Contents

Appendix A

Border Costs

A.1 Cost of Trading into a State

In Figures A.1 and A.2, we show for each state the costs for trading into and out of a state, respectively, relative to the average cost across all states (with the average cost fixed to be 0). For purposes of the maps, the states are grouped into quartiles to more clearly visualize distinctions between them and we exclude Hawaii and Alaska from the analysis. First, we show how we calculated the cost by state of crossing the state borders and trading into of each state. We omit the calculations for trading out of each state (they are similar as for trading into each state just in the opposite direction).

We aggregated all the data into CFS areas as in the main results and for 2012, ran the regression in (A.1). For this regression, we used Poisson Pseudo Maximum Likelihood to remain consistent with the main results.

$$\ln T_{i,j} = \beta_0 + \beta_1 \ln (GDP_j) + \beta_2 \ln (GDP_i) + \beta_3 \ln (d_{i,j}) + \beta_4 \delta_{i,j} \tag{A.1}$$

where $T_{i,j}$ is the trade flows from CFS Area i to area j, GDP_j is the GDP of Area j (and correspondingly for GDP_i), $\delta_{i,j}$ is an indicator, which is 1 if i and j are CFS Areas in the same state and 0 otherwise, and $d_{i,j}$ is the average distance between Areas i and j. Here, we take the average miles a shipment traveled for all shipments from i to j. This is not weighted by value or ton. If the average mile was not given in the data, we took the Haversine's distance instead.

After running the above regression, we ran the regression given in Equation (A.2) for each state j. This is equivalent to having one regression with $(1 - \delta_{i,j})E_j$ for each state j included as a covariate.

$$R_{i,j} = \beta(1 - \delta_{i,j})E_j \tag{A.2}$$

where the explanatory variable is an indicator, which is 1 if the areas are in different states and trade is going into the state j (thus, 1 if trade flows are from outside the state to the desired destination state). Here, E_j is an indicator for whether the destination state is the same as j. Thus, we get that

$$\beta_k = E_i[\ln (T_{i,j}) - E_j[\ln (T_{i,j})]|i \neq j \;\; \& \;\; j = k]$$

Figure A.1 gives a map for the coefficient value of the second regression for each state. The goal here is to see how much of the residuals in the gravity equation can be explained by the state we are entering. A negative β for the state implies that trade is lower than we would expect relative to the rest of the states (harder to trade into this state) whereas a positive β implies it is easier to trade into this state than other states. Note that these values are normalized so 0 is the average so the true cost of trading into each of these states is actually greater. We are using this formulation since we are mainly interested in regional differences in state borders.

We converted the coefficients into miles using the method outlined in the section, "Conversion to Miles." The numbers in the legends represent the maximum cost (in miles) that a state of that color has.

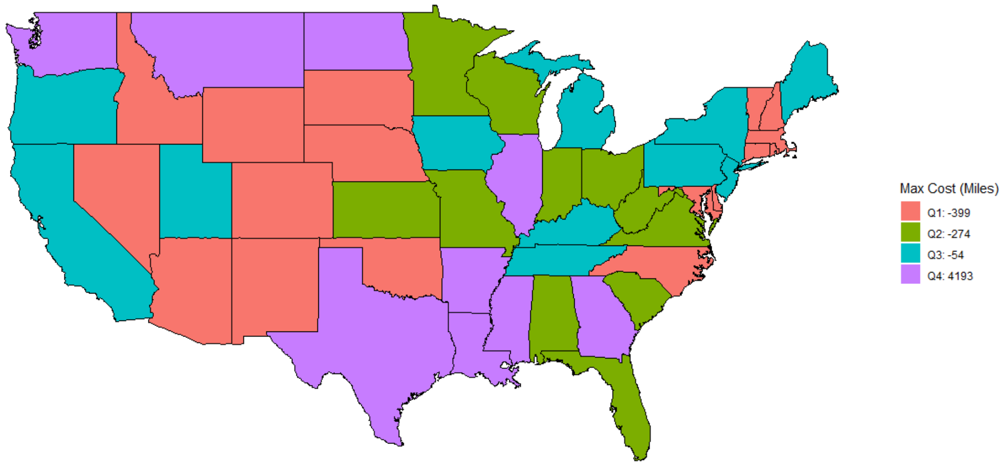


Figure A.1: Border cost of trading into each state, with 0 as the average.

While Figure A.1 appears to have a very large maximum cost in the top quartile, this is misleading since it gives the maximum cost, which is Montana's, which has very little trade. 47 out of the 49 states, in fact, have border costs between -740 and 301 miles for cost of trading into a state. The other state is Texas with a cost of 1,195 miles. Thus, we observe some heterogeneity in border costs across states, but most of the border costs are consistent across all states, if we use as reference **?**'s 3,589 miles as the average cost. There don't appear to be any clear regional patterns in border cost of trading into states.

Figure A.2 yields a map representing the border cost of trading out of each state. The minimum cost in this map (since it can't be seen) is -913 miles.

Regionally, it appears that the Midwest has lower outflows than would be expected given its GDP and trading partners.

