Separate Appendices with Supplemental Material for:

Roads and Trade: Evidence from the US

Gilles Duranton*
University of Pennsylvania
Peter M. Morrow‡
University of Toronto
Matthew A. Turner‡
University of Toronto

19 July 2013

Abstract: This document contains a set of appendices with supplemental material.

Keywords: interstate highways, transportation costs, trade and specialization.

JEL classification: F14, R41, R49

*Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, USA (e-mail: duranton@wharton.upenn.edu; website: https://real-estate.wharton.upenn.edu/profile/21470/). Also affiliated with the Center for Economic Policy Research and the Spatial Economic Centre at the London School of Economics.

‡Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: peter.morrow@utoronto.ca; website: http://www.economics.utoronto.ca/morrow/).

‡Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: mturner@chass.utoronto.ca; website: http://www.economics.utoronto.ca/mturner/index.htm).

Introduction

This document complements "Roads and Trade: Evidence from the US" by the same authors. It contains a number of extensions and many robustness checks not included in the main paper.

- Appendix D presents alternative first-step estimations.
- Appendix E explores the robustness of our railroad instrument in light of the possible long run effects of railroads on the production structure of cities. This is the first of six appendices presenting detailed results for the robustness checks on the second step of our estimation summarized in section 7.2 (main text).
- Appendix F considers a number of alternative control variables which may be correlated with our instruments and with the propensity of cities to export weight or value.
- Appendix G investigates the robustness of our results to alternate measures of interstate highways and to our choice of functional form for this variable.
- Appendix H assesses the robustness of our results to alternative measures of market potential and to the possible simultaneity of export and import market access.
- Appendix I considers alternative second-step dependent variables constructed from different first-step estimations.
- Appendix J compares the results our two-step estimation with those of a single step estimation.
- Appendix K provides supplementary results on the relationship between short and long distance trade and within-city roads.
- Appendix L recovers some key parameters of a simplified version of our model with additive trade costs. This allows us to assess the possible biases caused by our multiplicative specification for trade costs.
- Appendix M derives some predictions from our model for trade internal to cities.
- Appendix N presents additional specialization results.

Appendix D. Alternative first-step results

Appendix table 1 mirrors table 3 (main text) but uses all trade (i.e., road and rail trade) between pairs of cities. We note that the distance decay of trade is marginally smaller for all trade than for road trade. This is in large part due to the fact that road trade represents most of all trade in the data.

Appendix Table 1: Alternative first-step results, all trade

	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) OLS
Panel A. Dependent variable: Weight of bilateral trade flows, all trade.								
log(hwy. dist.)	-1.85***	-1.85***	-2.90***	-2.90***	-3.25	-1.23		-1.83***
	(0.024)	(0.023)	(0.17)	(0.17)	(6.57)	(6.31)		(0.21)
log(hwy. dist.)^2			0.084***	0.084***	0.62	0.13		
			(0.014)	(0.013)	(1.68)	(1.62)		
log(hwy. dist.)^3					-0.11	-0.056		
					(0.19)	(0.18)		
log(hwy. dist.)^4					0.0062	0.0043		
					(0.0075)	(0.0073)		
log(Euclid dist.)							-1.86***	-0.020
							(0.025)	(0.21)
Mean effect	-1.85	-1.85	-1.68	-1.68	-1.59	-1.58	-1.86	-1.85
Median effect	-1.85	-1.85	-1.66	-1.67	-1.68	-1.67	-1.86	-1.85
R^2	0.84	-	0.85	-	0.85	-	0.84	0.84
First-stage F		157,104		19,430		2,121		
Panel B. Dependent variable: Value of bilateral trade flows, all trade.								
log(hwy. dist.)	-1.24***	-1.24***	-2.66***	-2.65***	-7.45	-5.85		-1.45***
	(0.020)	(0.020)	(0.15)	(0.15)	(5.33)	(5.15)		(0.17)
log(hwy. dist.)^2			0.11***	0.11***	1.91	1.51		
			(0.011)	(0.011)	(1.35)	(1.30)		
log(hwy. dist.)^3					-0.26*	-0.22		
					(0.15)	(0.14)		
log(hwy. dist..)^4					0.013$^{\{**\}}$	0.011*		
					(0.0059)	(0.0058)		
log(Euclid dist.)							-1.23***	0.21
							(0.021)	(0.16)
Mean effect	-1.24	-1.24	-1.03	-1.03	-0.91	-0.91	-1.23	-1.24
Median effect	-1.24	-1.24	-1.00	-1.01	-1.05	-1.05	-1.23	-1.24
R^2	0.84	-	0.85	-	0.85	-	0.83	0.85
First-stage Stat.		167,755		19,518		2,265		

Notes: All regressions include importer and exporter fixed effects for all cities. The same regressions are run in both panels with different dependent variables: weight of trade flows for panel A and value of trade flows for panelB. Regressions in panel A are based on 2,639 observations and 3,299 observations in panel B. In column 2, 4 and 6, highway distance terms are instrumented by their corresponding 1947 planned highway and 1898 railroad distance terms. To measure the distance of a city to itself, we take $0.66(\text{area}/\pi)^{0.5}$ as Redding and Venables (2004). Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

Appendix table 2 also repeats table 3 (main text) but uses rail trade instead of road trade and adjusts the distance variables appropriately. We note that the results must be interpreted

Appendix Table 2: Alternative first-step results, rail trade

	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) OLS
Panel A. Dependent variable: Weight of bilateral trade flows, road trade.								
log(rail dist.)	-0.57***	-0.60***	-1.65**	-1.73***	31.7	42.1*		-1.47
	(0.094)	(0.070)	(0.79)	(0.59)	(32.8)	(23.9)		(1.38)
log(rail dist.)^2			0.091	0.096*	-9.10	-11.7*		
			(0.067)	(0.050)	(8.56)	(6.23)		
log(rail dist.)^3					1.09	1.36{*}		
					(0.97)	(0.70)		
log(rail dist.)^4					-0.047	-0.057{ ** }		
					(0.040)	(0.029)		
log(Euclid dist.)							-0.59***	0.93
							(0.099)	(1.44)
Mean effect	-0.57	-0.60	-0.32	-0.33	-0.58	-0.56	-0.59	-0.54
Median effect	-0.57	-0.60	-0.30	-0.32	-0.32	-0.28	-0.59	-0.54
R^2	0.75	-	0.75	-	0.76	-	0.75	0.75
First-stage F		10,828		1,115		346		
Panel B. Dependent variable: Value of bilateral trade flows, road trade.								
log(2004 railroad dist.)	-0.51***	-0.51***	-1.83**	-1.82***	85.9**	88.4***		-4.55***
	(0.088)	(0.067)	(0.73)	(0.56)	(41.2)	(31.7)		(1.09)
log(2004 railroad dist.)^2			0.11*	0.11**	-22.1**	-22.7***		
			(0.060)	(0.046)	(10.5)	(8.07)		
log(2004 railroad dist.)^3					2.43**	2.48***		
					(1.16)	(0.89)		
log(2004 railroad dist.)^4					-0.097**	-0.099***		
					(0.047)	(0.036)		
log(euclid dist.)							-0.50***	4.22***
							(0.093)	(1.14)
Mean effect	-0.51	-0.51	-0.22	-0.23	-0.26	-0.25	-0.50	-0.33
Median effect	-0.51	-0.51	-0.20	-0.21	0.01	0.02	-0.50	-0.33
R^2d	0.72	-	0.73	-	0.74	-	0.71	0.74
First-stage F		38526		8511		1096		

Notes: All regressions include importer and exporter fixed effects for all cities. The same regressions are run in both panels with different dependent variables: weight of trade flows for panel A and value of trade flows for panelB. Regressions in panel A are based on 245 observations and 269 observations in panel B. In column 2, 4 and 6, rail distance terms are instrumented by the corresponding 1898 railroad distance terms. To measure the distance of a city to itself, we take $0.66(\text{area}/\pi)^{0.5}$ as Redding and Venables (2004). Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

with caution since they are based on far fewer observations than the other first-stage tables. The coefficient on distance in the simple specifications is much lower than for road trade which is

Appendix Table 3: Alternative first-step results, road trade with no internal distance correction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	TSLS	OLS	TSLS	OLS	TSLS	OLS	OLS
Panel A. Dependent variable: Weight of bilateral trade flows, road trade.								
log(hwy. dist.)	-1.17***	-1.17***	-0.047	-0.046	0.51	0.35		0.78***
	(0.027)	(0.026)	(0.049)	(0.048)	(0.66)	(0.66)		(0.24)
log(hwy. dist.)^2			-0.13***	-0.13***	-0.013	0.066		
			(0.0052)	(0.0050)	(0.30)	(0.30)		
log(hwy. dist.)^3					-0.066	-0.079*		
					(0.046)	(0.046)		
log(hwy. dist.)^4					0.0055**	0.0061***		
					(0.0023)	(0.0023)		
log(Euclid dist.)							-1.20***	-1.99***
							(0.026)	(0.24)
Mean effect	-1.17	-1.17	-1.89	-1.89	-1.57	-1.57	-1.20	-1.21
Median effect	-1.17	-1.17	-1.93	-1.93	-1.70	-1.71	-1.20	-1.21
R^2	0.80	-	0.85	-	0.87	-	0.81	0.81
First-stage F		652,994		119,915		571		
Panel B. Dependent variable: Value of bilateral trade flows, road trade.								
log(hwy. dist.)	-0.88***	-0.88***	-0.037	-0.036	0.13	0.048		0.59***
	(0.021)	(0.021)	(0.041)	(0.040)	(0.60)	(0.60)		(0.18)
log(hwy. dist..)^2			-0.095***	-0.095***	0.14	0.18		
			(0.0044)	(0.0043)	(0.28)	(0.28)		
log(hwy. dist..)^3					-0.076*	-0.083**		
					(0.042)	(0.042)		
log(hwy. dist.)^4					0.0056***	0.0060***		
					(0.0021)	(0.0021)		
log(Euclid dist.)							-0.89***	-1.49***
							(0.021)	(0.18)
Mean effect	-0.88	-0.88	-1.40	-1.40	-1.14	-1.14	-0.89	-0.90
Median effect	-0.88	-0.88	-1.43	-1.43	-1.25	-1.25	-0.89	-0.90
R^2	0.78	-	0.82	-	0.83	-	0.79	0.79
First-stage F		645,892		126,389		605		

