INSTITUTE FOR ADVANCED STUDIES LUCCA

A Spatial Filtering Zero-Inflated Approach to the Estimation of the Gravity Model of Trade

Rodolfo Metulini, Roberto Patuelli, Daniel A. Griffith

ETSG 2015 - Paris

September 12, 2015

Contents

Introduction

- Non-linear estimation of the gravity model with PPML/ZIPML methods has become popular to model the international trade flows, as it permits to better account for large amount of zero flows
- Trade flows are not independent from each other due to spatial interaction. To overcome this problem, the eigenvector spatial filtering variants of the Poisson/Negative binomial (among various methods) has been proposed in the literature and is used here.

This paper contributes to the literature in two directions:

- First, by employing a stepwise selection criteria for spatial filters based on robust (sandwich) p-values.
- Second, by selecting a reduced set of spatial filters that properly account for importer-side and exporter-side specific spatial effects, at both stages of the process.

Traditional Gravity model

- The traditional gravity model for trade (Tinbergen,1962, Linneman, 1966) asserts that the volume between a country pair is proportional to the product of their GDPs and inversely related to a measure of distance.
- The log-linear specification along with the OLS estimation has been widely used in the empirical literature (Rose 2000, Franklen & Rose, 2002, Egger, 2002)
- Log-linearisation of the gravity model leads to inconsistent estimates in the presence of heteroscedasticity in trade levels (Helpman et al., 2008, Santos Silva & Tenreyro, 2006)
- Santos Silva and Tenreyro (2006, 2010, 2011) propose a Poisson specification of the gravity model along with the Poisson pseudo maximum likelihood (PPML) estimator.

Recent developments

We can roughly divide the available literature on incorporating spatial dependence and heterogeneity or network autocorrelation in gravity models into three main streams:

- **Linear spatial econometric models** (Baltagi et al. 2007; Fischer and Griffith 2008; LeSage and Pace 2008; Behrens et al. 2012; Koch and LeSage 2015): these models apply and adapt traditional (linear) spatial econometric techiques to the count data case by log-linearisation.
- **Spatial generalized linear models** (Lambert et al. 2010; Sellner et al. 2013): these models extend the previous approaches by allowing for estimation based on Poisson-type models, accommodating the concerns expressed in Santos Silva and Tenreyro (2006).
- **Non-parametric (eigenvector spatial filtering) models**(Chun 2008; Fischer and Griffith 2008; Scherngell and Lata 2013; Krisztin and Fischer 2015; Patuelli et al. 2015): these models take a non-parametric approach, by employing eigenvector spatial filtering within Poisson-type models.

Zero inflated gravity models of trade

- The ZINB ML gravity model provides a way to model gravity when the data generating process results in too many zeros (Martin and Pham, 2008, Burger et al., 2009, among others).
- Estimating the parameters by Poisson/Negative binomial ML methods would only be justified statistically if we believed that the trade flows were independent observations. Such an assumption is generally not valid: flows are spatially interconnected in nature.
- Several papers recently proposed to model spatial heterogeneity in the residuals by means of the three above-mentioned econometric techniques:

  (many focused on MTR, which can be considered as a main source of spatial heterogeneity: Behrens et al. 2012, Baier & Bergstrand 2009).
- Our approach implement **eigenvector spatial filters** to account for spatial heterogeneity in a **ZINB ML** framework, adopting ad ad-hoc **backward stepwise algorithm** to proper selecting the signficant spatial filters.

Spatial Filters

- Originally developed for area-based data by Griffith (2003), later extended to flow data (Chun 2008, Fischer and Griffth 2008, and Chun and Griffith 2011).
- Traditional advantage: including spatial filters (orthogonal and indipendent with the covariates), the model can be estimated by standard regression techniques, such as ordinary least squares (OLS) or Poisson regression.
- The first extracted eigenvector is the one showing the highest positive Moran I (MC, Cliff & Ord, 1972,1981). The subsequently extracted eigenvectors maximize MC while being orthogonal to and uncorrelated with the previously extracted eigenvectors.
- Usually, we select a subset of eigenvectors by means of the following threshold: $MC(e_i)/MC(e_1) > 0.25$.
- Since trade data are flows between points, the eigenvectors are linked to the flow data by means of Kronecker products. As a result, two sets of origin- and destination-specific filters are used (as in Patuelli et al. 2015).