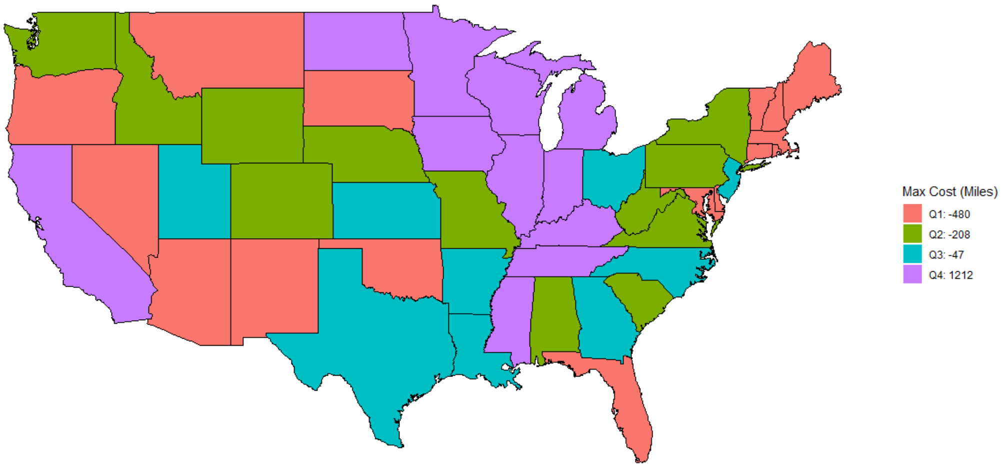


Figure A.2: Border cost of trading out of each state, with 0 as the average.

A.1.1 Conversion to Miles

To convert to miles, we took the β returned by Equation (A.2) and the β_3 returned by Equation (A.1) and let $x = e^{\frac{\beta_3}{\beta}}$. For each state, we then calculated the exponential of the mean log distance shipments to that state traveled (call this d_j) and returned $d_j(x-1)$. This then returned the expected border width to get the same difference in trade flows, relative to other states (hence why some values are negative). The proof for why this calculation gives the desired border cost in miles is as follows.

In the standard gravity equation formulation, we have:

$$T_{i,j} = G\frac{Y_i^{\alpha} Y_j^{\beta}}{(d_{i,j}x)^{\gamma}}$$

We want to solve for the value of x that would equate to the distance created by a shipment coming from out of state into the particular state in which area j lies.

In logs, our formulation (including the "Same State" indicator and indicator for whether the shipment is out of state into the particular state) is:

$$\ln(T_{i,j})) = \ln G + \alpha \ln Y_i + \beta \ln Y_j - \gamma \ln(d_{i,j}) + \psi I(\text{state}_i = \text{state}_j) - \chi_k I(\text{state}_i \neq \text{state}_j \;\; \& \;\; \text{state}_j = \text{stat}$$

Note we are performing the above specification in the code by taking the residuals of the standard gravity equation with a "Same State" indicator variable and then regressing the residuals on the last indicator, as in (A.2).

Exponentiating tells us that this regression returns:

$$T_{i,j} = G\frac{Y_i^{\alpha}Y_j^{\beta}}{d_{i,j}^{\gamma}}\frac{\exp\left(\psi I(\text{state}_i = \text{state}_j)\right)}{\exp\left(\chi_k I(i \neq j \ \ \& \ \ j = k)\right)}$$

Here, note that I have assumed the signs on χ_k and $d_{i,j}$ are negative and built that into the model.

We are only considering instances where $\text{state}_i \neq \text{state}_j$ and so $\exp\left(\psi I(\text{state}_i = \text{state}_j)\right) = 1$. Then, setting the above equation equal to our original set-up gives:

$$\begin{aligned}(d_{i,j}x)^{\gamma} &= d_{i,j}^{\gamma}\exp\left(\chi_k I(i \neq j \ \ \& \ \ j = k)\right)\\ x^{\gamma} &= e^{\chi_k I(i \neq j \ \& \ j=k)}\\ x &= \left(e^{\chi_k}\right)^{1/\gamma}\end{aligned}$$

We then return $(x-1)E_i[d_{i,k}]$. Note that this method is roughly in line with **?**, whose equation in our formulation would be as given in Equation (A.3).

$$((e^{-\psi})^{1/\gamma} - 1)E_{i,k}[d_{i,k}] \tag{A.3}$$

The differences are that we use a different mean distance for each state rather than an overall mean distance. Furthermore, we want to evaluate the cost of trading into or out of a state, so instead of looking at the coefficient on "Same State", we consider the coefficient on trade entering a state from outside the state (or for cost of trading out of a state, the coefficient on trade leaving a state). Thus, our border cost estimates a different (lower) border cost measure than those given by **?** and also lower than those of **?**.

A.2 Mode of Transportation

We perform a regression with the mode of transportation included to highlight that accounting for the lower cost of trading within a state doesn't justify the state-border effect. We run the regression using ordinary least squares, excluding CFS Areas with zero flows between them, and add in covariates for the mode of transportation. We interact distance with these modes since logically, we would want how the impact of a rail mile on trade flows differs from that of a truck mile. The equation for this part is given in Equation (A.4).

$$T_{ij} = \beta_0 + \beta_1 Y_j + \beta_2 Y_i + \beta_3 d_{ij} + \beta_4 \delta_{ij} + \gamma\chi_{mode} + \Psi\alpha_{ij} \tag{A.4}$$

Here, T_{ij} is log trade flows from Area i to Area j, Y_k represents the log GDP of area k for some k, δ_{ij} is an indicator for whether or not i and j are in the same state, χ_{mode} is a series of indicators for each of the modes in consideration (air, water, truck, rail, truck and water), d_{ij} is the log distance in miles between CFS areas i and j (as before, I take the average miles shipped between the two using the Census Bureau's GeoMiler and then used the Haversine distance between their centroids if that value was unavailable), and α_{ij} is a series of covariates, multiplying the distance by the indicator for each mode.

