Estimating Trade Flows: Trading Partners and Trading Volumes

Elhanan Helpman, Marc Melitz and Yona Rubinstein
QJE 2008

Feb 18, 2009

Problems with previous studies using Gravity Equations

- Sample selection bias:
  to estimate the determinants of trade volume but only use the sample of countries that have positive trade flows between them; however, countries which trade with each other are not randomly selected

- Symmetry problem:
  impose symmetry across countries that is inconsistent with the data

Solution

- Develop a theory that predicts positive as well as zero trade flows between countries: Melitz (2003)+asymmetry for demand, trade cost and entry cost
- Theory leads to structural estimation that using information from both trading and nontrading countries (Heckman correction), as well as accounting for heterogeneity of firms without using firm-level data

Facts about Trade Volume and Trade Status

- Half of the country pairs do not trade with one another
- the rapid growth of trade (7.5% per year) was mostly driven by trade between countries that traded with each other in both directions in 1970.
- the average volume of bilateral trade among countries that traded in 1970 is much lager than those did not
- intensive margin of trade was driven by variations in the fraction of trading firms

First glance

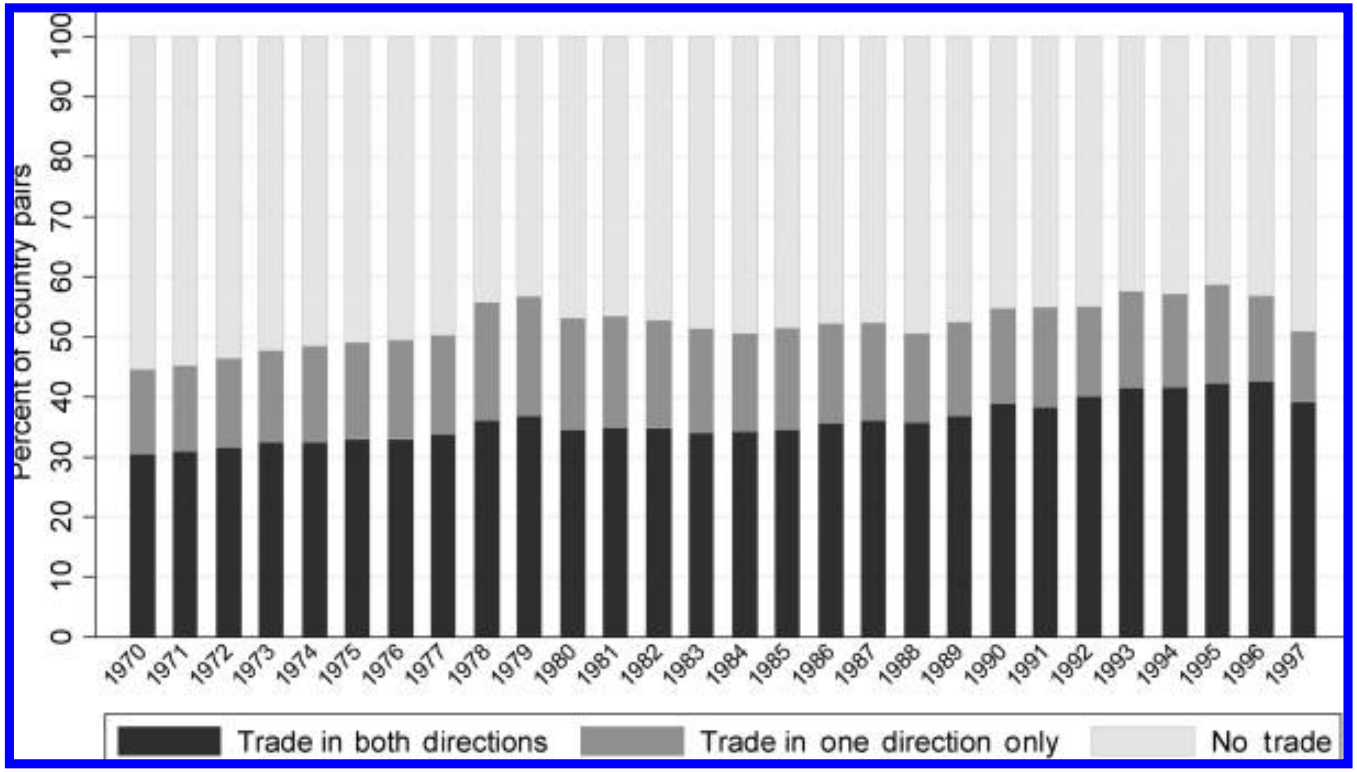

FIGURE I
Distribution of Country Pairs Based on Direction of Trade
Note. Constructed from 158 countries.

