WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

ePub^{WU} Institutional Repository

James P. LeSage and Manfred M. Fischer

Conventional versus network dependence panel data gravity model specifications

Paper

Conventional versus network dependence panel data gravity model specifications

James P. LeSage
Texas State University
San Marcos, TX, USA
james.lesage@txstate.edu

and

Manfred M. Fischer
Vienna University of Economics and Business
Vienna, Austria
manfred.fischer@wu.ac.at

Abstract

Past focus in the panel gravity literature has been on multidimensional fixed effects specifications in an effort to accommodate heterogeneity. After introducing conventional multidimensional fixed effects, we find evidence of cross-sectional dependence in flows.

We propose a simultaneous dependence gravity model that allows for network dependence in flows, along with computationally efficient Markov Chain Monte Carlo estimation methods that produce a Monte Carlo integration estimate of log-marginal likelihood useful for model comparison. Application of the model to a panel of trade flows points to network spillover effects, suggesting the presence of network dependence and biased estimates from conventional trade flow specifications. The most important sources of network dependence were found to be membership in trade organizations, historical colonial ties, common currency and spatial proximity of countries.

KEYWORDS: origin-destination panel data flows, cross-sectional dependence, log-marginal likelihood, gravity models of trade, sociocultural distance, convex combinations of interaction matrices.

JEL: C18, C33, C51

1 Introduction

The panel data gravity literature has focused on multi-indexed fixed effects specifications in an effort to accommodate heterogeneity in multi-indexed data on trade, foreign direct investment, migration, etc.[1] Using multi-indexed data on international trade flows, which represents a prominent example of such data, we show that even after introducing conventional fixed effects for *origins*, *destinations*, *origin-destination dyads*, and *time periods*, there is strong evidence of cross-sectional dependence in the dependent variable representing trade flows. Different sources of cross-sectional dependence based on various sociocultural factors such as common borders, language and currency, trade unions and colonial ties, are explored.

In a panel data model setting, distance as well as sociocultural factors (which we can view as generalized distance variables) are generally time invariant, so they are treated as fixed effects. Balazsi, Matyas and Wansbeek (2018) explore econometric implications of using a host of alternative multidimensional fixed effects in panel data gravity models, but assume trade flows to be independent. We apply two of the more widely used multidimensional fixed effects transformations from the empirical trade literature to a panel of $N = 70$ countries trade flows covering the $T = 38$ years from 1963 to 2000, in a model specification that allows for the presence of simultaneous cross-sectional dependence.

One approach uses N origin and N destination country fixed effects plus T time-specific effects proposed by Matyas (1997) as an extension of conventional fixed effects panel data model (e.g., Baltagi, 2005) to the multidimensional situation that arises in the case of gravity models. These models take an $N^2T \times 1$ vector of dependent variables reflecting the matrix of trade flows between the N countries (assuming flows between all countries) at each time period, resulting in a dummy matrix of fixed effects with column rank of $2N + T - 2$. The second approach makes use of fixed effects proposed by Cheng and Wall (2005) that introduce fixed effects for origin-destination dyads as well as time periods, resulting in a dummy matrix with column rank of $N^2 + T - 1$, frequently adopted in the empirical trade literature.

Apart from accommodating heterogeneity in flows using fixed effects, the dependent variable

[1] Note that gravity models are at least double-indexed, indexing a country (region) of origin, and a country (region) of destination. Pooling gravity equations across dyads of countries (regions) over time leads to a panel data structure of the data.

vector of $N^2 \times 1$ trade flows for each time period are assumed to be independent, so flows between countries that have a common currency, language, border or colonial ties are no more likely than flows between countries having nothing in common. Cross-sectional dependence in flows suggests that flows between countries with sociocultural similarity (e.g., common language, colonial ties, spatial neighbors, member of trade unions, etc.) are likely to exhibit dependence as opposed to independence. We set forth a model specification that allows for this type of dependence in flows across the N^2T country-time dyads. Vasilis and Wansbeek (2012) provide an overview of econometric specifications for dealing with cross-sectional dependence, consisting of two main approaches, spatial econometric and common factor models.

We take the spatial econometric approach here, but note that a common factor specification could also be employed to address the issue we raise. The nature of dependence that we model would be better labeled *network dependence* rather than spatial or cross-sectional dependence, because we introduce dependence between network nodes involving origin- and destination-dyads as well as covariance across these.[2] We use the terms network and cross-sectional dependence interchangeably here, but note that the network dependence specification introduced here reflects a special case of cross-sectional dependence that can arise in the case of origin-destination flows that has not received a great deal of attention in the literature.

Estimates from the network dependence model specification suggest that sociocultural proximity can reflect transmission channels that can be viewed as a source of cross-sectional dependence. Model specifications that accommodate network dependence are set forth, along with computationally efficient Markov Chain Monte Carlo (MCMC) estimation methods. Ignoring network dependence implies biased estimates from panel trade flow models that rely on fixed effects and the assumption of independence between flows.

An innovative aspect of our MCMC estimation approach is use of Metropolis-Hastings guided samples from the joint posterior distribution of the dependence parameters to construct a Monte Carlo integration estimate of the log-marginal likelihood useful for model comparison. Our MCMC estimation approach allows for estimation and posterior inference on a vector of dependence parameters that determines the relative importance of network dependence, as well as a

[2]Common factor cross-sectional dependence specifications would need to be extended to address the type of dependence that we consider here. See also Baltagi and Maasoumi (2013), who provide an introductory discussion for a series of articles in a special issue devoted to dependence in cross-section, time series and panel data models.

Monte Carlo integration estimate of the log-marginal likelihood which can be used for model comparison purposes.[3] In our case, we rely on Markov Chain Monte Carlo sampling to estimate the model parameters with the dependence parameters sampled using a Metropolis-Hastings procedure. Since this approach produces draws of the dependence parameters that are steered by Metropolis-Hastings accept/reject decisions to areas of high density of the joint posterior, we can produce an efficient Monte Carlo integration of the log-marginal likelihood.

Another methodological innovation is use of convex combinations of network dependence weight matrix structures (see Pace and LeSage, 2002; Debarsy and LeSage, 2017, 2018; Hazir, LeSage and Autant-Bernard, 2018; LeSage and Fischer, 2018). The weight matrix structures are constructed to reflect spatial proximity between countries, as well as numerous types of sociocultural proximity such as common currency, language, colonial ties, and so on. A convex combination of these multiple weight matrices is used to form a single weight matrix. This approach allows us to treat sociocultural factors (for example, common currency, common language, historical colonial relationships, trade agreements, and so on) as sources of network dependence in the panel gravity model.

Section 2 first introduces conventional panel gravity models as used in the empirical trade literature, along with a discussion of the two multidimensional fixed effects specifications that we explore,[4] and then discusses an extension of the conventional panel gravity model that allows for origin- and destination-based network dependence following ideas set forth by LeSage and Pace (2008). A computationally efficient approach to MCMC estimation is set forth. Section 3 sets forth computational challenges to estimation of the network dependence variant of the conventional panel gravity model, along with an MCMC estimation approach that overcomes these challenges. Section 4 applies our model to panel data on trade flows between 70 countries covering the 38 years from 1963 to 2000. In our application of the model we introduce an extension that allows for convex combinations of multiple sociocultural connectivity structures, that can be used in conjunction with log-marginal likelihood estimates to determine the relative importance of each type of connectivity. We find strong evidence of network dependence in

[3]Monte Carlo integration evaluates the expression to be integrated using random draws of the parameter values, but a drawback to this approach is inefficiency because many of the random draws for the parameters are not in areas of high density of the function being integrated.

[4]Choice of these two approaches from the myriad approaches available was based on their popularity in the empirical trade literature.

trade flows pointing to network spillover effects, and suggesting that ignoring the presence of this type of cross-sectional dependence will result in biased estimates from conventional trade flow specifications. Section 5 provides conclusions, and Appendix A presents information on data used as well as sources.

2 Multi-indexed panel gravity models

2.1 Conventional models

As noted above, Matyas (1997) made an early attempt to introduce multidimensional fixed effects for gravity model specifications such as that in (1).[5] The dependent variable y_{ijt} in (1) reflects an $n^2T \times 1$ vector of (logged) trade flows between N countries i and j at time t, so $i = 1, \ldots, n, j = 1, \ldots, n$, and $t = 1, \ldots, T$.

$$y_{ijt} \quad = \quad x_{ijt}\beta + \alpha_i + \gamma_j + \lambda_t + \varepsilon_{ijt}, \tag{1}$$

where β is a $2k \times 1$ vector of parameters on the $N^2T \times 2k$ (logged) covariates x_{ijt}, which are usually k measures of economic size of destinations and another k measures for origins in the country dyads (i, j) at time t. We note that distance between the countries is time-invariant and not in the set of covariates. The fixed effects parameters α_i, γ_j represent destination-specific and origin-specific country effects while λ_t are time-period specific fixed effects. It is assumed that ε_{ijt} are normal *i.i.d.* (independent, identically distributed) idiosyncratic disturbances with zero mean and scalar (σ_ε^2) variance.[6]

We can write the fixed effects as an $N^2T \times (2N + T)$ matrix:

$$D = \left(\begin{array}{ccc} I_N \otimes \iota_{NT}, & \iota_N \otimes I_N, & \iota_{N^2} \otimes I_T \end{array} \right),$$

with column rank $(2N + T - 2)$, where I_N is an identity matrix of dimension N and ι_N is an $N-$dimensional column vector of ones. One can use a projection matrix of size $N^2T \times N^2T$ to

[5] See Baltagi, Egger and Pfaffermayr (2015) for an explanation of theoretical models that give rise to this log-linear specification.

