IHS Economics Series
Working Paper 296
May 2013

Parameter Estimation and Inference with Spatial Lags and Cointegration

Jan Mutl
Leopold Sögner

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

Impressum

Author(s):
Jan Mutl, Leopold Sögner

Title:
Parameter Estimation and Inference with Spatial Lags and Cointegration

ISSN: Unspecified

2013 Institut für Höhere Studien - Institute for Advanced Studies (IHS)
Josefstädter Straße 39, A-1080 Wien
E-Mail: office@ihs.ac.at
Web: www.ihs.ac.at

All IHS Working Papers are available online: http://irihs.ihs.ac.at/view/ihs_series/
This paper is available for download without charge at: http://irihs.ihs.ac.at/2201/

296

Reihe Ökonomie

Economics Series

Parameter Estimation and Inference with Spatial Lags and Cointegration

Jan Mutl, Leopold Sögner

296
Reihe Ökonomie
Economics Series

Parameter Estimation and Inference with Spatial Lags and Cointegration

Jan Mutl, Leopold Sögner

May 2013

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Jan Mutl
EBS Business School
Gustav-Streseman-Ring 3
65189 Wiesbaden, GERMANY
email: jan.mutl@ebs.edu

Leopold Sögner
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, AUSTRIA
email: soegner@ihs.ac.at

Founded in 1963 by two prominent Austrians living in exile – the sociologist Paul F. Lazarsfeld and the economist Oskar Morgenstern – with the financial support from the Ford Foundation, the Austrian Federal Ministry of Education and the City of Vienna, the Institute for Advanced Studies (IHS) is the first institution for postgraduate education and research in economics and the social sciences in Austria. The **Economics Series** presents research done at the Department of Economics and Finance and aims to share "work in progress" in a timely way before formal publication. As usual, authors bear full responsibility for the content of their contributions.

Das Institut für Höhere Studien (IHS) wurde im Jahr 1963 von zwei prominenten Exilösterreichern – dem Soziologen Paul F. Lazarsfeld und dem Ökonomen Oskar Morgenstern – mit Hilfe der Ford-Stiftung, des Österreichischen Bundesministeriums für Unterricht und der Stadt Wien gegründet und ist somit die erste nachuniversitäre Lehr- und Forschungsstätte für die Sozial- und Wirtschaftswissenschaften in Österreich. Die **Reihe Ökonomie** bietet Einblick in die Forschungsarbeit der Abteilung für Ökonomie und Finanzwirtschaft und verfolgt das Ziel, abteilungsinterne Diskussionsbeiträge einer breiteren fachinternen Öffentlichkeit zugänglich zu machen. Die inhaltliche Verantwortung für die veröffentlichten Beiträge liegt bei den Autoren und Autorinnen.

Abstract

We study dynamic panel data models where the long run outcome for a particular cross-section is affected by a weighted average of the outcomes in the other cross-sections. We show that imposing such a structure implies several cointegrating relationships that are nonlinear in the coefficients to be estimated. Assuming that the weights are exogenously given, we extend the dynamic ordinary least squares methodology and provide a dynamic two-stage least squares estimator. We derive the large sample properties of our proposed estimator and investigate its small sample distribution in a simulation study. Then our methodology is applied to US financial market data, which consist of credit default swap spreads, firm specific and industry data. A "closeness" measure for firms is based on input-output matrices. Our estimates show that this particular form of spatial correlation of credit default spreads is substantial and highly significant.

Keywords

Dynamic ordinary least squares, cointegration, credit risk, spatial autocorrelation

JEL Classification

C31, C32

Comments

The authors thank Manfred Deistler, Justinas Pelenis, Benedikt Pötscher, Stefan Schneeberger, Martin Wagner, Derya Uysal, the participants of the CFE 2012 Conference, the 3rd Humboldt-Copenhagen Conference, as well as research seminar participants at the IHS, the European Business School and the University of Duisburg-Essen for interesting discussions and comments.

Contents

1 Introduction **1**

2 The Model **3**

3 Estimation Procedure and Large Sample Results **9**

4 Monte Carlo Simulations **14**

5 Empirical Illustration **20**

5.1 Data 24

5.2 Results 31

6 Conclusions **32**

A Proof of Theorem I **33**

B Instruments and the Rank Condition **45**

C Data **53**

References **56**

1 Introduction

Periods with a high number of defaults have shown that contagion can play a substantial role when pricing defaultable assets. The breakdowns of Lehman brothers and AIG are prominent examples for the effects arising with interlinked firms. Additionally, the European Central Bank reported a very high market concentration for the credit default swap (CDS) market, such that financial distress of one bank is expected to have impacts on the financial status of other banks (see ECB (2009)). Based on these observations, recent finance literature has drawn more attention to the correlation of credit risk and on credit risk contagion (see e.g. Tarashev and Zhu (2008)). One possibility to account for cross-sectional spillover effects in a statistical model is to include spatial lags following Cliff and Ord (1973). Additional complications arise due to the time series properties of the the economic variables of interest. Since credit default swap time series, used as a measure for credit risk, as well as some financial time series often used to predict or explain credit risk can be considered to be endogenous as well as integrated of order one, the empirical methodology used to investigate these data has to allow for possible regressor endogeneity as well as autocorrelation of the disturbances. In addition to this kind of endogeneity typically dealt with in panel cointegration models (see e.g. Mark et al. (2005)), the spatial lag results in further regressor endogeneity of a different type. To address these issues, this article considers a high dimensional cointegrating system including spatial lags.

Different approaches have emerged in the literature to estimate cointegrating relationships and to perform statistical inference. One possibility is to use a simple estimation routine, e.g. *ordinary least squares* (*OLS*) and then work out the (sometimes complicated) large sample distribution of the estimated parameters, e.g. Phillips and Hansen (1990), Phillips and Loretan (1991). Another opportunity is to adjust the estimation routine, such that the large sample distribution is either simpler or free of nuisance parameters. Examples along these lines are the *fully modified least squares* estimator (see e.g. Phillips and Hansen (1990), Phillips and Moon (1999), Pedroni (2000)), the *integrated modified least squares* estimator (see Vogelsang and Wagner (2011), where integrated modified least squares estimation is linked to fixed-b inference) and the *dynamic least squares* approach. Dynamic least squares estimation includes time-series leads and lags of the first differences of the regressors to control for the serial correlation and regressor endogeneity. This kind of estimator has been proposed by Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993). It has been applied to panel data e.g. in Kao and Chiang (2000), Mark and Sul (2003) and Mark et al. (2005).

Motivated by our application in empirical finance, we develop an econometric tool suitable for investigating situations where the long run outcome for a particular cross-section cannot be assumed to be independent of the outcomes of the other cross-sections and, at the same time, autocorrelation of the disturbances and regressor endogeneity are present. We do so in a context of a model that includes non-standard cointegrating relationships implied by the inclusion of peer or neighborhood effects, which are modeled as spatial lags. Since existing estimation procedures do not cope adequately with the endogeneity of the spatial lags, we propose to use a *dynamic two-stage least squares* ($D2SLS$) estimator, which combines *dynamic least squares* ($DOLS$) and *two stage least squares* ($2SLS$) estimation. In addition to the serial leads and lags used by $DOLS$, our estimation procedure uses cross-sectional (or spatial) lags of the regressors as instruments to control for the endogeneity of the spatial lags in the cointegrating vectors. We derive the large sample distribution of our estimator and show how to correctly conduct inference. We apply our methodology to our financial dataset, where we construct the economic space by using a "closeness" measure for firms based on input-output matrices. The weights matrix obtained from input-output data should approximate possible correlation patterns due to technology and demand shocks working their way through the economy. We find that our particular form of cross-sectional spillovers is substantial and highly significant.

In the rest of the paper we first describe our model and the formal assumptions in Section 2. Section 3 provides the $D2SLS$ estimation procedure and states our large sample results. We then investigate the small sample properties of the $D2SLS$ estimator in Section 4 and provide an illustrative application to modeling correlation of credit default swaps in Section 5. Finally, Section 6 offers conclusions.

2 The Model

Suppose that the data are generated from

$$y_{it} = \rho \sum_{j=1}^{n} W_{ij} y_{jt} + \beta' \mathbf{x}_{it} + \alpha_i + u_{it}^{\dagger} = \rho y_{it}^{*} + \beta' \mathbf{x}_{it} + \alpha_i + u_{it}^{\dagger} \ , \tag{1}$$

where y_{it} is the scalar response random variable and $\mathbf{x}_{it}$ is a $k \times 1$ vector of prediction random variables. Next, $t = 1, \ldots, T$ is the time index and $i = 1, \ldots, n$ is the cross-sectional index. We keep the cross-sectional dimension n fixed throughout the following analysis and take the limits for $T \to \infty$. The term $y_{it}^{*} = \sum_{j=1}^{n} W_{ij} y_{jt}$ is referred to as a *spatial lag* (see e.g. Cliff and Ord (1973), Kelejian and Prucha (1998), Kelejian and Prucha (1999) or Kapoor et al. (2007)) and represents the long-run impact of the neighboring observations on y_{it}. We collect the weights W_{ij} into an $n \times n$ spatial weights matrix $\mathbf{W}$.[1] We follow the spatial econometrics literature and maintain the following assumptions regarding the cross-sectional (or spatial) structure of the model:

Assumption 1. [Spatial Lag] *The spatial weights W_{ij} are non-stochastic and observable with $W_{ii} = 0$ and $\mathbf{W} \neq \mathbf{0}_{n \times n}$. The parameter ρ is such that largest absolute eigenvalue of $\rho \mathbf{W}$ is smaller than one.*

The restriction that $W_{ii} = 0$ is a normalization of the model, which requires that no observation is its own neighbor. The second part of the assumption guarantees that the matrix $(\mathbf{I}_n - \rho \mathbf{W})$ is invertible (see e.g. Corollary 5.6.16 in Horn and Johnson (1985)); $\mathbf{I}_n$ stands for the identity matrix of dimension n.[2] The invertibility of the matrix $\mathbf{I}_n - \rho \mathbf{W}$ is needed in order to provide a unique solution of the model and rule out multiple solutions for y_{it} that would be consistent with the explanatory variables and disturbances. The inverse $\mathbf{K} := (\mathbf{I}_n - \rho \mathbf{W})^{-1}$ is used in the consistency proof of the *D2SLS* estimator developed in this article.

The disturbance term is assumed to include an individual-specific effect α_i and an idiosyncratic com-

[1] Throughout the analysis we only consider one spatial lag term. However, the theory considered in this article can also be applied to a model where $y_{it} = \rho_1 \sum_{j=1}^{n} W_{1,ij} y_{jt} + \cdots + \rho_{k_\rho} \sum_{j=1}^{n} W_{k_\rho,ij} y_{jt} + \ldots$ in a straightforward way. The restriction that only one matrix $\mathbf{W}$ is included is used to keep the notation simple.

[2] The spectral radius is the lower bound for every induced matrix norm (cf. Theorem 5.6.9 in Horn and Johnson (1985)). Our assumption will, for example, be satisfied when the maximum absolute row or column sums of $\rho \mathbf{W}$ are less than one. Regarding notation $\mathbf{0}_{a \times b}$ stands for an $a \times b$ matrix of zeros, $\mathbf{0}_a$ is an a-dimensional column vector of zeros.

ponent $u_{it}^{\dagger}$ that is independent across i but possibly dependent across t. Analogically to Saikkonen (1991) the prediction random variable $\mathbf{x}_{it}$ is assumed to be integrated of order one, $I(1)$, and to be generated from

$$\Delta \mathbf{x}_{it} = \mathbf{v}_{it} \ . \tag{2}$$

In order to fully specify the model, we augment our set of assumptions by defining the process generating the disturbances:

Assumption 2. [Error Dynamics I; see Mark and Sul (2003), Mark et al. (2005), Phillips (2006)] *Let us define the stacked vector* $\mathbf{w}_{it}^{\dagger} = \left(u_{it}^{\dagger}, \mathbf{v}_{it}'\right)'$. *Then* $(\mathbf{w}_{it}^{\dagger})$ *has a moving average representation*

$$\mathbf{w}_{it}^{\dagger} = \mathbf{\Psi}_i^{\dagger}(L)\,\varepsilon_{it}^{\dagger} \ ,$$

where $\varepsilon_{it}^{\dagger}$ *is independent over both* i *and* t *with mean vector* $\mathbf{0}_{k+1}$, $k+1\times k+1$ *positive definite covariance matrix* $\mathbf{\Sigma}_{\varepsilon i}$ *and finite fourth moments.* $\mathbf{\Psi}_i^{\dagger}(L) = \sum_{j=0}^{\infty} \mathbf{\Psi}_{ij}^{\dagger} L^j$ *is a* $k+1 \times k+1$ *dimensional matrix polynomial in the lag operator* L, *with* $\mathbf{\Psi}_{i0}^{\dagger} = \mathbf{I}_{k+1}$ *and* $\sum_{j=0}^{\infty} j |\left[\mathbf{\Psi}_{ij}^{\dagger}\right]_{(m,n)}| < \infty$ *where* $\left[\mathbf{\Psi}_{ij}^{\dagger}\right]_{(m,n)}$ *is the* (m,n)*-th element of the matrix* $\mathbf{\Psi}_{ij}^{\dagger}$.

We shall denote the short-run $k+1 \times k+1$ covariance matrix of $\mathbf{w}_{it}^{\dagger}$ by $\mathbf{\Gamma}_{i0}^{\dagger}$, and the autocovariance matrices by $\mathbf{\Gamma}_{ij}^{\dagger}$, where

$$\mathbf{\Gamma}_{i0}^{\dagger} = \mathbb{E}\left(\mathbf{w}_{it}^{\dagger}\mathbf{w}_{it}^{\dagger\prime}\right) \text{ and } \mathbf{\Gamma}_{ij}^{\dagger} = \mathbb{E}\left(\mathbf{w}_{it}^{\dagger}\mathbf{w}_{i,t-j}^{\dagger\prime}\right) \ . \tag{3}$$

We will also use the following notation: $\Gamma_{uu,ij}^{\dagger}$ is the $(1,1)$ element of $\mathbf{\Gamma}_{ij}^{\dagger}$, $\mathbf{\Gamma}_{uv,ij}^{\dagger}$ corresponds to $\left[\mathbf{\Gamma}_{ij}^{\dagger}\right]_{(2:k+1,1)}$, $\mathbf{\Gamma}_{vu,ij}^{\dagger}$ corresponds to $\left[\mathbf{\Gamma}_{ij}^{\dagger}\right]_{(1,2:k+1)}$, while $\mathbf{\Gamma}_{vv,ij}$ corresponds to the $k \times k$ submatrix $\left[\mathbf{\Gamma}_{ij}^{\dagger}\right]_{(2:k+1,2:k+1)}$. The notation $(a:b, c:d)$ stands for "from row a to b and column c to d".

Let us define $\mathbf{w}_t^{\dagger} = \left(\mathbf{w}_{1t}^{\dagger\prime}, \ldots, \mathbf{w}_{nt}^{\dagger\prime}\right)'$, $\mathbf{u}_t^{\dagger} = \left(u_{1t}^{\dagger}, \ldots, u_{nt}^{\dagger}\right)'$ and $\mathbf{v}_t = (\mathbf{v}_{1t}', \ldots, \mathbf{v}_{nt}')'$. Then the $(k+1)\cdot n \times (k+1)\cdot n$ covariance matrices $\mathbf{\Gamma}_0^{\dagger} = \mathbb{E}\left(\mathbf{w}_t^{\dagger}\mathbf{w}_t^{\dagger\prime}\right)$ and $\mathbf{\Gamma}_j^{\dagger} = \mathbb{E}\left(\mathbf{w}_t^{\dagger}\mathbf{w}_{t-j}^{\dagger\prime}\right)$ are block diagonal with the blocks $\mathbf{\Gamma}_{i0}^{\dagger}$ and $\mathbf{\Gamma}_{ij}^{\dagger}$ along the main diagonal $(i = 1, \ldots, n)$. The $k+1 \times k+1$ long run covariance matrix

$\boldsymbol{\Omega}_i^\dagger$ of $\mathbf{w}_{it}^\dagger$ is given by

$$\begin{aligned}\boldsymbol{\Omega}_i^\dagger &= \sum_{j=-\infty}^{\infty} \mathbb{E}\left(\mathbf{w}_{it}^\dagger \mathbf{w}_{i,t+j}^{\dagger\prime}\right) = \boldsymbol{\Psi}_i^\dagger(1)\boldsymbol{\Sigma}_{\varepsilon i}\boldsymbol{\Psi}_i^\dagger(1)' = \boldsymbol{\Gamma}_{i0}^\dagger + \sum_{j=1}^{\infty}\left(\boldsymbol{\Gamma}_{ij}^\dagger + \boldsymbol{\Gamma}_{ij}^{\dagger\prime}\right) \quad (4)\\ &= \begin{pmatrix} \Omega_{uu,i}^\dagger & \boldsymbol{\Omega}_{vu,i}^\dagger \\ \boldsymbol{\Omega}_{uv,i}^\dagger & \boldsymbol{\Omega}_{vv,i} \end{pmatrix} = \begin{pmatrix} \Gamma_{uu,i0}^\dagger & \boldsymbol{\Gamma}_{vu,i0}^\dagger \\ \boldsymbol{\Gamma}_{uv,i0}^\dagger & \boldsymbol{\Gamma}_{vv,i0} \end{pmatrix} + \sum_{j=1}^{\infty}\left[\begin{pmatrix} \Gamma_{uu,ij}^\dagger & \boldsymbol{\Gamma}_{vu,ij}^\dagger \\ \boldsymbol{\Gamma}_{uv,ij}^\dagger & \boldsymbol{\Gamma}_{vv,ij} \end{pmatrix} + \begin{pmatrix} \Gamma_{uu,ij}^\dagger & \boldsymbol{\Gamma}_{vu,ij}^\dagger \\ \boldsymbol{\Gamma}_{uv,ij}^\dagger & \boldsymbol{\Gamma}_{vv,ij} \end{pmatrix}'\right].\end{aligned}$$

The long-run covariance matrix of $\mathbf{w}_t^\dagger$, denoted as $\boldsymbol{\Omega}^\dagger$, is then also block diagonal with the blocks $\boldsymbol{\Omega}_i^\dagger$ along the main diagonal. Analogically, the matrices $\boldsymbol{\Omega}_{uu}^\dagger$ and $\boldsymbol{\Omega}_{vv}$ contain the scalars $\Omega_{uu,i}^\dagger$ and the $k \times k$ blocks $\boldsymbol{\Omega}_{vv,i}$ along their main diagonal, where $i = 1, \ldots, n$.

Given the covariance structure, we want to exclude cointegration relationships between the terms of $\mathbf{x}_{it}$. In addition, we also want to guarantee that y_{it} is $I(1)$. Therefore we impose the following assumption:

Assumption 3. [Error Dynamics II; see Phillips (2006)]
$\boldsymbol{\Psi}_i^\dagger(1)$ *is non-singular and* $\boldsymbol{\Omega}_{vv,i}$ *has full rank* k*. Furthermore,* $\beta \neq \mathbf{0}_k$.

Note that by Assumption 3 and the independence across i assumption (i.e. Assumption 2), the rank of $\boldsymbol{\Omega}_{vv}$ is nk and $\mathbf{x}_{it}$ is a full rank integrated process. In addition, observe that if $\beta = \mathbf{0}_k$, the variable y_{it} becomes $I(0)$, see e.g. equation (1) and equation (14) below.

Assumption 2 implies that potentially all leads and lags of of $\Delta\mathbf{x}_{it}$ are correlated with $u_{it}^\dagger$. In the next step we follow *DOLS* literature and remove the serial correlation by projecting on the leads and lags of $\Delta\mathbf{x}_{it}$. For each sample size, *DOLS* estimation uses a finite number leads and lags, denoted by p in the following, to control for this correlation. Using such a truncation scheme will result in a specific truncation error e_{it}. However, under the conditions provided in Saikkonen (1991) this error will disappear asymptotically. In particular, the projection of $u_{it}^\dagger$ on the p leads and lags of $\Delta\mathbf{x}_{it}$ yields a truncation component $\sum_{s=-p}^{+p}\delta_{i,s}'\Delta\mathbf{x}_{i,t-s}$, a truncation error $\mathsf{e}_{it} = \sum_{s>p,s<-p}\delta_{i,s}'\Delta\mathbf{x}_{i,t-s}$ plus a *new* disturbance u_{it}, such that

$$u_{it}^\dagger = \sum_{s=-p}^{+p}\delta_{i,s}'\Delta\mathbf{x}_{i,t-s} + \sum_{s>p,s<-p}\delta_{i,s}'\Delta\mathbf{x}_{i,t-s} + u_{it} = \delta_i'\boldsymbol{\zeta}_{it} + \mathsf{e}_{it} + u_{it} = \delta_i'\boldsymbol{\zeta}_{it} + \underline{u}_{it}\ . \quad (5)$$

$\Delta\mathbf{x}_{i,t-s}$ and $\delta_{i,s}$ are vectors of dimension $k \times 1$, while the $(2p+1)k \times 1$ dimensional vectors of projection variables and projection coefficients are given by

$$\boldsymbol{\zeta}_{it} = \left(\Delta\mathbf{x}'_{i,t-p}, \ldots, \Delta\mathbf{x}'_{i,t}, \ldots, \Delta\mathbf{x}'_{i,t+p}\right)' = \left(\mathbf{v}'_{i,t-p}, \ldots, \mathbf{v}'_{i,t}, \ldots, \mathbf{v}'_{i,t+p}\right)' \text{ and } \delta_i = \left(\delta'_{i,-p}, ..., \delta'_{i,+p}\right)' . \quad (6)$$

$\boldsymbol{\zeta}_{it}$ is by construction orthogonal to the noise term u_{it}. The term $\underline{u}_{it} = \mathsf{e}_{it} + u_{it}$ can still be correlated with $\Delta\mathbf{x}_{it}$ for some $p < \infty$. Now we impose an additional restriction on the error dynamics that will guarantee that the truncation error e_{it} converges to zero:

Assumption 4. [Error Dynamics III; see Saikkonen (1991), Mark et al. (2005)]
Suppose that $p = p(T)$. Then $p(T)$ has to fulfill $\frac{p(T)^3}{T} \to 0$ and $\sqrt{T}\sum_{|s|>p(T)} \|\delta_{i,s}\|_2 \to 0$ as $T \to \infty$, where $\|.\|_2$ stands for the Euclidian norm.

Assumption 4 requires that $p(T)$ does not grow too fast, while the second part restricts the dependence between the noise term and the regressors. Based on Assumptions 2 to 4 and equation (5), if T becomes large then – due to the increase in the number of leads and lags $p(T)$ – the truncation error e_{it} becomes small. As a result, the difference between $\underline{u}_{it}$ and u_{it} becomes small and $\underline{u}_{it}$ becomes orthogonal to $\boldsymbol{\zeta}_{it}$ as $T \to \infty$.[3] Hence we arrive at the *new* covariance stationary process $\mathbf{w}_{it} = (u_{it}, \mathbf{v}'_{it})' = \boldsymbol{\Psi}_i(L)\,\varepsilon_{it}$ which has mean zero, covariance matrix $\boldsymbol{\Gamma}_{i0}$ and autocovariance $\boldsymbol{\Gamma}_{ij}$. These matrices have the structure

$$\boldsymbol{\Gamma}_{ij} = \mathbb{E}\left(\mathbf{w}_{it}\mathbf{w}'_{i,t-j}\right) = \begin{pmatrix} \Gamma_{uu,ij} & \mathbf{0}_{1\times k} \\ \mathbf{0}_k & \boldsymbol{\Gamma}_{vv,ij} \end{pmatrix}, \quad (7)$$

where $\Gamma_{uu,ij} = \mathbb{E}\left(u_{it}u_{i,t-j}\right)$, $\boldsymbol{\Gamma}_{vv,ij} = \mathbb{E}\left(\mathbf{v}_{it}\mathbf{v}'_{i,t-j}\right)$ and $j \in \mathbb{Z}$.

In addition, our model includes a full set of individual specific effects and hence a set of individual dummies α_i has been included to the regression (1) (*fixed effects specification*). In order to simplify the algebra, we shall use the within transformation and derive the asymptotic distribution of the estimates of the slope coefficients ρ and β using within-transformed data. In a linear regression, these estimated slope coefficients are algebraically equivalent to the *least squares dummy variable estimates* (see e.g. Baltagi

[3]For a short discussion on the truncation error and the Assumption 4 we refer the reader to Saikkonen (1991) and to Lütkepohl (2006)[Remark 1, p. 533]. For more technical details see Saikkonen (1991)[Theorem 4.1/Lemma A.5].

