Ninth ARTNeT Capacity Building Workshop for Trade Research
"Trade Flows and Trade Policy Analysis"

June 2013

Bangkok, Thailand

Cosimo Beverelli and Rainer Lanz

(World Trade Organization)

The gravity model in international trade

Content

a. What is it?
b. Naïve gravity estimation
c. Theoretical foundations
d. Mistakes to avoid
e. Estimating theoretically-founded gravity equation
f. Recent theoretical developments
g. References

a. What is it?

- Econometric model (ex-post analysis)
- Models many social interactions (migration, tourism, trade, FDI)
- For decades, social scientists have been using a modified version of Isaac Newton's Law of Gravitation to predict movement of people, information, and commodities between cities and even continents
- The gravity model takes into account the population size of two places and their distance. Since larger places attract people, ideas, and commodities more than smaller places and places closer together have a greater attraction, the gravity model incorporates these two features
- Initially, no theoretical foundations
- Why so popular?
  - High explanatory power (R^2 between 0.65 and 0.95)
  - Easy access to relevant data
  - Estimation standards and benchmarks clearly established

Newton's universal law of gravitation and the gravity specification in trade

Newton

$$F_{ij} = G \frac{M_i M_j}{D_{ij}^2}$$

- Where *F* is the attraction force, *G* is the gravitational constant, *M* is mass, *D* is distance, *i* and *j* index point masses

