Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

The Space of Gravity: Spatial Filtering Estimation of a Gravity Model for Bilateral Trade

Roberto Patuelli
Gert-Jan Linders
Rodolfo Metulini
Daniel A. Griffith

The Space of Gravity: Spatial Filtering Estimation of a Gravity Model for Bilateral Trade

Roberto Patuelli,[a] Gert-Jan Linders,[b] Rodolfo Metulini[c] and Daniel A. Griffith[d]

[a] Department of Economics, University of Bologna, Italy; The Rimini Centre for Economic Analysis (RCEA), Italy
[b] Ministry of Social Affairs and Employment, The Netherlands
[c] IMT Institute for Advanced Studies Lucca, Italy
[d] School of Economic, Political & Policy Sciences, The University of Texas at Dallas, USA

ABSTRACT

Bilateral trade flows traditionally have been analysed by means of the spatial interaction gravity model. Still, (auto)correlation of trade flows has only recently received attention in the literature. This paper takes up this thread of emerging literature, and shows that spatial filtering (SF) techniques can take into account the autocorrelation in trade flows. Furthermore, we show that the use of origin and destination specific spatial filters goes a long way in correcting for omitted variable bias in an otherwise standard empirical gravity equation. For a cross-section of bilateral trade flows, we compare an SF approach to two benchmark specifications that are consistent with theoretically derived gravity. The results are relevant for a number of reasons. First, we correct for autocorrelation in the residuals. Second, we suggest that the empirical gravity equation can still be considered in applied work, despite the theoretical arguments for its misspecification due to omitted multilateral resistance terms. Third, if we include SF variables, we can still resort to any desired estimator, such as OLS, Poisson or negative binomial regression. Finally, interpreting endogeneity bias as autocorrelation in regressor variables and residuals allows for a more general specification of the gravity equation than the relatively restricted theoretical gravity equation. In particular, we can include additional country-specific push and pull variables, besides GDP (e.g., land area, landlockedness, and per capita GDP). A final analysis provides autocorrelation diagnostics according to different candidate indicators.

1. Introduction

During the past two decades, scholars have shown renewed interest in the theoretical foundations and estimation of the gravity model for bilateral trade (e.g., Deardorff 1998; Anderson and van Wincoop 2003). The interest in modelling trade flows has increased with questions about the effectiveness of trade agreements (Baier and Bergstrand 2009) and the persistence of border and distance effects and largely unobserved trade costs (Anderson and van Wincoop 2004). The developments have re-affirmed the importance of accounting for relative trade costs in explaining patterns of trade. Yet, empirical application of the resulting gravity model framework that incorporates theoretically motivated multilateral resistance (MR) is not straightforward. The system of equations for MR involves non-linearities in the parameters and requires custom programming (Feenstra 2004).

An alternative specification that circumvents the need to consider the full system of equations includes country-specific effects to control for omitted country-specific MR variables. However, both the system approach and the alternative using fixed effects impose restrictions on the empirical specification of the gravity model. They allow identification of the impact of bilateral trade barriers, but preclude (at least in a cross-section) the analysis of country-specific covariates that may affect patterns of trade.

This paper aims to contribute to the literature in providing an alternative solution to deal with omitted MR, which allows for parameter identification for country-specific covariates in a cross-section analysis of trade patterns. This solution hinges on the interpretation of spatial autocorrelation (SAC)[1] in trade flows as reflecting unobserved country-specific heterogeneity due to MR. Our approach is complementary to a related recent strand of literature that starts from the same interpretation in that we offer an alternative methodology to deal with SAC in trade flows, called spatial filtering (SF) estimation.

The literature review about trade costs by Anderson and van Wincoop (2004) suggests that the application of spatial econometric techniques in modelling origin-destination trade flows needs further exploration, to take into account the (auto)correlation in trade flows. Although the gravity model is essentially a model of spatial interaction, little attention has been paid to

---

[1] Spatial autocorrelation is the correlation that occurs among the values of a georeferenced variable, and that can be attributed to the proximity of the units. The concept of SAC can be related to the first law of geography, stating that 'everything is related to everything else, but near things are more related than distant things' (Tobler 1970, p. 236).

flows autocorrelation in the trade literature (Porojan 2001, is an exception). In part, this lack of attention was due to technical reasons. Spatial econometric modelling of origin-destination flows is complex and computationally taxing. Estimation of spatial lag and spatial error models in this context has long been impossible due to computing power limitations. Applications of spatial interaction modelling in regional science have recently made progress on this issue (see Fischer and Griffith 2008; LeSage and Pace 2008; Sellner et al. 2013). Applications in empirical trade and FDI modelling have followed shortly thereafter (see Baltagi et al. 2007; Behrens et al. 2012). These contributions show the relevance of autocorrelation in trade flows. However, spatial econometric origin-destination flow models remain complex and relatively taxing to apply empirically. In response to these concerns, several studies have applied an alternative spatial econometric technique, SF, which deals with autocorrelation in a different but equally effective way. The technique of SF has recently been applied to the origin-destination flow context in other fields, such as commuting and patent citations (Fischer and Griffith 2008; Griffith 2009). Instead of accounting for autocorrelation by spatial modelling, SF estimation deals with it by filtering the residuals. Because only an origin-specific and a destination-specific filter are needed in order to account for autocorrelation, the dimensionality of estimation is much less demanding than in the case of a spatial lag or spatial error origin-destination model.

This paper follows up on this development by applying SF estimation to bilateral trade flows. We argue that the application of origin-specific and destination-specific filtering of residuals corresponds well to the theoretically expected importance of omitted origin-specific and destination-specific MR terms. Empirical results show that SF estimation can account well for autocorrelation in trade flows.

Moreover, SF estimation of an otherwise standard empirical gravity equation appears to go a long way in correcting for bias due to the origin- and destination-specific omitted variables predicted by the theoretical gravity model. The regression coefficients are close to the benchmark values in a specification using origin- and destination-specific indicator variables. This implies that SF estimation provides a relatively simple alternative to spatial econometric origin-destination flow models and custom-programmed non-linear estimation of the theoretical gravity model, which can be estimated using standard techniques such as ordinary least squares (OLS) or Poisson regression.[2]

---

[2] The estimates presented in this paper have been carried out with the R statistical software (R Core Team 2015). The script necessary for running the SF estimations is available for download from the first author's personal homepage.

Finally, the SF approach allows for a greater flexibility in the empirical specification of the gravity equation. Unlike the specification using indicator variables, we can include country-specific variables – so called push and pull factors – in the model. Moreover, a SF model is a significant improvement in terms of parsimony and efficiency compared to the indicator variables model. Compared to the theoretical gravity framework, we can relax the assumption that total trade depends exclusively and proportionately on the gross domestic product (GDP) of the trading countries. Other potential push and pull factors, such as landlockedness, land area, or per capita income can be included as well, and we do not have to assume a proportional relation between trade and GDP. Thus, SF estimation entails greater flexibility in specification choice compared to the stylized theoretical gravity model.

