Parameter Estimation and Inference with Spatial Lags and Cointegration

Jan Mutl
EBS Business School
Gustav-Streseman-Ring 3
65189 Wiesbaden, Germany

Leopold Sögner
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria

December 4, 2012

Abstract

We study dynamic panel data models with a cointegration relationship that includes a spatial lag. Such a relationship postulates that the long run outcome for a particular cross-section is affected by a weighted average of the outcomes in the other cross-sections. This model produces outcomes that are non-linear in the coefficients to be estimated. It also renders the existing estimation methods that typically assume cross-sectional independence inapplicable. Assuming that the weights are exogenously given, we extend the dynamic ordinary least squares ($DOLS$) methodology and provide a dynamic two-stage least squares estimator ($D2SLS$). We derive the large sample properties of our proposed estimator under a set of low-level assumptions and investigate its small sample distribution in a simulation study. We illustrate our estimation methodology by investigating the structure of spatial correlation of firm-level implied credit risk. Credit default swaps, firm specific data and industry data from the US market are used to estimate spatial correlation models. We construct the economic space based on a "closeness" measure for firms based on input-output matrices. Our estimates show that the spatial correlation of credit default spreads is substantial and highly significant.

Keywords: Dynamic Least Squares, Cointegration, Credit Risk.

JEL: C31, C32.

1 Introduction

Estimating the parameters of a cointegrating relationship is perceived to be a relatively straightforward task. The challenges arise when we attempt to provide guidance for statistical inference and hypothesis testing. Two broad approaches have emerged in the literature. One possibility is to use a simple estimation routine, i.e. 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). Using such approach might involve making assumptions (e.g. no serial correlation of the disturbances) that are possibly violated in some applications.

Another opportunity is to adjust the estimation routine, such that the large sample distribution is either simpler, or free of nuisance parameters. An example along these lines is the *fully modified OLS* estimator (see e.g. Phillips and Hansen (1990), Phillips and Moon (1999), Pedroni (2000); or Vogelsang and Wagner (2011) for a fixed-b perspective). The necessity to modify the OLS estimation arises from the presence of endogeneity and serial correlation of the errors. A further alternative where the estimation routine is augmented is the *dynamic OLS* estimation ($DOLS$). Here the serial correlation and the regressor endogeneity is controlled for by including time-series leads and lags of the regressors (cf. Phillips and Loretan (1991), Saikonnen (1991), Kao and Chiang (2000), or Mark and Sul (2003)). This approach will also be used and augmented in this article.

This paper investigates the estimation of a non-standard cointegrating relationship under the presence of regressor endogeneity and serial correlation in the disturbances. Our cointegrating vector in a panel setting is extended to include peer or neighborhood effects which are modeled as spatial lags following Cliff and Ord (1973). The spatially lagged variables in the cointegrating relationship are endogenous, such that the endogeneity cannot be controlled for by the dynamic OLS modification ($DOLS$). Therefore, we propose to use a dynamic two-stage least squares ($D2SLS$) estimator, which combines $DOLS$ literature and two stage least squares estimation. In addition, to the serial lags used by DOLS our estimation procedure uses cross-sectional lags of the regressors to control for the endogeneity of the spatial lags in the cointegrating vector.

In the rest of the paper we first describe our model and the formal assumptions in Section 2. Section 3 describes the $D2SLS$ estimation 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

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

where y_{it} is the scalar response random variable, $\mathbf{x}_{it}$ is a $k \times 1$ vector of prediction random variables. $t = 1, \ldots, T$ is the time index, $i = 1, \ldots, n$ is the cross-sectional dimension. n is kept fixed throughout the following analysis. The term $y_{it}^{*} = \sum_{j=1}^{n} W_{ij} y_{jt}$ is referred to as a *spatial lag* 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 $\mathbf{W}$ and 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)).[1] 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 disturbance term is assumed to include an individual-specific effect α_i and an idiosyncratic component $u_{it}^{\dagger}$ that is independent across i but possibly dependent across t. Analogically to Mark and Sul (2003)

[1] Observe that the requirement for invertibility is that there exists some matrix norm $\|.\|$ such that $\|\rho \mathbf{W}\| < 1$. Given that the spectral radius is the smallest matrix norm (cf. Theorem 5.6.9 in Horn and Johnson (1985)), our assumption is in this sense least restrictive. It will, for example, be satisfied when the maximum absolute row or column sums of $\rho \mathbf{W}$ are less than one.

the prediction random variable $\mathbf{x}_{it}$ is assumed to be integrated of order one, $I(1)$, and to be generated from

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

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)] Let us define the stacked vector $\mathbf{w}_{it}^{\dagger} = \left(u_{it}^{\dagger}, v_{it}'\right)'$. Then $(\mathbf{w}_{it}^{\dagger})$ is independent across $i = 1, \ldots, n$ and has a moving average representation

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

where $\varepsilon_{it}^{\dagger}$ is *iid* with mean zero, covariance $\mathbf{I}_k$ 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 lag polynomial in the lag operator L, where $\sum_{j=0}^{\infty} j \left[\mathbf{\Psi}_{ij}^{\dagger}\right]_{(m,n)} < \infty$; $\left[\mathbf{\Psi}_{ij}^{\dagger}\right]_{(m,n)}$ is the m, n element of the matrix $\mathbf{\Psi}_{ij}^{\dagger}$.

The short-run $k+1 \times k+1$ covariance matrix $\Gamma_{i0}^{\dagger}$ and the autocovariance matrix $\Gamma_{ij}^{\dagger}$ are

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

We will use the following notation: $\Gamma_{uu,ij}^{\dagger}$ is the $\mathtt{1,1}$ element of $\mathbf{\Gamma}_{ij}^{\dagger}$, $\mathbf{\Gamma}_{uv,j}^{\dagger}$ corresponds to $\left[\mathbf{\Gamma}_{ij}^{\dagger}\right]_{(\mathtt{2:k+1,1})}$, $\mathbf{\Gamma}_{vu,j}^{\dagger}$ corresponds to $\left[\mathbf{\Gamma}_{ij}^{\dagger}\right]_{(\mathtt{1,2:k+1})}$ while $\mathbf{\Gamma}_{vv,j}$ corresponds to $k \times k$ submatrix $\left[\mathbf{\Gamma}_{ii}^{\dagger}\right]_{(\mathtt{2:k+1,2:k+1})}$; regarding notation $(a:b, c:d)$ stands for "form 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 $\mathbf{\Omega}_{i}^{\dagger}$ of $\mathbf{w}_{it}^{\dagger}$ is given by

$$\begin{aligned}\boldsymbol{\Omega}_i^{\dagger} &= \sum_{j=-\infty}^{\infty} \mathbb{E}\left(\mathbf{w}_{it}^{\dagger}\mathbf{w}_{i,t+j}^{\dagger\prime}\right) = \boldsymbol{\Psi}_i^{\dagger}(1)\mathbf{I}_k\boldsymbol{\Psi}_i^{\dagger}(1)' = \boldsymbol{\Gamma}_{i0}^{\dagger} + \sum_{j=1}^{\infty}\left(\boldsymbol{\Gamma}_{ij}^{\dagger} + \boldsymbol{\Gamma}_{ij}^{\dagger\prime}\right) \qquad (4)\\ &= \begin{pmatrix} \Omega_{uu,i}^{\dagger} & \boldsymbol{\Omega}_{vu,i}^{\dagger} \\ \boldsymbol{\Omega}_{uv,i}^{\dagger} & \boldsymbol{\Omega}_{vv,i} \end{pmatrix} = \begin{pmatrix} \Gamma_{uu,i0}^{\dagger} & \boldsymbol{\Gamma}_{vu,i0}^{\dagger} \\ \boldsymbol{\Gamma}_{uv,i0}^{\dagger} & \boldsymbol{\Gamma}_{vv,i0} \end{pmatrix} + 2\sum_{j=1}^{\infty} \begin{pmatrix} \Gamma_{uu,ij}^{\dagger} & \boldsymbol{\Gamma}_{vu,ij}^{\dagger} \\ \boldsymbol{\Gamma}_{uv,ij}^{\dagger} & \boldsymbol{\Gamma}_{vv,ij} \end{pmatrix}.\end{aligned}$$

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 scalars $\Omega_{uu,i}^{\dagger}$ and blocks $\boldsymbol{\Omega}_{vv,i}$ along their main diagonal. Given such covariance structure, we want to exclude a cointegration relationship between the terms of $\Delta\mathbf{x}_{it}$. In addition our large sample results will require that $\Omega_{uu,i}^{\dagger}$ is invertible. Therefore we impose the following assumption:

Assumption 3. [Error Dynamics II; see Phillips (2006)]
$\boldsymbol{\Psi}_i^{\dagger}(1)$ is non-singular such that $\mathbf{x}_{it}$ is a full rank integrated process and $\boldsymbol{\Omega}_{vv,i}$ has full rank k.

By the independence across i assumption (i.e. Assumption 2), this implies that the rank of $\boldsymbol{\Omega}_{vv}$ is nk. In the next step we impose an additional restriction on the error dynamics:

XXX check whether is is correct - to be completed by JAN

Assumption 4. [Error Dynamics III; see Saikkonen (1991), Mark et al. (2005)]
Given the Assumptions 2 and 3, $p(T)$ fulfills $\frac{p(T)^3}{T} \to 0$ and $\sqrt{T}\sum_{k>p(T)} \delta_{kt} \to 0$ as $T \to \infty$.

