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

Rodolfo Metulini,[a] Roberto Patuelli,[b] Daniel A. Griffith[c]

[a] Institute of Economics (LEM), Scuola Superiore Sant'Anna, Italy

[b] Department of Economics, University of Bologna, Italy; The Rimini Centre for Economic Analysis (RCEA), Italy

[c] School of Economic, Political & Policy Sciences, The University of Texas at Dallas, USA

ABSTRACT

Nonlinear estimation of the gravity model with Poisson/negative binomial methods has become popular to model international trade flows, as it permits a better accounting for large numbers of zero flows. Nevertheless, as trade flows are not independent of each other due to spatial autocorrelation, those methods lead to biased parameter estimates. To overcome this problem, eigenvector spatial filtering variants of the Poisson/Negative binomial specification has been proposed in the literature of gravity modelling of trade. This paper contributes to the literature in two ways. First, by employing a stepwise selection criterion for spatial filters which is based on robust (sandwich) p-values and does not require likelihood-based indicators. In this respect, we develop an *ad hoc* backward stepwise function in R. Second, using this function, we select a reduced set of spatial filters that properly accounts for importer-side and exporter-side specific spatial effects, both at the count and the logit process. Applying this estimation strategy to a cross-section of bilateral trade flows between a set of worldwide countries for the year 2000, we find that our specification outperforms the benchmark models, in terms of model fitting, both considering the AIC and in predicting zero (and small) flows.

JEL codes: C14, C21, F10

Keywords: bilateral trade; unconstrained gravity model; eigenvector spatial filtering; zero flows; backward stepwise.

1. Introduction

A traditional gravity model for trade in its simple form (Tinbergen 1962; Linnemann 1966) asserts that the volume of trade between a country pair is proportional to the product of their gross domestic products and inversely related to a measure of distance separating them, where distance is broadly defined as a function of several variables that can be viewed as trade resistance factors. The log-linear specification of the gravity model along with ordinary least square (OLS) estimation has been widely used in the empirical literature (Rose 2000; Egger 2002; Frankel and Rose 2002) mostly because of its good empirical performance, but, in later years, for the strong theoretical foundations provided in papers such as Anderson (1979), and Anderson and van Wincoop (2003). However, most recent contributions stressed the fact that zero trade flows are to be specifically taken into account. Helpman et al. (2008) proved that disregarding countries that do not trade with each other generates biased estimates. Moreover, Santos Silva and Tenreyro (2006) have shown that log-linearization of the gravity model leads to inconsistent estimates in the presence of heteroscedasticity in trade levels. They propose a Poisson-type specification of the gravity model along with the Poisson pseudo-maximum likelihood (PPML) estimator, somehow similarly to the Poisson approach firstly proposed in Flowerdew and Aitkin (1982). Santos Silva and Tenreyro (2006; 2011) also provide simulation evidence that the PPML estimator is well behaved even when the conditional variance is far from being proportional to the conditional mean. Nowadays, several empirical studies of trade have applied the PPML estimator (see Linders et al. 2008; Burger et al. 2009; Westerlund and Wilhelmsson 2009; Martínez-Zarzoso 2013; Martin and Pham 2015; Patuelli et al. 2016). In order to correct for overdispersion, a negative binomial (NB) regression model is frequently employed, which belongs to the family of Poisson models, and allows for the dispersion parameter to differ from 1.

The zero-inflated model (Lambert 1992; Greene 1994; Long 1997), applied to negative binomial models (ZINB), permits a better estimate of a huge amount of zero flows, because it considers the existence of two groups within the population: the first having strictly zero counts and the second having a non-zero probability of having a trade flow greater than zero.

Burger et al. (2009) stressed the fact that some variables may be more important in determining the profitability of bilateral trade (decision to trade) rather than the potential volume of bilateral trade. However, so far, it is not so clear which variables determine the decision to trade.

Moreover, it is now well known that trade flows are not independent of each other (Griffith 2007; LeSage and Pace 2008) and that possible sources of spatial autocorrelation (SAC) among countries should be taken into account (Behrens et al. 2012; Sellner et al. 2013). With this paper we aim to better analyse the dynamic of the decision to trade (intensive margin) and the volume of trade (extensive margin), and, in particular, what is the contribution of SAC in both of these two processes. We focus on an eigenvector spatial filtering (ESF) approach (Griffith 2003), within a ZINB framework, using two sets of origin and destination spatial filters (Griffith 2007; Fischer and Griffith 2008), one accounting for the SAC in the logit process, and the other accounting for the SAC in the count process. In this regard, we prepared an *ad hoc* function which applies a backward stepwise algorithm which aims to properly detect the significant filters. The algorithm we propose has the advantage that, at each step, it drops the filter with the larger p-value, indifferently if this filter is in the count or in the logit process. We compare the results of this estimation with two benchmarks, namely, the ZINB and the NB with origin and destination spatial filters, and with the quasi-ML Poisson ESF. We conducted a comparison in terms of estimated coefficients and in terms of the goodness of fit (Akaike information Criteria, AIC and prediction of the zero and small flows). We found that our specification outperforms the comparison models, both in terms of AIC and in terms of prediction of small trade flows.

