FRE 502 - Gravity Model

Isaac Qi

November 5th, 2025

1 Questions 1-6

1.1 Descriptive Statistics

Table 1: Descriptive Statistics for Business Sector Trades in 2005

Variable	Mean	Min	Max	SD
Trade	78.22	0.00	10,740.00	511.02
Distance	5,764.48	59.62	19,586.18	4,667.77
GDP Importer	591,645.80	299.62	10,936,700.00	1,587,973.45
GDP Exporter	591,645.80	299.62	10,936,700.00	1,587,973.45
Contiguity	0.04	0.00	1.00	0.20
Common Language	0.07	0.00	1.00	0.25
Common Colonizer	0.02	0.00	1.00	0.12
Colony	0.04	0.00	1.00	0.18

Note: GDP are reported in millions.

1.1.1 Identify the number of countries and country pairs in the data who show non-zero trade in 2005.

The number of countries that traded Business Services in 2005 is 72, and the number of non-directional country pairs is 688.

1.1.2 What is the share of countries who trade Business Services in 2005?

The share of countries who traded Business Services in 2005 is 41.6%.

1.2 Graphical Correlation Analysis

1.2.1 Trade and Distance

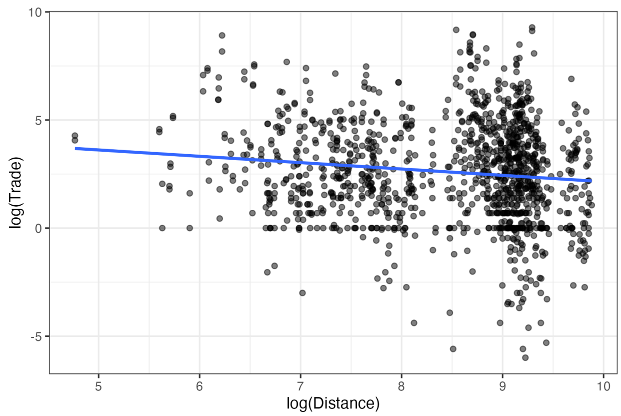

Figure 1: log(Trade) vs log(Distance)

There seems to graphically exist a relationship between trade and distance, where an increase in distance leads to a decrease in trade. This intuitively matches our hypothesis, where countries that are closer to each other would trade more due to proximity.

1.2.2 Trade and Combined GDP

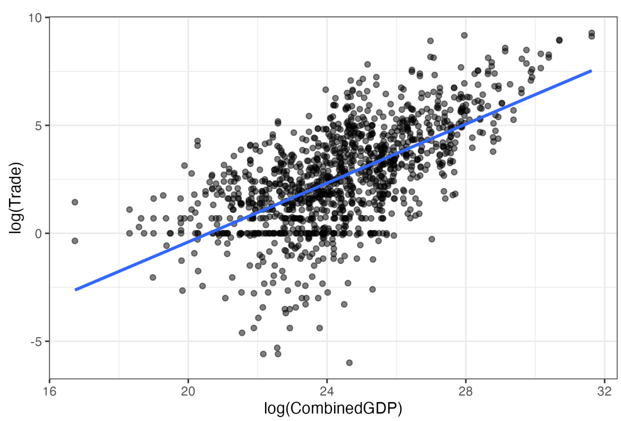

Figure 2: log(Trade) vs log(CombinedGDP)

As the product of the importing and exporting country's GDP increases, trade increases. This also makes sense intuitively, where when countries are economically larger, they are able to engage in higher levels of trade.