The paper proceeds as follows. In Section 2, we specify a theoretical gravity model following Anderson and van Wincoop (2003) and discuss some practical limitations of applying the theoretical framework. In Section 3, we illustrate the link between theoretical gravity and autocorrelation in trade flows. We present the approach of SF estimation to control for autocorrelation, and motivate that it allows controlling for unobserved MR. Section 4 outlines the empirical specifications and estimators that we compare, while Section 5 discusses the SAC tests that we use for post-estimation diagnostics. In Section 6, after an overview of data used, we turn to the estimation results and diagnostics. Section 7 concludes the paper.

2. The Gravity Model and Autocorrelation

We can divide the discourse over trade gravity modelling in two parts, regarding the theoretical and empirical approaches to the problem, respectively. The following sections attempt to provide such a discussion.

2.1 Theoretical gravity

Gravity equations for analysing bilateral trade flows have been estimated since the 1960s (e.g., Tinbergen 1962; Pöyhönen 1963). The model describes the volume of bilateral trade as a function of push and pull factors, such as economic size of origins and destinations, and transactional distance between trade partners. It has been deployed for various purposes, such as analysing the determinants of trade patterns, testing trade theories, forecasting future flows or estimating missing data, and comparative static analysis of changes in trade costs. Recent applications increasingly emphasize the importance of estimating a gravity equation that is

4

consistent with theoretical gravity (e.g., Anderson and van Wincoop 2003; Baier and Bergstrand 2009). The theoretical framework that is most influential has been developed by Anderson and van Wincoop (2003), in their paper on consistent estimation and assessment of the border effect in U.S.-Canadian regional trade flows.[3]

Anderson and van Wincoop derive a reduced-form gravity equation, assuming an *N*-country endowment economy, constant elasticity of substitution (CES) preferences, and symmetric bilateral trade costs. Their model explicitly takes into account the role played by country-specific price indices (MR terms). The gravity equation that results is specified as:

$$x_{ij} = \frac{y_i y_j}{y_w} \left(\frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma}, \qquad (1)$$

where x_{ij} is the value of the flow of goods from country *i* to country *j*, *y* is GDP (*w* stands for world) and t_{ij} is the bilateral trade cost factor. Finally, two variables enter that we discuss in greater detail later. Π_i measures outward MR of country *i*, and P_j measures inward MR of country *j*. The term σ is the elasticity of substitution ($\sigma > 1$).

Equation (1) shows that bilateral exports would be proportional to the size of the exporting market and the share of the import market in total demand, in the absence of bilateral trade costs (t_{ij}). Trade costs are of the iceberg cost type, and we define trade costs as a mark-up on the 'mill price' p_i ($t_{ij} \geq 1$). Hence, ($t_{ij} - 1$) is the ad-valorem tariff equivalent of bilateral trade costs. The bilateral delivered prices (p_{ij}) then equal:

$$p_{ij} = t_{ij} \cdot p_i. \qquad (2)$$

A wide variety of covariates in the literature is used to represent bilateral trade costs. We include some of the most common bilateral explanatory variables. A multiplicative formulation of bilateral trade costs (see Deardorff 1998; Anderson and van Wincoop 2004) yields:

$$t_{ij} = D_{ij}^{\beta_1} \cdot e^{\beta_2 \cdot (1 - CB_{ij})} \cdot e^{\beta_3 \cdot (1 - CL_{ij})} \cdot e^{\beta_4 \cdot (1 - CH_{ij})} \cdot e^{\beta_5 \cdot (1 - FTA_{ij})} \cdot b_{ij}, \qquad (3)$$

---

[3] Related theoretical derivations of a gravity equation for trade can be found in earlier literature as well, such as Bergstrand (1985) and Bröcker (1989).

where *D* stands for geographical distance; *CB* stands for an indicator variable equal to 1 if two countries share a (land) border (and zero otherwise); *CL*, *CH* and *FTA* are a set of similar indicator variables indicating whether or not two countries share a common official language, common colonial history, and/or common free-trade agreement. The parameter b_{ij} reflects the impact of all remaining bilateral trade barriers on the bilateral trade cost factor, assumed independent from the included covariates. Based on economic intuition, we expect positive parameters for the covariates in the trade cost function.

Bilateral export does not depend on only bilateral trade cost and the (exogenously given) size of the trading economies. It also depends on the weighted average trade costs that an exporter and importer face in their export and import market, respectively. This is reflected by the MR terms entering the denominator of equation (1). Anderson and van Wincoop (2003) derive the set of equations for the MR terms Π_i and P_j,

$$\Pi_i^{1-\sigma} = \sum_{j=1}^{N}\left(\theta_j\left(\frac{t_{ij}}{P_j}\right)^{1-\sigma}\right), \tag{4}$$

$$P_j^{1-\sigma} = \sum_{i=1}^{N}\left(\theta_i\left(\frac{t_{ij}}{\Pi_i}\right)^{1-\sigma}\right), \tag{5}$$

where $\theta_i = y_i/y_w, \forall i.$

Note that the outward (inward) resistance term includes the GDP-share-weighted average of bilateral trade costs relative to the inward (outward) resistance terms across destinations (origins). Given bilateral trade costs t_{ij}, a high value for MR implies that other countries *k* are less attractive trading partners. Hence countries *i* and *j* will trade more with each other, as shown in equation (1).

2.2 Practical Gravity

The theoretical gravity model conveys an important message. Trade flows are not mutually independent. For a consistent econometric estimation of the parameters in the model, problems emerge if the regressor variables are correlated with the residuals. The theoretical model shows that this endogeneity bias is likely to emerge if we do not control for country-specific MR.

Despite the prominent position of this theoretical framework over the past years, many empirical studies continued to rely on a more pragmatic empirical gravity equation instead. Several plausible explanations for this come to mind. Estimating a theoretically consistent

gravity equation involves dealing with Equations (4) and (5), which are nonlinear in parameters. Developing the required estimation procedures involve some restrictive assumptions (see Baldwin and Taglioni 2006; Balistreri and Hillberry 2007), and work on deriving an analytical solution has only recently emerged (e.g., Straathof 2008).

Furthermore, the theoretical framework puts restrictions on the empirical specification that follows from the stylized model rather than from practical considerations. In fact, trade depends proportionately on the GDP of an origin and destination. Moreover, GDP variables are the only push and pull factor in the model to explain total external trade. While the theoretical model requires total exports to sum to an exporter's GDP, and total imports to sum to an importer's GDP, these constraints do not hold in practical applications.[4] First, trade and GDP are measured in different units. While trade is measured in gross output values, GDP is a measure of value added. Moreover, the model includes intranational trade while most practical applications only consider international trade flows in estimating the gravity equation, due to data limitations. This context implies that theoretically imposed constraints in the model are not generally valid in estimation. Second, the share of external trade in total expenditure and gross output may be different from the predictions in the theoretical model. The theoretical gravity model predicts that larger economies are less open to international trade and allocate a larger share of their expenditure on intranational trade, but the share of international trade on GDP is often constrained to a constant by imposing proportionality between the former and the latter.