Notes: All regressions include importer and exporter fixed effects for all cities. The same regressions are run in both panels with different dependent variables: weight of trade flows for panel A and value of trade flows for panelB. Regressions in panel A are based on 2,476 observations and 2,705 observations in panel B. In column 2, 4 and 6, highway distance terms are instrumented by their corresponding 1947 planned highway and 1898 railroad distance terms. All distances of cities to themselves are set to unity. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

consistent with the notion that rail is used for trade over longer distances.

Appendix table 3 mirrors table 3 (main text) but does not correct for internal distances when trade flows internal to cities are considered. Recall that in table 3 we measure the distance of a city to itself as $0.66(\text{area}/\pi)^{0.5}$. This has become standard practice in gravity estimation following Redding and Venables (2004). Because cities trade a lot with themselves (main text table 2) this may have an effect on the results.

A comparison with table 3 (main text) immediately reveals that ignoring internal distances leads to a muted effect of distance on both the weight and the value of trade. For instance, ignoring internal distances leads to a coefficient on distance of -1.17 in column 2 without internal distances. With internal distances taken into account, this coefficient is -1.90. The coefficient on distance is thus fairly sensitive to whether internal distances are taken into account. Fortunately, the exporter and importer fixed effects are not. The pairwise correlations between the exporter and importer value fixed effects of table 3 (main text) and their corresponding estimations in appendix table 3 are all above 0.95 and yield similar second-stage results. Hence, even though gravity estimates are sensitive to the inclusion of internal distances, our estimates for the importer and exporter fixed effects are not.

Appendix E. 1898 railroads and the persistence of manufacturing

The results of table 6 panel A (main text) that 1898 railroad instrument yields somewhat higher coefficients for within-city highways than the other two instruments, old exploration routes in particular. Panel B of the same table also shows that these minor discrepancies mostly disappear when we control for the log share of contemporaneous manufacturing employment. Nonetheless, it remains possible that 2007 export specialization patterns may not be caused by contemporaneous within-city highways but instead reflect a legacy of specialization from 1898 railroads and perhaps old exploration routes. If so, in the worst case scenario, the 1947 planned highways might simply reflection of these early patterns of manufacturing specialization and be invalid as an instrument as well. Fortunately, our data permit us to examine these possibilities.

Appendix table 4 augments our main TSLS specifications reported in table 5 (main text) by adding 1956 manufacturing variables to control for persistence in patterns of manufacturing specialization. Panel A uses the log share of manufacturing employment in 1956. Interestingly this coefficient is highly significant in all specifications and is large in magnitude. A city with twice the share of employment in manufacturing in 1956 exports 40% more weight in 2007 with the coefficient on manufacturing employment of 0.49 in column 4 or more than 60% more with the coefficient of 0.70 in column 3. This shows the strength of the manufacturing persistence. However, controlling for 1956 manufacturing employment has only a small effect on the coefficients of within-city highways. In our preferred specification of column 3, we obtain an elasticity of 0.35 instead of 0.47, a difference that is not statistically significant. For the most complete specification, in column 4, we obtain the marginally lower elasticity of 0.35 instead of 0.39. The results are unchanged for trade in value: the effect of within-city highways remains insignificant.

Panel B of appendix table 4 repeats a similar exercise but uses an index of 1956 manufacturing weight instead of 1956 manufacturing employment. This index is computed by multiplying 1956 employment shares for each manufacturing sector by 2007 weight per unit value for these sectors

Appendix Table 4: Second-step results, OLS for exporter fixed effects with controls for 1956 employment

Exporter fixed effect	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
Panel A. Adding 1956 employment.								
log highway km	1.05***	0.36***	0.35***	0.35***	0.98***	-0.075	-0.080	-0.082
	(0.15)	(0.12)	(0.12)	(0.12)	(0.17)	(0.13)	(0.13)	(0.12)
log 1956 % manuf. emp.	0.34***	0.67***	0.70***	0.49***	0.54***	0.80***	0.95***	0.72***
	(0.12)	(0.11)	(0.14)	(0.18)	(0.13)	(0.095)	(0.14)	(0.16)
Controls.	0	1	2	3	0	1	2	3
Overid. p-value	0.37	0.68	0.28	0.33	0.42	0.93	0.62	0.67
First-stage Stat.	120	82.6	73.9	79.3	120	82.6	73.9	79.3
Panel B. Adding 1956 manufacturing weights.								
log highway km	1.13***	0.57***	0.45***	0.37***	1.05***	0.19	0.095	-0.0030
	(0.15)	(0.16)	(0.13)	(0.13)	(0.19)	(0.15)	(0.13)	(0.12)
1956 manuf. weight index	-0.00068	0.0075	0.12	0.13	-0.38**	-0.26*	-0.19	-0.17**
	(0.13)	(0.13)	(0.10)	(0.11)	(0.18)	(0.15)	(0.12)	(0.074)
Overid. p-value	0.10	0.043	0.16	0.37	0.069	0.064	0.21	0.42
First-stage Stat.	74.1	89.8	76.3	80.9	74.1	89.8	76.3	80.9
Panel C. TSLS estimations with controls for manufacturing employment.								
log highway km	1.06***	0.33***	0.32***	0.32***	0.97***	-0.063	-0.067	-0.065
	(0.15)	(0.12)	(0.12)	(0.12)	(0.17)	(0.12)	(0.13)	(0.12)
log 1956 % manuf. emp.	0.38***	0.73***	0.76***	0.58***	0.44***	0.77***	0.93***	0.67***
	(0.13)	(0.12)	(0.14)	(0.19)	(0.17)	(0.092)	(0.14)	(0.16)
1956 manuf. weight index	0.11	0.18**	0.21**	0.19**	-0.25	-0.074	-0.071	-0.096*
	(0.12)	(0.086)	(0.087)	(0.097)	(0.18)	(0.061)	(0.052)	(0.055)
Overid. p-value	0.40	0.87	0.41	0.45	0.31	0.87	0.52	0.56
First-stage Stat	109	79.7	70.6	75.6	109	79.7	70.6	75.6

Notes: 66 observations per column. All regressions include a constant. The set controls 0 is a constant; 1 adds log 2007 employment and export market access; 2 further adds log population for 1920, 1950 and 2000; 3 also considers log share of 2003 manufacturing employment. All regressions use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

and summing across sectors. Since manufacturing weights per unit value have changed since 1956, this variable must be viewed with caution. However, as the results in this panel show, including the index of 1956 manufacturing weight makes little difference to the estimates of the elasticity of weight exported with respect to within-city highways. Panel C of appendix table 4 considers both 1956 manufacturing employment and our 1956 manufacturing weight index. The results are similar to those of panel A. For exported weight, we find that there is a strong persistence in the patterns of manufacturing employment. If anything, the effect of 1956 manufacturing employment is stronger than in panel A and our weight index becomes significant. At the same time, the coefficient of within-city highways changes only slightly, from 0.35 to 0.32. In sum, we find no evidence of an effect of our historical instruments on the propensity to export weight except through their effect on roads.

Given that manufacturing patterns are persistent between 1956 and 2007, the results above require that 1898 railroads do not predict 1956 manufacturing patterns. Although it is difficult to provide strong evidence for the absence of an effect, we can replicate the specialisation exercise conducted in section 9 (main text) using historical rather than contemporary data. For this exercise we first regress the log of 1956 employment in each city on the log of 1898 railroads across all cities for each sector in turn. This allows us to estimate the elasticity of 1956 employment for each manufacturing sector with respect to the extent of railroads in 1898. We next use this elasticity as a dependent variable and regress it on weight per unit value. This allows us to assess whether sectors that produce heavier goods employed relatively more workers in 1956 in those cities with more railroads in 1898. Replicating the specifications of table 10 (main text) using 1956 (instead of 2007) manufacturing employment and 1898 railroads (instead of 2007 within-city highways) fails to yield a significant coefficient in any of the specifications. Overall our findings suggest that despite some persistence in manufacturing patterns since the beginning of the interstate highway system in 1956, the effect of our instruments on 1956 manufacturing patterns are too weak to affect our results.

