

## Economics 714 - Computational Methods in Economics

University of Pennsylvania

Fall 2021

---

# Replication

Javier Tasso

## Introduction

Originally we planned to replicate Eaton & Kortum (2002) using data from Kortum's website [here](#). We couldn't open most of the trade data needed to do the replication, so we choose to (try to) replicate Waugh (2010) instead. In the first section we verify that the gravity data Waugh has in his repository matches the relationship found in Eaton & Kortum and also that it is consistent with our intuition. In section 2 we combine datasets, plot some information and estimate the preliminary value of  $\theta$  which is our parameter of interest. Section 3 tries to replicate the results of the geographical barriers of Waugh. We couldn't replicate them, but we added a slight modification and tried to say something about the size of the economy and trade shares using his methods. Finally section 4 tries to diagnose what went wrong by looking at the data and the constructed measures we used in previous sections. Even though we weren't successful in the replication, we wanted to present the results anyways to show our work.

- Gravity data and country names were obtained from [this repository](#) that belongs to Mike Waugh.
- Price data as well as other inputs were downloaded from [here](#).
- Click [here](#) for the repository. Every input needed should be there and the codes themselves generate all the output for the tex file.

## 1 Preliminary Data

For this section we took advantage of the file *gravity\_data.txt* located at [this repository](#). Code *1\_replication\_Eaton\_Kortum.R* loads the data and generates figure (1). This figure should be similar to figure (1) in Eaton & Kortum (2002). We see a negative relationship between the normalized import share and distance. There are three countries missing (relative to Eaton Kortum (2002)) so the figure is not exactly the same. This relationship we see in figure (1) makes sense: as distance increases, the normalized import shares falls because things become more expensive.

