Testing the Correct Specification of a Spatial Dependence Panel Model for Stock Returns

Tim Kutzker* and Dominik Wied†

February 24, 2021

Abstract

This paper provides model specification tests for the spatial autoregressive panel model of order $m$ that also allows for exogenous covariates. The asymptotic null distributions are derived and size as well as power properties in finite samples are analyzed by Monte Carlo simulations. In the empirical application, we analyze Euro Stoxx 50 returns in two different time spans looking for insights how well models with different specifications of the spatial weighting matrices (local, country, industry and country-industry specific dependencies including interaction effects) fit to the data. The analyzes also demonstrate the ability of the tests to detect inaccurate Value-at-Risk forecasts.

Keywords: heteroscedasticity, method of moments, spatial dependence, stock returns, Value-at-Risk

JEL Classification Numbers: C13, C51, G12.

*Corresponding Author, Institute of Econometrics and Statistics, University of Cologne, Address: Albertus Magnus Platz, D- 50923 Cologne, Germany, telephone: +49 221 470 6561, e-mail: tim.kutzker@uni-koeln.de.

†Institute of Econometrics and Statistics, University of Cologne, 50923 Cologne, Germany, email: dwied@uni-koeln.de
1 Introduction and Summary

The purpose of this paper is a contribution to the literature of specification testing in spatial econometric models, in particular in a panel context. We consider a spatial autoregressive model in which a multivariate random variable is explained by exogenous regressors and by spatially weighted lags of itself. In contrast to other approaches in the literature, we allow for more than one spatial matrix and for cross-sectional heterogeneity of the error variances. The tests proposed in the paper give evidence if the model is a good fit to a particular data set.

A classical reference for the question if there is spatial dependence in a given data is Moran’s I (Moran, 1950) which is often used just as a descriptive measure, but is also used for tests of spatial dependence in linear and non-linear panel models (Kelejian and Prucha, 2001). Li et al. (2007) provide an alternative measure to Moran’s I. Other diagnostic tests for spatial dependence are the LM-tests by Anselin (1996) and the regression-based tests by Born and Breitung (2011). Su and Qu (2017) propose specification tests for SAR models. However, neither of the mentioned tests consider a panel context, which is different for Baltagi et al. (2003) who propose LM-tests for spatial dependence and for Millo (2017) who proposes a randomization test in a factor-augmented panel. Kelejian and Piras (2016) propose a J-test procedure for testing a null model against non-nested alternatives for a fixed effects spatial panel data framework. A special feature that our test has compared to the listed panel approaches is that we explicitly allow for heterogeneity in the error variances. This is possible because a two-step approach is used for estimating the model parameters.

A related branch of the specification literature consists of testing procedures to choose the best spatial weighting matrix among a set of potential candidates (Herrera et al., 2019; Kelejian and Piras, 2011; LeSage and Pace, 2014, e.g.). While our test can be used for such model selection issues as well (and is in fact in our application), the focus of the present paper is slightly different.

As application of our test, we focus on modeling the spatial dependence of stock returns. In recent years the literature in economics and finance has found some interest in the connection between spatial dependence and stock returns. To give some examples, Asgharian et al. (2013) investigate in which way stock market co-movements are determined by countries’ economic and geographical relations. Tam (2014) analyzes equity market linkages in East Asia, Blasques et al. (2016) extend the spatial Durbin model by a time-varying spatial dependence parameter, Selan and Kalatzis (2017) analyze peer effects in Brazil.

More relevant for our work is Arnold et al. (2013) who propose a spatial autoregressive (SAR) panel model for stock returns in order to capture local dependencies and dependencies within industrial branches. Wied (2013) considers structural breaks in
these models and Schmitt et al. (2016) combine the approach with local normalization techniques. Gong and Weng (2016) use the model for value at risk forecasts in the Chinese stock market. Catania and Billé (2017) generalize the SAR model with autoregressive and heteroscedastic disturbances by including methods from score-driven models. Various empirical analyses in the aforementioned papers show that the SAR panel model is generally suitable for Value-at-Risk (VaR) forecasts and outperforms, e.g., the one-factor model.

The model in the present paper is a generalization of the model from Arnold et al. (2013) in the sense that we allow for an arbitrary fixed number of spatial matrices and for exogenous regressors. We propose two methods on how to check the model fit. The basic idea stems from the model assumption that spatial weighting matrices capture all spatial dependence and that the remaining error terms are spatially uncorrelated. Therefore, we consider the model residuals such that the tests keep the null hypothesis of model fit if the covariance matrix of the residuals is basically diagonal, i.e., its off-diagonal elements are close to zero. We derive the asymptotic distribution of our test statistics and show in simulations that the tests have reasonable power properties against sparse error term covariance matrices. In the simulations, we also account for conditional heteroscedasticity, a feature that is considered to be important if spatial models are used for VaR predictions (see, e.g., Zhang et al. (2018)). In an empirical application on stock data from the Euro Stoxx 50, we test the model fit for different spatial weighting matrices and analyze in which sense the tests’ results are related to the quality of VaR forecasts. We consider the time spans around the global financial crisis in 2008 and the COVID-19 crisis in 2020.

This paper is organized as follows: Section 2 describes the classical spatial error model, discusses its assumptions and efficient estimation procedures and provides two model specification tests. Section 3 presents an extensive Monte Carlo simulation study and Section 4 an empirical application. Finally, Section 5 concludes.

2 A Cross Sectional Correlation Based Specification Test for SEM($m$) Panel Models

In this section, we first introduce the general spatial error panel model of order $m$, SEM($m$), with $m \in \mathbb{N}$ and discuss briefly its assumptions and efficient parameter estimation procedures. We then present two model specification tests, with the latter showing better power properties as demonstrated in Section 3.
2.1 SEM(m): Assumptions and Parameter Estimations

We consider the following panel data regression model:

\[ y_{ti} = X_{ti}' \beta + u_{ti}, \quad i = 1, \ldots, n \text{ and } t = 1, \ldots, T, \]  

(1)

where \( y_{ti} \) is the observation on the \( i \)th individual for the \( t \)th time period, \( X_{ti} \) denotes the \( k \times 1 \) vector of the covariates, \( \beta \) the \( k \times 1 \) parameter vector and \( u_{ti} \) is the regression disturbance. We assume that the disturbance \( u_{ti} \) can be modeled by a spatially autoregressive model of order \( m \in \mathbb{N}, \text{SAR}(m) \), i.e.

\[ u_{t} = \sum_{i=1}^{m} \rho_i W_i u_{t} + \varepsilon_t \]  

(2)

where \( u_t' = (u_{t1}, \ldots, u_{tn}) \) and \( \varepsilon_t' = (\varepsilon_{t1}, \ldots, \varepsilon_{tn}) \). The parameters \( \rho_i \) with \( i = 1, \ldots, m \) are the scalar spatial autoregressive coefficients and \( W_i \) for \( i = 1, \ldots, m \) with \( m \) fixed\(^1\) the known pre-specified \( n \times n \) spatial weighting matrices\(^2\) that does not vary over time\(^3\).

Using the vector representation \(^2\), we obtain a simplified version of model \((1)\), i.e.

\[ y_{t} = X_{t}\beta + u_{t} \quad \text{with} \quad u_{t} = (I_n - \sum_{i=1}^{m} \rho_i W_i)^{-1} \varepsilon_t, \]  

(3)

where \( y_{t} \) is of dimension \( n \times 1 \), \( X_{t} \) is \( n \times k \), \( \beta \) is \( k \times 1 \) and \( u_{t}, \varepsilon_{t} \) is \( n \times 1 \). The observations in the cross-sectional dimension \( i \) are assumed to be fixed. In order to ensure a consistent estimation of the parameters \( \beta_i \) with \( i = 1, \ldots, k \), \( \rho_j \) with \( j = 1, \ldots, m \) and \( \sigma_k^2 \) with \( k = 1, \ldots, n \), we impose additionally the following model assumptions:

**Assumption 1.**

1. For \( t \in \mathbb{Z} \), the regressor matrix \( X_{t} \) is strictly exogenous with \( \mathbb{E} [u_{t} | X_{t}] = 0 \). Additionally \( n^{-1} X_{t}'X_{t} \) has full rank for all \( n \).

2. The sequence of random vectors \( \{\varepsilon_{t}\}_{t \in \mathbb{N}} \) has zero mean, is stationary and ergodic.

\(^1\)In the application later on, we will introduce four different spatial matrices which are assumed to capture the structure of daily stock returns. The first part covers a general dependence which affects all subjects equally. The second part captures dependencies among industrial branches and national effects are included with the help of the third dependency structure. Finally, the fourth spatial matrix represents interactions effects of stocks that are affiliated with the same country and industry.

\(^2\)It would also be conceivable to include random or fixed effects terms in the error term. However, since we are finally developing a specification test for the idiosyncratic error term, we refrain from a more complex representation here with the hint that our test approach could also be adapted for SEM(\(m\)) with random or fixed effects.

\(^3\)An overview of commonly used spatial matrices is given in J.P. Elhorst (2012).
3. For $i \in \{1, \ldots, m\}$, $r = 1, \ldots, n$, $s = 1, \ldots, n$, $W_{i,rs} \geq 0$, $W_{i,rr} = 0$.

4. For $i \in \{1, \ldots, m\}$ and $r = 1, \ldots, n$, $\sum_{s=1}^{n} W_{i,rs} = 1$.

5. The parameter space $S_{\rho}$ is defined as $S_{\rho} = \{\rho \in \mathbb{R}^m : ||\rho||_1 < 1\}$ where $|| \cdot ||_1$ defines the $L^1$-norm, i.e. $||\rho||_1 := \sum_{i=1}^{m} |\rho_i|$.

6. The spatial parameter vector $\rho$ is uniquely identified.

7. For $t \in \mathbb{Z}$, $\text{Cov}(\varepsilon_t) = \text{diag}\{\sigma_1^2, \ldots, \sigma_n^2\} =: \Sigma \in \mathbb{R}^n$.

8. Each element of the vector $\left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \varepsilon_t' \right)_{i<j}$ meets the assumption of a central limit theorem and the corresponding long-term covariances

$$\sum_{s,t \in \mathbb{N}} \text{Cov}(\varepsilon_t \varepsilon_j, \varepsilon_k \varepsilon_s)$$

are finite for every $i < j$ and $k < l$, where we interpret $(\cdot)_{i<j}$ as the stacked vector of the upper triangular matrix.

