Appendix for Online Publication

Spinning the Web: Codifiability, Information Frictions and Trade

Réka Juhász[1] and Claudia Steinwender[2]

[1]Columbia University & CEPR
[2]MIT Sloan, NBER, CEPR & CEP

October 5, 2018

Contents

A.1 Poisson Pseudo-Maximum Likelihood versus OLS estimation

Panel A in Table A.6 shows estimation of the effects of interest by OLS with country and year fixed effects. All coefficients are insignificant. There are two potential explanations for the differences in estimates between OLS and poisson pseudo maximum likelihood (PPML): 1) OLS excludes observations with zero trade, 2) the error terms are heteroskedastic. In the following, we investigate both explanations, and conclude that the more likely reason for the differences between OLS and Poisson in our case is that heteroskedastic error terms bias the OLS estimates, as is often the case with trade data (Santos Silva and Tenreyro, 2006; Head and Mayer, 2013).

A.1.1 Zero trade observations

Panel A in Table A.6 uses the log of imports as the dependent variable, which means observations with 0 trade are dropped. In our sample, 19.3% of observations are zero for yarn, 5.6% for plain cloth, and 5.7% for finished cloth. Dropping 0 trade observations biases OLS estimates towards 0. Using simulations, we found that this bias is larger, the larger the share of 0 trade observations, which could potentially explain why the effect is small and insignificant for yarn in OLS, but negative and significant in PPML.

We can use median regression to test whether 0 trade observations are driving the results. As Angrist and Pischke (2008) discuss, as long as there are less than 50% of observations with 0 trade, median regressions provide consistent estimates in the case of a censored dependent variable . Panel B of Table A.6 shows median regressions, where the log of zero observations were replaced with a large negative value (in our case, -500; but the exact value does not matter).

However, median regression results are very similar to the OLS results, i.e., there are no statistically different effects of communication time on imports of any good and the estimated magnitudes are also similar. This leads us to the conclusion that the differential share of 0 trade observations cannot explain the differences between PPML and OLS.

A.1.2 Heteroskedastic error terms

Santos Silva and Tenreyro (2006) have shown that, with heteroskedastic error terms, OLS estimation results in biased estimates (because of Jensen's inequality), and suggest poisson pseudo-maximum likelihood (PPML) as an alternative estimator. While PPML is a relatively efficient estimator, an alternative estimator correcting for heteroskedasticity is non-linear least squares (NLS). Panel C in Table A.6 implements NLS. The results are in line with the PPML results, i.e., the effect of communication time on trade is largest and significant on yarn, followed by plain cloth (though not significant in NLS; consistent with PPML being the more efficient estimator the PPML standard errors are smaller), and finished cloth has a very small and insignificant estimate. This suggests that taking into account heteroskedasticity is important when estimating the effects, in line with what has been commonly found to be the case for trade data (Santos Silva and Tenreyro, 2006; Head and Mayer, 2013).

In addition, we also follow Head and Mayer (2013) who provide guidance in order to decide between different specifications. In our case, the recommended MaMu test statistic is significantly smaller than 2 (1.718 in the pooled specification; equality of the MaMu test statistics to 2 can be rejected with a p-value<0.001), which makes PPML the preferred estimation.

Overall, we conclude that heteroskedastic error terms rather than 0 trade observations are causing bias in OLS, and therefore PPML or NLS are our preferred estimation strategy.

A.2 Variable definitions

This section defines all variables used in the empirical analysis and reports the sources.

1. **Trade data** (Primary data)
   Source: British Customs Records, Series CUST 9 (1845-1880), The National Archives (United Kingdom)
   Data on quantity and value exported for all cotton textile categories, 1845-1880. The empirical analysis uses quantity exported to each country in our dataset.

2. **Communication time and mail shipping time** (Primary data)
   Source: Lloyd's List 1845-1880 (March-May), The British Library
   Communication time defined as the minimum of the information lag to London from a given port i in year t. Mail shipping time defined as the median of the information lag to London for a given port i in year t.

