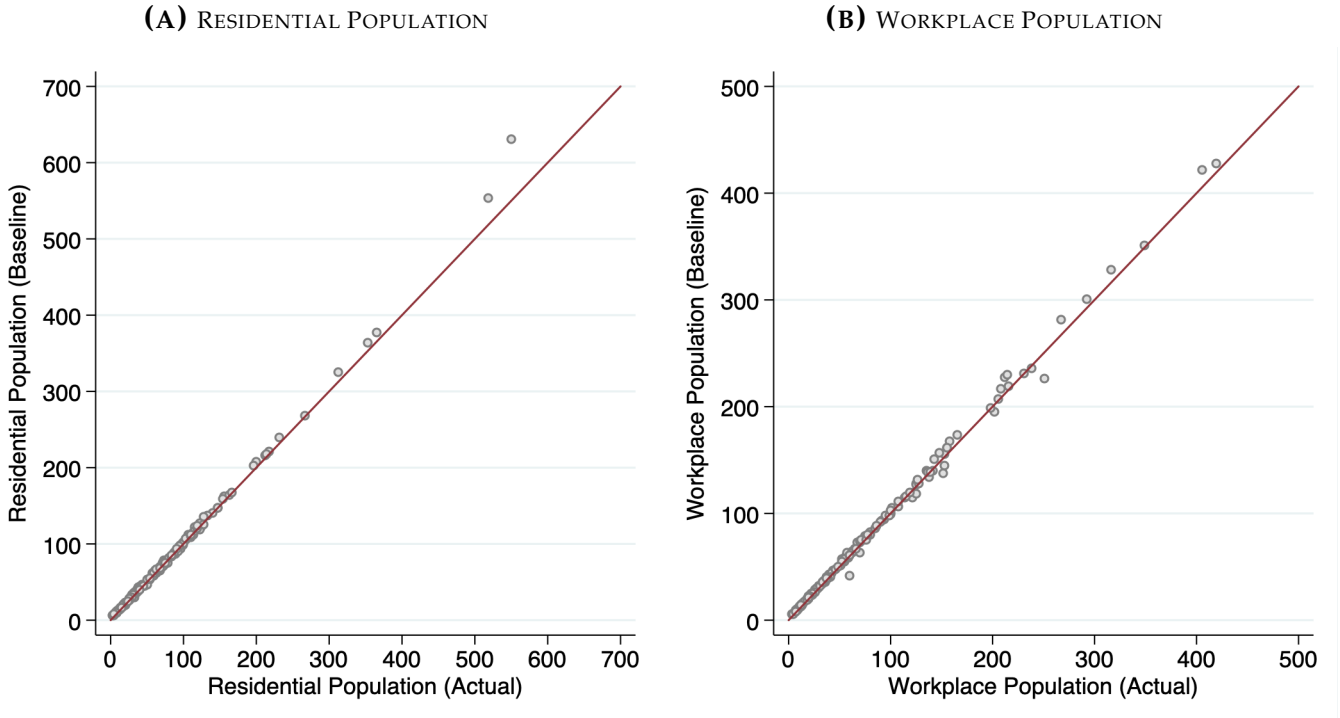
Notes: This table reports changes in average delay (i.e. $\widehat{\text{delay}}_{kl} = \text{delay}'_{kl}/\text{delay}_{kl}$, where delay_{kl} is expressed in minutes per km) from different counterfactuals listed in the row title. The columns report different measures of changes in average delay taken across different transport networks. For details on how changes in relative delay are calculated, see Appendix Section E.10. These results correspond to those presented in Table 8, and the parameters for the simulation can be found in Appendix Table A.22, Panel A.

Table A.24: Counterfactuals: Welfare ($\widehat{\overline{W}}$, Varying Parameters)

		Preferred	Tsivanidis (2019)	Allen Arkolakis (2020)	PPML Gravity
	Description	(1)	(2)	(3)	(4)
1	Baseline	1.00000	1.00000	1.00000	1.00000
2	Removing the BRT	0.99957	1.00058	0.99959	0.99957
3	Keeping BRT but Removing Spillovers	1.00082	1.00275	1.00072	1.00017
4	Improving the BRT, Corridor 1 Standards	1.00083	1.00277	1.00074	1.00017
5	Improving the BRT, TransMilenio Standards	1.00148	1.00389	1.00133	1.00048
6	Building Planned Lines, Low Quality	0.99999	0.99998	1.00005	0.99999
7	Building Planned Lines, High Quality	1.00106	1.00185	1.00099	1.00050
	Parameter Values	(1)	(2)	(3)	(4)
	θ	5.364	3.020	6.830	12.091
	α	-0.120	-0.120	-0.120	-0.120
	β	-0.100	-0.100	-0.100	-0.100
	λ	0.055	0.098	0.071	0.025
	γ_1	0.646	1.147	0.507	0.286
	γ_2	0.652	1.158	0.512	0.289

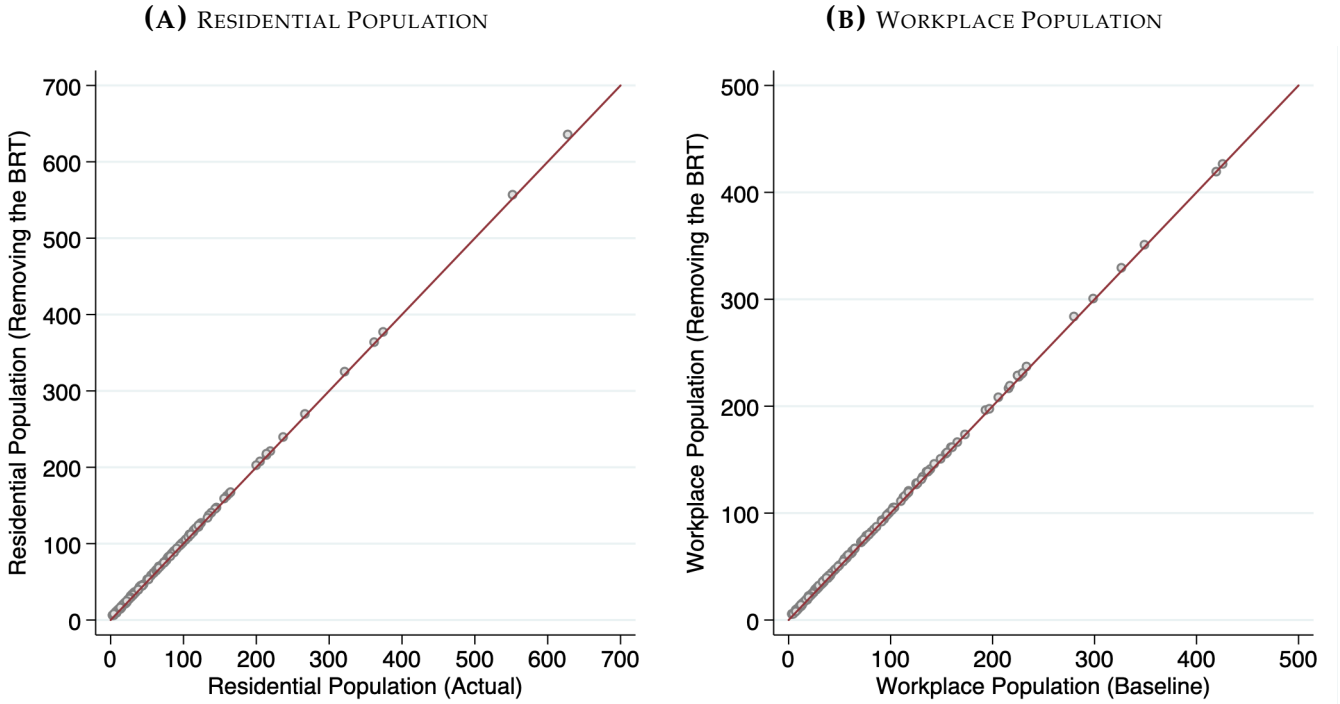
Notes: This table reports the changes in welfare from different counterfactuals listed in the row title under different parameter assumptions. The columns in this table refer to different parameter sets, taken from the respective panels of Appendix Table A.22. Panel A, our prefered set of tables, is also reported in Table 8.

Figure A.8: Baseline Model: Actual vs. Predicted Population



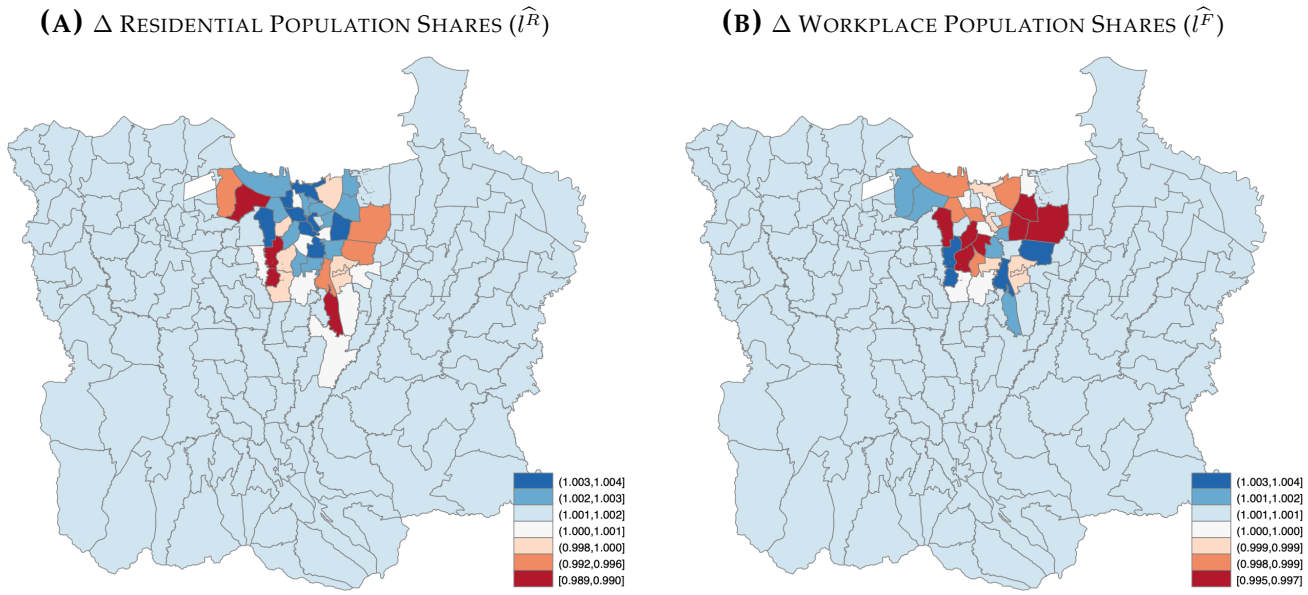
Notes: These scatterplots report a comparison of actual and model baseline residential and workplace population counts, where the baseline is simulated under no changes in transport costs (i.e. $\widehat{t_{ij}} = 1$). The red lines indicate the 45 degree lines. For the residential population graph, the bivariate regression coefficient is $\widehat{\beta} = 1.05$, with an $R^2 = 0.9965$. For the workplace population graph, the bivariate regression coefficient is $\widehat{\beta} = 1.01$, with an an $R^2 = 0.9971$.

Figure A.9: Baseline vs. Removing the BRT: Actual vs. Predicted Population



Notes: These scatterplots report a comparison of residential and workplace population counts between the baseline model and the removing the BRT counterfactual, where the baseline is simulated under no changes in transport costs (i.e. $\widehat{\tilde{t}_{ij}} = 1$). The red lines indicate the 45 degree lines. For the residential population graph, the bivariate regression coefficient is $\widehat{\beta} = 1.0041$, with an $R^2 = 0.99998$. For the workplace population graph, the bivariate regression coefficient is $\widehat{\beta} = 1.0009$, with an an $R^2 = 0.99997$.

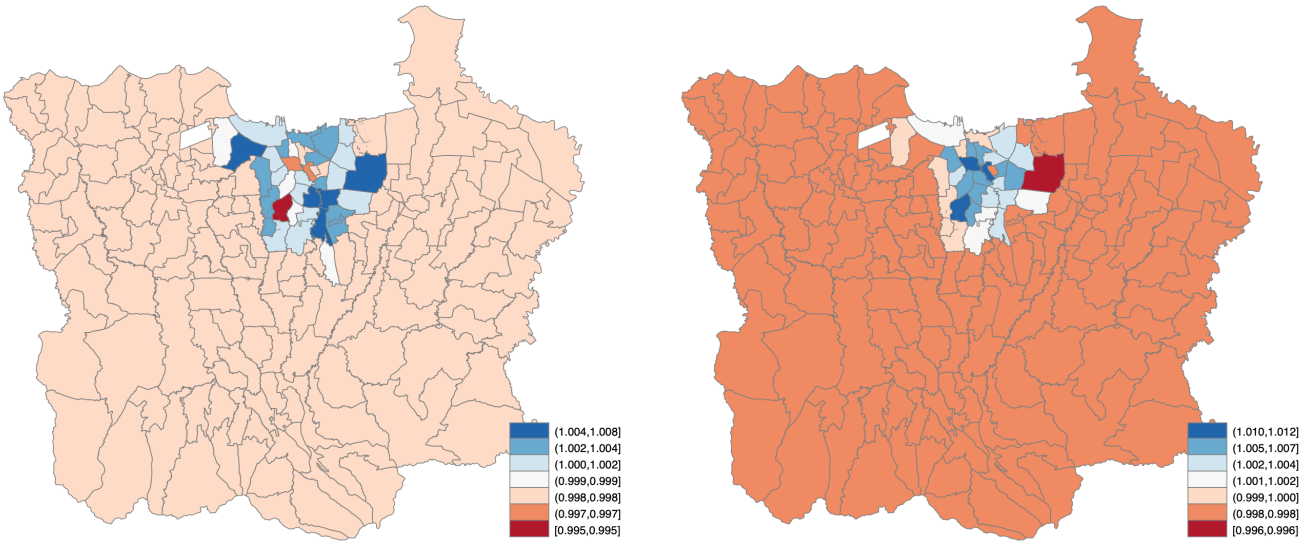
Figure A.10: Counterfactual Δ Residential and Workplace Population Shares: Removing the BRT



Notes: These panels report changes in workplace and residential population shares from the counterfactual corresponding to Table 8, row 2.

Figure A.11: Counterfactual Δ Residential and Workplace Population Shares: Keeping the BRT but Removing Spillovers

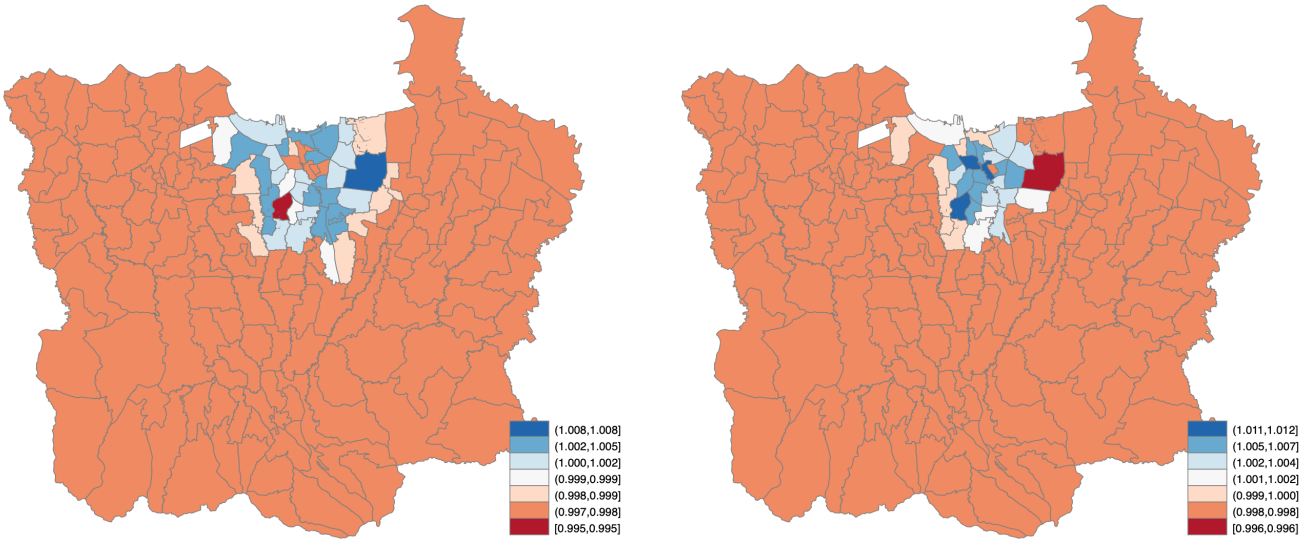
(A) Δ Residential Population Shares ($\widehat{l^R}$) **(B)** Δ Workplace Population Shares ($\widehat{l^F}$)



Notes: These panels report changes in workplace and residential population shares from the counterfactual corresponding to Table 8, row 3.

Figure A.12: Counterfactual Δ Residential and Workplace Population Shares: Corridor 1 Standards

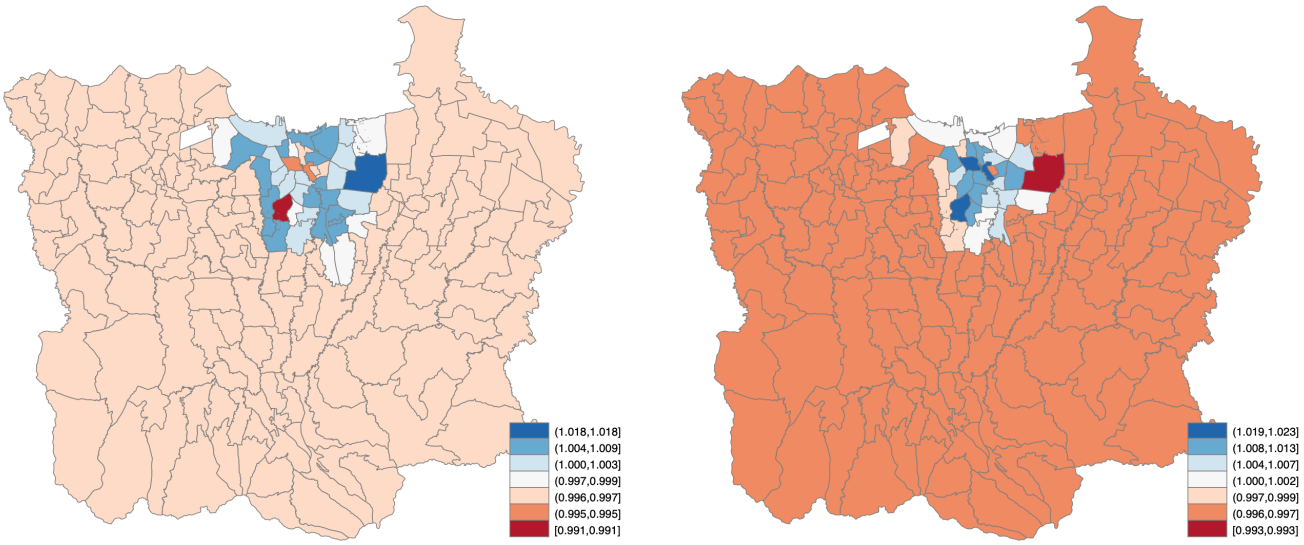
(A) Δ Residential Population Shares ($\widehat{l^R}$) **(B)** Δ Workplace Population Shares ($\widehat{l^F}$)



Notes: These panels report changes in workplace and residential population shares from the counterfactual corresponding to Table 8, row 4.

Figure A.13: Counterfactual Δ Residential and Workplace Population Shares: TransMilenio Standards

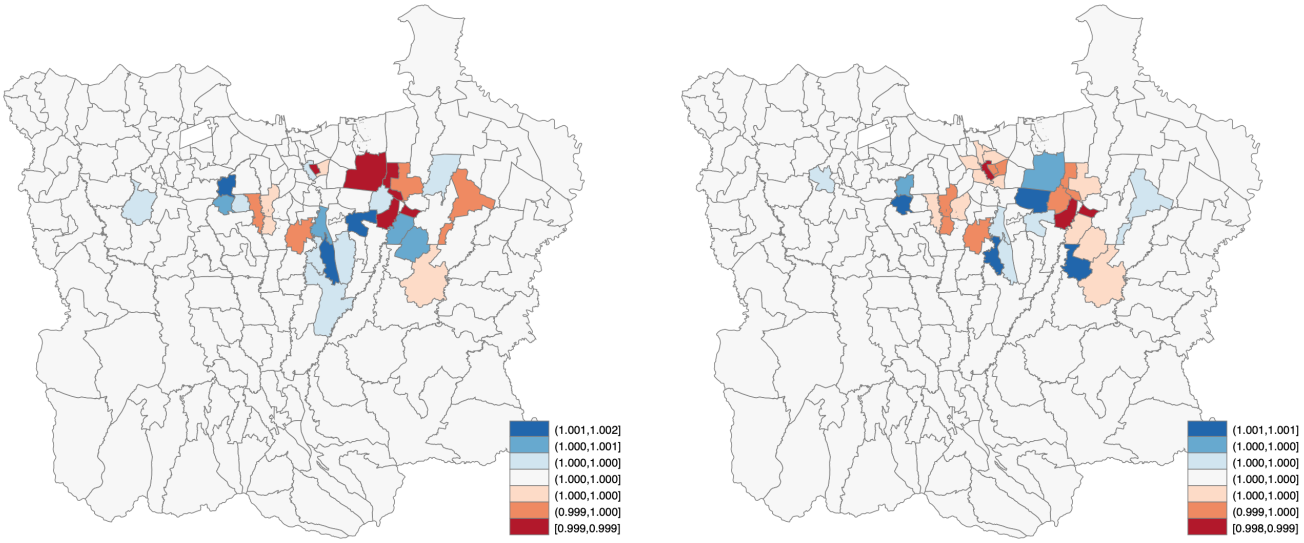
(A) Δ Residential Population Shares ($\widehat{l^R}$) **(B)** Δ Workplace Population Shares ($\widehat{l^F}$)



Notes: These panels report changes in workplace and residential population shares from the counterfactual corresponding to Table 8, row 5.