Appendix F. Alternative controls

In appendix table 5, we experiment with including a variety of additional control variables to our preferred estimates of the effect of within-city highways on the propensity to export.

It is natural to suspect that geography might affect both our instruments and productivity: our three instruments could be correlated with some geographical feature of cities which also affects their propensity to trade. In columns 1 and 2 of panel A of table 5, we duplicate our preferred OLS specification with an additional geographical control, log kilometers to the nearest Ocean, Gulf or Great Lake ('Water') and an index of the average land gradient. In panel B, we repeat this for our preferred TSLS specification. In panel C, we repeat this TSLS specification but also include the log share of manufacturing employment as a control. While the proximity to a body of water leads to a slightly smaller coefficient for within-city highways, adding average slope makes virtually no difference. In column 3, we take another approach to geography and introduce dummy variables for each census region. This leads to slightly smaller coefficients for within-city highways in all three panels relative to the corresponding estimates in table 4 or 5 (main text).

Another possibility is that our three instruments also affect socio-economic characteristics of cities and that these characteristics in turn affect trade. To investigate this possibility, in columns 4 and 5, we control for the log of the share of college population with at least a college degree and log income per capita. Including either of these variables leads to slightly smaller coefficients on within-city highways, although the change is not statistically significant. It is also possible that cities with more highways export more weight because they serve as logistical centres for

Appendix Table 5: Second-step results for weight exporter fixed effects with extra controls

Added var.	(1) Water	(2) Slope	(3) Census div.	(4) % college	(5) Income p.c.	(6) % wholesale	(7) Traffic	(8) All
Panel A. OLS estimations.								
log highway km	0.30*	0.38**	0.28*	0.26*	0.31**	0.42***	0.37*	0.19
	(0.15)	(0.15)	(0.16)	(0.14)	(0.15)	(0.14)	(0.19)	(0.17)
R^2	0.79	0.79	0.81	0.81	0.80	0.82	0.79	0.90
Panel B. TSLS estimations.								
log highway km	0.37**	0.47***	0.39***	0.36***	0.42***	0.51***	0.49**	0.45***
	(0.16)	(0.14)	(0.15)	(0.13)	(0.14)	(0.13)	(0.19)	(0.17)
Overid. p-value	0.11	0.20	0.13	0.19	0.14	0.36	0.12	0.46
First-stage Stat.	70.9	86.6	65.5	83.9	81.0	78.5	56.9	27.5
Panel C. TSLS estimations with controls for manufacturing employment.								
log highway km	0.34**	0.41***	0.35***	0.33***	0.36***	0.43***	0.42**	0.45***
	(0.15)	(0.13)	(0.13)	(0.12)	(0.13)	(0.11)	(0.16)	(0.17)
Overid. p-value	0.27	0.46	0.10	0.25	0.25	0.49	0.27	0.46
First-stage Stat.	79.9	90.3	70.0	87.0	82.7	83.6	62.3	27.5

Notes: 66 observations per column. All regressions include a constant, log 2007 employment, export market access, and log population for 1920, 1950 and 2000. All TSLS regressions in panel B use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. Robust standard errors in parentheses. The TSLS regressions in panel C also include the log of the share of manufacturing employment in 2003.
As extra controls, column 1 includes average distance to the nearest body of water. Column 2 includes average land gradient. Column 3 includes dummy variables for census regions. Column 4 includes the log share of the fraction of adult population with a college degree or more. Column 5 includes the log of average income per capita. Column 6 includes the log of the share of employment in wholesale trade. Column 7 includes the log of average daily traffic on the interstate highways in 2005. Column 8 adds all the extra variables of columns 1 to 7 together. ***, **, *: significant at 1%, 5%, 10%.

the trucking and warehousing industries (although it would be odd to observe such an effect on the weight but not value of trade). In column 6, we introduce the log share of employment in wholesale trade as a control. Relative to the results of tables 4 and 5 (main text), we find that this extra control leads to slightly larger coefficients for within-city highways.

In column 7, we control for congestion on within-city highways by introducing a control that measures the log of average daily traffic on these roads. The coefficients on within-city highways are almost the same in all three panels. In column 8, we consider all of these extra variables together. While the coefficient on within-city highways in the OLS specification of panel A is smaller and becomes insignificant, the TSLS coefficient estimates of panels B and C remain highly significant and close to their corresponding values in columns 3 and 4 of table 5 (main text) without the extra controls. These regressions pass overidentification tests easily and the instruments are not weak.

Appendix Table 6: Second-step results for weight exporter fixed effects with extra controls

Added var.	(1) Water	(2) Slope	(3) Census div.	(4) % college	(5) Income p.c.	(6) % wholesale	(7) Traffic	(8) All
Panel A. TSLS estimations using log 1947 planned highway km.								
log highway km	0.34**	0.39***	0.32**	0.29**	0.35**	0.41***	0.42**	0.39**
	(0.16)	(0.14)	(0.14)	(0.13)	(0.14)	(0.12)	(0.18)	(0.18)
First-stage Stat.	127	131	117	113	121	134	77.7	52.0
Panel B. TSLS estimations using log 1898 railroad km.								
log highway km	0.66**	0.57***	0.62***	0.56**	0.63***	0.59***	1.00*	0.73**
	(0.26)	(0.20)	(0.17)	(0.23)	(0.22)	(0.21)	(0.52)	(0.35)
First-stage Stat.	25.3	44.5	58.0	36.8	37.0	39.4	13.4	19.7
Panel C. TSLS estimations using log 1528-1850 exploration routes index.								
log highway km	0.24	0.46**	0.35*	0.41**	0.35*	0.49***	0.34	0.58**
	(0.15)	(0.13)	(0.13)	(0.12)	(0.13)	(0.11)	(0.16)	(0.17)
First-stage Stat.	11.0	22.7	22.7	29.7	26.5	23.3	44.6	15.6

Notes: 66 observations per column. All regressions include a constant, log 2007 employment, export market access, log population for 1920, 1950 and 2000 and the log share of 2003 manufacturing employment. Robust standard errors in parentheses.
As extra controls, column 1 includes average distance to the nearest body of water. Column 2 includes land gradient. Column 3 includes dummy variables for census regions. Column 4 includes the log share of the fraction of adult population with a college degree or more. Column 5 includes the log of average income per capita. Column 6 includes the log of the share of employment in wholesale trade. Column 7 includes the log of average daily traffic on the interstate highways in 2005. Column 8 adds all the extra variables of columns 1 to 7 together. ***, **, *: significant at 1%, 5%, 10%.

In appendix table 6 , we perform an even more demanding exercise where we replicate panel C of appendix table 5 using just one instrument at a time. We find that in most cases, none of the instruments is sensitive to these added variables. Even when these extra control variables are added together in column 8, each and every instrument remains strong and leads to a positive and significant coefficient between 0.4 and 0.7 for within-city highways.

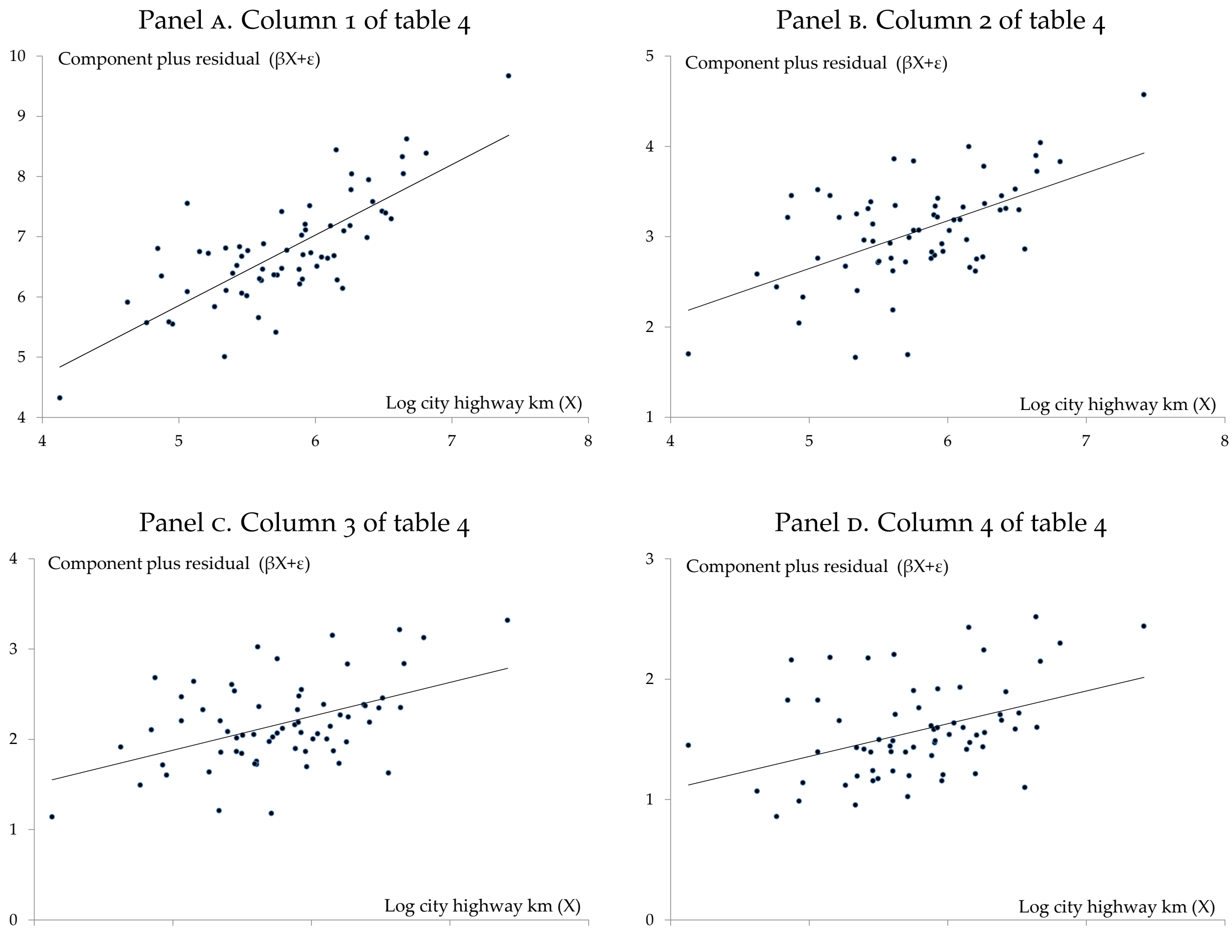
Appendix G. Alternate highway measures

We here investigate the effects of alternate functional forms and alternate measures for within-city highways.