Table A.1: Gravity Regression for Trade with Mode of Transportation

	Dependent variable:
	Log Trade Flows
Log Destination GDP	1.020***
	(0.025)
Log Origin GDP	0.902***
	(0.026)
Same State	3.722***
	(0.126)
Truck and Water	0.242
	(0.294)
Truck	4.941***
	(0.227)
Rail	0.043
	(0.803)
Water	0.595
	(5.727)
Log Distance	0.020
	(0.015)
Truck and Water, Log Distance	0.070
	(0.084)
Truck, Log Distance	−0.385***
	(0.034)
Rail, Log Distance	0.328**
	(0.133)
Water, Log Distance	0.134
	(0.750)
Constant	−7.877***
	(0.180)
Observations	12,670
R^2	0.271
Adjusted R^2	0.270
Residual Std. Error	2.415 (df = 12657)
F Statistic	391.478*** (df = 12; 12657)
Note:	*p<0.1; **p<0.05; ***p<0.01

In Equation (A.4), air is the default mode. Note the results are unsurprising in that initially, if we can use truck, trade is going to be very high (since truck is cheapest over short distances) and rail and water are also preferred to air if we can use that. However, adding distances significantly negatively impacts trade flows if we are using trucks while having less of a negative impact on rail and water, as we would expect.

Most relevantly, the coefficient on being in the same state has increased dramatically, suggesting that once we can control for mode of transportation, being in the same state has an even more positive effect than it did before. A possible explanation of this would be that there is a lot of trade between parts of the state that are very far apart from each other (think Pittsburgh and Philadelphia), but much of this trade occurs via rail rather than truck and if it were to occur by truck, there would be even more of it. However, this explanation is false since almost all trade between CFS areas in the same state is listed as being truck-based.

Thus, the increase in the coefficient estimate for "Same State" is likely anomalous, but it does suggest that the state-border effect is not driven primarily by the mode of transportation used.

A.3 Natural Borders

While it has never seriously been considered in the literature, we also evaluate whether the border effect for both migration and trade is a byproduct of geography. Specifically, state borders aren't random; they often lie on geographic features such as rivers. It is possible that those features hinder trade and migration as crossing the river is costly. This explanation has received scant attention in the literature since it doesn't seem plausible given modern technology making movement across rivers easy.

Still, to test the hypothesis, we manually collected a variable, "Natural Border," that is an indicator variable, one for if the origin and destination states share a border that has a river flowing through most of it and zero otherwise. If the two states are the same, then we code it as 0. We add this variable into the regressions from Equation (**??**) for migration, estimating with Poisson Pseudo Maximum Likelihood and using robust variance.

Note that we don't consider any other geographic features except rivers (although this is not a significant assumption since the vast majority of borders that are determined by geography are based on rivers). More crucially, if states do not border each other, then the variable is automatically coded as zero even if rivers need to be crossed between the two states.

We first note the results for migration, which can be found in Table A.2. Remember that we are still running the regression on county-county flows even though the natural border is determined at the state level. So, two counties are considered to have a natural border between them if the states the counties are in share a border that is mostly a river.

As the results of Table A.2 show, there is little effect of including an indicator for "Natural Border" on the state-border effect of migration. The coefficient estimate remains about the same as before for the counties being in the same state. Don't be fooled by the statistically significant estimate on "Natural Border." We would

Table A.2: Migration Regression with Natural Border - PPML

	Coefficient
Intercept	-14.14^{***}
	(0.02)
Log Distance	-0.88^{***}
	(0.00)
Log Destination Population	0.85^{***}
	(0.00)
Log Origin Population	0.85^{***}
	(0.00)
Same State	1.86^{***}
	(0.01)
Neighbor	1.97^{***}
	(0.01)
Natural Border	0.27^{***}
	(0.01)
Same State, Neighbor	-0.82^{***}
	(0.01)
Neighbor State	-0.03^{*}
	(0.01)
Deviance	$31,208,297.10$
Num. obs.	$9,382,536$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A.3: Trade Regression with Natural Border - PPML

	Coefficient
Intercept	4.64^{***}
	(0.31)
Log Distance	-0.52^{***}
	(0.04)
Log Destination GDP	0.59^{***}
	(0.03)
Log Origin GDP	0.59^{***}
	(0.03)
Natural Border	0.24^{*}
	(0.10)
Same MA	0.33^{**}
	(0.11)
Same ST	0.71^{***}
	(0.12)
Neighbor State	0.05
	(0.13)
Deviance	$1,475,775.14$
Num. obs.	836

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

have expected the estimate to go in the negative direction if the natural border was hindering trade.

The coefficient estimate on “Natural Border” are also unlikely to just be picking up the benefits of states being neighbors instead, since we added in an indicator “Neighbor State”, which is 1 is the flows are between two counties in neighboring states to adjust for this. Regardless, we should have seen some drop in the coefficient estimate for “Same State” if natural borders were the primary cause behind the state-border effect.

Next, we run the gravity regression for trade from Equation (**??**) using Poisson Pseudo Maximum Likelihood Estimation and adding in the indicator variable for state border. The result is in Table A.3.

Again, we see that including an indicator for sharing a natural border doesn’t really impact the coefficient estimate on being in the same state, suggesting that the state-border effect on trade is not being driven by natural borders forming many state borders. Furthermore, there is still a sizable difference in coefficient estimates between “Same MA” and “Same ST”. Again, it is unlikely that this regression is just picking up the effect of states being neighbors in “Natural Border” since we added an indicator variable, “Neighbor State” to adjust for this.

A.4 Ordinary Least Squares

A.4.1 Excluding Zeros

Instead of Poisson Pseudo Maximum Likelihood (PPML) estimation, we employ two other strategies to run the gravity equations in the main results. We first use a dataset where all CFS Area-CFS Area pairs (and for migration, all county pairs) that have zero flows are not even included as observations in the data. This lets us take the log of trade (or migration) flows in the Equation (**??**). We first run the Trade Gravity Regression Results, the results of which are given in Table A.4. Compared to Table **??**, I add in another indicator for whether the CFS Areas are the same. However, inclusion or exclusion of this leaves the qualitative results the same.