1.2.3 Contiguity

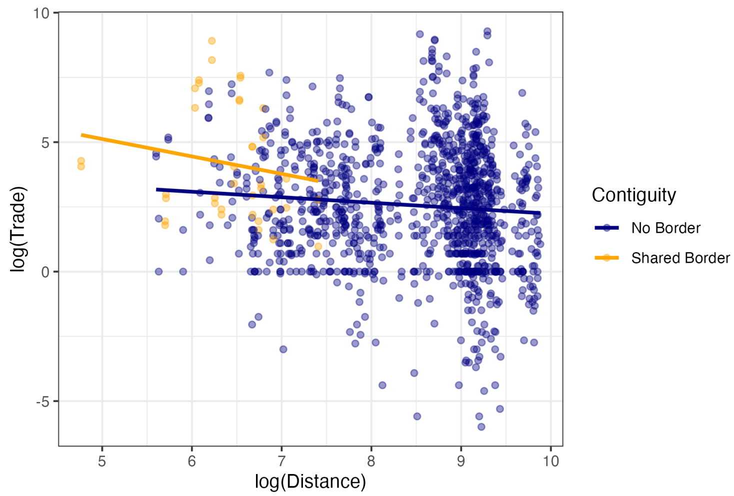

Figure 3: log(Trade) vs log(Distance) by Contiguity

There appears to be a relationship where when two countries share a border and engages in trade, distance has a larger effect where the shorter the distance the higher the amount of trade. This likely reflects factors such as ease of transport in and out of the countries directly next to each other.

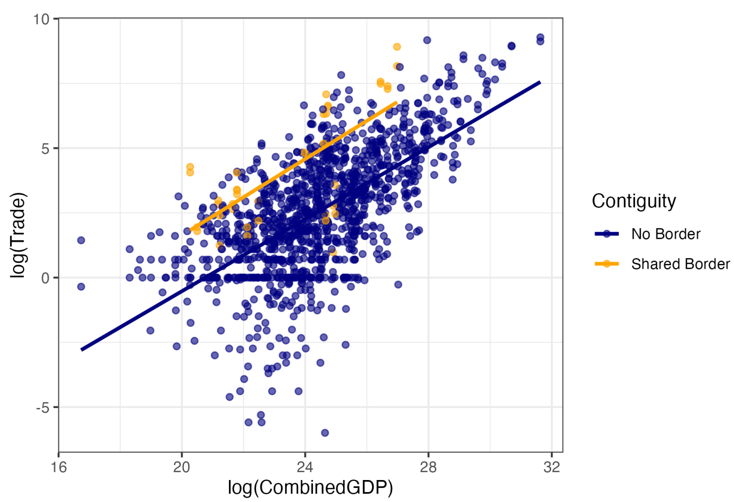

Figure 4: log(Trade) vs log(CombinedGDP) by Contiguity

In terms of contiguity's effect on GDP and trade, countries that share a border trade more than non-contiguous pairs at any given level of combined GDP. The similar slopes indicate that GDP affects trade similarly across both groups, while the upward intercept shift reflects lower trade costs among neighbouring countries.

1.2.4 Common Language

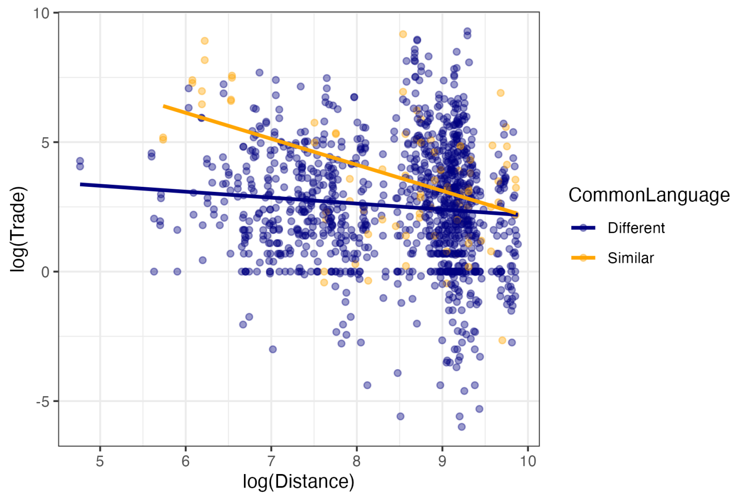

Figure 5: log(Trade) vs log(Distance) by Common Language

Countries that share a common language trade more at all short/medium distances, characterized by the orange regression line being above the navy regression line at all distances. However, the steeper negative slope indicates that the trade advantage associated with a common language diminishes as distance increases, with the two groups converging at larger distances.