Assumption 1.1. is a sufficient condition for the existence of an unbiased estimator $\hat{\beta}$ for $\beta$. The zero mean and stationarity condition in Assumption 1.2. is plausible especially in the context of daily stock returns (see [Aue et al., 2009]). In case of a violation of this assumption, trend adjustment or centering could ensure that it is fulfilled. To exclude self-neighbors, the diagonal elements of $W_i$ with $i = 1, \ldots, m$ are conventionally set equal to zero (Assumption 1.3.). Additionally, Assumption 1.3. claims that all elements are non-negative, which is natural, as distances are measured. Assumption 1.4. ensures that the matrices are bounded and standardized. Assumption 1.5. restricts the parameter space such that the sum of the absolute values of the elements of $\rho \in \mathbb{R}^m$ is smaller than 1. While the assumption could slightly be generalized (cf. [J.P. Elhorst, 2012]), we follow the notation of [Arnold et al., 2013] as it guarantees that the matrix $(I_n - \sum_{i=1}^{m} \rho_i W_i)$ is non-singular. Finally, Assumption 1.6. yields a high-level identification assumption which is specified in the Appendix in Assumption 3.1. It rules out certain combinations of spatial weighting matrices, e.g., these matrices must be pairwisely distinct. Hence, Assumptions 1.1. – 1.6. ensure that model (1) is well defined.

However, for the estimation of the parameters $\beta_i$ with $i = 1, \ldots, k$, $\rho_j$ with $j = 1, \ldots, m$ and $\sigma_l^2$ with $l = 1, \ldots, n$ in SEM($m$) different techniques are available. Depending on the underlying estimation procedure, the assumptions of Assumption [1] have to be augmented (cf. [Elhorst, 2003; Lee and Yu, 2010; Elhorst, 2010]). Generally, since

---

4The matrix $(I_n - \sum_{i=1}^{m} \rho_i W_i)$ is strictly diagonally dominant.
the covariance matrix of the error terms $u_t$ of our model (3) is given by

$$\text{Cov} [u_t] = (I_n - \sum_{i=1}^{m} \rho_i W_i)^{-1} \text{Cov} [\varepsilon_t] \left( (I_n - \sum_{i=1}^{m} \rho_i W_i)^{-1} \right)'$$

(4)

$$= (I_n - \sum_{i=1}^{m} \rho_i W_i)^{-1} \Sigma \left( (I_n - \sum_{i=1}^{m} \rho_i W_i)^{-1} \right)'$$

(5)

it is crucial to use more elaborate approaches for an efficient estimation of $\beta$:

If $\rho_i$ with $i = 1, \ldots, m$, e.g., is known or if we have consistent estimates, one of the possible approaches for estimating the regression coefficients $\beta$ of $\text{SEM}(m)$ is generalized least squares (GLS). If we additionally assume a (normal) distribution of the idiosyncratic errors, the ML approach could also be considered. Another estimation method is the GMM technique. [Yildirim and Kantar (2020)](2020) summarizes all these three estimation procedures for the case of the homoscedastic SEM(1), which can be easily applied to the more general (heteroskedastic) SEM($m$).

In case that there are no covariates in model (3) or that a consistent estimator for $\beta$ is available, the two step GMM procedure of [Arnold et al. (2013)](2013) could also be employed in order to obtain efficient estimates for the parameters $\rho_i$ with $i = 1, \ldots, m$ and $\sigma^2_l$ with $l = 1, \ldots, n$. Since we base our specification test in particular on the design of the error terms, we henceforth assume that the following assumption holds:

**Assumption 2.** Let $\hat{\beta}$ be an initial and consistent estimator of $\beta$ in (3).

An estimator that satisfies Assumption 2, e.g., is the ordinary least squares (OLS) estimator of model (3). While the classical OLS approach does not provide an efficient estimate of $\beta$, this estimation procedure provides consistent estimates and requires no assumptions other from those in Assumption 1. Given $\hat{\beta}_{OLS}$, one can apply the estimator by [Arnold et al. (2013)](2013) discussed below on the residuals $\hat{u}_t = y_t - X_t \hat{\beta}_{OLS} = u_t - X_t (\hat{\beta}_{OLS} - \beta)$ in order to obtain consistent estimates for $\rho_i$, $i = 1, \ldots, m$ and $\sigma^2_l$, $l = 1, \ldots, n$. If our model (3) is free of covariates ($k = 0$), we will also refer to $y_t$ by the residuals $\hat{u}_t$. This allows for treating the cases that either Assumption 2 is met or $k = 0$ simultaneously with a uniform notation. All subsequent calculations and asymptotic analyzes are conditioned on the estimator $\hat{\beta}$. The two-step GMM procedure by [Arnold et al. (2013)](2013) consists of the following two steps:

First, we estimate the spatial parameters $\rho_i$, $i = 1, \ldots, m$ by the method of moments. Due to its construction, this step does not depend on the parameters of variance $\hat{\sigma}^2_l$ with $l = 1, \ldots, n$. Secondly, given the spatial estimates, the estimation of the parameters of variance is simply achieved by averaging over the estimated $\hat{\varepsilon}^2_i$ with $i = 1, \ldots, n$. Under some regularity assumptions, the GMM estimator $\hat{\rho}$ is consistent and as asymptotically normal (cf. Theorem A.2). While this is worked out in
Arnold et al. (2013) for the special case of $m = 3$, a detailed derivation for the GMM estimator in the general case is presented in the Appendix [A] i.e. $m \in \mathbb{N}$ is finite and fixed.

However, assuming that we have efficient and unbiased estimates for our model, it is crucial to check whether the underlying data set also meets the model assumptions. The crucial requirement on which we will base our specification test is that the covariance matrix of the error terms $\varepsilon_t$ is diagonal which corresponds to Assumption 1.7. Assumption 1.8. guarantees that the limiting distribution of our suggested test statistic is not degenerated.

2.2 The Specification Test

We propose a testing procedure in case of validity of Assumption 1 and 2, i.e. $\hat{\beta}$ is a consistent estimate for $\beta$, where the results of our simulation study in Section 3.2 indicate that the testing procedure is also applicable if the error terms are replaced by GARCH residuals. So subsequently, the word data set can be regarded either as the original or GARCH adjusted data.

Following the discussion given in the previous subsection, our proposed testing procedure therefore checks if Assumption 1.6. is met. Even if the course of action seems technical, the idea behind the test statistic is straightforward: We do not reject the null hypothesis if the covariance matrix of the idiosyncratic error is basically a diagonal matrix, i.e. its off-diagonal elements should not deviate too far from zero.

Let $\hat{H} \in \mathbb{R}^{n \times n}$ denote the empirical covariance matrix of the residuals times the square root of the time horizon $T$, i.e. $\hat{H} := \sqrt{T} \text{Cov} [\hat{\varepsilon}_t]$ and $\hat{H}_{ij}$ its elements with $i, j \in \{1, 2, \ldots, n\}$. Let $\sigma_{ij}^2$ denote the $(i, j)$-th element of the theoretical counterpart $\Sigma$, i.e. the error covariance matrix. Since $\hat{H}$ and $\Sigma$ are symmetric, it is sufficient to consider only the elements of the upper triangle of the matrix $\Sigma$. Hence, the null hypothesis is given by

$$H_0 : \sigma_{ij}^2 = 0 \text{ for all } i < j \quad \text{vs.} \quad H_1 : \exists i, j \text{ with } i < j : \sigma_{ij}^2 \neq 0. \quad (6)$$

We opt to use $\chi^2$-type tests for this testing problem. Instead of considering each element or the maximum of the absolute value of all off-diagonals, we take the sum of each element squared into account. Thus, the naive test statistic is given by

$$S := \sum_{i < j, i,j=1,\ldots,n} (\hat{H}_{ij})^2. \quad (7)$$

The aim of the following theorem is to decompose the limit in distribution of the empirical covariance matrix conditioned on a consistent estimator $\hat{\beta}$ times $\sqrt{T}$ into three matrices, which allow to determine the limit in distribution of the sum of the
elements of the upper triangular matrix. Here and in the following \( \lim \) denotes limit in distribution and \( \equiv \) equality in distribution.

**Theorem 2.1.** Under the null hypothesis \( H_0 : \sigma_{ij}^2 = 0 \) for all \( i < j \) and the assumptions of Theorem A.2 the following holds for \( 1 \leq i, j \leq n \)

\[
\lim_{T \to \infty} \sqrt{T} \text{Cov} [\hat{\epsilon}_t] = A + B + B^T \in \mathbb{R}^{n \times n}
\]  

(8)

with \( A_{ii} = \lim_{T \to \infty} \sqrt{T} \sum_{t=1}^{T} \sigma_{ii}^2 = \infty \) and the off-diagonal elements of \( A \) are jointly normally distributed for \( i \neq j \) with \( A_{ij} \sim N(0, \lim_{T \to \infty} \text{Var} [\frac{1}{\sqrt{T}} \sum_{t=1}^{\infty} \epsilon_t \epsilon_{jt}]) \) and \( \text{Cov}[A_{ij}, A_{kl}] = 0 \) for \( i \neq j \) and \( k \neq l \) with \( (i, j) \neq (k, l) \).

Moreover, \( B \equiv (\sum_{i=1}^{m} X_i W_i)(I_n - \sum_{i=1}^{m} \rho_i W_i)^{-1} \Sigma, \) where \( X := (X_1, \ldots, X_m) \sim N(0, d^{-1} S_W(d^{-1})^T) \in \mathbb{R}^{1 \times m} \) and \( S_W = \sum_{t=-\infty}^{\infty} \text{E}[f(u, \rho)f(u, \rho)'f(u, \rho)] \) for \( f(u, \rho) = (\epsilon_1 W_1 \epsilon_1, \ldots, \epsilon_m W_m \epsilon_1)' \) with \( d \) defined in Assumption 3.

Three remarks about Theorem 2.1 are in order. First, the leading elements of matrix \( A \) diverge to infinity. However, the tests considers only the off-diagonal elements \( (i \neq j, i, j = 1, \ldots, n) \) which are finite by Assumption 1.7. This in turn ensures that the test is well defined. Secondly, since \( (I_n - \sum_{i=1}^{m} \rho_i W_i) \) is strictly diagonally dominant, the inverse exists. Thirdly, we note that the matrices \( B \) and its transposed appear in the limit. This is due to the fact of estimating \( \rho \) instead of using the unknown population quantity. The analysis of such a residual effect (see [Demetrescu and Wied, 2019]) is somewhat complicated, since the additional terms need different standardizing factors in the proof\(^5\). However, all terms in the limiting distribution are based on the same error terms, thus, the convergence is jointly and the limiting distribution in (8) is multivariate normal. If we additionally assume serially independence in the error vector, the variance of the elements in the limiting matrix \( A \) simplifies to a product shown in the following remark.

