Network dependence in multi-indexed data on international trade flows

Manfred M. Fischer¹ · James P. LeSage²

Received: 14 April 2020 / Accepted: 22 September 2020 / Published online: 23 October 2020
© The Author(s) 2020

Abstract

Faced with the problem that conventional multidimensional fixed effects models only focus on unobserved heterogeneity, but ignore any potential cross-sectional dependence due to network interactions, we introduce a model of trade flows between countries over time that allows for network dependence in flows, based on sociocultural connectivity structures. We show that conventional multidimensional fixed effects model specifications exhibit cross-sectional dependence between countries that should be modeled to avoid simultaneity bias. Given that the source of network interaction is unknown, we propose a panel gravity model that examines multiple network interaction structures, using Bayesian model probabilities to determine those most consistent with the sample data. This is accomplished with the use of computationally efficient Markov Chain Monte Carlo estimation methods that produce a Monte Carlo integration estimate of the log-marginal likelihood that can be used 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 · MCMC estimation · Log-marginal likelihood · Gravity models of trade · Sociocultural distance

JEL Classification C18 · C33 · C51

✉ Manfred M. Fischer
manfred.fischer@wu.ac.at
<http://sites.google.com/view/professormanfredmfischer>
James P. LeSage
james.lesage@txstate.edu

¹ Vienna University of Economics and Business, Vienna, Austria

² Texas State University, San Marcos, TX, USA

1 Introduction

In recent years fixed effects model specifications have enjoyed great popularity in the panel gravity literature as a way to take into account the specific three-dimensional nature of international trade flow data sets.¹ These models, differing in the specification of the fixed effects, focus on unobserved heterogeneity, but ignore any potential cross-sectional dependence due to network interactions. Faced with this problem, we propose a panel gravity model that examines multiple network interaction structures, using Bayesian model probabilities to determine those most consistent with the sample data. This is accomplished with the use of computationally efficient Markov Chain Monte Carlo (MCMC) estimation methods that produce a Monte Carlo integration estimate of the log-marginal likelihood that can be used for model comparison.

Empirical international trade modeling has focused on multilateral resistance terms based on Anderson and van Wincoop (2003, 2004), which Feenstra (2002, 2004) argued could be proxied by inclusion of origin- and destination-specific fixed effects parameters in the model specification. Ranjan and Tobias (2007) model the effects parameters as a (semi-parametric) function of an exogenous variable measuring bilateral trade resistance between origin-destination dyads. Koch and LeSage (2015) make a convincing argument that all of these models relax multilateral aspects of the dependence structure in favor of bilateral dependence, or a specification based on heterogeneity rather than simultaneous/multilateral dependence. They argue this is important given Behrens et al. (2012) demonstration that model specifications that control for regional heterogeneity do not necessarily capture the full structure of interdependence between trade flows implied by theory. Behrens et al. (2012) derive a dual version of the Anderson and van Wincoop (2003) model using a model based on quantity competition that also implies a simultaneous structure of interdependence between observed trade flows. Koch and LeSage (2015) show how linearizing the system of nonlinear multilateral resistance relationships around a free trade world leads to a structured random effects interdependence structure that can be used as a Bayesian prior to produce statistical estimates of the inward and outward multilateral resistance indices. They argue their statistical approach has advantages over the non-stochastic numerical approach used by Anderson and van Wincoop (2003) to solve for the multilateral resistance indices.

In contrast to the empirical trade literature, the regional science literature regarding modeling of flows between regions has placed more emphasis on use of spatial autoregressive models that include spatial lags of flows from regions neighboring the origin and destination, an idea promoted by LeSage and Pace (2008), who devote an entire chapter to these models in their book (LeSage and Pace 2009, Chapter 8). This paper extends this type of spatial autoregressive models to a panel data model setting. The literature discussed above did not deal with panel

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.

data models where we have a series of flow matrices covering numerous time periods, but rather more conventional gravity models consisting of a single flow matrix.

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 et al. (2018) explore econometric implications of using a host of alternative multidimensional fixed effects in panel 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 Mátyás (1997) as an extension of the 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 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, membership in trade organizations, 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 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, along with computationally efficient MCMC estimation methods.²

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. We use the terms network and cross-sectional dependence interchangeably, but note that the network dependence specification introduced here reflects a special case of cross-sectional dependence that can arise

² Sarafidis 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. In this paper we take the spatial econometric approach, but note that a common factor specification could also be employed to address the issue we raise. 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.

in the case of origin-destination flows that has not received a great deal of attention in the literature.

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. The MCMC estimation approach allows for estimation and posterior inference on a vector of dependence parameters that determines the relative importance of network dependence. 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.³

Another methodological innovation is use of multiple network interaction structures 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. An equally weighted combination of these multiple weight matrices is used to form a single weight matrix. This approach allows us to treat sociocultural factors as sources of network dependence in the panel gravity model.

The remainder of this paper is organized as follows. Section 2 outlines 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 Section 3 introduces network dependence in a conventional multidimensional fixed effects model specification, and describes the prior setup of the MCMC estimation approach. Section 4 sets forth computational challenges to estimation along with MCMC procedures to overcome these challenges. Section 5 applies our model to panel data on trade flows between 70 countries covering the 38 years from 1963 to 2000. In the application of the model we consider 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 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 model specifications. Section 6 concludes the article and a technical appendix provides further details on the MCMC estimation approach.

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.

2 Multi-indexed panel gravity models

In multidimensional panel data sets, the dependent variable of a panel gravity model is observed along three indices, such as y_{ijt} , $i = 1, \dots, N_i$, $j = 1, \dots, N_j$, $t = 1, \dots, T$.

2.1 The Mátyás model

Mátyás (1997) made an early attempt to introduce multidimensional fixed effects for (log-linear) 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 $i = 1, \dots, N$, $j = 1, \dots, N$, and $t = 1, \dots, T$.

$$y_{ijt} = X_{ijt} \beta + \alpha_i + \gamma_j + \lambda_t + \varepsilon_{ijt}, \quad (1)$$

where the α_i , γ_j parameters are the origin and destination country-specific fixed effects, and λ_t are time-specific fixed effects. β is a $2K \times 1$ vector of parameters on the $N^2T \times 2K$ (logged) covariates $X_{ijt} = (X_{it}^1, X_{jt}^1, \dots, X_{it}^K, X_{jt}^K)$ with X_{it}^k (X_{jt}^k) denoting the k th measure for the economic size of the origin country i (destination country j) 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. It is assumed that ε_{ijt} are normal *i.i.d.* idiosyncratic disturbances with zero mean and scalar (σ_ε^2) variance.⁶

2.2 The Cheng and Wall model

There are of course other specifications for the fixed effects. For example, Egger and Pfaffermayr (2003) propose bilateral specific fixed effects γ_{ij} . A variant of this, proposed by Cheng and Wall (2005), popular in the empirical trade literature is shown in (2),

$$y_{ijt} = X_{ijt} \beta + \gamma_{ij} + \lambda_t + \varepsilon_{ijt}, \quad (2)$$

or in matrix form $y = X\beta + \Delta_\gamma \gamma + \Delta_\lambda \lambda + \varepsilon$ with corresponding dummy design matrices $\Delta_\gamma = (I_N \otimes I_N \otimes I_T)$ and $\Delta_\lambda = (I_N \otimes I_N \otimes I_T)$. Balazsi et al. (2018) point out that the model in (1) represents a special case of that in (2), and there is an analogy of this 3-dimensional (3D) situation in (2) to 2-dimensional (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, membership in trade organizations, border or colonial ties are no more likely than flows between countries having nothing in common. Network dependence in flows

⁵ See Baltagi et al. (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.

suggests that flows between countries with sociocultural similarity 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.

3 The econometric framework

3.1 The model

We set forth an extension of the conventional panel gravity model that allows for origin-, destination- and origin-destination-based network dependence. The matrix expressions in (3) represent the *network dependence panel gravity model for origin-destination flows* introduced in matrix form.

$$y = \rho_o I_T \otimes (W \otimes I_N) y + \rho_d I_T \otimes (I_N \otimes W) y + \rho_w I_T \otimes (W \otimes W) y + X_{ot} \beta_o + X_{dt} \beta_d + \Delta_\gamma \gamma + \Delta_\lambda \lambda + \varepsilon, \quad (3)$$

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 $N^2T \times K$ matrix X_{ot} contains origin-specific covariates with the associated $K \times 1$ parameter vector β_o , where $X_{ot} = X_t \otimes \iota_N$, and similarly for $X_{dt} = \iota_N \otimes X_t$ with X_t being an $N \times K$ matrix of covariates measuring the (economic) size of each destination country at time t , with associated β_d . The $N^2T \times 1$ vector ε represents the normally distributed *i.i.d.* scalar variance disturbances. The matrices of fixed effects $\Delta_\gamma \gamma + \Delta_\lambda \lambda$ are as defined after (2), or we could use fixed effects from Mátyás (1997) shown in (1) as well.

The model in (3) indicates that flows at *each time period* $t = 1, \dots, T$ exhibit dependence on flows of countries that are *sociocultural neighbors* to 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 $N \times N$ weight matrix that defines neighboring countries based on sociocultural connectivity structures. 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 $(W \otimes I_N) y_t$ reflecting a linear combination of trade flows at time t from countries that are sociocultural neighbors to the origin country.

The model also allows for dependence of flows in each time period from countries that are sociocultural neighbors to 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 (sociocultural) neighbors. The matrix $(I_N \otimes W)$ defines neighbors to the destination, the matrix $(W \otimes I_N)$ identifies those 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 sociocultural neighbors to both the origin and destination countries. Following ideas expressed in LeSage and Pace (2008) we motivate this type of dependence using the (cross-sectional) specification in (4), where they argue that the matrix A can be viewed as a filter. For notational convenience in (4) (as in the sequel), we have defined the matrix Z to include X_{ot}, X_{dt} and the fixed effects, with the parameters β adjusted appropriately.

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

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 $W_w y$. LeSage and Pace (2008) note that this filter implies a restriction that $\rho_w = -\rho_o \rho_d$, but argue this restriction needs not be imposed during estimation, so we address the more general case here and allow for an unrestricted parameter ρ_w .⁷

3.2 MCMC estimation

The model can be written in a computationally efficient form as shown in (5).

$$\begin{aligned} \tilde{y} \omega &= Z\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_{N^2T}), \\ \tilde{y} &= [y \quad I_T \otimes (W \otimes I_N)y \quad I_T \otimes (I_N \otimes W)y \quad I_T \otimes (W \otimes W)y], \\ \omega &= (1 \quad -\rho_o \quad -\rho_d \quad -\rho_w)'. \end{aligned} \quad (5)$$

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.

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 $e'e$.

⁷ 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.

$$f(y; \omega, \sigma^2, \beta) = |R(\omega)| (2\pi\sigma^2)^{-N^2T/2} \exp \left(-\frac{1}{2\sigma^2} e'e \right), \quad (6)$$

$$e = \tilde{y} \omega - Z\beta,$$

$$R(\omega) = I_{N^2T} - \rho_o I_T \otimes (W \otimes I_N) - \rho_d I_T \otimes (I_N \otimes W) - \rho_w I_T \otimes (W \otimes W).$$

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 Ξ for the set of parameters $(\omega, \sigma^2, \beta)$ is: $\Xi := \Xi_\omega \times \Xi_{\sigma^2} \times \Xi_\beta = \Xi_{-1,1} \times (0, \infty) \times \mathcal{R}^{2K+M}$ (where M is the number of fixed effects included in the matrix Z).⁸

Since our focus is on large samples involving N^2T observations, we rely on uninformative priors for the parameters β and σ^2 , as these would not likely impact posterior estimates. By uninformative prior we mean a “flat” prior that is within the range of valid values for these parameters, so all intervals are of equal length, indicating all values are equally likely.⁹ 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, β can be represented as $p(\omega) \propto 1, p(\beta) \propto 1$. 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: $p(\omega) \times p(\sigma^2) \times p(\beta) \propto 1/\sigma^2$. While this flat prior is improper since the integral over the parameter space Ξ 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 β, σ^2 , with details regarding this integration found in Hepple (1995a, b).

To derive the joint posterior for the dependence parameters ω , begin with the full joint posterior $p(\omega, \beta, \sigma^2)$ and analytically integrate out β, 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 $\Gamma(\cdot)$ denotes the Gamma function.

$$f(y; \omega, \sigma^2, \beta) \propto |R(\omega)| (2\pi\sigma^2)^{-N^2T/2} \exp \left(-\frac{1}{2\sigma^2} \omega' u' u \omega \right), \quad (7)$$

$$u = \tilde{y} - Z\beta_d,$$

$$\beta_d = (Z'Z)^{-1}Z'\tilde{y},$$

⁸ The number of fixed effects M in the matrix Z differs for the Mátyás (1997) specification where $M = 2N + T - 2$ and the Cheng and Wall (2005) specification where $M = N^2 + T - 1$.

⁹ See Zellner (1971, pp. 41–53) for a discussion of uninformative priors.

$$p(\omega, \beta|y) \propto |R(\omega)| \int_0^\infty \sigma^{-(N^2T+1)} \exp\left(-\frac{1}{2\sigma^2} \omega' u' u \omega\right) d\sigma \quad (8)$$

$$= 2^{(N^2T-2)/2} \Gamma(N^2T/2) |R(\omega)| (\omega' u' u \omega)^{-N^2T/2}. \quad (9)$$

To integrate out the $2K + M$ different β 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 $-(N^2T - 2K - M)/2$ arising from this integration (see Hepple 1995a, b). This expression must be numerically integrated to arrive at the log-marginal likelihood for these models.

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

Dittrich et al. (2017) show there is no problem regarding posterior propriety for the cross-sectional model case and this result carries over to our model in a relatively straightforward and trivial way. The conditional distributions for the model parameters β, σ^2, ω needed to implement MCMC estimation are described in “Appendix A”.

A motivation for having analytically integrated out the parameters β, σ^2 is that further integration of the joint conditional posterior over the set of dependence parameters in ω , yields 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 (described later in Sect. 4.1). 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 ($\log m_i$), we can calculate: $\text{prob}(m_i) = \exp(\log m_i) / \sum_{i=1}^Q \exp(\log m_i)$ 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 weight matrices, or different fixed effects specifications.

4 Computational challenges to estimating the model

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 $-1 < \rho_o + \rho_d + \rho_w < 1$.¹⁰

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 specifications 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 and could be demanding and quickly hit a computational bottleneck. “Appendix B” shows how to side step this computational bottleneck.

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 in Sect. 4.1. 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 for the model specifications considered here.

Another aspect of our network dependence model relates to 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 r th explanatory variable. We take this issue up in Sect. 4.2. MCMC estimation proceeds by sampling sequentially from the conditional distributions of each parameter (or set of parameters).

4.1 Block sampling the dependence parameters

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

¹⁰ 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.

Metropolis-Hastings scheme for the dependence parameters. Block sampling means that a vector of dependence parameters in ω is 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.

Debarys 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, Debarys and LeSage (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 (see also LeSage et al. 2019).

We rely here on a reversible jump procedure to produce proposal values for the vector of parameters ρ_o , ρ_d , ρ_w . In particular, we rely on a three-headed coin flip for each scalar parameter. 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, σ_{ρ_o} , σ_{ρ_d} , σ_{ρ_w} for each parameter are calculated based on the first 1000 draws (and updated thereafter using an interval of $m = 1000$ 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 (2020) provide Monte Carlo evidence on this type of procedure.

4.2 Interpreting the model

The partial derivatives used to interpret how changes in (say the r th) 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 (11), where we record the $N \times N$ matrices of changes in (logged) flows arising from changing the r th 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 r th variable in each country i . We 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.

$$\begin{pmatrix} \partial Y_1 / \partial X_1^r \\ \partial Y_2 / \partial X_2^r \\ \vdots \\ \partial Y_N / \partial X_N^r \end{pmatrix} = (I_{N^2} - \rho_o W_o - \rho_d W_d - \rho_w W_w)^{-1} \begin{pmatrix} Jd_1 \beta_d^r + Jo_1 \beta_o^r \\ Jd_2 \beta_d^r + Jo_2 \beta_o^r \\ \vdots \\ Jd_N \beta_d^r + Jo_N \beta_o^r \end{pmatrix}. \quad (11)$$

In (11), Jd_i ($i = 1, \dots, N$) is an $N \times N$ matrix of zeros with the i th row equal to $\iota_N' \beta_d$, and Jo_i is an $N \times N$ matrix of zeros with the i th 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 $X_i^r, i = 1, \dots, N$) resulting in an $N^2 \times N$ matrix of partial derivatives reflecting the total effect on flows from changing the r th characteristic of all N regions, which in the gravity modeling literature is labelled the *total effect* (LeSage and Thomas-Agnan 2015, and LeSage and Fischer 2016).

LeSage and Thomas-Agnan (2015) provide a motivation for the expression in (11), 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 $i = 1, \dots, N$ 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. While 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 (11) 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 use an average of the cumulative off-diagonal elements from each row of the N different $N \times N$ matrices in (11) to summarize network effects arising from changes in (say) income in a typical country. *Network effects* are measured by 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 (11). 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.¹¹

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.

5 Application of the model

5.1 Context and different interaction structures

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)

¹¹ Given sparse matrices W it would not be difficult to calculate the matrix inverse for situations involving the typical sample of 100–200 countries used in trade flow models.

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 6 in "Appendix C") was constructed for which GDP, population and trade flows were available over the 38 years.¹²

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 Mátyás (1997) model in (1), and $N^2 + T - 1 = 4,937$ fixed effects parameters in the Cheng 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 Tables 7, 8, 9 and 10 in "Appendix C".¹³

5.2 Model comparison of alternative structures of network dependence

Estimates from the model in (12) where the parameters ρ_o, ρ_d, ρ_w are (significantly) different from zero point to the existence of cross-sectional dependence.

$$y = \rho_o W_1 y + \rho_d W_2 y + \rho_w W_3 y + GDP_o \beta_o + GDP_d \beta_d + \Delta_i \gamma + \Delta_i \lambda + \varepsilon, \\ W_1 = I_T \otimes (W \otimes I_N), \quad W_2 = I_T \otimes (I_N \otimes W), \quad W_3 = I_T \otimes (W \otimes W). \quad (12)$$

In the presence of dependence, estimates from conventional models that ignore cross-sectional/network dependence can be shown to be biased and inconsistent (see LeSage and Fischer 2020). 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 spatial

¹² 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 from our sample.

¹³ A reviewer pointed out that matrices such as $W_{currency}, W_{trade}$ might be viewed as variable over time. Introducing weight matrices that vary over time greatly complicates estimation as well as interpretation as shown in Lee and Yu (2012). Consider that in this case, the partial derivative effect of a change in X_{ot} on the outcome variable vector y would require that we take into account any (non-linear) impact arising from change in the connectivity structure over time.

Table 1 Log-marginal likelihood estimates for alternative models

Model	Model with Mátyás fixed effects		Model with Cheng and Wall fixed effects	
	log-marginal likelihood	$\rho_o + \rho_d + \rho_w$	log-marginal likelihood	$\rho_o + \rho_d + \rho_w$
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

autoregressive 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 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 Mátyás (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 model specifications based on W_{trade} have the highest log-marginal likelihood, those based on W_{space} are next highest, followed by $W_{language}$. Turning to estimates for the dependence parameters, we see that the sums of these are substantially positive, pointing to the presence of network dependence.

5.3 Parameter estimates and partial derivative impacts for the best model

Table 2 presents estimates for the best models based on W_{trade} using both the Mátyás (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 5000 retained MCMC draws (with an initial 5000 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% 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 in Sect. 4.2, 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 be calculated using the non-linear matrix expressions for the own- and cross-partial derivatives. The

Table 2 Estimates for the W_{trade} models

Parameter	Mátyás (1997) fixed effects				
	Mode	Mean	Median	MC error	Geweke
Constant	−9.0215	−9.0177	−9.0187	0.00660826	0.997389
β_o	0.2863	0.2865	0.2864	0.00030498	0.997029
β_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
Parameter	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
Parameter	Cheng and Wall (2005) fixed effects				
	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
Parameter	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

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 Sect. 4.2.

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

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

<i>Effects</i>	Mátyás (1997) fixed effects		Cheng and Wall (2005) fixed 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

and network destination effects than those from the Mátyás (1997) fixed effects specification, but smaller origin-specific direct and network effects than those from the Mátyás (1997) fixed effects specification.

Least-squares estimates were $\hat{\beta}_o = 0.9818$, $\hat{\beta}_d = 1.1562$ for the Mátyás (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.

5.4 Model comparison of extended versions of the model

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. Note that we (implicitly) assign equal weight to all matrices in each model, so a model containing three matrices (e.g., W_{space} , W_{trade} , $W_{language}$) assigns a weight of 1/3 to each of these.¹⁴ Log-marginal likelihoods are presented for these models in Table 4 for the specification based on Mátyás (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 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 $W_{space} + W_{currency} + W_{colony} + W_{trade}$.

¹⁴ Debarsy and LeSage (2018) discuss estimating individual weights in a simple cross-sectional setting, but only for the case of two matrices. Adding additional weight parameters would greatly complicate estimation and not allow the calculation of log-marginal likelihoods as done here.

Table 4 Model comparison of combinations of W -matrices (based on fixed effects from Mátyás 1997)

Model	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

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 Tables 4 and 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.

Table 5 Model comparison of combinations of W -matrices (based on fixed effects from Cheng and Wall 2005)

Model	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

6 Conclusions

A computationally efficient approach to MCMC estimation of a multi-indexed network dependence panel gravity model was set forth and used to examine the presence of a specific type of network dependence in trade flows. The network dependence model specifications are based on a combination of multiple weight matrices reflecting different sources of network dependence such as common currency, language, colonial ties, membership in trade unions or spatial proximity of countries. 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 Mátyás (1997) or Cheng and Wall (2005), there is evidence that significant 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 panel gravity model 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) with $k \neq i, l \neq j$), 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 ties, common currency and spatial proximity of countries.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funding Open access funding provided by Vienna University of Economics and Business (WU).

Appendix A: Conditional distributions for the model parameters

Since our focus is on large samples N^2T , we can rely on uninformative priors for the parameters β , as these would not likely impact posterior estimates. By uninformative prior we mean a “flat” prior that is within the range of valid values for these parameters, so all intervals are of equal length, indicating all values are equally likely. (See Zellner (1971, pp. 41–53) for a discussion of uninformative priors.) 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 in Sect. 4. Given the limited prior information, the conditional distribution for the parameters β of the model in (5) takes the form of a multivariate normal with mean and variance-covariance shown in (A.1).

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

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 β (given values for the parameters σ^2, ω) can take place in a rapid, computationally efficient way.

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

$$\begin{aligned}
 p(\sigma^2|\beta, \omega) &\propto (\sigma^2)^{-(\frac{N^2T}{2})} \exp\left(-\frac{1}{2\sigma^2}(e'e)\right), \\
 e &= (\tilde{y} \mid \omega - Z\beta) \\
 &\sim \mathcal{IG}(\tilde{a}, \tilde{b}), \\
 \tilde{a} &= N^2 T/2, \\
 \tilde{b} &= (e'e)/2.
 \end{aligned}
 \tag{A.2}$$

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

$$\begin{aligned}
 \log p(\omega|y, Z, W) &\propto \log[D(\omega)] - (N^2T/2) \log(\omega'F\omega), \\
 F &= (\tilde{y} - Z\beta_d)'(\tilde{y} - Z\beta_d), \\
 \beta_d &= Z(Z'Z)^{-1}\tilde{y},
 \end{aligned}
 \tag{A.3}$$

where $\log[D(\omega)]$ is a Taylor series approximation to the log-determinants in the model, described in detail in “Appendix B”. We note here that this log-determinant term depends on the dependence parameters in the vector ω , indicated by $D(\omega)$. 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 .

Appendix B: A Taylor’s series approximation to the log-determinant term

In “Appendix A”, we have motivated that (A.3) represents a computationally efficient expression for the joint posterior, but this involves the log-determinant term $\log[D(\omega)]$ in (B.1), 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 inherently demanding and quickly hit a computational bottleneck.

$$\log|I_{N^2T} - \rho_o W_1 - \rho_d W_2 - \rho_w W_3|. \quad (\text{B.1})$$

One way to side step this computational bottleneck is by approximation to the log-determinant term. Pace and LeSage (2002) set forth a Taylor series approximation for the log-determinant of a matrix like our expression: $\log|I_{NT} - \tilde{W}|$, where $\tilde{W} = (\rho_o W_1 + \rho_d W_2 + \rho_w W_3)$. They show that for a *symmetric* nonnegative weight matrix $\tilde{W}$ with eigenvalues $\lambda_{\min} \geq -1$, $\lambda_{\max} \leq 1$, and $\text{tr}(\tilde{W}) = 0$ where tr represents the trace:

$$\begin{aligned} \log|I_{N^2T} - \rho_o W_1 - \rho_d W_2 - \rho_w W_3| &= \log|I_{N^2T} - \tilde{W}|, \\ \tilde{W} &= (\rho_o W_1 + \rho_d W_2 + \rho_w W_3) \\ &= \text{tr}(\log(I_{N^2T} - \tilde{W})), \\ \log(I_{N^2T} - \tilde{W}) &= - \sum_{i=1}^{\infty} \tilde{W}^i / i, \\ \log|I_{N^2T} - \tilde{W}| &= - \sum_{i=1}^{\infty} \text{tr}(\tilde{W}^i) / i \end{aligned} \quad (\text{B.2})$$

$$\begin{aligned} &\simeq - \sum_{j=1}^q \text{tr}(\tilde{W}^j) / j, \\ \text{tr}(\tilde{W}) &= \text{tr}(\rho_o W_1 + \rho_d W_2 + \rho_w W_3) \\ &= \rho_o \text{tr}(W_1) + \rho_d \text{tr}(W_2) + \rho_w \text{tr}(W_3). \end{aligned} \quad (\text{B.3})$$

Golub and van Loan (1996, p. 566) provide the expression in (B.2), while (B.3) arises from linearity of the trace operator. Note that the first-order $\text{tr}(\tilde{W})$ is zero, given the definitions of W_1, W_2, W_3 . Let $\eta = (\rho_o \quad \rho_d \quad \rho_w)'$, and first consider the case of symmetric matrices W_1, W_2, W_3 , which allows the second-order trace to be expressed as a quadratic form in (B.5) involving the vector of parameters η and all pairwise multiplications of the individual matrices in $\tilde{W}$ as shown in (B.4).

$$\text{tr}(\tilde{W}^2) = \eta' \begin{pmatrix} \text{tr}(W_1^2) & \text{tr}(W_1 W_2) & \text{tr}(W_1 W_3) \\ \text{tr}(W_2 W_1) & \text{tr}(W_2^2) & \text{tr}(W_2 W_3) \\ \text{tr}(W_3 W_1) & \text{tr}(W_3 W_2) & \text{tr}(W_3^2) \end{pmatrix} \eta \quad (\text{B.4})$$

$$= \eta' Q^2 \eta. \quad (\text{B.5})$$

LeSage and Pace (2009) point out that accelerated computation of traces can be accomplished using sums of matrix Haddamard products, $Q_{ij}^2 = \sum_i^3 \sum_j^3 W_i \odot W_j$, $i = 1, 2, 3; j = 1, 2, 3$. 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 $(\eta \otimes \eta) \text{vec}(Q^2)$, 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 $\log|I_n - \tilde{W}|$ takes the form in (B.6).

$$\begin{aligned} \log|I_{N^2T} - \tilde{W}| &\simeq -(\eta \otimes \eta) \text{vec}(Q^2)/2 \\ &\quad - (\eta \otimes \eta) \otimes \eta \text{vec}(Q^3)/3 \\ &\quad - ((\eta \otimes \eta) \otimes \eta) \otimes \eta \text{vec}(Q^4)/4. \end{aligned} \quad (\text{B.6})$$

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.

Appendix C: Countries and network interaction structures

The countries used are listed in Table 6, while Tables 7, 8, 9 and 10 provide information on the sociocultural network connectivity structures.

Table 6 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 7 Language ties: common official and second languages (Krisztin and Fischer 2015)

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

Table 8 Free trade and stronger forms of agreements in 2000 (Krisztin and Fischer 2015)

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

Table 8 continued

Trade Agreement	Countries
<i>Bilateral Treaties</i>	Canada-Chile, Canada-Israel, Chile-Mexico, Colombia-Mexico, Fiji-Papua New Guinea, Israel-Mexico
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 9 Common currency ties

<i>Euro</i>	Austria, Belgium, France, Finland, Ireland, Italy, Netherlands, Portugal, Spain
<i>US Dollar</i>	United States, Bahamas ¹ , Panama
<i>West African CFA Franc</i> ^{2,4}	Benin, Burkina Faso, Ivory Coast, Niger, Senegal, Togo
<i>Central African CFA Franc</i> ^{3,4}	Cameroon, Central African Republic, Chad, Republic of Congo, Gabon

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 10 Direct colonial ties

<i>United Kingdom</i>	Australia, Bahamas, Cameroon, Fiji, Ghana, Hong Kong, Ireland, Israel, Jamaica, Kenya, Malaysia, Nigeria, Pakistan, Sierra Leone, South Africa, Sri Lanka, Sudan, Trinidad and Tobago, Uganda, United States
<i>France</i>	Algeria, Benin, Burkina Faso, Cameroon, Central African Rep., Chad, Congo, Dem. Rep., Congo, Rep., Gabon, Madagascar, Mauritania, Morocco, Niger, Senegal, Togo
<i>Spain</i>	Bolivia, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Honduras, Mexico, Netherlands, Nicaragua, Panama, Peru
<i>Portugal</i>	Brazil
<i>Belgium</i>	Congo, Dem. Rep.

References

- Anderson JE, van Wincoop E (2003) Gravity with gravitas: a solution to the border puzzle. *Am Econ Rev* 93(1):170–192
- Anderson JE, van Wincoop E (2004) Trade costs. *J Econ Lit* 42(3):691–751
- Balazsi L, Mátyás L, Wansbeek TJ (2018) The estimation of multidimensional fixed effects panel data models. *Econom Rev* 37(3):212–227
- Baltagi BH (2005) *Econometric analysis of panel data*, 3rd edn. Wiley, Chichester

- Baltagi B, Maasoumi E (2013) An overview of dependence in cross-section, time-series, and panel data. *Econom Rev* 32(5–6):543–546
- Baltagi BH, Egger P, Pfaffermayr M (2015) Panel data gravity models of international trade. In: Baltagi BH (ed) *The Oxford handbook of panel data*. Oxford University Press, Oxford, pp 322–346
- Behrens K, Ertur C, Koch W (2012) Dual gravity: using spatial econometrics to control for multilateral resistance. *J Appl Econ* 27(5):773–794
- Cheng IH, Wall HJ (2005) Controlling for heterogeneity in gravity models of trade and integration. *Federal Res Bank St Louis Rev* 87(1):49–63
- Debarsy N, LeSage JP (2018) Flexible dependence modeling using convex combinations of different types of connectivity structures. *Reg Sci Urban Econ* 69(2):46–68
- Debarsy N, LeSage JP (2020) Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity structures. *J Bus Econ Stat* (**forthcoming**)
- Dittrich D, Leenders TAJ, Mulder J (2017) Bayesian estimation of the network autocorrelation model. *Soc Netw* 48(1):213–236
- Egger P, Pfaffermayr M (2003) The proper panel econometric specification of the gravity equation: a three-way model with bilateral interaction effects. *Empir Econ* 28(3):571–580
- Feenstra RC (2002) Border effects and the gravity equation: consistent methods for estimation. *Scott J Polit Econ* 49(5):491–506
- Feenstra RC (2004) *Advanced international trade*. Princeton University Press, Princeton
- Feenstra RC, Lipsey RE, Deng H, Ma AC, 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*, 3rd edn. The Johns Hopkins University Press, Baltimore
- Hazir CS, LeSage JP, Autant-Bernard C (2018) The role of R&D collaboration networks on regional innovation performance. *Papers Reg Sci* 97(3):549–567
- Hepple LW (1995a) Bayesian techniques in spatial and network econometrics: 1. Model comparison and posterior odds. *Environ Plan A* 27(3):447–469
- Hepple LW (1995b) Bayesian techniques in spatial and network econometrics: 2. Computational methods and algorithms. *Environ Plan A* 27(4):615–644
- Koch W, LeSage JP (2015) Latent multilateral trade resistance indices: theory and evidence. *Scott J Polit Econ* 62(3):264–290
- Krisztin T, Fischer MM (2015) The gravity model for international trade: specification and estimation issues. *Spat Econ Anal* 10(4):451–470
- Lee L, Yu J (2012) QML estimation of spatial dynamic panel data models with time varying spatial weights matrices. *Spat Econ Anal* 7(1):31–74
- LeSage JP (2020) Fast MCMC estimation of multiple W-matrix spatial regression models and Metropolis-Hastings Monte Carlo log-marginal likelihoods. *J Geogr Syst* 22(1):47–75
- LeSage JP, Fischer MM (2020) Cross-sectional dependence model specifications in a static trade panel data setting. *J Geogr Syst* 22(1):5–46
- LeSage JP, Fischer MM (2016) Spatial regression-based model specifications for exogenous and endogenous spatial interaction. In: Patuelli R, Arbia G (eds) *Spatial interaction modelling*. Springer, Berlin, pp 15–36
- LeSage JP, Pace RK (2008) Spatial econometric modeling of origin-destination flows. *J Reg Sci* 48(5):941–967
- LeSage JP, Pace RK (2009) *Introduction to spatial econometrics*. Taylor Francis/CRC Press, Boca Raton
- LeSage JP, Thomas-Agnan C (2015) Interpreting spatial econometric origin-destination flow models. *J Reg Sci* 55(2):188–208
- LeSage JP, Chih Y-Y, Vance C (2019) Markov Chain Monte Carlo estimation of spatial dynamic panel models for large samples. *Comput Stat Data Anal* 138:107–125
- Mátyás L (1997) Proper econometric specification of the gravity model. *World Econ* 20(3):363–368
- Pace RK, LeSage JP (2002) Semiparametric maximum likelihood estimates of spatial dependence. *Geograph Anal* 34(1):75–90
- Ranjan R, Tobias JL (2007) Bayesian inference for the gravity model. *J Appl Econ* 22(4):817–838
- Sarafidis V, Wansbeek T (2012) Cross-sectional dependence in panel data analysis. *Econom Rev* 31(5):483–531
- World Bank (2002) *World development indicators*. The World Bank, Washington
- WTO (2014) *WTO regional trade agreements database*. WTO, Geneva
- Zellner A (1971) *Introduction to Bayesian inference in econometrics*. Wiley, New York