Starting with functional forms, we note that if we add higher order terms for within-city highways in the OLS and TSLS specifications of table 4 and 5 (main text) we find higher order terms are sometimes significant (details not reported).

The four panels of appendix figure 1 display residuals plots that correspond to columns 1 to 4 of table 4 (main text). For these four OLS regressions, the X-axis is log city highway kilometers whereas the Y-axis is the residual of the regression plus the component associated with log city highway kilometers.[1] In all four graphs, a linear specification provides a good description of the relationship between log city highway kilometers and its augmented residual.

Appendix Figure 1: Residual OLS plots

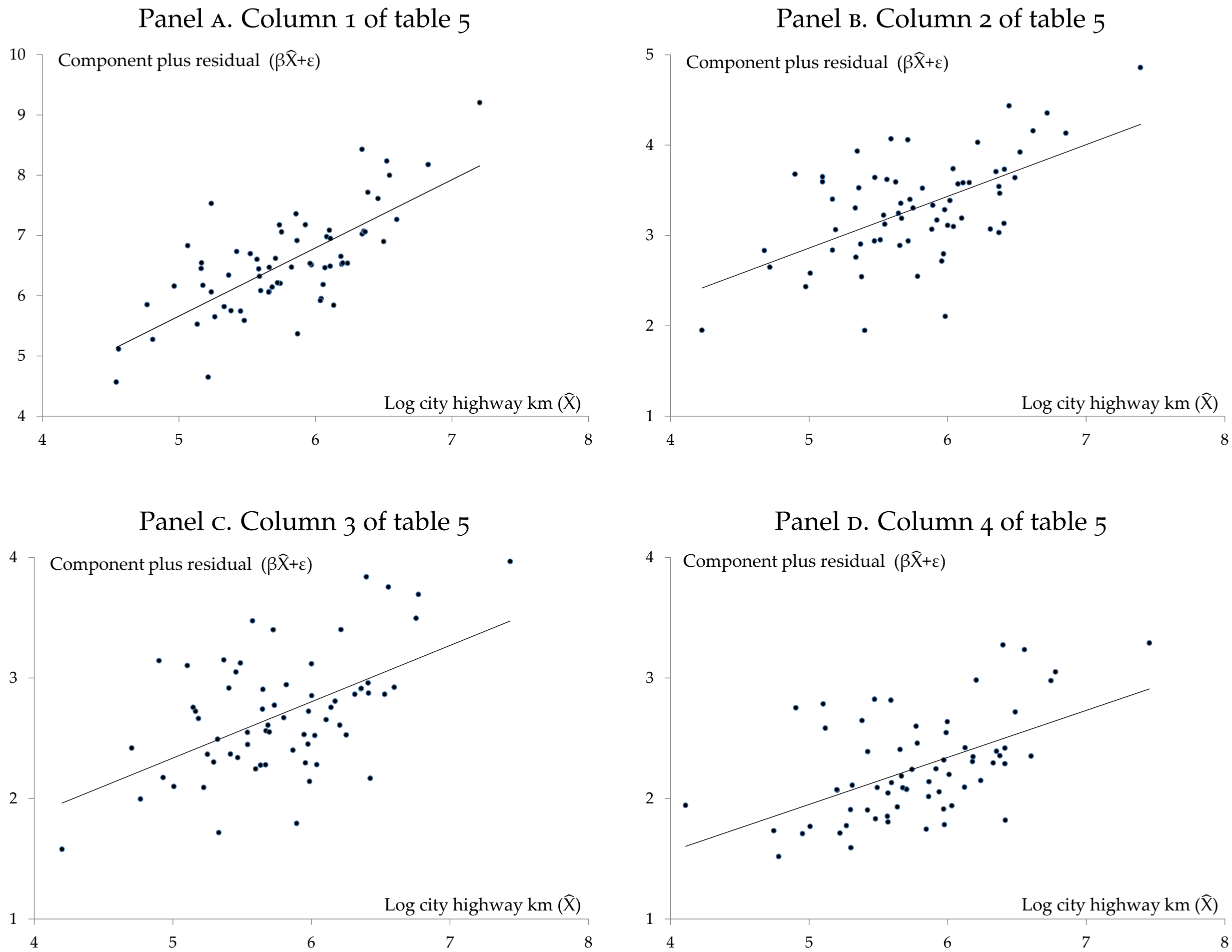


Notes: The horizontal axis plots log city highway kilometers. The vertical axis plots the residual of an OLS regression plus log city highway kilometers multiplied by its estimated coefficient.

The four panels of appendix figure 2 display residuals plots that correspond to columns 1 to 4 of table 5 (main text). For these four TSLS regression, the X-axis is log city highway kilometers as predicted by the first-stage regression whereas the Y-axis is the residual of the regression plus the component associated with predicted log city highway kilometers. For all four graphs, as in appendix figure 1, a linear specification provides a good description of the relationship between predicted log city highway kilometers and its augmented residual.

[1] That is, the quantity on the Y-axis is log city highway kilometers times its coefficient estimated in the corresponding column of table 4 plus the residual of the same regression.

Appendix Figure 2: Residual TSLS plots



Notes: The horizontal axis plots log city highway kilometers. The vertical axis plots the residual of a TSLS regression plus log city highway kilometers as predicted by the first stage of the regression multiplied by its estimated coefficient.

In appendix table 7 we investigate the effects of alternate measures of within-city roads on trade. In column 1 of panels A and B, we duplicate our preferred specifications from column 3 of tables 4 and 5 (main text) but use the log of the number of radial rays of interstate highways to measure within-city roads. The OLS coefficient in panel A is slightly smaller than its corresponding coefficient in table 4 main text). To estimate the TSLS coefficient in panel B, we use the number of rays of 1898 railroads and 1947 planned interstate highways as instruments. This TSLS coefficient is also slightly smaller than its counterpart in table 5 (main text). In column 2, we retain the same road variable and add the log share of manufacturing employment as a control. The OLS and TSLS results are close to the corresponding results of tables 4 and 5 (main text).

In columns 3 and 4, we use lane kilometers of within-city highway (instead of highway kilometers) as our measure of roads. While the OLS coefficients in panel A are smaller than the corre-

Appendix Table 7: Second-step results for weight exporter fixed effects with alternate measures of city highways

Measure of city roads	(1) log highway rays	(2)	(3) log highway lane km, 2007	(4)	(5) log highway urban km, 2007	(6)	(7) log highway km, 1987	(8)
Panel A. OLS estimations.								
Road var.	0.30***	0.33***	0.25**	0.21*	0.25	0.27*	0.11*	0.11**
	(0.11)	(0.094)	(0.12)	(0.11)	(0.16)	(0.15)	(0.057)	(0.050)
% manuf. emp.	N	Y	N	Y	N	Y	N	Y
R^2	0.79	0.86	0.78	0.84	0.77	0.84	0.78	0.84
Panel B. TSLS estimations.								
Road var.	0.41***	0.40***	0.55***	0.44***	0.95***	0.77***	0.41***	0.32***
	(0.15)	(0.12)	(0.17)	(0.15)	(0.34)	(0.29)	(0.13)	(0.12)
% manuf. emp.	N	Y	N	Y	N	Y	N	Y
Overid. p-value	0.63	0.14	0.47	0.68	0.20	0.32	0.18	0.20
First-stage Stat.	74.8	73.5	21.4	20.4	5.09	4.99	19.2	17.7

Notes: 66 observations per column. All regressions include a constant and log 2007 employment, log market access and log past populations in 1920, 1950 and 2000 as controls. The TSLS regressions of columns 1 and 2 in panel B use log rays of 1947 planned highway km and log rays of 1898 railroad km as instruments for log rays of interstate highways. The other TSLS regressions use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

sponding coefficients in tables 4 and 5 (main text), the TSLS coefficients are close to our preferred estimates. In columns 5 and 6, our measure of roads is the log of within-city interstate kilometers in urbanized areas within the city. While the OLS coefficients are smaller than the corresponding coefficients in table 4 (main text) and marginally significant, the TSLS results are much larger than in table 5 (main text). Since our instruments are weak for highways in urbanized areas, no conclusion should be drawn from these estimates. Finally, in columns 7 and 8 we lag our preferred measure of roads by 20 years. As expected, OLS coefficients are slightly smaller than for contemporaneous highways. TSLS results are much closer, since TSLS corrects for mismeasurement.