**Remark 2.2.** Suppose the assumptions of Theorem 2.1 hold. If \( \{\epsilon_t\}_{t \in \{1, \ldots, T\}} \) is serially independent, then

\[
A_{ij} \sim N(0, \sigma_i^2 \sigma_j^2)
\]

for \( i \neq j \).

(9)

In accordance with our test statistic \( S \) \(^7\), we can reformulate the test in vectorial

\(^5\)For a detailed analysis of the convergence rate we refer to Lemma B.1 in the corresponding Appendix.
notation, i.e.

\[ S = \hat{\alpha}' \hat{\alpha}, \]

(10)

where \( \hat{\alpha} \) represents the vector of the upper triangle of the empirical covariance matrix of the residuals times \( \sqrt{T} \), i.e. \( \hat{H} \). Since the empirical covariance matrix consists of \( n^2 \) elements, the upper triangular matrix vector consists of \( n(n-1)/2 \) elements and has the following form:

\[
\hat{\alpha} := \lim_{T \to \infty} \left( \sqrt{T} \text{Cov} \left[ \hat{\varepsilon}_t \right] \right)_{i < j, i, j = 1, \ldots, n}
\]

\[
= \lim_{T \to \infty} \left( \frac{1}{\sqrt{T}} \sum_{i,j=1}^{n} \hat{\varepsilon}_i \hat{\varepsilon}_j' \right)_{i < j, i, j = 1, \ldots, n} = \lim_{T \to \infty} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\varepsilon}_t' \in \mathbb{R}^{n(n-1)/2}
\]

with \( \hat{\varepsilon}_t^* := (\hat{\varepsilon}_1 \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_1 \hat{\varepsilon}_n, \hat{\varepsilon}_2 \hat{\varepsilon}_3, \ldots, \hat{\varepsilon}_2 \hat{\varepsilon}_n, \ldots, \hat{\varepsilon}_{(n-1)} \hat{\varepsilon}_n)' \).

By means of Slutzky’s theorem we define the theoretical counterpart

\[
\alpha := (A)_{i < j, i, j = 1, \ldots, n}
\]

\[
= \lim_{T \to \infty} \left( \frac{1}{\sqrt{T}} \sum_{i,j=1}^{n} \varepsilon_i \varepsilon_j' \right)_{i < j, i, j = 1, \ldots, n} = \lim_{T \to \infty} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t' \in \mathbb{R}^{n(n-1)/2}
\]

with \( \varepsilon_t^* := (\varepsilon_1 \varepsilon_2, \ldots, \varepsilon_1 \varepsilon_n, \varepsilon_2 \varepsilon_3, \ldots, \varepsilon_2 \varepsilon_n, \ldots, \varepsilon_{(n-1)} \varepsilon_n)' \).

which stacks the upper triangular matrix of the covariance matrix of the errors times \( \sqrt{T} \) in a vector. Analogously, \( \delta \) defines the vector of the stacked upper triangular matrix of \( B \) and \( \delta^* \) of \( B' \), respectively, i.e. for

\[
Z_W := \lim_{T \to \infty} \sum_{g=1}^{m} \sqrt{T}(\rho_g - \hat{\rho}_g) W_G
\]

we define

\[
\delta := (B)_{i < j, i, j = 1, \ldots, n} = \left( Z_W(I_n - \sum_{g=1}^{m} \rho_g W_G)^{-1} \Sigma \right)_{i < j, i, j = 1, \ldots, n} \in \mathbb{R}^{n(n-1)/2},
\]

\[
\delta^* := (B')_{i < j, i, j = 1, \ldots, n} = \left( \Sigma' (I_n - \sum_{g=1}^{m} \rho_g W_G')^{-1} Z_W' \right)_{i < j, i, j = 1, \ldots, n} \in \mathbb{R}^{n(n-1)/2}.
\]

The vectors \( \delta \) and \( \delta^* \) are well defined, since \( B \) is not necessarily symmetric.

**Lemma 2.3.** \( \delta \) represents the vector of the upper triangle and \( \delta^* \) the vector of the lower triangle of the matrix \( Z_W(I_n - \sum_{g=1}^{m} \rho_g W_G)^{-1} \Sigma \), i.e. for \( i, j \in \{1, \ldots, n\} \)

\[
\delta^* = \left( Z_W(I_n - \sum_{g=1}^{m} \rho_g W_G)^{-1} \Sigma \right)_{i > j, i, j = 1, \ldots, n} \in \mathbb{R}^{n(n-1)/2}. \quad (11)
\]
The next Lemma provides the limit distribution of our test statistic $S^{10}$.

**Lemma 2.4.** Suppose the assumptions of Theorem 2.1 hold. Then the test statistic $S^{7}$ is asymptotically distributed as

$$S = \hat{\alpha}' \frac{d}{T \to \infty} (\alpha + \delta + \delta^*)' (\alpha + \delta + \delta^*),$$

where the covariance matrix for $\alpha$ is given by

$$\text{Cov}[\alpha] = \begin{pmatrix}
\lim_{T \to \infty} \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{1t} \varepsilon_{2t} \right] & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \lim_{T \to \infty} \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{(n-1)t} \varepsilon_{nt} \right]
\end{pmatrix}. $$

Consequently, the critical value for the test statistic $S^{7}$ can be derived by drawing independently from the limiting distribution given in Lemma 2.4 and computing the corresponding quantile. As shown in Section 3, our proposed test takes care of size demands and has good power properties. The next subsection presents a modification of the proposed specification test with a simpler limit distribution.

### 2.3 Simplified Tests

In Theorem 2.1 we have shown that the elements of the limiting distribution follow a multivariate normal distribution. Thus, if we standardize the test statistic $S^{10}$ by its covariance matrix, we get a new test statistic $S^*_\chi$ which is $\chi^2$-distributed, i.e.

$$S^*_\chi := \hat{\alpha}' (\text{Cov}[\alpha + \delta + \delta^*])^{-1} \hat{\alpha} \sim \chi^2_{\frac{\chi^2 (n-1)}{2}}. $$

The quantiles of this limit distributions can be easily obtained, but the implementation of test statistic is more complicated than that of (7). The following discussion shows how the implementation can be simplified.

The terms $\delta$ and $\delta^*$ can be regarded as additional noise which comes from the estimation procedure. This additional noise can be extracted by decomposing the covariance matrix given in (12) into two parts. Thus, we have

$$\text{Cov}[\alpha + \delta + \delta^*] = \text{Cov}[\alpha] + \Psi$$

with $\Psi := \text{Cov} [\delta + \delta^*] + \text{Cov} [\alpha, \delta + \delta^*] + \text{Cov} [\alpha, \delta + \delta^*]'$. The first part $\text{Cov}[\alpha]$ covers the underlying variance structure while the second part $\Psi$ can be considered as a noise term.\footnote{If we additionally assume serial independence, the covariance matrix of $\alpha$ can easily be implemented, since only the variances need to be estimated, cf. Lemma 2.3. Otherwise, the covariance matrix of $\alpha$ is given in Lemma 2.4.}
If either \( \| (\text{Cov}[\alpha])^{-1}\Psi \| < 1 \) or \( \| \Psi(\text{Cov}[\alpha])^{-1} \| < 1 \) holds\(^7\), the inverse of the covariance matrix \(^8\) can be estimated by means of a Taylor series approximation and a telescoping sum. This yields

\[
(\text{Cov}[\alpha + \delta + \delta^*])^{-1} = (\text{Cov}[\alpha])^{-1} - (\text{Cov}[\alpha])^{-1}\Psi(\text{Cov}[\alpha])^{-1} \\
+ (\text{Cov}[\alpha])^{-1}\Psi(\text{Cov}[\alpha])^{-1}\Psi(\text{Cov}[\alpha])^{-1} - ... \\
\leq (\text{Cov}[\alpha])^{-1}.
\]

Thus, \( (\text{Cov}[\alpha])^{-1} \) is an upper bound for the inverse of the covariance matrix \(^8\). Hence, for

\[
S_{\chi} := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t^* (\text{Cov}[\alpha])^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^*,
\]

we have the algebraic relation \( S_{\chi} \geq S_{\chi}' \sim \chi^2_{n(n-1)} \). Therefore, using the test statistic \( S_{\chi} \) with the quantiles of the \( \chi^2_{n(n-1)} \) leads to a liberal test. In order to study the behavior of \( S \) and \( S_{\chi} \) in finite samples, we perform an extensive Monte Carlo simulation study which can be found in the next section.

### 3 Monte Carlo Simulation

#### 3.1 Serial Independence

The Monte Carlo (MC) simulation study consists of three main simulations, and in all simulations we assume a SEM\((m)\) for \( m = 3, 4 \) with no covariates, since the main focus of this chapter is to investigate the size and power properties of our proposed tests \( S \) and \( S_{\chi} \). While the first two simulations assume serial independence, the third simulation examines the behavior of the test in the case of GARCH(1,1) driven errors. The first less comprehensive simulation depicts a SEM(3) model with no exogenous regressors, i.e.

\[
y_t = \rho_1 W_1 y_t + \rho_2 W_2 y_t + \rho_3 W_3 y_t + \varepsilon_t,
\]

\(^7\)In our Monte Carlo simulation we observed that this is usually the case whenever the variance of \( \varepsilon_{it} \) is greater than 1 for all \( i = 1, ..., n \).