First glance

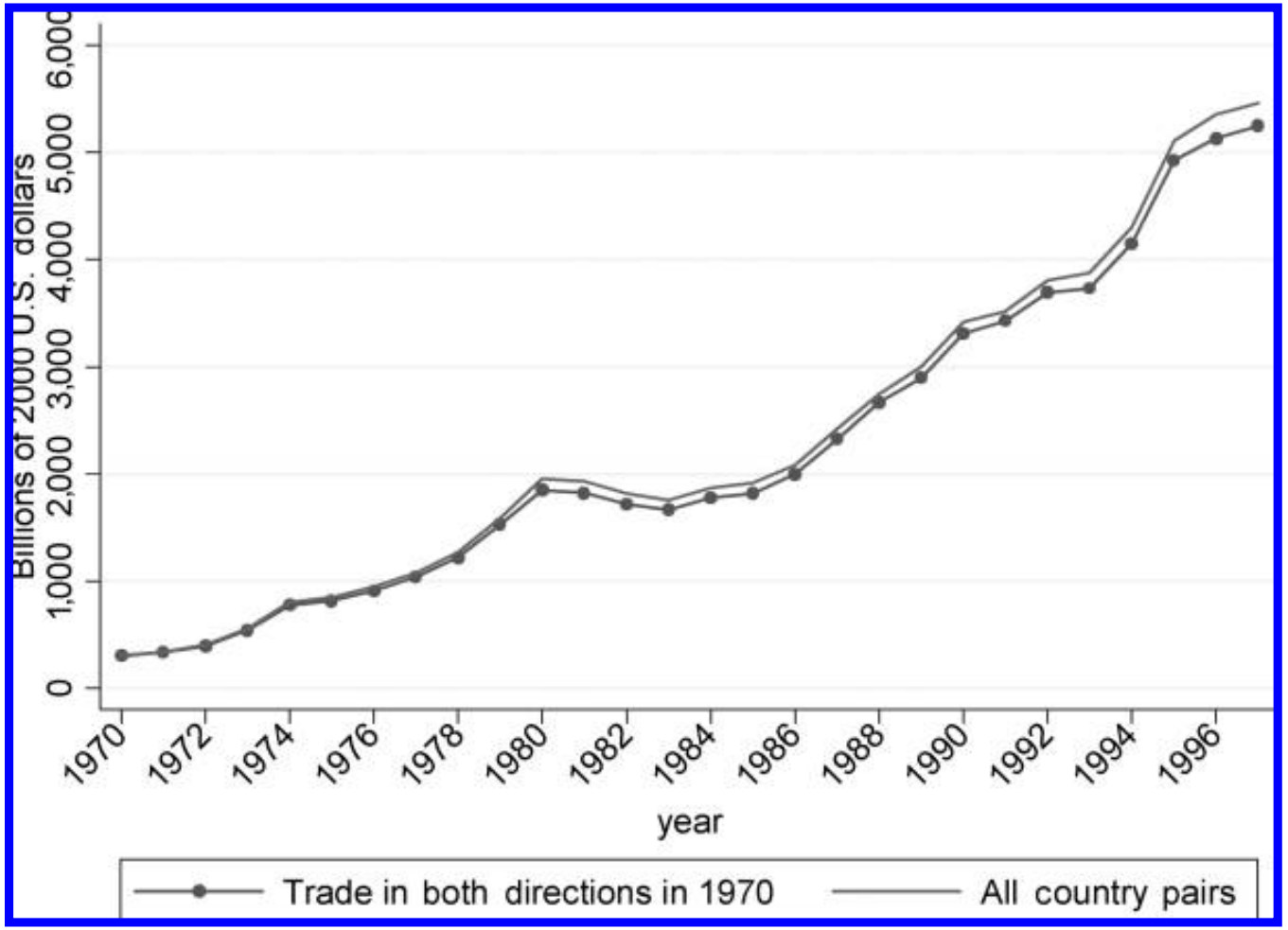

FIGURE II
Aggregate Volume of Exports of All Country Pairs and of Country Pairs That Traded in Both Directions in 1970

Theory - Melitz with asymmetric countries

Setup

- country j, product l, N_j monopolistic competitive firms
- utility $u_j = [\int_{l \in B_j} x_j(l)^\alpha dl]^{1/\alpha}$, elasticity of substitution $\epsilon = 1/(1-\alpha)$
- demand $x_j(l) = \frac{p_j(l)^{-\epsilon}}{P_j^{1-\epsilon}} Y_j$
- price index $P_j = [\int_{l \in B_j} p_j(l)^{1-\epsilon} dl]^{1/(1-\epsilon)}$
- per unit production cost $c_j a$
- $1/a$ represents firm specific productivity, cdf $G(a) : [a_L, a_H]$. same distribution across countries

