Residual Spatial Correlation in Two-Way Error Panel Data Models

with An Application of a Spatial Panel Data Model to Municipal Unemployment in Southern Sweden

Adam Flöhr

Master's Thesis in Statistics, STAN40
Department of Statistics
Lund University
Fall 2018
Supervisor: Krzysztof Nowicki

Abstract

This thesis examines the spatial autocorrelation in residuals of two-way error panel data models. Three types of models are examined: the standard linear panel data model, the dynamic panel data model, and the spatial lag panel data model. A known theoretical result for the linear model, that the within estimator applied to independent observations results in a spatial correlation in the residuals which is proportional to the inverse of the number of observed individual units, is supported in a Monte Carlo study. Similar Monte Carlo results are shown for the dynamic and spatial models. The Monte Carlo study shows the effect of residual correlation on the maximum likelihood estimation of the spatial model and on residual tests for spatial correlation. Randomization tests for spatial correlation are formulated and their properties are evaluated. The results suggest that a randomization test for local spatial autocorrelation is the most suitable test for samples large in number of individual regions and small in number of time points.

The model estimation methods and tests for residual spatial autocorrelation are applied in an empirical examination of regional unemployment in Southern Sweden. The study shows that the spatial lag model is sensitive to choice of spatial weight matrix and indicates the presence of a spatial structure which is not fully captured by the applied models and weight matrices.

Sammanfattning

Denna uppsats behandlar förekomsten av rumslig autokorrelation i residualer till paneldatamodeller med tvåvägsfel. Tre modelltyper undersöks: den linjära paneldatamodellen, en dynamisk paneldatamodell och en spatial paneldatamodell. Ett känt resultat för den linjära modellen, att minsta kvadrat-skattaren tillämpad på oberoende observationer resulterar i en rumslig korrelation i residualserien vars storlek är proportionell till inversen av antalet observerade individer, undersöks i en Monte Carlo-studie. Liknande simuleringsresultat ges för de dynamiska och rumsliga modellerna. Monte Carlo-studien visar effekten av residualkorrelation på maximum likelihood-skattningen av den rumsliga modellen och på hypotestester för rumslig korrelation. Randomiseringstest för rumslig korrelation formuleras och testas. Resultaten pekar mot att ett randomiseringstest för lokal rumslig autokorrelation är det lämpligaste testet för stickprov över ett stort antal regioner och ett lågt antal tidpunkter.

De tre typerna av paneldatamodeller skattas i en empirisk undersökning av regional arbetslöshet i södra Sverige (Skåne). Tester för rumslig autokorrelation tillämpas på modellernas residualer. Studien visar att den rumsliga paneldatamodellen är känslig för val av viktmatris och pekar på förekomsten av en rumslig struktur som inte helt fångas av de tillämpade modellerna och viktmatriserna.

Contents

1 Introduction

Recent years have seen growing interest in various models which extend the *linear panel data model*. In particular, *dynamic panel data models* have been used to model temporal autocorrelation, i.e. correlation between observations of the same individual unit at different time points, and *spatial panel data models* have been used to model spatial autocorrelation, i.e. correlation between different individuals (or spatial units) at the same time point. This paper concerns these three types of models - the linear, dynamic and spatial panel data models - with special focus on the structure of the residuals of each model. This introduction gives a brief overview of the subject and presents the purpose of the paper.

Panel data is defined as data which covers some variables for a set of individuals over a series of time points. It differs from cross-sectional data, i.e. data which covers a fixed time point, in that each variable is a time series for each individual. Spatiotemporal data is a specific form of panel in which the individuals are spatial units with some geographical connection. A typical example of the latter is economical data for a set of regions over a number of years.

The basic model for a panel is the linear panel data model, where an explained variable is assumed to depend linearly on a set of explanatory variables and an error term. The error term can be further decomposed to account for different sources of variance. The focus of this paper is on the model with a *two-way fixed effects error term*, in which the error depends on both individual and time, and these individual and time components are considered fixed parameters, rather than the outcome of a random process. This two-way error linear panel data model may be formulated as a general linear model and is typically estimated using ordinary least squares (OLS).

One property of panel data, in relation to a single time series or cross-sectional data, is that it allows for testing for remaining temporal and spatial autocorrelation by analyzing the residuals of the estimated model. A simple analysis of spatial autocorrelation could for example begin with an estimation of the empirical correlation between spatial units. However, there is an obstacle to testing for spatial correlation in that the OLS estimation results in a correlation structure in the residuals, so that even in a data set where there is no temporal or spatial correlation, the residuals of the panel data model will be correlated. This issue of residual correlation is present in any model estimated by OLS, but is especially interesting in the two-way fixed effects model for two reasons: firstly, the correlation is in both time and space; and secondly, the correlation can be expressed as a function of the number of individuals and time points in a simple closed form. Wooldridge (2002) addresses the residual correlation structure in the one-way error model by suggesting modifications to existing tests of autocorrelation in time. Mao (2015) analyzes the residuals of the two-way error model and proposes a transformation which back-transforms the residual correlation structure, making it possible to apply standard tests on the transformed residuals.

The *dynamic panel data model* is formed by expanding the linear panel data model with a lagged explained variable, so that the value of the explained variable at a specific time point is dependent on the explained variable in a previous time point, as well as a set of explanatory variables. The addition of a lagged variable may result in serious estimation problems if the model is estimated using OLS. Neyman and Scott (1948) analyze the problem of inconsistency in models where the number of estimated parameters increases with the sample size (the incidental parameter problem) and Nickell (1981) shows that the standard OLS estimator gives significantly biased results in panels with a small number of time periods (the Nickell bias). A number of alternative estimation methods have been suggested. Arellano and Bond (1991) developed a generalized method of moment estimator, an approach which has seen further theoretical development (Arellano and Bover, 1995) and much practical use. There are now a number of different GMM estimators available (Baltagi, 2013). Later studies point to some drawbacks of the GMM estimators, showing that the estimator is inconsistent if the dynamics is misspecified or if the data set is small (Alvarez and Arellano, 2003). In more recent years, there has been development in the factor analytical approach of Bai (2013).

The *spatial panel data model* (specifically the *spatial lag panel data model*) is the result of expanding the linear panel data model with a spatial term given by a weighted mean of the explained variable of the other individuals. The weights are chosen to capture the spatial relation between individuals, so that a close spatial connection would correspond to a higher weight. As for the dynamic model, OLS estimation applied to the spatial model would yield inconsistent results and there has therefore been development in alternative estimators through applications of two-stage least squares and maximum-likelihood methods. Baltagi and Liu (2011) give a two-stage least squares estimator for the spatial model with a random error. Mutl and Pfaffermayr (2011) present an instrumental variables estimator for the fixed effects model with spatial lag. Baltagi and Liu (2016) derive a two-stage least squares estimator for the same model. Anselin and Hudak (1992) form a maximum-likelihood estimator for the fixed effects model in the cross-sectional case. The cross-sectional estimator is extended to the panel data case in Elhorst (2003b) and further refined using a bias-correction in Lee and Yu (2010).

The ML estimator of the spatial panel data model applies the OLS estimator as part of the estimation procedure and it can therefore be expected that the residuals of the model have a similar correlation structure as the residuals of the linear model. Since there is an explicit spatial component in the model, it is furthermore possible that the correlation structure in the two-way error model influences the estimated spatial parameter.

The theoretical part of this paper concerns the three panel data models - the linear, dynamic and spatial - with special focus on the issue of residual autocorrelation in models with a two-way error component. The empirical part of the paper gives an application of the models to the case of regional unemployment in Southern Sweden.

Unemployment is one of several macroeconomic variables which has clear temporal and spatial components. The case of Southern Sweden, with municipalities as regions, is illustrated in Figure 1. The map shows a spatial pattern

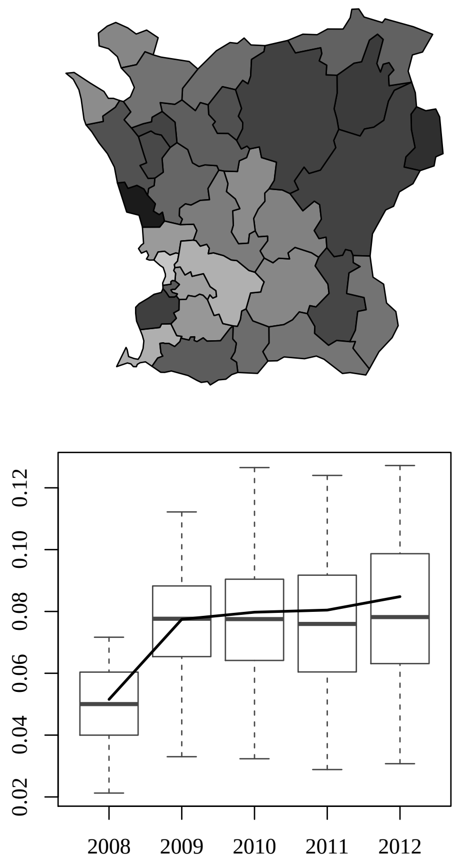

Figure 1: (Top) Map of unemployment levels for the south of Sweden, 2012. (Bottom) Box plot time series of regional unemployment for the south of Sweden 2008-2012, and unemployment in the total region (black line).

with higher unemployment in the eastern parts, and the time series shows a dependency between years.

There are multiple studies on regional unemployment and a wide range of statistical methods have been applied. Elhorst (2003a) gives a survey of studies published up until 1997 and links regional unemployment to differences in demography, income, education, and other factors. Gilmartin and Korobilis (2012) give an example of a panel data model for regional unemployment. The spatial dimension is treated by a data-based clustering method which is used to partition the regions into smaller groups, thus reducing the effect of spatial autocorrelation. Cracolici et al. (2009) apply a model for regional unemployment in Italy. The model is, through the use of a spatially weighted explained variable, closely connected to the spatial panel data model but is only applied to cross-sectional data. Lottmann (2012) applies a spatial dynamic panel data model, i.e. a model which incorporates both a time-lagged term and a spatially lagged term, to regional unemployment in Germany.

This paper seeks to analyze and apply the panel data models. The purpose of the paper is to investigate the spatial autocorrelation of the residuals of the linear, dynamic, and spatial panel data models with two-way fixed effects error, and to apply the models in an analysis of regional unemployment in Southern Sweden.

The theoretical investigation of the residual correlation structure is carried out in a Monte Carlo study where panel data is simulated, the models are estimated, and the properties of the residuals are analyzed. Here, the focus lies on investigating the existence and magnitude of residual correlation in the dynamic and spatial models, on the properties of parametric tests and randomization tests for spatial autocorrelation, and on the connection between the residual correlation structure and the spatial parameter in the spatial model.

The empirical application uses data on 33 municipalities in the south of Sweden (the *Scania* region) for 2008-2012. The explained variable is given as the proportion of unemployed in each municipality and the explanatory variables are chosen based on previous studies. A set of panel data models is estimated and the outcomes are compared using measures for goodness-of-fit and tests for remaining temporal or spatial correlation.

The paper is structured as follows. Chapter 2 describes the panel data models, the residual correlation structure, and hypothesis tests. Chapter 3 presents the results of the Monte Carlo investigation of residual correlation and its effects on residual tests and parameter estimation. Chapter 4 describes the data and presents exploratory analysis for the study of regional unemployment. Chapter 5 presents the estimated models and test results. Chapter 6 gives a summary and discussion of method and results.

2 Method

2.1 Introduction

This chapter describes the panel data models used in the paper and the problems concerning residual autocorrelation. Sections 2.2 and 2.3 define the linear panel data model and the dynamic panel data model. Section 2.4 describes the spatial weight matrix and the spatial lag panel data model. Section 2.5 outlines parameter estimation methods for the different models. Section 2.6 describes the residual autocorrelation resulting from OLS estimation of the linear two-way error panel data model and presents methods to test for spatial autocorrelation. Section 2.7 presents tests of additional model assumptions. Section 2.8 discusses methods to compare the results of the estimated models through measures of goodness-of-fit.

The presentation of models and estimation methods (sections 2.2 to 2.5) are based on Baltagi (2013) if not stated otherwise.

Figures, data handling, and estimations are done in the R core package (R Core Team, 2013) if not stated otherwise.

2.2 The Linear Panel Data Model

Panel data is characterized as data having an individual component and a time component, so that data covers N individuals for T time points. The total number of observations is given by NT, which will be denoted by M where convenient. The general form of a linear panel data model with P explanatory variables is given by

$$y_{it} = \alpha + \sum_{p=1}^{P} \beta_p x_{pit} + u_{it} \qquad (i = 1, ..., N; t = 1, ..., T), \tag{1}$$

where y_{it} is the explained variable for individual i at time point t, α is a constant, x_{pit} is the value of the pth explanatory variable for the ith individual at time t, β_p for $p = 1, ..., P$ is a parameter, and u_{it} is an error term.

The form of the error term will results in a specific panel data model. This paper focuses on the *two-way error model*, in which the error term contains both a time component and an individuals component. The error term is written

$$u_{it} = \mu_i + \lambda_t + v_{it} \qquad (i = 1, ..., N; t = 1, ..., T), \tag{2}$$

where μ_i is an individual-specific effect, λ_t is a time-specific effect, and v_{it} is the remaining error. Hence, the full two-way error linear panel data model is given by

$$y_{it} = \alpha + \sum_{p=1}^{P} \beta_p x_{pit} + \mu_i + \lambda_t + v_{it} \qquad (i = 1, ..., N; t = 1, ..., T). \tag{3}$$

The remaining errors v_{it} are assumed to be identically and independently drawn observations from some distribution, i.e. $v_{it} \in IID(0, \sigma_v^2)$ for $i = 1, ..., N$ and

$t = 1, ..., T$. An equivalent definition of the two-way error model is thereby to specify $u_{it} \in IID(\mu_i + \lambda_t, \sigma_v^2)$ in (1). For the constant α, the individual effects, and the time effects to be uniquely determined it is necessary to set conditions on the parameters, usually that the N individual effects and the T time effects both sum to zero. Uniqueness also fails if some explanatory variable x_p is constant over all individuals or all time periods.

Setting $\mu_i = 0$ for all i results in the *one-way error model with time effects*. Similarly, setting $\lambda_t = 0$ for all t gives the *one-way error model with individual effects*.

The two-way error linear model is given in matrix form by

$$\boldsymbol{y} = \boldsymbol{\alpha} + \boldsymbol{X\beta} + \boldsymbol{u} \tag{4}$$

with the error term

$$\boldsymbol{u} = \boldsymbol{Z}_\mu \boldsymbol{\mu} + \boldsymbol{Z}_\lambda \boldsymbol{\lambda} + \boldsymbol{v}. \tag{5}$$

Here, $\boldsymbol{y}$ is the explained variable given as a column vector of length M (an *M-column*), $\boldsymbol{y} = (y_{11}, ..., y_{N1}, y_{12}, ..., y_{N2}, ..., y_{1T}, ..., y_{NT})'$, $\boldsymbol{\alpha}$ is an M-column in which all elements are α, $\boldsymbol{X}$ is an $M \times P$ matrix containing the explanatory variables, $\boldsymbol{\beta}$ is a parameter vector of length P, $\boldsymbol{Z}_\mu$ is an $M \times N$ matrix containing dummy variables for individuals, $\boldsymbol{\mu}$ is a column vector of length N defined by $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)'$, $\boldsymbol{Z}_\lambda$ is an $M \times T$ matrix containing dummy variables for time periods, $\boldsymbol{\lambda}$ is a column vector of length T defined by $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_T)'$, and $\boldsymbol{v} = (v_{11}, ..., v_{NT})'$.

2.3 The Dynamic Panel Data Model

The general linear panel data model can be extended to incorporate a time-lagged explained variable in the explanatory variables. The resulting *dynamic panel data model* is defined by

$$y_{it} = \delta y_{i,t-1} + \sum_{p=1}^{P} \beta_p x_{pit} + u_{it} \qquad (i = 1, ..., N; t = 2, ..., T), \tag{6}$$

where δ is a parameter and other terms are defined as in (1). The two-way error dynamic panel data model is defined by setting the error term u_{it} equal to (2), including the IID assumption. The full two-way error model is hence given by

$$y_{it} = \delta y_{i,t-1} + \sum_{p=1}^{P} \beta_p x_{pit} + \mu_i + \lambda_t + v_{it} \qquad (i = 1, ..., N; t = 2, ..., T). \tag{7}$$

The model is written in matrix form as

$$\begin{aligned} \boldsymbol{y} &= \delta \boldsymbol{y}^{(-1)} + \boldsymbol{X\beta} + \boldsymbol{u}, \\ \boldsymbol{u} &= \boldsymbol{Z}_\mu \boldsymbol{\mu} + \boldsymbol{Z}_\lambda \boldsymbol{\lambda} + \boldsymbol{v}. \end{aligned} \tag{8}$$

where the explained variable is a vector of length $N(T-1)$ given by $\boldsymbol{y} = (y_{12}, ..., y_{N2}, y_{13}, ..., y_{N3}, ..., y_{1T}, ..., y_{NT})'$, δ is a scalar parameter, $\boldsymbol{y}^{(-1)}$ is the lagged explained variable as a vector of length $N(T-1)$ given by $\boldsymbol{y}^{(-1)} = (y_{11}, ..., y_{N1}, y_{12}, ..., y_{N2}, ..., y_{1(T-1)}, ..., y_{N(T-1)})'$. The remaining terms are similar to corresponding terms in (4): $\boldsymbol{X}$ is an $N(T-1) \times P$ matrix where the index corresponds to the index of $\boldsymbol{y}$ and $\boldsymbol{u}$ is a column vector of length $N(T-1)$ with the same shifted index as $\boldsymbol{y}$. Due to the lagged term, there is a data loss of N terms compared to the general model, as data belonging to the initial time period cannot be modelled.