3. **Year of connection to the global telegraph network**
   Source: Wenzlhuemer (2013)
   For each country c at time t, the variable takes the value 1 if the country is connected to the global telegraph network, defined as having an active connection to London. Wenzlhuemer (2013) contains only data on submarine connections, which were supplemented with data on terrestrial connections from country-specific sources.

4. **Product level tariffs**
   Source: Tena-Junguito et al. (2012)
   Ad-valorem tariffs on British imports of cotton yarn, plain cotton cloth and finished cotton cloth in country c in year t. The authors collected tariffs in place for British imports for 23 manufactured products, grouped in 11 categories, in 41 countries, colonies and dominions for selected years between 1846-1880 from British Parliamentary Papers. Specific duties were converted to ad-valorem rates by applying British export prices. These data were matched to our country categorization.

5. **GDP**
   Source: Madison Project Database (Bolt et al., 2018) and Fouquin and Hugot (2016)
   GDP in country c at time t. Data are from two sources; GDP per capita data from Bolt et al. (2018) multiplied by population data from Fouquin and Hugot (2016) yields one estimate of GDP. Fouquin and Hugot (2016) also provide GDP estimates from alternative sources. We interpolate data for missing years. Fouquin and Hugot (2016) report GDP in current pound sterling, while Bolt et al. (2018) report GDP in 1990 International dollars. We harmonize

the two using exchange rates from Fouquin and Hugot (2016) and historical US CPI from the Handbook of Labor Statistics.[1]

6. **International Merchants** (Primary data)
   Source: The Liverpool Commercial List, 1871-1872
   Number of Liverpool merchants with an affiliate in country i. The commercial list provided listings for merchants active in Liverpool including information on whether they had affiliate offices in other cities around the world. We extracted this latter information and matched cities in which British merchants had affiliates to our country categorization.

7. **Underwater elevation profile**
   Source: General Bathymetric Chart of the Oceans (GEBCO)
   Worldwide submarine elevation levels of 30-arc second grid.

8. **Population data for countries**
   Source: Fouquin and Hugot (2016) and Pascali (2017)
   Population of country c at time t. The baseline data are from Fouquin and Hugot (2016) and have been supplemented using Pascali (2017) for Persia. Data on spinning capacity and imported spinning machinery are normalized by population in 1845 (which is the start of the sample period). GDP data constructed for some countries using GDP per capita data from the Madison project database multiplied by population data from Fouquin and Hugot (2016).

[1] Data source for US CPI: https://www.minneapolisfed.org/community/financial-and-economic-education/cpi-calculator-information/consumer-price-index-1800.

A.3 Data construction

A.3.1 Trade data

Bilateral trade data for cotton textiles between Great Britain and other countries worldwide were collected from handwritten British Customs records for the years 1845-1880.[2] The data distinguish between product categories for cotton yarn, plain cotton cloth and finished cotton cloth. The main challenge in constructing these data was the harmonization of countries across time as reporting units changed (in initial years some countries were reported at more aggregate levels) and there were border changes (for example, the "scramble for Africa" led to numerous border changes). The regions thus constructed were kept at the lowest possible level of disaggregation.[3] Using this methodology yields a strongly balanced sample of 75 countries.

A.3.2 Communication and mail shipping time data

This section provides additional details on data construction for communication and mail shipping time. As discussed in the main text, data on shipping time and communication time relative to the UK were collected from a shipping newspaper, the Lloyd's List. This London-based publication printed what was considered to be the most up to date shipping information for ports worldwide every day of the week except Sunday (Wright and Fayle, 1928). Each day, Lloyd's published new information it had received about ship movements from ports around the world. Crucially for our purposes, Lloyd's used the telegraph to communicate with ports that were connected to the network (Huurdeman, 2003). Shipping time should be understood as the fastest mode of transportation (not restricted to sea-shipping) along a given route.[4]

Figure A.1 gives an example of the information published for a typical observation for the port of New York both prior to and after the connection to London was established in 1866. Panel A shows an excerpt for the May 5, 1865 edition. On this day, the Lloyd's List printed ship movements at the port of New York for the dates April 19-22, 1865. This meant that the latest information in London from New York was 14 days old. Panel B shows an excerpt of the edition for May 1, 1868. The latest ship movements from the port of New York date to April 30, 1865, implying an information lag that was 2 days. Notice that this data is consistent with the information lags between New York and Britain that Steinwender (2018) constructs using different publications as sources.