Hence, practical considerations may provide a valid motivation to choose an unconstrained empirical gravity equation, which allows more flexibility in specification. An empirical gravity equation can include additional push and pull factors to capture variation in openness to international trade. For example, we may think of per capita income, landlockedness, and land area as factors determining a country's openness to international trade. Many of these variables have been used in empirical specifications of the gravity model for international trade (e.g., Frankel 1997; Raballand 2003; De Groot et al. 2004).

Taking theoretical and practical insight seriously, ideally we would need to combine the flexibility of the empirical gravity equation and the insights about omitted variable bias due to MR of the theoretical foundation of gravity. An often used practical solution to deal with country-specific omitted variable bias is to include country-specific indicator variables in the

---

[4] The MR terms obtained impose the constraints: $\sum_i x_{ij} = y_j$ and $\sum_j x_{ij} = y_i$. In similar applications of the model in regional science, this type of specification is known as a doubly-constrained gravity model (e.g., Wilson 1970; Fotheringham and O'Kelly 1989).

gravity equation (Bröcker and Rohweder 1990). As argued by Feenstra (2004), a model specification that includes origin- and destination-specific intercepts is consistent with theoretical concerns. Moreover, this solution has been widely applied in regional science to deal with the practical problems of estimating a gravity equation in which the total flows are not known (Sen and Smith 1995).[5] This solution is not completely satisfactory, though. It is rather drastic medicine to cure the patient. First, including origin- and destination-specific indicator variables reduces the statistical efficiency of econometric estimation. Second, it precludes the analysis of country-specific determinants of trade, which are interesting for empirical applications, bthey explain cross-country variation in openness to international trade.

2.3 Consistent Estimation and Autocorrelation

The main insight from theoretically derived gravity is that regressor variables and residuals in the unconstrained gravity equation are likely to be correlated, because bilateral trade barriers also appear in the omitted MR terms. In empirical estimations, failure to control for MR might result in omitted variable bias in the parameter estimates of the bilateral regressors.

This paper proposes an alternative estimation approach that allows for the estimation of an unconstrained empirical specification of the gravity model, including push and pull factors, while offering a correction for origin- and destination-specific omitted variable bias. The approach starts from a specific interpretation of endogeneity bias as resulting from autocorrelation in trade flows. The argument for this interpretation has been made before in Behrens et al. (2012) and in Koch and LeSage (2009), and more generically relates to the recent revival in modelling SAC in bilateral flow data in the previously mentioned regional science literature. To the best of our knowledge, however, this paper is the first to link the theoretical MR effects to origin- and destination-specific filters, and to make use of SF techniques to accommodate autocorrelation in trade flows.

The argument starts by inspecting Equations (4) and (5). We propose that countries located in close spatial proximity tend to have similar MR. Similar geographical location implies similar geographical distance to trade partners across the world and a higher probability of shared neighbours. Likewise, shared languages tend to be more similar for countries closely

---

[5] Although total international trade by country is generally known, or can be proxied by summing available bilateral flows, we do not have comparable direct observations for intranational trade. Hence, we would need to proxy for openness to trade of each country in estimating the gravity equation. This can be done either by including (additional) push and pull factors in the specification, or by using country-specific intercepts.

located in space. Also, the logic of regional integration implies a higher likelihood of proximate countries being part of shared trade agreements with surrounding countries. This context implies that these spatial patterns in MR would induce autocorrelation in the residuals of the unconstrained gravity equation. As a result, the residuals and the bilateral trade cost variables are correlated, because similar reasoning to the preceding discussion suggests SAC would be in the regressor variables distance, contiguity, language and trade agreement. Omitted variable bias would result.

3. Recent Developments in Estimating the Theoretical Gravity Model of Trade

The theoretical gravity model shows that consistent estimation of the parameters requires us to take into account the price indices. As discussed in Feenstra (2004), the computational complexity of the non-linear estimation procedure has prevented its widespread use in the applied international trade literature. Still, Anderson and van Wincoop (2003) show that estimation of the more traditional empirical gravity equation (omitting the MR terms) yields inconsistent parameter estimates for the key regressor variables. A simple solution that results in consistent parameter estimates is to use a set of country-specific indicator variables for the exporting and importing countries (Bröcker and Rohweder 1990; Feenstra 2004). The indicator variables capture the country-specific MR terms, and control for omitted variable bias related to the country-specific intercepts. The main advantage of this formulation is that the resulting specification can be estimated by familiar methods such as OLS or Poisson regression.

However, the disadvantage of this solution is that the parameters of country-specific determinants of trade cannot be estimated. Variables such as GDP, per capita income, landlockedness, and land area are captured by the country-specific indicator variables. Still, empirical estimation of the effect of these variables may be relevant depending on the topic under investigation. Hence, a solution that would share the basic simplicity of estimation with the indicator variable specification, while allowing retention of the country-specific regressors, is needed.

Several recent developments in the trade gravity model literature focus on combining consistent estimation and flexibility in the specification of the gravity equation. Egger (2005) argues that a Hausman-Taylor approach, which allows for country-specific covariates, is consistent even if unobserved country-specific heterogeneity exists. This formulation provides an alternative to the indicator variables specification that controls for omitted variable bias due

to omitted MR terms, and allows for the estimation of the parameters related to the country-specific variables. The method is based upon an approach similar to instrumental variables, which relies on instruments from inside the model.

In contrast, Baier and Bergstrand (2009) log-linearize the MR terms using a first-order Taylor series approximation. This yields *exogenous* bilateral multilateral-world-resistance (MWR) variables that proxy the endogenous country-specific MR variables in Anderson and van Wincoop (2003). The resulting reduced-form gravity equation can be estimated with OLS. This method is termed *bonus vetus* ('good-old') OLS (BV-OLS). The approach yields log-linear approximations of the MR terms, using Taylor series expansion around a centre of identical and symmetric trade costs, $t_{ij} = t$, but differing economic sizes ($\theta_i = y_i/y_w$).

Starting from a reformulated Equation (1):

$$\ln x_{ij} = -\ln y_w + \ln y_i + \ln y_j - (\sigma-1)\ln t_{ij} + (\sigma-1)\ln P_i + (\sigma-1)\ln P_j, \tag{6}$$

the equation that Baier and Bergstrand derive is:

$$\ln x_{ij} = -\ln y_w + \ln y_i + \ln y_j - (\sigma-1)\ln t_{ij} + (\sigma-1)\left[\left(\sum_{j=1}^{N}\theta_j \ln t_{ij}\right) - \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\theta_i\theta_j \ln t_{ij}\right)\right]$$
$$+ (\sigma-1)\left[\left(\sum_{i=1}^{N}\theta_i \ln t_{ji}\right) - \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\theta_i\theta_j \ln t_{ij}\right)\right]. \tag{7}$$

The terms in square brackets are the MR terms. They contain a first component that captures multilateral trade frictions for each exporting or importing country, relative to a second part that reflects world trade costs.

A third approach to the consistent cross-sectional estimation of the gravity model is proposed in Behrens et al. (2012). Their approach is closely related to our approach. Starting from the Anderson and van Wincoop formulation of the theoretical gravity equation, they show that the MR terms can be shown to reflect a correlation structure between trade flows that can be modelled similarly to SAC. They suggest a spatial-autoregressive moving-average specification for the gravity model, which results in consistent estimates of the standard gravity equation parameters. At the same time, they argue that the baseline fixed-effects specification discussed previously does not fully succeed in capturing the MR dependencies in the error structure

introduced by the general equilibrium nature of trade patterns modelling, and that its residuals still show a significant amount of autocorrelation (Behrens et al. 2012).