(2008)[p. 11]). Here it is important to note that while the time index t goes from 1 to T in (1), after the projection facility is applied only the observations $p+1,\ldots,T-p$ can be used. We still use t as the index for the time period which now runs from 1 to $T_\star$, where $T_\star = T-2p$. The variables in deviations from their individual means are

$$\begin{aligned}
\widetilde{y}_{it} &= y_{it} - \frac{1}{T_\star}\sum_{t=1}^{T_\star} y_{it}, \quad \widetilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \frac{1}{T_\star}\sum_{t=1}^{T_\star} \mathbf{x}_{it}, \quad \widetilde{y}^*_{it} = \sum_{j=1}^{n} W_{ij}\widetilde{y}_{it} \ , \\
\widetilde{\boldsymbol{\zeta}}_{it} &= \boldsymbol{\zeta}_{it} - \frac{1}{T_\star}\sum_{t=1}^{T_\star} \boldsymbol{\zeta}_{it} \ ,
\end{aligned} \tag{8}$$

such that (1) after applying the within transform and the projection facility reads as follows:

$$\begin{aligned}
\widetilde{y}_{it} &= \rho\sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt} + \beta'\widetilde{\mathbf{x}}_{it} + \widetilde{u}^\dagger_{it} = \rho\tilde{\mathbf{y}}^*_{it} + \beta'\widetilde{\mathbf{x}}_{it} + \widetilde{u}^\dagger_{it} \\
&= \rho\sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt} + \beta'\widetilde{\mathbf{x}}_{it} + \delta_i'\tilde{\boldsymbol{\zeta}}_{it} + \widetilde{\underline{u}}_{it} = \rho\tilde{\mathbf{y}}^*_{it} + \beta'\widetilde{\mathbf{x}}_{it} + \delta_i'\tilde{\boldsymbol{\zeta}}_{it} + \widetilde{\underline{u}}_{it} \ .
\end{aligned} \tag{9}$$

Given the assumptions on the error dynamics, the *functional central theorem* (see e.g. Karatzas and Shreve (1991)[Chapter 4] or Davidson (1994)[Chapters 27-30]) can be applied. If $T_\star \to \infty$ then

$$\frac{1}{\sqrt{T_\star}}\sum_{t=1}^{[T_\star r]} \mathbf{w}_{it} \xrightarrow{d} \mathcal{B}_i(r) = \mathbf{\Omega}_i^{1/2}\mathcal{W}_i(r) \ , \tag{10}$$

where $r \in [0,1]$ and $\xrightarrow{d}$ stands for *weak convergence / convergence in distribution.* $\mathcal{B}_i(r) = (\mathcal{B}_{ui}(r), \mathcal{B}_{vi}(r)')'$, where $\mathcal{B}_{ui}$ and $\mathcal{B}_{vi}$ are *independent* Brownian motions, in $\mathbb{R}$ and $\mathbb{R}^k$, respectively. While $\mathcal{B}_i$ stands for a Brownian motion with covariance matrix $\mathbf{\Omega}_i$, $\mathcal{W}_i$ stands for a standard Brownian motion, where $\mathcal{W}_i(r) = (\mathcal{W}_{ui}(r), \mathcal{W}_{vi}(r)')'$. $[T_\star r]$ denotes the integer part of $T_\star r$.[4] $\mathbf{\Omega}_i$ is the $k+1 \times k+1$ long-run variance-covariance

[4]In some of the following expressions we omit the borders of integration as well as the continuous time index r of the Brownian motion, i.e. we write $\int \mathcal{W}$ instead of $\int_0^1 \mathcal{W}(r)dr$, while $\int_0^1 \mathcal{W}(r)d\mathcal{W}(r)$ is abbreviated by $\int \mathcal{W}d\mathcal{W}$.

matrix of $\mathbf{w}_{it}$. Due to the independence of $\mathcal{B}_{ui}(t)$ and $\mathcal{B}_{vi}(t)$, this matrix is of the structure

$$\mathbf{\Omega}_i = \begin{pmatrix} \Omega_{uu,i} & \mathbf{0}_{1\times k} \\ \mathbf{0}_k & \mathbf{\Omega}_{vv,i} \end{pmatrix} = \mathbf{\Gamma}_{i0} + \sum_{j=1}^{\infty} \left(\mathbf{\Gamma}_{ij} + \mathbf{\Gamma}'_{ij}\right) , \tag{11}$$

where the matrices $\mathbf{\Gamma}_{ij}$ are given by (7). For the demeaned term $\tilde{v}_{it} = v_{it} - \frac{1}{T_\star}\sum_{t=1}^{T_\star} v_{it}$ we get $\frac{1}{\sqrt{T_\star}}\sum_{t=1}^{[T_\star r]} \tilde{v}_{it} = \frac{1}{\sqrt{T_\star}}\sum_{t=1}^{[T_\star r]} \left(v_{it} - \frac{1}{T_\star}\sum_{t=1}^{T_\star} v_{it}\right) \xrightarrow{d} \mathcal{B}_{vi}(r) - r\mathcal{B}_{vi}(1)$. $\mathcal{B}_{vi}(r) - r\mathcal{B}_{vi}(1)$ is a *Brownian bridge.* Since $\mathbf{x}_{it}$ is an $I(1)$ process, $\mathbf{x}_{it}$ arises from a partial sum process. Then $\tilde{\mathbf{x}}_{it} = \sum_{\iota=1}^{t} v_{i\iota} - \frac{1}{T_\star}\sum_{t=1}^{T_\star}\sum_{\iota=1}^{t} v_{i\iota}$. By the continuous mapping theorem (see Klenke (2008)[p. 257], Davidson (1994)[Theorem 26.13 & 30.2]) the $T_\star \to \infty$ limit is given by the *demeaned* Brownian motion

$$\frac{1}{\sqrt{T_\star}}\tilde{\mathbf{x}}_{it} \xrightarrow{d} \mathcal{B}_{vi}(r) - \int_0^1 \mathcal{B}_{vi}(s)ds \ .$$

$\mathcal{B}_{vi}(r) - \int_0^1 \mathcal{B}_{vi}(s)ds$ will be abbreviated by $\tilde{\mathcal{B}}_{vi}(r)$. Davidson (1994)[Theorem 30.2] shows that $\frac{1}{T_\star^2}\sum_{t=1}^{T_\star} \tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}'_{it}$ converges in distribution to $\int_0^1 \mathcal{B}_{vi}(s)\mathcal{B}'_{vi}(s)ds$. Last but not least, Davidson (1994)[Theorem 30.13] and some algebra results in $\frac{1}{T_\star}\sum_{t=1}^{T_\star} \tilde{\mathbf{x}}_{it}\tilde{u}_{it} \xrightarrow{d} \sqrt{\mathbf{\Omega}_{uu,i}}\int_0^1 \tilde{\mathcal{B}}_{vi}(r)d\mathcal{W}_{ui}(r)$.

Before we proceed with the estimation part we would like to discuss our model for the $n = 2$ case. Here we observe that the cointegration equations are non-linear and due to the spatial lag component an additional source of endogeneity arises.

Remark 1. Consider (1) for the two-dimensional case, i.e. $n = 2$. Due to Assumption 1 the matrix $\mathbf{I}_2 - \rho\mathbf{W}$ has to be invertible, such that

$$\left[\mathbf{I}_2 - \rho\begin{pmatrix} 0 & W_{12} \\ W_{21} & 0 \end{pmatrix}\right]^{-1} = \frac{1}{1+\rho^2 W_{12}W_{21}} \cdot \begin{pmatrix} 1 & -\rho W_{21} \\ -\rho W_{12} & 1 \end{pmatrix} . \tag{12}$$

Combining (1) and (12) now results in

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \frac{1}{1+\rho^2 W_{12}W_{21}} \cdot \begin{pmatrix} \beta'\mathbf{x}_{1t} & -\rho W_{21}\beta'\mathbf{x}_{2t} & +u^\dagger_{1t} & -\rho W_{21}u^\dagger_{2t} & +\alpha_1 - \rho W_{21}\alpha_2 \\ -\rho W_{12}\beta'\mathbf{x}_{1t} & +\beta'\mathbf{x}_{2t} & +u^\dagger_{2t} & -\rho W_{12}u^\dagger_{1t} & +\alpha_2 - \rho W_{12}\alpha_1 \end{pmatrix} . \tag{13}$$

Equation (13) shows the $n = 2$ coinintegrating equations. The cointegrating relationships do not have the usual linear form in the sense that the solution for y_{it} is a nonlinear function of the parameter ρ.

Assumption 1 guarantees that $\mathbf{I}_n - \rho\mathbf{W}$ has the full rank n. Together with Assumptions 2-4 we observe that for an arbitrary but fixed $n \in \mathbb{N}$ the following equation (14) constitute n cointegrating relationships:

$$\begin{pmatrix} y_{1t} \\ \vdots \\ y_{nt} \end{pmatrix} = (\mathbf{I}_n - \rho\mathbf{W})^{-1} \left(\begin{pmatrix} \beta'\mathbf{x}_{1t} \\ \vdots \\ \beta'\mathbf{x}_{nt} \end{pmatrix} + \begin{pmatrix} \alpha_1 + u_{1t}^{\dagger} \\ \vdots \\ \alpha_n + u_{nt}^{\dagger} \end{pmatrix} \right) . \tag{14}$$

Summing up, when we consider the data generated by (1) we observe that: (i) $\mathbf{x}_{it}$ and $u_{it}^{\dagger}$ are correlated by the assumptions on $\Psi_i^{\dagger}$ and $\mathbf{\Sigma}_{\varepsilon i}$. (ii) For $\rho \neq 0$, y_{jt} depends on y_{it} and vice versa. (iii) $u_{it}^{\dagger}$ and $u_{jt}^{\dagger}$ are independent by Assumption 2. (iv) Since y_{jt} depends on y_{it} we know that $\rho W_{ij} y_{jt}$ and $u_{it}^{\dagger}$ have to be correlated (also for the within transformed data the same correlation structure is observed). Therefore the standard *DOLS* method is not sufficient to remove all the correlation between the regressors and the noise.

In the following section we shall construct an estimator where we account for "serial" endogeneity by means of the *DOLS* projection facility. In addition endogeneity enters via the spatial correlation modeled by $\rho\mathbf{W}$. To account for this kind of "spatial" endogeneity we follow the $2SLS$ approach. Combining these concepts will provide us with an estimator which accounts for both sources of endogeneity.

3 Estimation Procedure and Large Sample Results

The goal of the following analysis is to construct the $D2SLS$ estimator and to show that it leads to consistent estimates of the parameters ρ and β. We then provide the large sample distribution of the $D2SLS$ estimator. The parameters δ will be shown to be *nuisance parameters*. In order to write down our estimator in a compact way, we first define the model in a stacked notation. For notational simplicity we drop the tilde notation in the stacked model and define

$$
\begin{aligned}
\mathbf{y} &= (\widetilde{y}_{11},\ldots,\widetilde{y}_{1T_\star},\ldots,\widetilde{y}_{n1},\ldots,\widetilde{y}_{nT_\star})',\\
\mathbf{y}^* &= \left(\widetilde{y}^*_{11},\ldots,\widetilde{y}^*_{1T_\star},\ldots,\widetilde{y}^*_{n1},\ldots,\widetilde{y}^*_{nT_\star}\right)',\\
\mathbf{x} &= \left(\widetilde{\mathbf{x}}'_{11},\ldots,\widetilde{\mathbf{x}}'_{1T_\star},\ldots,\widetilde{\mathbf{x}}'_{n1},\ldots,\widetilde{\mathbf{x}}'_{nT_\star}\right)',\\
\underline{\mathbf{u}} &= \left(\widetilde{\underline{u}}_{11},\ldots,\widetilde{\underline{u}}_{1T_\star},\ldots,\widetilde{\underline{u}}_{n1},\ldots,\widetilde{\underline{u}}_{nT_\star}\right)' \ , \qquad (15)
\end{aligned}
$$

where $\mathbf{y}$, $\mathbf{y}^*$ and $\underline{\mathbf{u}}$ are of dimension $nT_\star \times 1$, while $\mathbf{x}$ is an $nT_\star \times k$ matrix. Furthermore, we have

$$
\boldsymbol{\zeta}\delta = \begin{pmatrix} \delta'_{11}\widetilde{\boldsymbol{\zeta}}_{11} \\ \vdots \\ \delta'_{nT_\star}\widetilde{\boldsymbol{\zeta}}_{nT_\star} \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{\zeta}}'_{11} & \mathbf{0}_{1\times(2p+1)k} & \mathbf{0}_{1\times(2p+1)k} \\ \vdots & & \\ \widetilde{\boldsymbol{\zeta}}'_{1T_\star} & \mathbf{0}_{1\times(2p+1)k} & \mathbf{0}_{1\times(2p+1)k} \\ \mathbf{0}_{1\times(2p+1)k} & \widetilde{\boldsymbol{\zeta}}_{21} & \mathbf{0}_{1\times(2p+1)k} \\ & & \ddots \\ \mathbf{0}_{1\times(2p+1)k} & \mathbf{0}_{1\times(2p+1)k} & \widetilde{\boldsymbol{\zeta}}_{nT_\star} \end{pmatrix} \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix} . \qquad (16)
$$

$\boldsymbol{\zeta}$ is a $nT_\star \times (2p+1)k \cdot n$ matrix, while (given δ_i of dimension $(2p+1)k$) δ is of dimension $(2p+1)k \cdot n \times 1$. This provides us with model (9) in stacked form

$$
\mathbf{y} = \rho\mathbf{y}^* + \mathbf{x}\beta + \boldsymbol{\zeta}\delta + \underline{\mathbf{u}} = (\mathbf{y}^*, \mathbf{x})\,\gamma + \boldsymbol{\zeta}\delta + \underline{\mathbf{u}} = \mathbf{X}\left(\gamma', \delta'\right)' + \underline{\mathbf{u}} \ , \qquad (17)
$$

where $\gamma = (\rho, \beta')'$. The right hand side variables are collected in $\mathbf{X} = (\mathbf{y}^*, \mathbf{x}, \boldsymbol{\zeta})$.

We shall estimate the model by using instruments for the endogenous variable $\widetilde{y}^*_{it} = \sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt}$. Here, we could proceed in an abstract way by assuming that q_ρ instruments are available to fulfill the properties necessary for instrumental variable estimation (see e.g. Kitamura and Phillips (1997)). In contrast to this high level assumption, we follow Kelejian and Prucha (1998) and base the instruments on the spatial lags of the explanatory variables. Our model can be solved as

$$
\mathbf{y} = \left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho\mathbf{W})^{-1}\right](\mathbf{x}\beta + \boldsymbol{\zeta}\delta + \underline{\mathbf{u}}) \ . \qquad (18)
$$

The matrix $(\mathbf{I}_n - \rho\mathbf{W})^{-1}$ can then be expanded as (see e.g. Corollary 5.6.16 in Horn and Johnson (1985)):

$$(\mathbf{I}_n - \rho\mathbf{W})^{-1} = \sum_{s=0}^{\infty} (\rho\mathbf{W})^s \ . \tag{19}$$

This implies that variables of the form $\sum_{j=1}^n W_{ij}\widetilde{x}_{jtv}, \sum_{j=1}^n W_{ij}^2\widetilde{x}_{jtv}, \ldots$ are suitable instruments for $\mathbf{Wy}$. $\tilde{x}_{jtv}$ is the element v of $\widetilde{\mathbf{x}}_{jt}$. Note that these instruments have an intuitive interpretation: we instrument the W_{ij} weighted sum of the neighbors/peers $\widetilde{y}_{jt}$ by the W_{ij} weighted sum of the characteristics of the neighbors (their $\widetilde{\mathbf{x}}_{it}$ values). The higher order spatial lags as instruments then use the characteristics of the neighbors of the neighbors, etc. Hence we assume that the following set of instruments is used:

Assumption 5. [Valid Instruments; see Kitamura and Phillips (1997)] *The instruments are* $\widetilde{x}^*_{itv} = \sum_{j=1}^n W_{ij}^{\tau_v}\widetilde{x}_{jtv}$, *where* $v = 1, \ldots, q_\rho$ *and* $\tau_v \in \mathbb{N}$. $\widetilde{\mathbf{x}}^*_{it} = \left(\widetilde{x}^*_{it1}, \ldots, \widetilde{x}^*_{itq_\rho}\right)'$ *is a vector of dimension* q_ρ. *We assume that these instruments fulfill the requirements for instrumental variable estimation as stated e.g. in Ruud (2000)[Chapter 20], Phillips and Hansen (1990) and Kitamura and Phillips (1997). I.e. (i) the number of instruments is larger or equal to the number of parameters (*order condition*), (ii) the* $T_\star$*-limit of* $\frac{1}{T_\star^2}\sum_{t=1}^{T_*}(y^*_{it}, \tilde{\mathbf{x}}'_{it})'((\widetilde{\mathbf{x}}^*_{it})', \tilde{\mathbf{x}}'_{it})$ *is of rank* $k+1$ *(almost surely) and (iii) the* $T_\star$*-limit of* $\frac{1}{T_\star^2}\sum_{t=1}^{T_\star}((\widetilde{\mathbf{x}}^*_{it})', \tilde{\mathbf{x}}'_{it})'((\widetilde{\mathbf{x}}^*_{it})', \tilde{\mathbf{x}}'_{it})$ *is of rank* $k + q_\rho$ *(almost surely).*

Appendix B shows that with $\tau_v = 1$ and some regularity conditions on $\mathbf{W}$ the *rank conditions* (ii) and (iii) are satisfied. To keep the notation simple, we consider - as already stated at the beginning of Section 2 - a model with one spatial lag ($k_\rho = 1$). With $q_\rho = 1$ we are in the just identified case, while if $q_\rho > 1$ we consider the over-identified case. We collect our instruments in the $nT_\star \times q_\rho$ matrix

$$\mathbf{x}^* = \left(\widetilde{\mathbf{x}}^{*\prime}_{11}, \ldots, \widetilde{\mathbf{x}}^{*\prime}_{1T_\star}, \ldots, \widetilde{\mathbf{x}}^{*\prime}_{n1}, \ldots, \widetilde{\mathbf{x}}^{*\prime}_{nT_\star}\right)' \ . \tag{20}$$

The set of our instruments is then $\mathbf{Z} = (\mathbf{x}^*, \mathbf{x}, \boldsymbol{\zeta})$. While the matrix of explanatory variables $\mathbf{X}$ is of dimension $T_\star n \times 1 + k + (2p+1)k \cdot n$, the dimension of $\mathbf{Z}$ is $T_\star n \times q_\rho + k + (2p+1)k \cdot n$.[5] Before we present our estimator, let us discuss why e.g. DOLS and *two-stage least squares* ($2SLS$) do not provide us with

[5] A variant of our model is $\mathbf{\Psi}_i^\dagger(L) = \mathbf{\Psi}^\dagger(L)$ for $i = 1, \ldots, n$. Then $\mathbf{X}$ is of dimension $T_\star n \times 1 + k + (2p+1)k$ while $\mathbf{Z}$ is of dimension $T_\star n \times q_\rho + k + (2p+1)k$.

consistent estimators:

Remark 2. [Endogeneity] Let us consider (17). From the discussion in the last paragraph of Remark 1 we already know that $\tilde{u}^{\dagger}_{it}$ is correlated with $\tilde{y}^{*}_{it}$ and with $\tilde{\mathbf{x}}_{it}$. Given the $k+1+(2p+1)k \cdot n$ dimensional vector of regressors $\mathbf{X}_{it} = \left(\tilde{y}^{*}_{it}, \tilde{\mathbf{x}}'_{it}, \mathbf{0}_{1\times(2p+1)k\cdot(i-1)}, \tilde{\delta}'_{it}, \mathbf{0}_{1\times(2p+1)k\cdot(n-i-1)}\right)'$ and the $k+q_{\rho}+(2p+1)k \cdot n$ dimensional vector of instruments $\mathbf{Z}_{it} = \left(\tilde{\mathbf{x}}^{*\prime}_{it}, \tilde{\mathbf{x}}'_{it}, \mathbf{0}_{1\times(2p+1)k\cdot(i-1)}, \tilde{\delta}'_{it}, \mathbf{0}_{1\times(2p+1)k\cdot(n-i-1)}\right)'$ we observe that $\widetilde{u}_{it}$ is still correlated with the first component of $\mathbf{X}_{it}$. Therefore $DOLS$ does not result in consistent estimates. When applying $2SLS$ we get $\mathbf{X}^{2SLS}_{it} = (\tilde{y}^{*}_{it}, \tilde{\mathbf{x}}'_{it})'$ and $\mathbf{Z}^{2SLS}_{it} = (\tilde{\mathbf{x}}^{*\prime}_{it}, \tilde{\mathbf{x}}'_{it})'$. Since no projection facility is used with $2SLS$, the residual term is given by $\widetilde{u}^{\dagger}_{it}$. Therefore, the term $\tilde{\mathbf{x}}_{it}$ contained in $\mathbf{Z}^{2SLS}_{it}$, is still correlated with $\widetilde{u}^{\dagger}_{it}$ and $2SLS$ is not consistent.

In analogy to a standard regression setting with endogenous regressors, we now construct a two stage-least square procedure for our panel setting where leads and lags of $\Delta\tilde{\mathbf{x}}_{it}$ are included. Let us define the project operator $\mathbf{P}_H$ projecting on the space spanned by $\mathbf{Z}$ (see e.g. Ruud (2000)[Chapter 3]). In formal terms

$$\mathbf{P}_H := \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}' \ . \tag{21}$$

Since $\mathbf{Z}$ is a $T_{\star}n \times q_{\rho}+k+(2p+1)k \cdot n$ matrix, $\mathbf{P}_H$ has to be a $T_{\star}n \times T_{\star}n$ matrix. With two-stage least squares the initial stage results in the projected values

$$\widehat{\mathbf{y}^{*}} = \mathbf{P}_H\mathbf{y}^{*} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}\mathbf{y}^{*} \ , \tag{22}$$

while $\mathbf{P}_H\mathbf{x} = \mathbf{x}$ and $\mathbf{P}_H\boldsymbol{\zeta} = \boldsymbol{\zeta}$. The second stage estimator is

$$\begin{pmatrix} \widehat{\rho} \\ \widehat{\beta} \\ \widehat{\delta} \end{pmatrix}_{D2SLS} = \begin{pmatrix} \widehat{\mathbf{y}^{*}}'\widehat{\mathbf{y}^{*}} & \mathbf{x}'\widehat{\mathbf{y}^{*}} & \boldsymbol{\zeta}'\widehat{\mathbf{y}^{*}} \\ \widehat{\mathbf{y}^{*}}'\mathbf{x} & \mathbf{x}'\mathbf{x} & \boldsymbol{\zeta}'\mathbf{x} \\ \widehat{\mathbf{y}^{*}}'\boldsymbol{\zeta} & \mathbf{x}'\boldsymbol{\zeta} & \boldsymbol{\zeta}'\boldsymbol{\zeta} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\mathbf{y}^{*}}' \\ \mathbf{x}' \\ \boldsymbol{\zeta}' \end{pmatrix} \mathbf{y}. \tag{23}$$

In the first stage we project the endogenous variable $\widetilde{y}^{*}_{it}$ on $\mathbf{Z}$. In contrast to usual two stage least squares estimates, the projected values $\widehat{\widetilde{y}^{*}_{it}}$ are still correlated with $\underline{\widetilde{u}}_{jt}$ and $\widetilde{u}_{jt}$. To see this, we consider $\widetilde{y}^{*}_{it} = \sum_{j=1} W_{ij}\widetilde{y}_{jt} = \sum_{j=1} W_{ij}\sum_{l=1}^{n} K_{jl}\left(\beta'\tilde{\mathbf{x}}_{lt} + \tilde{u}^{\dagger}_{lt}\right) = \sum_{j=1} W_{ij}\sum_{l=1}^{n} K_{jl}\left(\beta'\tilde{\mathbf{x}}_{lt} + \delta'_{l}\boldsymbol{\zeta}_{lt} + \underline{\tilde{u}}_{lt}\right)$; K_{jl} is the (j,l) element of the matrix $\mathbf{K} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}$. Since in general $K_{jl} \neq 0$, this also holds for $l = i$ such that the projected values $\widehat{\widetilde{y}^*_{it}}$ can still be correlated with the noise. Next we observe that by the construction of $\mathbf{Z}$, for each component i only the *own* leads and lags are considered, i.e. only $\Delta\tilde{\mathbf{x}}_{it\pm p}$ are included in $\mathbf{Z}_{it}$. From the above calculations it follows that $\widetilde{y}^*_{it} = \sum_{j=1} W_{ij}\widetilde{y}_{jt} = \ldots$ includes the terms $\tilde{\mathbf{x}}_{lt}$ and $\tilde{u}^\dagger_{lt}$, which are correlated as well by the model assumptions. Therefore, a priori it need not be clear whether we obtain a Gaussian mixture distribution when $T_\star \to \infty$. For example, one could potentially include *all* the leads and lags $\Delta\widetilde{\mathbf{x}}_{lt\pm p}$, $l = 1,\ldots,n$, to get rid of this type of correlation. This would increase the number of nuisance parameters enormously (the dimension of $\mathbf{Z}$ would increase from $Tn \times q_\rho + k + (2p+1)k \cdot n$ to $Tn \times q_\rho + k + (2p+1)k \cdot n^2$). However, in the proof of Theorem 1 we shall observe that due to the fact that "$\mathbf{Z}_{it,1:q_\rho+k}$ over $T_\star$" and "$\mathbf{X}_{it,1:1+k}$ over $T_\star$" are considered, this type of correlation becomes neglectable when taking limits. Therefore, we still attain a normal mixture distribution.[6] Based on this discussion, we can now compactly write the *dynamic two-stage least squares* estimator of $(\rho,\beta',\delta')' = (\gamma',\delta')'$ as

$$\begin{aligned}\widehat{(\gamma',\delta')}'_{D2SLS} &= \left(\mathbf{X}'\mathbf{P}_H\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_H\mathbf{y} \\ &= (\gamma',\delta')' + \left(\mathbf{X}'\mathbf{P}_H\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_H\underline{\mathbf{u}}\,. \end{aligned} \tag{24}$$

With $q_\rho = 1$ we are in the just identified case, where the estimator is given by

$$\widehat{(\gamma',\delta')}'_{D2SLS} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'\mathbf{y} = (\gamma',\delta')' + \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'\underline{\mathbf{u}}\,. \tag{25}$$

Given the definition of the *D2SLS* estimator in (24), we summarize its large sample properties in the following result:

[6] In Appendix A we shall observe that in $\mathbf{M}_{nTi} = \frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\mathbf{X}_{it,1:1+k}\mathbf{Z}'_{it,1:q_\rho+k}$ the impact of the terms including $\widetilde{u}^\dagger_{lt}$ goes to zero due to normalization with $1/T_\star^2$ for $T_\star \to \infty$. In addition for the "last term" in (24) we get $\frac{1}{T_\star}\sum_{t=1}^{T_\star}\mathbf{Z}_{it,1:q_\rho+k}\underline{\tilde{u}_{it}}$. Since $\tilde{x}^*_{it} = \sum_{l=1}^{n} W_{il}\tilde{x}_{lt\iota}$, with $W_{ii} = 0$ by Assumption 1, is independent of $\underline{\tilde{u}_{it}}$ by the model Assumption 2, no further correlation terms arise when taking the limit. This results in the term $\mathbf{m}_{niZu}$ presented in (57).

If, however, $\underline{\tilde{u}_{it}}$ and the limit $\tilde{u}_{it}$ were correlated with $\tilde{\mathbf{x}}_{lt}$, for $i \neq l$, a projection in the own leads and lags would not be sufficient. To see this, the first q_ρ elements of $\frac{1}{T_\star}\sum_{t=1}^{T_\star}\mathbf{Z}_{it,1:q_\rho+k}\underline{\tilde{u}_{it}}$ are given by $\frac{1}{T_\star}\sum_{t=1}^{T_\star}\sum_{l=1}^{n} W_{il}\tilde{x}_{lt\iota}\underline{\tilde{u}_{it}}$, where $\iota = 1,\ldots,q_\rho$. By Davidson (1994)[Theorem 30.13], $\frac{1}{T_\star}\sum_{t=1}^{T_\star}\tilde{\mathbf{x}}_{lt\iota}\underline{\tilde{u}_{it}} \xrightarrow{d} \sqrt{\Omega_{uu,i}}\int\tilde{\mathcal{B}}_{vl}d\mathcal{W}_{ui} + \underline{\Delta}_{vu,li,\iota}$. The 1×1 correlation term $\underline{\Delta}_{vu,li,\iota}$ is given by $\mathbb{E}\left(\Delta\tilde{\mathbf{x}}_{lt\iota}\tilde{u}_{it}\right) + \sum_{j=1}^{\infty}\mathbb{E}\left(\Delta\tilde{\mathbf{x}}_{lt\iota}\tilde{u}_{it-j}\right)$. If $\tilde{u}_{it}$ and $\tilde{\mathbf{x}}_{lt\iota}$ are correlated, then $\underline{\Delta}_{vu,li,\iota} \neq 0$. This was excluded by Assumption 2 in our analysis.

Theorem 1 (Limits for $D2SLS$ Estimation)**.** *Consider the fixed effects spatial correlation model (1) and the estimator (24) based on the within-transformed model (9). Suppose that the Assumptions 1 to 5 hold.* $T_\star = T - 2p(T)$. *Then for* n *fixed and* $T \to \infty$ *it follows that*

1. $T_\star(\hat{\gamma}_{D2SLS} - \gamma)$ *and* $\sqrt{T_\star}(\hat{\delta}_{D2SLS,i} - \delta_i)$ *are asymptotically independent for each* $i = 1, \ldots, n$.

2. $\sqrt{n}T_\star\left(\hat{\gamma}_{D2SLS} - \gamma\right)$ *converges in distribution to* $\mathbf{M}_n^{-1}\mathbf{m}_n$, *where* $\mathbf{m}_n$ *and* $\mathbf{M}_n$ *are given by (59) and (60).*

3. *Given a* $s \times k + 1$ *restriction matrix* $\mathbf{R}$, *the Wald statistic* $S_{\gamma,nT}$ *defined in (67) converges to a* χ^2 *random variable with* s *degrees of freedom.*

Remark 3. By Assumption 4, if $T \to \infty$, then $T_\star \to \infty$. In Remark 2 we already observe that the two-stage least squares estimator and the $DOLS$ estimator are special cases of the dynamic two-stage least squares estimator. Hence, the Wald-statistic presented in Appendix A can be used to obtain the Wald statistic for the two-stage least squares estimator and the $DOLS$ estimator.

4 Monte Carlo Simulations

This section investigates the small sample properties of the $D2SLS$ estimator as well as the size and power of the Wald tests defined in Theorem 1. We generate the data based on an error process that follows from Assumptions 2-4. To operationalize this we need to specify the lag polynomials $\mathbf{\Psi}_i^\dagger(L)$. In particular, we have to specify the error dynamics of the vector $\mathbf{w}_{it}^\dagger$. Here we assume the same error dynamics for all cross sections $i = 1, \ldots, n$. We use two explanatory variables $\mathbf{x}_{it}$ such that $k = 2$ and set $\beta = (1, 1)'$. The number of instruments is $q_\rho = 2$.