Backward stepwise algorithm

- Used to select variables in a regression model (proposed by Efroymson, 1960)
- Usually, this takes the form of a sequence of F-tests or t-tests, but other techniques are possible, such as adjusted R-square, AIC, Bayesian information criterion (BIC), or simply based on p-values.
- Backward elimination: involves starting with all candidate variables, testing the deletion of each variable, deleting the variable (if any) that improves the model the most by being deleted, and repeating this process until no further improvement is possible.
- Backward elimination procedure in the framework of count models and zero inflated count models has been implemented in many routines. In R, the *be.zeroinfl* (*mpath*, Wang et al. 2015) function performs a backward elimination (and forward selection) stepwise **based on log likelihood**.

Model specification

- For estimation, we follow a standard specification of the gravity equation of bilateral trade.
- We employ some variables commonly mentioned in the literature (see, e.g., Frankel 1997; Raballand 2003).
- We use the following standard specification of the gravity equation and we estimate it, after a proper over-dispersion test, by mean of a Zero inflated negative binomial (ZINB) maximum likelihood:

$$\Pr(Tr_{ij}=1)=\alpha_1 dist_{ij}+\alpha_2 comcur+\alpha_3 contig+\alpha_4 hist+\alpha_5 fta+\beta_1 area_i+\beta_2 area_j+\beta_3 gdp_i+\beta_4 gdp_j+\beta_5 gdpcap_i+\beta_6 gdpcap_j+\beta_7 Island_i+\beta_8 Island_j+\beta_9 landl_i+\beta_{10} landl_j+\varepsilon_{ij}$$

$$Vol(Tr_{ij})=\alpha_1 dist_{ij}+\alpha_2 comcur+\alpha_3 contig+\alpha_4 hist+\alpha_5 fta+\beta_1 area_i+\beta_2 area_j+\beta_3 gdp_i+\beta_4 gdp_j+\beta_5 gdpcap_i+\beta_6 gdpcap_j+\beta_7 Island_i+\beta_8 Island_j+\beta_9 landl_i+\beta_{10} landl_j+\varepsilon_{ij}$$

Data

- The trade data are from the World Trade Database compiled on the basis of COMTRADE data by Feenstra et al. (2005).
- GDP and per capita GDP data are from the World Banks WDI database.
- Distance, language, colonial history, landlocked countries, and land area data are from the CEPII institute.
- Whether pairs of countries take part in a common regional integration agreement (FTA) has been determined on the basis of OECD data about major regional integration agreements.
- Data on island status have been kindly provided by Hildegunn Kyvik-Nordas (from Jansen and Nords 2004).

Details on stepwise algorithm

- We modified the *be.zeroinfl* function from *mpath* package (Wang et al. 2015.
- This function performs a **backward** stepwise on a **zero inflated model**, and it uses **p-values** as a model choice criteria.
- In each step, the function drops the variable with the largest p-value, indifferently of whether the variable is in the count (2nd step) or in the logit (1st step) part of the model. This stepwise process ends up when no more non significant variables are considered.
- We want to preserve all the non spatial variables as we consider them as benchmark: we set those variables as fixed, even if their estimated coefficient is not significant.

Estimation results - I

- Our model specification, based on different diagnostics, seems to outperforms the benchmark models (Spat filt NB ML and ZINB ML without filters). We based these diagnostics on Akaike information criteria and on the loglikelihood.
- In terms of AIC (slide 12), our spatial filters ZINB ML model have a smaller value (47026.13), meaning it performs better than the others (48370.13 for the ZINB, 48431.86 for the spatial filter NB).
- The same holds for the log likelihood: our model presents the smaller value (-2.32e+04 compared to -2.42e+04 for both the benchmark models).
- We compare the observed number of small flows with the estimated counterpart using our selected model, the ZINB and the NB with selected filters (slides 13 and 14): Spatial filtering ZINB predict 92 out of 484 zero flows, while Spatial Filters NB predict 19 zero flows, and ZINB without filters only predict 2 zero flows.