[6] It is also assumed that the covariates and the disturbance terms are uncorrelated, ruling out endogeneity of the measures of country size.

eliminate the fixed effects in D, corresponding to the usual scalar transformation involving what have been labeled "Within" transformations.

There are of course other specifications for the fixed effects. For example, Egger and Pfaffermayr (2003) propose bilateral specific fixed effects γ_{ij}, where $D = \begin{pmatrix} I_N \otimes I_N \otimes \iota_T \end{pmatrix}$, of size $N^2T \times N^2$ with full column rank N^2. A variant of this, proposed by Cheng and Wall (2005), that is popular in the empirical trade literature is shown in (2),

$$y_{ijt} = x_{ijt}\beta + \gamma_{ij} + \lambda_t + \varepsilon_{ijt}, \tag{2}$$

where $D = \begin{pmatrix} I_N \otimes I_N \otimes \iota_T & I_N \otimes \iota_N \otimes I_T \end{pmatrix}$ of size $N^2T \times (N^2 + T)$, with column rank $(N^2 + T - 1)$. Of course, there is a projection matrix and corresponding scalar "Within-type" transformation that can be used to eliminate this more extensive set of fixed effects. Balazsi, Matyas and Wansbeek (2018) point out that the model in (1) represents a special case of that in (2), and there is an analogy of this 3D situation in (2) to 2D panel data models, where individuals in the 2D situation are treated as (i, j) pairs in the 3D setting. In other words, individual effects are now assigned to (i, j) dyads.

We note that the model specifications in (1) and (2) assume that the dependent variable vector of $N^2 \times 1$ trade flows for each time period are statistically independent, so flows between countries that have a common currency, language, border or colonial ties are no more likely than flows between countries having nothing in common. Network dependence in flows suggests that flows between countries with sociocultural similarity (e.g., common language, colonial ties, spatial neighbors, member of trade unions, etc.) are likely to exhibit dependence as opposed to independence. In the next section, we set forth a model specification that allows for this type of dependence in flows across the N^2T country-time dyads.

2.2 Extension for network dependence

We set forth an extension of the conventional panel gravity model that allows for origin- and destination-based network dependence. The matrix expressions in (3) represent a panel data extension of the cross-sectional *gravity model for origin-destination flows* introduced in LeSage and Pace (2008).

$$y \;=\; \rho_o I_T \otimes (W \otimes I_N)y + \rho_d I_T \otimes (I_N \otimes W)y + \rho_w I_T \otimes (W \otimes W)y + Z\delta + \varepsilon, \qquad (3)$$

where y is the $N^2T \times 1$ dependent variable vector of origin-destination flows for each time period, organized with t being the slow index for elements y_{ijt} in the vector y. The $N^2T \times 2k$ matrix Z contains covariates with the associated $2k \times 1$ parameter vector δ, and the $N^2T \times 1$ vector ε represents the normally distributed *i.i.d.* scalar variance disturbances.[7]

The model in (3) indicates that flows at *each time period* $t = 1, \ldots, T$ exhibit dependence on flows of countries *neighboring* the origin country captured by the $N^2T \times 1$ vector $I_T \otimes (W \otimes I_N)y$ with the associated scalar parameter ρ_o measuring the strength of that dependence. The matrix W is an $N \times N$ matrix that defines neighbors and for now, we define neighboring countries as those with common borders (spatial neighbors), as in the model of LeSage and Pace (2008).[8] A neighboring country is indicated by a non-zero (i, j) element in the $N \times N$ matrix W, which has zeros on the main diagonal. The matrix W is normalized to have row-sums of unity, resulting in the $N^2 \times 1$ vector $(W \otimes I_N)y$ reflecting a linear combination of trade flow values from countries that are *neighbors* to the origin country.

The model also allows for dependence of flows in each time period from countries neighboring the destination country, captured by the vector $I_T \otimes (I_N \otimes W)y$, with associated scalar parameter ρ_d, and we note that this vector relies on the same matrix W used to define (spatial) neighbors. LeSage and Pace (2008) point out that while the matrix $(I_N \otimes W)$ defines neighbors to the destination, the matrix $(W \otimes I_N)$ identifies neighbors to the origin, when the vector of flows for time t arises from a conventional $N \times N$ origin-destination flow matrix, organized with dyads (i, j) representing flows from origin j to destination i.

Another type of dependence is also included in the model, reflected by the $N^2T \times 1$ vector $I_T \otimes (W \otimes W)y$ and associated scalar parameter ρ_w, which captures dependence of flows from countries that are neighbors to both the origin and destination countries. LeSage and Pace (2008) motivate this type of dependence using the (cross-sectional) specification in (4), where

[7]For notational convenience, we assume that the matrix D of fixed effects has been eliminated from the model through the use of a scalar transformation of the type described in Balazsi, Matyas and Wansbeek (2018).

[8]We consider more general definitions of neighboring countries based on other connectivity constructs such as common currency, common language, etc. later.

they argue that the matrix A can be viewed as a spatial filter.

$$
\begin{aligned}
Ay &= \alpha \iota_{N^2} + Z\delta + \varepsilon, \\
A &= (I_{N^2} - \rho_d W_d)(I_{N^2} - \rho_o W_o) \\
&= (I_{N^2} - \rho_d W_d - \rho_o W_o + \rho_w W_w), \\
W_d &= I_N \otimes W, \\
W_o &= W \otimes I_n, \\
W_w &= W_d \otimes W_o = W_o \otimes W_d = W \otimes W.
\end{aligned} \tag{4}
$$

The argument is that the existence of origin- and destination-based dependence between trade flows $(W_o y, W_d y)$, logically implies a covariance between these two types of dependence which is reflected in $W_w y$. They note that this filter implies a restriction that $\rho_w = -\rho_o \rho_d$, but argue this restriction need not be imposed during estimation, so we address the more general case here and allow for an unrestricted parameter ρ_w.[9]

LeSage and Pace (2008) also point out that the matrix of covariates reflecting origin-destination dyads can be written as $Z = \begin{pmatrix} X_{ot} & X_{dt} \end{pmatrix}$, where $X_{ot} = X_t \otimes \iota_N$ and $X_{dt} = \iota_N \otimes X_t$ with X_t being an $N \times k$ matrix of covariates measuring the (economic) size of each country at time t.

3 Estimating the network dependence panel data gravity model

The model can be written as shown in (5).

$$
\begin{aligned}
\tilde{y}\omega &= Z\delta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_{N^2 T}), \\
\tilde{y} &= \begin{bmatrix} y & I_T \otimes (I_N \otimes W)y & I_T \otimes (W \otimes I_N)y & I_T \otimes (W \otimes W)y \end{bmatrix}, \\
\omega &= \begin{pmatrix} 1 \\ -\rho_o \\ -\rho_d \\ -\rho_w \end{pmatrix}.
\end{aligned} \tag{5}
$$

[9]Of course, given unrestricted estimates of ρ_o, ρ_d, ρ_w one could test if the restriction $\rho_w = -\rho_o \rho_d$ is consistent with the sample data.

A key feature of $\tilde{y}$ is that this expression separates dependence parameters to be estimated from sample data describing the simultaneous dependence, with the scalar dependence parameters in the vector ω. We assume normally distributed, zero mean, constant variance (σ^2) disturbances.

3.1 Likelihood and priors

The likelihood is shown in (6), where $|R(\omega)|$ is the determinant that depends on the dependence parameters in ω, as does the expression $e'e(\omega)$.

$$\begin{aligned} f(y;\omega,\sigma^2\delta) &= |R(\omega)|(2\pi\sigma^2)^{-N^2T/2}\exp(-\frac{1}{2\sigma^2}e'e(\omega)), \\ e &= \tilde{y}\omega - Z\beta, \\ R(\omega) &= I_{N^2T} - \rho_o I_T \otimes (I_N \otimes W) - \rho_d I_T \otimes (W \otimes I_N) - \rho_w I_T \otimes (W \otimes W). \end{aligned} \tag{6}$$

To ensure that $R(\omega)$ is non-singular, restrictions need to be placed on the dependence parameters ω to ensure that $R(\omega)^{-1}$ exhibits an underlying stationary process. Specifically, $\rho_o + \rho_d + \rho_w < 1$. The parameter space for the set of parameters $(\omega, \sigma^2, \delta)$ is: $\Delta := \Delta_\omega \times \Delta_{\sigma^2} \times \Delta_\delta = \Delta_{-1,1} \times (0,\infty) \times \mathcal{R}^{2k}$.