Table A.4: Trade Gravity Regression Results - Excluding Zeros

	Dependent variable:		
	Log Trade Outflows		
	(1)	(2)	(3)
Log Distance	-1.132^{***}	-1.269^{***}	-1.135^{***}
	(0.016)	(0.040)	(0.016)
Log Destination GDP	0.870^{***}	0.803^{***}	0.880^{***}
	(0.014)	(0.044)	(0.014)
Log Origin GDP	0.962^{***}	0.848^{***}	0.973^{***}
	(0.013)	(0.044)	(0.014)
Same State	0.844^{***}	0.361^{**}	
	(0.060)	(0.156)	
Same MA			0.247^{***}
			(0.031)
Same ST			0.778^{***}
			(0.062)
Same CFS			1.073^{***}
			(0.117)
Constant	4.529^{***}	6.460^{***}	4.410^{***}
	(0.118)	(0.313)	(0.122)
Observations	9,696	1,681	9,696
R^2	0.640	0.750	0.642
Adjusted R^2	0.640	0.750	0.642
Residual Std. Error	1.062 (df = 9691)	1.081 (df = 1676)	1.060 (df = 9689)
F Statistic	$4{,}311.635^{***}$ (df = 4; 9691)	$1{,}258.998^{***}$ (df = 4; 1676)	$2{,}896.518^{***}$ (df = 6; 9689)
Note:			*p<0.1; **p<0.05; ***p<0.01

A linear hypothesis test comparing the coefficient estimates for "Same MA" and "Same ST" yielded an F score of 51.86 and given the difference in the two estimates, this result lets us conclude that the coefficient estimate on "Same ST" is larger with at least 99.99% confidence.

We next test whether area-specific attributes are truly constant across metropolitan areas as we assume and are the primary cause behind the state-border effect.

Table A.5: GAM for Gravity Eq. with Demographics - Excluding Zeros

A. Parametric Coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	4.1054	0.1641	25.0212	< 0.0001
Log Distance	-1.1208	0.0184	-60.9578	< 0.0001
Log Destination GDP	0.8235	0.0208	39.6429	< 0.0001
Log Origin GDP	1.0663	0.0189	56.2711	< 0.0001
Same MA	0.0537	0.0412	1.3049	0.1920
Same ST	0.7808	0.0844	9.2524	< 0.0001
Same CFS	0.9110	0.1322	6.8894	< 0.0001
B. Smooth Terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Rent)	4.6971	5.8392	41.6278	< 0.0001
s(Income)	2.1663	2.5602	0.1405	0.9415
s(Unemp)	3.0475	3.9563	9.2555	< 0.0001
s(Gini)	5.1515	6.3747	29.6039	< 0.0001
s(Bachelor)	3.1418	4.0149	24.7310	< 0.0001
s(Age)	4.2034	5.3009	2.5587	0.0215
s(Democrat)	2.1663	2.5602	0.1405	0.9415
s(Marry)	2.1663	2.5602	0.1405	0.9415
s(AQI_90)	5.7663	6.9833	2.9039	0.0050
s(Tot_Crime)	8.5169	8.9332	5.4998	< 0.0001

We add in the same area-specific attributes as in Table **??** but also add in two other attributes: total crime rate per 100,000 persons and the 90th percentile Air Quality Index for the CFS Area (where high is worse), aggregating over counties by taking the mean. Before inputting all the attributes into the regression, we take the ratio between the destination and origin values and take logs of the ratios. For each explanatory variable, we allow the degrees of freedom to take on any value and choose the optimal value via cross-validation.

As we can see in Table A.5, controlling for demographic factors nearly eliminates the positive effect on trade of being in the same metro area but in a different state. The coefficient estimate for "Same ST", on the other hand, remains roughly similar. This provides further evidence that there is in fact a state border effect on trade independent of any demographic effects.

Next, we add an indicator for whether the trade is between two different CFS Areas, one of which is the nearest urban area to the other (note then that the other must be a rural area since otherwise, it would itself be its own nearest urban area). Second, we run a regression with an indicator for whether the trade is within the same state between rural and urban CFS Areas. The results in Table A.6.

These results are consistent with those of Table **??**, showing that there is still a significant state-border effect even when these factors are accounted for. However, it does appear that rural-urban trade is especially high, suggesting that the hypothesis does have some merit, even if it is insufficient to explain the state-border effect.

Next, we run the gravity regression for migration using the dataset excluding zeros in A.7. The analog to this table in the main paper with PPML estimation is Table **??**.

Again, we see a strong state-border effect for migration in all three regressions. In fact, we also see that for border counties, migration is much likelier to occur for

Table A.6: Regressions controlling for trade going through a nearby city - Excluding Zeros

	Dependent variable:	
	Log Trade Flows	
	(1)	(2)
Log Destination GDP	0.882***	0.870***
	(0.013)	(0.014)
Log Origin GDP	0.949***	0.962***
	(0.013)	(0.013)
Log Distance	−1.110***	−1.136***
	(0.015)	(0.016)
Same State	0.852***	0.707***
	(0.060)	(0.074)
Nearest City, Different CFS	0.085	
	(0.129)	
Same State, Rural-Urban		0.360***
		(0.091)
Constant	4.410***	4.547***
	(0.113)	(0.118)
Observations	9,356	9,696
R^2	0.654	0.641
Adjusted R^2	0.653	0.640
Residual Std. Error	1.018 (df = 9350)	1.062 (df = 9690)
F Statistic	3,527.604*** (df = 5; 9350)	3,455.235*** (df = 5; 9690)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table A.7: Gravity Regression for Migration - Excluding Zeros