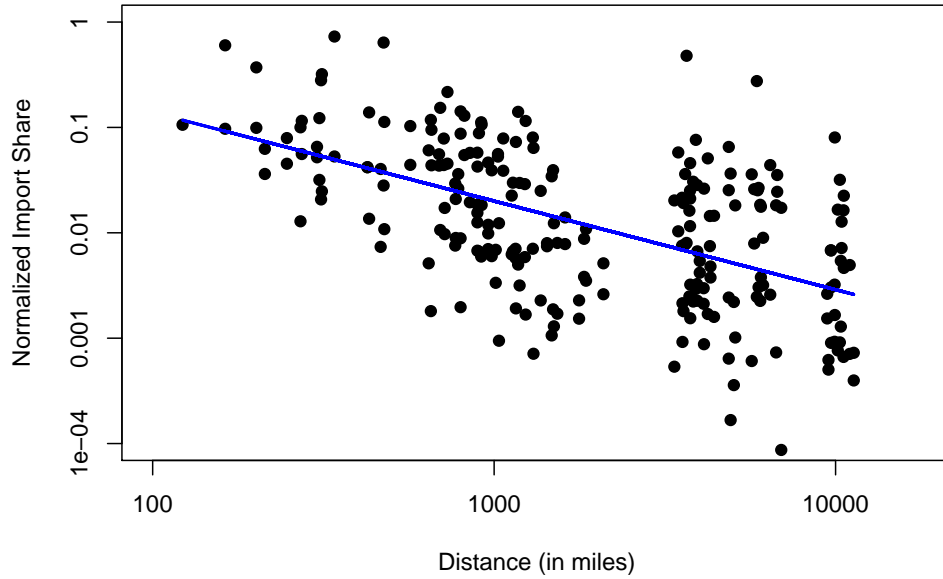


Figure 1: Distance and Normalized Import Shares

## 2 Construction of the Price Measure

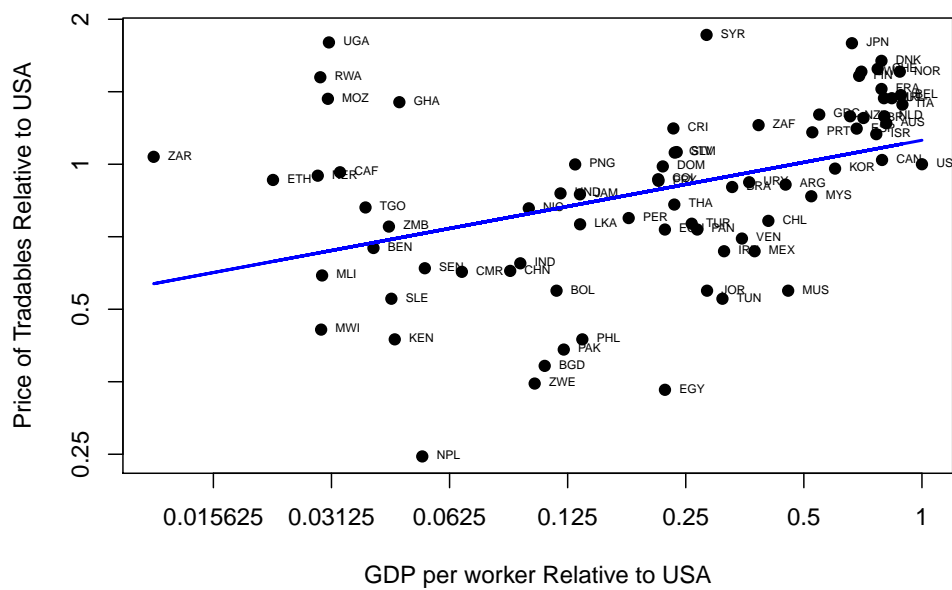
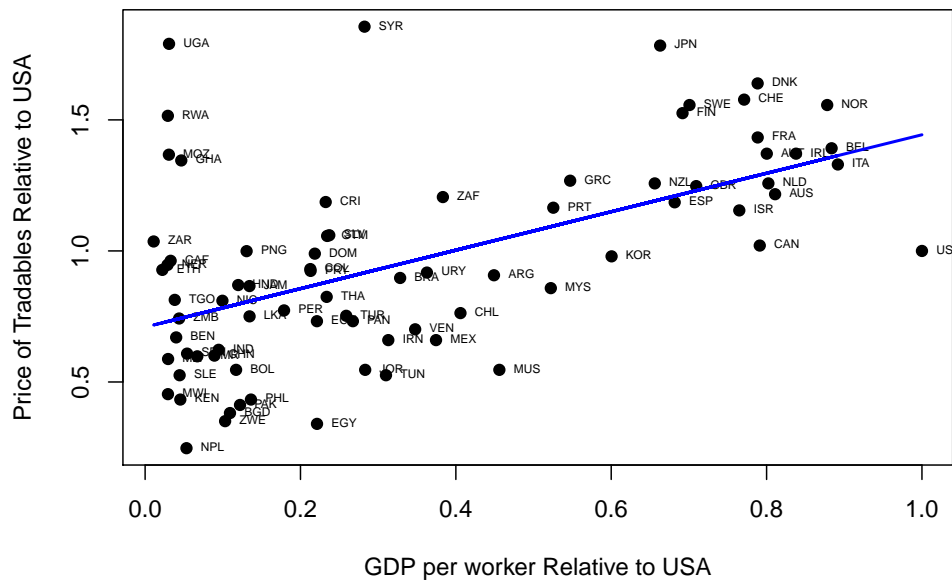
To construct the price measure we follow Waugh (2010). All the data can be downloaded from [this link](#). Code `2_replication_Waugh.R` generates all the output of this section. Table (1) corresponds to table (1) in the paper and presents the trade shares in 1996 for selected countries.

	USA	CAN	CHN	JPN	MEX	MWI	SEN	ZAR
USA	83.25	39.74	3.63	2.27	31.63	1.58	2.16	2.93
CAN	3.79	49.21	0.32	0.22	0.72	0.67	0.57	0.51
CHN	1.79	1.42	77.61	1.45	0.31	2.51	2.70	6.81
JPN	3.04	2.01	7.00	92.56	1.60	2.65	1.34	0.82
MEX	1.89	1.34	0.06	0.02	61.10	0.00	0.01	0.01
MWI	0.00	0.00	0.00	0.00	0.00	41.52	0.00	0.00
SEN	0.00	0.00	0.00	0.00	0.00	0.00	52.68	0.00
ZAR	0.00	0.01	0.00	0.00	0.00	0.00	0.00	51.53

Table 1: 1996 Trade Shares for Selected Countries

Figures (2) and (3) are equivalent to figure (1) in the paper. Note that figure (3) has a different scale on the axis.