Since our focus is on large samples involving N^2T observations, we rely on uninformative priors for the parameters δ and σ^2, as these would not likely impact posterior estimates. Since the dependence parameters in ω are a focus of inference, uniform priors for these dependence parameters are used, which must obey a stability constraint.

Mathematically, the flat or uniform prior for ω, δ can be represented as $p(\omega) \propto 1, p(\delta) \propto 1$. The noise variance σ^2, is restricted to positive values, with a flat prior assigned to the log-transformed value which is denoted $p(\sigma^2) \propto 1/\sigma^2$. Given this prior information, and prior independence, one can write: $p(\omega) \times p(\sigma^2) \times p(\delta) \propto 1/\sigma^2$. While this flat prior is improper since the integral over the parameter space Δ is not finite, the joint posterior distribution for the dependence parameters ω is proper under relatively unrestrictive assumptions. This joint posterior is derived by analytically integrating out the parameters δ, σ^2, with details regarding this integration found in Hepple (1995a, 1995b).

To derive the joint posterior for the dependence parameters ω, begin with the full joint posterior $p(\omega, \delta, \sigma^2)$ and analytically integrate out δ, σ^2. This relies on standard techniques from the Bayesian regression literature (Zellner, 1971). Combining the likelihood function in (7) with the flat priors (and ignoring the constant $2\pi^{N^2T/2}$) leads to the joint posterior in (8), from which σ can be integrated out, leading to (9), where $\Gamma(.)$ denotes the Gamma function.

$$
\begin{aligned}
f(y; \omega, \sigma^2, \delta) &\propto |R(\omega)|(2\pi\sigma^2)^{-N^2T/2}\exp(-\frac{1}{2\sigma^2}\omega' u' u\omega), & (7)\\
u &= \tilde{y} - Z\delta_d, \\
\delta_d &= (Z'Z)^{-1}Z'\tilde{y}, \\
p(\omega, \delta|y) &\propto |R(\omega)| \int_0^\infty \sigma^{-(N^2T+1)}\exp\left(-\frac{1}{2\sigma^2}\omega' u' u\omega\right) d\sigma & (8)\\
&= 2^{(N^2T-2)/2}\Gamma(N^2T/2)|R(\omega)| \left(\omega' u' u\omega\right)^{-N^2T/2}. & (9)
\end{aligned}
$$

To integrate out the $2k$ different δ parameters, properties of the multivariate $t-$distribution in conjunction with 'completing the square' are used (see Zellner, 1971). This leads to a joint distribution for the dependence parameters ω shown in (10), with the term $|Z'Z|^{-1/2}$ and the exponent $-(N^2T - 2k)/2$ arising from this integration (see Hepple (1995a, 1995b). This expression must be numerically integrated to arrive at the log-marginal likelihood for these models. This is accomplished using Monte Carlo integration discussed later.

$$
p(\omega|y) = 2^{(N^2T-2)/2}\Gamma(N^2T/2)|R(\omega)||Z'Z|^{-1/2}(\omega' u' u\omega)^{-(N^2T-2k)/2}. \quad (10)
$$

A question that has been explored in the literature is whether this conditional posterior distribution is proper and can be integrated over the parameter space for the case of the flat priors described here. Dittrich et al. (2017) tackle the spatial autoregressive model (SAR) for the case of a single weight matrix W and a single cross-section consisting of N regions. Their conclusion is that propriety of this distribution requires that: (i) $N > k$, (ii) $(Z'Z)^{-1}$ exists, and (iii) $(y'MWy)^2 \neq y'W'MWyy'My$, where $M = I_N - Z(Z'Z)^{-1}Z'$.

Conditions (i) and (ii) are not restrictive, and condition (iii) arises from noting that the term

$u'u$ in the case of the single weight matrix model can be written as in (11).[10]

$$
\begin{aligned}
u'u &= \begin{pmatrix} e_y'e_y & e_{Wy}'e_y \\ e_y'e_{Wy} & e_{Wy}'e_{Wy} \end{pmatrix}, \\
e_y &= y - Z\delta, \\
e_{Wy} &= Wy - WZ\delta, \\
|u'u| > 0 &\rightarrow e_y'e_ye_{Wy}'e_{Wy} > (e_y'e_{Wy})^2, \\
|u'u| = 0 &\rightarrow e_y'e_ye_{Wy}'e_{Wy} = (e_y'e_{Wy})^2.
\end{aligned} \tag{11}
$$

This condition essentially means that a valid error covariance matrix exists for the SAR model. This should not be an important restriction to ensure posterior propriety of the joint distribution for the dependence parameter ρ in the case of the single weight matrix model. This condition would be met in all cases where the single weight W is defined so that all observations are connected and the usual normalization is applied.[11]

Extending this result to the multi-indexed panel gravity model considered here involving multiple weight matrices and associated dependence parameters in the vector ω results in a similar expression that implies the error covariance matrix of the model exists. This can be written as shown in (12), where the definition of $\tilde{y}$ from (5) is used.

$$
\begin{aligned}
u'u &= \begin{pmatrix} e_y'e_y & \tilde{e}_y'e_y \\ e_y'\tilde{e}_y & \tilde{e}_y'\tilde{e}_y \end{pmatrix}, \\
e_y &= y - Z\beta, \\
\tilde{e}_y &= \tilde{y} - Z\delta_d, \\
\delta_d &= (Z'Z)^{-1}Z'\tilde{y}, \\
|u'u| > 0 &\rightarrow e_y'e_y\tilde{e}_y'\tilde{e}_y > (e_y'\tilde{e}_y)^2, \\
|u'u| = 0 &\rightarrow e_y'e_y\tilde{e}_y'\tilde{e}_y = (e_y'\tilde{e}_y)^2.
\end{aligned} \tag{12}
$$

Again, this condition essentially means that a valid error covariance matrix exists for our

[10] The single matrix W is assumed symmetric for simplicity here.

[11] Multiple approaches to normalizing the matrix W have appeared in the spatial econometrics literature, all of which ensure that the maximum eigenvalue of the matrix W is one.

models. This should not be an important restriction to ensure posterior propriety of the joint distribution in the case of the multiple dependence parameter models.

3.2 Computational challenges

One issue that arises when considering estimation of the model in (3) is that multiple dependence parameters ρ_o, ρ_d, ρ_w would require use of a multivariate optimization routine to produce estimates based on maximum likelihood. It is also the case that the dependence parameters are (well) defined over the $(-1, 1)$ interval, meaning that constrained optimization would be required to ensure values $-1 < \rho_o + \rho_d + \rho_w < 1$.[12]

Another challenge to maximum likelihood estimation is the log-determinant term that arises in the (log) likelihood function, specifically (log): $|I_{N^2T} - \rho_o I_T \otimes (W \otimes I_N) - \rho_d I_T \otimes (I_N \otimes W) - \rho_w I_T \otimes (W \otimes W)|$. In the case of conventional spatial regression models involving a single weight matrix, there is a great deal of literature on approaches to efficiently calculating or approximating the log-determinant term that appears in the (log) likelihood $|I_N - \rho W|$, (see LeSage and Pace, 2008, Chapter 4). These approaches are not directly applicable to the model considered here, complicating maximum likelihood estimation, since the log-determinant expression needs to be evaluated for multiple dependence parameter values during optimization. In the case of Markov Chain Monte Carlo estimation, the log-determinant term appears in the conditional distribution for the dependence parameters requiring multiple evaluations during sampling.

Because of the issues outlined above, we set forth estimation based on Markov Chain Monte Carlo, with no prior distributions assigned to the parameters β, σ^2. Parameter restrictions are imposed on the dependence parameters during MCMC sampling using methods described later. Since emphasis is on modeling situations involving large samples of observations, prior information would not play a role in determining posterior estimates of the parameters, so MCMC is used as a computational device to produce estimates that should be identical to those from maximum likelihood estimation. MCMC estimation involves sequentially sampling each parameter (or set of parameters) from their conditional distributions (or joint conditional distribution in the case of a set of parameters). Expressions for the conditional distributions are frequently easier to calculate than those required to evaluate the (log) likelihood, which is true

[12]The lower bound of -1 is typically used for convenience in applied practice and ensures the existence of the matrix inverse for the reduced form of the model.

for the models considered here.

Another aspect of this model regards proper interpretation of the partial derivative impacts on the dependent variable vector arising from changes in the explanatory variables, e.g., $\partial E(y)/\partial X^r$ for the rth explanatory variable. We take this issue up in a later section.

MCMC estimation proceeds by sampling sequentially from the conditional distributions of each parameter (or set of parameters). The conditional distributions for the model parameters δ, σ^2, ω needed to implement MCMC estimation are set forth next.