[2] One feature of these data should be noted; the unit of observation was given in the trade returns by country of shipment. This means that landlocked countries do not show up in our data by definition and exports destined for these countries will be included in exports to the country through which shipments entered (Mitchell, 1988).

[3] The one exception to this methodology was the case of trading ports such as Aden, Gibraltar and Hong-Kong. These were merged with the surrounding countries for the reason that much of the trade passing through these ports was intended for the hinterland.

[4] Lloyd's of London's business model was based on having the most up to date information. Information on ship movements for Lloyd's thus took routes and modes of transportation that would be too expensive for merchandise trade. An example of this is the overland route across Egypt prior to the building of the Suez canal.

To understand why we use the median and the minimum of the information lag to measure communication time and mail shipping time respectively, it is instructive to examine Figure A.2. This shows the histograms of the information lag with New York before (1865) and after (1867) the telegraph was adopted. While shipping was the only method of communication with London, the distribution is single-peaked. However, once the telegraph is adopted and both methods of communication (shipping and telegraphy) are in use, the distribution becomes bi-modal, with much more mass at observations within a few days information lag of London.[5]

Our data extraction method was partially automated using OCR technology and text-matching tools. However, as we needed to extract dates for each port, our measure of shipping, and in particular communication time was sensitive to even small mistakes in text-matching. For example, mistaking the month of March for May would lead to a 60 day error for a given observation. For this reason, we manually verified all influential outliers. We matched ports to the countries and regions in our data based on historical maps. For each country, we have used the port with the largest number of observations in our sample.[6]

[5]We use the minimum of the information lag for each year to capture communication time as for ports with a relatively low number of observations, it is often very difficult to pick up an observations transmitted by the telegraph.

[6]A small number of countries have a very small number of observations per year. For these countries, we have pooled all observations across all ports of that country together in order to increase sample size.

A.4 Additional tables

Table A.1: Poisson specifications with different FE

VARIABLES	(1) Yarn	(2) Plain cloth	(3) Finished cloth
Panel A. No fixed effects			
ln(comm time)	-0.785***	-0.376***	-0.295***
	(0.141)	(0.061)	(0.073)
Panel B. Year fixed effects			
ln(comm time)	-0.910***	-0.364***	-0.303***
	(0.177)	(0.074)	(0.089)
Panel C. Year and country fixed effects			
ln(comm time)	-0.183***	-0.097***	0.006
	(0.041)	(0.029)	(0.044)
Observations	2,150	2,150	2,150
Nr of countries	72	72	72

Notes: ln(commtime) defined as the natural logarithm of communication time (in days) to London. Standard errors clustered by country in parentheses. Notation for statistical significance; *** p<0.01, ** p<0.05, * p<0.1.

Table A.2: Pooled specifications

Dep var: imports Estimation:	(1) Poisson	(2) IV1: Riley	(3) IV2: Norm cable
ln(comm time)*(Yarn dummy)	-0.183***	-0.272***	-0.269***
	(0.041)	(0.076)	(0.077)
ln(comm time)*(Plain cloth dummy)	-0.097***	-0.128**	-0.110*
	(0.029)	(0.065)	(0.065)
ln(comm time)*(Finished cloth dummy)	0.006	-0.030	0.007
	(0.044)	(0.066)	(0.064)
Observations	6,450	6,450	6,450
P-values from tests on equality of coefficients:			
Yarn = Plain: p-value	0.067	0.013	0.003
Plain = Finished: p-value	0.024	0.319	0.164
Yarn = Finished: p-value	0	0.036	0.014