Figure A.14: Counterfactual Δ Residential and Workplace Population Shares: Building Planned Lines, Low Quality

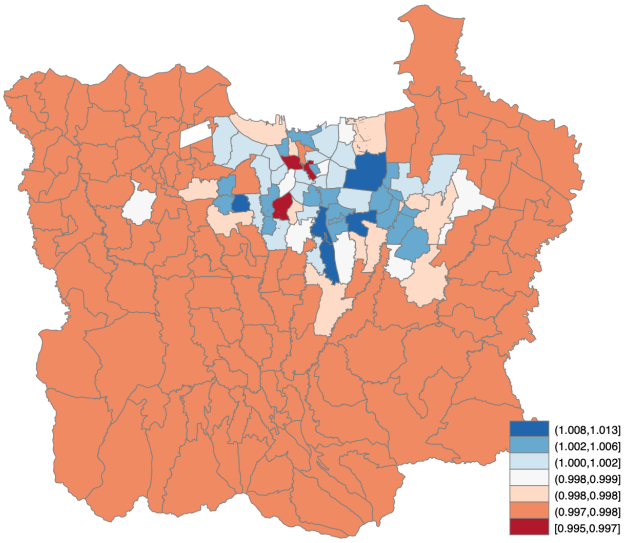
(A) Δ Residential Population Shares ($\widehat{l^R}$) **(B)** Δ Workplace Population Shares ($\widehat{l^F}$)



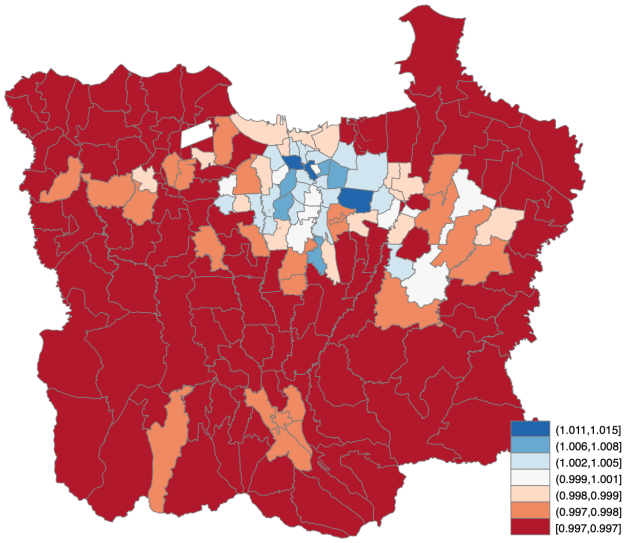
Notes: These panels report changes in workplace and residential population shares from the counterfactual corresponding to Table 8, row 6.

Figure A.15: Counterfactual Δ Residential and Workplace Population Shares: Building Planned Lines, High Quality

(A) Δ Residential Population Shares ($\widehat{l^R}$)



(B) Δ Workplace Population Shares ($\widehat{l^F}$)



Notes: These panels report changes in workplace and residential population shares from the counterfactual corresponding to Table 8, row 7.

B Data Appendix

Commuter Travel Surveys. Our trip-level data come from two cross-sectional surveys of households conducted by the Japan International Cooperation Agency (JICA) study teams in 2002 and 2010. The 2002 survey, known as the Home Visit Survey (HVS), was part of the *Study on Integrated Transportation Master Plan* (SITRAMP). SITRAMP was a technical assistance project intended to anticipate future transportation challenges in the Jabodetabek metropolitan area. HVS was collected, among others, to establish a person-trip origin-destination (PTOD) matrices for urban transportation planning purposes (JICA, 2004b).

The 2002 HVS's sampling strategy was designed to capture trip characteristic information for approximately 3 percent of the Jabodetabek population. Ultimately, its sample included 163,334 households across 1,485 communities (*kelurahan*) (JICA, 2004b, Table 1.8.3). HVS collected information on the demographic and socio-economic characteristics of households and their members, as well as trip information of household members. The trip information module included questions on the origin and destination of trips as well as their modes of transport (including transfers), and departure and arrival times. According to background reports, the 2002 survey was a massive undertaking, with 2,418 enumerators each making approximately 70 home visits over a 3 month period (July-September) (JICA, 2004a).

For 2010, the trip-level data come from the 2010 Commuter Survey (CS). This survey was also collected by the JICA study team under the *Jabodetabek Urban Transport Policy Integration Project* (JUTPI). The 2010 CS was similarly designed to capture trip characteristics for 3 percent of the Jabodetabek population (or around 180,000 households in 2010). Its final sample included around 180,000 households in 1,499 communities (OCAC, 2011, Table 1.2.1).

As an update to the HVS data, the survey contained modules that were similar to the 2002 HVS (OCAC, 2011). The survey was also another large data collection effort, employing 1,800 enumerators, each of whom surveyed approximately 100 households over a 6 month period (March-August). The 2010 field team also consisted of 65 supervisors, 13 field coordinators, and 4 region chiefs to administer the survey work (OCAC, 2011).

Census Data. We use the 2000 and 2010 population census microdata to construct aggregate community-level characteristics. These population censuses provide complete enumerations of the Indonesian population. These censuses collect individual-level demographic information such as age, gender, education, and migration status in the past five years. In addition, they also collect information about the physical characteristics of buildings and individual residences. These datasets were collected by Indonesia's national statistical agency, *BPS Indonesia*.

***Potensi Desa* (*Podes*).** To examine the pre-treatment trends in vehicle ownership between treated and almost-treated communities before BRT implementation, we use the 1996 and 2000 village census data known as *Potensi Desa* or Podes. Podes collects socio-economic and demographic characteristics from every single Indonesian village by interviewing its village official (typically, the village head). It is collected triennially by Indonesia's national statistical agency, *BPS Indonesia*.

NOAA Data on Light Intensity. To proxy for economic activities at the local level, we make use of an innovative technique, developed by Henderson et al. (2012), which uses satellite data on nighttime lights. Daily between 8:30 PM and 10:00 PM local time, satellites from the United States Air Force Defense Meteorological Satellite Program (DMSP) record the light intensity of every 30-arc-second-square of the Earth's surface (corresponding to roughly 0.86 square kilometers). DMSP cleans this daily data, dropping anomalous observations, and provides the public with annual averages of light intensity from multiple satellites. After averaging the data across multiple satellites, we obtain annual estimates of light intensity for every 30-arc-second square of the Earth's surface in 1992 and 2002. We construct log average night light intensity in 1992 as a pre-treatment baseline measure of each community's development and we measure the log difference between average night light intensity in 1992 and 2002 as a proxy

for urban growth.[51]

Elevation. We construct the topographical variables using raster data from the Harmonized World Soil Database (HWSD), Version 1.2 (Fischer et al., 2008). The digital elevation map records the median elevation (in meters) of each 5 arc minute $\times$ 5 arc minute pixel (roughly 10 km^2).[52]

Ruggedness. A 30 arc-second ruggedness raster was computed for Indonesia according to the methodology described by Sappington et al. (2007). The authors propose a Vector Ruggedness Measure (VRM), which captures the distance or dispersion between a vector orthogonal to a topographical plane and the orthogonal vectors in a neighborhood of surrounding elevation planes. To calculate the measure, one first calculates the x, y, and z coordinates of vectors that are orthogonal to each 30-arc second grid of the Earth's surface. These coordinates are computed using a digital elevation model and standard trigonometric techniques.

Given this, a resultant vector is computed by adding a given cell's vector to each of the vectors in the surrounding cells; the neighborhood or window is supplied by the researcher. Finally, the magnitude of this resultant vector is divided by the size of the cell window and subtracted from 1. This results in a dimensionless number that ranges from 0 (least rugged) to 1 (most rugged).[53]

For example: on a (3×3) flat surface, all orthogonal vectors point straight up, and each vector can be represented by $(0, 0, 1)$ in the Cartesian coordinate system. The resultant vector obtained from adding all vectors is equal to $(0, 0, 9)$, and the VRM is equal to $1 - (9/9) = 0$. As the (3×3) surface deviates from a perfect plane, the length of the resultant vector gets smaller, and the VRM increases to 1.

BRT Capital and Operating Costs. To obtain cost parameters for counterfactual exercises, we used a capital cost value for Corridor 1 of \$1.512 million per km, inflating the figure reported in Hidalgo and Graftieaux (2008) to 2010 dollars. This same paper also reports a value for TransMilenio's capital costs of \$8.86 million per km (adjusted to 2010 dollars).

For other corridor capital costs, local news sources reported the costs for 36.88 km of Corridors 2-3 to be \$66.78 million, or 1.81 million per km in 2004 dollars.[54] The capital costs for 54.33 km of Corridors 4-7 were reported to be \$95.29 million, or 1.75 million per km in 2006 dollars.[55] The capital costs for 74.9 km of Corridors 8-10 were reported to be \$35.36 million, or 0.47 million per km in 2008 dollars.[56] Converting these figures to 2010 dollars and taking a weighted average by km, we obtain an average cost per km value of \$1.297 for other corridor capital costs. Note that the difference between actual and counterfactual construction costs can be tricky to interpret; for instance, one reason why the other corridor construction costs were relatively similar to Corridor 1 costs could have been that it was easier to construct Corridor 1 because it was originally a wider route than the other areas.

For all counterfactuals, we also assumed that the system's operating costs would be \$45.01 million per year in 2010 dollars. This figure is consistent with data from Jakarta's Department of Public Transportation (*Dinas Per-*

[51] The DMSP-OLS Nighttime Lights Time Series Version 4 datasets can be downloaded here: http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html.

[52] Raster files from the HWSD project are publicly available at http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/.

[53] The authors have generously provided a Python script for computing their Vector Ruggedness Measure (VRM) in ArcView. The script and detailed instructions for installation can be found here: https://www.arcgis.com/home/item.html?id=9e4210b3ee7b413bbb1f98fb9c5b22d4.

[54] https://news.detik.com/berita/d-158050/biaya-pembangunan-koridor-ii-dan-iii-busway-rp-600-miliar, accessed on 5/4/2021.

[55] https://news.detik.com/berita/d-514013/sutiyoso-ultimatum-menhub-canangkan-subway-tahun-ini-juga, accessed on 5/4/2021.

[56] https://megapolitan.kompas.com/read/2008/01/16/21411358/dibangun.63.halte.busway, accessed on 5/4/2021.

hubungan), which reports a 71 percent cost recovery set against a total revenue of $63.392 million in 2010 dollars.[57] While we do not vary operating costs in simulations, these costs could either increase or decrease depending on how the BRT adjustments impacted ridership, the number of stations, and the number of buses being operated.

[57] *Dinas Perhubungan Dalam Angka 2011* from https://jakarta.go.id/dokumen/46/dinas-perhubungan-dalam-angka-tahun-2011, accessed on 5/2/2021.

C Characterizing Greater Jakarta's Urban Form

This section describes Greater Jakarta's evolving spatial structure and trends in commuting patterns, providing context for the potential impacts of a BRT system. We first summarize economic and demographic characteristics of the metropolitan area. We then link these characteristics to a description of residents' commuting mode choices. Finally, we summarize their commuting trips.

Residential and Workplace Locations. Employment in Jakarta is largely service-sector oriented, and most employers are located in DKI Jakarta. Appendix Figure C.1, Panel A presents a map of employer locations, showing the probability that an individual works in a community based on the 2010 CS data. The greatest employment probabilities are found in the center of the city; however, some employment has shifted away from the city center and other centers are located in different places in the metropolitan region. From 2000 to 2010, Greater Jakarta also experienced rapid population growth, adding 7 million more people to its total population. This amounts to an annual population growth rate of 3.6 percent per year. However, growth was more pronounced in the peripheral regions of the metropolitan area, outside of DKI Jakarta borders (depicted in thick black) and symptomatic of urban sprawl. Panel B of Appendix Figure C.1 depicts population growth across communities, with darker areas corresponding to faster growth.

Because such a large share of Greater Jakarta's population does not reside in DKI Jakarta but instead lives in the surrounding municipalities, the TransJakarta BRT system does not adequately connect workers and firms. In Appendix Table C.1, we report the share of the population in 2002 that lives and works within different distances of BRT stations (built by 2010). In 2002, only 12.7 percent of the population of Greater Jakarta lived and worked within 1 km of a BRT station. This share fell to 9.4 percent in 2010, as workers and firms decentralized away from the central city where BRT stations are located.

Vehicle Ownership and Commuting Mode Choice. Between 2002 and 2010, the number of workers living farther from their jobs increased, stimulating demand for travel (Turner, 2012). Over the same period, Jakarta also experienced a dramatic increase in vehicle ownership, especially motorcycles. Panel A of Appendix Figure C.2 shows that the share of households owning at least one motorcycle more than doubled, from 37.0 percent in 2002 to a staggering 75.8 percent in 2010. Although some of the expansion in motorcycle ownership could be explained by per-capita income gains, another explanation may be new loan schemes and expanded consumer credit, which enabled even the lowest income households to own motorcycles (Yagi et al., 2012).[58] Car ownership also increased from 2002 to 2010, but not as dramatically as motorcycle ownership.[59]

To measure mode choice, we rely on a question in both surveys that asks the respondent to name the mode they most commonly use for intra-city travel purposes.[60] Using this definition, Panel B of Appendix Figure C.2 shows how changes in vehicle ownership significantly altered mode choice patterns. In 2002, the traditional public bus system was the most popular commuting mode with a 52.3 percent share, but by 2010, this share had fallen to 23.4 percent.[61] By 2010, the most popular commuting mode was private motorcycles. During the sample period, motorcycle's mode share more than doubled, rising from 21.5 percent in 2002 to 50.8 percent in 2010. In a congested

[58] Yagi et al. (2012) also document rising motorcycle ownership but use a different data source: the number of registered vehicles in DKI Jakarta. During the same time, the number of registered cars doubled, while the number of registered motorcycles more than quadrupled. Note also that according to the 2010 CS over 20 percent of households owned more than one motorcycle, and nearly one third of the lowest-income households surveyed owned a motorcycle.

[59] Income growth also explains some of these findings, and Senbil et al. (2007) shows that the share of motorcycle and car ownership in Greater Jakarta increases with income. However, unlike the case of cars, the share of motorcycle ownership actually declines for the top 3 income groups in the JICA data. See Appendix Figure C.3 for more detail.

[60] Other measures, such as those constructed from trip data to calculate the mode consuming the most distance or the most time during an individual's trips, yield similar results.

[61] The traditional public bus system consists of several independent private bus operators with fleets of minivans and small buses. Although they follow set routes in Jakarta, traditional public buses do not keep a fixed schedule, and drivers are compensated on a per-fare basis, so that they compete for riders. Traditional buses make stops anywhere they want to pick up and drop off customers, instead of using designated bus stops, which the city has not provided (Radford, 2016).

traffic environment, motorcycles may offer commuters a way to weave through traffic and reduce travel times. In 2010, the large share of motorcycles substantially dwarfs the small portion of commuters who mainly ride the TransJakarta BRT system (4.3 percent).[62]

Commuting Characteristics. Appendix Table C.2 contains summary statistics for all well-defined trips. Overall, the average trip in 2002 had a distance of 4 km, with an average travel time of over 30 minutes and a slow speed of just over 8 km per hour. By 2010, trip distances had increased slightly, to an average of 4.7 km, travel times fell slightly to an average of 29 minutes, and average speeds increased to nearly 12 km per hour. In 2002, 50 percent of trips in the data took place within a single community, and this share increased slightly in 2010.

In Appendix Table C.3, we regress log travel times on trip characteristics and a year indicator to measure overall changes in travel times. Column 1, which only controls for distance, demonstrates that overall travel times fell by 11.6 percent between 2002 and 2010. However, nearly all of this reduction can be explained by differences in trip characteristics, departure times, mode choices, and the mix of origins and destinations. When we control for these effects in Column 4, the time savings fall to 3.2 percent. Although statistically significant, this effect is not economically meaningful, representing only a minute of time savings for an average trip duration.

[62]Note that BRT's small mode share also persists in data collected subsequent to our sample, even after TransJakarta was transferred to private management and service quality improved. In 2014, BPS conducted a commuter survey in Greater Jakarta and found that the overall TransJakarta mode share was only 2.5 percent (see Appendix Figure C.4). Witoelar et al. (2017) also find similar mode share results for a survey of females in DKI Jakarta in 2016.

Table C.1: Residence and Workplace Distance to BRT Stations (2002)

		Workplace Distance to BRT Stations (km)					
		0-1	**1-2**	**2-3**	**3-4**	**4-5**	**5+**
Residential Distance to BRT Stations (km)	**0-1**	12.7	2.7	1.0	0.5	0.4	1.4
	1-2	2.8	7.0	0.8	0.4	0.3	0.8
	2-3	1.1	0.9	3.4	0.3	0.2	0.5
	3-4	0.6	0.5	0.3	2.3	0.2	0.4
	4-5	0.5	0.3	0.2	0.2	1.7	0.6
	5+	2.5	1.1	0.6	0.4	0.7	49.7

Notes: This table reports the percent of the population in 2002 and living and working within different distances of BRT stations built by mid 2010, based on the JICA data. The rows and columns list different bins of distances to BRT stations, with distances measured in kilometers.

Table C.2: Summary Statistics on Well-Defined Trips

	2002 (N = 653,814)	**2010 (N = 541,630)**	Δ
Panel A: All Trips	**Mean (sd)**	**Mean (sd)**	p**-Value**
Distance from Origin to Destination (km)	4.00 (5.73)	4.69 (6.87)	0.000
Trip within Kelurahan (0 1)	0.50 (0.50)	0.51 (0.50)	0.000
Travel time (min)	31.56 (27.49)	28.70 (24.49)	0.000
Speed (km / hour)	8.29 (10.13)	11.80 (32.63)	0.000

Notes: Authors' calculations on well-defined trips, using the 2002 HVS and the 2010 CS trip data. The sample of well-defined trips consists of all trips that contain information on travel times, origin and destination communities (*kelurahan*), modes, and trip purposes. Each observation is a trip, and means are computed using survey weights. The p-values in this table are computed by conducting a two-sided equality of means t-test between years.

