ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

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

Rodolfo Metulini
Roberto Patuelli
Daniel A. Griffith

Quaderni - Working Paper DSE N°1081

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] Department of Economics and Management, University of Brescia, 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, because it permits a better accounting for zero flows and extreme values in the distribution tail. Nevertheless, as trade flows are not independent from each other due to spatial autocorrelation, these methods may lead to biased parameter estimates. To overcome this problem, eigenvector spatial filtering variants of the Poisson/negative binomial specification have been proposed in the literature of gravity modelling of trade. However, no specific treatment has been developed for cases in which many zero flows are present. This paper contributes to the literature in two ways. First, by employing a stepwise selection criterion for spatial filters that 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 processes of zero-inflated methods. 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; zero-inflation.

1. Introduction

A traditional gravity model describing trade in its simple form (Linnemann 1966; Tinbergen 1962) 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 squares (OLS) estimation has been widely used in the empirical literature (Egger 2002; Frankel and Rose 2002; Rose 2000), mostly because of its good empirical performance and, 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 stress that null trade flows are to be specifically taken into account. Helpman et al. (2008) prove that disregarding countries that do not trade with each other generates biased estimates. Moreover, Santos Silva and Tenreyro (2006) show 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 initially proposed by 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. Several empirical studies of trade have applied the PPML estimator (see Burger et al. 2009; Linders et al. 2008; Martin and Pham 2015; Martínez-Zarzoso 2013). Alternatively, in order to correct for overdispersion, a negative binomial (NB) regression model, which belongs to the family of Poisson models, and allows for the dispersion parameter to differ from 1, is employed. A wider discussion regarding the choice between Poisson and NB estimators (for the pseudo-ML case in particular), can be found, for example, in Bosquet and Boulhol (2014) and Head and Mayer (2015).

The zero-inflated specification (Greene 1994; Lambert 1992; Long 1997) applied to NB models (ZINB) permits a better estimate in the presence of a large number of zero flows, because it considers the existence of two groups within the population: one having strictly zero counts, and another having a non-zero probability of having a trade flow greater than zero.

Burger et al. (2009) stress 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, which variables determine the decision to trade is not so clear.

Today, a well-known feature of trade flows is that they 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 (extensive margin) and the volume of trade (intensive margin), and, in particular, what the contribution of SAC is in both of these 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 (Fischer and Griffith 2008; Griffith 2007), one accounting for SAC in the logit part, and the other accounting for SAC in the count part. In this regard, we devise an *ad hoc* function that applies a backward stepwise algorithm aiming to properly identify the significant spatial filters. Our proposed algorithm has the advantage that, at each step, it drops the eigenvector with the largest p-value, regardless of whether it is in the count or in the logit part. We compare the results of this estimation with two methodologically nested benchmarks, namely a ZINB and a NB with origin and destination spatial filters (the former employing spatial filters only in the count part). We conduct a comparison in terms of both estimated coefficients and goodness of fit (Akaike information Criteria, AIC and prediction of zero and small flows). We find that our specification outperforms the comparison models, in terms of both AIC and prediction of small trade flows. An alternative analysis based on a ZIP specification is provided.

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 recently experienced 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 with, for example, the popular paper by Santos Silva and Tenreyro (2006), have pushed the envelope in the field, and a number of researchers are actively pursuing further methodological advances pertaining to, in particular, 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 (2016) compare the suitability of various estimation approaches under an international economics general equilibrium perspective. Baltagi and Egger (2016) 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 their network's topological characteristics) in gravity models (Patuelli and Arbia 2016), trade being a frequent application. While the relevance of spatial autocorrelation originally was suggested for trade models in Anderson and van Wincoop (2004), and much earlier within spatial interaction modelling (Curry 1972; Curry et al. 1975; Sheppard et al. 1976), this issue attracted significant attention only in recent years. Studies by Behrens et al. (2012), Fischer and Griffith (2008), and LeSage and Pace (2008) provide, 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; Behrens et al. 2012; Fischer and Griffith 2008; Koch and LeSage 2015; LeSage and Pace 2008): these models apply and adapt traditional (linear) spatial econometric techniques to the count data case.
- Spatial generalized linear models (GLMs) (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; Krisztin and Fischer 2015; Patuelli et al. 2016; Scherngell and Lata 2013): 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 in Section 3.2) is a spatial statistics technique based on the decomposition of spatial weights matrices. The available studies employing this technique demonstrate how spatial filters can be used successfully at the intercept level as ‘interceptors’ of (i.e., proxies for) unobserved spatial heterogeneity. This paper aims to further investigate the use of ESF, by allowing for separate spatial filter sets in zero-inflated models.

