ARTNeT- KRI Capacity Building Workshop on Trade Policy Analysis:
Evidence-based Policy Making and Gravity Modelling for Trade Analysis
18-20 August 2015, Kuala Lumpur

Session 5: Advanced issues

Dr. Witada Aunkoonwattaka

Trade and Investment Division, ESCAP

anukoonwattaka@un.org

UNITED NATIONS ESCAP
Economic and Social Commission for Asia and the Pacific

ARTNeT

KHAZANAH RESEARCH INSTITUTE

Overview of the workshop

Day 1 AM: Introduction to evidence-based policy making (EBPM) and tools

Day 1 PM: Introduction to gravity model for trade analysis

- Its concepts, applications, weakness, and development

Day 2 AM: Estimating the gravity models in STATA

- Estimating intuitive gravity models
- Estimating trade potential using a gravity model
- Problems of intuitive gravity models

Day 2 PM: Theoretical gravity models

- Theoretical gravity models
- Econometric approaches to estimating theoretical gravity models

Day 3: Advanced issues and consolidation

- Recent development in gravity techniques
- Group exercises and presentation
- Workshop wrap-up

Dealing with endogeneity

- Endogeneity (the non-zero correlation between Xi's and the error term) is usually a serious problem for the applied economist.
- The main causes are:

  -Missing variables (from the regression model)

  -Simultaneous relationships

  -Reverse causality

  -(Systematic) measurement error

  -Self-selection

  -Serial autocorrelation

  -Stochastic/deterministic trend

- Use of instrumental variables (2SLS and 3SLS), dynamic models (panel data in first difference), BB approach, etc.

First-stage regression of tariff_exp_sim:

OLS estimation

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on distw

Number of clusters (distw) = 520		Number of obs =	858
		F(9, 519) =	10.32
		Prob > F =	0.0000
Total (centered) SS =	24361.77561	Centered R2 =	0.1054
Total (uncentered) SS =	73193.39336	Uncentered R2 =	0.7022
Residual SS =	21794.88273	Root MSE =	5.07

tariff_exp~m	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_gdp_exp	-.5629992	.0991486	-5.68	0.000	-.7577812	-.3682172
ln_gdp_imp	.0838952	.0699862	1.20	0.231	-.0535959	.2213863
ln_distance	-.5332194	.3246971	-1.64	0.101	-1.171102	.1046628
contig	1.35537	.6347469	2.14	0.033	.1083814	2.60236
comlang_off	-1.918886	.5189189	-3.70	0.000	-2.938326	-.8994466
colony	-.9304101	.908853	-1.02	0.306	-2.715893	.8550728
comcol	1.080064	.6265997	1.72	0.085	-.1509196	2.311047
ln_lat_exp	.7109096	.1916194	3.71	0.000	.3344646	1.087355
ln_lat_imp	-.2880081	.2000156	-1.44	0.150	-.6809479	.1049316
_cons	23.49424	3.92257	5.99	0.000	15.78817	31.2003

Included instruments: ln_gdp_exp ln_gdp_imp ln_distance contig comlang_off colony comcol ln_lat_exp ln_lat_imp