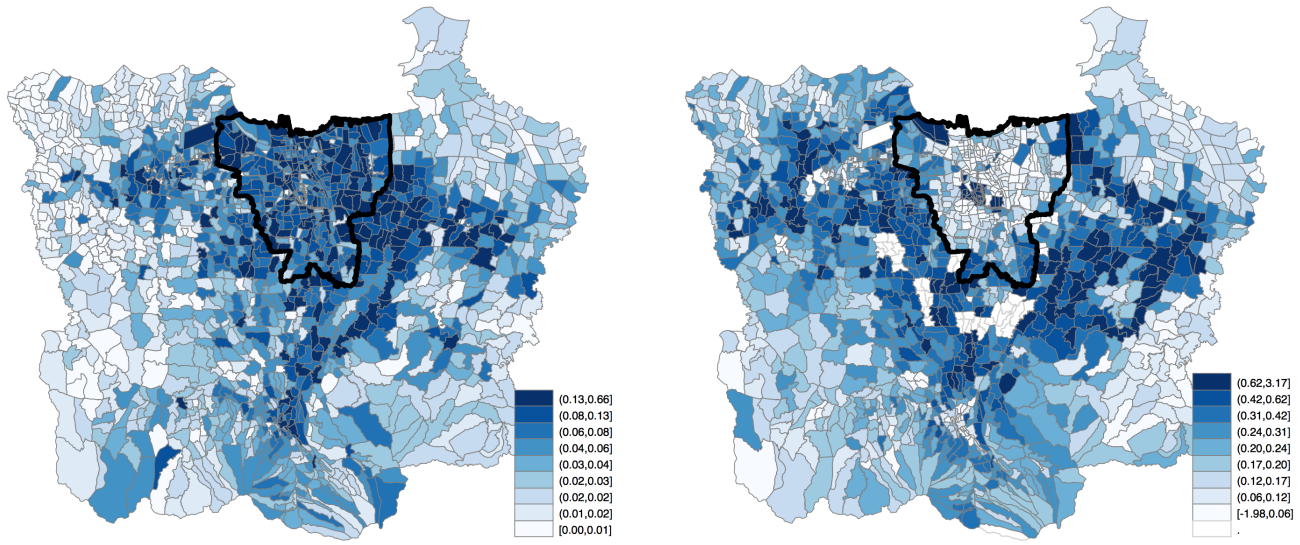
Table C.3: Log Travel Time Regressions

	(1)	(2)	(3)	(4)
Year is 2010 (0 1)	-0.116***	-0.107***	-0.087***	-0.080***
	(0.007)	(0.009)	(0.010)	(0.009)
Distance from Origin to Destination (km)	0.074***	0.068***	0.063***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Train		-0.058***	-0.029**	-0.005
		(0.013)	(0.014)	(0.013)
Other Public Transport (Bus / Van)		-0.098***	-0.038**	-0.001
		(0.014)	(0.015)	(0.014)
Taxi / Ojek / Bajaj		-0.195***	-0.069***	-0.012
		(0.018)	(0.017)	(0.016)
Private Car		0.095***	0.058***	0.014
		(0.017)	(0.016)	(0.015)
Private Motorcycle		-0.108***	-0.089***	-0.085***
		(0.014)	(0.015)	(0.014)
Non-Motorized Transit		-0.119***	-0.089***	-0.034*
		(0.019)	(0.021)	(0.020)
To School		-0.093***	-0.088***	-0.003
		(0.005)	(0.007)	(0.005)
From Work		0.003	0.040***	0.063***
		(0.006)	(0.007)	(0.006)
From School		-0.044***	-0.016*	0.073***
		(0.007)	(0.008)	(0.006)
N	1137900	1137900	1137900	1137900
Adjusted R^2	0.236	0.268	0.315	0.447
Adjusted R^2 (Within)			0.216	0.032
Departure Hour FE		Yes	Yes	Yes
Origin FE			Yes	
Destination FE			Yes	
Origin × Destination FE				Yes

Notes: This table reports the results of a regression of log travel times on trip characteristics, pooling the 2002 HVS / 2010 CS trip data. Column 1 is the unadjusted comparison, including only distance and a 2010 year dummy. Column 2 includes several different trip characteristics (with coefficients reported), while Column 3 includes separate origin and destination fixed effects. Column 4 includes fixed effects for origin-by-destination pairs; identification of the distance coefficient comes from variation in trip distances within an origin-destination route. All columns include separate purpose-by-year effects and separate indicators for each possible departure hour. Robust standard errors, two-way clustered by origin and destination community, are reported in parentheses. */**/*** denotes significance at the 10% / 5% / 1% levels.

Figure C.1: Employment Shares and Population Growth by Community

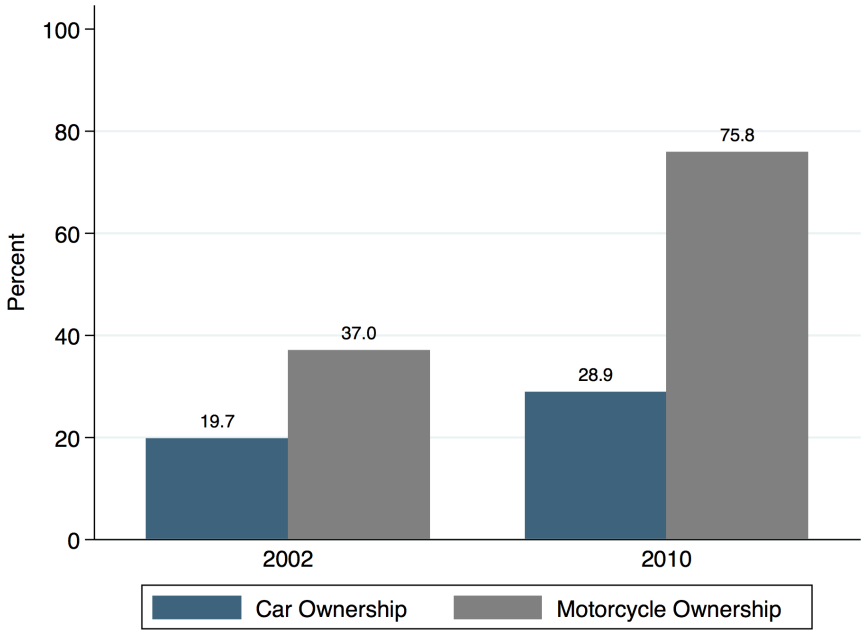
(A) Employment Share (2010) **(B)** Population Growth (2000-2010)



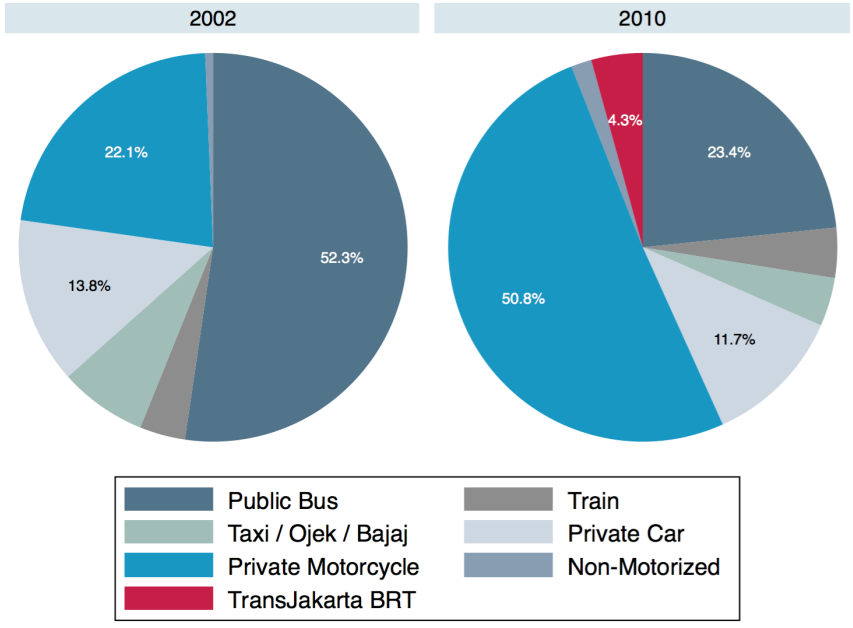
Notes: Authors' calculations, using data from the 2010 CS data in Panel A and the 2000 and 2010 population censuses in Panel B. Darker shading corresponds to greater unconditional employment probabilities (Panel A) and more rapid population growth, in percent changes (Panel B). The thick dark border denotes the boundaries of DKI Jakarta, the special capital region.

Figure C.2: Changes in Vehicle Ownership and Mode Choice

(A) PANEL A: VEHICLE OWNERSHIP



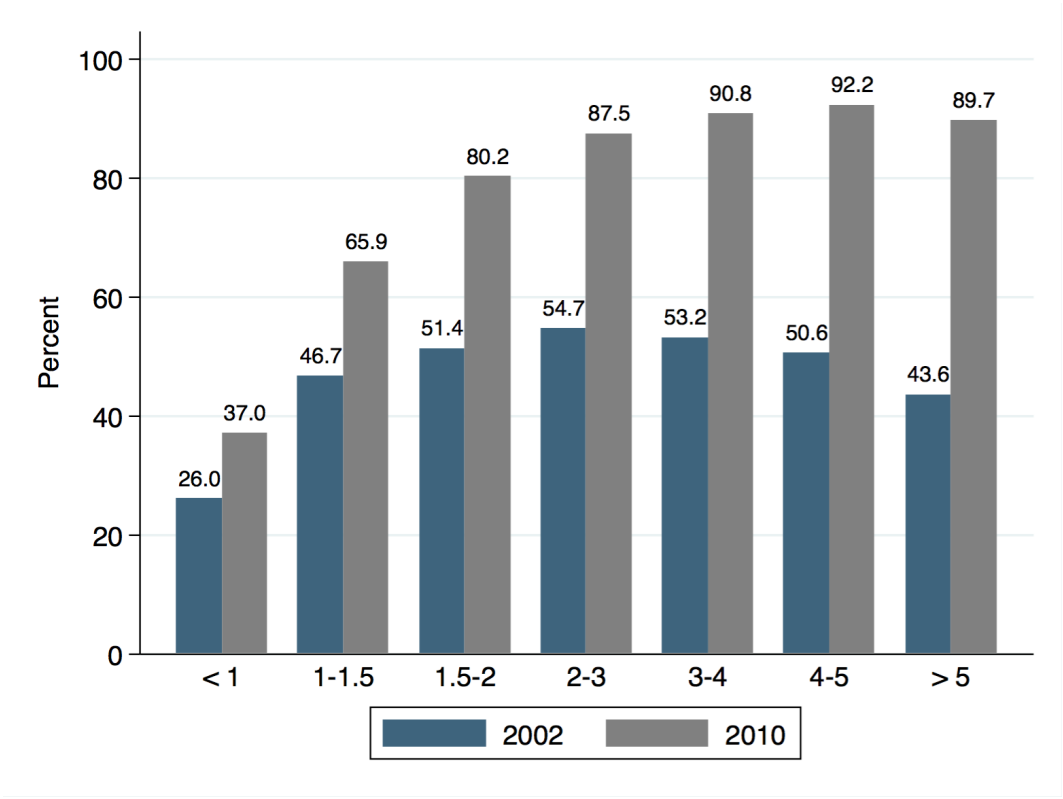
(B) PANEL B: MODE CHOICE



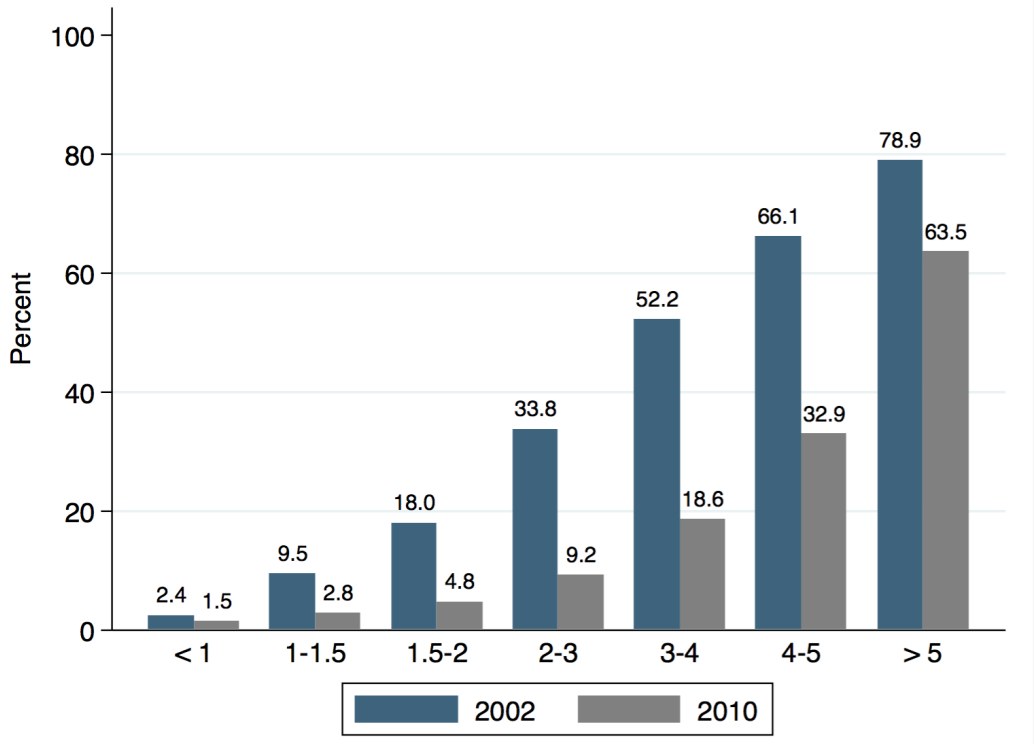
Notes: Authors' calculations, using data from the 2002 and 2010 JICA surveys. All percentages are calculated using survey weights.

Figure C.3: Vehicle Ownership by Income, 2002-2010

(A) MOTORCYCLES

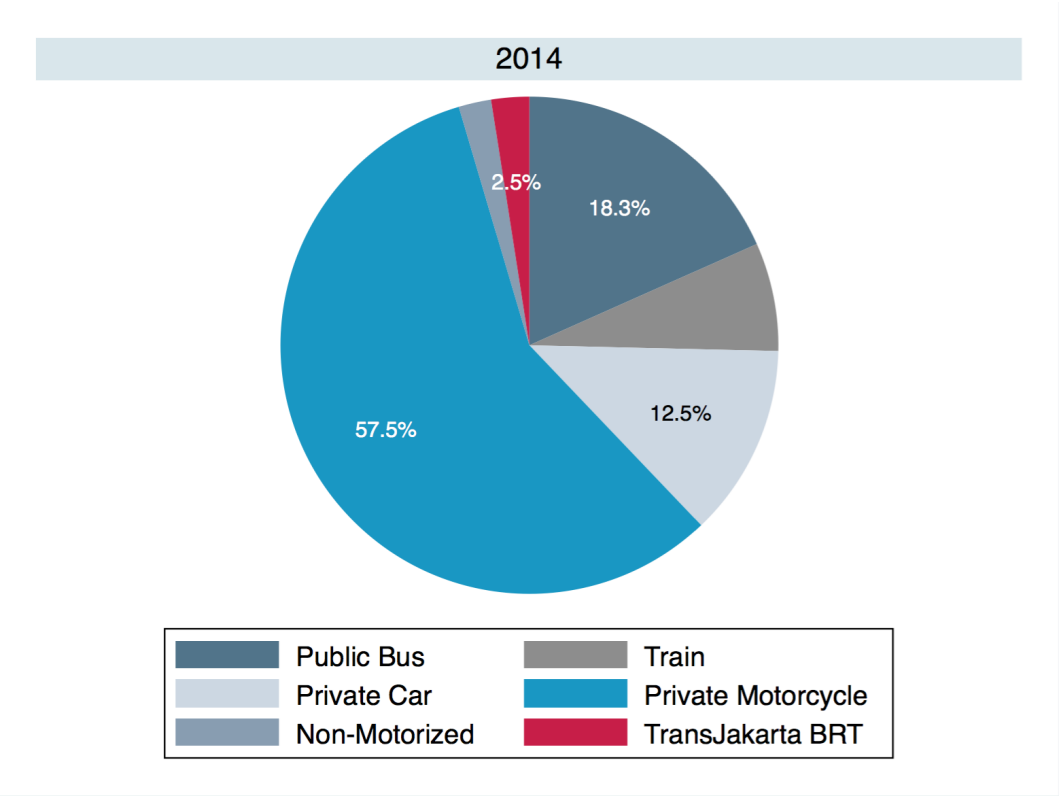


(B) CARS



Notes: This figure plots the share of households owning different types of motor vehicles by nominal income (in millions of Indonesian Rupiah) for 2002 and 2010, based on the JICA data.

Figure C.4: Mode Choice in 2014



Notes: Authors calculations, based on BPS (2014). Note that these data are not completely comparable to those reported in Figure C.2, Panel B, because BPS (2014) did not report the use of taxi/ojek/bajaj separately in tabulating the results of the 2014 Jakarta Commuter Survey.

D ATT Effects of Station Proximity: Demographics and Housing

In Appendix Table D.1, we report similar results of the effects of TransJakarta BRT station proximity on demographic and housing outcomes. We find no significant effects on changes in population density or average years of schooling. We also find no significant effects on changes in total employment, where employment is measured from the 2002 and 2010 JICA data as estimates of the total number of people who work in each community. The estimates on changes in the shares of recent migrants are also insignificant, though the results are fairly imprecisely estimated. The growth of lower-middle income residents (monthly incomes between 1 and 1.5 million Rp.) increased by 4.3 percent in response to the BRT system, but the effects of the BRT system on the growth of the other income shares were insignificantly different from zero.

To construct housing outcomes, we use the 2000 and 2010 Census data, which include information on the household's building type. These building types are aggregated into three categories: (1) single family homes; (2) multi-family units (containing 2-4 households); or (3) larger residential structure with 5 or more households (which we term a "high rise" residence). Housing characteristics were also recorded, including whether or not the home had access to piped water, its own or a shared toilet, and state-provided electricity.

We find that almost treated areas experienced a 14.1 percent increase in residential building construction growth relative to treated areas. This growth seems strongest for single family homes, as there were no differences larger multi-family or high rise residential structures, and on the whole no differences in population density. This evidence suggests that instead of attracting higher density, transit-oriented development in areas near stations, lower density sprawl seems to have taken place in non-treated areas.

These results on demographic and housing outcomes survive a similar set of robustness checks as do our main results on mode choice and vehicle ownership (results available upon request).

Table D.1: ATT Estimates of BRT Proximity: Demographic, Employment, and Housing

	Treated vs. Planned			
	(1)	(2)	(3)	(4)
Δ population density	-0.123**	-0.002	0.025	-0.003
	(0.051)	(0.028)	(0.029)	(0.039)
Δ Log Total Employment	-0.167*	-0.082	-0.168	-0.028
	(0.095)	(0.117)	(0.151)	(0.129)
Δ % recent migrants from w/in Jakarta	4.572***	0.288	0.856	0.870
	(1.473)	(0.796)	(0.763)	(1.164)
Δ % recent migrants from outside Jakarta	4.111***	0.086	0.827	0.299
	(1.384)	(0.787)	(0.664)	(1.042)
Δ Average Years of Schooling	-0.282***	0.068*	0.052	0.072
	(0.096)	(0.038)	(0.049)	(0.064)
Δ Log Number of Residential Buildings	-0.475***	-0.137*	-0.008	-0.141*
	(0.115)	(0.069)	(0.028)	(0.082)
Δ Log Number of Single Family Buildings	-1.461***	-0.612**	-0.107*	-0.434
	(0.516)	(0.301)	(0.065)	(0.331)
Δ Log Number of Multi-Family Buildings	-0.702***	0.009	0.051	-0.016
	(0.182)	(0.103)	(0.048)	(0.105)
Δ Log Number of High Rise Buildings	0.066	0.017	-0.015	-0.040
	(0.131)	(0.089)	(0.130)	(0.130)
Δ % Buildings with clean water	-0.003	0.013	0.020	-0.004
	(0.025)	(0.017)	(0.016)	(0.019)
Δ % Buildings with electricity	-0.009	0.008**	0.005*	0.006
	(0.006)	(0.004)	(0.003)	(0.005)
Δ % Buildings with own toilet	-0.042*	0.003	0.004	-0.009
	(0.023)	(0.009)	(0.009)	(0.016)
Monthly income < Rp. 1 mil, Delta	-0.032	0.015*	0.014	0.010
	(0.024)	(0.009)	(0.009)	(0.011)
Monthly income Rp. 1-1.5 mil, Delta	0.030	0.045**	0.047***	0.043*
	(0.019)	(0.019)	(0.017)	(0.022)
Monthly income Rp. 1.5-2 mil, Delta	-0.028*	-0.024	-0.028	-0.024
	(0.017)	(0.018)	(0.022)	(0.022)
Monthly income Rp. 2-3 mil, Delta	-0.011	0.001	0.014	0.017
	(0.018)	(0.015)	(0.017)	(0.018)
Monthly income Rp. 3-4 mil, Delta	0.006	-0.006	-0.003	-0.008
	(0.012)	(0.014)	(0.012)	(0.015)
Monthly income Rp. 4-5 mil, Delta	0.015	-0.008	-0.006	0.000
	(0.010)	(0.011)	(0.018)	(0.013)
Monthly income > Rp. 5 mil, Delta	0.025	-0.019	-0.034	-0.025
	(0.021)	(0.022)	(0.023)	(0.024)
N	241	241	241	241
Controls	.	X	X	X
Logistic Reweighting	.	.	X	.
Oaxaca-Blinder	.	.	.	X

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column) on an indicator for whether or not the community is within 1 km of a BRT station. Columns 1-4 restrict the non-treated sample to include only almost-treated communities. Column 2 includes pre-treatment controls, and Columns 3 reports a double-robust specification that both includes controls and reweights non-treated communities by $\hat{\kappa} = \hat{P}/(1-\hat{P})$, where $\hat{P}$ is the estimated probability that the community is within 1 km of a BRT station. Columns 4 reports a control function specification based on a Oaxaca-Blinder decomposition, described in Kline (2011). Robust standard errors, clustered at the sub-district level, are reported in parentheses and are estimated using a bootstrap procedure, with 1000 replications, in column 3 to account for the generated $\hat{\kappa}$ weights. Sample sizes vary slightly across outcomes but include as many 132 "treated" community and 109 "almost-treated" community. */**/*** denotes significant at the 10% / 5% / 1% levels.