3. A Methodological Approach

The proposed approach is described in this section.

3.1 Zero-Inflated Gravity Models of Trade

In recent years, an increasing recognition is that the level of trade between countries frequently is zero. Small countries may not have trade relations with all possible trading partners, or statistical offices may not report trade flows below a certain threshold. Moreover, the issue of zero flows is more pronounced when analysing sector-disaggregated trade flows. Zero-inflated gravity models provide one way to model an excess of zero flows. Martin and Pham (2015) and Burger et al. (2009) propose the zero-inflated extension of the Poisson gravity model for situations where the data-generating process (DGP) results in too many zeros. The model may be viewed as a "two-part" extension, in which the distribution of the outcome variable is approximated by mixing two component distributions. The zero-inflation part of the model consists of a qualitative-dependent model to determine the probability of whether a particular origin-destination trade flow is zero or positive. The second part contains the standard Poisson (or NB) 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) apply zero-inflated count models for the analysis of international trade.

Estimating the parameters of the NB gravity model (with or without zero-inflation) by standard non-spatial methods only is justified statistically if we believed that trade flows are independent observations. However, such an assumption generally is not valid because flows fundamentally are spatial in nature. Several recent papers propose modelling the spatial heterogeneity in the residuals by means of different econometric techniques. Among those works, many focus 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 in the Poisson gravity model by means of spatial autoregressive techniques (Lambert et al. 2010; Sellner et al. 2013). Another is ESF (Griffith 2003). It is considered here because it allows for greater flexibility in modelling, and can be applied seamlessly to any estimation framework. In their recent work, Patuelli et al. (2016) apply spatial filters with NB as a way to filter out spatial heterogeneity due to MTRs. However, residual heterogeneity could be present both for the logit and the count process, whereas the previously mentioned works only account for SAC in 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. In particular, their approach implies using a network autocorrelation spatial filter in the count part of the model. This work follows a similar approach used by Krisztin and Fischer, but we introduce an *ad hoc* backward stepwise procedure to properly select the filters. Moreover, we perform diagnostics in order 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

ESF originally was developed for area-based data by Griffith (2003), and later extended to flow data (Chun 2008; Chun and Griffith 2011; Fischer and Griffith 2008; Griffith 2009). One traditional advantage, when including eigenvectors as additional origin- and destination-specific regressors, is that the model can be estimated within standard regression frameworks, such as OLS or Poisson regression, which are common in the literature about spatial interaction. 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 the 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 **W**. The subsequently extracted eigenvectors maximize MC while being orthogonal to and uncorrelated with the previously extracted eigenvectors. Finally, the last extracted eigenvector maximizes negative MC.

Having extracted the eigenvectors of Equation (1), a spatial filter is constructed as a linear combination of a judiciously selected subset of these n eigenvectors. In detail, for our empirical application, we select a first subset of eigenvectors (which we call the 'candidate eigenvectors') by means of the following threshold: $\mathrm{MC}(e_i)/\mathrm{MC}(e_1) > 0.25$. This threshold yields a spatial filter whose variance attributable to SAC is at least roughly 95% (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 given data) to just the subset of eigenvectors that are statistically significant as regressors in the model to be evaluated. The linear combination of the resulting group of eigenvectors is what we call our 'spatial filter'. This estimation technique has been applied in 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 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 gross domestic product (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 (see also Griffith and Chun 2016).

The new challenge here is that we want to account for SAC in both the logit and in the count parts of zero-inflated models, so we use two sets of filters at the logit level, and two sets

[1] In this regard, Chun et al. (2016) formulate an estimation equation, based on residual SAC, to predict the ideal size of the candidate set, and demonstrate that the optimal size of the set of candidate eigenvectors is positively related to the amount of spatial autocorrelation to account for.

of filters at the count level. This choice allows us to account for potentially different omitted variables related to the intensive and extensive margins of trade. Moreover, the selection of different eigenvectors in the two parts (i.e., exclusion restrictions) may help obtain identification as well, consistent with Papadopoulos and Santos Silva (2012).

3.3 A Backward Stepwise Algorithm

A stepwise procedure is an algorithm used to choose variables in a regression model, first proposed by Efroymson (1960). It usually takes the form of a sequence of F- or t-tests, but other criteria are possible, such as (adjusted) R-squared, AIC, Bayesian information criterion (BIC), or simply based on p-values.