We now proceed to discuss the methodology followed in this paper. The alternative we propose, SF, combines two attractive features. First, it is fairly simple to apply, much like OLS with indicator variables; second, it takes into account the general equilibrium interdependence of trade flows that can be modelled as SAC, like spatial econometric origin-destination specifications.

4. Proposed Methodology: Spatial Filtering Estimation

The theoretical gravity model includes origin- and destination-specific MR variables that reflect the export and import accessibility of countries. Omitting these endogenous MR variables from the specification results in potential omitted variable bias, both for the trade cost variables and for the size variables in the gravity equation. Consistent estimation requires some way to capture the endogeneity between MR terms and standard regressors. We propose to make use of the fact that this dependency structure is likely to manifest as SAC in the residuals of the traditional specification of the gravity model. The reasoning is that many trade cost variables, such as geographical distance, adjacency, trade agreements, and common language, are spatially correlated: countries close in space are more likely to share the same (or similar) characteristics. This context likewise implies that both inward and outward accessibility are spatially correlated: close countries are likely to have more similar accessibility. We deal with SAC by using an origin- and a destination-specific spatial filter, which serve to capture the spatially autocorrelated parts of the residuals.

When including these spatial filters as additional origin- and destination-specific regressors (much like the origin and destination specific MR variables), 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 (see Griffith 2007; Fischer and Griffith 2008).

Basically, SF estimation of georeferenced data regressions (such as international trade) can reduce to defining a geographically varying mean and a variance on the basis of an exogenous spatial weights matrix. In other words, the spatially correlated residuals from an otherwise non-

spatial regression model are partitioned into two synthetic variables: (i) a spatial filter which captures latent SAC; and, (ii) a non-spatial variable (free of SAC), which will be the newly obtained residuals. The workhorse for this SF decomposition is a transformation procedure based upon eigenvector extraction from the matrix

$$(\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/n) \, \mathbf{W} \, (\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/n), \tag{8}$$

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 Moran's coefficient (MC), which is a commonly used measure of SAC (see Section 5).

The eigenvectors of Equation (8) 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 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 negative MC.

Having extracted the eigenvectors of Equation (8), a spatial filter is constructed by judiciously selecting a subset of these n eigenvectors. In detail, for our empirical application, we select a first subset of eigenvectors (which we will call 'candidate eigenvectors') by means of the following threshold: $\mathrm{MC}(e_i)/\mathrm{MC}(e_1) > 0.25$. This threshold yields a spatial filter that approximately replicates the amount of variance explained by a spatial autoregressive model (SAR) (Griffith 2003).[6] 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, both in autoregression and in traditional modelling terms, 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).

---

[6] 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.

The added challenge, with regard to the case at hand, is that trade data do not represent points in space, but flows between points. Therefore, the eigenvectors are linked to the flow data by means of Kronecker products: the product $\mathbf{E_K} \otimes \mathbf{1}$, where $\mathbf{E_K}$ 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 $\mathbf{1} \otimes \mathbf{E_K}$ may be linked to destination-specific information (again, for example, the GDP of importing countries) (Fischer and Griffith 2008). As a result, we have two sets of origin- and destination-specific variables, 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 main advantages of the proposed estimation method are: (a) this approach can be applied to *any* type of regression, including simple OLS and generalized linear models (GLMs) such as Poisson or negative binomial regressions (although auto-Poisson and auto-negative binomial specifications cannot describe positive spatial dependence), for which usually dedicated spatial econometric applications do not exist; (b) by avoiding the use of indicator variables, we are able to save degrees of freedom, and, (c) the approach can be used to estimate regression parameters for origin- and destination-specific variables, such as GDP or trade agreements indicators.

For our case study, because of the nature of trade data, as suggested by Santos Silva and Tenreyro (2006), we estimate a count data model. While the natural choice would be Poisson regression, in order to take into account overdispersion in the data due to unobserved heterogeneity (which results in a sample variance which is much greater than the sample mean), we estimate a negative binomial model, which can explicitly account for such overdispersion by iteratively estimating the dispersion parameter. In subsequent comparisons regarding residual spatial autocorrelation, we consider, for the SF models, quasi-Poisson estimations as well.

5. **Spatial Autocorrelation Diagnostics**

When employing GLMs, traditional SAC indices may not be appropriate, as discussed below. In this section, we review the available alternatives.

In linear regression contexts, when analysing model residuals, an adapted Moran test (Cliff and Ord 1972; 1981) is commonly used, under a standard assumption of normality. A *t* test can

be used to test the null hypothesis of spatial randomness of the residuals. The formula for the MC computed on the residuals is the following:

$$I = \frac{n\sum_{i,j} w_{ij}\varepsilon_i\varepsilon_j}{S_0\sum_i \varepsilon_i^2},$$ (9)

where w_{ij} is the (i,j) element of a chosen spatial weight matrix $\mathbf{W}$, ε_i and ε_i are the related model residuals, and S_0 is the sum of all elements of $\mathbf{W}$. The expected value of this index is:

$$E(I) = \frac{n\operatorname{tr}(\mathbf{A})}{S_0(n-k)},$$ (10)

where $\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{X}$ corresponds to the factor that accounts for the effect of the independent variables. $\mathbf{X}$ is the n x k matrix containing the values of the k independent variables included in the regression model.

A permutation-based Moran test has also been proposed (Cliff and Ord 1981) in order to improve the results of the approximate t test and to gain insights in its sampling distribution under spatial randomness.

Because the Moran test has been developed for linear models and normally distributed residuals, the use of MC calculated on the residuals of count data (Poisson, negative binomial) regression models is questionable (Schabenberger and Gotway 2005, p. 377), despite recent literature agrees that it possesses good power against a wide array of autoregressive models and different distributions of the residuals (Anselin and Rey 1991).

Griffith (2010) studies the behaviour of the MC for non-normal random variables, and shows that, above moderate values of n (25–100), the MC is a suitable indicator in these cases as well. However, Griffith does not study the case of SAC diagnostics for regression residuals, in which we can consider the effect of the independent variables in the model.

Further, Moran's test may not be properly applied to the residuals of Poisson or negative binomial regression, whose distributional properties are not well known. In addition, because the test does not consider the heterogeneity of observations, its standard moments may not be appropriate under heteroscedasticity. For more details, one can refer to Oden (1995), who discusses this problem.

Lin and Zhang (2007) suggest that the MC can be used to test the residuals of a Poisson model by employing Pearson or deviance residuals under an asymptotic normality assumption. This

approach is followed, among others, by Scherngell and Lata (2013), which employ a panel SF modelling approach. However, this permutation test once again does not incorporate the effect of the independent variables of the model in constructing a reference distribution.

