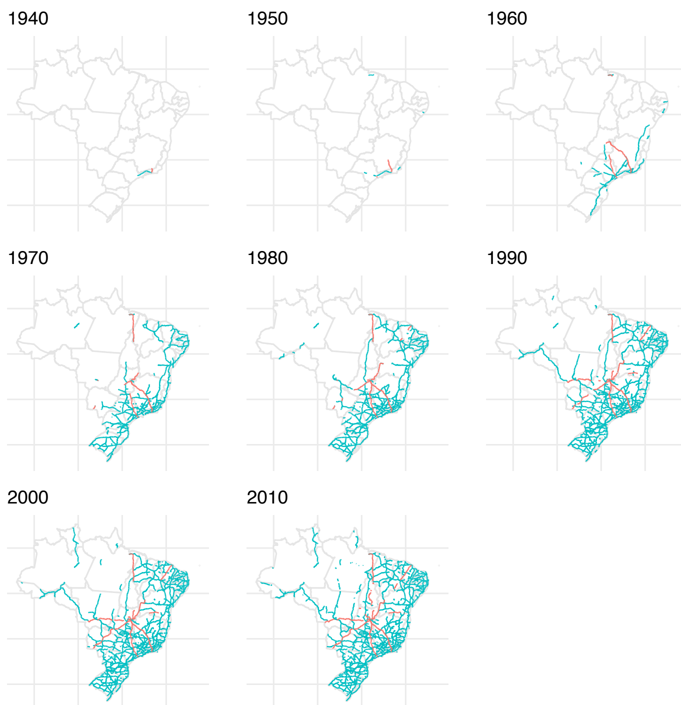
ONLINE APPENDIX

The effects of roads on trade and migration: Evidence from a planned capital city

Melanie Morten and Jaqueline Oliveira

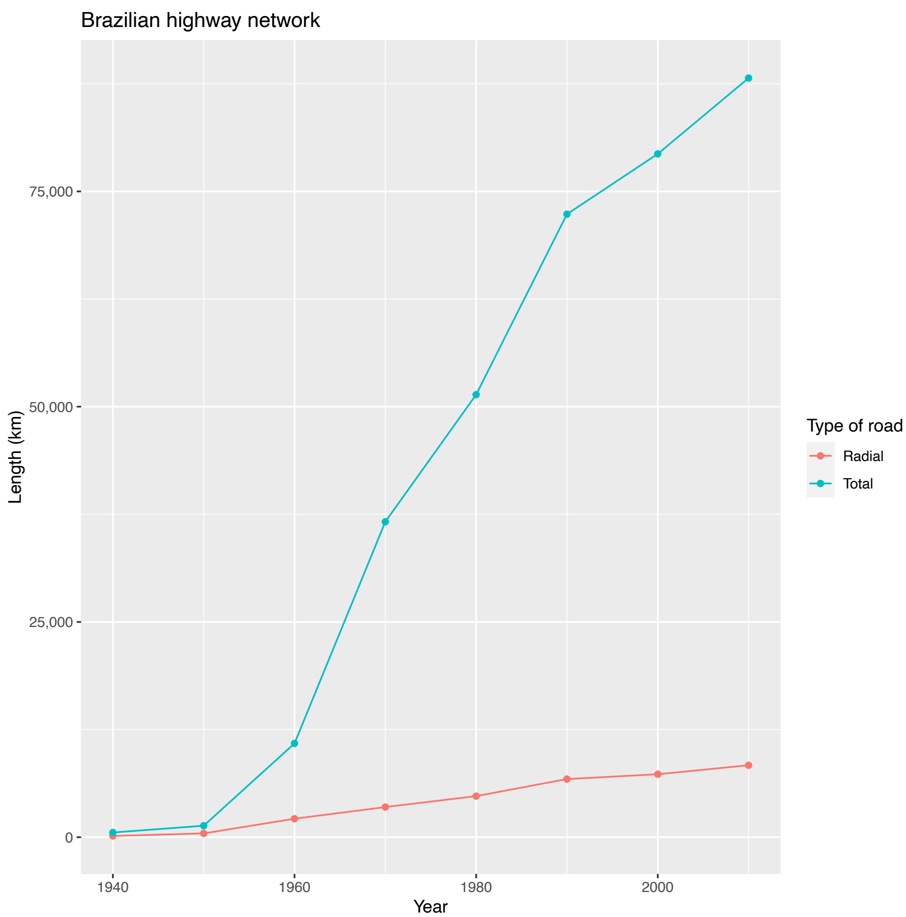
Appendix Figures and Tables

Appendix Figure A1. : Evolution of Brazil's federal highway system, 1940-2010



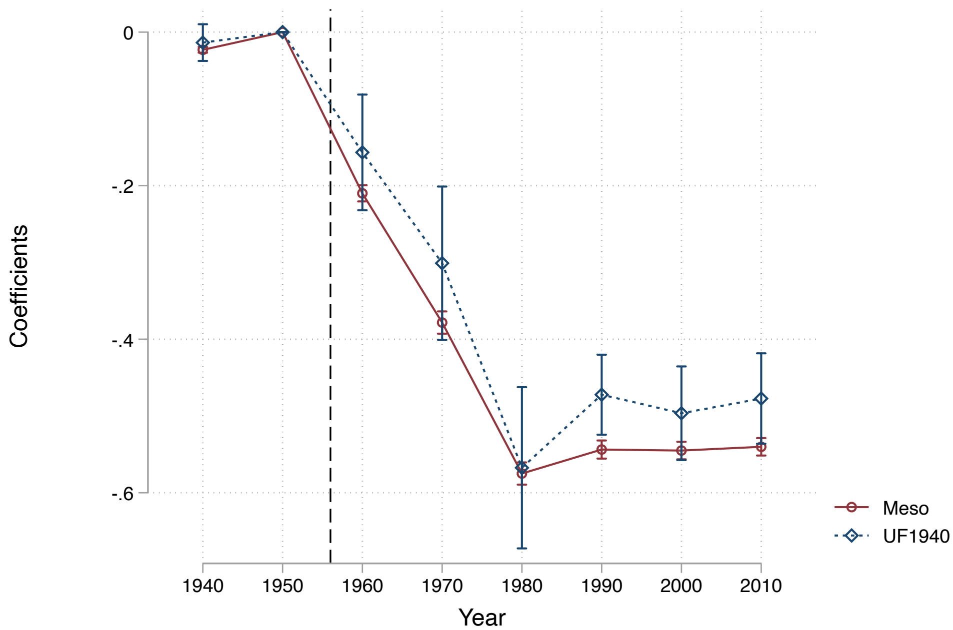
Notes: Figure is a map of the Brazilian road network. The map shows the non-radial highways in blue and radial highways in pink. Consistent state boundaries appear in the background of the map. Source: authors' calculations based on maps obtained from the Brazilian Ministry of Transportation.

Appendix Figure A2. : The growth of the road network



Notes: Figure shows the total length of paved highways in Brazil. The figure shows non-radial highways in blue and radial highways in pink. Source: authors' calculations based on maps obtained from the Brazilian Ministry of Transportation.

Appendix Figure A3. : Time-varying effects of the predicted road network (State vs Meso geo-units)



Notes: The graphs plot the estimated α_t coefficients from Equation (2) for state pairs (UF1940) and meso pairs (meso). The coefficient corresponding to 1950, the year immediately before Brasilia, is normalized to zero. For all other years, the coefficient measures the difference in the MST-induced changes in the outcome of interest between that year and the year 1950. Standard errors are clustered at the pair level.