Let  $X_{ij}$  be the the amount of goods (aggregated somehow) that country  $j$  imports from country  $i$  and let  $X_{jj}$  be the amount of goods that country  $j$  buys from itself. Next define  $p_j$  and  $p_i$  price indexes and  $\tau_{ij}$  a parameter that governs trade costs. Then we have the following relationship which is equation (4) in the paper. .



$$\frac{X_{ij}}{X_{jj}} = \left( \tau_{ij} \cdot \frac{p_j}{p_i} \right)^{-\frac{1}{\theta}}$$

Note that the right hand side is a factor that takes into account the relative prices between home and country  $i$ , the trade costs possible arising from distance and other barriers and a parameter  $\theta > 0$ .  $\theta$  is a parameter that controls the strength of the comparative advantage of the countries and we want to estimate it. Take logs to get the following.

$$\ln \left( \frac{X_{ij}}{X_{jj}} \right) = -\frac{1}{\theta} \cdot \ln(\hat{\tau}_{ij}) \quad (1)$$

The left hand side is observed. We will recover from data  $\ln(\hat{\tau}_{ij})$  and then run a regression with no intercept to get an estimate for  $\theta$ . The next step is to construct this measure that involves prices. Let  $l$  be an index for goods from price information we will construct that measure as follows.  $\max^2$  represents the second highest value.

$$\ln(\hat{\tau}_{ij}) = \max_l^2 \{ \ln(p_i(l)) - \ln(p_j(l)) \}$$

For each pair of countries  $(i, j)$  we are choosing the good that makes the relative price the highest, using the second maximum to get a better measure. That's going to give us an idea of the trade costs between countries  $(i, j)$ . We expect a negative relationship between normalized trade shares (measured in logs) and the measure of  $\tau$  constructed from price data (also measured in logs). We see that in figure (4).

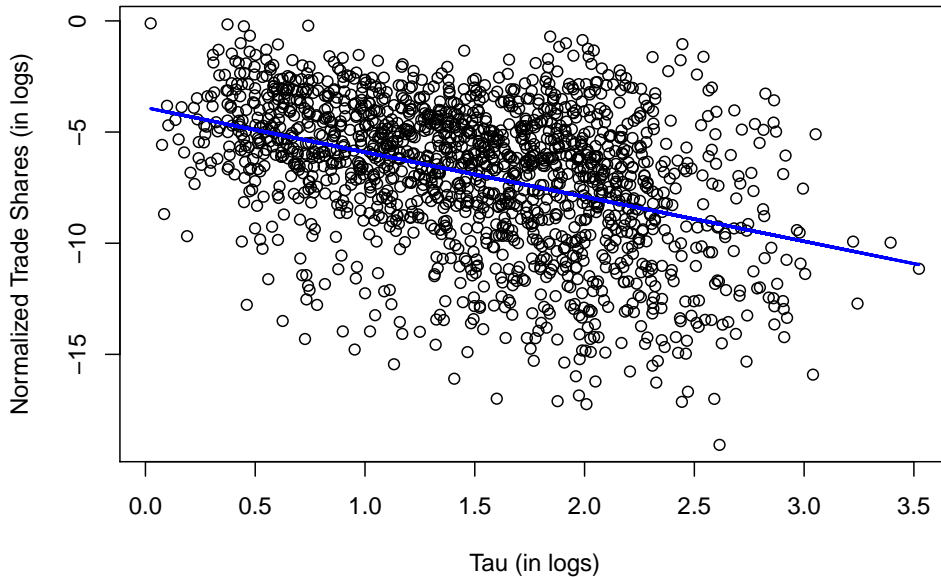


Figure 4: Trade Shares and  $\tau$

Figure (4) was generated in `2_replication_Waugh.R`. The code reshapes and merges different datasets that will be used in next section. Next we regress the normalized trade

shares in our measure of  $\hat{\tau}$  without intercept to get the results of table (2). These results imply a  $\hat{\theta} \simeq 0.24$  which differs from the one found in Waugh (2010). We will continue with the analysis and then see if we can find the reason why we get different results.