Regarding the error dynamics we use the stationary designs of Binder et al. (2005) to generate the data for the vector $\mathbf{w}_{it}^\dagger$. The innovations $\varepsilon_{it}^\dagger$ are generated as independent draws from $\varepsilon_{it}^\dagger \sim N\left(0, \mathbf{\Sigma}_{\varepsilon i}\right)$. For $\mathbf{\Sigma}_{\varepsilon i}$ we use (I) $[\mathbf{\Sigma}_{\varepsilon i}]_{jj} = 1$ for $j = 1, \ldots, 3$, the remaining elements are -0.2, (II) $\mathbf{\Sigma}_{\varepsilon i} = \mathbf{I}_3$ and (III) $[\mathbf{\Sigma}_{\varepsilon i}]_{jj} = 1$ for $j = 1, \ldots, 3$, while the other elements are 0.2. In the first three designs we generate $\mathbf{w}_{it}^\dagger$ by means of the first order vector autoregressive system ($VAR(1)$)

$$\mathbf{w}_{it}^{\dagger} = \mathbf{\Phi}\mathbf{w}_{i,t-1}^{\dagger} + \varepsilon_{it}^{\dagger} \ , \tag{26}$$

where the 3×3 matrix $\mathbf{\Phi}$ comes from one of we use the following designs:

Design $DGP = 1$: A stationary $VAR(1)$ with maximum eigenvalue of 0.6, where

$$\mathbf{\Phi} = \begin{pmatrix} 0.4 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.4 \end{pmatrix} . \tag{27}$$

Design $DGP = 2$: A stationary $VAR(1)$ with maximum eigenvalue of 0.8, where

$$\mathbf{\Phi} = \begin{pmatrix} 0.6 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.6 \end{pmatrix} . \tag{28}$$

Design $DGP = 3$: A stationary $VAR(1)$ with maximum eigenvalue of 0.95, where

$$\mathbf{\Phi} = \begin{pmatrix} 0.75 & 0.1 & 0.1 \\ 0.1 & 0.75 & 0.1 \\ 0.1 & 0.1 & 0.75 \end{pmatrix} . \tag{29}$$

In addition we consider a finite-order vector moving average (MA) processes of the form

$$\mathbf{w}_{it}^{\dagger} = \varepsilon_{it}^{\dagger} + \sum_{l=1}^{q} \mathbf{\Psi}_{il}^{\dagger} \varepsilon_{i,t-l}^{\dagger}, \tag{30}$$

where we choose: Design $DGP = 4$, which is a first-order MA process where

$$\mathbf{\Psi}_{i1}^{\dagger} = \begin{pmatrix} 0.4 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.4 \end{pmatrix}, \tag{31}$$

and Design $DGP = 5$, where $\mathbf{w}_{it}^{\dagger}$ follows a second-order MA process with

$$\mathbf{\Psi}_{i1}^{\dagger} = \begin{pmatrix} 0.6 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.6 \end{pmatrix} \text{ and } \mathbf{\Psi}_{i2}^{\dagger} = \begin{pmatrix} 0.4 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.4 \end{pmatrix}. \tag{32}$$

Recall that the disturbance in the equation for y_{it} is given by the first element of the vector $\mathbf{w}_{it}^{\dagger}$, while its remaining elements contain $\delta \mathbf{x}_{it}$. The maximum numbers of leads and lags of the explanatory variables that are conditionally correlated with the disturbances is equal to *one* in the Designs 1-3, while for the Designs 4 and 5 all lags of the explanatory variables are conditionally correlated with the disturbances. In the case of the VAR(1) models, we generate the initial values for the process $\mathbf{w}_{it}^{\dagger}$ from the implied stationary distribution. Note that by backward substitution, we obtain

$$\mathbf{w}_{i0}^{\dagger} = \sum_{j=0}^{\infty} \mathbf{\Phi}^{j} \varepsilon_{i,-j}^{\dagger} \tag{33}$$

and hence $\mathbf{w}_{i0}^{\dagger}$ is a random variable that is independent from $\varepsilon_{it}^{\dagger}$ for $t > 0$. When the innovations $\varepsilon_{it}^{\dagger}$ are normally distributed, it also follows that $\mathbf{w}_{i0}^{\dagger}$ is normally distributed. Furthermore, it has a mean of zero and $k + 1 \times k + 1$ variance-covariance matrix $E\left(\mathbf{w}_{it}^{\dagger}\mathbf{w}_{it}^{\dagger\prime}\right) = \mathbf{\Gamma}_{i0}^{\dagger}$ where

$$\mathbf{\Gamma}_{i0}^{\dagger} = \mathbb{E}\left[\left(\sum_{j=0}^{\infty} \mathbf{\Phi}^{j} \varepsilon_{i,-j}^{\dagger}\right)\left(\sum_{j=0}^{\infty} \mathbf{\Phi}^{j} \varepsilon_{i,-j}^{\dagger}\right)'\right] = \sum_{j=0}^{\infty} \mathbf{\Phi}^{j} \mathbf{\Sigma}_{\varepsilon i} \mathbf{\Phi}'^{j} . \tag{34}$$

The above expression implies

$$\mathbf{\Phi}\mathbf{\Gamma}_{i0}^{\dagger}\mathbf{\Phi}' = \sum_{j=0}^{\infty}\mathbf{\Phi}^{j+1}\mathbf{\Sigma}_{\varepsilon i}\mathbf{\Phi}'^{j+1} = \mathbf{\Sigma}_{\varepsilon i} + \sum_{j=0}^{\infty}\mathbf{\Phi}^{j}\mathbf{\Sigma}_{\varepsilon i}\mathbf{\Phi}'^{j} = \mathbf{\Sigma}_{\varepsilon i} + \mathbf{\Gamma}_{i0}^{\dagger}. \quad (35)$$

After vectorizing and solving for $\mathbf{\Gamma}_{i0}^{\dagger}$ we obtain (see e.g. Lütkepohl (2006)[p. 29])

$$vec\left(\mathbf{\Gamma}_{i0}^{\dagger}\right) = \left(\mathbf{I}_{(k+1)^2} - \mathbf{\Phi}\otimes\mathbf{\Phi}\right)^{-1} vec\left(\mathbf{\Sigma}_{\varepsilon i}\right) \ . \quad (36)$$

The remaining parameters of the model are chosen as follows: We generate the individual effects α_i from $\alpha_i \sim N(\mathbf{0}_3, \mathbf{I}_3)$. The spatial correlation parameter ρ is chosen from the set $\{-0.95, -0.5, -0.1, 0, 0.1, 0.5, 0.95\}$. The choice of $\mathbf{W}$ is based on Kapoor et al. (2007). In more details we consider: (i) A "one step ahead-one step behind circular world" with corresponding entries 1/2. I.e. $W_{i,i+1} = 0.5$ and $W_{i+1,i} = 0.5$ for $i = 1, \ldots, n-1$. $W_{1,n} = 0.5$ and $W_{n,1} = 0.5$, the other entries are zero. (ii) A "three step ahead-three step behind circular world" with corresponding entries 1/6. (iii) A "five step ahead-five step behind circular world" with corresponding entries 1/10. (iv) A "one step ahead-one step behind Rook constellation" with corresponding entries 1/2. This design is non-circular. Here $W_{i,i+1} = 0.5$ and $W_{i+1,i} = 0.5$ for $i = 1, \ldots, n-1$; the other entries are zero. (v) A "two step ahead-two step behind Queen constellation". In this non-circular design $W_{i,i+1} = 0.3$, $W_{i,i+2} = 0.2$, $W_{i+1,i} = 0.3$ and $W_{i+2,i} = 0.2$ for $i = 1, \ldots, n-2$; the other entries are zero. Thus we have in total 525 different data generating processes (3, 5, 7, 5 different settings for $\mathbf{\Sigma}_{\varepsilon i}$, the autoregressive structure of $\mathbf{w}_{it}^{\dagger}$, the spatial correlation parameter ρ and the spatial correlation matrix $\mathbf{W}$, respectively).

For the estimation of the long run covariance $\Omega_{uu,i}$ we applied the Bartlett and the truncated kernel,[7] i.e. $\hat{\Omega}_{uu,i} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathtt{k}\left(\frac{t-s}{b_T}\right)\hat{\hat{u}}_{it}^2$, where $\mathtt{k}(.)$ is a kernel function with bandwidth b_T and $\hat{\hat{u}}_{it}$ are the residuals. The truncated kernel exhibits a better performance than the Bartlett kernel. For the truncated kernel we either kept the number of lags fixed for all $i = 1, \ldots, n$ or stopped the summation for component i at lag s if the mode of the autocorrelation of the residuals $\{\hat{\hat{u}}_{it}, \hat{\hat{u}}_{i,t-s}\}$ becomes smaller

[7]Given these kernels and our model assumptions the conditions for consistent estimation provided in Jansson (2002) are satisfied.

than the $1.96/\sqrt{T}$ bound. The later choice has lead to a better performance in our simulations. When implementing the $D2SLS$ estimator, the number of leads and lags p included in the regression has to be chosen. Recent literature proposed to choose p by means of information criteria (see e.g. Kejriwal and Perron (2008) and Kurozumi and Tuvaandorj (2010)). With small T and n the implementation of such criteria is straightforward. However, since a dataset with (relatively) large n and T is going to be considered, working with small p becomes necessary due to computational restrictions. In particular, we set $p = 2$ for all components $i = 1, \ldots, n$. For all designs working with $p = 2$ performed better than working with $p = 1$. Last but not least $\mathtt{M} = 1000$ is the number of Monte Carlo steps and $\mathtt{m}$ is the index of the corresponding iteration. To obtain the $DOLS$ estimates $\mathbf{Z}$ is replaced by $\mathbf{X}$ in (25). For the $2SLS$ estimator $\mathbf{Z}$ and $\mathbf{X}$ do not contain any leads and lags, i.e. $\boldsymbol{\zeta} = \emptyset$, while for OLS we use $\mathbf{Z} = \mathbf{X}$ with $\boldsymbol{\zeta} = \emptyset$.

Tables 1 to 8 present results from the simulation runs. We consider the cases where $n = 5$, $T = 200$ and $n = 50$, $T = 200$. In the first three columns the numbers abbreviate the different designs: $\mathbf{W} = (i)$ for the one-steps ahead setting, ..., $\mathbf{W} = (v)$ for the Queen constellation, $DGP = 1, \ldots, 5$ for the autoregressive model used to generate $\mathbf{w}_{it}$, $\boldsymbol{\Sigma}_{\varepsilon i} = (I), (II), (III)$ stands for the covariance matrix used. With different cross-sectional dimensions n we investigated the size of the Wald statistic and obtained the percentages of the simulation runs where the true null hypothesis $\rho = 0$ has been rejected at $\alpha_c = \{0.01, 0.05, 0.1\}$ significance levels. For $\rho = \{-0.95, -0.5, -0.1, 0.1, 0.5, 0.95\}$ the *false* null-hypothesis of $\rho = 0$ has been rejected in all of the simulation runs for most simulation settings. The exceptions appear with $n = 5$, $\rho = \pm 0.1$ and the moving average designs $DGP = 4$ and $DGP = 5$. Here for different $\mathbf{W}$ and $\boldsymbol{\Sigma}_{\varepsilon i}$ we sometimes observe rejection rates smaller than 99%. The smallest rejection rate was 0.45 in the case of the $MA(1)$ process, where $\mathbf{W}$ is of type (iii) and $\boldsymbol{\Sigma}_{\varepsilon i}$ of type (I). For $n = 50$ we observed that the false null hypothesis $\rho = 0$ has always been always rejected.

Tables 1, 2, 5 and 6 present the rejection rates of the Wald test for the true null hypothesis $\rho = 0$. Since for $\rho = 0$ the $DOLS$ estimation procedure is one of the theoretically correct tools presented in literature, the comparison of $D2SLS$ to $DOLS$ is of special interest. With $n = 5$, for designs $DGP = 1 - 3$ the oversizing remains modest for $DOLS$ and $D2SLS$. The rejection rates observed are very similar although the $D2SLS$ uses the instrumental variables, where the numerical complexity is increased. With the moving average process stronger oversizing effects are obtained, for $DOLS$ these effect can become very strong. The highest oversizing with our $D2SLS$ estimator is observed with $\mathbf{\Sigma}_{\varepsilon i}$ of type (I) - where the "serial" endogeneity is stronger - and $\mathbf{W} = (iii)$, where many spatial lags are used. For $n = 50$ the performance of $DOLS$ is very close to the performance of our $D2SLS$ estimator; here in some settings undersizing is observed.

According to the $Bias = \rho - \frac{1}{\mathtt{M}}\sum_{\mathtt{m}=1}^{\mathtt{M}} \hat{\rho}_{\mathtt{m}}$ and the bias adjusted root mean squared error $RMSE = \sqrt{Bias^2 + \left(\frac{1}{\mathtt{M}-1}\sum_{\mathtt{m}=1}^{\mathtt{M}} \hat{\rho}_{\mathtt{m}}^2 - \left(\frac{1}{\mathtt{M}}\sum_{\mathtt{m}=1}^{\mathtt{M}} \hat{\rho}_{\mathtt{m}}\right)^2\right)}$ hardly any differences are observed between $DOLS$ and $D2SLS$ with $n = 5$ for the designs $DGP = 1 - 3$, while substantial differences can be observed with $DGP = 4$ and 5 (see Tables 3 and 4 for $\rho = 0.5$). In addition these effects are more pronounced with $\mathbf{\Sigma}_{\varepsilon i}$ of type (I) and $\mathbf{W} = (iii)$. Surprisingly, $2SLS$ - sometimes also OLS - worked quite well for the different designs considered although these methods do not cope with "serial" endogeneity still arising with $\rho = 0$. We explain this phenomenon as follows: The Wald statistic $S_{\gamma,nT}$ includes estimates of the long-run covariance matrix $\Omega_{uu,i}$ and the parameter estimates. We observed that the estimates of the covariance matrix with OLS and $2SLS$ are in almost all cases much larger than the estimates with $DOLS$ and $D2SLS$. However, also the volatility of the estimator, e.g. measured by the RMSE, increases. Here we claim that these effects can cancel out, such that the rejection rates are not too far away from the significance level α_c considered. With $D2SLS$ the rejection rates are in a lot of the cases even closer to α_c than with $DOLS$ and $D2SLS$. Oversizing of Wald tests based on $DOLS$ estimation (which also require the calculation of the long run covariance matrix) have already been extensively reported in literature (e.g. Mark and Sul (2003)[Table 1]). For $n = 50$ the bias and the $RMSE$ become smaller, as can be expected when using more data (see Tables 7 and 8 for $\rho = 0.95$). Although most differences are small, we observe that the bias and the $RMSE$ of our $D2SLS$ estimator is smaller or equal to the bias and the $RMSE$ of the $DOLS$ estimator. In addition - except for $\mathbf{W} = (v)$, $DPG = 4$ and $\mathbf{\Sigma}_{\varepsilon i}$ of type (I), $D2SLS$ dominates the $2SLS$ estimator. $D2SLS$ always dominates OLS in terms of the bias and the $RMSE$.

Since $2SLS$ did not perform so bad in a lot of the cases considered above, we also increased the degree of serial "endogeneity". We additionally investigate a $\mathbf{\Sigma}_{\varepsilon i}$ where $[\mathbf{\Sigma}_{\varepsilon i}]_{jj} = 1$ for $j = 1, \ldots, 3$, and the other elements are 0.9. For $n = 5$, even with $\rho = 0$ the bias with $2SLS$ is a factor two to ten larger than the bias observed with $D2SLS$. For $\rho \neq 0$ the bias increases. Things are even worse with OLS. Our $D2SLS$ estimator exhibits the usual oversizing behavior with $\rho = 0$ as well. Regarding the empirical analysis carried out in the next section, it is important to note that even with $\rho = 0$, the bias with $2SLS$ for the estimates of the regressors β is approximately a factor two to three of the bias observed with $D2SLS$. This can be explained by the fact that $\tilde{\mathbf{x}}_{it}$ is part of $\mathbf{Z}_{it}$, where $\tilde{\mathbf{x}}_{it}$ is correlated with the noise term by the construction of the model. For $n = 50$ the bias in most cases is the smallest one with $D2SLS$, however the differences in absolute terms become small with $n = 50$.

Summing up, we observe that the estimator developed in Section 3 exhibits (in most cases) some oversizing behavior as already observed in the literature where dynamic least squares estimation has been applied. However, even with the true null-hypothesis $\rho = 0$, where no spatial endogenetiy is present, the $D2SLS$ estimator in most cases outperforms the $DOLS$ estimator. By increasing, both - serial and spatial endogeneity - we observe that the $D2SLS$ estimator performs reasonably well. With the larger data sets, where $T = 200$ and $n = 50$, the bias and the $RMSE$ decreases compared to $T = 200$ and $n = 5$.

5 Empirical Illustration

In this section we apply the tools developed in the former sections to credit risk data. Quantitative finance literature has mainly focused on the default risk of an entity (see e.g. Eom et al. (2004), Crosbie and Bohn (2003), Collin-Dufresne et al. (2001), Campbell and Taksler (2003), Ericsson et al. (2009), Longstaff et al. (2008), among others). In their seminal paper Collin-Dufresne et al. (2001) looked at the residuals – arising from regressing bond spreads on usual credit risk factors – by means of a principal component analysis, where the authors detected a strong factor in the residuals. While the coefficients of determination in the initial regressions are surprisingly low, this factor has a higher explanatory power than the regressors obtained from economic literature. Collin-Dufresne et al. (2001) claim that the strong factor is driven by liquidity risk or other joint market behavior. Based on these findings a lot of articles also looked on joint determinants of credit spreads (see e.g. Zhou (2001), Collin-Dufresne et al. (2003), Jorion and Zhang (2007) and Norden and Weber (2009)). In the following a spatial correlation matrix $\mathbf{W}$ will be derived from input-output data. Equipped with this matrix $\mathbf{W}$ we shall estimate model (9) by the

			$\alpha_c = 0.01$				$\alpha_c = 0.05$				$\alpha_c = 0.10$			
W	DGP	$\boldsymbol{\Sigma}_{\varepsilon i}$	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*
i	1	I	1.0	1.0	3.0	2.9	5.4	5.7	9.2	9.4	10.8	11.0	15.1	15.2
i	1	II	1.2	1.2	2.4	2.2	5.3	5.3	8.7	8.8	10.9	11.2	15.1	14.9
i	1	III	1.2	1.2	3.6	3.5	6.0	5.8	8.9	9.0	11.1	11.2	14.9	14.7
i	2	I	1.7	1.6	2.5	2.5	4.7	4.7	7.8	7.7	8.3	8.4	12.0	12.1
i	2	II	1.8	1.8	3.4	3.2	5.5	5.6	10.1	10.3	9.0	9.1	16.8	16.5
i	2	III	2.6	2.6	3.8	3.8	6.3	6.3	11.8	12.0	11.7	11.7	18.1	17.5
i	3	I	1.6	1.6	1.9	2.0	4.4	4.5	6.8	6.7	8.3	8.4	10.9	11.0
i	3	II	1.8	1.8	2.3	2.3	5.7	5.6	9.2	9.0	8.8	8.9	15.2	15.2
i	3	III	3.0	2.9	3.6	3.6	6.2	6.4	11.4	11.4	11.5	11.4	17.7	17.6
i	4	I	1.4	1.4	1.6	1.6	4.8	5.0	6.6	6.8	9.7	8.1	12.8	12.9
i	4	II	1.2	1.3	2.0	2.1	4.5	4.7	7.1	6.9	8.7	8.7	12.5	12.5
i	4	III	1.3	1.3	2.0	2.0	4.9	4.8	7.0	7.2	10.4	10.4	12.6	12.6
i	5	I	3.8	3.1	7.2	7.8	10.5	9.2	17.2	16.7	18.6	15.4	23.5	23.8
i	5	II	3.1	3.0	5.0	5.4	9.6	9.5	13.8	13.5	16.8	16.8	21.7	21.6
i	5	III	3.2	3.2	5.0	5.0	10.2	10.4	12.5	12.5	16.5	16.6	21.4	21.4
ii	1	I	0.9	0.9	3.1	2.9	5.5	5.2	9.9	9.6	10.7	10.6	16.3	16.1
ii	1	II	1.4	1.3	3.0	2.6	6.3	6.0	9.7	9.6	12.3	12.5	17.4	17.3
ii	1	III	2.0	1.9	2.7	3.1	6.9	6.9	10.2	10.3	13.0	13.4	17.3	17.3
ii	2	I	1.1	1.1	3.2	3.0	3.3	3.2	8.2	8.0	6.8	6.8	14.7	14.4
ii	2	II	1.0	1.0	3.1	3.0	4.3	4.4	9.2	9.0	7.2	7.1	16.6	16.1
ii	2	III	1.5	1.6	3.6	3.7	5.7	5.7	10.0	9.6	10.3	10.4	16.2	16.4
ii	3	I	1.0	1.0	2.5	2.6	3.5	3.4	7.4	7.4	6.9	6.9	13.4	13.7
ii	3	II	1.0	1.0	2.6	2.6	4.3	4.1	8.8	8.9	6.7	6.9	15.2	14.8
ii	3	III	1.7	1.7	3.1	3.1	5.5	5.4	9.1	9.2	9.3	9.4	16.0	16.2
ii	4	I	1.3	1.2	2.0	1.8	4.8	3.9	7.2	7.5	10.4	8.8	12.6	11.9
ii	4	II	1.0	0.8	1.7	2.0	4.1	4.0	7.1	7.3	9.1	9.5	12.7	13.0
ii	4	III	1.4	1.3	2.2	2.2	4.7	4.9	7.5	7.4	10.6	10.6	12.3	12.6
ii	5	I	4.2	2.7	7.2	7.1	11.9	9.3	17.4	17.9	18.9	16.5	25.8	27.6
ii	5	II	3.2	3.2	5.5	5.2	10.3	10.6	13.9	14.2	17.4	17.3	21.3	21.8
ii	5	III	3.2	3.2	5.7	5.5	10.5	10.6	14.0	14.0	18.2	18.2	20.4	20.9
iii	1	I	2.1	0.9	4.6	2.9	8.3	5.2	13.0	9.7	15.0	10.7	19.6	16.1
iii	1	II	2.3	1.3	4.9	2.7	9.3	6.0	13.4	9.7	16.2	12.5	19.8	17.4
iii	1	III	3.0	1.8	5.4	3.1	11.0	6.9	15.9	10.5	18.3	13.4	22.7	17.3
iii	2	I	1.2	1.1	3.1	3.0	3.7	3.3	9.1	8.0	7.1	6.7	15.2	14.4
iii	2	II	1.1	1.0	3.9	3.0	4.3	4.4	10.9	8.9	7.8	7.4	17.7	16.2
iii	2	III	1.6	1.6	4.5	3.7	6.7	5.8	11.2	9.5	11.5	10.4	18.8	16.3
iii	3	I	1.0	1.0	2.6	2.6	3.8	3.4	8.5	7.4	6.7	6.9	13.7	13.6
iii	3	II	1.1	1.0	3.3	2.6	4.6	4.1	10.2	8.8	7.4	6.8	16.9	14.9
iii	3	III	2.0	1.7	4.1	3.1	6.2	5.4	10.7	9.2	10.5	9.3	17.8	16.4
iii	4	I	68.5	1.0	47.3	1.8	78.9	4.0	61.1	7.0	84.0	8.8	67.6	11.7
iii	4	II	5.0	0.9	6.0	2.0	11.4	4.0	14.4	7.3	18.7	9.3	21.8	13.0
iii	4	III	1.9	1.3	2.4	2.2	7.0	4.9	7.9	7.4	13.2	10.6	14.1	12.6
iii	5	I	75.5	2.4	61.0	7.1	84.9	9.1	69.9	17.7	88.7	16.2	76.1	27.5
iii	5	II	9.2	3.2	12.3	5.3	21.2	10.6	22.1	14.3	28.2	17.3	31.7	21.9
iii	5	III	4.6	3.2	6.2	5.5	13.4	10.6	14.4	14.0	22.3	18.2	22.5	20.9

Table 1: Size for the parameter ρ: Rejections rates of the Wald test for the true null hypothesis $\rho = 0$ in percentage terms, given the significance levels $\alpha_c = \{0.01, 0.05, 0.1\}$. $\mathtt{M} = 1000$ simulation runs. Cross-sectional dimension $n = 5$, time series dimension $T = 200$.

$D2SLS$ approach.

Similar to Berndt et al. (2008) the left hand side variable is the CDS spread, while firm specific credit risk proxies, interest rate data and the VIX volatility index are used as the right hand side variables. By

			$\alpha_c = 0.01$				$\alpha_c = 0.05$				$\alpha_c = 0.10$			
$\mathbf{W}$	DGP.	$\mathbf{\Sigma}_{\varepsilon i}$	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*
iv	1	I	0.9	0.9	3.1	3.1	4.6	4.7	8.6	8.2	10.0	9.9	15.5	15.7
iv	1	II	1.4	1.4	2.7	2.5	4.6	4.6	7.9	7.9	10.5	10.4	14.7	14.8
iv	1	III	1.3	1.3	3.0	2.9	5.0	5.2	9.0	9.6	11.0	11.1	15.1	15.1
iv	2	I	1.8	1.8	1.9	1.8	5.7	5.9	7.4	7.5	9.1	8.9	13.6	13.3
iv	2	II	2.3	2.1	3.2	3.1	6.6	6.6	10.6	10.6	10.2	10.2	17.9	18.2
iv	2	III	2.6	2.6	5.1	4.9	7.4	7.6	13.0	12.9	12.5	12.3	18.8	19.0
iv	3	I	1.7	1.7	1.7	1.7	5.4	5.3	7.0	7.2	8.7	9.0	11.7	11.6
iv	3	II	2.2	2.1	2.8	2.8	6.3	6.5	9.5	9.6	9.8	9.9	16.4	16.6
iv	3	III	2.4	2.4	4.1	4.5	7.0	6.9	11.6	11.8	11.4	11.2	17.7	17.8
iv	4	I	1.3	1.5	2.1	2.3	4.8	5.0	7.7	7.0	10.9	9.3	13.7	12.4
iv	4	II	1.0	1.0	2.2	2.5	5.2	5.5	7.1	7.5	10.6	10.4	13.8	14.4
iv	4	III	1.5	1.5	2.3	2.3	6.1	6.2	8.4	8.5	11.3	11.2	13.9	14.2
iv	5	I	4.5	3.9	7.0	7.7	11.4	10.1	17.0	16.9	19.2	16.6	23.8	23.7
iv	5	II	4.1	4.0	6.1	6.1	10.8	10.7	13.8	13.8	17.3	17.2	21.2	20.9
iv	5	III	4.2	4.2	6.0	5.9	11.2	11.4	13.9	14.2	18.1	18.2	21.0	20.9
v	1	I	0.9	0.9	3.2	3.3	5.1	5.0	9.4	9.4	9.9	10.0	15.8	15.9
v	1	II	1.1	1.1	2.6	2.6	5.5	5.6	8.9	8.9	10.8	10.7	14.5	14.4
v	1	III	1.4	1.4	2.7	2.7	6.6	6.8	9.9	10.1	12.3	12.3	17.4	17.3
v	2	I	1.4	1.4	2.5	2.5	5.1	5.4	7.7	7.9	9.6	9.6	13.2	12.8
v	2	II	1.4	1.4	3.1	3.1	6.1	6.1	11.2	11.3	9.7	9.4	17.6	18.0
v	2	III	2.3	2.3	3.9	3.8	7.3	7.2	12.6	12.6	13.0	12.9	19.2	19.2
v	3	I	1.0	1.0	2.0	2.0	5.1	5.1	6.6	6.6	9.6	9.4	11.9	11.8
v	3	II	1.4	1.4	2.7	2.7	5.9	5.6	9.4	9.5	9.8	9.8	15.9	16.2
v	3	III	2.1	2.1	3.4	3.3	6.8	6.7	12.0	12.2	12.9	12.7	18.7	18.8
v	4	I	1.7	1.0	2.7	2.4	5.2	3.9	7.6	7.0	10.0	8.1	13.6	12.6
v	4	II	0.7	0.7	2.2	2.1	4.9	4.6	7.1	7.0	9.7	9.8	12.7	13.4
v	4	III	0.4	0.4	2.0	2.0	5.2	5.1	7.4	7.3	11.3	11.4	13.1	13.2
v	5	I	3.9	2.9	7.6	7.8	9.9	9.2	16.9	17.1	17.2	16.9	25.7	24.9
v	5	II	2.8	3.0	5.7	5.9	11.0	10.7	14.3	13.8	17.8	17.6	22.3	22.4
v	5	III	3.0	2.9	6.0	6.0	11.8	11.8	14.1	14.1	18.9	18.7	22.0	22.2

Table 2: Size for the parameter ρ: Rejections rates of the Wald test for the true null hypothesis $\rho = 0$ in percentage terms, given the significance levels $\alpha_c = \{0.01, 0.05, 0.1\}$. $\mathtt{M} = 1000$ simulation runs. Cross-sectional dimension $n = 5$, time series dimension $T = 200$.

means of the matrix $\mathbf{W}$ we model a specific form of default risk correlation. The Wald test developed in Theorem 1 checks whether the impact of spatial correlation described by $\mathbf{W}$ is significant. Although our approach cannot "solve" the economic problem highlighted by Collin-Dufresne et al. (2001), the following analysis tries to add a further part to the puzzle of modeling credit spreads.