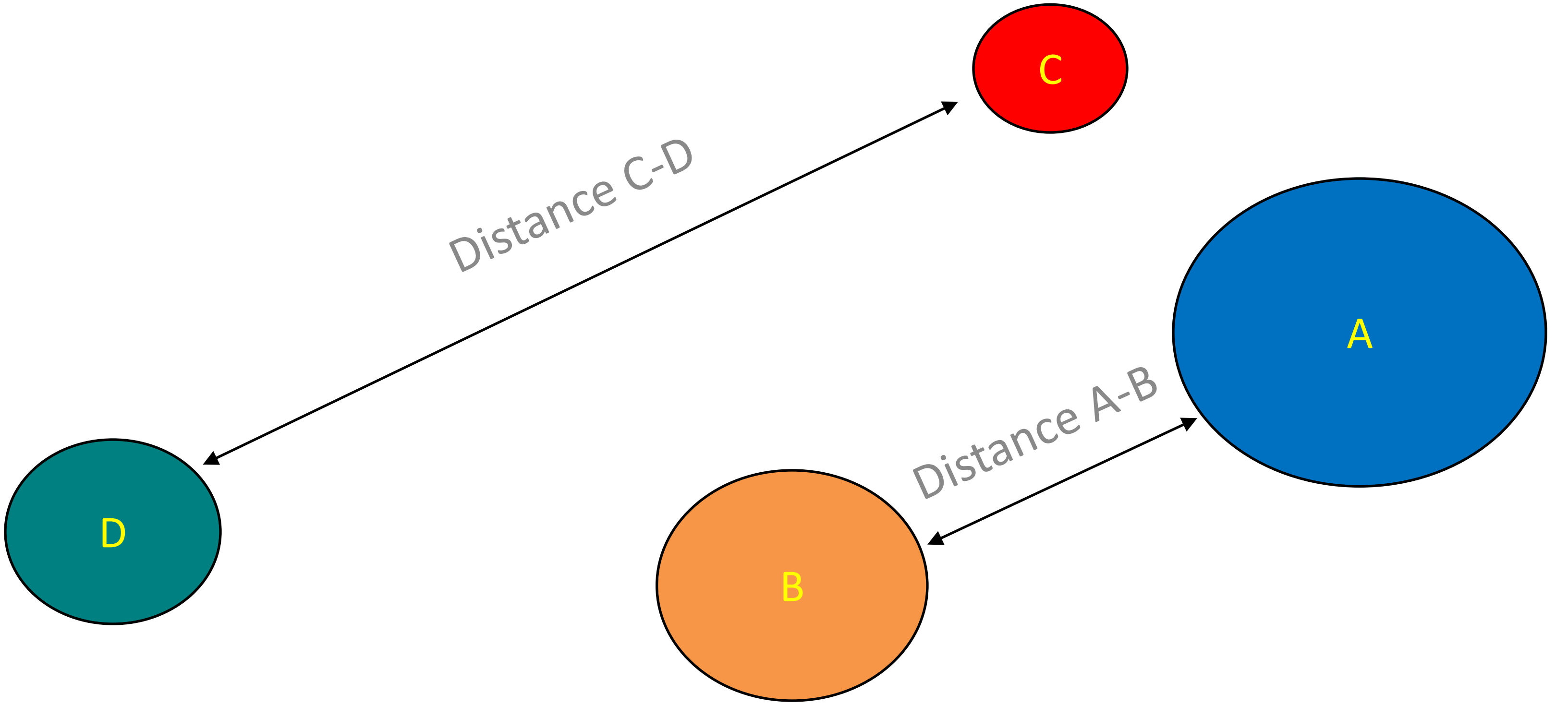
Gravity

$$X_{ij} = G \frac{Y_i^{\alpha} Y_j^{\beta}}{T_{ij}^{\theta}}$$

- Where X_{ij}= exports from country *i* to *j* or total trade; *Y*=economic size (GDP, POP) and *T*= Trade costs

The effects of size and distance

- More bilateral trade the larger the countries
- Less bilateral trade the more the distance between them
- A and B are predicted to trade more than C and D



Proxies for trade costs

- Distance
- Contiguity
- Common language
- Colonial links
- Common currency
- Island, landlocked country
- Institutions, infrastructure, migration flows,…
- Surprisingly enough, tariffs are very often omitted
- Main reason: endogeneity issue
  - Do tariffs reduce trade or does more trade induce tariff reductions? (reverse causality problem)

b. Naïve gravity estimation

- Naïve estimation of the gravity regression is

$$\ln(Trade_{ij}) = \alpha + \beta_1 \ln(GPD_i) + \beta_2 \ln(GPD_j) + \beta_3 \ln(dist_{ij}) + \varepsilon_{ij}$$

- This regression fits the data very well
  - R-squared of 0.7 in cross-section data (*ij* dataset)

- However this naïve version can lead to very biased results
  - Serious omitted variable bias: any *i*- or *j*- characteristic that correlates both with trade and GDP ends up in the error term. The basic OLS assumption of orthogonality between the error term and the explanatory variables is violated

c. Theoretical foundations

- Deardorff (1998) *"I suspect that just about any plausible model of trade would yield something very like the gravity equation"*