3.3 Conditional distributions for the model parameters

Since our focus is on large samples N^2T, we can rely on uninformative priors for the parameters δ, as these would not likely impact posterior estimates. For the same reason, we rely on an uninformative prior $p(\sigma^2) \propto 1/\sigma^2$ for σ^2. Since the dependence parameters in ω are at the center of inference, we employ uniform priors for these dependence parameters which are constrained to lie in the open interval $(-1, 1)$. There is also the need to impose stability restrictions on these parameters discussed later. Given the limited prior information, the conditional distribution for the parameters δ of the model in (5) takes the form of a multivariate normal with mean and variance-covariance shown in (13).

$$\begin{aligned} p(\delta|\sigma^2, \omega) &= \mathcal{N}(\tilde{\delta}, \tilde{\Sigma}_\delta), \\ \tilde{\delta} &= (Z'Z)^{-1}(Z'\tilde{y}\omega), \\ \tilde{\Sigma}_\delta &= \sigma^2(Z'Z)^{-1}. \end{aligned} \tag{13}$$

We note that $(Z'Z)^{-1}Z'\tilde{y}$ consists of only sample data information, so this expression can be calculated once prior to MCMC sampling, and this is true of $(Z'Z)^{-1}$ as well. This means that sampling new values of the parameters δ (given values for the parameters σ^2, ω) can take place in a rapid, computationally efficient way.

The conditional posterior for σ^2 (given δ, ω) takes the form in (14), with the uninformative prior.

$$
\begin{aligned}
p(\sigma^2|\delta,\omega) &\propto (\sigma^2)^{-(\frac{N}{2})}\exp\left(-\frac{1}{2\sigma^2}(e'e)\right), \\
e &= (\tilde{y}\omega - Z\delta) \\
&\sim \mathcal{IG}(\tilde{a},\tilde{b}), \\
\tilde{a} &= N/2, \\
\tilde{b} &= (e'e)/2.
\end{aligned}
\tag{14}
$$

The joint conditional distribution for the dependence parameters in ω can be obtained by analytically integrating out δ, σ^2 leading to a (log kernel) expression for the joint posterior of the dependence parameters in ω.

$$
\begin{aligned}
log\ p(\omega|y,Z,W) &\propto log[D(\omega)] - (N^2T/2)log(\omega' F\omega), \\
F &= (\tilde{y} - Z\delta_d)'(\tilde{y} - Z\delta_d), \\
\delta_d &= Z(Z'Z)^{-1}\tilde{y},
\end{aligned}
\tag{15}
$$

where $log[D(\omega)]$ is a Taylor series approximation to the log-determinants in the model, described in detail later. For now we note that this log-determinant term depends on the dependence parameters in the vector ω, indicated by $D(\omega)$. We note that F consists of only sample data, so this expression can be calculated prior to MCMC sampling, leading to a computationally efficient expression reflecting a quadratic form: $log(\omega' F\omega)$, that can be easily evaluated for any vector of dependence parameters ω.

One motivation for working with the joint conditional posterior distribution for the dependence parameters is the need to impose stability restrictions on these parameters. Specifically, $-1 < \rho_o + \rho_d + \rho_w < 1$. Working with the joint conditional posterior distribution for these parameters allows us to adopt a block sampling Metropolis-Hastings (M-H) scheme for the dependence parameters (described in detail later). Block sampling means that a vector of dependence parameters in ω are proposed and compared to the current vector of dependence parameters. The proposed vector is either accepted or rejected. This allows proposals of dependence parameters that obey the stability restriction, so any vectors that are accepted by the Metropolis-Hastings procedure will always obey the needed restrictions.

A second motivation is that having analytically integrated out the parameters δ, σ^2, further integration of the joint conditional posterior over the set of dependence parameters in ω, would yield the log-marginal likelihood for these models. We can use Monte Carlo integration to accomplish this task. Monte Carlo integration evaluates the expression to be integrated using random draws of the parameter values. A drawback to this approach is inefficiency because many of the random draws for the parameters are not in areas of high density of the function being integrated. In our case, the Metropolis-Hastings sampling procedure used to produce draws of the dependence parameters steers these parameter values to areas of high density of the joint posterior. This allows us to produce an efficient Monte Carlo integration of the log-marginal likelihood.

Given an estimate of the log-marginal likelihood for a model M_i ($logM_i$), we can calculate: $prob(M_i) = \exp(logM_i)/\sum_{i=1}^{Q} \exp(logM_i))$ (in the case of Q different models). Of course, there is a great deal of interest in comparing alternative models, for example, models based on different spatial weight matrices, or different fixed effects specifications.

3.4 A Taylor's series approximation to the log-determinant term

We have motivated that (15) represents a computationally efficient expression for the joint posterior, but this involves the log-determinant term $log[D(\omega)]$ in (16), where: $W_1 = I_T \otimes (W \otimes I_N), W_2 = I_T \otimes (I_N \otimes W), W_3 = I_T \otimes (W \otimes W)$, which could be difficult and slow to calculate.

$$\ln|I_{N^2 \times T} - \rho_o W_1 - \rho_d W_2 - \rho_w W_3|. \tag{16}$$

An approximation to the log-determinant term works to preserve the computational efficiency of the expression (15). Pace and LeSage (2002) set forth a Taylor series approximation for the log-determinant of a matrix like our expression: $\ln|I_{NT} - \tilde{W}|$, where $\tilde{W} = \left(\rho_o W_o + \rho_d W_d + \rho_w W_w \right)$. They show that for a *symmetric* nonnegative weight matrix $\tilde{W}$

with eigenvalues $\lambda_{\min} \geq -1, \lambda_{\max} \leq 1$, and $tr(\tilde{W}) = 0$ (where tr represents the trace):

$$\begin{aligned}
\ln|I_{N^2T} - \rho_o W_o - \rho_d W_d - \rho_w W_W| &= \ln|I_{N^2T} - \tilde{W}|, \\
\tilde{W} &= \left(\rho_o W_o + \rho_d W_d + \rho_w W_w \right) \\
&= tr(\ln(I_{N^2T} - \tilde{W})), \\
\ln(I_{N^2T} - \tilde{W}) &= -\sum_{i=1}^{\infty} \tilde{W}^i / i, \\
\ln|I_{N^2T} - \tilde{W}| &= -\sum_{i=1}^{\infty} tr(\tilde{W}^i)/i \qquad (17) \\
&\simeq -\sum_{j=1}^{q} tr(\tilde{W}^j)/j, \\
tr(\tilde{W}) &= tr(\rho_o W_o + \rho_d W_d + \rho_w W_w) \\
&= \rho_o tr(W_o) + \rho_d tr(W_d) + \rho_w tr(W_w). \qquad (18)
\end{aligned}$$

Golub and van Loan (1996, p. 566) provide the expression in (17), while (18) arises from linearity of the trace operator. Note that the first-order $tr(\tilde{W})$ is zero, given the definitions of W_o, W_d, W_w. Let $\eta = \left(\rho_o \quad \rho_d \quad \rho_w \right)'$, and first consider the case of symmetric matrices W_o, W_d, W_w, which allows the second-order trace to be expressed as a quadratic form in (20) involving the vector of parameters η and all pairwise multiplications of the individual matrices in $\tilde{W}$ as shown in (19).

$$\begin{aligned}
tr(\tilde{W}^2) &= \eta' \begin{pmatrix} tr(W_o^2) & tr(W_o W_d) & tr(W_o W_w) \\ tr(W_d W_o) & tr(W_d^2) & tr(W_d W_w) \\ tr(W_w W_o) & tr(W_w W_d) & tr(W_w^2) \end{pmatrix} \eta \qquad (19) \\
&= \eta' Q^2 \eta. \qquad (20)
\end{aligned}$$

LeSage and Pace (2009) point out that accelerated computation of traces can be accomplished using sums of matrix Haddamard products, $Q^2_{ij} = \sum_i^3 \sum_j^3 W_i \odot W_j, i = o, d, w; j = o, d, w$. For the case of asymmetric matrices, matrix products $\sum_i^3 \sum_j^3 W_i \odot W_j'$ can be used, and the weight matrices in the multi-indexed panel gravity data model would be an example of asymmetric matrices. Note that this formulation separates the parameters in the vector η from the matrix of traces, which allows pre-calculation of the matrix of traces prior to MCMC sampling. A more efficient computational expression is $(\eta \otimes \eta)vec(Q^2)$, where $\otimes$ is the Kronecker product and vec the operator that stacks the columns of the matrix Q^2.

Using this approach leads to a similar expression for the third-order trace, which involves $3^3 = 27$ matrix products, and a fourth-order trace with $3^4 = 81$ matrix products. A fourth-order Taylor series approximation to the log-determinant $\ln|I_n - \tilde{W}|$ takes the form in (21).

$$
\begin{aligned}
\ln|I_{N^2T} - \tilde{W}| \quad \simeq \quad & -(\eta \otimes \eta)vec(Q^2))/2 \\
& -(\eta \otimes \eta) \otimes \eta \; vec((Q^3))/3 \\
& -((\eta \otimes \eta) \otimes \eta) \otimes \eta \; (vec(Q^4)/4.
\end{aligned} \tag{21}
$$

A key aspect of these calculations is that traces of products of the weight matrices can be pre-calculated prior to MCMC sampling. This means that updating the log-determinant expression for any set of dependence parameters (ρ_o, ρ_d, ρ_w) involves simple multiplications, where the dependence parameters in η can be separated from these matrix products.

3.5 Block sampling the dependence parameters ω

As noted above, a second computational challenge for estimation of these models is the need to impose stability restrictions on the dependence parameters $(-1 < \rho_o + \rho_d + \rho_w < 1)$. Debarsy and LeSage (2018) set forth a block-sampling approach that proposes a vector of candidate values for a similar set of dependence parameters in the context of a model involving a convex combination of weight matrices. Dependence parameters that do not meet the stability restriction can be rejected, so any values accepted are consistent with stability.