W	DGP	$\Sigma_{\varepsilon i}$	Bias *OLS*	Bias *2SLS*	Bias *DOLS*	Bias *D2SLS*	RMSE *OLS*	RMSE *2SLS*	RMSE *DOLS*	RMSE *D2SLS*
i	1	I	-5.3E-4	2.4E-5	-4.3E-4	1.6E-5	2.16E-3	2.04E-3	2.04E-3	1.95E-3
i	1	II	-5.9E-4	1.0E-5	-5.7E-4	2.0E-5	2.32E-3	2.15E-3	2.15E-3	2.36E-3
i	1	III	-5.9E-4	1.1E-4	-6.0E-4	1.4E-4	2.44E-3	2.37E-3	2.37E-3	2.61E-3
i	2	I	-3.7E-4	-9.7E-5	-2.0E-4	-6.2E-5	2.44E-3	2.60E-3	2.60E-3	1.25E-3
i	2	II	-4.3E-4	-1.5E-4	-2.9E-4	-1.0E-4	2.49E-3	2.63E-3	2.63E-3	1.55E-3
i	2	III	-4.4E-4	-1.1E-4	-4.2E-4	-1.3E-4	2.66E-3	2.80E-3	2.80E-3	2.11E-3
i	3	I	-3.6E-4	-9.0E-5	-1.7E-4	-4.8E-5	2.38E-3	2.55E-3	2.55E-3	1.04E-3
i	3	II	-4.2E-4	-1.4E-4	-2.4E-4	-7.7E-5	2.40E-3	2.54E-3	2.54E-3	1.27E-3
i	3	III	-4.2E-4	-9.6E-5	-3.5E-4	-1.1E-4	2.50E-3	2.64E-3	2.64E-3	1.71E-3
i	4	I	-4.2E-2	-8.0E-4	-2.0E-2	-6.2E-4	5.93E-2	1.86E-2	1.86E-2	2.95E-2
i	4	II	-8.4E-4	-7.2E-5	-8.4E-4	-6.9E-5	2.50E-3	2.10E-3	2.10E-3	2.57E-3
i	4	III	-1.4E-4	-1.3E-5	-1.0E-4	-1.9E-5	7.18E-4	7.15E-4	7.15E-4	6.81E-4
i	5	I	-4.2E-2	-8.5E-4	-2.0E-2	-7.0E-4	5.97E-2	1.84E-2	1.84E-2	2.96E-2
i	5	II	-8.5E-4	-7.8E-5	-8.4E-4	-8.3E-5	2.50E-3	2.10E-3	2.10E-3	2.57E-3
i	5	III	-1.4E-4	-1.5E-5	-9.7E-5	-2.7E-5	7.17E-4	7.10E-4	7.10E-4	6.89E-4
ii	1	I	-9.7E-4	-2.9E-4	-5.7E-4	3.4E-5	4.98E-3	4.66E-3	4.66E-3	4.44E-3
ii	1	II	-8.9E-4	-9.6E-5	-8.1E-4	-1.3E-5	5.44E-3	5.21E-3	5.21E-3	5.73E-3
ii	1	III	-8.1E-4	3.0E-4	-1.2E-3	-1.2E-4	6.11E-3	6.01E-3	6.01E-3	6.83E-3
ii	2	I	-7.1E-4	-4.7E-4	-3.1E-4	-1.5E-4	5.12E-3	5.06E-3	5.06E-3	2.96E-3
ii	2	II	-7.9E-4	-5.1E-4	-4.9E-4	-2.7E-4	5.26E-3	5.17E-3	5.17E-3	3.60E-3
ii	2	III	-6.7E-4	-3.0E-4	-7.0E-4	-3.7E-4	5.68E-3	5.63E-3	5.63E-3	4.69E-3
ii	3	I	-7.0E-4	-4.7E-4	-2.5E-4	-1.1E-4	4.96E-3	4.90E-3	4.90E-3	2.45E-3
ii	3	II	-7.7E-4	-5.0E-4	-4.0E-4	-2.0E-4	5.03E-3	4.95E-3	4.95E-3	2.95E-3
ii	3	III	-6.6E-4	-3.2E-4	-5.7E-4	-2.9E-4	5.28E-3	5.22E-3	5.22E-3	3.79E-3
ii	4	I	-5.2E-2	7.0E-3	-2.5E-2	-1.5E-4	7.88E-2	5.40E-2	5.40E-2	4.48E-2
ii	4	II	-1.7E-3	-3.9E-4	-1.5E-3	-1.5E-4	6.53E-3	5.81E-3	5.81E-3	6.60E-3
ii	4	III	-5.4E-4	-3.9E-4	-2.1E-4	-6.9E-5	2.20E-3	2.07E-3	2.07E-3	1.96E-3
ii	5	I	-5.2E-2	6.5E-3	-2.6E-2	-3.4E-5	7.83E-2	5.22E-2	5.22E-2	4.39E-2
ii	5	II	-1.7E-3	-3.8E-4	-1.4E-3	-8.1E-6	6.59E-3	5.91E-3	5.91E-3	6.61E-3
ii	5	III	-5.4E-4	-3.8E-4	-9.3E-5	5.4E-5	2.23E-3	2.09E-3	2.09E-3	2.00E-3
iii	1	I	-4.1E-3	-5.0E-4	-3.1E-3	5.8E-5	1.09E-2	8.05E-3	8.05E-3	9.20E-3
iii	1	II	-4.5E-3	-1.7E-4	-4.3E-3	-2.2E-5	1.19E-2	9.01E-3	9.01E-3	1.22E-2
iii	1	III	-5.4E-3	5.1E-4	-5.9E-3	-2.0E-4	1.41E-2	1.04E-2	1.04E-2	1.58E-2
iii	2	I	-2.3E-3	-8.2E-4	-1.1E-3	-2.7E-4	9.29E-3	8.76E-3	8.76E-3	5.29E-3
iii	2	II	-2.5E-3	-8.9E-4	-1.7E-3	-4.7E-4	9.62E-3	8.94E-3	8.94E-3	6.55E-3
iii	2	III	-2.6E-3	-5.3E-4	-2.4E-3	-6.4E-4	1.03E-2	9.74E-3	9.74E-3	8.62E-3
iii	3	I	-2.3E-3	-8.2E-4	-9.5E-4	-1.8E-4	9.00E-3	8.48E-3	8.48E-3	4.38E-3
iii	3	II	-2.5E-3	-8.8E-4	-1.4E-3	-3.5E-4	9.22E-3	8.56E-3	8.56E-3	5.38E-3
iii	3	III	-2.5E-3	-5.5E-4	-2.0E-3	-5.0E-4	9.62E-3	9.04E-3	9.04E-3	7.00E-3
iii	4	I	-2.7E-1	1.2E-2	-1.3E-1	-1.2E-4	3.80E-1	9.70E-2	9.70E-2	1.88E-1
iii	4	II	-7.6E-3	-6.8E-4	-7.3E-3	-2.6E-4	1.77E-2	1.01E-2	1.01E-2	1.73E-2
iii	4	III	-1.5E-3	-6.8E-4	-8.6E-4	-1.2E-4	4.49E-3	3.59E-3	3.59E-3	3.69E-3
iii	5	I	-2.7E-1	1.1E-2	-1.3E-1	9.2E-5	3.81E-1	9.34E-2	9.34E-2	1.88E-1
iii	5	II	-7.7E-3	-6.5E-4	-7.3E-3	-1.3E-5	1.80E-2	1.02E-2	1.02E-2	1.73E-2
iii	5	III	-1.5E-3	-6.6E-4	-6.9E-4	9.3E-5	4.54E-3	3.61E-3	3.61E-3	3.64E-3

Table 3: Bias and $RMSE$ for the parameter estimates for $\rho = 0.5$. Cross-sectional dimension $n = 5$, time series dimension $T = 200$. $\mathtt{M} = 1000$ Monte Carlo steps.

W	DGP	$\mathbf{\Sigma}_{\varepsilon i}$	Bias *OLS*	Bias *2SLS*	Bias *DOLS*	Bias *D2SLS*	RMSE *OLS*	RMSE *2SLS*	RMSE *DOLS*	RMSE *D2SLS*
iv	1	I	-5.6E-4	3.0E-5	-4.5E-4	2.7E-5	2.57E-3	2.42E-3	2.42E-3	2.27E-3
iv	1	II	-5.9E-4	6.3E-5	-5.6E-4	7.5E-5	2.67E-3	2.54E-3	2.54E-3	2.68E-3
iv	1	III	-6.1E-4	1.3E-4	-6.2E-4	1.5E-4	2.90E-3	2.79E-3	2.79E-3	3.11E-3
iv	2	I	-4.2E-4	-1.4E-4	-2.5E-4	-9.6E-5	2.93E-3	3.02E-3	3.02E-3	1.46E-3
iv	2	II	-4.9E-4	-2.1E-4	-3.5E-4	-1.4E-4	3.07E-3	3.13E-3	3.13E-3	1.89E-3
iv	2	III	-5.0E-4	-1.6E-4	-4.7E-4	-1.7E-4	3.27E-3	3.32E-3	3.32E-3	2.57E-3
iv	3	I	-4.0E-4	-1.3E-4	-2.1E-4	-7.8E-5	2.87E-3	2.96E-3	2.96E-3	1.21E-3
iv	3	II	-4.6E-4	-1.9E-4	-2.8E-4	-1.1E-4	2.96E-3	3.01E-3	3.01E-3	1.53E-3
iv	3	III	-4.8E-4	-1.5E-4	-4.0E-4	-1.4E-4	3.08E-3	3.12E-3	3.12E-3	2.08E-3
iv	4	I	-4.5E-2	-1.6E-3	-2.1E-2	-8.4E-4	6.40E-2	2.20E-2	2.20E-2	3.22E-2
iv	4	II	-9.7E-4	-1.9E-4	-9.2E-4	-1.4E-4	2.92E-3	2.47E-3	2.47E-3	2.92E-3
iv	4	III	-1.9E-4	-6.3E-5	-1.3E-4	-4.4E-5	8.76E-4	8.35E-4	8.35E-4	8.20E-4
iv	5	I	-4.5E-2	-1.6E-3	-2.2E-2	-8.7E-4	6.42E-2	2.17E-2	2.17E-2	3.24E-2
iv	5	II	-9.6E-4	-1.8E-4	-9.1E-4	-1.4E-4	2.88E-3	2.45E-3	2.45E-3	2.92E-3
iv	5	III	-1.8E-4	-6.0E-5	-1.2E-4	-4.4E-5	8.68E-4	8.28E-4	8.28E-4	8.21E-4
v	1	I	-7.5E-4	-1.6E-4	-5.4E-4	-5.0E-5	3.49E-3	3.28E-3	3.28E-3	3.07E-3
v	1	II	-7.4E-4	-5.6E-5	-6.6E-4	1.2E-5	3.78E-3	3.60E-3	3.60E-3	3.77E-3
v	1	III	-6.3E-4	1.7E-4	-7.2E-4	1.0E-4	4.28E-3	4.12E-3	4.12E-3	4.62E-3
v	2	I	-5.2E-4	-2.6E-4	-3.2E-4	-1.8E-4	4.06E-3	4.21E-3	4.21E-3	2.09E-3
v	2	II	-5.9E-4	-3.0E-4	-4.4E-4	-2.4E-4	4.23E-3	4.35E-3	4.35E-3	2.62E-3
v	2	III	-5.5E-4	-1.8E-4	-5.7E-4	-2.5E-4	4.46E-3	4.56E-3	4.56E-3	3.51E-3
v	3	I	-5.0E-4	-2.4E-4	-2.6E-4	-1.4E-4	3.97E-3	4.12E-3	4.12E-3	1.72E-3
v	3	II	-5.6E-4	-2.8E-4	-3.6E-4	-1.8E-4	4.08E-3	4.21E-3	4.21E-3	2.13E-3
v	3	III	-5.4E-4	-1.8E-4	-4.7E-4	-2.1E-4	4.20E-3	4.31E-3	4.31E-3	2.84E-3
v	4	I	-4.7E-2	-2.4E-4	-2.2E-2	-1.0E-3	6.81E-2	2.98E-2	2.98E-2	3.70E-2
v	4	II	-1.2E-3	-3.2E-4	-1.0E-3	-1.8E-4	3.91E-3	3.46E-3	3.46E-3	3.88E-3
v	4	III	-3.0E-4	-1.7E-4	-1.6E-4	-6.7E-5	1.27E-3	1.22E-3	1.22E-3	1.14E-3
v	5	I	-4.7E-2	-3.1E-4	-2.3E-2	-1.0E-3	6.84E-2	2.92E-2	2.92E-2	3.66E-2
v	5	II	-1.2E-3	-3.1E-4	-1.0E-3	-1.5E-4	3.86E-3	3.43E-3	3.43E-3	3.85E-3
v	5	III	-2.9E-4	-1.7E-4	-1.2E-4	-3.5E-5	1.26E-3	1.21E-3	1.21E-3	1.14E-3

Table 4: Bias and $RMSE$ for the parameter estimates for $\rho = 0.5$. Cross-sectional dimension $n = 5$, time series dimension $T = 200$. $\mathtt{M} = 1000$ Monte Carlo steps.

5.1 Data

In this analysis CDS spreads are used to describe the implied credit risk of a firm.[8] The insurance premium the buyer has to pay to the seller is the CDS premium. The CDS premium is the amount payable per year to insure against the event of default of any underlying with notational amount 1; it is usually measured in basis points. With the usual quarterly frequency, the buyer pays $premium/(4 \cdot 10000)$ times the nominal value stipulated in the contract to the seller. The probability of default and the loss given default (one minus the recovery rate) should be the main driving forces of the CDS spreads (see e.g. Hull (2006),

[8] With a CDS contract a protection buyer acquires insurance against the default of a specified company. The protection seller declares his willingness to compensate the protection buyer for a loss arising in the case of default of the specified entity. For more details on the specification of credit default swap contracts we refer the reader to the International Securities and Derivatives Association (ISDA); `www.isda.org`.

			$\alpha_c = 0.01$				$\alpha_c = 0.05$				$\alpha_c = 0.10$			
W	DGP	$\mathbf{\Sigma}_{\varepsilon i}$	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*
i	1	I	1.3	1.3	1.9	1.9	5.3	5.6	8.8	9.2	10.9	10.9	14.2	14.0
i	1	II	1.4	1.4	2.2	2.4	6.5	6.5	8.9	8.8	11.6	11.7	15.8	15.8
i	1	III	1.4	1.4	2.9	3.0	6.6	6.6	9.5	9.3	11.4	11.4	14.5	14.6
i	2	I	1.2	1.1	0.8	0.9	5.1	5.2	4.7	4.9	9.7	9.6	10.2	10.4
i	2	II	1.0	1.1	2.1	2.3	5.4	5.5	8.2	8.3	10.5	10.4	14.1	13.8
i	2	III	1.6	1.7	4.0	3.9	6.7	6.8	10.3	10.2	13.5	13.7	16.6	16.8
i	3	I	1.1	1.1	0.4	0.5	4.9	4.9	4.0	4.0	9.8	9.5	7.9	7.9
i	3	II	0.9	1.0	1.2	1.2	5.2	5.3	7.1	7.3	10.2	10.4	12.2	12.5
i	3	III	1.6	1.6	3.3	3.2	6.7	6.5	9.8	9.9	13.5	13.5	15.6	15.5
i	4	I	0.2	0.2	0.8	0.8	3.2	3.0	6.2	6.0	7.8	6.8	12.4	12.2
i	4	II	0.0	0.0	0.6	0.8	3.2	3.2	5.8	6.0	8.4	8.4	11.8	11.8
i	4	III	0.4	0.4	1.0	1.0	4.0	4.0	6.4	6.4	8.8	8.8	12.0	12.0
i	5	I	2.4	2.0	5.6	6.2	8.2	7.6	14.2	14.2	14.4	13.2	21.0	20.8
i	5	II	2.0	2.0	4.8	4.6	8.4	8.6	13.4	12.4	15.6	15.4	21.0	21.2
i	5	III	2.2	2.2	4.0	4.0	8.2	8.2	11.8	11.8	15.4	15.4	20.0	20.2
ii	1	I	1.3	1.3	2.0	2.1	5.6	5.5	8.2	8.2	11.2	11.0	16.0	16.0
ii	1	II	1.3	1.2	1.7	1.5	6.3	6.3	9.5	9.4	11.2	11.3	15.9	16.1
ii	1	III	1.2	1.2	1.9	2.0	6.3	6.0	9.4	9.7	11.9	12.0	16.2	16.4
ii	2	I	0.8	0.8	1.1	1.0	4.1	4.1	4.9	4.7	8.5	8.4	10.4	10.6
ii	2	II	1.4	1.4	1.8	1.8	4.3	4.3	7.0	7.0	9.0	9.4	13.8	13.8
ii	2	III	1.8	1.8	2.5	2.4	6.4	6.6	10.0	9.8	11.2	11.0	15.1	15.6
ii	3	I	0.9	0.9	0.9	0.9	3.8	3.9	3.8	3.8	8.3	8.3	8.2	8.0
ii	3	II	1.3	1.2	2.0	2.0	4.6	4.4	6.4	6.3	8.4	8.6	12.2	12.1
ii	3	III	1.6	1.6	2.2	2.2	6.5	6.5	8.8	9.0	11.3	11.2	14.4	14.8
ii	4	I	0.0	0.4	0.6	1.2	4.0	3.8	6.0	5.0	7.8	7.2	9.6	9.8
ii	4	II	0.2	0.2	0.4	0.4	3.8	3.8	6.0	5.6	7.8	7.8	10.4	10.2
ii	4	III	0.2	0.2	0.6	0.6	4.6	4.6	5.6	5.6	7.8	8.0	11.2	11.2
ii	5	I	3.6	2.8	4.8	4.6	9.2	8.0	12.8	13.4	16.0	13.6	22.8	22.2
ii	5	II	3.6	3.4	5.0	4.6	8.4	8.6	11.2	11.0	15.4	15.0	19.0	18.8
ii	5	III	3.0	3.0	4.4	4.4	9.2	9.0	11.2	11.0	16.0	15.8	20.0	20.2
iii	1	I	1.6	1.7	2.4	2.6	5.6	5.7	8.0	7.9	13.0	13.0	12.9	12.9
iii	1	II	1.3	1.3	2.3	2.4	6.6	6.7	7.5	7.6	12.9	12.7	13.7	13.8
iii	1	III	1.2	1.3	2.2	2.3	6.8	6.6	9.3	9.0	12.4	12.3	13.7	13.6
iii	2	I	1.0	0.9	1.5	1.5	4.0	4.2	4.2	4.2	8.0	7.9	8.9	8.9
iii	2	II	1.4	1.3	2.3	2.4	5.1	5.1	8.0	7.9	10.0	9.8	13.7	13.5
iii	2	III	1.6	1.6	3.3	3.3	6.3	6.6	9.9	9.9	12.2	12.1	15.1	15.2
iii	3	I	0.9	1.0	1.0	0.9	4.3	4.2	3.7	3.6	8.0	8.0	7.0	7.0
iii	3	II	1.1	1.2	1.8	1.9	5.3	5.1	7.3	7.4	9.9	10.4	11.9	12.1
iii	3	III	1.4	1.4	2.8	2.8	7.0	6.8	8.7	8.5	12.5	12.6	15.4	15.3
iii	4	I	0.6	0.6	2.2	2.4	4.0	3.6	6.0	5.0	9.0	7.8	12.0	10.8
iii	4	II	0.6	0.6	1.0	1.0	4.2	4.2	6.6	6.6	9.8	9.6	12.4	12.4
iii	4	III	0.4	0.4	1.0	1.0	4.6	4.6	7.2	7.2	9.0	9.0	11.8	11.8
iii	5	I	3.6	2.6	4.8	5.0	10.2	10.0	13.6	13.6	16.4	15.8	23.2	21.2
iii	5	II	3.0	3.0	5.6	5.2	10.8	11.2	14.0	13.6	18.2	18.6	22.6	22.2
iii	5	III	3.2	3.2	5.2	5.2	9.8	9.8	14.4	14.4	17.4	17.6	22.4	22.4

Table 5: Size for the parameter ρ: Rejections rates of the Wald test for the true null hypothesis $\rho = 0$ in percentage terms, given the significance levels $\alpha_c = \{0.01, 0.05, 0.1\}$. $\mathtt{M} = 1000$ simulation runs. Cross-sectional dimension $n = 50$, time series dimension $T = 200$.

Schönbucher (2003)).

We utilize the dataset already used in Schneider et al. (2010), where CDS spreads of 278 firms obtained from the *Markit Group* have been investigated. We focus on the five year maturities which are typically

			$\alpha_c = 0.01$				$\alpha_c = 0.05$				$\alpha_c = 0.10$			
W	DGP	$\boldsymbol{\Sigma}_{\varepsilon i}$	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*
iv	1	I	1.3	1.4	1.7	1.7	5.5	5.4	8.6	8.6	11.0	10.7	15.2	15.0
iv	1	II	1.6	1.5	2.3	2.4	6.4	6.4	9.1	9.1	12.1	12.1	15.9	15.8
iv	1	III	1.5	1.4	3.3	3.1	6.9	6.9	8.7	8.9	11.7	11.6	15.1	15.0
iv	2	I	1.0	1.0	0.8	0.8	5.2	5.1	5.0	4.8	9.5	9.5	9.8	9.6
iv	2	II	1.3	1.2	1.9	2.0	5.2	5.3	7.5	7.6	10.4	10.2	13.9	13.8
iv	2	III	1.6	1.6	3.5	3.7	6.9	7.0	10.1	9.8	13.0	13.1	16.4	16.4
iv	3	I	1.0	1.0	0.5	0.5	4.7	4.7	3.6	3.6	9.5	9.5	8.2	8.2
iv	3	II	1.1	1.1	1.2	1.1	5.0	4.9	6.8	6.6	10.6	10.6	12.2	12.0
iv	3	III	1.6	1.6	3.2	3.2	6.5	6.5	8.9	9.0	12.9	12.9	15.7	15.4
iv	4	I	0.2	0.2	0.8	0.8	3.6	3.8	5.8	5.8	7.4	7.0	11.4	11.6
iv	4	II	0.2	0.2	1.0	1.0	3.8	3.8	7.0	7.0	7.8	8.0	12.8	12.8
iv	4	III	0.6	0.6	1.2	1.2	4.2	4.2	6.8	6.8	8.8	8.8	13.2	13.0
iv	5	I	2.6	2.4	5.6	5.8	8.6	7.6	13.8	14.4	14.2	15.2	22.2	22.4
iv	5	II	2.4	2.4	4.8	4.8	8.2	8.4	13.2	13.2	16.4	16.2	20.8	20.8
iv	5	III	2.2	2.2	3.8	3.8	9.0	9.0	12.6	12.6	15.6	15.4	19.2	19.6
v	1	I	1.3	1.3	1.7	1.6	4.5	4.6	7.9	7.7	10.4	10.3	14.0	13.9
v	1	II	1.2	1.2	2.1	2.1	4.7	4.7	7.7	7.5	10.7	10.5	14.9	14.6
v	1	III	1.1	1.2	2.1	2.1	5.4	5.4	8.0	7.9	10.7	10.6	15.3	15.1
v	2	I	0.7	0.7	0.8	0.9	4.6	4.8	3.9	3.9	9.4	9.4	9.1	9.0
v	2	II	1.0	0.9	1.6	1.6	5.0	5.1	7.2	7.5	10.9	11.1	13.6	13.4
v	2	III	1.8	1.6	3.1	3.1	6.9	7.0	10.5	10.4	11.3	11.4	16.5	17.0
v	3	I	0.6	0.6	0.6	0.6	4.7	4.7	2.8	2.8	9.5	9.3	6.9	7.1
v	3	II	0.9	1.0	1.3	1.3	4.6	4.8	6.5	6.4	10.3	10.3	12.3	12.1
v	3	III	2.0	1.9	2.6	2.6	7.0	7.0	9.4	9.4	11.0	11.1	16.5	16.4
v	4	I	0.0	0.2	1.4	1.4	3.4	4.0	6.4	5.4	8.0	6.8	10.4	11.0
v	4	II	0.4	0.4	2.0	2.0	4.8	5.0	6.0	6.2	7.6	7.6	13.0	12.6
v	4	III	0.2	0.2	2.0	2.0	5.0	5.0	6.4	6.4	8.4	8.2	11.6	11.6
v	5	I	1.0	1.7	4.0	4.0	6.0	4.7	10.3	10.0	11.7	9.0	20.0	18.0
v	5	II	2.3	2.3	4.3	4.3	5.7	5.7	10.3	10.0	12.3	12.3	17.0	17.3
v	5	III	2.0	2.0	3.7	3.7	6.7	6.7	10.3	10.3	13.0	12.3	17.7	18.0

Table 6: Size for the parameter ρ: Rejections rates of the Wald test for the true null hypothesis $\rho = 0$ in percentage terms, given the significance levels $\alpha_c = \{0.01, 0.05, 0.1\}$. `M` = 1000 simulation runs. Cross-sectional dimension $n = 5$, time series dimension $T = 200$.

the most liquid ones (see e.g. Hull et al. (2004)). The observation period is January 2, 2001 to May 30, 2008. In line with a bulk of quantitative finance literature we stick to weekly data, such that $T = 230$. Using weekly data instead of daily observations is often done to avoid day of the week effects.