E Model Appendix

In this appendix, we present our modest extension to the quantitative spatial framework introduced by Allen and Arkolakis (2020) to allow for multiple congestible transit networks, generalizing the two-network case presented in Fan et al. (2021). The model is presented here in detail, elaborating on the treatment discussed in the text. Derivations to many of the key results can be found below in Appendix Section E.

E.1 Setup: Individuals and Routes

Let $\nu \in [0,1]$ index individuals, and let $i \in \{1,2,...,N\}$ index locations (sub-districts, or *kecamatan*) in Jabodetabek. The city has an exogenous (fixed) aggregate population of commuters, given by $\overline{L}$. An individual who lives in location i, works in block j, and commutes via route r of length K_r has the following indirect utility:

$$V_{ij,r}(\nu) = \left[\frac{u_i w_j}{\prod_{l=1}^{K_r} t_{r_{l-1},r_l}\left(m_{r_{l-1},r_l}\right)}\right] \varepsilon_{ij,r}(\nu) \tag{E.1}$$

In this expression, u_i denotes the amenity value of residential location i, w_j denotes the wage at workplace location j, and $\varepsilon_{ij,r}(\nu)$ is a Fréchet distributed idiosyncratic preference shock, specific to each location's origin, destination, and route choice, with shape parameter $\theta > 0$.[62] The shape parameter governs commuter heterogeneity; as θ grows large, commuters become less differentiated and make more similar choices given the same values of u_i and w_j. Individuals choose where to live, work, and which route to take to work to maximize $V_{ij,r}(\nu)$.

Locations in the city are arranged on multiple transport networks, and workers use these networks for commuting between different locations. Let $m = 1,...,M$ index the different transport networks available for commuting in the city, and let $t_{kl}(m) \geq 1$ measure the disutility of moving directly from location k to location l using network m. In our application, we assume that Greater Jakarta has three transport networks: (1) BRT; (2) trains; and (3) surface (used by motorcycles, cars, and traditional public buses), so that $M = 3$, but we present results for the general case of M transport networks.[63] As we will see, $t_{kl}(m)$ will be determined in equilibrium and will depend on the route and location choices of all workers in the model. We can collect these equilibrium iceberg transportation costs in a series of $(N \times N)$ transportation matrices, defined as follows:

$$\mathbf{T}_1 = [t_{kl}(1)] \quad \mathbf{T}_2 = [t_{kl}(2)] \quad ... \quad \mathbf{T}_M = [t_{kl}(M)]$$

If no direct link exists for network m, we set $t_{kl}(m) = \infty$ for all k and m, and to exclude self loops, we also set $t_{kk}(m) = \infty$ for all m.

A route from i to j represents a sequence of edges along different networks that link origin i to destination j. For example, a commuter from i to j may take the following route r on their morning commute:

$$i \overset{\text{surface}}{\longrightarrow} k \overset{\text{BRT}}{\longrightarrow} l \overset{\text{BRT}}{\longrightarrow} n \overset{\text{surface}}{\longrightarrow} j$$

In this example, the length of route r, denoted by K_r, is equal to 4. Transport costs are multiplicative, so the total cost along this path is given by:

$$TC(r) = t_{ik}(\text{surface}) \times t_{kl}(\text{BRT}) \times t_{ln}(\text{BRT}) \times t_{nj}(\text{surface})$$

[62] The CDF of $\varepsilon_{ij,r}(\nu)$ is given by:

$$F(\varepsilon_{ij,r}) = Pr\{\varepsilon \leq \varepsilon_{ij,r}\} = \exp\left\{-\varepsilon_{ij,r}^{-\theta}\right\}$$

[63] This approach ignores different mode choices within the surface network, and it also assumes that traffic congestion depends on the total volume of commuters on the surface network, and not on the mix of vehicles.

For a more general route r of length K_r, the total cost from commuting is given by:

$$TC(r) = \prod_{l=1}^{K_r} t_{r_{l-1},r_l} \left(m_{r_{l-1},r_l}\right)$$

where m_{r_{l-1},r_l} indexes the network chosen on link l of route r. Let $\mathcal{R}_{ij}$ denote the (countably infinite) set of all of possible routes from i to j.

E.2 Commuter Flows and Welfare

From properties of the Fréchet distribution, we can show that the probability that a worker choses to live in i, work in j and commute by $r \in \mathcal{R}_{ij}$ is given by:

$$\pi_{ij,r} = \frac{u_i^\theta w_j^\theta \left(\prod_{l=1}^{K_r} t_{r_{l-1},r_l} \left(m_{r_{l-1},r_l}\right)^{-\theta}\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{r'\in\mathcal{R}_{ij}} u_i^\theta w_j^\theta \left(\prod_{l=1}^{K_{r'}} t_{r'_{l-1},r'_l} \left(m_{r'_{l-1},r'_l}\right)^{-\theta}\right)} \tag{E.2}$$

By summing (E.2) across all routes, we can show that the total number of workers who live in i and work in j is given by:

$$L_{ij} = u_i^\theta w_j^\theta \tau_{ij}^{-\theta} \overline{L}\, \overline{W}^{-\theta} \tag{E.3}$$

In this equation, τ_{ij} measures the (endogenous) transportation costs from i to j, given by:

$$\tau_{ij} \equiv \left[\sum_{r\in\mathcal{R}_{ij}} \left(\prod_{l=1}^{K_r} t_{r_{l-1},r_l} \left(m_{r_{l-1},r_l}\right)^{-\theta}\right)\right]^{-\frac{1}{\theta}} \tag{E.4}$$

This a sum of the total costs of commuting across all routes from i to j, where those costs are determined in equilibrium by each of the least cost routing choices made by different workers in the city. In (E.3), $\overline{W}$ is also the expected welfare of a resident in the city, given by:

$$\overline{W} = \mathbb{E}\left[\max_{i,j,r} V_{ij,r}\left(\nu\right)\right] = \left(\sum_{i=1}^{N}\sum_{j=1}^{N} \tau_{ij}^{-\theta} u_i^\theta w_j^\theta\right)^{\frac{1}{\theta}} \tag{E.5}$$

Note that equation (E.3) shows that the model admits a gravity model formulation of bilateral commuting flows, as in McDonald and McMillen (2010) and Ahlfeldt et al. (2015).[64]

E.3 Wages, Productivity, and Residential Amenities

We assume that each location j produces a homogenous product that can be costlessly traded between locations in the city. Labor is the only input into production, and production takes place under perfect competition and constant returns to scale at the level of the firm. We also assume that the productivity of location i is given by:

$$A_i = \overline{A}_i \left(L_i^F\right)^\alpha \tag{E.6}$$

Here, $\overline{A}_i$ measures location fundamentals that contribute to productivity in i, and L_i^F measures the total number of workers who work in location i, so that α measures the strength of (net) local agglomeration externalities. Because of perfect competition, workers are paid their marginal product, and in equilibrium, $w_i = A_i$.

[64] See Appendix E.11.1 for a derivation of (E.3).

Similarly, we also assume that residential amenities are affected by the size of the residential population, as follows:

$$u_i = \overline{u}_i \left(L_i^R\right)^{\beta} \tag{E.7}$$

where L_i^R measures the total number of residents in location i, $\overline{u}_i$ measures exogenous amenities in location i, and β measures the size of (net) residential externalities. Note that α and β could both be negative because there is no direct land consumption in the model.

E.4 Characterizing Equilibrium

To characterize the model's equilibrium, we need a few tools. First, we derive a matrix representation for the equilibrium transport costs from i to j, which will be convenient to work with analytically. Next, we explain how the model's market clearing conditions can be used to express the gravity commuting flow equations, (E.3), in market access terms. Then, we develop an expression for how frequently links are used in equilibrium, and we use that to characterize traffic patterns.

Matrix Representation of Transportation Costs. In Appendix Section E.11.2, we prove that we can express transport costs from i to j, τ_{ij}, as follows:

$$\tau_{ij}^{-\theta} = \sum_{K=0}^{\infty} A_{ij}^K \tag{E.8}$$

where $\mathbf{A} = [A_{ij}]$ is an $(N \times N)$ sum of weighted adjacency matrices, given by:

$$\mathbf{A} = \left[\sum_{m=1}^{M} t_{ij}(m)^{-\theta}\right] \tag{E.9}$$

The $(i,j)^{\text{th}}$ element of $\mathbf{A}$ is equal to the sum across networks of the transport cost of moving from i to j along network m, raised to the power $-\theta$ if a direct link exists. Otherwise, entries are equal to zeros. The proof of (E.8) uses induction, following Fan et al. (2021) who focus on the case of two networks.

As long as the spectral radius of $\mathbf{A}$ is less than 1, we can write this geometric sum as follows:

$$\sum_{k=0}^{\infty} \mathbf{A}^K = (\mathbf{I} - \mathbf{A})^{-1} \equiv \mathbf{B}$$

where $\mathbf{B} = [b_{ij}]$ is the Leontief inverse of the sum of the weighted adjacency matrices. This allows us to express τ_{ij} in a parsimonious form: $\tau_{ij}^{-\theta} = b_{ij}$. Allen and Arkolakis (2020) note that the spectral radius of $\mathbf{A}$ will be less than 1 if $\tau_j t_{ij}^{-\theta} < 1$ for all i. This condition will be satisfied when either of the following conditions hold: (1) if transport costs between locations are sufficiently large; or (2) if θ is sufficiently large; or (3) if the adjacency matrix has many infinite elements (so that most locations are only indirectly connected).

Market Access. Next, we impose two market clearing conditions to express the gravity commuting flows equation, (E.3), in market access terms. We first require that the total number of residents in location i, L_i^R, is equal to the number of commuters who flow from i. We also require that the number of commuters who work in location j, L_j^F, is equal to the sum of commuters from all other locations to location j:

$$L_i^R = \sum_{j=1}^{N} L_{ij} \qquad L_j^F = \sum_{i=1}^{N} L_{ij} \tag{E.10}$$

Using (E.10), we can write (E.3) as follows:

$$L_{ij} = \tau_{ij}^{-\theta} \left[\frac{L_i^R}{\Pi_i^{-\theta}} \right] \left[\frac{L_j^F}{P_j^{-\theta}} \right] \tag{E.11}$$

where Π_i measures the inverse of commuting market access that residents in i have to firms in all other locations

$$\Pi_i = \left(\sum_{j=1}^{N} \tau_{ij}^{-\theta} L_j^F P_j^{\theta} \right)^{-\frac{1}{\theta}} = u_i \left[L_i^R \right]^{-\frac{1}{\theta}} \left[\frac{\overline{L}}{\overline{W}^{\theta}} \right]^{\frac{1}{2\theta}} \tag{E.12}$$

and P_j measures the inverse of market access that firms in j have to workers in all other locations

$$P_j = \left(\sum_{i=1}^{N} \tau_{ij}^{-\theta} L_j^R \Pi_i^{\theta} \right)^{-\frac{1}{\theta}} = w_j \left[L_j^F \right]^{-\frac{1}{\theta}} \left[\frac{\overline{L}}{\overline{W}^{\theta}} \right]^{\frac{1}{2\theta}} \tag{E.13}$$

Link Intensity and Traffic Flows. We can define the *link intensity* for a link (k, l) on route (i, j) as the expected number of times that link (k, l) is used in commuting between i and j:

$$\pi_{ij}^{kl} = \sum_{r \in \mathcal{R}_{ij}} \left[\frac{\pi_{ij,r}}{\sum_{r' \in \mathcal{R}_{ij}} \pi_{ij,r'}} \right] n_r^{kl} \tag{E.14}$$

In this expression, the term in brackets is the probability that route r is chosen among all routes linking i to j and n_r^{kl} is the number of times the link (k, l) is used along route r (allowing for the possibility that routes double back).

In Appendix Section E.11.3, we discuss how results in Allen and Arkolakis (2020) can be used to show that π_{ij}^{kl} can be written as:

$$\pi_{ij}^{kl} = \sum_{m=1}^{M} \left(\frac{\tau_{ij}}{\tau_{ik} t_{kl}(m) \tau_{lj}} \right)^{\theta} = \sum_{m=1}^{M} \pi_{ij}^{kl}(m) \tag{E.15}$$

where $\pi_{ij}^{kl}(m) = \left(\tau_{ij} / \tau_{ik} t_{kl}(m) \tau_{lj} \right)^{\theta}$ measures the intensity of using network m for link (k, l) on routes from i to j. Equation (E.15) states that the total link intensity for (k, l) along routes from i to j can be written as the sum of network-specific link intensities. The more out of the way the transportation link (k, l) is from the optimal path between i and j, the greater the denominator (which measures optimally going from i to k, going k to l using route m, and going optimally from l to j), and the less the link will be used.

The total traffic over link (k, l) across network m is defined as:

$$\Xi_{kl}(m) \equiv \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{r \in \mathcal{R}_{ij}} \pi_{ij,r} n_r^{k,l} \overline{L} \tag{E.16}$$

This is simply equal to the number of commuters who use link (k, l) along a route from i to j, summed up over all bilateral origin and destination pairs and routes connecting them. Using the link intensity expression (E.15) and substituting in the market access version of the gravity relationship between commuting flows, (E.11), we can show that the total traffic over link (k, l) across all networks can be written as:

$$\Xi_{kl} = \sum_{m=1}^{M} t_{kl}(m)^{-\theta} P_k^{-\theta} \Pi_l^{-\theta} = \sum_{m=1}^{M} \Xi_{kl}(m) \tag{E.17}$$

Here, Ξ_{kl} follows a gravity equation, and traffic flows are determined by economic conditions at the beginning and

the end of the link, as reflected in the market access terms.[65] As commuter market access for location l and firm market access to location k increase, traffic also increases. Equation (E.17) also shows that the total traffic along link (k,l) can be written as the sum of traffic on (k,l) across networks, where network-specific traffic is given by $\Xi_{kl}(m) = t_{kl}(m)^{-\theta} P_k^{-\theta} \Pi_l^{-\theta}$.

Traffic Congestion. For simplicity, we assume that all networks are congestible. Congestion for public transit modes, for example BRT buses or trains, might occur as the number of riders increases. We also assume that the direct cost of traversing a particular link using network m depends on the total traffic flowing through that link using network m, as follows:

$$t_{kl}(m) = \bar{t}_{kl}(m) \left[\Xi_{kl}(m)\right]^{\lambda} \tag{E.18}$$

where $\bar{t}_{kl}(m)$ measures the exogenous component of transport costs for network m, and $\lambda > 0$ measures the strength of traffic congestion. By substituting in the mode-specific traffic expression from (E.17), we obtain:

$$t_{kl}(m) = \left[\bar{t}_{kl}(m)\right]^{\frac{1}{1+\theta\lambda}} \left[P_k\right]^{\frac{-\theta\lambda}{1+\theta\lambda}} \left[\Pi_l\right]^{\frac{-\theta\lambda}{1+\theta\lambda}} \tag{E.19}$$

This equation explains how the distribution of economic activity affects equilibrium transport costs through traffic congestion. In equilibrium, the cost of traversing link (k,l) grows larger as commuter market access for location l and firm market access to location k increase. The parameter λ affects the strength of these market access forces.

Plugging (E.19) expression back into the network-specific traffic expression from (E.17), we obtain:

$$\Xi_{kl}(m) = \left[\bar{t}_{kl}(m)\right]^{\frac{-\theta}{1+\theta\lambda}} \left[P_k\right]^{\frac{-\theta}{1+\theta\lambda}} \left[\Pi_l\right]^{\frac{-\theta}{1+\theta\lambda}} \tag{E.20}$$

This equation relates the distribution of economic activity to equilibrium traffic flows.

Parameterizing Transport Costs. We assume that for networks $m = 1, 2, 3$, we have:

$$t_{kl}(m) = \left[\text{distance}_{kl}(m) \times \text{speed}_{kl}^{-1}(m)\right]^{\delta_0} \left[\Gamma(m)\right]^{\gamma_m}, \tag{E.21}$$

where δ_0 is the time elasticity of transport costs, and $\Gamma(m) \equiv \exp\{\mathbb{1}_m\}$ is a network-specific indicator, raised to the power γ_m. We set $\delta_0 = 1/\theta$ to imply a distance elasticity for commuting flows of negative one, following Allen and Arkolakis (2020) and a large gravity literature (Disdier and Head, 2008; Chaney, 2018).

We also set $\gamma_3 = 0$ for the surface network ($m = 3$), so that $\Gamma(m)^{\gamma_m}$ measures the disutility of commuting via network m relative to the surface network, holding travel times constant. If BRT or train networks create greater disutility from travel than surface based modes, perhaps because of waiting times or lower comfort, we would expect that $\gamma_1 > 0$ and $\gamma_2 > 0$.[66]

We also assume that $\text{speed}_{kl}^{-1}(m)$ is given by:

$$\text{speed}_{kl}^{-1}(m) = \left[m_0 \times \left[\frac{\Xi_{kl}(m)}{\text{lanes}_{kl}(m)}\right]^{\delta_1} \times \varepsilon_{kl}(m)\right], \tag{E.22}$$

where $\text{lanes}_{kl}(m)$ measures the number of lanes for network m from k to l. Combining (E.21) and (E.22), we obtain

[65] See Appendix Section E.11.4 for more details.

[66] In January 2014, a UN-sponsored survey of TransJakarta BRT riders found that nearly 30 percent of riders considered the BRT buses to either be "uncomfortable" or "very uncomfortable" (Sayeg and Lubis, 2014), and such findings were corroborated in a recent survey of females in DKI Jakarta (Witoelar et al., 2017).

equation (E.18) with

$$\bar{t}_{kl}(m) = \left[\text{distance}_{kl}(m) \times m_0 \times \text{lanes}_{kl}(m)^{-\delta_1} \times \varepsilon_{kl}(m)\right]^{\frac{1}{\theta}} \left[\Gamma(m)\right]^{\gamma_m}$$

and $\lambda = \delta_1/\theta$.

E.5 Equilibrium

Given location fundamentals, $\left\{\overline{A}_i, \overline{u}_i\right\}$ for $i = 1, ..., N$, the aggregate labor endowment, $\overline{L}$, and the exogenous components of transport costs for all networks: $\mathbb{T}_m = \left[\bar{t}_{kl}(m)\right]$ for $m = 1, ..., M$, an *equilibrium* is a distribution of economic activity, $\left\{l_i^F, l_i^R\right\}$ and inverse aggregate welfare, χ, such that:

1. The equilibrium distribution of economic activity ensures that commuting markets clear (i.e. equations (E.10) hold).
2. Given equilibrium transport costs, $\mathbb{T}_m$ for $m = 1, .., M$, agents optimally choose routes through the network:

$$\tau_{ij} = b_{ij}^{-1/\theta} \quad \text{where} \quad \mathbf{B} = [b_{ij}] = (\mathbf{I} - \mathbf{A})^{-1} \quad \text{and} \quad \mathbf{A} = \left[\sum_{m=1}^{M} t_{ij}(m)^{-\theta}\right]$$

3. Given the equilibrium distribution of economic activity, the exogenous components of transport costs for all networks, and agents optimal route choices, equilibrium transport costs are determined by traffic congestion, i.e. $\mathbb{T}_m = [t_{kl}(m)]$ where $t_{kl}(m)$ is given by equation (E.19).

A strictly positive equilibrium is one where $l_i^F > 0$ and $l_i^R > 0$ for all $i = 1, ..., N$. In Appendix Section E.11.6, we show that in equilibrium, the following system of equations will hold:

$$\begin{aligned}\left[l_i^R\right]^{1-\beta\theta} \left[l_i^F\right]^{\frac{(1-\theta\alpha)\theta\lambda}{1+\theta\lambda}} &= \chi \left(\overline{A}_i\right)^{\theta} \left(\overline{u}_i\right)^{\theta} \left(l_i^F\right)^{\frac{\theta(\alpha+\lambda)}{1+\theta\lambda}} \\ &+ \chi^{\frac{\theta\lambda}{1+\theta\lambda}} \sum_{j=1}^{N} \left\{ \left[\overline{A}_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\overline{u}_i\right]^{\theta} \left[\overline{u}_j\right]^{\frac{-\theta}{1+\theta\lambda}} \left[\overline{L}\right]^{\frac{-\lambda\theta}{1+\theta\lambda}} \left[l_j^R\right]^{\frac{1-\beta\theta}{1+\theta\lambda}} \left[\sum_{m=1}^{M} \left[\bar{t}_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right] \right\}\end{aligned} \tag{E.23}$$

$$\begin{aligned}\left[l_i^R\right]^{\frac{\theta\lambda(1-\beta\theta)}{1+\theta\lambda}} \left[l_i^F\right]^{(1-\theta\alpha)} &= \chi \left(\overline{A}_i\right)^{\theta} \left(\overline{u}_i\right)^{\theta} \left(l_i^R\right)^{\frac{\theta(\beta+\lambda)}{1+\theta\lambda}} \\ &+ \chi^{\frac{\theta\lambda}{1+\theta\lambda}} \sum_{j=1}^{N} \left\{ \left[\overline{A}_i\right]^{\theta} \left[\overline{A}_j\right]^{\frac{-\theta}{1+\theta\lambda}} \left[\overline{u}_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\overline{L}\right]^{\frac{-\lambda\theta}{1+\theta\lambda}} \left[l_j^F\right]^{\frac{1-\alpha\theta}{1+\theta\lambda}} \left[\sum_{m=1}^{M} \left[\bar{t}_{ji}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right] \right\}\end{aligned} \tag{E.24}$$

These equations determine the equilibrium distribution of economic activity as a function of the model's parameters, $\{\alpha, \beta, \theta, \lambda\}$, location fundamentals, and the transportation networks, after accounting for optimal routing choices and the resulting traffic patterns and congestion they imply. They are identical to the equilibrium expressions for the quantitative urban model in Allen and Arkolakis (2020) (specifically, equations 30 and 31), except that the transport costs term reflects the sum of costs across all networks, instead of a single network's transport cost.