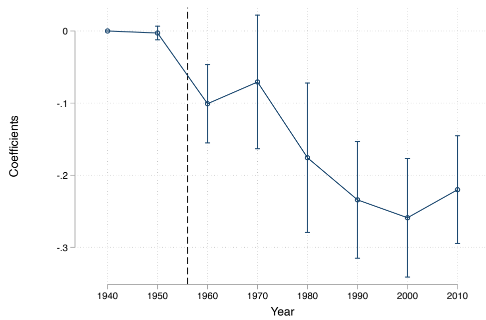
Appendix Table A1—: Correlation of MST-induced reduction in travel time with geographic variables

Dep. variable: reduction in travel time	(1)	(2)
Log distance	0.281	0.334
	(0.015)***	(0.013)***
Log distance origin to coast		0.020
		(0.016)
Log distance destination to coast		0.020
		(0.016)
Log distance origin to Brasilia		-0.152
		(0.019)***
Log distance destination to Brasilia		-0.158
		(0.019)***
Log distance origin nearest state capital		-0.082
		(0.023)***
Log distance destination nearest state capital		-0.083
		(0.023)***
N	420	420

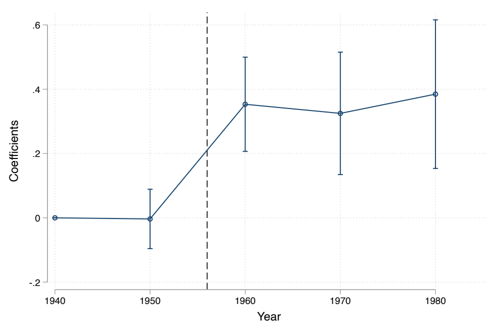
Notes: Table shows regression coefficients form a regression of MST-induced reduction in travel time on geographical characteristics of the pair. The unit of analysis is a state-state pair. Regression is unweighted. Standard errors are unclustered as there is only one observation per pair. * p<0.1, ** p<0.05, *** p<0.01

Appendix Figure A4. : Time-varying effects of the predicted road network (Imputation Estimator)

(a) Roads



(b) Migration



(c) Trade

Notes: The graphs plot the estimated α_t coefficients from Equation (2) for (a) travel time on the actual road network, (b) migration, and (c) trade using the did imputation estimator of Borusyak, Jaravel and Spiess (2021). did imputation needs a discrete treatment and a region is defined to be treated if the reduction of its travel time on the MST road network is above median.

Appendix Table A2—: Time-varying effects of the predicted road network.

	Travel Time on Roads		Migration		Trade	
	(1) Before/After	(2) Year dummies	(3) Before/After	(4) Year dummies	(5) Before/After	(6) Year dummies
MST-induced Reduction in TT x After 1950	-0.392		0.746		0.879	
	(0.031)***		(0.174)***		(0.455)*	
MST-induced Reduction in TT x 1940		-0.014		-0.090		1.655
		(0.012)		(0.103)		(0.409)***
MST-induced Reduction in TT x 1945						1.094
						(0.342)***
MST-induced Reduction in TT x 1955						
MST-induced Reduction in TT x 1960		-0.157		0.414		1.276
		(0.038)***		(0.093)***		(0.429)***
MST-induced Reduction in TT x 1965						1.586
						(0.438)***
MST-induced Reduction in TT x 1970		-0.301		0.347		1.226
		(0.051)***		(0.144)**		(0.453)***
MST-induced Reduction in TT x 1975						4.900
						(1.386)***
MST-induced Reduction in TT x 1980		-0.568		0.589		
		(0.054)***		(0.166)***		
MST-induced Reduction in TT x 1985						2.327
						(0.521)***
MST-induced Reduction in TT x 1990		-0.472		0.796		
		(0.027)***		(0.188)***		
MST-induced Reduction in TT x 2000		-0.496		0.992		1.500
		(0.031)***		(0.202)***		(0.545)***
N	2940	2940	2939	2939	2913	2913

Notes: An observation is a state-pair-year. **Travel time**: travel time between state o and state d centroids on the actual federal highway network in year t, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation. **MST-induced Reduction in TT**: change in travel times on the predicted road network (PRN), relative to travel times on an empty map. The PRN is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state o living in destination state d in time t. Data available decennially from 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Migration data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from population census. **Trade**: Measure of trade value is the value of trade from state o to destination d in time t. Data is available annnually from 1942-1949, 1959-1974, 1985, and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or because it was not reported. Furthermore, to ease vizualization we have averaged the trade data over five years. Trade data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). The omitted year in columns (2), (4), and (6) is 1950. Dependent variables are in logs. After 1950 is an indicator for whether the observation comes from years after 1950. Regressions are weighted by the migration and normalized trade flows so that each individual migration move or good transaction has equal weight. All regressions include state origin-year, state destination-year, and state pair fixed effects. Standard errors clustered at the pair level are reported in paretheses. Migration regression weighted by migration share. Trade regression weighted by trade share. * p<0.1, ** p<0.05, *** p<0.01

Appendix Table A3—: Migration elasticities to travel time on roads - three-way clustered standard errors

	Migration			
	(1) OLS	(2) IV	(3) PPML	(4) PPML
Log Travel Time on Roads	-0.332	-1.753	-0.560	-2.123
	(0.242)	(0.703)**	(0.283)**	(0.735)***
N	2939	2939	2940	2940
F-stat	1.883	6.219		
First-stage F stat		54.617		
First-stage residuals control			no	yes

Notes: An observation is a state-pair-year. **Log travel time on roads**: travel time between state o and state d centroids on the actual federal highway network in year t, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation. Log travel time on roads is instrumented by the change in travel times on the predicted road network, relative to travel times on an empty map, interacted with year dummies (omitted year is 1950). The predicted road network is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state o living in destination state d in time t. Data is decennial covering 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from Brazilian Population Census. **Trade**: Measure of trade value is the value of trade from state o to destination d in time t. Data is annual covering 1942-1949, 1959-1974, 1985 and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or it was not reported. Data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). All regressions include state origin-year, state destination-year, and state pair fixed effects. The dependent variables are in logs in columns (1), (2), (5), and (6), and in levels in columns (3), (4), (7), and (8). The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Columns (3), (4), (7) and (8) present Poisson Pseudo Maximum Likelihood (PPML) estimates. Columns (4) and (8) add control for the first-stage regression residuals on the entire sample, including those with zero flows. Standard errors clustered at the origin, destination, and year levels reported in paretheses. OLS/IV regressions are weighted by the migration and normalized trade flows so that each individual migration move or good transaction has equal weight. Regressions are unweighted for the PPML regressions. * p<0.1, ** p<0.05, *** p<0.01.

Appendix Table A4—: Migration and Trade elasticities to travel time on roads: First Stage

	Migration	Trade
	(1) OLS	(2) OLS
MST-induced Reduction in TT x 1940	0.042 (0.050)	
MST-induced Reduction in TT x 1942		-0.116 (0.060)*
MST-induced Reduction in TT x 1943		-0.116 (0.057)**
MST-induced Reduction in TT x 1944		-0.042 (0.042)
MST-induced Reduction in TT x 1945		-0.011 (0.041)
MST-induced Reduction in TT x 1946		-0.015 (0.036)
MST-induced Reduction in TT x 1947		-0.012 (0.038)
MST-induced Reduction in TT x 1948		-0.063 (0.043)
MST-induced Reduction in TT x 1959		0.185 (0.065)***
MST-induced Reduction in TT x 1960	-0.012 (0.075)	-0.131 (0.114)
MST-induced Reduction in TT x 1961		-0.132 (0.109)
MST-induced Reduction in TT x 1962		-0.130 (0.112)
MST-induced Reduction in TT x 1963		-0.138 (0.093)
MST-induced Reduction in TT x 1964		-0.119 (0.109)
MST-induced Reduction in TT x 1965		-0.042 (0.090)
MST-induced Reduction in TT x 1966		-0.136 (0.112)
MST-induced Reduction in TT x 1967		0.006 (0.087)
MST-induced Reduction in TT x 1968		-0.145 (0.097)
MST-induced Reduction in TT x 1969		0.037 (0.091)
MST-induced Reduction in TT x 1970	-0.041 (0.106)	0.089 (0.149)
MST-induced Reduction in TT x 1971		-0.053 (0.105)
MST-induced Reduction in TT x 1972		-0.098 (0.103)
MST-induced Reduction in TT x 1973		-0.014 (0.130)
MST-induced Reduction in TT x 1974		0.078 (0.187)
MST-induced Reduction in TT x 1980	-0.331 (0.114)***	
MST-induced Reduction in TT x 1985		-0.414 (0.096)***
MST-induced Reduction in TT x 1990	-0.311 (0.089)***	
MST-induced Reduction in TT x 1998		-0.384 (0.081)***
MST-induced Reduction in TT x 1999		-0.394 (0.082)***
MST-induced Reduction in TT x 2000	-0.294 (0.085)***	
N	2939	7451
First-stage F stat	22.086	11.217