XXX rewrite this paragraph Note that Assumption 4 should be interpreted as a restriction on the coefficients of the lag polynomial $\boldsymbol{\Psi}_i^{\dagger}(L)$ that implies that at most p leads and lags of $\mathbf{v}_{it}$ are conditionally correlated with $u_{it}^{\dagger}$. Equipped with Assumption 4 a dynamic least squares estimator can be constructed.

Remark 1. Let us briefly discuss how Assumption 4 is related to $ARMA(p,q)$ models. Assumption 4 restricts the autocorrelation/autocovariance structure of the process $(w_{it}^{\dagger})$. The conditional covariance between $w_{t\pm s}^{\dagger}$ and $w_t^{\dagger}$, conditional on $w_{t-p}^{\dagger}, \ldots, w_t^{\dagger}, \ldots, w_{t+p}^{\dagger}$, becomes zero for $s > p$. For the univariate case this corresponds to a partial autocorrelations of zero for orders higher than p.

From literature (e.g. Brockwell and Davis (2006)[Chapter 11]) it follows that if $w_t^\dagger \in \mathbb{R}^d$ follows a causal[2] $ARMA(p,q)$ process (where WN stands for white noise)

$$\begin{aligned} w_t^\dagger - \Phi_1 w_{t-1}^\dagger + \dots \Phi_p w_{t-p}^\dagger &= \varepsilon_t^\dagger + \Theta_1 \varepsilon_{t-1}^\dagger + \dots + \Theta_q \varepsilon_{t-q}^\dagger \\ \Phi(B) w_t^\dagger &= \Theta(B)\varepsilon_t^\dagger \text{ where } \varepsilon_t^\dagger \sim WN(0, \Gamma_0^\dagger) \ , \end{aligned} \tag{5}$$

then (5) has a unique stationary solution

$$w_t^\dagger = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}^\dagger \text{ where } \Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j = \Phi^{-1}(z)\Theta(z) \ . \tag{6}$$

I.e. a causal multivariate ARMA process has a $MA(\infty)$ representation (a more general treatment of linear processes and its Wold representation is provided in Hannan and Deistler (2012)[Chapter 1]). If $(w_t^\dagger)$ follows an $ARMA(p,0)$ process the conditional covariance becomes zero for orders larger than p. Since $\varepsilon_{it}^\dagger$ is iid, conditionally uncorrelated implies conditionally independent such that the condition $\mathbb{E}\left(u_{it}^\dagger v_{i,t+s} \middle| v_{i,t+p}, .., v_{i,t-p}\right) = \mathbf{0}_{k\times 1}$ for all $|s| > p$ is met. In constrast, an $ARMA(0,1)$ does not fulfill the requirement of Assumption 4.

Equipped with Assumptions 1 to 4, we can follow Mark and Sul (2003) and Mark et al. (2005) and remove the serial correlation by projecting on leads and lags of $\Delta \mathbf{x}_{it}$. In particular, Assumption 4 implies that at most p leads and lags of $\Delta \mathbf{x}_{it}$ are correlated with $u_{it}^\dagger$. Hence we have that the projection of $u_{it}^\dagger$ on the p leads and lags of $\Delta \mathbf{x}_{it}$ yields a new disturbance u_{it} that by construction is orthogonal to $\Delta \mathbf{x}_{it}$:

$$u_{it}^\dagger = \sum_{s=-p}^{+p} \delta_{i,s}' \Delta \mathbf{x}_{i,t-s} + u_{it} = \delta_i' \zeta_{it} + u_{it} \ , \tag{7}$$

where $\Delta \mathbf{x}_{i,t-s}$ and $\delta_{i,s}$ are a vectors of dimension k, such that the $(2p+1)k \times 1$ dimensional vectors of

[2] I.e. $det(\Phi(z)) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.

projection variables and projection coefficients are given by

$$\zeta_{it} = \left(\Delta \mathbf{x}'_{i,t+p}, ..., \Delta \mathbf{x}'_{i,t-p}\right)' \text{ and } \delta_i = \left(\delta'_{i,-p}, ..., \delta'_{i,+p}\right)' \ . \tag{8}$$

Based on Assumptions 2 to 4 and (7), the process $(\mathbf{w}_{it}) = (u_{it}, \mathbf{v}'_{it})'$ is covariance stationary with

$$\mathbf{w}_{it} = \mathbf{\Psi}_i(L)\varepsilon_{it} \ , \quad \mathbf{\Psi}_i(L) = \begin{pmatrix} \Psi_{uu,i}(L) & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{\Psi}_{vv,i}(L) \end{pmatrix}, \tag{9}$$

where $\Psi_{uu,i}(L)$ is a scalar and $\mathbf{\Psi}_{vv,i}(L)$ is a $k \times k$ matrix lag polynomial.

Given our assumptions, a functional central theorem (see e.g. Johansen (1995)[Appendix], Karatzas and Shreve (1991)[Chapter 4] or Davidson (1994)[Chapters 27-30]) can be applied. This yields

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \mathbf{w}_{it} \xrightarrow{d} \mathcal{B}_i(r) = \mathbf{\Psi}_i\left(1\right) \mathcal{W}_i(r) \ , r \in [0,1] \ , \tag{10}$$

with $\mathcal{B}_i(r) = (\mathcal{B}_{ui}(r), \mathcal{B}_{vi}(r)')'$, 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 Ω_i, $\mathcal{W}_i$ stands for a standard Brownian motion, where $\mathcal{W}_i(r) = (\mathcal{W}_{ui}(r), \mathcal{W}_{vi}(r)')'$. $[Tr]$ denotes the integer part of rT.

[3]

The $k+1 \times k+1$ long-run variance-covariance matrix of $\mathbf{w}_{it}$ which we denote $\mathbf{\Omega}_i$ is given by

$$\mathbf{\Omega}_i = \begin{pmatrix} \Psi_{uu,i}(1)^2 & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times k} & \mathbf{\Psi}_{vv,i}(1)\mathbf{\Psi}_{vv,i}(1)' \end{pmatrix} = \begin{pmatrix} \Omega_{uu,i} & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{\Omega}_{vv,i} \end{pmatrix} = \mathbf{\Gamma}_{i0} + \sum_{j=1}^{\infty} \left(\mathbf{\Gamma}_{ij} + \mathbf{\Gamma}'_{ij}\right), \tag{11}$$

where the matrices $\mathbf{\Gamma}_{ij}$ are given by

$$\mathbf{\Gamma}_{ij} = \mathbb{E}\left(\mathbf{w}_{it}\mathbf{w}'_{i,t-j}\right) = \begin{pmatrix} \Gamma_{uu,ij} & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{\Gamma}_{vv,ij} \end{pmatrix}, \tag{12}$$

[3] In addition we follow the econometrics literature in this field (see e.g. Mark et al. (2005)) and omit the borders of integration as well as the dependence of the Brownian motion on r, i.e. we write $\int W$ instead of $\int_0^1 W(r)dr$, while $\int_0^1 W(r)dW(r)$ is abbreviated by $\int WdW$.

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

Observe that our model includes a full set of individual specific effects and hence a set of individual dummies α_i is typically included to the regression (1) (*fixed effects specification*). In order to simplify the algebra, we will 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 LSDV estimates (see e.g. Baltagi (2008)[p. 11]). We thus define the variables in deviations from their individual means as

$$\begin{aligned}
\widetilde{y}_{it} &= y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}, \;\; \widetilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}, \;\; \widetilde{y}^*_{it} = \sum_{j=1}^{n} W_{ij}\widetilde{y}_{it} \\
\widetilde{\zeta}_{it} &= \zeta_{it} - \frac{1}{T}\sum_{t=1}^{T}\zeta_{it}, \;\; \widetilde{u}_{it} = u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it} \;, \qquad (13)
\end{aligned}$$

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

$$\begin{aligned}
\widetilde{y}_{it} &= \rho\sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt} + \beta'\widetilde{\mathbf{x}}_{it} + \widetilde{u}^{\dagger}_{it} = \rho\tilde{\mathbf{y}}^*_{it} + \beta'\widetilde{\mathbf{x}}_{it} + \widetilde{u}^{\dagger}_{it} \\
&= \rho\sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt} + \beta'\widetilde{\mathbf{x}}_{it} + \delta'_{it}\tilde{\zeta}_{it} + \widetilde{u}_{it} = \rho\tilde{\mathbf{y}}^*_{it} + \beta'\widetilde{\mathbf{x}}_{it} + \delta'_{it}\tilde{\zeta}_{it} + \widetilde{u}_{it} \;. \qquad (14)
\end{aligned}$$

Remark 2. An important point is that u_{it} and $\widetilde{u}_{it}$ have similar limit properties (see e.g. the arguments in Mark and Sul (2003)[p. 663]). In (10) we observed that $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} v_{it} \xrightarrow{d} \mathcal{B}_i(r) = \mathbf{\Omega}_{vv,i}\mathcal{W}_{vi}(r)$. For the demeaned term $\tilde{v}_{it} = v_{it} - \frac{1}{T}\sum_{t=1}^{T} v_{it}$ we get

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\tilde{v}_{it} = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\left(v_{it} - \frac{1}{T}\sum_{t=1}^{T} v_{it}\right) \xrightarrow{d} \mathcal{B}_{vi}(r) - r\mathcal{B}_{vi}(1).$$