Forward selection involves starting with no variables in a model, testing the addition of each variable, adding the variable (if any) that improves the model the most, and repeating this process until no more (significant) improvement is possible. Backward elimination involves starting with all candidate variables, sequentially testing the deletion of each of them, deleting the variable (if any) whose deletion improves the model fit the most, and repeating this process until no further improvements are possible. Backward elimination procedures are implemented in many routines. Chun and Griffith (2013) list `R` code for stepwise selection in GLMs based on SAC minimization. In the *mpath* package (Wang et al. 2015), the *be.zeroinfl* function performs a backward elimination (and forward selection) based on maximum likelihood criteria, and can be applied to zero-inflated models.

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

Our algorithm (see Appendix A.1) is inspired by the *be.zeroinfl* function, but has at least two advantages vis-à-vis it. First, at each step of our algorithm, we compute robust standard errors and we select the variable to be removed based on the related p-values. Second, our algorithm is constructed in order to be able to drop the variables with the largest p-values, regardless of whether they belong to the count or the logit part. We also structured the function so that a minimum model (*minmodel*) can be defined. In other words, we let the algorithm drop only the eigenvectors, because we consider included standard explanatory variables to have substantive meaning.

4. An Empirical Application

The data for trade analysed in this paper 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. Whether pairs of countries take part in a common regional integration agreement (FTA) was determined on the basis of OECD data about major regional integration agreements.[2] An indicator variable measures whether a pair of countries has (membership in) at least one common FTA. Data about island status have been kindly provided by Hildegunn Kyvik-Nordas (from Jansen and Nordås 2004). Internal flows are excluded from our analyses because they typically deserve special treatment in trade models (see, e.g., LeSage and Fischer 2016). Their treatment within our modelling framework is left for future research.

4.1 The Model Specification

For estimation, we follow a standard specification of the gravity equation of bilateral trade, and we employ some variables commonly used in the literature (see, e.g., Frankel 1997; Raballand 2003). We use the following standard specification of the gravity equation, which we estimated by means of a ZINB (and ZIP as a sensitivity analysis):

$$\begin{aligned}
\Pr(Tr_{ij}=1) &= \alpha_1 dist_{ij} + \alpha_2 comcur + \alpha_3 contig + \alpha_4 hist + \alpha_5 fta \\
&+ \beta_1 area_i + \beta_2 area_j + \beta_3 gdp_i + \beta_4 gdp_j + \beta_5 gdpcap_i + \beta_6 gdpcap_j \\
&+ \beta_7 Island_i + \beta_8 Island_j + \beta_9 landl_i + \beta_{10} landl_j + \varepsilon_{ij};
\end{aligned} \tag{2a}$$

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

where *Tr* represents trade flows, *gdp* represents the GDP (in logs), *gdpcap* represents per capita GDP (in logs), *Island* is an indicator variable that equals 1 if a country is an island, *Area* is the land area of a country (in logs), and *landl* equals 1 for landlocked countries. The other variables are country-pair indicators, identifying whether a pair of countries share a currency (*comcur*), a common border (*contig*), a common history (*hist*), or engage in free

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

trade agreements (*fta*), and *dist* is a measure of the geographical distance between them (in logs).

4.2 Estimation Results

We estimate Equation (2) and select spatial filters as a ZINB, using a cross-section of 64 countries (4,032 country pairs) for the year 2000 (ZINB ESF, Table 1). We estimate the same model using two benchmarks that methodologically can be considered as special cases of our proposed model: a ZINB using spatial filters only in the count part (ZINB ESFc), and an NB with spatial filters (NB ESF).

Looking to the count part (second step) of the ZINB ESF, distance has a negative significant effect, the country-pair indicator variables all present positive and significant coefficients, and country mass variables a positive one, as expected: GDP positively affects trade flows, at both the exporter and importer country side. The area size as well as the GDP per capita of the exporter country negatively affect trade, while the values for importing countries have smaller positive coefficients. When comparing the findings of Model (1) with the ones of the benchmarks, the coefficients do not change much, with some exceptions for Model (3), most likely because of compensation for the lack of the zero-inflation part. In general, coefficients that are significant in the ZINB ESF also are significant in our benchmark models.

