The Margins of Trade

Appendices

Jonathan Eaton and Ana Cecília Fieler*

July 22, 2019

Contents

A Appendix: Data Description

This appendix describes the sources of our data and our procedures for cleaning them. Data on GDP and population are from the World Development Indicators. Data on geographical characteristics are from CEPII. Our trade data are from the United Nations COMTRADE database. We restrict our analysis to merchandise trade among the fifty largest countries in terms of GDP in the 2007 cross section. We focus on this subsample to avoid zero bilateral trade flows and to ensure sufficient overlap in HS6 products across country pairs. We combine China with Hong Kong, Malaysia with Singapore, and Belgium with Luxembourg. The list of countries appears in Table A.1. For each country, the table reports GDP, GDP per capita, and population. Even though we selected countries on the basis of total GDP, our sample includes poor countries such as Pakistan, India, and Nigeria.

*Eaton: The Pennsylvania State University and NBER. Fieler: Yale University and NBER.

Table A.1: Summary of Countries in Sample

	country	GDP (US$ MI)	GDP per capita	population (000)
1	United States	13,751,400	45,642	301,290
2	Japan	4,384,250	34,313	127,771
3	China, Hong-Kong	3,589,339	2,708	1,325,237
4	Germany	3,317,370	40,324	82,268
5	United Kingdom	2,772,030	45,442	61,001
6	France	2,589,840	41,970	61,707
7	Italy	2,101,640	35,396	59,375
8	Spain	1,436,890	32,017	44,879
9	Brazil	1,333,270	7,013	190,120
10	Canada	1,329,880	40,329	32,976
11	Russian Federation	1,290,080	9,079	142,100
12	India	1,176,890	1,046	1,124,787
13	Korea, Rep.	1,049,240	21,653	48,456
14	Mexico	1,022,820	9,715	105,281
15	Australia	820,974	39,066	21,015
16	Netherlands	765,818	46,750	16,381
17	Turkey	655,881	8,984	73,004
18	Belgium, Luxembourg	502,213	45,221	11,106
19	Sweden	454,310	49,662	9,148
20	Indonesia	431,933	1,914	225,630
21	Poland	424,790	11,143	38,121
22	Switzerland	424,367	56,207	7,550
23	Norway	388,412	82,480	4,709
24	Saudi Arabia	383,587	15,879	24,157
25	Austria	373,192	44,880	8,315
26	Malaysia, Singapore	353,669	11,358	31,138
27	Greece	313,354	27,995	11,193
28	Denmark	311,579	57,051	5,461
29	Iran, Islamic Rep.	286,058	4,028	71,021
30	South Africa	283,743	5,930	47,851
31	Argentina	262,451	6,644	39,503
32	Ireland	259,018	59,324	4,366
33	Finland	244,661	46,261	5,289
34	Thailand	236,615	3,533	66,979
35	Venezuela, RB	228,071	8,299	27,483
36	Portugal	222,758	20,998	10,608
37	Colombia	207,786	4,724	43,987
38	Czech Republic	173,958	16,833	10,334
39	Romania	165,976	7,703	21,547
40	Nigeria	165,921	1,121	147,983
41	Israel	163,957	22,835	7,180
42	Chile	163,878	9,875	16,595
43	Philippines	144,043	1,624	88,718
44	Pakistan	142,893	879	162,481
45	Ukraine	142,719	3,069	46,509
46	Hungary	138,757	13,799	10,056
47	New Zealand	135,667	32,086	4,228
48	Algeria	134,304	3,967	33,853
49	Egypt	130,476	1,630	80,061
50	Peru	107,291	3,763	28,508