$\mathcal{B}_{vi}(r) - r\mathcal{B}_{vi}(1)$ is a Brownian bridge. Since $\mathbf{x}_{it}$ is an $I(1)$ process, $\mathbf{x}_{it}$ arises from the partial sum process $\sum_{\iota=1}^{t} v_{i\iota}$. Then $\tilde{\mathbf{x}}_{it} = \sum_{\iota=1}^{t} v_{i\iota} - \frac{1}{T}\sum_{t=1}^{T}\sum_{\iota=1}^{t} v_{i\iota}$. By the continuous mapping theorem (see Klenke

(2008)[p. 257], Davidson (1994)[Theorem 26.13 & 30.2]) the $T \to \infty$ limit is given by

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

$\mathcal{B}_{vi}(r) - \int_0^1 \mathcal{B}_{vi}(s)ds$ will be abbreviated by $\tilde{\mathcal{B}}_{vi}(r)$. By Davidson (1994)[Theorem 30.2] we also obtain the result that $\frac{1}{T^2}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}'$ converges in distribution to $\int_0^1 \mathcal{B}_{vi}(s)\mathcal{B}_{vi}'(s)ds$. In addition Davidson (1994)[Theorem 30.13] and some algebra shows that $\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}\tilde{\mathbf{u}}_{it} \xrightarrow{d} \sqrt{\mathbf{\Omega}_{uu,i}}\int_0^1 \tilde{\mathcal{B}}_{vi}(r)d\mathcal{W}_{ui}(r)$.

XXX begin two dimensional system

Remark 3. 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. In more details we get

$$\left[\mathbf{I}_2 - \rho\begin{pmatrix} 0 & W_{12} \\ W_{21} & 0 \end{pmatrix}\right]^{-1} = \frac{1}{1+\rho^2 W_{12}W_{21}} \cdot \begin{pmatrix} 1 & -\rho W_{21} \\ -\rho W_{12} & 1 \end{pmatrix}. \qquad (15)$$

Equations (1) and (15) with $n = 2$ now result in

$$\begin{pmatrix} y_{1t} \\ y_{21t} \end{pmatrix} = \frac{1}{1+\rho^2 W_{12}W_{21}} \cdot \begin{pmatrix} \beta^\top x_{1t} & -\rho W_{21}\beta^\top x_{2t} & +u_{1t}^\dagger & -\rho W_{21}u_{2t}^\dagger & +\alpha_1 \\ -\rho W_{12}\beta^\top x_{1t} & +\beta^\top x_{2t} & +u_{2t}^\dagger & -\rho W_{12}u_{1t}^\dagger & +\alpha_1 \end{pmatrix}. \qquad (16)$$

Equation 16 also shows the $n = 2$ coinintegration relationships. These are described by the first and the second row of this equation. The cointegrating relationship does not have the usual linear form in the sense that the solution for y_{it} is a nonlinear function of the parameter ρ. For an arbitrary but fixed $n \in \mathbb{N}$ we obtain n cointegration relationships. This can be seen by looking at $\mathbf{y}_{it} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}\left(\beta^\top x_{it} + u_{it}^\dagger\right)$. By considering the terms $\left(\beta^\top x_{it} + u_{it}^\dagger\right)$ and Assumptions 2-4 we have n cointegration relationships. Since Assumption 1 guarantees that $\mathbf{I}_n - \rho\mathbf{W}$ has the full rank n, the cointegration space remains of dimension n.

In addition we observe that: (i) x_{it} and $u_{it}^\dagger$ are correlated by the assumptions on $\Psi_i^\dagger$. (ii) 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 $u^{\dagger}_{it}$ have to be correlated. We account for "serial" endogeneity by means of DOLS as proposed in Mark and Sul (2003), Mark et al. (2005). In addition endogeneity enters via the spatial correlation modelled by $\rho\mathbf{W}$. To account for this kind of "spatial" endogeneity the DOLS approach has to be augmented.

XXX end two dimensional system

3 Estimation Procedure and Large Sample Results

The goal of the following analysis is to construct the $D2SLS$ estimator and show that it leads to consistent estimates of the parameters ρ and β. Then we provide the large sample distribution of the $D2SLS$ estimator. The parameters δ are nuisance parameters. By our analysis it should also become more clear why applying DOLS is not sufficient to obtain consistent parameter estimates. In order to write down our estimator, we first define the model in a stacked notation. For notational simplicity we drop the tilde notation in the stacked model and define

$$\begin{aligned} \mathbf{y} &= \left(\widetilde{y}_{11}, ..., \widetilde{y}_{1T}, ..., \widetilde{y}_{n1}, ..., \widetilde{y}_{nT}\right)', \\ \mathbf{y}^* &= \left(\widetilde{y}^*_{11}, ..., \widetilde{y}^*_{1T}, ..., \widetilde{y}^*_{n1}, ..., \widetilde{y}^*_{nT}\right)', \\ \mathbf{x} &= \left(\widetilde{\mathbf{x}}'_{11}, ..., \widetilde{\mathbf{x}}'_{1T}, ..., \widetilde{\mathbf{x}}'_{n1}, ..., \widetilde{\mathbf{x}}'_{nT}\right)', \\ \mathbf{u} &= \left(\widetilde{u}_{11}, ..., \widetilde{u}_{1T}, ..., \widetilde{u}_{n1}, ..., \widetilde{u}_{nT}\right)' \end{aligned} \tag{17}$$

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

$$\zeta\delta = \begin{pmatrix} \delta'_{11}\widetilde{\zeta}_{11} \\ \vdots \\ \delta'_{nT}\widetilde{\zeta}_{nT} \end{pmatrix} = \begin{pmatrix} \widetilde{\zeta}'_{11} & \mathbf{0}_{1\times(2p+1)k} & \mathbf{0}_{1\times(2p+1)k} \\ \vdots & & \\ \widetilde{\zeta}'_{1T} & \mathbf{0}_{1\times(2p+1)k} & \mathbf{0}_{1\times(2p+1)k} \\ \mathbf{0}_{1\times(2p+1)k} & \widetilde{\zeta}_{21} & \mathbf{0}_{1\times(2p+1)k} \\ & & \ddots \\ \mathbf{0}_{1\times(2p+1)k} & \mathbf{0}_{1\times(2p+1)k} & \widetilde{\zeta}_{nT} \end{pmatrix} \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix}. \tag{18}$$

ζ is a $nT \times (2p+1)k \cdot n$ 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 (14) in stacked form:

$$\mathbf{y} = \rho \mathbf{y}^* + \mathbf{x}\beta + \zeta\delta + \mathbf{u} = (\mathbf{y}^*, \mathbf{x})\,\gamma + \zeta\delta + \mathbf{u} = \mathbf{X}\left(\gamma', \delta'\right)' + \mathbf{u}\ , \tag{19}$$

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

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 instruments of dimension $q_\rho \geq k_\rho = 1$ are available fulfilling the properties necessary for instrumental variable estimation. In contrast to this, we follow Kelejian and Prucha (XX) and base the instruments on the spatial lags of the explanatory variables. Observe that our model can be solved as

$$\mathbf{y} = \left[\mathbf{I}_T \otimes (\mathbf{I}_n - \rho\mathbf{W})^{-1}\right](\mathbf{x}\beta + \zeta\delta + \mathbf{u})\,, \tag{20}$$

where the inverse exists by our Assumption 1. 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)):

$$(\mathbf{I}_n - \rho\mathbf{W})^{-1} = \sum_{s=0}^{\infty} (\rho\mathbf{W})^s\,, \tag{21}$$

This implies that variables of the form $\sum_{j=1}^n W_{ij}x_{jtv}, \sum_{j=1}^n W_{ij}^2 x_{jtv}, \ldots$ are suitable instruments for $\mathbf{Wy}$. We hence assume that the following set of instruments is used:

Assumption 5. [Valid Instruments] The instruments are $\widetilde{\mathbf{x}}^*_{itv} = \sum_{j=1}^n W_{ij}^{\tau_v}\widetilde{\mathbf{x}}_{jtv}$; $v = 1, \ldots, q_\rho$ and $\tau_v \in \mathbb{N}$. We assume that these instruments fulfill the necessary requirements for instrumental variable estimation (see e.g. Ruud (2000)[Chapter 20], Phillips and Hansen (1990) and Kitamura and Phillips (1997)).

By the above assumptions $\widetilde{\mathbf{x}}^*_{itv}$ is correlated with $\widetilde{y}^*_{it}$. The independence of $\widetilde{\mathbf{x}}^*_{itv}$ and $\widetilde{u}_{it}$ follows from the construction of $\widetilde{u}_{it}$ which implies that $\widetilde{\mathbf{x}}_{it}$ and $\widetilde{u}_{it}$ are independent. Appendix B shows that with $\tau_v = 1$ and some regularity conditions on $\mathbf{W}$ the rank condition is satisfied.