Appendix table 7 shows that our main results do not depend on a specific road variable or on the exact functional form specification we use. We obtain similar results for several different measures of within-city highways. Experiments with specifications which jointly estimate the effect of more than one road measure were inconclusive due to strong collinearity between the two variables. Thus, our data provide strong evidence that expansions of the within-city highway network increases the weight of exports, but cannot shed light on the precise aspect of the road network that causes this increase.

Appendix H. Market access

We now focus on two problems posed by the export market access variables. First, the calculation of market access for city i in equation (14 - main text) sums the importer effect of all cities including i. Since exporter and importer market access are strongly correlated, this raises an obvious endogeneity concern. Second, market access for city i depends on the importer effect of city j, which is itself affected by city i. Ignoring own-city effects in the construction of market access is an imperfect solution.

Appendix Table 8: Second-step results, robustness checks for market access

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) TSLS	(6) TSLS	(7) TSLS	(8) TSLS
log highway km	0.53**	0.28*	0.33**	0.76***	0.44***	0.44***	0.44***	0.33
	(0.21)	(0.14)	(0.14)	(0.19)	(0.14)	(0.13)	(0.13)	(0.40)
Market access (export)		-0.52***	-0.96***	0.57***	-0.74***	-0.78***	-0.74***	-0.75
		(0.10)	(0.18)	(0.14)	(0.16)	(0.16)	(0.16)	(0.66)
R^2	0.69	0.81	0.80	0.74				
Overid. p-value					0.17	0.18	0.31	0.024
First-stage Stat.					45.5	41.7	35.7	1.46

Notes: 66 observations per column. All regressions also use log 2007 employment and log population for 1920, 1950 and 2000 as explanatory variables and include a constant. All TSLS regressions use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways.
In column 2, market access is constructed by summing the income of neighbouring cities weighted by their inverse distance. In column 2, we use 1920 population instead of income. In column 4, we add a city's own fixed effect in the summation of the market access. In column 5, we use our standard measure of market access but instrument it along with city highways. Our instruments are log 1947 planned highway km, log 1898 railroad km, log 1528-1850 exploration routes index and our ad-hoc measure of market access based on income. In column 6, we replace this instrument by the ad-hoc market access term computed from 1920 populations. In column 7, we use these two instruments together. In column 8, the instrumented market access contains a city's own fixed effect. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

Appendix table 8 checks the robustness of our results with respect to the export market access variable. Columns 1 to 4 duplicate our preferred OLS specification from column 3 of table 4 (main text), but make changes to the market access variable. Column 1 omits it. This leads to a higher coefficient, by about one standard deviation, for within-city highway kilometers. Column 2 replaces our theory driven market access variable with an ad-hoc market access measure. This ad hoc market access measure is defined as the sum of aggregate income in other cities with the contribution of each city weighted by distance to the power -0.9.[2] This lowers the coefficient on within-city highway kilometers and the coefficient on market access is smaller in magnitude than in column 3 of table 4 (main text). Column 4 introduces an alternative ad-hoc market access, this time computed from 1920 populations. The coefficient on within-city highways increases slightly

[2]-0.9 is the mean of all distance decay rates for trade in Disdier and Head's (2008) meta analysis.

and the market access coefficient nearly doubles. In column 4 we include a city's own fixed effect in our calculation of market access. Since most cities have large internal markets, the resulting market access variable is endogenous. As a result the coefficient on market access changes sign and becomes positive and large. The coefficient on highway kilometers also increases.

Columns 5 to 8 duplicate our preferred OLS specification of column 3 of table 5 (main text) but also instrument the market access variable. In column 5, our standard market access variable is instrumented by the ad-hoc market access variables computed from city incomes. The coefficient on within-city highways is close to that of column 3 of table 5 (main text). The coefficient on market access is larger in magnitude. Column 6 repeats this specification using instead ad-hoc market access computed from 1920 populations. The results are almost exactly the same. Column 7 considers these two market access instruments jointly and the results are again the same. Finally, column 8 uses as an endogenous explanatory variable a market access term which includes a city's own fixed effect. Unfortunately our two external market access instruments do not predict this total market access term well and the instruments are extremely weak. Although insignificant, the point estimates for both within-city highway kilometers and market access remain close to those in the previous specifications.

Appendix I. Alternative specifications for the first step

Panel A of appendix table 9 duplicates table 4 (main text) but uses exporter fixed effects estimated using weighted least squares (in the first step) as the dependent variable. Weighting the first step estimation can justified by the notion that the true unit of observation is a shipment of a given size. For trade in weight in columns 1 to 4, the results are close to those of table 4. For trade in value, the coefficients are higher. For the TSLS results of panel B, the coefficients on within-city highways are slightly higher for trade both in weight and value than the corresponding coefficients in table 5 (main text). We note that these higher coefficients are consistent with the results of table 7 (main text) which show that the effects of highways is stronger for short-distance trade. This is because our use of WLS in the first step puts more weight on the larger trade flows at shorter distances.

Panel C of appendix table 9 duplicates table 4 (main text) but uses as dependent variable exporter fixed effects estimated using a type-2 TOBIT (a.k.a., HECKIT) in the first step. Taking censoring explicitly into account can be justified by the fact that although there are few very true zeroes, censoring can bias the estimated coefficients on distance and thus indirectly affect the fixed effects. We are limited by the absence of a clear censoring threshold in the CFS data and by the absence of a an appropriate instrument that satisfies the exclusion restriction associated with this estimation. Simply put, we have no variable that would explain censoring but not the weight or value of trade. As a result, selection in the first stage is identified only from (non-linear) functional forms. Panel B performs the same exercise but applies it to the second-stage TSLS estimates of table 5 (main text). This yields slightly lower coefficients on within-city highway kilometers for trade in weight and similar coefficients for trade in volume, but these coefficients are generally statistically indistinguishable across specifications.

Appendix Table 9: Second-step estimation with alternative first-step estimations 1

Exporter fixed effect	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
Panel A. OLS with WLS first step.								
log highway km	1.08***	0.40**	0.38**	0.31*	1.26***	0.31*	0.21	0.10
	(0.14)	(0.17)	(0.18)	(0.17)	(0.15)	(0.17)	(0.14)	(0.13)
Controls.	0	1	2	3	0	1	2	3
R^2	0.59	0.77	0.79	0.82	0.62	0.82	0.85	0.91
Panel B. TSLS with WLS first step.								
log highway km	1.07***	0.54***	0.53***	0.48***	1.12***	0.28*	0.21*	0.13
	(0.14)	(0.17)	(0.16)	(0.16)	(0.16)	(0.15)	(0.13)	(0.11)
Overid. p-value	0.38	0.31	0.21	0.34	0.13	0.11	0.39	0.76
First-stage Stat.	97.5	89.0	79.5	84.3	97.5	89.0	79.5	84.3
Panel C. OLS with HECKIT first step.								
log highway km	0.80***	0.33**	0.25*	0.20	1.05***	0.11	0.000042	-0.095
	(0.12)	(0.13)	(0.13)	(0.12)	(0.15)	(0.17)	(0.15)	(0.15)
R^2	0.48	0.67	0.70	0.73	0.52	0.77	0.80	0.86
Panel D. TSLS with HECKIT first step.								
log highway km	0.79***	0.40***	0.36***	0.32***	0.92***	0.093	0.018	-0.049
	(0.12)	(0.13)	(0.13)	(0.12)	(0.16)	(0.15)	(0.14)	(0.13)
Overid. p-value	0.78	0.52	0.37	0.27	0.21	0.19	0.48	0.59
First-stage Stat.	97.5	90.3	80.4	85.4	97.5	90.3	80.4	85.4

Notes: 66 observations per column. The set controls 0 is a constant; 1 adds log 2007 employment and export market access; 2 further adds log population for 1920, 1950 and 2000; 3 also considers log share of manufacturing employment. All TSLS regressions use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

In the same spirit as the previous table, panels A and B of appendix table 10 also duplicate tables 4 and 5 (main text) but for the dependent variable uses exporter fixed effects estimated without correcting for internal distances in the case of trade internal to the city. This check is needed because the correcting for internal distance in table 3 of the main text (relative to not correcting for them and setting them to zero in log as done in appendix table 3 above) has a sizeable effect on the distance coefficient. The reported coefficients on within-city highway kilometers are all extremely close to their corresponding coefficients in tables 4 and 5 (main text). The differences between them are less than 0.03.

In panels C and D of appendix table 10, we use exporter fixed effects estimated in a first stage that excludes completely trade flows internal to cities. The results are again similar.