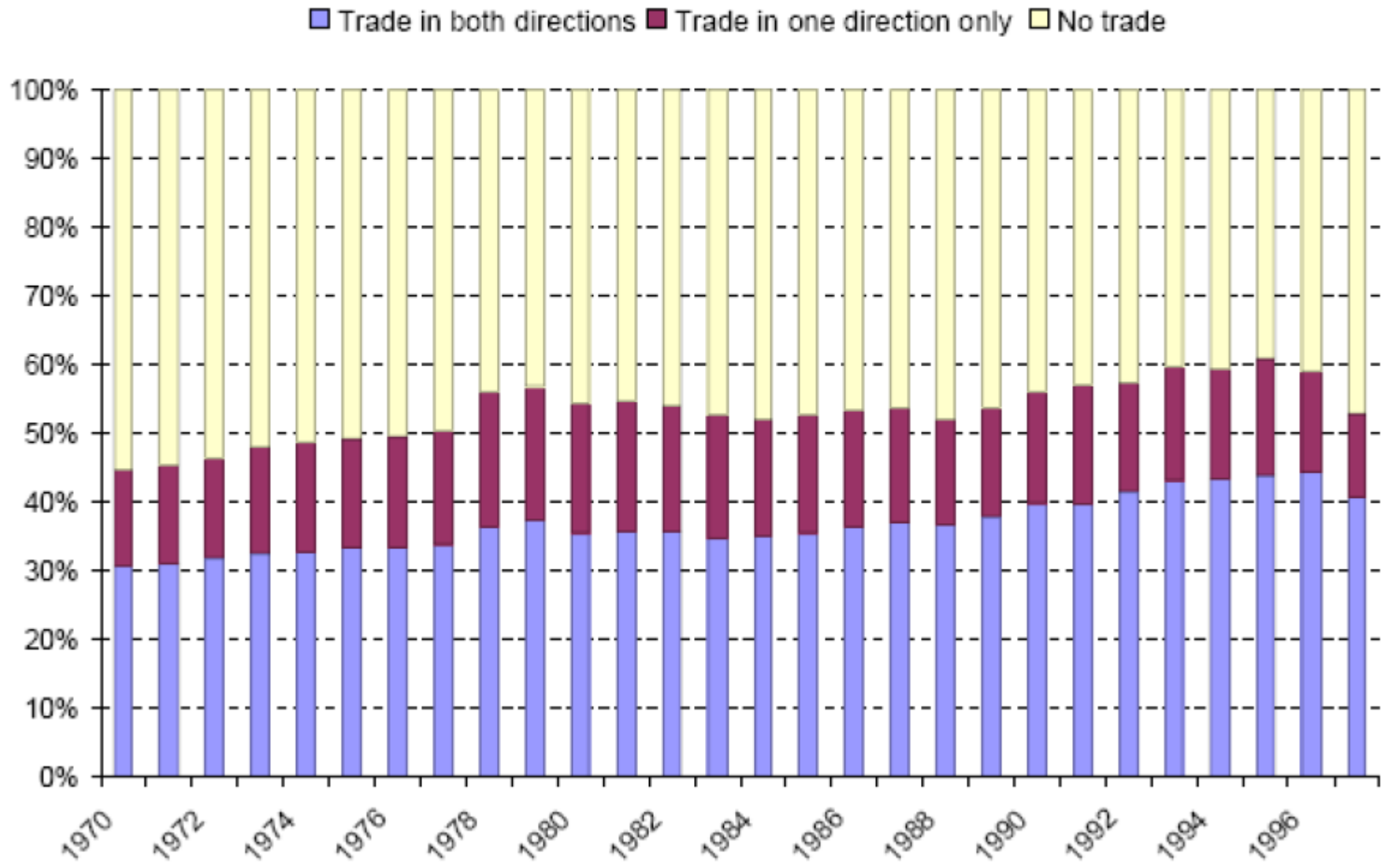
Partial R-squared of excluded instruments: 0.0212
Test of excluded instruments:
F(2, 519) = 9.13
Prob > F = 0.0001