The conditional distributions for the current and proposed dependence vectors that we can label ω^c, ω^p are evaluated with a Metropolis-Hastings step used to either accept or reject the newly proposed vector ω^p. Block sampling the dependence parameter vector ω has the virtue that accepted vectors will obey any restrictions and reduce autocorrelation in the MCMC draws for these parameters. However, block sampling is known to produce lower acceptance rates which may require more MCMC draws in order to collect a sufficiently large sample of draws for posterior inference regarding ω. To address this issue, Debarsy and LeSage (2018) as well as

LeSage, Chih and Vance (2018) propose a hybrid approach that begins with a reversible jump sampling procedure and switches to a tuned random-walk proposal procedure for proposing vectors ω after some initial number of start-up samples are drawn.

We rely on a reversible jump procedure to produce proposal values for the vector of parameters ρ_o, ρ_d, ρ_w. For each scalar parameter we rely on a three-headed coin flip. By this we mean a uniform random number on the open interval *coin flip* $= U(0,1)$, with head #1 equal to a value smaller or equal to 1/3, head #2 a value larger than 1/3, but smaller or equal to 2/3 and head #3 a value larger than 2/3 and smaller than one. Given a head #1 result, we set a proposal ρ_o^p using a uniform random draw on the open interval $(-1 < \rho_o^p < \rho_o^c)$, where ρ_o^c is the current value. A head #2 results in setting the proposal value equal to the current value $(\rho_o^p = \rho_o^c)$, while a head #3 selects a proposal value based on a uniform random draw on the open interval $(\rho_o^c < \rho_o^p < 1)$. Of course, a similar approach is used to produce proposals for the parameters ρ_d, ρ_w. Proposed vectors of these parameters inconsistent with the stability restrictions are eliminated via rejection sampling.

The reversal jump approach to proposing the block of dependence parameters has the virtue that accepted vectors will obey the stability restriction and will also reduce autocorrelation in the MCMC draws for these parameters. However, proposals from the reversible jump procedure based on the large intervals between $(-1 < \rho_o^c)$ and $(\rho_o^c < 1)$ will not produce candidates likely to be accepted when these parameters are estimated with a great deal of precision, as would be the case for problems involving large N^2T. This can result in a failure to move the chain adequately over the parameter space. To address this issue, standard deviations, $\sigma_{\rho_o}, \sigma_{\rho_d}, \sigma_{\rho_w}$ for each parameter are calculated based on the first 1,000 draws (and updated thereafter using an interval of $m = 1,000$ draws). These are used in a tuned random-walk procedure to produce candidate/proposed values. Specifically, we use a tuning scalar c for each parameter that is adjusted based on acceptance rates for each parameter. This is used in conjunction with the standard deviations to produce proposals: $\rho_o^p = \rho_o^c + c\mathcal{N}(0,1)\sigma_{\rho_o}$, with the same approach used for ρ_d, ρ_w.

The proposed estimation method relies on a great many approximations, raising the issue of whether resulting estimates have desirable properties such as small bias and mean-squared error as well as good coverage. By coverage we mean that the (say) 2.5% and 97.5% intervals from the empirical distributions of the effects estimates on which practitioners base conclusions regarding statistical significance of the effects estimates cover the true values 95% of the time.

Debarsy and LeSage (2018) present results from Monte Carlo experiments for the case of a cross-sectional convex combination of weights SAR model that relies on the same fourth-order Taylor series approximation to the log-determinant and the reversible-jump, hybrid tuned random-walk procedure for estimating the spatial dependence parameters. They show small bias and mean-squared error as well as good coverage across a range of negative and positive dependence parameters.

LeSage, Chih and Vance (2018) show results from Monte Carlo experiments for the dynamic space-time panel data model, which also involves three spatial weight matrices like the model described here. They also report Monte Carlo results with small bias and mean-squared error as well as good coverage across a range of negative and positive dependence parameters. LeSage (2018) discusses the commonality of the cross-sectional convex combination model of Debarsy and LeSage (2018), LeSage, Chih and Vance (2018), and the model described here, as well as Monte Carlo results.

3.6 Interpreting the network dependence panel data gravity model

The partial derivatives used to interpret how changes in (say the rth) explanatory variable of the model impacts changes in the dependent variable vector are non-linear matrix expressions. The sequence of partial derivatives for this model are shown in (22), where we record the $N \times N$ matrices of changes in (logged) flows arising from changing the rth variable in each country i X_i^r using $Y_i, i = 1, \dots, N$, to denote the $N \times N$ flow matrices associated with changing the rth variable in each country i. We define $\tilde{W}_o = (W \otimes I_N), \tilde{W}_d = (I_N \otimes W), \tilde{W}_w = (W \otimes W)$ to simplify notation in (22), and note that because the matrix W does not change over time in our static panel data model, we have a set of $N^2 \times N$ matrices describing the partial derivative impacts.

$$\begin{pmatrix} \partial Y_1/\partial X_1^r \\ \partial Y_2/\partial X_2^r \\ \vdots \\ \partial Y_N/\partial X_N^r \end{pmatrix} = (I_{N^2} - \rho_o W_o - \rho_d W_d - \rho_w W_w)^{-1} \begin{pmatrix} Jd_1\beta_d^r + Jo_1\beta_o^r \\ Jd_2\beta_d^r + Jo_2\beta_o^r \\ \vdots \\ Jd_N\beta_d^r + Jo_N\beta_o^r \end{pmatrix}. \qquad (22)$$

In (22), Jd_i ($i = 1, \ldots, N$) is an $N \times N$ matrix of zeros with the ith row equal to $\iota_N'\beta_d$, and Jo_i is an $N \times N$ matrix of zeros with the ith column equal to $\iota_N\beta_o$, where β_o and β_d denote parameters associated with origin and destination size measures. We have N sets of $N \times N$ outcomes, (one for each change in $X_i^r, i = 1, \ldots, N$) resulting in an $N^2 \times N$ matrix of partial derivatives reflecting the total effect on flows from changing the rth characteristic of all N regions, which LeSage and Thomas-Agnan (2015) label the *total effect*.

These authors provide a motivation for the expression in (22), noting that changes in the (size) characteristics of a single country i will (potentially) produce impacts on all elements of the $N \times N$ flow matrix. Intuitively, a change in (say) income of a single country can impact trade flows involving immediate trading partners, as well as, trade flows involving partners to the trading partners, partners to the partners of the trading partners, and so on, potentially impacting the entire $N \times N$ flow matrix.

Since regression models typically consider changes in characteristics (say income) of all $i = 1, \ldots, N$ observations/countries, this produces a set of N different $N \times N$ matrices of partial derivatives associated with changes in *each* explanatory variable in the model. LeSage and Thomas-Agnan (2015) propose scalar summary measures for the various types of effects that average over certain dimensions of the sequence of N different $N \times N$ matrices. We adopt a simpler strategy here for producing scalar summary measures of the partial derivative impacts. We take an average of the diagonal elements of the N different $N \times N$ matrices in (22) as a measure of own-partial derivative impacts reflecting own-country changes in flows arising from changes in (say) the typical country's income. And we use an average of the cumulative off-diagonal elements from each row of the N different $N \times N$ matrices in (22) to summarize network effects arising from changes in (say) income in a typical country. Network effects represent a scalar summary measure of the spillover impacts on other countries associated with changes in an explanatory variable in the model (say income). The scalar summary averages over all countries, and since the model is a static panel data model, over all time periods as well. We can delineate between origin and destination specific effects using the expressions involving β_o, β_d, which allows us to determine the relative importance of changes in (say) income at origin versus destination countries on trade flows.

In addition to point estimates of the partial derivative impacts, there might also be a need to calculate empirical measures of dispersion for the effects that could be used for inference. An empirical distribution of the scalar own- and cross-partial derivatives (labeled direct and network effects here) can be constructed using MCMC draws for the parameters $\rho_o, \rho_d, \rho_w, \beta_o, \beta_d$ in expression (22). However, this would require inversion of the $N^2 \times N^2$ matrix $(I_{N^2} - \rho_o W_o - \rho_d W_d - \rho_w W_w)$ thousands of times for each set of draws for ρ_o, ρ_d, ρ_w, making this computationally intensive.[13]

A compromise approach would be to use posterior means of the estimated parameters ρ_o, ρ_d, ρ_w to calculate a single matrix inverse: $(I_{N^2} - \rho_o W_o - \rho_d W_d - \rho_w W_w)^{-1}$ in conjunction with the MCMC draws for the parameters β_o, β_d. However, this would ignore stochastic variation in the effects estimates that arise from the fact that there is uncertainty regarding the parameters ρ_o, ρ_d, ρ_w. Ideally, we would like to use draws for these dependence parameters from their posterior distributions when simulating the empirical distribution of effects estimates.