This paper is structured as follows. Section 2 presents a review of the gravity of trade, from the traditional models to recent developments. In Section 3 we define our proposed model and the stepwise algorithm we adopted. Section 4 presents the empirical application, together with results. Section 5 concludes the paper.

2. The Gravity Model of Trade: Recent Developments

The scientific community has recently gained a renewed interest in both the theoretical and empirical aspects of the gravity model of trade. In particular, the aforementioned theoretical developments on multilateral resistance terms generated the need for consistent estimation approaches that would fit such advancements. The vastly increased computational power available for econometric analysis played an additional role, allowing more complex and data-intensive (i.e., nonlinear and panel) estimation efforts.

Several studies, starting, for example, from the popular paper by Santos Silva and Tenreyro (2006), have pushed the envelope of the field, and a number of researchers are actively pursuing further methodological advancements pertaining in particular to the estimation of the gravity model of trade. Egger and Tarlea (2015) propose a multi-way clustering approach to consistently estimate regression coefficients pertaining to preferential trade agreements. Egger and Staub (2015) compare the suitability of various estimation approaches under an international economics general equilibrium perspective. Baltagi and Egger (2015) develop a quantile regression structural estimation solution for the gravity model.

Within the aforementioned econometric developments, a niche of its own is emerging pertaining to the incorporation of spatial dependence and heterogeneity or network autocorrelation (i.e., the correlation of flow data based on its network's topological characteristics) in gravity models, trade still being a frequent application. While the relevance of spatial correlations was originally suggested for trade in Anderson and van Wincoop (2004), and much earlier within spatial interaction modelling (Curry 1972; Curry et al. 1975; Sheppard et al. 1976), it is only in recent years that this issue attracted significant attention, thanks to recent developments in econometrics and computational power. Studies by Behrens et al. (2012) (available online for several years before actual publication), Fischer and Griffith (2008) and LeSage and Pace (2008) provided, from different perspectives (economic theory, spatial econometrics, spatial statistics), the necessary stepping stones for analysing SAC aspects in flow data. We can roughly divide the available literature 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 techniques to the count data case.
- 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, therefore accommodating the concerns expressed in Santos Silva and Tenreyro (2006).
- Non-parametric (ESF) models (Chun 2008; Fischer and Griffith 2008; Scherngell and Lata 2013; Krisztin and Fischer 2015; Patuelli et al. 2016): these models take a non-parametric approach, by employing ESF within Poisson-type models.

This paper is concerned with this latest class of models. ESF (Griffith 2003) (described in more detail below in Section 3.2) is a spatial statistics technique based on the decomposition of spatial weights matrices. It is considered here because it allows for greater flexibility in modelling, and can seamlessly be applied to any estimation framework. The available studies employing this technique demonstrate how (sets of) spatial filters can be successfully used at the intercept level as ‘interceptors’ (i.e., proxies for) unobserved spatial heterogeneity. This paper aims to further investigate the use of ESF, by allowing for separate spatial filters sets in zero-inflated models. The proposed approach is described in the next section.

3. Methodological Approach

3.1 Zero-Inflated Gravity Models of Trade

In recent years, it has become increasingly recognized that the level of trade between countries is frequently zero. Small countries may not have trade relations with all possible trading partners or because statistical offices do not report trade flows below a certain threshold. Moreover, the issue of zero flows is more pronounced when analysing sector disaggregate trade flows. The ZINB gravity model provides one way to model an excess number of zero flows. Martin and Pham (2015) and Burger et al. (2009) have proposed the zero-inflated extension of the Poisson gravity model for situations where the data generating process results in too many zeros. The model may be viewed as a "two-part" extension, in which the distribution of the outcome is approximated by mixing two component distributions. The zero-inflated part of the model consists of a qualitative-dependent model to determine the probability of whether a particular origin-destination trade flow will be zero or positive. The second part contains the standard Poisson gravity model to estimate the relationship between trade flows and explanatory variables for each trade flow that has a non-zero probability (Leung and Yu 1996). Among others, Xiong and Beghin (2012) and Philippidis et al. (2013) applied zero-inflated count models for the analysis of international trade.