Estimation results - II

Table: Estimation results (1) ZINB ML with spatial filters (2) ZINB ML (3) NB ML with spatial filters. (1st step coeffs dropped here due to lack of space)

2nd step: count	**(1)**	**(2)**	**(3)**
Distance	-0.84***	-0.65***	-0.71***
Common language	0.46***	0.37***	0.44***
Contiguity	0.54***	0.71***	0.65***
Common history	0.77***	0.57***	0.81***
Free trade agreements	0.48***	0.66***	0.75***
Area exporter	-0.20***	-0.25***	-0.22***
Area importer	0.07***	-0.02	-0.04*
GDP per cap. exporter	-0.24***	-0.26***	-0.13***
GDP per cap. importer	0.16***	0.02	-0.12***
GDP exporter	1.06***	1.10***	0.99***
GDP importer	0.63***	0.75***	0.82***
Island exporter	0.43***	0.55***	0.30*
Island importer	-0.70***	-0.16	0.02
Landlocked exporter	-0.21***	-0.09	-0.44***
Landlocked importer	0.32***	-0.18*	0.35***
Costant	-28.71	-30.93	-29.8
# selected filters 1st step - exporter	11	-	-
# selected filters 1st step - importer	24	-	-
# selected filters 2nd step - exporter	11	-	8
# selected filters 2nd step - importer	8	-	12
Theta	0.859	0.726	0.595
AIC	**47026.13**	**48370.13**	**48431.86**
Log likelihood	-2.32E+04	-2.42E+04	-2.42E+04
Observations	4032	4032	4032

Estimation results- III

Table: Count of observed versus predicted values. Model comparison

Trade flow	0	1	2	3	4	5	6	7	8	9
Observed	484	136	112	76	64	39	42	49	35	29
Spat Filt ZINB	92	59	46	42	46	39	45	34	31	38
ZINB	2	22	23	31	31	37	32	33	26	29
Spat filt NB	19	42	39	45	44	35	37	41	35	38