\(^8\)The sum and product of two symmetric positive semidefinite (psd) matrices is still psd.
where \( \mathbf{y}_t, \mathbf{\varepsilon}_t \in \mathbb{R}^n \) for \( t = 1, \ldots, T \) and \((W_1)_{ij} = \frac{1}{n-1}\) for all \( i \neq j \) and \((W_1)_{ii} = 0\). The spatial matrices \( W_2 \) and \( W_3 \) are defined as

\[
(W_2)_{ij} = \begin{cases} 
1, & \text{if } j \text{ even and } i \neq j \\
1, & \text{if } j - 1 = i \\
0, & \text{otherwise}
\end{cases}
\]

\[
(W_3)_{ij} = \begin{cases} 
1/(n/2 - 1), & \text{if } i, j \leq \frac{n}{2} \\
0, & \text{otherwise}
\end{cases}
\]

where additionally the matrix \( W_2 \) is row standardized by its row sum \( \sum_j (W_2)_{ij} \) for \( i = 1, \ldots, n \). The expression \( i, j \leq \frac{n}{2} \) in the definition of \( W_3 \) indicates that both \( i \) and \( j \) are either smaller or equal or greater than \( \frac{n}{2} \). In terms of interpretation, the matrix \( W_1 \) can be regarded as a weighting matrix, where each firm has the same weight with respect to a portfolio. Thus, the matrix \( W_1 \) captures a general effect, e.g. global crisis, market performance in the past etc.\(^9\) The spatial matrix \( W_2 \) is an example of an asymmetric weighting matrix. \( W_3 \) may be regarded as the dichotomous component of the market which divides the market into two different fields (e.g. the beneficiaries of a given change, e.g. fiscal reform, aid payments, etc. and those who are not affected).

In the first part, the vector of observation \( \mathbf{y}_t \) is generated by a multivariate normal error vector \( \mathbf{\varepsilon}_t \) with zero mean and covariance matrix \( \Sigma := \sigma^2 I_n \), where \( I_n \) represents the \( n \)-dimensional identity matrix and the model representation in (15). The parameter of spatial dependence is given by \( \rho = (0.45, 0.3, 0.15) \) and the homoscedastic variance equals \( \sigma^2 = 2 \).

To calculate the power of our tests, we do not simulate the errors from a multivariate normal with a diagonal covariance matrix. Instead, we use the following misspecification: If we consider a market with \( n \) participants, then there are \( n(n-1)/2 \) possible pairs representing the off-diagonal elements of the covariance matrix. After a simple transformation these off-diagonal elements can be considered as participants that are correlated with each other. The parameter \( \zeta \) describes the portion of how many pairs we wish to consider\(^1\) the parameter \( \kappa^2 \) describes their correlation. E.g. if we consider a market that consists of \( n = 20 \) actors, then there are \( n(n-1)/2 = 190 \) different pairs. If \( \zeta = 0.1 \) and \( \kappa = 0.2 \), we assume that there are 19 pairs that have a correlation coefficient that is equal to 0.04. No further assumptions are made about the structure of correlation. However, the correlation structure is completely random\(^11\) i.e. the correlation of specific pairs is not predetermined; only

\(^{9}\)Even if \( W_1 \) is equally weighted, \( \rho_1 \) cannot be considered as a fixed affect which affects market participant equally, since fixed effects are time independent. SAR models try to capture this time dependence structure with fixed weighting matrices.

\(^{10}\)In case that \( \zeta \cdot n(n-1)/2 \) is odd we round down.

\(^{11}\)This procedure of misspecification ensures that the moment conditions \((17)\) are violated, thus, the GMM estimator is biased \((\text{Hansen}1982)\).
the proportion of correlated pairs and their correlation. To examine size and power properties of the first test $S(2.1)$, we draw $B = 300$ times from the asymptotic limit distribution given in Lemma (2.4). The overall number of MC repetitions is equal to 701. We begin by studying the size of the first test for $n = 20, 50$ and $T = 50, 100, 200, 500$. Results are presented in Table 1. Collectively, the test has good size. Similar properties are derived for the power analysis of the test. Whenever the ratio of $T$ over $n$ is small and the dependence structure in the error term is more or less negligible (cf. $\kappa = \zeta = 0.05$) the power of the test is low. However, if there are sufficient observations (i.e. $\frac{T}{n} > 10$) and if the dependence structure in the data set is not negligible ($\kappa, \zeta \geq 0.1$), the test provides good power properties. All in all we observe an increasing power whenever the dependence structure ($\kappa$ or $\zeta$) or the number of observations ($n$ or $T$) increases.

Similar results are obtained for the second test $S\chi$ which can be found in Table 1. In small samples, $S\chi$ performs worse than $S$ in terms of size and power. This is due to the fact that we are using the empirical approximation for the inverse covariance matrix that is employed in $S\chi$, which is biased in small samples. Consequently, as $T$ tends to infinity the size of the test $S\chi$ converges to the desired nominal level of 5% and the power increases as the level of misspecification rises.

However, additional simulations show that the tests’ power decreases in the case of too large $\zeta$, i.e. the case of a highly non-sparse covariance matrix. Here, the population moment conditions of the GMM estimate (cf. equation (17) given in the Appendix) are severely violated so that the model is misspecified and the behavior of the model estimators $\hat{\rho}$ is unclear [Fleming, 2004].

To summarize, both tests show good size and power properties whenever the ratio $T$ over $n$ is greater or equal to 10. Based on the simple limiting distribution of $S\chi^*$, the test $S\chi$ is also very easy to implement since the test statistic $S\chi$ requires only the empirical covariance matrix of the residuals.

The second MC simulation extends the analyses. Here, we consider a SAR(4) model

$$y_t = \rho_1 W_1 y_t + \rho_2 W_2 y_t + \rho_3 W_3 y_t + \rho_4 W_4 y_t + \epsilon_t, \ t = 1, \ldots, T,$$

where $W_1$ is a group interaction matrix of the first two-thirds (the off-diagonal elements of the first two-third upper sub-matrix are set to $1/(\frac{2}{3} n - 1)$ and all other elements to zero), $W_2$ is a group interaction matrix of the last one-third (cf. $W_1$), $W_3$ a binary contiguity matrix of the third-order neighbors assuming the observations $1, \ldots, n$ are arranged in a circular pattern, e.g., 2 is a neighbor of $n - 1, n, 1, 3, 4, 5$.

12 The second test is applicable since for every simulation it holds true that either $\|\text{(Cov } \alpha)^{-1}\Psi\| < 1$ or $\|\Psi(\text{Cov } \alpha)^{-1}\| < 1$. 


and

$$(W_4)_{ij} = \begin{cases} \frac{1}{2^{n-1}}, & \text{if } i \text{ is even and } j \text{ odd or vice versa} \\ 0, & \text{otherwise.} \end{cases}$$

The vector of autoregressive parameters $\rho$ is given by $\rho = (-0.2 \ 0.05 \ 0.1 \ 0.5)$. Moreover, we presuppose heteroscedastic normal error terms, i.e. $\sigma_i \sim N(0,1)$ for $i = 1,\ldots,n$. In order to analyze the power in case of misspecification, we choose $\zeta$ and $\kappa$ likewise to the first MC simulation. To determine the size and power we draw $B = 300$ times from the asymptotic limit distribution given in Lemma (2.4). The overall number of MC repetitions is equal to 701. The results of the tests can be found in Table 3.

Even if the results of the second analysis are not one-to-one comparable with those from the first simulation, it is clearly observable that the tests $S$ and $S_{\chi}$ hold the size level. The power increases if either the correlation structure ($\kappa$ or $\zeta$) or the number of observation increases ($n$ or $T$). Thus, the results presented in the second, more complex study are in line with those given in the first simulation.

In summary, the MC study has shown that the test is also applicable in case of small samples as long as the vector of observations is sufficiently large compared to the cross sectional dimension $n$. The next section shows that the test even holds size and power demands if the error terms follow a GARCH process.

### 3.2 GARCH(1,1)

One of many problems researchers and practitioners face when analyzing financial data is its volatile structure. Volatility of financial data has been extensively studied in the last twenty years. An important aspect of the analysis is volatility clustering, where conditional heteroskedasticity, which leads to an increase in the probability of rare events, can be modeled with GARCH errors. Since the SEM($m$) model with no covariates (cf. SAR($m$)) is a powerful instrument in modeling financial data, the third Monte Carlo simulation for our proposed test statistic (10) assumes that the errors of the data generating process (DGP) are driven by a GARCH(1,1) model.
i.e. for $t = 1, \ldots, T$ and $i = 1, \ldots, n$

$$y_{it} = \sigma_{it}(I_n - \rho_1 W_1 - \rho_2 W_2 - \rho_3 W_3)^{-1} \epsilon_{it},$$

$$\sigma_{it}^2 = 0.33 + 0.33\sigma_{i(t-1)}^2 + 0.075y_{i(t-1)}^2,$$

$$\epsilon_{it} \sim \text{i.i.d. } N(0, 1).$$

To receive comparable results, the spatial matrices $W_1, W_2, W_3$ are the same to those of the first MC simulation of Section 3. The size and power results are presented in Table 4. At first, it should be noted that the number of observation of a GARCH adjusted data set needs to be significantly higher compared to a data set with no GARCH adjustment, since for the case of a GARCH adjustment an initial estimate needs to be conducted and a primarily high error of estimation violates the stationarity assumption. However, with a sufficiently large set of observations, the test $S(10)$ performs also well with reference to size and power.
Table 1 Size and Power of $S$ for $m = 3$

<table>
<thead>
<tr>
<th>n = 20</th>
<th>$\zeta = 0.05$</th>
<th>$\zeta = 0.1$</th>
<th>$\zeta = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa = 0.00$</td>
<td>$\kappa = 0.05$</td>
<td>$\kappa = 0.1$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>0.031</td>
<td>0.339</td>
<td>0.060</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>0.039</td>
<td>0.056</td>
<td>0.956</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>0.042</td>
<td>0.069</td>
<td>0.220</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>0.034</td>
<td>0.114</td>
<td>0.576</td>
</tr>
<tr>
<td>$T = 1000$</td>
<td>0.045</td>
<td>0.257</td>
<td>0.929</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 20</th>
<th>$\zeta = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa = 0.00$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>0.034</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>0.038</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>0.041</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>0.034</td>
</tr>
<tr>
<td>$T = 1000$</td>
<td>0.045</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 20</th>
<th>$\zeta = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa = 0.00$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>0.033</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>0.036</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>0.029</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>0.037</td>
</tr>
<tr>
<td>$T = 1000$</td>
<td>0.045</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 50</th>
<th>$\zeta = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa = 0.00$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>0.048</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>0.041</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>0.051</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>0.059</td>
</tr>
<tr>
<td>$T = 1000$</td>
<td>0.046</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 50</th>
<th>$\zeta = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa = 0.00$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>0.048</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>0.041</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>0.051</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>0.059</td>
</tr>
<tr>
<td>$T = 1000$</td>
<td>0.045</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 50</th>
<th>$\zeta = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa = 0.00$</td>
</tr>
<tr>
<td>$T = 50$</td>
<td>0.049</td>
</tr>
<tr>
<td>$T = 100$</td>
<td>0.041</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>0.051</td>
</tr>
<tr>
<td>$T = 500$</td>
<td>0.059</td>
</tr>
<tr>
<td>$T = 1000$</td>
<td>0.046</td>
</tr>
</tbody>
</table>

Size and power analysis of the test statistic $S$ [7] with $\rho = (0.45, 0.3, 0.15) \in \mathbb{R}^3$. The DGP follows a multivariate normal distribution where $\zeta$ describes the expected portion of pairs that are correlated with each other with correlation $\kappa^2$ and variance $\sigma^2_i = 2$ for all $i \in \{1, \ldots, n\}$. The number of draws from the limit distribution is set to $B = 300$ by 701 Monte Carlo repetitions.
The number of draws from the limit distribution is set to $T = 1000$.