Table 2: First Estimate of Theta

<i>Dependent variable:</i>	
Trade Share (in logs)	
Tau (in logs)	-4.189*** (0.054)
Observations	1,592
R <sup>2</sup>	0.805
Adjusted R <sup>2</sup>	0.804
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

### 3 Different Regressions

Code `3_replication_Waugh.R` does everything for this section. In the first part of the code we generate dummy variables for different levels of distance (in miles) this will help us divide the sample in 6 groups depending on how close or how far apart are two countries. The groups are  $\text{dist} < 375$ ,  $375 \leq \text{dist} < 750$ ,  $750 \leq \text{dist} < 1500$ ,  $1500 \leq \text{dist} < 3000$ ,  $3000 \leq \text{dist} < 6000$  and  $6000 \leq \text{dist}$ . We weren't able to replicate the results from table (2) of Waugh's paper, but we will present some estimations in this section and then try to discover what went wrong. We want to extend equation (1) allowing for different intercepts (fixed effects for importers and exporters), and controlling for distance to see the behaviour of the coefficient next to the constructed  $\ln(\hat{\tau}_{ij})$ . Table (3) shows the results. From table (3) we learn that controlling for distance and allowing for different intercepts changes the estimates of  $\theta$  by a lot. The theory predicts that we should not have an intercept, but it does not seem to fit the data so well. We could already see that in figure (4) where the intercept of the line does not seem to be zero at all. Following Waugh we want to check the correlation between exporter fixed effects (or importer fixed effects) and GDP per worker. Figure (5) plots the exporter fixed effects against GDP. This figure should be similar to figure (2) in Waugh (2010).

To get a better understanding of what is going on here, we are estimating a relationship that has the following form.

$$\ln\left(\frac{X_{ij}}{X_{ii}}\right) = -S_i - \frac{1}{\theta} \ln(\hat{\tau}_{ij}) \quad (2)$$

From our estimates we recover  $-S_i$  for each country and then we are correlating that with GDP. Trade cost have the same impact on normalized trade shares (this is given by  $-1/\theta$ ), but the relationship has different constants and these constants are correlated with GDP. A similar relationship is true for importer fixed effects as we can see in figure (6).

Table 3: Estimates of Theta - Different Models

	<i>Dependent variable:</i>			
	Trade Share (in logs)			
	No FE	FE exporter	FE importer	FE both
	(1)	(2)	(3)	(4)
Tau (in logs)	−1.703*** (0.123)	−0.402*** (0.079)	−2.015*** (0.128)	−0.301*** (0.083)
Dist 1	−1.019 (0.964)	−7.503*** (0.594)	1.328 (1.039)	−3.601*** (0.576)
Dist 2	−2.364*** (0.651)	−7.001*** (0.433)	0.288 (0.779)	−1.972*** (0.448)
Dist 3	−2.812*** (0.313)	−7.777*** (0.323)	−0.143 (0.592)	−2.844*** (0.371)
Dist 4	−3.613*** (0.270)	−8.178*** (0.308)	−0.975* (0.567)	−3.109*** (0.357)
Dist 5	−4.502*** (0.255)	−8.716*** (0.308)	−2.431*** (0.566)	−4.043*** (0.355)
Dist 6	−4.778*** (0.220)	−9.618*** (0.296)	−2.212*** (0.545)	−4.802*** (0.339)
Shared Border	0.102 (0.587)	0.380 (0.317)	0.118 (0.540)	0.639** (0.272)
Observations	1,557	1,557	1,557	1,557
R <sup>2</sup>	0.853	0.960	0.883	0.972
Adjusted R <sup>2</sup>	0.852	0.958	0.879	0.970