Next the CDS data are matched with firm specific characteristics obtained from *Thomson Datastream* and *Compustat* data. We construct the KMV distance to default, DD_{it}, from firm specific data by following Crosbie and Bohn (2003). Moreover, we calculate the debt to value ratio, DVR_{it}. This firm specific data was available for 176 out of the 278 firms. Following Berndt et al. (2008) we also include the VIX volatility index from the *Chicago Board Options Exchange* (`http://www.cboe.com/micro/VIX/vixintro.aspx`) as an explanatory variable. Additionally, we include a short run and a long run interest rate obtained from the *Federal Reserve* (`http://federalreserve.gov/releases/h15/data.htm`). In more detail we use

W	DGP	$\Sigma_{\varepsilon i}$	Bias *OLS*	Bias *2SLS*	Bias *DOLS*	Bias *D2SLS*	RMSE *OLS*	RMSE *2SLS*	RMSE *DOLS*	RMSE *D2SLS*
i	1	I	-3.1E-5	-2.1E-6	-2.8E-5	-3.4E-6	6.96E-5	8.49E-5	8.49E-5	6.00E-5
i	1	II	-3.4E-5	-2.2E-6	-3.5E-5	-4.1E-6	7.54E-5	9.02E-5	9.02E-5	7.29E-5
i	1	III	-3.9E-5	-1.4E-6	-3.9E-5	-3.1E-6	8.10E-5	9.51E-5	9.51E-5	8.52E-5
i	2	I	-3.6E-5	-8.6E-6	-1.5E-5	-6.9E-7	1.19E-4	1.72E-4	1.72E-4	5.36E-5
i	2	II	-3.8E-5	-5.1E-6	-2.1E-5	1.4E-8	1.23E-4	1.74E-4	1.74E-4	6.80E-5
i	2	III	-3.9E-5	1.3E-6	-2.9E-5	6.3E-7	1.25E-4	1.74E-4	1.74E-4	8.86E-5
i	3	I	-3.5E-5	-8.5E-6	-1.3E-5	-9.2E-7	1.18E-4	1.70E-4	1.70E-4	4.52E-5
i	3	II	-3.7E-5	-5.2E-6	-1.8E-5	-2.0E-7	1.20E-4	1.71E-4	1.71E-4	5.64E-5
i	3	III	-3.7E-5	1.2E-6	-2.4E-5	6.1E-7	1.21E-4	1.68E-4	1.68E-4	7.24E-5
i	4	I	-2.7E-3	-2.1E-5	-1.3E-3	-2.9E-5	3.39E-3	6.25E-4	6.25E-4	1.65E-3
i	4	II	-3.3E-5	-2.5E-6	-3.3E-5	-2.9E-6	6.14E-5	6.52E-5	6.52E-5	6.44E-5
i	4	III	-4.7E-6	-7.2E-7	-3.9E-6	-8.1E-7	1.64E-5	2.20E-5	2.20E-5	1.65E-5
i	5	I	-2.7E-3	-2.3E-5	-1.3E-3	-3.1E-5	3.40E-3	6.22E-4	6.22E-4	1.65E-3
i	5	II	-3.3E-5	-2.6E-6	-3.3E-5	-2.4E-6	6.17E-5	6.53E-5	6.53E-5	6.41E-5
i	5	III	-4.8E-6	-7.5E-7	-3.7E-6	-6.8E-7	1.65E-5	2.21E-5	2.21E-5	1.65E-5
ii	1	I	-4.0E-5	-4.5E-6	-3.5E-5	-6.8E-6	1.12E-4	1.25E-4	1.25E-4	9.73E-5
ii	1	II	-4.2E-5	-2.0E-6	-4.2E-5	-5.9E-6	1.18E-4	1.32E-4	1.32E-4	1.12E-4
ii	1	III	-4.4E-5	3.2E-6	-4.6E-5	-2.1E-6	1.20E-4	1.33E-4	1.33E-4	1.26E-4
ii	2	I	-4.1E-5	-1.8E-5	-1.5E-5	-1.1E-8	1.73E-4	2.29E-4	2.29E-4	8.23E-5
ii	2	II	-4.6E-5	-1.7E-5	-2.1E-5	3.8E-7	1.82E-4	2.31E-4	2.31E-4	1.04E-4
ii	2	III	-4.4E-5	-6.0E-6	-3.1E-5	2.5E-6	1.85E-4	2.31E-4	2.31E-4	1.35E-4
ii	3	I	-4.1E-5	-1.8E-5	-1.3E-5	-1.6E-7	1.71E-4	2.27E-4	2.27E-4	6.93E-5
ii	3	II	-4.6E-5	-1.8E-5	-1.8E-5	2.4E-8	1.79E-4	2.28E-4	2.28E-4	8.56E-5
ii	3	III	-4.4E-5	-6.6E-6	-2.6E-5	2.1E-6	1.79E-4	2.23E-4	2.23E-4	1.10E-4
ii	4	I	-3.1E-3	-3.2E-5	-1.4E-3	-6.7E-5	4.29E-3	8.83E-4	8.83E-4	2.12E-3
ii	4	II	-3.9E-5	-5.0E-6	-3.8E-5	-6.1E-6	9.11E-5	9.20E-5	9.20E-5	9.37E-5
ii	4	III	-6.5E-6	-1.6E-6	-4.3E-6	-1.7E-6	2.64E-5	3.13E-5	3.13E-5	2.61E-5
ii	5	I	-3.1E-3	-4.2E-5	-1.4E-3	-6.6E-5	4.31E-3	8.79E-4	8.79E-4	2.10E-3
ii	5	II	-4.0E-5	-6.0E-6	-3.7E-5	-5.3E-6	9.19E-5	9.19E-5	9.19E-5	9.25E-5
ii	5	III	-6.8E-6	-2.0E-6	-3.6E-6	-1.4E-6	2.64E-5	3.16E-5	3.16E-5	2.59E-5
iii	1	I	-5.3E-5	-9.0E-6	-4.6E-5	-7.1E-6	1.64E-4	1.71E-4	1.71E-4	1.41E-4
iii	1	II	-5.6E-5	-8.9E-6	-5.5E-5	-8.0E-6	1.73E-4	1.83E-4	1.83E-4	1.63E-4
iii	1	III	-5.7E-5	-4.4E-6	-5.9E-5	-5.5E-6	1.71E-4	1.85E-4	1.85E-4	1.77E-4
iii	2	I	-4.6E-5	-2.1E-5	-1.5E-5	-8.4E-7	2.12E-4	2.84E-4	2.84E-4	1.08E-4
iii	2	II	-5.2E-5	-2.0E-5	-2.3E-5	1.5E-6	2.28E-4	2.99E-4	2.99E-4	1.39E-4
iii	2	III	-4.6E-5	-5.2E-6	-3.4E-5	-8.4E-7	2.32E-4	3.20E-4	3.20E-4	1.76E-4
iii	3	I	-4.7E-5	-2.2E-5	-1.3E-5	-7.8E-7	2.10E-4	2.80E-4	2.80E-4	9.01E-5
iii	3	II	-5.3E-5	-2.1E-5	-1.9E-5	1.4E-6	2.23E-4	2.94E-4	2.94E-4	1.14E-4
iii	3	III	-4.6E-5	-6.5E-6	-2.9E-5	-4.0E-7	2.24E-4	3.09E-4	3.09E-4	1.44E-4
iii	4	I	-3.6E-3	-4.9E-5	-1.7E-3	-7.0E-5	5.45E-3	1.27E-3	1.27E-3	2.68E-3
iii	4	II	-5.0E-5	-1.1E-5	-4.6E-5	-8.9E-6	1.29E-4	1.38E-4	1.38E-4	1.30E-4
iii	4	III	-9.3E-6	-4.4E-6	-4.9E-6	-2.8E-6	3.78E-5	4.64E-5	4.64E-5	3.63E-5
iii	5	I	-3.6E-3	-6.5E-5	-1.7E-3	-7.2E-5	5.46E-3	1.26E-3	1.26E-3	2.64E-3
iii	5	II	-5.1E-5	-1.3E-5	-4.5E-5	-8.0E-6	1.32E-4	1.38E-4	1.38E-4	1.27E-4
iii	5	III	-9.6E-6	-4.9E-6	-3.7E-6	-2.2E-6	3.82E-5	4.68E-5	4.68E-5	3.59E-5

Table 7: Bias and $RMSE$ for the parameter estimates for $\rho = 0.95$. Cross-sectional dimension $n = 50$, time series dimension $T = 200$. $\mathtt{M} = 1000$ Monte Carlo steps.

two year and ten year US treasury yields, denoted by r_{2t} and r_{10t}, respectively. Since a firm's cost of capital is usually affected by interest rates, government bond yields are often included when credit risk is investigated. A more detailed description of the data and the construction of the explanatory variables is provided in Appendix C.

$\mathbf{W}$	DGP	$\mathbf{\Sigma}_{\varepsilon i}$	Bias				RMSE			
			OLS	*2SLS*	*DOLS*	*D2SLS*	*OLS*	*2SLS*	*DOLS*	*D2SLS*
iv	1	I	-3.2E-5	-2.8E-6	-2.9E-5	-3.9E-6	7.52E-5	8.86E-5	8.86E-5	6.35E-5
iv	1	II	-3.5E-5	-2.9E-6	-3.7E-5	-4.7E-6	8.05E-5	9.38E-5	9.38E-5	7.66E-5
iv	1	III	-4.0E-5	-2.2E-6	-4.1E-5	-3.9E-6	8.46E-5	9.98E-5	9.98E-5	8.89E-5
iv	2	I	-3.7E-5	-9.1E-6	-1.6E-5	-6.4E-7	1.28E-4	1.80E-4	1.80E-4	5.63E-5
iv	2	II	-3.9E-5	-5.8E-6	-2.2E-5	-2.8E-7	1.30E-4	1.81E-4	1.81E-4	7.15E-5
iv	2	III	-4.0E-5	3.2E-7	-3.0E-5	1.1E-8	1.31E-4	1.82E-4	1.82E-4	9.20E-5
iv	3	I	-3.6E-5	-9.1E-6	-1.4E-5	-9.2E-7	1.27E-4	1.78E-4	1.78E-4	4.76E-5
iv	3	II	-3.8E-5	-5.9E-6	-1.9E-5	-4.7E-7	1.28E-4	1.78E-4	1.78E-4	5.93E-5
iv	3	III	-3.8E-5	2.3E-7	-2.5E-5	1.3E-7	1.26E-4	1.76E-4	1.76E-4	7.52E-5
iv	4	I	-2.9E-3	-2.8E-5	-1.4E-3	-2.7E-5	3.64E-3	6.75E-4	6.75E-4	1.76E-3
iv	4	II	-3.6E-5	-3.5E-6	-3.6E-5	-3.7E-6	6.58E-5	6.98E-5	6.98E-5	6.87E-5
iv	4	III	-5.3E-6	-1.1E-6	-4.4E-6	-1.1E-6	1.74E-5	2.35E-5	2.35E-5	1.73E-5
iv	5	I	-2.9E-3	-3.1E-5	-1.4E-3	-2.9E-5	3.65E-3	6.73E-4	6.73E-4	1.76E-3
iv	5	II	-3.6E-5	-3.7E-6	-3.6E-5	-3.2E-6	6.60E-5	6.98E-5	6.98E-5	6.82E-5
iv	5	III	-5.4E-6	-1.2E-6	-4.1E-6	-9.4E-7	1.75E-5	2.35E-5	2.35E-5	1.73E-5
v	1	I	-3.6E-5	-4.4E-6	-3.2E-5	-6.5E-6	9.37E-5	1.06E-4	1.06E-4	7.87E-5
v	1	II	-3.9E-5	-5.3E-6	-4.0E-5	-8.3E-6	9.92E-5	1.11E-4	1.11E-4	9.35E-5
v	1	III	-4.2E-5	-4.3E-6	-4.3E-5	-7.3E-6	1.02E-4	1.17E-4	1.17E-4	1.08E-4
v	2	I	-3.9E-5	-1.1E-5	-1.7E-5	1.3E-6	1.57E-4	2.00E-4	2.00E-4	7.08E-5
v	2	II	-4.2E-5	-8.8E-6	-2.3E-5	2.3E-6	1.62E-4	2.00E-4	2.00E-4	8.97E-5
v	2	III	-4.2E-5	-3.4E-6	-3.2E-5	3.5E-6	1.63E-4	2.02E-4	2.02E-4	1.17E-4
v	3	I	-3.9E-5	-1.2E-5	-1.4E-5	6.2E-7	1.55E-4	1.98E-4	1.98E-4	5.99E-5
v	3	II	-4.1E-5	-9.6E-6	-2.0E-5	1.3E-6	1.59E-4	1.96E-4	1.96E-4	7.43E-5
v	3	III	-4.1E-5	-4.2E-6	-2.7E-5	2.9E-6	1.57E-4	1.94E-4	1.94E-4	9.52E-5
v	4	I	-3.1E-3	-3.0E-6	-1.5E-3	-1.8E-5	4.06E-3	7.96E-4	7.96E-4	1.98E-3
v	4	II	-3.9E-5	-2.3E-6	-3.9E-5	-2.5E-6	8.13E-5	8.20E-5	8.20E-5	8.37E-5
v	4	III	-6.1E-6	-9.0E-7	-4.6E-6	-7.0E-7	2.26E-5	2.75E-5	2.75E-5	2.21E-5
v	5	I	-3.0E-3	-5.1E-6	-1.4E-3	-1.9E-6	3.92E-3	7.04E-4	7.04E-4	1.86E-3
v	5	II	-3.7E-5	-2.5E-6	-3.6E-5	-1.8E-6	7.89E-5	7.41E-5	7.41E-5	7.90E-5
v	5	III	-6.1E-6	-1.1E-6	-4.1E-6	-6.7E-7	2.22E-5	2.53E-5	2.53E-5	2.09E-5

Table 8: Bias and $RMSE$ for the parameter estimates for $\rho = 0.95$. Cross-sectional dimension $n = 50$, time series dimension $T = 200$. $\mathtt{M} = 1000$ Monte Carlo steps.

To apply and estimate the spatial autocorrelation model, the spatial weights matrix $\mathbf{W}$ has to be constructed. We use the industry-by-industry total requirements matrix for the year 2002 provided by the *Bureau of Labor Statistics* (BLS) and match each firm in our data to a particular BLS industry. In this data set the total requirements matrix contains for each industry $\mathtt{i}$ the proportion of inputs ultimately stemming from each other industry $\mathtt{j}$ relative to its own sales. We use this to proxy for possible correlation of shocks coming through the supply chain. In more formal terms we consider the weights $\underline{\mathtt{W}}_{\mathtt{ij}} = \{$Inputs from industry $\mathtt{j}$ in US\$$\}/\{$Total sales in industry $\mathtt{i}$ in US\$ $\}$. If firm i operates in industry $\mathtt{i}$ and firm j in industry $\mathtt{j}$, the weights $\underline{W}_{ij}$ are set equal to $\underline{\mathtt{W}}_{\mathtt{ij}}$. This results in the $n \times n$ matrix $\underline{\mathbf{W}}$. Then we set the elements of the matrix $\underline{\mathbf{W}}$ along the main diagonal equal to zero, which yields the matrix $\underline{\underline{\mathbf{W}}}$. To improve the numerical properties and to be able to interpret the estimated coefficients, we

	2	3	4	5	6	7	Total
AA	0	4	1	5	0	0	10
A	4	19	5	16	0	1	45
BBB	16	25	18	8	1	4	72
BB	1	6	6	1	0	0	14
B	1	3	2	0	1	0	7
Total	22	57	32	30	2	5	148

Table 9: Distribution of firms according to industry and rating. Horizontally first digit of the firm's NAIC Code. Cross-sectional dimension $n = 148$.

normalize the matrix $\underline{\underline{\mathbf{W}}}$ by its largest absolute eigenvalue. As a result we get our matrix $\mathbf{W}$. The range for the spatial autocorrelation parameter is bounded by one and, as a result, ρ can be interpreted in a manner comparable to the time series autocorrelation parameter. With the zeros in the main diagonal and $\rho \in (-1, 1)$, the requirements of Assumption 1 are fulfilled.

After matching the CDS data with the data collected from *Thomson Datastream*, *Compustat* and the *Bureau of Labor Statistics* and correcting for firms where we detected problems in the data (e.g. extreme spikes, missing values, unclear industry affiliation, etc) we arrived at a cross-section of $n = 148$ firms. A clustering of the data toward the first digit of the NAICs industry classification and the S&P rating results in Table 9. A NAICs code starting with 2 stands for mining, utilities or construction, 3 for manufacturing, 4 for trade and transportation, 5 for information, banking and finance, 6 for educational services, health care and social assistance, while 7 stands for arts, entertainment, and accommodation and food services. For more details see `http://www.naics.com`.

Before we proceed with the econometric model, let us briefly discuss the expected impacts (expected based on economic theory, intuition and literature). The reader should note that the CDS spread is often used as an indicator for the probability of default of a firm. Since the distance to default measures the distance to the default boundary, we expect a lower spread if the distance to default increases. A raise in the firm's leverage should increase the default probability and therefore the CDS spread. If the interest rate increases the cost of capital increases for a leveraged firm. This should drive up the CDS spread. With the volatility measure VIX we expect higher spreads in periods of higher volatility. The rating of a firm should also reflect the probability of default. Rating effects should be included in the fixed effects

α_i, $i = 1, \ldots, n$ in model (1). Last but not least, due to possible credit risk correlation and contagion effects we expect that the CDS spreads are positively correlated. So we expect a positive ρ when (9) is estimated. Finally, we want to remark that this econometric specification includes the "common variables" $\mathbf{x}_{ct} = (\Delta r_{2t}, \Delta r_{10t}, \Delta VIX_t)'$, where $\mathbf{w}_{ct} = (\Delta r_{2t}, \Delta r_{10t}, \Delta VIX_t)'$. To include $\mathbf{x}_{ct}$ in model (1) we only have to augment Assumption 2 by the assumption that $\mathbf{w}_{ct} = \mathbf{\Psi}_c(L)\,\varepsilon_{ct}$ is independent of $\mathbf{w}_{it}$ for all $i = 1, \ldots, n$, and Assumption 3 by assuming that also $\mathbf{x}_{ct}$ is a full rank integrated process with $\mathbf{\Omega}_c > 0$. In this case $\tilde{\mathbf{x}}_{ct}$ can be included into $\mathbf{X}$ and $\mathbf{Z}$, where $\mathbf{x}_{ct}$ is part of each $\mathbf{X}_{it}$ and $\mathbf{Z}_{it}$. Theorem 1 still continues to hold.[9]

By Assumption 2 the explanatory variables $\mathbf{x}_{it}$ should be $I(1)$. The question arises whether our model assumptions are compatible with the data observed. Our variables include the distance to default which should follow a geometric Brownian motion as long as the firm does not default based on the model assumptions (see e.g. Crosbie and Bohn (2003), Schönbucher (2003)). Observe that by construction, it must be that $DD_{it} \geq 0$. Translated to discrete time the distance to default should follow a random walk with an absorbing barrier. Only firms that do not hit this barrier are observed in the sample. Among the remaining variables, the debt to value ratio lives on the interval $[0, 100]$, the VIX index measures volatility and is therefore be non-negative. Following applied literature, we run augmented Dickey-Fuller tests for a unit root for these data and the CDS spreads themselves and find that the null of a unit root is not rejected for almost all time series with a five percent significance level. We also used the Im, Pesaran and Shin tests provided in in the *EViews* package and arrived at the same results. For the distance to default the null of a unit root is rejected, although the serial correlation is quite high. For the debt-to-value ratios, the VIX and the interest rates there is strong evidence for the presence of a unit root. Given these results, we conjecture that our theoretical model considered in Section 2 provides a useful approximation of the (unknown) data generating process of the empirical data considered.

[9] Maybe the independence assumption between $\mathbf{w}_{it}$ and $\mathbf{w}_{ct}$ may be a strong assumption. By relaxing this assumption and allowing for correlation between all components would require to project on all leads and lags as already discussed in the paragraph below equation (23). For large n we would suffer from the curse of dimensionality arising with the nuisance parameters. This problem will be subject to further research.

$\hat{\gamma}$	*OLS*		*2SLS*		*DOLS*		*D2SLS*	
ρ	0.5211	< 0.001	0.3920	0.0331	0.4936	< 0.001	0.3928	< 0.001
β_{DVR}	-14.6912	0.0092	-15.2656	0.4317	-21.5765	< 0.001	-22.2903	< 0.001
β_{DD}	5.3419	< 0.001	5.4718	< 0.001	5.1182	< 0.001	5.2172	< 0.001
β_{r_2}	4.6310	0.4589	4.2952	0.0025	9.9653	0.0256	9.9547	0.0260
$\beta_{r_{10}}$	-39.5642	0.0024	-42.4143	0.8117	-49.7624	< 0.001	-52.4572	< 0.001
β_{VIX}	-0.0472	0.8319	-0.0233	< 0.001	-0.1985	0.2076	-0.1869	0.2293

Table 10: Parameter Estimates: Model (9) applied to CDS data. y_{it} is the CDS spread on a firm level. The explanatory variables are the distance to default, DD_{it}, the debt to value ratio, DVR_{it}, a two year bond yield r_{2t}, a ten year bond yield r_{10t}, and the VIX volatility index VIX_t. $T = 230$, $n = 148$, $p = 2$ and $q_\rho = 2$.

5.2 Results

Using our data set, we estimate the parameter vector γ by means of two-stage least squares, $DOLS$, OLS and $D2SLS$. The results are presented in Table 10. Based on the theoretical considerations above, only the $D2SLS$ estimator should be used. The results from the other estimation methods are included only for comparison. When instrumental variables are used in the estimation, the debt-to-value ratio and the VIX are used in $\sum_{i=1}^{n} \mathbf{W}\tilde{x}_{it\iota}$, i.e. $q_\rho = 2$. For these two variables we observed the highest correlation with $\sum_{i=1}^{n} \mathbf{W}\tilde{y}_{it}$. All the p-values presented in Table 10 are obtained by means of a Wald test as described in Theorem 1. For the distance to default and the debt to value ratio the parameters are highly significant and have the expected signs. Both interest rates are significant but work in oppsite directions. Whereas the short term interest rate r_{2t} increases the CDS spread, the long term interest rate decreases the spread. In contrast to some results obtained in literature, the VIX volatility index is not significant when $D2SLS$ estimation is performed and default significance levels (1%, 5%, 10%) are applied. The additional parameter which has been investigated in our analysis is the spatial correlation ρ. With the dynamic two stage least squares estimator the spatial correlation parameter ρ is positive as expected and highly significant. I.e. in addition to the methodological results obtained in the former sections, our model allows to include and to test for spatial correlation. Here we observed a significant effect.

6 Conclusions

In this paper we have studied panel data models with a cointegration relationship including a spatial lag. Due to this spatial lag standard estimation techniques do not provide us with appropriate tools to estimate the parameters and to perform inference. Based on this problem we stick to the usual assumptions used in the dynamic least squares estimation and develop a dynamic two stage least squares estimator. We show that the parameter vector of interest is asymptotically independent of the nuisance parameters. Moreover, we derive the asymptotic distribution of the parameters, which also allows constructing a Wald test to perform statistical inference. Our estimation methodology is applied to simulated data to investigate the small sample properties and to financial data to test for the impact of spatial correlation on credit default swap spreads. Given this financial data set and a spatial correlation matrix obtained from input-output data, our analysis shows that spatial correlation is highly significant.

A Proof of Theorem 1

The two stage least squares estimator is given by (24) and hence

$$
\begin{aligned}
\widehat{(\gamma',\delta')}'_{D2SLS} - (\gamma',\delta')' &= \left(\mathbf{X}'\mathbf{P}_H\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_H\underline{\mathbf{u}}, \qquad (37)\\
&\quad \text{where with } \mathbf{P}_H = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}' \text{ we get}\\
&= \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\underline{\mathbf{u}}\ .
\end{aligned}
$$

$\mathbf{X}$ is a $Tn \times 1 + k + (2p+1)k \cdot n$ matrix while $\mathbf{Z}$ is of dimension $Tn \times q_\rho + k + (2p+1)k \cdot n$. Note that the orthogonal projection $\mathbf{P}_H$ on $(\mathbf{x}_t, \boldsymbol{\zeta}_t)$ is $(\mathbf{x}_t, \boldsymbol{\zeta}_t)$. This yields

$$
\mathbf{X}'\mathbf{P}_H = \begin{pmatrix} \mathbf{y}^* \\ \text{---} \\ \mathbf{x}' \\ \text{---} \\ \boldsymbol{\zeta}' \end{pmatrix} \mathbf{P}_H = \begin{pmatrix} \mathbf{y}^*\mathbf{P}_H \\ \text{-------} \\ \mathbf{x}' \\ \text{-------} \\ \boldsymbol{\zeta}' \end{pmatrix}. \qquad (38)
$$

$\mathbf{Z}_{it}$ was defined in Remark 2 as the transpose of the row of $\mathbf{Z}$ corresponding to the index it. It is of dimension $q_\rho + k + (2p+1)kn \times 1$. $\mathbf{Z}_{it,1:d_z}$ consists of the first d_z elements of $\mathbf{Z}_{it}$, where $d_z = q_\rho + k \times 1$. The remaining elements of $\mathbf{Z}_{it}$ contain $\boldsymbol{\zeta}_{it}$. The "non-it elements" of this vector are zero. In the same way we obtained $\mathbf{X}_{it}$ which is of dimension and $1 + k + (2p+1)kn \times 1$. The first d_x elements are $\mathbf{X}_{it,1:d_x}$, where $d_x = k + 1$. $d_x \le d_z$ hold throughout the following analysis.

Due to Assumption 4 the number of leads and lags used in the projection is $p = p(T)$. Note that $T_\star = T - 2p(T)$, where $T_\star \to \infty$ if $T \to \infty$. Therefore it is not necessary to distinguish between the $T_\star \to \infty$ and $T \to \infty$ when taking limits in the following analysis. In addition note that $\underline{u}_{it}$ converges to u_{it} (this follows from Saikkonen (1991)[Theorem 4.1/ Lemma A.5]). u_{it} is orthogonal to $\boldsymbol{\zeta}_{it}$.

`Step 1`: Let us consider the term $\mathbf{X}'\mathbf{P}_H\mathbf{X}$. We normalize the elements of $\mathbf{Z}$ and $\mathbf{X}$ as follows: expand the first d_z and d_x elements by $\frac{1}{T_\star}$, the remaining terms (accounting for $\boldsymbol{\zeta}_t$) are multiplied by $\frac{1}{\sqrt{T_\star}}$. Based on this we arrive at:

Definition 1. The $d_x + (2p+1)k \cdot n \times d_x + (2p+1)k \cdot n$ matrix $\mathbf{M}^{\star}_{nT}$ is given by:

$$\mathbf{M}^{\star}_{nT} = \begin{pmatrix} \mathbf{y}^{*\prime}\mathbf{x}^{*}/(T_\star^2 n) & \mathbf{y}^{*\prime}\mathbf{x}/(T_\star^2 n) & \mathbf{y}^{*\prime}\boldsymbol{\zeta}/(T_\star\sqrt{T_\star}) \\ \mathbf{x}^{\prime}\mathbf{x}^{*}/(T_\star^2 n) & \mathbf{x}^{\prime}\mathbf{x}/(T_\star^2 n) & \mathbf{x}^{\prime}\boldsymbol{\zeta}/(T_\star\sqrt{T_\star}) \\ \boldsymbol{\zeta}^{\prime}\mathbf{x}^{*}/(T_\star\sqrt{T_\star}) & \boldsymbol{\zeta}^{\prime}\mathbf{x}/(T_\star\sqrt{T_\star}) & \boldsymbol{\zeta}^{\prime}\boldsymbol{\zeta}/(\sqrt{T_\star}\sqrt{T_\star}) \end{pmatrix} \cdot$$
$$\cdot \begin{pmatrix} \mathbf{x}^{*\prime}\mathbf{x}^{*}/(T_\star^2 n) & \mathbf{x}^{*\prime}\mathbf{x}/(T_\star^2 n) & \mathbf{x}^{*\prime}\boldsymbol{\zeta}/(T_\star\sqrt{T_\star}) \\ \mathbf{x}^{\prime}\mathbf{x}^{*}/(T_\star^2 n) & \mathbf{x}^{\prime}\mathbf{x}/(T_\star^2 n) & \mathbf{x}^{\prime}\boldsymbol{\zeta}/(T_\star\sqrt{T_\star}) \\ \boldsymbol{\zeta}^{\prime}\mathbf{x}^{*}/(T_\star\sqrt{T_\star}) & \boldsymbol{\zeta}^{\prime}\mathbf{x}/(T_\star\sqrt{T_\star}) & \boldsymbol{\zeta}^{\prime}\boldsymbol{\zeta}/(\sqrt{T_\star}\sqrt{T_\star}) \end{pmatrix}^{-1} \cdot$$
$$\cdot \begin{pmatrix} \mathbf{x}^{*\prime}\mathbf{y}/(T_\star^2 n) & \mathbf{x}^{*\prime}\mathbf{x}^{*}/(T_\star^2 n) & \mathbf{x}^{*}\boldsymbol{\zeta}^{\prime}/(T_\star\sqrt{T_\star}) \\ \mathbf{x}^{\prime}\mathbf{y}^{*}/(T_\star^2 n) & \mathbf{x}^{\prime}\mathbf{x}/(T_\star^2 n) & \mathbf{x}^{\prime}\boldsymbol{\zeta}/(T_\star\sqrt{T_\star}) \\ \boldsymbol{\zeta}^{\prime}\mathbf{y}^{*}/(T_\star\sqrt{T_\star}) & \boldsymbol{\zeta}^{\prime}\mathbf{x}/(T_\star\sqrt{T_\star}) & \boldsymbol{\zeta}^{\prime}\boldsymbol{\zeta}/(\sqrt{T_\star}\sqrt{T_\star}) \end{pmatrix}, \quad (39)$$

where the $T \to \infty$ and therefore the $T_\star \to \infty$ limit of $\mathbf{M}^{\star}_{nT}$ is denoted by $\mathbf{M}^{\star}_{n}$. In addition we define the $d_x \times d_x$ matrices $\mathbf{M}_{nTi}$ and $\mathbf{M}_{nT}$:

$$\begin{aligned} \mathbf{M}_{nTi} &= \left(\frac{1}{T_\star^2 n}\sum_{i=1}^{n}\sum_{t=1}^{T_\star}\mathbf{X}_{it,1:d_x}\mathbf{Z}^{\prime}_{it,1:d_z}\right)\left(\frac{1}{T_\star^2 n}\sum_{i=1}^{n}\sum_{t=1}^{T_\star}\mathbf{Z}_{it,1:d_z}\mathbf{Z}^{\prime}_{it,1:d_z}\right)^{-1}\frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\mathbf{Z}^{\prime}_{it,1:d_z}\mathbf{X}_{it,1:d_x}\ , \\ \mathbf{M}_{nT} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}_{nTi}\ . \end{aligned} \quad (40)$$

We denote their $T \to \infty$ limits in distribution by $\mathbf{M}_{ni}$ and $\mathbf{M}_{n}$, respectively.