Costs and prices

- Domestic market: production cost $c_j a$
- Foreign market i: $c_j a$, fixed production cost $c_j f_{ij}$ and "iceberg" cost τ_{ij}
- price: $p_j(l) = \tau_{ij} \frac{c_j a}{\alpha}$
- cutoff productivity a_{ij}:

$$\pi_{ij}(a_{ij}) = (1-\alpha)(\frac{\tau_{ij} c_j a_{ij}}{\alpha P_i})^{1-\epsilon} Y_i - c_j f_{ij} = 0$$

Trade Volume

- Imports of i from j: $M_{ij} = (\frac{c_j \tau_{ij}}{\alpha P_i})^{1-\epsilon} Y_i N_j V_{ij}$
  - $V_{ij} = \int_{a_L}^{a_{ij}} a^{1-\epsilon} dG(a)$ for $a_{ij} \geq a_L$, otherwise $V_{ij} = 0$
  - Pareto $G(a) = (a^k - a_L^k)/(a_H^k - a_L^k)$
  - $V_{ij} = bW_{ij}$
  - fraction of firms that export: $W_{ij} = \max\{(\frac{a_{ij}}{a_L})^{k-\epsilon+1} - 1, 0\}$
- Suppose $\tau_{ij}^{\epsilon-1} = D_{ij}^{\gamma} e^{-u_{ij}}$, unmeasured trade friction $u_{ij} \sim N(0, \sigma_u^2)$
- Estimating equation: $m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + w_{ij} + u_{ij}$

  previous gravity equation studies:
  - confound $\gamma d_{ij} + w_{ij} \Rightarrow$ upward bias γ
  - omit zero trade, $corr(u, d) > 0 \Rightarrow$ downward bias γ
    – country pairs with large (observed) d_{ij} that trade with each other are likely to have low (unobserved) u_{ij}

Controlling selection effect u_{ij} and endogenous numbers of exporters w_{ij}

Estimating w_{ij}

- Assume stochastic fixed export cost
  $f_{ij} = \exp(\phi_{EX,j} + \phi_{IM,i} + \kappa\phi_{ij} - \nu_{ij}), \nu_{ij} \sim N(0, \sigma_\nu^2)$

- ratio of export profits for the most productive firm to fixed export costs
  $Z_{ij} = \frac{(1-\alpha)(P_i \frac{\alpha}{c_j \tau_{ij}})^{\epsilon-1} Y_i a_L^{1-\epsilon}}{c_j f_{ij}} = (\frac{a_L}{a_{ij}})^{1-\epsilon}$, $Z_{ij} > 1$ if j exports to i
  $W_{ij} = Z_{ij}^{(k-\epsilon+1)/(\epsilon-1)} - 1$

- $z_{ij} = \gamma_0 + \xi_j + \zeta_i - \gamma d_{ij} - \kappa\phi_{ij} + \eta_{ij}$,
  $\eta_{ij} = u_{ij} + \nu_{ij} \sim N(0, \sigma_u^2 + \sigma_\nu^2)$

Heckman correction procedure

Suppose a researcher wants to estimate the determinants of wage offers, but only has access to wage for those who work.

- First stage: probability of working (based on theory) $\Pr(D = 1|Z) = \Phi(Z\gamma)$, Φ is unit-normal distribution (Heckman assume normality).
- Second stage: correct for self-selection by incorporating a transformation using the predicted probability as an additional explanatory variable
  - $w^* = X\beta + u$
  - $E[w|X, D = 1] = X\beta + E[u|X, D = 1]$
  - error terms are jointly normal: $E[w|X, D = 1] = X\beta + \rho\sigma_u\lambda(Z\gamma)$, λ is the inverse Mills ratio.
    (sample selection can be viewed as a form of omitted-variables bias)
  - disadvantage: normality assumption. *Exclusion restriction* is required: there must be at least one variable which appears with a non-zero coefficient in the selection equation but does not appear in the equation of interest, essentially an instrument

Consistent Estimation of Log-Linear Equation

steps:

1. Estimate Probit $\rho_{ij} = \Pr(T_{\{z_{ij}>0\}} = 1|\text{observed variables})$
   $= \Phi(\gamma_0^* + \xi_j^* + \zeta_i^* - \gamma^* d_{ij} - \kappa^* \phi_{ij})$,
2. predict $\hat{\rho}_{ij}$
3. estimate $\hat{z}_{ij}^* = \Phi^{-1}(\hat{\rho}_{ij})$, $\hat{\bar{\eta}}_{ij}^* = \phi(\hat{z}_{ij}^*)/\Phi(\hat{z}_{ij}^*)$ – inverse Mills' ratio
4. rewrite $\hat{\bar{w}}_{ij}^* = \ln\{\exp[\delta(\hat{z}_{ij}^* + \hat{\bar{\eta}}_{ij}^*)] - 1\}$, a consistent estimate for $E[w_{ij}|., T_{\{z_{ij}>0\}} = 1]$
5. rewrite $\hat{\bar{u}}_{ij} = \beta_{u\eta}\hat{\bar{\eta}}_{ij}^*$, a consistent estimate for $E[u_{ij}|., T_{\{z_{ij}>0\}} = 1]$

estimating equation:

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \ln\{\exp[\delta(\hat{z}_{ij}^* + \hat{\bar{\eta}}_{ij}^*)] - 1\} + \beta_{u\eta}\hat{\bar{\eta}}_{ij}^* + e_{ij}$$

Traditional Estimates - unidirectional trade value

TABLE I

BENCHMARK GRAVITY AND SELECTION INTO TRADING RELATIONSHIPS

	1986		1980s			
Variables	m_{ij}	(Probit) T_{ij}	m_{ij}	(Probit) T_{ij}	m_{ij}	(Probit) T_{ij}
Distance	−1.176**	−0.263**	−1.201**	−0.246**	−1.200**	−0.246**
	(0.031)	(0.012)	(0.024)	(0.008)	(0.024)	(0.008)
Land border	0.458**	−0.148**	0.366**	−0.146**	0.364**	−0.146**
	(0.147)	(0.047)	(0.131)	(0.032)	(0.131)	(0.032)
Island	−0.391**	−0.136**	−0.381**	−0.140**	−0.378**	−0.140**
	(0.121)	(0.032)	(0.096)	(0.022)	(0.096)	(0.022)
Landlock	−0.561**	−0.072	−0.582**	−0.087**	−0.581**	−0.087**
	(0.188)	(0.045)	(0.148)	(0.028)	(0.147)	(0.028)
Legal	0.486**	0.038**	0.406**	0.029**	0.407**	0.028**
	(0.050)	(0.014)	(0.040)	(0.009)	(0.040)	(0.009)
Language	0.176**	0.113**	0.207**	0.109**	0.203**	0.108**
	(0.061)	(0.016)	(0.047)	(0.011)	(0.047)	(0.011)
Colonial ties	1.299**	0.128	1.321**	0.114	1.326**	0.116
	(0.120)	(0.117)	(0.110)	(0.082)	(0.110)	(0.082)
Currency union	1.364**	0.190**	1.395**	0.206**	1.409**	0.206**
	(0.255)	(0.052)	(0.187)	(0.026)	(0.187)	(0.026)
FTA	0.759**	0.494**	0.996**	0.497**	0.976**	0.495**
	(0.222)	(0.020)	(0.213)	(0.018)	(0.214)	(0.018)
Religion	0.102	0.104**	−0.018	0.099**	−0.038	0.098**
	(0.096)	(0.025)	(0.076)	(0.016)	(0.077)	(0.016)
WTO (none)					−0.068	−0.056**
					(0.058)	(0.013)
WTO (both)					0.303**	0.093**
					(0.042)	(0.013)
Observations	11,146	24,649	110,697	248,060	110,697	248,060
R^2	0.709	0.587	0.682	0.551	0.682	0.551