Fortunately, the standardized t statistic of Jacqmin-Gadda et al. (1997) can be applied in this context. This t statistic can be considered as an extension of standard SAC statistics into the domain of GLMs. It is derived in an analogous way to a score test based on generalized estimating equations (Prentice and Zhao 1991). As the condition of validity of the above test does not always hold, since the computation is intractable for large samples, a test based on the permutation distribution has been also proposed by the same authors.

Under the null hypothesis of no spatial autocorrelation, the t statistic is defined as:

$$t = \sum_{i=1}^{n} \sum_{j \neq i} w_{ij} (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_j), \tag{11}$$

or, in matrix notation:

$$t = (\mathbf{Y} - \hat{\boldsymbol{\mu}})^T \mathbf{W}(\mathbf{Y} - \hat{\boldsymbol{\mu}}), \tag{12}$$

where $\mathbf{Y}$ is the n x 1 vector of the observations of the dependent variable, and $\hat{\boldsymbol{\mu}}$ is the n x 1 vector of the estimated means.

Using a first-order Taylor series expansion for the deviation of estimated means from the true means, Jacqmin-Gadda et al. (1997) show that the index's expectation and variance are as follows:

$$E(t) = \text{tr}(\mathbf{RD}); \tag{13}$$

$$\text{var}(t) = \sum_{i=1}^{n} R_{ii}^2 (\mu_{i(4)} - \mu_{i(2)}^2) + 2\text{tr}(\mathbf{RDRD}), \tag{14}$$

where $\mathbf{R} = \mathbf{M}^T\mathbf{WM}$, $\mathbf{M} = \mathbf{I} - \mathbf{DX}(\mathbf{X}^T\mathbf{DX})^{-1}\mathbf{X}^T$, and $\mathbf{D}$ is the diagonal matrix whose elements are the variance of each observation. Consequently, R_{ii}^2 is the ith diagonal element of matrix $\mathbf{R}$, while $\mu_{i(2)}$ and $\mu_{i(4)}$ are the second and the fourth central moments of the ith observation, respectively. Jacqmin-Gadda et al. (1997) show that the standardized t statistic asymptotically follows the standard normal distribution.

The Jacqmin-Gadda (JG) test is a development of the statistic developed by le Cessie and van Houwelingen (1995), similarly derived as a score test in the spirit of Prentice and Zhao (1991), but not accounting for the effect of the independent variables. In fact, referring to Equation (13), the component **R** in the le Cessie (LC) test is reduced to **R** = **W**^{T}**W**, while **D** = cov(**Y**)**.** In other words, the LC test does not incorporate the adjustment of estimating parameters, that is, the effect of independent variables is not considered in constructing a reference distribution. In summary, using the JG standardized t statistic, a test for spatial autocorrelation in the context of GLMs can be carried out.

6. Empirical application

We apply the SF estimation to a cross-section of bilateral trade flows between 64 (major trading) countries for the year 2000 (a full list of countries is provided in the Appendix, Table A.**1**). In this section, we discuss the empirical specification, data and the estimation results.

6.1 Data and Model Specification

For estimation, we follow a standard specification of the gravity equation of bilateral trade. Starting from the trade costs variables identified in equation (3), we further extend the specification with additional variables commonly mentioned in the literature (see, e.g., Frankel 1997; Raballand 2003). We use the following standard specification of the gravity equation:

$$
\begin{aligned}
\ln X_{ij} - \ln(GDP_i \cdot GDP_j) &= \alpha_0 + \alpha_1 \cdot \ln(GDPCAP_i \cdot GDPCAP_j) + \beta_1 \cdot \ln(D_{ij}) \\
&+ \beta_2 \cdot CB_{ij} + \beta_3 \cdot CL_{ij} + \beta_4 \cdot CH_{ij} + \beta_5 \cdot FTA_{ij} + \beta_6 \cdot ISL_i + \beta_7 \cdot ISL_j + \beta_8 \cdot \ln(Area)_i \\
&+ \beta_9 \cdot \ln(Area)_j + \beta_{10} \cdot LL_i + \beta_{11} \cdot LL_j + \delta_2 \cdot MWRCB_{ij} + \delta_3 \cdot MWRCL_{ij} \\
&+ \delta_4 \cdot MWRCH_{ij} + \delta_5 \cdot MWRFTA_{ij} + \varepsilon_{ij},
\end{aligned}
\tag{15}
$$

where *GDPCAP* represents per capita GDP, *ISL* is an indicator variable that equals 1 if the country is an island, *Area* is the land area of a country, and *LL* equals 1 for landlocked countries, and in which:

$$
MWRCB_{ij} = \left[\left(\sum_{j=1}^{N} \theta_j CB_{ij} \right) + \left(\sum_{i=1}^{N} \theta_i CB_{ij} \right) - \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \theta_i \theta_j CB_{ij} \right) \right],
\tag{16}
$$

and likewise for the remaining *MWR* variables. The other variables are as defined earlier. The product of origin and destination GDPs is used as an offset variable.

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.[7] 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.[8] 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).

We first estimate Equation (15) using negative binomial regression including country-specific indicator variables. GDP is used as an offset, which implies we move the log-sum of GDP to the left handside, assuming it has a proportional effect on trade with elasticity equal to 1 (Anderson and van Wincoop 2003). This is our first benchmark model, which, according to Feenstra (2004), yields consistent parameter estimates, but is criticized by Behrens et al. (2012). Secondly, we estimate Equation (15), extending it with approximations of MR terms obtained using the Taylor series approximation proposed by Baier and Bergstrand (2009). This is our second benchmark model. These results, as well as the ones for the SF approach, are discussed in Section 6.2.

6.2 Estimation Results: Spatial Filtering and Benchmark Models

The first benchmark model includes origin- and destination-specific indicator variables. As shown in Anderson and van Wincoop (2003) and Feenstra (2004), this specification accounts for MR terms, and yields consistent parameter estimates. The disadvantage is that country-specific variables cannot be included, as their effect cannot be identified separately. This implies that explanatory variables that are potentially relevant for explaining variation in bilateral trade patterns, such as GDP per capita, land area and landlockedness, cannot be investigated empirically (if not ex post, by, e.g., regressing the indicator variable coefficients on them). A second disadvantage is the loss of degrees of freedom for estimation, because a substantial number of indicator variables ($2n - 2$) is needed. Usually, however, the degrees of freedom remain large enough, since observations are bilateral (i.e., $n^2 - n$).

---

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

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

The second benchmark model is the specification developed in Baier and Bergstrand (2009), which includes first-order Taylor series approximations of the MR terms. This specification follows from Equation (6). Further manipulation [substituting Equation (3) for bilateral trade costs] allows us to combine both terms between square brackets into a set of bilateral variables, one for each bilateral trade costs variable determining trade costs (such as geographical distance). The reduced-form double-log gravity equation is as follows:

$$
\begin{aligned}
\ln X_{ij} - \ln(GDP_i \cdot GDP_j) = {} & \alpha_0 + \alpha_1 \cdot \ln(GDPCAP_i \cdot GDPCAP_j) + \beta_1 \cdot \ln(D_{ij}) + \beta_2 \cdot CB_{ij} + \beta_3 \cdot CL_{ij} \\
& + \beta_4 \cdot CH_{ij} + \beta_5 \cdot FTA_{ij} + \beta_6 \cdot ISL_i + \beta_7 \cdot ISL_j + \beta_8 \cdot \ln(Area)_i + \beta_9 \cdot \ln(Area)_j + \beta_{10} \cdot LL_i + \beta_{11} \cdot LL_j \\
& + \delta_2 \cdot MWRCB_{ij} + \delta_3 \cdot MWRCL_{ij} + \delta_4 \cdot MWRCH_{ij} + \delta_5 \cdot MWRFTA_{ij} + \varepsilon_{ij},
\end{aligned}
$$

(17)

Baier and Bergstrand (2009) show that theory imposes the restrictions $\delta_k = -\beta_k$ for each k.