Notes: Regressors are ln(commtime) defined as the natural logarithm of communication time (in days) to London interacted with a binary variable that takes the value of one if the product traded is yarn, plain or finished cotton cloth, respectively. Column (1) presents the baseline Poisson specification. Column (2) presents the IV-Poisson specification using the Riley ruggedness instrument. Column (3) presents the IV-Poisson specification using the normalized cable length instrument. Industry*country and industry*year FEs included. Standard errors clustered by country in parentheses. Notation for statistical significance; *** p<0.01, ** p<0.05, * p<0.1.

Table A.3: Robustness: Cable Length Measure

VARIABLES	(1) Yarn	(2) Plain cloth	(3) Finished cloth
Panel A. Baseline IV (Cable length measure)			
ln(comm time)	-0.269***	-0.110*	0.007
	(0.077)	(0.065)	(0.064)
Observations	2,150	2,150	2,150
Nr of countries	72	72	72
Panel B. Drop Civil War			
ln(comm time)	-0.266***	-0.109	-0.004
	(0.077)	(0.068)	(0.071)
Observations	1,845	1,845	1,845
Nr of countries	72	72	72
Panel C. Drop British colonies			
ln(comm time)	-0.416***	-0.185**	0.073
	(0.131)	(0.084)	(0.093)
Observations	1,632	1,632	1,632
Nr of countries	55	55	55
Panel D. Control for tariffs			
ln(comm time)	-0.277***	-0.069	0.002
	(0.070)	(0.056)	(0.054)
tariff rate	-0.428	0.020	-0.019***
	(0.276)	(0.289)	(0.002)
Observations	1,096	1,096	1,096
Nr of countries	36	36	36
Panel E. Control for GDP			
ln(comm time)	-0.429**	-0.031	0.142*
	(0.198)	(0.087)	(0.072)
ln(GDP)	0.083	-0.755***	-0.757***
	(0.423)	(0.158)	(0.259)
Observations	920	920	920
Nr of countries	38	38	38
Panel F. Import values			
ln(comm time)	-0.134	-0.059	0.052
	(0.105)	(0.067)	(0.067)
Observations	2,150	2,150	2,150
Nr of countries	72	72	72

Notes: ln(commtime) defined as the natural logarithm of communication time (in days) to London. The instrument used across all specifications is a binary variable that takes the value of one including and after the predicted year of connection based on the normalized cable length measure interacted with the closest neighbor's drop in communication time after connection to the telegraph. Controls: ad-valorem product specific tariffs from Tena-Junguito et al. (2012), current price annual (log) GDP values from Bolt et al. (2018) and Hugot and Dajud (2016). Appendix A.2 contains a detailed discussion of the construction of each variable. Year and country FEs included. Standard errors clustered by country in parentheses. Notation for statistical significance; *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: Mechanism: Cable Length Measure

VARIABLES	(1) Yarn	(2) Plain cloth	(3) Finished cloth
Panel A. Interaction with international merchants			
ln(comm time)	-0.333***	-0.203***	-0.007
	(0.102)	(0.066)	(0.081)
ln(comm time)*	-0.002	0.026**	0.020**
number int merchants	(0.022)	(0.010)	(0.010)
Observations	2,019	2,019	2,019
Nr of countries	68	68	68
Panel B. Control for mail shipping time			
ln(comm time)	-0.303***	-0.108	0.023
	(0.095)	(0.067)	(0.067)
ln(mail ship time)	0.196	-0.084	-0.237
	(0.154)	(0.122)	(0.144)
Observations	2,150	2,150	2,150
Nr of countries	72	72	72