By Proposition 1 of Allen and Arkolakis (2020), if location fundamentals are positive (i.e. $\overline{A}_i > 0$ and $\overline{u}_i > 0$ for all i), if the city's aggregate labor endowment is positive ($\overline{L} > 0$), and if the infrastructure matrix is such that there is a path with finite costs between any two locations, then an equilibrium will exist, and it will be unique if:

$$\alpha \leq \frac{1}{2}\left(\frac{1}{\theta} - \lambda\right) \qquad \beta \leq \frac{1}{2}\left(\frac{1}{\theta} - \lambda\right) \tag{E.25}$$

These conditions state that productivity and amenities must be weakly net dispersive in order to ensure uniqueness, where the size of that dispersion is regulated by θ and λ.

E.6 Counterfactuals

To determine how the distribution of economic activity and welfare would change under different configurations of the transportation network, we follow the "exact hat algebra" approach (Dekle et al., 2008). Let $\widehat{\cdot}$ denote an operator that indicates the counterfactual change in a variable, so that for example, $\widehat{l_i^R} = l_i^{R'}/l_i^R$, where l_i^R measures the actual residential population share, and $l_i^{R'}$ measures the counterfactual residential population share. In Appendix E.11.7, we show how to express the system of equilibrium equations, (E.23) and (E.24), in changes form, as follows:

$$\left[\widehat{l_i^F}\right]^{\frac{\theta\lambda(1-\theta\alpha)}{1+\theta\lambda}}\left[\widehat{l_i^R}\right]^{1-\theta\beta} = \widehat{\chi}\left[\frac{L_i^F}{L_i^F+\sum_{j=1}^N \Xi_{ij}}\right]\left[\widehat{l_i^F}\right]^{\frac{\theta(\alpha+\lambda)}{1+\theta\lambda}} + \widehat{\chi}^{\frac{\theta\lambda}{1+\theta\lambda}}\sum_{j=1}^N\left(\frac{\Xi_{ij}}{L_i^F+\sum_{j=1}^N \Xi_{ij}}\right)\left(\widehat{\bar{t}_{ij}}\left[\widehat{l_j^R}\right]^{\frac{1-\theta\beta}{1+\theta\lambda}}\right) \tag{E.26}$$

$$\left[\widehat{l_i^F}\right]^{1-\theta\alpha}\left[\widehat{l_i^R}\right]^{\frac{\theta\lambda(1-\theta\beta)}{1+\theta\lambda}} = \widehat{\chi}\left[\frac{L_i^R}{L_i^R+\sum_{j=1}^N \Xi_{ji}}\right]\left[\widehat{l_i^R}\right]^{\frac{\theta(\beta+\lambda)}{1+\theta\lambda}} + \widehat{\chi}^{\frac{\theta\lambda}{1+\theta\lambda}}\sum_{j=1}^N\left(\frac{\Xi_{ji}}{L_i^R+\sum_{j=1}^N \Xi_{ji}}\right)\left(\widehat{\bar{t}_{ji}}\left[\widehat{l_j^F}\right]^{\frac{1-\theta\alpha}{1+\theta\lambda}}\right) \tag{E.27}$$

where $\widehat{\bar{t}_{ij}}$ is defined as follows:

$$\widehat{\bar{t}_{ij}} \equiv \left[\frac{\sum_{m=1}^M \left[\bar{t}'_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}}{\sum_{m=1}^M \left[\bar{t}_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}}\right] \tag{E.28}$$

Note that (E.26) and (E.27) are nearly identical to equations (38) and (39) in Allen and Arkolakis (2020), except for the terms $\widehat{\bar{t}_{ij}}$. These terms account for improvements to potentially multiple transport networks, while Allen and Arkolakis (2020) only consider a single network in their analysis.

E.7 Estimating λ

Our reduced form estimates of the travel time spillovers from the BRT from Table 7 identify what happens when the BRT takes away a lane of traffic from traditional surface based modes. That is, our reduced form regressions estimate the following parameter:

$$\widehat{\beta}_{RF} = \frac{-\partial \ln\left(\text{travel time}_{kl}\right)}{\partial \text{lanes}_{kl}(m)} = \frac{-\partial \ln\left(\text{speed}_{kl}^{-1}(m)\right)}{\partial \text{lanes}_{kl}(m)}$$

where we use a negative sign to indicate that the BRT took away lanes of traffic (as opposed to adding them), and where we express the change in terms of the inverse of speed (because distance is held constant). Taking logs of (E.22), we can write δ_1 as:

$$-\delta_1 = \frac{\partial \ln \text{speed}_{kl}^{-1}(m)}{\partial \ln \text{lanes}_{kl}(m)}$$

We convert the semi-elasticity, δ_1, into the reduced form elasticity, $\widehat{\beta}_{RF}$, as follows:

$$-\delta_1 = \frac{\partial \ln \text{speed}_{kl}^{-1}(m)}{\partial \ln \text{lanes}_{kl}(m)}$$

$$-\delta_1 = \frac{\partial \ln \text{speed}_{kl}^{-1}(m)}{\partial \text{lanes}_{kl}(m)/\text{lanes}_{kl}(m)}$$

$$
\begin{aligned}
-\delta_1 &= \frac{\partial \ln \text{speed}_{kl}^{-1}(m)}{\partial \text{lanes}_{kl}(m)} \text{lanes}_{kl}(m) \\
-\delta_1 &= -\widehat{\beta}_{RF} \times \text{lanes}_{kl}(m) \\
\implies \quad \delta_1 &= \widehat{\beta}_{RF} \times \text{lanes}_{kl}(m)
\end{aligned}
$$

Evaluating $\delta_1 = \widehat{\beta}_{RF} \times \text{lanes}_{kl}(m)$ at a mean $\text{lanes}_{kl}(m)$ of 3 (as discussed in Section 2 and presented in Appendix Figure A.1), and recalling that $\delta_1 = \theta\lambda$, our estimate of λ is given by:

$$
\lambda = \frac{3\widehat{\beta}_{RF}}{\theta}
$$

Our baseline estimate of $\lambda = 0.055$ implies that a 10 percent increase in traffic increases transport costs by 0.5 percent.[67] This estimate is comparable to the value of $\lambda = 0.071$ used by Allen and Arkolakis (2020). This parameter will vary depending on the value of θ we work with, as presented in Appendix Table A.22.

E.8 Estimating γ_1 and γ_2

In this subsection, we describe how to derive equation (15) in the paper, which enables us to obtain estimates of γ_1 and γ_2. Note that we can express the probability that network m is chosen on the link (k, l) through the route from i to j as a ratio of network-specific link intensities, from (E.15):

$$
\begin{aligned}
p_{ij}^{kl}(m) &= \frac{\pi_{ij}^{kl}(m)}{\sum_{s=1}^{M} \pi_{ij}^{kl}(s)} \\
&= \frac{\left(\frac{\tau_{ij}}{\tau_{ik} t_{kl}(m) \tau_{lj}}\right)^{\theta}}{\sum_{s=1}^{M} \left(\frac{\tau_{ij}}{\tau_{ik} t_{kl}(s) \tau_{lj}}\right)^{\theta}} \\
&= \frac{t_{kl}(m)^{-\theta}}{\sum_{s=1}^{M} t_{kl}(s)^{-\theta}}
\end{aligned}
$$

Substituting in (E.21), we obtain:

$$
\begin{aligned}
p_{ij}^{kl}(m) &= \frac{\left[\left[\text{distance}_{kl}(m) \times \text{speed}_{kl}^{-1}(m)\right]^{\frac{1}{\theta}} \left[\Gamma(m)\right]^{\gamma_m}\right]^{-\theta}}{\sum_{s=1}^{M} \left[\left[\text{distance}_{kl}(m) \times \text{speed}_{kl}^{-1}(s)\right]^{\frac{1}{\theta}} \left[\Gamma(s)\right]^{\gamma_s}\right]^{-\theta}} \\
&= \frac{\text{travel time}_{kl}(m)^{-1} \left[\Gamma(m)\right]^{-\gamma_m \theta}}{\sum_{s=1}^{M} \text{travel time}_{kl}(s)^{-1} \left[\Gamma(s)\right]^{-\gamma_s \theta}}
\end{aligned}
$$

Defining $y_{ij}(m) \equiv \ln\left(p_{ij}^{kl}(m)\right) - \ln\left(p_{ij}^{kl}(\text{surface})\right)$ and plugging in (13), we obtain:

$$
\begin{aligned}
y_{ij}(m) &\equiv \ln\left(p_{ij}^{kl}(m)\right) - \ln\left(p_{ij}^{kl}(\text{surface})\right) \\
&= -\ln(\text{travel time}_{kl}(m)) - \gamma_m \theta \ln\left(\Gamma(m)\right) + \ln(\text{travel time}_{kl}(\text{surface})) + \gamma_3 \theta \ln\left(\Gamma(\text{surface})\right) \\
&= \gamma_3 \theta \ln\left(\Gamma(\text{surface})\right) - \gamma_m \theta \ln\left(\Gamma(m)\right) + \left\{ \ln\left(\text{travel time}_{kl}(\text{surface})\right) - \ln\left(\text{travel time}_{kl}(m)\right) \right\}
\end{aligned}
$$

[67] Note that if traffic and lanes are positively correlated, and if our controls in Table 7 do not fully capture traffic variation, estimates from this procedure should underestimate the congestion elasticity.

For estimation, we proxy $y_{ij}(m)$ using the log of the share of people commuting from i to j who use m as their primary transport network, minus the log of the share of people commuting from i to j who take surface based modes. Instead of measuring travel time differences directly between k and l, we use average speeds by mode from i to j (multiplied by the appropriate distances). We also condition on origin-by-destination fixed effects, so the variation in travel times is due to speeds. Further, recall that $\Gamma(m) = \exp\{\mathbb{1}_m\}$, and that we normalized $\gamma_3 = 0$. So, our regression is the following:

$$\ln(p_{ij}(m)) - \ln(p_{ij}(\text{surface})) = \widetilde{\gamma_m} \ln(\Gamma(m)) + \delta[\ln(\text{travel time}_{ij}(\text{surface})) - \ln(\text{travel time}_{ij}(m))] + \boldsymbol{\theta}_{ij} + \varepsilon_{ijm}$$

where $\widetilde{\gamma_m} \equiv -\theta\gamma_m$, the term $\boldsymbol{\theta}_{ij}$ denotes origin-by-destination fixed effects, and ε_{ijm} is an error reflecting measurement error. On the right hand side, $\ln(\Gamma(m)) \equiv \mathbb{1}_m$, so we simply regress the inverted share on network indicators, network travel time differences, and origin-and-destination fixed effects. This is equation (15) as stated in the main text.

E.9 Approximating Traffic

Note that to calculate $\widetilde{\Xi_{kl}}(m)$ from equation (16) in the paper, we can form the following matrices:

$$\widetilde{\mathbf{T}}(m) \equiv \left[\widetilde{t_{kl}}(m)^{-\theta}\right] \qquad \mathbf{Lb} \equiv \left[L_{ij}(m)/\widetilde{b_{ij}}\right]$$

Then, we can calculate the matrix of traffic flows for network m, $\widetilde{\boldsymbol{\Xi}}(m) \equiv \left[\widetilde{\Xi}_{kl}(m)\right]$ as follows:

$$\widetilde{\boldsymbol{\Xi}}(m) = \widetilde{\mathbf{T}}(m) \odot \left[\widetilde{\mathbf{B}} \cdot \mathbf{Lb} \cdot \widetilde{\mathbf{B}}\right]$$

where $\odot$ indicates the Hadamard (or element-wise) product, and $\cdot$ indicates the usual matrix product. This formula can be verified by simple (but tedious) matrix multiplication. For example, in the (3×3) case, and dropping the network-specific m subscripts, we have:

$$\begin{aligned}
\widetilde{\mathbf{T}} \odot \left[\widetilde{\mathbf{B}} \cdot \mathbf{Lb} \cdot \widetilde{\mathbf{B}}\right] &= \widetilde{\mathbf{T}} \odot \begin{bmatrix} \widetilde{b}_{11} & \widetilde{b}_{12} & \widetilde{b}_{13} \\ \widetilde{b}_{21} & \widetilde{b}_{22} & \widetilde{b}_{23} \\ \widetilde{b}_{31} & \widetilde{b}_{32} & \widetilde{b}_{33} \end{bmatrix} \begin{bmatrix} \frac{L_{11}}{\widetilde{b}_{11}} & \frac{L_{12}}{\widetilde{b}_{12}} & \frac{L_{13}}{\widetilde{b}_{13}} \\ \frac{L_{21}}{\widetilde{b}_{21}} & \frac{L_{22}}{\widetilde{b}_{22}} & \frac{L_{23}}{\widetilde{b}_{23}} \\ \frac{L_{31}}{\widetilde{b}_{31}} & \frac{L_{32}}{\widetilde{b}_{32}} & \frac{L_{33}}{\widetilde{b}_{33}} \end{bmatrix} \begin{bmatrix} \widetilde{b}_{11} & \widetilde{b}_{12} & \widetilde{b}_{13} \\ \widetilde{b}_{21} & \widetilde{b}_{22} & \widetilde{b}_{23} \\ \widetilde{b}_{31} & \widetilde{b}_{32} & \widetilde{b}_{33} \end{bmatrix} \\
&= \widetilde{\mathbf{T}} \odot \begin{bmatrix} \sum_{i=1}^3 \widetilde{b}_{1i}\left(L_{i1}/\widetilde{b}_{i1}\right) & \sum_{i=1}^3 \widetilde{b}_{1i}\left(L_{i2}/\widetilde{b}_{i2}\right) & \sum_{i=1}^3 \widetilde{b}_{1i}\left(L_{i3}/\widetilde{b}_{i3}\right) \\ \sum_{i=1}^3 \widetilde{b}_{2i}\left(L_{i1}/\widetilde{b}_{i1}\right) & \sum_{i=1}^3 \widetilde{b}_{2i}\left(L_{i2}/\widetilde{b}_{i2}\right) & \sum_{i=1}^3 \widetilde{b}_{2i}\left(L_{i3}/\widetilde{b}_{i3}\right) \\ \sum_{i=1}^3 \widetilde{b}_{3i}\left(L_{i1}/\widetilde{b}_{i1}\right) & \sum_{i=1}^3 \widetilde{b}_{3i}\left(L_{i2}/\widetilde{b}_{i2}\right) & \sum_{i=1}^3 \widetilde{b}_{3i}\left(L_{i3}/\widetilde{b}_{i3}\right) \end{bmatrix} \begin{bmatrix} \widetilde{b}_{11} & \widetilde{b}_{12} & \widetilde{b}_{13} \\ \widetilde{b}_{21} & \widetilde{b}_{22} & \widetilde{b}_{23} \\ \widetilde{b}_{31} & \widetilde{b}_{32} & \widetilde{b}_{33} \end{bmatrix} \\
&= \widetilde{\mathbf{T}} \odot \begin{bmatrix} \sum_i \sum_j \widetilde{b}_{1i}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j1} & \sum_i \sum_j \widetilde{b}_{1i}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j2} & \sum_i \sum_j \widetilde{b}_{1i}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j3} \\ \sum_i \sum_j \widetilde{b}_{2i}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j1} & \sum_i \sum_j \widetilde{b}_{2i}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j2} & \sum_i \sum_j \widetilde{b}_{2i}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j3} \\ \sum_i \sum_j \widetilde{b}_{3i}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j1} & \sum_i \sum_j \widetilde{b}_{3i}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j2} & \sum_i \sum_j \widetilde{b}_{3i}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j3} \end{bmatrix} \\
&= \begin{bmatrix} \sum_i \sum_j \widetilde{b}_{1i}t_{11}^{-\theta}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j1} & \sum_i \sum_j \widetilde{b}_{1i}t_{12}^{-\theta}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j2} & \sum_i \sum_j \widetilde{b}_{1i}t_{13}^{-\theta}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j3} \\ \sum_i \sum_j \widetilde{b}_{2i}t_{21}^{-\theta}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j1} & \sum_i \sum_j \widetilde{b}_{2i}t_{22}^{-\theta}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j2} & \sum_i \sum_j \widetilde{b}_{2i}t_{23}^{-\theta}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j3} \\ \sum_i \sum_j \widetilde{b}_{3i}t_{31}^{-\theta}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j1} & \sum_i \sum_j \widetilde{b}_{3i}t_{32}^{-\theta}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j2} & \sum_i \sum_j \widetilde{b}_{3i}t_{33}^{-\theta}\left(L_{ij}/\widetilde{b}_{ij}\right)\widetilde{b}_{j3} \end{bmatrix}
\end{aligned}$$

Close inspection reveals that each element of the resulting matrix is equivalent to the expression from $\widetilde{\Xi_{kl}}(m)$ from equation (16) in the paper.

E.10 Calculating Changes in Delay

To calculate changes in delay for network m from k to l, we first express network-specific traffic from k to l in the changes form, using (E.20):

$$\widehat{\Xi}_{kl}(m) = \widehat{t}_{kl}(m)^{\frac{-\theta}{1+\theta\lambda}} \widehat{P}_k^{\frac{-\theta}{1+\theta\lambda}} \widehat{\Pi}_l^{\frac{-\theta}{1+\theta\lambda}}$$

Plugging in the result from equation (E.45) derived above, we can write:

$$\widehat{\Xi}_{kl}(m) = \widehat{t}_{kl}(m)^{\frac{-\theta}{1+\theta\lambda}} \left[\widehat{l_k}^F\right]^{\frac{1-\theta\alpha}{1+\theta\lambda}} \left[\widehat{l_l}^R\right]^{\frac{1-\theta\beta}{1+\theta\lambda}} \left[\widehat{\overline{W}}\right]^{\frac{\theta}{1+\theta\lambda}} \tag{E.29}$$

This expression can be calculated after solving the exact hat system of equations, (E.26) and (E.27), for changes in residential populations, workplace populations, and welfare. To convert the expression above into changes in speed, we simply substitute (E.29) into an expression for changes in inverse speeds, which can be easily derived from equation (E.22):

$$\widehat{\text{speed}}_{kl}^{-1}(m) = \left[\frac{\widehat{\Xi}_{kl}(m)}{\widehat{\text{lanes}}_{kl}(m)}\right]^{\delta_1}$$

E.11 Model Derivations

These model derivations closely follow arguments presented by Allen and Arkolakis (2020), extending results for a single network to the multiple network case presented above.

E.11.1 Bilateral Commuting Flows

To show (E.3), we sum commuting probabilities across all routes and multiply by the aggregate population:

$$\begin{aligned}
L_{ij} &= \sum_{r\in\mathcal{R}_{ij}} L_{ij,r} = \sum_{r\in\mathcal{R}_{ij}} \pi_{ij,r}\overline{L} \\
&= \frac{u_i^\theta w_j^\theta \overline{L}\left[\sum_{r\in\mathcal{R}_{ij}} \prod_{l=1}^{K_r} t_{r_{l-1},r_l}\left(m_{r_{l-1},r_l}\right)^{-\theta}\right]}{\sum_{i=1}^N\sum_{j=1}^N\left[\sum_{r\in\mathcal{R}_{ij}} \prod_{l=1}^{K_r} t_{r_{l-1},r_l}\left(m_{r_{l-1},r_l}\right)^{-\theta}\right] u_i^\theta w_j^\theta} \\
&= \frac{\overline{L} u_i^\theta w_j^\theta \tau_{ij}^{-\theta}}{\sum_{i=1}^N\sum_{j=1}^N \tau_{ij}^{-\theta} u_i^\theta w_j^\theta} \\
&= \overline{L} u_i^\theta w_j^\theta \tau_{ij}^{-\theta} \overline{W}^{-\theta}
\end{aligned}$$

where $\overline{W}$ is defined in equation (E.5).