Remark 4. In Remark 2 we already noted that the two-stage least squares estimator and the *DOLS* estimator are special cases of the dynamic two-stage least squares estimator. When we consider $\mathbf{M}_{nTi}$ and assume that $\mathbf{x}^* = \mathbf{y}^*$, the product of the first two terms has to result in the identity matrix. In this case $\mathbf{M}_{nTi} = \frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\mathbf{Z}^{\prime}_{it,1:d_z}\mathbf{X}_{it,1:d_x} = \frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\mathbf{X}^{\prime}_{it,1:d_z}\mathbf{X}_{it,1:d_x}$. This term exactly corresponds to the term $\mathbf{M}_{nTi}$ in the *DOLS* paper of Mark and Sul (2003). The same argument holds with $\mathbf{m}_{nTi}$.

In the following steps we observe that $\mathbf{M}_n$ is a submatrix of $\mathbf{M}^{\star}_{n}$. To obtain the $T \to \infty$ limit of $\mathbf{M}^{\star}_{nT}$, we are confronted with the terms

$$\frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\tilde{\mathbf{x}}_{it\lambda}\tilde{\mathbf{x}}_{jt\iota}\xrightarrow{d}\int\tilde{\mathcal{B}}_{vi\lambda}\tilde{\mathcal{B}}_{vj\iota}\;,\;\frac{1}{T_\star^\kappa}\sum_{t=1}^{T_\star}\tilde{\boldsymbol{\zeta}}_{it\lambda}\tilde{\mathbf{x}}_{jt\iota}\xrightarrow{p}0\;,\;\frac{1}{T_\star^\kappa}\sum_{t=1}^{T_\star}\underline{\tilde{u}}_{it\lambda}\tilde{\mathbf{x}}_{jt\iota}\xrightarrow{p}0\text{ for }\kappa\geq\frac{3}{2}. \tag{41}$$

The terms of the form of $\frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it}\xrightarrow{d}\int\tilde{\mathcal{B}}_{vi}\tilde{\mathcal{B}}'_{vi}$.[10] In addition we meet terms of the structure

$$\frac{1}{\sqrt{T_\star}\sqrt{T_\star}}\sum_{t=1}^{T_\star}\tilde{\boldsymbol{\zeta}}_{it}\tilde{\boldsymbol{\zeta}}_{it}\xrightarrow{p}\begin{cases}\boldsymbol{\Gamma}^{\boldsymbol{\zeta}}_{vv,ij} & \text{for } j\in\mathbb{Z}\text{ and}\\ \mathbf{0}_{k\times k} & \text{else.}\end{cases} \tag{42}$$

$\frac{1}{\sqrt{T_\star}\sqrt{T_\star}}\sum_{t=1}^{T_\star}\tilde{\boldsymbol{\zeta}}_{it}\tilde{\boldsymbol{\zeta}}_{jt}$ converges to a matrix of zeros by the independence across i assumption (i.e. Assumption 2). For each fixed $i=1,\ldots,n$, the matrix $\Gamma^{\boldsymbol{\zeta}}_{vv,ij}$ contains the $k\times k$ covariance matrices $\Gamma_{vv,ij}$, where $j\in\mathbb{Z}$.

Consider the now terms in (39). By the above arguments each of the three matrices converges to a block diagonal matrix. For the first matrix we obtain a non-zero block in the north-west of dimension $d_x\times d_x$, and a non-zero block consisting of $\boldsymbol{\Gamma}_{vv,ij}$. The south-west and the north-east blocks are zero. With the second matrix we observe almost the same effect. The non-zero north-west block is of dimension $d_z\times d_z$, the south-east block is the same as the south-east block of the first matrix. The south-west and the north-east blocks are zero. The third matrix is the transpose of the first matrix. Therefore, the limit matrix $\mathbf{M}_n^\star$ is block diagonal. From $\mathbf{M}_n^\star$ we can extract the matrix $\mathbf{M}_{ni}^\star$ focusing on the index i. The limit of the submatrix $[\mathbf{M}_{ni}^\star]_{(1:k+1,1:k+1)}$ is $\mathbf{M}_{ni}$ while the limit of $[\mathbf{M}_n^\star]_{(1:k+1,1:k+1)}$ is

[10] The second and the third term converge to zero in probability. This also follows from Johansen (1995)[Chapter 13 & Appendix], Saikkonen (1991) and Davidson (1994). We already know that (e.g. Davidson (1994)[Theorem 30.13])

$$\frac{1}{T_\star}\sum_{t=1}^{T_\star}\tilde{\boldsymbol{\zeta}}_{it\lambda}\tilde{\mathbf{x}}_{lt\iota}\xrightarrow{d}\int d\tilde{\mathcal{B}}_{vi\lambda}\tilde{\mathcal{B}}_{vl\iota}\text{ and }\frac{1}{T_\star}\sum_{t=1}^{T_\star}\tilde{\mathbf{x}}_{lt\iota}\underline{\tilde{u}}_{it}\xrightarrow{d}\sqrt{\Omega_{uu,i}}\int\tilde{\mathcal{B}}_{vl\iota}d\mathcal{W}_{ui}+\underline{\Delta}_{vu,li,\iota}\;.$$

The correlation term $\underline{\Delta}_{vu,li,\iota}$ is derived by means of $\mathbb{E}\left(\Delta\tilde{\mathbf{x}}_{lt\iota}\tilde{u}_{it}\right)+\sum_{j=1}^{\infty}\mathbb{E}\left(\Delta\tilde{\mathbf{x}}_{lt\iota}\tilde{u}_{it-j}\right)$. If e.g. u_{it} and $\mathbf{v}_{it}$ are independent, then $\underline{\Delta}_{vu,ii,\iota}=0$ for $\iota=1,\ldots,k$. With independent components (i,l) all $\underline{\Delta}_{vu,li,\iota}=0$ for $i\neq l$. A random variable convergent in distribution is bounded in probability, or $\mathcal{O}_p(1)$ in Landau notation (see e.g. White (2001)[Lemma 4.5]). We can now consider $\frac{1}{T_\star^{3/2}}\sum_{t=1}^{T_\star}\tilde{\boldsymbol{\zeta}}_{it\lambda}\mathbf{x}_{jt\iota}$ as the product $a\cdot b$, where $a=\frac{1}{\sqrt{T_\star}}$ and $b=\frac{1}{T_\star}\sum_{t=1}^{T_\star}\tilde{\boldsymbol{\zeta}}_{it\lambda}\mathbf{x}_{jt\iota}$. Since a is converging to zero, it is $o(1)$ and therefore also $o_p(1)$. b converges in distribution and therefore (its Euclidian norm) is $\mathcal{O}_p(1)$. We obtain convergence in probability to zero since the product $o_p(1)\mathcal{O}_p(1)$ behaves like $o_p(1)$. Landau symbols are e.g. discussed in Poirier (1995)[page 196].

$\mathbf{M}_n$. $[\mathbf{M}^\star_{ni}]_{(k+2:k+1+(2p(T)+1)k\cdot n,k+2:k+1+(2p(T)+1)k\cdot n)}$ is a block diagonal matrix consisting of $\mathbf{\Gamma}^{\boldsymbol{\zeta}}_{vv,ij}$. The elements in the north-eastern and the south-western blocks of $\mathbf{M}^\star_{ni}$ and $\mathbf{M}^\star_n$ are zero. By this result, in the limit *only* the first d_x and d_z columns have an impact of the estimates of γ, while the remaining non-zero block affects the estimates of δ. Next we consider $\mathbf{X}'\mathbf{P}_H\underline{\mathbf{u}}$. Let us define the following terms:

Definition 2. Consider the $1 + k + (2p(T) + 1)k \cdot n$ dimensional vector

$$\mathbf{m}^\star_{nT} = \begin{pmatrix} \mathbf{y}^{*\prime}\mathbf{x}^*/(T_\star^2 n) & \mathbf{y}^{*\prime}\mathbf{x}/(T_\star^2 n) & \mathbf{y}^{*\prime}\boldsymbol{\zeta}/(T_\star\sqrt{T_\star}) \\ \mathbf{x}'\mathbf{x}^*/(T_\star^2 n) & \mathbf{x}'\mathbf{x}/(T_\star^2 n) & \mathbf{x}'\boldsymbol{\zeta}/(T_\star\sqrt{T_\star}) \\ \boldsymbol{\zeta}'\mathbf{x}^*/(T_\star\sqrt{T_\star}) & \boldsymbol{\zeta}'\mathbf{x}/(T_\star\sqrt{T_\star}) & \boldsymbol{\zeta}'\boldsymbol{\zeta}/(\sqrt{T_\star}\sqrt{T_\star}) \end{pmatrix} \cdot \tag{43}$$

$$\cdot \begin{pmatrix} \mathbf{x}^{*\prime}\mathbf{x}^*/(T_\star^2 n) & \mathbf{x}^{*\prime}\mathbf{x}/(T_\star^2 n) & \mathbf{x}^{*\prime}\boldsymbol{\zeta}/(T_\star\sqrt{T_\star}) \\ \mathbf{x}'\mathbf{x}^*/(T_\star^2 n) & \mathbf{x}'\mathbf{x}/(T_\star^2 n) & \mathbf{x}'\boldsymbol{\zeta}/(T_\star\sqrt{T_\star}) \\ \boldsymbol{\zeta}'\mathbf{x}^*/(T_\star\sqrt{T_\star}) & \boldsymbol{\zeta}'\mathbf{x}/(T_\star\sqrt{T_\star}) & \boldsymbol{\zeta}'\boldsymbol{\zeta}/(\sqrt{T_\star}\sqrt{T_\star}) \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{x}^{*\prime}/(T_\star\sqrt{n}) \\ \mathbf{x}'/(T_\star\sqrt{n}) \\ \boldsymbol{\zeta}'/\sqrt{T_\star} \end{pmatrix} \underline{\mathbf{u}} \cdot$$

The $T \to \infty$ limit is denoted by $\mathbf{m}^\star_n$. In addition we define

$$\begin{aligned} \mathbf{m}_{nTi} &= \left(\frac{1}{T_\star^2 n}\sum_{i=1}^{n}\sum_{t=1}^{T_\star}\mathbf{X}_{it,1:d_x}\mathbf{Z}'_{it,1:d_z}\right)\left(\frac{1}{T_\star^2 n}\sum_{i=1}^{n}\sum_{t=1}^{T_\star}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}\right)^{-1}\frac{1}{T_\star}\sum_{t=1}^{T_\star}\mathbf{Z}_{it,1:d_z}\underline{u}_{it}\ , \\ \mathbf{m}_{nT} &= \frac{1}{\sqrt{n}}\mathbf{m}_{nTi}\ . \end{aligned} \tag{44}$$

We denote their $T \to \infty$ limits by $\mathbf{m}_{ni}$ and $\mathbf{m}_n$ respectively.

The first and the second matrix have been considered in (39), where we have already observed that only the first d_x and d_z elements affect γ as $T \to \infty$. The product of these two matrices is multiplied with $\sum_{t=1}^{T_\star}(\mathbf{Z}_{it}\cdot D_T)\underline{u}_{it}$, where $D_T = \left((\frac{1}{T_\star}\cdot\mathbf{1}_{d_z})', (\frac{1}{\sqrt{T_\star}}\cdot\mathbf{1}_{(2p(T)+1)k\cdot n})'\right)'$ and $\mathbf{1}_a$ is a vector of ones of dimension a. Last but not least Saikkonen (1991)[Theorem 4.1/ Lemma A.5] has shown that for the truncation error e_{it} we observe that $\|\sum_{t=1}^{T_\star}(\mathbf{Z}_{it}\cdot D_T)\mathsf{e}_{it}\|_2 = o_p(p^{1/2})$ such that $\underline{u}_{it}$ converges in probability to u_{it} as $T \to \infty$. u_{it} is uncorrelated with $\tilde{\boldsymbol{\zeta}}_{it} = \mathbf{Z}_{it,d_z+1:d_z+(2p(T)+1)k\cdot n}$. By this we observe that only the first d_x components of $\mathbf{X}_{it}$ and the first d_z components of $\mathbf{Z}_{it}$ enter into the limit of the estimator γ. Therefore we have arrived at the *first result*: When $T \to \infty$ the limit distribution of γ is given by the inverse of $\mathbf{M}_n$ times $\mathbf{m}_n$. By the block diagonal structure obtained in the above paragraphs, elements of $\mathbf{m}^{\star}_{nT}$ and $\mathbf{M}^{\star}_{nT}$ outside $(1 : d_x)$ and $(1 : d_x \times 1 : d_x)$ do not affect the asymptotic distribution of γ. Hence γ and δ are asymptotically independent. The submatrix $[\mathbf{M}^{\star}_{nT}]_{(k+2:k+1+(2p(T)+1)k\cdot n,k+2:k+1+(2p(T)+1)k\cdot n)}$ converges to a matrix consisting of $\mathbf{\Gamma}_{vv,ij}$, γ and δ_i are independent for $i = 1, \ldots, n$.

`Step 2`: Based on this asymptotic independence result we are permitted to focus on the matrix $\mathbf{M}_{nT}$ and on the $k+1$ dimensional vector $\mathbf{m}_{nT}$ to investigate the limit behavior of γ. In more detail

$$\mathbf{M}_{ZZ,nTi} := \frac{1}{T_\star^2} \sum_{t=1}^{T_\star} \mathbf{Z}_{it,1:d_z} \mathbf{Z}'_{it,1:d_z} \tag{45}$$
$$= \frac{1}{T_\star^2} \sum_{t=1}^{T_\star} (\sum_{j=1}^{n} W_{ij}^{\tau_1} \widetilde{x}_{jt1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}} \widetilde{x}_{jtq_\rho}, \widetilde{\mathbf{x}}'_{it})'(\sum_{j=1}^{n} W_{ij}^{\tau_1} \widetilde{x}_{jt1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}} \widetilde{x}_{jtq_\rho}, \widetilde{\mathbf{x}}'_{it}) \ .$$

Using the results for the functional central limit we derive

$$\frac{1}{T_\star^2} \sum \mathbf{Z}_{it,1:d_z} \mathbf{Z}'_{it,1:d_z} \stackrel{d}{\to} \mathbf{M}_{ZZ,ni} \tag{46}$$

where

$$\mathbf{M}_{ZZ,ni,(1:q_\rho,1:q_\rho)} :=$$
$$\begin{pmatrix} \int \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{\mathcal{B}}_{vj1} \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{\mathcal{B}}_{vj1}, \ldots, \int \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{\mathcal{B}}_{vj1} \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{\mathcal{B}}_{vjq_\rho} \\ \vdots \ddots \vdots \\ \int \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{\mathcal{B}}_{vjq_\rho} \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{\mathcal{B}}_{vj1}, \ldots, \int \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{\mathcal{B}}_{vjq_\rho} \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{\mathcal{B}}_{vjq_\rho} \end{pmatrix},$$

$$\mathbf{M}_{ZZ,ni,(1:q_\rho,q_\rho+1:q_\rho+k)} := \begin{pmatrix} \int \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{\mathcal{B}}_{vj1} \tilde{\mathcal{B}}_{vi1}, \ldots, \int \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{\mathcal{B}}_{vj1} \tilde{\mathcal{B}}_{vik} \\ \vdots \\ \int \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{\mathcal{B}}_{vjq_\rho} \tilde{\mathcal{B}}_{vi1}, \ldots, \int \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{\mathcal{B}}_{vjq_\rho} \tilde{\mathcal{B}}_{vik} \end{pmatrix},$$

$$\mathbf{M}_{ZZ,ni,(q_\rho+1:q_\rho+k,1:q_\rho)} := \mathbf{M}'_{ZZ,ni,(1:q_\rho,q_\rho+1:q_\rho+k)} \ ,$$
$$\mathbf{M}_{ZZ,ni,(q_\rho+1:q_\rho+k,q_\rho+1:q_\rho+k)} := \int \tilde{\mathcal{B}}_{vi} \tilde{\mathcal{B}}'_{vi} \ .$$

Based on (46) we arrive at

$$\mathbf{M}_{ZZ,Ti} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}_{ZZ,nTi} \xrightarrow{d} \mathbf{M}_{ZZ,n} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}_{ZZ,ni}\ . \tag{47}$$

In a similar way we derive the limit of

$$\mathbf{X}_{it,1:d_x}\mathbf{Z}'_{it,1:d_z} = (\sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt}, \widetilde{\mathbf{x}}'_{it})'(\sum_{j=1}^{n} W_{ij}^{\tau_1}\widetilde{x}_{jt1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}}\widetilde{x}_{jtq_\rho}, \widetilde{\mathbf{x}}'_{it})\ . \tag{48}$$

Based on our model assumptions we obtain $y_{jt} = \sum_{l=1}^{n} K_{jl}\left(\beta'\mathbf{x}_{lt} + \tilde{u}_{lt}^{\dagger}\right)$, where $K_{jl} = \left[(\mathbf{I} - \rho\mathbf{W})^{-1}\right]_{(jl)}$. Then

$$\sum_{j=1}^{n} W_{ij}\tilde{y}_{jt} \;=\; \sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}K_{jl}\left(\beta'\tilde{\mathbf{x}}_{lt} + \tilde{u}_{lt}^{\dagger}\right)\ . \tag{49}$$

By means of (48) and (49) we arrive at the $k+1 \times k+q_\rho$ matrix

$$\mathbf{M}_{XZ,nTi} = \frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\left(\sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt}, \widetilde{\mathbf{x}}_{it}'\right)'\left(\sum_{j=1}^{n} W_{ij}^{\tau_1}\widetilde{\mathbf{x}}_{jt1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}}\widetilde{\mathbf{x}}_{jtq_\rho}, \widetilde{\mathbf{x}}_{it}'\right) \tag{50}$$

$$= \left(\begin{array}{c|c} \mathbf{M}_{XZ,nTi,(1:1,1:q_\rho)} & \mathbf{M}_{XZ,nTi,(1:1,q_\rho+1:k+q_\rho)} \\ \hline \mathbf{M}_{XZ,nTi,(2:k+1,1:q_\rho)} & \mathbf{M}_{XZ,nTi,(2:k+1,q_\rho+1:k+q_\rho)} \end{array}\right),$$

where

$$\mathbf{M}_{XZ,nTi,(1:1,1:q_\rho)} :=$$

$$\frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\left(\begin{array}{c} \sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}K_{jl}\left(\beta'\widetilde{\mathbf{x}}_{lt} + \tilde{u}_{lt}^{\dagger}\right)\cdot \sum_{\kappa=1}^{n} W_{i\kappa}^{\tau_1}\widetilde{x}_{\kappa t1} \\ \vdots \\ \sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}K_{jl}\left(\beta'\widetilde{\mathbf{x}}_{lt} + \tilde{u}_{lt}^{\dagger}\right)\cdot \sum_{j=1}^{n} W_{i\kappa}^{\tau_{q_\rho}}\widetilde{x}_{\kappa tq_\rho} \end{array}\right)',$$

$$\mathbf{M}_{XZ,nTi,(1:1,q_\rho+1:k+q_\rho)} := \frac{1}{T_\star^2}\sum_{t=1}^{T_\star} \sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}K_{jl}\left(\beta'\widetilde{\mathbf{x}}_{lt} + \tilde{u}_{lt}^{\dagger}\right)\cdot \widetilde{\mathbf{x}}_{it}',$$

$$\mathbf{M}_{XZ,nTi,(2:k+1,1:q_\rho)} := \frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\left(\widetilde{\mathbf{x}}_{it}\cdot\sum_{\kappa=1}^{n} W_{i\kappa}^{\tau_1}\widetilde{x}_{\kappa t1}, \ldots, \widetilde{\mathbf{x}}_{it}\cdot\sum_{\kappa=1}^{n} W_{i\kappa}^{\tau_{q_\rho}}\widetilde{x}_{\kappa tq_\rho}\right),$$

$$\mathbf{M}_{XZ,nTi,(2:k+1,q_\rho+1:k+q_\rho)} := \frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it}'.$$

The $T \to \infty$ limit of the matrix $\mathbf{M}_{XZ,nTi}$ is given by:

$$\mathbf{M}_{XZ,ni} = \left(\begin{array}{c|c} \mathbf{M}_{XZ,ni,(1:1,1:q_\rho)} & \mathbf{M}_{XZ,ni,(1:1,q_\rho+1:k+q_\rho)} \\ \hline \mathbf{M}_{XZ,ni,(2:k+1,1:q_\rho)} & \mathbf{M}_{XZ,ni,(2:k+1,q_\rho+1:k+q_\rho)} \end{array} \right) ,$$

where

$$\mathbf{M}_{XZ,nTi,(1:1,1:q_\rho)} \xrightarrow{d} \left(\begin{array}{c} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_1} \sum_{j=1}^n \sum_{l=1}^n W_{ij} K_{jl} \left(\beta' \tilde{\mathcal{B}}_{vl} \tilde{\mathcal{B}}_{v\kappa 1} \right) \\ \vdots \\ \sum_{\kappa=1}^n W_{i\kappa}^{\tau_{q_\rho}} \sum_{j=1}^n \sum_{l=1}^n W_{ij} K_{jl} \left(\beta' \tilde{\mathcal{B}}_{vl} \tilde{\mathcal{B}}_{v\kappa q_\rho} \right) \end{array} \right)' = \mathbf{M}_{XZ,ni,(1:1,1:q_\rho)} , \tag{51}$$

$$\mathbf{M}_{XZ,nTi,(1:1,q_\rho+1+q_\rho+k)} \xrightarrow{d} \left(\begin{array}{l} \sum_{j=1}^n \sum_{l=1}^n W_{ij} K_{jl} \left(\beta' \tilde{\mathcal{B}}_{vl} \tilde{\mathcal{B}}_{vi1} \right) \\ \sum_{j=1}^n \sum_{l=1}^n W_{ij} K_{jl} \left(\beta' \tilde{\mathcal{B}}_{vl} \tilde{\mathcal{B}}_{vi2} \right) \\ \vdots \\ \sum_{j=1}^n \sum_{l=1}^n W_{ij} K_{jl} \left(\beta' \tilde{\mathcal{B}}_{vl} \tilde{\mathcal{B}}_{vik} \right) \end{array} \right)' = M_{XZ,ni,(1:1,q_\rho+1+q_\rho+k)} , \tag{52}$$

$$\mathbf{M}_{XZ,nTi,(2+k+1,1:q_\rho)} \xrightarrow{d} \tag{53}$$

$$\left(\sum_{\kappa=1}^n W_{i\kappa}^{\tau_1} \int \tilde{\mathcal{B}}_{vi} \tilde{\mathcal{B}}_{v\kappa 1}, \ldots, \sum_{\kappa=1}^n W_{i\kappa}^{\tau_{q_\rho}} \int \tilde{\mathcal{B}}_{vi} \tilde{\mathcal{B}}_{v\kappa q_\rho} \right) = \mathbf{M}_{XZ,ni,(2:k+1,1:q_\rho)} , \tag{54}$$

$$\mathbf{M}_{XZ,nTi,(2+k+1,q_\rho+1:q_\rho+k)} \xrightarrow{d}$$

$$\int \tilde{\mathcal{B}}_{vi} \tilde{\mathcal{B}}_{vi}' = M_{XZ,ni,(2:k+1,q_\rho+1:+q_\rho+k)} . \tag{55}$$

When taking limits we also meet terms of the structure described in (42). Since $\frac{1}{T_\star^{3/2}}$ yields terms bounded in probability, with $\frac{1}{T_\star^2}$ these terms converge to zero in probability. Note that $\tilde{\mathcal{B}}_{vj\cdot}$ is a scalar while $\tilde{\mathcal{B}}_{vi}$ a k dimensional vector. Summing up we arrive at

$$\begin{aligned} \mathbf{M}_{XZ,nTi} &\xrightarrow{d} \mathbf{M}_{XZ,ni} , \\ \mathbf{M}_{XZ,nT} = \frac{1}{n} \sum_{i=1}^n \mathbf{M}_{XZ,nTi} &\xrightarrow{d} \mathbf{M}_{XZ,n} = \frac{1}{n} \sum_{i=1}^n \mathbf{M}_{XZ,ni} . \end{aligned} \tag{56}$$

Remark 5. By Assumption 5 we have assumed that the matrix $\mathbf{M}_{XZ,n}$ has rank $k+1$, while the matrix $\mathbf{M}_{ZZ,n}$ has rank $k+q_\rho \geq k+1$. Therefore $\mathbf{M}_{XZ,n}\left(\mathbf{M}_{ZZ,n}\right)^{-1}\mathbf{M}_{ZX,n}$ has rank $k+1$. Lemma 1 shows that this assumption is non-empty. If the conditions of Lemma 1 hold, then the matrices $\mathbf{M}_{XZ,n}$ and $\mathbf{M}_{ZZ,n}$ have rank, $k+1$ and $k+q_\rho$, respectively.