Table 2: Size and Power of $S_\chi$ for $m = 3$

<table>
<thead>
<tr>
<th>$n = 20$</th>
<th>$\zeta = 0.05$</th>
<th>$\kappa$</th>
<th>$0.00$</th>
<th>$0.05$</th>
<th>$0.1$</th>
<th>$0.15$</th>
<th>$0.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 50$</td>
<td>0.028</td>
<td>0.025</td>
<td>0.040</td>
<td>0.057</td>
<td>0.220</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 100$</td>
<td>0.042</td>
<td>0.030</td>
<td>0.072</td>
<td>0.175</td>
<td>0.679</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 200$</td>
<td>0.038</td>
<td>0.057</td>
<td>0.158</td>
<td>0.503</td>
<td>0.965</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 500$</td>
<td>0.047</td>
<td>0.121</td>
<td>0.567</td>
<td>0.988</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 1000$</td>
<td>0.055</td>
<td>0.233</td>
<td>0.922</td>
<td>0.998</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 20$</th>
<th>$\zeta = 0.1$</th>
<th>$\kappa$</th>
<th>$0.00$</th>
<th>$0.05$</th>
<th>$0.1$</th>
<th>$0.15$</th>
<th>$0.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 50$</td>
<td>0.028</td>
<td>0.037</td>
<td>0.045</td>
<td>0.092</td>
<td>0.561</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 100$</td>
<td>0.042</td>
<td>0.057</td>
<td>0.133</td>
<td>0.414</td>
<td>0.987</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 200$</td>
<td>0.038</td>
<td>0.060</td>
<td>0.384</td>
<td>0.912</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 500$</td>
<td>0.047</td>
<td>0.238</td>
<td>0.937</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 1000$</td>
<td>0.055</td>
<td>0.591</td>
<td>0.998</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 20$</th>
<th>$\zeta = 0.2$</th>
<th>$\kappa$</th>
<th>$0.00$</th>
<th>$0.05$</th>
<th>$0.1$</th>
<th>$0.15$</th>
<th>$0.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 50$</td>
<td>0.028</td>
<td>0.033</td>
<td>0.060</td>
<td>0.253</td>
<td>0.954</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 100$</td>
<td>0.043</td>
<td>0.063</td>
<td>0.346</td>
<td>0.844</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 200$</td>
<td>0.039</td>
<td>0.113</td>
<td>0.831</td>
<td>0.997</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 500$</td>
<td>0.047</td>
<td>0.483</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 1000$</td>
<td>0.055</td>
<td>0.894</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 50$</th>
<th>$\zeta = 0.05$</th>
<th>$\kappa$</th>
<th>$0.00$</th>
<th>$0.05$</th>
<th>$0.1$</th>
<th>$0.15$</th>
<th>$0.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 50$</td>
<td>0.005</td>
<td>0.014</td>
<td>0.016</td>
<td>0.065</td>
<td>0.709</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 100$</td>
<td>0.018</td>
<td>0.027</td>
<td>0.156</td>
<td>0.601</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 200$</td>
<td>0.030</td>
<td>0.771</td>
<td>0.617</td>
<td>0.991</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 500$</td>
<td>0.033</td>
<td>0.369</td>
<td>0.998</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 1000$</td>
<td>0.047</td>
<td>0.829</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 50$</th>
<th>$\zeta = 0.1$</th>
<th>$\kappa$</th>
<th>$0.00$</th>
<th>$0.05$</th>
<th>$0.1$</th>
<th>$0.15$</th>
<th>$0.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 50$</td>
<td>0.005</td>
<td>0.010</td>
<td>0.0411</td>
<td>0.330</td>
<td>0.995</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 100$</td>
<td>0.018</td>
<td>0.047</td>
<td>0.045</td>
<td>0.989</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 200$</td>
<td>0.023</td>
<td>0.164</td>
<td>0.982</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 500$</td>
<td>0.033</td>
<td>0.773</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 1000$</td>
<td>0.047</td>
<td>0.999</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 50$</th>
<th>$\zeta = 0.2$</th>
<th>$\kappa$</th>
<th>$0.00$</th>
<th>$0.05$</th>
<th>$0.1$</th>
<th>$0.15$</th>
<th>$0.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 50$</td>
<td>0.005</td>
<td>0.013</td>
<td>0.203</td>
<td>0.900</td>
<td>0.997</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 100$</td>
<td>0.018</td>
<td>0.106</td>
<td>0.927</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 200$</td>
<td>0.030</td>
<td>0.435</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 500$</td>
<td>0.033</td>
<td>0.997</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$T = 1000$</td>
<td>0.047</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

Size and power analysis of the test statistic $S_\chi$ where $\zeta$ describes the expected portion of pairs that are correlated with each other with correlation $\kappa^2$ and variance $\sigma_i^2 = 2$ for all $i \in \{1, \ldots, n\}$.

The number of draws from the limit distribution is set to $B = 300$ by 701 Monte Carlo repetitions.
Table 3 *Size and Power of S for m = 4*

<table>
<thead>
<tr>
<th>n = 60</th>
<th>ζ = 0.05</th>
<th>ζ = 0.1</th>
<th>ζ = 0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>κ = 0.00</td>
<td>κ = 0.05</td>
<td>κ = 0.1</td>
</tr>
<tr>
<td>T = 50</td>
<td>0.036</td>
<td>0.049</td>
<td>0.073</td>
</tr>
<tr>
<td>T = 100</td>
<td>0.043</td>
<td>0.050</td>
<td>0.114</td>
</tr>
<tr>
<td>T = 200</td>
<td>0.036</td>
<td>0.069</td>
<td>0.227</td>
</tr>
<tr>
<td>T = 500</td>
<td>0.035</td>
<td>0.117</td>
<td>0.681</td>
</tr>
<tr>
<td>T = 1000</td>
<td>0.050</td>
<td>0.329</td>
<td>0.960</td>
</tr>
</tbody>
</table>

Size and power analysis of the test statistic $S$ with $\rho = (-0.2 \ 0.05 \ 0.1 \ 0.5)$. The errors are heteroscedastic, i.e. $\sigma_i \sim N(0,1)$, $i = 1,...,n$. The parameter $\zeta$ describes the portion of expected pairs of firms that are correlated to each other with correlation intensity $\rho^2$. The number of draws from the limit distribution is set to $B = 300$ by 701 Monte Carlo repetitions.
Table 4 Size and Power of \( S \) under GARCH model for \( m = 3 \)

<table>
<thead>
<tr>
<th>( n = 50 )</th>
<th>( \zeta = 0.02 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa = 0.00 )</td>
<td>( \kappa = 0.05 )</td>
</tr>
<tr>
<td>( T = 1000 )</td>
<td>0.086</td>
</tr>
<tr>
<td>( T = 1500 )</td>
<td>0.078</td>
</tr>
<tr>
<td>( T = 2000 )</td>
<td>0.062</td>
</tr>
<tr>
<td>( T = 2500 )</td>
<td>0.068</td>
</tr>
<tr>
<td>( T = 3000 )</td>
<td>0.044</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( n = 50 )</th>
<th>( \zeta = 0.04 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa = 0.00 )</td>
<td>( \kappa = 0.05 )</td>
</tr>
<tr>
<td>( T = 1000 )</td>
<td>0.086</td>
</tr>
<tr>
<td>( T = 1500 )</td>
<td>0.078</td>
</tr>
<tr>
<td>( T = 2000 )</td>
<td>0.062</td>
</tr>
<tr>
<td>( T = 2500 )</td>
<td>0.068</td>
</tr>
<tr>
<td>( T = 3000 )</td>
<td>0.044</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( n = 50 )</th>
<th>( \zeta = 0.1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa = 0.00 )</td>
<td>( \kappa = 0.05 )</td>
</tr>
<tr>
<td>( T = 1000 )</td>
<td>0.086</td>
</tr>
<tr>
<td>( T = 1500 )</td>
<td>0.078</td>
</tr>
<tr>
<td>( T = 2000 )</td>
<td>0.062</td>
</tr>
<tr>
<td>( T = 2500 )</td>
<td>0.068</td>
</tr>
<tr>
<td>( T = 3000 )</td>
<td>0.044</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( n = 80 )</th>
<th>( \zeta = 0.02 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa = 0.00 )</td>
<td>( \kappa = 0.05 )</td>
</tr>
<tr>
<td>( T = 1000 )</td>
<td>0.073</td>
</tr>
<tr>
<td>( T = 1500 )</td>
<td>0.070</td>
</tr>
<tr>
<td>( T = 2000 )</td>
<td>0.043</td>
</tr>
<tr>
<td>( T = 2500 )</td>
<td>0.060</td>
</tr>
<tr>
<td>( T = 3000 )</td>
<td>0.050</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( n = 80 )</th>
<th>( \zeta = 0.04 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa = 0.00 )</td>
<td>( \kappa = 0.05 )</td>
</tr>
<tr>
<td>( T = 1000 )</td>
<td>0.073</td>
</tr>
<tr>
<td>( T = 1500 )</td>
<td>0.070</td>
</tr>
<tr>
<td>( T = 2000 )</td>
<td>0.043</td>
</tr>
<tr>
<td>( T = 2500 )</td>
<td>0.060</td>
</tr>
<tr>
<td>( T = 3000 )</td>
<td>0.050</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( n = 80 )</th>
<th>( \zeta = 0.1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa = 0.00 )</td>
<td>( \kappa = 0.05 )</td>
</tr>
<tr>
<td>( T = 1000 )</td>
<td>0.073</td>
</tr>
<tr>
<td>( T = 1500 )</td>
<td>0.070</td>
</tr>
<tr>
<td>( T = 2000 )</td>
<td>0.043</td>
</tr>
<tr>
<td>( T = 2500 )</td>
<td>0.060</td>
</tr>
<tr>
<td>( T = 3000 )</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Size and power analysis of the test statistic \( S \) with \( \rho = (0.45, 0.3, 0.15) \) under a GARCH model. The data generating process is GARCH(1,1) with constant and GARCH parameter equal to 0.33 and ARCH parameter equal to 0.075 with standard normal errors. \( \zeta \) describes the expected portion of pairs that are correlated with each other with correlation \( \kappa^2 \) and variance \( \sigma_i^2 = 2 \) for all \( i \in \{1, \ldots, n\} \).
4 Empirical Analysis

In the empirical application, we analyze the spatial dependencies of the daily stock returns of the Euro Stoxx 50 members for the period from April 2003 to December 2009 and from January 2018 to December 2020, using the composition of January 2010 and January 2018\textsuperscript{16}. The stock prices are adjusted and transferred to log returns.