The matrix formulation of the error term of the dynamic model is similar to (5), but due to the lost initial time period $\boldsymbol{u}$ is a column vector of length $N(T-1)$ with the same indices as $\boldsymbol{y}$, $\boldsymbol{Z}_\mu$ is an $N(T-1) \times N$ matrix, $\boldsymbol{\mu}$ is a column vector of length N, $\boldsymbol{Z}_\lambda$ is an $N(T-1) \times (T-1)$ matrix, $\boldsymbol{\lambda}$ is a column vector of length $T-1$ given by $\boldsymbol{\lambda} = (\lambda_2, ..., \lambda_T)'$, and $\boldsymbol{v}$ is a column vector of length $N(T-1)$ containing the remaining error.

2.4 The Spatial Panel Data Model

2.4.1 Weight matrices

A weight matrix is an $N \times N$ matrix used to formalize a relation between units and to incorporate a spatial component in panel data models (Cressie, 1991). The value of a specific cell w_{ij} represents a distance between unit i and unit j, so that a high value relates to a shorter distance. These values are the weights of the matrix. The diagonal is set to zero for all elements, and in applications the weight matrix is typically normalized so that all rows sum to one. The product $\boldsymbol{W}\boldsymbol{y}$, with $\boldsymbol{W}$ a normalized weight matrix and $\boldsymbol{y}$ a vector of length N, is an N-vector where each element is a weighted average, with the weights for element i given by the ith row of the matrix.

The empirical part of this study uses two of the common ways to define the weights: the *contiguity matrix*, where the weights are based on the existence of a shared border between regions, and the *inverted distance matrix*, where the weights are based on the distance between the central points of two regions.

To define the contiguity matrix, let $A(i,j)$ denote the minimum number of units passed on a path from unit i to unit j, not including units i and j themselves. The contiguity matrix is then given by

$$(\boldsymbol{W}_c)_{ij} = \begin{cases} 1 & A(i,j) = 0 \\ 0 & A(i,j) \neq 0, \end{cases} \tag{9}$$

so that the value of the cell is one if the units share a border (including sharing a single border point) and zero otherwise. The normalized contiguity matrix is denoted $\boldsymbol{W}_C$.

The weights of the inverted distance matrix are defined by the inverted Euclidean distance between a point in unit i and a point in unit j. In this study, the x coordinate of a unit is calculated as the mean of the westernmost and easternmost points of the unit, and the y coordinate is calculated as the mean of the southernmost and northernmost points. For unit i the central point is denoted (x_i, y_i), and the weight matrix is thus defined by

$$(\boldsymbol{W}_d)_{ij} = \frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}. \tag{10}$$

The normalized inverted distance matrix is denoted $\boldsymbol{W}_D$. In the empirical part of the study, weights between regions that are further than 50 kilometers apart are set to zero. This transformation, which is done to simplify calculations, is equivalent to the assumption that those regions lack a direct impact on each other.

2.4.2 The spatial lag panel data model

The *spatial lag panel data model* (in coming sections called the spatial panel data model or the spatial model) is formed by adding a weighted average of values of the explained variable in neighbouring units to the explanatory variables (Elhorst, 2014). The model expands the linear model in (1) and is written

$$y_{it} = \gamma \sum_{j=1}^{N} w_{ij} y_{jt} + \sum_{p=1}^{P} \beta_p x_{pit} + u_{it} \qquad (i = 1, ..., N; t = 1, ..., T), \tag{11}$$

where γ is a parameter and w_{ij} are weights such that $\sum_{j=1}^{N} w_{ij} = 1, i = 1, ..., N$. Note that the weights w_{ij} for fixed i is a row in a weight matrix $\boldsymbol{W}$ and that the sum of products may be seen as the cross product of that row and the vector of y-values at a fixed t.

The two-way error spatial lag model is defined in analogue to the two-way error linear model, i.e.

$$u_{it} = \mu_i + \lambda_t + v_{it} \qquad (i = 1, ..., N; t = 1, ..., T), \tag{12}$$

with $v_{it} \in IID(0, \sigma_v^2)$. The full two-way error spatial panel data model is thereby written

$$y_{it} = \gamma \sum_{j=1}^{N} w_{ij} y_{jt} + \sum_{p=1}^{P} \beta_p x_{pit} + \mu_i + \lambda_t + v_{it} \qquad (i = 1, ..., N; t = 1, ..., T). \tag{13}$$

In vector form, the two-way error spatial model is given by

$$\begin{aligned} \boldsymbol{y} &= \gamma \boldsymbol{W}_M \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \\ \boldsymbol{u} &= \boldsymbol{Z}_\mu \boldsymbol{\mu} + \boldsymbol{Z}_\lambda \boldsymbol{\lambda} + \boldsymbol{v}, \end{aligned} \tag{14}$$

where $\boldsymbol{y}$, $\boldsymbol{X}$ and $\boldsymbol{\beta}$ are defined as in (4), $\boldsymbol{Z}_\mu$, $\boldsymbol{\mu}$, $\boldsymbol{Z}_\lambda$, $\boldsymbol{\lambda}$ are defined as in (5), γ is a scalar parameter, and $\boldsymbol{W}_M$ is an $M \times M$ matrix formed by reproducing a weight matrix T times and setting those on the diagonal of a block matrix of $T \times T$ blocks. By solving for $\boldsymbol{y}$, with the identity matrix of size M as $\boldsymbol{I}_M$, the model can be written

$$\boldsymbol{y} = (\boldsymbol{I}_M - \gamma \boldsymbol{W}_M)^{-1} (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_\mu \boldsymbol{\mu} + \boldsymbol{Z}_\lambda \boldsymbol{\lambda} + \boldsymbol{v}). \tag{15}$$

2.5 Parameter Estimation

2.5.1 Estimation of the linear panel data model

Having defined the three models we will be examining, we move on to describe the estimation methods of the models.

The linear panel data model with a two-way error may be expressed as

$$\boldsymbol{y} = \boldsymbol{\alpha} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_\mu \boldsymbol{\mu} + \boldsymbol{Z}_\lambda \boldsymbol{\lambda} + \boldsymbol{v} \tag{16}$$

by inserting (5) into (4). Recall that $\boldsymbol{Z}_\mu$ is a set of N dummy variables for individuals and that $\boldsymbol{Z}_\lambda$ is a set of T dummy variables for time. Since the indices are in increasing order from 1 to NT, with time as the faster index, the dummy matrix for individuals is an $M \times N$ matrix consisting of T identity matrices of dimension N stacked in a column of blocks. Similarly, the dummy matrix for time is an $M \times T$ matrix consisting of T blocks, where the ith block is 1 in the ith column and 0 otherwise.

The model can be expressed in a standard linear form and estimated directly using OLS, but due to the large number of parameters it is more common to transform the data and apply the OLS estimator to the transformed matrix. This transformation removes the individual and time effects, but does not impact the estimate of $\boldsymbol{\beta}$.

The intuitive approach to eliminating individual and time effects is to estimate the effects by the mean value for each individual and each time, and subtracting the means from both sides of (16). In order to avoid subtracting the total mean twice, a total mean must be added. This approach can also be shown to be formally correct, using results on partial regression.

The desired transformation can be expressed as a matrix operator (see Mao, 2015, for a similar formulation). This formulation will be used in section 2.6.1 to examine the covariance structure of the model residuals. We let $\boldsymbol{Z}_{N\cdot}$ denote an $M \times M$ matrix in $T \times T$ blocks, with each block an identity matrix of size N. If $\boldsymbol{y}$ is as in (16), i.e. an M-vector with time as the faster index, then $\boldsymbol{Z}_{N\cdot}\boldsymbol{y}$ gives the sums of all observations for each individual, repeated T times. The first element of the vector is, for example, the sum of all observations of the first individual. Similarly, let $\boldsymbol{Z}_{\cdot T}$ denote an $M \times M$ matrix in $T \times T$ blocks such that the blocks on the diagonal consists of ones and all other blocks consists of zeroes. Then, $\boldsymbol{Z}_{\cdot T}\boldsymbol{y}$ is an M-vector containing sums over all individuals for each time period. For example, the T first elements will be the sum of all observations of the first time period. Finally, let $\boldsymbol{Z}_{\cdot\cdot}$ denote an $M \times M$ matrix containing only ones. Then, $\boldsymbol{Z}_{\cdot\cdot}\boldsymbol{y}$ is an M-vector where every element is the total sum of all observations.

Using the defined matrices, the transformation operator is given by

$$\boldsymbol{Q} = \boldsymbol{I}_M - \frac{1}{T}\boldsymbol{Z}_{N\cdot} - \frac{1}{N}\boldsymbol{Z}_{\cdot T} + \frac{1}{NT}\boldsymbol{Z}_{\cdot\cdot}. \tag{17}$$

The transformed explained variable is $\boldsymbol{Q}\boldsymbol{y}$ and the transformed explanatory variables are $\boldsymbol{Q}\boldsymbol{X}$. For $\boldsymbol{Q}\boldsymbol{y}$, the element connected to the ith individual and

jth time is the original value from $\boldsymbol{y}$ minus the mean of all observations for individual i, minus the mean of all observation for time j, plus the mean of all observation regardless of individual and time. The same goes for each of the explanatory variables in $\boldsymbol{X}$.

Applying the transformation to (16) gives

$$\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{v}, \tag{18}$$

as $\boldsymbol{Q}\boldsymbol{\alpha} = \boldsymbol{Q}\boldsymbol{Z}_\mu = \boldsymbol{Q}\boldsymbol{Z}_\lambda = \boldsymbol{0}$. Standard OLS results now give the estimator of β as

$$\hat{\boldsymbol{\beta}}_{FE} = ((\boldsymbol{Q}\boldsymbol{X})'(\boldsymbol{Q}\boldsymbol{X}))^{-1}(\boldsymbol{Q}\boldsymbol{X})'\boldsymbol{Q}\boldsymbol{y}, \tag{19}$$

the estimates for the residual variance

$$\widehat{s_v^2} = \frac{(\boldsymbol{Q}\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{X}\hat{\boldsymbol{\beta}}_{FE})'(\boldsymbol{Q}\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{X}\hat{\boldsymbol{\beta}}_{FE})}{NT - N - T - P}, \tag{20}$$

and the covariance matrix for the parameter vector

$$\widehat{\mathbb{V}}(\hat{\boldsymbol{\beta}}_{FE}) = \widehat{s_v^2}((\boldsymbol{Q}\boldsymbol{X})'(\boldsymbol{Q}\boldsymbol{X}))^{-1}. \tag{21}$$

The estimator $\hat{\boldsymbol{\beta}}_{FE}$ is known as the *fixed effects* or *within* estimator.

Since the within estimator is an OLS estimator, the estimator properties follow from OLS results. Hence, $\hat{\boldsymbol{\beta}}_{FE}$ is unbiased, efficient and normally distributed under the assumptions of zero-mean for the error term v_{it}, independence between explanatory variables and error (exogeneity), independence between different errors (no serial or cross-sectional correlation), constant variance (homoskedasticity), non-multicollinearity in explanatory variables, and normally distributed errors (Verbeek, 2004). The estimator is unbiased given the assumptions on zero-mean errors and independence between explanatory variables and error. The independence assumption can be relaxed as the estimator is consistent if the error terms and explanatory variables are uncorrelated. The panel model also requires a poolability assumption, i.e. that the parameters β_p are constant over individual and time.

Although the assumptions on the error term do not hold for the transformed model $\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{v}$ (as will be shown in section 2.6.1), unbiasedness and normality of $\boldsymbol{\beta}$ hold if the OLS assumptions hold for the original error vector $\boldsymbol{v}$, as OLS on the transformed model is equivalent to OLS on the original model with dummies for individuals and time periods.

Estimation is implemented in the R package plm (Croissant and Millo, 2008).

2.5.2 Estimation of the dynamic panel data model

Due to the correlation between the error term $\boldsymbol{v}$ and the lagged explanatory variable $\boldsymbol{y}^{(-1)}$, the dynamic model of (8) cannot be estimated using an OLS method without introducing bias and inconsistency. There has therefore been developments of alternative estimators with relaxed assumptions and better properties. This study uses the *Arellano-Bond estimator*, a generalized method of moments

(GMM) estimator where instruments for the lagged explanatory variable (i.e. a variable which is correlated to the original explanatory variable but uncorrelated with the error term) are derived from greater lags of the variable (Arellano and Bond, 1991).

Let Δ denote a time difference operator such that $\Delta y_{it} = y_{it} - y_{i,t-1}$. Applying the operator to the one-way error dynamic model cancels the individual effects:

$$y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + (\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1})'\boldsymbol{\beta} + (\mu_i - \mu_i) + (v_{it} - v_{i,t-1}), \quad (22)$$

or, expressed with the difference operator,

$$\Delta y_{it} = \delta \Delta y_{i,t-1} + \Delta \boldsymbol{x}'_{it}\boldsymbol{\beta} + \Delta v_{it}. \quad (23)$$

For this differenced model, there is correlation between $\Delta y_{i,t-1}$ and Δv_{it}, however, $y_{i,t-2}$ is correlated to $\Delta y_{i,t-1}$ but uncorrelated to Δv_{it}, and can thereby act as an instrumental variable. This holds for $y_{i,t-s}$ for any $s \geq 2$ and the number of instruments thereby increases with t. The GMM estimator is formed by defining a set of parameter dependent moment conditions, approximating the conditions with their empirical counterparts, and setting the parameters so that the difference between the theoretical moments and the estimated moments is minimized. For the Arellano-Bond estimator, the conditions are given by the assumption that the instruments and the explanatory variables are uncorrelated with the error terms.

The expressions for the estimator and estimated variance are simplified by a matrix construction. In this part the notation $[\boldsymbol{x}, \boldsymbol{y}]$ is used for concatenation of matrices with equal number of rows. In particular, $[\boldsymbol{x}, \boldsymbol{y}] = (x_1, ..., x_n, y_1, ..., y_m)$ for row vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of lengths n and m. The vector $\mathbf{0}_n$ denotes a row vector of zeroes of length n. For any $i = 1, ..., N$, define the matrix

$$\boldsymbol{Z}_i = \begin{pmatrix} [y_{i1}, \Delta \boldsymbol{x}'_{i3}] & \mathbf{0}_{2+P} & \cdots & \mathbf{0}_{T-2+P} \\ \mathbf{0}_{1+P} & [(y_{i1}, y_{i2}), \Delta \boldsymbol{x}'_{i4}] & \cdots & \mathbf{0}_{T-2+P} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1+P} & \mathbf{0}_{2+P} & \cdots & [(y_{i1}, ..., y_{i,T-2}), \Delta \boldsymbol{x}'_{i,T}] \end{pmatrix} \quad (24)$$

and the $(T-2)$ vector

$$\Delta \boldsymbol{v}_i = \begin{pmatrix} v_{i3} - v_{i2} \\ v_{i4} - v_{i3} \\ \vdots \\ v_{iT} - v_{i,T-1} \end{pmatrix}. \quad (25)$$

The number of columns of $\boldsymbol{Z}_i$ is given by $(T-2)\left(\frac{T-1}{2} + P\right)$, and corresponds to the number of instruments.

The vector of theoretical moment conditions is given by $E(\boldsymbol{Z}'_i \Delta \boldsymbol{v}_i) = \mathbf{0}$, a typical element of which is $E(y_{is}\Delta v_{it}) = 0$ with $t = 3, ..., T$ and $s = 1, ..., t-2$ for the lagged explanatory variable, and $E(\Delta x_{itp} \Delta v_{it}) = 0$ for the explained variables. The full matrix of instruments is $\boldsymbol{Z} = [\boldsymbol{Z}_1', \boldsymbol{Z}_2', ..., \boldsymbol{Z}_N']'$, a matrix of dimension $N(T-2) \times (T-2)(\frac{T-1}{2} + P)$. The moment conditions are dependent on the parameters δ and $\boldsymbol{\beta}$ through the error term $\boldsymbol{v}$ and the empirical forms of the conditions are for each time period given by mean values over the N individuals.

The variance of the estimator can be obtained using that $\Delta\boldsymbol{v}_i$ is an MA(1) process with parameter -1. By setting an index ranging from 1 to $T-2$ and a variance σ_v^2, such process has a covariance matrix given by $E(\Delta\boldsymbol{v}_i\Delta\boldsymbol{v}_i') = \sigma_v^2\boldsymbol{G}$, where $\boldsymbol{G}$ is a $(T-2) \times (T-2)$ matrix with 2 on the diagonal, -1 on the super and sub-diagonals, and 0 for all other entries. For the parameter vector $\boldsymbol{\theta} = [\delta, \boldsymbol{\beta}']$ and the data matrix $\boldsymbol{T} = [\boldsymbol{y}^{(-1)}, \boldsymbol{X}]$, the Arellano-Bond estimator is given through a two-step procedure. The one-step estimator is given by

$$\hat{\boldsymbol{\theta}}_1 = [(\Delta\boldsymbol{T})'\boldsymbol{Z}\boldsymbol{V}^{-1}\boldsymbol{Z}'(\Delta\boldsymbol{T})]^{-1}[(\Delta\boldsymbol{T})'\boldsymbol{Z}\boldsymbol{V}^{-1}\boldsymbol{Z}'(\Delta\boldsymbol{y})], \tag{26}$$

where $\boldsymbol{V} = \sum_{i=1}^{N} \boldsymbol{Z}_i'\boldsymbol{G}\boldsymbol{Z}_i$. The estimated parameters $\hat{\boldsymbol{\theta}}_1$ are used to estimate the error vector, $\hat{\boldsymbol{v}}$, and the two-step estimator $\hat{\boldsymbol{\theta}}_2$ is given by replacing $\boldsymbol{V}$ in (26) with

$$\hat{\boldsymbol{V}}_N = \sum_{i=1}^{N} \boldsymbol{Z}_i'(\Delta\hat{\boldsymbol{v}}_i)(\Delta\hat{\boldsymbol{v}}_i)'\boldsymbol{Z}_i. \tag{27}$$

The covariance matrix of the estimator is

$$\widehat{\mathbb{V}}(\hat{\boldsymbol{\theta}}_2) = [(\Delta\boldsymbol{T})'\boldsymbol{Z}\hat{\boldsymbol{V}}_N^{-1}\boldsymbol{Z}'(\Delta\boldsymbol{T})]^{-1}. \tag{28}$$

The two-way error model may be estimated by incorporating time dummy variables in the data matrix $\boldsymbol{X}$.

Estimation of the dynamic panel data model is implemented in the plm package (Croissant and Millo, 2008).

2.5.3 Estimation of the spatial panel data model

In the spatial panel model (14), the spatially lagged variable is correlated with the error term and any OLS estimator is inconsistent. There are several alternative estimators. This study uses a maximum likelihood estimator due to Elhorst (2014), based on an estimator for cross-sectional spatial models (Anselin and Hudak, 1992). The estimator uses a two-step approach where the parameter $\boldsymbol{\beta}$ is estimated using OLS and the spatial parameter γ is estimated using maximum likelihood.

Starting with the OLS step, note that the spatial model can be written

$$\boldsymbol{y} - \gamma\boldsymbol{W}_M\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_\mu\boldsymbol{\mu} + \boldsymbol{Z}_\lambda\boldsymbol{\lambda} + \boldsymbol{v} \tag{29}$$

by collecting the explained variable on the left-hand side, and that this model can be simplified by applying the transformation operator $\boldsymbol{Q}$ from the estimation of the linear model, giving

$$\boldsymbol{Q}\boldsymbol{y} - \gamma\boldsymbol{Q}\boldsymbol{W}_M\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{v}. \tag{30}$$

This model has the same right-hand side as the transformed linear panel data model (18) and, if γ was known, the parameter $\boldsymbol{\beta}$ could be estimated using OLS.

Next, the parameter $\boldsymbol{\beta}$ is decomposed as $\boldsymbol{\beta} = \boldsymbol{\beta}_0 - \gamma\boldsymbol{\beta}_1$ where the terms are the parameters from the models

$$\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{Q}\boldsymbol{v}_0, \tag{31}$$

and

$$\boldsymbol{Q}\boldsymbol{W}_M\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta}_1 + \boldsymbol{Q}\boldsymbol{v}_1. \tag{32}$$

Both of these models may be estimated using OLS, giving the estimates $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$, as well as the two residual series $\hat{\boldsymbol{v}}_0$ and $\hat{\boldsymbol{v}}_1$. Using these estimates, the estimation of γ is reduced to a one-dimensional optimization problem.

The spatial parameter γ is estimated using maximum likelihood estimation and the assumption that the error terms are independent and identically normal, i.e. $\boldsymbol{v} \in N(0, \boldsymbol{I}_M\sigma_v^2)$. The alternative matrix formulation in (15) gives that the random vector $\boldsymbol{y}$ is multivariate normal with known expectation and variance,

$$\begin{aligned} \boldsymbol{y} \in N((&\boldsymbol{I}_M - \gamma\boldsymbol{W}_M)^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_\mu\boldsymbol{\mu} + \boldsymbol{Z}_\lambda\boldsymbol{\lambda}), \\ &(\boldsymbol{I}_M - \gamma\boldsymbol{W}_M)^{-1}\sigma_v^2((\boldsymbol{I}_M - \gamma\boldsymbol{W}_M)^{-1})'), \end{aligned} \tag{33}$$

and the log-likelihood, calculated by taking the log of the density function and summing the elements, is given by

$$l(\boldsymbol{y}) = -\frac{NT}{2}\log(2\pi\sigma_v^2) + T\log|\boldsymbol{I}_N - \gamma\boldsymbol{W}| - \frac{1}{2\sigma_v^2}\boldsymbol{v}'\boldsymbol{v}. \tag{34}$$

The expression $|\boldsymbol{I}_N - \gamma\boldsymbol{W}|$ is a determinant stemming from the transformation from $\boldsymbol{v}$ as random variable to $\boldsymbol{y}$ as random variable. The calculation of the multiple of the term $\log|\boldsymbol{I}_N - \gamma\boldsymbol{W}|$ is based on the construction of $\boldsymbol{W}_M$ as a block matrix, so that $|\boldsymbol{I}_M - \gamma\boldsymbol{W}_M| = |\boldsymbol{I}_N - \gamma\boldsymbol{W}|^T$.

Next, the maximum likelihood estimate of the variance, $\sigma_v^2 = \frac{1}{NT}\boldsymbol{v}'\boldsymbol{v}$, and the decomposition of the error term, $\hat{\boldsymbol{v}} = \hat{\boldsymbol{v}}_0 - \gamma\hat{\boldsymbol{v}}_1$, is used to simplify the expression to a concentrated likelihood depending only on γ:

$$\begin{aligned} l(\boldsymbol{y}) &= -\frac{NT}{2}\log(2\pi\boldsymbol{v}'\boldsymbol{v}) + T\log|\boldsymbol{I}_N - \gamma\boldsymbol{W}| - \frac{NT}{2} \\ &= -\frac{NT}{2}\log{(\hat{\boldsymbol{v}}_0 - \gamma\hat{\boldsymbol{v}}_1)'(\hat{\boldsymbol{v}}_0 - \gamma\hat{\boldsymbol{v}}_1)} + T\log|\boldsymbol{I}_N - \gamma\boldsymbol{W}| + C \end{aligned} \tag{35}$$

where C includes all terms constant in γ. The parameter γ is estimated by numerically maximizing the final expression in (35) and the parameter $\hat{\boldsymbol{\beta}}$ is retrieved from the decomposition as

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_0 - \hat{\gamma}\hat{\boldsymbol{\beta}}_1. \tag{36}$$

The variance is estimated by the standard maximum likelihood estimate

$$\widehat{\sigma_v^2} = \frac{1}{NT}\hat{\boldsymbol{v}}'\hat{\boldsymbol{v}} \tag{37}$$

and the asymptotic covariance matrix of the estimates is

$$\mathbb{V}(\hat{\boldsymbol{\beta}}, \hat{\gamma}, \widehat{\sigma_v^2}) =$$
$$\begin{pmatrix} \frac{(\boldsymbol{QX})'(\boldsymbol{QX})}{\sigma_v^2} & & \\ \frac{(\boldsymbol{QX})'\tilde{\boldsymbol{W}}_M(\boldsymbol{QX})\boldsymbol{\beta}}{\sigma_v^2} & \frac{\sigma_v^2 T\cdot tr(\tilde{\boldsymbol{W}}\tilde{\boldsymbol{W}}+\tilde{\boldsymbol{W}}'\tilde{\boldsymbol{W}})+\boldsymbol{\beta}'(\boldsymbol{QX})'\tilde{\boldsymbol{W}}_M'\tilde{\boldsymbol{W}}_M(\boldsymbol{QX})\boldsymbol{\beta}}{\sigma_v^2} & \\ 0 & \frac{T\cdot tr(\tilde{\boldsymbol{W}})}{\sigma_v^2} & \frac{NT}{2\sigma_v^4} \end{pmatrix}^{-1}, \quad (38)$$

where $\tilde{\boldsymbol{W}} = \boldsymbol{W}(\boldsymbol{I}_N - \gamma\boldsymbol{W})^{-1}$, $\tilde{\boldsymbol{W}}_M$ is an $M \times M$ matrix constructed in the same way as $\boldsymbol{W}_M$ in (14), and tr denotes the trace of the matrix.

The estimation procedure is implemented in the R package splm (Millo and Piras, 2012).

2.6 Testing for Residual Spatial Autocorrelation

2.6.1 Correlation in OLS residuals

In the case of spatial data, it can be of interest to test the residual series for spatial autocorrelation, as this might indicate that the model should incorporate a spatial term, for example by applying a spatial lag model. There are several available tests of residual spatial autocorrelation, both of global nature, i.e. tests which do not require a predetermined geographical structure on the residuals, and of local nature, i.e. tests which use a weight matrix to formalize geographical distance.

Tests for spatial autocorrelation typically apply the null hypothesis that there is no autocorrelation. This creates a problem in the linear panel data model with two-way error estimated with the standard OLS method (as presented in section 2.5.1), namely that the estimation method results in a spatial structure on the estimated residuals, even if the original variables are uncorrelated (Mao, 2015). To calculate the residual correlation, recall that the OLS estimator of $\boldsymbol{\beta}$ in the linear model is equivalent to OLS on the transformed model

$$\boldsymbol{Qy} = \boldsymbol{QX\beta} + \boldsymbol{Qv}, \quad (39)$$

where

$$\boldsymbol{Q} = \boldsymbol{I}_M - \frac{1}{T}\boldsymbol{Z}_{N\cdot} - \frac{1}{N}\boldsymbol{Z}_{\cdot T} + \frac{1}{NT}\boldsymbol{Z}_{\cdot\cdot}. \quad (40)$$

Note that $\boldsymbol{Q}$ is symmetric ($\boldsymbol{Q} = \boldsymbol{Q}'$) and invariant under multiplication with itself ($\boldsymbol{QQ} = \boldsymbol{Q}$). Under the assumption that the error terms in $\boldsymbol{v}$ are independent with variance σ_v^2, applying the transformation operator $\boldsymbol{Q}$ to the error in the linear model gives

$$\begin{aligned} \mathbb{V}(\boldsymbol{Qv}) &= \mathbb{E}(\boldsymbol{Qv}(\boldsymbol{Qv})') = \boldsymbol{Q}\mathbb{E}(\boldsymbol{vv}')\boldsymbol{Q}' \\ &= \boldsymbol{Q}\mathbb{V}(\boldsymbol{v})\boldsymbol{Q}' = \boldsymbol{Q}\sigma_v^2\boldsymbol{I}_M\boldsymbol{Q}' = \sigma_v^2\boldsymbol{Q}. \end{aligned} \quad (41)$$

This directly gives the variance-covariance matrix of the transformed error terms

$$(\mathbb{V}(\boldsymbol{Qv}))_{it,js} = \begin{cases} \frac{(N-1)(T-1)}{NT}\sigma_v^2 & i=j, t=s \\ -\frac{N-1}{NT}\sigma_v^2 & i=j, t\neq s \\ -\frac{T-1}{NT}\sigma_v^2 & i\neq j, t=s \\ \frac{1}{NT}\sigma_v^2 & i\neq j, t\neq s \end{cases} \tag{42}$$

and the corresponding correlation matrix

$$(\rho(\boldsymbol{Qv}))_{it,js} = \begin{cases} 1 & i=j, t=s \\ -\frac{1}{T-1} & i=j, t\neq s \\ -\frac{1}{N-1} & i\neq j, t=s \\ \frac{1}{(N-1)(T-1)} & i\neq j, t\neq s. \end{cases} \tag{43}$$

Hence, the correlation between two transformed error terms is different from zero, even under the assumption that the original error terms are independent, and this correlation is large for small N.

This result holds for the estimated error, i.e. the residuals $\hat{\boldsymbol{v}}$, if the number of observations is large and the error is assumed to be uncorrelated with the explanatory variables. This follows from noting that the variance of the residuals is

$$\begin{aligned} \mathbb{V}(\widehat{\boldsymbol{Qv}}) &= \mathbb{V}(\boldsymbol{Qy} - \boldsymbol{QX}\hat{\boldsymbol{\beta}}) = \mathbb{V}(\boldsymbol{QX}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \boldsymbol{Qv}) \\ &= \mathbb{V}(\boldsymbol{QX}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) + \mathbb{V}(\boldsymbol{Qv}), \end{aligned} \tag{44}$$

which tends to $\mathbb{V}(\boldsymbol{Qv})$ as $\hat{\boldsymbol{\beta}}$ tends to $\boldsymbol{\beta}$ with N or T.

This gives that any test of residual spatial correlation which relies on the null hypothesis that the residuals are uncorrelated will give false rejections.

2.6.2 Effect on tests for spatial autocorrelation

The effect of the residual correlation structure is exemplified using two spatial tests: the Pesaran test, and the bias-corrected LM test.

The Pesaran test is a test of global cross-sectional dependence based on estimations of correlations between all individuals (Pesaran, 2004). For a vector of residuals, the test statistic is given by

$$CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_j), \tag{45}$$

where $\hat{\rho}(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_j)$ is the estimated correlation between individuals i and j, i.e. $\hat{\rho}(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_j) = \hat{\boldsymbol{v}}_i'\hat{\boldsymbol{v}}_j / \sqrt{\hat{\boldsymbol{v}}_i'\hat{\boldsymbol{v}}_i\hat{\boldsymbol{v}}_j'\hat{\boldsymbol{v}}_j}$ for the vectors $\boldsymbol{v}_i' = (\hat{v}_{i1}, ..., \hat{v}_{iT})$ and $\boldsymbol{v}_j' = (\hat{v}_{j1}, ..., \hat{v}_{jT})$. Given the null hypothesis that there is no cross-sectional correlation, CD tends to a standard normal distribution as N tends to infinity.

The LM test is a test for global spatial autocorrelation (Baltagi, 2013). The test statistic of the bias-corrected version of the test is given by

$$LM_{BC} = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (T\hat{\rho}(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_j)^2 - 1) - \frac{N}{2(T-1)}, \tag{46}$$

where the last term is the bias-correction. LM_{BC} tends to a standard normal distribution as N and T tends jointly to infinity with N/T tending to some non-zero constant.

The asymptotic results for the Pesaran test and the LM test rely on the null hypothesis of zero spatial correlation. If the variable $\boldsymbol{v}$ is formed from OLS residuals and thereby carries the correlation structure from (43), this assumption of zero correlation is violated. For small samples it is clear that the tests are questionable for residuals with the correlation structure in (43), as the expected values of the test statistics are non-zero.

2.6.3 Parametric tests for correlated residuals

This study examines two methods of testing for spatial correlation for data with residual correlation structure, the first being parametric tests after an inverse-like back-transform of the residuals as suggested by Mao (2015) and the second being an application of a randomization test. This section covers the parametric approach, starting with a description of a general inversion method for symmetric matrices and ending with descriptions of the specific tests. The section follows Mao (2015), with some minor changes in terminology and presentation.

As noted above, the matrix $\boldsymbol{Q}$, defined in (17), is idempotent, i.e. $\boldsymbol{QQ} = \boldsymbol{Q}$. As such, the following facts about idempotent matrices hold for $\boldsymbol{Q}$: (i) $\boldsymbol{Q}$ is non-invertible or the identity matrix, (ii) the trace of $\boldsymbol{Q}$ is equal to the rank of $\boldsymbol{Q}$, (iii) the eigenvalues of $\boldsymbol{Q}$ are zero or one and the sum of eigenvalues is equal to the trace, (iv) if the eigenvectors corresponding to eigenvalue one are set in an $n \times tr(Q)$ matrix, denoted $\boldsymbol{D}$, then $\boldsymbol{DD}' = \boldsymbol{Q}$ and $\boldsymbol{D}'\boldsymbol{D} = \boldsymbol{I}_{tr(\boldsymbol{Q})}$. These properties of idempotent matrices make it possible to transform the estimated residuals to eliminate the residual correlation structure. For the full set of residuals, the transformation takes the following form: given a vector $\boldsymbol{v}$ of length M and variance matrix $\boldsymbol{I}_M\sigma_v^2$, the vector $\boldsymbol{Qv}$ has the structure from (42); if $\boldsymbol{D}$ denotes the matrix of eigenvectors of $\boldsymbol{Q}$, then the variance of $\boldsymbol{D}'\boldsymbol{Qv}$ is given by

$$\begin{aligned} \mathbb{V}(\boldsymbol{D}'\boldsymbol{Qv}) &= \mathbb{E}(\boldsymbol{D}'\boldsymbol{Qvv}'\boldsymbol{Q}'\boldsymbol{D}) = \mathbb{E}(\boldsymbol{D}'\boldsymbol{DD}'\boldsymbol{vv}'\boldsymbol{DD}'\boldsymbol{D}) \\ &= \mathbb{E}(\boldsymbol{D}'\boldsymbol{vv}'\boldsymbol{D}) = \boldsymbol{D}'\mathbb{V}(\boldsymbol{v})\boldsymbol{D} = \boldsymbol{D}'\boldsymbol{I}_M\sigma_v^2\boldsymbol{D} \\ &= \boldsymbol{I}_{(N-1)(T-1)}\sigma_v^2. \end{aligned} \tag{47}$$

Hence, the transformation results in a vector of length $(N-1)(T-1)$ which has uncorrelated elements if the original errors $\boldsymbol{v}$ are uncorrelated. Unfortunately, this inversion of the residuals has serious drawbacks as there is a data loss of $N + T - 1$ observations and a loss of original local structure, making it difficult to calculate empirical correlations between the original spatial units.

Mao (2015) suggests a transformation which uses the same decomposition of an idempotent matrix, but keeps more of the original structure. For any natural number m, define a matrix $\boldsymbol{E}_m$ by $\boldsymbol{E}_m = \boldsymbol{I}_m - m^{-1}\boldsymbol{J}_m$, where $\boldsymbol{J}_m$ is an $m \times m$ matrix of ones. This matrix is idempotent with $tr(\boldsymbol{E}_m) = m - 1$, and can thereby be decomposed as $\boldsymbol{E}_m = \boldsymbol{D}_m\boldsymbol{D}_m'$ for some $m \times (m-1)$ matrix of eigenvectors. Given an NT-vector of residuals, form an $N \times T$ matrix $\boldsymbol{U}$ by setting the values for individual i as the ith row. The transformation is now given by $\boldsymbol{S} = \boldsymbol{D}_N'\boldsymbol{U}\boldsymbol{D}_T$, where $\boldsymbol{D}_N$ and $\boldsymbol{D}_T$ come from the decomposition of $\boldsymbol{E}_N$ and $\boldsymbol{E}_T$ respectively. The resulting $(N-1)\times(T-1)$ matrix $\boldsymbol{S}$ contains the back-transformed residuals, and these are uncorrelated if the original errors are uncorrelated. For the transformed residuals, the test statistic for the Pesaran test is given by

$$CD_Q = \sqrt{\frac{2(T-1)}{(N-1)(N-2)}} \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \hat{\rho}(\hat{\boldsymbol{s}}_i, \hat{\boldsymbol{s}}_j), \tag{48}$$

and for the LM test by

$$LM_Q = \sqrt{\frac{T+1}{(N-1)(N-2)(T-2)}} \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} (T\hat{\rho}(\hat{\boldsymbol{s}}_i, \hat{\boldsymbol{s}}_j)^2 - 1), \tag{49}$$

where $\hat{\rho}(\hat{\boldsymbol{s}}_i, \hat{\boldsymbol{s}}_j) = \hat{\boldsymbol{s}}_i'\hat{\boldsymbol{s}}_j / \sqrt{\hat{\boldsymbol{s}}_i'\hat{\boldsymbol{s}}_i\hat{\boldsymbol{s}}_j'\hat{\boldsymbol{s}}_j}$ for column vectors like $\hat{\boldsymbol{s}}_i = (\hat{s}_{i1}, ..., \hat{s}_{i,(T-1)})$. Under the null hypothesis that the original errors are uncorrelated, CD_Q and LM_Q tend to standard normal distributions as N and T tend to infinity with N/T tending to a constant.

2.6.4 Randomization tests of spatial autocorrelation

An alternative to the inversion approach is given by various randomization tests, where a statistic on the true data is compared to that statistic on randomly resampled data. This has the advantage of allowing for tests on local spatial correlation and of avoiding the use of asymptotic results which may be imprecise for small samples. There are a number of randomization tests of spatial autocorrelation available for cross-sectional data (Good, 2005). In this study, two randomization tests of spatial autocorrelation for panel data are formulated, tested and applied. The first is a local test based on the estimated correlation between a variable and the locally weighted means of that variable. The second is a global test based on the LM-statistic.

The local test is based on the correlation between the original series and the series of locally weighted means. The null hypothesis is that there is no spatial autocorrelation. Given a panel data series $\boldsymbol{y}$, possibly but not necessarily a set of residuals, and a weight matrix $\boldsymbol{W}_M$ similar in form to that in (11), the test statistic is given by the correlation estimate between the series $\boldsymbol{y}$ and $\boldsymbol{W_M y}$:

$$C_{obs}^{loc} = \hat{\rho}(\boldsymbol{y}, \boldsymbol{W}_M\boldsymbol{y}). \tag{50}$$

The data is resampled by reordering individuals while keeping the time component intact. The reordered vector is denoted $\boldsymbol{y}_p^{loc}$ and the comparative value for the test statistic is $C_p^{loc} = \hat{\rho}(\boldsymbol{y}_p^{loc}, \boldsymbol{W}_M \boldsymbol{y}_p^{loc})$. The resampling procedure is repeated K times for a set of comparative values $\{C_p^{loc}\}_{p=1}^K$. The number of values higher than or equal to C_{obs}^{loc} and the number of values lower than or equal to C_{obs}^{loc} are denoted by C_H and C_L respectively. The p-value for the two-sided randomization test is calculated as

$$\hat{p} = \frac{2\min(C_H, C_L) + 2}{K + 2}. \tag{51}$$

If the sample of individuals N is small, the number of permutations[1], and hence possible comparative values, will be low. This greatly impacts the size of the test.

The global randomization test is based on the LM-statistic in (46). The null hypothesis is that there is no spatial autocorrelation. The observed value is calculated as the LM-statistic for the original series $\boldsymbol{y}$, i.e.

$$C_{obs}^{glob} = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (T\hat{\rho}(\hat{\boldsymbol{y}}_i, \hat{\boldsymbol{y}}_j)^2 - 1) \tag{52}$$

Next, the data is reorded by permuting within each time point, and for each random permutation a comparative value is calculated by applying (52) to the reordered series $\boldsymbol{y}_p^{glob}$. This gives a set of K comparative vales, $\{C_p^{glob}\}_{p=1}^K$, from which values of C_H, C_L and a p-value are calculated according to the same definition as in the local randomization test.

2.7 Additional Hypothesis Tests

The empirical part of this study uses a set of tests of the underlying assumptions of the applied models.

The fixed effects linear model estimated with the within estimator rely on the standard OLS assumptions (zero-mean errors, independence between explanatory variables and error (exogeneity), independence between different errors (no serial or cross-sectional correlation), constant variance (homoskedasticity), non-multicollinearity in explanatory variables, and normally distributed errors) and a poolability assumption (Verbeek, 2004).

Testing the first assumptions, that the error terms are zero-mean and independent of the explanatory variables, requires a set of alternative forms of given explanatory variables or a suitable set of instruments, usually found through economical theory. For this study, it is assumed that the explanatory variable have suitable functional form and are exogenous.

Besides spatial correlation, tests for which have already been described, panel model residuals may be correlated in time. Such serial correlation is tested using an AR(1)-Wooldridge test (Wooldridge, 2002). The test only checks for first

[1]Typically $N!$ but possibly smaller, depending on the weight matrix $\boldsymbol{W}_M$

order serial correlation, but has better performance for short panels (i.e. panels with a low T) than alternative tests. The null hypothesis is that there is no serial correlation in the residuals. The test is implemented in the plm package (Croissant and Millo, 2008)

Homoskedasticity is tested using the studentized version of the Breusch-Pagan test (Koenker, 1981). The null hypothesis is that the errors are homoskedastic. The test is implemented using the lmtest package (Zeileis and Hothorn, 2002)

Near-perfect multicollinearity is tested by estimating each explanatory variable as a function of all other explanatory variables and calculating the goodness-of-fit. A strong connection between explanatory variables may result in inflated standard errors and indicates that the variable can be removed without losing explanatory power (Verbeek, 2004).

Normality of the errors is tested using a Shapiro-Wilk test on the residuals (Shapiro and Wilk, 1965). The null hypothesis of the test is that the errors follow a normal distribution with the estimated mean and variance. Normality of the residuals is also investigated using QQ-plots and histograms.

Under the assumption of normality, the standard Wald test is applicable and may be used to test for significance in particular explanatory variables and to investigate the presence of individual and time effects by testing for $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\lambda} = \mathbf{0}$ (Verbeek, 2004).

Poolability is tested using a Chow test which compares the pooled model with individual and time effects to a model where parameters $\boldsymbol{\beta}$ vary with individuals or time (Baltagi, 2013). For the test to be applicable, the number of observations used in the estimation must outnumber the number of estimated parameters, so if the parameters are allowed to vary between individuals, the number of time periods must be greater than the number of explanatory variables. The null hypothesis is that parameters are constant. The test is implemented in the plm package (Croissant and Millo, 2008).

In the dynamic model, serial correlation is tested using an Arellano-Bond test (Arellano and Bond, 1991). The test, which uses the null hypothesis that serial correlation is not present, is applied to both first and second order correlation, i.e. dependence on a once lagged respectively a twice lagged explanatory variable. Due to the additional time lag in the estimation method, the test is expected to reject the null for first order correlation and accept for second order correlation in cases where the data follows the dynamic model.

The GMM method for the dynamic model is based on more moment conditions than there are estimated parameters. This makes it possible to estimate the degree to which the conditions are met by performing a Sargan test (Arellano and Bond, 1991). The alternative hypothesis of the test is that some moment condition is violated, indicating that one or several instrumental variables are invalid. The Arellano-Bond test and the Sargan test are both applied using function in the plm package (Croissant and Millo, 2008).

Wald tests for effect of explanatory variables and time effects, as well as the Shapiro-Wilk test for normality of the errors, are applied to the dynamic model as to the linear model.

For the spatial model, the effect of explanatory variables and time dummies is tested by a likelihood-ratio test, using the null hypothesis that there is no explanatory effect (Verbeek, 2004). Normality of the error terms is tested using the Shapiro-Wilk test.

2.8 Goodness-of-Fit

Two common measures of goodness-of-fit for panel data are calculated for each estimated model. The first measure is given by the portion of total variance in $\boldsymbol{y}$ which is explained by the full set of explanatory variables, including the individual and time effects (the *total* measure). The second measure is given by the proportion of variance after individual and time effects are removed which is explained by the explanatory variables $\boldsymbol{X}$ (the *within* measure).

Let $\hat{e}_{it}$ denote the OLS residuals in the model which has individual and time effects but no explanatory variables, i.e. $\hat{e}_{it} = y_{it} - \hat{\alpha} - \hat{\mu}_i - \hat{\lambda}_t$, and let $\hat{v}_{it}$ denote the residuals in a model with individual and time effects and explanatory variables. The vector equivalents are denoted $\hat{\boldsymbol{e}}$ and $\hat{\boldsymbol{v}}$. The measure for the total effect of explanatory variables and individual and time effects is now calculated by

$$R_t^2 = 1 - \frac{\hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}}{(\boldsymbol{y} - \boldsymbol{\iota}\bar{y}_{..})'(\boldsymbol{y} - \boldsymbol{\iota}\bar{y}_{..})}, \tag{53}$$

where $\boldsymbol{\iota}$ is an M-vector containing ones and $\bar{y}_{..}$ is the mean of y over all individuals and time periods. The within measure for effect of explanatory variables given the effect of individual and time effects is calculated by

$$R_w^2 = 1 - \frac{\hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}}{\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}}. \tag{54}$$

Given the two measures, the proportion of total variance explained by individual and time effects alone (the *effects* measure) is calculated by

$$R_e^2 = 1 - \frac{\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}}{(\boldsymbol{y} - \boldsymbol{\iota}\bar{y}_{..})'(\boldsymbol{y} - \boldsymbol{\iota}\bar{y}_{..})} = \frac{R_t^2 - R_w^2}{1 - R_w^2}. \tag{55}$$

Since the dynamic and spatial models are not estimated by OLS, the R^2 measures are not as well-behaved as in the standard case. Most noticeably, R^2 for the dynamic and spatial cases are not necessarily larger than R^2 for the linear case, even though both contain all variables of the linear model.

2.9 Summary of Method

In the method chapter, three panel data models and corresponding estimation methods have been described. The linear panel data model is estimated using standard OLS, which in the panel setting is performed using the within estimator, a transformation method where individual and time means are removed. The dynamic panel data model adds a temporal component to the linear model and is estimated using a generalized method of moments estimator. The spatial panel data model adds a spatially lagged component to the linear model and is estimated using a maximum likelihood-method closely related to the estimator for the linear model.

Next, it was noted that the OLS method creates a correlation structure in the residuals, even if the individual units are uncorrelated in the original data. This correlation structure affects tests for residual autocorrelation. Two possible ways to compensate for this is the back-transform suggested in previous literature (Mao, 2015) and randomization test.

The next chapter presents a Monte Carlo study of the connection between residual autocorrelation, results of residual tests, and the estimates of the spatial model.

3 Monte Carlo Study of Residual Correlation

3.1 Introduction

This chapter presents the results of the Monte Carlo study. Section 3.2 describes the setup used to simulate data and weight matrices. Section 3.3 presents the results of calculating correlations in the residuals of estimated models. Section 3.4 examines the properties of tests for spatial autocorrelation in the residuals. Section 3.5 presents the results of applying randomization tests. Section 3.6 gives results on the estimated value of the spatial parameter γ in the spatial panel data model.

3.2 Setup for Generated Data

Pseudo-random numbers are used to form datasets using two different setups. The first setup generates linear panel data and is given by

$$\begin{aligned}
&\boldsymbol{y} = 0.1\boldsymbol{x}_1 + 0.2\boldsymbol{x}_2 + \boldsymbol{Z}_\mu\boldsymbol{\mu} + \boldsymbol{Z}_\lambda\boldsymbol{\lambda} + \boldsymbol{v}\\
&\boldsymbol{v} \in N(\boldsymbol{0}, 0.1\boldsymbol{I}_M),\\
&\\
&\boldsymbol{x}_1 \in N(\boldsymbol{0}, \boldsymbol{I}_M), \boldsymbol{x}_2 \in N(\boldsymbol{0}, \boldsymbol{I}_M),\\
&\boldsymbol{\mu} \in N(\boldsymbol{0}, \boldsymbol{I}_N), \boldsymbol{\lambda} \in N(\boldsymbol{0}, \boldsymbol{I}_T),
\end{aligned} \tag{56}$$

where $\boldsymbol{y}$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{v}$ are M-vectors, $\boldsymbol{\mu}$ is an N-vector, $\boldsymbol{\lambda}$ is a T-vector, and $\boldsymbol{Z}_\mu$ and $\boldsymbol{Z}_\lambda$ are defined as in (5).

The second setup generates spatial panel data and is given by

$$\begin{aligned}
&\boldsymbol{y} = (\boldsymbol{I}_M - 0.3\boldsymbol{W}_M)^{-1}(0.1\boldsymbol{x}_1 + 0.2\boldsymbol{x}_2 + \boldsymbol{Z}_\mu\boldsymbol{\mu} + \boldsymbol{Z}_\lambda\boldsymbol{\lambda} + \boldsymbol{v})\\
&\boldsymbol{v} \in N(\boldsymbol{0}, 0.1\boldsymbol{I}_M),\\
&\\
&\boldsymbol{x}_1 \in N(\boldsymbol{0}, \boldsymbol{I}_M), \boldsymbol{x}_2 \in N(\boldsymbol{0}, \boldsymbol{I}_M),\\
&\boldsymbol{\mu} \in N(\boldsymbol{0}, \boldsymbol{I}_N), \boldsymbol{\lambda} \in N(\boldsymbol{0}, \boldsymbol{I}_T),
\end{aligned} \tag{57}$$

where $\boldsymbol{W}_M$ is constructed from a weight matrix $\boldsymbol{W}$ in the same way as in (14), i.e. as a $T \times T$ block matrix with the matrix $\boldsymbol{W}$ in the diagonal blocks.

In the study, N and T takes values in the set $\{5, 10, 25, 50\}$, giving a total of 16 combinations. The dynamic model can only be estimated where $N > T$, and is hence restricted to six combinations.

The weight matrix $\boldsymbol{W}$ used in the setup for spatial data is generated as a symmetric matrix where every row has the same number of non-zero cells, every row sums to one, and the elements on the diagonal are all zero. The generated matrix is thereby similar to a normalized contiguity matrix. The number of non-zero weights is set to two for simulations where $N = \{5, 10\}$ and four for simulations where $N = \{25, 50\}$.

3.3 Spatial Correlation in Residuals

The presence of residual correlation is studied by first forming a random linear panel using the setup in (56). Next, the three panel data models are estimated and the residuals of each model calculated. Finally, the mean empirical correlation between individuals is calculated using the following formula:

$$\frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}(\hat{\boldsymbol{v}}_i,\hat{\boldsymbol{v}}_j), \tag{58}$$

where $\hat{\rho}(\hat{\boldsymbol{v}}_i,\hat{\boldsymbol{v}}_j)$ is the correlation between the residuals of individual i and the residuals of individual j, given by

$$\hat{\rho}(\hat{\boldsymbol{v}}_i,\hat{\boldsymbol{v}}_j)=\frac{\hat{\boldsymbol{v}}_i'\hat{\boldsymbol{v}}_j}{\sqrt{\hat{\boldsymbol{v}}_i'\hat{\boldsymbol{v}}_i\hat{\boldsymbol{v}}_j'\hat{\boldsymbol{v}}_j}} \tag{59}$$

for the vectors $\boldsymbol{v}_i'=(\hat{v}_{i1},...,\hat{v}_{iT})$ and $\boldsymbol{v}_j'=(\hat{v}_{j1},...,\hat{v}_{jT})$. This procedure is repeated 1000 times for each N and T, and the mean value of the measure in (58) is taken for each N, T, and model.

The results are presented in Table 1. The first column gives the theoretical correlation calculated in (42), $-(N-1)^{-1}$, and the remaining three columns give Monte Carlo mean values of the correlation measures for each model. The mean values for the linear and spatial cases are close to the theoretical correlation for all sample sizes, and for fixed N the empirical means approach the theoretical correlation as T increases. The latter effect is possibly due to a downwards bias in the empirical correlation estimate. It is furthermore clear that the outcome for the linear model is very close to that of the spatial model, an expected result since the estimation of β in the spatial model is based on the OLS estimator and since the proportion of non-zero weights is comparatively smaller for larger N. The results for the dynamic model differ greatly from the theoretical OLS correlation. The mean empirical correlation is positive and decreases as the number of observations increase. A standard z-test of the mean values indicate that all mean correlations differ from zero on the five percent level except for the dynamic model at $(N,T)=(50,25)$.

The estimation of the spatial model relies on a weight matrix simulated according to the description in 3.2. The matrix is re-simulated for each of the 1000 runs. The use of a weight matrix in the estimation of the spatial model makes it possible to look specifically at mean residual correlation between individuals with a non-zero weight. Such pairs will be called connected, since the matrix may be seen as a simulated contiguity matrix. The measure is similar to that in (58), but instead of all possible pairs of individuals only pairs of connected individuals are used. The opposite restriction gives the correlation between non-connected individuals. The result of this decomposition is given in Table 2. The column for total gives the mean correlation when all possible pairs are used. The residual correlation between connected individuals is smaller in absolute than both the theoretical correlation and the total correlation. Correspondingly, the

N	T	Theoretical	Linear	Dynamic	Spatial
5	5	-0.2500	-0.2223		-0.2224
5	10	-0.2500	-0.2369		-0.2373
5	25	-0.2500	-0.2450		-0.2454
5	50	-0.2500	-0.2476		-0.2478
10	5	-0.1111	-0.0983	0.1075	-0.0983
10	10	-0.1111	-0.1052		-0.1053
10	25	-0.1111	-0.1089		-0.1089
10	50	-0.1111	-0.1100		-0.1100
25	5	-0.0417	-0.0369	0.0395	-0.0369
25	10	-0.0417	-0.0394	0.0127	-0.0394
25	25	-0.0417	-0.0408		-0.0408
25	50	-0.0417	-0.0412		-0.0412
50	5	-0.0204	-0.0180	0.0182	-0.0180
50	10	-0.0204	-0.0193	0.0046	-0.0193
50	25	-0.0204	-0.0200	-0.0000	-0.0200
50	50	-0.0204	-0.0202		-0.0202

Table 1: Mean empirical residual correlation for 1000 MC samples where the simulated data is a linear panel and the linear, dynamic, and spatial two-way error models are estimated.

correlation between non-connected individuals is larger in absolute than the theoretical correlation. Both the correlation between connected and the correlation between non-connected decrease with the number of observations. These results indicate that there is non-zero correlation in the residuals of the spatial model and that there is remaining residual correlation both between individuals which the weight matrix identify as connected and between individuals which are identified as non-connected.

3.4 Parametric Tests of Spatial Autocorrelation

Sections 2.6.2 and 2.6.3 introduce four tests for spatial autocorrelation in the residuals of the linear panel data model: the Pesaran and LM tests applied to the original residuals of the model, and the Pesaran and LM tests applied to the back-transformed residuals (using the method presented in Mao (2015)). The Monte Carlo setup for linear data (56) is used to examine the properties of these tests under the hypothesis that the original errors v_{it} are spatially independent. A linear panel is simulated and the residuals of an estimated two-way error linear panel data model are used to calculate the test statistics and p-values of the four tests. The procedure is repeated for 1000 MC draws.

Tables 3 and 4 give the empirical α-levels of the tests, i.e. the proportion of draws where the p-value of the test is lower than the theoretical α-level. The tests based on the model residuals are expected to have distorted sizes due to the presence of residual correlation. The Pesaran test shows large distortions in

N	T	Theoretical	Total	Connected	Non-connected
5	5	-0.2500	-0.2224	-0.0726	-0.3722
5	10	-0.2500	-0.2373	-0.0830	-0.3915
5	25	-0.2500	-0.2454	-0.0887	-0.4021
5	50	-0.2500	-0.2478	-0.0885	-0.4071
10	5	-0.1111	-0.0983	-0.0268	-0.1187
10	10	-0.1111	-0.1053	-0.0272	-0.1276
10	25	-0.1111	-0.1089	-0.0283	-0.1319
10	50	-0.1111	-0.1100	-0.0268	-0.1338
25	5	-0.0417	-0.0369	-0.0103	-0.0423
25	10	-0.0417	-0.0394	-0.0098	-0.0453
25	25	-0.0417	-0.0408	-0.0091	-0.0472
25	50	-0.0417	-0.0412	-0.0095	-0.0476
50	5	-0.0204	-0.0180	-0.0027	-0.0194
50	10	-0.0204	-0.0193	-0.0044	-0.0206
50	25	-0.0204	-0.0200	-0.0045	-0.0214
50	50	-0.0204	-0.0202	-0.0043	-0.0216

Table 2: Mean empirical residual correlation for 1000 MC samples where the simulated data is a spatial panel and the spatial two-way error model is estimated. Residual correlation given in total and decomposed into correlation between connected individuals and non-connected individuals.

size. For samples small in T, the test will give a large p-value, and for large T, the test will result in a low p-value, below the smallest presented level of 0.001. The LM test outperforms the Pesaran test for most N and T: the observed sizes are close to the desired α-level when N, T, and the ratio N/T are all relatively large.

As for the tests based on the back-transformed residuals, the Pesaran test gives empirical sizes that are close to the theoretical αs, at least for larger data sets, while the LM test performs well for cases where the ratio of N to T is small, but poorly for cases where N is close to or greater than T.

Next, the strengths of the tests are examined using the Monte Carlo setup for spatial data (57). For 1000 MC draws, a random spatial panel is simulated. The weight matrix is re-simulated for each draw. Next, the linear model is estimated and the test statistics and p-values of the tests are calculated. The resulting empirical α-values are given in Table 5 for the Pesaran and LM tests applied to the model residuals directly and in Table 6 for the Pesaran and LM tests on the back-transformed residuals.

The Pesaran test on the model residuals shows the same great distortions as in the study on size, while the LM test performs well in cases where T is large compared to N. For the Pesaran test on back-transformed residuals the results differ only slightly from the results on the linear data, suggesting that the test performs poorly under the alternative hypothesis. The LM test on back-transformed residuals performs comparatively better and shows high chances of rejecting a faulty null hypothesis for several of the sample sizes where the test also performed well under the null. A comparison between the LM test on model residuals and the LM test on back-transformed residuals indicates that the former performs better in size and that the tests are equal in strength.

In conclusion, the tests and sample sizes which result in α-levels close to the theoretical value and reasonable strength are the LM test on model residuals when N is greater than T and both are large, and the LM test on back-transformed residuals when N is smaller than T and both are large. The Pesaran test on model residuals suffers from distorted size due to the OLS residuals correlation and the Pesaran test on back-transformed residuals has low strength. As for the case in the coming empirical study, where N is large relative T and T is small, none of the studied tests show strong results.

3.5 Randomization Tests of Spatial Autocorrelation

Section 2.6.4 presented two randomization tests for residual autocorrelation. This section presents the MC results on size and strength of those tests.

In order to estimate the sizes, the tests are applied to data simulated by the Monte Carlo setup for linear panels (56). The Monte Carlo procedure is run 1000 times for each combination of (N, T), for each simulated data set a two-way error linear panel data model is estimated, and the randomization tests are applied to the residuals $\hat{\boldsymbol{v}}$. The number of random permutations K is set to 1000 for each test, giving that the lowest possible value of $\hat{p}$ is $\frac{2}{502} \approx 0.004$.

The estimated sizes of the local and global permutation tests are given in Table 7. As stated, the tests never give a p-value below 0.001. The local test performs poorly when N is small, an expected outcome due to the low number of possible permutations, but produces better results for higher N. The global test gives reasonable estimation of sizes for all (N, T), although there is indication that the size is underestimated for cases where T is small.

The strength of the tests are examined by simulating spatial panel data using the setup in (57). Random data and a random weight matrix are drawn 1000 times for each (N, T). For each simulation, the two-way error linear panel data model is estimated, and the local and global permutation tests (with $K = 1000$) are applied to the residuals series $\hat{\boldsymbol{v}}$.

A summary of the estimated p-values is presented in Table 8. The local test performs well for all (N, T), with increasing accuracy when N or T increases. However, recall that the local test is not appropriate for small N due to faulty size. It is also necessary to keep in mind that the weight matrix used in the local test is the same as the weight matrix used in the data generation. This is likely a major reason for the strong results for the local test. The results for the global test shows a dependence on T, with poor performance for small T and increasing strength as T increases. This is possibly an outcome of large variance in the correlation estimate for small samples.

3.6 The Spatial Parameter in the Two-Way Error Model

This section presents results on the estimated value of γ, the spatial parameter in the spatial panel data model. First, the setup in (56) is used to generate linear panel data and then the spatial model, with a simulated weight matrix, is estimated. The data generation and estimation is repeated 1000 times for each value of N and T. Next, the procedure is repeated using the setup in (57) for spatial panel data. The weight matrix is re-simulated for each run and the same matrix is used for both the data generation and the model estimation. Finally, a one-way error spatial model is estimated on simulated one-way error data, i.e. data with $\boldsymbol{\lambda} = \mathbf{0}$. These results are used for a comparison between the two-way error model and the one-way error model.

Table 9 presents the means of the γ estimates from the Monte Carlo sample. Each row gives a combination of N and T and each column gives one of the data setups (linear data, spatial data with two-way error, and spatial data with one-way error). For the linear data simulated by (56), γ is clearly affected by the residual correlation in the two-way error model. The parameter estimate decreases as N increases and takes values close to the theoretical correlation $-(N-1)^{-1}$. The results are less clear for the spatial data simulated by (57) and estimated with the two-way error spatial model. For small N and T the estimate is close to the theoretical correlation of OLS residuals and, as N and T increase, the parameter estimate approaches the value from the simulated data, 0.3. For the one-way spatial data estimated with the one-way error spatial model, the estimate of the spatial parameter is close to the parameter value of the simulation, even for small N. For each N the estimate approaches 0.3 as T increases.

The MC setup for a random spatial panel (57) is also used to estimate the mean residual correlation between individuals. For each draw a random panel is simulated and estimated using the two-way error spatial model. Table 10 gives the total mean correlation between all possible pairs of individuals, the mean correlation between connected individuals as given by the weight matrix, and the mean correlation between non-connected individuals. The results are similar to the results for the simulated linear data (see Table 2): the total correlation is close to the theoretical correlation, the correlation between connected individuals is close to zero but exhibits some structure in that all mean values are negative, and the correlation between non-connected individuals is highly negative. These results suggest that the residuals of a two-way error spatial model will show global and local spatial autocorrelation even if the model is a good fit for the data.

3.7 Summary of Monte Carlo Study

The Monte Carlo study focuses on four questions: the presence and magnitude of spatial correlation in the residuals of the three models (linear, dynamic and spatial), the performance of parametric tests for autocorrelation, the performance of randomization tests for autocorrelation, and the effect of residual correlation on the estimate of the spatial parameter in the spatial model.

The study indicates that residuals of the linear and spatial models are spatially correlated. This supports the theoretical calculation on the correlation structure in the residuals. The residuals of the dynamic model also show high spatial autocorrelation. There is however a clear difference in magnitude.

The study on parametric tests show a clear effect of the residual correlation structure. If the correlation structure in the residuals is not compensated for, the parametric tests can show large distortions in size and strength. The best performing tests are the LM-test and the LM-test on back-tranformed residuals, and only in the case of a large number of observed units N and a large number of time points T. Performance of the LM-test on back-transformed residuals also depends on the ratio T/N being large.

The randomization tests show a clear dependence on the number of observations. The local test does not give reliable results for small N, but performs better for larger datasets. The global test shows low strength for small T, but performs better as T increases.

Finally, the Monte Carlo study shows that the spatial parameter γ in the spatial model is clearly connected to the correlation in residuals of OLS estimated models. The results show that tests of the spatial parameter will give significant results even if there is no spatial component in the data.

Test	N	T	0.001	0.01	0.05	0.25	0.5
Pesaran	5	5	0.000	0.000	0.000	0.971	0.996
	5	10	0.000	0.000	0.993	1.000	1.000
	5	25	1.000	1.000	1.000	1.000	1.000
	5	50	1.000	1.000	1.000	1.000	1.000
	10	5	0.000	0.000	0.000	0.967	1.000
	10	10	0.000	0.000	0.986	1.000	1.000
	10	25	1.000	1.000	1.000	1.000	1.000
	10	50	1.000	1.000	1.000	1.000	1.000
	25	5	0.000	0.000	0.000	0.958	1.000
	25	10	0.000	0.000	0.982	1.000	1.000
	25	25	1.000	1.000	1.000	1.000	1.000
	25	50	1.000	1.000	1.000	1.000	1.000
	50	5	0.000	0.000	0.000	0.954	0.997
	50	10	0.000	0.000	0.990	1.000	1.000
	50	25	1.000	1.000	1.000	1.000	1.000
	50	50	1.000	1.000	1.000	1.000	1.000
LM	5	5	0.018	0.044	0.113	0.279	0.531
	5	10	0.068	0.146	0.239	0.455	0.650
	5	25	0.528	0.801	0.956	1.000	1.000
	5	50	1.000	1.000	1.000	1.000	1.000
	10	5	0.013	0.035	0.067	0.220	0.468
	10	10	0.020	0.040	0.091	0.278	0.518
	10	25	0.052	0.136	0.270	0.537	0.730
	10	50	0.339	0.584	0.771	0.949	0.985
	25	5	0.007	0.016	0.037	0.188	0.447
	25	10	0.006	0.024	0.063	0.242	0.466
	25	25	0.005	0.028	0.075	0.298	0.536
	25	50	0.017	0.085	0.201	0.467	0.680
	50	5	0.001	0.013	0.037	0.174	0.442
	50	10	0.004	0.022	0.056	0.240	0.502
	50	25	0.003	0.011	0.047	0.229	0.480
	50	50	0.007	0.025	0.077	0.295	0.544

Table 3: Results of MC study of sizes of tests for residual spatial autocorrelation. Tests are on the residuals of the linear panel data model applied to simulated linear data. For each test and (N, T) the results specify the proportion of simulated p-values below the α-level given in the header.

Test	N	T	0.001	0.01	0.05	0.25	0.5
Pesaran, back-transformed	5	5	0.013	0.045	0.083	0.310	0.589
	5	10	0.005	0.020	0.066	0.299	0.562
	5	25	0.002	0.017	0.047	0.232	0.483
	5	50	0.001	0.008	0.041	0.249	0.525
	10	5	0.012	0.027	0.056	0.252	0.555
	10	10	0.006	0.019	0.044	0.268	0.538
	10	25	0.003	0.014	0.058	0.260	0.512
	10	50	0.000	0.012	0.062	0.270	0.528
	25	5	0.015	0.036	0.064	0.231	0.544
	25	10	0.011	0.021	0.052	0.244	0.507
	25	25	0.004	0.010	0.042	0.245	0.504
	25	50	0.000	0.009	0.041	0.244	0.489
	50	5	0.008	0.028	0.071	0.242	0.564
	50	10	0.005	0.020	0.050	0.266	0.532
	50	25	0.004	0.016	0.058	0.266	0.504
	50	50	0.001	0.009	0.045	0.236	0.494
LM, back-transformed	5	5	0.042	0.075	0.157	0.338	0.575
	5	10	0.012	0.031	0.070	0.251	0.494
	5	25	0.012	0.027	0.058	0.234	0.503
	5	50	0.008	0.025	0.054	0.209	0.486
	10	5	0.131	0.242	0.420	0.734	0.899
	10	10	0.025	0.053	0.108	0.285	0.511
	10	25	0.004	0.021	0.056	0.237	0.496
	10	50	0.004	0.011	0.052	0.267	0.521
	25	5	1.000	1.000	1.000	1.000	1.000
	25	10	0.089	0.207	0.381	0.682	0.843
	25	25	0.003	0.031	0.097	0.289	0.525
	25	50	0.002	0.013	0.068	0.271	0.526
	50	5	1.000	1.000	1.000	1.000	1.000
	50	10	0.551	0.808	0.944	0.996	1.000
	50	25	0.029	0.090	0.185	0.474	0.679
	50	50	0.009	0.028	0.099	0.329	0.585

Table 4: Results of MC study of sizes of tests for residual spatial autocorrelation. Tests are on the residuals of the linear panel data model applied to simulated linear data. For each test and (N, T) the results specify the proportion of simulated p-values below the α-level given in the header.

Test	N	T	0.001	0.01	0.05	0.25	0.5
Pesaran	5	5	0.000	0.000	0.000	0.979	0.999
	5	10	0.000	0.000	0.997	1.000	1.000
	5	25	1.000	1.000	1.000	1.000	1.000
	5	50	1.000	1.000	1.000	1.000	1.000
	10	5	0.000	0.000	0.000	0.967	0.999
	10	10	0.000	0.000	0.988	1.000	1.000
	10	25	1.000	1.000	1.000	1.000	1.000
	10	50	1.000	1.000	1.000	1.000	1.000
	25	5	0.000	0.000	0.000	0.943	1.000
	25	10	0.000	0.000	0.983	1.000	1.000
	25	25	1.000	1.000	1.000	1.000	1.000
	25	50	1.000	1.000	1.000	1.000	1.000
	50	5	0.000	0.000	0.000	0.954	0.997
	50	10	0.000	0.000	0.985	1.000	1.000
	50	25	1.000	1.000	1.000	1.000	1.000
	50	50	1.000	1.000	1.000	1.000	1.000
LM	5	5	0.023	0.064	0.129	0.322	0.559
	5	10	0.135	0.238	0.363	0.593	0.747
	5	25	0.844	0.955	0.995	1.000	1.000
	5	50	1.000	1.000	1.000	1.000	1.000
	10	5	0.015	0.033	0.074	0.253	0.508
	10	10	0.070	0.142	0.244	0.469	0.657
	10	25	0.532	0.718	0.834	0.938	0.972
	10	50	0.994	0.999	1.000	1.000	1.000
	25	5	0.009	0.017	0.037	0.184	0.446
	25	10	0.025	0.061	0.135	0.344	0.575
	25	25	0.151	0.290	0.466	0.720	0.851
	25	50	0.717	0.855	0.941	0.985	0.995
	50	5	0.011	0.022	0.069	0.232	0.498
	50	10	0.019	0.050	0.107	0.321	0.545
	50	25	0.123	0.250	0.437	0.699	0.841
	50	50	0.649	0.824	0.924	0.980	0.996

Table 5: Results of MC study of strength of tests for residual spatial autocorrelation when simulated data is a spatial panel with $\gamma = 0.3$. For each test and (N, T), the results specify the proportion of simulated p-values below the α-level given in the header.