Next we derive $\mathbf{m}_{ni}$ and $\mathbf{m}_n$. For the term $\frac{1}{T_\star}\sum_{t=1}^{T_\star}\mathbf{X}_{it,2:d_x}\underline{u}_{it} = \frac{1}{T_\star}\sum_{t=1}^{T_\star}\tilde{\mathbf{x}}_{it,1:d_x}\underline{u}_{it}$ the $T_\star \to \infty$ limit is already given by $\sqrt{\Omega_{uu,i}}\int\tilde{\mathcal{B}}_{vi}d\mathcal{W}_{ui}$. By using the functional central limit theorem and Saikkonen (1991)[Theorem 4.1/ Lemma A.5] (e.g. $\|\sum_{t=1}^{T_\star}(\mathbf{Z}_{it}\cdot D_T)\mathsf{e}_{it}\|_2 = o_p(p^{1/2})$ such that $\underline{u}_{it} = u_{it}$ as $T \to \infty$) the term $\frac{1}{T_\star}\sum_{t=1}^{T_\star}\mathbf{Z}_{it,1:d_z}\underline{u}_{it}$ converges in distribution to

$$\mathbf{m}_{niZu} = \sqrt{\Omega_{uu,i}}\left(\sum_{j=1}^{n}W_{ij}^{\tau_1}\int\tilde{\mathcal{B}}_{vj1}d\mathcal{W}_{ui},\ldots,\sum_{j=1}^{n}W_{ij}^{\tau_{q_\rho}}\int\tilde{\mathcal{B}}_{vjq_\rho}d\mathcal{W}_{ui},\left(\int\tilde{\mathcal{B}}_{vi}d\mathcal{W}_{ui}\right)'\right)'. \tag{57}$$

To obtain the first term of $\mathbf{m}_n$ we have to combine $\mathbf{M}_{ZZ,n}$ provided by (47), $\mathbf{M}_{XZ,ni}$ given by (56) and $\mathbf{m}_{niZu}$. Then the continuous mapping theorem yields

$$\mathbf{m}_{nT} \xrightarrow{d} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sqrt{\Omega_{uu,i}}\left(\mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}\begin{pmatrix}\sum_{j=1}^{n}W_{ij}^{\tau_1}\int\tilde{\mathcal{B}}_{vj1}d\mathcal{W}_{ui}\\ \vdots\\ \sum_{j=1}^{n}W_{ij}^{\tau_{q_\rho}}\int\tilde{\mathcal{B}}_{vjq_\rho}d\mathcal{W}_{ui}\\ \int\tilde{\mathcal{B}}_{vi}d\mathcal{W}_{ui}\end{pmatrix}\right). \tag{58}$$

$\mathbf{M}_{ZZ,n}$ is a $k+q_\rho \times k+q_\rho$ matrix, while $\mathbf{m}_{niZu}$ as well as

$$\left(\sum_{j=1}^{n}\int W_{ij}^{\tau_1}\tilde{\mathcal{B}}_{vj1}d\mathcal{W}_{ui},\ldots,\sum_{j=1}^{n}W_{ij}^{\tau_{q_\rho}}\int\tilde{\mathcal{B}}_{vjq_\rho}d\mathcal{W}_{ui},\left(\int\tilde{\mathcal{B}}_{vi}d\mathcal{W}_{ui}\right)'\right)'$$

are vectors of dimension $q_\rho+k$. The elements 2 to $k+1$ of $\mathbf{m}_n$ are given by a sum of the k dimensional vectors $\int\tilde{\mathcal{B}}_{vi}d\mathcal{W}_{ui}$. Since the application of the projection operator $\mathbf{P}_H$ on $\mathbf{X}_{i,2:d_x}$ is $\mathbf{X}_{i,2:d_x}$ (see equation (38)), the rows $(2:k+1)$ have to be equal to the limit of $\frac{1}{T_\star}\sum_{t=1}^{T_\star}\mathbf{X}_{it,2:d_x}\underline{u}_{it} = \frac{1}{T_\star}\sum_{t=1}^{T_\star}\mathbf{x}_{it}\underline{u}_{it}$. This yields

$$\mathbf{m}_{nT} \xrightarrow{d} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sqrt{\Omega_{uu,i}}\left(\begin{array}{c}\left[\mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}\begin{pmatrix}\sum_{j=1}^{n}W_{ij}^{\tau_1}\int\tilde{\mathcal{B}}_{vj1}d\mathcal{W}_{ui}\\ \vdots \\ \sum_{j=1}^{n}W_{ij}^{\tau_{q_\rho}}\int\tilde{\mathcal{B}}_{vjq_\rho}d\mathcal{W}_{ui}\\ \int\tilde{\mathcal{B}}_{vi}d\mathcal{W}_{ui}\end{pmatrix}\right]_{(1,1)}\\ \hdashline \int\tilde{\mathcal{B}}_{vi}d\mathcal{W}_{ui}\end{array}\right) = \mathbf{m}_n \ . \tag{59}$$

It remains to calculate the limit distribution of (39). By the asymptotic independence arguments for γ and δ, we are allowed to restrict to $\mathbf{X}'_{1:d_x}\mathbf{Z}_{1:d_z}\left(\mathbf{Z}'_{1:d_z}\mathbf{Z}_{1:d_z}\right)^{-1}\mathbf{Z}'_{1:d_z}\mathbf{X}_{1:d_x}$ (weighted by $1/T_\star$ and $1/\sqrt{n}$). Using the above results and the continuous mapping theorem we get

$$\begin{aligned}\mathbf{M}_{ni} &= \mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}\mathbf{M}'_{XZ,ni},\\ \mathbf{M}_n &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}_{ni} = \frac{1}{n}\mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}\mathbf{M}'_{XZ,n} \ .\end{aligned} \tag{60}$$

This yields the *second result*: $(\gamma', \delta')'$ can be consistently estimated. $\sqrt{n}T_\star\left(\hat{\gamma}_{D2SLS}-\gamma\right)$ converges in distribution to $\mathbf{M}_n^{-1}\mathbf{m}_n$ as $T\to\infty$, where $\mathbf{m}_n$ and $\mathbf{M}_n$ are given by (59) and (60), respectively. With these estimates we can derive the residuals, which allow us to consistently estimate $\Omega_{uu,i}$.

`Step 3`: Finally we construct the Wald statistic $S_{\gamma,n}$. We follow Phillips and Hansen (1990), Johansen (1995) and Park and Phillips (1988) to derive the so called *observed Wald-statistic* $S_{\gamma,nT}$ and its limit $S_{\gamma,n}$. Consider the $s\times k+1$ restriction matrix $\mathbf{R}$. Since S-ancillarity is implied by strong exogeneity as observed in our model, the ancillarity results presented in Johansen (1995) can be used. With $\mathcal{B}_{vi}$ fixed for all $i=1,\ldots,n$: (i) the terms $\mathbf{M}_{ni}$ and $\mathbf{M}_n$ are constant matrices; (ii) $\mathbf{m}_n$ is a mixed Gaussian vector with mean zero and variance $\mathbf{V}_n$ where[11]

$$\mathbf{V}_n = \frac{1}{n}\sum_{i=1} V_{ni} \text{ where } \mathbf{V}_{ni} = \Omega_{uu,i} \cdot \tilde{\Upsilon}_{ni} \ . \tag{61}$$

$$\tilde{\Upsilon}_{ni} = \int \left[\begin{pmatrix} \left[\mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}\mathbf{m}_{Z,ni}\right]_{(1,1)} \\ \tilde{\mathcal{B}}_{vi} \end{pmatrix} \begin{pmatrix} \left[\mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}\mathbf{m}_{Z,ni}\right]_{(1,1)} \\ \tilde{\mathcal{B}}_{vi} \end{pmatrix}' \right] ,$$

$$\mathbf{m}_{Z,n} = \left(\sum_{j=1}^{n} W_{ij}^{\tau_1} \tilde{\mathcal{B}}_{vj1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}} \tilde{\mathcal{B}}_{vjq_\rho}, \tilde{\mathcal{B}}_{vi}' \right)' . \tag{62}$$

The term $\mathbf{m}_{Z,ni}$ is given by

$$\mathbf{m}_{Z,ni} = \begin{pmatrix} \sum_{j=1}^{n} W_{ij}^{\tau_1} \int \tilde{\mathcal{B}}_{vj1} \\ \vdots \\ \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}} \int \tilde{\mathcal{B}}_{vjq_\rho} \\ \int \tilde{\mathcal{B}}_{vi} \end{pmatrix} . \tag{63}$$

Then the asymptotic covariance matrix of $\sqrt{n}T_\star \left(\gamma_{D2SLS} - \gamma\right)$ becomes

$$\mathbf{D}_n = \mathbf{M}_n^{-1}\mathbf{V}_n\mathbf{M}_n^{-1} \ . \tag{64}$$

$\mathbf{V}_{nT}$ provides an estimate of $\mathbf{V}_n$, which is derived by means of means of

$$\mathbf{V}_{nT} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\Omega}_{uu,i} \frac{1}{{T_\star}^2} \sum_{t=1}^{T_\star} \Upsilon_{nTi}\Upsilon_{nTi}'$$

$$\Upsilon_{nTi} = \begin{pmatrix} \left[\left(\sum_{l=1}^{n}\sum_{t=1}^{T_\star} \mathbf{X}_{lt,1:d_x}\mathbf{Z}_{lt,1:d_z}'\right) \left(\sum_{j=1}^{n}\sum_{t=1}^{T_\star} \mathbf{Z}_{jt,1:d_z}\mathbf{Z}_{jt,1:d_z}'\right)^{-1} \mathbf{Z}_{it,1:d_z} \right]_{(1,1)} \\ \mathbf{x}_{it} \end{pmatrix} . \tag{65}$$

[11]Note that as with $\mathbf{m}_{nT}$ the term $\mathbf{M}_{XZ,Tn}\mathbf{M}_{ZZ,Tn}^{-1}\frac{1}{{T_\star}^2}\sum_{t=1}^{T_\star}\left(\sum_{j=1}^{n} W_{ij}^{\tau_1} Z_{it,1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}} Z_{it,q_\rho}, \tilde{\mathbf{x}}_{it}'\right)'$ is equal to $\left(\left[\mathbf{M}_{XZ,Tn}\mathbf{M}_{ZZ,Tn}^{-1}\frac{1}{{T_\star}^2}\sum_{t=1}^{T_\star}\left(\sum_{j=1}^{n} W_{ij}^{\tau_1} Z_{it,1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}} Z_{it,q_\rho}, \tilde{\mathbf{x}}_{it}'\right)'\right]_{(1,1)}, \tilde{\mathbf{x}}_{it}'\right)'$ by the fact the projection $\mathbf{P}_H$ applied to $\tilde{\mathbf{x}}_{it}$ is $\tilde{\mathbf{x}}_{it}$.

Combining (65) and $\mathbf{M}_{nT}$, which is an estimate of $\mathbf{M}_n$, we arrive at an estimate of the covariance matrix

$$\mathbf{D}_{nT} = \mathbf{M}_{nT}^{-1}\mathbf{V}_{nT}\mathbf{M}_{nT}^{-1} \ . \tag{66}$$

Equipped with these terms we obtain

$$\begin{aligned} S_{\gamma,nT} &= \left(\sqrt{n}T_\star \mathbf{R}\left(\widehat{\gamma}_{D2SLS} - \gamma\right)\right)' \left(\mathbf{R}\mathbf{D}_{nT}\mathbf{R}'\right)^{-1} \left(\sqrt{n}T_\star \mathbf{R}\left(\widehat{\gamma}_{D2SLS} - \gamma\right)\right) \ , \\ S_{\gamma,nT} &\overset{d}{\to} S_{\gamma,n} = \left(\sqrt{n}T_\star \mathbf{R}\left(\widehat{\gamma}_{D2SLS} - \gamma\right)\right)' \left(\mathbf{R}\mathbf{D}_{n}\mathbf{R}'\right)^{-1} \left(\sqrt{n}T_\star \mathbf{R}\left(\widehat{\gamma}_{D2SLS} - \gamma\right)\right) \ . \end{aligned} \tag{67}$$

Under the null hypothesis the Wald statistic $S_{\gamma,nT}$ follows a χ^2 distribution with s degrees of freedom. This yields the *third result*: $S_{\gamma,nT} \overset{d}{\to} S_{\gamma,n}$; $\mathbf{D}_{nT}$ provides us with an estimate of the asymptotic covariance of the estimator $\mathbf{D}_n$.

Remark 6 (Identification)**.** From Deistler and Seifert (1978)[Theorem 4] it follows that if there exists a consistent estimator for the internal characteristics C_I of the model, the model is identifiable (i.e. an *identifying function* exists). In our setup the internal characteristics can be described by means of the parameters ρ and β and the covariances of the noise. Based on our Theorem 1 we have obtained a (super) consistent estimator for the parameters ρ, β. The covariance matrices can be estimated consistently due to Jansson (2002). The reasons why $\hat{\gamma}_{D2SLS}$ becomes consistent are Assumption 5 on the instruments (Appendix B shows that this assumption is non-empty, there also Assumption 1 becomes important; in addition Assumption 3 is important to obtain the corresponding ranks of in limits required in Assumption 5) and Assumption 4 such that the truncation error goes to zero.

B Instruments and the Rank Condition

Consider the instruments $\widetilde{x}^*_{itv} = \sum_{j=1}^n W_{ij}^{\tau_v}\widetilde{x}_{jtv}$, $v = 1, \ldots, q_\rho$. In the following we show that the requirement of valid instruments (see also Assumption 5) can be fulfilled under fairly mild restrictions. It is important to note that this implies that our "high level" Assumption 5 is non-empty. For the $\mathbf{W}$ used in the applied part we use the following lemma:

Lemma 1. Given the model Assumptions 1 to 4 of Section 2. Suppose that $\widetilde{x}^*_{itv} = \sum_{j=1}^n W_{ij}^{\tau_v}\widetilde{x}_{jtv}$, $v = 1, \ldots, q_\rho$ and $\tau_v = 1$ and $W_{ij} \neq 0$ for at least one j, $j \neq i$, for each row i. Then the $q_\rho + k \times q_\rho + k$ matrices $\mathbf{M}_{ZZ,ni}$ and $\mathbf{M}_{ZZ,n}$ have rank $q_\rho + k$ almost surely. Additionally, the rank of the $k+1 \times q_\rho + k$ matrices $\mathbf{M}_{XZ,ni}$ and $\mathbf{M}_{XZ,n}$ is $k+1$ almost surely.

Proof. We consider the vectors

$$\mathbf{Z}_{it,1:d_z} = \begin{pmatrix} \sum_{j=1}^n W_{ij}^{\tau_1}\tilde{x}_{jt1} \\ \vdots \\ \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}}\tilde{x}_{jtq_\rho} \\ \tilde{x}_{it1} \\ \vdots \\ \tilde{x}_{itk} \end{pmatrix} \text{ and}$$

$$\mathbf{X}_{it,1:d_x} = \begin{pmatrix} \sum_{j=1}^n W_{ij}\tilde{y}_{jt} \\ \tilde{x}_{it1} \\ \vdots \\ \tilde{x}_{itk} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n \sum_{l=1}^n W_{ij}K_{jl}\left(\beta'\tilde{\mathbf{x}}_{lt} + \delta_i'\boldsymbol{\zeta}_{lt} + \underline{\tilde{u}}_{lt}\right) \\ \tilde{x}_{it1} \\ \vdots \\ \tilde{x}_{itk} \end{pmatrix} \tag{68}$$

of dimension $q_\rho + k$ and $1+k$, respectively. $\sum_{j=1}^n W_{ij}\tilde{y}_{it}$ follows from (49), where $K_{jl} = \left[(\mathbf{I} - \rho\mathbf{W})^{-1}\right]_{(jl)}$. In the following we calculate the limits of $\frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}$, $\frac{1}{n}\sum_{i=1}^n\frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}$, $\frac{1}{T_\star^2}\sum_{t=1}^{T_\star}\mathbf{X}_{it,1:d_x}\mathbf{Z}'_{it,1:d_z}$ and $\frac{1}{T_\star^2 n}\sum_{i=1}^n\sum_{t=1}^{T_\star}\mathbf{X}_{it,1:d_x}\mathbf{Z}'_{it,1:d_z}$. Let us start with $\sum_{t=1}^{T_\star}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}$ where we get

$$\sum_{t=1}^{T_\star} \mathbf{Z}_{it,1:d_z} \mathbf{Z}'_{it,1:d_z} = \tag{69}$$

$$\sum_{t=1}^{T_\star} \left(\begin{array}{ccc|ccc} \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} & \dots & \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} & \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} \tilde{x}_{it1} & \dots & \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} \tilde{x}_{itk} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} & \dots & \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} & \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} \tilde{x}_{it1} & \dots & \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} \tilde{x}_{itk} \\ \hline \tilde{x}_{it1} \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} & \dots & \tilde{x}_{it1} \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} & \tilde{x}_{it1} \tilde{x}_{it1} & \dots & \tilde{x}_{it1} \tilde{x}_{itk} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{itk} \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} & \dots & \tilde{x}_{itk} \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} & \tilde{x}_{itk} \tilde{x}_{it1} & \dots & \tilde{x}_{itk} \tilde{x}_{itk} \end{array} \right).$$

Since $\sum_{j=1}^n W_{ij}^{\tau_v} \sum_{j=1}^n W_{ij}^{\tau_w} \tilde{x}_{jtv} \tilde{x}_{jtw} = \sum_{j=1}^n \sum_{l=1}^n W_{ij}^{\tau_v} W_{il}^{\tau_w} \tilde{x}_{jtv} \tilde{x}_{ltw}$ the matrix (69) can be written as:

$$\sum_{t=1}^{T_\star} \mathbf{Z}_{it,1:d_z} \mathbf{Z}'_{it,1:d_z} = \tag{70}$$

$$\sum_{t=1}^{T_\star} \left(\begin{array}{ccc|ccc} \sum_{j=1}^n \sum_{l=1}^n W_{ij}^{\tau_1} W_{il}^{\tau_1} \tilde{x}_{jt1} \tilde{x}_{lt1} & \dots & \sum_{j=1}^n \sum_{l=1}^n W_{ij}^{\tau_1} W_{il}^{\tau_{q_\rho}} \tilde{x}_{jt1} \tilde{x}_{ltq_\rho} & \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} \tilde{x}_{it1} & \dots & \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} \tilde{x}_{itk} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^n \sum_{l=1}^n W_{ij}^{\tau_{q_\rho}} W_{il}^{\tau_1} \tilde{x}_{jtq_\rho} \tilde{x}_{lt1} & \dots & \sum_{j=1}^n \sum_{l=1}^n W_{ij}^{\tau_{q_\rho}} W_{il}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} \tilde{x}_{ltq_\rho} & \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} \tilde{x}_{it1} & \dots & \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} \tilde{x}_{itk} \\ \hline \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} \tilde{x}_{it1} & \dots & \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} \tilde{x}_{it1} & \tilde{x}_{it1} \tilde{x}_{it1} & \dots & \tilde{x}_{it1} \tilde{x}_{itk} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^n W_{ij}^{\tau_1} \tilde{x}_{jt1} \tilde{x}_{itk} & \dots & \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \tilde{x}_{jtq_\rho} \tilde{x}_{itk} & \tilde{x}_{itk} \tilde{x}_{it1} & \dots & \tilde{x}_{itk} \tilde{x}_{itk} \end{array} \right).$$

(70) is a symmetric matrix. Now we set $\tau_v = 1$ and take the limit of "(70) divided by ${T_\star}^2$". With $\widetilde{x}^*_{itv} = \sum_{j=1}^n W_{ij}^{\tau_v} \widetilde{x}_{jtv}$, $v = 1, \ldots, q_\rho$ and $\tau_v = 1$, we observe that $q_\rho \leq k$. By the functional central limit theorem we obtain

$$\mathbf{M}_{ZZ,ni} = \lim_{t\to\infty} \frac{1}{{T_\star}^2} \sum_{t=1}^{T_\star} \mathbf{Z}_{it,1:d_z} \mathbf{Z}'_{it,1:d_z} = \tag{71}$$

$$\left(\begin{array}{ccc|ccc}
\int \sum_{j=1}^n \sum_{l=1}^n W_{ij} W_{il} \tilde{\mathcal{B}}_{vj1} \tilde{\mathcal{B}}_{vl1} & \cdots & \int \sum_{j=1}^n \sum_{l=1}^n W_{ij} W_{il} \tilde{\mathcal{B}}_{vj1} \tilde{\mathcal{B}}_{vlq_\rho} & \int \sum_{j=1}^n W_{ij} \tilde{\mathcal{B}}_{vj1} \tilde{\mathcal{B}}_{vi1} & \cdots & \int \sum_{j=1}^n W_{ij} \tilde{\mathcal{B}}_{vj1} \tilde{\mathcal{B}}_{vik} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\int \sum_{j=1}^n \sum_{l=1}^n W_{ij} W_{il} \tilde{\mathcal{B}}_{vjq_\rho} \tilde{\mathcal{B}}_{vl1} & \cdots & \int \sum_{j=1}^n \sum_{l=1}^n W_{ij} W_{il} \tilde{\mathcal{B}}_{vjq_\rho} \tilde{\mathcal{B}}_{vlq_\rho} & \int \sum_{j=1}^n W_{ij} \tilde{\mathcal{B}}_{vjq_\rho} \tilde{\mathcal{B}}_{vi1} & \cdots & \int \sum_{j=1}^n W_{ij} \tilde{\mathcal{B}}_{vjq_\rho} \tilde{\mathcal{B}}_{vik} \\
\hline
\int \sum_{j=1}^n W_{ij} \tilde{\mathcal{B}}_{vj1} \tilde{\mathcal{B}}_{vi1} & \cdots & \int \sum_{j=1}^n W_{ij} \tilde{\mathcal{B}}_{vjq_\rho} \tilde{\mathcal{B}}_{vi1} & \int \tilde{\mathcal{B}}_{vi1} \tilde{\mathcal{B}}_{vi1} & \cdots & \int \tilde{\mathcal{B}}_{vi1} \tilde{\mathcal{B}}_{vi1} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\int \sum_{j=1}^n W_{ij} \tilde{\mathcal{B}}_{vj1} \tilde{\mathcal{B}}_{vik} & \cdots & \int \sum_{j=1}^n W_{ij} \tilde{\mathcal{B}}_{vjq_\rho} \tilde{\mathcal{B}}_{vik} & \int \tilde{\mathcal{B}}_{vik} \tilde{\mathcal{B}}_{vi1} & \cdots & \int \tilde{\mathcal{B}}_{vik} \tilde{\mathcal{B}}_{vik}
\end{array}\right).$$

When we consider (70) we observe that for any fixed t the row v is a linear combination of $\mathbf{Z}'_{it,1:d_z}$ with the element $\left[\mathbf{Z}_{it,1:d_z}\right]_v$. This also translates to the limit (71), where we observe that in each row each element includes a term arising from $\left[\mathbf{Z}_{it,1:d_z}\right]_v$. E.g. $\sum_{j=1}^n W_{ij} \tilde{\mathcal{B}}_{vj1}$ for the first row, $\sum_{j=1}^n W_{ij} \tilde{\mathcal{B}}_{vjq_\rho}$ for row q_ρ, $\tilde{\mathcal{B}}_{vi1}$ for row $q_\rho + 1$, ...and $\tilde{\mathcal{B}}_{vik}$ for row $d_z = q_\rho + k$. The important ingredients to have a matrix $\mathbf{M}_{ZZ,ni}$ of full rank are (i) Assumption 3 (based on Phillips (2006)) to consider terms where $i = j$, (ii) Phillips and Hansen (1990)[Lemma A.3] to consider terms where $i \neq j$ and (iii) the assumption that $W_{ij} \neq 0$. By $W_{ij} \neq 0$ the terms in the first q_ρ rows and the first q_ρ columns of (71) are non-zero almost surely. Note that by Assumption 3 the matrix $\int \tilde{\mathcal{B}}_{vi} \tilde{\mathcal{B}}'_{vi}$ is positive definite and of full rank k. By Phillips and Hansen (1990)[Lemma A.3] the $k \times k$ matrix $\int \tilde{\mathcal{B}}_{vi} \tilde{\mathcal{B}}'_{vj}$ is also positive definite and of rank k (almost surely).[12] By our assumption (in Lemma 1) that $\mathbf{W}_{i,1:n}$ is not equal to a row vector of zeros, we meet a sum of matrics $\int \tilde{\mathcal{B}}_{vj} \tilde{\mathcal{B}}'_{vl}$ weighted by $W_{ij} W_{il}$ in the north-western part of the matrix $\mathbf{M}_{ZZ,ni}$ in (71). $W_{ij} \neq 0$ and $W_{il} \neq 0$ for some $j, l \neq i$ by Assumption 1; note that $W_{ij} W_{il} \neq 0$ for $j = l$. For each pair j and l the term $\int \tilde{\mathcal{B}}_{vj,1:q_\rho} \tilde{\mathcal{B}}'_{vl,1:q_\rho}$ has rank q_ρ by Assumption 3 or Phillips and Hansen (1990)[Lemma A.3]. Additionally, these terms are positive definite. Therefore, the term $\int \sum_{j=1}^n \sum_{l=1}^n W_{ij} W_{il} \tilde{\mathcal{B}}_{vj,1:q_\rho} \tilde{\mathcal{B}}'_{vl,1:q_\rho}$ has rank q_ρ (see also Horn and Johnson (1985)[Obs. 7.1.3, page 398]). The south-eastern part has rank k by Assumption 3. The south-western part and the north-eastern part mix $\tilde{\mathcal{B}}_{vl,1:q_\rho}$ with $\tilde{\mathcal{B}}'_{vi}$. Since $q_\rho \leq k$ the ranks of these matrices are $\leq k$. Since the rank of the $k + q_\rho \times q_\rho$ submatrix $[\mathbf{M}_{ZZ,ni}]_{(1:k+q_\rho,1:q_\rho)}$ is $\leq k$, there exist some scalars $\lambda_{\ell 1}, \ldots, \lambda_{\ell k}$ to express the rows ℓ, $\ell = 1, \ldots, q_\rho$, by means of a linear combination

[12] Given the notation in Phillips and Hansen (1990)[Lemma A.3]. Our $\mathbf{v}_i$ corresponds to Δx_2, while $\mathbf{v}_j$ corresponds to Δx_3. If $i = j$ in $\int \tilde{\mathcal{B}}_{vi} \tilde{\mathcal{B}}'_{vj}$, then the full rank process assumption is used, with $i \neq j$ the Lemma A.3 can be applied.

of the other k rows. I.e. in the north-western part of the matrix $\mathbf{M}_{ZZ,ni}$ we meet mixtures of components $\tilde{\mathcal{B}}_{vj}$ with components of $\tilde{\mathcal{B}}_{vl}$, $l = 1, \ldots, n$ where $l \neq i$ ($W_{ii} = 0$ by Assumption 1 while some $W_{il} \neq 0$ as assumed above). In the north-eastern part of the matrix $\mathbf{M}_{ZZ,ni}$ we meet mixtures of the components $\tilde{\mathcal{B}}_{vj}$ with the components of $\tilde{\mathcal{B}}_{vi}$. $\tilde{\mathcal{B}}_{vi} \neq \tilde{\mathcal{B}}_{vl}$ (almost surely) by the properties of Brownian motion. Therefore the corresponding rows of the north-eastern submatrix cannot be reconstructed by the mixture weights $\lambda_{\ell 1}, \ldots, \lambda_{\ell k}$. Note that in each of the rows $1, \ldots, q_\rho$ we meet different mixtures, from $\tilde{\mathcal{B}}_{vj1}$ in the first row to $\tilde{\mathcal{B}}_{vjq_\rho}$ in the q_ρth row. That is to say we cannot rebuild rows ℓ, $\ell \in \{1, \ldots, q_\rho\}$ from the other rows. Since the matrix is symmetric this also holds for the columns. Summing up, the rank of the $k+q_\rho \times k+q_\rho$ matrix $\mathbf{M}_{ZZ,ni}$ is $k+q_\rho$ (almost surely).

The limit of $\frac{1}{n}\sum_{i=1}^{n} \frac{1}{T_\star{}^2} \sum_{t=1}^{T_\star} \mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}$ still has rank $q_\rho + k$. In the same way as for some $\mathbf{M}_{ZZ,ni}$, we cannot reconstruct the columns/rows from the other columns/rows when the sum $\frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{M}_{ZZ,ni}$ is considered.

Remark 7. Note that $\mathbf{M}_{ZZ,nTi}$, $\mathbf{M}_{ZZ,nT}$, $\mathbf{M}_{ZZ,ni}$ are $\mathbf{M}_{ZZ,n}$ symmetric matrices.