Notes: An observation is a state-pair-year. Dependent variable in first-stage regressions is the log of travel time on roads. **Log travel time on roads**: travel time between state o and state d centroids on the actual federal highway network in year t, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation.**MST-induced Reduction in TT**: change in travel times on the predicted road network, relative to travel times on an empty map. The predicted road network is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state o living in destination state d in time t. Data is decennial covering 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from Brazilian Population Census. **Trade**: Measure of trade value is the value of trade from state o to destination d in time t. Data is annual covering 1942-1949, 1959-1974, 1985 and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or it was not reported. Data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). All regressions include state origin-year, state destination-year, and state pair fixed effects. Omitted year is 1950. The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Standard errors clustered at the pair level are reported in paretheses. Regressions are weighted by the migration and normalized trade flows so that each individual migration move (or good transaction) has equal weight. * p<0.1, ** p<0.05, *** p<0.01.

Appendix Table A5—: Elasticities to travel time on roads: PPML robustness to higher-order first-stage residual polynomials

	Migration					Trade				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	2nd	4th	6th	8th	10th	2nd	4th	6th	8th	10th
Log Travel Time on Roads	-2.155	-2.157	-2.098	-2.158	-2.168	-1.908	-1.938	-1.960	-1.585	-1.556
	(0.336)***	(0.336)***	(0.333)***	(0.321)***	(0.316)***	(0.442)***	(0.434)***	(0.440)***	(0.468)***	(0.471)***
N	2940	2940	2940	2940	2940	8960	8960	8960	8960	8960

Notes: Each column shows PPML estimates that add controls for polynomials of first-stage residuals up to the order displayed on the column heading. An observation is a state-pair-year. **Log travel time on roads**: travel time between state o and state d centroids on the actual federal highway network in year t, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation. Log travel time on roads is instrumented by the change in travel times on the predicted road network, relative to travel times on an empty map, interacted with year dummies (omitted year is 1950). The predicted road network is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state o living in destination state d in time t. Data is decennial covering 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from Brazilian Population Census. **Trade**: Measure of trade value is the value of trade from state o to destination d in time t. Data is annual covering 1942-1949, 1959-1974, 1985 and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or it was not reported. Data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). All regressions include state origin-year, state destination-year, and state pair fixed effects. Standard errors clustered at the pair level are reported in paretheses. OLS/IV regressions are weighted by the migration and normalized trade flows so that each individual migration move or good transaction has equal weight. Regressions are unweighted for the PPML regressions. * p<0.1, ** p<0.05, *** p<0.01.

Appendix Table A6—: Elasticity of trade flows to distance

	(1)	(2)	(3)	(4)
Dep. var: log trade flows				
Log distance	-1.477			-0.509
	(0.053)***			(0.207)**
Log traveltime		-1.633	-0.978	-1.101
		(0.054)***	(0.139)***	(0.218)***
Pair FE			✓	
N	7453	7453	7451	7453
F stat	790.839	927.396	49.771	463.823
widstat				

Notes: An observation is a state-pair-year. The dependent variable is log trade flows. Standard errors clustered at the pair level are reported in paretheses. * p<0.1, ** p<0.05, *** p<0.01

Appendix Table A7—: Migration and Trade elasticities to travel time on roads - robustness to dropping Goias

	Migration				Trade			
	(1) OLS	(2) IV	(3) PPML	(4) PPML	(5) OLS	(6) IV	(7) PPML	(8) PPML
Log Travel Time on Roads	-0.419	-2.409	-0.676	-2.481	-0.786	-2.307	0.021	-1.866
	(0.119)***	(0.536)***	(0.125)***	(0.336)***	(0.158)***	(0.556)***	(0.222)	(0.451)***
N	2660	2660	2660	2660	6734	6734	8132	8132
F-stat	12.474	20.214			24.735	17.227		
First-stage F stat		18.681				14.036		
First-stage residuals control			no	yes			no	yes

Notes: Sample excludes pairs involving Goias, the state which houses Brasilia. An observation is a state-pair-year. **Log travel time on roads**: travel time between state o and state d centroids on the actual federal highway network in year t, calculated using a Fast-Marching Algorithm. Data available decennially from 1940-2000. Road data source: Brazilian Ministry of Transportation. Log travel time on roads is instrumented by the change in travel times on the predicted road network, relative to travel times on an empty map, interacted with year dummies (omitted year is 1950). The predicted road network is created by using a Minimum Spanning Tree (MST) algorithm aiming to connect Brasilia's centroid to those of all other existing state capitals within each one of the eight pie slices defined by the cardinal and intercardinal directions. **Migration**: Measure of migration is the stock of people born in state o living in destination state d in time t. Data is decennial covering 1940-2000. The maximum number of observations is 2940 (21 states*20 states*7 years). Data source: pre-1990: digitized from historical yearbooks. 1991-2000: microdata from Brazilian Population Census. **Trade**: Measure of trade value is the value of trade from state o to destination d in time t. Data is annual covering 1942-1949, 1959-1974, 1985 and 1998-1999. The maximum number of observations is 11340 (21 states*20 states*27 years) but some pairs are missing trade data because there was no trade between them or it was not reported. Data source: pre-1998: digitized from historical yearbooks. 1998-99: de Vasconcelos and de Oliveira (2006). All regressions include state origin-year, state destination-year, and state pair fixed effects. The dependent variables are in logs in columns (1), (2), (5), and (6), and in levels in columns (3), (4), (7), and (8). The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Columns (3), (4), (7) and (8) present Poisson Pseudo Maximum Likelihood (PPML) estimates. Columns (4) and (8) add control for the first-stage regression residuals on the entire sample, including those with zero flows. Standard errors clustered at the pair level are reported in paretheses. OLS/IV regressions are weighted by the migration and normalized trade flows so that each individual migration move or good transaction has equal weight. Regressions are unweighted for the PPML regressions. * p<0.1, ** p<0.05, *** p<0.01.

Appendix Table A8—: Estimating trade elasticity

	(1) OLS	(2) IV	(3) First stage
Change log wage	0.890	-70.793	
	(1.524)	(544.661)	
Bartik instrument			0.383
			(2.136)
Implied theta	-1.124	0.014	
	(1.925)	(0.109)	
N			21
F stat	0.341	0.015	
First stage F stat		0.016	0.032

Notes: Unit of analysis is a state and each observation is the 1999-1985 difference in estimated origin fixed effect. Instruments for changes in wages are state-level Bartik shocks. The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Standard errors are unclustered. Regressions unweighted. * p<0.1, ** p<0.05, *** p<0.01.