The equations specify the model in double-logarithmic transformation. We estimated the benchmark models multiplicatively, using negative binomial regressions, aside from the BV model, which is estimated linearly. This method allows a direct treatment of the non-negative values of trade flows and of the zeros, and enables us to correct for overdispersion of trade flows (see Santos Silva and Tenreyro 2006).

The empirical estimation results are presented in Table 1. Model (1) presents the regression results for the first benchmark model, including country-specific indicator variables. Following Anderson and van Wincoop (2003), we estimate the model using GDP as an offset variable (i.e., restricting the coefficient of GDP variables to equal 1). The parameter estimates are in line with the findings elsewhere in the literature (see, e.g., Anderson and van Wincoop 2004; Disdier and Head 2008). Geographical distance has a negative effect on trade, with an estimated elasticity of –1.30. The effect of proximity on trade is reinforced by a positive and (marginally) significant effect of contiguity on trade. Proximity in terms of language and colonial links also positively affects bilateral trade, while preferential trade policy (i.e., enjoying common FTAs), appears to have a counterintuitive negative effect. These results – with the exception of the latter – confirm previous findings about the importance of these dimensions of transactional distance on trade (e.g., Obstfeld and Rogoff 2000; Loungani et al. 2002).

Model (3) compares these findings with the regression outcomes for the second benchmark model, the Baier-Bergstrand estimation. This method proxies for the endogenous and unobserved MR terms by including exogenous linear approximations based upon bilateral trade costs variables. Provided that the approximation is sufficiently adequate, this specification

results in consistent estimates (Baier and Bergstrand 2009). Once again, GDP has been used as an offset variable, and the model is estimated by OLS. The obtained parameter estimates are comparable to the estimates for the first benchmark model [Model (1)], including the negative effect found for free-trade blocs. Additionally, on the one hand, the Baier-Bergstrand specification has an advantage, because it enables us to include country-specific regressors explicitly; on the other hand, the results do not always appear to be satisfactory.

Table 1. Estimation results

	(1) Fixed effects (GDP offset)	(2) Spatial filter	(3) BB-estimation (GDP offset)	(4) BB-estimation
Distance	-1.30^{***}	-1.23^{***}	-1.25^{***}	-1.22^{***}
Common border	0.24^{*}	0.33^{**}	0.23	0.25
Common language	0.36^{***}	0.33^{***}	0.32^{***}	0.37^{***}
Common history	0.86^{***}	0.71^{***}	0.79^{***}	0.80^{***}
Free trade	-0.14^{**}	0.41	-0.27^{***}	-0.22^{**}
GDP exporter	–	0.75^{***}	–	0.91^{***}
GDP importer	–	0.92^{***}	–	1.15^{***}
GDP per cap. exporter	–	0.13^{***}	-0.06^{**}	0.02
GDP per cap. importer	–	0.12^{***}	-0.04^{*}	-0.16^{***}
Island exporter	–	-0.41^{***}	-0.29^{***}	-0.28^{***}
Island importer	–	-0.31^{***}	0.08	0.20^{*}
Area exporter	–	-0.00	-0.11^{***}	-0.07^{***}
Area importer	–	-0.17^{***}	-0.22^{***}	-0.28^{***}
Landlocked exporter	–	0.23^{*}	0.30^{**}	0.26^{**}
Landlocked importer	–	-0.58^{***}	0.07	0.19^{*}
Constant	-29.60^{***}	-27.42^{***}	-34.01^{***}	-34.97^{***}
AIC	101,713	47,805	102,485	102,436
Observations	4032	4032	4032	4032

Notes: BB stands for Baier-Bergstrand, and AIC for Akaike information criterion. ***, **, * denote parameter estimates statistically significant at 1%, 5% and 10%, respectively.

Closer inspection of the Baier-Bergstrand estimation, dropping the offset assumption on the product of exporter and importer GDP in Model (4), yields qualitatively similar –and in some cases more plausible (e.g., for landlocked importers) – results, and a slightly better fit. For example, although a negative effect of GDP per capita variables on trade is not uncommon in some specifications (see, e.g., Anderson and Marcouiller 2002), the effect in Model (3) seems to be driven mainly by offsetting GDP, which imposes a GDP elasticity of trade (of 1), which empirically is too high.

Summarizing, the two benchmark models yield somewhat different results. Although, as mentioned, some effects may be more plausible in the Baier-Bergstrand estimation results, the more traditional specification using country-specific indicator variables results in a slightly better model likelihood, as shown by the Akaike information criterion (AIC). The disadvantages of this model, though, are the loss of country-specific variables, and a diminished precision in the determination of the significance of variables, resulting from the loss of degrees of freedom in the model estimation.

Results emerging from the SF estimation of the gravity model, which combines the consistent estimation of the first benchmark model with the flexibility of specification of the second benchmark model, are shown for Model (2) in Table 1. The results presented here are obtained for a simmetrized k-nearest neighbours[9] spatial weights matrix **C**, and for a negative binomial estimation, employed in order to cope with overdispersion in the trade flows. With regard to the coefficients of bilateral resistance variables, we note that with the exception of the one for FTA, they are highly significant, and their values are consistent with the ones found for Model (1). The FTA coefficient not being significantly negative anymore might be seen as a result that is more consistent with theoretical expectations. With regard to the importer- and exporter-specific variables, we are able to identify highly significant and positive coefficients for GDP, and GDP per capita is now significant and positive as well in both cases. This result is in contrast with the ones for the Baier-Bergstrand benchmarks [Models (3) and (4)], in which the same variable is either not significant or significantly negative. The SF estimation also allows us to estimate significant parameters for the variables identifying the geographical characteristics of importer and exporter countries. The signs obtained are mostly consistent with the ones found

---

9   For the k-nearest neighbours definition of proximity, each country's neighbours are defined by selecting the k closest countries. Distance between the geographical centroids of the countries was used (Great Circle), setting $k = 3$ and forcing, for computational reasons, symmetry of the spatial weights matrix. As a result, the minimum number of neighbours per country is 3, while the maximum number is not constrained. Alternative definitions of proximity based upon, for example, simple rook contiguity or distance decay could be tested in order to assess the sensitivity of the model to the choice of spatial specification.

for the Baier-Bergstrand benchmarks. They show that larger and both landlocked and island countries tend to trade less. Noteworthy differences between the SF model and the benchmarks regard the negative and significant coefficients obtained for the importing patterns of island and landlocked economies (it was marginally positive or non-significant for the benchmarks). For islands, it may seem counterintuitive to find this result, although it should be considered that the sample of countries used excludes, because of non-reporting, most micro-island countries, while includes all large island countries like the UK or Japan. In contrast, in the case of landlocked countries, a negative importing coefficient is more consistent with theoretical expectations.

