ASOCIACION ARGENTINA DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

LIII Reunión Anual

Noviembre de 2018

ISSN 1852-0022
ISBN 978-987-28590-6-0

Network effects error components models

Montes Rojas Gabriel

Network effects error components models

Gabriel Montes-Rojas*

June 14, 2018

Abstract

This paper develops a random effects error components structure for network data regression models. In particular, it allows for edge and triangle specific components, which serve as a basal model for modeling network effects. It then evaluates the potential effects of ignoring network effects in the estimation of the variance-covariance matrix. Network effects will typically imply heteroskedasticity, and as with the Moulton factor, the key role is given by the joint consideration of the intra-network correlation of the error term(s) and the covariates. Then it proposes consistent estimator of the variance components and Lagrange Multiplier tests for evaluating the appropriate model of random components in networks. Monte Carlo simulations show the tests have very good performance in finite samples.

Keywords: Networks; Clusters; Moulton factor.
JEL Classification: C2; C12.

*Instituto Interdisciplinario de Economía Politica de Buenos Aires (IIEP-BAIRES-UBA) and CONICET, Facultad de Ciencias Económicas, Universidad de Buenos Aires, Córdoba 2122, 2do piso, C1120AAQ, Ciudad Autónoma de Buenos Aires, Argentina, e-mail: gabriel.montes@fce.edu.ar

1 Introduction

Statistical inference when data are grouped into clusters is an important issue in empirical work, and failure to control for within-cluster correlation can lead to misleadingly small standard errors (see the discussion in Cameron and Miller [18]). This is especially important when using aggregate variables on micro units in which ordinary least-squares (OLS) standard errors are seriously underestimated. The seminal work of Moulton ([40, 41, 42]) allows for a quantification of this potential pitfall, a fact that has been emphasized in the Angrist and Pischke ([1], ch.8) textbook among many others (see Montes-Rojas [38]).

A particular data structure related to cluster effects is that of networks. Matched data, where the interaction among agents is observed, are one type of such network data, where the information on who is in direct or indirect contact with whom matters. This has attracted a considerable attention with regards to spillover effects in education, production, financial markets, trade and many others. See de Paula [19] for a recent selective review of the literature.

Within a given network observations are not independent and the dependence structure is related to the network position of the observation. There is no obvious pattern to construct clusters or groups. Network models differ from cluster ones in the heterogeneity of the groups which need to be defined ad-hoc within the network as there are no obvious way to group observations. The most obvious type of intra-network correlation arises when we consider observations given by vertices or nodes that have a common edge or link. If we consider a link-specific effect, this would result in a specific factor that arises for linked nodes and not for others. Nodes that share a link might be correlated with each other.

We are mostly concerned with a linear regression model where observations are the nodes and specifically with the correct estimation of the variance-covariance structure. Thus we explore error components structure where the components depend on local network features of the observations. In particular, for a given graph we construct the error components model when considering link and triangle specific effects.

The main purpose of this exercise is that the empirical researcher starts from a standard variance-covariance structure (i.e., independent error components), and then tests sequentially for potential components' patterns (i.e., edges, triangles, diamonds, cycles, etc.).

First, contrary to the standard error components models, network effects will typically imply heteroskedastiticy. Take for instance the vertice&edge-only error components model where each vertice will have a vertice specific random component and an edge specific random component. Vertices that have one link are different from those that have two or more. The edge specific component will in fact generate a higher variance for vertices with more links.

Second, as with the Moulton factor, the key role is given by the joint consideration of the intra-network correlation of the error term(s) and the covariates. More formally, given an intra-network covariance structure of the error term and one of the covariates, the potential effect of misspecifying the variance-covariance of the estimators will depend comparison of the different correlation model will depend on the sample intra-network covariance between the covariance factors of the error term and the covariates.

In most empirical settings, both covariance factors are positively correlated (i.e., a high correlation between two unobservables usually corresponds to a high correlation between the covariates), and thus this determines that the OLS estimator variance that do not consider the potential network effects will underestimate the true variance. In particular, in the special case of covariates with no intra-network correlation, the standard OLS variance is correct.

This paper differs from the literature in several ways. First, many network related contributions focus on dyadic data structures where the unit of observation is the pair, i.e. the link, rather than the node. Among these, Hoff et al. [25, 24] develop likelihood models. Second, most of the linear regression network models using nodes as the unit of observations build upon spatial regression models. The seminal contribution is Manski [36]. Spatial models have the advantage of estimating fewer parameters (the spatial autoregressive parameter) than our proposed network random effects components. Moreover, they will not face the restrictions determined by the nonnegativity contraints, but they face problems of their own. Many network features that can be modeled from imposing additional parameters on the powers of the adjacency matrix.

2 Network error components model

2.1 Network definitions and notation

Consider an undirected graph $G = (V, L)$ as a mathematical structure consisting of a set V of vertices (also commonly called nodes) and a set L of edges (also commonly called links). Unless otherwise specified the graph is undirected where elements of L are unordered pairs (i, s) of distinct vertices $(i, s) \in V \times V$. If the graph were directed where the elements of L are ordered pairs $(i, s) \in V \times V$. The number of vertices is $N = |V|$ and the number of edges is $M = |L|$. Without loss of generality, we will label the vertices simply with the integers $1, \ldots, N$, and the edges, $1, \ldots, M$. Note that $M \leq N(N-1)/2$ for undirected graphs (and $M \leq N(N-1)$ for directed ones).

For our purposes consider a set of triangles in undirected graphs as $Triangles = \{(i, s, r) \in V^3, i < s < r, (i, s), (s, r), (i, r) \in L^3\}$, the number of triangles is $T \leq N(N-1)(N-2)/6$. The set of triangles could be defined differently for directed graphs.

The fundamental connectivity of a graph G may be captured in an $N \times N$ binary matrix A with entries

$$a_{is} = \begin{cases} 1 & \text{if vertices } \{i, s\} \in L \\ 0 & \text{otherwise} \end{cases},$$

the edge-incidence matrix B, an $N \times M$ binary matrix with entries

$$b_{ij} = \begin{cases} 1 & \text{if vertex } i \text{ is incident to edge } j \\ 0 & \text{otherwise} \end{cases},$$

and the triangle incidence matrix C, an $N \times T$ binary matrix with entries

$$c_{ik} = \begin{cases} 1 & \text{if vertex } i \text{ is incident to triad } k \\ 0 & \text{otherwise} \end{cases}.$$

For an undirected network A is symmetric and we can define the vertices' degree $\{d_i\}_{i=1}^N$ which can be obtained by $diag(BB^\top)$, and vertices' triangles $\{t_i\}_{i=1}^N$ which can be obtained by $diag(CC^\top)$.

The definitions above correspond to unweighted networks. We could extend this to weighted networks by defining an $N \times N$ binary matrix w with entries

$$w_{is} = \begin{cases} w_{is} & \text{if vertices } \{i, s\} \in L \\ 0 & \text{otherwise} \end{cases}.$$

The matrices B and C need to be constructed accordingly.

2.2 Random effects in the undirected graph model

Consider the following assumption on the probability space.

Assumption 1:

Let $G \in \mathcal{G}_N$ be a space of graphs of size N and $x \in \mathcal{X}_N$ the domain of covariates, $\sigma(\mathcal{G}_N, \mathcal{X}_N)$ a σ-algebra in the sample space $(\mathcal{G}_N, \mathcal{X}_N)$, and $\mathcal{P}_N$ a probability space on the measurable space on $(\mathcal{G}_N, \mathcal{X}_N), \sigma(\mathcal{G}_N, \mathcal{X}_N)$. Then $[(\mathcal{G}_N, \mathcal{X}_N), \sigma(\mathcal{G}_N, \mathcal{X}_N), \mathcal{P}_N]$ form a probability space.

Assumption 2:

Let ν, μ and δ be mutually independent random vectors of size N, M and T, respectively.

Correct mean specification: $\forall_{i,j,t}$ $\mathrm{E}(\nu_i \mid x, G) = \mathrm{E}(\mu_{ij} \mid x, G) = \mathrm{E}(\delta_{ijt} \mid x, G) = 0$.

Variance: $\forall_{i,j,t}$ $\mathrm{Var}(\nu_i \mid x, G) = \sigma^2_\nu$, $\mathrm{Var}(\mu_{ij} \mid x, G) = \sigma^2_\mu$, $\mathrm{Var}(\delta_{ijt} \mid x, G) = \sigma^2_\delta$.

Consider the error components regression model for an unweighted undirected graph network structure,

$$y_i = x_i\beta + \varepsilon_{ijk}, \tag{1}$$

$$\varepsilon_i := \mathrm{E}[y_i - E(y_i \mid x, G)] = \nu_i + \sum_{j=1}^{M} b_{ij}\mu_j + \sum_{k=1}^{T} c_{ik}\delta_k,$$

$$i = 1, 2, ..., N.$$

The error components can also be written as

$$\varepsilon_i = \nu_i + \sum_{i=1}^{N}\sum_{s>i}^{N} a_{is}\mu_{(is)} + \sum_{i=1}^{N}\sum_{s>i}^{N}\sum_{r>s}^{N} a_{is}a_{sr}a_{ir}\delta_{(isr)},$$

where $\mu_{(is)}$ and $\delta_{(isr)}$ correspond to the common edge and triangle effects, respectively.

In matrix notation the model above can be written as $y = x\beta + \varepsilon$, where y and ε are $N \times 1$ vectors, x is $N \times K$ matrix, and β is a $K \times 1$ vector. Then consider

$$\varepsilon = \nu + B\mu + C\delta,$$

and

$$\Omega := \mathrm{E}[\varepsilon\varepsilon^\top \mid x, G] = E[\nu\nu^\top + B\mu\mu^\top B^\top + C\delta\delta^\top C^\top \mid x, G] \quad (2)$$
$$= \sigma_\nu^2 I_N + \sigma_\mu^2 BB^\top + \sigma_\delta^2 CC^\top,$$

where ν is a $N \times 1$ random vector, μ is a $M \times 1$ random vector, δ is a $T \times 1$ random vector.

Note that this model allows for the covariates x to be dependent on the network structure. Thus for instance, vertice-specific features such as network centrality (degree, betweeness, clustering, etc.) may be covariates of the model.

Consider the OLS estimator $\hat{\beta} = (x^\top x)^{-1} x^\top y$, and consider the goal of estimating $Var[\hat{\beta} \mid x, G]$. Given the assumptions of the model, then consider Then

$$Var[\hat{\beta} \mid x, G] = (x^\top x)^{-1}(x^\top \Omega x)(x^\top x)^{-1}.$$

Note that Ω acts as a selector and weighting matrix, which selects which row and columns of x should be considered and weights them accordingly.

In the case with no network effects, defined as the vertex-only model,

$$\Omega_v = \sigma_\nu^2 I_N,$$

and thus only the xs that correspond to the same vertices i are considered. Thus

$$x^\top \Omega_v x = \sigma_\nu^2 \sum_{i=1}^{N} x_i x_i^\top.$$

The random-effects vertice&edge-only incidence model would have

$$\Omega_{ve} = \sigma_\nu^2 I_N + \sigma_\mu^2 BB^\top.$$

Thus

$$x^\top \Omega_{ve} x = \sum_{i=1}^{N} (\sigma_\nu^2 + d_i \sigma_\mu^2) x_i x_i^\top + 2\sigma_\mu^2 \sum_{i=1}^{N-1} \sum_{s>i}^{N} a_{is} x_i x_s^\top.$$
$$= \sum_{i=1}^{N} (\sigma_\nu^2 + d_i \sigma_\mu^2) x_i x_i^\top + 2\sigma_\mu^2 \sum_{i=1}^{N-1} \sum_{s>i}^{N} (\sum_{j=1}^{M} b_{ij} b_{sj}) x_i x_s^\top.$$

Two things are important to notice from this variance-covariance. First, note that the model implies an heteroskedastic structure, where the diagonal elements are proportional to the degree d_i of each vertex. Second, the off-diagonal elements that have a role are those of vertices that have a common link, which in this case have a maximum of one.

The random-effects vertice&triangle-only incidence model would have

$$\Omega_{vt} = \sigma_\nu^2 I_N + \sigma_\delta^2 C C^\top.$$

Thus

$$x^\top \Omega_{vt} x = \sum_{i=1}^{N} (\sigma_\nu^2 + t_i \sigma_\delta^2) x_i x_i^\top + 2\sigma_\delta^2 \sum_{i=1}^{N-2} \sum_{s>i}^{N-1} \sum_{r>s=1}^{N} a_{ir} a_{sr} a_{is} x_i x_s^\top$$
$$= \sum_{i=1}^{N} (\sigma_\nu^2 + t_i \sigma_\delta^2) x_i x_i^\top + 2\sigma_\delta^2 \sum_{i=1}^{N-1} \sum_{s>i}^{N} (\sum_{k=1}^{T} c_{ik} c_{sk}) x_i x_s^\top.$$

In the same way as the vertice&edge-only model the model has an heteroskedastic structure that depends on the number of triangles each vertice belongs to. Moreover the off diagonal elements are proportional to the number of triangles each edge belongs to (maximum $N-2$).

Joining both models gives

$$x^\top \Omega_{vet} x = \sum_{i=1}^{N} (\sigma_\nu^2 + d_i \sigma_\mu^2 + t_i \sigma_\delta^2) x_i x_i^\top + 2 \sum_{i=1}^{N-1} \sum_{s>i}^{N} \left[\sigma_\mu^2 \sum_{j=1}^{M} b_{ij} b_{sj} + \sigma_\delta^2 \sum_{k=1}^{T} c_{ik} c_{sk} \right] x_i x_s^\top.$$

2.3 Extension to weighted networks

The results above can be easily extended to weighted networks where A is replaced by W, and the B and C matrices are also constructed using the weighted components. Note that for weighted networks the potential misspecification problems in estimating the variance-covariance components are likely to be more severe if $w_{is} \propto x_i x_s'$.

3 Consistent variance components estimators

Here we consider simple consistent estimators of the variance components using ANOVA-type decompositions.

Consider the following statistics:

$$S_1 = \frac{1}{N}\sum_{i=1}^{N} u_i^2,$$

$$S_2 = \frac{1}{M}\sum_{i=1}^{N-1}\sum_{s>i}^{N} a_{is}u_i u_s,$$

$$S_3 = \frac{1}{T}\sum_{i=1}^{N-2}\sum_{s>i}^{N-1}\sum_{r>s}^{N} a_{is}a_{sr}a_{ir}u_i u_s.$$

S_1 contains the usual sum of squared errors. Note that for each vertex there will be at most $N-1$ edges to which it belongs and $N-2$ triangles. Moreover, each edge will be repeated twice for undirected graphs, one for each vertex, and each triangle will be repeated three times, one for each vertex. Then,

$$E[S_1 \mid x, G] = \sigma_\nu^2 + \sigma_\mu^2\frac{2M}{N} + \sigma_\delta^2\frac{3T}{N}.$$

$E[S_1 \mid x, G]$ is the (conditional) variance of a vertex.

S_2 contains the cross products of the error terms, which corresponds to the number of edges M. This corresponds to the existing active links (i.e.,

$a_{is} = 1$, $s > i$). For each active link, there could be at most $N - 2$ triangles that can be formed from it. Now, each triangle will be repeated three times for each link. That is, for S_2 if we have an edge, say (i, s), that belongs to a triangle, say (i, s, r), such that $r > s > i$, then the triangle effect $\delta_{(i,s,r)}$ will appear in the edges (i, s), (s, r), and (i, s). Thus, each triangle will be found 3 times for every edge. Thus,

$$E[S_2 \mid x, G] = \sigma_\mu^2 + \sigma_\delta^2 \frac{3T}{M}.$$

$E[S_2 \mid x, G]$ is the (conditional) covariance of two vertices that have a common edge.

Finally, S_3 computes the cross products for active triangles (i.e., $a_{is} = a_{sr} = a_{ir} = 1$, $r > s > i$). Note that for S_3 if we have a triangle, say (i, s, r), then two nodes, say i and s, must share both $\mu_{(i,s)}$ and $\delta_{(i,s,r)}$. Then,

$$E[S_3 \mid x, G] = \sigma_\mu^2 + \sigma_\delta^2.$$

$E[S_3 \mid x, G]$ is the (conditional) covariance of two vertices that have common edge and triangle(s).

In the absence of triangle effects, i.e., $\sigma_\delta^2 = 0$, the model simplifies to

$$\sigma_\nu^2 = E[S_1 \mid x, G] - E[S_2 \mid x, G]\frac{2M}{N},$$
$$\sigma_\mu^2 = E[S_2 \mid x, G],$$

such that the nonnegativity restrictions are $E[S_2 \mid x, G] \geq 0$ and $\frac{E[S_1|x,G]}{E[S_2|x,G]} \geq \frac{2M}{N}$, such that the ratio of the variance of a vertex to the covariance of two random vertices needs to be bigger than the average number of edges per vertex. First, take for instance a cycle graph, a 2-regular graph with all vertices of degree 2 such that $M = N$. For this case the variance of the vertices need to be at least twice the covariance. Second, consider a complete graph with $M = N(N-1)/2$. In this case, the ratio of variance to covariance needs to grow faster than the number of vertices.

In the absence of edge effects, $\sigma_\mu^2 = 0$, the model simplifies to

$$\sigma_\nu^2 = E[S_1 \mid x, G] - E[S_3 \mid x, G]\frac{3T}{N},$$
$$\sigma_\delta^2 = E[S_3 \mid x, G],$$

such that the nonnegativity restrictions are $E[S_3 \mid x, G] \geq 0$ and $\frac{E[S_1|x,G]}{E[S_3|x,G]} \geq \frac{3T}{N}$, such that the ratio of the variance of a vertex to the covariance of two random vertices needs to be bigger than the average number of triangles per vertex.

For the edge and triangle effects model, solving for $(\sigma_\nu^2, \sigma_\mu^2, \sigma_\delta^2)$ gets

$$\sigma_\nu^2 = E[S_1 - \frac{E[S_2 \mid x, G] - E[S_3 \mid x, G]\frac{3T}{M}}{1 - \frac{3T}{M}}\frac{2M}{N} - \frac{E[S_3] - E[S_2]}{1 - \frac{3T}{M}}\frac{3T}{N},$$
$$\sigma_\mu^2 = \frac{E[S_2 \mid x, G] - E[S_3 \mid x, G]\frac{3T}{M}}{1 - \frac{3T}{M}},$$
$$\sigma_\delta^2 = \frac{E[S_3 \mid x, G] - E[S_2 \mid x, G]}{1 - \frac{3T}{M}}.$$

For this case the nonnegativity restrictions imply: (i) $\frac{E[S_3|x,G]}{E[S_2|x,G]} \geq 1$, (ii) $\frac{E[S_2|x,G]}{E[S_3|x,G]} \geq \frac{3T}{M}$, (iii) $\frac{E[S_1|x,G]}{E[S_2|x,G]} \geq \frac{2M}{N}$, and (iv) $\frac{E[S_1]}{E[S_3]} \geq \frac{3T}{MN}$. Restriction (i) implies that the covariance among vertices that belong to a triangle must be larger than the covariance of vertices that share a link. Restriction (ii) states that the ratio $\frac{E[S_3|x,G]}{E[S_2|x,G]}$ cannot exceed the average number of triangles per edge. Restriction (iii) correspond to the number average number of links per vertex. Restriction (iv) is a combination of the above with no clear interpretation.

The consistent estimators are then constructed by defining $\hat{S}_1$, $\hat{S}_2$, and $\hat{S}_3$, where the OLS residuals $\hat{u}$ are used, and the nonnegativity constraints are imposed.

4 Specification tests for undirected unweighted graphs

The log likelihood function for this problem is given by

$$\mathrm{L}(\beta,\theta) \;\propto\; -\frac{1}{2}\ln|\Omega| - \frac{1}{2}\varepsilon^\top\Omega^{-1}\varepsilon,$$

with $\theta = (\sigma^2_\nu, \sigma^2_\mu, \sigma^2_\delta)$, $\varepsilon = y - x\beta$, and Ω is given by equation (2). In this model we have that the Fisher information matrix is block diagonal in terms of β and θ. This feature also applies to non-Gaussian error components, where in fact OLS estimators for β are consistent. In turn, this simplifies the subsequent algebra where we only consider θ for constructing our LM tests.

Let $\theta \in \Theta \subseteq \mathbb{R}^p$, where p is the dimension of θ. Using the formulas in Harville ([23], p.326) (see also [3]) the score functions can be expressed as

$$s_r(\theta) = \partial L/\partial\theta_r = -\frac{1}{2}\mathrm{tr}(\Omega^{-1}\partial\Omega/\partial\theta_r) + \frac{1}{2}\{u^\top\Omega^{-1}(\partial\Omega/\partial\theta_r)\Omega^{-1}u\},$$

for $1 \leq r \leq p$. The information matrix $\mathcal{J}$ can be obtained for for $1 \leq r, k \leq p$. as

$$\begin{aligned}\partial^2 L/\partial\theta_r\partial\theta_k \;&=\; \frac{1}{2}\mathrm{tr}\left(\Omega^{-1}\left\{\frac{\partial^2\Omega}{\partial\theta_r\partial\theta_k} - \frac{\partial\Omega}{\partial\theta_r}\Omega^{-1}\frac{\partial\Omega}{\partial\theta_k}\right\}\right)\\ &\quad+\frac{1}{2}u^\top\Omega^{-1}\left(\frac{\partial\Omega}{\partial\theta_r\partial\theta_k} - 2\frac{\partial\Omega}{\partial\theta_r}\Omega^{-1}\frac{\partial\Omega}{\partial\theta_r}\right)\Omega^{-1}u,\end{aligned}$$

and

$$\mathcal{J}_{rk}(\theta) \equiv -\mathrm{E}(\partial^2 L/\partial\theta_r\partial\theta_k) = \frac{1}{2}tr\left(\Omega^{-1}\frac{\partial\Omega}{\partial\theta_r}\Omega^{-1}\frac{\partial\Omega}{\partial\theta_k}\right).$$

Note that

$$\partial\Omega/\partial\sigma^2_\nu = I_N, \tag{3}$$

$$\partial\Omega/\partial\sigma_\mu^2 = BB^\top, \tag{4}$$

$$\partial\Omega/\partial\sigma_\delta^2 = CC^\top. \tag{5}$$

In order to construct LM tests, first note that the block diagonality between β and θ allow us to focus on the scores corresponding to θ only. Second, consistent estimators of θ under the null can be obtained using an ANOVA-type analysis as in Section 3. Hence our tests will be based on Neyman's $C(\alpha)$ principle, which produces tests that are asymptotically equivalent to likelihood based LM tests under $\sqrt{N}$-consistent non-maximum likelihood estimation of the nuisance parameters. See Bera and Bilias [10] for a dicussion.

Consider a partition of $\theta = (\theta_1^\top, \theta_2^\top)^\top$, where θ_2 contains the parameters under the corresponding null hypothesis $H_0^2 : \theta_2 = 0$, and θ_1 the nuisance parameters that need to be estimated. In our particular case, θ will be partitioned into either $\theta_1 = \sigma_\nu^2, \theta_2 = \sigma_\mu^2$ when we want to test for the presence of edge network effects assuming $\sigma_\delta^2 = 0$, $\theta_1 = \sigma_\nu^2, \theta_2 = \sigma_\delta^2$ when we want to test for the presence of edge and triangle network effects assuming $\sigma_\mu^2 = 0$, $\theta_1 = \sigma_\nu^2, \theta_2 = (\sigma_\mu^2, \sigma_\delta^2)$ when we want to test for the presence jointly of edge and triangle network effects, $\theta_1 = (\sigma_\nu^2, \sigma_\mu^2), \theta_2 = \sigma_\delta^2$ when we want to test for the presence pf triangle effects assuming edge effects or $\theta_1 = (\sigma_\nu^2, \sigma_\delta^2), \theta_2 = \sigma_\mu^2$ when we want to test for the presence pf triangle effects assuming edge effects. Correspondingly, the score will be partitioned as $s(\theta) = (s_1(\theta)^\top, s_2(\theta)^\top)^\top$, and the information matrix as $\mathcal{J}(\theta) = \begin{pmatrix} \mathcal{J}_{11}(\theta) & \mathcal{J}_{12}(\theta) \\ \mathcal{J}_{21}(\theta) & \mathcal{J}_{22}(\theta) \end{pmatrix}$.

Conditional LM statistics for H_0^2 under maximum likelihood estimation are defined as

$$LM_2(\theta) = s_2(\theta)^\top\{\mathcal{J}_{22}(\theta) - \mathcal{J}_{21}(\theta)\mathcal{J}_{11}^{-1}(\theta)\mathcal{J}_{12}(\theta)\}^{-1}s_2(\theta).$$

Neyman's C(α) adjusted scores are defined as

$$s_{2\cdot1}(\theta) \equiv s_2(\theta) - \mathcal{J}_{21}(\theta)\mathcal{J}_{11}^{-1}(\theta)\mathcal{J}_{12}(\theta)s_1(\theta).$$

Then, the Neyman's C(α) LM statistic is

$$LM_{2\cdot 1}(\theta) = s_{2\cdot 1}(\theta)^\top \{\mathcal{J}_{22}(\theta) - \mathcal{J}_{21}(\theta)\mathcal{J}_{11}^{-1}(\theta)\mathcal{J}_{12}(\theta)\}^{-1} s_{2\cdot 1}(\theta).$$

A well known result is that $LM_{2\cdot 1}(\hat{\theta}) \xrightarrow{d} \chi^2_{dim(\theta_2)}$, where $\hat{\theta}$ is a $\sqrt{N}$-consistent estimator under the corresponding null hypothesis. Note that when we estimate the parameters under the joint null $\sigma_n u^2 = \sigma_\delta^2 = 0$, the ML estimators of β and σ_ν^2 coincide with the least-squares estimators.

Consider now Bera and Yoon [15] locally size-robust type statistics. For this, consider a new partition of $\theta = (\theta_1, \theta_2, \theta_3)' = (\theta_1, \theta_{23})'$ where we want to test for the null hypothesis H_0^2, we consider θ_1 as nuisance parameters to be estimated, but the validity of the test is affected by the validity of $H_0^3 : \theta_3 = 0$. Global valid tests for H_0^2 would require consistent estimators of θ_3 as in the construction of the conditional LM statistics above. In practice, however, estimators of θ_3 may be cumbersome or it might suffer identification conditions under the null. Thus, Bera and Yoon [15] has been successfully implemented to test one particular null without estimating the other nuisance parameter θ_3. This procedure is valid under $\sqrt{N}$-*local* deviations of H_0^3, but different empirical studies confirmed its validity for non-local deviations too. In our particular case, the parameter will be partitioned as $\theta_1 = \sigma_\nu^2, \theta_2 = \sigma_\mu^2, \theta_3 = \sigma_\delta^2$. This procedure thus allows us to test for triangle effects but without estimating edge effects variance, even when we are estimating under the joint null hypothesis $H_0^2 \& H_0^3 : \sigma_\mu^2 = \sigma_\delta^2 = 0$, which is just least-squares estimation. The statistic is constructed as in Bera, Montes-Rojas and Sosa-Escudero [12, 13] for non-maximum likelihood estimation as

$$LM_{2(3)\cdot 1}(\theta) = s_{2(3)\cdot 1}(\theta)'[\mathcal{J}_{2(3)\cdot 1}(\theta)]^{-1} s_{2(3)\cdot 1}(\theta),$$

where

$$s_{2(3)\cdot 1}(\theta) = s_{2\cdot 1}(\theta) - \mathcal{J}_{23\cdot 1}(\theta)\mathcal{J}_{33\cdot 1}^{-1}(\theta) s_{3\cdot 1}(\theta),$$

$$\mathcal{J}_{2(3)\cdot 1}(\theta) = \mathcal{J}_{22\cdot 1}(\theta) - \mathcal{J}_{23\cdot 1}(\theta)\mathcal{J}_{33\cdot 1}^{-1}(\theta)\mathcal{J}_{32\cdot 1}(\theta),$$

$$\mathcal{J}_{22\cdot 1}(\theta) = \mathcal{J}_{22}(\theta) - \mathcal{J}_{21}(\theta)\mathcal{J}_{11}^{-1}(\theta)\mathcal{J}_{12}(\theta),$$

$$\mathcal{J}_{33\cdot 1}(\theta) = \mathcal{J}_{33}(\theta) - \mathcal{J}_{31}(\theta)\mathcal{J}_{11}^{-1}(\theta)\mathcal{J}_{13}(\theta),$$

$$\mathcal{J}_{23\cdot 1}(\theta) = \mathcal{J}_{23}(\theta) - \mathcal{J}_{23,1}(\theta)\mathcal{J}_{11}^{-1}(\theta)\mathcal{J}_{1,23}(\theta).$$

Then $LM_{2(3)\cdot 1}(\hat{\theta}) \xrightarrow{d} \chi^2_{dim(\theta_2)}$ for $\hat{\theta}$ being a consistent estimator under the joint null hypothesis $H_0^2 \& H_0^3 : \sigma_\mu^2 = \sigma_\delta^2 = 0$ and for $\theta_3 = \sigma_\delta^2 = o(1/\sqrt{N})$.

In sum, the LM tests considered are:

- LM_μ: LM test for $H_0 : \sigma_\mu^2 = 0$ when σ_ν^2 is estimated as MSE after OLS estimation and $\sigma_\delta^2 = 0$ is assumed.

- LM_δ: LM test for $H_0 : \sigma_\delta^2 = 0$ when σ_ν^2 is estimated as MSE after OLS estimation and $\sigma_\mu^2 = 0$ is assumed.

- $LM_{\mu,\delta}$: LM test for $H_0 : \sigma_\mu^2 = \sigma_\delta^2 = 0$ when σ_ν^2 is estimated as MSE after OLS estimation.

- $LM_{\mu(\delta)}$: BY test for $H_0 : \sigma_\mu^2 = 0$ when σ_ν^2 is estimated as MSE after OLS estimation and $\sigma_\delta^2 = 0$ is allowed to have local deviations.

- $LM_{\delta(\mu)}$: BY test for $H_0 : \sigma_\delta^2 0$ when σ_ν^2 is estimated as MSE after OLS estimation and $\sigma_\mu^2 = 0$ is allowed to have local deviations.

- $LM_{\delta-\mu}$: LM test for $H_0 : \sigma_\delta^2 = 0$ when $(\sigma_\nu^2, \sigma_\mu^2)$ is estimated as in Section 3 after OLS estimation.

5 Monte Carlo experiments

This section explores the small sample performance of the proposed tests through a Monte Carlo experiment. We will consider the following simple regression model:

$$y_i = x_i\beta + \varepsilon_i,$$
$$\varepsilon_i = \nu_i + \sum_{i=1}^{N}\sum_{s>i}^{N} a_{is}\mu_{(is)} + \sum_{i=1}^{N}\sum_{s>i}^{N}\sum_{r>s}^{N} a_{is}a_{sr}a_{ir}\delta_{(isr)},$$
$$i = 1, 2, ..., N,$$

where $A = \{a_{ir}\}$ is an adjacent contiguity matrix. We assume $x_i \sim \ iid\ N(0, 1)$, $\beta = 1$, $\nu_i \sim \ iid\ N(0, 10)$, $\mu_{(is)} \sim \ iid\ N(0, \sigma^2_\mu)$ and $\delta_{(isr)} \sim \ iid\ N(0, \sigma^2_\delta)$.

We simulate two types of networks. First, we consider an Erdös-Rényi random graph where links are randomly generated with a given probability p_N, i.e., $Prob(a_{ir} = 1) = p_N$, $i, r = 1, \ldots, N, i \neq r$. For the Erdös-Rényi graphs we have on average a constant proportion of vertices and edges, N/M, using $p_{100} = 0.05, p_{225} = 0.05 \times 100/225, p_{400} = 0.05 \times 100/400$. In this case, the number of triangles per node is also constant on average. Second, a queen-type spatial structure where edges are generated according to queen contiguity, i.e., for a squared board with number of rows and columns $n = \sqrt{N}$, for $i = 1, \ldots, N$, $a_{ir} = 1$ if $r \in \{i-1, i+1, i-n-1, i-n, i-n+1, i+n-1, i+n, i+n+1\}$ with $1 \leq r \leq N$, and $a_{ir} = 0$ otherwise. Note that the considered spatial-type model has a similar number of triangles and edges for each node, i.e. 8 edges and triangles for a node that is not on the border of the board. We consider $N \in \{100, 225, 400\}$.

First, we consider the empirical size results where $\sigma^2_\mu = \sigma^2_\delta = 0$ in Table 6. In all cases, marginal, joint and robust tests have the appropriate size, for all levels of significance.

Second, we consider the empirical power and robustness for $(\sigma^2_\mu, \sigma^2_\delta) \in \{0, 1, \ldots, 10\}^2$ in figures 1 and 2.

The former figure report the tests for detecting edge heterogeneity, $\sigma_\mu > 0$. Note that the marginal tests LM_μ has the largest power performance for changes in σ_μ (figures 1-(a) and 1-(c)), followed by the joint tests $LM_{\mu\delta}$. However the marginal test also rejects in the direction of $\sigma_\delta > 0$, as Figures 1-(b) and 1-(d) show, that is, it is not robust to the presence of triangle effects.

The BY robust test have good power performance in figure 1-(a), close to the joint test, but it has low power in the Queen spatial more complex network model, as shown in figure 1-(c). In fact, the BY is robust to deviations in $\sigma_\delta > 0$, as seen in figures 1-(b) and 1-(d).

Tests for triangle effects have a similar performance to those of edge effects. As in the previous paragraph, the tests have the expected rejection rates in the direction of $\sigma_\delta > 0$, and the BY robust test have correct size for $\sigma_\mu > 0$. Note that the conditional test $LM_{\delta-\mu}$ estimates σ_μ, and as such it should be robust to misspecification in edge effects. For this case the BY robust tests outperforms it in terms of size and power in the Erdös-Rényi random graph model, and it is very close to the conditional tests in the Queen spatial structure.

6 Conclusion

This paper develops a simple model of network random effects that can be used to estimate the variance-covariance matrix in a linear OLS set up. It focuses on evaluating the appropriate level of effects, using the example of links and triangles effects as random components.

References

[1] J. Angrist, J.-S. Pischke, Mostly harmless econometrics: An empiricist's companion, Princeton University Press, Princeton 2009.

[2] T. Amemiya, The estimation of the variances in a variance-components mode, International Economic Review 12 (1971) 1–13.

[3] B.H. Baltagi, Econometric Analysis of Panel Data, Wiley, NJ, 2013.

[4] B.H. Baltagi, Q. Li, A transformation that will circumvent the problem of autocorrelation in an error-component model, J. Econometrics 48 (1991) 385–393.

[5] B.H. Baltagi, Q. Li, Testing AR(l) against MA(l) disturbances in an error component model, J. Econometrics 68 (1995) 133–151.

[6] B.H. Baltagi, S.H. Song, B.C. Jung, The unbalanced nested error component regression model, J. Econometrics 101 (2001) 357–381.

[7] B.H. Baltagi, S.H. Song, B.C. Jung, LM tests for the unbalanced nested panel data regression model with serially correlated errors, Annales d'économie et de statistique 65 (2002) 219–268.

[8] B.H. Baltagi, S.H. Song, B.C. Jung, B.C., Simple LM tests for the unbalanced nested error component regression model, Econometric Reviews 21 (2002) 167–187.

[9] B.H. Baltagi, P.X. Wu, Unequally spaced panel data regressions with AR(1) disturbances, Econometric Theory 15 (1999) 814–823.

[10] A. Bera, Y. Bilias, Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: An essay on historical developments and some new results, J. Statist. Plann. Inf. 97 (2001) 9–44.

[11] A. Bera, C. Jarque, Model specification tests: A simultaneous approach, J. Econometrics 20 (1982) 59–82.

[12] A. Bera, G. Montes-Rojas, W. Sosa-Escudero, General specification testing with locally misspecified models, Econom. Theory 26 (2010) 1838–1845.

[13] A. Bera, G. Montes-Rojas, W. Sosa-Escudero, A new robust and most powerful test in the presence of local misspecification Communications in Statistics - Theory and Methods, 46:16 (2017), 8187–8198

[14] A. Bera, W. Sosa-Escudero, M. Yoon, Tests for the error component model in the presence of local misspecification, J. Econometrics 101 (2001) 1–23.

[15] A. Bera, M. Yoon, Specification testing with locally misspecified alternatives, Econom. Theory 9 (1993) 649–658.

[16] M. Bertrand, E. Duflo, S. Mullainathan, How much should we trust differences-in-differences estimates? Quarterly J. Economics 119 (2004) 249–275.

[17] T.S. Breusch, A.R. Pagan, The Lagrange multiplier test and its applications to model specification in econometrics, Review of Economic Studies 47 (1980) 239–253.

[18] C. Cameron, D.L. Miller, A practitioner's guide to cluster-robust inference, J. Human Resources 50 (2015) 317–372.

[19] A. de Paula, Econometrics of network models, CEMMAP Working Paper CWP52/15.

[20] M. Evans, Robustness of size of tests of autocorrelation and heteroscedasticity to nonnormality, J. Econometrics 51 (1992) 7–24.

[21] M. Furno, LM tests in the presence of non-normal error distributions, Econometric Theory 16 (2000) 249–261.

[22] E.A. Hanushek, L. Woessmann, The economics of international differences in educational achievement, In: E.A. Hanushek, S. Machin, L. Woessmann (Eds.), Handbook of the Economics of Education 3, North Holland, Amsterdam, 2011, pp. 89–200.

[23] D.A. Harville, Maximum likelihood approaches to variance component estimation and to related problems, J. Amer. Statist. Assoc. 72 (1977) 320–340.

[24] Peter D. Hoff, Bilinear mixed-effects models for dyadic data J. Amer. Statist. Assoc. 100 (2005) 286–295.

[25] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock Latent Space Approaches to Social Network Analysis J. Amer. Statist. Assoc. 97 (2002) 1090–1098.

[26] Y. Honda, Testing the error components model with non-normal disturbances, Rev. Economic Studies 52 (1985) 681–690.

[27] A. Inoue, G. Solon, A Portmanteau test for serially correlated errors in fixed effects models, Econometric Theory 22 (2006) 835–851.

[28] M. Jackson, Social and Economic Networks, Princeton University Press (2009).

[29] G. Kézdi, Robust standard error estimation in fixed-effects panel models, Hungarian Statistical Review Special Number 9 (2004) 95–116.

[30] G. King, M.E. Roberts, How robust standard errors expose methodological problems they do not fix, and what to do about it, Political Analysis 23 (2015) 159–179.

[31] E. Kolaczyk, Statistical Analysis of Network Data, Springer-Verlag (2009).

[32] L.F. Lee, Identification and estimation of econometric models with group interactions, contextual factors and fixed effects, Journal of Econometrics 140 (207), 333–374.

[33] L.F. Lee, X. Liu, X. Lin, Specification and estimation of social interaction models with network structures, Econometrics Journal 13 (2010) 145–176.

[34] L.F. Lee, J. Yu, Estimation of spatial autoregressive panel data models with fixed effects, Journal of Econometrics 154 (2010) 165–185.

[35] X. Lin, L.F. Lee, GMM estimation of spatial autoregressive models with unknown heteroskedasticity, Journal of Econometrics 157 (2010) 34–52.

[36] C.F. Manski, Identification of endogenous social effets: The reflection problem, The Review of Economic Studies 60 (1993) 531–542.

[37] L. Matyas, Error components models, In: L. Matyas, P. Sevestre (Eds.), The Econometrics of Panel Data, 2nd edition, Kluwer Academic Publishers, Dordrecht, 2012, pp. 50–76.

[38] G. Montes-Rojas, An equicorrelation Moulton factor in the presence of arbitrary intra-cluster correlation, Economics Letters 45 (2016) 221–224.

[39] G. Montes-Rojas, W. Sosa-Escudero, Robust tests for heteroskedasticity in the one-way error components model, J. Econometrics 160 (2011) 300–310.

[40] B.R. Moulton, Random group effects and the precision of regression estimates, J. Econometrics 32 (1986) 385–397.

[41] B.R. Moulton, Diagnostics for group effects in regression analysis, J. Bus. Econom. Statist. 5 (1987) 275–282.

[42] B.R. Moulton, An illustration of a pitfall in estimating the effects of aggregate variables in micro units, Review of Economics and Statistics 72 (1990) 334–338.

[43] T.D. Wallace, A. Hussain, The use of error components models in combining cross section with time series data, Econometrica 37 (1969) 55–72.

[44] J.M. Wooldridge, Econometric analysis of cross section and panel data, 2nd edition, MIT Press, Boston, MA, 2010.

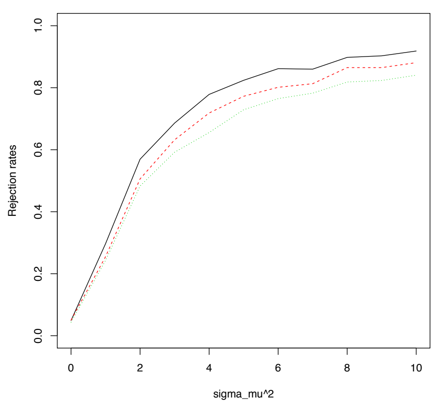
Table 1: Empirical size

N	LM_{μ}	LM_{δ}	$LM_{\mu,\delta}$	$LM_{\mu(\delta)}$	$LM_{\delta(\mu)}$	$LM_{\delta-\mu}$
Erdös-Rényi random graph						
Size 1%						
100	0.009	0.016	0.0145	0.009	0.0165	0.0115
225	0.012	0.0115	0.015	0.013	0.012	0.009
400	0.013	0.012	0.0085	0.0095	0.0075	0.007
Size 5%						
100	0.043	0.05	0.0465	0.042	0.052	0.041
225	0.052	0.0485	0.0495	0.052	0.0495	0.041
400	0.047	0.0475	0.049	0.046	0.046	0.0435
Size 10%						
100	0.082	0.0885	0.0855	0.089	0.092	0.0765
225	0.1045	0.092	0.102	0.098	0.0995	0.0875
400	0.089	0.087	0.093	0.0965	0.099	0.0915
Spatial queen structure						
Size 1%						
100	0.0115	0.0105	0.0105	0.01	0.011	0.0115
225	0.0075	0.0065	0.012	0.0145	0.0135	0.014
400	0.0085	0.0085	0.0095	0.012	0.011	0.011
Size 5%						
100	0.0475	0.0515	0.047	0.048	0.044	0.046
225	0.045	0.039	0.0565	0.0595	0.052	0.0525
400	0.046	0.0465	0.049	0.0535	0.049	0.0505
Size 10%						
100	0.0965	0.0975	0.0955	0.094	0.09	0.097
225	0.0965	0.09	0.1	0.1085	0.1115	0.1125
400	0.0935	0.0965	0.098	0.096	0.0995	0.1015

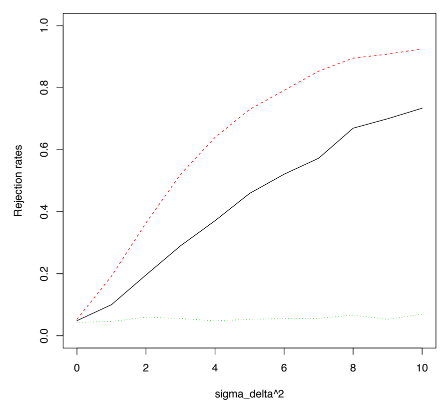
Figure 1: LM tests for σ_μ^2

Erdös-Rényi random graph

(a)

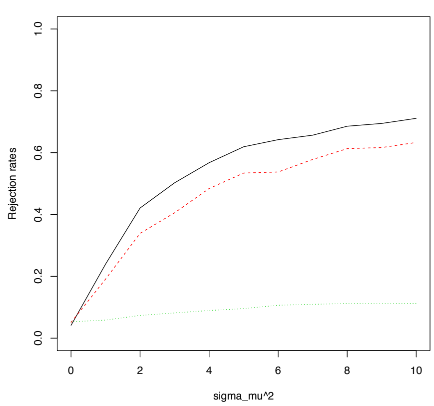


(b)

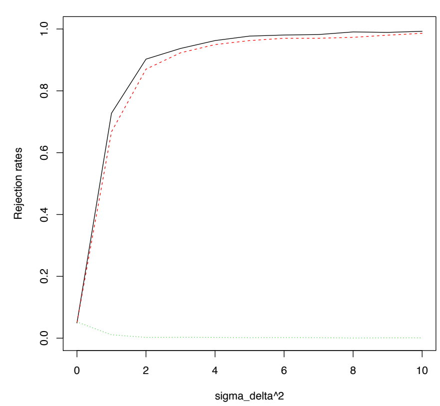


Queen spatial structure

(c)



(d)

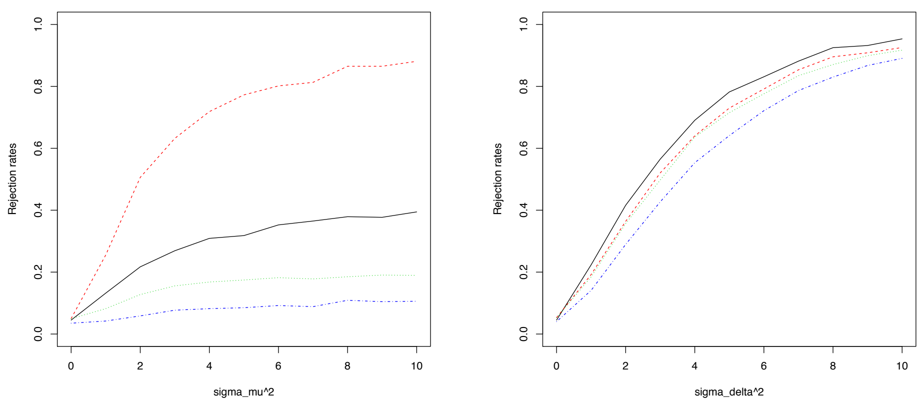


Notes: Monte carlo experiments based on 2000 replications. Solid black line: LM_μ. Dashed red line: $LM_{\mu\delta}$. Dotted green line: $LM_{\mu(\delta)}$.

Figure 2: LM tests for σ_δ^2

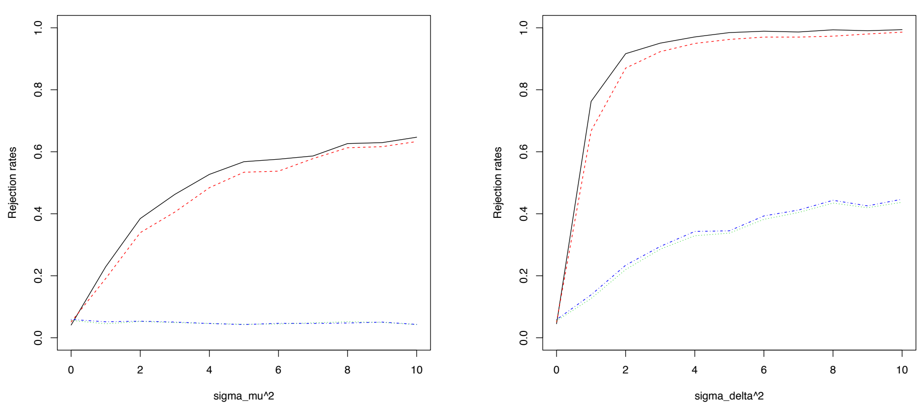
Erdös-Rényi random graph

(a) (b)



Queen spatial structure

(c) (d)



Notes: Monte carlo experiments based on 2000 replications. Solid black line: LM_δ. Dashed red line: $LM_{\mu\delta}$. Dotted green line: $LM_{\delta(\mu)}$. Dash-dot blue line: $LM_{\delta-\mu}$.