Estimation results- IV

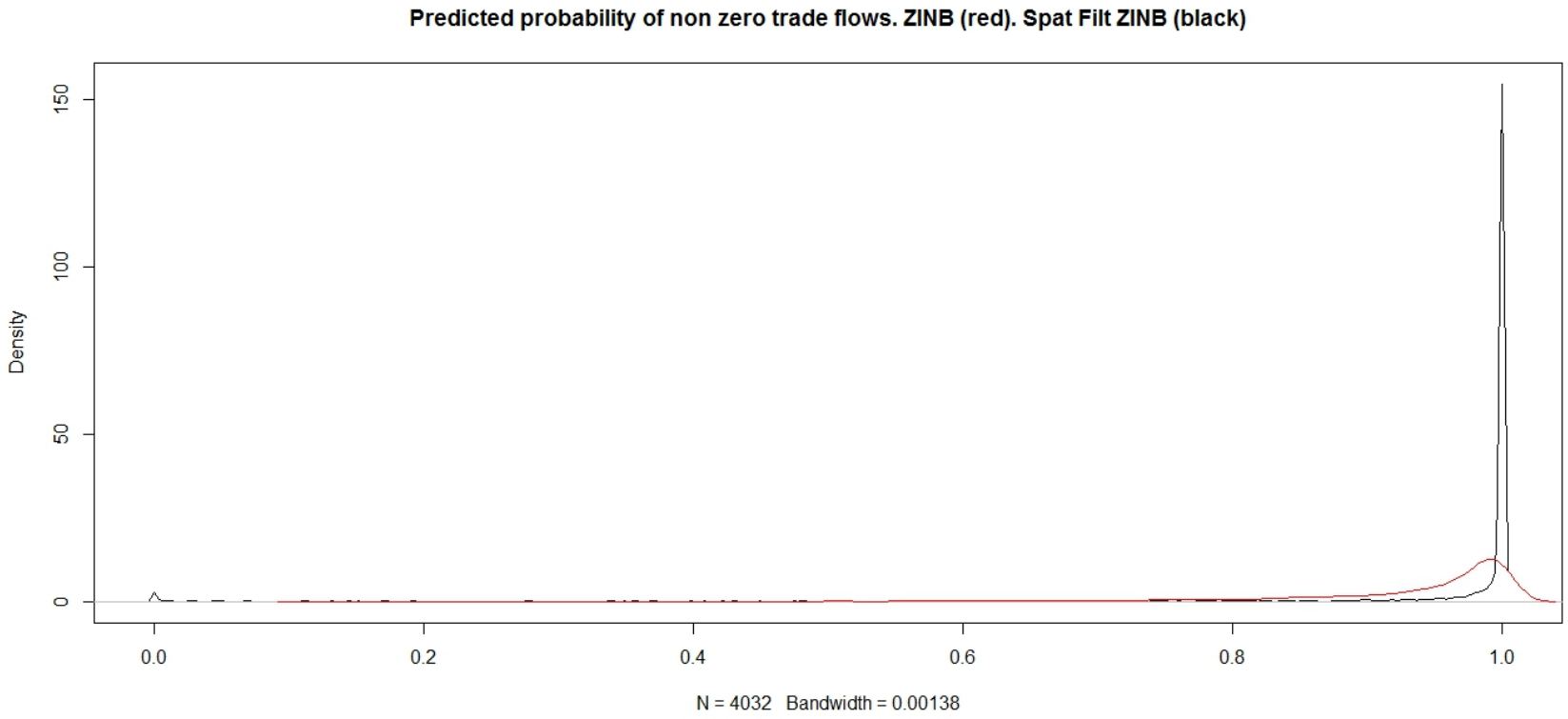

Figure: Estimated probability of trade for non-zero trade flows observations. ZINB(red) vs. Spat Filt ZINB(black)