Finally, appendix table 11 explores more broadly the robustness of our results to the exact

Appendix Table 10: Second-step estimation with alternative first-step estimations 2

Exporter fixed effect	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
Panel A. OLS with first step with no internal distance correction.								
log highway km	1.15***	0.51***	0.35**	0.25*	1.24***	0.22	0.079	-0.050
	(0.14)	(0.17)	(0.14)	(0.13)	(0.16)	(0.19)	(0.15)	(0.15)
Controls.	0	1	2	3	0	1	2	3
R^2	0.58	0.73	0.79	0.84	0.55	0.77	0.81	0.88
Panel B. TSLS with first step with no internal distance correction.								
log highway km	1.11***	0.55***	0.44***	0.36***	1.09***	0.16	0.054	-0.043
	(0.14)	(0.16)	(0.14)	(0.12)	(0.17)	(0.16)	(0.14)	(0.13)
Overid. p-value	0.077	0.033	0.13	0.27	0.077	0.066	0.27	0.54
First-stage Stat.	97.5	90.3	80.4	85.2	97.5	90.3	80.4	85.2
Panel C. OLS with first step excluding internal trade.								
log highway km	1.17***	0.52***	0.35**	0.25*	1.27***	0.22	0.071	-0.058
	(0.14)	(0.19)	(0.15)	(0.14)	(0.16)	(0.20)	(0.16)	(0.15)
R^2	0.57	0.72	0.79	0.83	0.54	0.76	0.80	0.87
Panel D. TSLS with first step excluding internal trade.								
log highway km	1.13***	0.55***	0.43***	0.36***	1.11***	0.15	0.041	-0.057
	(0.15)	(0.17)	(0.14)	(0.13)	(0.18)	(0.16)	(0.15)	(0.13)
Overid. p-value	0.066	0.024	0.088	0.19	0.080	0.071	0.25	0.51
First-stage Stat.	97.5	90.3	80.4	85.2	97.5	90.3	80.4	85.2

Notes: 66 observations per column. The set controls 0 is a constant; 1 adds log 2007 employment and export market access; 2 further adds log population for 1920, 1950 and 2000; 3 also considers log share of manufacturing employment. All TSLS regressions use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

specification for the first-step estimation. The eight columns of panel A reproduce our preferred OLS specification from column 3 of table 4 (main text) for each first-step specification described by the corresponding column in table 3 (main text). The control variables are log 2007 employment, log population for 1920, 1950 and 2000, and the appropriate market access term computed from the exporter fixed effects of the corresponding first-step estimation. Panel B performs a similar exercise for our preferred TSLS specification from column 3 of table 5 (main text). The coefficients on within-city highway kilometers are between 0.30 and 0.38 with OLS and 0.41 and 0.47 with TSLS. They always differ from the corresponding specification in table 4 or 5 (main text) by less than one standard deviation.

Appendix Table 11: Second-step estimation with alternative first-step estimations 3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. OLS with first-step variants from table 3 in the main text.								
log highway km	0.32**	0.32**	0.36**	0.36**	0.38***	0.38**	0.30**	0.31**
	(0.15)	(0.15)	(0.14)	(0.15)	(0.14)	(0.14)	(0.15)	(0.15)
R^2	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.80
Panel B. TSLS with first-step variants from table 3 in the main text.								
log highway km	0.42***	0.42***	0.46***	0.46***	0.47***	0.47***	0.41***	0.38***
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.13)
Overid. p-value	0.32	0.32	0.20	0.20	0.15	0.15	0.33	0.36
First-stage Stat.	81.9	81.9	81.6	81.6	80.6	80.4	82.7	80.9

Notes: 66 observations per column. The dependent variable is an exporter fixed effect for trade in weight estimated from the corresponding column of table 3 in the main text. All regressions use log 2007 highway kilometers, log 2007 employment, log population for 1920, 1950 and 2000, and the appropriate market access term as explanatory variables and include a constant. All TSLS regressions use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

Appendix J. One-step specifications

While our theoretical model maps naturally into a two-step estimation procedure, we can also estimate of the effects of highways on trade in one step with the following regression

$$\log X_{ij} = \lambda_0 + \rho_R^X \log R_i + \lambda^{X'} C_i + \alpha' O_l(\log R_{ij}) + \rho_R^M \log R_j + \lambda^{M'} C_j + v_{ij}\,. \tag{J1}$$

As previously, C_i and C_j are city level controls, $O_l(\log R_{ij})$ is an order l polynomial in pairwise highway distance, and R_i and R_j are within-city highways in the exporting and importing city. In our two-step estimation strategy, we estimate our theoretically derived measure of market access from the first-step gravity equation for use in the second-step estimation. By construction, this approach is not immediately possible in a one-step estimation. Instead, we use the ad-hoc measure of market access defined above.

When estimating equation (J1), we need to account for the fact that the error term v_{ij} has three components: one associated with city i, another associated with city j, and one associated with the pair ij. This suggests two-way clustering and we follow the procedure developed by Cameron, Gelbach, and Miller (2010) to implement this error structure. In addition, we try to keep our instrumentation strategy simple. In our TSLS estimations of equation (J1) we only instrument within-city highways but not highway distances between cities.

Appendix table 12 presents results for one-step estimations. Panel A corresponds to the OLS results of table 4 (main text) and those of table 13 panel A in Appendix C (main text). Panel B of table 12 corresponds to the TSLS results of table 5 and those of table 13 panel B in Appendix C (main text). In all cases the one-step estimates are within one standard error of their two-step counterparts and

Appendix Table 12: Results for one-step estimations

Bilateral trade flows	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
Panel A. OLS estimations.								
log highway km	1.10***	0.43***	0.31**	0.21	1.18***	0.18	0.091	-0.030
for exporter	(0.13)	(0.15)	(0.13)	(0.14)	(0.13)	(0.15)	(0.14)	(0.12)
log highway km	0.83***	0.091	0.054	0.13	0.84***	0.064	0.019	0.052
for importer	(0.23)	(0.15)	(0.14)	(0.13)	(0.19)	(0.13)	(0.12)	(0.11)
log employment		0.64***	-0.66	-0.51		0.69***	-0.36	-0.27
for exporter		(0.10)	(0.35)	(0.32)		(0.10)	(0.48)	(0.37)
log employment		0.51***	0.64***	-0.69		0.58***	0.69***	-0.38
for importer		(0.12)	(0.61)	(0.47)		(0.11)	(0.52)	(0.45)
Market access		-0.47***	-0.82***	-0.75***		-0.12	-0.38*	-0.28*
for exporter		(0.16)	(0.21)	(0.19)		(0.15)	(0.20)	(0.17)
Market access		-1.12***	-1.12***	-1.17***		-0.89***	-0.98***	-1.02***
for importer		(0.19)	(0.19)	(0.22)		(0.17)	(0.17)	(0.20)
log populations 20, 50, 00	N	N	Y	Y	N	N	Y	Y
log % manuf. emp.	N	N	N	Y	N	N	N	Y
R^2	0.70	0.77	0.78	0.80	0.59	0.71	0.72	0.75
Panel B. TSLS estimations.								
log highway km	1.08***	0.51***	0.41***	0.35***	1.04***	0.17	0.095	0.026
for exporter	(0.13)	(0.14)	(0.13)	(0.12)	(0.15)	(0.15)	(0.14)	(0.11)
log highway km	0.81***	0.13	0.092	0.14	0.78***	0.058	0.020	0.034
for importer	(0.21)	(0.14)	(0.14)	(0.15)	(0.17)	(0.13)	(0.13)	(0.13)
log employment		0.56***	0.47	0.35		0.89***	0.96**	0.72*
for exporter		(0.095)	(0.36)	(0.33)		(0.096)	(0.48)	(0.38)
log employment		0.62***	-0.67	-0.50		0.69***	-0.36	-0.26
for importer		(0.13)	(0.61)	(0.46)		(0.12)	(0.52)	(0.45)
Market access		-0.46***	-0.79***	-0.72***		-0.13	-0.38*	-0.26
for exporter		(0.16)	(0.21)	(0.19)		(0.15)	(0.21)	(0.16)
Market access		-1.11***	-1.15***	-1.22***		-0.89***	-0.98***	-1.02***
for importer		(0.19)	(0.19)	(0.21)		(0.16)	(0.17)	(0.20)
log populations 20, 50, 00	N	N	Y	Y	N	N	Y	Y
log % manuf. emp.	N	N	N	Y	N	N	N	Y
Overid. p-value	0.24	0.11	0.23	0.37	.	0.18	0.39	.
First-stage Stat.	69.3	47.6	58.6	55.8	70.3	45.4	44.1	53.1

Notes: 2,476 observation for columns 1 to 4 and 2,705 for columns 5 to 8. All regressions include a constant and quartic function of 2005 highway distance between i and j. Columns 3-4 and 7-8 include log population in 1920, 1950 and 2000 for both exporters and importers. Columns 4 and 8 include log of 2003 manufacturing employment for both exporters and importers. In panel B, we use (same city) log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways for both the importer and exporter cities. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%. In some cases, the overidentification statistics is not produced because the estimated covariance matrix of moment conditions is not of full rank due to an insufficient number of clusters using the procedure of Cameron *et al.* (2010).

often differ from them only in the second decimal place. The coefficient on exporter highways using the one-step version of our preferred specification is 0.41. This is close to its value of 0.47 in the two-step estimation and reassuring regarding the possible correlations between the first- and second-step explanatory variables.