Finally, the AIC of the SF model appears to vastly improve on the ones of the benchmark models, because of the high amount of variance explained by the origin- and destination-specific spatial filters, which are also highly significant from a statistical viewpoint (not shown in Table 1). In summary, the proposed SF approach to the estimation of a gravity model of trade allows identification of the regression parameters related to the bilateral variables, as well as those related to the origin- and destination-specific variables. Moreover, the model has a better likelihood (leading to improved AIC) than the competing models tested, and uses a limited number of degrees of freedom.

6.3 Testing for Spatial Autocorrelation

In Section 5, we discussed SAC statistics based on the score test [by le Cessie and van Houwelingen (1995) and Jacqmin-Gadda et al. (1997)], which are alternative to the traditional MC in case of GLMs, since the MC statistical distribution theory has been developed under linear regression assumptions.

Having an n^2 x n^2 spatial weight matrix (obtained as $\mathbf{W} \otimes \mathbf{W}$) and the t statistic by Jacqmin-Gadda et al. (1997), residual SAC in Poisson and negative binomial regressions can be modelled by eigenvector SF within the same framework as standard spatial autocorrelation in regression residuals. The eigenvectors employed in Model (2) (see preceding section) represent a certain level of SAC, given a spatial connectivity pattern, and by including them as proxy variables for such spatial autocorrelation, SAC that is not explained by independent variables is expected to be filtered out (at least partially) of the residuals, and transferred to the mean response.

Because eigenvectors are introduced as independent variables in a (forward or backward) stepwise manner, the adjustment of estimating parameters for independent variables developed in the Jacqmin-Gadda test seems to be desirable. Chun (2008) performs the test to evaluate SAC in a Poisson model in an analysis of migration flows. To the best of our knowledge, no one so

far has used the test on a negative binomial model. We performed both the aforementioned score tests described in Section 5 to empirically detect the presence (or the absence) of SAC. We compare the tests on the model augmented with selected SF variables with the ones on the non-filtered model to verify if the introduction of the selected spatial filters lets the SAC be filtered out of the residuals. The tests are calculated on both quasi-Poisson[10] and negative binomial model residuals (estimating or offsetting GDP benchmark variables).

A further relevant question is whether adjusting the test for the presence of independent variables considerably changes SAC detection outcomes, or if this correction has just marginal effects. Table 2 presents the results for the different SAC tests. We start by reporting, in the first and second row of the table, the value of MC computed on the residuals as developed by Cliff and Ord for linear models. In the first row, we show the results of the basic, stand-alone, MC, while in the second row, the test accounts for the effect of independent variables. The presence of SAC is never rejected, even when we introduce the spatial filters in the model (despite the scores decreasing). Performing also the discussed MC permutation test, our findings do not change: the permutation score decreases adding the spatial filters, but we never reject the SAC hypothesis.[11] In the third and the fourth rows of Table 2, the values for the LC and JG tests are reported. Using these tests, developed for GLMs, we can note how the SAC is effectively filtered out by the introduction of the selected spatial filters. The tests show significant SAC in the baseline model, which is filtered out by the spatial filter eigenvectors, especially when using negative binomial regression, for which the p-value stands to 0.239 (0.230 with offsets). Moreover, the results from LC and JG are quite similar, highlighting that the introduction of the correction for the independent variables in JG test does not considerably change the test results. The general increase in *t*-scores obtained when the right-hand-side variables are taken into account may be explained by the fact that their inclusion pulls expected values slightly to the left (towards negative values).

These results seem to be comforting, and they lead to a positive confirmation of the initial theorized idea that we can account for spatial autocorrelation in the model by filtering out the

---

[10] Quasi-Poisson models are equivalent to standard Poisson models in terms of coefficient estimation, but because a dispersion parameter is estimated from the data, inference differs. For the purposes of eigenvector selection, AIC- or BIC-based selection is not possible (quasi-Poisson models have no likelihood), so it is manually performed by backward eliminating (iteratively) the eigenvector with the highest p-value.

[11] Results of the Moran permutation tests are available upon request.

residual spatial component by means of the selected spatial filters, and that this is detectable only using correct SAC test (specifically designed for GLMs).

Table 2. SAC with different statistics, for different models

		Quasi-Poisson		Negative binomial		Negative Binomial (offset)	
		Non-spatial	Spatial filter	Non-spatial	Spatial filter	Non-spatial	Spatial filter
MC	Score	0.212	0.129	0.185	0.043	0.158	0.035
	t	39.99	24.61	34.93	8.08	30.06	6.74
	p-value	<2.2e–16	<2.2e–16	<2.2e–16	<2.2e–16	<2.2e–16	8.09e–12
MC	Score	0.168	0.119	0.429	0.277	0.375	0.283
(res.)	t	31.206	21.853	79.823	54.496	72.249	55.086
	p-value	< 2.2e–16	< 2.2e–16	< 2.2e–16	< 2.2e–16	< 2.2e–16	< 2.2e–16
LC	t	4.962	1.971	3.218	0.652	4.601	0.683
	p-value	3.49e–07	0.024	0.001	0.257	2.10e–06	0.247
JG	t	5.125	2.111	3.336	0.708	4.766	0.737
	p-value	1.49e–07	0.017	< 0.001	0.239	9.41e–07	0.230

Notes: MC stands for the standalone Moran's I test, MC (res.) for the Moran's I test on regression residuals, LC for the le Cessie test, and JG for the Jacqmin-Gadda test.

7. Conclusions

Recent contributions to the modelling of bilateral trade have shown the importance of sound theoretical underpinnings for obtaining consistent parameter estimates for the determinants of trade in the gravity model of bilateral trade. This paper addresses the issue of how to achieve empirical consistency without the need to estimate a full general equilibrium system of equations, and without the loss of specification flexibility that results from the use of origin- and destination-specific indicator variables. We argue that endogeneity of regressors and residuals – due to omitted MR variables in the traditional gravity model – is likely to manifest in the form of autocorrelation in both regressors and residuals. By including an origin-specific and a destination-specific spatial filter as additional regressors, SF estimation of the gravity equation enables us to filter SAC out of the residuals, as demonstrated by the results obtained implementing appropriate SAC tests for nonlinear models. As a result, the residuals and the regressors are no longer correlated, and standard estimation methods can be applied to obtain consistent parameter estimates for the determinants of bilateral trade. We demonstrate the use

of SF estimation in a negative binomial estimation of the gravity equation of bilateral trade. The comparison with two benchmark models, which are theoretically consistent in estimation, reveals that SF yields results that are highly comparable to the estimation using country-specific indicator variables. Moreover, SF estimation does not suffer from the drawbacks of using indicator variables. It allows explicit estimation of the effect of country-specific variables that are potentially important determinants of bilateral trade, such as GDP, per capita GDP and landlockedness.