	Dependent variable:		
	Log Migrant Flows		
	(1)	(2)	(3)
Log Destination Population	0.330***	0.389***	0.355***
	(0.003)	(0.003)	(0.005)
Log Origin Population	0.330***	0.385***	0.344***
	(0.003)	(0.003)	(0.005)
Log Distance	−0.505***	−0.394***	−0.353***
	(0.004)	(0.004)	(0.006)
Same State	0.493***	0.550***	0.521***
	(0.009)	(0.008)	(0.015)
Neighbor		1.292***	1.290***
		(0.028)	(0.032)
Same State, Neighbor		0.019	0.090***
		(0.028)	(0.034)
Constant	−1.302***	−3.518***	−2.822***
	(0.047)	(0.051)	(0.090)
Observations	89,532	89,532	27,062
R^2	0.360	0.486	0.452
Adjusted R^2	0.360	0.486	0.452
Residual Std. Error	0.797 (df = 89527)	0.714 (df = 89525)	0.733 (df = 27055)
F Statistic	12,588.950*** (df = 4; 89527)	14,124.930*** (df = 6; 89525)	3,715.902*** (df = 6; 27055)
Note:			*p<0.1; **p<0.05; ***p<0.01

neighboring counties within the same state than for neighboring counties in another state, although migration across neighboring counties is even more significant than within the same state. To test whether the migration effect is being driven by area-specific attributes like demographics, we run a GAM, allowing the degrees of freedom for each explanatory variable to take on any value and choosing this value via cross-validation. The results are in Table A.8. The analogous table in the main results is **??**.

Just as in the main results, Table A.8 shows that even accounting for demographics and other area attributes doesn't change the main result of there being a significant state border effect.

A.4.2 Including Zeros

We next include all possible CFS-CFS Area and county-county pairs, even those with zero flows between them. We code each Area pair with zero trade flows as having a flow of $1. This should hardly make a difference in the results as trade flows tend to be in the billions and furthermore, almost every CFS Area in America has trades with almost every other CFS Area. For county-county pairs, this potentially has a greater impact since migration flows tend to be smaller and most county-county pairs have no migration between them. As we will note, however, the qualitative results don't change.

As we can see in Table A.9, the coefficient estimates are not very different even when we include CFS Area-CFS Area pairs without any trade flows between them as having a flow of 1. The results match those in **??**. When including zero-flow

Table A.8: Migration Gravity Regression with Demographics - Excluding Zeros

A. Parametric Coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	-8.3256	0.1195	-69.6667	< 0.0001
Log Destination Population	0.4961	0.0848	5.8488	< 0.0001
Log Origin Population	0.6416	0.0847	7.5742	< 0.0001
Log Distance	-0.4030	0.0059	-67.9584	< 0.0001
Same State	0.8460	0.0164	51.5520	< 0.0001
Neighbor	1.4942	0.0765	19.5414	< 0.0001
Same State, Neighbor	0.2433	0.0823	2.9554	0.0031
B. Smooth Terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Rent)	5.3518	6.5569	38.4062	< 0.0001
s(Income)	8.1912	8.8253	14.7922	< 0.0001
s(Unemp)	4.8711	6.0448	1.3517	0.2186
s(Gini)	3.4517	4.4415	6.6247	< 0.0001
s(Bachelor)	6.5898	7.7656	14.6952	< 0.0001
s(Age)	7.9965	8.7334	13.7911	< 0.0001
s(Democrat)	3.0291	3.9160	4.4428	0.0015
s(Marry)	7.5361	8.4842	7.2090	< 0.0001
s(AQI_90)	8.0349	8.7450	4.2544	< 0.0001
s(Tot_Crime)	8.9041	8.9967	53.7183	< 0.0001

pairs, there is also little difference in the coefficient estimates derived from running regressions controlling for trade going through a nearby city, as seen in Table A.10. These results match those of **??** in the main results.

Lastly, we display the table for county-county migration flows. Note that this time, there are significant changes in the coefficient estimates. The reason for this is that we have added many more flows close to 0, which is going to tend to reduce the estimates produced by the model. The qualitative analysis of the results, however, remains the same. Table A.11 shows the regression results corresponding to the gravity equation regressions for migration. The first and second columns contain all county-county pairs, whereas the third column only contains cases where the origin county is a border county.

These results are broadly consistent with our results in Table **??**, suggesting that using OLS or PPML doesn't impact the presence of a large state-border effect.

Table A.9: Trade Gravity Regression Results - Including Zeros

	Dependent variable:		
	Log Trade Flows		
	(1)	(2)	(3)
Log Destination GDP	0.846***	0.749***	0.862***
	(0.013)	(0.039)	(0.013)
Log Origin GDP	0.968***	0.928***	0.984***
	(0.012)	(0.036)	(0.012)
Same MA			0.301***
			(0.026)
Same ST			0.779***
			(0.061)
Same CFS			1.066***
			(0.114)
Log Distance	−1.131***	−1.271***	−1.139***
	(0.014)	(0.034)	(0.015)
Same State	0.837***	0.312**	
	(0.058)	(0.147)	
Constant	4.609***	6.399***	4.460***
	(0.114)	(0.301)	(0.117)
Observations	11,991	2,457	11,991
R^2	0.648	0.725	0.651
Adjusted R^2	0.647	0.725	0.651
Residual Std. Error	1.100 (df = 11986)	1.113 (df = 2452)	1.095 (df = 11984)
F Statistic	5,505.445*** (df = 4; 11986)	1,616.813*** (df = 4; 2452)	3,725.634*** (df = 6; 11984)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A.10: Regressions controlling for trade going through a nearby city - Including Zeros