4 Application of the network dependence panel model

We consider panel model specifications that use a panel of trade flows as the dependent variable vector y over the 38 years from 1963 to 2000. The (single) explanatory variable is (logged) gross domestic product per capita (GDP) lagged one year to cover the period from 1962 to 1999. The trade flows are from Feenstra et al. (2005), while the GDP data at market prices (current US$) and population data come from World Bank's (2002) World Development Indicators. A usable sample of 70 countries (see Table A.1 in Appendix A) was constructed for which GDP, population and trade flows were available over the 38 years.[14]

[13]Given sparse matrices W it would not be difficult to calculate the matrix inverse for situations involving the typical sample of 100 to 200 countries used in trade flow models.

[14]We eliminated countries from our sample that had one or more zero rows in any of the five weight matrices. This resulted in a few countries such as South Korea, Japan and India for which data was available to be excluded

Given our sample of 70 countries and 38 years, this results in $N^2T = 186,200$; with $2N + T - 2 = 176$ fixed effects parameters for the case of the Matyas (1997) model in (1), and $N^2 + T - 1 = 4,937$ fixed effects parameters in the Chen and Wall (2005) approach set forth in the model from (2).

We used five different definitions for the matrix W describing alternative structures of network dependence, specifically, W_{space} based on the three nearest spatial neighboring countries, $W_{language}$ based on countries sharing a common language, $W_{currency}$ based on common currency, W_{colony} based on countries with direct historical colonial ties, and W_{trade} based on membership in the same trade union (excluding the WTO). Details regarding countries with common borders, language, currency, colonial ties and trade union membership can be found in Appendix A.

Estimates from the model in (23) where the parameters ρ_o, ρ_d, ρ_w are (significantly) different from zero point to the existence of cross-sectional dependence.

$$\begin{aligned} y &= \rho_o W_o y + \rho_d W_d y + \rho_w W_w y + GDP_o \beta_o + GDP_d \beta_d + \varepsilon, \\ W_o &= I_T \otimes (W \otimes I_N), \quad W_d = I_T \otimes (I_N \otimes W), \quad W_w = I_T \otimes (W \otimes W). \end{aligned} \tag{23}$$

In the presence of cross-sectional dependence, estimates from conventional models that ignore cross-sectional/network dependence can be shown to be biased and inconsistent (see LeSage and Fischer, 2018). The presence of network dependence also implies spillover impacts arising from changes in neighboring countries $j \neq i$ income on country i's trade flows. In our model, neighbors are defined to include spatial neighbors in the case where W_{space} is used when estimating the model. More broadly, sociocultural neighbors arise when the matrix W used is based on common language, currency, trade union membership or direct colonial ties. Specifically, changes in income of countries j that have spatial, common language, currency, trade agreements, or colonial ties with country i will impact flows in the SAR model, provided that the scalar dependence parameters ρ_o, ρ_d, ρ_w are different from zero and the parameters β_o, β_d are non-zero.

Table 1 shows log-marginal likelihood function values for models based on the alternative

from our sample.

Table 1: Log-marginal likelihood estimates for alternative models

Model	Matyas (1997) fixed effects	$\rho_o + \rho_d + \rho_w$	Cheng and Wall (2005) fixed effects	$\rho_o + \rho_d + \rho_w$
W-trade	−5.1401e+05	0.8988	−5.0487e+05	0.8708
W-language	−5.5419e+05	0.6597	−5.3048e+05	0.6741
W-colony	−5.5610e+05	0.5726	−5.3548e+05	0.5911
W-currency	−5.6827e+05	0.6091	−5.4406e+05	0.5708
W-space	−5.2307e+05	0.8459	−5.0988e+05	0.8137

definitions of the weight matrix as well as the two alternative approaches to including fixed effects. The sum of posterior means for $\rho_o + \rho_d + \rho_w$ are also reported, since non-zero values of these parameters point to significant network dependence. From the table, we see that models using the Cheng and Wall (2005) fixed effects have higher log-marginal likelihoods than the corresponding Matyas (1997) model based on the same weight matrix, indicating these models are more consistent with our sample data. A second finding indicated by the estimated log-marginal likelihoods in the table is that the rank-ordering of preferred models for the various types of weight matrices is very similar for both types of fixed effects. Specifically, the weight matrix based on W_{trade} has the highest log-marginal likelihood, W_{space} is next highest, followed by $W_{language}$. Turning to estimates for the dependence parameters, we see that the sum of these are substantially positive, pointing to the presence of network dependence.

4.1 Estimates for the best models

Table 2 presents estimates for the best models based on W_{trade} using both the Matyas (1997) fixed effects and those of Cheng and Wall (2005). The table presents the mode of the parameter estimates evaluated using the joint posterior distribution as well as the mean and median based on 5,000 retained MCMC draws (with an initial 5,000 excluded for burn-in of the sampler). Monte Carlo (MC) error estimates are reported along with Geweke's diagnostic that compares draws from the first ten percent of the MCMC sampling (after burn-in) and the last 50 percent of the draws. The test is whether the batched means are equal, which indicates convergence.

From the estimates we see that the dependence parameters ρ_o, ρ_d, ρ_w are different from zero based on the credible intervals calculated from the MCMC draws. As noted in the discussion of model interpretation, the parameters β_o, β_d do not represent partial derivative impacts of the elasticity response of trade flows to changes in origin and destination-country GDP. These need to

Table 2: Estimates for the W_{trade} models

	Matyas (1997) fixed effects				
Parameter	Mode	Mean	Median	MC error	Geweke
Constant	−9.0215	−9.0177	−9.0187	0.00660826	0.997389
$beta_o$	0.2863	0.2865	0.2864	0.00030498	0.997029
$beta_d$	0.3182	0.3178	0.3177	0.00024616	0.995985
ρ_o	0.6811	0.6815	0.6816	0.00032737	0.997605
ρ_d	0.6342	0.6339	0.6340	0.00045538	0.998746
ρ_w	−0.4165	−0.4166	−0.4162	0.00061022	0.996369
Variable	Lower 0.01	Lower 0.05	Mean	Upper 0.95	Upper 0.99
Constant	−9.3420	−9.2677	−9.0177	−8.7669	−8.6804
β_o	0.2753	0.2779	0.2865	0.2949	0.2973
β_d	0.3065	0.3093	0.3178	0.3266	0.3289
ρ_o	0.6753	0.6770	0.6815	0.6862	0.6875
ρ_d	0.6272	0.6287	0.6339	0.6386	0.6408
ρ_w	−0.4248	−0.4234	−0.4166	−0.4109	−0.4097

	Cheng and Wall (2005) fixed effects				
Parameter	Mode	Mean	Median	MC error	Geweke
Constant	−12.4699	−12.5429	−12.5429	0.00780577	0.998862
β_o	0.2964	0.2985	0.2985	0.00019371	0.998628
β_d	0.4891	0.4920	0.4920	0.00044417	0.997119
ρ_o	0.4706	0.4671	0.4671	0.00055234	0.995611
ρ_d	0.6375	0.6347	0.6347	0.00033448	0.998707
ρ_w	−0.2363	−0.2308	−0.2308	0.00062930	0.992945
Variable	Lower 0.01	Lower 0.05	Mean	Upper 0.95	Upper 0.99
Constant	−13.6988	−13.3789	−12.5429	−11.6729	−11.3569
β_o	0.2877	0.2901	0.2985	0.3067	0.3086
β_d	0.4656	0.4727	0.4920	0.5115	0.5181
ρ_o	0.4584	0.4600	0.4671	0.4720	0.4748
ρ_d	0.6299	0.6310	0.6347	0.6382	0.6399
ρ_w	−0.2387	−0.2370	−0.2308	−0.2236	−0.2221

be calculated using the non-linear matrix expressions for the own- and cross-partial derivatives. The results from doing this are presented in Table 3, where we see substantial network effects. The network effects reflect cumulated off-diagonal elements of the matrix of partial derivatives (cross-partial derivatives) averaged over all countries as described in our discussion of model interpretation.

These estimates show larger direct and network impacts arising from changes in destination country than origin country income on trade flows in the case of both types of fixed effects. The Cheng and Wall (2005) fixed effects lead to larger direct and network destination effects than those from the Matyas (1997) fixed effects specification, but smaller origin-specific direct and network effects than those from the Matyas (1997) fixed effects specification.

Least-squares estimates were $\hat{\beta}_o = 0.9818, \hat{\beta}_d = 1.1562$ for the Matyas (1997) specification, and $\hat{\beta}_o = 0.9822, \hat{\beta}_d = 1.3032$ for the Cheng and Wall (2005) specification. The total effects estimates from the cross-sectional dependence models would be comparable to the least-squares estimates, and we see that ignoring network effects that arise from cross-sectional dependence lead to a substantial downward bias in the least-squares estimates.