Appendix Table A9—: Comparing response across geography and stock/flow

	Working men			Entire population			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep var: log migration flow	Meso flow	UF flow	UF stock	Meso flow	UF flow	UF stock	Hist. UF stock
Panel a: Distance							
Log distance	-1.352	-1.296	-1.505	-1.564	-1.389	-1.538	-1.947
	0.014***	0.074***	0.086***	0.014***	0.076***	0.086***	0.096***
Panel b: Traveltime on road							
Log traveltime road	-1.602	-1.636	-1.936	-1.863	-1.756	-1.986	-2.108
	0.019***	0.098***	0.111***	0.019***	0.101***	0.110***	0.099***
Panel c: Traveltime on mst							
Log traveltime mst	-1.694	-1.683	-1.965	-1.978	-1.806	-2.005	-2.525
	0.018***	0.105***	0.122***	0.019***	0.109***	0.122***	0.146***
N	43640	1670	1680	51511	1677	2100	2099
r2	0.600	0.822	0.829	0.610	0.821	0.830	0.798
origFE	✓	✓	✓	✓	✓	✓	✓
destFE	✓	✓	✓	✓	✓	✓	✓
histsample							✓
microsample	✓	✓	✓	✓	✓	✓	

Notes: Table shows regression results from estimating gravity equations on log migration. The dependent variable is either meso-meso migration over the last five years (meso flow), state-state migration over the last five years (UF flow), or migration outside state of birth (UF stock). Datasource is the Brazilian census microdata between 1980 and 2010 for columns (1)-(6). The sample for column (7) is the historical data for 1940-1980 digitized from historical yearbooks. Working men are men between 20-65 with non-zero income. All is the entire Brazilian population. Standard errors clustered at the orig-dest pair. Regression is unweighted. * p<0.1, ** p<0.05, *** p<0.01

Appendix Table A10—: Estimating distance coefficient (meso level)

	Poisson		Linear	
	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
Log distance	-1.705	-1.698	-1.596	-1.597
	(0.035)***	(0.035)***	(0.014)***	(0.014)***
Estimated change in mig costs	170.367		-148.573	
	(71.609)**		(179.058)	
N	55896	55896	37939	37939

Notes: An observation is a meso-pair year in the range 1980-2000. The estimated change in mig costs are the results implied from the state-level gravity regressions. In columns (2) and (4) the value of the estimated change in mig (trade) costs is constrained to be 1 and is not reported. Standard errors are clustered at the pair level. Regressions are unweighted. * p<0.1, ** p<0.05, *** p<0.01.

Appendix Table A11—: Estimating distance coefficient (state-level)

	Migration		Trade	
	(1)	(2)	(3)	(4)
Log distance	-1.230	-1.277	-0.996	-1.187
	(0.160)***	(0.124)***	(0.088)***	(0.056)***
Estimated change in mig costs	0.793			
	(0.509)			
Estimated change in trade costs			0.194	
			(0.257)	
N	1260	1260	1260	1260

Notes: An observation is a state-pair year in the range 1980-2000. The estimated change in mig (trade) costs are the results implied from the state-level gravity regressions. In columns (2) and (4) the value of the estimated change in mig (trade) costs is constrained to be 1 and is not reported. Standard errors are clustered at the pair level. Regressions are unweighted. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Appendix Table A12—: First stage: migration and housing elasticity

	Mig elasticity		Housing elasticity
	(1)	(2)	(3)
	Cost	Nocost	
Bartik shock	1.801	1.801	-0.347
	(0.461)***	(0.461)***	(1.042)
Predicted change in log pop, IV	0.097	0.097	0.282
	(0.036)***	(0.036)***	(0.079)***
N	137	137	137
First-stage F stat	16.103	16.103	6.395

Notes: Unit of analysis is a meso-region and each observation is the 2010-1980 difference. The dependent variable in columns (1) and (2) is the 2010-1980 change in (log) indirect utility of the meso-region. The indirect utilities for years 2010 and 1980 are obtained as the set of (meso) destination-year fixed effects after estimating a gravity equation from meso-to-meso bilateral migration flows, accounting for bilateral migration costs (as a linear function of distance and travel times in logs) and (meso) origin-year fixed effects. The dependent variable in columns (3) and (4) are also the 2010-1980 change in indirect utilities from meso-to-meso flows, but assuming zero bilateral migration costs. The dependent variable in columns (5) and (6) is the change in (log) rental prices. Rents are calculated from census micro-data in each year as the meso-specific average after netting out housing characteristics such as number of rooms and bedrooms, electricity access, walls, roof, and floor quality, among others. Instruments for changes in wages and housing expenditure are meso-level Bartik shocks and a model-based measure of changes in labor as a function of Bartik shocks of all meso-regions weighted by migration costs. The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Standard errors clustered at the state level. Regressions unweighted. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Appendix Table A13—: Imputed prices: migration and housing elasticity

	Baseline		Imputed prices		Raw rents		F.S. baseline	F.S. prices		F.S. raw. rents
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7)	(8)	(9)	(10)
Change in wages, adjusted for residualized rents	0.029	4.515	1.239	5.841						
	(0.295)	(0.763)***	(0.500)**	(1.136)***						
Change log imputed prices			0.197	7.983						
			(1.496)	(3.925)*						
Change in wages, adjusted for rents					1.553	5.653				
					(0.465)***	(0.815)***				
Bartik shock							1.801	1.471	0.001	1.416
							(0.461)***	(0.421)***	(0.033)	(0.431)***
Predicted change in log pop, IV							0.097	0.090	0.001	0.115
							(0.036)***	(0.035)**	(0.002)	(0.031)***
Change log imputed prices IV								-0.381	0.359	
								(0.160)**	(0.019)***	
N	137	137	137	137	137	137	137	137	137	137
F stat	0.009	35.000	3.286	13.329	11.132	48.112				
SW F		16.103		8.423		21.687	16.103	8.423	8.423	21.687

Notes: Unit of analysis is a meso-region and each observation is the 2010-1980 difference. The dependent variable in columns (1) and (2) is the 2010-1980 change in (log) indirect utility of the meso-region. The indirect utilities for years 2010 and 1980 are obtained as the set of (meso) destination-year fixed effects after estimating a gravity equation from meso-to-meso bilateral migration flows, accounting for bilateral migration costs (as a linear function of distance and travel times in logs) and (meso) origin-year fixed effects. The dependent variable in columns (3) and (4) are also the 2010-1980 change in indirect utilities from meso-to-meso flows, but assuming zero bilateral migration costs. The dependent variable in columns (5) and (6) is the change in (log) rental prices. Rents are calculated from census micro-data in each year as the meso-specific average after netting out housing characteristics such as number of rooms and bedrooms, electricity access, walls, roof, and floor quality, among others. Instruments for changes in wages and housing expenditure are meso-level Bartik shocks and a model-based measure of changes in labor as a function of Bartik shocks of all meso-regions weighted by migration costs. The Kleibergen-Paap F statistic (First-stage F stat) for weak identification is reported for IV estimates. Standard errors clustered at the state level. Regressions unweighted. * p<0.1, ** p<0.05, *** p<0.01.

Appendix Table A14—: Robustness of welfare to elasticity and distance elasticity: equalizing coefficients

	Trade dist. elast.	Trade elast.	Mig dist. elast	Mig elast.	Change trade tau	Change mig tau	Change welfare	Share welfare trade
Baseline	0.47	4.00	0.39	4.51	0.10	0.08	1.03	0.76
Set to trade coeff.	0.47	4.00	0.47	4.00	0.10	0.10	1.03	0.71
Set to mig. coeff.	0.39	4.51	0.39	4.51	0.08	0.08	1.02	0.72

Notes: Table shows the relative change. All values are relative to a baseline value of 1. The numbers are computed by simulating the structural model under different elasticities. Table equalizes the migration and trade distance coefficients (and migration and trade elasticities) either to the estimated migration values or the estimated trade values.

Appendix Table A15—: Robustness of welfare to elasticity and distance elasticity

	Half			Baseline			Double		
	0.5x	1x	2x	0.5x	1x	2x	0.5x	1x	2x
Trade elasticity/distance elasticity									
Elasticity of trade costs to roads	0.24	0.24	0.24	0.47	0.47	0.47	0.94	0.94	0.94
Trade elasticity	2.00	4.00	8.00	2.00	4.00	8.00	2.00	4.00	8.00
Change in welfare	1.02	1.02	1.02	1.03	1.03	1.03	1.05	1.06	1.08
Implied reduction in trade costs	0.05	0.05	0.05	0.10	0.10	0.09	0.19	0.18	0.18
Share of welfare gain due to trade	0.57	0.59	0.61	0.74	0.76	0.79	0.86	0.88	0.91
Migration elasticity/distance elasticity									
Elasticity of mig costs to roads	0.19	0.19	0.19	0.39	0.39	0.39	0.78	0.78	0.78
Migration elasticity	2.26	4.51	9.03	2.26	4.51	9.03	2.26	4.51	9.03
Change in welfare	1.02	1.02	1.02	1.03	1.03	1.03	1.04	1.04	1.05
Implied reduction in trade costs	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Share of welfare gain due to trade	0.88	0.87	0.86	0.77	0.76	0.71	0.60	0.55	0.42

Notes: Table shows the relative change. All values are relative to a baseline value of 1. The numbers are computed by simulating the structural model under different elasticities. The first two rows of each panel report the value of the elasticity considered. The results reported in Table 4 correspond to column (5) i.e., Baseline elasticty of trade (migration) costs to roads and 1x the value of the trade (migration) elasticity.