To model the spatial relations, we consider four spatial matrices $W_G$, $W_C$, $W_B$ and $W_I$, which are designed to capture the dependencies in different ways. The matrix $W_G$ represents general dependencies\textsuperscript{17} and the matrices $W_C$ and $W_B$ local and industry-specific dependencies, respectively\textsuperscript{18}. Finally, the matrix $W_I$ captures the spatial interaction effect of countries and industries, i.e. we affiliate two stocks that belong to the same country and industry. Note that due to construction, none of the spatial matrices $W_G$, $W_C$, $W_B$ and $W_I$ are symmetric.

In our analysis, we examine 15 different SAR models, i.e. we consider SAR models, where the number and composition of spatial matrices varies between the models.

In the simplest case, only one spatial matrix constitutes our SAR model, e.g., $y_t = \rho G W_G y_t + \varepsilon_t$; in the most complex case, all four spatial matrices are included, i.e., $y_t = \rho G W_G y_t + \rho C W_C y_t + \rho B W_B y_t + \rho I W_I y_t + \varepsilon_t$. Thus, for $m \in \{1, 2, 3, 4\}$ the general model for the log stock returns on day $t = 1, ..., T$ is given by

\[ y_t = \sum_{i=1}^{m} \rho_i W_i y_t + \varepsilon_t \text{ with } W_I \in \{W_G, W_C, W_B, W_I\} \]

where $y_t$ is the vector of log stock returns on day $t$ and $\rho_i$ the corresponding spatial parameter. In this model, we can interpret the terms $W_i y_t$ as specific factors, where external information (about countries and industries) is included. The term $W_G y_t$ represents the market factor. Our empirical analysis is motivated by two key questions: First, how well can the different SAR models explain the movement of stocks and which spatial matrices do we need for a sufficiently good model? Secondly, can we use the model specification test applied to the 15 different models to ascertain which spatial matrix provides the greatest explanatory power?

To answer these questions, we consider a rolling window of the length of one quarter...
<table>
<thead>
<tr>
<th>Branch</th>
<th>2010</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finance</strong></td>
<td>Aegon, Allianz, AXA, BCO Bilbao, BCO Santander, BNP, Crédit Agricole, Deutsche Bank, Deutsche Börse, Generali, ING, Intesa, Münchener Rück, Société Générale, Unicredit</td>
<td>Allianz, AXA, BCO Bilbao, BCO Santander, BNP Paribas, Deutsche Bank, Deutsche Post, Société Générale, ING Group, Intesa Sanpaolo, Münchener Rück, Unibail-Rodamco</td>
</tr>
<tr>
<td><strong>Automobil</strong></td>
<td>Daimler, Renault, VW</td>
<td>BMW, Daimler, Volkswagen</td>
</tr>
<tr>
<td><strong>Energy</strong></td>
<td>Alstom, E.ON, ENEL, ENI, Iberdrola, Repsol, BWE, SUEZ, Total</td>
<td>ENEL, Engie, ENI, E.ON, Iberdrola, Total</td>
</tr>
<tr>
<td><strong>Telecom and Media</strong></td>
<td>Dt. Telekom, France Telecom, Telecom Italia, Telefonica, Vivendi</td>
<td>Dt. Telekom, Nokia, Orange, Telefonica, Vivendi</td>
</tr>
<tr>
<td><strong>Pharma and Chemicals</strong></td>
<td>Air Liquide, BASF, Bayer, Sanofi</td>
<td>Air Liquide, BASF, Bayer, Essilor, Fresenius, Sanofi</td>
</tr>
<tr>
<td><strong>Consumer Electronics</strong></td>
<td>Nokia, Philips, SAP, Siemens, Schneider</td>
<td>Airbus, ASML, Philips Electronics, Safran, Schneider Electric, Siemens</td>
</tr>
<tr>
<td><strong>Consumer retail</strong></td>
<td>Anheuser Busch, Carrefour, Danone, L’Oréal, LVMH, Unilever</td>
<td>AB Inbev, Adidas, Ahold Delhaize, Danone, Inditex, L’Oréal, LVMH, Unilever</td>
</tr>
<tr>
<td><strong>Basic Industry</strong></td>
<td>Arcelor Mittal, CRH, Saint Gobain, Vinci</td>
<td>CRH, Saint Gobain, Vinci</td>
</tr>
<tr>
<td><strong>Benelux</strong></td>
<td>Aegon, Anheuser Busch, Arcelor, ING, Philips, Unilever</td>
<td>AB Inbev, Ahold Delhaize, ASML ING Group, Philips Electronics, Unilever</td>
</tr>
<tr>
<td><strong>France</strong></td>
<td>Air Liquide, Alstom, AXA, BNP, Carrefour, Crédit, Agricole, France Telecom, Danone, L’Oréal, LVMH, Saint Gobain, Sanofi, Schneider, Société Générale, SUEZ, Total, Vinci, Vivendi</td>
<td>Air Liquide, Airbus, AXA, BNP, Danone, Engie, Essilor, Société Générale, L’Oréal, LVMH, Orange, Safran, Saint Gobain, Sanofi, Schneider Electric, Total, Unibail-Rodamco, Vinci, Vivendi</td>
</tr>
<tr>
<td><strong>Germany</strong></td>
<td>Allianz, BASF, Bayer, Daimler, Deutsche Bank, Deutsche Börse, Dt. Telekom, E.ON, Münchener Rück, RWE, SAP, Siemens, VW</td>
<td>Adidas, Allianz, BASF, Bayer, BMW, Daimler, Deutsche Bank, Deutsche Post, Dt. Telekom, E.ON, Fresenius, Münchener Rück, SAP, Siemens, Volkswagen</td>
</tr>
<tr>
<td><strong>Italy</strong></td>
<td>Generali, ENEL, ENI, Intesa, Telecom Italia, Unicredit</td>
<td>ENEL, ENI, Intesa Sanpaolo</td>
</tr>
<tr>
<td><strong>Spain</strong></td>
<td>BCO Bilbao, BCO Santander, Iberdrola, Repsol, Telefonica</td>
<td>BCO Bilbao, BCO Santander Iberdrola Inditex TELEFONICA</td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td>CRH, Nokia</td>
<td>CRH, Nokia</td>
</tr>
</tbody>
</table>

The table depicts the Euro Stoxx 50 members in the composition of January 2010 and January 2018. The upper table shows the country and the lower table the sector affiliation. The spatial matrices $W_B$ and $W_C$ are constructed in the following way: The off-diagonal elements are set to 1 if the corresponding stocks belong to the same branch ($W_B$) or country ($W_C$). The off-diagonal elements of the interaction matrix $W_I$ are set to 1 if the corresponding stocks belong to the same branch and country. Finally, the rows are standardized according to the market capitalization.

$(T = 63)$, i.e., we check whether our model specification test rejects the null hypothesis that the log returns can be modeled by an SAR($m$) model for $m \in \{1, 2, 3, 4\}$. Furthermore, we use the observation period of the rolling window to calculate one-
day Value-at-Risk (VaR) forecasts for each model to see if our new specification test
could also be used as a backtest in the spirit of Ziggel et al. (2014) among others.
A high proportion of incorrect VaR forecasts speaks additionally against the under-
lying model. All analyses are performed on a significance level of 5%. The number
of draws from the limit distribution is set to $B = 300$. The results for the time
periods from 2003 to 2009 and 2018 to 2020 are depicted in Figure 1 and Figure 2,
respectively.
Figure 1 shows the results of our suggested model specification test applied to the
15 models from (16) in a rolling window for the period from April 2003 to December
2009 of the Euro Stoxx 50. In order to apply the model specification test we first
calculate the spatial parameters of the underlying model using a two step GMM
estimation procedure (cf. Appendix A) and then apply our model specification test.
The blue line in the subplots of Figure 1 describes the ratio of the test statistic
and its 95%-quantile. Thus, if the value is greater than 1, there is no statistical
evidence to reject the null hypothesis. A value smaller than 1, however, provides
evidence against the underlying model. In addition, we generate VaR forecasts with
standard normally distributed errors for the minimum-variance portfolio based on
the underlying model on a rolling window of size $T = 63$. The proportion of incorrect
VaR forecasts is represented by the number in each subplot. In Figure 1, a total of
four core observations can be made:
First, it seems that more complex models with at least two spatial matrices de-
scribe the data better. On the one hand, the period in which the null hypothesis
of a correct model assumption cannot be rejected is longer and on the other hand,
false VaR predictions are close to the significance level $\alpha = 5\%$. Secondly, Figure
1 illustrates that in periods of economic crisis the spatial model (16) is applica-
table to a lesser extent, since in all models the ratio (blue line) is on a significantly
lower level. Particularly, this can be observed in the period of the financial crisis
beginning in summer 2007, which is also consistent that in times of bear markets
the correlation among market participants increases significantly. Thus, the result-
ing extensive dependency structure cannot be well captured by a simple SAR($m$)
model. Accordingly, the results of our test provide evidence that the effects of the
dot-com bubble crisis around 2000 last until summer 2004, since the test rejects the
application of model (16). In the two following years (2004-2006), however, subplot
6 of Figure 1 depicts evidence to apply the model, since the blue line is often greater
than 1. Finally, in the remaining observation period the test indicates that a spa-
tial model is inappropriate, which overlaps with the period of the financial crisis.
Thirdly, however, applying the model specification test to the 15 different models
also clearly demonstrates that the spatial matrix $W_G$, which describes a general
influence, provides the strongest explanatory power. Country and industry specific
effects seem to have a minor impact ($W_C$ and $W_B$) when considering the ratio and the number of VaR forecast violations (0.1804 for subplot 2 and 0.2372 for subplot 3, respectively). Fourthly, the interaction effect of industry and country affiliation seems to play a minor role in the Euro Stoxx 50 modeling, as both the ratios and the VaR forecast violations of subplots 11 and 15 are similar (0.0546 and 0.0579, respectively). Taking the ratio of VaR forecast violations, the time period at which the null hypothesis cannot be rejected and sparsity in the number of variables as goodness of fit criteria, the SAR(2) incorporating general and industry-specific dependencies, i.e. $y_t = p_G W_G + p_B W_B + \varepsilon_t$ (cf. subplot 6 of Figure 1), seems to be the most appropriate.