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

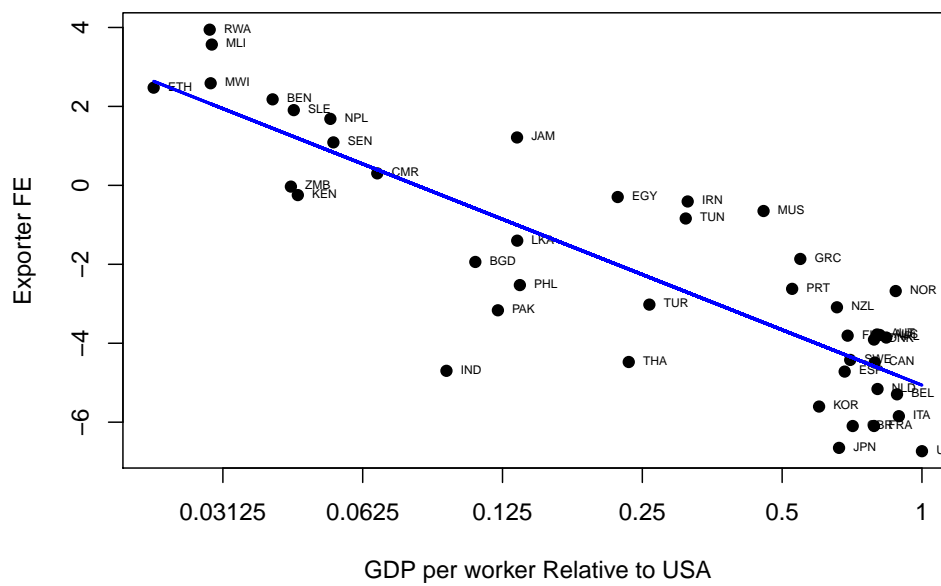


Figure 5: Exporter FE and GDP

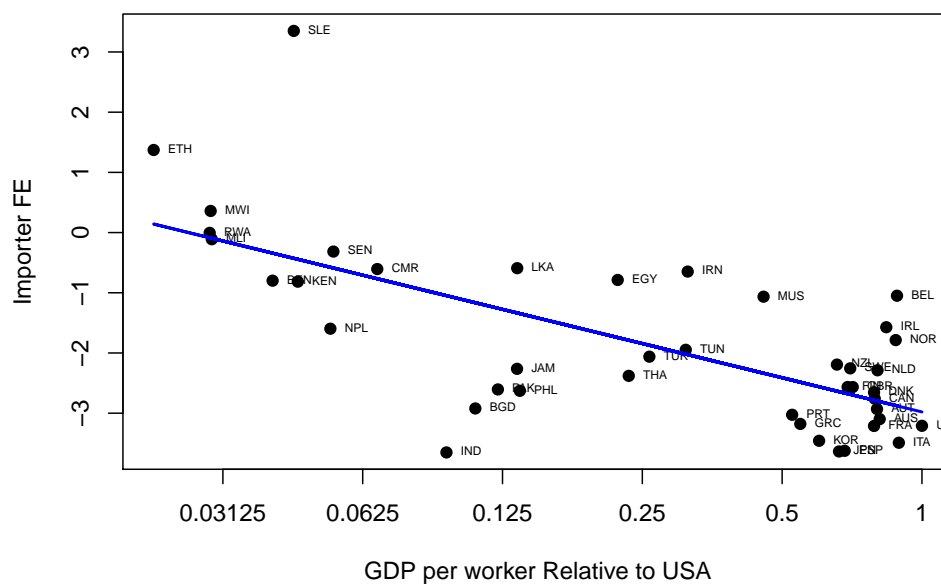


Figure 6: Importer FE and GDP

Here we are estimating a model that looks like the following equation, but including dummies for distance and border. GDP correlates positively with these measures too.

$$\ln\left(\frac{X_{ij}}{X_{ii}}\right) = S_j - \frac{1}{\theta} \ln(\hat{\tau}_{ij}) \quad (3)$$

We believe that Waugh (2010) does not include  $\tau$  in the regression when trying to explain the geographical barriers of table 2. Table (4) here is the equivalent of that analysis. We see that a higher distance is associated with a smaller normalized trade share, but we weren't able to replicate the same results. Next section will try to identify mistakes in our analysis.