- Baldwin and Taglioni (2006) *"The emergency of the new trade theory in the late 1970s and early 1980s started a trend where the gravity model passed from having too few theoretical foundations to having too many"*

Major contributions

- Anderson (1979)
  - Armington hypothesis (goods differentiated by country of origin)
- Bergstrand (1990)
  - Monopolistic competition
  - Price indices are those used in practice, not suggested by the theory
- Anderson and Van Wincoop (2003)
  - Monopolistic competition
  - No complete account of price indices => practical way to estimate gravity coefficients in cross country
- Helpman et al. (2008)
  - Trade model with heterogeneous firms: structural treatment of zeros

- See also Head and Mayer (2013)

Deriving the gravity equation

a. Step 1: the (Dixit-Stiglitz) demand function

$$x_{ij} = Y_j \frac{p_{ij}^{-\sigma}}{P_j^{1-\sigma}} \qquad (1)$$

Where *i* indexes exporter, *j* indexes importer

- LHS = nominal demand by *j*'s consumers for *i*'s goods
- Y_j is *j*'s nominal income
- p_{ij} is imports price

$$P_j \equiv \left[\sum_i \left(p_{ij} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \qquad (2)$$

is the ideal CES price index in *j,* σ > 1 is the elasticity of substitution across varieties

Deriving the gravity equation (ct'd)

- (1) can be rewritten in terms of value:

$$p_{ij}x_{ij} \equiv T_{ij} = Y_j \left(\frac{p_{ij}}{P_j} \right)^{1-\sigma} \qquad (3)$$