	Dependent variable:	
	Log Trade Flows	
	(1)	(2)
Log Destination GDP	0.859***	0.846***
	(0.013)	(0.013)
Log Origin GDP	0.966***	0.969***
	(0.012)	(0.012)
Log Distance	−1.111***	−1.134***
	(0.014)	(0.014)
Same State	0.837***	0.702***
	(0.059)	(0.073)
Nearest City, Different CFS	0.097	
	(0.124)	
Same State, Rural-Urban		0.359***
		(0.090)
Constant	4.445***	4.627***
	(0.109)	(0.114)
Observations	11,453	11,991
R^2	0.660	0.648
Adjusted R^2	0.660	0.648
Residual Std. Error	1.059 (df = 11447)	1.100 (df = 11985)
F Statistic	4,447.455*** (df = 5; 11447)	4,409.947*** (df = 5; 11985)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table A.11: Gravity Regression for Migration - Including Zeros

	Dependent variable:		
	Log Trade Flows		
	(1)	(2)	(3)
Log Destination Population	0.045***	0.044***	0.041***
	(0.0002)	(0.0002)	(0.0003)
Log Origin Population	0.045***	0.044***	0.039***
	(0.0002)	(0.0002)	(0.0003)
Log Distance	−0.062***	−0.038***	−0.046***
	(0.0004)	(0.0003)	(0.001)
Same State	0.475***	0.339***	0.273***
	(0.003)	(0.002)	(0.004)
Neighbor		2.902***	2.879***
		(0.062)	(0.062)
Same State, Neighbor		0.982***	1.149***
		(0.065)	(0.073)
Constant	−0.485***	−0.627***	−0.496***
	(0.004)	(0.003)	(0.005)
Observations	9,382,536	9,382,536	3,199,749
R^2	0.102	0.213	0.208
Adjusted R^2	0.102	0.213	0.208
Residual Std. Error	0.406 (df = 9382531)	0.380 (df = 9382529)	0.367 (df = 3199742)
F Statistic	265,374.300*** (df = 4; 9382531)	423,385.800*** (df = 6; 9382529)	140,141.600*** (df = 6; 3199742)
Note:			*p<0.1; **p<0.05; ***p<0.01

Appendix B

Migration Costs

B.1 Employment Share Changes by Industry

We perform the same analysis as in the main results to study changes in employment shares but grouping by industries instead of occupations. Table B.1 lists the industry groups and their ranks in each of the years whose results we present.

Table B.1: Industry Ranks by Income

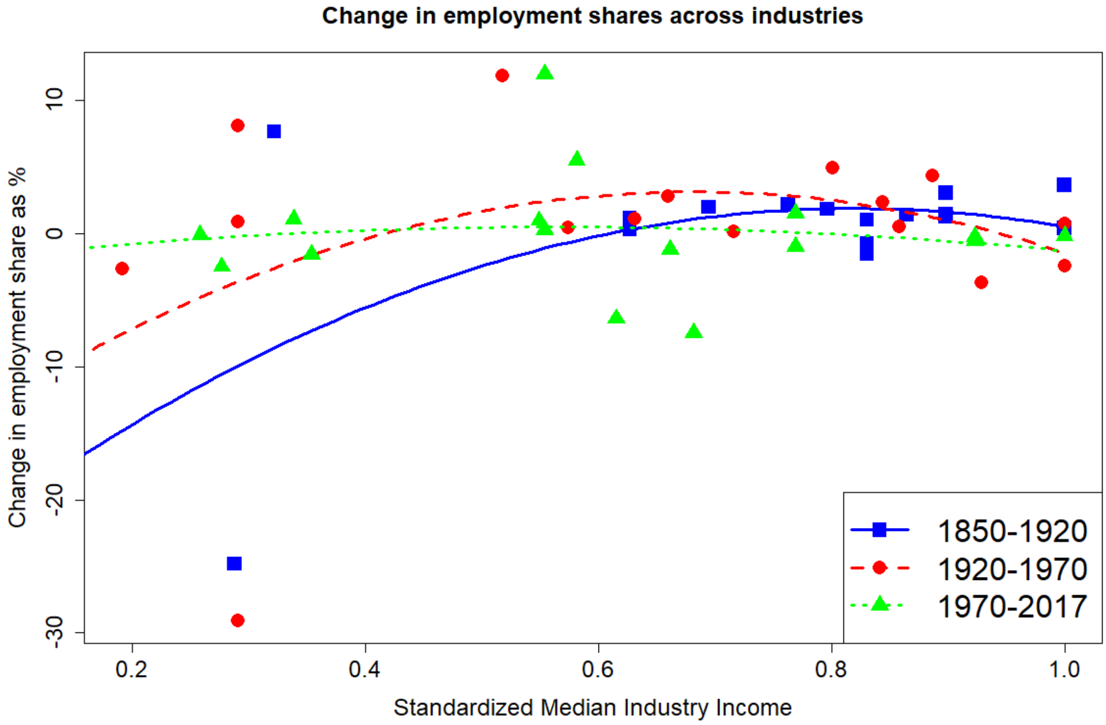
	Industry	1850-1950	1970	2000	2017
1	Retail Trade	3	3	1	1
2	Personal Services	2	1	1	2
3	Entertainment and Recreation	3	3	4	3
4	Agriculture	1	3	3	4
5	Construction	9	12	7	5
6	Transportation	15	14	10	6
7	Professional and Related	5	5	5	6
8	Business and Repair	9	7	6	8
9	Manufacturing, Nondurable goods	6	6	7	9
10	Wholesale Trade	13	11	9	10
11	Manufacturing, Durable goods	13	13	12	11
12	Finance, Insurance, Real Estate	7	8	11	12
13	Public Administration	13	10	13	12
14	Mining	11	15	14	14
15	Telecommunications	9	9	16	14
16	Utilities and Sanitary	15	15	15	16

Figure B.1 show how the change in employment shares varies across industries.