To keep the notation simple we consider a model with $k_\rho = 1$. With $q_\rho = k_\rho = 1$ we are in the just identified case, while if $k \geq q_\rho > k_\rho = 1$ we consider the over-identified case. We collect our instruments in the matrix

$$\mathbf{x}^* = \left(\widetilde{\mathbf{x}}_{11}^{*\prime}, ..\widetilde{\mathbf{x}}_{1T}^{*\prime}, ...\widetilde{\mathbf{x}}_{n1}^{*\prime}, ..\widetilde{\mathbf{x}}_{nT}^{*\prime}\right)', \tag{22}$$

which is of dimension $nT \times q_\rho$. The set of our instruments is then $\mathbf{Z} = (\mathbf{x}^*, \mathbf{x}, \zeta)$. The matrix of explanatory variables $\mathbf{X}$ is of dimension $Tn \times 1 + k + (2p+1)k \cdot n$, while the dimension of $\mathbf{Z}$ is $Tn \times q_\rho + k + (2p+1)k \cdot n$.[4] Before we present our estimator, let us discuss why e.g. DOLS does not provide us with a consistent estimator:

Remark 4 (Endogeneity)**.** Let us consider (19). If $\widetilde{u}_{it}$ and the explanatory variables $(\mathbf{y}^*, \mathbf{x}, \zeta)$ were orthogonal, the results of Mark and Sul (2003) and Mark et al. (2005) could be applied to obtain consistent estimates. However, $\widetilde{y}_{it}$ is influenced by $\widetilde{y}_{jt}$ (if some $W_{ij} \neq 0$). Since $\widetilde{y}_{jt}$ is affected by the noise $\widetilde{u}_{jt}$ it follows that $(\mathbf{y}^*, \mathbf{x}, \zeta)$ and $\mathbf{u}$ are not orthogonal. Therefore, the DOLS estimate does not provide consistent estimates.

In analogy to a standard regression setting with endogenous regressors, we now construct a two stage-least square procedure for our panel setting. With two-stage least squares the initial stage results in projected values

$$\widehat{\mathbf{y}^*} = \mathbf{x}^* \left(\mathbf{x}^{*\prime}\mathbf{x}^*\right)^{-1} \mathbf{x}^{*\prime}\mathbf{y}^*, \tag{23}$$

while the second stage estimator is

$$\begin{pmatrix} \widehat{\rho} \\ \widehat{\beta} \\ \widehat{\delta} \end{pmatrix}_{D2SLS} = \begin{pmatrix} \widehat{\mathbf{y}^*}'\widehat{\mathbf{y}^*} & \mathbf{x}'\widehat{\mathbf{y}^*} & \zeta'\widehat{\mathbf{y}^*} \\ \widehat{\mathbf{y}^*}'\mathbf{x} & \mathbf{x}'\mathbf{x} & \zeta'\mathbf{x} \\ \widehat{\mathbf{y}^*}'\zeta & \mathbf{x}'\zeta & \zeta'\zeta \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\mathbf{y}^*}' \\ \mathbf{x}' \\ \zeta' \end{pmatrix} \mathbf{y}. \tag{24}$$

Since $\widetilde{\mathbf{x}}_{it}$ is orthogonal to $\widetilde{u}_{it}$, we will also have (for a fixed n) that our instruments $\widetilde{\mathbf{x}}_{it}^*$ are also orthogonal to $\widetilde{u}_{it}$. Since in the first stage we project the endogenous variable $\widetilde{y}_{it}^*$ on $\widetilde{\mathbf{x}}_{it}^*$, the projected values denoted by $\widehat{\widetilde{y}_{it}^*}$ will also be orthogonal to $\widetilde{u}_{it}$. Our analysis will show that this results in a consistent estimator in the second stage regression. Based in this discussion, we can now compactly write the *dynamic two-stage*

[4] If $\mathbf{\Psi}_i^\dagger(L) = \mathbf{\Psi}^\dagger(L)$ then $\mathbf{X}$ is of dimension $Tn \times 1 + k + (2p+1)k$ while $\mathbf{Z}$ is of dimension $Tn \times q_\rho + k + (2p+1)k$.

least squares estimator of $(\rho, \beta', \delta')' = (\gamma', \delta')'$ as

$$\begin{aligned} \widehat{(\gamma', \delta')}'_{D2SLS} &= \left(\mathbf{X}' P_H \mathbf{X}\right)^{-1} \mathbf{X}' P_H \mathbf{y} \\ &= (\gamma', \delta')' + \left(\mathbf{X}' P_H \mathbf{X}\right)^{-1} \mathbf{X}' P_H \mathbf{u} \ . \end{aligned} \tag{25}$$

P_H is the project operator (see e.g. Ruud (2000)[Chapter 3]) , projecting on the space spanned by $\mathbf{Z}$, i.e.

$$P_H = \mathbf{Z} \left(\mathbf{Z}'\mathbf{Z}\right)^{-1} \mathbf{Z}' \ . \tag{26}$$

Since $\mathbf{Z}$ is a $Tn \times q_\rho + k + (2p+1)k \cdot n$ matrix, P_H has to by a $Tn \times Tn$ matrix. With $q_\rho = 1$, we are in the just identified case, where the estimator is given by:

$$\widehat{(\gamma', \delta')}'_{D2SLS} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1} \mathbf{Z}'\mathbf{y} = (\gamma', \delta')' + \left(\mathbf{Z}'\mathbf{X}\right)^{-1} \mathbf{Z}'\mathbf{u} \ . \tag{27}$$

Equipped with the $D2SLS$ estimator (25) we obtain the following result:

Theorem 1 (Limits for $D2SLS$ Estimation)**.** *Consider the fixed effects spatial correlation model (14) and the estimator (25). Then for n fixed and $T \to \infty$ it follows that*

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

2. *$\sqrt{n}T\left(\hat{\gamma}_{D2SLS} - \gamma\right)$ converges in distribution to $M_n^{-1} m_n$ where m_n and M_n are given by (61) and (62).*

3. *Given a $s \times k + 1$ restriction matrix $\mathbf{R}$, the Wald statistic $S_{\gamma,nT}$ defined in (69) converges to a χ^2 random variable with s degrees of freedom.*

Remark 5. The reader should note that the two-stage least squares estimator and the $DOLS$ estimator are special cases of the dynamic two-stage least squares estimator. With $p = 0$ the matrix ζ becomes empty, yielding the two-stage least squares estimator. If $\mathbf{x}^* = \mathbf{y}^*$, then we obtain the $DOLS$ estimator.

The Wald-statistic presented in Section 4, can therefore also 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 x_{it} such that $k = 2$ and set $\beta = (1, 1)'$. The number of instruments is $k_\rho = 2$.

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 $\varepsilon_{it}^\dagger \sim N(0, \Sigma_\varepsilon)$. For Σ_ε we use $\Sigma_\varepsilon = diag(1, 0.1, 0.1)$, $\Sigma_\varepsilon = \mathbf{I}_3$ and $\Sigma_\varepsilon = diag(1, 10, 10)$; $\mathbf{I}_3$ stands for the three-dimensional identity matrix, while $diag(1, 0.1, 0.1)$ stands for a diagonal matrix with entries 1, 0.1 and 0.1 along the main diagonal. In a first step we generate $\mathbf{w}_{it}^\dagger$ by means of the first order vector autoregressive system ($VAR(1)$)

$$\mathbf{w}_{it}^\dagger = \mathbf{\Phi}\mathbf{w}_{i,t-1}^\dagger + \varepsilon_{it}^\dagger \, . \tag{28}$$

For the 3×3 matrix $\mathbf{\Phi}$ we use the following designs:

Design 1: Stationary $VAR(1)$ with maximum eigenvalue of 0.6

$$\mathbf{\Phi} = \begin{pmatrix} 0.4 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.4 \end{pmatrix} . \tag{29}$$

Design 2: Stationary $VAR(1)$ with maximum eigenvalue of 0.8

$$\mathbf{\Phi} = \begin{pmatrix} 0.6 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.6 \end{pmatrix}. \tag{30}$$

Design 3: Stationary $VAR(1)$ with maximum eigenvalue of 0.95

$$\mathbf{\Phi} = \begin{pmatrix} 0.75 & 0.1 & 0.1 \\ 0.1 & 0.75 & 0.1 \\ 0.1 & 0.1 & 0.75 \end{pmatrix}. \tag{31}$$

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

$$\mathbf{w}_{it}^{\dagger} = \varepsilon_{it}^{\dagger} + \sum_{l=1}^{q} \mathbf{\Psi}_{il}^{\dagger} \varepsilon_{i,t-l}^{\dagger}, \tag{32}$$

where we choose:

Design 4: First-order MA, $q = 1$

$$\mathbf{\Psi}_{i1}^{\dagger} = \begin{pmatrix} 0.4 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.4 \end{pmatrix}. \tag{33}$$

Design 5: Second-order MA, $q = 2$

$$\mathbf{\Psi}_{i1}^{\dagger} = \begin{pmatrix} 0.6 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.6 \end{pmatrix}, \quad \mathbf{\Psi}_{i2}^{\dagger} = \begin{pmatrix} 0.4 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.4 \end{pmatrix}. \tag{34}$$

Recall that the disturbance of the model is given by the first element of the vector $\mathbf{w}_{it}^{\dagger}$, while its remaining elements are the innovations of the explanatory variables. Therefore, the maximum numbers of leads and lags of the explanatory variables that are conditionally correlated with the disturbances is 1 in Designs

1-3, while the Designs 4 and 5 fail to satisfy Assumption 4.