Estimating the parameters of the negative binomial gravity model (without or with zero-inflation) by pseudo-maximum likelihood methods would only be justified statistically if we believed that the trade flows were independent observations. Such an assumption, however, is generally not valid, because flows are fundamentally spatial in nature. Several papers recently proposed modelling the spatial heterogeneity in the residual by means of different econometric techniques. Among those works, many focused on the issue of multilateral trade resistances (MTR), which can be considered as a main source of spatial heterogeneity (Baier and Bergstrand 2009; Behrens et al. 2012). One way to relax this independence assumption is by incorporating spatial dependence into the Poisson version of the gravity model by means of spatial autoregressive techniques (Lambert et al. 2010; Sellner et al. 2013). Another is ESF, originally developed for area-based data (Griffith 2003) and later extended to flow data (Chun 2008; Fischer and Griffith 2008; Chun and Griffith 2011). In their recent work, Patuelli et al. (2016) applied spatial filters with NB, as a way to filter out spatial heterogeneity due to MTRs. However, residual heterogeneity should be present both for the logit and the count process, while the above mentioned works only account for SAC for the count process. Krisztin and Fischer (2015) have very recently applied network-autocorrelation SFs to a trade model, by including, among others, zero-inflated specifications. This work follows a similar approach used by Krisztin and Fischer, but we introduced an *ad hoc* backward stepwise procedure to properly select the filters. Moreover, we perform some diagnostics to: i) compare our model with other benchmarks, and ii) evaluate the fitting of our specification in predicting zero (and small) trade flows.

3.2 Spatial Filters

SF has SF has been originally developed for area-based data by Griffith (2003), and later extended to flow data (Chun 2008; Fischer and Griffith 2008; Chun and Griffith 2011). One traditional advantage when including these spatial filters as additional origin- and destination-specific regressors is that the model can be estimated by standard regression techniques, such as OLS or Poisson regression, which are common in the literature about spatial interaction patterns. The parameters of the standard regressor variables are unrelated to the remaining residual term, and standard estimation yields consistent parameter estimates as a result. We refer to this estimation method as SF estimation of origin-destination models.

The workhorse for this SF decomposition is a transformation procedure based upon eigenvector extraction from the matrix

$$(\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/n)\ \mathbf{W}\ (\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/n) \tag{1}$$

where $\mathbf{W}$ is a generic n x n spatial weights matrix; $\mathbf{I}$ is an n x n identity matrix; and, $\mathbf{1}$ is an n x 1 vector containing 1s. The spatial weights matrix $\mathbf{W}$ defines the relationships of proximity between the n georeferenced units (e.g., points, regions, and countries). The transformed matrix appears in the numerator of the Moran I coefficient (MC).

The eigenvectors of Equation (1) represent distinct map pattern descriptions of SAC underlying georeferenced variables (Griffith 2003). Moreover, the first extracted eigenvector, say e_1, is the one showing the highest positive MC (Cliff and Ord 1972; 1981) that can be achieved by any spatial recombination induced by $\mathbf{W}$. The subsequently extracted eigenvectors maximize MC while being orthogonal to and uncorrelated with the previously extracted eigenvectors. Finally, the last extracted eigenvector maximizes the negative MC.

Having extracted the eigenvectors of Equation (1), a spatial filter is constructed as a linear combination of a judiciously selected a subset of these *n* eigenvectors. In detail, for our empirical application, we select a first subset of eigenvectors (which we will call the 'candidate eigenvectors') by means of the following threshold: MC(e_i)/MC(e_1) > 0.25. This threshold yields a spatial filter that approximately replicates at least the amount of variance explained by a pure spatial autoregressive model (SAR) (Griffith 2003).[1] Subsequently, a stepwise regression model may be employed to further reduce the first subset (whose eigenvectors have not yet been related to the data) to just the (smaller) subset of eigenvectors that are statistically significant as additional regressors in the model to be evaluated. The resulting group of eigenvectors is what we call our 'spatial filter'. This estimation technique has been applied to various fields, including labour markets (Patuelli 2007), innovation (Grimpe and Patuelli 2011), economic growth (Crespo Cuaresma and Feldkircher 2013) and ecology (Monestiez et al. 2006).

Because trade data do not represent points in space, but flows between points, the eigenvectors are linked to the flow data by means of Kronecker products: the product $E_K \otimes \mathbf{1}$, where E_K is the n x k matrix of the candidate eigenvectors, may be linked to the origin-specific information (for example, GDP per exporting countries), while the product $\mathbf{1} \otimes E_K$ may be linked to destination-specific information (again, for example, the GDP of importing countries) (Fischer and Griffith 2008). As a result, two sets of origin- and destination-specific variables are used (Patuelli et al. 2016), which aim to capture the SAC patterns commonly accounted for by the indicator variables of a doubly-constrained gravity model (Griffith 2009), therefore avoiding omitted variable bias.

The new challenge here is that we want to account for SAC in both the logit and in the count processes, so we used two sets of filters at the logit level and (the same) two sets of filters at the count level.