Two-Stage Estimation

		m_{ij}				
					Indicator variables	
Variables	(Probit) T_{ij}	Benchmark	NLS	Polynomial	50 bins	100 bins
Distance	−0.213**	−1.167**	−0.813**	−0.847**	−0.755**	−0.789**
	(0.016)	(0.040)	(0.049)	(0.052)	(0.070)	(0.088)
Land border	−0.087	0.627**	0.871**	0.845**	0.892**	0.863**
	(0.072)	(0.165)	(0.170)	(0.166)	(0.170)	(0.170)
Island	−0.173*	−0.553*	−0.203	−0.218	−0.161	−0.197
	(0.078)	(0.269)	(0.290)	(0.258)	(0.259)	(0.258)
Landlock	−0.053	−0.432*	−0.347*	−0.362+	−0.352+	−0.353+
	(0.050)	(0.189)	(0.175)	(0.187)	(0.187)	(0.187)
Legal	0.049**	0.535**	0.431**	0.434**	0.407**	0.418**
	(0.019)	(0.064)	(0.065)	(0.064)	(0.065)	(0.065)
Language	0.101**	0.147+	−0.030	−0.017	−0.061	−0.036
	(0.021)	(0.075)	(0.087)	(0.077)	(0.079)	(0.083)
Colonial ties	−0.009	0.909**	0.847**	0.848**	0.853**	0.838**
	(0.130)	(0.158)	(0.257)	(0.148)	(0.152)	(0.153)
Currency union	0.216**	1.534**	1.077**	1.150**	1.045**	1.107**
	(0.038)	(0.334)	(0.360)	(0.333)	(0.337)	(0.346)
FTA	0.343**	0.976**	0.124	0.241	−0.141	0.065
	(0.009)	(0.247)	(0.227)	(0.197)	(0.250)	(0.348)
Religion	0.141**	0.281*	0.120	0.139	0.073	0.100
	(0.034)	(0.120)	(0.136)	(0.120)	(0.124)	(0.128)
Regulation costs	−0.108**	−0.146				
	(0.036)	(0.100)				
R. costs (days & proc.)	−0.061*	−0.216+				
	(0.031)	(0.124)				
δ (from $\hat{w}^*_{ij}$)			0.840**			
			(0.043)			
$\hat{\eta}^*_{ij}$			0.240*	0.882**		
			(0.099)	(0.209)		
$\hat{z}^*_{ij}$				3.261**		
				(0.540)		
$\hat{z}^{*2}_{ij}$				−0.712**		
				(0.170)		
$\hat{z}^{*3}_{ij}$				0.060**		
				(0.017)		
Observations	12,198	6,602	6,602	6,602	6,602	6,602
R^2	0.573	0.693		0.701	0.704	0.706

Bias

- higher trade volumes are not just direct consequence of lower trade barriers; greater proportion of exporters to a particular destination also matters
- previous studies upward biases the effect of trade friction.
- large bais on FTAs, common language and religion

Decomposing the Effect

BIAS DECOMPOSITION

	1986 full sample			
Variables	Benchmark	NLS	Firm heterogeneity	Heckman selection
Distance	−1.176**	−0.798**	−0.769**	−1.214**
	(0.031)	(0.039)	(0.038)	(0.031)
Land border	0.458**	0.834**	0.855**	0.436**
	(0.147)	(0.132)	(0.142)	(0.149)
Island	−0.391**	−0.169	−0.164	−0.425**
	(0.121)	(0.120)	(0.118)	(0.120)
Landlock	−0.561**	−0.447**	−0.433*	−0.565**
	(0.188)	(0.172)	(0.187)	(0.187)
Legal	0.486**	0.387**	0.381**	0.488**
	(0.050)	(0.048)	(0.049)	(0.050)
Language	0.176**	0.023	0.023	0.223**
	(0.061)	(0.062)	(0.060)	(0.061)
Colonial ties	1.299**	1.001**	0.979**	1.311**
	(0.120)	(0.204)	(0.119)	(0.123)
Currency union	1.364**	1.023**	0.996**	1.391**
	(0.255)	(0.273)	(0.260)	(0.257)
FTA	0.759**	0.380*	0.314+	0.737**
	(0.222)	(0.182)	(0.168)	(0.235)
Religion	0.102			
	(0.096)			
δ (from $\hat{w}^*_{ij}$)		0.871**		
		(0.028)		
$\hat{\eta}^*_{ij}$		0.372**		0.265**
		(0.069)		(0.070)
$\hat{z}^*_{ij}$			0.892**	
			(0.051)	
Observations	11,146	11,146	11,146	11,146
R^2	0.709		0.716	0.710

Asymmetries

ASYMMETRIES

	1986 full sample				
		$m_{ij} - m_{ji}$			
Variable	$T_{ij} - T_{ji}$	NLS		Polynomial	
$\hat{\rho}_{ij} - \hat{\rho}_{ji}$	0.999**				
	(0.0169)				
$\hat{\bar{w}}^*_{ij} - \hat{\bar{w}}^*_{ji}$		1.187**	1.251**		
		(0.042)	(0.266)		
$\hat{v}(\hat{\bar{z}}^*_{ij}) - \hat{v}(\hat{\bar{z}}^*_{ji})$				1.012**	0.703**
				(0.035)	(0.143)
Country fixed effects	No	No	Yes	No	Yes
Observations	12,246	4,517	4,517	4,517	4,517
R^2	0.219	0.153	0.324	0.157	0.325