IVs are strongly correlated with the potentially endogenous, as required.

```
Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on distw

Number of clusters (distw) = 520                  Number of obs =       858
                                                  F(  9,    519) =     17.64
                                                  Prob > F      =    0.0000
Total (centered) SS     =  13093.23428            Centered R2   =   -1.4615
Total (uncentered) SS   =  106004.5678            Uncentered R2 =    0.6960
Residual SS             =  32228.78661            Root MSE      =     6.129

------------------------------------------------------------------------------
             |               Robust
    ln_trade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
tariff_exp_sim |  -.8499463   .3305808    -2.57   0.010    -1.497873   -.2020198
tariff_imp_sim |  -.6322651   .3791732    -1.67   0.095    -1.375431    .1109008
    ln_gdp_exp |   .8106766   .1972706     4.11   0.000     .4240332    1.19732
    ln_gdp_imp |   .8022187   .0930646     8.62   0.000     .6198155    .984622
   ln_distance |  -2.408098   .5565373    -4.33   0.000    -3.498891   -1.317305
        contig |   1.395972   .9669588     1.44   0.149    -.4992327    3.291176
   comlang_off |  -1.629753   1.666283    -0.98   0.328    -4.895608    1.636102
        colony |  -.6709804   1.598405    -0.42   0.675    -3.803797    2.461836
        comcol |   2.353651   1.231581     1.91   0.056    -.0602033    4.767505
         _cons |   .8951502   13.75358     0.07   0.948    -26.06138    27.85168
------------------------------------------------------------------------------

Underidentification test (Kleibergen-Paap rk LM statistic):                7.254
                                                  Chi-sq(1) P-val =    0.0071
------------------------------------------------------------------------------

Weak identification test (Kleibergen-Paap rk Wald F statistic):            3.968
Stock-Yogo weak ID test critical values: 10% maximal IV size               7.03
                                         15% maximal IV size               4.58
                                         20% maximal IV size               3.95
                                         25% maximal IV size               3.63
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------

Hansen J statistic (overidentification test of all instruments):           0.000
                                                  (equation exactly identified)
-endog- option:
Endogeneity test of endogenous regressors:                                49.232
                                                  Chi-sq(2) P-val =    0.0000
Regressors tested:   tariff_exp_sim tariff_imp_sim
```

Ho: no endog. is rejected

Incidence of zero trade



Dealing with zero trade flows

- Problem of zero trade flows: Ln(0) = ?
  - Concern in particular large data samples (many counries and sectoral data)
  - Throwing observations
  - Ln(Xij + 0.0001)
  - Tobit with (Xij + 1) as a dependent variable (inconsistent estimator)

- Solutions:
  - Pseudo-Maximum Likelihood (PML).
  - Heckmann

ppml tradeMN ln_distance contig comlang_off colony comcol /// exp_dum* imp_dum*, cluster(dist)

```
Number of parameters: 88
Number of observations: 1224
Number of observations dropped: 0
Pseudo log-likelihood: -413143.41
R-squared: .95165482
                                        (Std. Err. adjusted for 818 clusters in distw)
```

tradeMN	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_distance	-.6658154	.0938784	-7.09	0.000	-.8498136	-.4818171
contig	.6362194	.1191128	5.34	0.000	.4027627	.8696762
comlang_off	.4758099	.1811773	2.63	0.009	.120709	.8309108
colony	-.3234539	.2873625	-1.13	0.260	-.8866741	.2397662
comcol	.1181997	.2092555	0.56	0.572	-.2919335	.5283328

Note:

1) we regress “trade”, NOT “ln_trade”.

2) R-squared increased by pick up significant features of the data.

- PPML is exactly equivalent to running a type of NLS on the original equation (to pick the heteroskedasticity of the type of multiplictive errors).

Heckman selection model
(regression model with sample selection)

Number of obs	=	2196
Censored obs	=	1205
Uncensored obs	=	991
Wald chi2(76)	=	4488.62
Prob > chi2	=	0.0000

Log pseudolikelihood = -2265.04

(Std. Err. adjusted for 1040 clusters in distw)

ln_trade	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_trade						
ln_distance	-2.464538	.1220474	-20.19	0.000	-2.703746	-2.225329
contig	.1439684	.3218181	0.45	0.655	-.4867834	.7747203
comlang_off	.5355219	.2255096	2.37	0.018	.0935312	.9775127
colony	-.0297133	.8099366	-0.04	0.971	-1.61716	1.557733
comcol	.8477501	.2242911	3.78	0.000	.4081476	1.287353
exp_dum_1	0	(omitted)				
exp_dum_2	.0283343	.6996024	0.04	0.968	-1.342861	1.39953
select						
ln_distance	.3450178	.1098436	3.14	0.002	.1297283	.5603074
contig	1.730593	.7114701	2.43	0.015	.3361375	3.125049
comlang_off	2.095941	.407317	5.15	0.000	1.297614	2.894268
colony	5.222824	.4534588	11.52	0.000	4.334061	6.111587
comcol	.8812726	.2314637	3.81	0.000	.4276121	1.334933
ent_cost_both	-.0001027	.0000289	-3.55	0.000	-.0001593	-.000046
exp_dum_1	-6.504016	.9829342	-6.62	0.000	-8.430532	-4.577501
/athrho	.0000667	.0867514	0.00	0.999	-.169963	.1700964
/lnsigma	.5310664	.0364328	14.58	0.000	.4596593	.6024734
rho	.0000667	.0867514			-.1683451	.1684747
sigma	1.700745	.0619629			1.583534	1.826631
lambda	.0001134	.147542			-.2890636	.2892903