Considering the logit part (first step), the probability of a country pair to be involved in trade negatively depends on distance, positively depends on the importer and exporter country areas, but, surprisingly, considering Model (1), negatively depends on GDP. Moreover, the coefficients resulting from the alternative zero-inflated specification [Model (2), which does not include spatial filters in the logit part] often differ from the ones for Model (1), suggesting that the inclusion of the spatial filters has a relevant role. These results highlight the need to better analyse the determinants of trade decisions.

Based on the AIC and the log-likelihood values, our model specification outperforms the benchmarks. In terms of AIC, the ZINB ESF has the lowest value (47,026), meaning it performs better than the benchmarks (47,566 for the ZINB ESFc, and 48,414 for the NB ESF). The same holds for the log-likelihood ($-2.32 * 10^4$ compared to $-2.37 * 10^4$ and $-2.42 * 10^4$, respectively).

Appendix A.2 summarizes Poisson estimation results. Results appearing in Table A.1 confirm that, similarly to the NB case, the zero-inflated Poisson (ZIP) ESF outperforms the two benchmark models (ZIP and Poisson ESF) in terms of both AIC and likelihood.

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

	(1) ZINB ESF	(2) ZINB ESFc	(3) NB ESF
First Step			
Distance	–1.18***	0.23**	–
Common language	–1.34**	0.50**	–
Contiguity	1.85*	0.14	–
Common history	–2.28*	–1.54	–
Free trade agreements	–0.86	–1.43***	–
Area importer	2.77***	–0.05	–
Area exporter	0.23**	0.36***	–
GDP per cap. importer	–0.69***	–0.53***	–
GDP per cap. exporter	0.97***	0.45***	–
GDP importer	–5.25***	–0.14	–
GDP exporter	–2.82***	–1.26***	–
Island importer	17.04***	–0.99	–
Island exporter	–1.97	–2.01***	–
Landlocked importer	39.70***	0.12	–
Landlocked exporter	3.00***	–1.37***	–
Constant	126.48***	27.53***	–
Eigenvectors (exp)	11	–	–
Eigenvectors (imp)	24	–	–
Second Step			
Distance	–0.84***	–0.83***	–0.71***
Common language	0.46***	0.44***	0.42***
Contiguity	0.54***	0.54***	0.66***
Common history	0.77***	0.76***	0.83***
Free trade agreements	0.48***	0.48***	0.77***
Area importer	–0.20***	–0.20***	–0.23***
Area exporter	0.07***	0.08***	–0.03
GDP per cap. importer	–0.24***	–0.26***	–0.14***
GDP per cap. exporter	0.16***	0.19***	–0.10***
GDP importer	1.06***	1.06***	1.00***
GDP exporter	0.63***	0.62***	0.81***
Island importer	0.43***	0.36***	0.34***
Island exporter	–0.70***	–0.75***	–0.02
Landlocked importer	–0.21**	–0.28***	–0.27***
Landlocked exporter	0.32***	0.24**	0.37***
Constant	–28.71***	–28.65***	–30.06***
Eigenvectors (exp)	11	9	8
Eigenvectors (imp)	8	9	12
Theta	0.86	0.92	0.59
AIC	47,026	47,566	48,414
Log-likelihood	–2.32e+04	–2.37e+04	–2.42e+04
McFadden’s pseudo-R2	0.1312	0.1196	0.1022
Observations	4032	4032	4032
Residual dof	3945	3981	3995

***, **, * denote statistical significance at the 1, 5, 10 per cent level.

We also can analyse the robustness of our model in terms of fitting small trade flows. We compare the observed frequencies of small flows with their estimated counterparts (fitted values rounded to integers) obtained for all the models. Because one advantage of our model specification is that it should better predict small flows, we expect it to outperform the two benchmark models in this regard, especially if small flows are spatially autocorrelated. Results reported in Table 2 confirm this expectation. The ZINB ESF predicts 440 out of 484 zero flows, whereas the NB ESF predicts only 281 zero flows. The ZINB ESFc, using only count-level spatial filters, predicts 480 zero flows, but it is less efficient in predicting other small flows, compared to the ZINB ESF. Appendix A.2, Table A.2 reports similar results for predictions of small flows using Poisson-based models. Despite ZIP models adequately predicting zero flows, both they and the standard Poisson model lack efficiency in predicting small flows. In this respect, NB models appear to outperform Poisson models.

Table 2. Counts of observed versus predicted values

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 ESFc	480	79	68	61	55	50	47	43	41	38
NB ESF	281	156	117	95	82	72	64	58	53	49

The spatial part of the model, with the ZINB ESF we select in the logit part, comprises 11 exporter-side and 24 importer-side eigenvectors. In the count part, the number of significant eigenvectors is 11 for the exporter countries and 8 for the importer countries.