Figure 2 shows the results of the second application of our proposed model specification test for the period of 2018-2020, where our proceeding is identical to that from the first empirical application: First, we compute the spatial model parameters using a two-step GMM approach. Then, in a second step, we calculate the test statistics of our model specification test. The results of the second application seem to corroborate the findings of the first application: Even if the 15 considered SAR models are rejected by our model specification test in the period 2018-2020, it is striking that models which incorporate at least two spatial matrices model the Euro Stoxx 50 better. This can be seen from the fact that in models with at least two appropriate spatial matrices ($W_G$ should be incorporated) both the ratio of the test statistic and its quantile is larger and VaR forecast violations correspond approximately to the significance level of $\alpha = 5\%$. Furthermore, the beginning of the COVID-19 pandemic in March 2020 is also clearly visible in every subplot of Figure 2 as the ratio (blue line) decreases significantly. In summary, in terms of sparsity with regard to the number of variables included in a model, the ratio of the test statistic and the quantile and the proportion of VaR forecast violations the SAR(2) model incorporating general and industry-specific effects (cf. subplot 6 in Figure 2) seems to be the most suitable modeling the Euro Stoxx 50 in the periods considered.
Rolling window parameter estimation of size $T = 63$ in a data set of size 1861 and dimension $n = 50$. 15 different models are considered. The first figure describes the SAR(1) model incorporating only a single general dependence structure ($W_G$). The last figure depicts the results of the SAR(4) model incorporating the spatial matrices $W_G, W_C, W_B, W_I$. The number of draws from the limit distribution is set to $B = 300$. The blue line depicts the ratio of the $95\%$-quantile of the limit distribution given in Lemma 2.4 over the test statistic $S$ from (7). A value below $1$ indicates that the assumed model is not correct. The orange line is the accumulated spatial dependence parameter in the $L^1$-norm. The number in each subplot represents the proportion of VaR forecast violations.
Rolling window parameter estimation of size $T = 63$ in a data set of size 653 and dimension $n = 50$. 15 different models are considered. The first figure describes the SAR(1) model incorporating only a single general dependence structure ($W_G$). The last figure depicts the results of the SAR(4) model incorporating the spatial matrices $W_G, W_C, W_B, W_I$. The number of draws from the limit distribution is set to $B = 300$. The blue line depicts the ratio of the 95%-quantile of the limit distribution given in Lemma (2.4) over the test statistic $S$ from (7). A value below 1 indicates that the assumed model is not correct. The orange line is the accumulated spatial dependence parameter in the $L^1$-norm. The number in each subplot represents the proportion of VaR forecast violations.
5 Conclusion

We propose specification tests for spatial models and analyze the size and power of these tests. The proposed tests show good size and power properties in finite samples for both initial data and GARCH adjusted data. An empirical analysis of the Euro Stoxx 50 between 2003 and 2009 and 2018 and 2020 substantiates that bull markets provide statistical evidence to apply a SAR model which models general and industry-specific dependencies. However, in bear markets a simple spatial model captures the extensive structure of relations and market dependencies to a lesser extent. Accordingly, our proposed testing procedure provides statistical evidence that the model fit is worse in the time after the dot-com bubble, the time around the Lehman Brothers bankruptcy and the time of the COVID-19 pandemic.

References


A Two Step GMM Estimation Procedure for SAR($m$) Models

In the following, we assume that Assumption 1 and 2 is fulfilled, i.e., in particular $\hat{\beta}$ is a consistent estimator for $\beta$. Due to readability, we again suppress the conditioning
on $\hat{\beta}$ in the following. The covariance matrix of $\boldsymbol{u}_t = (I_n - \sum_{i=1}^{m} \rho_i W_i)^{-1} \boldsymbol{e}_t$ is given by

$$
\text{Cov}[\boldsymbol{u}_t] = \left(I_n - \sum_{i=1}^{m} \rho_i W_i\right)^{-1} \Sigma \left(I_n - \sum_{i=1}^{m} \rho_i W_i'\right)^{-1} =: V.
$$

For the estimation conditioned on a consistent estimate $\hat{\beta}$, a two step procedure is considered: First, we estimate the correlation parameters by the method of moments which does not depend on the parameters of variance. Secondly, we estimate the variance parameters.

The moment estimator for the correlation parameters uses the following $m$-moment conditions:

$$
E[\boldsymbol{e}_t' W_i \boldsymbol{e}_t] = \text{tr}(W_i \Sigma) = 0 \quad \text{for} \quad i = 1, \ldots, m.
$$

 Clearly, the variance parameters $\sigma_i^2$ for $i = 1, \ldots, m$ do not enter the moment conditions. Replacing $\boldsymbol{e}_t$ by

$$
\boldsymbol{e}_t = \left(I_n - \sum_{i=1}^{m} \rho_i W_i\right) \boldsymbol{u}_t
$$

and averaging over $t$ gives the theoretical system of equations

$$
\Gamma \lambda + \gamma = 0,
$$

where $\lambda := \lambda(\rho)$ is a functional vector of $\rho := (\rho_1, \ldots, \rho_m)$ of dimension $M := \binom{m}{1} + \binom{m+2-1}{2}$, $\binom{\cdot}{\cdot}$ denoting the binomial coefficient, such that

$$
\lambda_i = \rho_i \quad \text{for} \quad i = 1, \ldots, m
$$

$$
\lambda_{m+i} = \rho_i^2 \quad \text{for} \quad i = 1, \ldots, m
$$

$$
\lambda_{2m + \# \{ij \mid i < j, i < l, j \leq k\}} = \rho_l \rho_k \quad \text{for} \quad l, k = 1, \ldots, m,
$$

where $\# \{ij \mid i < j, i < l, j \leq k\}$ represents the number of integer pairs $ij$ such that the conditions $i < j, i < l$ and $j \leq k$ are fulfilled for $l, k = 1, \ldots, m$. The elements of
$\Gamma \in \mathbb{R}^{m \times M}$ and $\gamma \in \mathbb{R}^{m}$ are defined by for $i,j = 1,\ldots,m$,

$$
\Gamma_{i,j} = \mathbb{E} \left[ -\frac{1}{T} \sum_{t=1}^{T} u_i'(W_i + W_i') W_j u_t \right],
$$
(21)

$$
\Gamma_{i,m+j} = \mathbb{E} \left[ \frac{1}{T} \sum_{t=1}^{T} u_i' W_j' W_i u_t \right],
$$
(22)

$$
\Gamma_{i,2m+\#\{ij|i<j,i<l,j\leq k\}} = \mathbb{E} \left[ \frac{1}{T} \sum_{t=1}^{T} u_i' W_j' \left( W_i + W_i' \right) W_k u_t \right],
$$
(23)

$$
\gamma_i = \mathbb{E} \left[ \frac{1}{T} \sum_{t=1}^{T} u_i' W_i u_t \right].
$$

Let $G$ and $\mathbf{g}$ be the empirical counterparts of $\Gamma$ and $\gamma$, i.e. the expectation operator is left out. The moment estimator for $\rho = (\rho_1,\ldots,\rho_m)'$ is defined as

$$
\hat{\rho} := (\hat{\rho}_1,\ldots,\hat{\rho}_m)' := \arg\min_{\rho \in S} ||G\rho + \mathbf{g}||
$$

where $||\cdot||$ represents the euclidean norm.

**Remark A.1.** For $k,l \in \{1,\ldots,m\}$, the entries of $\mathbb{E}[G] = \Gamma$ given in (21)-(23) can be calculated as

$$
\Gamma_{k,l} = \text{tr} \left( \left( W_k + W_k' \right) W_l V \right),
$$

$$
\Gamma_{k,m+l} = \text{tr} \left( W_l' W_k W_i V \right),
$$

$$
\Gamma_{i,2m+\#\{ij|i<j,i<l,j\leq k\}} = \text{tr} \left( W_i' \left( W_i + W_i' \right) W_k V \right).
$$

Since the theoretical term $\Gamma\lambda + \gamma$ is equal to zero for the true parameter values, the moment estimator for $\hat{\rho}$ minimizes the corresponding empirical system $G\lambda + \mathbf{g}$. *Arnold et al.* [2013] prove consistency and asymptotic normality of the moment estimator (cf. Theorem A.2) for $T \to \infty$, for which an additional assumption is needed.

**Assumption 3.**

1. The true parameter $\rho \in S$ is the unique solution of the theoretical system of equations, i.e.

$$
\Gamma\lambda + \gamma = 0 \Leftrightarrow \hat{\rho} = \rho.
$$

2. The matrix $\mathbb{E} \left( \frac{\partial (G\lambda + \mathbf{g})}{\partial \rho} (y_t,\rho) \right) =: \mathbf{d} = \Gamma^{(1)}$ exists, is finite and has full rank.
with $\lambda^{(1)}$ a $(M \times m)$ dimensional matrix defined as

$$
\lambda^{(1)}(l,l) = 1, \quad \lambda^{(1)}(m+l,l) = 2\rho_l, \quad \lambda^{(1)}(2m+\#\{ij \mid i < j, i < l, j \leq k\}, l) = \rho_k
$$

for all $l,k = 1,\ldots,m$.

3. For

$$
f(u_t, \rho) = \begin{pmatrix}
\varepsilon_t' W_1 \varepsilon_t \\
\vdots \\
\varepsilon_t' W_m \varepsilon_t
\end{pmatrix}
$$

it holds that, for $j \to \infty$, $\mathbb{E} [f(u_t, \rho) | f(u_{t-j}, \rho), f(u_{t-j-1}, \rho), \ldots] \}$ converges in mean square to zero and that, for

$$
v_j := \mathbb{E} [f(u_t, \rho) | f(u_{t-j}, \rho), f(u_{t-j-1}, \rho), \ldots] - \mathbb{E} [f(u_t, \rho) | f(u_{t-j-1}, \rho), f(u_{t-j-2}, \rho), \ldots]
$$

the infinite sum $\sum_{t=-\infty}^{\infty} \mathbb{E} [(v_j^t)^2]$ is finite.