- Equation (3) could be estimated directly, but researchers often lack good data on trade prices

b. Step 2: adding the pass-through equation

$$p_{ij} = p_i t_{ij} \qquad (4)$$

Where p_i is the producer price in country *i* and t_{ij} are all trade costs

Deriving the gravity equation (ct'd)

c. Step 3: Market clearing condition (aggregate supply equals aggregate demand) to eliminate the nominal price

$$Y_i = \sum_j T_{ij} \Leftrightarrow \frac{Y_i}{Y_W} \equiv \theta_i = \sum_j \frac{T_{ij}}{Y_W} \qquad (5)$$

- Where Y_W is world nominal income and $\theta_i \equiv Y_i/Y_W$ is the share of *i*'s nominal income in world nominal income
- Using (3) and (4) into (5) we obtain:

$$\theta_i = \left(p_i\right)^{1-\sigma} \Omega_i^{1-\sigma} \qquad (6)$$

In this expression,

$$\Omega_i = \left[\sum_j \left(\frac{t_{ij}}{P_j} \right)^{1-\sigma} \theta_j \right]^{\frac{1}{1-\sigma}} \qquad (7)$$

is the multilateral resistance (openness of *i*'s exports to world markets)

Deriving the gravity equation (ct'd)

- We now want to get rid of producer prices
- Substitute (6) back into (3) and use (4) to obtain:

$$T_{ij} = \frac{Y_i Y_j}{Y_W} \left(\frac{\tau_{ij}}{\Omega_i P_j} \right)^{1-\sigma} \qquad (8)$$

- Theoretically-founded gravity equation
- Major contribution of Anderson and Van Wincoop (2003): bilateral trade is determined by relative trade costs
- Using (6) into (2), one can derive the price index as:

$$P_j = \left[\sum_j \left(\frac{\tau_{ij}}{\Omega_i} \right)^{1-\sigma} \theta_i \right]^{\frac{1}{1-\sigma}} \qquad (9)$$

- Equations (7) and (9) can be solved for all *P*'s and *Ω*'s

Deriving the gravity equation (ct'd)

- Anderson and Van Wincoop (2003) show that if trade costs are symmetric ($\tau_{ij} = \tau_{ji}$), $\Omega_i = P_i$ and the gravity equation (9) becomes:

$$X_{ij} = \frac{Y_i Y_j}{Y_W} \left(\frac{t_{ij}}{P_i P_j} \right)^{1-\sigma} \qquad (10)$$

- Taking $dist_{ij}$ as proxy for t_{ij}, rewrite (10) to look like as Newton's physical law of gravity:

$$\text{bilateral trade} = G \frac{Y_i Y_j}{(dist_{ij})^{\sigma-1}}$$

Where $G \equiv (1/P_i P_j)^{1-\sigma}$ is not a constant as it is in the physical world, but rather *ij* specific (also *t* if we have time variation)

Deriving the gravity equation (ct'd)

- In the gravity literature it is in general assumed that trade costs take the form:

$$t_{ij} = d_{ij}^{\delta_1} \exp\left(\delta_2 cont_{ij} + \delta_3 lang_{ij} + \delta_4 ccol_{ij} + \delta_5 col_{ij} + \delta_6 llock_{ij} + \delta_7 RTA_{ij}\right)$$

- Where d_{ij} is bilateral distance, and $cont_{ij}$, $lang_{ij}$, $ccol_{ij}$, col_{ij}, $llock_{ij}$, RTA_{ij} are dummy variables denoting respectively whether the two countries have a common border, common language, common colonizer, whether one was a colony of the other at some point in time, whether one of the two is a landlocked country, whether the two countries are member of and RTA

d. Mistakes to avoid

- Standard estimating procedure

$$\ln(x_{ij}) = \alpha + \beta_1 \ln(Y_i) + \beta_2 \ln(Y_j) + \beta_3 \ln(dist_{ij}) + \varepsilon_{ij}$$

- Gold medal mistake: omitting the un-constant term *G*
  - Omitted variable bias, since Ω_i and P_j (in error term) include t_{ij} (in matrix of explanatory variables)

Other common mistakes include:

- Silver medal mistake: averaging the reciprocal trade flows (exports from *i* to *j* and exports from *j* to *i*) before taking logs (theory tells that one should take the average of logs, not the log of the average, of X_{ij} and X_{ji})

- Bronze medal mistake: deflating the nominal flows by US price index
  - Gravity is an expenditure function allocating nominal GDP into nominal imports. If use US GDP deflator, need to include time dummies to control for inflation

e. Estimating theoretically-founded gravity equation

- Normal trade with resistances

$$\ln(x_{ij}) = \alpha + \beta_1 \ln(Y_i) + \beta_2 \ln(Y_j) + \beta_3 \ln(dist_{ij}) + MTR_i + MTR_j + \varepsilon_{ij}$$

- MTR is the multilateral trade resistance term…
- …it is unobservable
- Three ways to take MTR into account:

1. Use an iterative method to solve MTR as function of observables (see Anderson and Van Wincoop, 2003) or linear approximation of MTR (Baier and Bergstrand, 2009)
2. Proxy MTR using remoteness REM (trade/GDP weighted average distances from the rest of the world)

$$REM_i = \sum_j \frac{dist_{ij}}{GDP_j / GDP_W}$$

3. Fixed Effects (FE)

Fixed effects estimation

- Importer (exporter) fixed effect is a (0, 1) dummy that denotes the importer (exporter)
- Fixed effects control for unobserved characteristics of a country, i.e. any country characteristic that affect its propensity to import (export)
- They are used to proxy each country remoteness
- They do not control for unobserved characteristics of pair of countries, e.g. if they have a RTA in place (need country-pair fixed effects for this)
- Cross section vs. panel datasets

a. Fixed effects estimation in a cross section

- In cross country analysis MTR are fixed. Therefore, using country fixed effects yields consistent estimation

$$\ln(x_{ij}) = \alpha + \beta_1 \ln(\tau_{ij}) + \sum_i \gamma_i D_i + \sum_j \gamma_j D_j + \varepsilon_{ij}$$

  Where

  - τ_{ij} is bilateral trade cost (e.g. distance)
  - *D* are country specific dummies
  - There are 2n dummies
  - Total observations = n(n-1)

- It is impossible to estimate the coefficient for GDP and other country-specific variables (they are perfectly collinear with the country fixed effect)

b. Fixed effects estimation in a panel

- It is now possible to estimate the coefficient for GDP and other country-specific variables

$$\ln(x_{ijt}) = \alpha + \beta_1 \ln(Y_{it}) + \beta_2 \ln(Y_{jt}) + \beta_3 \ln(\tau_{ijt}) + \sum_i \gamma_i D_i + \sum_j \gamma_j D_j + \varepsilon_{ij}$$

Where

- There are 2n country-specific dummies
- Total observations = n(n-1)T

- It is still not possible to estimate time-invariant country-specific characteristics (e.g. island, landlockedness…)

- There may be a bias due to the variation over time of MTRs

b. Fixed effects estimation in a panel (ct'd)

- In the case of a panel, MTRs may change over time (variation in transport costs or composition of trade)

$$\ln(x_{ijt}) = \alpha + \beta_1 \ln(\tau_{ijt}) + \sum_i \gamma_{it} D_{it} + \sum_j \gamma_{jt} D_{jt} + \sum_t \delta_t K_t + \varepsilon_{ij}$$

  Where
  - *D* are time-varying country-specific dummies
  - *K* is a time dummy (to take global inflation trends, business cycles and the like into account)
  - There are 2nT+T dummies, where T denotes the time period

- Impossible to estimate the coefficient of GDP (perfectly collinear with country-time dummy)

b. Fixed effects estimation in a panel (ct'd)

- Address the bias due to the correlation between the bilateral trade barriers and the MTRs

$$\ln(x_{ijt}) = \alpha + \beta_1 \ln(Y_{it}) + \beta_2 \ln(Y_{jt}) + \beta_3 \ln(\tau_{ijt}) + \sum_i \gamma_{ij} D_{ij} + \varepsilon_{ij}$$

Where

  - I= are country-pair dummies
  - There are n(n-1)/2 such dummies

- Disadvantage: coefficients of bilateral variables are estimated on the time dimension of the panel and cannot estimate coefficient for distance, common border, common language (they do not vary over time)
- A solution: Use random effects and the Hausman test (*xtoverid* in Stata) to choose between random and fixed effect estimation

How to proceed?