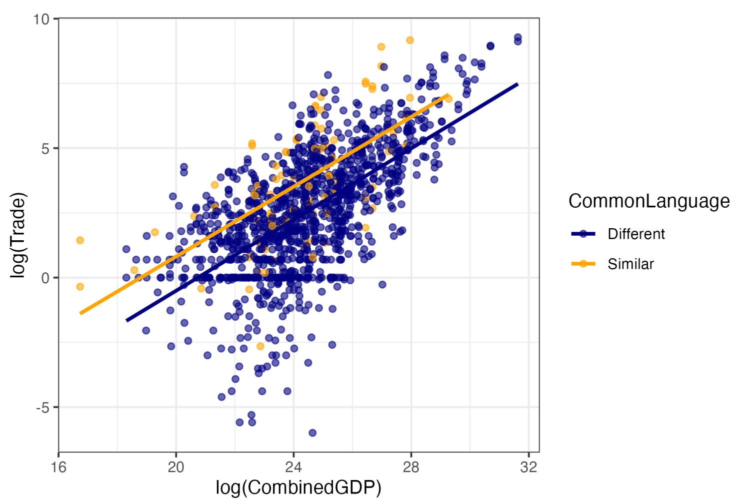

Figure 6: log(Trade) vs log(CombinedGDP) by Common Language

The relationship between trade and combined GDP is similar across both groups, as shown by the nearly parallel slopes. However, the shared-language regression line is shifted upward, indicating that for any given level of combined GDP, country pairs with a common language consistently trade more.

1.2.5 Correlation Matrix

Table 2: Correlation Matrix for Business Services Trade Variables

	Trade	Distance	CombinedGDP	Contiguity	CommonLanguage
Trade	1	-0.031	0.751	0.098	0.095
Distance	-0.031	1	0.059	-0.233	0.037
CombinedGDP	0.751	0.059	1	-0.026	-0.026
Contiguity	0.098	-0.233	-0.026	1	0.100
CommonLanguage	0.095	0.037	-0.026	0.100	1

The correlation matrix supports the graphical analysis prior. Trade is strongly positively correlated with CombinedGDP (0.751), consistent with the result that larger combined economic size is associated with higher trade volumes. The correlations between Trade and Contiguity (0.098) and between Trade and CommonLanguage (0.095) are positive but small, which aligns with the upward intercept shifts observed in the plots, where shared borders and shared language increase baseline trade.

The correlation matrix did not capture the interaction effects characterized by the steeper slopes in the contiguity and common language figures, as it only shows pairwise linear relationships.

1.3 Intuitive Gravity Model

Table 3: Intuitive Gravity Model

	Dependent variable:
	log(Trade)
log(GDP Importer)	0.894***
	(0.030)
log(GDP Exporter)	0.787***
	(0.028)
log(Distance)	−0.944***
	(0.059)
Contiguity	0.249
	(0.300)
Common Language	1.320***
	(0.204)
Common Colonizer	0.243
	(0.951)
Constant	−33.258***
	(1.240)
R^2	0.561
Adjusted R^2	0.559
F Statistic	196.5*** (df = 6; 687)
Note:	*p<0.1; **p<0.05; ***p<0.01
	Robust SEs clustered by Distance.

1.3.1 Model Quality

The intuitive gravity model provides an acceptable fit to the data; the R^2 and adjusted R^2 value of 0.561 and 0.559, respectively, indicates that approximately 56% of the variation in bilateral trade flows is explained by the regression variables. The F-statistic of 196.5 ($p < 0.01$) strongly rejects the null hypothesis that all slope coefficients are jointly equal to zero. The model as a whole is statistically significant, meaning that the included explanatory variables collectively have explanatory power.

1.3.2 Distance on Trade

The coefficient on log(Distance) is -0.944 and is statistically significant at the 1% level. Given that this is a log-log model, the model estimates that a 1% increase in Distance is approximately a 0.944% decrease in bilateral trade flows, holding all other factors constant.

This result is intuitively consistent where we expect greater physical distance to increase transportation costs, reduces information exchange, thus making trade more difficult to carry out. The near 1 to -1 relationship implies that trade is highly sensitive to distance, meaning that even moderate increases in geographic separation can reduce trade volumes by a large margin.

1.3.3 Relevance for Trade Partners

i. to have a common border: The coefficient on Contiguity is positive but not statistically significant, suggesting that sharing a land border does not have any effect on trade in this dataset once other factors are controlled for.

ii. to speak the same language: The Common Language variable is positive and highly significant. Trade partners who speak a common language trade more with each other, even after controlling for distance and GDP.

iii. to have been in some sort of colonial relationship with the same colonizing nation: The Common Colonizer variable is not statistically significant, where there is no clear evidence that countries once colonized by the same empire trade more with each other once we control for language, distance, and GDP.