Under the Assumptions 1, 2 and 3 the GMM estimator $\hat{\rho}$ is consistent and asymptotic normal as the following theorem shows:

**Theorem A.2.** Let Assumptions 1, 2 and 3 hold. Then, for $S_W = \sum_{t=-\infty}^{\infty} \mathbb{E} [f(u_t, \rho) f(u_t, \rho)' ]$ and $T \to \infty$ it holds:

1. $\hat{\rho} \overset{p}{\to} \rho$

2. $\sqrt{T}(\hat{\rho} - \rho) \overset{d}{\to} N(0, d^{-1} S_W (d^{-1})' )$ .

**B Proofs**

Theorem 2.1 is proved by means of the following lemmata.

**Lemma B.1.** Let $I_n$ denote the $n$-dimensional identity matrix and $W$ the stack of spatial matrices, i.e. $W' = (W_1', \ldots, W_m')$ with $W_i \in \mathbb{R}^{n \times n}$ for $i = 1,\ldots,m$. Under Assumption 1 and given that $\{\varepsilon_t\}_{t \in \{1,\ldots,T\}}$ is serially independent the following holds for $\rho := (\rho_1, \ldots, \rho_m)$ and $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_m)$

$$
\sqrt{T} \text{ Cov} [\varepsilon_t] = \frac{1}{\sqrt{T}} \sum \varepsilon_t \varepsilon_t' + \frac{1}{T} \sum \Delta_T \varepsilon_t \Delta_t' + \frac{1}{T} \sum \varepsilon_t \varepsilon_t' \Delta_T' + \frac{1}{T} \sum \Delta_T \varepsilon_t \varepsilon_t' \frac{\Delta_T'}{\sqrt{T}}
$$
with $\Delta_T := \sqrt{T}((\rho - \hat{\rho}) \otimes I_n)W(I_n - (\rho \otimes I_n)W)^{-1}$, where $\otimes$ represents the Kronecker product.

Proof. It holds:

$$\sqrt{T} \text{Cov} [\hat{e}_t] = \sqrt{T} \hat{E} [\hat{e}_t \hat{e}_t']$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{e}_t \hat{e}_t' = \frac{1}{\sqrt{T}} \sum (I_n - (\hat{\rho} \otimes I_n)W)u_t u_t'(I_n - (\hat{\rho} \otimes I_n)W)'$$

$$= \frac{1}{\sqrt{T}} \sum (I_n - (\hat{\rho} \otimes I_n)W)(I_n - (\rho \otimes I_n)W)^{-1} \epsilon_t \epsilon_t'$$

$$= \frac{1}{\sqrt{T}} \sum (I_n - (\rho \otimes I_n)W + (\rho \otimes I_n)W - (\hat{\rho} \otimes I_n)W)(I_n - (\rho \otimes I_n)W)^{-1} \epsilon_t \epsilon_t'$$

$$= \frac{1}{\sqrt{T}} \sum [(I_n - (\rho \otimes I_n)W + (\rho \otimes I_n)W - (\hat{\rho} \otimes I_n)W)(I_n - (\rho \otimes I_n)W)^{-1}]'$$

$$= \frac{1}{T} \sum [\sqrt{T} I_n + \sqrt{T}((\rho - \hat{\rho}) \otimes I_n)W(I_n - (\rho \otimes I_n)W)^{-1}] \epsilon_t \epsilon_t'$$

$$= \frac{1}{T} \sum [\sqrt{T} I_n + \Delta_T] \epsilon_t \epsilon_t' [I_n + \frac{\Delta_T}{\sqrt{T}}]'$$

$$= \frac{1}{T} \sum [\sqrt{T} \epsilon_t \epsilon_t' + \Delta_T \epsilon_t \epsilon_t' \frac{\Delta_T}{\sqrt{T}} + \sqrt{T} \epsilon_t \epsilon_t' + \Delta_T \epsilon_t \epsilon_t']$$

$$= \frac{1}{\sqrt{T}} \sum \epsilon_t \epsilon_t' + \frac{1}{T} \sum \Delta_T \epsilon_t \epsilon_t' + \frac{1}{T} \sum \epsilon_t \epsilon_t' \Delta_T + \frac{1}{T} \sum \Delta_T \epsilon_t \epsilon_t' \Delta_T$$

$$= \lim_{T \to \infty} \frac{1}{T} \sum \epsilon_t \epsilon_t' + \lim_{T \to \infty} \frac{1}{T} \sum \Delta_T \epsilon_t \epsilon_t' + \lim_{T \to \infty} \frac{1}{T} \sum \epsilon_t \epsilon_t' \Delta_T + \lim_{T \to \infty} \frac{1}{T} \sum \Delta_T \epsilon_t \epsilon_t' \Delta_T$$

$$= \lim_{T \to \infty} \frac{1}{T} \sum \epsilon_t \epsilon_t' + \lim_{T \to \infty} \frac{1}{T} \sum \Delta_T \epsilon_t \epsilon_t' + \lim_{T \to \infty} \frac{1}{T} \sum \epsilon_t \epsilon_t' \Delta_T + \lim_{T \to \infty} \frac{1}{T} \sum \Delta_T \epsilon_t \epsilon_t' \Delta_T$$

The claim in Theorem 2.1 is achieved by standard arguments and an adjustment of Theorem 2.1. in [Arnold et al.] (2013).

Lemma B.2. If presume the same Assumptions as in Lemma B.1, then $\alpha = (A)_{i \neq j, i \neq j} = \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_{1t} \epsilon_{2t}\right)_{i < j, i \neq j}$ has expectation zero and the following covariance matrix

$$\text{Cov} [\alpha] = \begin{pmatrix}
\lim_{T \to \infty} \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_{1t} \epsilon_{2t}\right] & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \lim_{T \to \infty} \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_{(n-1)t} \epsilon_{nt}\right]
\end{pmatrix}.$$
Proof. The zero mean statement follows directly from the cross-sectional uncorrelatedness for every $t = 1, \ldots, T$. Furthermore, we observe

$$
\text{Cov} [\alpha] = \lim_{T \to \infty} \begin{pmatrix}
\text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{1t} \varepsilon_{2t} \right] & \cdots & \frac{1}{T} \text{Cov} \left[ \sum_{t=1}^{T} \varepsilon_{1t} \varepsilon_{2t} \sum_{t=1}^{T} \varepsilon_{1s} \varepsilon_{ns} \right] \\
\vdots & \cdots & \vdots \\
\frac{1}{T} \text{Cov} \left[ \sum_{t=1}^{T} \varepsilon_{1t} \varepsilon_{nt} \sum_{t=1}^{T} \varepsilon_{1s} \varepsilon_{2s} \right] & \cdots & \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{(n-1)t} \varepsilon_{nt} \right]
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\lim_{T \to \infty} \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{1t} \varepsilon_{2t} \right] & \cdots & 0 \\
\vdots & \cdots & \vdots \\
0 & \cdots & \lim_{T \to \infty} \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{(n-1)t} \varepsilon_{nt} \right]
\end{pmatrix}
$$

$$
\in \mathbb{R}^{n(n-1) \times n(n-1)}.
$$
Lemma B.3. If presume the same Assumptions as in Lemma B.1 and \( \{\varepsilon_t\}_{t \in \{1, \ldots, T\}} \) being serially independent, then \( \alpha = (A)_{i<j, i \neq j} = \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \varepsilon_t' \right)_{i<j, i \neq j} \) is multivariate normally distributed with expectation zero and

\[
\text{Cov} \left[ \left( \varepsilon_t \varepsilon_t' \right)_{i<j, i \neq j} \right] = \text{Cov} [\alpha] = \text{diag} \left( \sigma_1^2, \sigma_2^2, \ldots, \sigma_{n-1}^2, \sigma_n^2 \right).
\]

Proof. The vector \( \alpha \) can be rewritten as \( \sqrt{T} \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \varepsilon_t' \right) \). By Assumption 1.5 and the multivariate central limit theorem we obtain that \( \alpha \) is normally distributed with expectation zero. Since we assume uncorrelatedness in the cross-section for every \( t = 1, \ldots, T \), we have for \( i \neq j \neq k \neq i \)

\[
\begin{align*}
\text{Cov} [\varepsilon_t \varepsilon_t', \varepsilon_t \varepsilon_j] &= E \left[ \varepsilon_t^2 \varepsilon_j^2 \right] - 0 = \sigma_i^2 \sigma_j^2, \quad \text{(24)} \\
\text{Cov} [\varepsilon_t \varepsilon_t', \varepsilon_t \varepsilon_k] &= E \left[ \varepsilon_t^2 \varepsilon_k^2 \right] - 0 = E \left[ \varepsilon_t^2 \right] E \left[ \varepsilon_t \varepsilon_k \right] = 0. \quad \text{(25)}
\end{align*}
\]

Thus, the covariance matrix for the limiting normal distribution is given by

\[
\text{Cov} [\alpha] = \\
\begin{pmatrix}
\text{Cov} \left[ \varepsilon_{1t} \varepsilon_{2t}, \varepsilon_{1t} \varepsilon_{2t} \right] & \cdots & \text{Cov} \left[ \varepsilon_{1t} \varepsilon_{2t}, \varepsilon_{(n-1)t} \varepsilon_{nt} \right] \\
\text{Cov} \left[ \varepsilon_{1t} \varepsilon_{3t}, \varepsilon_{1t} \varepsilon_{2t} \right] & \cdots & \text{Cov} \left[ \varepsilon_{1t} \varepsilon_{3t}, \varepsilon_{(n-1)t} \varepsilon_{nt} \right] \\
\vdots & \ddots & \vdots \\
\text{Cov} \left[ \varepsilon_{(n-1)t} \varepsilon_{nt}, \varepsilon_{1t} \varepsilon_{2t} \right] & \cdots & \text{Cov} \left[ \varepsilon_{(n-1)t} \varepsilon_{nt}, \varepsilon_{(n-1)t} \varepsilon_{nt} \right] \\
\sigma_1^2 & \cdots & 0 \\
0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \sigma_{(n-1)}^2 \sigma_n^2
\end{pmatrix} \in \mathbb{R}^{n(n-1)/2 \times n(n-1)/2}.
\]