A Moran test can be conducted on each of the four spatial filters, which are obtained as the linear combinations of the selected eigenvectors multiplied by their respective estimated coefficients. The one 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 count part ones (MC = 0.160, with 8 eigenvectors, and MC = 0.298, with 11 eigenvectors, for importer- and exporter-side, respectively). 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.

A further dimension to be investigated is the differentiated use of the eigenvectors in the construction of the spatial filters, at the importer/exporter and logit/count levels, which can provide hints regarding the extent of omitted explanatory variables and their overlap across contexts. A comparison of importer and exporter spatial filters (Table 3) implies that more common eigenvectors are present in the logit part of the model. This finding suggests that (omitted) trade determinants are more differentiated, in terms of emissiveness and attractiveness, on the intensive margin. When looking at differences between the logit and count parts of the model, the same number of common eigenvectors can be found for the exporter and importer sides, showing that a moderate amount of omitted information is relevant for both extensive and intensive margins. More generally, only one eigenvector (e9) is common to all four spatial filters, while out of the top three eigenvectors (e1–e3), only e1 (the one implying the spatial pattern with the highest level of SAC) appears in more than one spatial filter. These final findings lead us to believe that (omitted) trade patterns are mostly idiosyncratic or tied to specific areas, rather than linked to larger geographical agglomerations.

5. Conclusions

Eigenvector spatial filtering (ESF) variants of nonlinear gravity models of trade (such as Poisson or NB specifications) have been proposed in the literature, because trade flows are not independent and contain spatial autocorrelation (SAC). Using a zero-inflated negative binomial (ZINB) approach, this paper contributes to the existing literature in two ways. First, we present a zero-inflated stepwise selection procedure for constructing spatial filters based on robust p-values. Second, we identify spatial filters that properly account for importer- and exporter-side specific unexplained spatial patterns, in both the logit and count parts. 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 ESFc and NB ESF) in terms of model fitting, both considering AIC and log-likelihood values, and in predicting zero (and small) flows.

Future research should compare this model with further ZINB specifications that account for SAC differently, and evaluate the contribution of the logit and the count parts of the model in terms of explained variance based on different DGPs. Attention should be devoted to a specific treatment of internal flows as well. Moreover, a similar analysis, taking care of appropriate changes, should be applied to a panel data setting to evaluate, for example, possible trade-offs between the spatial filters and individual (dyadic) fixed effects.

Table 3. Common and unique eigenvectors

	Comparison	Eigenvectors	Comparison	Eigenvectors
Exporter vs Importer	Exporter/importer, logit (common)	e4, e5, e8–e10, e12, e14, e17, e23	Exporter/importer, count (common)	e1, e9, e20
	Exporter, logit (unique)	e1, e11	Exporter, count (unique)	e3, e4, e10, e11, e15, e17, e19, e23
	Importer, logit (unique)	e2, e7, e13, e16, e18–e20, e22, e24–e30	Importer, count (unique)	e5, e7, e14, e22, e25
Logit vs count	Logit/count, exporter (common)	e1, e4, e9–e11, e17, e23	Logit/count, importer (common)	e5, e7, e9, e14, e20, e22, e25
	Logit, exporter (unique)	e5, e8, e12, e14	Logit, importer (unique)	e2, e4, e8, e10, e12, e13, e16–e19, e23, e24, e26–e30
	Count, exporter (unique)	e3, e15, e19, e20	Count, importer (unique)	e1

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 (2016) Estimation of Structural Gravity Quantile Regression Models. Empirical Economics 50 (1):5-15

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
Bosquet C, Boulhol H (2014) Applying the GLM Variance Assumption to Overcome the Scale-Dependence of the Negative Binomial QGPML Estimator. Econometric Reviews 33 (7):772-84
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
Chun Y, Griffith DA, Lee M, Sinha P (2016) Eigenvector Selection with Stepwise Regression Techniques to Construct Eigenvector Spatial Filters. Journal of Geographical Systems 18 (1):67-85
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 134:144-7
Egger PH, Staub KE (2016) GLM Estimation of Trade Gravity Models with Fixed Effects. Empirical Economics 50 (1):137-75
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 D, Chun Y (2016) Evaluating Eigenvector Spatial Filter Corrections for Omitted Georeferenced Variables. Econometrics 4 (2):29