E.11.2 Induction Proof for the Matrix Representation of Transport Costs

We want to show that we can write

$$\begin{aligned}
\tau_{ij}^{-\theta} &= \left(\sum_{r\in\mathcal{R}_{ij}}\left[\prod_{l=1}^{K_r} t_{r_{l-1},r_l}^{-\infty}\right]\right) \\
&= \sum_{K=0}^{\infty}\left[\sum_{K_1=1}^{\infty}\sum_{K_2=1}^{\infty}\cdots\sum_{K_{K-1}=1}^{\infty} a_{i,k_1}\cdot a_{k_1,k_2}\cdot\ldots\cdot a_{k_{K-2},k_{K-1}}\cdot a_{k_{K-1},j}\right] \\
&= \sum_{K=0}^{\infty} A_{ij}^K
\end{aligned}$$

where $\mathbf{A}$ is an $(N \times N)$ matrix, given by:

$$\mathbf{A} = \left[\sum_{m=1}^{M} t_{ij}(m)^{-\theta} \right]$$

To do so, note that we have:

$$\tau_{ij}^{-\theta} = \sum_{r \in \mathcal{R}_{ij}} \left(\prod_{l=1}^{K_r} t_{r_{l-1},r_l} \left(m_{r_{l-1},r_l} \right)^{-\theta} \right)$$

Also, define the mode-specific $\mathbf{A}$ matrices:

$$\mathbf{A}_1 = \left[t_{ij}(1)^{-\theta} \right] \quad \mathbf{A}_2 = \left[t_{ij}(2)^{-\theta} \right] \quad \dots \quad \mathbf{A}_M = \left[t_{ij}(M)^{-\theta} \right]$$

Proof. We proceed by induction. Consider the transport cost of moving from o to d along routes with only one network segment. To write this, let $\mathcal{R}_{ij}(K=1)$ denote the set of all routes of length 1 between i and j. Explicitly enumerating all possible routes, we have:

$$\begin{aligned}
\tau_{od,1}^{-\theta} &= \sum_{r \in \mathcal{R}_{ij}(K=1)} \left(\prod_{l=1}^{K=1} t_{i,j} \left(m_{i,j} \right)^{-\theta} \right) \\
&= \tau_{od,1}(1)^{-\theta} + \tau_{od,1}(2)^{-\theta} + ... + \tau_{od,1}(M)^{-\theta} \\
&= \sum_{m=1}^{M} \tau_{od,1}(m)^{-\theta} \\
&= A_{od}
\end{aligned}$$

Now, suppose that we can write:

$$\tau_{od,k}^{-\theta} = \left[\mathbf{A}^k \right]_{od}$$

That is, suppose the cost of travelling from o to d in k steps is given by the (o,d) element of the matrix $\mathbf{A}$ raised to the power k. To find $\tau_{od,k+1}^{-\theta}$, we simply add a trip of length 1 on each separate mode:

$$\begin{aligned}
\tau_{od,k+1}^{-\theta} &= \left[\mathbf{A}^k \mathbf{A}_1 \right]_{od} + \left[\mathbf{A}^k \mathbf{A}_2 \right]_{od} + \ldots + \left[\mathbf{A}^k \mathbf{A}_1 \right]_{od} \\
&= \left[\mathbf{A}^k \left(\mathbf{A}_1 + \mathbf{A}_2 + \ldots + \mathbf{A}_M \right) \right]_{od} \\
&= \left[\mathbf{A}^k \mathbf{A} \right]_{od} \\
&= \left[\mathbf{A}^{k+1} \right]_{od}
\end{aligned}$$

Therefore, by induction, we can write:

$$\tau_{od}^{-\theta} = \sum_{K=0}^{\infty} A_{ij}^K$$

where A_{ij} is the (i,j) the element of $\mathbf{A}$, and

$$\mathbf{A} = \left(\mathbf{A}_1 + \mathbf{A}_2 + \ldots + \mathbf{A}_M \right) = \left[\sum_{m=1}^{M} t_{ij}(m)^{-\theta} \right]$$

□

E.11.3 Link Intensity

In Appendix A.2, Allen and Arkolakis (2020) use a matrix calculus argument to show that the link intensity, π_{ij}^{kl}, can be written as:

$$\pi_{ij}^{kl} = \frac{b_{ik} a_{kl} b_{lj}}{\tau_{ij}^{-\theta}}$$

However, we work with a different definition of $\mathbf{A}$, so our expression is slightly different from their paper. Given our definition of $\mathbf{A}$ from (E.9), we can write this as:

$$\begin{aligned}
\pi_{ij}^{kl} &= \frac{b_{ik} a_{kl} b_{lj}}{\tau_{ij}^{-\theta}} \\
&= \frac{b_{ik} \left[\sum_{m=1}^{M} t_{kl}(m)^{-\theta} \right] b_{lj}}{\tau_{ij}^{-\theta}} \\
&= \sum_{m=1}^{M} \frac{b_{ik} t_{kl}(m)^{-\theta} b_{lj}}{\tau_{ij}^{-\theta}} \\
&= \sum_{m=1}^{M} \frac{\tau_{ik}^{-\theta} t_{kl}(m)^{-\theta} \tau_{lj}^{-\theta}}{\tau_{ij}^{-\theta}} \\
&= \sum_{m=1}^{M} \left(\frac{\tau_{ij}}{\tau_{ik} t_{kl}(m) \tau_{lj}} \right)^{\theta} \\
&= \sum_{m=1}^{M} \pi_{ij}^{kl}(m)
\end{aligned}$$

So, the total link intensity for link (k, l) along routes from i to j can be written as the sum of mode-specific link intensities for link (k, l) along routes from i to j. The more out of the way the transportation link (k, l) is from the optimal path between i and j, the greater the denominator (which measures optimally taking routes from i to k, then using route m to get from k to l, and finally taking an optimal route from l to j). As the denominator grows, π_{ij}^{kl} declines, and usage of the link along routes from i to j decreases.

E.11.4 Traffic

Let Ξ_{kl} denote the total traffic over link (k, l), across all routes. We can write:

$$\begin{aligned}
\Xi_{kl} &\equiv \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_{ij,r} \sum_{r \in \mathcal{R}_{ij}} n_r^{k,l} \overline{L} \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_{ij}^{kl} L_{ij} \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{b_{ik} a_{kl} b_{lj}}{\tau_{ij}^{-\theta}} \right) L_{ij}
\end{aligned}$$

Substituting for L_{ij} using the market access version of the gravity relationship between commuting flows, from equation (E.11), we have:

$$\Xi_{kl} = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{b_{ik} a_{kl} b_{lj}}{\tau_{ij}^{-\theta}} \right) \left(\tau_{ij}^{-\theta} \left[\frac{L_i^R}{\Pi_i^{-\theta}} \right] \left[\frac{L_j^F}{\Pi_j^{-\theta}} \right] \right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{\tau_{ik}^{-\theta}a_{kl}\tau_{lj}^{-\theta}}{\tau_{ij}^{-\theta}}\right)\left(\tau_{ij}^{-\theta}\left[\frac{L_i^R}{\Pi_i^{-\theta}}\right]\left[\frac{L_j^F}{P_j^{-\theta}}\right]\right)\\
&= a_{kl}\underbrace{\left(\sum_{i=1}^{N}\tau_{ik}^{-\theta}\left[\frac{L_i^R}{\Pi_i^{-\theta}}\right]\right)}_{P_k^{-\theta}}\underbrace{\left(\sum_{j=1}^{N}\tau_{lj}^{-\theta}\left[\frac{L_j^F}{P_j^{-\theta}}\right]\right)}_{\Pi_l^{-\theta}}\\
&= a_{kl}P_k^{-\theta}\Pi_l^{-\theta}\\
&= \left[\sum_{m=1}^{M}t_{kl}(m)^{-\theta}\right]P_k^{-\theta}\Pi_l^{-\theta}\\
&= \sum_{m=1}^{M}t_{kl}(m)^{-\theta}P_k^{-\theta}\Pi_l^{-\theta}\\
&= \sum_{m=1}^{M}\Xi_{kl}(m)
\end{aligned}
$$

So, we can write the total traffic along link (k,l) as the sum of traffic on link (k,l) across different modes, where mode-specific traffic, $\Xi_{kl}(m)$, is given by:

$$\Xi_{kl}(m) = t_{kl}(m)^{-\theta}P_k^{-\theta}\Pi_l^{-\theta}$$

E.11.5 Traffic Congestion

We assume that $t_{kl}(m)$ is given by equation (E.18):

$$t_{kl}(m) = \bar{t}_{kl}(m)\left[\Xi_{kl}(m)\right]^{\lambda}$$

Using this relationship, we can write:

$$
\begin{aligned}
t_{kl}(m) &= \bar{t}_{kl}(m)\left[\Xi_{kl}(m)\right]^{\lambda}\\
t_{kl}(m) &= \bar{t}_{kl}(m)\left[t_{kl}(m)^{-\theta}P_k^{-\theta}\Pi_l^{-\theta}\right]^{\lambda}\\
\implies t_{kl}(m)^{1+\theta\lambda} &= \bar{t}_{kl}(m)P_k^{-\theta\lambda}\Pi_l^{-\theta\lambda}
\end{aligned}
$$

So, we have:

$$t_{kl}(m) = \left[\bar{t}_{kl}(m)\right]^{\frac{1}{1+\theta\lambda}}\left[P_k\right]^{\frac{-\theta\lambda}{1+\theta\lambda}}\left[\Pi_l\right]^{\frac{-\theta\lambda}{1+\theta\lambda}}$$

Plugging this expression into the mode-specific traffic formula in equation (E.17), we obtain:

$$\Xi_{kl}(m) = \left[\bar{t}_{kl}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\left[P_k\right]^{\frac{-\theta}{1+\theta\lambda}}\left[\Pi_l\right]^{\frac{-\theta}{1+\theta\lambda}}$$

E.11.6 Equilibrium

To derive the model's equilibrium conditions, we first rewrite workplace and residential populations as shares and express them in terms of transport costs, amenities, and wages. We then use the matrix product expressions for transport costs to derive the equilibrium expressions, equations (E.23) and (E.24).

Rewriting Workplace and Residential Populations as Shares After combining equations (E.6) and (E.7) with the gravity commuting equation (E.11), we can use the market clearing conditions (E.10) to obtain the follow-

ing:

$$\begin{aligned}\left(l_i^R\right)^{1-\beta\theta} &= \chi \sum_{j=1}^{N} \tau_{ij}^{-\theta} \left(\overline{u}_i \overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \\ \left(l_i^F\right)^{1-\alpha\theta} &= \chi \sum_{j=1}^{N} \tau_{ji}^{-\theta} \left(\overline{u}_j \overline{A}_i\right)^{\theta} \left(l_j^R\right)^{\beta\theta}\end{aligned} \tag{E.30}$$

where $\chi = \overline{L}^{(\alpha+\beta)\theta} / \overline{W}^{\theta}$ measures the inverse of aggregate welfare in the system. Given transportation costs $\{\tau_{ij}\}$, productivities $\{\overline{A}_i\}$, and amenities, $\{\overline{u}_i\}$, we can solve the equations in (E.30) to determine the equilibrium distribution of where people live and where they work.

To show (E.30), we substitute the productivity and amenity equations, (E.6) and (E.7), into the gravity commuting equation, (E.3):

$$\begin{aligned}L_i^R = \sum_{j=1}^{N} L_{ij} &= \sum_{j=1}^{N} u_i^{\theta} w_j^{\theta} \tau_{ij}^{-\theta} \overline{L}\, \overline{W}^{-\theta} \\ &= \left[\frac{\overline{L}}{\overline{W}^{\theta}}\right] \sum_{j=1}^{N} \tau_{ij}^{-\theta} \left[\overline{u}_i \left(L_i^R\right)^{\beta}\right]^{\theta} \left[\overline{A}_j \left(L_j^F\right)^{\alpha}\right]^{\theta} \\ &= \left[\frac{\overline{L}}{\overline{W}^{\theta}}\right] \sum_{j=1}^{N} \tau_{ij}^{-\theta} \overline{u}_i^{\theta} \left(L_i^R\right)^{\beta\theta} \overline{A}_j^{\theta} \left(L_j^F\right)^{\alpha\theta} \\ \implies \left(L_i^R\right)^{1-\beta\theta} &= \left[\frac{\overline{L}}{\overline{W}^{\theta}}\right] \sum_{j=1}^{N} \tau_{ij}^{-\theta} \left(\overline{u}_i \overline{A}_j\right)^{\theta} \left(L_j^F\right)^{\alpha\theta}\end{aligned}$$

Now, setting $l_i^R = L_i^R / \overline{L}$ and $l_i^F = L_i^F / \overline{L}$, we can write:

$$\begin{aligned}\left(\overline{L} l_i^R\right)^{1-\beta\theta} &= \left[\frac{\overline{L}}{\overline{W}^{\theta}}\right] \sum_{j=1}^{N} \tau_{ij}^{-\theta} \left(\overline{u}_i \overline{A}_j\right)^{\theta} \left(\overline{L} l_j^F\right)^{\alpha\theta} \\ \left(l_i^R\right)^{1-\beta\theta} &= \left[\frac{\overline{L}^{(\alpha+\beta)\theta}}{\overline{W}^{\theta}}\right] \sum_{j=1}^{N} \tau_{ij}^{-\theta} \left(\overline{u}_i \overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta}\end{aligned}$$

which is the first equation of (E.30) where we set $\chi = \overline{L}^{(\alpha+\beta)\theta} / \overline{W}^{\theta}$. The second equation follows similarly.

Deriving the Equilibrium Equations We proceed by writing the equilibrium conditions characterizing the distribution of economic activity as functions of endogeneous transport costs using a matrix product.

$$\begin{aligned}\left(l_i^R\right)^{1-\beta\theta} &= \chi \sum_{j=1}^{N} \tau_{ij}^{-\theta} \left(\overline{u}_i \overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \\ \implies \quad \left(\overline{u}_i\right)^{-\theta} \left(l_i^R\right)^{1-\beta\theta} &= \chi \sum_{j=1}^{N} \tau_{ij}^{-\theta} \left(\overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta}\end{aligned}$$

This relationship holds for all locations $i = 1, ..., N$, so we can write:

$$\begin{bmatrix} (\overline{u}_1)^{-\theta} \left(l_1^R\right)^{1-\beta\theta} \\ (\overline{u}_2)^{-\theta} \left(l_2^R\right)^{1-\beta\theta} \\ \vdots \\ (\overline{u}_N)^{-\theta} \left(l_N^R\right)^{1-\beta\theta} \end{bmatrix} = \begin{bmatrix} \chi \sum_{j=1}^N \tau_{1j}^{-\theta} \left(\overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \\ \chi \sum_{j=1}^N \tau_{2j}^{-\theta} \left(\overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \\ \vdots \\ \chi \sum_{j=1}^N \tau_{Nj}^{-\theta} \left(\overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \end{bmatrix}$$

$$\begin{bmatrix} (\overline{u}_1)^{-\theta} \left(l_1^R\right)^{1-\beta\theta} \\ (\overline{u}_2)^{-\theta} \left(l_2^R\right)^{1-\beta\theta} \\ \vdots \\ (\overline{u}_N)^{-\theta} \left(l_N^R\right)^{1-\beta\theta} \end{bmatrix} = \chi \begin{bmatrix} \tau_{11}^{-\theta} & \tau_{12}^{-\theta} & \dots & \tau_{1N}^{-\theta} \\ \tau_{21}^{-\theta} & \tau_{22}^{-\theta} & \dots & \tau_{2N}^{-\theta} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{N1}^{-\theta} & \tau_{N2}^{-\theta} & \dots & \tau_{NN}^{-\theta} \end{bmatrix} \begin{bmatrix} \left(\overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \\ \left(\overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \\ \vdots \\ \left(\overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \end{bmatrix}$$

So, we can write this as:

$$\begin{aligned} [\mathbf{uL}_R] &= \chi \mathbf{B} [\mathbf{aL}_F] \\ [\mathbf{uL}_R] &= \chi (\mathbf{I} - \mathbf{A})^{-1} [\mathbf{aL}_F] \\ (\mathbf{I} - \mathbf{A}) [\mathbf{uL}_R] &= \chi [\mathbf{aL}_F] \\ [\mathbf{uL}_R] - \mathbf{A} [\mathbf{uL}_R] &= \chi [\mathbf{aL}_F] \end{aligned}$$

$$\begin{bmatrix} (\overline{u}_1)^{-\theta} \left(l_1^R\right)^{1-\beta\theta} \\ (\overline{u}_2)^{-\theta} \left(l_2^R\right)^{1-\beta\theta} \\ \vdots \\ (\overline{u}_N)^{-\theta} \left(l_N^R\right)^{1-\beta\theta} \end{bmatrix} - \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix} \begin{bmatrix} (\overline{u}_1)^{-\theta} \left(l_1^R\right)^{1-\beta\theta} \\ (\overline{u}_2)^{-\theta} \left(l_2^R\right)^{1-\beta\theta} \\ \vdots \\ (\overline{u}_N)^{-\theta} \left(l_N^R\right)^{1-\beta\theta} \end{bmatrix} = \chi \begin{bmatrix} \left(\overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \\ \left(\overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \\ \vdots \\ \left(\overline{A}_j\right)^{\theta} \left(l_j^F\right)^{\alpha\theta} \end{bmatrix}$$

So, we can write:

$$(\overline{u}_i)^{-\theta} \left(l_i^R\right)^{1-\beta\theta} - \sum_{j=1}^N a_{ij} (\overline{u}_j)^{-\theta} \left(l_j^R\right)^{1-\beta\theta} = \chi \left(\overline{A}_i\right)^{\theta} \left(l_i^F\right)^{\alpha\theta}$$

$$\implies \boxed{(\overline{u}_i)^{-\theta} \left(l_i^R\right)^{1-\beta\theta} = \chi \left(\overline{A}_i\right)^{\theta} \left(l_i^F\right)^{\alpha\theta} + \sum_{j=1}^N a_{ij} (\overline{u}_j)^{-\theta} \left(l_j^R\right)^{1-\beta\theta}} \tag{E.31}$$

Similarly, we can show that:

$$\begin{aligned} \left(l_i^F\right)^{1-\alpha\theta} &= \chi \sum_{j=1}^N \tau_{ji}^{-\theta} \left(\overline{u}_j \overline{A}_i\right)^{\theta} \left(l_j^R\right)^{\beta\theta} \\ \left(\overline{A}_i\right)^{\theta} \left(l_i^F\right)^{1-\alpha\theta} &= \chi \sum_{j=1}^N \tau_{ji}^{-\theta} (\overline{u}_j)^{\theta} \left(l_j^R\right)^{\beta\theta} \\ \implies \quad [\mathbf{aL}_F] &= \chi \mathbf{B}^T [\mathbf{uL}_R] \\ [\mathbf{aL}_F] &= \chi \left(\mathbf{I} - \mathbf{A}^T\right)^{-1} [\mathbf{uL}_R] \\ \left(\mathbf{I} - \mathbf{A}^T\right) [\mathbf{aL}_F] &= \chi [\mathbf{uL}_R] \\ [\mathbf{aL}_F] &= \chi [\mathbf{uL}_R] + \mathbf{A}^T [\mathbf{aL}_F] \end{aligned}$$

$$\implies \begin{bmatrix} \left(\overline{A}_1\right)^{\theta}\left(l_1^F\right)^{1-\alpha\theta} \\ \left(\overline{A}_2\right)^{\theta}\left(l_2^F\right)^{1-\alpha\theta} \\ \vdots \\ \left(\overline{A}_N\right)^{\theta}\left(l_N^F\right)^{1-\alpha\theta} \end{bmatrix} = \chi \begin{bmatrix} \left(\overline{u}_1\right)^{\theta}\left(l_1^R\right)^{\beta\theta} \\ \left(\overline{u}_2\right)^{\theta}\left(l_2^R\right)^{\beta\theta} \\ \vdots \\ \left(\overline{u}_N\right)^{\theta}\left(l_N^R\right)^{\beta\theta} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} & \dots & a_{N1} \\ a_{12} & a_{22} & \dots & a_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1N} & a_{2N} & \dots & a_{NN} \end{bmatrix} \begin{bmatrix} \left(\overline{A}_1\right)^{\theta}\left(l_1^F\right)^{1-\alpha\theta} \\ \left(\overline{A}_2\right)^{\theta}\left(l_2^F\right)^{1-\alpha\theta} \\ \vdots \\ \left(\overline{A}_N\right)^{\theta}\left(l_N^F\right)^{1-\alpha\theta} \end{bmatrix}$$

So, we can write:

$$\boxed{\left(\overline{A}_i\right)^{\theta}\left(l_i^F\right)^{1-\alpha\theta} = \chi\left(\overline{u}_i\right)^{-\theta}\left(l_i^R\right)^{\beta\theta} + \sum_{j=1}^{N} a_{ji}\left(\overline{A}_j\right)^{\theta}\left(l_j^F\right)^{1-\alpha\theta}} \tag{E.32}$$