In the case of the VAR 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

$$\mathbf{w}_{i0}^{\dagger} = \sum_{j=0}^{\infty} \mathbf{\Phi}^{j} \varepsilon_{i,-j}^{\dagger}. \tag{35}$$

and hence $\mathbf{w}_{i0}^{\dagger}$ is a random variable that is independent from $\varepsilon_{it}^{\dagger}$ for $t > 0$. When 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 $E\left(\mathbf{w}_{it}^{\dagger}\mathbf{w}_{it}^{\dagger\prime}\right) = \mathbf{\Gamma}_{i0}$ where

$$\mathbf{\Gamma}_{i0} = E\left(\sum_{j=0}^{\infty} \mathbf{\Phi}^{j} \varepsilon_{i,-j}^{\dagger}\right)\left(\sum_{j=0}^{\infty} \mathbf{\Phi}^{j} \varepsilon_{i,-j}^{\dagger}\right)' = \sum_{j=0}^{\infty} \mathbf{\Phi}^{j} \mathbf{\Phi}'^{j}. \tag{36}$$

The above expression implies

$$\mathbf{\Phi}\mathbf{\Gamma}_{i0}\mathbf{\Phi}' = \sum_{j=0}^{\infty} \mathbf{\Phi}^{j+1} \mathbf{\Phi}'^{j+1} = \mathbf{I}_{(k+1)} + \sum_{j=0}^{\infty} \mathbf{\Phi}^{j} \mathbf{\Phi}'^{j} = \mathbf{I}_{(k+1)} + \mathbf{\Gamma}_{i0}. \tag{37}$$

After vectorizing and solving for $\mathbf{\Gamma}_{i0}$ we obtain

$$vec\left(\mathbf{\Gamma}_{i0}\right) = \left(\mathbf{I}_{(k+1)^2} - \mathbf{\Phi} \otimes \mathbf{\Phi}\right)^{-1} vec\left(\mathbf{I}_{(k+1)}\right) \ . \tag{38}$$

The remaining parameters of the model are chosen as follows: We generate the individual effects α_i from $\alpha_i \sim N\left(\mathbf{0}_{3\times 1}, \mathbf{I}_3\right)$. 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. (iv) 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 (DGP) (3, 5, 7, 5 different settings for Σ_ε, the autoregressive structure of $\mathbf{w}_{it}^\dagger$, the spatial correlation parameter ρ and the spatial correlation matrix $\mathbf{W}$, respectively).

Table 4 presents the fraction of the simulation runs for a "three step ahead-three step behind circular world", where the true null hypothesis $\rho = 0$ has been rejected by applying $\alpha_c = \{0.01, 0.05, 0.1\}$ significance levels. The sample size is given by $n = 50$ and $T = 200$. The results are obtained by means of $\texttt{M} = 1000$ simulation steps. By using the four other designs for the spatial correlation matrix $\mathbf{W}$ we observe very similar results. In more details we get the following results: When increasing the serial correlation in the $\mathbf{w}_{it}$ (from Design 1 to Design 3), the percentages where the true null is rejected increase. A similar effect is observed when $\Sigma_\varepsilon = diag(1, 1, 1)$ or $\Sigma_\varepsilon = diag(1, 10, 10)$ is used instead of $\Sigma_\varepsilon = diag(1, 0.1, 0.1)$ (abbreviated by the index 2, 3 and 1 in the second column of Table 4). While the rejection rates are too low with the covariance matrix with the low variance in the components driving $\mathbf{x}_{it}$, these rates are too high with the two other alternatives. For $\rho = \{-0.95, -.5, -0.1, 0.1, 0.5, 0.95\}$ the rejection rates are close to 100%, therefore we do not present the results in this Table.

		$\alpha_c = 0.01$			$\alpha_c = 0.05$			$\alpha_c = 0.10$		
Design	Σ_ε	*2SLS*	*DOLS*	*D2SLS*	*2SLS*	*DOLS*	*D2SLS*	*2SLS*	*DOLS*	*D2SLS*
1	1	1.80	0.30	0.30	6.70	2.20	2.30	12.50	4.80	4.50
1	2	1.10	1.30	1.40	6.60	6.70	6.50	11.60	13.80	13.80
1	3	1.20	2.00	2.00	6.00	7.20	7.30	11.90	13.30	13.30
2	1	1.60	0.40	0.30	5.70	2.20	2.10	11.30	3.70	4.10
2	2	1.70	2.10	2.00	6.80	8.30	8.50	13.10	15.50	15.50
2	3	1.30	2.20	2.30	7.60	8.40	8.50	13.50	15.10	15.10
3	1	2.70	1.10	1.30	10.80	4.30	4.20	17.60	7.60	7.60
3	2	5.10	2.90	2.90	13.70	12.00	12.00	22.00	19.00	19.10
3	3	5.40	5.60	5.60	13.50	14.90	14.70	21.60	21.90	22.10
4	1	1.10	0.50	0.50	6.40	3.50	3.60	11.20	7.80	8.20
4	2	1.20	1.60	1.60	6.60	7.70	7.90	11.10	11.90	12.20
4	3	1.50	1.80	1.80	6.50	7.60	7.70	10.70	12.80	12.70
5	1	1.30	0.60	0.50	6.10	3.50	3.30	10.90	7.20	7.40
5	2	1.20	1.70	1.70	6.40	7.80	7.70	10.40	13.20	13.20
5	3	1.40	1.80	1.90	6.50	7.70	7.70	11.60	13.20	13.10

Table 1: Size for the parameter ρ for three step ahead-three step behind circular world: Rejections of the true null hypothesis $\rho = 0$ in percentage terms, given the significance levels $\alpha_c = \{0.01, 0.05, 0.1\}$. 1000 Simulation runs. Cross-sectional dimension $n = 50$, time series dimension $T = 200$. $\texttt{M} = 1000$ Monte Carlo runs.

We also estimated the bias by means of $\rho - \sum_{\mathtt{m}=1}^{\mathtt{M}} \hat{\rho}_{\mathtt{M}}$, where $\mathtt{M} = 1000$ is the number of Monte Carlo steps and $\mathtt{m}$ is the index of the corresponding Monte Carlo step. Table 4 presents the results for the Designs 3 and 5, the results for the other designs are very similar. In Table 4 we observe that the biases of the corresponding estimation approaches are quite small. Only for the unbalanced covariance matrix with smaller errors in the explanatory variables (indexed by one), a slightly larger bias is observed with the *DOLS* approach.

Design	Σ_ε	ρ	*2SLS*	*DOLS*	*D2SLS*
3	1	-0.95	1.66E-05	1.59E-04	-6.01E-06
3	1	-0.5	1.59E-05	8.82E-05	-6.47E-06
3	1	-0.1	1.21E-05	1.23E-05	-6.39E-06
3	1	0	1.08E-05	-4.93E-06	-6.27E-06
3	1	0.1	9.27E-06	-2.05E-05	-6.10E-06
3	1	0.5	2.12E-06	-5.61E-05	-4.82E-06
3	1	0.95	-6.12E-06	-1.33E-05	-1.07E-06
3	2	-0.95	-6.89E-05	8.47E-05	-8.77E-05
3	2	-0.5	-6.55E-05	1.02E-05	-8.58E-05
3	2	-0.1	-5.72E-05	-5.77E-05	-7.66E-05
3	2	0	-5.44E-05	-7.16E-05	-7.32E-05
3	2	0.1	-5.12E-05	-8.34E-05	-6.93E-05
3	2	0.5	-3.50E-05	-1.00E-04	-4.88E-05
3	2	0.95	-7.38E-06	-2.59E-05	-1.20E-05
3	3	-0.95	-7.27E-05	2.71E-05	-7.53E-05
3	3	-0.5	-7.20E-05	-2.09E-05	-7.61E-05
3	3	-0.1	-6.50E-05	-5.90E-05	-6.99E-05
3	3	0	-6.23E-05	-6.61E-05	-6.72E-05
3	3	0.1	-5.92E-05	-7.19E-05	-6.41E-05
3	3	0.5	-4.23E-05	-7.67E-05	-4.67E-05
3	3	0.95	-9.18E-06	-2.47E-05	-1.16E-05
5	1	-0.95	4.11E-05	1.25E-02	-2.94E-04
5	1	-0.5	4.25E-05	7.04E-03	-2.96E-04
5	1	-0.1	3.75E-05	1.14E-03	-2.70E-04
5	1	0	3.53E-05	-2.21E-04	-2.59E-04
5	1	0.1	3.27E-05	-1.47E-03	-2.47E-04
5	1	0.5	1.73E-05	-4.56E-03	-1.81E-04
5	1	0.95	-1.10E-05	-1.53E-03	-5.47E-05
5	2	-0.95	-8.80E-06	2.73E-04	-2.29E-05
5	2	-0.5	-9.74E-06	1.45E-04	-2.36E-05
5	2	-0.1	-9.83E-06	9.86E-06	-2.22E-05
5	2	0	-9.72E-06	-2.09E-05	-2.16E-05
5	2	0.1	-9.55E-06	-4.90E-05	-2.07E-05
5	2	0.5	-8.11E-06	-1.16E-04	-1.59E-05
5	2	0.95	-3.65E-06	-3.89E-05	-5.28E-06
5	3	-0.95	-3.91E-06	2.86E-05	-6.14E-06
5	3	-0.5	-4.46E-06	1.30E-05	-6.52E-06
5	3	-0.1	-4.58E-06	-2.61E-06	-6.30E-06
5	3	0	-4.54E-06	-6.07E-06	-6.15E-06
5	3	0.1	-4.47E-06	-9.17E-06	-5.96E-06
5	3	0.5	-3.79E-06	-1.59E-05	-4.71E-06
5	3	0.95	-1.49E-06	-4.91E-06	-1.60E-06

Table 2: Bias for the parameter estimates of ρ for three step ahead-three step behind circular world for the Designs 2 and 5. Cross-sectional dimension $n = 50$, time series dimension $T = 200$. $\mathtt{M} = 1000$ Monte Carlo steps.