Further analyses aimed at measuring the extent to which SAC is filtered out in SF estimation. We tested three different SAC tests, either from the linear modelling tradition (Moran's I tests) or specifically developed for GLMs (the le Cessie and Jacqmin-Gadda tests) on both quasi-Poisson and negative binomial model estimations. Our results confirm the 'filtering' effect of the spatial filters on the residuals. Such finding is mostly evident on the GLM tests, which can be expected to be more suitable for analysing our models' residuals. On the other hand, the inclusion of right-hand-side variables in the computation of the SAC tests does not appear to considerably change our findings.

Future research should focus, on the methodological side, on expanding the analyses above to the SF network-autocorrelation approach first suggested by Chun (2008) and further employed in a panel framework (see, e.g., Scherngell and Lata 2013). Furthermore, quasi- or pseudo-Poisson estimation could be considered more extensively (as suggested in Section 7.3), by employing stepwise selection criteria which do not require likelihood-based indicators. In this regard, Krisztin and Fischer (2015) have very recently applied network-autocorrelation SF to a trade model, by including, among others, zero-inflated specifications. On the empirical side, it would be desirable to exploit the methodology proposed toward investigating specific research questions in the trade field, while a simulation study could help further extend the presented evidence on the adequacy of the SF approach for cross-sectional spatial interaction/gravity models.

International Conference on Econometrics and the World Economy (Fukuoka); SSES Annual

Meeting 2009 (Geneva).

References

Anderson JE, Marcouiller D (2002) Insecurity and the Pattern of Trade: An Empirical Investigation. Review of Economics and Statistics 84 (2):342-52

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

Anselin L, Rey S (1991) Properties of Tests for Spatial Dependence in Linear Regression Models. Geographical Analysis 23 (2):112-31

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

Baldwin R, Taglioni D (2006) Gravity for Dummies and Dummies for Gravity Equations. NBER Working Paper Series, National Bureau of Economic Research, Cambridge

Balistreri EJ, Hillberry RH (2007) Structural Estimation and the Border Puzzle. Journal of International Economics 72 (2):451-63

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

Bergstrand JH (1985) The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence. Review of Economics and Statistics 67 (3):474-81

Bröcker J (1989) How to Eliminate Certain Defects of the Potential Formula. Environment and Planning A 21 (6):817-30

Bröcker J, Rohweder HC (1990) Barriers to International Trade. Ann Reg Sci 24 (4):289-305

Chun Y (2008) Modeling Network Autocorrelation within Migration Flows by Eigenvector Spatial Filtering. Journal of Geographical Systems 10 (4):317-44

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

De Groot HLF, Linders G-J, Rietveld P, Subramanian U (2004) The Institutional Determinants of Bilateral Trade Patterns. Kyklos 57 (1):103-23

Deardorff AV (1998) Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World? In: Frankel JA (ed) The Regionalization of the World Economy. The University of Chicago Press, Chicago London, pp. 7-31

Disdier A-C, Head K (2008) The Puzzling Persistence of the Distance Effect on Bilateral Trade. The Review of Economics and Statistics 90 (1):37-48

Egger P (2005) Alternative Techniques for Estimation of Cross-Section Gravity Models. Review of International Economics 13 (5):881-91

Feenstra RC (2004) Advanced International Trade: Theory and Evidence. Princeton University Press, Princeton

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

Fotheringham AS, O'Kelly ME (1989) Spatial Interaction Models: Formulations and Applications. Studies in Operational Regional Science. Kluwer Academic Publishers, Dordrecht

Frankel JA (1997) Regional Trading Blocs in the World Economic System. Institute for International Economics, Washington

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

Griffith DA (2010) The Moran Coefficient for non-Normal Data. Journal of Statistical Planning and Inference 140 (11):2980-90

Grimpe C, Patuelli R (2011) Regional Knowledge Production in Nanomaterials: A Spatial Filtering Approach. Ann Reg Sci 46 (3):419-51

Jacqmin-Gadda H, Commenges D, Nejjari C, Dartigues J-F (1997) Tests of Geographical Correlation with Adjustment for Explanatory Variables: An Application to Dyspnoea in the Elderly. Statistics in Medicine 16 (11):1283-97

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 (2009) Latent Multilateral Trade Resistance Indices: Theory and Evidence. Université de Bourgogne, Auxerre

Krisztin T, Fischer MM (2015) The Gravity Model for International Trade: Specification and Estimation Issues. Vienna University of Economics and Business, Vienna

le Cessie S, van Houwelingen HC (1995) Testing the Fit of a Regression Model Via Score Tests in Random Effects Models. Biometrics 51 (2):600-14

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

Lin G, Zhang T (2007) Loglinear Residual Tests of Moran's I Autocorrelation and their Applications to Kentucky Breast Cancer Data. Geographical Analysis 39 (3):293-310

Loungani P, Mody A, Razin A (2002) The Global Disconnect: The Role of Transactional Distance and Scale Economies in Gravity Equations. Scottish Journal of Political Economy 49 (5):526-43

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

Obstfeld M, Rogoff K (2000) The Six Major Puzzles in International Macroeconomics: Is There a Common Cause? NBER Macroeconomics Annual 15:339-90

Oden N (1995) Adjusting Moran's I for Population Density. Statistics in Medicine 14 (1):17-26

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

Porojan A (2001) Trade Flows and Spatial Effects: The Gravity Model Revisited. Open Economies Review 12 (3):265-80

Pöyhönen P (1963) A Tentative Model for the Volume of Trade between Countries. Weltwirtschaftliches Archiv 90 (1):93-9

Prentice RL, Zhao LP (1991) Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses. Biometrics 47 (3):825-39

R Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna

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

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

Schabenberger O, Gotway CA (2005) Statistical Methods for Spatial Data Analysis. Chapman & Hall/CRC, Boca Raton

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

Sen A, Smith TE (1995) Gravity Models of Spatial Interaction Behavior. Advances in Spatial and Network Economics series. Springer, Heidelberg and New York

Straathof SM (2008) Gravity with Gravitas: Comment. CPB Discussion Paper 111, Netherlands Bureau for Economic Policy Analysis, The Hague

Tinbergen J (1962) Shaping the World Economy: Suggestions for an International Economic Policy. Twentieth Century Fund, New York

Tobler W (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography 46 (supplement):234-40

Wilson AG (1970) Entropy in Urban and Regional Modelling. Pion, London

Appendix

Table A.1. List of the countries used in the empirical application

Algeria	Angola	Argentina
Australia	Austria	Belgium
Brazil	Bulgaria	Canada
Chile	China	Colombia
Czech Republic	Denmark	Dominican Republic
Ecuador	Finland	France
Germany	Greece	Hungary
India	Indonesia	Iran
Ireland	Israel	Italy
Japan	Kazakhstan	Kuwait
Libya	Malaysia	Mexico
Morocco	Netherlands	New Zealand
Nigeria	Norway	Oman
Pakistan	Peru	Philippines
Poland	Portugal	Qatar
Romania	Russia	Saudi Arabia
Singapore	Slovakia	Slovenia
South Africa	South Korea	Spain
Sweden	Switzerland	Thailand
Tunisia	Turkey	United Arab Emirates
United Kingdom	United States	Venezuela
Vietnam		

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS