CEIS Centre for Economic and International Studies
TOR VERGATA

CEIS Tor Vergata

RESEARCH PAPER SERIES

Vol. 14, Issue 3, No. 369 – February 2016

The W Matrix in Network and Spatial Econometrics: Issues Relating to Specification and Estimation

Luisa Corrado and Bernard Fingleton

The **W** Matrix in Network and Spatial Econometrics: Issues Relating to Specification and Estimation

Luisa Corrado* and Bernard Fingleton†

This Version: September 2015

Abstract

Network and spatial econometric models commonly embody a so-called W matrix which defines the connectivity between nodes of a network. The reason for the existence of W is that it facilitates parsimonious parametrization of inter-nodal interaction which would otherwise be very difficult to achieve from a practical modelling perspective. The problem considered in this paper is the effect of misspecifying W. The paper demonstrates the effect in the context of two types of model, the dynamic spatial autoregressive panel model and the multilevel spatial autoregressive panel model, both of which include W as part of the model specification and use W in estimation. Monte Carlo results are presented showing the impact on bias and RMSE of misspecification of W. The paper highlights the need for careful attention to the correct structure of W in spatial econometric and network modelling.

Key-Words: Networks, Multilevel Modelling, Fixed Effects, Dynamic Spatial Autoregressive Panel Model, Multilevel Spatial Autoregressive Panel Model.

*Associate Professor, University of Rome Tor Vergata, DEDI Department, Italy and University of Cambridge, Faculty of Economics. E-mail: lc242@econ.cam.ac.uk. Luisa Corrado acknowledges the Marie-Curie ERG EWEBE (276682).

†Director of Research, University of Cambridge Department of Land Economy, bf100@cam.ac.uk.

1 Introduction

The W matrix is the term commonly used in network and spatial econometrics to describe a matrix defining the connectivity between N entities that are located in two-dimensional space. As described in Corrado and Fingleton (2012), one can envisage the N entities to be nodes of a network and the quantitative values or weights assigned to the cells of W to be indicators of the existence of, and perhaps importance of, a link between each pair of nodes. Typically W, denoted by $\mathbf{W}_N$ below, is of dimension N by N and in many spatial econometric applications, the N entities will be places on the surface of the earth, for example cities, regions or countries, and connectivity between these localities will be some function of 'distance' between them. Its role is to provide a parsimonious parametrization for interdependence between observations on a variable. As explained by LeSage and Pace (2009), once we allow for dependence relations between a set of N entities on a single variable, for example as represented by the N by 1 vector Y, there are potentially $N^2 - N$ parameters that define individual interdependence, such as the relation between y_i and y_j, having excluded dependence of an observation on itself. This leads to an over-parametrization problem, which can be solved by imposing an a priori structure, or weights matrix W, on the interdependence relations, thus reducing the number of parameters to be estimated from $N^2 - N$ to one. The quantity in cell (i, j) of $\mathbf{W}_N$, w_{ij}, may be a function of great circle distance, a binary indicator of distance based on whether i and j are contiguous or not, or a much more complex measure of connectivity. Typically, in a spatial econometric model we capture the interdependence among the entities represented by the cells of vector $\mathbf{Y}$ by the N by 1 vector $\mathbf{W}_N\mathbf{Y}$ which is commonly referred to as the spatial lag. The $i'th$ element $\sum_{j=1}^{N} w_{ij}y_j$ of $\mathbf{W}_N\mathbf{Y}$ is a linear combination of 'neighbouring' values of $\mathbf{Y}$, weighted according to the $i'th$ row of matrix $\mathbf{W}_N$ so that $\mathbf{W}_N\mathbf{Y}$ is typically one of a number of explanatory variables in a spatial econometric model which is associated with a single parameter, say ρ. Hence, by imposing structure on interdependence via $\mathbf{W}_N$, only this single parameter, rather than $N^2 - N$, is assumed to be unknown. Given $\rho \neq 0$, the presence of $\mathbf{W}_N\mathbf{Y}$ in a spatial econometric model is a signal that mutual interdependence among the elements of $\mathbf{Y}$ is important and the presence of the spatial lag is required to avoid misspecification and biased parameter estimates.

Increasingly there is a growing awareness of the importance of connectivity in modelling which extends beyond its roots in spatial econometrics and economic geography. Nodes need not necessarily be places and there is a burgeoning wider literature that provides additional support to the concept of network interaction. Network interaction has become an important issue in economic research, for example, connections via networks enable the beneficial exchange of goods, information and services. The manifestations of connections are spillovers. Spillovers have diverse origins, and therefore one would anticipate that the way to model them takes various forms. For instance, they may be the outcome of network economics, commuting, migration, displaced demand and supply effects in the housing market, localized information flows through social networks, strategic interaction between policy makers, tax competition between local authorities, or even simply arbitrary boundaries causing spatial autocorrelation. Some of these networks are dynamic and endogenous, but modelling these is beyond the scope of the present paper. Here we restrict attention to the case in which W describes a fixed, exogenously determined interaction structure. Our illustrations take a subset of the many models to show some of the implications of models incorporating network interdependence.

While the cell values of W are assumed to be known, in reality often they are not known and a model is estimated using a W matrix that is inaccurate. This leads us to the principal problem under discussion in this paper, what are the implications for model outcomes of misspecifying the W matrix in the model structure? The paper focuses on two specific types of model embodying the W matrix, firstly a dynamic spatial panel model and secondly a multilevel model with spatial effects. In section 2 of the paper we outline the specification of the dynamic spatial panel model. Section 3 outlines the estimation method. Section 4 gives Monte Carlo simulation results in which we discuss the implications for estimation and interpretation of model estimates as W departs from

the true W used in the data generating process (DGP). Section 5 gives specification details of the second model we consider, which is a multilevel model with spatial dependence. Section 6 describes estimation techniques for this model. Section 7 focuses on interpretation of the multilevel model outcomes under W matrices that are different from that used in the data generating process of a Monte Carlo simulation. In section 8 we briefly discuss how the true W might be approximated and the problems this entails. Section 9 concludes.

2 The Dynamic Spatial Autoregressive Panel Model

The set-up for spatial panel models is the existence of separate time-series, one for each of N interacting entities (such as cities, regions, countries) with distinct locations. We refer to the N entities in the panel context generally as individuals. The model we discuss is dynamic, so that the individual i's observed value of the dependent variable at time t depends in part on the value at $t-1$. Thus in the model specification

$$\mathbf{y}_t = \gamma \mathbf{y}_{t-1} + \rho_1 \mathbf{W}_N \mathbf{y}_t + \mathbf{x}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \tag{1}$$

In which $\mathbf{y}_t = (y_{1t}, ..., y_{Nt})$ denotes the N by 1 vector of observations at time t of a regressand, $\mathbf{x}_t$ is an N by k matrix of regressors and $\boldsymbol{\varepsilon}_t$ is the N by 1 vector of errors. Also $\mathbf{W}_N$ is the N by N connectivity matrix, and $\mathbf{W}_N \mathbf{y}_t$ is an N by 1 vector thus giving a spatial autoregressive process involving the spatial lag of $\mathbf{y}_t$. Also γ and $\rho_1$are scalar temporal lag and spatial lag parameters respectively and $\boldsymbol{\beta}$ is a k by 1 vector of parameters relating to the k regressors. Additionally the model includes a spatial autoregressive error process is given by

$$\begin{aligned} \boldsymbol{\varepsilon}_t &= \rho_2 \mathbf{M}_N \boldsymbol{\varepsilon}_t + \mathbf{u}_t \\ \mathbf{u}_t &= \boldsymbol{\mu} + \boldsymbol{\nu}_t \end{aligned} \tag{2}$$

in which ρ_2 is a scalar parameter, and $\mathbf{M}_N$ is a connectivity matrix specific to the errors. For simplicity of exposition we assume that $\mathbf{M}_N = \mathbf{W}_N$. Regarding the two components of $\mathbf{u}_t$, $\boldsymbol{\mu}$ is an N by 1 vector of individual-specific (node-specific) time invariant effects and $\boldsymbol{\nu}_t$ is a remainder term. These error components capture unobservable variables and are treated here as random effects, so that $\mu_i \sim iid(0, \sigma_\mu^2)$, $\nu_{it} \sim iid(0, \sigma_\nu^2)$ and $cov(\mu_i, \nu_{it}) = 0$ hence μ_i and ν_{it} are independent of themselves and each other. The individual-specific component $\boldsymbol{\mu}$ imparts time dependency to the error process and accounts for spatially autocorrelated inter-node heterogeneity.

3 Estimation

Baltagi et al (2013) give an estimation method for this model which reflects the presence of the spatial lag $\mathbf{W}_N \mathbf{y}_t$. Here we provide a brief summary. The starting point for estimation is approach of Arellano and Bond (1991) for the dynamic panel model which is applied to the difference equation for the dynamic spatial panel model, which is

$$\Delta \mathbf{y}_t = \gamma \Delta \mathbf{y}_{t-1} + \rho_1 \mathbf{W}_N \Delta \mathbf{y}_t + \Delta \mathbf{x}_t \boldsymbol{\beta} + \Delta \boldsymbol{\nu}_t \tag{3}$$

Following Anderson and Hsiao (1981, 1982) and Arellano and Bond(1991), by differencing a dynamic model one eliminates the unobserved individual effects μ_i, which are correlated with the time-lagged dependent variable, and this leads to orthogonality conditions and hence valid instruments, which are levels of the endogenous and explanatory variables which are uncorrelated with the differenced errors, assuming no serial correlation in the errors. Baltagi et al (2013) (see also Bouayad-Agha and Védrine, 2010) add to the orthogonality conditions of Arellano and Bond (1991) to take account of the presence of the endogenous variable $\mathbf{W}_N \mathbf{y}_t$, so that, assuming the variables $\mathbf{x}_t$ are exogenous, the

orthogonality conditions are

$$\begin{array}{l} E(y_{il}\Delta\nu_{it}) = 0 \qquad \forall i, l = 1, 2, ..., t-2; t = 3, 4, ..., T \\ E(x_{k,im}\Delta\nu_{it}) = 0 \qquad \forall i, k\ m = 1, 2, ..., T; t = 3, 4, ..., T \\ E(\sum_{i\neq j} w_{ij}y_{jl}\Delta\nu_{it}) = 0 \quad \forall i,\ l = 1, 2, ..., t-2; t = 3, 4, ..., T \\ E(\sum_{i\neq j} w_{ij}x_{k,jm}\Delta\nu_{it}) = 0 \quad \forall i,\ k\ m = 1, 2, ..., T; t = 3, 4, ..., T \end{array} \tag{4}$$

Assuming exogenous $\mathbf{x}$, the assumption is that the regressors are uncorrelated with all past, present and future ν. Therefore we can use all lags of the $\mathbf{x}s$ and $\mathbf{W}_N\mathbf{x}s$ as instruments for every t. Making a weaker assumption that the $\mathbf{x}_t$ are predetermined would entail one deeper lag to ensure orthogonality and assuming $\mathbf{x}_t$ comprises endogenous variables means that the lag has to be two deep as in the case of endogenous $\mathbf{y}_t$, as explained by Bond (2002). In these cases the admissible set of instruments changes with t, as t increases more lags become orthogonal and can be included in the instrument set. In this example we impose exogeneity in the DGP, but the significant point here is that what is assumed for $\mathbf{W}_N$ affects the instruments, different $\mathbf{W}_N$ lead to different instruments.

Following Kapoor et al(2007) and Kelejian and Prucha (1998), some conditions need to be imposed on $\mathbf{W}_N$ to give consistent estimation. $\mathbf{W}_N$ should be uniformly bounded in absolute value, which means that a constant c exists such that

$$\begin{array}{l} \max_{1\leq j\leq N} \sum_{h=1}^{N} |w_{jh}| \leq c \leq \infty \\ \max_{1\leq h\leq N} \sum_{h=1}^{N} |w_{jh}| \leq c \leq \infty \end{array} \tag{5}$$

And this should hold for all N, so conceptually as N increases the bound c still holds.
Having obtained consistent parameter estimates using appropriate instruments, the next step in the estimation process involves estimation of the parameters of the error process, ρ_2, σ_ν^2 and subsequently σ_μ^2, which also depends on $\mathbf{M}_N = \mathbf{W}_N$. The method adopted follows the GM approach of Kapoor et al(2007).
To summarise, following Baltagi et al (2013), there are four steps in the estimation process.
In step 1, we first eliminate, by first differencing, the individual effects. Then on the basis of the differenced data we use an IV or GMM estimator to obtain consistent estimates of $\gamma, \rho_1$and $\boldsymbol{\beta}$.
In step 2, resulting consistent IV or GMM residuals are used to obtain consistent GM estimates of ρ_2, σ_ν^2 and $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$ hence σ_μ^2. This is achieved by solving sample moments involving $\mathbf{M}_N = \mathbf{W}_N$ using nonlinear least squares. We provide some detail of this to highlight the role played by $\mathbf{M}_N = \mathbf{W}_N$.
For the GMM estimation at step 2, following Kapoor et al (2007),

$$\boldsymbol{\Gamma}\boldsymbol{\Phi}' - \boldsymbol{\eta} = 0 \tag{6}$$

and

$$\tilde{\boldsymbol{\Gamma}}\boldsymbol{\Phi}' - \tilde{\boldsymbol{\eta}} = 0 \tag{7}$$

where $\boldsymbol{\Gamma}$ and $\tilde{\boldsymbol{\Gamma}}$ are 3 by 4 matrices, $\boldsymbol{\eta}$ and $\tilde{\boldsymbol{\eta}}$ are 3 by 1 vectors and $\boldsymbol{\Phi} = \begin{bmatrix} \rho_2 & \rho_2^2 & \sigma_\nu^2 & \sigma_1^2 \end{bmatrix}$ is a vector of parameters. Using the estimated disturbances $\hat{\boldsymbol{\varepsilon}}$, one obtains sample counterparts $\mathbf{g}$ and $\tilde{\mathbf{g}}$ of vectors $\boldsymbol{\eta}$ and $\tilde{\boldsymbol{\eta}}$, and sample counterparts $\mathbf{G}$ and $\tilde{\mathbf{G}}$ of matrices $\boldsymbol{\Gamma}$ and $\tilde{\boldsymbol{\Gamma}}$. Thus, given $\mathbf{Q}_0 = (\mathbf{I}_T - \frac{\mathbf{J}_T}{T}) \otimes \mathbf{I}_N$ and $\mathbf{Q}_1 = \frac{\mathbf{J}_T}{T} \otimes \mathbf{I}_N$ in which $\mathbf{J}_T$ is a $T \times T$ matrix of 1s, $\mathbf{I}_T$ is a $T \times T$ diagonal matrix , $\mathbf{I}_N$ is a $N\times$ N diagonal matrix,

$$\mathbf{G} = \begin{bmatrix} \frac{2}{N(T-1)}\hat{\boldsymbol{\varepsilon}}'\mathbf{Q}_0\hat{\boldsymbol{\varepsilon}}_{-1} & \frac{-1}{N(T-1)}\hat{\boldsymbol{\varepsilon}}'_{-1}\mathbf{Q}_0\hat{\boldsymbol{\varepsilon}}_{-1} & 1 & 0 \\ \frac{2}{N(T-1)}\hat{\boldsymbol{\varepsilon}}'_{-2}\mathbf{Q}_0\hat{\boldsymbol{\varepsilon}}_{-1} & \frac{-1}{N(T-1)}\hat{\boldsymbol{\varepsilon}}'_{-2}\mathbf{Q}_0\hat{\boldsymbol{\varepsilon}}_{-2} & \frac{1}{N}t_1 & 0 \\ \frac{1}{N(T-1)}(\hat{\boldsymbol{\varepsilon}}'\mathbf{Q}_0\hat{\boldsymbol{\varepsilon}}_{-2} + \hat{\boldsymbol{\varepsilon}}'_{-1}\mathbf{Q}_0\hat{\boldsymbol{\varepsilon}}_{-1}) & \frac{-1}{N(T-1)}\hat{\boldsymbol{\varepsilon}}'_{-1}\mathbf{Q}_0\hat{\boldsymbol{\varepsilon}}_{-2} & 0 & 0 \end{bmatrix} \tag{8}$$

$$\mathbf{g} = \begin{bmatrix} \frac{1}{N(T-1)}\hat{\varepsilon}'\mathbf{Q}_0\hat{\varepsilon} \\ \frac{1}{N(T-1)}\hat{\varepsilon}'_{-1}\mathbf{Q}_0\hat{\varepsilon}_{-1} \\ \frac{1}{N(T-1)}\hat{\varepsilon}'\mathbf{Q}_0\hat{\varepsilon}_{-1} \end{bmatrix} \tag{9}$$

$$\mathbf{G}\begin{bmatrix} \rho_2 & \rho_2^2 & \sigma_\nu^2 & \sigma_1^2 \end{bmatrix}' - \mathbf{g} = \boldsymbol{\zeta}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2) \tag{10}$$

$$\tilde{\mathbf{G}} = \begin{bmatrix} \frac{2}{N}\hat{\varepsilon}'\mathbf{Q}_1\hat{\varepsilon}_{-1} & \frac{-1}{N}\hat{\varepsilon}'_{-1}\mathbf{Q}_1\hat{\varepsilon}_{-1} & 0 & 1 \\ \frac{2}{N}\hat{\varepsilon}'_{-2}\mathbf{Q}_1\hat{\varepsilon}_{-1} & \frac{-1}{N}\hat{\varepsilon}'_{-2}Q_1\hat{\varepsilon}_{-2} & 0 & \frac{1}{N}t_1 \\ \frac{1}{N}(\hat{\varepsilon}'\mathbf{Q}_1\hat{\varepsilon}_{-2} + \hat{\varepsilon}'_{-1}\mathbf{Q}_1\mathbf{Q}_1\hat{\varepsilon}_{-1}) & \frac{-1}{N}\hat{\varepsilon}'_{-1}\mathbf{Q}_1\hat{\varepsilon}_{-2} & 0 & 0 \end{bmatrix} \tag{11}$$

$$\tilde{\mathbf{g}} = \begin{bmatrix} \frac{1}{N}\hat{\varepsilon}'\mathbf{Q}_1\hat{\varepsilon} \\ \frac{1}{N}\hat{\varepsilon}'_{-1}\mathbf{Q}_1\hat{\varepsilon}_{-1} \\ \frac{1}{N}\hat{\varepsilon}'\mathbf{Q}_1\hat{\varepsilon}_{-1} \end{bmatrix} \tag{12}$$

$$\tilde{\mathbf{G}}\begin{bmatrix} \rho_2 & \rho_2^2 & \sigma_\nu^2 & \sigma_1^2 \end{bmatrix}' - \tilde{\mathbf{g}} = \tilde{\boldsymbol{\zeta}}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2) \tag{13}$$

Here we see the role played by the error connectivity matrix $\mathbf{M}_N = \mathbf{W}_N$ in that $t_1 = tr(\mathbf{W}'_N\mathbf{W}_N)$, $\hat{\varepsilon}_{-1} = (\mathbf{I}_T \otimes \mathbf{W}_N)\hat{\varepsilon}$ and $\hat{\varepsilon}_{-2} = (\mathbf{I}_T \otimes \mathbf{W}_N)\hat{\varepsilon}_{-1}$.
Given that $\boldsymbol{\zeta}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2)$, $\tilde{\boldsymbol{\zeta}}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2)$ are vectors of residuals, the nonlinear least squares estimators are given by

$$\begin{array}{l}(\hat{\rho}_2, \hat{\sigma}_\nu^2, \hat{\sigma}_1^2) = \arg\min\{\boldsymbol{\zeta}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2)'\ \boldsymbol{\zeta}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2) + \\ \tilde{\boldsymbol{\zeta}}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2)'\ \tilde{\boldsymbol{\zeta}}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2)\ \}\end{array} \tag{14}$$

Estimation is via unconstrained non-linear least squares estimation using a modified Newton-Raphson method allowing minimisation based on numerical differences.
In general the variances of $\boldsymbol{\zeta}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2)'\ \boldsymbol{\zeta}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2)$ and $\tilde{\boldsymbol{\zeta}}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2)'\ \tilde{\boldsymbol{\zeta}}(\quad \rho_2 \quad \sigma_\nu^2 \quad \sigma_1^2)$ differ, but for simplicity we have not introduced differential weighting. Kapoor et al (2007) note that giving equal weight does give consistent estimates.
Step 3 uses estimated to give preliminary one-step consistent spatial GMM estimates thus

$$\hat{\boldsymbol{\delta}}_1 = (\Delta\tilde{\mathbf{x}}'^*\hat{\mathbf{A}}_N\mathbf{Z}^{*'}\Delta\tilde{\mathbf{x}})^{-1}\Delta\tilde{\mathbf{x}}'^*\hat{\mathbf{A}}_N\mathbf{Z}^{*'}\Delta\mathbf{y} \tag{15}$$

in which $\mathbf{W}_N$ again has a role since $\Delta\tilde{\mathbf{x}} = (\Delta\mathbf{y}_{-1}, (\mathbf{I}_{T-2} \otimes \mathbf{W}_N)\Delta\mathbf{y}, \Delta\mathbf{x})$, $\boldsymbol{\delta}_1 = (\gamma, \rho_1, \beta')$ and $\hat{\mathbf{A}}_N = \left[\mathbf{Z}^{*'}(\mathbf{I}_{T-2} \otimes \hat{\mathbf{H}}_N)(\mathbf{G} \otimes \mathbf{I}_N)(\mathbf{I}_{T-2} \otimes \hat{\mathbf{H}}'_N)\mathbf{Z}^*\right]^{-1}$ with $\hat{\mathbf{H}}_N = (\mathbf{I}_N - \hat{\rho}_2\mathbf{W}_N)^{-1}$. In this $\mathbf{Z}^*$ is the matrix of instruments and $\mathbf{G}$ is a matrix of constants used to obtain the weights matrix of moments $\mathbf{A}_N$ as detailed in Baltagi et al(2013).
In step 4, the final two step-spatial GMM estimates are given by replacing $\hat{\mathbf{A}}_N$ by $\hat{\mathbf{V}}_N$ hence

$$\hat{\boldsymbol{\delta}}_2 = (\Delta\tilde{\mathbf{x}}'^*\hat{\mathbf{V}}_N\mathbf{Z}^{*'}\Delta\tilde{\mathbf{x}})^{-1}\Delta\tilde{\mathbf{x}}'^*\hat{\mathbf{V}}_N\mathbf{Z}^{*'}\Delta\mathbf{y} \tag{16}$$

In which $\hat{\mathbf{V}}_N = \left[\mathbf{Z}^{*'}(\mathbf{I}_{T-2} \otimes \hat{\mathbf{H}}_N)(\Delta\tilde{\boldsymbol{\nu}})(\Delta\tilde{\boldsymbol{\nu}})'(\mathbf{I}_{T-2} \otimes \hat{\mathbf{H}}'_N)Z^*\right]^{-1}$ where $\Delta\tilde{\boldsymbol{\nu}}$ are differenced residuals obtain from step 3.

4 Monte Carlo Simulation

The aim of these Monte Carlo simulation is to demonstrate the effect of misspecifying $\mathbf{W}_N$. This is done by obtaining data by a DGP based on a true $\mathbf{W}_N$ and then estimating the model using a misspecified connectivity matrix $\tilde{\mathbf{W}}_N$. We take different approaches to W matrix specification. First, to facilitate simulation, the spatial matrix "j ahead j behind" given by Kelejian and Prucha (1999) is adopted, where $\mathbf{W}_N$ is based on $j = 2, 5$ and 10. The result is N by N matrices with zeros on

the main diagonal and with non-zero weights equal to 0.25, 0.1 and .05 summing to 1 both across rows and down columns. Data $\mathbf{y}_t$ are generated via $\mathbf{y}_t = \gamma\mathbf{y}_{t-1} + \rho_1\mathbf{W}_N\mathbf{y}_t + \mathbf{x}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t$, $t = 1, ..., T$, commencing with $\mathbf{y}_0 = (\mathbf{I}_N - \rho_1\mathbf{W}_N)^{-1}(\mathbf{x}_0\boldsymbol{\beta} + \boldsymbol{\varepsilon}_0)$.

Following Baltagi et al (2013),

$$x_{it} = \delta x_{it-1} + \xi_t \tag{17}$$

with $x_{i0} = 0$, $\xi_t \sim N(0, \sigma_\xi^2)$. To operationalise the simulations, we utilise the following parameter values $\delta = 0.6$, $\sigma_\xi^2 = 5.0$, $\gamma = 0.2$, $\beta = 1.0$ and $\rho_1 = 0.7$, $N = 100$ at $T = 7$.
For the error process

$$\begin{aligned} &\boldsymbol{\varepsilon}_t = \rho_2\mathbf{M}_N\boldsymbol{\varepsilon}_t + \mathbf{u}_t; \qquad t = 0, ..., T \\ &\mathbf{u}_t = \boldsymbol{\mu} + \boldsymbol{\nu}_t \end{aligned} \tag{18}$$

with $\rho_2 = 0.4$, $\mathbf{M}_N = \mathbf{W}_N$, $\mu_i \sim N(0, \sigma_\mu^2)$, $\nu_{it} \sim N(0, \sigma_\nu^2)$, $\sigma_\mu^2 = 0.8$ and $\sigma_\nu^2 = 0.2$.
Estimation is based on $\tilde{\mathbf{W}}_N$ generated assuming $j = 2, 5$ and 10, so given j, $\tilde{\mathbf{W}}_N$, with equal weights $0.25, 0.1$ and $.05$, is substituted for $\mathbf{W}_N$ in Section 3. In the simulations, so as to reduce the importance of the initial values, we ignore the outcome of the first 10 iterations.
Observe that there are some conditions attached to valid parameter estimates, as given by Baltagi et al (2013). For this model to be dynamically stable and stationary, the parameter space requires $(\mathbf{I}_N - \rho_1\mathbf{W}_N)$ to be non-singular, where $\mathbf{I}_N$ is the (N x N) identity matrix. Non-singularity occurs when $\rho_1 \neq 1/r_i$ for all r_i where r_i denotes an eigenvalue of $\mathbf{W}_N$ and so is guaranteed when $1/r_{\min} < \rho_1 < 1/r_{\max}$ where $r_{\min}$is the most negative purely real characteristic root of $\mathbf{W}_N$. Thus we require ρ_1 to fall within these bounds. It is convenient to observe that with row normalisation of $\mathbf{W}_N$, as in the case here since rows sum to 1.0, $r_{\max} = 1$. Also dynamic stability and stationarity occurs when $|\gamma| < 1 - \rho_1 r_{\max}$ and $\rho_1 > 0$, and $|\gamma| < 1 - \rho_1 r_{\min}$ when $\rho_1 < 0$ (though here we do not exclude complex eigenvalues) and $|\gamma| < 1$. In addition, the conditions imposed on ρ_2 mirror those for ρ_1 but with respect to $\mathbf{M}_N$ rather than $\mathbf{W}_N$, although here $\mathbf{M}_N = \mathbf{W}_N$. Note that for $j = 10$, $r_{\max} = 1$ and $r_{\min} = -0.2782$, for $j = 5$, $r_{\max} = 1$ and $r_{\min} = -0.3457$, and for $j = 2$ $r_{\max} = 1$and $r_{\min} = -0.5625$. Hence the structure of $\mathbf{W}_N$ affects the conditions defining valid parameter estimates.
The outcomes described below are the result of 1000 Monte Carlo samples, where the mean is the mean of the 1000 parameter estimates taken over 1000 replications, the bias is the difference between the estimate and the true value averaged over 1000 replications, and RMSE is the square root of the mean of 1000 squared biases.
The outcomes summarised in Table 1 show clearly that when the true $\mathbf{W}_N$ of the DGP is used in estimation, so that $\tilde{\mathbf{W}}_N = \mathbf{W}_N$, then the estimates are comparatively unbiased with relatively small RMSEs. However when $\tilde{\mathbf{W}}_N \neq \mathbf{W}_N$ bias and RMSE increase, and the means in some cases violate the stability and stationarity conditions, suggesting the same for individual simulations. The ramifications of this are particularly evident when ones examines the 'true' derivatives given in Tables 4, 5 and 6. As pointed out by LeSage and Pace (2009), Debarsy et al (2012) and Elhorst (2013), unlike standard regression, the derivatives $\frac{\partial y}{\partial x_k} \neq \beta_k$. This is as a consequence of feedback effects caused by the lagged endogenous variable and leads to the matrices of derivatives given in equations (5) and (6). Equation (5) relates to the instantaneous effect at time t, hence the total short-run effect is the effect on $\mathbf{y}_t$ at time t of a one unit change in $\mathbf{x}_t$ in each of N nodes. The total long-run effect is the effect on $\mathbf{y}_T$ at time T (T very large) of a one unit change in $\mathbf{x}_t$ in each of N nodes which remains through all times from t to T. LeSage and Pace (2009) suggest summarising these by using the mean of the main diagonal of (5) and (6) to represent mean direct effects, and the means of the off-diagonal cells to represent indirect effects. Total effects are the sum of mean direct and mean indirect effects. We calculate these using the simulation means given in Tables 1. The results in Table 2 show that when $\tilde{\mathbf{W}}_N = \mathbf{W}_N$ there is a large amount of spillover as reflected by the size of the indirect effects especially in the long-run. Thus in the long-run, the total effect of a unit increase is maintained through time in each of the N cells of $\mathbf{x}_t$, $t = 1, ..., T$; $T \to \infty$ is an equilibrium increment in the level of $\mathbf{y}_T$ of about 11, depending on j hence $\mathbf{W}_N$. In the cases in which $\tilde{\mathbf{W}}_N \neq \mathbf{W}_N$, especially when

the means violate stability and stationarity conditions, somewhat 'deviant' outcomes occur which are very dissimilar to the 'true' values obtained when the estimates are based on the true $\mathbf{W}_N$ of the DGP.

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_{1k}} & . & \frac{\partial y_1}{\partial x_{Nk}} \\ . & . & . \\ \frac{\partial y_N}{\partial x_{1k}} & . & \frac{\partial y_N}{\partial x_{Nk}} \end{bmatrix}_t = (\mathbf{I}_N - \rho_1 \mathbf{W}_N)^{-1} \begin{bmatrix} \beta_k & . & 0 \\ . & \beta_k & . \\ 0 & . & \beta_k \end{bmatrix} \tag{19}$$

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_{1k}} & . & \frac{\partial y_1}{\partial x_{Nk}} \\ . & . & . \\ \frac{\partial y_N}{\partial x_{1k}} & . & \frac{\partial y_N}{\partial x_{Nk}} \end{bmatrix} = ((1-\gamma)\mathbf{I}_N - \rho_1 \mathbf{W}_N)^{-1} \begin{bmatrix} \beta_k & . & 0 \\ . & \beta_k & . \\ 0 & . & \beta_k \end{bmatrix} \tag{20}$$

The second approach to W matrix generation is to apply a random process, so that each of $N = 100$ nodes connects at random to other nodes. We assume that were the network to extend so that $N > 100$ additional nodes would be isolated and unconnected to the network so that $\mathbf{W}_N$ would be uniformly bounded in absolute value. Figures 1 and 2 illustrate two alternative networks that provide the basis for comparison of misspecification effects. We refer to Figure 1 as a dense network, and Figure 2 as a sparse network. On the sparse network there are 409 inter-nodal connections with a network density equal to 0.041[1] and on the dense network there are 618 with a network density equal to 0.062. The sparsity of the networks is conditioned somewhat by a wish to otherwise avoid isolated nodes, so in both networks each node has at least one connection to another node. As with the j-ahead and j-behind networks, the main diagonal comprises zeros and we normalise the initial matrix of 1s and 0s so that the resulting matrices $\mathbf{W}_N$ and $\tilde{\mathbf{W}}_N$ have rows which sum to 1, so each row has the same non-zero cell values summing to 1, with values determined by the number of connections for each row (node). Tables 3 again illustrates the effect of misspecifying the true matrix $\mathbf{W}_N$ so that $\tilde{\mathbf{W}}_N \neq \mathbf{W}_N$, summarising the outcomes from 1000 Monte Carlo samples. Using the true matrix, so that $\tilde{\mathbf{W}}_N = \mathbf{W}_N$, generally produces relatively unbiased estimates and small RMEs. Note that for the dense matrix, $r_{\max} = 1$ reflecting row normalisation, but many of the characteristic roots are complex numbers. Of the real characteristic roots, and $r_{\min} = -0.3746$. Likewise for the sparse matrix, $r_{\max} = 1$ and the most negative real characteristic root is $r_{\min} = -0.4721$. Table 4 shows that in terms of short and long-run, direct, indirect and total effects, again using the means from the Monte Carlo simulations, the derivatives are quite similar if we assume the true $\mathbf{W}_N$, regardless of the density. However wrongly assuming a dense matrix in particular makes the total long-run effect much larger than the 'true' derivative, and wrongly assuming a sparse matrix makes the total long-run effect much smaller than the 'true' derivative.

5 The Multilevel Spatial Autoregressive Panel Model

So far we have discussed the implications of misspecifying the matrix $\mathbf{W}_N$ in the context of Dynamic Panel SAR models. However, there is also a growing interest in determining how misspecifying $\mathbf{W}_N$ affects the observed individual outcome when variables are organized in a nested hierarchy, as in Multilevel SAR Panel models, which are becoming increasingly popular across the range of the social sciences.[2] While network effects are invariably ignored, we see many applications of multilevel modelling in educational research where there exist a number of well defined groups organized within a hierarchical structure, such as class within schools, leading to the analysis of effects on individual

[1] The density of a network is defined as the proportion of actually observed ties among the potentially observable ones. In a directed, binary network with actors there could be $N(N-1)$ ties hence the density is defined as $D(N-1)/N$ where D is the number of ties.

[2] Over the past decade there has been a development of methods which have enabled researchers to model hierarchical data. Examples of these methods include multilevel models (see, for example, Goldstein (1998), random coefficient models (Longford, 1993) and hierarchical multilevel models proposed by Goldstein (1986) based on iterative generalized least squares (IGLS).

pupil behaviour coming from different hierarchical levels. In geographical studies, we can often envisage a hierarchy of effects at the level of cities, regions containing cities, and countries containing regions. Failure to recognize these effects emanating from different hierarchical levels can lead to an incorrect interpretation of causality. However, while accounting for multilevel effects goes some way to providing proper inferential analysis, nevertheless this will be flawed if individuals are also connected via a network. This section combines multilevel random effects and simultaneous interactions between individuals caused by their interconnection in a network and examines the effects of misspecifying the matrix $\mathbf{W}_N$ in this setting.

Multilevel models with group effects are generally defined as economic environments where the payoff function of a given agent takes as direct arguments the choice of other agents (Brock and Durlauf, 2001). A typical example is the emergence of social networks where it is often observed that people belonging to the same group tend to behave similarly (Manski, 2000) and the propensity that a person behaves in a certain way varies positively with the dominant behaviour in the group (Bernheim, 1994; Kandori, 1992).[3] As an example, consider students within classrooms, with students' scores y_{ij} dependent on individual student attributes, $\mathbf{x}_{ij}$. In addition other unmeasured causes of individual score variation are represented by random group (classroom) effects α_j and by individual random effects, μ_{ij}, capturing individual latent traits. With $\rho_1 \neq 0$ scores may also be endogenously determined so that a higher score level achieved by one student spills over (via $\mathbf{W}_N\mathbf{y}_t$) to other students in the class. The problem in linear multilevel models is the identification of the parameters in the presence of what Manski (2000) calls reflection problems insofar individual behaviour is affected by the behaviour of the group (s)he belongs to. If the groups are small and the interactions are confined within each group the identification of the endogenous group effect, ρ_1, can be challenging in particular with cross-sectional data. As stressed by Brock and Durlauf (2001) panels allow, instead, for intertemporal interactions and for a richer notion of belief formation which ameliorates identification of the coefficients in the structural model.[4] In addition a more general network structure allows for spillovers both *within* and *between* groups which further facilitates identification.

We now examine the particular case of a two-level hierarchical network model where the inter-individual network interactions spill over across group boundaries. Our model again commences with a matrix $\mathbf{W}_N$ which maps the connections between N individuals in a network. We take account of the existence of groups of individuals because of the existence of levels by the following notation. $\mathbf{W}_N$ is an N by N matrix of connectivity, with cells values 1 and 0 according to whether a pair of individuals is connected (more generally, we can assign values other than 1 and 0 to allow for the strength of the connection between pairs of individuals). Assume individuals are also nested into G groups within a dense and a sparse network structure as shown in Figures 3 and 4, so each group has several individuals within it and spillovers spread across groups. The number of links of each individual varies in the two network structures according to whether it is sparse or dense, nonetheless we assume that each node has at least one connection. We represent the existence of both within and between group effects by writing the matrix $\mathbf{W}_N$ as:

$$\mathbf{W}_N = \begin{bmatrix} \mathbf{W}_{11} & ... & \mathbf{W}_{1q} & ... & \mathbf{W}_{1J} \\ \mathbf{W}_{21} & ... & \mathbf{W}_{2q} & ... & \mathbf{W}_{2J} \\ ... & ... & ... & ... & ... \\ \mathbf{W}_{J1} & & \mathbf{W}_{Jq} & & \mathbf{W}_{JJ} \end{bmatrix} \tag{21}$$

Given group p and group q the generic element $\mathbf{W}_{pq}$, $p = 1, ..., J$; $q = 1, ..., J$,is an N_p by N_q matrix defining the interconnectivity between individuals in group p and those in group q. The matrix

[3] Other influences are the so called peer influence effects which have been extensively examined both in education (Bénabou (1993)), in the psychology literature (Brown (1990) and Brown, Clasen, and Eicher (1986)) and in the occurrence of social pathologies (Bauman and Fisher (1986); Krosnick and Judd (1982); Jones (1994)).

[4] Of course, identification only becomes a problem in linear-in-means models, and any nonlinearity, for instance as in an expanded spatial Durbin model (see Gibbons and Overman, 2010), automatically solves the problem.

$\mathbf{W}_N$ thus defines how individuals are connected to other individuals, and these could be in the same group, or in different groups. In the absence of consideration of network effects, much of the multilevel literature assumes that inter-individual interaction is restricted to within group boundaries which is the main cause of lack of identification of the endogenous effect. Thus, from a spatial perspective it is assumed, somewhat differently from conventional spatial econometrics involving autoregressive processes, that inter-individual interactions are restricted spatially.

A Multilevel SAR Panel model can be represented as:

$$y_{ij,t} = \rho_1 \sum_{j=1}^{J} \sum_{i=1}^{N_j} W_{ij,gh} y_{gh,t} + \mathbf{x}_{ij,t}\boldsymbol{\beta} + e_{ij,t} \tag{22}$$

In the above, we assume a hierarchical structure with J groups ($j = 1, ..., J$). Nested within group j there are N_j individuals ($i = 1, ..., N_j$). And for each individual there are T observations ($t = 1, ..., T$). Hence, the response $y_{ij,t}$ partly depends on the endogenous spatial lag with typical element of the weights matrix $\mathbf{W}$ denoted by $W_{ij,gh} = W_{k,l}(k = ij, l = gh)$ where ij denotes individual i within group j. Thus k, $l = 1, ..., N_j$ with $N = \sum_{j=1}^{J} N_j$ and $\mathbf{W}_N$ is the N by N network matrix.

The hierarchical random effects operate via the errors. Accordingly, the error components for N individuals distributed amongst G groups are considered to be:

$$e_{ij,t} = \alpha_j + \mu_{ij} + \nu_{ij,t} \tag{23}$$

in which α_j is the unobservable group specific effect which is constant over time, and assumed to be i.i.d. $(0, \sigma_\alpha^2)$, μ_{ij} is the time invariant unobservable individual effect distributed as i.i.d. $(0, \sigma_\mu^2)$, and $\nu_{ij,t}$ is the remainder disturbance distributed as i.i.d. $(0, \sigma_\nu^2)$. Each of α_j, μ_{ij} and $\upsilon_{ij,t}$ are independent of each other and also internally independent.

We can see how the identification in presence of endogeneity can pose serious challenges in this setting by considering the following additional regression:

$$\widehat{y}^w_{ij,t} = \frac{1}{1-\rho_1}\left(\beta_0 + \mathbf{x}^w_{ij,t}\boldsymbol{\beta}_1\right) \quad \text{where } \widehat{y}^w_{ij,t} = \sum_{j=1}^{J}\sum_{i=1}^{N_j} W_{ij,gh} y_{gh,t} \quad \text{and} \quad \mathbf{x}^w_{ij,t} = \sum_{j=1}^{J}\sum_{i=1}^{N_j} W_{ij,gh}\mathbf{x}_{gh,t} \tag{24}$$

which is derived directly taking the spatial average of (22). We note that the set of regressors $(1, Y^w_{ij,t}, \mathbf{x}_{ij,t})$ in the structural model (22) requires the estimation of $(2 + k)$ parameters. While in the reduced form implied by (22) and by (24) the set of regressors[5] $(1, \mathbf{x}_{ij,t}, \mathbf{x}^w_{ij,t})$ allows one to identify $(1 + 2k)$ parameters. Assuming $k = 1$ all parameters in the structural equation (22) are therefore just exactly identified.

5.1 Identification and Endogeneity in SAR Multilevel Panel Data

The main assumption behind the identification of the endogenous effect is that there no correlation between the error terms at the individual and group levels and the regressors. This, of course, may be a limiting assumption. For example, students or school unobservable effects may be correlated with some of the observable characteristics at the school of individual level. To account for this potential correlation we consider the following extended model:

$$\begin{aligned} y_{ij,t} &= \beta_0 + \rho_1 y^w_{ij,t} + \mathbf{x}_{ij,t}\boldsymbol{\beta} + e_{ij,t} \\ e_{ij,t} &= \widetilde{\alpha}_j + \widetilde{\mu}_{ji} + \nu_{ij,t} \end{aligned} \tag{25}$$

[5] If we replace the instrument (24) in (22), we obtain the following reduced form $Y_{ij,t} = \beta_0 + \mathbf{x}_{ij,t}\beta + \frac{\rho_1}{1-\rho_1}\mathbf{x}^w_{ij,t}\beta + \alpha_j + \mu_{ij} + \nu_{ijt}$ with a set of regressors given by $(1, \mathbf{x}_{ij,t}, \mathbf{x}^w_{ij,t})$.

where $e_{ij,t}$ is an error containing both individual effects, $\widetilde{\alpha}_j$, group effects, $\widetilde{\mu}_{ij}$, and a remaining disturbance term, $\nu_{ij,t}$, which is iid across i, j and t.

We now extend Mundlak's (1978) approach to a multilevel setting. According to Mundlak (see also Debarsy, 2012) the random effect specification ignores the possible correlation between individual effects and regressors $\sigma_{\alpha x} \neq 0$ and $\sigma_{\mu x} \neq 0$ (see also Ebbes et al, 2004). By controlling for this correlation in an auxiliary regression for $\widetilde{\alpha}_j$ and $\widetilde{\mu}_{ij}$ the assumption of independence between the random effects and the controls can be restored. In fact, as $\widetilde{\alpha}_j$ and $\widetilde{\mu}_{ij}$ are time invariant, they should only be correlated with the time-invariant part of the explanatory variables:

$$\begin{aligned} \widetilde{\alpha}_j &= \overline{\mathbf{x}}_{ij}\boldsymbol{\beta}_3 + \alpha_j \\ \widetilde{\mu}_{ij} &= \overline{\mathbf{x}}_{j}\boldsymbol{\beta}_4 + \mu_{ij} \end{aligned} \tag{26}$$

where $\overline{\mathbf{x}}_{ij} = \sum_{t=1}^{T} \mathbf{x}_{ij,t}$ denotes the average over time of $\mathbf{x}$ for the unit i, $\overline{\mathbf{x}}_{j} = \sum_{t=1}^{T} \mathbf{x}_{j,t}$ denotes the average over time for group j and $\alpha_j \sim iid(0, \sigma_\alpha)$ and $\mu_{ij} \sim iid(0, \sigma_\mu)$. By substituting (26) in (25):

$$y_{ij,t} = \beta_0 + \rho_1 y^w_{ij,t} + \mathbf{x}_{ij,t}\boldsymbol{\beta} + \overline{\mathbf{x}}_{ij}\boldsymbol{\beta}_3 + \overline{\mathbf{x}}_{j}\boldsymbol{\beta}_4 + \alpha_j + \mu_{ij} + \nu_{ijt} \tag{27}$$

We can see how the identification problem is relaxed in this setting by considering the following instrumental regression for $y^w_{ij,t}$ which is derived by taking the spatial average of (27):

$$Ey^w_{ij,t} = \frac{1}{1-\rho_1}\left(\beta_0 + \mathbf{x}^w_{ij,t}\boldsymbol{\beta}_1 + \overline{\mathbf{x}}^w_{ij}\boldsymbol{\beta}_3 + \overline{\mathbf{x}}^w_{j}\boldsymbol{\beta}_4\right) \tag{28}$$

where $\overline{\mathbf{x}}^w_{ij} = \sum_{t=1}^{T} \overline{\mathbf{x}}^w_{ij,t}$ denotes the average over time of the spatial lag for the individual effect and $\overline{\mathbf{x}}^w_{j} = \sum_{t=1}^{T} \overline{\mathbf{x}}^w_{j,t}$ is the average over time of the spatial lag for the group effect. The model avoids the linear dependence between $y^w_{j,t}$, $\mathbf{x}^w_{ij,t}$ since we also have in the instrumental equation for $y^w_{ij,t}$ the average over time of the spatially lagged individual and group effects, $\overline{\mathbf{x}}^w_{ij}$ and $\overline{\mathbf{x}}^w_{j}$. This implies that $y^w_{ij,t}$ depends on the entire history of the spatial interactions embedded in $\overline{\mathbf{x}}^w_{ij}$ and $\overline{\mathbf{x}}^w_{j}$resolving the contemporaneous correlation with the same variables. If we replace (28) in (27), we obtain:

$$y_{ij,t} = \beta_0 + \mathbf{x}_{ij,t}\boldsymbol{\beta} + \overline{\mathbf{x}}_{ij}\boldsymbol{\beta}_3 + \overline{\mathbf{x}}_{j}\boldsymbol{\beta}_4 + \frac{\rho_1}{1-\rho_1}\left(\mathbf{x}^w_{ij,t}\boldsymbol{\beta} + \overline{\mathbf{x}}^w_{ij}\boldsymbol{\beta}_3 + \overline{\mathbf{x}}^w_{j}\boldsymbol{\beta}_4\right) + \alpha_j + \mu_{ij} + \nu_{ijt} \tag{29}$$

We can state the following results for the identification of the social interaction effects in a panel data setting:

- The set of regressors $(1, y^w_{ij,t}, \mathbf{x}_{ij,t}, \overline{\mathbf{x}}_{ij}, \overline{\mathbf{x}}_{j})$ in the structural equation (27) requires the estimation of $(2+3r)$ parameters.
- In the reduced form (29) the set regressors $(1, \mathbf{x}_{ij,t}, \mathbf{x}^w_{ij,t}, \overline{\mathbf{x}}^w_{ij}, \overline{\mathbf{x}}^w_{j}, \overline{\mathbf{x}}_{ij}, \mathbf{v}_j)$ allows to identify $(1+6r)$ parameters. The five parameters in the structural equation (25) are now over-identified and we can recover ρ_1.

Hence, by augmenting the random effects specification with variables capturing the correlation between (time-varying) regressors and individual and group effects we are able to efficiently estimate all the parameters in the structural model and identify the endogenous effects. There are several advantages of the proposed method. First, in the presence of network interconnectivity we can

account for cross-group behaviour and achieve identification of the endogenous social interaction effects using the additional information stemming from cross-group spillovers. Second, the model avoids linear dependence between $y^{\mathbf{w}}_{ij,t}$, and $\mathbf{x}^{\mathbf{w}}_{ij,t}$ which may be not provide sufficient information for the identification of the endogenous spatial effects (this would be an exactly identified model as shown in the section above). Now the equation for the spatial endogenous effect, $y^{\mathbf{w}}_{ij,t}$, is linearly dependent also on the individual average and on the group average over time of the spatial lag of the $\mathbf{x}$ variables, $\overline{\mathbf{x}}^{\mathbf{w}}_{ij}$ and $\overline{\mathbf{x}}^{\mathbf{w}}_{j}$, which are time invariant and so are correlated with the (time invariant) individual and group effects. Hence, by introducing the two additional instruments $\overline{\mathbf{x}}^{\mathbf{w}}_{ij}$ and $\overline{\mathbf{x}}^{\mathbf{w}}_{j}$ one can also remove any residual correlation between the group and individual effects α_j, μ_{ij} and the regressors (which is the other source of endogeneity in the model) while continuing to use a random effect specification.

However, the identification process rests on the proper specification of the network matrix $\mathbf{W}_N$ and any misspecification, as we will see in the simulations below, will translate into a bias in the estimation of the endogenous effect ρ_1. Again, the significant point here, as in the dynamic panel data estimation, is that what is assumed for $\mathbf{W}_N$ affects the internal instruments, and different $\mathbf{W}_N$ lead to different internal instruments which affect the individual outcome.

6 IGLS Estimation

The univariate model (22) written in matrix notation is:

$$\mathbf{y}_t = \rho_1 \mathbf{W}_N \mathbf{y}_t + \mathbf{x}_t \boldsymbol{\beta} + \mathbf{e}_t \tag{30}$$

in which $\mathbf{y}_t$ is of dimension (N x 1), $\mathbf{x}_t$ is an (N x k) matrix of explanatory variables and $\mathbf{e}_t$ is the error term of dimension (N x 1) comprising nested error components, with the suffix $t = 1, ..., T$ denoting time. Also ρ_1 is a scalar and $\boldsymbol{\beta}$ is a (k x 1) vector of parameters to be estimated. Stacking over time, the model is

$$\begin{aligned} \mathbf{y} &= \rho_1 (\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y} + \mathbf{x}\boldsymbol{\beta} + \mathbf{e} \\ \mathbf{y} &= \mathbf{B}^{-1}(\mathbf{x}\boldsymbol{\beta} + \mathbf{e}) \end{aligned} \tag{31}$$

in which $\mathbf{B} = \mathbf{I}_T \otimes (\mathbf{I}_{M_i} - \rho_1 \mathbf{W}_N)$. The consequence of this is that $(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}$ is not uncorrelated with $\mathbf{e}$, since $E[(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}\mathbf{e}^T] = E[(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{B}^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{e})\mathbf{e}^T] = (\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{B}^{-1}\boldsymbol{\Omega} \neq \mathbf{0}$. The OLS estimator is therefore not consistent.

The hierarchical two-stage method for estimating the fixed and random effects in (30) originally proposed by Goldstein (1986),[6] is based upon an Iterative Least Squares (IGLS) method that results in consistent and asymptotically efficient estimates of $\boldsymbol{\beta}$. We first rewrite (30) in compact form as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\delta} + \boldsymbol{\theta}_{\mathbf{e}}\mathbf{e} \tag{32}$$

where $\mathbf{X} = [(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}\ \mathbf{x}]$, $\boldsymbol{\delta} = [\rho_1\ \boldsymbol{\beta}]$ and $\boldsymbol{\theta}_{\mathbf{e}}$ is the design matrix of the random effects that will be used in the estimation to derive the estimates for σ^2_α, $\widehat{\sigma}^2_\mu$ and $\widehat{\sigma}^2_v$.

First one obtains starting values for $\boldsymbol{\delta}$, $\widetilde{\boldsymbol{\delta}}$ by performing OLS in a standard single level system assuming the variance at higher level of the model to be zero. Conditioned upon $\widetilde{\boldsymbol{\delta}}$, we form the vector of residuals which we use to construct an initial estimate, $\mathbf{V}$, the covariance matrix for the response variable $\mathbf{y}$. Then one iterates the following procedure first estimating $\widehat{\boldsymbol{\delta}}$ in a GLS regression as:

$$\widehat{\boldsymbol{\delta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}\right) \tag{33}$$

[6] The method is currently implemented in the software RunMLwiN.

and again calculating residuals $\widehat{\mathbf{r}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\delta}}$. We vectorize the residual covariance matrix by stacking the columns one on top of another into a vector, i.e. $\mathbf{R}^* = vec(\widehat{\mathbf{r}}\,\widehat{\mathbf{r}}^T)$ regressing it on a design matrix so that the estimated vector of regression coefficients $\widehat{\boldsymbol{\theta}}_{\mathbf{e}}$ comprises the initial estimates of $\widehat{\sigma}^2_\alpha$, $\widehat{\sigma}^2_\mu$ and $\widehat{\sigma}^2_v$. Hence, we estimate the random component variances as:

$$\widehat{\boldsymbol{\theta}}_{\mathbf{e}} = (\boldsymbol{\theta}^{*T}_{\mathbf{e}}\mathbf{V}^{*-1}\boldsymbol{\theta}^{*}_{\mathbf{e}})^{-1}\left(\boldsymbol{\theta}^{*T}_{\mathbf{e}}\mathbf{V}^{*-1}\mathbf{R}^{*}\right) \tag{34}$$

where $\mathbf{V}^* = 2\,(\mathbf{V}\otimes\mathbf{V})$ is the covariance matrix of $\mathbf{R}^*$ and $\mathbf{V} = E(\widehat{\mathbf{r}}\,\widehat{\mathbf{r}}^T)$ is the covariance matrix of $\widehat{\mathbf{r}}$ while $\boldsymbol{\theta}^*_{\mathbf{e}}$ is the design matrix of the random effects. Given $\mathbf{V}^*$ new coefficients $\widehat{\boldsymbol{\delta}}$ are estimated once again. This procedure is repeated until some convergence criteria are met. Given the consistency of the starting values, $\widehat{\boldsymbol{\delta}}_{GLS}$, since at convergence $\widehat{\boldsymbol{\delta}}_{IGLS} = \widehat{\boldsymbol{\delta}}_{GLS}$, the estimator in (33) provides a consistent (maximum likelihood under Normality) estimator of $\boldsymbol{\delta}$. We also obtain consistent estimates of the fixed effects standard errors using:

$$\text{SE}(\widehat{\boldsymbol{\delta}}) = (\mathbf{X}^T\mathbf{V}^{*-1}\mathbf{X})^{-1} \tag{35}$$

Given that the IGLS method used in the context of random multilevel modelling is equivalent to a maximum likelihood method under multivariate normality this in turn may lead to biased estimates (see Rice et al, 1998). To produce unbiased estimates we use a Restricted Iterative Generalized Least Squares (RIGLS) method which, after the convergence is achieved, turns out to be equivalent to a Restricted Maximum Likelihood Estimate (REML). One advantage of the latter method is that, in contrast to IGLS, estimates of the variance components take into account the loss of the degrees of freedom resulting from the estimation of the regression parameters. Hence, while the IGLS estimates for the variance components have a downward bias, the RIGLS estimates don't.

6.1 Multivariate IGLS

We also consider a simultaneous estimation of $\mathbf{y}_t$ and $\mathbf{W}_N\mathbf{y}_t$ were we allow for a correlation between the error terms at the individual and group levels and the regressors. This is done via a simultaneous estimation of equations (27) and (28) which in compact form read as:

$$\begin{aligned}
\mathbf{y} &= \rho_1(\mathbf{I}_T\otimes\mathbf{W}_N)\mathbf{y} + \mathbf{x}\boldsymbol{\beta} + \mathbf{J}_T\widetilde{\mathbf{x}}\boldsymbol{\beta}_2 + \mathbf{J}_{GT}\widetilde{\mathbf{x}}\boldsymbol{\beta}_3 + \mathbf{e}_1 \\
(\mathbf{I}_T\otimes\mathbf{W}_N)\mathbf{y} &= (\mathbf{I}_T\otimes\mathbf{W}_N)\mathbf{x}\boldsymbol{\gamma}_1 + \mathbf{J}_T(\mathbf{I}_T\otimes\mathbf{W}_N)\widetilde{\mathbf{x}}\boldsymbol{\gamma_2} + \mathbf{J}_{GT}(\mathbf{I}_T\otimes\mathbf{W}_N)\widetilde{\mathbf{x}}\boldsymbol{\gamma}_3 + \mathbf{e}_2
\end{aligned} \tag{36}$$

where $\mathbf{J}_T = \overline{\mathbf{J}}_T\otimes diag(\mathbf{I}_{N_j})$, $\overline{\mathbf{J}}_T = \frac{\boldsymbol{\iota}_T\boldsymbol{\iota}_T^{'}}{T}$ is the operator computing averages of observations over time with $\boldsymbol{\iota}_T$ denoting a T-dimensional vector of ones and $\mathbf{J}_{GT} = \overline{\mathbf{J}}_T\otimes diag(\overline{\mathbf{J}}_{N_j})$ is the operator computing averages over groups and over time in which $\overline{\mathbf{J}}_{N_j} = \frac{\boldsymbol{\iota}_{N_j}\boldsymbol{\iota}_{N_j}^{'}}{N_j}$. T is the number of time periods and there are J groups each containing N_j units so that the total number of observations is $\sum_{j=1}^{J} N_j = N$. We denote with $\widetilde{\mathbf{x}}$ the set of controls excluding the constant term.

The equation for $\mathbf{y}$, which normally would simply comprise the endogenous variable $(\mathbf{I}_T\otimes\mathbf{W})\mathbf{y}$ and the exogenous regressors $\mathbf{x}$ is thus extended to include, in parallel with Mundlak (1978), the individual average and the group average over time of the $\mathbf{x}$ variables, $\mathbf{J}_T\widetilde{\mathbf{x}}$ and $\mathbf{J}_{\mathbf{GT}}\widetilde{\mathbf{x}}$. The auxiliary equation for $\mathbf{W}_N\mathbf{y}_t$, which would normally comprise only $(\mathbf{I}_T\otimes\mathbf{W})\widetilde{\mathbf{X}}$, also includes the individual average and the group average over time of the spatial lag of the $\mathbf{x}$ variables, $\mathbf{J}_T(\mathbf{I}_T\otimes\mathbf{W})\widetilde{\mathbf{x}}$ and $\mathbf{J}_{GT}(\mathbf{I}_T\otimes\mathbf{W})\widetilde{\mathbf{x}}$. These additional variables are assumed to be linearly related to the unobservable random effects and required to produce consistent estimates.

System (36) can be written in compact form as:

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{X}_1\boldsymbol{\delta}_1 + \boldsymbol{\theta}_{\mathbf{e}_1}\mathbf{e}_1 \\
\mathbf{y}_2 &= \mathbf{X}_2\boldsymbol{\delta}_2 + \boldsymbol{\theta}_{\mathbf{e}_2}\mathbf{e}_2
\end{aligned} \tag{37}$$

where $\mathbf{y}_1 = \mathbf{y}$, $\boldsymbol{\delta}_1 = [\boldsymbol{\delta}\ \boldsymbol{\beta}_2\ \boldsymbol{\beta}_3]$ $\mathbf{X}_1 = [\mathbf{X}\ \ \mathbf{J}_T\widetilde{\mathbf{x}}\ \ \mathbf{J}_{GT}\widetilde{\mathbf{x}}]$, $\mathbf{y}_2 = (\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}$, $\boldsymbol{\delta}_2 = [\gamma_1\ \boldsymbol{\gamma_2}\ \boldsymbol{\gamma_3}]$ and $\mathbf{X}_2 = [(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{x}\ \ \mathbf{J}_T(\mathbf{I}_T \otimes \mathbf{W}_N)\widetilde{\mathbf{x}}\ \ \mathbf{J}_{GT}(\mathbf{I}_T \otimes \mathbf{W}_N)\widetilde{\mathbf{x}}]$**.** The estimation of $\boldsymbol{\delta}_1$,$\boldsymbol{\delta}_2$, $\boldsymbol{\theta}_{\mathbf{e}_1}$ and $\boldsymbol{\theta}_{\mathbf{e}_2}$ is also done via IGLS. Where the measurements have a Normal distribution IGLS provides maximum likelihood estimates.

7 Monte Carlo Simulations

The aim of these Monte Carlo simulations is to demonstrate the effect of misspecifying $\mathbf{W}_N$ and how this affects the identification of the endogenous effect ρ_1. This is done by obtaining data by a DGP based on a true $\mathbf{W}_N$ and then estimating the model using a misspecified connectivity matrix $\tilde{\mathbf{W}}_N$. Specifically, data $\mathbf{y}_t$ are generated via

$$y_{ij,t} = \rho_1 \sum_{j=1}^{J} \sum_{i=1}^{N_j} W_{ij,gh} y_{gh,t} + x_{ij,t}\beta + e_{ij,t} \tag{38}$$

where $t = 1, ..., T$ and the observations are organised in a nested hierarchy with J groups ($j = 1, ..., J$) and N_j individuals in each group ($i = 1, ..., N_j$). In our example we have 100 nodes organised in 10 groups and each node has 7 observations in time hence $N = 100, G = 10$ and $T = 7$. We also assume that $\mathbf{W}_N$ is a time invariant network matrix. We set $\rho_1 = 0.7$, $\beta = 1$ and we assume that the variable $x_{ij,t}$ is drawn from a uniform distribution, $x_{ij,t} \in (0, 1)$. For the error process we assume:

$$e_{ij,t} = \alpha_i + \mu_{ij} + \nu_{ij,t} \tag{39}$$

where the error components are also drawn from a uniform distribution, hence $\alpha_i \in (0, 1)$, $\mu_{ij} \in (0, 1)$ and $\nu_{ij,t} \in (0, 1)$.

To define $\mathbf{W}_N$ we consider the same sparse and dense network structures defined in section four. The difference in the network structure is that now nodes are nested within groups and nodes interconnections are not confined within the same group but spillover across groups, as shown in Figures 3 and 4. Table 5 illustrate the effect of misspecifying the true matrix $\mathbf{W}_N$ so that $\tilde{\mathbf{W}}_N \neq \mathbf{W}_N$ when we estimate equation (30) via IGLS, summarising the outcomes from 1000 Monte Carlo samples. Using the true matrix, so that $\tilde{\mathbf{W}}_N = \mathbf{W}_N$, generally produces relatively unbiased estimates of ρ_1 and β and small RMEs. When a dense matrix is assumed but the true matrix is sparse leads to larger bias in the estimation of both ρ_1 and β.

We now consider the outcome of multivariate estimation of both $\mathbf{y}_t$ and $\mathbf{W}_N\mathbf{y}_t$ to account for the endogeneity of $\mathbf{W}_N\mathbf{y}_t$. We first consider a multivariate model where there is no correlation between the error terms and the regressors ($\boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = \boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_3 = 0$). This shown in Table 6. The bias in estimating the endogenous effect ρ_1 when the true matrix $\mathbf{W}_N$ is misspecified is much larger than in the univariate estimation. This happens since $\mathbf{y}_t$ and $\mathbf{W}_N\mathbf{y}_t$ are jointly estimated but the equation for $\mathbf{W}_N\mathbf{y}_t$ is misspecified since the internal instrument $(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{x}$ uses the wrong matrix $\mathbf{W}_N$ which translates into a larger bias in the estimation of ρ_1 in the equation for $\mathbf{y}_t$. We note that the endogenous effect ρ_1 will be largely downsized especially when a sparse matrix in used in the DGP but a dense matrix is assumed in the estimation. When we estimate the multivariate model assuming correlation between the error terms and the regressors ($\boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_3 \neq \boldsymbol{\gamma}_2 \neq \boldsymbol{\gamma}_3 \neq 0$), see Table 7, the bias in the estimation of the endogenous effect, ρ_1, is even larger especially, again, when the true matrix $\mathbf{W}_N$ of the DGP is sparse but a dense $\mathbf{W}_N$ is assumed in the estimation. Since the equation for $\mathbf{W}_N\mathbf{y}_t$ includes also the individual average and the group average over time of the spatial lag of the $\mathbf{x}$ variables $\mathbf{J}_T(\mathbf{I}_T \otimes \mathbf{W_N})\widetilde{\mathbf{x}}$ and $\mathbf{J}_{GT}(\mathbf{I}_T \otimes \mathbf{W_N})\widetilde{\mathbf{x}}$**,** which are also a function of $\mathbf{W_N}$, if $\mathbf{W_N}$ is misspecified the bias in the estimation of the endogenous effect, ρ_1, is further amplified.

7.1 Interpretation

In the previous simulation study we have assumed a sparse $\mathbf{W}_N$ matrix, condition likely to be met in a number of scenarios, for example when interaction pools are limited. In order to obtain a closer representation of the spatial interaction process in $\mathbf{W}_N$ matrix construction choices, Anselin (2010) suggests greater focus on modelling agents involved in social and economic interactions which can indeed imply a network structure with a small number of interconnections. Assuming in the estimation a dense network structure when the true underlying structure is sparse leads to serious bias in estimating the strength of the endogenous effect ρ_1 which may invalidate the analysis. In multilevel analysis is also good practice to account for potential correlation between the error terms and the regressors by including spatial lags of $\mathbf{x}$ in the auxiliary equation for $\mathbf{W}_N\mathbf{y}_t$. In case the matrix $\mathbf{W}_N$ is misspecified the inclusion of these spatial lags, which are a function of $\mathbf{W}_N$, further increases the bias in the estimation of the endogenous effect, ρ_1.

The second issue related to $\mathbf{W}_N$ as a representation of a network involving nodes (people or places) and links between nodes is that these can be seen as dynamic evolving entities, and we can envisage network development to be a response to costs and benefits in being a node or a link on the network. Hence the misspecification of $\mathbf{W}_N$ may simply come from ignoring its dynamical evolution in time. Some networks might be dynamic and ephemeral, some networks in a stable equilibrium, and some network slowly evolving. Following Goyal (2009), it may be the case that ephemeral and dynamic networks manifest themselves when there are payoffs. This leads to a theory of network formation, thus 'A game of network formation specifies a set of players, the link formation actions available to each player and the payoffs to each player from the networks that arise out of individual linking decisions', and 'A network is said to be strategically stable or an equilibrium if there are no incentives for individual players (either acting alone or in groups) to form or delete links and thereby alter the network' (Goyal, 2009). A quasi-stable network is similar to what is typically assumed in the regional science or economic geography literature, where a network is fixed or only very slowly evolving as a result of the sunk capital embodied in transport infrastructure investment which defines the inter-nodal links. The emerging literature on endogenous network dynamics considers dynamic stochastic network formation games which could give a more rational basis for the structure of the $\mathbf{W}_N$ matrix (see also Corrado and Fingleton, 2012).

The potential for dynamic $\mathbf{W}_N$ matrices introduces some problems for estimation, given the assertion that $\mathbf{W}_N$ is necessarily a fixed entity. This may be of little consequence in cross-sectional modelling, where at a given snapshot in time a fixed $\mathbf{W}_N$ may be a reasonable approximation, but with the extension of spatial econometrics to include panel data modeling there is the possibility that $\mathbf{W}_N$ evolves, and interacts with the regression variables. The possibility of endogenous interaction leads Anselin (2010) to remarks that 'an endogenous spatial weights matrix would jointly determine who interacts (and why) and how that interaction affects the rest of the model. Much progress remains to be made...'. Nevertheless one can have a dynamic $\mathbf{W}_N$ matrix if it is part of a simulation, with no consequence for estimation, as in Fingleton (2001).

8 Conclusions

We have demonstrated that the estimates produced are contingent on what is assumed about $\mathbf{W}_N$, given that in practice the true $\mathbf{W}_N$ is likely to be unknown. It is evident that an estimate of true $\mathbf{W}_N$ should be based on data which gives an accurate measure of the interdependence between nodes. One relatively simple but surprisingly popular approach, which is purely assumption rather than evidence based, is that nodes only form network connections or interact with immediate neighbours. In this case one might base $\mathbf{W}_N$ on a simple 1,0 metric of distance, in which 1 indicates that a pair of nodes are contiguous (near), and 0 that they are not immediate neighbours (far). However this is a very debatable assumption of what determines connectivity and a very crude measure of distance, which

may not stand up to empirical scrutiny. It is simply a theory and it would seem somewhat risky to base inference on the untested assumptions embodied within this contiguity matrix approach. A slightly more sophisticated measure of distance would be straight line distance, travel time or some related measure, so that non-contiguous nodes could be assumed to interact, with the strength of interaction governed by the distance between pairs of nodes. However without evidence to support this supposition, inference would seem to be on shaky ground. Treating areal units (counties, regions etc) as nodes, Cliff and Ord (1973) go a step further by suggesting that $w_{ij} = d_{ij}^a / s_{ij}^b$ in which d_{ij} is a measure of distance between nodes i and j and s_{ij} is the proportion of i's boundary shared with j. However one would still have to assume values for a and b and therefore one would still be basing $\mathbf{W}_N$ on assumption.

Model-based approaches to connectivity such as Cliff and Ord (1973) relate connectivity, which may be difficult to measure per se, to easy-to-measure variables such as geographic distance. The concept of economic distance is similar, but partially relates connectivity to the size of the nodes (economies), thus $w_{ij} = Q_i^{\alpha} Q_j^{\beta} d_{ij}^{-\gamma}$ in which N by 1 vector $\mathbf{Q}$ has elements which are measures of the size of N economies (say cities, regions or countries), as given by employment or gross value added, d_{ij} is the geographic distance separating nodes i and j, and α, β and γ are parameters. If α, β and γ were known, then $\mathbf{W}_N$ could be constructed, but this is unlikely.

It appears that the limitations of modelling as described above should lead us to what appears to be a more practicable way to proceed which is to estimate directly the spatial matrix. Lam and Souza (2015) has proposed a direct estimation of the spatial weight matrix $\mathbf{W}_N$ along with spatial parameters, with recourse to a regularization based on the adaptive lasso with instrumental variables which uses exogenous covariates as internal instruments. However, the direct estimation of $\mathbf{W}_N$ is clearly going to break down when the dataset are large.

An alternative route is to observe directly the interaction between nodes. Data measuring internodal connectivity are freely available in diverse forms, for example international or interregional migration flows, traffic flows within and between cities, regions and countries, international trade flows, and more recently internet communications at various scale from the level of individual people upwards. These various flow data have at least two characteristics in common. One is that they are dynamic rather than static, and secondly there is an overabundance of data, particular with regard to electronic communication; we are in the era of big data. Neither of these characteristics are in practice particularly attractive when it comes to using such data as a basis for $\mathbf{W}_N$, which needs to be a static matrix in order to underpin our currently available estimation technology. Big data needs to be reduced to a workable format, and this entails some form of reduction and modelling strategy to extract usable information about connectivity on the network of interest. It is clear that in modelling spatial phenomena, and probably generally, as ever there is a need for good judgement as to the appropriate variables and how they are measured.

To sum up, network and spatial interactions have become an important issue in economic research. Correctly modeling and estimating such effects in spatial and network econometrics is, therefore, of paramount importance. Particular attention should be paid to the definition of the $\mathbf{W}_N$ matrix which is what such network and spatial interdependence translates into. While the cell values of $\mathbf{W}_N$ are assumed to be known, in reality often they are not known and when a model is estimated using a $\mathbf{W}_N$ matrix that is inaccurate this will lead to biased estimates. We show this in the context of dynamic and multilevel SAR panel data models. The results of the Monte Carlo simulations show a large bias in estimating the strength of the endogenous effect ρ_1 when the spatial lag or the interconnectivity assumed for $\mathbf{W}_N$ in the estimation differs from that of the true DGP. This implication is not far from being the norm in many econometric analysis involving spatial and network effects of various forms where the estimates produced are conditional on what is assumed about $\mathbf{W}_N$, given that in practice the true $\mathbf{W}_N$ is likely to be unknown.

References

[1] Akerlof, George A. (1997). 'Social Distance and Social Decisions', *Econometrica*, 65, pp. 1005–1028.

[2] Anderson, T.W. and Hsiao, C. (1981). 'Estimation of dynamic models with error components', *Journal of the American Statistical Association*, Vol. 76, pp. 598–606.

[3] Anderson, T. W. and Hsiao, C. (1982). 'Formulation and estimation of dynamic models using panel data', *Journal of Econometrics*, Vol. 18, pp. 47–82.

[4] Arellano, M. and Bond, S. (1991). 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment', *Review of Economic Studies*, Vol. 58, pp. 277–297.

[5] Baltagi, B.H., Fingleton, B. and Pirotte, A. (2013). 'Estimating and Forecasting with a Dynamic Spatial Panel Model', *Oxford Bulletin of Economics and Statistics,* Vol. 76, pp. 112–138.

[6] Baltagi, B., B. Fingleton, and A. Pirotte (2014). 'Spatial Lag Models with Nested Random Effects: an IV Procedure with Application to English House Prices', *Journal of Urban Economics*, Vol. 80, pp. 76–86.

[7] Bauman, Karl, and Lynn Fisher (1986). 'On the Measurement of Friend Behavior in Research on Friend Influence and Selection: Findings from Longitudinal Studies of Adolescent Smoking and Drinking,' *Journal of Youth and Adolescence*, Vol. 15, pp. 345–353.

[8] Bénabou, Roland (1993). 'Workings of a City: Location, Education, and Production,' *Quarterly Journal of Economics*,' CVIII, pp. 619–652.

[9] Bernheim, Douglas (1994). 'A Theory of Conformity,' *Journal of Political Economy*, Vol. 102, pp. 841–877.

[10] Bond, S. (2002). 'Dynamic panel data models: a guide to micro data methods and practice', *Portuguese Economic Journal*, Vol. 1, pp. 141-162.

[11] Bouayad-Agha, S. and Védrine, L. (2010). 'Estimation strategies for a spatial dynamic panel using GMM. A new approach to the convergence issue of European regions', *Spatial Economic Analysis*, Vol. 5, pp. 205-227.

[12] Brock, William A. and Stephen N. Durlauf (2001). 'Interactions-Based Models,' in J. J. Heckman, and E. Leamer (eds), Handbook of Econometrics, Vol. 5. Amsterdam: Elsevier Science, pp. 3297-3380.

[13] Brown, B. (1990). 'Peer Groups and Peer Cultures,' in S. Feldman, and G. Elliott (eds.), At the Threshold: the Developing Adolescent, Cambridge MA: Harvard University Press, pp. 171-196.

[14] Brown, Bradford, Clasen, D. and Eicher, S. (1986). 'Perceptions of Peer Pressure, Peer Conformity Dispositions and Self-Reported Behavior Among Adolescents,' *Developmental Psychology*, Vol. 22, pp. 521–530.

[15] Corrado, L. and Fingleton, B. (2012). 'Where is the economics in spatial econometrics?,' *Journal of Regional Science*, Vol. 52, pp. 210-239.

[16] Debarsy, N. (2012). 'The Mundlak Approach in the Spatial Durbin Panel Data Model,' *Spatial Economic Analysis*, 7(1), pp. 109–131.

[17] Debarsy, N., Ertur, C. and LeSage, J. P. (2012). 'Interpreting dynamic space-time panel data models,' Statistical Methodology, 9, pp. 158-171.

[18] Ebbes, P., Böckenholt, U. and Wedel, M. (2004). 'Regressor and Random-Effects Dependencies in Multilevel models,' *Statistica Neerlandica*, Vol. 58(2), pp. 161–178.

[19] Elhorst. J. P. (2014). 'Spatial Econometrics : from Cross-Sectional Data to Spatial Panels,' Springer-Verlag, Heidelberg.

[20] Gibbons, Stephen, and Henry Overman (2010). 'Mostly Pointless Spatial Econometrics?,' *Journal of Regional Science*, 52(2), pp. 172–191.

[21] Goldstein, H. (1986). 'Multilevel Mixed Linear Model Analysis Using Iterative Generalised Least Squares', *Biometrika*, 73, pp. 46–56.

[22] Goldstein, H. (1998). 'Multilevel Models for Analysing Social Data,' in Encyclopaedia of Social Research Methods. Newbury Park CA: Sage Publications.

[23] Jones, Andrew (1994). 'Health, Addiction, Social Interaction, and the Decision to Quit Smoking,' *Journal of Health Economics*, Vol. 13, pp. 93-110.

[24] Kandori, Michihiro (1992). 'Social Norms and Community Enforcement,' *Review of Economic Studies*, Vol. 59, pp. 63-80.

[25] Kapoor, M., Kelejian, H.H. and Prucha, I.R. (2007). 'Panel data models with spatially correlated error components,' *Journal of Econometrics*, Vol. 140, pp. 97-130.

[26] Kelejian, H. H. and Prucha, I. R. (1999). 'A generalized moments estimator for the autoregressive parameter in a spatial model', *International Economic Review*, Vol. 40, pp. 509–533.

[27] Krosnick, Jon A., and Charles M. Judd (1982). 'Transitions in Social Influence in Adolescence: Who Induces Cigarette Smoking,' *Developmental Psychology*, Vol. 81, pp. 359-368.

[28] Lam C. and Pedro CL Souza (2015). One-Step Regularized Spatial Weight Matrix and Fixed Effects Estimation with Instrumental Variables, mimeo.

[29] LeSage, J. P. and Pace, R. K. (2009). Introduction to Spatial Econometrics, Chapman & Hall/CRC Press, Boca Raton.

[30] Longford, Nicholas T. (1993). Random Coefficient Models. London: Claredon Press.

[31] Manski, Charles F. (1993). 'Identification of Endogenous Social Effects: The Reflection Problem,' *Review of Economic Studies*, Vol. 60, pp. 531-542.

[32] Manski, Charles F. (2000). 'Economic Analysis of Social Interactions,' *Journal of Economic Perspectives*, Vol. 3, pp. 115-136.

[33] Mundlak, Y. (1978). 'On the Pooling of Time Series and Cross-sectional Data,' *Econometrica*, Vol. 46, pp. 69-86.

[34] Rice, N., Jones A.M., and Goldstein, H. (1998). 'Multilevel Models where the Random Effects are Correlated with the Fixed Predictors: A Conditioned Iterative Generalised Least Squares Estimator (CIGLS),' CHETechnical Paper 10 University of York.

Table 1. Mean, bias and RMSE of dynamic spatial panel parameter estimates.

True **W** matrix is $j = 5$ ahead and $j = 5$ behind, 1000 Monte Carlo replications.

	ASSUMED	$j=2$			$j=5$			$j=10$		
		γ	ρ_1	β	γ	ρ_1	β	γ	ρ_1	β
	Mean	0.2046	0.4948	1.0411	0.1958	0.7156	1.0009	0.2089	0.852	1.0172
	Mean bias	0.0046	-0.2052	0.0411	-0.0042	0.0156	0.0009	0.0089	0.152	0.0172
	RMSE	0.0117	0.5061	1.0412	0.0089	0.2848	1.001	0.0135	0.1613	1.0172
TRUE $j=5$		ρ_2	σ^2_μ	σ^2_v	ρ_2	σ^2_μ	σ^2_v	ρ_2	σ^2_μ	σ^2_v
	Mean	0.3614	1.2925	0.4452	0.3061	0.795	0.2302	0.3754	1.2137	0.5361
	Mean bias	-0.0386	0.4925	0.2452	-0.0939	-0.005	0.0302	-0.0246	0.4137	0.3361
	RMSE	0.1613	0.546	0.2484	0.2246	0.1217	0.0344	0.3481	0.5057	0.3406

True **W** matrix is $j = 10$ ahead and $j = 10$ behind, 1000 Monte Carlo replications

	ASSUMED	$j=2$			$j=5$			$j=10$		
		γ	ρ_1	β	γ	ρ_1	β	γ	ρ_1	β
	Mean	0.2130	0.3572	1.0470	0.2159	0.5374	1.0177	0.1963	0.7202	1.0004
	Mean bias	0.0130	-0.3428	0.0470	0.0159	-0.1626	0.0177	-0.0037	0.0202	0.0004
	RMSE	0.0172	0.6446	1.0471	0.0183	0.4648	1.0178	0.009	0.2805	1.0004
TRUE $j=10$		ρ_2	σ^2_μ	σ^2_v	ρ_2	σ^2_μ	σ^2_v	ρ_2	σ^2_μ	σ^2_v
	Mean	0.3703	1.1610	0.4436	0.2819	0.9144	0.3039	0.2291	0.7942	0.2213
	Mean bias	-0.0297	0.3610	0.2436	-0.1181	0.1144	0.1039	-0.1709	-0.0058	0.0213
	RMSE	0.1841	0.4244	0.2525	0.2843	0.1907	0.1110	0.3560	0.1226	0.0355

True **W** matrix is $j = 2$ ahead and $j = 2$ behind, 1000 Monte Carlo replications.

	ASSUMED	$j=2$			$j=5$			$j=10$		
		γ	ρ_1	β	γ	ρ_1	β	γ	ρ_1	β
	Mean	0.1958	0.7102	1.0012	0.2454	0.7477	1.0537	0.2894	0.7776	1.0967
	Mean bias	-0.0042	0.0102	0.0012	0.0454	0.0477	0.0537	0.0894	0.0776	0.0967
	RMSE	0.0083	0.2900	1.0013	0.0483	0.2531	1.0538	0.0915	0.2255	1.0968
TRUE $j=2$		ρ_2	σ^2_μ	σ^2_v	ρ_2	σ^2_μ	σ^2_v	ρ_2	σ^2_μ	σ^2_v
	Mean	0.3589	0.7910	0.2409	0.1115	2.8426	0.9776	-0.1770	5.5118	0.9900
	Mean bias	-0.0411	-0.009	0.0409	-0.2885	2.0426	0.7776	-0.5770	4.7118	0.7900
	RMSE	0.1274	0.1226	0.0414	0.4052	2.1559	0.7780	0.7154	5.0434	0.7900

Table 2. Short and Long-run direct, indirect and total effects using mean dynamic spatial panel parameter estimates.

		True **W** matrix is $j = 5$ ahead and $j = 5$ behind								
	ASSUMED	$j = 2$			$j = 5$			$j = 10$		
		direct	indirect	total	direct	indirect	total	direct	indirect	total
TRUE	Short-run	1.131	0.9295	2.061	1.119	2.401	3.520	1.157	5.715	6.872
$j = 5$	Long-run	1.523	1.941	3.464	1.673	9.623	11.3	9.37	-26.07	-16.7

		True **W** matrix is $j = 2$ ahead and $j = 2$ behind								
	ASSUMED	$j = 2$			$j = 5$			$j = 10$		
		direct	indirect	total	direct	indirect	total	direct	indirect	total
TRUE	Short-run	1.254	2.201	3.455	1.201	2.975	4.176	1.192	3.740	4.932
$j = 2$	Long-run	2.116	8.54	10.66	4.187	148.8	153	-0.97	-15.38	-16.35

		True **W** matrix is $j = 10$ ahead and $j = 10$ behind								
	ASSUMED	$j = 2$			$j = 5$			$j = 10$		
		direct	indirect	total	direct	indirect	total	direct	indirect	total
TRUE	Short-run	1.088	0.5405	1.629	1.066	1.134	2.20	1.064	2.512	3.576
$j = 10$	Long-run	1.423	1.013	2.436	1.429	2.697	4.126	1.485	10.5	11.99

Table 3. Mean, bias and RMSE of dynamic spatial panel parameter estimates. True **W** matrix is either sparse or dense, 1000 Monte Carlo replications.

	ASSUMED	SPARSE			DENSE		
		γ	ρ_1	β	γ	ρ_1	β
	Mean	0.1983	0.7054	1.0015	0.2291	0.7406	1.0070
	Mean bias	-0.0017	0.0054	0.0015	0.0291	0.0406	0.0070
	RMSE	0.0086	0.2950	1.0015	0.0309	0.2604	1.0071
TRUE SPARSE		ρ_2	σ^2_μ	σ^2_v	ρ_2	σ^2_μ	σ^2_v
	Mean	0.3788	0.7917	0.2302	0.1434	1.2044	0.7515
	Mean bias	-0.0212	-0.0083	0.0302	-0.2566	0.4044	0.5515
	RMSE	0.1727	0.1213	0.0325	0.3562	0.4797	0.5539

	ASSUMED	SPARSE			DENSE		
		γ	ρ_1	β	γ	ρ_1	β
	Mean	0.1933	0.4223	0.9799	0.1985	0.7061	1.0012
	Mean bias	-0.0067	-0.2777	-0.0201	-0.0015	0.0061	0.0012
	RMSE	0.0117	0.5788	0.9800	0.0086	0.2945	1.0012
TRUE DENSE		ρ_2	σ^2_μ	σ^2_v	ρ_2	σ^2_μ	σ^2_v
	Mean	0.2969	1.1989	0.5202	0.3734	0.7919	0.2217
	Mean bias	-0.1031	0.3989	0.3202	-0.0266	-0.0081	0.0217
	RMSE	0.2577	0.4681	0.3234	0.2334	0.1208	0.029

Table 4. Short and Long-run direct, indirect and total effects using mean dynamic spatial panel parameter estimates. True **W** matrix either sparse or dense.

	ASSUMED	SPARSE			DENSE		
		Direct	Indirect	Total	Direct	Indirect	Total
TRUE	Short-run	1.018	2.381	3.400	1.028	2.854	3.882
SPARSE	Long-run	1.329	9.072	10.4	1.612	31.56	33.18

	ASSUMED	SPARSE			DENSE		
		Direct	Indirect	Total	Direct	Indirect	Total
TRUE	Short-run	0.983	0.7133	1.696	1.018	2.388	3.407
DENSE	Long-run	1.222	1.328	2.549	1.331	9.169	10.5

Table 5. Mean, bias and RMSE of SAR-RIGLS Multilevel Panel data parameter estimates. True **W** matrix is either sparse or dense, 1000 Monte Carlo replications.

	ASSUMED	SPARSE			DENSE		
		ρ_1	β		ρ_1	β	
	Mean	0.7000	0.9986		0.4083	0.9300	
	Mean bias	0.0006	-0.0013		-0.2916	-0.0609	
	RMSE	0.0004	0.0318		0.2916	0.0649	
TRUE SPARSE							
		σ^2_α	σ^2_μ	σ^2_v	σ^2_α	σ^2_μ	σ^2_v
	Mean	0.5131	0.4896	0.5133	0.4891	0.5017	0.4923
	S.E.	0.0414	0.0142	0.0048	0.0446	0.0462	0.0707

	ASSUMED	SPARSE			DENSE		
		ρ_1	β		ρ_1	β	
	Mean	0.6438	1.0574		0.6999	1.0022	
	Mean bias	-0.0561	0.0574		-0.0001	0.0022	
	RMSE	0.0561	0.0652		0.0018	0.0342	
TRUE DENSE							
		σ^2_α	σ^2_μ	σ^2_v	σ^2_α	σ^2_μ	σ^2_v
	Mean	0.4891	0.5016	0.4926	0.5133	0.4896	0.5123
	S.E.	0.0560	0.0617	0.1713	0.0414	0.0142	0.0048

Table 6. Mean, bias and RMSE of SAR-RIGLS Multivariate Multilevel Panel data parameter estimates. Regressors and error terms are orthogonal ($\boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = \boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_3 = 0$). True **W** matrix is is either sparse or dense. 1000 Monte Carlo replications.

	ASSUMED	SPARSE			DENSE		
		ρ_1	β		ρ_1	β	
	Mean	0.7001	1.0005		0.2747	0.8866	
	Mean bias	0.0001	0.0005		-0.4252	-0.1133	
	RMSE	0.0019	0.0335		0.4252	0.1134	
TRUE SPARSE							
		σ^2_α	σ^2_μ	σ^2_v	σ^2_α	σ^2_μ	σ^2_v
	Mean	0.5133	0.4896	0.5126	0.4891	0.5017	0.4933
	S.E.	0.0375	0.01425	0.0004	0.0419	0.0406	0.0810

	ASSUMED	SPARSE			DENSE		
		ρ_1	β		ρ_1	β	
	Mean	0.6317	1.0798		0.7000	0.9995	
	Mean bias	-0.0682	0.0798		0.0001	-0.0004	
	RMSE	0.0684	0.0826		0.0102	0.0318	
TRUE DENSE							
		σ^2_α	σ^2_μ	σ^2_v	σ^2_α	σ^2_μ	σ^2_v
	Mean	0.4891	0.5017	0.4925	0.5133	0.4896	0.5126
	S.E.	0.0566	0.0803	0.1695	0.0375	0.0142	0.0050

Table 7. Mean, bias and RMSE of SAR-RIGLS Multivariate Multilevel Panel data parameter estimates. Regressors and error terms are correlated ($\boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_3 \neq \boldsymbol{\gamma}_2 \neq \boldsymbol{\gamma}_3 \neq 0$). True $\mathbf{W}$ matrix is is either sparse or dense. 1000 Monte Carlo replications.

	ASSUMED	SPARSE			DENSE		
		ρ_1	β		ρ_1	β	
	Mean	0.6993	0.999		0.2577	0.9215	
	Mean bias	-0.0006	-0.0001		-0.4422	-0.0784	
	RMSE	0.0213	0.0324		0.4422	0.0793	
TRUE SPARSE							
		σ^2_α	σ^2_μ	σ^2_v	σ^2_α	σ^2_μ	σ^2_v
	Mean	0.4891	0.5017	0.4930	0.4891	0.5017	0.4930
	S.E.	0.0330	0.0141	0.0050	0.0358	0.0396	0.0853
	ASSUMED	SPARSE			DENSE		
		ρ_1	β		ρ_1	β	
	Mean	0.6435	1.0834		0.7008	1.0006	
	Mean bias	-0.0564	0.0834		0.0008	0.0006	
	RMSE	0.0679	0.0841		0.0099	0.0340	
TRUE DENSE							
		σ^2_α	σ^2_μ	σ^2_v	σ^2_α	σ^2_μ	σ^2_v
	Mean	0.4891	0.5017	0.4930	0.5133	0.4896	0.5133
	S.E.	0.0637	0.0492	0.1700	0.0330	0.0141	0.0050

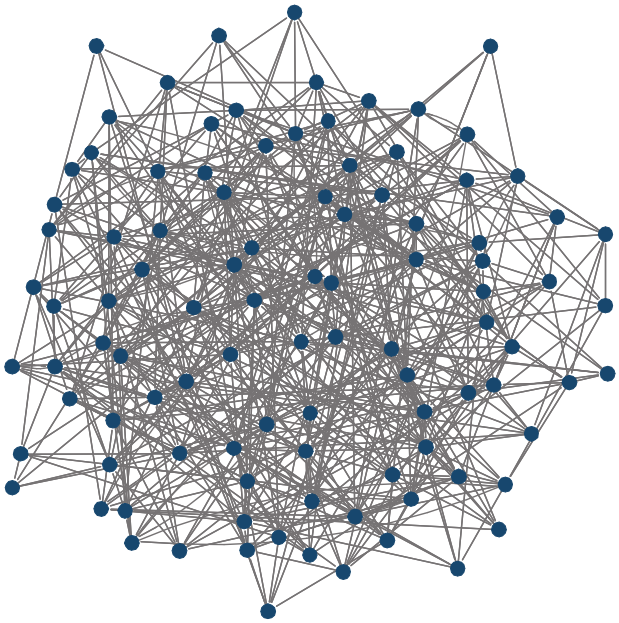


Figure. Dense Network

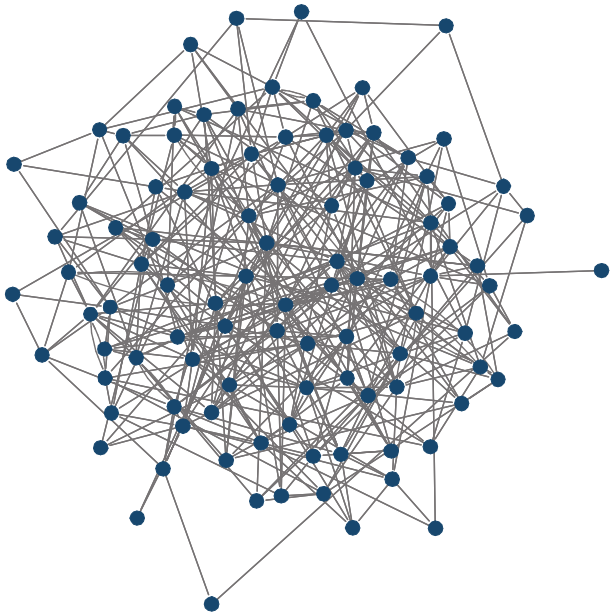
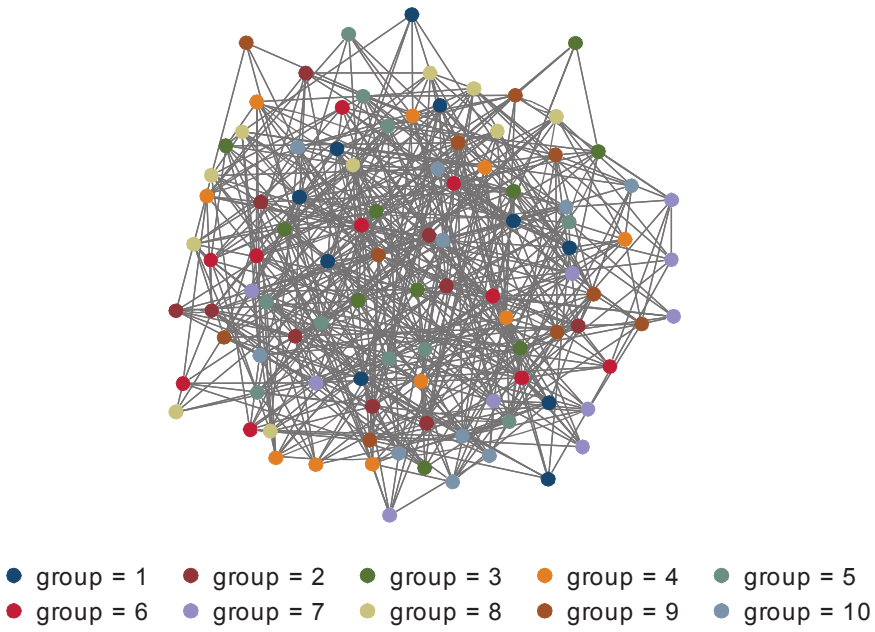
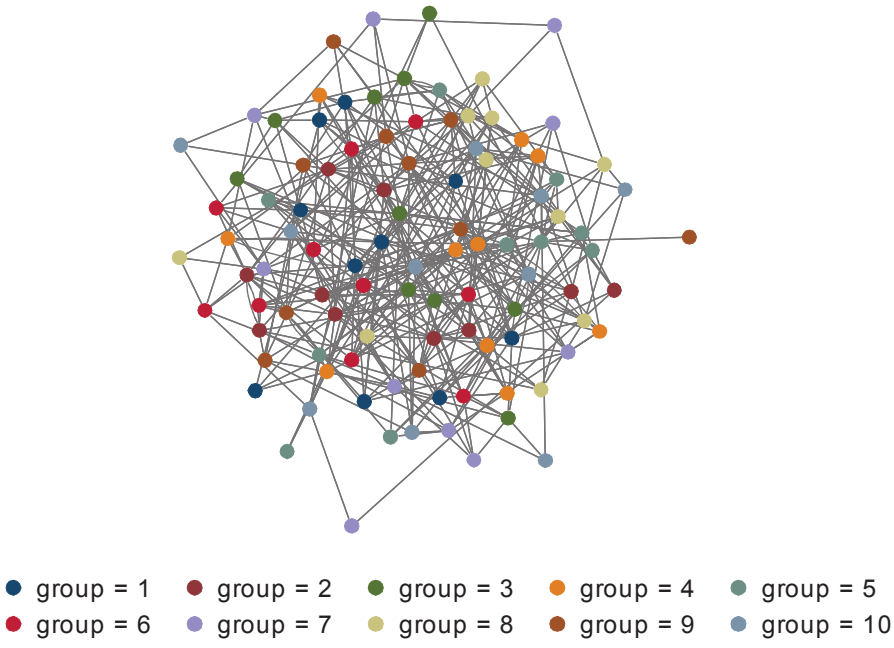


Figure 2. Sparse Network



Dense Network with 100 nodes and 10 groups.



Sparse Network with 100 nodes and 10 groups.

RECENT PUBLICATIONS BY *CEIS Tor Vergata*

Fishing the Corporate Social Responsibility Risk Factors
Leonardo Becchetti, Rocco Ciciretti and Ambrogio D'Alò
CEIS Research Paper, 368 , February 2016

Modeling public health care expenditure using patient level data: Empirical evidence from Italy
Vincenzo Atella, Federico Belotti, Valentina Conti, Claudio Cricelli, Joanna Kopinska and Andrea Piano Mortari
CEIS Research Paper, 367 , February 2016

Tax Reforms and the Underground Economy: A Simulation-Based Analysis
Barbara Annicchiarico and Claudio Cesaroni
CEIS Research Paper, 366 , February 2016

The Effect of Survey Design on Extreme Response Style: Rating Job Satisfaction
Luisa Corrado and Majlinda Joxhe
CEIS Research Paper, 365 , February 2016

Till Mortgage Do Us Part: Refinancing Costs and Mortgage Shopping
Marianna Brunetti, Rocco Ciciretti and Ljubica Djordjevic
CEIS Research Paper, 364 , February 2016

Other regarding preferences and reciprocity: insights from experimental findings and satisfaction data
Leonardo Becchetti, Vittorio Pelligra and Serena F. Taurino
CEIS Research Paper, 363 , February 2016

The Model Confidence Set package for R
Mauro Bernardi and Leopoldo Catania
CEIS Research Paper, 362 ,November 2015

The Effect of Discretion on Procurement Performance
Decio Coviello, Andrea Guglielmo and Giancarlo Spagnolo
CEIS Research Paper, 361 ,November 2015

Current Account and Real Effective Exchange Rate misalignments in Central Eastern EU Countries: an update using the Macroeconomic Balance approach
Mariarosaria Comunale
CEIS Research Paper, 360 ,November 2015

The Vote With the Wallet as a Multiplayer Prisoner's Dilemma
Leonardo Becchetti and Francesco Salustri
CEIS Research Paper, 359 ,November 2015

DISTRIBUTION

Our publications are available online at www.ceistorvergata.it

DISCLAIMER

The opinions expressed in these publications are the authors' alone and therefore do not necessarily reflect the opinions of the supporters, staff, or boards of CEIS Tor Vergata.

COPYRIGHT

MEDIA INQUIRIES AND INFORMATION

For media inquiries, please contact Barbara Piazzi at +39 06 72595652/01 or by e-mail at ufficiostampa@ceis.uniroma2.it. Our web site, www.ceistorvergata.it, contains more information about Center's events, publications, and staff.

DEVELOPMENT AND SUPPORT

For information about contributing to CEIS Tor Vergata, please contact Carmen Tata at +39 06 72595615/01 or by e-mail at segr.ceis@economia.uniroma2.it