GIS and Road Data

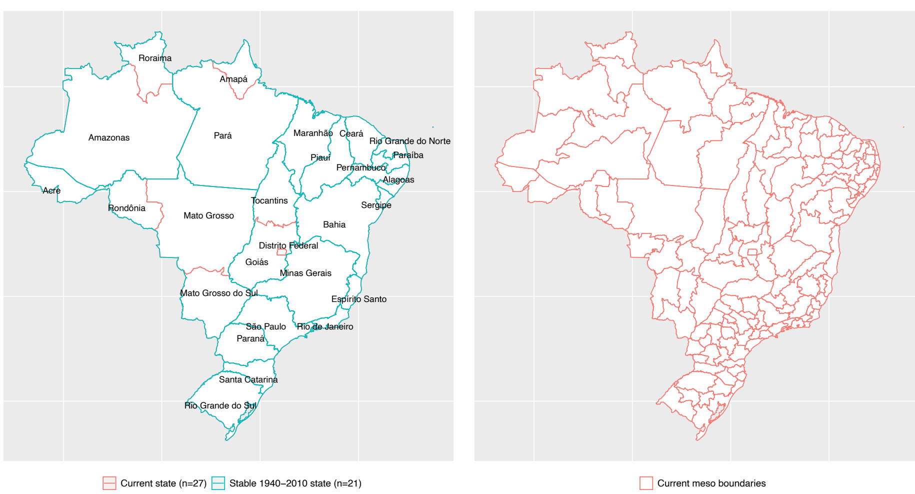
B1. *Geographical units*

For all GIS data, we use the SIRGAS 2000 (EPSG:4674) coordinate system. To compute distances, we project the shape file using the Brazil SIRGAS 2000 Polyconic (EPSG:5880) projection.

The analysis in the paper occurs at either the state level or at the meso region level. State boundaries change over time: to generate consistent boundaries from 1940 we construct a set of 21 consistent states.[32] The second unit of analysis are meso regions, of which there are 137 in Brazil.

Appendix Figure B1. : Geographical units

(a) States (b) Meso regions



Notes: Figure is a map of Brazil. Map (a) shows the current state boundaries (n=27) and how the consistent state boundaries are constructed (n=21). Map (b) shows the mesoregion boundaries (n=137).

We create a crosswalk between the 2010 meso boundaries and the municipalities in each year between 1970 and 2010. This crosswalk is based on the official division of the Brazilian territory (*Diviso Territorial Brasileira*) produced by IBGE. The URL to download these files (as of August 2021) is: `https://www.ibge.gov.br/`

[32]There were 21 states in Brazil in 1940 and there are 27 states in Brazil now. Those 21 states still exist and 6 states, including the Federal District (Brasilia), split from those states. Three states split from and returned to the 1940 states between 1940 and 2010. The old Federal District became the State of Guanabara which merged into the State of Rio de Janeiro (one of the 1940 states).

geociencias/organizacao-do-territorio/estrutura-territorial/23701-divisao-territorial-b html?edicao=23704&t=downloads. We cross-referenced our crosswalk with the crosswalk created by Ehrl (2017), available from download (as of August 2021) at: https://sites.google.com/site/philippehrl/research.

B2. Road data

1960-2010. — The main data source for roads between 1960 and 2010 is the Brazilian Ministry of Transportation. We downloaded the shapefiles for each year between 1960 and 2010 from https://www.gov.br/infraestrutura/pt-br/assuntos/dados-de-transportes/bit/bitmodosmapas#maprodo on 3/21/2021. The shapefiles are appended together to generate the cumulative road network for each year. We call this data source the MOT database.

The MOT database contains a variable that indicates the road surface status (leg_multim). This variable describes whether the road is planned, paved, or unpaved. Our analysis sample is all roads that are either paved or unpaved (drop leg_multim contains "Planejada"). The paved sample is roads that are paved (leg_multim contains the strings "Duplicada" (double-paved road) or "Pavimentada" (paved road)).[33]

1940-1950. — The only paved road in the 1940s was the road connecting Rio de Janeiro to Sao Paulo (Rodrigues, 2008). We identified paved roads built in the 1950s using (Planos Nacionais de Viacao) which contains a map showing paved roads in 1956. We selected the 1960 paved roads that matched the road from Rio de Janeiro to Sao Paulo and the paved roads in the 1956 map to construct the 1940 and 1950 road data.

Verification against OSM. — To check for the completeness of the road network we verify it against Open Street Maps.[34] We downloaded the entire OSM road network for Brazil from http://download.geofabrik.de/south-america/brazil-latest.osm.pbf on 7/22/2021. From this database, we keep road segments that have a highway classification as motorway, trunk, primary, or secondary. The OSM data contains both federal and state roads. To match the MOT data, which is only federal highways, we create a crosslisted dataset which contains all road segments that include the prefix for a federal highway in Brazil

[33]The variable ds_superfi also indicates whether a road is paved, unpaved, or planned. This variable appears to designate the federal status of the road and not the status of the road itself. For example, an existing unpaved road may be a planned paved road, or an existing paved state highway may be a planned paved federal highway. We use the leg_multim variable instead to be consistent with the actual road that exists.

[34]We cannot systematically use Google maps for the verification exercise because the data are not downloadable.

(a "BR-").[35] The road surface (i.e., paved or unpaved) is not consistently marked in the OSM database. To account for small differences in the two shape files we consider a 1km buffer when intersecting.

Share of MOT roads in OSM

- 1.5% of the paved 2010 MOT roads are not in the 2021 OSM database
- 5.3% of the paved 2010 MOT roads are not in the crosslisted 2021 OSM database

This is a fairly high match rate and the small degree of mismatch when we restrict the OSM to the crosslisted roads is likely due to the imperfect inclusion of highway codes.

Share of OSM roads in MOT

- 12.4% of the crosslisted 2021 OSM roads are not in the 2010 MOT analysis data
- 12.1% of the crosslisted 2021 OSM roads are not in the 2010 MOT data (including planned roads)

This number is slightly high. One reason that OSM roads could be missing from the 2010 database is that they were constructed after 2010 and not necessarily planned in 2010. To check if this is the explanation, we use an additional dataset available from the Brazilian Ministry of Transportation (at the same website listed above) that has the 2019 highway system. 3.8% of the crosslisted 2021 OSM roads are not in the 2019 MOT analysis data The remaining mismatch is primarily due to roads built between 2019 and 2021: if we expand the 2019 MOT to include planned roads, only 2.9% of the crosslisted 2021 OSM roads are not in the 2019 MOT data.

Overall, the mismatch rate is around 5% between the MOT and OSM database. Given this number, we feel confident that our road data matches the existing highway system in Brazil.

[35]Some roads are coincidental state and federal highways (for example, the BR-010 highway runs from Belm to Braslia and contains crosslisted state and federal segments including GO-118 (in the state of Goias), DF-345 (in the Distrito Federal), and TO-387 (in the state of Tocantins)).