For the x-axis, we divide the median income for each industry group in the latter year of the period by the maximum industry in that period. We use this standardized income measure instead of rank for plotting. The qualitative results are identical if we used rank; the primary reason we use standardized income here is to indicate the range of incomes, so that industries closer together in income are treated as being closer together on the plots.

Note the results contrast with those obtained using occupational groups as there appears to be no difference in the change in employment shares over income using industry. As noted in the main results, this doesn't nullify the earlier findings since

Figure B.1: Change in Employment Shares using Income by Industry



it is merely a reflection of the large within-industry variations in incomes. One way to interpret these results in light of the occupational ones is that there has been little re-allocation across industries in terms of workers (at least in any way that's related to the industries' incomes), but within each industry, there has been significant re-allocation of employees, so that middle-class jobs are being cut.

To give an analogy, using the burger flippers example, the share of people working in the restaurant industry has not changed much, but those who do work in it now are either service workers "flipping burgers" or well-compensated professionals/executives with fewer mid-level workers.

B.1.1 Migrant Status

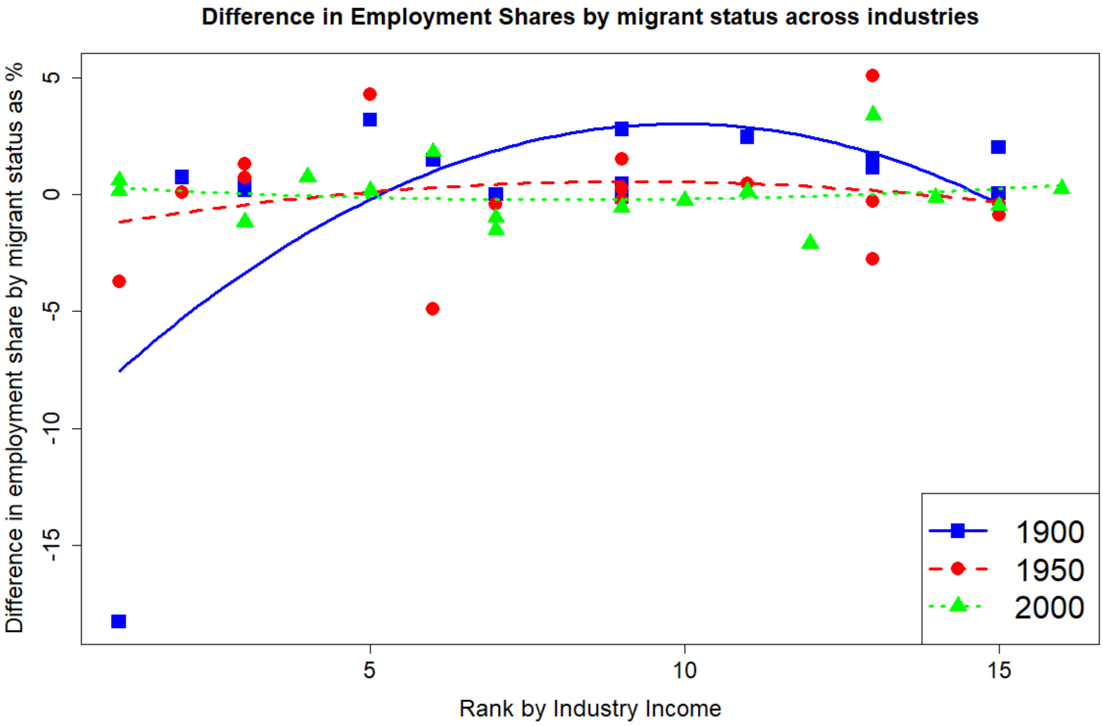
We look at differences in employment shares by migrant status as in the main paper, but this time for industry groups, as seen in Figure B.2.

The results are consistent with our earlier results, showing that migrants tended to have higher-paying jobs in 1900, but that this effect is hardly discernible in recent years.

B.2 Intergenerational Mobility by Industry

We next run the same methods for intergenerational mobility as in the main results but for industry. In Figure B.3, we took the median out-switching cost for each industry in each year and plotted the line graphs.

Figure B.2: Difference in Employment Shares using Industry Income by Migrant Status



We can see that the trend of low and declining intergenerational migration costs that we found with occupations also holds with industries. We also note that agriculture tends to have very high migration costs, especially early on, before declining considerably in recent years. This is also in line with the earlier results, suggesting that switching out of agriculture was very difficult, particularly, from 1850 to 1980 or so. Lastly, we can see that retail trade has had an increase in switching costs, also in line with the rise in migration costs for service workers we saw earlier.

Since the legend in Figures B.3 and B.4 use abbreviations for each industry, we have included a table with the industry's corresponding names in Table B.2.

Lastly, the results from the heat maps for four years using industry yields similar results as for the occupations, as can be see in Figure B.4. First, note that due to the high costs for agriculture until 1990, we exclude agriculture from the heat maps until then and so in the figure, agriculture is only included for 2017. This is because agriculture's migration costs dwarf all other costs, so the heat maps are not illustrative with its inclusion.

The findings are the same using industries as they were with occupations. Migration costs are much higher for lower-paying occupations than for higher-paying ones, suggesting a lack of intergenerational upward mobility relative to the intergenerational downward mobility. Furthermore, by 2017, the costs of switching out of retail trade are very large. Lastly, costs to switching out of mining also appear large in the historical data, so that children of miners often ended up becoming miners themselves.