3.3 Backward Stepwise Algorithm

The stepwise procedure is an algorithm used to choice the variables in a regression model, and it has been proposed by (Efroymson 1960) usually, this takes the form of a sequence of F-

[1] Ongoing research by Griffith and collaborators is looking into formulating an estimation equation, based on residual SAC, to predict the ideal size of the candidate set.

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.

Forward selection involves starting with no variables in the model, testing the addition of each variable using a chosen model comparison criterion, adding the variable (if any) that improves the model the most, and repeating this process until no more improvements are possible. The backward elimination, which involves starting with all candidate variables, testing the deletion of each variable using a chosen model comparison criterion, deleting the variable (if any) that improves the model the most by being deleted, and repeating this process until no further improvements are possible. A backward elimination procedure in the framework of count models and zero inflated count models has been implemented in many routines. Chun and Griffith (2013) list R code for stepwise selection for GLMs. In the *mpath* package (Wang et al. 2015), the *be.zeroinfl* function performs a backward elimination (and forward selection) stepwise based on maximum likelihood criteria.

Here, we are interested in using a stepwise algorithm to define the proper set of eigenvectors to include in the regression model in order to account for SAC.

Although the algorithm is inspired by *be.zeroinfl* code but, it has at least two advantages. First, at each step of the algorithm, we compute robust standard errors and we select the variable to be removed out based on the robust p-values. Second, the algorithm is constructed in order to be able to drop the variables with the biggest p-values, regardless of whether if they come from the count or the logit process. We also structured the function so that a *minmodel* can be defined. In other words, we let the algorithm drop only the eigenvectors, because we consider the non-spatial explanatory variables to have substantive meaning.

4. Empirical Application

4.1 Data and 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 by mean of a ZINB:

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

$$Vol(Tr_{ij}) = \alpha_1 dist_{ij} + \alpha_2 comcur + \alpha_3 contig + \alpha_4 hist + \alpha_5 fta + \beta_1 area_i + \beta_2 area_j + \beta_3 gdp_i + \beta_4 gdp_j + \beta_5 gdpcap_i + \beta_6 gdpcap_j + \beta_7 Island_i + \beta_8 Island_j + \beta_9 landl_i + \beta_{10} landl_j + \varepsilon_{ij}, \tag{2}$$

where *Tr* represents trade flows, *gdp* represents the gross domestic product, *gdpcap* represents per capita GDP, *Island* is an indicator variable that equals 1 if a country is an island, *Area* is the land area of a country, and *landl* equals 1 for landlocked countries. The other variables are country pair related dummies, and they indicate if a pair share a currency (*comcur*), a common border (*contig*), are in free trade agreements (*fta*) and/or history (*hist*) as well as a measure of geographical distance between them (*dist*).

The data for trade 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 Bank's WDI database. Distance, language, colonial history, landlocked countries, and land area data are from the CEPII institute.[2] 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.[3] A dummy variable indicates whether a pair of countries has (membership in) at least one common FTA. Data on island status have been kindly provided by Hildegunn Kyvik-Nordas (from Jansen and Nordås 2004).

4.2 Details on the Stepwise Algorithm

We modified *be.zeroinfl* function from *mpath* package (Wang et al. 2015). This function performs a stepwise on a zero inflated model, and it uses p-values as a model choice criteria[4]. In each step, the function drops the variable with the biggest p-value, indifferently if the variable is in the count or in the zero process. This stepwise process ends up when no more non-significant variables are considered. However, *be.zeroinfl* do or permit to define a so called *minmodel,* in other words, this function selects the variable to be dropped among the full list of explanatories considered in the model. Since we want to preserve all the non-spatial variables as we consider them as "benchmark", we want to set them as nonerasable, even if their estimated coefficient is not significant. Moreover, we desire a more robust algorithm, so, we use, in each step of the process, the p-values from robust estimation of standard errors instead from simple standard errors. Having defined in details which is the desired stepwise

[2] See http://www.cepii.fr.

[3] See http://www.oecd.org/dataoecd/39/37/1923431.pdf.

[4] Other stepwise algorithms use Akaike information criteria (AIC) and Bayesian information criteria (BIC) in place of the p-Value to choose the variable to be dropped.

algorithm specification. We called this function *be.zeroinfl.filt.robust*. The detailed description and the full code are reported in Appendix A.1.

4.3 Estimation Results

We estimated the model in equation (2) using ZINB with selected spatial filters to a cross section of totally 4032 country pairs, for the year 2000 (first column, table 1). We also estimated the same model using two benchmarks: a ZINB without spatial filters (second column, table 1), and an NB with selected spatial filters (that only models the second line of equation (2), third column, table 1). First, we describe the estimated coefficients.