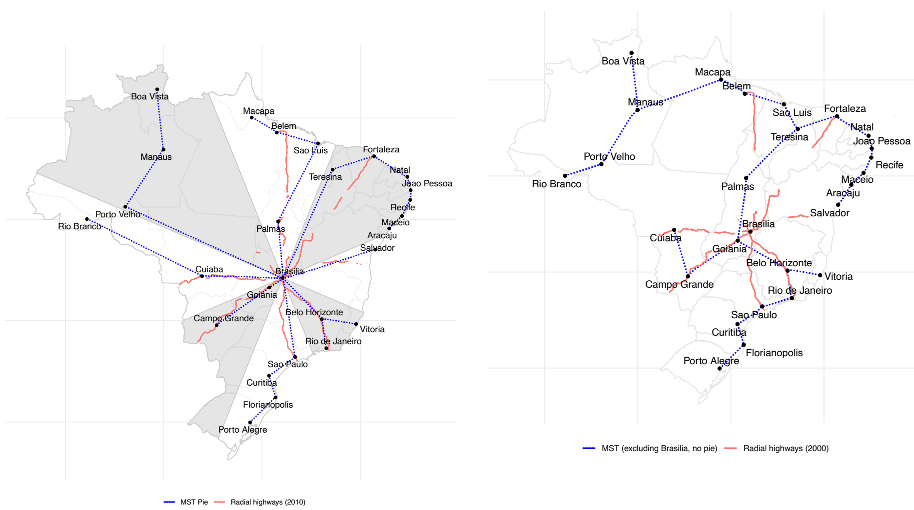
B3. *Construction of MST network*

We use ArcGIS to compute the minimum spanning tree (MST) network. We use latitude-longitude coordinates of the geo-units' centroids to create point features representing the location of Brasilia and the 26 state capitals. Next, we divide the country into 8 exogenous equal-sized slices, centered around 0N. We then use the *Optimal Region Connection* tool in ArcGIS Pro to connect the cities within the pie slice. Figure B2 illustrates this process.

To construct the counterfactual Rio network, we eliminate the Brasilia city point and then connect the remaining 26 cities using the *Optimal Region Connection* tool in ArcGIS Pro.

Appendix Figure B2. : Construction of MST networks

(a) MST pie network (b) MST network without Brasilia



Notes: Figure is a map of the Brazilian road network indicating Brasilia and the 26 state capitals. The map shows radial highways out of Brasilia and the straight-line instrument for roads. Panel (a) shows the eight equal-sized pie slices used to construct the MST instrument centered around Brasilia. Panel (b) shows the alternative MST network. Consistent state boundaries appear in the background of the map. Source: authors' calculations based on maps obtained from the Brazilian Ministry of Transportation.

B4. Construction of road travel times

To construct measures of the distance between origin-destination pairs taking into account the actual road coverage, we use the fast marching algorithm, following the approach used in Allen and Arkolakis (2014).The fast marching algorithm finds the solution to the Eikonal equation used to characterize the propagation of wave fronts. The algorithm uses a search pattern for grid points in computing the arrival times (distances) that is similar to the Dijkstra shortest path algorithm (Hassouna and Farag (2007)). However, because the fast marching algorithm is applied to a continuous graph, it reduces the grid bias and generates more accurate bilateral distances. We compute the fast marching method in R using the algorithm **fastmaRching.R**, available on CRAN at `https://cran.r-project.org/package=fastmaRching`. To run the algorithm we assign a speed of 10 to non-paved roads and a speed of 100 to paved roads.

Validation of road travel time measures. — To validate the road travel time measure, we calculated travel time between pairs of 398 populated areas in Brazil using our travel time measure on the 2010 road network and travel time scraped from Google Maps. Google Maps requires a location to be relatively close to a road network to compute travel times. The thus use a list of 398 populated places and run our fast marching algorithm and compare it to Google maps travel times. Appendix Table B1 shows that the two measures of travel time are highly correlated, even after controlling for the log distance between the two locations.

Datasources

C1. Historical state-to-state trade and migration flows

We draw state-to-state trade flow data from the statistical yearbooks produced by the IBGE. These data are available annually, spanning the periods 1942–1949, 1954-1975, and 1985. The yearbooks report the value of total exports of each state to other states across the country. The data was sourced by the Technical Council of Economics and Finance. The yearbook data reported imports and exports. However, not all states are reported. We differentiate between missing data and zero flows as follows. If we observe a reported import or reported export from a location we assume that location reports all its data and so treat any missing data as true zeros. If we do not observe any reported imports or exports we treat the data as missing. Appendix Table C1 summarizes the pair-level coverage of the trade data. We drop any year where more than 75% of the pair-level data is missing. Years that are dropped (1954-1958 and 1975) are indicated in the final column of the table.

For 1998 and 1999, interstate bilateral trade flow data are derived from information on state tax on the movement of goods and services (*Imposto sobre*

Appendix Table B1—: Validation of travel time measure with google travel times

	All cities		Exclude close to capital	
	(1)	(2)	(3)	(4)
Dep var: log google time				
Log road travel time	1.121	0.183	1.144	0.178
	0.001***	0.002***	0.001***	0.002***
Log dist km		0.793		0.800
		0.001***		0.002***
origFE	✓	✓	✓	✓
destFE	✓	✓	✓	✓
N	133605	133605	74265	74265
r2	0.958	0.989	0.956	0.989

Notes: Table shows regression results from estimating a gravity equation where the dependent variable is the log travel time between two cities computed by Google. Log road travel time is computed on the 2010 road network. An observation is a city-city pair. The all cities sample is travel time between 398 populated cities. Excluding close to capital excludes any origin or destination within 100 km of a state capital. Standard errors are not clustered as there is only one observation per pair. Regression is unweighted.

Circulacao de Mercadorias e Servicos). We use the study produced by de Vasconcelos (2001) as a data source. The data are collected by double accounting, converting reported imports and reported exports. If we observe a reported import or reported export from a location we assume that location reports all its data and so treat any missing data as true zeros. If we do not observe any reported imports or exports we treat the data as missing. Appendix Table C1 summarizes the pair-level coverage of the trade data. We drop any year where more than 75% of the pair-level data is missing. Years that are dropped (1954-1958 and 1975) are indicated in the final column of the table.

Data on state-to-state migration flows are also available from the statistical yearbooks on a decennial basis for the years 1940-1980. The books report the number of residents in all states by state of birth. Therefore, we are able to construct these flows for origin of birth. The data come from the decennial Censuses conducted by the IBGE.

C2. Census database, 1970-2010

We construct a regional database of migration, wages and roads at the meso level between 1970–2010. Summary statistics for the regional database are presented in Appendix Table C2). The primary datasource is the individual data files from the Brazilian Census, 1970–2010, collected by the Brazilian Institute of Geography and Statistics (IBGE).[36]

Our sample of interest is males aged 20–65 who report non-zero earnings in their main occupation. All nominal variables are converted into constant 2010 prices; the exchange rate between the USD and Real is approximately 1 USD = 2.3 BRL. [37]

Employment and wages. — Wage data are sourced from the census. The census asks both the average earnings per month in the main occupation,[38] as well as the usual hours worked. We use earnings from main occupation and the hours worked to construct an equivalent hourly wage rate. Assuming a standard 2000

[36]For the purposes of sampling, the national territory is divided in "setores censitarios" (census blocks). Within each sector, a fraction of the households is randomly selected and the questionnaires are administered. The fraction of households sampled within a sector has varied across census years. In 1970 and 1980, 25% of households were drawn from the population. In 1991 and 2000, the fractions varied according to municipality size. For municipalities with up to 15,000 inhabitants, 20% of their population was sampled; the fraction was 10% for municipalities with more than 15,000 people. In 2010, there were five fractions. In municipalities with up to 2,500 inhabitants, 50% were sampled; in municipalities with more than 2,500 but less than 8,000 inhabitants, the sampling fraction was 33%; in municipalities with more than 8,000 and less than 20,000 inhabitants, 20% were sampled; in municipalities with more than 20,000 and less than 500,000 people, 10% were sampled; and in municipalties with mora than 500,000 inhabitants, the fraction was 5%.

[37]We constructed a modified consumer price index that accounts for changes in the Brazilian currency that occurred within the period under analysis. All nominal variables were converted to 2010 BRL. See http://www.ipeadata.gov.br/ for the factors of conversion for the Brazilian currency.

[38]The exception is 1970, where only total earnings, rather than earnings in the main occupation, is asked.