1.3.4 Effect of GDP on Trade

Both the importer's and the exporter's GDP have positive and highly significant effects on trade. The coefficients on log(GDPImporter) and log(GDPExporter) can be interpreted as elasticities given that this is a log-log model:

1) A 1% increase in the importer's GDP is associated with approximately a 0.89% increase in bilateral trade, holding all else constant, and
2) A 1% increase in the exporter's GDP is associated with approximately a 0.79% increase in bilateral trade, holding all else constant.

1.3.5 Joint Hypothesis

The joint hypothesis test that Contiguity, Common Language, and Common Colonizer have no effect on trade is rejected at the 1% significance level ($F(3, 1234) = 15.56$, $p < 0.001$). This indicates that the group of dichotomous (dummy) variables is collectively relevant for explaining bilateral trade patterns. However, the previous regression result suggests that the significance may be primarily driven by common language, as the individual effects of contiguity and shared colonizer are not statistically significant once we control for economic size and distance.

1.4 Structural Gravity Model

Table 4: Comparison of Intuitive and Structural Gravity Models

	Model 1: Intuitive (log(Trade))	Model 2: Structural FE (log(Trade))
log(GDP Importer)	0.894^{***}	
	(0.030)	
log(GDP Exporter)	0.787^{***}	
	(0.028)	
log(Distance)	-0.944^{***}	-1.216^{***}
	(0.059)	(0.076)
Contiguity	0.249	0.318
	(0.300)	(0.264)
Common Language	1.320^{***}	0.085
	(0.204)	(0.182)
Common Colonizer	0.243	-1.648^{***}
	(0.951)	(0.548)
Fixed Effects		
Exporter FE	No	Yes
Importer FE	No	Yes
Observations	1,241	1,227
R^2	0.561	0.810
F Statistic	196.5^{***} (df = 6; 687)	—
Joint Wald Test	—	85.4^{***} (df = 4; 1100)
Note:	Standard errors clustered by Distance in parentheses. *p<0.1; **p<0.05; ***p<0.01	

The structural gravity model shows a much better fit compared to the intuitive gravity model in Question 3. The R^2 increased from 0.56 to 0.81, indicating that importer and exporter fixed effects explain a much larger share of variation in bilateral trade flows. This improvement is likely due to country fixed effects absorbing many determinants of trade that are unique to the country.

Overall, The structural model is a stronger and more aligned model compared to the intuitive gravity model.

1.4.1 Exclusion of GDP

When we include importer and exporter fixed effects, we are allowing each country to have its own intercept that captures all time-invariant country specific characteristics that affect its overall level of trade. Since GDP is a country-specific characteristic, its influence is absorbed by the fixed effects, leaving no remaining independent variation for the GDP variables to explain. If I were to include log(GDP) in a model that already has importer and exporter fixed effects, the GDP variables would likely be perfectly collinear.

The importer and exporter fixed effects also capture any other country-specific factors that do not vary across partners. Some examples being population size, institutional quality, and willingness to trade.

1.4.2 Model Comparison

The estimated elasticity of trade with respect to distance is larger in this model, from -0.94 to -1.22. This likely implies that once we control for country specific characteristics, trade is even more sensitive to geographical distance. Contiguity and Common Language are both insignificant in this model, likely captured by the exporter and importer fixed effects. Interestingly, common colonizer is now very negative and significant, showing that once country fixed effects are accounted for, historical colonial ties affect trade patterns significantly.