Table 3: Partial derivative impacts for the W_{trade} models

	Matyas (1997) fixed effects		Cheng and Wall (2005) fixed effects	
Effects	GDP_o	GDP_d	GDP_o	GDP_d
direct	0.3475	0.3855	0.3395	0.5589
network	2.3160	2.5695	1.8371	3.0240
total	2.6635	2.9550	2.1766	3.5828

Table 4: Model comparison of convex combinations of $W-$matrices (Matyas, 1997, fixed effects)

Models	Log-marginal likelihood	Model probability	W_{space}	$W_{currency}$	$W_{language}$	W_{colony}	W_{trade}
Model 1	−519529.635	0.000	1	1	NA	NA	NA
Model 2	−514041.460	0.000	1	NA	1	NA	NA
Model 3	−512428.703	0.000	1	NA	NA	1	NA
Model 4	−507705.640	0.000	1	NA	NA	NA	1
Model 5	−534178.348	0.000	NA	1	1	NA	NA
Model 6	−539338.264	0.000	NA	1	NA	1	NA
Model 7	−513959.160	0.000	NA	1	NA	NA	1
Model 8	−535185.602	0.000	NA	NA	1	1	NA
Model 9	−512443.984	0.000	NA	NA	1	NA	1
Model 10	−503071.944	0.000	NA	NA	NA	1	1
Model 11	−510951.761	0.000	1	1	1	NA	NA
Model 12	−509278.763	0.000	1	1	NA	1	NA
Model 13	−507252.885	0.000	1	1	NA	NA	1
Model 14	−511418.909	0.000	1	NA	1	1	NA
Model 15	−505617.056	0.000	1	NA	1	NA	1
Model 16	−505331.781	0.000	1	NA	NA	1	1
Model 17	−521467.837	0.000	NA	1	1	1	NA
Model 18	−511281.957	0.000	NA	1	1	NA	1
Model 19	−502105.215	1.000	NA	1	NA	1	1
Model 20	−510301.764	0.000	NA	NA	1	1	1
Model 21	−508705.425	0.000	1	1	1	1	NA
Model 22	−504953.579	0.000	1	1	1	NA	1
Model 23	−503487.829	0.000	1	1	NA	1	1
Model 24	−509310.719	0.000	1	NA	1	1	1
Model 25	−508467.724	0.000	NA	1	1	1	1
Model 26	−504584.898	0.000	1	1	1	1	1

4.2 Extended versions of the network dependence models

We produced estimates for models based on averages of all 26 possible combinations of two or more weight matrices. For example, we define the combined weight matrix: $W_c = W_{space} + W_{trade} + W_{language} + W_{currency} + W_{colony}$, where W_c is row-normalized to have row-sums of unity. Log-marginal likelihoods are presented for these models in Table 4, for the specification based on Matyas (1997) fixed effects, and in Table 5 for the Cheng and Wall (2005) fixed effects specification.

In the Table 4 results, model #19 dominates all others leading to a posterior model probability of one assigned to this specification, based on $W_{currency} + W_{colony} + W_{trade}$. We also note that a comparison of the log-marginal likelihood for the best single weight matrix model from Table 1

Table 5: Model comparison of convex combinations of $W-$matrices (Cheng and Wall, 2005, fixed effects)

Models	Log-marginal likelihood	Model probability	W_{space}	$W_{currency}$	$W_{language}$	W_{colony}	W_{trade}
Model 1	−507048.345	0.000	1	1	NA	NA	NA
Model 2	−503766.884	0.000	1	NA	1	NA	NA
Model 3	−502532.222	0.000	1	NA	NA	1	NA
Model 4	−499358.674	0.000	1	NA	NA	NA	1
Model 5	−518174.284	0.000	NA	1	1	NA	NA
Model 6	−523989.738	0.000	NA	1	NA	1	NA
Model 7	−504428.222	0.000	NA	1	NA	NA	1
Model 8	−520322.487	0.000	NA	NA	1	1	NA
Model 9	−503695.091	0.000	NA	NA	1	NA	1
Model 10	−497304.291	0.000	NA	NA	NA	1	1
Model 11	−501405.745	0.000	1	1	1	NA	NA
Model 12	−500300.054	0.000	1	1	NA	1	NA
Model 13	−498846.687	0.000	1	1	NA	NA	1
Model 14	−501643.920	0.000	1	NA	1	1	NA
Model 15	−498027.006	0.000	1	NA	1	NA	1
Model 16	−495841.701	0.000	1	NA	NA	1	1
Model 17	−510634.416	0.000	NA	1	1	1	NA
Model 18	−502613.916	0.000	NA	1	1	NA	1
Model 19	−496487.855	0.000	NA	1	NA	1	1
Model 20	−500274.233	0.000	NA	NA	1	1	1
Model 21	−499479.361	0.000	1	1	1	1	NA
Model 22	−497376.837	0.000	1	1	1	NA	1
Model 23	−495177.846	1.000	1	1	NA	1	1
Model 24	−496643.681	0.000	1	NA	1	1	1
Model 25	−499171.712	0.000	NA	1	1	1	1
Model 26	−495836.711	0.000	1	1	1	1	1

shows that combinations of weight matrices produce a specification more consistent with our sample data. That is, the log-marginal likelihood for the model based on W_{trade} alone was $-5.1401e{+}05$, compared to that for model #19 based on three weight matrices of $-5.0210e{+}05$. The next best model was model #10 based on $W_{colony} + W_{trade}$ and the 3rd best model was model #23 based on $W_{space} + W_{currency} + W_{colony} + W_{trade}$.

The Table 5 results are based on the extended set of fixed effects from Cheng and Wall (2005) where we see that the best model (#23) is one based on $W_{space} + W_{currency} + W_{colony} + W_{trade}$, and the next best model (#26) included all five weight matrices, with the third-best model (#16) including $W_{space} + W_{colony} + W_{trade}$. What seems clear from the results in Table 4 and Table 5 is that membership in trade unions and historical colonial ties are an important source of interaction between countries' trade flows. The results from Table 5 place emphasis on W_{space} not found for the model based on simpler fixed effects. Recall that the model based on Cheng and Wall (2005) fixed effects whose results are presented in Table 5 represents the preferred model as it has higher log-marginal likelihood values.

5 Conclusions

A computationally efficient approach to MCMC estimation of a network dependence gravity model specification was set forth and used to examine the presence of a specific type of cross-sectional dependence in trade flows. The alternative simultaneous network dependence gravity model specification here is based on a spatial econometric specification set forth by LeSage and Pace (2008) that allows trade flows to be dependent on flows between countries that are spatial neighbors to the origin and destination countries. The extension set forth here allows for more general types of network dependence such as common currency, language, colonial ties or membership in trade unions. In cross-sectional gravity models these are typically treated as generalized distance variables, with the interpretation being that they reflect heterogeneity impacting the intercept term. In a panel data specification, these types of commonality between countries reflect time-invariant factors that are thought to be modeled by fixed effects.

We show that after including commonly used fixed effects of the type suggested by Maytas (1997) or Cheng and Wall (2005), there is evidence that network dependence in trade flows remains. Conventional gravity models assume the variable vector of $N^2 \times 1$ trade flows for each time period are independent, so trade flows between countries that have a common currency, language, border, colonial ties or are members of a trade union are no more likely than flows between countries having nothing in common.

Our specification allows these sociocultural factors to represent a basis for trade interaction between countries, with more similar flows between countries that share common borders, currency, language etc. Application of the model to a panel of trade flows covering 38 years and 70 countries provides evidence that this is the case. Network dependence produces simultaneous dependence, which means that flows from country dyad (i, j) depend on flows from other country dyads (say (k, l)), where the dependence structure is based on sociocultural factors. The most important sources of cross-sectional dependence were found to be trade organizations, historical colonial tries, common currency and spatial proximity of countries.

References

Balazsi L, Mátyás L and Wannsbeek, TJ (2018) The estimation of multidimensional fixed effects panel data models, *Econometric Reviews* 37(3), 212-227

Baltagi BH (2005) *Econometric Analysis of Panel Data.* Third edition. John Wiley & Sons, Chichester

Baltagi, B and Maasoumi, E (2013) An overview of dependence in cross-section, time-series, and panel data *Econometric Reviews* 32(5-6), 543-546

Baltagi BH, Egger P and Pfaffermayr M (2015) Panel data gravity models of international trade, in Baltagi BH (ed.) *The Oxford Hanbook of Panel Data*, Oxford University Press, pp. 322-346.