Given the above results with the simulation designs proposed in literature, it seems to be the case that although there are theoretical concerns regarding the $DOLS$ and the two-stage least squares estimator, working the with the $DOLS$ or $2SLS$ estimator and the corresponding Wald statistic is sufficient when the model described by (14) should be estimated. However, in a slightly parallel way as to construct examples where the ordinary least squares and the two stage least squares estimator do not result in the same inference, we simply have to construct examples where the noise term $\mathbf{u}$ and the explanatory variables $\mathbf{X}$ are correlated. This can be done as follows: (i) Increase the correlation between $\tilde{\mathbf{y}}^*_{it}$ and u_{it}: We can stick to the above designs for $\mathbf{w}_{it}$ and Σ_ε and replace the spatial correlation matrix designs with a matrix $\mathbf{W}$ having non-zero entries in all off diagonal positions. As an example we take the input-output data used in Section 5. From this 148×148 matrix obtained in Section 5 we take the first $n \times n$ components and normalize this matrix such that the largest eigenvalue of this normalized matrix is equal to one. The first part of Table 4 already shows that here the $D2SLS$ estimator outperforms the $DOLS$ estimator. (ii) Increase the correlation between $\tilde{\mathbf{x}}_{it}$ and u_{it}: E.g. we can use the $VAR(1)$ designs 6 and 7:

$$\mathbf{\Phi} = \begin{pmatrix} 0 & 0.95 & 0 \\ 0.95 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \mathbf{\Phi} = \begin{pmatrix} 0 & 0.45 & 0.45 \\ 0.45 & 0 & 0.45 \\ 0.45 & 0.45 & 0 \end{pmatrix}. \tag{39}$$

The results obtained in a Monte Carlo study are presented in last two lines of Table 4, where we observe that with these designs the performance of the $D2SLS$ is slightly better than the performance of the $2SLS$ estimator. We call the performance slightly better since for a critical value $\alpha_c = 0.01$ the number of rejections is much too low for the $D2SLS$ estimator. By these final simulation runs we conclude that although the size with the $D2SLS$ does not perfectly fit to the critical levels, we can also construct data generating processes where the performance of the $D2SLS$ estimator is also superior in the finite sample compared to the $DOLS$ and the $2SLS$ estimator. These last simulation runs justify the application of the $D2SLS$ estimator when model (14) is applied.

Design	Σ_ε	$\alpha_3 = 0.01$ *DOLS*	*D2SLS*	*D2SLS*	$\alpha_c = 0.05$ *DOLS*	*D2SLS*	*D2SLS*	$\alpha_c = 0.10$ *DOLS*	*D2SLS*	*D2SLS*
1	2	1.00	9.00	5.00	7.00	22.00	7.00	9.00	35.00	15.00
2	2	2.00	8.00	3.00	8.00	18.00	10.00	11.00	25.00	16.00
3	2	5.00	4.00	2.00	8.00	11.00	11.00	18.00	18.00	17.00
4	2	4.00	12.00	3.00	8.00	24.00	7.00	11.00	31.00	14.00
5	2	4.00	15.00	4.00	8.00	22.00	10.00	14.00	35.00	18.00
6	2	3.00	2.00	1.00	12.00	7.00	3.00	25.00	16.00	8.00
7	2	2.00	3.00	0.01	9.00	10.00	4.00	11.00	13.00	9.00

Table 3: Size for the parameter ρ for $\mathbf{W}$ obtained from empirical data. Rejections of the true null hypothesis $\rho = 0$ in percentage terms, given the significance levels $\alpha_c = \{0.01, 0.05, 0.1\}$. 1000 Simulation runs. Cross-sectional dimension $n = 50$, time series dimension $T = 200$. $\mathtt{M} = 1000$ Monte Carlo runs.

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 the 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 (14) by the $D2SLS$ approach. Similar to Berndt et al. (2008) the left hand side variable is the CDS spread, while firm specific credit risk proxies and the VIX volatility index are used as the right hand side variables. By means of the matrix $\mathbf{W}$ we model some 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. While 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.

5.1 Data

In this analysis CDS spreads are used to describe the implied credit risk of a firm.[5] 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), 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 said to be 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 details we use 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.

To apply and estimate the spatial autocorrelation model the spatial weights matrix $\mathbf{W}$ has to be

[5] With a CDS contract a protection buyer acquires insurance against the default of a specified entity. 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`.

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

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 the proportion of inputs ultimately stemming from each other industry relative to its own sales. We use this to proxy for possible correlation of shocks coming through the supply chain. The resulting weights matrix thus approximates the possible correlation patterns due to technology and demand shocks working their way through the economy. The elements along the main diagonal are set to zero. To improve the numerical properties and to be able to interpret the estimated coefficients, we normalize our spatial weights matrix by its largest absolute eigenvalue. As a result, the spatial autocorrelation parameter in our model is bounded by one from above and thus has the usual range as in the time series autocorrelation models. I.e. with the zeros in the main diagonal and $\rho \in (-1, 1)$, the requirements of Assumption 1 are fulfilled. ρ can be interpreted like a usual correlation parameter.

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 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 4. 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. With the interest rate the effect is not so clear. If the interest rate increases the cost of capital increases for a leveraged firm. This should drive up the CDS spread. However, also banks are in our data set, where this impact could also be different. 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 we expect that CDS spreads are positively correlated, so we expect a positive ρ when (14) is estimated.

By Assumption 2 the explanatory variables $\mathbf{x}_{it}$ should be $I(1)$. The question arises whether this assumption is met by the data. For the variables considered the distance to default 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)). By construction $DD_{it} \geq 0$. Translated to discrete time the distance to default should follow a random walk with an absorbing barrier; only firms where this barrier is not hit are observed in the data set. By construction the debt to value ration lives on the interval $[0, 100]$, the VIX index measures volatility and is therefore be non-negative. Following applied literature and running augmented Dickey-Fuller tests on a unit root for these data show the following: The null of a unit root is not rejected for almost all time series on a five percent significance level for the CDS spreads. Im, Pesaran and Shin tests implemented in the *EViews* package provide us with 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 stronger evidence for the presence of a unit root. Based on this discussion the data considered cannot exactly match the conditions of Assumption 2. Based on the time series properties we suppose that the model considered in Section 2 still provides us with a useful approximation of the (unknown) data generating process of the empirical data considered.

$\hat{\gamma}$	*2SlS*		*DOLS*		*D2SlS*	
ρ	0.3920	0.0331	0.5094	< 0.001	0.3951	< 0.001
β_{DD}	-15.2656	0.4317	-19.4116	< 0.001	-20.1287	< 0.001
β_{DVR}	5.4718	< 0.001	5.1892	< 0.001	5.3026	< 0.001
β_{r_2}	4.2952	0.0025	8.1290	0.0103	8.0172	0.0159
$\beta_{r_{10}}$	-42.4143	0.8117	-45.8192	0.0000	-48.6698	< 0.001
β_{VIX}	-0.0233	< 0.001	-0.1584	0.0673	-0.1436	0.2097

Table 5: Parameter Estimates: Model (14) applied to CDS data. y_{it} is CDS data 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. $T = 230$, $N = 148$, $k = 5$ and $k_\rho = 2$.

5.2 Results

Equipped with our data set we estimate the parameter vector γ by means of two-stage least squares, *DOLS* and the *D2SLS*. The results are presented in Table 5. Based on the theoretical considerations of the former chapters the *D2SLS* estimator should be used, the results from the other estimation methods are for comparison. When instrumental variables are used in the estimation, the debt-to-value ratio and the VIX are used in $\sum_{i=1}^{n} \mathbf{W}\tilde{\mathbf{x}}_{it\iota}$, i.e. $k_\rho = 2$. 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 5 are obtained by means of a Wald test. For the distance to default and the debt to value ratio the parameters are highly significant, also the signs expected have been realized. Both interest rates are significant, where the short term interest rate r_{2t} increases the CDS spread, while the long term interest rate decreases the spread. In contrast to 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 small, 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 that include 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 to construct 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. Our analysis shows that this spatial correlation is highly significant.

A Proof of Theorem 1

The two stage least squares estimator is given by (25) and hence

$$
\begin{aligned}
(\widehat{\gamma', \delta'}')'_{D2SLS} - (\gamma', \delta')' &= \left(\mathbf{X}'\mathbf{P}_H\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_H\mathbf{u}, \\
&\quad \text{where with } \mathbf{P}_H = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}' \text{ we get} \\
&= \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{u}.
\end{aligned} \tag{40}
$$

$\mathbf{X}$ 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 of $\mathbf{P}_H(x_t, \zeta_t) = (x_t, \zeta_t)$. This yields

$$
\mathbf{X}'\mathbf{P}_H = \begin{pmatrix} \mathbf{y}^* \\ \text{---} \\ \mathbf{x}' \\ \text{---} \\ \zeta' \end{pmatrix} \mathbf{P}_H = \begin{pmatrix} \mathbf{y}^*\mathbf{P}_H \\ \text{-------} \\ \mathbf{x}' \\ \text{-------} \\ \zeta' \end{pmatrix}. \tag{41}
$$

Additionally, $\mathbf{Z}_{it}$ is 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 $\mathbf{Z}_{it}$ containt ζ_{it}. The "non-it elements" of this vector are zero. In the same way we obtain $\mathbf{X}_{it}$ 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 $d_x = k + 1$. $d_x \leq d_z$ hold throughout the following analysis.