Test	N	T	0.001	0.01	0.05	0.25	0.5
Pesaran, back-transformed	5	5	0.010	0.022	0.061	0.299	0.579
	5	10	0.003	0.014	0.052	0.287	0.550
	5	25	0.007	0.019	0.082	0.324	0.557
	5	50	0.004	0.030	0.108	0.356	0.587
	10	5	0.019	0.034	0.070	0.282	0.566
	10	10	0.007	0.026	0.064	0.280	0.570
	10	25	0.004	0.032	0.097	0.326	0.590
	10	50	0.013	0.046	0.128	0.362	0.627
	25	5	0.017	0.036	0.070	0.242	0.567
	25	10	0.010	0.029	0.055	0.286	0.556
	25	25	0.004	0.022	0.051	0.261	0.516
	25	50	0.008	0.020	0.058	0.291	0.539
	50	5	0.021	0.051	0.075	0.259	0.557
	50	10	0.010	0.022	0.053	0.276	0.535
	50	25	0.006	0.015	0.052	0.272	0.529
	50	50	0.007	0.015	0.046	0.262	0.521
LM, back-transformed	5	5	0.049	0.096	0.170	0.373	0.603
	5	10	0.029	0.057	0.127	0.326	0.548
	5	25	0.084	0.145	0.252	0.446	0.649
	5	50	0.254	0.385	0.507	0.719	0.831
	10	5	0.149	0.290	0.475	0.771	0.922
	10	10	0.061	0.107	0.225	0.477	0.663
	10	25	0.150	0.280	0.425	0.652	0.800
	10	50	0.517	0.691	0.810	0.934	0.971
	25	5	1.000	1.000	1.000	1.000	1.000
	25	10	0.183	0.348	0.526	0.788	0.900
	25	25	0.131	0.268	0.440	0.682	0.827
	25	50	0.382	0.600	0.761	0.913	0.971
	50	5	1.000	1.000	1.000	1.000	1.000
	50	10	0.698	0.900	0.971	0.999	1.000
	50	25	0.273	0.487	0.674	0.870	0.944
	50	50	0.569	0.773	0.891	0.973	0.992

Table 6: Results of MC study of strength of tests for residual spatial autocorrelation when simulated data is a spatial panel with $\gamma = 0.3$. For each test and (N, T), the results specify the proportion of simulated p-values below the α-level given in the header.

Test	N	T	0.001	0.01	0.05	0.25	0.5
Local	5	5	0.000	0.000	0.000	0.183	0.425
	5	10	0.000	0.000	0.000	0.176	0.426
	5	25	0.000	0.000	0.000	0.154	0.407
	5	50	0.000	0.000	0.000	0.164	0.413
	10	5	0.000	0.009	0.049	0.267	0.529
	10	10	0.000	0.002	0.041	0.262	0.500
	10	25	0.000	0.009	0.038	0.248	0.510
	10	50	0.000	0.008	0.034	0.229	0.491
	25	5	0.000	0.008	0.043	0.250	0.493
	25	10	0.000	0.007	0.040	0.241	0.489
	25	25	0.000	0.006	0.039	0.245	0.482
	25	50	0.000	0.008	0.042	0.253	0.491
	50	5	0.000	0.008	0.052	0.253	0.495
	50	10	0.000	0.011	0.051	0.271	0.514
	50	25	0.000	0.007	0.038	0.238	0.477
	50	50	0.000	0.008	0.043	0.241	0.495
Global	5	5	0.000	0.006	0.036	0.208	0.455
	5	10	0.000	0.005	0.037	0.248	0.475
	5	25	0.000	0.011	0.059	0.246	0.472
	5	50	0.000	0.012	0.040	0.231	0.486
	10	5	0.000	0.002	0.027	0.205	0.473
	10	10	0.000	0.005	0.049	0.239	0.482
	10	25	0.000	0.015	0.049	0.221	0.490
	10	50	0.000	0.009	0.056	0.249	0.497
	25	5	0.000	0.013	0.051	0.195	0.476
	25	10	0.000	0.012	0.043	0.252	0.485
	25	25	0.000	0.015	0.052	0.255	0.515
	25	50	0.000	0.010	0.050	0.240	0.495
	50	5	0.000	0.007	0.039	0.205	0.468
	50	10	0.000	0.002	0.043	0.242	0.503
	50	25	0.000	0.008	0.041	0.233	0.489
	50	50	0.000	0.010	0.045	0.261	0.511

Table 7: Results of MC study of sizes of randomization tests for residual spatial autocorrelation. Tests are on the residuals of the linear panel data model applied to simulated linear data. For each test and (N, T) the results specify the proportion of simulated p-values below the α-level given in the header.

Test	N	T	0.001	0.01	0.05	0.25	0.5
Local	5	5	0.000	0.000	0.000	0.252	0.532
	5	10	0.000	0.000	0.000	0.392	0.662
	5	25	0.000	0.000	0.000	0.653	0.869
	5	50	0.000	0.000	0.000	0.885	0.976
	10	5	0.000	0.138	0.324	0.637	0.807
	10	10	0.000	0.355	0.624	0.881	0.953
	10	25	0.000	0.879	0.973	1.000	1.000
	10	50	0.000	1.000	1.000	1.000	1.000
	25	5	0.000	0.344	0.593	0.819	0.920
	25	10	0.000	0.696	0.887	0.968	0.990
	25	25	0.000	0.991	0.998	1.000	1.000
	25	50	0.000	1.000	1.000	1.000	1.000
	50	5	0.000	0.698	0.851	0.971	0.991
	50	10	0.000	0.984	1.000	1.000	1.000
	50	25	0.000	1.000	1.000	1.000	1.000
	50	50	0.000	1.000	1.000	1.000	1.000
Global	5	5	0.000	0.008	0.037	0.216	0.472
	5	10	0.000	0.008	0.052	0.258	0.533
	5	25	0.000	0.070	0.170	0.455	0.676
	5	50	0.000	0.252	0.488	0.775	0.889
	10	5	0.000	0.004	0.036	0.221	0.487
	10	10	0.000	0.029	0.091	0.304	0.543
	10	25	0.000	0.182	0.389	0.654	0.811
	10	50	0.000	0.692	0.861	0.967	0.988
	25	5	0.000	0.011	0.041	0.244	0.459
	25	10	0.000	0.016	0.060	0.276	0.515
	25	25	0.000	0.096	0.248	0.535	0.738
	25	50	0.000	0.542	0.736	0.899	0.959
	50	5	0.000	0.007	0.036	0.205	0.457
	50	10	0.000	0.016	0.076	0.293	0.559
	50	25	0.000	0.144	0.306	0.577	0.767
	50	50	0.000	0.603	0.800	0.944	0.981

Table 8: Results of MC study of strength of randomization tests for residual spatial autocorrelation when simulated data is a spatial panel with $\gamma = 0.3$. For each test and (N, T) the results specify the proportion of simulated p-values below the α-level given in the header.

		Linear data	Spatial data	
N	T	Two-way	Two-way	One-way
5	5	-0.2364	-0.0492	0.2736
5	10	-0.2398	-0.0519	0.2918
5	25	-0.2473	-0.0469	0.3003
5	50	-0.2515	-0.0440	0.3002
10	5	-0.0943	0.1768	0.2842
10	10	-0.1012	0.1821	0.2942
10	25	-0.0977	0.1864	0.2975
10	50	-0.0972	0.1861	0.2985
25	5	-0.0830	0.2156	0.2877
25	10	-0.0733	0.2151	0.2955
25	25	-0.0740	0.2232	0.2959
25	50	-0.0765	0.2253	0.2978
50	5	-0.0378	0.2650	0.2939
50	10	-0.0361	0.2618	0.2979
50	25	-0.0360	0.2666	0.2989
50	50	-0.0341	0.2673	0.3001

Table 9: Mean of estimates of γ over 1000 MC samples for the simulated linear data estimated with a two-way error spatial model, for the simulated two-way error spatial data estimated with a two-way error spatial model, and for the simulated one-way error data estimated with a one-way error spatial model.

N	T	Induced	Total	Connected	Non-connected
5	5	-0.2500	-0.2209	-0.0467	-0.3951
5	10	-0.2500	-0.2384	-0.0483	-0.4284
5	25	-0.2500	-0.2456	-0.0540	-0.4371
5	50	-0.2500	-0.2479	-0.0562	-0.4396
10	5	-0.1111	-0.0982	-0.0099	-0.1234
10	10	-0.1111	-0.1053	-0.0109	-0.1322
10	25	-0.1111	-0.1088	-0.0131	-0.1362
10	50	-0.1111	-0.1100	-0.0104	-0.1385
25	5	-0.0417	-0.0367	-0.0059	-0.0429
25	10	-0.0417	-0.0394	-0.0044	-0.0464
25	25	-0.0417	-0.0408	-0.0039	-0.0482
25	50	-0.0417	-0.0412	-0.0043	-0.0486
50	5	-0.0204	-0.0181	-0.0008	-0.0196
50	10	-0.0204	-0.0193	-0.0025	-0.0208
50	25	-0.0204	-0.0200	-0.0026	-0.0215
50	50	-0.0204	-0.0202	-0.0024	-0.0218

Table 10: Mean empirical correlation for the residuals of the two-way error spatial model applied to simulated spatial data. Mean correlation in total and decomposed into correlation between connected individuals and non-connected individuals.

4 Data for Empirical Study

4.1 Data Selection and Transformations

The collected data covers eight variables for 33 municipalities in southern Sweden and 5 years, from 2008 to 2012.

The explained variable unemployment (*une*) is the number of people in unemployment at the end of the year divided by the number of people of working age. The data is collected from the Swedish employment agency, Arbetsförmedlingen (2016), and is, for each municipality and year, calculated as the sum of people openly unemployed and people in active programs, divided by the population in the 20-64 age span.

Of the seven explanatory variables, three are connected to demography: population, proportion of young adults, and proportion of people of working age. All three variables are collected from the demography statistics of Statistics Sweden (2016b). The log-population (*pop*) is the logarithm of the number of people living in the municipality at the end of the year. The proportion of young adults (*you*) is the number of people in the 20-24 age span divided by the population, and proportion of people of working age (*wor*) is the number of people in the 25-64 age span divided by the population.

The remaining four explanatory variables are connected to socio-economic factors and are given by income levels, proportion of population from another country, crime levels, and education levels. The log-income (*inc*) is taken as the logarithm of the mean income of employment for the specific municipality and year. Data is collected from Statistics Sweden (2016a). The proportion of the population with a background in a country other than Sweden (*foe*) is calculated as the number of people who are either born in another country or have both parents born in another country, divided by the population; data from Statistics Sweden (2016b). The level of crime (*cri*) is taken as the ratio between the number of reported crimes and the population. Crime statistics are collected from The Swedish National Council for Crime Prevention (2016). Education (*edu*) is measured as the proportion of the population with some tertiary education, i.e. university level or similar. Education data is from Statistics Sweden (2016d).

Maps and weight matrices are constructed from geographical data in the RT90 coordinate system (Statistics Sweden, 2016c).

4.2 Descriptive Statistics

Table 12 presents the mean value across municipalities for each variable and each year; Table 13 presents the corresponding standard deviations. Several variables show a strictly increasing or decreasing yearly trend, indicating a time component in the data. Unemployment shows a large increase in mean between 2008 and 2009, with a corresponding increase in standard deviation. This increase is evident in Figure 2, which also displays a spatial pattern where municipalities to the north and east show high unemployment figures.

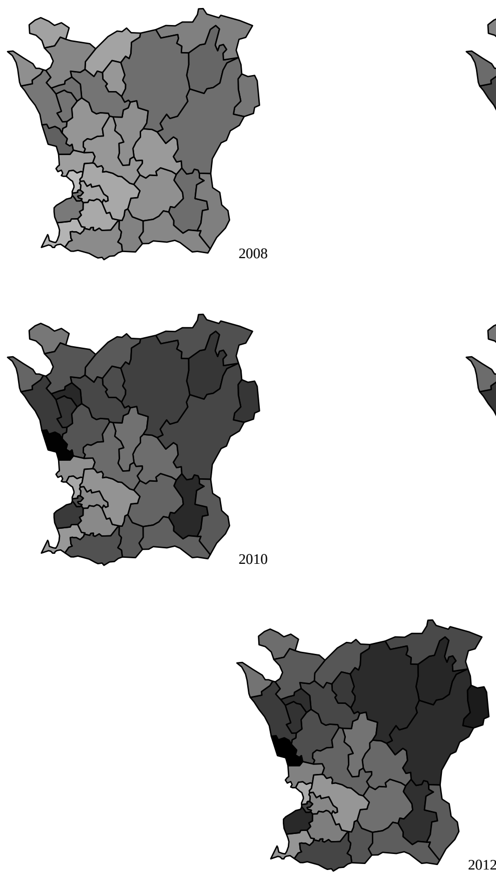

Figure 2: Maps of unemployment is southern Sweden 2008-2012.

The nature of panel data - the presence of both a temporal and a cross-sectional dimension - makes it possible to examine correlation between variables in three different ways: using the full dataset with NT observations; using datasets of N regions for a fixed year; and using datasets of T years for a fixed region.

Table 14 presents correlation for the full dataset. Several variable pairings show high correlation, including unemployment which is positively correlated to year, proportion with foreign background, and criminality, and negatively correlated with income and education levels.

Variable	Name	Mean	StDev
Unemployment	une	0.071	0.022
Log-population	pop	10.101	0.794
Youth	you	0.059	0.014
Working age	wor	0.504	0.016
Log-income	inc	5.160	0.106
Foreign background	foe	0.164	0.078
Criminality	cri	0.122	0.033
Education	edu	0.192	0.065

Table 11: Explained variable (une) and explanatory variables.

	2008	2009	2010	2011	2012
une	0.049	0.075	0.076	0.075	0.079
pop	10.087	10.096	10.102	10.107	10.110
you	0.055	0.057	0.060	0.061	0.062
wor	0.512	0.507	0.503	0.500	0.497
inc	5.123	5.137	5.149	5.180	5.210
foe	0.154	0.160	0.165	0.168	0.172
cri	0.124	0.128	0.118	0.124	0.115
edu	0.183	0.187	0.192	0.197	0.201

Table 12: Mean over municipalities for each year and variable.

	2008	2009	2010	2011	2012
une	0.013	0.019	0.021	0.022	0.023
pop	0.798	0.802	0.804	0.806	0.810
you	0.014	0.014	0.014	0.014	0.014
wor	0.015	0.015	0.015	0.016	0.017
inc	0.100	0.101	0.102	0.105	0.103
foe	0.076	0.078	0.080	0.081	0.082
cri	0.027	0.034	0.032	0.036	0.033
edu	0.064	0.065	0.065	0.066	0.066

Table 13: Standard deviations over municipalities for each year and variable.

Table 15 shows the mean of the five correlation matrices given by estimating correlations for each year separately. Unemployment shows the same correlation pattern as for the full dataset with high positive correlation with proportion of population with foreign background and criminality, and high negative correlation with income and education. This estimate of correlation, where time is fixed, gives a measure of correlation between regions and is therefore sensitive to extreme values. This is a possible explanation for some of the higher correlations.

Finally, Table 16 presents the mean of the 33 correlation matrices calculated by setting the region fixed. The results show very large correlations: unemployment has an absolute correlation above 0.6 for all variables but one. This is likely an effect of the small amount of data - five observations for each region - and the linear trend which was noted in the time series on means and is evident in that all variables are correlated with year.

The relation between unemployment and income levels is further illustrated in Figure 3. The full data and the data for each single year indicates negative correlation, while the lines connecting the earliest and latest observations for each municipality indicating positive correlation. Hence, for any fixed year, unemployment and income are negatively correlated, but for any fixed region unemployment and income are positively correlated. This further supports the notion that the high correlations for several variables pairs when region is fixed (Table 16), is due to a time trend.

The temporal aspect of the data also makes it possible to examine the relation between volatility in unemployment rate and a variable. Figure 4 (left) gives the standard deviation of the unemployment rate for each municipality as a function of the unemployment rate in 2012. The plot shows a clear positive relation. Furthermore, there is a possible connection between volatility of the unemployment rate and the population size. Figure 4 (right) shows the standard deviation of the unemployment rate for each municipality as a function of the logarithm of the population size in 2012. The plot indicates a weak negative relation between the variables.

	year	une	pop	you	wor	inc	foe	cri	edu
year	1.00								
une	0.38	1.00							
pop	0.01	0.04	1.00						
you	0.17	0.24	0.48	1.00					
wor	-0.32	-0.04	0.48	0.29	1.00				
inc	0.29	-0.50	0.21	-0.23	-0.30	1.00			
foe	0.08	0.45	0.45	0.51	0.50	-0.37	1.00		
cri	-0.10	0.42	0.39	0.49	0.53	-0.55	0.81	1.00	
edu	0.10	-0.47	0.63	0.34	0.11	0.71	0.04	-0.11	1.00

Table 14: Correlation matrix for the full variable set and year.

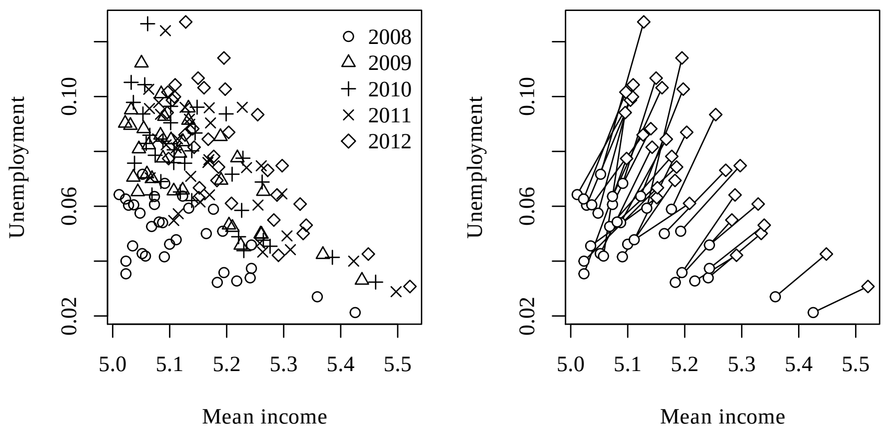

Figure 3: (Left) Unemployment as a function of income level separated by year. (Right) Unemployment as a function of income level for 2008 and 2012 with lines connecting the observations for each municipality.

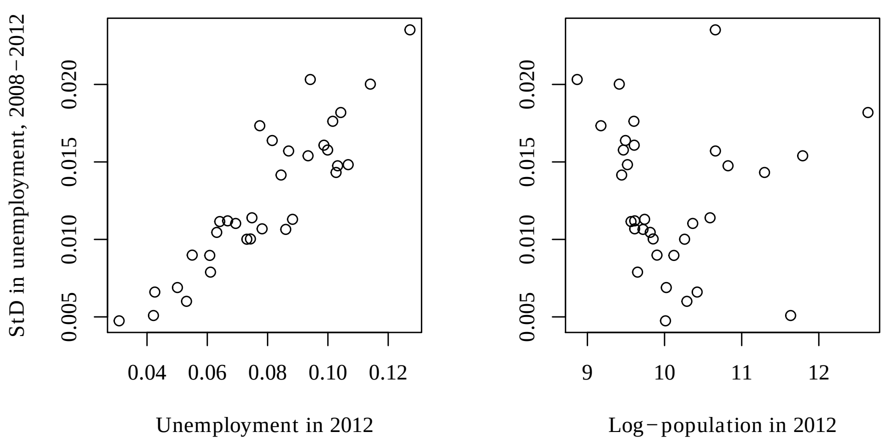

Figure 4: (Left) Standard deviation of unemployment rate for each municipality as a function of unemployment rate. (Right) Standard deviation of unemployment rate for each municipality as a function of log-population.

	une	pop	you	wor	inc	foe	cri	edu
une	1.00							
pop	0.05	1.00						
you	0.20	0.49	1.00					
wor	0.09	0.51	0.37	1.00				
inc	-0.72	0.21	-0.29	-0.22	1.00			
foe	0.48	0.45	0.50	0.56	-0.41	1.00		
cri	0.52	0.40	0.53	0.53	-0.55	0.84	1.00	
edu	-0.59	0.64	0.33	0.16	0.72	0.03	-0.10	1.00

Table 15: Mean of correlation matrices when correlation is calculated for each year separately.

	year	une	pop	you	wor	inc	foe	cri	edu
year	1.00								
une	0.72	1.00							
pop	0.78	0.65	1.00						
you	0.89	0.69	0.75	1.00					
wor	-0.93	-0.67	-0.76	-0.85	1.00				
inc	0.97	0.60	0.72	0.80	-0.89	1.00			
foe	0.98	0.75	0.79	0.90	-0.92	0.93	1.00		
cri	-0.36	-0.18	-0.24	-0.31	0.29	-0.33	-0.36	1.00	
edu	0.99	0.71	0.79	0.89	-0.92	0.96	0.97	-0.36	1.00

Table 16: Mean of correlation matrices when correlation is calculated for each municipality separately.

5 Results of Empirical Study

5.1 Estimation Results for the Linear Model

The linear model (4) is estimated using the within estimator. Estimated parameters $\hat{\boldsymbol{\beta}}$ and goodness-of-fit are presented in Table 17. The outcome of hypothesis tests are presented in Table 18 and in Table 19. Diagnostic plots for residuals are given in Figure 5.

The t-tests of parameters show that population size, size of work force, and income have significant impact ($\alpha = 0.05$), with size of work force being the single significant variable with a positive relation to unemployment. Measure for goodness-of-fit give a high R_t^2. Most of the explanatory power of the model is due to individual and time effects ($R_e^2 = 0.95$).

Hypothesis testing suggests that several model assumptions are violated: the Wooldridge test points to the presence of serial dependence and the Breusch-Pagan test strongly indicates heteroskedasticity. Furthermore, the goodness-of-fit presented in Table 19 show high collinearity among the explanatory variables and that some explanatory variables are to a great extent explained by region and time alone. This suggests that some explanatory variables can be dropped without significant loss of explanatory power. The Wald tests for explanatory variables, individual effects, and time effects all reject the null hypothesis, indicating that all three factors have a significant contribution.

The previous Monte Carlo study gives that the only test for spatial autocorrelation which has both reasonable size and strength for the current sample size is the local randomization test. When applied to the linear model, the local randomization test accepts the null hypothesis for both the contiguity weight matrix $\boldsymbol{W}_C$ and the inverted distance matrix $\boldsymbol{W}_D$, indicating that there is no spatial autocorrelation in the residuals. Recall however, that the Monte Carlo results on the strength of the local test is for the ideal case where the weight matrix of the data is perfectly known.

The residual plots in Figure 5 show that the residuals follow a normal distribution rather well. The maps do not show any clear remaining spatial correlation.

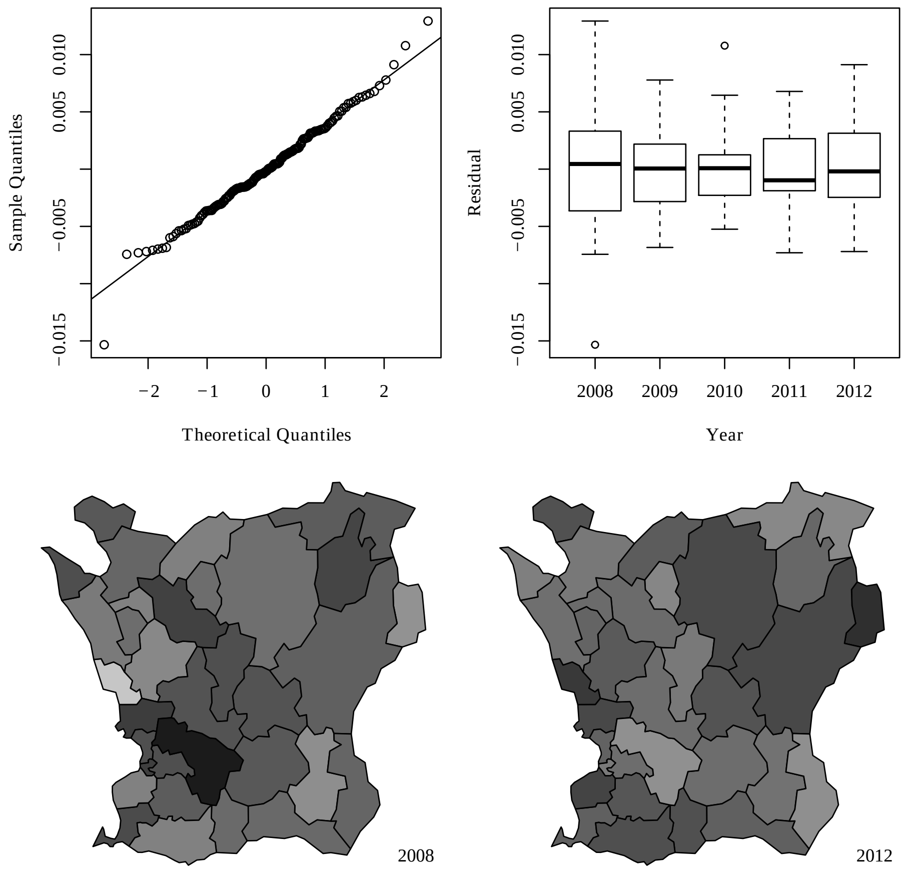

Figure 5: Plots for residuals of the linear panel data model with individual and time effects: QQ-plot; box plot for each year; maps for 2008 and 2012.

	Param	St.err.	P-value
pop	-0.1202	0.0528	0.0245
you	0.1286	0.3099	0.6790
wor	0.5636	0.1675	0.0010
inc	-0.2643	0.0644	0.0001
foe	0.0648	0.1686	0.7013
cri	0.0051	0.0357	0.8869
edu	-0.3765	0.2807	0.1823
σ_v	0.0046		
R_t^2	0.9685		
R_w^2	0.3933		
R_e^2	0.9480		

Table 17: Estimation results for the linear panel data model with individual and time effects.

Assumption	Test	P-value
Serial indep.	AR(1)-Wooldridge	0.0058
Zero spatial correlation	Rand. test, local $\boldsymbol{W}_C$	0.0838
Zero spatial correlation	Rand. test, local $\boldsymbol{W}_D$	0.3593
Homoskedasticity	Breusch-Pagan	0.0110
Normality	Shapiro-Wilk	0.1659
Zero explanatory effect	Wald	0.0000
Zero individual effect	Wald	0.0000
Zero time effect	Wald	0.0000
Time-constant param.	Chow	0.2446

Table 18: Hypothesis tests for the linear panel data model with individual and time effects.

	R_t^2	R_w^2	R_e^2
pop	0.9999	0.1986	0.9999
you	0.9933	0.1553	0.9921
wor	0.9827	0.3564	0.9731
inc	0.9972	0.3985	0.9953
foe	0.9993	0.5081	0.9985
cri	0.9038	0.0640	0.8972
edu	0.9996	0.0979	0.9996

Table 19: R^2-measures for each variable when explained by all other explanatory variables. Models estimated using the OLS estimator.

5.2 Estimation Results for the Dynamic Model

The parameter estimation results for the dynamic model (8) are presented in Table 20, the results of hypothesis tests are given in Table 21, and plots for residuals are shown in Figure 6.

The t-tests for parameter significance give no significant variables, but relatively low p-values for population size, relative size of the young population, and criminality. All three have a negative relation to unemployment. The parameter of the lagged variable δ is clearly non-significant. The goodness-of-fit measures are low relative the corresponding measures in the linear model.

Hypothesis tests indicate that the assumptions on the lagged structure are incorrect. If the data fits the model, the first order Arellano test should, due to the first order lag in the estimation method, reject the null of no serial correlation, but here both the first and second order test accepts the null, although narrowly in the second order case. The Wald tests show that the explanatory variables are significant, but that the time dummies are not.

The local randomization test is applied using the two weight matrices. The resulting p-values are 0.481 for the contiguity weight matrix $\boldsymbol{W}_C$ and 0.029 for the distance weight matrix $\boldsymbol{W}_D$.

The plots in Figure 6 illustrate that the residuals deviate from the normal distribution in both tails. The maps suggest that there is remaining spatial autocorrelation: for 2010 eastern municipalities show lower residuals than western municipalities, while the opposite holds for 2012.

	Param	St.err.	P-value
δ	0.1708	0.3101	0.5818
pop	-0.1561	0.1147	0.1734
you	-1.1849	0.6117	0.0527
wor	-0.0548	0.1956	0.7794
inc	-0.0274	0.1125	0.8075
foe	0.2048	0.2715	0.4506
cri	-0.0605	0.0373	0.1050
edu	0.3861	0.4795	0.4207
σ_v	0.0051		
R_t^2	0.9445		
R_w^2	0.0658		

Table 20: Estimation results for the dynamic panel data model with individual and time effects.

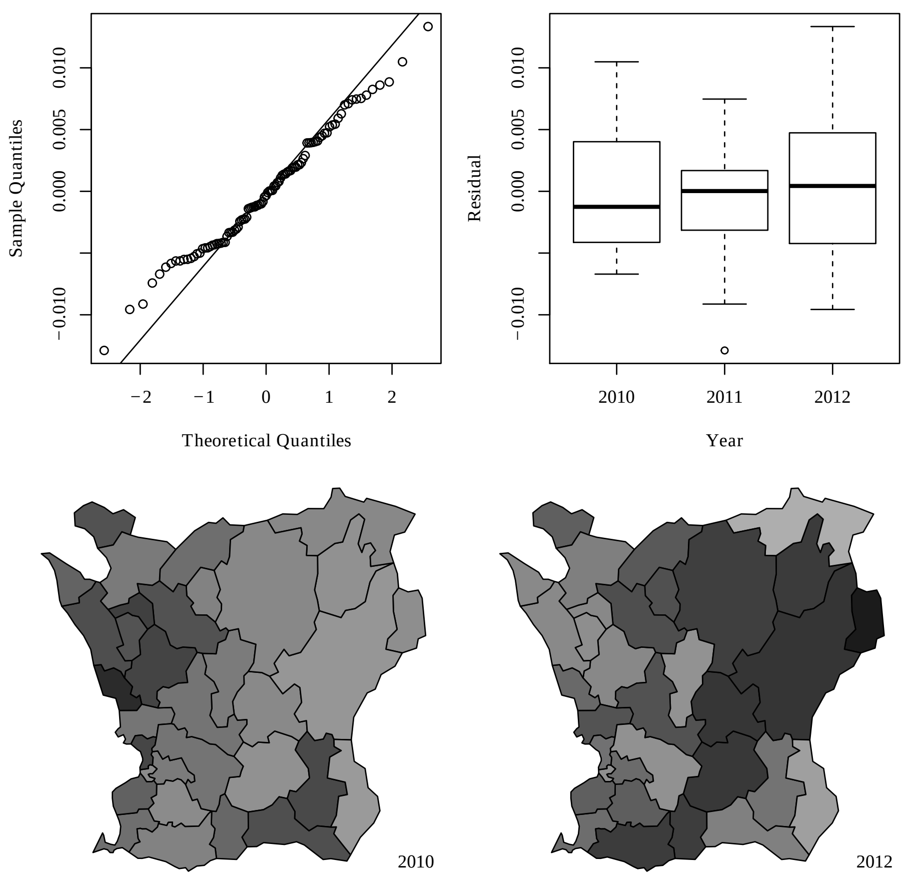

Figure 6: Plots for residuals of the dynamic panel data model with individual and time effects: QQ-plot; box plot for each year; maps for 2010 and 2012.

Assumption	Test	P-value
Serial indep.	Arellano, deg. 1	0.4427
Serial indep.	Arellano, deg. 2	0.0510
Zero spatial correlation	Rand. test, local $\boldsymbol{W}_C$	0.0299
Zero spatial correlation	Rand. test, local $\boldsymbol{W}_D$	0.0519
Normality	Shapiro-Wilk	0.4806
Zero explanatory effect	Wald	0.0290
Zero time effects	Wald	0.2447
Suitable instruments	Sargan	0.1037

Table 21: Hypothesis tests for the dynamic panel data model with individual and time effects.

5.3 Estimation Results for the Spatial Models

The spatial model specified in (14) is estimated using two different weight matrices: the contiguity matrix $\boldsymbol{W}_C$, defined in (9), and the inverted distance matrix $\boldsymbol{W}_D$, defined in (10). The estimation results are presented in Table 22 and Table 23. Residuals plots are given in Figure 7 and Figure 8.

For both models, the t-tests show significance for population size, relative size of work force, and income. The spatially lagged parameter is significant in the model with contiguity matrix ($\hat{\gamma} = 0.2$ with a p-value in a t-test of 0.03), but not significant in the model with the distance matrix ($\hat{\gamma} = 0.12$ with a p-value of 0.35). The parameter estimations differ greatly from the Monte Carlo results for simulated linear data estimated using a spatial model (see Table 9), suggesting that the empirical data of the study has structure which differs from the simulated linear data.

The goodness-of-fit measures show values close to the results from the linear model. The estimate of error variance σ_v^2 is smaller than the corresponding estimate in the linear model. This is due to the difference in estimation method: the maximum likelihood estimation of the variance does not correct for the loss of degrees of freedom.

The Shapiro-Wilk test does not indicate non-normality of the residuals in either model, and the LR tests give that explanatory variables and time effects are significant in both models.

The residual plots in Figure 7 and Figure 8 show great similarities between the two estimations. In both cases, the residuals deviate from the normal distribution in the tails. The residual maps do not indicate any clear remaining spatial autocorrelation.

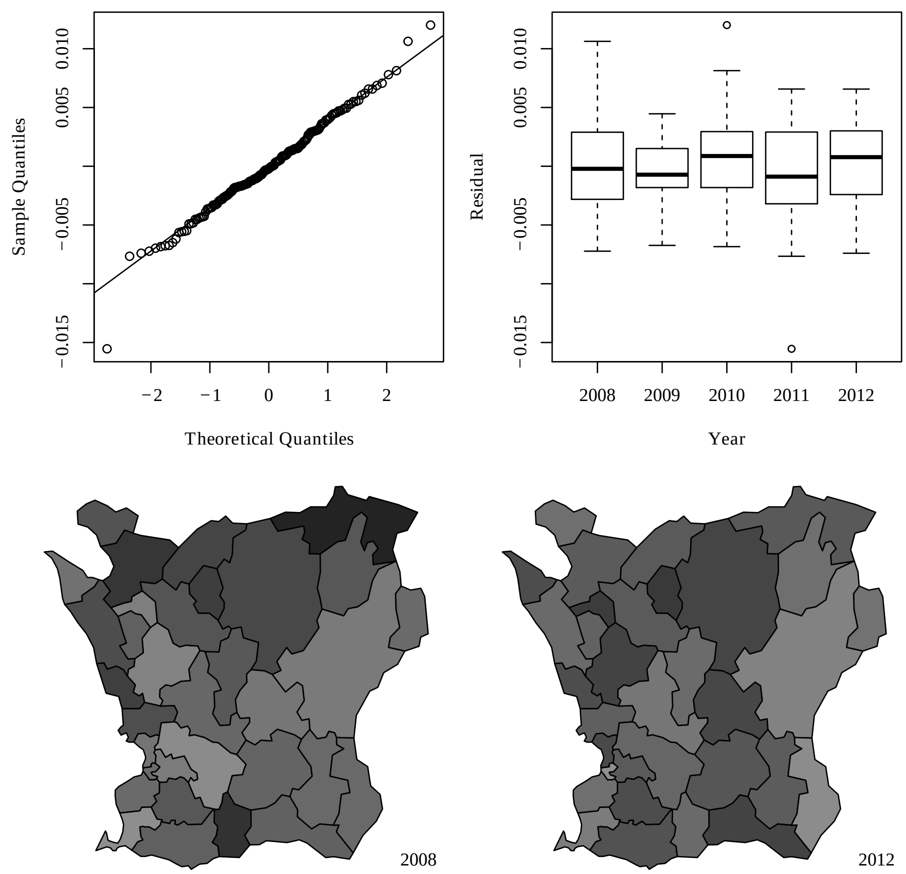

Figure 7: Plots for residuals of the spatial panel data model with contiguity weight matrix $\boldsymbol{W}_C$ and individual and time effects: QQ-plot; box plot for each year; maps for 2008 and 2012.

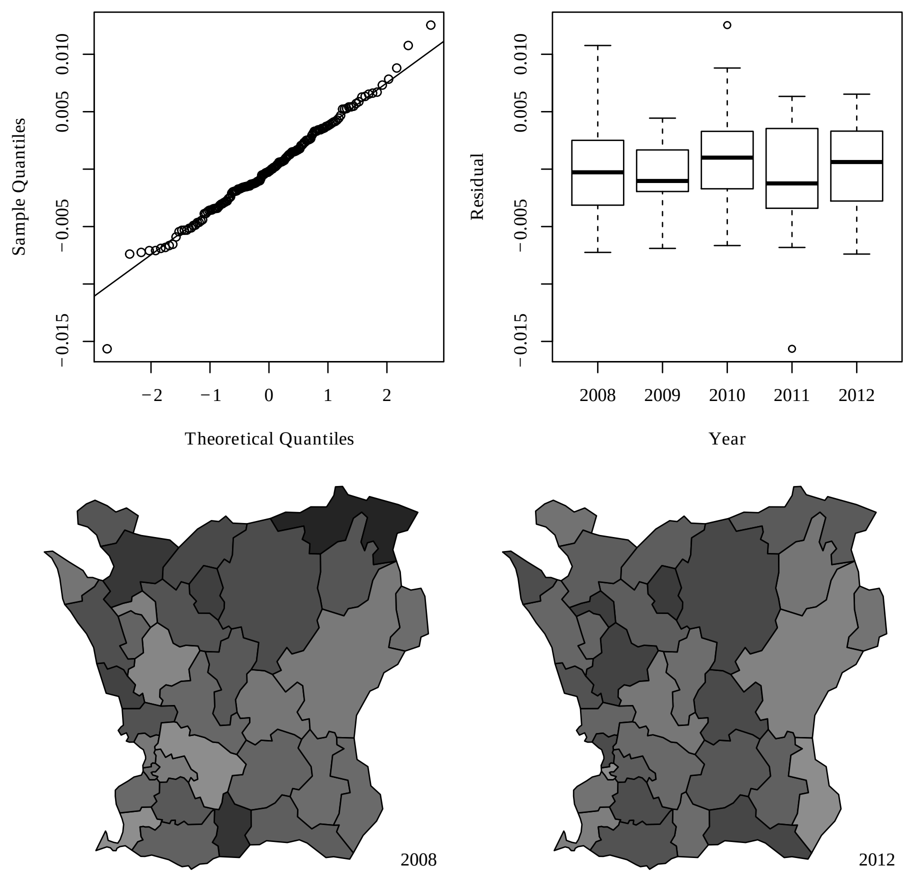

Figure 8: Plots for residuals of the spatial panel data model with distance weight matrix $\boldsymbol{W}_D$ and individual and time effects: QQ-plot; box plot for each year; maps for 2008 and 2012.

	Param	St.err.	P-value
γ	0.1977	0.0920	0.0317
pop	-0.0896	0.0454	0.0483
you	0.1696	0.2611	0.5161
wor	0.5886	0.1412	0.0000
inc	-0.2388	0.0548	0.0000
foe	0.0515	0.1421	0.7173
cri	0.0003	0.0301	0.9908
edu	-0.3205	0.2370	0.1762
σ_v	0.0039		
R_t^2	0.9695		
R_w^2	0.4130		

Table 22: Estimation results for the spatial panel data model with contiguity weight matrix $\boldsymbol{W}_C$ and individual and time effects.

	Param	St.err.	P-value
γ	0.1222	0.1303	0.3485
pop	-0.1091	0.0462	0.0183
you	0.1405	0.2645	0.5954
wor	0.5741	0.1433	0.0001
inc	-0.2498	0.0561	0.0000
foe	0.0627	0.1441	0.6634
cri	0.0045	0.0305	0.8836
edu	-0.3562	0.2400	0.1377
σ_v	0.0039		
R_t^2	0.9687		
R_w^2	0.3975		

Table 23: Estimation results for the spatial panel data model with distance weight matrix $\boldsymbol{W}_D$ and individual and time effects.

Assumption	Test	P-value
Normality	Shapiro-Wilk	0.1444
Zero individual effect	LR test	0.0000
Zero time effect	LR test	0.0000

Table 24: Hypothesis tests for the spatial panel data model with contiguity weight matrix $\boldsymbol{W}_C$ and individual and time effects.

Assumption	Test	P-value
Normality	Shapiro-Wilk	0.1179
Zero individual effect	LR test	0.0000
Zero time effect	LR test	0.0000

Table 25: Hypothesis tests for the spatial panel data model with distance weight matrix $\boldsymbol{W}_D$ and individual and time effects.

5.4 Summary of Empirical Study

Table 26 presents a summary of the parameter estimates from the four estimated models. As expected from the strong connection between the OLS estimate for the linear model and the ML estimate for the spatial model, the estimates of the linear model are similar to those from the two spatial models. This is evident for both point estimates and for statistical inference. The variables for population size, proportion of population in working age, and level of income are significant in the linear and spatial models. The dynamic model does not indicate any significant variables. The spatial model with the contiguity weight matrix gives a significant result for the spatial component.

The measures of goodness-of-fit follow the same pattern as the estimate of $\boldsymbol{\beta}$ in that the linear model and the two spatial models have similar results, while the dynamic model differs from the other three. The linear and spatial models have a total variance explained, R_t^2 of about 97 percent. About 95 percent of the total variance is explained by individual and time effects alone. The dynamic model performs less well according to the measures, with a total variance explained at about 94 percent.

	Linear		Dynamic		Spatial, $\boldsymbol{W}_C$		Spatial, $\boldsymbol{W}_D$	
	Param	Sig.	Param	Sig.	Param	Sig.	Param	Sig.
pop	-0.1202	*	-0.1561		-0.0896	*	-0.1091	*
you	0.1286		-1.1849		0.1696		0.1405	
wor	0.5636	**	-0.0548		0.5886	***	0.5741	***
inc	-0.2643	***	-0.0274		-0.2388	***	-0.2498	***
foe	0.0648		0.2048		0.0515		0.0627	
cri	0.0051		-0.0605		0.0003		0.0045	
edu	-0.3765		0.3861		-0.3205		-0.3562	
δ			0.1708					
γ					0.1977	*	0.1222	

Table 26: Summarized parameter estimates for the four estimated models.

6 Summary and Discussion

The theoretical part of the paper examines four questions: the presence and magnitude of correlation in the residuals of linear, dynamic and spatial two-way error panel data models; the accuracy of parametric tests of spatial correlation; the accuracy of randomization tests for spatial correlation; and the connection between residual correlation and the estimation of the spatial parameter in the spatial model.

The Monte Carlo study indicates that there is spatial autocorrelation in the residuals of all three model, even when the original data lacks a spatial component. The correlation is similar in size for the linear and spatial models, where it coincides with the theoretical results for OLS residuals, but differs for the dynamic model, where there is a clear dependence on the number of time periods rather than the number of individuals. For the spatial model, a decomposition into pairs of neighbours and pairs of non-neighbours shows non-zero autocorrelation in both cases and a difference in the degree of correlation.

Mao (2015) presents a method to back-transform the residual correlation and modifies the Pesaran and LM tests to the inverted residuals. This study produces some Monte Carlo results on these modified tests as well as the Pesaran and LM tests on the original residuals. The simulations give that the LM test on original residuals give fair results on size and and strength for setups where N T are large and the LM test on back-transformed residuals for the case where both N and T are large and the ratio T/N is large.

The third part of the Monte Carlo study presents simulation results on local and global randomization tests. The local test has weak strength for small N but performs well in other cases. The results improve as N and T increase. The strong results on the strength of the local randomization test must however be interpreted in light of the test being based on the same weight matrix as the simulated data, a situation which is unlikely in empirical applications. The global test performs poorly when the ratio N/T is large, and increasingly well as the ratio decreases and N and T increase. In conclusion, of all the available tests, the local randomization test is the most suitable for data where N and T are small, and the ratio N/T is large, such as the situation in the empirical part of this study.

The final part of the Monte Carlo study gives that the spatial parameter γ of the spatial model is clearly connected to the theoretical correlation in OLS residuals, resulting in estimation results which differ from zero even when there is no spatial component in the data, and which, at least for cases where N is small, give false rejections in a t or z-test. A comparison between estimation results for different combinations of N and T shows that for the maximum likelihood estimator of γ to be accurate, a large sample of individuals is necessary, but that the estimator is less influenced by the number of time periods.

The empirical study of regional unemployment lacks conclusive results. The results for the linear panel data model give that population size, relative size of work force, and income levels have a significant effect on unemployment. This result is supported in the spatial panel data model, for both the contiguity weight matrix W_C and the inverse distance matrix W_D. For the linear model, the local randomization test for spatial autocorrelation accepts the null hypothesis that there is no spatial autocorrelation. However, the estimations of the spatial parameter γ in the spatial models differ greatly from the expected value given the results from the Monte Carlo study. This might suggest that there is spatial autocorrelation which is not captured by the model, either due to model specification or to choice of weight matrix. The impact of the choice of weight matrix is further exemplified by the differences between the two spatial models, in particular that the spatial parameter is significant in the model based on the contiguity matrix $\boldsymbol{W}_C$, but not in the model based on the distance matrix $\boldsymbol{W}_D$.

The dynamic two-way error model gives no significant explanatory variables. Hypothesis tests for serial autocorrelation indicate that the model is not a suitable fit. The difference between the results for the dynamic model and the spatial or linear models, and the relative similarities of the latter two, highlights the similarity of the maximum likelihood estimation of the spatial model and the OLS estimator of the linear model.

The study points to some possible problems and areas which may be of interest for future studies. The section on descriptive statistics briefly touches upon the existence of linear trend in several explanatory variables and the test of multicollinearity (Table 19) shows that some variables are almost completely explained by other explanatory variables or by individual and time effects. The linear trend is a lesser problem in the linear and spatial models, since the estimation methods are based on a demeaning method which removes linear trend, but may have an impact on estimation of the dynamic model. This is one possible explanation for the differences in results for the different models. The results on multicollinearity strongly suggests that some variables can be dropped to form alternative variable sets.

There are alternatives to the maximum likelihood estimator for the spatial model (Mutl and Pfaffermayr, 2011; Baltagi and Liu, 2016), which are possibly not as closely linked to the OLS estimator. Given the Monte Carlo results on residual correlation for the dynamic model, which is estimated by a GMM estimator and shows a lower correlation which decreases with T, it may be of further interest to investigate residual correlation for the alternative estimators.

The Monte Carlo study shows strong results for the local randomization test. However, these results are likely boosted by the fact that the test is based on the same weight matrix as the simulated data. It could be of some interest to investigate the robustness of the results to differences between the weight matrix used in data simulation and the weight matrix used in testing.

The Monte Carlo study on residual correlation in the spatial model shows that the correlation between neighbouring individuals differ from the correlation between non-neighbouring individuals. The residual spatial correlation in the spatial model thereby differs from that in the linear model, meaning that the randomization tests suggested here may not be applied. A possible approach towards a randomization test for spatial correlation in residuals of the spatial model would be to permute within, and not across, the groups of neighbours and non-neighbours.

References

Alvarez, J. and Arellano, M. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, (4):1121–1159, 2003.

Anselin, L. and Hudak, S. Spatial econometrics in practice. a review of software options. *Regional Science and Urban Economics*, 22:509–536, 1992.

Arbetsförmedlingen. Sökande per kommun 1996-2015, 2016. URL `http://www.arbetsformedlingen.se/Om-oss/Statistik-prognoser/Tidigare-statistik.html`.

Arellano, M. and Bond, S. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, (2):277, 1991.

Arellano, M. and Bover, O. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51, 1995.

Bai, J. Fixed-effects dynamic panel models, a factor analytical method. *Econometrica*, (1):285, 2013.

Baltagi, B. H. *Econometric Analysis of Panel Data.* John Wiley & Sons, Inc., Chichester, 2013.

Baltagi, B. H. and Liu, L. Instrumental variable estimation of a spatial autoregressive panel model with random effects. *Economics Letters*, 111:135–137, 2011.

Baltagi, B. H. and Liu, L. Random effects, fixed effects and hausman's test for the generalized mixed regressive spatial autoregressive panel data model. *Econometric Reviews*, 35(4):638, 2016.

Cracolici, M. F., Cuffaro, M., and Nijkamp, P. A spatial analysis on Italian unemployment differences. *Statistical Methods and Applications*, 18(2):275–291, 2009.

Cressie, N. A. C. *Statistics for Spatial Data.* Wiley, New York, 1991.

Croissant, Y. and Millo, G. Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27(2), 2008.

Elhorst, J. P. The mystery of regional unemployment differentials: Theoretical and empirical explanations. *Journal of Economic Surveys*, 17(5):709–748, 2003a.

Elhorst, J. P. Specification and estimation of spatial panel data models. *International Regional Science Review*, 26(3):244, 2003b.

Elhorst, J. P. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels.* Springer, Berlin, Heidelberg, 2014.

Gilmartin, M. and Korobilis, D. On regional unemployment: An empirical examination of the determinants of geographical differentials in the UK. *Scottish Journal of Political Economy*, 59(2):179–195, 2012.

Good, P. I. *Permutation, Parametric and Bootstrap Tests of Hypotheses.* Springer Series in Statistics. Springer, New York, 2005.

Koenker, R. A note on studentizing a test for heteroskedasticity. *Journal of Econometrics*, 17(1):107–112, 1981.

Lee, L.-f. and Yu, J. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154:165–185, 2010.

Lottmann, F. Explaining regional unemployment differences in Germany: A spatial panel data analysis. 2012.

Mao, G. Testing for sphericity in a two-way error components panel data model. *Econometric Reviews*, 2015.

Millo, G. and Piras, G. splm: Spatial panel data models in R. *Journal of Statistical Software*, 47(1):1–38, 2012.

Mutl, J. and Pfaffermayr, M. The hausman test in a cliff and ord panel model. *Econometrics Journal*, 14(1):48–76, 2011.

Neyman, J. and Scott, E. L. Consistent estimates based on partially consistent observations. *Econometrica*, (1):1, 1948.

Nickell, S. Biases in dynamic models with fixed effects. *Econometrica*, (6):1417, 1981.

Pesaran, M. General diagnostic tests for cross section dependence in panels. Cambridge Working Papers in Economics 0435, Faculty of Economics, University of Cambridge, 2004. URL `https://ideas.repec.org/p/cam/camdae/0435.html`.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`.

Shapiro, S. S. and Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965.

Statistics Sweden. Inkomster och skatter, 2016a. URL `http://www.scb.se/he0110`.

Statistics Sweden. Befolkningsstatistik, 2016b. URL `http://www.scb.se/be0101`.

Statistics Sweden. Regionala indelningar, 2016c. URL `http://www.scb.se/hitta-statistik/regional-statistik-och-kartor/regionala-indelningar/`.

Statistics Sweden. Befolkningens utbildning, 2016d. URL `http://www.scb.se/uf0506`.

The Swedish National Council for Crime Prevention. Anmälda brott, 2016. URL `https://bra.se/statistik/kriminalstatistik/anmalda-brott.html`.

Verbeek, M. *A Guide to Modern Econometrics.* Wiley, Chichester, 2 edition, 2004.

Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Cambridge, Mass., 2002.

Zeileis, A. and Hothorn, T. Diagnostic checking in regression relationships. *R News*, 2(3):7–10, 2002.