Figure B.3: Intergenerational Occupation Switching Costs over Time

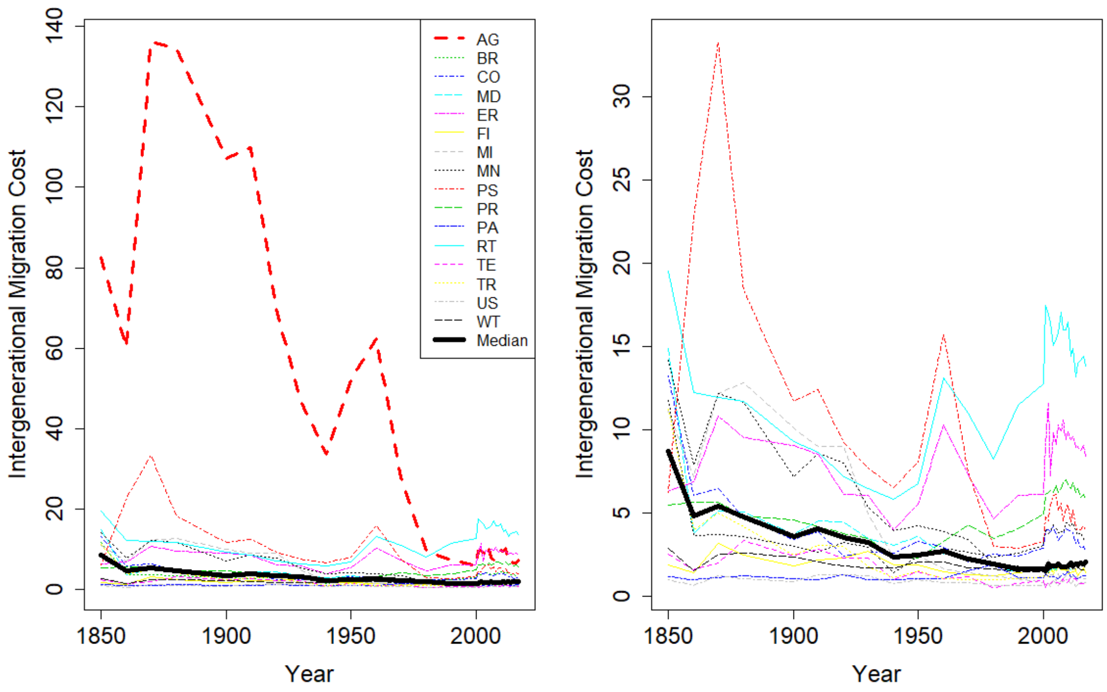
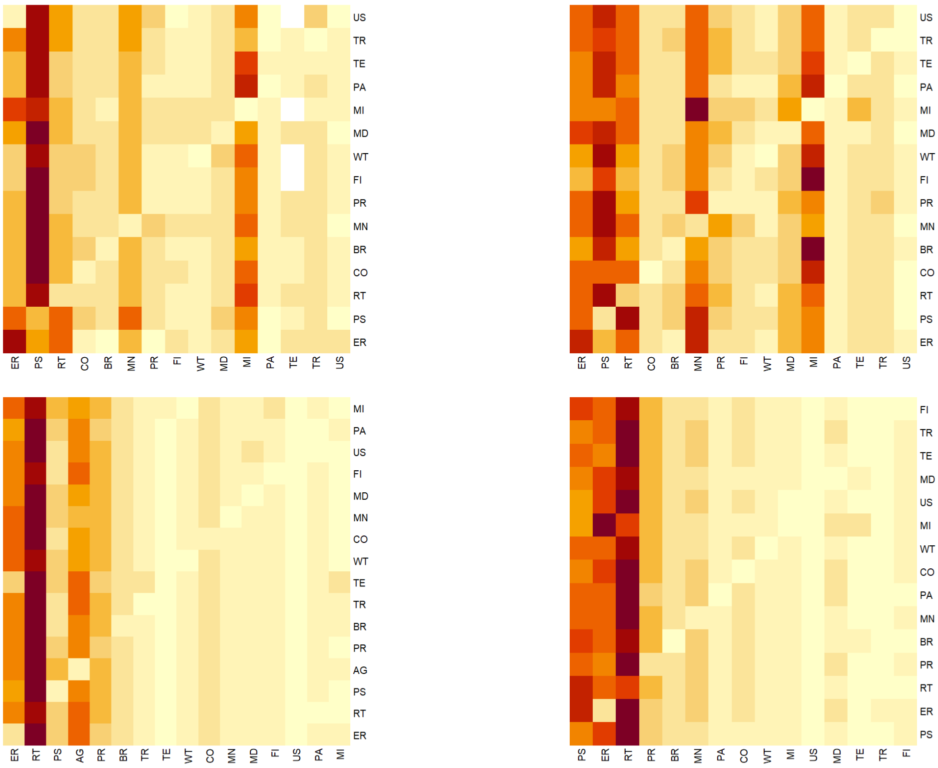


Table B.2: Industry Abbreviations

Name	Abbreviation
Agriculture	AG
Mining	MI
Construction	CO
Manufacturing, Durable goods	MD
Manufacturing, Nondurable goods	MN
Transportation	TR
Telecommunications	TE
Utilities and Sanitary	US
Wholesale Trade	WT
Retail Trade	RT
Finance, Insurance, Real Estate	FI
Business and Repair	BR
Personal Services	PS
Entertainment and Recreation	ER
Professional and Related	PR
Public Administration	PA
Common or general laborer	LA

Figure B.4: Intergenerational Industry Switching Costs Heat Maps



Note: Years in clockwise order from the top-left: 1870, 1920, 1970, 2017