Wald test of indep. eqns. (rho = 0): chi2(1) = 0.00 Prob > chi2 = 0.9994

Ho: rho = 0
No sample selection bias

Ho is accepted

ppml vs heckman

PPML

- Dealing with heteroskedasticity
- Assuming the zero and non-zero observations are produced by the same data generating process.

Heckman

- Dealing with sample selection bias
- Assuming the zero and non-zero observations are produced by different data generating process.

ARTNeT- KRI Capacity Building Workshop on Trade Policy Analysis:
Evidence-based Policy Making and Gravity Modelling for Trade Analysis
18-20 August 2015, Kuala Lumpur

Session 6: Wrap-up

Dr. Witada Aunkoonwattaka

Trade and Investment Division, ESCAP

anukoonwattaka@un.org

UNITED NATIONS ESCAP
Economic and Social Commission for Asia and the Pacific

ARTNeT

KHAZANAH RESEARCH INSTITUTE

Using gravity models for policy research

1. Make explicit reference to the theory and recent literature as much as possible
   - At least, appropriate dimensions of FE must be included in the gravity models.
2. OLS may not give an efficient estimates if
   - existence of zero trade flow
   - endogeneity problems
   - multi-dimension data
3. Make all possible attempts to correct the endogeneity problems.

Using gravity models for policy research

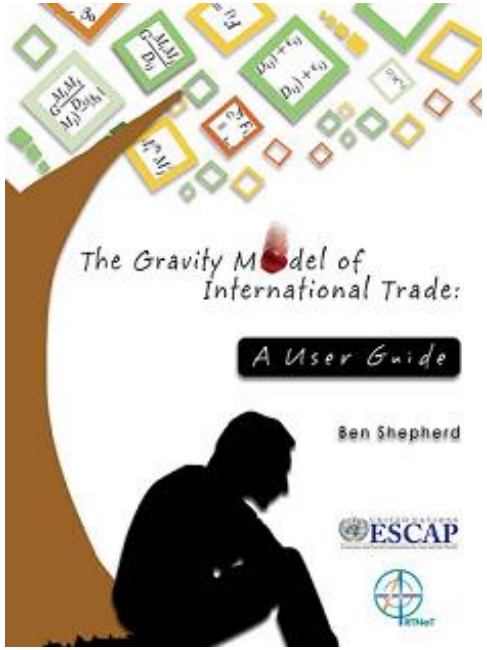
4. Asking questions that gravity can answer well
   - Gravity model should be used for describing sensitivity of trade flows to particular factors, not for discussing economic welfare
5. Being modest when using the gravity results in your analysis
   - Should not to be a primary tool for forecast / counterfactuals
   - Can give ideas about the relationships, but do not expect precise number
   - If an analysis really need number, do it for a range rather than a single number (±SD)
   - Treat the results as association, not causal links

Using the gravity model for policy research

4. Should apply gravity analysis in conjunction with other approaches such as CGE if discussing the general equilibrium implications (ie. welfare, resource allocations,…).

5. Research that does not use the latest models and techniques does not represent a sound basis for drawing policy conclusions.
   - Do robustness check to ensure results are robust to estimation using different techniques

Thank you and good luck



http://www.unescap.org/tid/publication/tipub2645.asp