In the next step we investigate $\mathbf{M}_{XZ,nTi}$. Hence we consider

$$\sum_{t=1}^{T_\star} \mathbf{X}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z} = \tag{72}$$
$$\sum_{t=1}^{T_\star} \left(\begin{array}{ccc|ccc} \sum_{j=1}^{n} W_{ij}\tilde{y}_{jt} \sum_{\kappa=1}^{n} W_{i\kappa}^{\tau_1}\tilde{x}_{\kappa t1} & \cdots & \sum_{j=1}^{n} W_{ij}\tilde{y}_{jt} \sum_{\kappa=1}^{n} W_{i\kappa}^{\tau_{q_\rho}}\tilde{x}_{\kappa t q_\rho} & \sum_{j=1}^{n} W_{ij}\tilde{y}_{jt}\tilde{x}_{it1} & \cdots & \sum_{j=1}^{n} W_{ij}\tilde{y}_{jt}\tilde{x}_{itk} \\ \hline \tilde{x}_{it1}\sum_{\kappa=1}^{n} W_{i\kappa}^{\tau_1}\tilde{x}_{\kappa t1} & \cdots & \tilde{x}_{it1}\sum_{\kappa=1}^{n} W_{i\kappa}^{\tau_{q_\rho}}\tilde{x}_{\kappa t q_\rho} & \tilde{x}_{it1}\tilde{x}_{it1} & \cdots & \tilde{x}_{it1}\tilde{x}_{itk} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{itk}\sum_{\kappa=1}^{n} W_{i\kappa}^{\tau_1}\tilde{x}_{\kappa t1} & \cdots & \tilde{x}_{itk}\sum_{\kappa=1}^{n} W_{i\kappa}^{\tau_{q_\rho}}\tilde{x}_{\kappa t q_\rho} & \tilde{x}_{itk}\tilde{x}_{it1} & \cdots & \tilde{x}_{itk}\tilde{x}_{itk} \end{array} \right).$$

Since $\sum_{j=1}^{n} W_{ij}\tilde{y}_{jt} \sum_{\kappa=1}^{n} W_{i\kappa}^{\tau_w}\tilde{x}_{\kappa tw} = \sum_{j=1}^{n}\sum_{\kappa=1}^{n} W_{ij}W_{i\kappa}^{\tau_w}\tilde{y}_{jt}\tilde{x}_{\kappa tw}$ the matrix (72) can be written as:

$$\sum_{t=1}^{T_\star} \mathbf{X}_{it,1:d_z} \mathbf{Z}'_{it,1:d_z} = \tag{73}$$

$$\sum_{t=1}^{T_\star} \left(\begin{array}{ccc|ccc} \sum_{j=1}^n \sum_{\kappa=1}^n W_{ij} W_{i\kappa}^{\tau_1} \tilde{y}_{jt} \tilde{x}_{\kappa t1} & \dots & \sum_{j=1}^n \sum_{\kappa=1}^n W_{ij} W_{i\kappa}^{\tau_{q_\rho}} \tilde{y}_{jt} \tilde{x}_{\kappa t q_\rho} & \sum_{j=1}^n W_{ij} \tilde{x}_{jt} \tilde{x}_{it1} & \dots & \sum_{j=1}^n W_{ij} \tilde{y}_{jt} \tilde{x}_{itk} \\ \hline \tilde{x}_{it1} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_1} \tilde{x}_{\kappa t1} & \dots & \tilde{x}_{it1} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_{q_\rho}} \tilde{x}_{\kappa t, q_\rho} & \tilde{x}_{it1} \tilde{x}_{it1} & \dots & \tilde{x}_{it1} \tilde{x}_{itk} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{itk} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_1} \tilde{x}_{\kappa t1} & \dots & \tilde{x}_{itk} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_{q_\rho}} \tilde{x}_{\kappa t, q_\rho} & \tilde{x}_{itk} \tilde{x}_{it1} & \dots & \tilde{x}_{itk} \tilde{x}_{itk} \end{array} \right).$$

The limit of the sum of "(73) divided by ${T_\star}^2$" provides us with

$$\mathbf{M}_{XZ,ni} = \lim_{t \to \infty} \frac{1}{{T_\star}^2} \sum_{t=1}^{T_\star} \mathbf{X}_{it,1:d_x} \mathbf{Z}'_{it,1:d_z} = \tag{74}$$

$$\left(\begin{array}{ccc|ccc} \int \sum_{j=1}^n \sum_{l=1}^n W_{ij} K_{jl} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_1} \beta' \tilde{\mathcal{B}}_{vl} \tilde{\mathcal{B}}_{v\kappa 1} & \dots & \int \sum_{j=1}^n \sum_{l=1}^n W_{ij} K_{jl} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_{q_\rho}} \beta' \tilde{\mathcal{B}}_{vl} \tilde{\mathcal{B}}_{v\kappa q_\rho} & \int \sum_{j=1}^n \sum_{l=1}^n W_{ij} K_{jl} \beta' \tilde{\mathcal{B}}_{vl} \tilde{\mathcal{B}}_{vi1} & \dots & \int \sum_{j=1}^n \sum_{l=1}^n W_{ij} K_{jl} \beta' \tilde{\mathcal{B}}_{vl} \tilde{\mathcal{B}}_{vik} \\ \hline \int \tilde{\mathcal{B}}_{vi1} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_1} \tilde{\mathcal{B}}_{v\kappa 1} & \dots & \int \tilde{\mathcal{B}}_{vi1} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_{q_\rho}} \tilde{\mathcal{B}}_{v\kappa q_\rho} & \int \tilde{\mathcal{B}}_{vi1} \tilde{\mathcal{B}}_{vi1} & \dots & \int \tilde{\mathcal{B}}_{vi1} \tilde{\mathcal{B}}_{vik} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \int \tilde{\mathcal{B}}_{vjk} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_1} \tilde{\mathcal{B}}_{v\kappa 1} & \dots & \int \tilde{\mathcal{B}}_{vik} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_{q_\rho}} \tilde{\mathcal{B}}_{v\kappa q_\rho} & \int \tilde{\mathcal{B}}_{vik} \tilde{\mathcal{B}}_{vi1} & \dots & \int \tilde{\mathcal{B}}_{vik} \tilde{\mathcal{B}}_{vik} \end{array} \right).$$

Since $W_{i,1:n} \neq \mathbf{0}_{1\times n}$ (which is assumed in Lemma 1) and $\beta \neq \mathbf{0}_k$ (which follows from Assumption 3) the north-western term in (74) is not equal to zero, this also follows for $[\mathbf{M}_{XZ,ni}]_{(1,q_\rho+1:k+q_\rho)}$ and $[\mathbf{M}_{XZ,ni}]_{(2:2+k,1:q_\rho)}$. The rank of these matrices has to be $\leq k$. I.e. we can write the first row of the submatrix $[\mathbf{M}_{XZ,ni}]_{(1,q_\rho+1:k+q_\rho)}$ as superposition of the remaining k rows of this submatrix. Due to the fact that $W_{ij} \neq 0$ for some j, we observe that the first row of the matrix $\mathbf{M}_{XZ,ni}$ cannot be expressed by the other rows (we also meet the index i in the remaining submatrix). Therefore, the matrix (74) has rank $k+1$ almost surely.

Last but not least by considering the limit of the sum $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{{T_\star}^2}\sum_{t=1}^{T_\star}\mathbf{X}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}$ we obtain $\mathbf{M}_{XZ,n} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}_{XZ,ni}$. By the same reasoning as with $\mathbf{M}_{XZ,ni}$ we observe that this sum of matrices has rank $k+1$. $\square$

Remark 8. Note that $\mathbf{M}_{ZX,nTi} = \mathbf{M}'_{XZ,nTi}$, $\mathbf{M}_{ZX,ni} = \mathbf{M}'_{XZ,ni}$, $\mathbf{M}_{ZX,nT} = \mathbf{M}'_{XZ,nT}$ and $\mathbf{M}_{ZX,n} = \mathbf{M}'_{XZ,n}$.

If necessary the conditions of Lemma 1 can be replaced by less stringent assumptions. To keep the notation simple and by the fact that $\tau_v = 1$ in the applied part, Lemma 1 assumes $\tau_v = 1$. One way to extend Lemma 1 is as follows:

Lemma 2. Given the model Assumptions 1 to 4 of Section 2 and the instruments $\widetilde{x}^*_{itv} = \sum_{j=1}^{n} W_{ij}^{\tau_v}\widetilde{x}_{jtv}$, $v = 1,\ldots,q_\rho$ and $q_\rho \leq k$. Suppose that $W_{ij}^{\tau_v} \neq 0$ for at least one j where $j \neq i$. Then the $q_\rho + k \times q_\rho + k$ matrix $\mathbf{M}_{ZZ,n}$ has rank $q_\rho + k$ almost surely. Additionally, the rank of the $k+1 \times q_\rho + k$ matrix $\mathbf{M}_{XZ,n}$ is $k+1$ almost surely.

Proof. By taking the sum over the components $i = 1,\ldots,n$ of the limits of "(70) divided by ${T_\star}^2 n$" we get $\mathbf{M}_{ZZ,n}$. $\mathbf{M}_{ZZ,n}$ is equal to

$$\mathbf{M}_{ZZ,n} = \frac{1}{n}\sum_{i=1}^{n}\lim_{t\to\infty}\frac{1}{{T_\star}^2}\sum_{t=1}^{T_\star}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z} = \left(\begin{array}{c|c} [\mathbf{M}_{ZZ,ni}]_{(1:q_\rho,1:q_\rho)} & [\mathbf{M}_{ZZ,ni}]_{(1:q_\rho,q_\rho+1:k+q_\rho)} \\ \hline [\mathbf{M}_{ZZ,ni}]_{(q_\rho+1:k+q_\rho,1:q_\rho)} & [\mathbf{M}_{ZZ,ni}]_{(q_\rho+1:k+q_\rho,q_\rho+1:k+q_\rho)} \end{array}\right)$$

where

$$[\mathbf{M}_{ZZ,ni}]_{(1:q_\rho,1:q_\rho)} =$$

$$\frac{1}{n}\cdot\begin{pmatrix} \int\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n W_{ij}^{\tau_1}W_{il}^{\tau_1}\tilde{\mathcal{B}}_{vj1}\tilde{\mathcal{B}}_{vl1} & \dots & \int\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n W_{ij}^{\tau_1}W_{il}^{\tau_{q_\rho}}\tilde{\mathcal{B}}_{vj1}\tilde{\mathcal{B}}_{vlq_\rho} \\ \vdots & \ddots & \vdots \\ \int\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n W_{ij}^{\tau_{q_\rho}}W_{il}^{\tau_1}\tilde{\mathcal{B}}_{vjq_\rho}\tilde{\mathcal{B}}_{vl1} & \dots & \int\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n W_{ij}^{\tau_{q_\rho}}W_{il}^{\tau_{q_\rho}}\tilde{\mathcal{B}}_{vjq_\rho}\tilde{\mathcal{B}}_{vlq_\rho} \end{pmatrix},$$

$$[\mathbf{M}_{ZZ,ni}]_{(q_\rho+1:k+q_\rho,1:q_\rho)} = \frac{1}{n}\cdot\begin{pmatrix} \int\sum_{i=1}^n\tilde{\mathcal{B}}_{vi1}\sum_{l=1}^n W_{il}^{\tau_1}\tilde{\mathcal{B}}_{vl1} & \dots & \int\sum_{i=1}^n\tilde{\mathcal{B}}_{vi1}\sum_{l=1}^n W_{il}^{\tau_{q_\rho}}\tilde{\mathcal{B}}_{vlq_\rho} \\ \vdots & \ddots & \vdots \\ \int\sum_{i=1}^n\tilde{\mathcal{B}}_{vik}\sum_{l=1}^n W_{il}^{\tau_1}\tilde{\mathcal{B}}_{vl1} & \dots & \int\sum_{i=1}^n\tilde{\mathcal{B}}_{vik}\sum_{l=1}^n W_{il}^{\tau_{q_\rho}}\tilde{\mathcal{B}}_{vlq_\rho} \end{pmatrix},$$

$$[\mathbf{M}_{ZZ,ni}]_{(1:q_\rho,q_\rho+1:k+q_\rho)} = \frac{1}{n}\cdot\begin{pmatrix} \int\sum_{i=1}^n\sum_{j=1}^n W_{ij}^{\tau_1}\tilde{\mathcal{B}}_{vj1}\tilde{\mathcal{B}}_{vi1} & \dots & \int\sum_{i=1}^n\sum_{j=1}^n W_{ij}^{\tau_1}\tilde{\mathcal{B}}_{vj1}\tilde{\mathcal{B}}_{vik} \\ \vdots & \ddots & \vdots \\ \int\sum_{i=1}^n\sum_{j=1}^n W_{ij}^{\tau_{q_\rho}}\tilde{\mathcal{B}}_{vjq_\rho}\tilde{\mathcal{B}}_{vi1} & \dots & \int\sum_{i=1}^n\sum_{j=1}^n W_{ij}^{\tau_{q_\rho}}\tilde{\mathcal{B}}_{vjq_\rho}\tilde{\mathcal{B}}_{vik} \end{pmatrix},$$

$$[\mathbf{M}_{ZZ,ni}]_{(q_\rho+1:k+q_\rho,q_\rho+1:k+q_\rho)} = \frac{1}{n}\cdot\int\sum_{i=1}^n\tilde{\mathcal{B}}_{vi}\tilde{\mathcal{B}}'_{vi}\,. \tag{75}$$

First, the south-eastern part $[\mathbf{M}_{ZZ,ni}]_{(q_\rho+1:k+q_\rho,q_\rho+1:k+q_\rho)}$ is a sum of positive definite matrices of rank k (see Assumption 3), by Assumption 2 these matrices are also independent. Therefore this block is of rank k almost surely. Second, we consider the submatrix $[\mathbf{M}_{ZZ,ni}]_{(1:q_\rho+k,1:q_\rho)}$ where we meet the terms $\tilde{\mathcal{B}}_{vi.}\tilde{\mathcal{B}}_{vj.}$, possibly with $i=j$. By the assumptions of this lemma at least one term $W_{ij}^{\tau_v}\neq 0$ with $i\neq j$. Therefore $[\mathbf{M}_{ZZ,ni}]_{(1:q_\rho+k,1:q_\rho)}\neq\mathbf{0}_{q_\rho+k,q_\rho}$. Since $q_\rho\leq k$ the rank of the submatrix $[\mathbf{M}_{ZZ,ni}]_{(1:q_\rho+k,1:q_\rho)}$ is $\leq\min\{k,q_\rho\}$ and therefore smaller or equal to k. I.e. we can express row $\ell\in\{1,\dots,q_\rho\}$ by some linear combination of the rows $q_\rho+1,\dots,k+q_\rho$ with weights $\lambda_{\ell 1},\dots,\lambda_{\ell k}$. Now we take these weights and the submatrix $[\mathbf{M}_{ZZ,ni}]_{(1:q_\rho+k,q_\rho+1:k+q_\rho)}$. Let us try to express row ℓ by means of a linear combination of the weights $\lambda_{\ell 1},\dots,\lambda_{\ell k}$ and the rows $q_\rho+1$ to $q_\rho+k$ of this submatrix. This would require to express terms of the structure $\tilde{\mathcal{B}}_{vl.}\tilde{\mathcal{B}}_{vj.}$, where $j,l\neq i$ for at least one summand, by means of the terms $\tilde{\mathcal{B}}_{vi.}\tilde{\mathcal{B}}_{vi.}$ which is not possible by the properties of the Brownian motion. In other words since $\tilde{\mathcal{B}}_{vi}\neq\tilde{\mathcal{B}}_{vj}$ (almost surely), this cannot work for row ℓ. In addition we observe that we mix with $\tilde{\mathcal{B}}_{vj1}$ in the first row, with $\tilde{\mathcal{B}}_{vj2}$ in the second row until $\tilde{\mathcal{B}}_{vjq_\rho}$ in row q_ρ. Therefore the argument that we cannot express row $\ell \in \{1, \ldots, q_\rho\}$ by some linear combination of the rows $q_\rho + 1, \ldots, k + q_\rho$ holds for all $\ell \in \{1, \ldots, q_\rho\}$. Since $\mathbf{M}_{ZZ,nT}$ is symmetric this also holds for the columns. Hence the rank of the $k + q_\rho \times k + q_\rho$ matrix $\mathbf{M}_{ZZ,n}$ is $k + q_\rho$ (almost surely).

If $W_{ij}^{\tau_v} \neq 0$ for some diagonal element(s), then $K_{ij} = \left[(\mathbf{I} - \rho\mathbf{W})^{-1}\right]_{(ij)}$ has some non-zero off diagonal element(s) as well. $\beta \neq \mathbf{0}_k$ makes the terms $\beta'\tilde{\mathcal{B}}_{vj}$ non-zero (almost surely). Thus, in the same way we have shown that the rank of $\mathbf{M}_{ZZ,n}$ is $k + q_\rho$, we can show that the rank of the $k + 1 \times k + q_\rho$ matrix $\mathbf{M}_{ZZ,n}$ is $k + 1$ (almost surely). □

C Data

CDS Data: We use the CDS dataset already used in Schneider et al. (2010), which was obtained from the *Markit Group*. After concentrating on the US market only and by excluding firms with a too large percentage of missing values, 278 firms had been used. The data set also includes the beginning of the financial crises.

Firm specific and industry data: To estimate a model, where the default probabilities are driven by firm and industry factors, the following data has been downloaded from *Thomson Datastream* and *Compustat*: (i) Share prices p_{it} (in US\$) and the number of shares $NumS_{it}$. The Value of preferred stock PS_{it}, where quarterly records are available. To get weekly data we follow literature and perform linear interpolations. 34 of 176 companies issued preferred stock. In this article we assign preferred stock to equity. Since PS_{it} is small compared to debt and the remaining equity, the impact of the assignment to equity is of minor importance, with both the debt to value ratio and the distance to default, respectively. (ii) Short term (SD_{it}) and long term debt (LD_{it}), quarterly records. To get weekly data, we follow literature and perform linear interpolations. As mentioned in Section 5.1, matching data form these different data sources provides us with 176 firms.

In addition the following data was collected: (iii) US treasury yields $r_{\cdot t}$ for the maturities 1, 2, 3, 5, 7, 10 and 30 years (in percentage terms). (iv) Data of the VIX index which is a volatility index obtained from implied Black-Scholes volatilities from the US stock market (for a description see `http://www.cboe.com/micro/VIX/vixintro.aspx`). (v) NAICs industry classification codes. (vi) Standard and Poors (S&P) ratings. (vii) Input-Output data from the BLS Employment Projection Program (`http://www.bls.gov/emp/ep_data_input_output_matrix.htm`). We excluded firms where we detected further data problems (e.g. extreme spikes, missing values, unclear industry), such that $n = 148$ firms were still remaining.

From the above balance sheet and stock market data we calculate the *debt to value ratio* measured in percentage terms:

$$DVR_{it} = \left[\frac{D_{it}}{S_{it} + D_{it}}\right] \cdot 100 \ , \tag{76}$$

where $S_{it} = p_{it}NumS_{it} + PS_{it}$ is the *market capitalization* and $D_{it} = SD_{it} + LD_{it}$ is the market *value of a firm's debt.* As usual in industry and applied academic research we assume that the market value of a firm's debt is equal to the corresponding book value available in the firm's balance sheets.

In Merton type models and in the financial industry the distance to default is frequently used to forecast the conditional probability of default (see e.g. Schönbucher (2003)). Intuitively, the distance to default is the number of standard deviations of the annual asset growth by which the firm's expected assets at a given maturity exceed a measure of book liabilities. The distance to default is usually derived by a calibration procedure that matches both market value of equity and equity volatility to the figures that can be observed in the market (for details see Crosbie and Bohn (2003)). In this paper the distance to default is derived from

$$DD_{it} = \frac{VA_{it} - DP_{it}}{VA_{it}\sigma_{Ait}} \ . \tag{77}$$

VA_{it} is the firm value. The *default point* DP_{it} is the sum of short-term liabilities +1/2 long-term liabilities, i.e. $DP_{it} = SD_{it} + 1/2LD_{it}$. σ_{Ait} is the standard deviation of the firm value; σ_{Eit} is a measure of the equity volatility. Based on Crosbie and Bohn (2003):

$$\begin{aligned}
VA_{it} &= VE_{it}\mathcal{N}(d_{1it}) + \exp(-y_{tm}M)(SD_{it} + LD_{it})\mathcal{N}(d_{2it}) \ , \\
\sigma_{Ait} &= \sigma_{Eit}\frac{VE_{it}}{VA_{it}} \ , \\
d_{1it} &= \frac{\log\left(VA_{it}/(SD_{it} + LD_{it})\right) + \left(y_{tm} + \frac{1}{2}\sigma^2_{Ait}\right)M}{\sqrt{\sigma^2_{Ait}M}} \ , \\
d_{2it} &= d_{1it} - \sqrt{\sigma^2_{Ait}M} \ .
\end{aligned} \tag{78}$$

The standard deviation of the firm value, σ_{Ait}, is derived by an implicit estimation from the Black/Scholes formula. We derived estimates of VA_{it} and σ_{Ait} by minimizing a weighted sum of the squared distances between the model implied value of equity, VE_{it}, and the market capitalization S_{it}, and the terms $\sigma_{Ait}VA_{it}$ and $\sigma_{Eit}VE_{it}$, respectively. Following industry praxis we set $M = 1$ and y_{tm} equal to the one year treasury yield r_{1t}. We have to point out that the minimization strongly depends on how all these values are scaled. σ^2_{Eit} is estimated from log asset returns. Here e.g. the sample variance $\hat{\sigma}^2_{iE}$ (resulting in a constant equity volatility σ_{Eit} over time t) can be used. In this article we follow Ericsson et al. (2009) and approximate the equity volatility by means of exponential smoothing, where

$$\hat{\sigma}^2_{Eit} = \lambda \hat{\sigma}^2_{Eit-1} + (1-\lambda)(\Delta \log p_{it})^2 \tag{79}$$

with $\lambda = 0.94$. This $\hat{\sigma}^2_{Eit}$ has been used in the calculation of the distance to default.

References

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data.* Whiley, Chichester, UK, 4 edition.

Berndt, A., Douglas, R., Duffie, D., Ferguson, M., and Schranz, D. (2008). Measuring default risk premia form default swap rates and EDFs. Working paper, Stanford University.

Binder, M., Hsiao, C., and Pesaran, M. H. (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory*, 21(04):795–837.

Campbell, J. Y. and Taksler, G. B. (2003). Equity volatility and corporate bond yields. *Journal of Finance*, 58(6):2321–2350.

Cliff, A. and Ord, J. (1973). *Spatial Autocorrelation.* Pion.

Collin-Dufresne, P., Goldstein, R. S., and Helwege, J. (2003). Is credit event risk priced? Modeling contagion via the updating of beliefs. Working paper, Carnegie Mellon University, Washington University and Ohio State University.

Collin-Dufresne, P., Goldstein, R. S., and Martin, J. S. (2001). The determinants of credit spread changes. *Journal of Finance*, 56(6):2177–2207.

Crosbie, P. and Bohn, J. (2003). Modeling default risk - Modeling methodology. *KMV corporation.*

Davidson, J. (1994). *Stochastic Limit Theory - An Introduction for Econometricians.* Oxford University Press, New York.

Deistler, M. and Seifert, H.-G. (1978). Identifiability and consistent estimability in econometric models. *Econometrica*, 46(4):969–980.

ECB (2009). Credit default swaps and counterparty risk. Technical report, European Central Bank, Financial Stability and Supervision.

Eom, Y. H., Helwege, J., and Huang, J.-Z. (2004). Structural models of corporate bond pricing: An empirical analysis. *Review of Financial Studies*, 17(2):499–544.

Ericsson, J., Jacobs, K., and Oviedo, R. A. (2009). The determinants of credit default swap premia. *Journal of Financial and Quantitative Analysis*, 44:109–132.

Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis.* Cambridge University Press, New York.

Hull, J. C. (2006). *Options, Futures and Other Derivatives.* Prentice Hall, Pearson Education, Upper Saddle River, New Jersey, 6 edition.

Hull, J. C., Predescu, M., and White, A. (2004). The relationship between credit default swap spreads, bond yields, and credit rating announcements. *Journal of Banking and Finance*, 28:2789–2811.

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory*, 18(06):1449–1459.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press, Oxford.

Jorion, P. and Zhang, G. (2007). Good and bad credit contagion: Evidence from credit default swaps. *Journal of Financial Economics*, 84(3):860–883.

Kao, C. and Chiang, M. (2000). On the estimation and inference of a cointegrated regression in panel data. In Baltagi, B., editor, *Nonstationary Panels, Panel Cointegration, and Dynamic Panels.* Elsevier.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140:97–130.

Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York, 2 edition.

Kejriwal, M. and Perron, P. (2008). Data dependent rules for selection of the number of leads and lags in the dynamic ols cointegrating regression. *Econometric Theory*, 24(05):1425–1441.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17:99–121.

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

Kitamura, Y. and Phillips, P. C. (1997). Fully modified iv, give and gmm estimation with possibly non-stationary regressions and instruments. *Journal of Econometrics*, 80:85–123.

Klenke, A. (2008). *Probability Theory - A Comprehensive Course*. Springer.

Kurozumi, E. and Tuvaandorj, P. (2010). Model selection criteria in multivariate models with multiple structural changes. Global COE Hi-Stat Discussion Paper Series gd10-144, Institute of Economic Research, Hitotsubashi University.

Longstaff, F., Singleton, K., Pan, J., and Pedersen, L. H. (2008). How sovereign is sovereign credit risk? *Working paper, UCLA Anderson School*.

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin Heidelberg.

Mark, N. C., Ogaki, M., and Sul, D. (2005). Dynamic seemingly unrelated cointegrating regressions. *Review of Economic Studies*, 72(3):797–820.

Mark, N. C. and Sul, D. (2003). Cointegration vector estimation by panel dols and long-run money demand. *Oxford Bulletin of Economics and Statistics*, 65(5):655–680.

Norden, L. and Weber, M. (2009). The co-movement of credit default swap, bond and stock markets: an empirical analysis. *European Financial Management*, 15(3):529–573.

Park, J. Y. and Phillips, P. C. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, 4(03):468–497.

Pedroni, P. (2000). Fully modified OLS for heterogeneous cointegrated panels. In Baltagi, B., editor, *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*. Elsevier.

Phillips, P. C. and Loretan, M. (1991). Estimating long run economic equilibria. *Review of Economic Studies*, 58:407–436.

Phillips, P. C. B. (2006). Optimal estimation of cointegrated systems with irrelevant instruments. Cowles Foundation Discussion Papers 1547, Cowles Foundation for Research in Economics, Yale University.

Phillips, P. C. B. and Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies*, 57(1):99–125.

Phillips, P. C. B. and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1112.

Poirier, D. J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach.* MIT Press, Cambridge, Mass.

Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory.* Oxford University Press, New York.

Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory*, 7(1):1–21.

Schneider, P., Sögner, L., and Veža, T. (2010). The economic role of jumps and recovery rates in the market for corporate default risk. *Journal of Financial and Quantitative Analysis*, 45:1517–1547.

Schönbucher, P. J. (2003). *Credit Derivates Pricing Models: Models, Pricing and Implementation.* Wiley Finance Series. John Wiley & Sons Ltd.

Stock, J. H. and Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61(4):783–820.

Tarashev, N. and Zhu, H. (2008). The pricing of correlated default risk evidence from the credit derivatives market. *Journal of Fixed Income*, 18(1):5–24.

Vogelsang, T. J. and Wagner, M. (2011). Integrated modified ols estimation and fixed-b inference for cointegrating regressions. Economics Series 263, Institute for Advanced Studies.

White, H. (2001). *Asymptotic Theory For Econometricians.* Emerald Group Publishing, Bingley, UK, revised edition.

Zhou, C. (2001). An analysis of default correlations and multiple defaults. *Review of Financial Studies*, 14(2):555–76.

Authors: Jan Mutl, Leopold Sögner

Title: Parameter Estimation and Inference with Spatial Lags and Cointegration

Reihe Ökonomie / Economics Series 296

ISSN: 1605-7996

ISSN: 1605-7996