Table 1. Estimation results for: (1) ZINB ESF; (2) ZINB; (3) NB ESF

	(1)	(2)	(3)
	ZINB ESF	ZINB	NB ESF
First Step			
Distance	–1.18***	0.35**	
Common language	–1.34*	0.51*	
Contiguity	1.85*	0.24*	
Common history	–2.28	–1.83	
Free trade agreements	–0.86	–1.71**	
Area importer	2.77***	–0.07	
Area exporter	0.233*	0.36***	
GDP per cap. importer	–0.69**	–0.61***	
GDP per cap. exporter	0.97***	0.47***	
GDP importer	–5.25***	–0.11	
GDP exporter	–2.82***	–1.35***	
Island importer	17.04***	–0.98	
Island exporter	–1.97	–2.39**	
Landlocked importer	39.70***	0.24	

Landlocked exporter	3.00 ***	–1.50***	
Constant	126.48***	28.43***	
Eigenvectors (exp)	11	–	–
Eigenvectors (imp)	24	–	–
Second Step			
Distance	–0.84***	–0.65***	–0.71***
Common language	0.46***	0.37***	0.42***
Contiguity	0.54***	0.71***	0.66***
Common history	0.77***	0.57***	0.83***
Free trade agreements	0.48***	0.66***	0.77***
Area importer	–0.20***	–0.25***	–0.23***
Area exporter	0.07***	–0.02	–0.03*
GDP per cap. importer	–0.24***	–0.26***	–0.14***
GDP per cap. exporter	0.16***	0.02	–0.10***
GDP importer	1.06***	1.10***	1.00***
GDP exporter	0.63***	0.75***	0.81***
Island importer	0.43***	0.55***	0.34*
Island exporter	–0.70***	–0.16	-0.02
Landlocked importer	–0.21***	–0.09	–0.27***
Landlocked exporter	0.32***	–0.18*	0.37***
Constant	–28.71	–30.93	–30.06
Number significant eigenvectors exporter	11	–	8
Number significant eigenvectors importer	8	–	12
Theta	0.86	0.73	0.59
AIC	47026.13	48370.13	48414

Log-likelihood	2.32e+04	2.42e+04	2.42e+04
McFadden's pseudo-R^2	0.1312	0.1022	0.1022
Observations	4032	4032	4032
Residual dof	3945	3999	3995

Looking to the count process (second step) of the spatial filters ZINB, distance have a negative significant effect, the country-pair dummies all present positive and significant coefficients and dimension, and country mass variables a positive one, as expected: GDPs positively affect the trade flow, both at the exporter country and at the importer country side. Both the area and the GDPs per capita of the exporter country negatively affect trade, while, the importing countries values do not affect trade flows. In all, coefficients on the count process do not change a lot among different model specification, and all the coefficients that are significant with the spatial filters ZINB, are still significant using our benchmarks.

Considering the first step (logit part) the probability of a country pair to be involved in trade negatively depends on the distance, positively depends on the importer and exporter country areas, but, surprisingly, negatively depends on the GDPs, considering our model specification in Column (1). Moreover, the estimated coefficient resulting from the alternative model specification [ZINB, Column (2)] are not always similar to the ones in column 1. These results highlight the need to better analyse the determinants of trade decision.

Results are in favour of our model specification, which, based on different diagnostics, seems to outperform the models we use as benchmark. We based these diagnostics on AIC and on the log-likelihood.

In terms of AIC, our SF ZINB model has a smaller value (47026), meaning it performs better than the others (48370 for the ZINB, 48414 for the SF 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).

However, we also report Poisson estimation results in Appendix A.2. Looking to the results in table A.1, we can conclude that Zero Inflated Poisson (ZIP) ESF outperforms the two benchmark models (Poisson ESF and ZIP) in terms of both likelihood and AIC.

In addition, we analyse the robustness of our model in terms of fitting of small trade flows. We compare the observed frequencies of small flows (rounded to integer) with their estimated counterpart using our selected model, the ZINB and the NB with selected filters. Because one advantage of our model specification is that it should better predict small flows, we expect that our model, which both accounts for a logit part and for SAC, outperforms the two benchmarks. Results reported in Table 2 confirm this expectation. ZINB ESF predicts 440 out of 484 zero flows, whereas NB ESF predicts 281 zero flows. ZINB without filters predicts 474 zero flows, but it is less efficient in predicting small flows, compared to ZINB ESF. Moreover, results reporting predictions of small flows using Poisson models are reported in Appendix A.2, Table A.2. Despite ZIP models perfectly predict zero flows, both them and the non-zero-inflated Poisson model lack efficiently in predicting small flows larger than zero. In this respect, negative binomial models outperform Poisson models.

Table 2. 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
ZINB ESF	440	88	75	66	59	54	50	46	43	40
ZINB	474	86	71	62	55	50	47	43	41	38
NB ESF	281	156	117	95	82	72	64	58	53	49

Going more in detail along the spatial part of the model, using the SF ZINB we select in the first step 11 exporter-side and 24 importer-side eigenvectors. In the second step, the number of significant eigenvectors is 11 for the exporter countries and 8 for the importer countries.