1.5 OECD

Table 5: Comparison of Intuitive and Structural Gravity Models (with OECD)

	Model 1: Intuitive (log(Trade))	Model 2: Structural (log(Trade))	Model 3: OECD (log(Trade))
log(GDP Importer)	0.894^{***}		
	(0.030)		
log(GDP Exporter)	0.787^{***}		
	(0.028)		
log(Distance)	-0.944^{***}	-1.216^{***}	-1.234^{***}
	(0.059)	(0.076)	(0.0769)
Contiguity	0.249	0.318	0.256
	(0.300)	(0.264)	(0.2727)
Common Language	1.320^{***}	0.085	0.089
	(0.204)	(0.182)	(0.1816)
Common Colonizer	0.243	-1.648^{***}	-1.807^{**}
	(0.951)	(0.548)	(0.5601)
OECD Membership			0.346
			(0.2275)
Fixed Effects			
Exporter FE	No	Yes	Yes
Importer FE	No	Yes	Yes
Observations	1,241	1,227	1,227
Adjusted R^2	0.559	0.811	0.789
F Statistic (OLS)	196.5^{***} (df = 6; 687)	—	—
Joint Wald Test	—	85.4^{***} (df = 4; 1100)	69.1^{***} (df = 5; 1099)

Notes: SEs clustered by Distance in parentheses.
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

1.5.1 Graphical Analysis

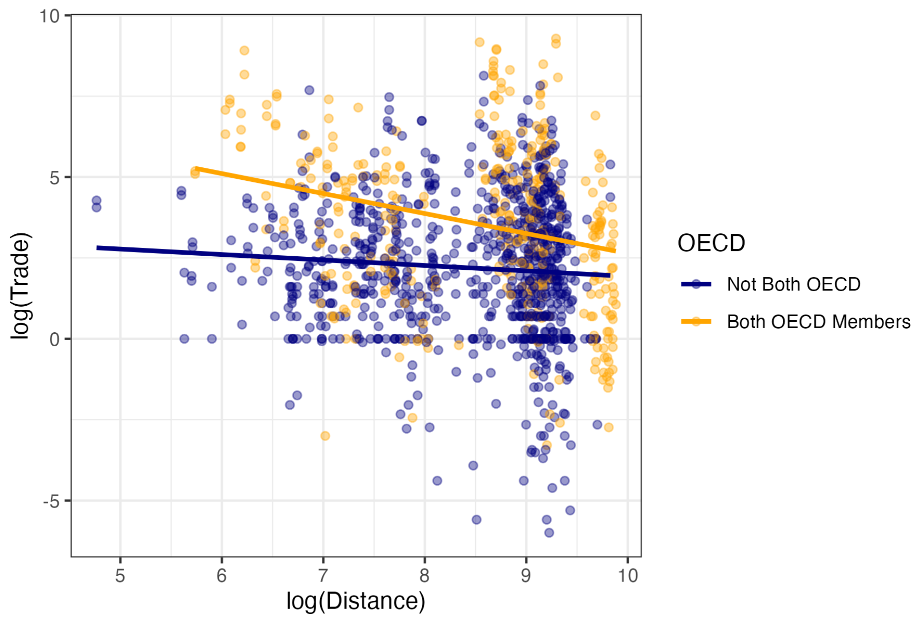

Figure 7: log(Trade) vs log(Distance) by OECD Membership

There appears to be a graphical relationship where when both country pairs are OECD members, more trade occur at all distances. This suggests that OECD membership represent some incentive for members to trade with each other.

1.5.2 Statistical Analysis

Statistically, OECD membership is positive but not significant. Model fit worsened slightly compared to the standard structural FE model. Distance is still significant and negative, indicating that for every 1% increase in distance, trade decreases by -1.23%.

1.6 WTO

Table 6: Comparison of Intuitive and Structural Gravity Models (OECD / WTO)

	Model 1: Intuitive (log(Trade))	Model 2: Structural (log(Trade))	Model 3: OECD (log(Trade))	Model 4: WTO (log(Trade))
log(GDP Importer)	0.894***			
	(0.030)			
log(GDP Exporter)	0.787***			
	(0.028)			
log(Distance)	−0.944***	−1.216***	−1.234***	−1.216***
	(0.059)	(0.076)	(0.0769)	(0.0762)
Contiguity	0.249	0.318	0.256	0.318
	(0.300)	(0.264)	(0.2727)	(0.2642)
Common Language	1.320***	0.085	0.089	0.085
	(0.204)	(0.182)	(0.1816)	(0.1819)
Common Colonizer	0.243	−1.648***	−1.807**	−1.648**
	(0.951)	(0.548)	(0.5601)	(0.5475)
OECD Pair			0.346	
			(0.2275)	
WTO Pair				*dropped (collinear)*
Fixed Effects				
Exporter FE	No	Yes	Yes	Yes
Importer FE	No	Yes	Yes	Yes
Observations	1,241	1,227	1,227	1,227
R^2	0.561	—	—	—
Adjusted R^2	0.559	—	0.789	0.789
Within R^2	—	0.810	0.285	0.282
F Statistic (OLS)	196.5*** (df = 6; 687)	—	—	—
Joint Wald Test	—	85.4*** (df = 4; 1100)	69.1*** (df = 5; 1099)	85.4*** (df = 4; 1100)