Now, note that $\mathbf{A} = [a_{ij}]$ with $a_{ij} = \sum_{m=1}^{M} t_{ij}(m)^{-\theta}$. Plugging in the expression for equilibrium travel costs, (E.19), we can write:

$$\begin{aligned} a_{ij} &= \sum_{m=1}^{M} t_{ij}(m)^{-\theta} \\ &= \sum_{m=1}^{M}\left[\left[\overline{t}_{ij}(m)\right]^{\frac{1}{1+\theta\lambda}}\left[P_i\right]^{\frac{-\theta\lambda}{1+\theta\lambda}}\left[\Pi_j\right]^{\frac{-\theta\lambda}{1+\theta\lambda}}\right]^{-\theta} \\ &= \sum_{m=1}^{M}\left[\overline{t}_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\left[P_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}}\left[\Pi_j\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \\ \implies & \boxed{a_{ij} = \left[P_i\Pi_j\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}}\left[\sum_{m=1}^{M}\left[\overline{t}_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right]} \end{aligned} \tag{E.33}$$

Now, we need to solve for P_i and Π_j as functions of data and parameters. Note that:

$$\begin{aligned} P_i &= A_i\left(L_i^F\right)^{-\frac{1}{\theta}}\left[\frac{\overline{L}}{\overline{W}^{\theta}}\right]^{\frac{1}{2\theta}} \\ &= \overline{A}_i\left(L_i^F\right)^{\alpha}\left(L_i^F\right)^{-\frac{1}{\theta}}\left[\frac{\overline{L}}{\overline{W}^{\theta}}\right]^{\frac{1}{2\theta}} \\ &= \overline{A}_i\left[L_i^F\right]^{\frac{(\theta\alpha-1)}{\theta}}\left[\frac{\overline{L}}{\overline{W}^{\theta}}\right]^{\frac{1}{2\theta}} \end{aligned}$$

Also, note that we have:

$$\begin{aligned} \Pi_j &= u_j\left(L_j^R\right)^{-\frac{1}{\theta}}\left[\frac{\overline{L}}{\overline{W}^{\theta}}\right]^{\frac{1}{2\theta}} \\ &= \overline{u}_j\left(L_j^R\right)^{\beta}\left(L_j^R\right)^{-\frac{1}{\theta}}\left[\frac{\overline{L}}{\overline{W}^{\theta}}\right]^{\frac{1}{2\theta}} \\ &= \overline{u}_j\left[L_j^R\right]^{\frac{(\theta\beta-1)}{\theta}}\left[\frac{\overline{L}}{\overline{W}^{\theta}}\right]^{\frac{1}{2\theta}} \end{aligned}$$

This implies that we have:

$$P_i\Pi_j = \left(\overline{A}_i\left[L_i^F\right]^{\frac{(\theta\alpha-1)}{\theta}}\left[\frac{\overline{L}}{\overline{W}^{\theta}}\right]^{\frac{1}{2\theta}}\right)\left(\overline{u}_j\left[L_j^R\right]^{\frac{(\theta\beta-1)}{\theta}}\left[\frac{\overline{L}}{\overline{W}^{\theta}}\right]^{\frac{1}{2\theta}}\right)$$

$$
\begin{aligned}
&= \overline{A}_i \overline{u}_j \left[L_i^F\right]^{\frac{(\theta\alpha-1)}{\theta}} \left[L_j^R\right]^{\frac{(\theta\beta-1)}{\theta}} \left[\frac{\overline{L}}{\overline{W}^\theta}\right]^{\frac{1}{\theta}} \\
&= \overline{A}_i \overline{u}_j \left[\overline{L} \cdot l_i^F\right]^{\frac{(\theta\alpha-1)}{\theta}} \left[\overline{L} \cdot l_j^R\right]^{\frac{(\theta\beta-1)}{\theta}} \left[\frac{\overline{L}}{\overline{W}^\theta}\right]^{\frac{1}{\theta}} \\
&= \overline{A}_i \overline{u}_j \left[\overline{L}\right]^{\frac{(\theta\alpha+\theta\beta-1)}{\theta}} \left[l_i^F\right]^{\frac{(\theta\alpha-1)}{\theta}} \left[l_j^R\right]^{\frac{(\theta\beta-1)}{\theta}} \overline{W}^{-1} \\
\implies \left[P_i \Pi_j\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} &= \left[\overline{A}_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\overline{u}_j\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\overline{L}\right]^{\frac{\theta\lambda(\theta\alpha+\theta\beta-1)}{1+\theta\lambda}} \left[l_i^F\right]^{\frac{(\theta\alpha-1)\theta\lambda}{1+\theta\lambda}} \left[l_j^R\right]^{\frac{(\theta\beta-1)\theta\lambda}{1+\theta\lambda}} \left[\overline{W}\right]^{\frac{-\theta^2\lambda}{1+\theta\lambda}}
\end{aligned}
$$

Plugging this expression into (E.33) above, we have:

$$
a_{ij} = \left[\overline{A}_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\overline{u}_j\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\overline{L}\right]^{\frac{\theta\lambda(\theta\alpha+\theta\beta-1)}{1+\theta\lambda}} \left[l_i^F\right]^{\frac{(\theta\alpha-1)\theta\lambda}{1+\theta\lambda}} \left[l_j^R\right]^{\frac{(\theta\beta-1)\theta\lambda}{1+\theta\lambda}} \left[\overline{W}\right]^{\frac{-\theta^2\lambda}{1+\theta\lambda}} \left[\sum_{m=1}^{M} \left[\overline{t}_{kl}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right]
$$

Plugging this expression into (E.31) and simplifying, we obtain:

$$
\begin{aligned}
(\overline{u}_i)^{-\theta} \left(l_i^R\right)^{1-\beta\theta} &= \chi \left(\overline{A}_i\right)^\theta \left(l_i^F\right)^{\alpha\theta} + \sum_{j=1}^{N} a_{ij} \left(\overline{u}_j\right)^{-\theta} \left(l_j^R\right)^{1-\beta\theta} \\
(\overline{u}_i)^{-\theta} \left(l_i^R\right)^{1-\beta\theta} &= \chi \left(\overline{A}_i\right)^\theta \left(l_i^F\right)^{\alpha\theta} \\
&\quad + \sum_{j=1}^{N} \left\{ \left[\overline{A}_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\overline{u}_j\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\overline{L}\right]^{\frac{\theta\lambda(\theta\alpha+\theta\beta-1)}{1+\theta\lambda}} \left[l_i^F\right]^{\frac{(\theta\alpha-1)\theta\lambda}{1+\theta\lambda}} \left[l_j^R\right]^{\frac{(\theta\beta-1)\theta\lambda}{1+\theta\lambda}} \left[\overline{W}\right]^{\frac{-\theta^2\lambda}{1+\theta\lambda}} \right. \\
&\quad \left. \times \left[\sum_{m=1}^{M} \left[\overline{t}_{kl}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right] \left(\overline{u}_j\right)^{-\theta} \left(l_j^R\right)^{1-\beta\theta} \right\} \\
(\overline{u}_i)^{-\theta} \left(l_i^R\right)^{1-\beta\theta} &= \chi \left(\overline{A}_i\right)^\theta \left(l_i^F\right)^{\alpha\theta} \\
&\quad + \chi^{\frac{\theta\lambda}{1+\theta\lambda}} \sum_{j=1}^{N} \left\{ \left[\overline{A}_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\overline{u}_j\right]^{\frac{-\theta}{1+\theta\lambda}} \left[\overline{L}\right]^{\frac{-\lambda\theta}{1+\theta\lambda}} \left[l_i^F\right]^{\frac{(\theta\alpha-1)\theta\lambda}{1+\theta\lambda}} \left[l_j^R\right]^{\frac{1-\beta\theta}{1+\theta\lambda}} \left[\sum_{m=1}^{M} \left[\overline{t}_{kl}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right] \right\}
\end{aligned}
$$

This implies that we have:

$$
\begin{aligned}
\left[l_i^R\right]^{1-\beta\theta} \left[l_i^F\right]^{\frac{(1-\theta\alpha)\theta\lambda}{1+\theta\lambda}} &= \chi \left(\overline{A}_i\right)^\theta \left(\overline{u}_i\right)^\theta \left(l_i^F\right)^{\frac{\theta(\alpha+\lambda)}{1+\theta\lambda}} \\
&\quad + \chi^{\frac{\theta\lambda}{1+\theta\lambda}} \sum_{j=1}^{N} \left\{ \left[\overline{A}_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left(\overline{u}_i\right)^\theta \left[\overline{u}_j\right]^{\frac{-\theta}{1+\theta\lambda}} \left[\overline{L}\right]^{\frac{-\lambda\theta}{1+\theta\lambda}} \left[l_j^R\right]^{\frac{1-\beta\theta}{1+\theta\lambda}} \left[\sum_{m=1}^{M} \left[\overline{t}_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right] \right\}
\end{aligned}
\tag{E.34}
$$

Similarly, we can repeat the same procedure with (E.32) to show:

$$
\begin{aligned}
\left[l_i^R\right]^{\frac{\theta\lambda(1-\beta\theta)}{1+\theta\lambda}} \left[l_i^F\right]^{(1-\theta\alpha)} &= \chi \left(\overline{A}_i\right)^\theta \left(\overline{u}_i\right)^\theta \left(l_i^R\right)^{\frac{\theta(\beta+\lambda)}{1+\theta\lambda}} \\
&\quad + \chi^{\frac{\theta\lambda}{1+\theta\lambda}} \sum_{j=1}^{N} \left\{ \left[\overline{A}_i\right]^\theta \left[\overline{A}_j\right]^{\frac{-\theta}{1+\theta\lambda}} \left[\overline{u}_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\overline{L}\right]^{\frac{-\lambda\theta}{1+\theta\lambda}} \left[l_j^F\right]^{\frac{1-\alpha\theta}{1+\theta\lambda}} \left[\sum_{m=1}^{M} \left[\overline{t}_{ji}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right] \right\}
\end{aligned}
\tag{E.35}
$$

Equation (E.34) is equation (E.23) above, and equation (E.35) is equation (E.24) above.

E.11.7 Deriving the Counterfactuals: Exact Hat Algebra

Recall from equations (E.10) and (E.11), we have:

$$L_i^R = \sum_{j=1}^{N} L_{ij} \qquad L_j^F = \sum_{i=1}^{N} L_{ij}$$

and

$$L_{ij} = \tau_{ij}^{-\theta} \left[\frac{L_i^R}{\Pi_i^{-\theta}} \right] \left[\frac{L_j^F}{P_j^{-\theta}} \right]$$

Substituting in the expression for L_{ij} into the expression for L_i^R, we have:

$$\begin{aligned}
L_i^R &= \sum_{j=1}^{N} L_{ij} \\
&= \sum_{j=1}^{N} \tau_{ij}^{-\theta} \left[\frac{L_i^R}{\Pi_i^{-\theta}} \right] \left[\frac{L_j^F}{P_j^{-\theta}} \right] \\
&= \sum_{j=1}^{N} b_{ij} \left[\frac{L_i^R}{\Pi_i^{-\theta}} \right] \left[\frac{L_j^F}{P_j^{-\theta}} \right] \\
\implies \Pi_i^{-\theta} &= \sum_{j=1}^{N} b_{ij} \left[\frac{L_j^F}{P_j^{-\theta}} \right] \\
\implies \begin{bmatrix} \Pi_1^{-\theta} \\ \Pi_2^{-\theta} \\ \vdots \\ \Pi_N^{-\theta} \end{bmatrix} &= \begin{bmatrix} b_{11} & b_{21} & \dots & b_{N1} \\ b_{12} & b_{22} & \dots & b_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1N} & b_{2N} & \dots & b_{NN} \end{bmatrix} \begin{bmatrix} L_1^F/P_1^{-\theta} \\ L_2^F/P_2^{-\theta} \\ \vdots \\ L_N^F/P_N^{-\theta} \end{bmatrix} \\
\implies [\mathbf{\Pi}] &= \mathbf{B} \left[\mathbf{L}^F / \mathbf{P} \right] \\
\implies [\mathbf{\Pi}] &= (\mathbf{I} - \mathbf{A})^{-1} \left[\mathbf{L}^F / \mathbf{P} \right] \\
\implies (\mathbf{I} - \mathbf{A}) [\mathbf{\Pi}] &= \left[\mathbf{L}^F / \mathbf{P} \right] \\
\implies [\mathbf{\Pi}] &= \left[\mathbf{L}^F / \mathbf{P} \right] + \mathbf{A} [\mathbf{\Pi}] \\
\implies \begin{bmatrix} \Pi_1^{-\theta} \\ \Pi_2^{-\theta} \\ \vdots \\ \Pi_N^{-\theta} \end{bmatrix} &= \begin{bmatrix} L_1^F/P_1^{-\theta} \\ L_2^F/P_2^{-\theta} \\ \vdots \\ L_N^F/P_N^{-\theta} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} & \dots & a_{N1} \\ a_{12} & a_{22} & \dots & a_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1N} & a_{2N} & \dots & a_{NN} \end{bmatrix} \begin{bmatrix} \Pi_1^{-\theta} \\ \Pi_2^{-\theta} \\ \vdots \\ \Pi_N^{-\theta} \end{bmatrix}
\end{aligned}$$

So, this implies:

$$\boxed{\Pi_i^{-\theta} = \frac{L_i^F}{P_i^{-\theta}} + \sum_{j=1}^{N} a_{ij} \Pi_j^{-\theta}} \tag{E.36}$$

Similarly, we can also show:

$$\begin{aligned}
L_j^F &= \sum_{i=1}^{N} L_{ij} \\
&= \sum_{i=1}^{N} \tau_{ij}^{-\theta} \left[\frac{L_i^R}{\Pi_i^{-\theta}} \right] \left[\frac{L_j^F}{P_j^{-\theta}} \right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} b_{ij} \left[\frac{L_i^R}{\Pi_i^{-\theta}}\right] \left[\frac{L_j^F}{P_j^{-\theta}}\right] \\
\implies P_j^{-\theta} &= \sum_{i=1}^{N} b_{ij} \left[\frac{L_i^R}{\Pi_i^{-\theta}}\right] \\
\implies \begin{bmatrix} P_1^{-\theta} \\ P_2^{-\theta} \\ \vdots \\ P_N^{-\theta} \end{bmatrix} &= \begin{bmatrix} b_{11} & b_{21} & \dots & b_{N1} \\ b_{12} & b_{22} & \dots & b_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1N} & b_{2N} & \dots & b_{NN} \end{bmatrix} \begin{bmatrix} L_1^R/\Pi_1^{-\theta} \\ L_2^R/\Pi_2^{-\theta} \\ \vdots \\ L_N^R/\Pi_N^{-\theta} \end{bmatrix} \\
\implies [\mathbf{P}] &= \mathbf{B}\left[\mathbf{L}^R/\mathbf{\Pi}\right] \\
\implies [\mathbf{P}] &= (\mathbf{I}-\mathbf{A})^{-1}\left[\mathbf{L}^R/\mathbf{\Pi}\right] \\
\implies (\mathbf{I}-\mathbf{A})[\mathbf{P}] &= \left[\mathbf{L}^R/\mathbf{\Pi}\right] \\
\implies [\mathbf{P}] &= \left[\mathbf{L}^R/\mathbf{\Pi}\right] + \mathbf{A}[\mathbf{P}] \\
\implies \begin{bmatrix} P_1^{-\theta} \\ P_2^{-\theta} \\ \vdots \\ P_N^{-\theta} \end{bmatrix} &= \begin{bmatrix} L_1^R/\Pi_1^{-\theta} \\ L_2^R/\Pi_2^{-\theta} \\ \vdots \\ L_N^R/\Pi_N^{-\theta} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} & \dots & a_{N1} \\ a_{12} & a_{22} & \dots & a_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1N} & a_{2N} & \dots & a_{NN} \end{bmatrix} \begin{bmatrix} P_1^{-\theta} \\ P_2^{-\theta} \\ \vdots \\ P_N^{-\theta} \end{bmatrix}
\end{aligned}
$$

So, this implies:

$$
\boxed{P_i^{-\theta} = \frac{L_i^R}{\Pi_i^{-\theta}} + \sum_{j=1}^{N} a_{ji} P_j^{-\theta}} \tag{E.37}
$$

So, we can write:

$$
\begin{aligned}
\widehat{\Pi_i}^{-\theta} &= \frac{\Pi_i^{-\theta'}}{\Pi_i^{-\theta}} \\
&= \frac{\frac{L_i^{F'}}{P_i^{-\theta'}} + \sum_{j=1}^{N} a'_{ij}\Pi_j^{-\theta'}}{\frac{L_i^F}{P_i^{-\theta}} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}} \\
&= \frac{\frac{L_i^{F'}}{P_i^{-\theta'}}}{\frac{L_i^F}{P_i^{-\theta}} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}} + \sum_{j=1}^{N}\left(\frac{a'_{ij}\Pi_j^{-\theta'}}{\frac{L_i^F}{P_i^{-\theta}} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}}\right) \\
&= \left[\frac{L_i^{F'}/P_i^{-\theta'}}{L_i^F/P_i^{-\theta} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}}\right]\left[\frac{L_i^F/P_i^{-\theta}}{L_i^F/P_i^{-\theta}}\right] + \sum_{j=1}^{N}\left(\frac{a'_{ij}\Pi_j^{-\theta'}}{L_i^F/P_i^{-\theta} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}}\right)\left(\frac{a_{ij}\Pi_j^{-\theta}}{a_{ij}\Pi_j^{-\theta}}\right) \\
&= \left[\frac{L_i^F/P_i^{-\theta}}{L_i^F/P_i^{-\theta} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}}\right]\left[\frac{L_i^{F'}/P_i^{-\theta'}}{L_i^F/P_i^{-\theta}}\right] + \sum_{j=1}^{N}\left(\frac{a_{ij}\Pi_j^{-\theta}}{L_i^F/P_i^{-\theta} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}}\right)\left(\frac{a'_{ij}\Pi_j^{-\theta'}}{a_{ij}\Pi_j^{-\theta}}\right) \\
&= \left[\frac{L_i^F/P_i^{-\theta}}{L_i^F/P_i^{-\theta} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}}\right]\left[\frac{\widehat{L_i}^F}{\widehat{P_i}^{-\theta}}\right] + \sum_{j=1}^{N}\left(\frac{a_{ij}\Pi_j^{-\theta}}{L_i^F/P_i^{-\theta} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}}\right)\left(\widehat{a_{ij}\Pi_j}^{-\theta}\right)
\end{aligned}
$$

Recalling that $\Xi_{ij} = a_{ij}\Pi_j^{-\theta}P_i^{-\theta}$, we have:

$$
\widehat{\Pi_i}^{-\theta} = \left[\frac{L_i^F/P_i^{-\theta}}{L_i^F/P_i^{-\theta} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}}\right]\left[\frac{\widehat{L_i}^F}{\widehat{P_i}^{-\theta}}\right] + \sum_{j=1}^{N}\left(\frac{a_{ij}\Pi_j^{-\theta}}{L_i^F/P_i^{-\theta} + \sum_{j=1}^{N} a_{ij}\Pi_j^{-\theta}}\right)\left(\widehat{a_{ij}\Pi_j}^{-\theta}\right)
$$

$$= \left[\frac{L_i^F}{L_i^F + \sum_{j=1}^N a_{ij} P_i^{-\theta} \Pi_j^{-\theta}}\right] \left[\frac{\widehat{L_i}^F}{\widehat{P_i}^{-\theta}}\right] + \sum_{j=1}^N \left(\frac{a_{ij} P_i^{-\theta} \Pi_j^{-\theta}}{L_i^F + \sum_{j=1}^N a_{ij} P_i^{-\theta} \Pi_j^{-\theta}}\right) \left(\widehat{a_{ij}} \widehat{\Pi_j}^{-\theta}\right)$$

$$\boxed{\widehat{\Pi_i}^{-\theta} = \left[\frac{L_i^F}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right] \left[\frac{\widehat{L_i}^F}{\widehat{P_i}^{-\theta}}\right] + \sum_{j=1}^N \left(\frac{\Xi_{ij}}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right) \left(\widehat{a_{ij}} \widehat{\Pi_j}^{-\theta}\right)} \tag{E.38}$$