An MC test can then be carried out on the weighted (based on estimated coefficients) mean of the selected eigenvectors, separately for each of the four sets of selected filters. The filter including the largest number of significant eigenvectors (24) appears to be the one with the lowest MC (0.036). The sets of eigenvectors with the highest MC values are the second-step importer (MC 0.160, 8 eigenvectors) and the second-step exporter (MC = 0.298, 11 eigenvectors). The relationship between the number of eigenvectors selected and the strength of the proxied SAC appears to require further investigation, in order to better interpret the modelled patterns and educate expectations about the number of degrees of freedom to be used for the computation of spatial filters.

4. Conclusions

The eigenvector spatial filtering (ESF) variants of nonlinear gravity models of trade (such as Poisson or negative binomial) have been proposed in the literature, because trade flows are not independent but contain) spatial autocorrelation (SAC). Using a zero-inflated negative binomial (ZINB) approach, this paper contributed to the existing literature in two ways. First, we developed a zero-inflated stepwise selection procedure for constructing spatial filters based on robust p-values. Second, we identified spatial filters that properly account for importer- and exporter-side specific unexplained spatial patterns, both for the count and the logit process. Results applied to a cross-section of bilateral trade flows between a set of worldwide countries showed that our specification outperforms the benchmark models (ZINB and NB with spatial filters) in terms of model fitting, both considering AIC and log-likelihood, and in predicting zero (and small) flows.

Future research should compare this model with further ZINB specifications that alternatively account for SAC, and evaluate the contribution of the logit and the count parts of the model in terms of explained variance. Moreover, a similar analysis, taking care of appropriate changes, should be applied to a panel data framework, so to evaluate, for example, trade-offs between the spatial filters and .

References

Anderson JE (1979) A Theoretical Foundation for the Gravity Equation. The American Economic Review 69 (1):106-16

Anderson JE, van Wincoop E (2003) Gravity with Gravitas: A Solution to the Border Puzzle. American Economic Review 93 (1):170-92

Anderson JE, van Wincoop E (2004) Trade Costs. Journal of Economic Literature 42 (3):691-751

Baier SL, Bergstrand JH (2009) Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects Using the Gravity Equation. Journal of International Economics 77 (1):77-85

Baltagi BH, Egger P (2015) Estimation of Structural Gravity Quantile Regression Models. Empirical Economics (forthcoming)

Baltagi BH, Egger P, Pfaffermayr M (2007) Estimating Models of Complex FDI: Are There Third-Country Effects? Journal of Econometrics 140 (1):260-81

Behrens K, Ertur C, Koch W (2012) 'Dual' Gravity: Using Spatial Econometrics to Control for Multilateral Resistance. Journal of Applied Econometrics 25 (2):773-94

Burger M, van Oort F, Linders G-J (2009) On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation. Spatial Economic Analysis 4 (2):167-90