In extracting values and quantities in bilateral trade flows from COMTRADE we proceed as follows. First, COMTRADE reports each bilateral trade flow twice: Once as reported by the exporter and once as reported by the importer. We consider only the importer reports, which, according to the UN COMTRADE website, are typically c.i.f. An exception is Iran, which does not report trade flows to the United Nations. We construct Iranian imports using other countries' reports of their exports to Iran. Of the possible 2450 importer-exporter pairs, COMTRADE reports no trade from Venezuela to Iran nor from Algeria to Israel. Second, COMTRADE reports total merchandise trade (HS0 level) as well as trade at the level of six-digit HS codes (HS6). For all importer-exporter pairs, the sum of the values of trade flows across all HS6 codes never exceeds the reported aggregate value. But for 240 importer-exporter pairs, the aggregate trade flow exceeds the sum of HS6 codes. Since the difference is at times significant, we use aggregate reports in all regressions involving aggregate trade flows.[1]

We now turn to the product dimension of the data. Among our original importer-exporter pairs (before merging the three country-pairs reported above), COMTRADE reports values for 3,239,484 importer-exporter-HS6 triads. Among these, 3,192,745 triads also have quantity data. Table A.2 summarizes our steps in cleaning these data and the remaining number of observations at each step. For some HS6 codes, different importer-exporter pairs report their quantity data using different units of measurement, e.g., number of pairs, length, and weight. We select the unit of measurement for each HS6 code that delivers the largest number of observations. Imposing this consistency decreases the number of quantity observations to 3,030,357. Merging Hong Kong with China, Singapore with Malaysia, and Luxembourg with Belgium reduces the number of observations with value to 3,008,407. The corresponding value of trade decreases much less as we lose trade only within these pairs. The number of observations with quantity data drops more because, for each product category, we lose the quantity observation whenever one of the two merging countries has missing quantity information. For each of the remaining importer-exporter-HS6 triads, we use the observations on value and quantity to calculate unit value. We drop observations for which the unit value is more than 20 times as large or less than 1/20 as large as the world average unit value for that HS6 code. Our resulting data set has 3,008,407 triads with observations on value (representing total trade in the amount of US$11.1 trillion in 4,975 HS6 codes), of which 2,611,700 triads also have observations on quantity (representing total trade in the amount of US$9.6 trillion

[1] The difference between HS0 and HS6 for the 240 country pairs is on average 5 percent. COMTRADE also reports trade flows at the 4-digit level. The sum of these flows is identical to the sum of 6-digit flows for all country pairs.

Table A.2: Data compilation and cleaning

	imp-exp-HS6 triads with quantity data		imp-exp-HS6 triads with value data	
	number of observations	total value (US$ trillion)	number of observations	total value (US$ trillion)
1. downloaded data	3,192,745	11.08	3,239,484	11.14
2. after selecting one unit for each HS6 codes	3,030,357	10.23	3,239,484	11.14
3. merging of CHN-HKG, MYS-SGP, BEL-LUX	2,804,336	9.98	3,008,407	11.12
4. after dropping quantity data with large price deviation	2,611,700	9.62	3,008,407	11.12

Table A.3: Units of measurement for observations with quantity data

	number of observations	total value (US$ trillion)
kilograms	2,095,041	6.96
number of items or number of pairs	469,344	2.44
other	47,315	0.22
total	**2,611,700**	**9.62**

in 4,945 HS6 codes).

Table A.3 summarizes the units for measuring quantity in our data set. About 80% of quantity observations are measured in weight, and most of the remaining observations are measured in number of items.[2] In value, trade flows whose quantities are measured in weight account for 72 percent (6.96/9.62) of the data.

B Appendix: Hummels and Klenow (2005) Decompositions

Table B.1 adds standard errors to the Hummels and Klenow decompositions presented in Table 1 of the main text. These standard errors are calculated clustering by importer and by exporter. All coefficients are significantly different from zero except for the coefficients on GDP and population for the price margin in the last column. Panel C shows the t-statistic testing for equality of the coefficients on GDP per capita and on population, first for exporter and then for importer, in the set of regressions in panel B. For values and extensive margins, we can't reject equality at any reasonable significance level. We reject equality at the 5% significance level for prices and quantities in the data and for prices in the model.