- Sensitivity analysis, test the robustness of the results to alternative specifications of the gravity equations

- Report the results for the different equations estimated

Interpretation of results

- Most of the variables are expressed in natural logarithms, so coefficients obtained from linear estimation can be read directly as elasticities
- The elasticity of trade to distance, for instance, is usually between -1 and -1.5, so a 10 per cent increase in distance between two countries cuts their trade, on average, by 10 to 15 per cent
- Elasticities with respect to importing-country GDPs are also typically unitary, suggesting unitary income elasticities of imports at the aggregate level

Interpretation of results (ct'd)

- The coefficients for the dummies (e.g. common border) are not elasticities
- They need to be transformed as follows to be interpreted as elasticities:

$$\widehat{p} = \exp(\widehat{\delta}) - 1$$

where $\hat{p}$ is the % change in the dependent variable and $\hat{\delta}$ is the estimated coefficient of the dummy variable

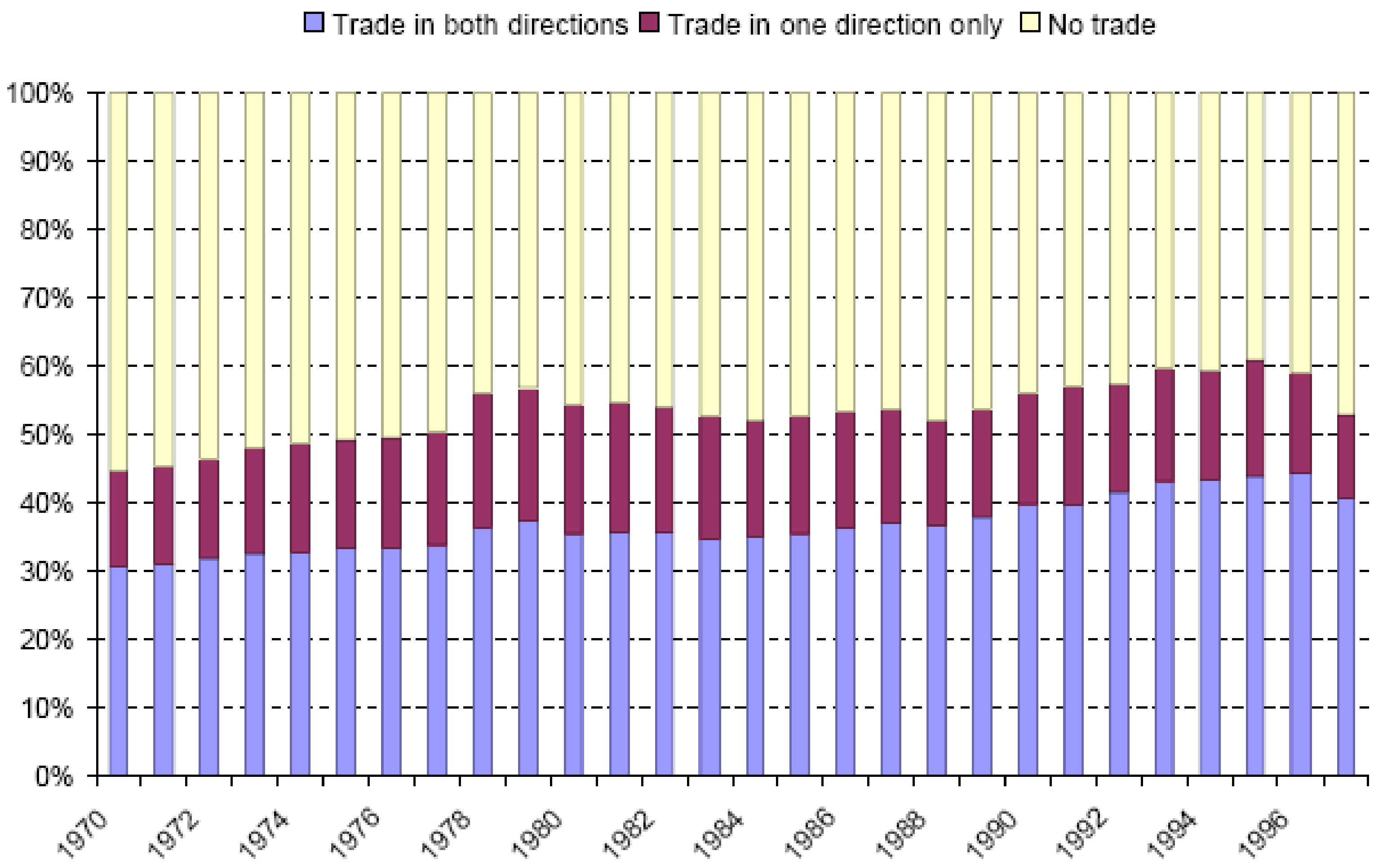
- To derive this formula:
  - Consider that $\ln X_{ij(1)}$ is the predicted value of trade when the dummy = 1 while $\ln X_{ij(0)}$ is the value of trade when dummy = 0
  - The difference $\ln X_{ij(1)} - \ln X_{ij(0)} = \delta$
  - $X_{ij(1)} / X_{ij(0)} = \exp(\delta)$, which in turn implies that the percentage change in trade value due to the dummy switching from 0 to 1 is: $X_{ij(1)} - X_{ij(0)} / X_{ij(0)} = \exp(\delta) - 1$