Appendix K. Supplementary results for short and long distance trade

Appendix table 13 mirrors table 7 (main text) but uses our preferred estimate of the importer fixed effect instead of the exporter fixed effect as dependent variable. Panel A shows weak evidence that within-city highways increase the weight of imports when the trade distance is below 1000 km. There is no effect for imported weight for distances above 1000 km or even 750 km. There is no effect on imported values regardless of distance. The significant coefficient for imported weight at short distances is in any case much lower than the corresponding estimate for exports.

Appendix Table 13: Second-step results for short and long distance trade

Importer fixed effect	(1) weight OLS	(2) weight OLS	(3) weight TSLS	(4) weight TSLS	(5) value OLS	(6) value OLS	(7) value TSLS	(8) value TSLS
Panel A. Short distance trade (less than 1000 km).								
log highway km	0.15	0.16*	0.18*	0.19*	0.067	0.086	0.044	0.052
	(0.10)	(0.096)	(0.11)	(0.11)	(0.12)	(0.12)	(0.13)	(0.13)
% manuf. emp.	N	Y	N	Y	N	Y	N	Y
R^2	0.84	0.84			0.78	0.78		
Overid. p-value			0.023	0.037			0.031	0.045
First-stage Stat.			80.8	85.5			80.8	85.5
Panel B. Long distance trade (more than 1000 km)								
log highway km	0.077	0.14	0.19	0.23	-0.028	0.019	0.016	0.044
	(0.21)	(0.20)	(0.21)	(0.21)	(0.18)	(0.16)	(0.18)	(0.17)
R^2	0.66	0.68			0.70	0.71		
Overid. p-value			0.090	0.095			0.12	0.11
First-stage Stat.			83.0	87.9			83.0	87.9
Panel C. Long distance trade (more than 750 km)								
log highway km	0.19	0.25	0.29	0.33	0.025	0.080	0.055	0.09
	(0.22)	(0.21)	(0.21)	(0.22)	(0.19)	(0.17)	(0.19)	(0.18)
R^2	0.65	0.67			0.70	0.71		
Overid. p-value			0.13	0.12			0.19	0.15
First-stage Stat.			86.2	90.6			86.2	90.6

Notes: 66 observations per column. All regressions use log 2007 employment, export market access and log population for 1920, 1950 and 2000 as explanatory variables, and include a constant. All TSLS regressions use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

Appendix L. Additive trade costs

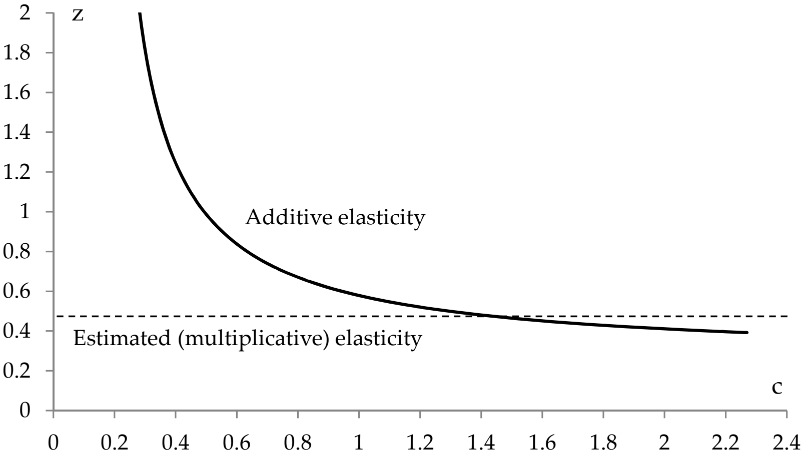
In a simplified partial equilibrium setting we consider a city i facing the following additive transportation costs for exports to city j

$$\tau_{ij} \equiv \left(D_{ij}\right)^x + c\, R_i^{-z}\,, \tag{L1}$$

where D_{ij} is the distance between i and j and R_i measures roads inside city i. x is the elasticity of the distance component of transportation costs with respect to distance. z is the elasticity of the exit cost component of transportation costs with respect to city roads.

In table 7 (main text) we estimate the effect of city roads on exports for distances below 1,000 kilometers and distances above 1,000 kilometers. For distance below 1,000 kilometers the median distance is about 200 kilometers. For distances above 1,000 kilometers the median distance is about

Appendix Figure 3: Additive city roads elasticities



Notes: The plain downwards sloping line represents z the elasticity of the cost of exiting cities as a function of c, the cost of exiting a city relative to the cost of (short) distance. These numbers are computed under the assumption of additive separability and satisfy equations (L2), (L3) and (L4). The dotted horizontal line is the estimated value of the same elasticity under the assumption of multiplicative trade costs.

1,600 kilometers, about 8 times as far. For distance both below and above 1,000 kilometers we estimate an elasticity of trade with respect to distance of 1.8.

After normalizing the cost of distance for short-distance trade to unity by choice of units, from equation (7 – main text) we can write the ratio of short to long distance trade as

$$\left(\frac{1 + cR_i^{-z}}{8^x + cR_i^{-z}}\right)^{1-\sigma} = 8^{1.8} \approx 42\,. \tag{L2}$$

Without loss of generality we also normalize roads in city i to unity so that c can be interpreted as the cost of exiting city relative to the cost of distance for short-distance trade. From table 7, panel A, and column 3 (main text), the coefficient on within-city highways of 0.81 implies that a city that has twice as much highways as city i exports $2^{0.81} \approx 1.75$ times as much as city i. If these two cities are otherwise identical, it should be the case from equation (7 – main text) that

$$\left(\frac{1 + 2^{-z}c}{1 + c}\right)^{1-\sigma} = 2^{0.81} \approx 1.75\,. \tag{L3}$$

For the same two cities now trading over long distance, using the coefficient on within-city highways estimated in table 7, panel B and column 3 we obtain

$$\left(\frac{8^x + 2^{-z}c}{8^x + c}\right)^{1-\sigma} = 2^{0.23} \approx 1.17\,. \tag{L4}$$

The three equations (L2), (L3) and (L4) involve four unknowns, c, x, z, and σ so that the system is under-identified. However simple algebra shows that there is unique solution for $\sigma \approx 4.1$. Then it is easy to obtain x and z as a function of c, the cost of exiting city i relative to the cost of short distance trade.

Appendix figure 3 plots the elasticity of the cost of exiting city i with respect to its roads as a function of c, the cost of exiting city i relative to the cost of short distance. The figure shows clearly that c needs to be above 1.5 for our (multiplicative) coefficient on within-city highways to overestimate an additive elasticity. Put differently, if the cost of exiting a city was additively separable, we would underestimate its elasticity with respect to city roads unless the cost of exiting the city were to represent more than 60% of the overall cost of transportation to another city located 200 kilometers away. Even if the cost of exiting city i were arbitrarily large relative to the cost of distance, z cannot be lower than 0.32.

While this calculation is admittedly crude, its intuition is nonetheless straightforward. When the cost of exiting a city is fairly small and adds to the cost of distance, this cost of exiting a city must be highly sensitive to city roads to match our estimates.

Appendix M. Trade internal to cities

This section demonstrates the degree to which assumptions about the costs associated with internal trade affect the precise estimating equation and interpretation of results for internal trade. Specifically, we consider two possibilities. First, we consider the case in which goods have to 'leave' the city and then re-enter. As such, a variety produced in industry k in city i leaves the city, incurring iceberg cost τ_i^k, travels on roads from i to i incurring cost τ_{ii} and the re-enters city i incurring the importing iceberg cost τ_i. In this case aggregate internal shipments from i to i are

$$X_{ii} = \left(\frac{A_i}{W_i}\right)^{\sigma-1} \left[\int_0^1 \left(\tau_i^k\right)^{1-\sigma} dk\right] \tau_{ii}^{1-\sigma} \tau_i^{1-\sigma} W_i N_i \mathbb{P}_i^{\sigma-1}. \quad \text{(M1)}$$

Using the definitions of importer market access and the labor market clearing condition (9), we have

$$X_{ii} = N_i^{\frac{2(\sigma-1)}{\sigma}} A_i^{\frac{2(\sigma-1)}{\sigma}} \left[\int_0^1 \left(\tau_i^k\right)^{1-\sigma} dk\right] \tau_{ii}^{1-\sigma} e^{\frac{2-\sigma}{\sigma} MA_i^X - MA_i^M}, \quad \text{(M2)}$$

which shows that more roads can increase the level of internal trade by either reducing the costs of exiting the city or by reducing the bilateral costs of shipping within the city τ_{ii}. Taking the ratio of internal shipments and total shipments gives

$$\frac{X_{ii}}{X_i} = \frac{\tau_{ii}^{1-\sigma} \tau_i^{1-\sigma} W_i N_i \mathbb{P}_i^{\sigma-1}}{\sum_j \left(\tau_{ij}^{1-\sigma} \tau_j^{1-\sigma} W_j N_j \mathbb{P}_j^{\sigma-1}\right)}. \quad \text{(M3)}$$

Substituting in both the importer and exporter market access terms MA_i^X and MA_i^M and the labor market clearing condition (9), we obtain

$$\frac{X_{ii}}{X_i} = \tau_{ii}^{1-\sigma} N_i^{\frac{\sigma-1}{\sigma}} A_i^{\frac{\sigma-1}{\sigma}} e^{\frac{1}{\sigma} MA_i^X - MA_i^M}. \quad \text{(M4)}$$

Hence, within-city roads affect internal shipments only to the degree that they affect the internal bilateral shipping costs holding all other variables constant.