Chun Y (2008) Modeling Network Autocorrelation within Migration Flows by Eigenvector Spatial Filtering. Journal of Geographical Systems 10 (4):317-44
Chun Y, Griffith DA (2011) Modeling Network Autocorrelation in Space–Time Migration Flow Data: An Eigenvector Spatial Filtering Approach. Annals of the Association of American Geographers 101 (3):523-36
Chun Y, Griffith DA (2013) Spatial Statistics and Geostatistics: Theory and Applications for Geographic Information Science and Technology. Sage, London
Cliff A, Ord K (1972) Testing for Spatial Autocorrelation Among Regression Residuals. Geographical Analysis 4 (3):267-84
Cliff AD, Ord JK (1981) Spatial Processes: Models & Applications. Pion, London
Crespo Cuaresma J, Feldkircher M (2013) Spatial Filtering, Model Uncertainty and the Speed of Income Convergence in Europe. Journal of Applied Econometrics 28 (4):720-41
Curry L (1972) A Spatial Analysis of Gravity Flows. Regional Studies 6 (2):131-47
Curry L, Griffith DA, Sheppard ES (1975) Those Gravity Parameters Again. Regional Studies 9 (3):289-96
Efroymson MA (1960) Multiple Regression Analysis. In: Ralston A, Wilf HS (eds) Mathematical Methods for Digital Computers. Wiley, New York, pp. 191-203
Egger P (2002) An Econometric View on the Estimation of Gravity Models and the Calculation of Trade Potentials. World Economy 25 (2):297-312
Egger P, Tarlea F (2015) Multi-Way Clustering Estimation of Standard Errors in Gravity Models. Economics Letters (forthcoming)
Egger PH, Staub KE (2015) GLM Estimation of Trade Gravity Models with Fixed Effects. Empirical Economics (forthcoming)
Feenstra RC, Lipsey RE, Deng H, Ma AC, Mo H (2005) World Trade Flows: 1962-2000. NBER Working Paper, National Bureau of Economic Research, Cambridge
Fischer MM, Griffith DA (2008) Modeling Spatial Autocorrelation in Spatial Interaction Data: An Application to Patent Citation Data in the European Union. Journal of Regional Science 48 (5):969-89
Flowerdew R, Aitkin M (1982) A Method of Fitting the Gravity Model Based on the Poisson Distribution. Journal of Regional Science 22 (2):191-202
Frankel J, Rose A (2002) An Estimate of the Effect of Common Currencies on Trade and Income. The Quarterly Journal of Economics 117 (2):437-66
Frankel JA (1997) Regional Trading Blocs in the World Economic System. Institute for International Economics, Washington
Greene WH (1994) Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models. NYU Working Paper, New York University Stern School of Business, New York, NY
Griffith DA (2003) Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization. Springer-Verlag, Berlin Heidelberg New York
Griffith DA (2007) Spatial Structure and Spatial Interaction: 25 Years Later. The Review of Regional Studies 37 (1):28-38
Griffith DA (2009) Spatial Autocorrelation in Spatial Interaction: Complexity-to-Simplicity in Journey-to-Work Flows. In: Reggiani A, Nijkamp P (eds) Complexity and Spatial Networks: In Search of Simplicity. Springer-Verlag, Berlin Heidelberg, pp. 221-37
Grimpe C, Patuelli R (2011) Regional Knowledge Production in Nanomaterials: A Spatial Filtering Approach. Ann Reg Sci 46 (3):419-51
Helpman E, Melitz M, Rubinstein Y (2008) Estimating Trade Flows: Trading Partners and Trading Volumes. The Quarterly Journal of Economics 123 (2):441-87
Jansen M, Nordås HK (2004) Institutions, Trade Policy and Trade Flows. CEPR Discussion Paper 4416, Centre for Economic Policy Research (CEPR), London

Koch W, LeSage JP (2015) Latent Multilateral Trade Resistance Indices: Theory and Evidence. Scottish Journal of Political Economy 62 (3):264-90

Krisztin T, Fischer MM (2015) The Gravity Model for International Trade: Specification and Estimation Issues. Spatial Economic Analysis (forthcoming)

Lambert D (1992) Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. Technometrics 34 (1):1-14

Lambert DM, Brown JP, Florax RJGM (2010) A Two-Step Estimator for a Spatial Lag Model of Counts: Theory, Small Sample Performance and an Application. Regional Science and Urban Economics 40 (4):241-52

LeSage JP, Pace RK (2008) Spatial Econometric Modeling of Origin-Destination Flows. Journal of Regional Science 48 (5):941-67

Leung SF, Yu S (1996) On the Choice between Sample Selection and Two-Part Models. Journal of Econometrics 72 (1–2):197-229

Linders G-JM, Burger MJ, van Oort FG (2008) A Rather Empty World: The Many Faces of Distance and the Persistent Resistance to International Trade. Cambridge Journal of Regions, Economy and Society 1 (3):439-58

Linnemann H (1966) An Econometric Study of International Trade Flows. North-Holland, Amsterdam

Long JS (1997) Regression Models for Categorical and Limited Dependent Variables. Sage Publications, Thousand Oaks

Martin W, Pham CS (2015) Estimating the Gravity Model when Zero Trade Flows Are Frequent and Economically Determined. vol Washington, DC. Policy Research Working Paper 7308, World Bank,

Martínez-Zarzoso I (2013) The Log of Gravity Revisited. Applied Economics 45 (3):311-27

Monestiez P, Dubroca L, Bonnin E, Durbec JP, Guinet C (2006) Geostatistical Modelling of Spatial Distribution of Balaenoptera Physalus in the Northwestern Mediterranean Sea from Sparse Count Data and Heterogeneous Observation Efforts. Ecological Modelling 193 (3-4):615-28

Patuelli R (2007) Regional Labour Markets in Germany: Statistical Analysis of Spatio-Temporal Disparities and Network Structures. Ph.D. Thesis, Department of Spatial Economics, VU University Amsterdam, Amsterdam

Patuelli R, Linders G-J, Metulini R, Griffith DA (2016) 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)

Philippidis G, Resano-Ezcaray H, Sanjuán-López AI (2013) Capturing Zero-Trade Values in Gravity Equations of Trade: An Analysis of Protectionism in Agro-Food Sectors. Agricultural Economics 44 (2):141-59

Raballand G (2003) Determinants of the Negative Impact of Being Landlocked on Trade: An Empirical Investigation Through the Central Asian Case. Comparative Economic Studies 45 (4):520-36