Appendix Table C1—: Classification of trade data by year

	(1) Positive	(2) Zero	(3) Missing	(4) Dropped from analysis
1942	0.52	0.48	0.00	0.00
1943	0.65	0.35	0.00	0.00
1944	0.58	0.28	0.14	0.00
1945	0.61	0.20	0.19	0.00
1946	0.60	0.25	0.14	0.00
1947	0.62	0.19	0.19	0.00
1948	0.53	0.28	0.19	0.00
1949	0.59	0.27	0.14	0.00
1954	0.14	0.00	0.86	1.00
1955	0.19	0.00	0.81	1.00
1956	0.05	0.00	0.95	1.00
1957	0.05	0.00	0.95	1.00
1958	0.19	0.00	0.81	1.00
1959	0.54	0.08	0.38	0.00
1960	0.68	0.08	0.24	0.00
1961	0.77	0.09	0.14	0.00
1962	0.75	0.10	0.14	0.00
1963	0.76	0.10	0.14	0.00
1964	0.73	0.13	0.14	0.00
1965	0.67	0.09	0.24	0.00
1966	0.75	0.06	0.19	0.00
1967	0.69	0.07	0.24	0.00
1968	0.68	0.08	0.24	0.00
1969	0.66	0.06	0.29	0.00
1970	0.52	0.05	0.43	0.00
1971	0.52	0.05	0.43	0.00
1972	0.52	0.10	0.38	0.00
1973	0.49	0.04	0.48	0.00
1974	0.35	0.08	0.57	0.00
1975	0.11	0.03	0.86	1.00
1985	0.99	0.01	0.00	0.00
1998	0.97	0.03	0.00	0.00
1999	0.99	0.01	0.00	0.00
Total	0.56	0.11	0.33	0.18
N	13860	13860	13860	14553

Notes: Classification of trade data into missing, zero, or positive flow. 420 total pairs (21 origin x 20 destination states as own-trade is never measured). We drop years from the analysis if 75% or more of the pair-level data is missing for that year.

Appendix Table C2—: Census summary stats

	(1) 1970	(2) 1980	(3) 1991	(4) 2010
Live in Brasilia	0.006	0.011	0.011	0.015
Live in Rio de Janeiro	0.085	0.088	0.074	0.068
Migrated between UF	.	0.054	0.043	0.032
Migrated between meso	.	0.095	0.071	0.053
Report previous meso	0.000	0.986	0.995	0.992
Report previous UF	0.000	0.986	0.995	0.992
Mean wage	2.979	6.124	5.184	8.403
Mean rent	219.033	295.956	302.135	364.283
Share renting	0.176	0.224	0.158	0.186
Avg years school	2.781	3.812	5.438	8.623
N	4733022	5896078	3540516	4471780

Notes: Summary stats computed on sample of men with non-zero earnings between 20-65 years old. All nominal variables converted to 2010 Brazilian reals. Migration measured based on location five years prior. Summary stats weighted by IBGE-provided individual weights. Data source: Brazilian census microdata/

hour work year, the annual wage of 6.4 BRL in 2010 would be equivalent to annual income of $5565. The per capita GDP figures for Brazil was $5600 in 2010 (World Development Indicators) and so the wage is the correct magnitude.

Migration. — The current location of the individual is coded to the municipality level. From 1980, location 5 years ago is also coded to the municipality level. To get consistent geographic boundaries over time we aggregate municipalities to match the 2010 meso and UF boundaries. Migration data is first reported in the 1980 census; we see that the interstate migration rate is between 5.3% in 1980 to 3.2% in 2010. The inter-meso migration rate is naturally higher, at 9.5% in 1980 and 5.3% in 2010.

Rental prices. — For rental rates we use census data on the rents paid for housing. The mean rental rate in 2010 is 321 BRL a month, equivalent to 50 hours of work at the mean wage. 18.5% of the population report paying rents for their housing in 2010. While this may seem low, the equivalent number for US houses in 2005 is 24%.

Appendix Table C3 tabulates the observable characteristics for housing quality collected in each year. Appendix Table C4 shows the hedonic regressions for rent on observables by year. The residuals from these regressions are used to construct the residualized rent for each meso region.

Appendix Table C3—: Summary statistics of hedonic variables

	(1) 1970	(2) 1980	(3) 1991	(4) 2010
Number of rooms	4.855	5.274	5.640	5.922
Number of bedrooms	2.190	2.216	2.150	2.216
Sanitation	0.122	0.288	0.346	0.577
Public water	0.311	0.560	0.696	0.836
Has electricity	0.460	0.709	0.867	0.992
Rural	0.453	0.294	0.234	0.126
House made from good materials	0.260			
House not apartment		0.934	0.922	0.996
Walls made from durable materials		0.701	0.793	0.978
Roof made from durable materials		0.914	0.945	
Floor made from durable materials		0.830		
Has toilet			0.783	0.954
N	4733048	5782473	3540519	4471780

Notes: An observation an individual from our sample of men with non-zero earnings aged between 20 and 65 years. We note that the sample includes individuals from all kinds of housing units, whether or not the unit is a rental. Table presents average values of available housing variables for each census year. Averages are weighted by IBGE-provided individual weights. Data source: Brazilian census microdata.

Appendix Table C4—: Hedonic Regressions

Dep. var: log rent	(1) 1970	(2) 1980	(3) 1991	(4) 2010
Number of rooms	0.154	0.178	0.162	0.159
	0.016***	0.007***	0.003***	0.004***
Number of bedrooms	-0.013	-0.003	-0.001	0.024
	0.006**	0.007	0.007	0.003***
Sanitation	0.339	0.280	0.243	0.262
	0.119***	0.015***	0.026***	0.011***
Public water	0.316	0.271	0.245	0.041
	0.043***	0.016***	0.016***	0.025*
Has electricity	0.351	0.429	0.390	0.373
	0.023***	0.021***	0.025***	0.030***
Rural	-0.271	-0.325	-0.469	-0.314
	0.040***	0.030***	0.042***	0.022***
House made from good materials	-0.183			
	0.024***			
House not apartment		-0.424	-0.369	0.000
		0.087***	0.032***	.
Walls made from durable materials		0.302	0.302	0.356
		0.029***	0.021***	0.017***
Roof made from durable materials		0.047	0.051	
		0.012***	0.020**	
Floor made from durable materials		0.287		
		0.020***		
Has toilet			0.244	0.291
			0.018***	0.044***
N	808468	1257066	498763	730640

Notes: Table reports regression results from running hedonic regressions. The dependent variable is log rent. Not all variables are available in each survey year. Standard errors are clustered at the mesoregion. Regression is unweighted. Data source: Brazilian census microdata.

Theoretical derivations

D1. Exact hat derivation

This appendix describes the procedure to compute the exact-hat changes in labor and prices:

Labor. — As given by Equation 5, labor in location d is given by:

$$\begin{aligned} L_{dt} &= \sum_o N_{odt} \\ &= \sum_o V_{dt}^{\epsilon} \kappa_{odt}^{-\epsilon} \Phi_{ot}^{-1} L_{ot} \end{aligned}$$

We can derive an expression in changes by first dividing and multiplying through by lagged terms, and then rearranging:

$$\begin{aligned} L_{dt} &= \sum_o V_{dt}^{\epsilon} \frac{V_{dt-1}^{\epsilon}}{V_{dt-1}^{\epsilon}} \kappa_{odt}^{-\epsilon} \frac{\kappa_{odt-1}^{-\epsilon}}{\kappa_{odt-1}^{-\epsilon}} \Phi_{ot}^{-1} \frac{\Phi_{ot-1}^{-1}}{\Phi_{ot-1}^{-1}} L_{ot} \frac{L_{ot-1}}{L_{ot-1}} \\ &= \sum_o \widehat{V}_{dt}^{\epsilon} \widehat{\kappa}_{odt}^{-\epsilon} \widehat{\Phi}_{ot}^{-1} \widehat{L}_{ot} V_{dt-1}^{\epsilon} \kappa_{odt-1}^{-\epsilon} \Phi_{ot-1}^{-1} L_{ot-1} \\ &= \sum_o \widehat{V}_{dt}^{\epsilon} \widehat{\kappa}_{odt}^{-\epsilon} \widehat{\Phi}_{ot}^{-1} \widehat{L}_{ot} V_{dt-1}^{\epsilon} N_{odt-1} \\ \frac{L_{dt}}{L_{dt-1}} &= \sum_o \left(\frac{N_{odt-1}}{L_{dt-1}} \right) \widehat{V}_{dt}^{\epsilon} \widehat{\kappa}_{odt}^{-\epsilon} \widehat{\Phi}_{ot}^{-1} \widehat{L}_{ot} \\ \widehat{L}_{dt} &= \sum_o \pi_{odt-1}^{d} \widehat{V}_{dt}^{\epsilon} \widehat{\kappa}_{odt}^{-\epsilon} \widehat{\Phi}_{ot}^{-1} \widehat{L}_{ot} \end{aligned}$$

Where π_{odt-1}^{d} is the destination share of labor from location o, $\frac{N_{odt-1}}{L_{dt-1}}$.