Table 4: Geographic Barriers

	<i>Dependent variable:</i>			
	Trade Share (in logs)			
	No FE	FE exporter	FE importer	FE both
	(1)	(2)	(3)	(4)
Dist 1	−1.901* (1.018)	−7.827*** (0.595)	0.255 (1.118)	−9.087*** (0.565)
Dist 2	−3.701*** (0.682)	−7.453*** (0.427)	−1.088 (0.835)	−7.506*** (0.429)
Dist 3	−4.390*** (0.308)	−8.264*** (0.311)	−1.786*** (0.629)	−8.398*** (0.352)
Dist 4	−5.691*** (0.238)	−8.760*** (0.289)	−3.209*** (0.592)	−8.727*** (0.334)
Dist 5	−7.359*** (0.158)	−9.449*** (0.275)	−5.617*** (0.570)	−9.762*** (0.327)
Dist 6	−7.475*** (0.108)	−10.348*** (0.261)	−5.398*** (0.546)	−10.566*** (0.306)
Shared Border	0.001 (0.621)	0.370 (0.320)	−0.011 (0.583)	0.639** (0.274)
Observations	1,557	1,557	1,557	1,557
R <sup>2</sup>	0.835	0.959	0.863	0.972
Adjusted R <sup>2</sup>	0.834	0.958	0.859	0.970

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01



## 4 Diagnosis

Something went wrong during the replication. In this section we want to see if we can spot the mistake. We consider three possible problems.

1. The measure  $\ln(\hat{\tau}_{ij})$  doesn't behave well with respect to distance.
2. The measure  $\ln(\hat{\tau}_{ij})$  doesn't behave well with respect to trade shares. We consider all the sample or only OECD countries to check for this issue.
3. Sample is different. To address this we repeat the analysis using only OECD countries.

First let's find the correlation between this variable and distance (also measured in logs). We should find a positive relationship, as distance increases costs are higher. That's what we see in figure (7).

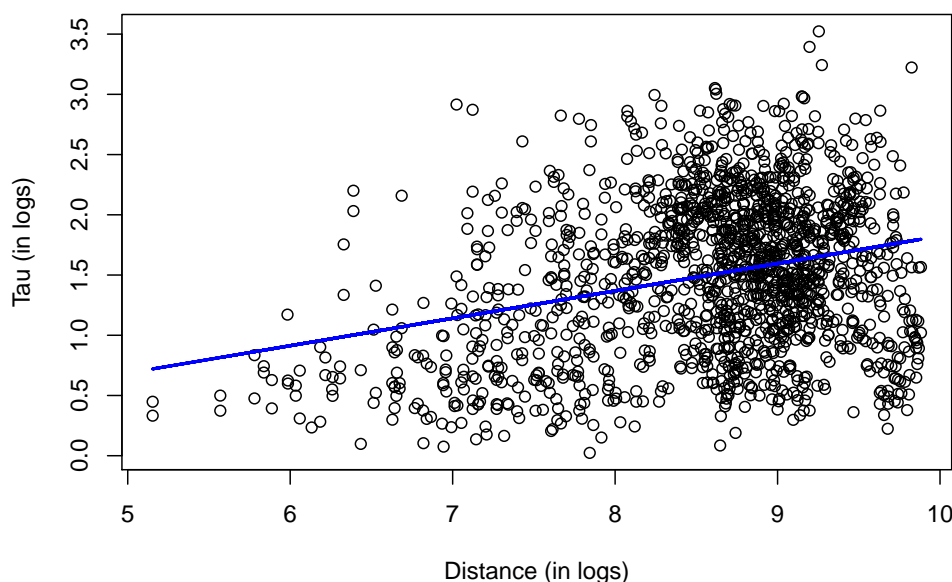


Figure 7: Correlation between Distance and  $\tau$

As  $\tau_{ij}$  increases the normalized trade share should decrease. Here we should get something similar to figure (2) of Eaton Kortum (2002). We think we get something similar as we saw in figure (4) before. If we restrict the attention to OECD countries we get something that has the flavour of EK (2002) plot but with a different measure for the  $D_{ni}$ . See figure (8).