`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 $\frac{1}{T}$, the remaining terms (accounting for ζ_t) are multiplied by $\frac{1}{\sqrt{T}}$. 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:

$$\begin{aligned}
\mathbf{M}^{\star}_{nT} \quad = \quad & \begin{pmatrix} \mathbf{y}^{*\prime}\mathbf{x}^*/(T^2n) & \mathbf{y}^{*\prime}\mathbf{x}/(T^2n) & \mathbf{y}^{*\prime}\zeta/(T\sqrt{T}) \\ \mathbf{x}'\mathbf{x}^*/(T^2n) & \mathbf{x}'\mathbf{x}/(T^2n) & \mathbf{x}'\zeta/(T\sqrt{T}) \\ \zeta'\mathbf{x}^*/(T\sqrt{T}) & \zeta'\mathbf{x}/(T\sqrt{T}) & \zeta'\zeta/(\sqrt{T}\sqrt{T}) \end{pmatrix} \cdot \\
& \cdot \begin{pmatrix} \mathbf{x}^{*\prime}\mathbf{x}^*/(T^2n) & \mathbf{x}^{*\prime}\mathbf{x}/(T^2n) & \mathbf{x}^{*\prime}\zeta/(T\sqrt{T}) \\ \mathbf{x}'\mathbf{x}^*/(T^2n) & \mathbf{x}'\mathbf{x}/(T^2n) & \mathbf{x}'\zeta/(T\sqrt{T}) \\ \zeta'\mathbf{x}^*/(T\sqrt{T}) & \zeta'\mathbf{x}/(T\sqrt{T}) & \zeta'\zeta/(\sqrt{T}\sqrt{T}) \end{pmatrix}^{-1} \cdot \\
& \cdot \begin{pmatrix} \mathbf{x}^{*\prime}\mathbf{y}/(T^2n) & \mathbf{x}^{*\prime}\mathbf{x}^*/(T^2n) & \mathbf{x}^*\zeta'/(T\sqrt{T}) \\ \mathbf{x}'\mathbf{y}^*/(T^2n) & \mathbf{x}'\mathbf{x}/(T^2n) & \mathbf{x}'\zeta/(T\sqrt{T}) \\ \zeta'\mathbf{y}^*/(T\sqrt{T}) & \zeta'\mathbf{x}/(T\sqrt{T}) & \zeta'\zeta/(\sqrt{T}\sqrt{T}) \end{pmatrix} , \qquad (42)
\end{aligned}$$

where the $T \to \infty$ limit of $\mathbf{M}^{\star}_{nT}$ is denoted by $\mathbf{M}^{\star}_{n}$. In addition we define the $d_x \times d_x$ matrix $\mathbf{M}_{nT}$ which is described by:

$$\begin{aligned}
\mathbf{M}_{nTi} \quad &= \quad \left(\frac{1}{T^2 n} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{X}_{it,1:d_x} \mathbf{Z}'_{it,1:d_z} \right) \left(\frac{1}{T^2 n} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbf{Z}_{it,1:d_z} \mathbf{Z}'_{it,1:d_z} \right)^{-1} \frac{1}{T^2} \sum_{t=1}^{T} \mathbf{Z}'_{it,1:d_z} \mathbf{X}_{it,1:d_x} \\
\mathbf{M}_{nT} \quad &= \quad \frac{1}{n} \sum_{i=1}^{n} \mathbf{M}_{nTi} \ . \qquad (43)
\end{aligned}$$

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

Remark 6. In Remark 5 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 $\mathbf{M}_{nTi} = \frac{1}{T^2} \sum_{t=1}^{T} \mathbf{Z}'_{it,1:d_z} \mathbf{X}_{it,1:d_x}$. 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

$$\frac{1}{T^2}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it\lambda}\tilde{\mathbf{x}}_{jt\iota} \xrightarrow{d} \int \tilde{\mathcal{B}}_{vi\lambda}\tilde{\mathcal{B}}_{vj\iota}\ ,\ \frac{1}{T^\kappa}\sum_{t=1}^{T}\tilde{\zeta}_{it\lambda}\tilde{\mathbf{x}}_{jt\iota} \xrightarrow{p} 0\ ,\ \frac{1}{T^\kappa}\sum_{t=1}^{T}\tilde{u}_{it\lambda}\tilde{\mathbf{x}}_{jt\iota} \xrightarrow{p} 0\ , \tag{44}$$

where $\kappa \geq \frac{3}{2}$. Thus the terms of the form of $\frac{1}{T^2}\sum_{t=1}^{T}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it} \xrightarrow{d} \int \tilde{\mathcal{B}}_{vi}\tilde{\mathcal{B}}'_{vi}$.[6] In addition we meet terms of the structure

$$\frac{1}{\sqrt{T}\sqrt{T}}\sum_{t=1}^{T}\zeta_{it}\zeta_{it} \xrightarrow{p} \begin{cases} \Gamma^{\zeta}_{vv,ij} & \text{for } j = -2p,\ldots,2p \text{ and} \\ \mathbf{0}_{k\times k} & \text{else.} \end{cases} \tag{45}$$

$\frac{1}{\sqrt{T}\sqrt{T}}\sum_{t=1}^{T}\zeta_{it}\zeta_{jt}$ converges to a matrix of zeros by the independence across i assumption (i.e. Assumption 2). For each fixed $i = 1,\ldots,n$, the $(2p+1)k \cdot n \times (2p+1)k \cdot n$ matrix $\Gamma^{\zeta}_{vv,ij}$ contains the $k \times k$ covariance matrices $\Gamma_{vv,ij}$.

Consider the now terms in (42). 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 $\Gamma_{vv,ij}$. 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 $\mathbf{M}^\star_n$ is block diagonal. From $\mathbf{M}^\star_n$ we can extract the matrix $\mathbf{M}^\star_{ni}$ focusing on the index i. The limit of the submatrix $[\mathbf{M}^\star_{ni}]_{(1:k+1,1:k+1)}$ is $\mathbf{M}_{ni}$ while the limit of $[\mathbf{M}^\star_n]_{(1:k+1,1:k+1)}$ is $\mathbf{M}_n$. $[\mathbf{M}^\star_{ni}]_{(k+2:k+1+(2p+1)k\cdot n,k+2:k+1+(2p+1)k\cdot n)}$ is a block diagonal matrix consisting of $\boldsymbol{\Gamma}^{\zeta}_{vv,ij}$. 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*

[6] 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

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\zeta}_{it\lambda}\tilde{\mathbf{x}}_{jt\iota} \xrightarrow{d} \int d\tilde{\mathcal{B}}_{vi\lambda}\tilde{\mathcal{B}}_{vj\iota} \text{ and }\ ,\ \frac{1}{T}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{\mathbf{x}}_{jt\iota} \xrightarrow{d} \sqrt{\Omega_{uu,i}}\int \tilde{\mathcal{B}}_{vj\iota}d\mathcal{W}_{ui}\ .$$

A random variable convergent in distribution is bounded in probability, or $\mathcal{O}_p(1)$ in Landau notation. We can now consider $\frac{1}{T^{3/2}}\sum_{t=1}^{T}\tilde{\zeta}_{it\lambda}\mathbf{x}_{jt\iota}$ as the product $a \cdot b$, where $a = \frac{1}{\sqrt{T}}$ and $b = \frac{1}{T}\sum_{t=1}^{T}\tilde{\zeta}_{it\lambda}\mathbf{x}_{jt\iota}$. Since a is converging to zero, it is $o(1)$ and therefore also $o_p(1)$. b converges in distribution and therefore it 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].

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 $\mathbf{X}'\mathbf{P}_H\mathbf{u}$. Let us define the following terms:

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

$$\mathbf{m}^\star_{nT} = \begin{pmatrix} \mathbf{y}^{*\prime}\mathbf{x}^*/(T^2 n) & \mathbf{y}^{*\prime}\mathbf{x}/(T^2 n) & \mathbf{y}^{*\prime}\zeta/(T\sqrt{T}) \\ \mathbf{x}'\mathbf{x}^*/(T^2 n) & \mathbf{x}'\mathbf{x}/(T^2 n) & \mathbf{x}'\zeta/(T\sqrt{T}) \\ \zeta'\mathbf{x}^*/(T\sqrt{T}) & \zeta'\mathbf{x}/(T\sqrt{T}) & \zeta'\zeta/(\sqrt{T}\sqrt{T}) \end{pmatrix} \cdot \tag{46}$$

$$\cdot \begin{pmatrix} \mathbf{x}^{*\prime}\mathbf{x}^*/(T^2 n) & \mathbf{x}^{*\prime}\mathbf{x}/(T^2 n) & \mathbf{x}^{*\prime}\zeta/(T\sqrt{T}) \\ \mathbf{x}'\mathbf{x}^*/(T^2 n) & \mathbf{x}'\mathbf{x}/(T^2 n) & \mathbf{x}'\zeta/(T\sqrt{T}) \\ \zeta'\mathbf{x}^*/(T\sqrt{T}) & \zeta'\mathbf{x}/(T\sqrt{T}) & \zeta'\zeta/(\sqrt{T}\sqrt{T}) \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{x}^{*\prime}/(T\sqrt{n}) \\ \mathbf{x}'/(T\sqrt{n}) \\ \zeta'/\sqrt{T} \end{pmatrix} \mathbf{u} \,.$$