We can proceed in a similar way to derive $\widehat{\Phi}_{ot}^{-1}$:

$$
\begin{aligned}
\Phi_{ot} &= \sum_d V_{dt}^{\epsilon} \tau_{odt}^{-\epsilon} \\
&= \sum_d \widehat{V}_{dt}^{\epsilon} \widehat{\tau}_{odt}^{-\epsilon} V_{dt-1}^{\epsilon} \tau_{odt-1}^{-\epsilon} \\
\frac{\Phi_{ot}}{\Phi_{ot-1}} &= \sum_d \left(\frac{V_{dt-1}^{\epsilon} \tau_{odt-1}^{-\epsilon}}{\Phi_{ot}} \right) \widehat{V}_{dt}^{\epsilon} \widehat{\tau}_{odt}^{-\epsilon} \\
\widehat{\Phi}_{ot} &= \sum_d \pi_{odt-1} \widehat{V}_{dt}^{\epsilon} \widehat{\tau}_{odt}^{-\epsilon}
\end{aligned}
$$

Putting the two together yields:

$$
\begin{aligned}
\widehat{L}_{dt} &= \sum_o \pi_{odt-1}^{d} \widehat{V}_{dt}^{\epsilon} \widehat{\kappa}_{odt}^{-\epsilon} \left(\sum_d \pi_{odt-1} \widehat{V}_{dt}^{\epsilon} \widehat{\tau}_{odt}^{-\epsilon} \right)^{-1} \widehat{L}_{ot} \\
&= \underbrace{\widehat{V}_{dt}^{\epsilon}}_{\text{Direct effect}} \sum_o \underbrace{\pi_{odt-1}^{d} \widehat{\kappa}_{odt}^{-\epsilon} \left(\sum_d \pi_{odt-1} \widehat{V}_{dt}^{\epsilon} \widehat{\tau}_{odt}^{-\epsilon} \right)^{-1} \widehat{L}_{ot}}_{\text{Origin labor market access effect}}
\end{aligned}
$$

Since we focus on the post-Brasilia period, we assume that there are no changes in migration costs i.e., $\widehat{\kappa}_{odt} = 1$. We also abstract from exogenous population growth at the origin $\widehat{L}_{ot} = 1$. That leaves the final instrument that labor change in destination d is a composition of the direct increase in the benefits of being in location d, given by $\widehat{V}_{dt}^{\epsilon}$, and then a market access term that shows that when migrants have easy access to other destinations with a large direct gain, they will be less likely to migrate to d.

$$
\widehat{L}_{dt} = \widehat{V}_{dt}^{\epsilon} \sum_o \pi_{odt-1}^{d} \left(\sum_d \pi_{odt-1} \widehat{V}_{dt}^{\epsilon} \right)^{-1}
$$

Price Index. — A similar procedure shows that the change in the price index for location d is given by:

$$P_{dt} \propto \left(\sum_{o'} A_{o't} w_{o't}^{-\theta} \tau_{o'dt}^{-\theta} \right)^{\frac{-1}{\theta}}$$

$$P_{dt} = \left(\sum_{o'} A_{o't-1} w_{o't-1}^{-\theta} \tau_{o'dt-1}^{-\theta} \hat{A}_{o't} \hat{w}_{o't}^{-\theta} \hat{\tau}_{o'dt}^{-\theta} \right)^{\frac{-1}{\theta}}$$

$$\widehat{P_{dt}} = \frac{1}{P_{dt-1}} \left(\sum_{o'} A_{o't-1} w_{o't-1}^{-\theta} \tau_{o'dt-1}^{-\theta} \hat{A}_{o't} \hat{w}_{o't}^{-\theta} \hat{\tau}_{o'dt}^{-\theta} \right)^{\frac{-1}{\theta}}$$

$$\widehat{P_{dt}} = \left(\sum_{o'} \frac{A_{o't-1} w_{o't-1}^{-\theta} \tau_{o'dt-1}^{-\theta}}{\sum_{o'} A_{o't} w_{o't}^{-\theta} \tau_{o'dt}^{-\theta}} \hat{A}_{o't} \hat{w}_{o't}^{-\theta} \hat{\tau}_{o'dt}^{-\theta} \right)^{\frac{-1}{\theta}}$$

$$\widehat{P_{dt}} = \left(\sum_{o'} \pi_{odt-1} \hat{A}_{o't} \hat{w}_{o't}^{-\theta} \hat{\tau}_{o'dt}^{-\theta} \right)^{\frac{-1}{\theta}}$$

D2. Dynamics

One other important component of the migration decision may be dynamic in nature: a location has benefits both today, but also in the future, given that the individual should be expecting to re-optimize location in the following period. While the main focus of our analysis is through the lens of a static model, it is easy to extend the model to incorporate dynamics. Our estimation strategy is robust to the presence of a dynamic component of utility (the "continuation value", below). However, our counterfactuals do not account for any dynamic benefits of roads. In that sense, our counterfactuals are an underestimate of the cumulative effect of roads through repeated migration decisions.

Following Artuç, Chaudhuri and McLaren (2010),[39] at a given time t, the (log) utility flow that worker i living in o and moving to d enjoys is:

$$\log U_{odt}(i) \equiv \tilde{U}_{odt}(i) = \tilde{B}_{dt} + \tilde{w}_{dt} + \tilde{b}_{dt}(i) - \tilde{\kappa}_{odt},$$

[40] where $\tilde{B}_{dt} = \log B_{dt}$, $\tilde{w}_{dt} = \log w_{dt} - \alpha \log P_{dt} - (1-\alpha) \log r_{dt}$ and $\tilde{b}_{dt}(i) = \log b_{dt}(i)$. Since $b_{dt}(i)$ has a Frechet distribution, $\tilde{b}_{dt}(i)$ has a Gumbel distribution.

[39] Caliendo and Parro (2015) adopts this methodology to study the dynamic effects in the US from increased trade with China.

[40] Artuç, Chaudhuri and McLaren (2010) assumes that a worker starting in location o at time t enjoys the real wage at the origin, $\tilde{w}_{ot}$ and then pays the cost of migrating to the destination d. Unlike them, we assume that the worker who chooses to migrate from o to d at time t enjoys the real wages at the destination, $\tilde{w}_{dt}$.

In a dynamic model, workers also take into account the expected future value of living in the destination d given their information set at time t, $\beta E_t V_{d,t+1}$. Therefore, the gross flow of workers that migrate from o to d at time t is given by:

$$M_{odt} = \frac{\exp(\tilde{B}_{dt} + \tilde{w}_{dt} + \beta E_t V_{d,t+1} - \tilde{\kappa}_{odt})}{\sum_{s \in N} \exp\left(\tilde{B}_{st} + \tilde{w}_{st} + \beta E_t V_{s,t+1} - \tilde{\kappa}_{ost}\right)} \times L_{ot}. \tag{D1}$$

The estimating equation from this model is exactly the same as our main gravity equation. The only difference from the original model is the continuation value $\beta E_t V_{d,t+1}$. The continuation value is isomorphic to an amenity value of location d and is included in the fixed effect terms, denoted as $\hat{\delta}^{\kappa}_{dt}$:

$$\log M_{odt} = \hat{\delta}^{\kappa}_{dt} + \hat{\delta}^{\kappa}_{ot} - \epsilon\mu \log(\text{road travel time}_{odt}) + \varepsilon^{\kappa}_{odt}.$$