- You can do better and estimate $\hat{p} = \frac{\exp[\widehat{\delta}]}{\exp[\frac{1}{2}var(\widehat{\delta})]} - 1$, which is consistent and (almost) unbiased

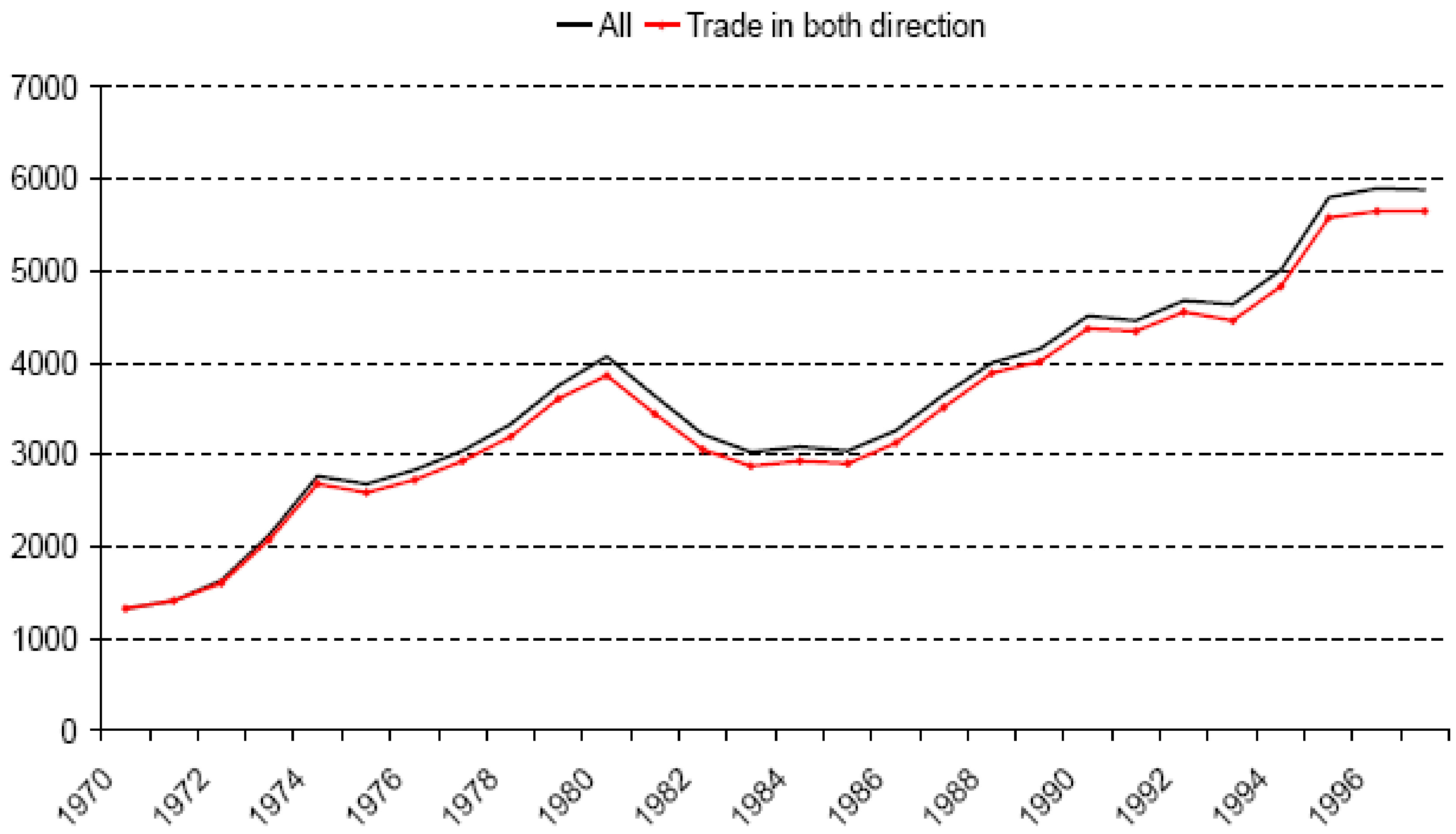
f. Recent theoretical developments

- Recently, Helpman et al. (2008) (HMR) derive a gravity equation from an heterogeneous firms model of trade
- The importance of this derivation relates to three issues that previous models of trade could not explain
  - Zero-trade observations
  - Asymmetric trade flows
  - The extensive margin of trade: more countries trade over time

Incidence of zero trade



Extensive margin of trade



- Trade growth in the extensive margin is minor

How to handle zero trade data?

a. Traditionally

- When taking logs, zero observations are dropped from the sample. Then, the OLS estimation is run on positive values
- Take the log(1+X_{ij}), but then use Tobit estimation as the OLS would provide biased results (distribution is censored at zero: all observations possibly negative are identified as equal to zero)

b. More recently

- The probability of having positive (non-zero) trade between two countries is likely correlated with unobserved characteristics of that country pair
- A selection model *à la* Heckman is called for. In this context, zero trade flows result from the firm decisions not to export in a certain market
  - The appropriate estimation procedure is therefore that to model these decisions, and correct the estimation on the volume of trade for this selection bias

How to handle zero trade data? (ct'd)

b. More recently (ct'd)

- HMR claim: the standard approach produces biased results
- In their model, differences in trade costs across countries and firms heterogeneity account for both asymmetric trade flows and zero trade
- Zero trade occurs when the productivity of all firms in country *i* is below the threshold that would make exporting to *j* profitable
- Problem: the probability of having positive trade between 2 countries is correlated with unobserved characteristics of that country pair
- These characteristics also affect the volume of their bilateral trade, given that they trade

How to handle zero trade data? (ct'd)

- Selection bias and omitted variable problem
- HMR provide the following solution: 2 stage estimation

1. Probit on the likelihood that 2 countries trade (i.e. probability that firms enter the export market)
2. Gravity model on positive trade values, where
   - The results of the first stage are use to correct for the sample selection bias introduced by omitting zero trade flows (the standard Heckman correction term, the inverse of the Mill's ratio)
   - An estimate of the (unobserved) share of firms selecting into the export market is added to the equation to control for firm heterogeneity

Helpman et al. (HMR) (2008) model

1. First-stage Probit regression

$$\rho_{ij} = \Pr(T_{ij} = 1) = \Phi(\gamma_0 + \xi_j + \zeta_i - \gamma d_{ij} - \kappa\varphi_{ij})$$