Rose AK (2000) One Money, One Market: The Effect of Common Currencies on Trade. Economic Policy 15 (30):8-45

Santos Silva JMC, Tenreyro S (2006) The Log of Gravity. Review of Economics and Statistics 88 (4):641-58

Santos Silva JMC, Tenreyro S (2011) Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator. Economics Letters 112 (2):220-2

Scherngell T, Lata R (2013) Towards an Integrated European Research Area? Findings from Eigenvector Spatially Filtered Spatial Interaction Models Using European Framework Programme Data. Papers in Regional Science 92 (3):555-77

Sellner R, Fischer MM, Koch M (2013) A Spatial Autoregressive Poisson Gravity Model. Geographical Analysis 45 (2):180-201
Sheppard ES, Griffith DA, Curry L (1976) A Final Comment on Mis-Specification and Autocorrelation in Those Gravity Parameters. Regional Studies 10 (3):337-9
Tinbergen J (1962) Shaping the World Economy: Suggestions for an International Economic Policy. Twentieth Century Fund, New York
Wang Z, Ma S, Wang C-Y (2015) Variable Selection for Zero-Inflated and Overdispersed Data with Application to Health Care Demand in Germany. Biometrical Journal 57 (5):867-84
Westerlund J, Wilhelmsson F (2009) Estimating the Gravity Model without Gravity Using Panel Data. Applied Economics 43 (6):641-9
Xiong B, Beghin J (2012) Does European Aflatoxin Regulation Hurt Groundnut Exporters from Africa? European Review of Agricultural Economics 39 (4):589-609

Appendix A.1. The *be.zeroinfl.filt.robust* Function

usage

be.zeroinfl.filt.rob = 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

Note 1: The sum of all the variables defined in the previous 4 inputs must be exactly equal to the list of explanatories contained in *object*.
Note2: The print commands in the second codes still should be changed (so far, it traces as output the results with the classical standard errors, and it wrongly write the name of the variable to be subsetted in each step).

Appendix A.2. Poisson-based Estimation Results

Table A.1. Estimation results for: (1) ZIP ESF; (2) ZIP; (3) Poisson ESF

	(1) ZIP ESF	(2) ZIP	(3) Poisson ESF
First Step			
Distance	0.76***	–0.36***	
Common language	–0.35*	0.28	
Contiguity	0.56	0.16	
Common history	–0.56	–1.42	
Free trade agreements	–0.87*	–1.43**	
Area importer	0.07	–0.05	
Area exporter	0.30***	0.28***	
GDP per cap. importer	–0.28***	–0.28***	
GDP per cap. exporter	0.22*	0.32***	
GDP importer	–0.53***	–0.45***	
GDP exporter	–1.16***	–1.16***	
Island importer	–0.31	–1.16*	
Island exporter	–1.20*	–1.73***	
Landlocked importer	3.31***	–0.14	
Landlocked exporter	–0.85 **	–1.06***	
Constant	126.48***	31.10***	
Eigenvectors (exp)	13	–	–
Eigenvectors (imp)	17	–	–
Second Step			
Distance	–0.54***	–0.39***	–0.41***
Common language	0.13***	0.39***	0.23***

	(1) ZIP ESF	(2) ZIP	(3) Poisson ESF
Contiguity	0.57***	0.66***	0.61***
Common history	0.17***	0.13***	0.21***
Free trade agreements	0.58***	0.72***	0.80***
Area importer	–0.19***	–0.07***	–0.21***
Area exporter	0.02***	–0.05***	0.01*
GDP per cap. importer	–0.18***	–0.04***	–0.06***
GDP per cap. exporter	0.04***	0.09***	–0.05***
GDP importer	0.95***	0.84***	0.91***
GDP exporter	0.72***	0.76***	0.71***
Island importer	–0.08***	–0.13***	0.29***
Island exporter	–0.58***	–0.23***	–0.34***
Landlocked importer	–0.01***	–0.05***	–0.24***
Landlocked exporter	0.13***	0.19***	0.20***
Constant	–29.48	–29.52	–29.10
Number significant eigenvectors exporter	7	–	11
Number significant eigenvectors importer	9	–	22
AIC	1,851,472	2,545,233	2,249,365
Log-likelihood	-8.983e+05	-1.27e+06	-1.12e+06
McFadden's pseudo-R^2	0.9186	0.9068	0.8881
Observations	4032	4032	4032
Residual dof	3954	4000	3983

Table A.2. Count of observed versus predicted values. Poisson models comparison

Trade flow	0	1	2	3	4	5	6	7	8	9
Observed	484	136	112	76	64	39	42	49	35	29
ZIP ESF	484	3	5	7	8	9	10	10	11	11
ZIP	484	0	1	1	2	3	4	5	6	7
Poisson ESF	2	4	8	10	13	15	16	18	18	19