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

Head K, Mayer T (2015) Gravity Equations: Workhorse, Toolkit, and Cookbook. In: Helpman E, Gopinath G (eds) Handbook of International Economics. vol 4. BV, Oxford, pp. 131-95

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 10 (4):451-70

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, Fischer MM (2016) Spatial Regression-Based Model Specifications for Exogenous and Endogenous Spatial Interaction. In: Patuelli R, Arbia G (eds) Spatial Econometric Interaction Modelling. Springer, Heidelberg Berlin (forthcoming),

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

Papadopoulos G, Santos Silva JMC (2012) Identification Issues in Some Double-Index Models for Non-Negative Data. Economics Letters 117 (1):365-7

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, Arbia G (eds) (2016) Spatial Econometric Interaction Modelling. Springer, Cham

Patuelli R, Linders G-JM, Metulini R, Griffith DA (2016) The Space of Gravity: Spatially Filtered Estimation of a Gravity Model for Bilateral Trade. In: Patuelli R, Arbia G (eds) Spatial Econometric Interaction Modelling. Springer International Publishing, Cham, pp. 145-69

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

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 =
  ("poisson", "negbin", "geometric"), alpha = 0.05, trace =
  FALSE, subset.zero, subset.count, minmod.zero, minmod.count).
```

Details

Conduct a backward stepwise variable elimination for zero inflated count regression with the `zeroinfl` function, providing a possibility to define a minimum model and using sandwich robust standard errors.

Value

An object of `zeroinfl` class with all variables having p-values less than the significance level alpha.

Arguments

`object`: an object from function `zeroinfl`

`data`: an argument controlling formula processing via `model.frame`

`dist`: one of the distributions in the `zeroinfl` function

`alpha`: the significance level for variable elimination

`trace`: logical value, if TRUE, generates printed detailed calculation results

`subset.zero`: a list of the variable names to be subset in the zero component

`subset.count`: a list of the variable names to be subset in the count component

`minmod.zero`: a list of the variable names not to be subset in the zero component

`minmod.count`: a list of the variable names not to be subset in the count component

Note: The sum of all the variables defined in the previous four inputs must be exactly equal to the list of explanatory variables contained in `object`.

Appendix A.2. Poisson-Based Estimation Results

Table A.1. Estimated coefficients for: (1) ZIP ESF; (2) ZIP ESFc; (3) Poisson ESF

	(1) ZIP ESF	(2) ZIP ESFc	(3) Poisson ESF
First Step			
Distance	0.76***	0.36***	–
Common language	–0.35	0.28	–
Contiguity	0.56	0.15	–
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***	–

	(1) ZIP ESF	(2) ZIP ESFc	(3) Poisson ESF
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	–1.17***	–
Island exporter	–1.20**	–1.73***	–
Landlocked importer	3.31***	–0.14	–
Landlocked exporter	–0.85***	–1.06***	–
Constant	126.48***	31.01***	–
Eigenvectors (exp)	13	–	–
Eigenvectors (imp)	17	–	–
Second Step			
Distance	–0.54***	–0.54***	–0.42***
Common language	0.13	0.13	0.23**
Contiguity	0.57***	0.57***	0.61***
Common history	0.17*	0.17*	0.21**
Free trade agreements	0.58***	0.58***	0.80***
Area importer	–0.19***	–0.19***	–0.21***
Area exporter	0.02	0.02	0.01
GDP per cap. importer	–0.18***	–0.18***	–0.06
GDP per cap. exporter	0.04	0.04	–0.05
GDP importer	0.95***	0.95***	0.91***
GDP exporter	0.72***	0.72***	0.71***
Island importer	–0.08	–0.08	0.29
Island exporter	–0.58***	–0.58***	–0.34***
Landlocked importer	–0.01	–0.01	–0.24
Landlocked exporter	0.13	0.13	0.20
Constant	–29.48***	–29.48***	–29.10***
Eigenvectors (exp)	7	7	11
Eigenvectors (imp)	9	9	22
AIC	1,851,472	1,851,887	2,249,365
Log-likelihood	–8.98e+05	–9.26 e+05	–1.12e+06
McFadden’s pseudo-*R*2	0.9186	0.9186	0.8881
Observations	4032	4032	4032
Residual dof	3954	3983	3983

***, **, * denote statistical significance at the 1, 5, 10 per cent level.

Table A.2. Counts of observed versus predicted values. A 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 ESFc	484	2	4	6	8	9	9	10	11	11
Poisson ESF	2	4	8	10	13	15	16	18	18	19

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it