Estimation results- V

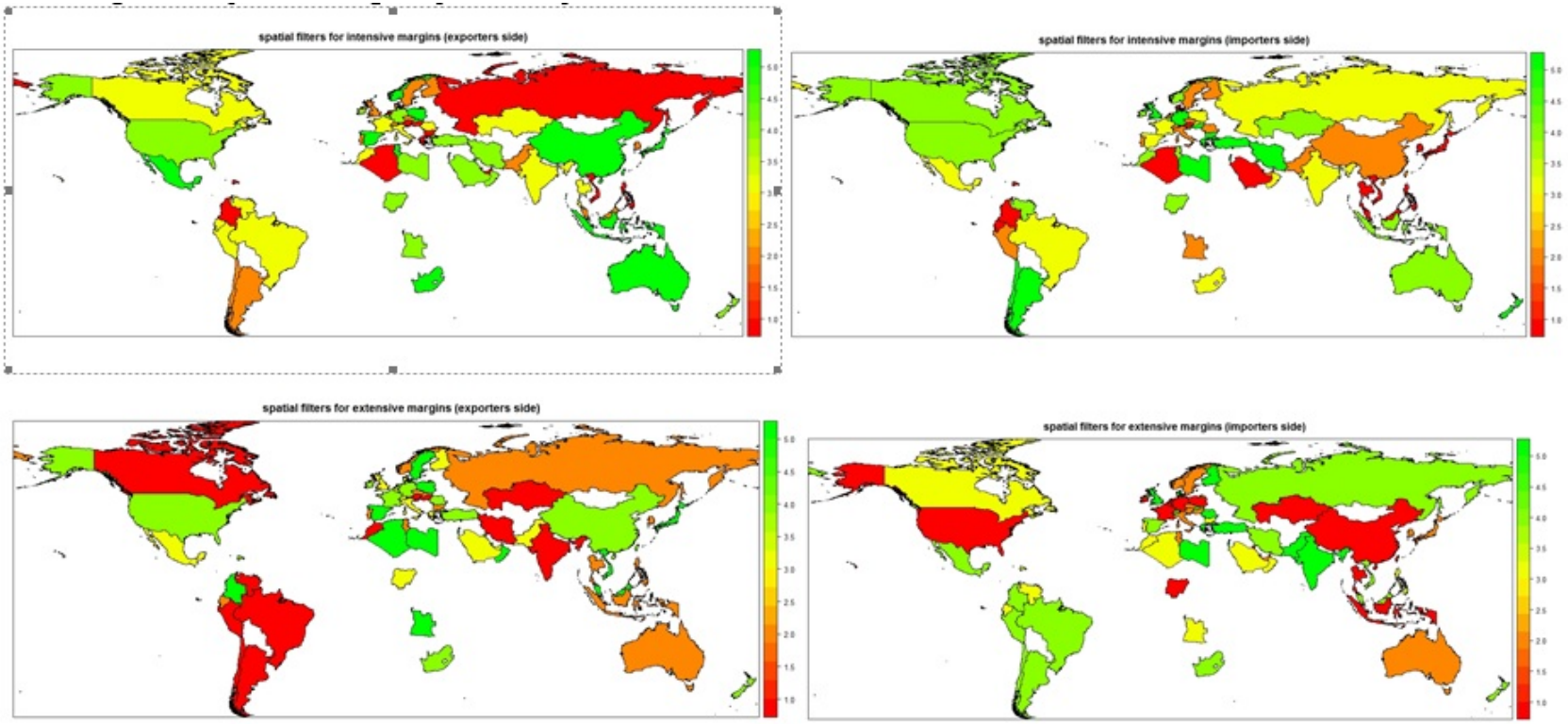

Figure: Mapped spatial filters (all selected filters weighted by the model coefficients).

Top-left: 2st step exporters (Moran I: 0.046. Selected filters: 11)
Top-right: 2nd step importers (Moran I: 0.160**. Selected filters: 8)
Bottom-left: 1st step exporters (Moran I: 0.298***. Selected filters: 11)
Bottom-right: 1st step - importers (Moran I: 0.036. Selected filters: 24)

Conclusions

- We employ a backward stepwise selection criteria for spatial filtering ZINB ML, based on robust p-values.
- Second, we select a reduced set of spatial filters that properly account for importer-side and exporter-side specific spatial effects, both at the count and the logit process.
- Results highlighted that our specification outperforms the benchmark models (ZINB and NB with spatial filters), in terms of model fitting, both considering the AIC and the log likelihood, and in predicting non zero (and small) flows.

Future directions

- Future research directions include comparing this model with other ZINB ML model specifications that alternatively account for SAC (Spatial generalized linear models, as in Lambert et al. 2010, Sellner et al. 2013).
- A caveat regards evaluating the contribution of the logit and the count part of the process, in terms of the explained variance of trade flows.
- Moreover, a similar analysis, making appropriate changes, can be applied to a panel data framework.

Main references

- Patuelli R, Linders G-J, Metulini R, Griffith DA (2015) **The Space of Gravity: Spatial Filtering Estimation of a Gravity Model for Bilateral Trade**. In: Patuelli R, Arbia G (eds) Spatial Econometric Interaction Modelling. Advances in Spatial Science. Springer, Berlin, p. (forthcoming)
- Krisztin T, Fischer MM (2015) **The Gravity Model for International Trade: Specification and Estimation Issues**. Vienna University of Economics and Business, Vienna
- Zhu Wang, Shuangge Ma and Ching-Yun Wang (2015) **Variable selection for zero-inflated and overdispersed data with application to health care demand in Germany**, Biometrical Journal. Article first published online June 8, 2015

backward stepwise algorithm function

usage
be.zeroinfl.filt = function (object, data, dist = c("poisson", "negbin", "geometric"), alpha = 0.05, trace = FALSE, subset.zero, subset.count, minmod.zero, minmod.count)

Details
conduct backward stepwise variable elimination for zero inflated count regression from zeroinfl function, providing a possibility to define a minmodel and using robust standard errors.

Value
an object of zeroinfl with all variables having p-values less than the significance level alpha

Arguments
Object: an object from function zeroinfl
Data: argument controlling formula processing via model.frame
Dist: one of the distributions in zeroinfl function
Alpha: significance level of variable elimination
Trace: logical value, if TRUE, print detailed calculation results
subset.zero: is a list of the variable names to be subsetted in the zero component
subset.count: is a list of the variable names to be subsetted in the count component
minmod.zero: is a list of the variable names to do not subset in the zero component

minmod.count: is a list of the variable names to do not subset in the count component

Thank you!

r.metulini@sssup.it