Notes: ln(commtime) defined as the natural logarithm of communication time (in days) to London. The instrument used across all specifications is a binary variable that takes the value of one including and after the predicted year of connection based on the normalized cable length measure interacted with the closest neighbor's drop in communication time after connection to the telegraph. Controls: international merchants defined as the number of British merchant houses that have an affiliate merchant house in the destination market, mail shipping time defined as the natural logarithm of mail shipping time (in days) to London. Appendix A.2 contains a detailed discussion of the construction of each variable. Year and country FEs included. Standard errors clustered by country in parentheses. Notation for statistical significance; *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: Robustness: Panel jackknife bias correction

VARIABLES	(1) Yarn	(2) Plain cloth	(3) Finished cloth
Panel A. IV: Riley measure			
ln(comm time)	-0.272***	-0.128**	-0.030
	(0.076)	(0.065)	(0.066)
Observations	2,150	2,150	2,150
Nr of countries	72	72	72
Panel B. Panel jackknife bias corr.: Riley measure			
ln(comm time)	-0.277***	-0.098	-0.035
	(0.076)	(0.065)	(0.066)
Observations	2,150	2,150	2,150
Nr of countries	72	72	72
Panel C. IV: Normalized cable length			
ln(comm time)	-0.269***	-0.110*	0.007
	(0.077)	(0.065)	(0.064)
Observations	2,150	2,150	2,150
Nr of countries	72	72	72
Panel D. Panel jackknife bias corr.: Norm. cable length			
ln(comm time)	-0.271***	-0.084	-0.000
	(0.077)	(0.065)	(0.064)
Observations	2,150	2,150	2,150
Nr of countries	72	72	72

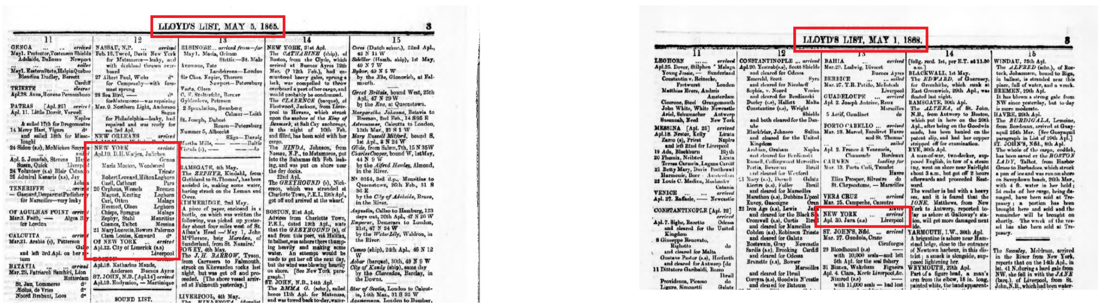
Notes: Panels A and C report the baseline IV-Poisson estimates for the Riley and normalized cable length instruments, respectively. Panels B and D report the panel jackknife bias correction for each specification. Jackknife bias correction method used is the split panel jackknife applied to the two dimensions of the panel. The SEs are derived from the corresponding baseline IV estimation. Cruz-Gonzalez et al. (2017) contains a further discussion of the correction. Year and country FEs included. Standard errors clustered by country in parentheses. Notation for statistical significance; *** p<0.01, ** p<0.05, * p<0.1.

Table A.6: Alternative specifications

VARIABLES:	(1) Yarn ln(imports)	(2) Plain cloth ln(imports)	(3) Finished cloth ln(imports)
Panel A. OLS			
ln(comm time)	0.005	-0.080	-0.070
	(0.091)	(0.052)	(0.045)
Observations	1,736	2,029	2,027
Nr of countries	71	72	72
Panel B. Median regression			
ln(comm time)	0.002	-0.042	-0.053
	(0.072)	(0.033)	(0.038)
Observations	2,150	2,150	2,150
Nr of countries	72	72	72
Panel C. Non-linear least squares estimation			
ln(comm time)	-0.303***	-0.078	-0.006
	(0.053)	(0.041)	(0.054)
Observations	2,150	2,150	2,150
Nr of countries	72	72	72

Notes: Dependent variable is ln(imports, in quantity). ln(commtime) defined as the natural logarithm of communication time (in days) to London. All regressions include country and year fixed effects. Observations with 0 trade are dropped in panel A, resulting in a smaller number of observations (in the case of yarn, there is even an entire country dropped which never imports yarn). Standard errors clustered by country in parentheses. Notation for statistical significance; *** p<0.01, ** p<0.05, * p<0.1.