Our baseline decompositions of trade flows into margins in Table B.1 define the extensive margin EM_{ni} as simply the number of product categories $|K_{ni}|$ that importer n sources from exporter i, where K_{ni} is the set of these products. The price margin P_{ni} is the simple weighted average across products of the log-deviation of prices with respect to world prices:

$$\log P_{ni} = \frac{1}{|K_{ni}|} \sum_{k \in K_{ni}} [\log(p_{nik}) - \log(p_{\text{world},k})]$$

where p_{nik} is the unit value of product k exported to n from i and $p_{\text{world},k}$ is the average

[2]Products measured in units include those measured in "Number of items," "Number of pairs," or "Thousands of items" in the COMTRADE nomenclature.

Table B.1: Decomposition of trade flows with standard errors

	Data				Model			
dependent variable →	value	extensive margin	quantity	price	value	extensive margin	quantity	price
Panel A								
exporter GDP	1.16	0.76	0.36	0.04	1.16	0.67	0.45	0.04
	(0.10)	(0.10)	(0.08)	(0.03)	(0.10)	(0.08)	(0.11)	(0.03)
importer GDP	1.11	0.34	0.73	0.05	1.07	0.28	0.73	0.06
	(0.06)	(0.04)	(0.06)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)
distance	-0.81	-0.43	-0.39	0.02	-0.80	-0.42	-0.35	-0.03
	(0.08)	(0.06)	(0.06)	(0.02)	(0.06)	(0.05)	(0.06)	(0.02)
number of observations	2448	2448	2448	2448	2448	2448	2448	2448
Panel B								
exporter GDP per capita	1.18	0.84	0.19	0.15	1.18	0.70	0.31	0.17
	(0.12)	(0.13)	(0.08)	(0.02)	(0.11)	(0.09)	(0.11)	(0.01)
exporter population	1.15	0.71	0.46	-0.02	1.15	0.65	0.53	-0.03
	(0.11)	(0.11)	(0.08)	(0.02)	(0.11)	(0.09)	(0.11)	(0.00)
importer GDP per capita	1.10	0.41	0.56	0.13	1.04	0.27	0.64	0.13
	(0.06)	(0.05)	(0.05)	(0.02)	(0.04)	(0.03)	(0.03)	(0.01)
importer population	1.12	0.30	0.82	0.00	1.08	0.28	0.77	0.03
	(0.07)	(0.04)	(0.05)	(0.02)	(0.03)	(0.03)	(0.02)	(0.01)
distance	-0.81	-0.37	-0.53	0.10	-0.80	-0.42	-0.43	0.05
	(0.08)	(0.06)	(0.06)	(0.01)	(0.06)	(0.05)	(0.06)	(0.01)
number of observations	2448	2448	2448	2448	2448	2448	2448	2448
Panel C: T-statistic for the hypothesis that the coefficient on GDP per capita equals the coefficient on population.								
exporter	0.15	0.60	-1.98	4.75	0.15	0.31	-1.09	22.9
importer	-0.17	1.35	-2.87	3.58	-0.62	-0.16	-2.66	5.69

All variables are in logs. Standard errors in parenthesis are clustered by importer and by exporter.

unit value of product k across all importer-exporter pairs in our sample. The quantity margin is defined as the residual $X_{ni}/(EM_{ni}P_{ni})$ where X_{ni} is the value of trade flow to importer n from exporter i.