Following a similar approach for $P_i^{-\theta}$, we have:

$$\begin{aligned}
\widehat{P_i}^{-\theta} &= \frac{P_i^{-\theta'}}{P_i^{-\theta}} \\
&= \frac{\frac{L_i^R}{\Pi_i^{-\theta'}} + \sum_{j=1}^N a'_{ji} P_j^{-\theta'}}{\frac{L_i^R}{\Pi_i^{-\theta}} + \sum_{j=1}^N a_{ji} P_j^{-\theta}} \\
&= \frac{\frac{L_i^{R'}}{\Pi_i^{-\theta'}}}{\frac{L_i^R}{\Pi_i^{-\theta}} + \sum_{j=1}^N a_{ji} P_j^{-\theta}} + \sum_{j=1}^N \left(\frac{a'_{ji} P_j^{-\theta'}}{\frac{L_i^R}{\Pi_i^{-\theta}} + \sum_{j=1}^N a_{ji} P_j^{-\theta}}\right) \\
&= \left[\frac{L_i^{R'}/\Pi_i^{-\theta'}}{L_i^R/\Pi_i^{-\theta} + \sum_{j=1}^N a_{ji} P_j^{-\theta}}\right] \left[\frac{L_i^R/\Pi_i^{-\theta}}{L_i^R/\Pi_i^{-\theta}}\right] + \sum_{j=1}^N \left(\frac{a'_{ij} P_j^{-\theta'}}{L_i^R/\Pi_i^{-\theta} + \sum_{j=1}^N a_{ji} P_j^{-\theta}}\right) \left(\frac{a_{ji} P_j^{-\theta}}{a_{ji} P_j^{-\theta}}\right) \\
&= \left[\frac{L_i^R/\Pi_i^{-\theta}}{L_i^R/\Pi_i^{-\theta} + \sum_{j=1}^N a_{ji} P_j^{-\theta}}\right] \left[\frac{L_i^{R'}/\Pi_i^{-\theta'}}{L_i^R/\Pi_i^{-\theta}}\right] + \sum_{j=1}^N \left(\frac{a_{ji} P_j^{-\theta}}{L_i^R/\Pi_i^{-\theta} + \sum_{j=1}^N a_{ji} P_j^{-\theta}}\right) \left(\frac{a'_{ji} P_j^{-\theta'}}{a_{ji} P_j^{-\theta}}\right) \\
&= \left[\frac{L_i^R/\Pi_i^{-\theta}}{L_i^R/\Pi_i^{-\theta} + \sum_{j=1}^N a_{ji} P_j^{-\theta}}\right] \left[\frac{\widehat{L_i}^R}{\widehat{\Pi_i}^{-\theta}}\right] + \sum_{j=1}^N \left(\frac{a_{ji} P_j^{-\theta}}{L_i^R/\Pi_i^{-\theta} + \sum_{j=1}^N a_{ji} P_j^{-\theta}}\right) \left(\widehat{a_{ji}} \widehat{P_j}^{-\theta}\right) \\
&= \left[\frac{L_i^R}{L_i^R + \sum_{j=1}^N a_{ji} \Pi_i^{-\theta} P_j^{-\theta}}\right] \left[\frac{\widehat{L_i}^R}{\widehat{\Pi_i}^{-\theta}}\right] + \sum_{j=1}^N \left(\frac{a_{ji} \Pi_i^{-\theta} P_j^{-\theta}}{L_i^R + \sum_{j=1}^N a_{ji} \Pi_i^{-\theta} P_j^{-\theta}}\right) \left(\widehat{a_{ji}} \widehat{P_j}^{-\theta}\right)
\end{aligned}$$

$$\boxed{\widehat{P_i}^{-\theta} = \left[\frac{L_i^R}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right] \left[\frac{\widehat{L_i}^R}{\widehat{\Pi_i}^{-\theta}}\right] + \sum_{j=1}^N \left(\frac{\Xi_{ji}}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right) \left(\widehat{a_{ji}} \widehat{P_j}^{-\theta}\right)} \tag{E.39}$$

Now, recall that we have:

$$\begin{aligned}
a_{ij} &= \sum_{m=1}^M t_{ij}(m)^{-\theta} \\
&= \sum_{m=1}^M \left[\left[\bar{t}_{ij}(m)\right]^{\frac{1}{1+\theta\lambda}} \left[P_i\right]^{\frac{-\theta\lambda}{1+\theta\lambda}} \left[\Pi_j\right]^{\frac{-\theta\lambda}{1+\theta\lambda}}\right]^{-\theta} \\
&= \left[P_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\Pi_j\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\sum_{m=1}^M \left[\bar{t}_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right]
\end{aligned}$$

This implies:

$$\begin{aligned}
\widehat{a_{ij}} &\equiv \frac{a'_{ij}}{a_{ij}} \\
&= \frac{\left[P'_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\Pi'_j\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\sum_{m=1}^M \left[\bar{t}'_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right]}{\left[P_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\Pi_h\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\sum_{m=1}^M \left[\bar{t}_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}\right]}
\end{aligned}$$

$$= \left[\widehat{P}_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\widehat{\Pi}_j\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\frac{\sum_{m=1}^{M}\left[\overline{t}'_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}}{\sum_{m=1}^{M}\left[\overline{t}_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}}\right]$$

$$\Longrightarrow \boxed{\widehat{a_{ij}} = \left[\widehat{P}_i\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \left[\widehat{\Pi}_j\right]^{\frac{\theta^2\lambda}{1+\theta\lambda}} \widehat{\overline{t}_{ij}}} \tag{E.40}$$

where $\widehat{\overline{t}_{ij}}$ is defined as follows:

$$\widehat{\overline{t}_{ij}} \equiv \left[\frac{\sum_{m=1}^{M}\left[\overline{t}'_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}}{\sum_{m=1}^{M}\left[\overline{t}_{ij}(m)\right]^{\frac{-\theta}{1+\theta\lambda}}}\right]$$

Substituting equation (E.40) into equation (E.38), we have:

$$\widehat{\Pi}_i^{-\theta} = \left[\frac{L_i^F}{L_i^F + \sum_{j=1}^{N}\Xi_{ij}}\right]\left[\frac{\widehat{L_i^F}}{\widehat{P}_i^{-\theta}}\right] + \sum_{j=1}^{N}\left(\frac{\Xi_{ij}}{L_i^F + \sum_{j=1}^{N}\Xi_{ij}}\right)\left(\widehat{a_{ij}}\widehat{\Pi}_j^{-\theta}\right)$$

$$\Longrightarrow \widehat{P}_i^{-\theta}\widehat{\Pi}_i^{-\theta} = \left[\frac{L_i^F}{L_i^F + \sum_{j=1}^{N}\Xi_{ij}}\right]\widehat{L_i^F} + \sum_{j=1}^{N}\left(\frac{\Xi_{ij}}{L_i^F + \sum_{j=1}^{N}\Xi_{ij}}\right)\left(\widehat{a_{ij}}\widehat{P}_i^{-\theta}\widehat{\Pi}_j^{-\theta}\right)$$

$$\Longrightarrow \boxed{\widehat{P}_i^{-\theta}\widehat{\Pi}_i^{-\theta} = \left[\frac{L_i^F}{L_i^F + \sum_{j=1}^{N}\Xi_{ij}}\right]\widehat{L_i^F} + \sum_{j=1}^{N}\left(\frac{\Xi_{ij}}{L_i^F + \sum_{j=1}^{N}\Xi_{ij}}\right)\left(\widehat{\overline{t}_{ij}}\widehat{P}_i^{\frac{-\theta}{1+\theta\lambda}}\widehat{\Pi}_j^{\frac{-\theta}{1+\theta\lambda}}\right)} \tag{E.41}$$

Similarly, substituting equation (E.40) into equation (E.39), we have:

$$\widehat{P}_i^{-\theta} = \left[\frac{L_i^R}{L_i^R + \sum_{j=1}^{N}\Xi_{ji}}\right]\left[\frac{\widehat{L_i^R}}{\widehat{\Pi}_i^{-\theta}}\right] + \sum_{j=1}^{N}\left(\frac{\Xi_{ji}}{L_i^R + \sum_{j=1}^{N}\Xi_{ji}}\right)\left(\widehat{a_{ji}}\widehat{P_j}^{-\theta}\right)$$

$$\Longrightarrow \widehat{P}_i^{-\theta}\widehat{\Pi}_i^{-\theta} = \left[\frac{L_i^R}{L_i^R + \sum_{j=1}^{N}\Xi_{ji}}\right]\widehat{L_i^R} + \sum_{j=1}^{N}\left(\frac{\Xi_{ji}}{L_i^R + \sum_{j=1}^{N}\Xi_{ji}}\right)\left(\widehat{a_{ji}}\widehat{P_j}^{-\theta}\widehat{\Pi}_i^{-\theta}\right)$$

$$\Longrightarrow \boxed{\widehat{P}_i^{-\theta}\widehat{\Pi}_i^{-\theta} = \left[\frac{L_i^R}{L_i^R + \sum_{j=1}^{N}\Xi_{ji}}\right]\widehat{L_i^R} + \sum_{j=1}^{N}\left(\frac{\Xi_{ji}}{L_i^R + \sum_{j=1}^{N}\Xi_{ji}}\right)\left(\widehat{\overline{t}_{ji}}\widehat{P_j}^{\frac{-\theta}{1+\theta\lambda}}\widehat{\Pi}_i^{\frac{-\theta}{1+\theta\lambda}}\right)} \tag{E.42}$$

Recall that from equation (E.12), we have:

$$\begin{aligned}
\Pi_i &= u_i\left[L_i^R\right]^{\frac{-1}{\theta}}\left[\frac{\overline{L}}{\overline{W}^{\theta}}\right]^{\frac{1}{2\theta}} \\
&= \overline{u}_i\left[L_i^R\right]^{\beta}\left[L_i^R\right]^{\frac{-1}{\theta}}\left[\overline{L}\right]^{\frac{1}{2\theta}}\left[\overline{W}\right]^{\frac{-1}{2}} \\
&= \overline{u}_i\left[l_i^R\right]^{\frac{\theta\beta-1}{\theta}}\left[\overline{L}\right]^{\frac{2\theta\beta-1}{2\theta}}\left[\overline{W}\right]^{\frac{-1}{2}} \\
\Longrightarrow \widehat{\Pi}_i &= \frac{\overline{u}_i\left[l_i^{R\prime}\right]^{\frac{\theta\beta-1}{\theta}}\left[\overline{L}\right]^{\frac{2\theta\beta-1}{2\theta}}\left[\overline{W'}\right]^{\frac{-1}{2}}}{\overline{u}_i\left[l_i^R\right]^{\frac{\theta\beta-1}{\theta}}\left[\overline{L}\right]^{\frac{2\theta\beta-1}{2\theta}}\left[\overline{W}\right]^{\frac{-1}{2}}}
\end{aligned}$$

$$\boxed{\widehat{\Pi}_i = \left[\widehat{l_i^R}\right]^{\frac{\theta\beta-1}{\theta}}\left[\widehat{\overline{W}}\right]^{\frac{-1}{2}}} \tag{E.43}$$

Similarly, recall that from equation (E.13), we have:

$$P_i = A_i\left[L_i^F\right]^{\frac{-1}{\theta}}\left[\frac{\overline{L}}{\overline{W}^{\theta}}\right]^{\frac{1}{2\theta}}$$

$$
\begin{aligned}
&= \overline{A}_i \left[L_i^F\right]^{\alpha} \left[L_i^F\right]^{\frac{-1}{\theta}} \left[L\right]^{\frac{1}{2\theta}} \left[W\right]^{\frac{-1}{2}} \\
&= \overline{A}_i \left[l_i^F\right]^{\frac{\theta\alpha-1}{\theta}} \left[L\right]^{\frac{2\theta\alpha-1}{2\theta}} \left[W\right]^{\frac{-1}{2}} \\
\implies \widehat{P}_i &= \frac{\overline{A}_i \left[l_i^{F'}\right]^{\frac{\theta\alpha-1}{\theta}} \left[L\right]^{\frac{2\theta\alpha-1}{2\theta}} \left[W'\right]^{\frac{-1}{2}}}{\overline{A}_i \left[l_i^F\right]^{\frac{\theta\alpha-1}{\theta}} \left[L\right]^{\frac{2\theta\alpha-1}{2\theta}} \left[W\right]^{\frac{-1}{2}}} \\
&\boxed{\widehat{P}_i = \left[\widehat{l}_i^F\right]^{\frac{\theta\alpha-1}{\theta}} \left[\widehat{W}\right]^{\frac{-1}{2}}}
\end{aligned}
\tag{E.44}
$$

The two boxed expressions above, equations (E.43) and (E.44), imply:

$$
\begin{aligned}
\widehat{\Pi}_i \widehat{P}_i &= \left[\widehat{l}_i^F\right]^{\frac{\theta\alpha-1}{\theta}} \left[\widehat{l}_i^R\right]^{\frac{\theta\beta-1}{\theta}} \left[\widehat{W}\right]^{-1} \\
\implies &\boxed{\widehat{\Pi}_i^{-\theta} \widehat{P}_i^{-\theta} = \left[\widehat{l}_i^F\right]^{1-\theta\alpha} \left[\widehat{l}_i^R\right]^{1-\theta\beta} \left[\widehat{W}\right]^{\theta}} \\
\implies &\boxed{\widehat{\Pi}_i^{\frac{-\theta}{1+\theta\lambda}} \widehat{P}_j^{\frac{-\theta}{1+\theta\lambda}} = \left[\widehat{l}_i^R\right]^{\frac{1-\theta\beta}{1+\theta\lambda}} \left[\widehat{l}_j^F\right]^{\frac{1-\theta\alpha}{1+\theta\lambda}} \left[\widehat{W}\right]^{\frac{\theta}{1+\theta\lambda}}}
\end{aligned}
\tag{E.45}
$$

Further, note that:

$$
\begin{aligned}
\chi &= \left(\frac{\overline{L}^{(\alpha+\beta)}}{\overline{W}}\right)^{\theta} \\
&= \overline{L}^{\theta(\alpha+\beta)} \overline{W}^{-\theta} \\
\implies \widehat{\chi} &= \frac{\overline{W}^{-\theta'}}{\overline{W}^{-\theta}} \\
\implies &\boxed{\widehat{\chi} = \widehat{\overline{W}}^{-\theta}}
\end{aligned}
$$

Plugging the expressions from (E.45) into equation (E.41) and rearranging, we obtain:

$$
\begin{aligned}
\widehat{P}_i^{-\theta} \widehat{\Pi}_i^{-\theta} &= \left[\frac{L_i^F}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right] \widehat{L_i}^F + \sum_{j=1}^N \left(\frac{\Xi_{ij}}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right) \left(\widehat{\overline{t}_{ij}} \widehat{P}_i^{\frac{-\theta}{1+\theta\lambda}} \widehat{\Pi}_j^{\frac{-\theta}{1+\theta\lambda}}\right) \\
\implies \left[\widehat{l}_i^F\right]^{1-\theta\alpha} \left[\widehat{l}_i^R\right]^{1-\theta\beta} \left[\widehat{W}\right]^{\theta} &= \left[\frac{L_i^F}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right] \widehat{l}_i^F + \sum_{j=1}^N \left(\frac{\Xi_{ij}}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right) \left(\widehat{\overline{t}_{ij}} \left[\widehat{l}_j^R\right]^{\frac{1-\theta\beta}{1+\theta\lambda}} \left[\widehat{l}_i^F\right]^{\frac{1-\theta\alpha}{1+\theta\lambda}} \left[\widehat{W}\right]^{\frac{\theta}{1+\theta\lambda}}\right) \\
\implies \left[\widehat{l}_i^F\right]^{1-\theta\alpha} \left[\widehat{l}_i^R\right]^{1-\theta\beta} &= \left[\widehat{W}\right]^{-\theta} \left[\frac{L_i^F}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right] \widehat{l}_i^F + \left[\widehat{W}\right]^{\frac{-\theta^2\lambda}{1+\theta\lambda}} \sum_{j=1}^N \left(\frac{\Xi_{ij}}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right) \left(\widehat{\overline{t}_{ij}} \left[\widehat{l}_j^R\right]^{\frac{1-\theta\beta}{1+\theta\lambda}} \left[\widehat{l}_i^F\right]^{\frac{1-\theta\alpha}{1+\theta\lambda}}\right) \\
\implies \left[\widehat{l}_i^F\right]^{1-\theta\alpha} \left[\widehat{l}_i^R\right]^{1-\theta\beta} &= \widehat{\chi} \left[\frac{L_i^F}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right] \widehat{l}_i^F + \widehat{\chi}^{\frac{\theta\lambda}{1+\theta\lambda}} \sum_{j=1}^N \left(\frac{\Xi_{ij}}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right) \left(\widehat{\overline{t}_{ij}} \left[\widehat{l}_j^R\right]^{\frac{1-\theta\beta}{1+\theta\lambda}} \left[\widehat{l}_i^F\right]^{\frac{1-\theta\alpha}{1+\theta\lambda}}\right) \\
\implies &\boxed{\left[\widehat{l}_i^F\right]^{\frac{\theta\lambda(1-\theta\alpha)}{1+\theta\lambda}} \left[\widehat{l}_i^R\right]^{1-\theta\beta} = \widehat{\chi} \left[\frac{L_i^F}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right] \left[\widehat{l}_i^F\right]^{\frac{\theta(\alpha+\lambda)}{1+\theta\lambda}} + \widehat{\chi}^{\frac{\theta\lambda}{1+\theta\lambda}} \sum_{j=1}^N \left(\frac{\Xi_{ij}}{L_i^F + \sum_{j=1}^N \Xi_{ij}}\right) \left(\widehat{\overline{t}_{ij}} \left[\widehat{l}_j^R\right]^{\frac{1-\theta\beta}{1+\theta\lambda}}\right)}
\end{aligned}
\tag{E.46}
$$

Similarly, plugging the expressions from (E.45) into equation (E.42) and rearranging, we obtain:

$$\widehat{P_i}^{-\theta}\widehat{\Pi_i}^{-\theta} = \left[\frac{L_i^R}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right]\widehat{L_i}^R + \sum_{j=1}^N \left(\frac{\Xi_{ji}}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right)\left(\widehat{t_{ji}}\widehat{P_j}^{\frac{-\theta}{1+\theta\lambda}}\widehat{\Pi_i}^{\frac{-\theta}{1+\theta\lambda}}\right)$$

$$\implies \left[\widehat{l_i}^F\right]^{1-\theta\alpha}\left[\widehat{l_i}^R\right]^{1-\theta\beta}\left[\widehat{W}\right]^{\theta} = \left[\frac{L_i^R}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right]\widehat{l_i}^R + \sum_{j=1}^N \left(\frac{\Xi_{ji}}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right)\left(\widehat{t_{ji}}\left[\widehat{l_i}^R\right]^{\frac{1-\theta\beta}{1+\theta\lambda}}\left[\widehat{l_j}^F\right]^{\frac{1-\theta\alpha}{1+\theta\lambda}}\left[\widehat{W}\right]^{\frac{\theta}{1+\theta\lambda}}\right)$$

$$\implies \left[\widehat{l_i}^F\right]^{1-\theta\alpha}\left[\widehat{l_i}^R\right]^{1-\theta\beta} = \left[\widehat{W}\right]^{-\theta}\left[\frac{L_i^R}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right]\widehat{l_i}^R + \left[\widehat{W}\right]^{\frac{-\theta^2\lambda}{1+\theta\lambda}}\sum_{j=1}^N \left(\frac{\Xi_{ji}}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right)\left(\widehat{t_{ji}}\left[\widehat{l_i}^R\right]^{\frac{1-\theta\beta}{1+\theta\lambda}}\left[\widehat{l_j}^F\right]^{\frac{1-\theta\alpha}{1+\theta\lambda}}\right)$$

$$\implies \left[\widehat{l_i}^F\right]^{1-\theta\alpha}\left[\widehat{l_i}^R\right]^{1-\theta\beta} = \widehat{\chi}\left[\frac{L_i^R}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right]\widehat{l_i}^R + \widehat{\chi}^{\frac{\theta\lambda}{1+\theta\lambda}}\sum_{j=1}^N \left(\frac{\Xi_{ji}}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right)\left(\widehat{t_{ji}}\left[\widehat{l_i}^R\right]^{\frac{1-\theta\beta}{1+\theta\lambda}}\left[\widehat{l_j}^F\right]^{\frac{1-\theta\alpha}{1+\theta\lambda}}\right)$$

$$\implies \boxed{\left[\widehat{l_i}^F\right]^{1-\theta\alpha}\left[\widehat{l_i}^R\right]^{\frac{\theta\lambda(1-\theta\beta)}{1+\theta\lambda}} = \widehat{\chi}\left[\frac{L_i^R}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right]\left[\widehat{l_i}^R\right]^{\frac{\theta(\beta+\lambda)}{1+\theta\lambda}} + \widehat{\chi}^{\frac{\theta\lambda}{1+\theta\lambda}}\sum_{j=1}^N \left(\frac{\Xi_{ji}}{L_i^R + \sum_{j=1}^N \Xi_{ji}}\right)\left(\widehat{t_{ji}}\left[\widehat{l_j}^F\right]^{\frac{1-\theta\alpha}{1+\theta\lambda}}\right)} \tag{E.47}$$

Equation (E.46) is identical to (E.26) in the text, while (E.47) is identical to (E.27) in the text. Note that (E.26) and (E.27) are nearly identical to equations (38) and (39) in Allen and Arkolakis (2020), except for the terms $\widehat{\bar{t}_{ij}}$. In this extended model, the exact hat equations need to account for improvements to potentially multiple transport modes, while in equations (38) and (39) of Allen and Arkolakis (2020), only one transport mode is considered.