Where the probability of where the probability of a positive trade flows between *i* and *j*, ρ_{ij}, depends on the importer and exporter dummies (ξ_j and ζ_i) and bilateral trade costs, where *d* denotes variable trade costs and φ denotes bilateral fixed costs of entry

2. Second-stage gravity regression

$$x_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \ln\{\exp[\delta(z_{ij} + \eta_{ij})] - 1\} + \beta_\eta \eta_{ij} + \varepsilon_{ij}$$

Where λ_j and χ_i are importer and exporter fixed effects and the term in the curly brackets is the estimator for the share of firms that export to *j;* z_{ij} are the fitted variable for the latent variable from the fist-stage Probit and η_{ij} is the inverse Mill's ratio

- Notice that the last equation has to be estimated by non-linear least squares (NLLS)

How to handle zero trade data? (ct'd)

b. More recently (ct'd)

- An alternative approach is to use (Pseudo) Poisson maximum likelihood (ML) estimator. This method can be applied on the levels of trade, thus estimating directly the non-linear form of the gravity model and avoiding dropping zero trade
- Santos Silva and Tenreyro (2006) highlight that, in the presence of heteroskedasticity (as usual in trade data), the PPML is a robust approach
- This approach has been used in a number of estimation of gravity equations
- See the "log of gravity" webpage for details, FAQs, Stata codes and references

g. Application: Gravity estimation of AVE of NTBs

- The gravity equation can be used in reverse to measure bilateral trade costs and to decompose trade costs into a tariff and non-tariff component (see for instance Jacks et al., 2011 and ESCAP/WB database)

a. Theoretical approach

- Take expression (8) and write it down for X_{ij}, X_{ji}, X_{ii} and X_{jj}, where X_{ii} and X_{jj} are the expressions for intra-national trade
- Rearrange trade costs

$$\frac{t_{ij}t_{ji}}{t_{ii}t_{jj}} = \left(\frac{X_{ii}X_{jj}}{X_{ij}X_{ji}}\right)^{\frac{1}{1-\sigma}} \quad (11)$$

- Tariff equivalent of bilateral trade costs relative to domestic trade costs can then be expressed as geometric average of trade barriers in both directions

$$\tau = \sqrt{\frac{t_{ij}t_{ji}}{t_{ii}t_{jj}}} - 1 = \left(\frac{X_{ii}X_{jj}}{X_{ij}X_{ji}}\right)^{\frac{1}{2(1-\sigma)}} - 1 \quad (12)$$

a. Theoretical approach (ct'd)

- Expression (11) is a derivation of overall trade costs from the gravity equation without imposing a cost function
- It is neither assumed that domestic trade costs are zero, nor that they are the same across countries (t_{ii} may differ from t_{jj}) nor that bilateral trade costs are symmetric (t_{ij} may differ from t_{ij})
- Under the specific assumption that domestic trade costs are zero and bilateral trade costs symmetric (as implied by taking the geometric average as measure of bilateral trade costs), it is also possible to decompose overall trade costs in their various cost components by assuming an arbitrary trade cost function (see slide 18)

a. Theoretical approach (ct’d)

- For example, it is possible to decompose overall trade costs in their tariff and non-tariff component by simply estimating

$$\ln(\tau_{ij}) = \delta_1 \ln(dist_{ij}) + \delta_2 \ln(tariff_{ij}) + \delta_3 \ln(NTB_{ij}) + \varepsilon_{ij}$$

where NTB is a dummy

- To compute the tariff equivalent of a quota, we only need to calculate what is the percentage change in tariff that has the same impact on trade costs as a quota
- Tariff equivalent = $exp(\delta_3/\delta_2) - 1$

b. Empirical implementation

- The difficulty in calculating (11) is to get figures for intra-national trade
- One approach is to estimate these figures as the difference between production and exports (see for instance Novy 2012)
  - The use of GDP instead of production data tends to overstate intra-national trade, and therefore trade costs, because a growing share of trade is services (largely non-tradable)
- One can use the TPP datasets (World Bank or CEPII)
- See Exercise 2 of Chapter 3 of the Practical Guide to Trade Policy Analysis