Finally we want to check if using only OECD data improve the results and make them closer to the ones in the paper. Table (5) presents the geographical barriers regressions. The estimates have the same sign but the magnitudes are different.

Table 5: Geographic Barriers

	<i>Dependent variable:</i>			
	Trade Share (in logs)			
	No FE	FE exporter	FE importer	FE both
	(1)	(2)	(3)	(4)
Dist 1	−2.211*** (0.717)	−0.584 (0.472)	−3.079*** (0.833)	−0.252 (0.443)
Dist 2	−3.228*** (0.577)	−0.297 (0.392)	−4.096*** (0.695)	0.281 (0.374)
Dist 3	−3.871*** (0.212)	−1.333*** (0.262)	−4.472*** (0.470)	−0.524* (0.303)
Dist 4	−5.056*** (0.197)	−1.883*** (0.257)	−5.631*** (0.462)	−0.878*** (0.296)
Dist 5	−6.026*** (0.271)	−2.994*** (0.248)	−6.537*** (0.442)	−1.970*** (0.279)
Dist 6	−5.962*** (0.130)	−3.674*** (0.223)	−6.360*** (0.404)	−2.951*** (0.232)
Shared Border	0.176 (0.472)	0.234 (0.265)	0.535 (0.465)	0.689*** (0.208)
Observations	378	378	378	378
R <sup>2</sup>	0.908	0.974	0.920	0.986
Adjusted R <sup>2</sup>	0.906	0.972	0.914	0.984

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

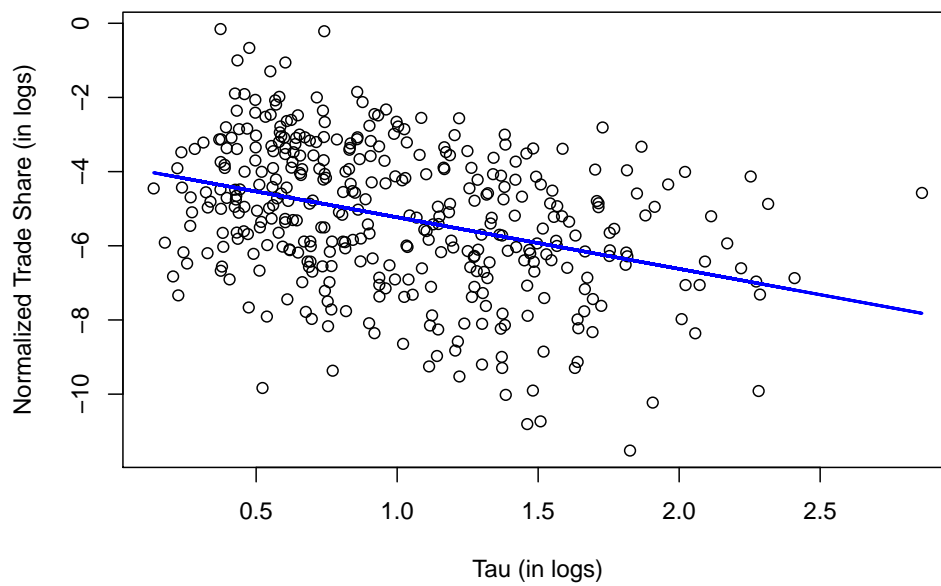


Figure 8: Correlation between Trade Shares and  $\tau$

## References

- Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5), 1741-1779.
- Waugh, M. E. (2010). International trade and income differences. *American Economic Review*, 100(5), 2093-2124.