A.5 Additional figures



(a) May 1865, Pre-telegraph

(b) May 1868, Telegraph

Figure A.1: Screen shots of the entries for New York in Lloyd's List issues for 1865 and 1868

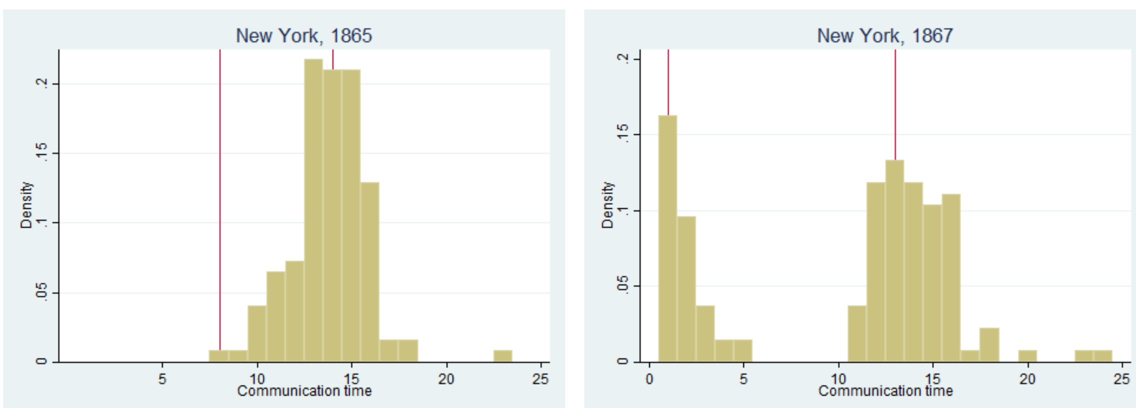


Figure A.2: Histogram of information lags observed for New York in 1865 and 1857

References

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Bolt, J., R. Inklaar, H. de Jong, and J. L. van Zanden (2018). Rebasing 'Maddison': new income comparisons and the shape of long-run economic development. *Madison Project Working Paper 10*.

Cruz-Gonzalez, M., I. Fernandez-Val, and M. Weidner (2017). Bias corrections for probit and logit models with two-way fixed effects. *The Stata journal 17*(3), 517–545.

Fouquin, M. and J. Hugot (2016). Two centuries of bilateral trade and gravity data: 1827-2014.

Head, K. and T. Mayer (2013). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4. Elsevier.

Hugot, J. and C. U. Dajud (2016). Trade costs and the Suez and Panama Canals. Technical report, CEPII.

Huurdeman, A. A. (2003). *The worldwide history of telecommunications.* John Wiley Sons.

Mitchell, B. R. (1988). *British Historical Statistics.* Cambridge University Press.

Pascali, L. (2017). The wind of change: Maritime technology, trade and economic development. *The American Economic Review 107*(9), 2821–54.

Santos Silva, J. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and Statistics 88*(4), 641–658.

Steinwender, C. (2018). Information frictions and the law of one price: "when the states and the kingdom became united". *American Economic Review 108*(3), 657–696.

Tena-Junguito, A., M. Lampe, and F. T. Fernandes (2012). How much trade liberalization was there in the world before and after Cobden-Chevalier? *The Journal of Economic History 72*(3), 708–740.

Wenzlhuemer, R. (2013). *Connecting the nineteenth-century world: The telegraph and globalization.* Cambridge University Press.

Wright, C. and C. E. Fayle (1928). *A History of Lloyd's from the Founding of Lloyd's Coffee House to the Present Day.* Macmillan and Company.