Cheng IH and Wall HJ (2005) Controlling for heterogeneity in gravity models of trade and integration, *Federal Reserve Bank of St. Louis Review* 87(1), 49-63

Debarsy N and LeSage JP (2017) Efficient estimation of spatial regression models based on convex combinations of different types of weight matrices, Paper presented at the North American Meetings of the Regional Science Association International, November, Vancouver, CA

Debarsy N and LeSage JP (2018) Flexible dependence modeling using convex combinations of different types of connectivity structures, *Regional Science & Urban Economics* 69(2), 46-68

Dittrich, D, Leenders, TAJ, and Mulder, J (2017) Bayesian estimation of the network autocorrelation model, *Social Networks* 48(1), 213-236

Egger P and Pfaffermayr M (2003) The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects, *Empirical Economics* 28(3), 571-580

Feenstra RC, Lipsey RE, Deng H, Ma AC and Mo H (2005) World trade flows: 1962-2000. NBER Working Paper Series 11040, http://www.nber.org/papers/w11040

Golub, GH, Van Loan, CF (1996) *Matrix Computation*, Third Edition. The Johns Hopkins University Press, Baltimore

Hazir CS, LeSage JP and Autant-Bernard C (2018) The role of R&D collaboration networks on regional innovation performance, *Papers in Regional Science* 97(3), 549-567

Hepple LW (1995a), Bayesian techniques in spatial and network econometrics: 1. model comparison and posterior odds, *Environment and Planning A* 27(3), 447-469

Hepple LW (1995b), Bayesian techniques in spatial and network econometrics: 2. computational methods and algorithms, *Environment and Planning A* 27(4), 615-644

Krisztin T and Fischer MM (2015) The gravity model for international trade: Specification and estimation issues, *Spatial Economic Analysis* 10(4), 451-470

LeSage JP (2018) Fast MCMC estimation of multiple W-matrix spatial regression models and Metropolis Hastings-Monte Carlo log-marginal likelihoods, submitted for publication in the Special Issue on Recent Developments in Spatial Econometrics, *Journal of Geographical Systems*

LeSage JP and Fischer MM (2018) Cross-sectional dependence model specifications in a static trade panel data setting, submitted for publication in the Special Issue on Recent Developments in Spatial Econometrics, *Journal of Geographical Systems*

LeSage JP and Pace RK (2008) Spatial econometric modeling of origin-destination flows, *Journal of Regional Science* 48(5), 941-967

LeSage JP and Pace RK (2009) *Introduction to Spatial Econometrics.* Taylor Francis/CRC Press, Boca Raton

LeSage JP and Thomas-Agnan C (2015) Interpreting spatial econometric origin-destination flow models, *Journal of Regional Science* 55(2), 188-208

LeSage JP, Chih Y-Y and Vance C (2018) Markov Chain Monte Carlo estimation of spatial dynamic panel models for large samples, Ruhr Economic Papers, forthcoming

Mátyás L (1997) Proper econometric specification of the gravity model, *The World Economy* 20(3), 363-368

Pace RK and LeSage JP (2002) Semiparametric maximum likelihood estimates of spatial dependence, *Geographical Analysis* 34(1), 75-90

Sarafidis V and Wansbeek T (2012) Cross-sectional dependence in panel data analysis, *Econometric Reviews* 31(5), 483-531

World Bank (2002) *World Development Indicators.* WTO, accessed on 2014-07-01

WTO (2014) *WTO Regional Trade Agreements Database.* WTO, Geneva

Zellner, A. (1971), *Introduction to Bayesian Inference in Econometrics.* John Wiley, New York.

APPENDIX A

Table A.1: List of countries

Algeria	Costa Rica	Kenya	South Africa
Australia	Denmark	Madagascar	Spain
Austria	Dominican Rep.	Malaysia	Sri Lanka
Bahamas	Ecuador	Mauritania	Sudan
Belgium	Fiji	Mexico	Suriname
Benin	Finland	Morocco	Sweden
Bolivia	France	Netherlands	Thailand
Brazil	Gabon	Nicaragua	Togo
Burkina Faso	Ghana	Niger	Trinidad and Tobago
Burundi	Greece	Nigeria	Uganda
Cameroon	Guatemala	Pakistan	United Kingdom
Canada	Guyana	Panama	United States
Central African Rep.	Honduras	Papua New Guinea	Uruguay
Chad	Hong Kong	Peru	
Chile	Ireland	Philippines	
China	Israel	Portugal	
Colombia	Italy	Senegal	
Congo, Dem. Rep.	Ivory Coast	Sierra Leone	
Congo, Rep.	Jamaica	Singapore	

Table A.2: Language ties: Common official and second languages (Krisztin and Fischer 2015)

English	**French**	**Spanish**	**Arabic**
Australia	Algeria	Bolivia	Algeria
Bahamas	Belgium	Chile	Chad
Cameroon	Benin	Colombia	Mauritania
Canada	Burkina Faso	Costa Rica	Morocco
Fiji	Burundi	Dominican Rep.	Sudan
Ghana	Cameroon	Ecuador	
Guyana	Canada	Guatemala	**Chinese**
Ireland	Cent. African Rep.	Honduras	China
Jamaica	Chad	Mexico	Hong Kong
Kenya	Congo, Dem. Rep.	Nicaragua	Malaysia
Nigeria	Congo, Rep.	Panama	Singapore
Pakistan	France	Peru	
Panama	Gabon	Spain	**Malay**
Papua New Guina	Ivory Coast	Uruguay	Malaysia
Philippines	Madagascar		Singapore
Sierra Leone	Morocco	**Dutch**	
Singapore	Niger	Belgium	
South Africa	Rwanda	Netherlands	
Sri Lanka	Senegal	Suriname	
Suriname	Togo		
Trinidad and Tobago			
Uganda	**Portuguese**		
United Kingdom	Brazil		
USA	Portugal		

Table A.3: Free trade and stronger forms of agreements in 2000 (Krisztin and Fischer 2015)

APTA	*CEMAC*	*EU*	Malaysia	*NAFTA*
Philippines	Burundi	Austria	Mexico	Canada
Sri Lanka	Cameroon	Belgium	Morocco	Mexico
	Central African Rep.	Demark	Nicaragua	USA
ASEAN [AFTA]	Chad	Finland	Pakistan	
Malaysia	Congo, Rep.	France	Peru	*PATCRA*
Philippines	Congo. Dem. Rep.	Greece	Philippines	Australia
Singapore	Gabon	Ireland	Singapore	Papua New Guinea
Thailand		Italy	Sri Lanka	
	COMESA	Netherlands	Sudan	*SICA*
CAN	Burundi	Portugal	Thailand	Costa Rica
Bolivia	Congo, Dem. Rep.	Spain	Trinidad and Tobago	Guatemala
Colombia	Kenya	Sweden		Honduras
Ecuador	Madagascar	United Kingdom	*LAIA*	Nicaragua
Peru	Sudan	Uruguay	Bolivia	
	Uganda		Brazil	*EU treaties*
CACM		*GSTP*	Chile	EU-Israel
Costa Rica	*ECOWAS*	Algeria	Colombia	EU-South Africa
Guatemala	Benin		Ecuador	
Honduras	Burkina Faso	Bolivia	Mexico	*Bilateral treaties*
Nicaragua	Ghana	Brazil	Panama	Canada-Chile
	Ivory Coast	Cameroon	Peru	Canada-Israel
CARICOM	Niger	Chile		Chile-Mexico
Bahamas	Nigeria	Colombia	*MERCOSUR*	Colombia-Mexico
Dominican Rep.	Senegal	Ecuador	Bolivia	Fiji-Papua New Guinea
Guyana	Sierra Leone	Ghana	Brazil	Israel-Mexico
Jamaica	Togo	Guyana	Chile	
Suriname			Uruguay	
Trinidad and Tobago				

Note: Asia Pacific Trade Agreement (APTA), Asian Free Trade Area (AFTA), Andean Community (CAN), Central American Common Market (CACM), Caribbean Community and Common Market (CARICOM), Economic Community of Central African States (CEMAC), Common Market for Eastern and Southern Africa (COMESA), Economic Community of West African States (ECOWAS), Global System of Trade Preferences among Developing Countries (GSTP), Latin American Integration Association (LAIA), Mercado Comun del Sur (MERCOSUR), North American Free Trade Agreement (NAFTA), Agreement on Trade between Australia and New Guinea (PATCRA), Central American Integration System (SICA) (Source: WTO (2014))

Table A.4: Common currency ties

Euro:	Austria, Belgium, France, Finland, Ireland, Italy, Netherlands, Portual, Spain
US Dollar:	United States, Bahamas[1], Panama
West African CFA Franc[2,4]:	Benin, Burkina Faso, Ivory Coast, Niger, Senegal, Togo
Central African CFA Franc[3,4]:	Cameroon, Central African Republic, Chad, Republic of Congo, Gabon

Notes: 1) The Bahamian dollar is bagged to the US dollar on a one-to one basis. 2) CFA stands for African Financial Community. It is issued by the Central Bank of the West African States, located in Dakar, Senegal, for the countries of the West African Economic and Monetary Union. 3) CFA stands for Financial Cooperation in Central Africa. It is issued by the Bank of Central African States, located in Yaoundé, Cameroon, for the countries of the Economic and Monetary Union of Central Africa. 4) The two CFA Franc currencies, although theoretically separate, are effectively interchangeable.

Table A.5: Direct colonial ties

UNITED KINGDOM	Nigeria	*FRANCE*	Morocco	Honduras
Australia	Pakistan	Algeria	Niger	Mexico
Bahamas	Sierra Leone	Benin	Senegal	Netherlands
Cameroon	South Africa	Burkina Faso	Togo	Nicaragua
Fiji	Sri Lanka	Cameroon		Panama
Ghana	Sudan	Central African Rep.	*SPAIN*	Peru
Hong Kong	Trinidad and Tobago	Chad	Bolivia	
Ireland	Uganda	Congo, Dem. Rep.	Chile	*BELGIUM*
Israel	United States	Congo, Rep.	Colombia	Congo, Dem. Rep.
Jamaica		Gabon	Costa Rica	
Kenya		Madagascar	Ecuador	*PORTUGAL*
Malaysia		Mauritania	Guatemala	Brazil