In contrast to our measures of the extensive and price margins, Hummels and Klenow (2005) use weighted definitions of margins in their decomposition of total trade flows. Their weighted extensive margin is

$$EM_{ni}^{W} = \frac{\sum_{k \in K_{ni}} x_{\text{world},k}}{\sum_{k \in K} x_{\text{world},k}}$$

where K is the total set of product categories in the data, and $x_{\text{world},k}$ is the value of world trade in product k. That is, each product category is weighted according to its representation in world trade flows. The price margin is

$$P_{ni}^{W} = \prod_{k \in K_{ni}} \left(\frac{p_{nik}}{p_{\text{world},k}} \right)^{v_{nik}}$$

where, following Sato (1976) and Vartia (1976), the weights are:

$$v_{nik} = \frac{(s_{nik} - s_{\text{world},nik})/(\log s_{nik} - \log s_{\text{world},nik})}{\sum_{k' \in K_{ni}} (s_{nik'} - s_{\text{world},nik'})/(\log s_{nik'} - \log s_{\text{world},nik'})}$$

where

$$s_{nik} = \frac{x_{nik}}{\sum_{k' \in K_{ni}} x_{nik'}}$$
$$s_{\text{world},nik} = \frac{x_{\text{world},k}}{\sum_{k' \in K_{ni}} x_{\text{world},k'}}$$

and x_{nik} is the value of trade flow that importer n sources from country i in product k. Feenstra (1994) shows that P_{ni}^{W} is the ideal price index under CES preferences when goods are homogeneous within product categories. In our model goods prices are heterogeneous with a distribution governed by the term Φ, so that that P_{ni}^{W} is no longer ideal.

Table B.2 repeats the results in B.1 applying HK's weighted definitions of margins to our data. The first column is of course identical in the two tables. Compared with our simpler measure, the weighted extensive margin is smaller. The reason is that the products that are traded by more country pairs receive more weight. Hence there is less variation in the extensive margin in the weighted than in the unweighted measures. Panel C shows that weighting doesn't affect the outcome of the tests for equality of coefficients on per capita GDP and population either for importers or for exporters.

Table B.2: Decomposition of trade flows with weighted definitions of margins

	Data				**Model**			
		extensive			extensive			
dependent variable →	value	margin	quantity	price	value	margin	quantity	price
Panel A								
exporter GDP	1.16	0.45	0.69	0.01	1.16	0.45	0.66	0.05
	(0.10)	(0.06)	(0.08)	(0.02)	(0.10)	(0.07)	(0.07)	(0.03)
importer GDP	1.11	0.25	0.85	0.02	1.07	0.20	0.81	0.06
	(0.06)	(0.03)	(0.05)	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)
distance	-0.81	-0.27	-0.54	0.00	-0.80	-0.27	-0.49	-0.03
	(0.08)	(0.04)	(0.06)	(0.01)	(0.06)	(0.04)	(0.05)	(0.02)
number of observations	2448	2448	2448	2448	2448	2448	2448	2448
Panel B								
exporter GDP per capita	1.18	0.54	0.56	0.08	1.18	0.48	0.52	0.18
	(0.12)	(0.08)	(0.07)	(0.01)	(0.11)	(0.08)	(0.05)	(0.00)
exporter population	1.15	0.40	0.77	-0.03	1.15	0.44	0.74	-0.02
	(0.11)	(0.07)	(0.08)	(0.02)	(0.11)	(0.08)	(0.07)	(0.00)
importer GDP per capita	1.10	0.31	0.72	0.07	1.04	0.19	0.73	0.12
	(0.06)	(0.04)	(0.04)	(0.01)	(0.04)	(0.03)	(0.03)	(0.01)
importer population	1.12	0.21	0.92	-0.01	1.08	0.20	0.86	0.02
	(0.07)	(0.03)	(0.05)	(0.01)	(0.03)	(0.03)	(0.03)	(0.01)
distance	-0.81	-0.21	-0.65	0.05	-0.80	-0.27	-0.58	0.04
	(0.08)	(0.04)	(0.07)	(0.01)	(0.06)	(0.05)	(0.06)	(0.01)
number of observations	2448	2448	2448	2448	2448	2448	2448	2448
Panel C: T-statistic for the hypothesis that the coefficient on GDP per capita equals the coefficient on population.								
exporter	0.15	1.07	-1.57	3.64	0.15	0.28	-2.05	27.8
importer	-0.17	1.43	-2.62	4.31	-0.62	-0.20	-2.60	5.87

All variables are in logs. Standard errors in parenthesis are clustered by importer and by exporter.

C Appendix: Attenuation Bias

This appendix estimates attenuation bias in the coefficients of Φ_n and Φ_i in the price regression (37) in the text. These coefficients are potentially biased toward zero because we estimate Φ in a previous stage using the gravity equation. Hence Φ has potential measurement error. If δ is the true coefficient on Φ the probability limit of the estimated coefficient $\hat{\delta}$ is

$$\text{plim}\,\hat{\delta} = \delta \frac{\sigma_\Phi^2}{\sigma_\Phi^2 + \sigma_e^2}$$

where σ_Φ^2 is the variance of Φ and σ_e^2 is the variance of the estimate of Φ.

To adjust for attenuation, we calculate

$$\sigma_\Phi^2 = N^{-1} \sum_{n=1}^{N} (\Phi_n - \overline{\Phi})^2$$

where $\overline{\Phi}$ is the average Φ across countries.

To calculate σ_e^2 we estimate the variance of Φ_n for each country n as follows. We start with the gravity equation (35) in the text

$$\log\left(\frac{\pi_{ni}}{\pi_{nn}}\right) = A_n + B_i + \delta^g \log dist_{ni} + \varepsilon_{ni}^X \tag{C.1}$$

where A_n is an importer fixed effect, B_i is an exporter fixed effect, δ^g is a parameter, and $dist_{ni}$ is the distance between the most populous cities in importer n and in exporter i. We order countries by total GDP and normalize the USA's importer fixed effect $A_1 = 0$.

Our estimate of Φ_n is

$$\hat{\Phi}_n = \exp(-\hat{A}_n) + \sum_{i \neq n} \exp(\hat{\delta}^g \log dist_{ni} + \hat{B}_i). \tag{C.2}$$

To calculate the variance of $\hat{\Phi}_n$, we define

$$\mathbf{x} = \{\hat{A}_2, \hat{A}_3, ..., \hat{A}_N, \hat{B}_1, \hat{B}_2, ..., \hat{B}_N, \hat{\delta}^g\},$$

the vector of coefficients of the gravity equation (C.1), and define $\Sigma_\mathbf{x}$ as the variance-covariance matrix of $\mathbf{x}$. Equation (C.2) defines $\hat{\Phi}_n$ as a function of $\mathbf{x}$. We use the first-order approximation:

$$\text{Var}(\hat{\Phi}_n(\mathbf{x})) = \nabla\hat{\Phi}_n(\mathbf{x})\Sigma_\mathbf{x}\nabla\hat{\Phi}_n(\mathbf{x})^T$$

where $\nabla\hat{\Phi}_n(\mathbf{x})$ is the gradient of $\hat{\Phi}_n$ evaluated at $\mathbf{x}$. The elements of $\nabla\hat{\Phi}_n(\mathbf{x})$ are:

$$\frac{\partial\hat{\Phi}_n}{\partial\hat{A}_i} = \begin{cases} -\exp(-\hat{A}_n) & \text{if } n = i \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial\hat{\Phi}_n}{\partial\hat{B}_i} = \begin{cases} 0 & \text{if } n = i \\ \exp(\hat{\delta}^g \log dist_{ni} + \hat{B}_i) & \text{otherwise} \end{cases}$$

$$\frac{\partial\hat{\Phi}_n}{\partial\hat{\delta}^g} = \sum_{i\neq n} \log(dist_{ni}) \exp(\hat{\delta}^g \log dist_{ni} + \hat{B}_i).$$

We take σ_e^2 as the simple mean

$$\sigma_e^2 = N^{-1} \sum_{n=1}^{N} \mathrm{Var}(\hat{\Phi}_n(\mathbf{x})).$$

D Appendix: Monte Carlo Simulations

A leap in our analysis has been between varieties in our model and products in the COMTRADE data. In particular, in Sections 4.2 and 4.3 we derived specifications for unit values and demand at the variety level but estimated at the product level. A question is whether we can identify parameters that apply at the variety level in our model with product-level data when products are groupings of varieties.

To help answer this question we repeat the estimation exercises in the main text using our model-generated data to check if we can recover the original parameters used in the simulations. As discussed in the text, we simulate the model to generate a data set with prices, quantities and value for 3,807 product categories with 2,319,837 importer-exporter-product tuples.[3] Using these simulated data we re-estimate the price regressions of Section 4.2 (Table 4) to get parameters ν, γ, and θ and we re-estimate the demand equation of Section 4.3 to get the parameter β.

Price regression. From Section 4.2 the price regression is

$$\log p_{nik} = \delta_k + \delta_{w,M} \log w_n + \delta_{\Phi,M} \log \Phi_n + \delta_{w,X} \log w_i + \delta_{\Phi,X} \log \Phi_i + \varepsilon^P_{nik}. \tag{D.1}$$

[3] The mean number of observations (importer-exporter pairs) per product is 609 in the simulated data and 605 in the original COMTRADE data.

where δ_k are product fixed effects and

$$\begin{aligned}
\delta_{w,M} &= \frac{\gamma}{1+\gamma}, \\
\delta_{\Phi,M} &= -\frac{1}{\theta(1+\gamma)}, \\
\delta_{w,X} &= \frac{\nu}{1+\gamma}, \\
\delta_{\Phi,X} &= \frac{\nu}{\theta(1+\gamma)}.
\end{aligned} \tag{D.2}$$

Table D.1 shows the results of the price regressions using simulated data. Column (1) reports the estimates of the regression (D.1). In column (2), we regress prices on product fixed effects and on importer-exporter fixed effects. The dependent variables in columns (3) and (4) are the importer-exporter fixed effects from column (2). Column (3) presents the unrestricted regression, and column (4) imposes the restrictions in (D.2). In all specifications, columns (1), (3) and (4), the estimates of γ and ν are very close to the original parameters $\gamma = 0.13$ and $\nu = 0.215$ used in the simulations. The point estimates of θ vary more, but they are also less precisely estimated. When we impose (D.2) in column (4), the estimate of $\theta = 4.09$ with standard error 0.09. Hence we cannot reject equality with the original parameter $\theta = 4$.

Demand regression. From Section 4.3 the demand regression is:

$$\log x_{nik} = \delta_n + \delta_i + \delta_k - \frac{\beta}{1-\beta} \log p_{nik} + \varepsilon^X_{nik} \tag{D.3}$$

where δ_n, δ_i, and δ_k are, respectively, importer, exporter, and product fixed effects, and ε^X_{nik} is a residual. The independent variable p_{nik} is the unit value of country n's imports from country i of product k and x_{nik} is the value. We instrument price p_{nik} with the average unit value at which exporter i sells product k to destinations other than n. We present results from both OLS and IV specifications using simulated data, but the IV is not necessary in the simulated data. With perfect competition, prices in the model are determined by technologies and economy-wide input costs, not by variety-specific demand.

Table D.2 reports the results. Columns (1) and (2) report the results in the original data set for reference. Columns (3) and (4) use simulated data. The coefficient on price is -1.61 (standard error 0.05) in the OLS specification of column (3), and it's -1.27 (standard error 0.05), both smaller in absolute magnitude than the coefficient -1.83 in the data. We conjecture that these simulation-based estimates are biased toward zero because product fixed effects in regression (D.3) absorb some of the relevant variation in prices in the model

Table D.1: Results from Price Regression Using Simulated Data

dependent variable → independent variable ↓	original	p_{nik} (1)	p_{nik} (2)	importer-exporter fixed effect from specification (2) (3)	same as specification (3) restricting $\theta_n = \theta_i$ (4)
importer per capita income (w_n)		0.116		0.115	0.116
		(0.002)		(0.002)	(0.002)
$\hat{\Phi}_n$		-0.215		-0.213	-0.216
		(0.005)		(0.005)	(0.005)
exporter per capita income (w_i)		0.181		0.180	0.183
		(0.004)		(0.005)	(0.004)
$\hat{\Phi}_i$		0.061		0.062	0.045
		(0.008)		(0.008)	(0.001)
importer-exporter fixed effect		no	yes	no	no
product category fixed effect		yes	yes	no	no
R-squared		0.60	0.61	0.98	0.97
number of observations		2,319,837	2,319,837	2,448	2,448
parameters implied by importer coefficients					
γ	0.13	0.13		0.130	0.132
		(0.002)		(0.002)	(0.002)
θ_n	4	4.112		4.15	4.09
		(0.09)		(0.1)	(0.09)
parameters implied by exporter coefficients					
ν	0.215	0.20		0.203	0.207
		(0.005)		(0.005)	(0.004)
θ_i	4	3.0		2.90	4.09
		(0.4)		(0.4)	(0.09)

since there are typically few exporters per product. When an exporter is very efficient in producing a particular variety, it exports a large value to a lot of countries. Then observations of the exporter-variety pair will have a large effect on the product aggregates and be partly absorbed by the product fixed effect.

Columns (5) and (6) repeat the regressions of columns (3) and (4) dropping the product fixed effects. The theory does not require product fixed effects in the simulated data, in which all varieties are measured in comparable units. In support of our conjecture, the new coefficients on price, -1.90 in the OLS of column (5) (standard error 0.08) and -1.99 of column (6) (standard error 0.13), are close to the original coefficient -1.83. The last row of the table shows the implied parameter β. In the OLS without product fixed effects of column (5), we recover the value of β, 0.65, used in the simulations.

In all, these Monte Carlo exercises show that we can recover the price-related parameters (ν, γ, θ) using the estimation procedure in the model. The parameter governing the distribution of structural errors, β, is less well identified. The estimation procedure may bias it toward zero because product fixed effects absorb part of the variation across varieties within products.[4]

[4]We also verify that the trade shares in the simulated data are very close to the predicted shares and that we can recover the extensive margin parameters $(\lambda_1, \lambda_2, \lambda_3)$ of Section 4.4. These results are not surprising since the aggregation of varieties into products do not affect predicted trade flows and the extensive margin parameters already take into account this aggregation.

Table D.2: Estimates of β Using Simulated Data

Dependent variable is the value of trade flow, $\log x_{nik}$

	Data		Simulated Data		Simulated Data	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
price $\log p_{nik}$	-0.252	-1.828	-1.611	-1.269	-1.896	-1.990
	(0.038)	(0.019)	(0.052)	(0.045)	(0.080)	(0.133)
importer fixed effect	yes	yes	yes	yes	yes	yes
exporter fixed effect	yes	yes	yes	yes	yes	yes
product fixed effect	yes	yes	yes	yes	no	no
number of observations	2,585,111	2,585,111	2,306,998	2,306,998	2,306,998	2,306,998
implied β	0.20	0.65	0.62	0.56	0.65	0.67
	(0.024)	(0.002)	(0.008)	(0.009)	0.010	(0.015)

References

Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *The American Economic Review* 84 (1), 157-177.

Hummels, D. and P. J. Klenow (2005). The variety and quality of a nation's exports. *The American Economic Review* 95 (3), 704-723.

Sato, K. (1976). The ideal log-change index number. *The Review of Economics and Statistics* 58 (2), 223-228.

Vartia, Y. O. (1976). Ideal log-change index numbers. *Scandinavian Journal of Statistics* 3 (3), 121-126.