Notes: SEs clustered by Distance in parentheses.
WTO is dropped in Model 4 for perfect collinearity with the FE structure.
*p<0.1; **p<0.05; ***p<0.01.

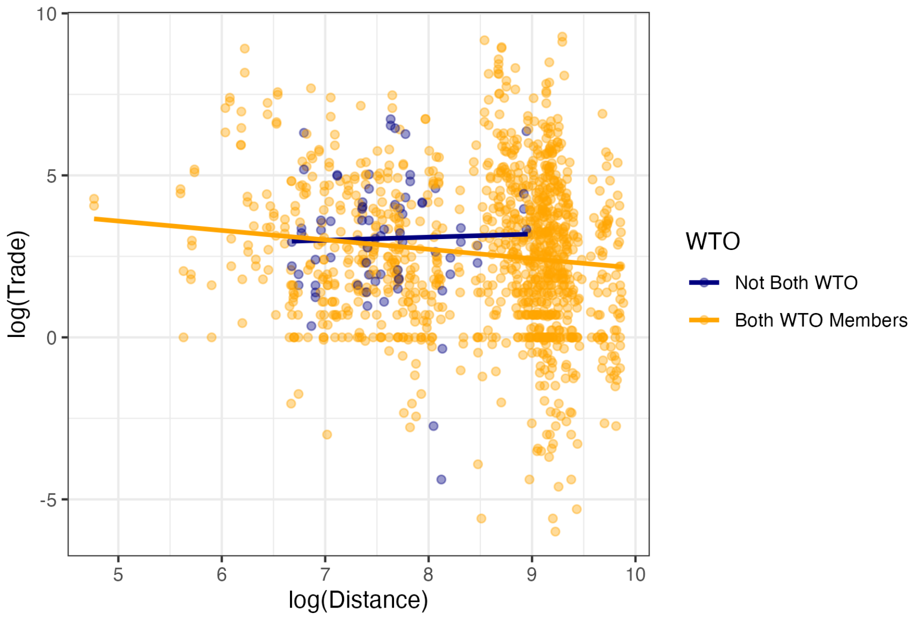

Figure 8: log(Trade) vs log(Distance) by WTO Membership

1.6.1 Statistical & Graphical Analysis

The model eliminated WTO membership when running the model, likely due to the high concentration of WTO members in this dataset leading to multicollinearity. This high concentration is evident graphically in Figure 8. A regression table is included but the results are not different than compared to the baseline structural fixed effects model.

1.7 Discussion

The most fundamental problem that biases the models is that OECD and WTO membership are not randomly assigned. Countries that trade more are more likely to join these organizations, creating reverse causality. The question asked by the assignment suggests that countries who join the organizations end up trading more, but in reality, countries likely joined the WTO/OECD because they wanted to trade more. This bidirectional relationship means my estimates are biased and cannot be interpreted causally.

Secondly, countries self-select into these organizations based on unobservable characteristics that also affect trade, which are captured by the country fixed effects. In addition, most countries in this dataset are WTO members, which led to the multicollinearity issue as it was impossible to identify WTO membership effect through the dummy. The OECD dummy has the opposite problem, where very few country pairs have both members in the OECD, limiting the model's statistical power despite seeing a clear trend graphically.

Third, there are policies that impact certain country pairs that are not captured. For example, it's possible that members of the WTO or OECD have agreements outside of what we're capturing, such as tax treaties or trade agreements. It's also the case that WTO membership affects different countries differently. For example, if both countries in the WTO pair are developed countries, this relationship may matter less than if one of the countries is a developing country. The single dummy variable is not capable of capturing all of these variations.

In conclusion, we cannot draw causal effects from these models. Many requirements for causal inference are not met. What we found here are correlations, not causation.