In the second case, we assume that a variety in industry k only incurs the exit cost τ_i^k but does not incur the bilateral cost τ_{ii} nor the importing iceberg cost τ_i. In this case the value of internal shipments is

$$X_{ii} = \left(\frac{A_i}{W_i}\right)^{\sigma-1} \left[\int_0^1 \left(\tau_i^k\right)^{1-\sigma}\right] W_i N_i \mathbb{P}_i^{\sigma-1}. \tag{M5}$$

Using the definitions of importer and exporter market access and the labor market clearing condition (9), we can transform this to

$$X_{ii} = N_i^{\frac{2(\sigma-1)}{\sigma}} A_i^{\frac{2(\sigma-1)}{\sigma}} \left[\int_0^1 \left(\tau_i^k\right)^{1-\sigma} dk\right] \tau_i^{\sigma-1} e^{\frac{2-\sigma}{\sigma} MA_i^X} e^{-MA_i^M}. \tag{M6}$$

Based on this expression, everything else equal, roads will increase internal shipments if the fall in aggregator costs of getting goods out of the city τ_i^k falls by more than the costs of getting goods into the city τ_i such that the cost of bringing goods into the city does not fall sufficiently to make internal trade relatively more expensive. Taking this value relative to total shipments gives the following expression

$$\frac{X_{ii}}{X_i} = \frac{W_i N_i \mathbb{P}_i^{\sigma-1}}{\sum_j \left(\tau_{ij}^{1-\sigma} \tau_j^{1-\sigma} W_j N_j \mathbb{P}_j^{\sigma-1}\right)}. \tag{M7}$$

Using our definitions of importer and exporter market access as well as the labor market clearing condition (9), we obtain

$$\frac{X_{ii}}{X_i} = N_i^{\frac{\sigma-1}{\sigma}} A_i^{\frac{\sigma-1}{\sigma}} \tau_i^{\sigma-1} e^{MA_i^M}. \tag{M8}$$

Here we get the unexpected result that more roads reduce internal shipments. In this case, it is because more roads lower the entry cost of goods from other cities and, therefore, increases the relative price of internal shipments.

Examining expressions (M2) and (M6), the theoretical sign of the effect of roads on the level of internal trade is positive in the first but ambiguous in the second. Comparing equations (M4) and (M8), the sign of the relationship of roads with the share of shipments that are internal changes. As such, we interpret any empirical relationship between roads and various measures of internal trade with caution.

Appendix N. Further specialization results

Appendix table 14 checks the robustness of the results of panel A of table 10 (main text) by estimating the effect of the interaction between unit weight and highways directly. Each regression regresses log sectoral employment in 2007 on log highway kilometers, log highway kilometers interacted with log unit value, log of 1920, 1950 and 2000 population, log 2007 employment, log market access, log 1956 employment in the same industry, and sector fixed effects. For the TSLS regressions of columns 2 to 8, we use the same three instruments as before (log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index) and their interaction with log unit value

Column 1 is estimated with OLS whereas columns 2-7 are estimated with TSLS using log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index and their

Appendix Table 14: Specialization results, one-step estimations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: industry specific coefficient on interstate highways								
estimated with:	OLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
using additional controls:	-	-	Water & Slope	Census Div.	% College	Mining	Income pc.	Wholesale
Dependent variable: $\Delta_{2007-1956}$ log employment								
log highway km ×	0.11***	0.15***	0.16***	0.15***	0.17***	0.099**	0.16***	0.16***
log tons per $	(0.041)	(0.045)	(0.045)	(0.044)	(0.044)	(0.047)	(0.044)	(0.044)
log highway km	0.28***	0.22***	0.12	0.24***	0.090	0.28***	0.17**	0.27***
	(0.061)	(0.069)	(0.084)	(0.073)	(0.073)	(0.079)	(0.070)	(0.069)
Overid. p-value	-	0.018	0.0035	0.064	0.48	0.34	0.64	0.076
First-stage Stat.	-	614	630	845	685	978	442	356

Notes: 1,262 observations per column (except column 6, 1,117). Standard errors in parentheses. The dependent variable is the industry employment in 2007. All regressions also include the log of 1920, 1950 and 2000 population, log 2007 employment, log market access, log 1956 employment in the same industry, and sector fixed effects. Column 1 is estimated with OLS whereas columns 2-7 are estimated with TSLS using log 1947 planned highway km, log 1898 railroad km, and log 1528-1850 exploration routes index and their respective products with log tons per $ as instruments for log kilometers of interstate highways and log kilometers of interstate highways multiplied by log tons per $. The regressions also include log distance to water and log slope (column 3), census region dummies (column 4), log share college graduates in 1980, 1990 and 2000 (column 5), log share mining employment (column 6), log share manufacturing employment in 1978 and 2003 (column 7), and log share employment in wholesale in 1956 and 2007 (column 8). Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

respective products with log tons per dollar, as instruments for log kilometers of interstate highways and log kilometers of interstate highways multiplied by log tons per dollar. The results are very close to those of table 10 (main text). We note that overidentifying restrictions are sometimes rejected when we use all six instruments.

Next, we exploit time series variation in the CBP to observe the dynamics of our results. To do this, appendix table 15 replicates panel D of table 10 (main text) controlling for employment increasingly recent in time. Panel A, adds 1970 CBP sectoral employment as an additional control variable in our estimation of equation (23 – main text) in the first step. Panel B includes 1956, 1970 and 1977 employment as controls. Analogously, panels C and D add 1987 employment and then both 1987 and 1997 employment, respectively, as controls. Hence panel D includes sectoral employment for 1956, 1970, 1977, 1987 and 1997 as additional control variables in the estimation of equation (22 – main text). Our instrument set still comprises 1947 planned highways, 1898 railroads and old exploration routes of the continent.

Although the pattern is pervasive down each column, we analyze column 2. $\hat{\rho}_R^{N,k}$ is monotonically declining as we include more recent employment controls. For example, when we control for 1970 employment in addition to 1956 employment in city i in industry k in our estimation of equation (23 – main text), $\hat{\rho}_R^{N,k}$ falls by 13% from 0.17 to 0.15. In panel D, the relationship between $\hat{\rho}_R^{N,k}$ and log unit weight is insignificant but remains significant in the OLS estimation of column 1.

Appendix Table 15: Specialization results, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: industry specific coefficient on interstate highways								
estimated with:	OLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
using additional controls:	-	-	Water & Slope	Census Div.	% College	Mining	Income pc.	Wholesale
Panel A: Control for 1956 and 1970 sectoral employment								
log weight per unit value	0.12*** (0.037)	0.15*** (0.038)	0.14*** (0.040)	0.12*** (0.042)	0.11*** (0.033)	0.16*** (0.045)	0.14*** (0.037)	0.15*** (0.038)
R^2	0.36	0.42	0.38	0.29	0.36	0.40	0.40	0.43
Panel B: Control for 1956, 1970, 1977 sectoral employment								
log weight per unit value	0.13*** (0.036)	0.13*** (0.039)	0.13** (0.048)	0.10** (0.044)	0.10*** (0.032)	0.16*** (0.037)	0.12*** (0.037)	0.13*** (0.038)
R^2	0.40	0.37	0.29	0.22	0.35	0.49	0.35	0.38
Panel C: Control for 1956, 1970, 1977, 1987 sectoral employment								
log weight per unit value	0.12*** (0.032)	0.11*** (0.035)	0.11** (0.043)	0.087** (0.039)	0.092** (0.033)	0.13*** (0.034)	0.10*** (0.034)	0.11*** (0.035)
R^2	0.42	0.33	0.24	0.20	0.28	0.41	0.32	0.32
Panel D: Control for 1956, 1970,1977,1987, 1997 sectoral employment								
log weight per unit value	0.064** (0.025)	0.044 (0.027)	0.027 (0.037)	0.028 (0.040)	0.028 (0.030)	0.057** (0.027)	0.038 (0.030)	0.042 (0.027)
R^2	0.25	0.12	0.03	0.02	0.04	0.18	0.07	0.11

Notes: 22 observations per column for each panel. Standard errors in parentheses. The dependent variable is the industry specific coefficient on interstate highways estimated from regression (23 – main text) using log kilometers of interstate highways, log 2007 employment, log 1920, 1950 and 2000 population, and log market access. Column 1 is estimated with OLS whereas columns 2-7 are estimated with TSLS using log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. The regression estimating the highways elasticity of employment also includes log distance to water and log slope (column 3), census region dummies (column 4), log share college graduates (column 5), log share mining employment (column 6), log share manufacturing employment (column 7) and log share employment in wholesale (column 8). Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

This suggests that highways only gradually affect the specialisation patterns of cities.

References

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2010. Robust inference with multi-way clustering. *Journal of Business and Economic Statistics* 29(2):238–249.

Disdier, Anne-Célia and Keith Head. 2008. The puzzling persistence of the distance effect on bilateral trade. *Review of Economics and Statistics* 90(1):37–48.

Redding, Stephen and Anthony J. Venables. 2004. Economic geography and international inequality. *Journal of International Economics* 62(1):63–82.