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

$$\begin{aligned} \mathbf{m}_{nTi} &= \left(\frac{1}{T^2 n}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{X}_{it,1:d_x}\mathbf{Z}'_{it,1:d_z}\right)\left(\frac{1}{T^2 n}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{it,1:d_z}u_{it}, \\ \mathbf{m}_{nT} &= \frac{1}{\sqrt{n}}\mathbf{m}_{nTi} \,. \end{aligned} \tag{47}$$

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

The first and the second matrix have already been considered with (42), 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 $\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{it}u_{it}$. 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. Since $[\mathbf{M}^\star_n]_{(k+2:k+1+(2p+1)k\cdot n,k+2:k+1+(2p+1)k\cdot n)}$ 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 details:

$$\mathbf{M}_{ZZ,nTi} := \frac{1}{T^2}\sum_{t=1}^{T}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z} \tag{48}$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}(\sum_{j=1}^{n} W_{ij}^{\tau_1} x_{jt1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}} \widetilde{x}_{jtq_\rho}, \widetilde{x}'_{it})'(\sum_{j=1}^{n} W_{ij}^{\tau_1} \widetilde{x}_{jt1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}} \widetilde{x}_{jtq_\rho}, \widetilde{\mathbf{x}}'_{it})$$

Using the functional central limit and the continuous mapping theorem we derive

$$\frac{1}{T^2}\sum \mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z} \xrightarrow{d} \mathbf{M}_{ZZ,ni} \tag{49}$$

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

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

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

$$\mathbf{M}_{ZZ,ni,(q_\rho+1:q_\rho+k,1:q_\rho)} := \mathbf{M}'_{ZZ,ni,(1:q_\rho,q_\rho+1:q_\rho+k)}$$

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

Based on (49) we arrive at

$$\mathbf{M}_{ZZ,Ti} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}_{ZZ,nTi} \xrightarrow{d} \mathbf{M}_{ZZ,n} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}_{ZZ,ni} \ . \tag{50}$$

In a similar way we derive the limit of

$$\mathbf{X}_{it,1:d_x}\mathbf{Z}'_{it,1:d_z} = (\sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt}, \widetilde{\mathbf{x}}'_{it})'(\sum_{j=1}^{n} W_{ij}^{\tau_1}\widetilde{x}_{jt1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}}\widetilde{x}_{jtq_\rho}, \widetilde{\mathbf{x}}'_{it}) \ . \tag{51}$$

Based on (51) we arrive at the $k+1 \times k+q_\rho$ matrix

$$\begin{aligned}
\mathbf{M}_{XZ,nTi} &= \frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{j=1}^{n} W_{ij}\widetilde{y}_{jt}, \widetilde{\mathbf{x}}'_{it}\right)'\left(\sum_{j=1}^{n} W_{ij}^{\tau_1}\widetilde{\mathbf{x}}_{jt1}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}}\widetilde{\mathbf{x}}_{jtq_\rho}, \widetilde{\mathbf{x}}'_{it}\right) \qquad (52)\\
&= \left(\begin{array}{c|c} \mathbf{M}_{XZ,nTi,(1:1,1:q_\rho)} & \mathbf{M}_{XZ,nTi,(1:1,q_\rho+1:k+q_\rho)} \\ \hline \mathbf{M}_{XZ,nTi,(2:k+1,1:q_\rho)} & \mathbf{M}_{XZ,nTi,(2:k+1,q_\rho+1:k+q_\rho)} \end{array}\right) \ , \\
&\mathbf{M}_{XZ,nTi,(1:1,1:q_\rho)} := \\
&\frac{1}{T^2}\sum_{t=1}^{T}\left(\begin{array}{c} \sum_{j=1}^{n} W_{ij}^{\tau_1}\widetilde{x}_{jt1}\cdot\sum_{j=1}^{n}\left[\sum_{l=1}^{n} W_{il}K_{lj}\right]\left(\beta'\widetilde{\mathbf{x}}_{jt}+\delta_j\widetilde{\zeta}_{jt}+\widetilde{u}_{jt}\right) \\ \vdots \\ \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}} x_{jtq_\rho}\cdot\sum_{j=1}^{n}\left[\sum_{l=1}^{n} W_{il}K_{lj}\right]\left(\beta'\widetilde{\mathbf{x}}_{jt}+\delta_j\widetilde{\zeta}_{jt}+\widetilde{u}_{jt}\right) \end{array}\right)' \ , \\
&\mathbf{M}_{XZ,nTi,(1:1,q_\rho+1:k+q_\rho)} := \frac{1}{T^2}\sum_{t=1}^{T} \ \widetilde{\mathbf{x}}'_{it}\cdot\sum_{j=1}^{n}\left[\sum_{l=1}^{n} W_{il}K_{lj}\right]\left(\beta'\widetilde{\mathbf{x}}_{jt}+\delta_j\widetilde{\zeta}_{jt}+\widetilde{u}_{jt}\right) \ , \\
&\mathbf{M}_{XZ,nTi,(2:k+1,1:q_\rho)} := \frac{1}{T^2}\sum_{t=1}^{T}\left(\widetilde{\mathbf{x}}_{it}\cdot\sum_{i=1}^{n} W_{ij}^{\tau_1}\widetilde{x}_{jt1}, \ldots, \widetilde{\mathbf{x}}_{it}\cdot\sum_{i=1}^{n} W_{ij}^{\tau_{q_\rho}}\widetilde{x}_{jtq_\rho}\right) \ , \\
&\mathbf{M}_{XZ,nTi,(2:k+1,q_\rho+1:k+q_\rho)} := \frac{1}{T^2}\sum_{t=1}^{T}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}'_{it} \ .
\end{aligned}$$

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

$$\mathbf{M}_{XZ,nTi,(1:1,1:q_\rho)} \overset{d}{\to}$$
$$\begin{pmatrix} \sum_{\kappa=1}^n W_{i\kappa}^{\tau_1} \sum_{j=1}^n \left[\sum_{l=1}^n W_{il} K_{lj}\right] \left(\beta' \tilde{\mathcal{B}}_{vj} \tilde{\mathcal{B}}_{v\kappa 1}\right) \\ \vdots \\ \sum_{\kappa=1}^n W_{i\kappa}^{\tau_{q_\rho}} \sum_{j=1}^n \left[\sum_{l=1}^n W_{il} K_{lj}\right] \left(\beta' \tilde{\mathcal{B}}_{vj} \tilde{\mathcal{B}}_{v\kappa q_\rho}\right) \end{pmatrix}' = \mathbf{M}_{XZ,ni,(1:1,1:q_\rho)} \ . \tag{53}$$
$$\mathbf{M}_{XZ,nTi,(1:1,q_\rho+1+q_\rho+k)} \overset{d}{\to}$$
$$\begin{pmatrix} \sum_{j=1}^n \left[\sum_{l=1}^n W_{il} K_{lj}\right] \left(\beta' \tilde{\mathcal{B}}_{vj} \tilde{\mathcal{B}}_{vi1}\right) \\ \sum_{j=1}^n \left[\sum_{l=1}^n W_{il} K_{lj}\right] \left(\beta' \tilde{\mathcal{B}}_{vj} \tilde{\mathcal{B}}_{vi2}\right) \\ \vdots \\ \sum_{j=1}^n \left[\sum_{l=1}^n W_{il} K_{lj}\right] \left(\beta' \tilde{\mathcal{B}}_{vj} \tilde{\mathcal{B}}_{vik}\right) \end{pmatrix}' = M_{XZ,ni,(1:1,q_\rho+1+q_\rho+k)} \ . \tag{54}$$
$$\mathbf{M}_{XZ,nTi,(2+k+1,1:q_\rho)} \overset{d}{\to} \tag{55}$$
$$\left(\sum_{j=1}^n W_{ij}^{\tau_1} \int \tilde{\mathcal{B}}_{vi} \tilde{\mathcal{B}}_{vj1}, \ldots, \sum_{j=1}^n W_{ij}^{\tau_{q_\rho}} \int \tilde{\mathcal{B}}_{vi} \tilde{\mathcal{B}}_{vjq_\rho}\right) = \mathbf{M}_{XZ,ni,(2:k+1,1:q_\rho)} \ . \tag{56}$$
$$\mathbf{M}_{XZ,nTi,(2+k+1,q_\rho+1:q_\rho+k)} \overset{d}{\to}$$
$$\int \tilde{\mathcal{B}}_{vi} \tilde{\mathcal{B}}_{vi}' = M_{XZ,ni,(2:k+1,q_\rho+1:+q_\rho+k)} \tag{57}$$

Note that $\tilde{\mathcal{B}}_{vj\cdot}$ is a scalar, $\tilde{\mathcal{B}}_{vi}$ a k dimensional vector. Summing up we arrive at

$$\begin{aligned} \mathbf{M}_{XZ,nTi} &\overset{d}{\to} \mathbf{M}_{XZ,ni} \ , \\ \mathbf{M}_{XZ,nT} = \frac{1}{n}\sum_{i=1}^n \mathbf{M}_{XZ,nTi} &\overset{d}{\to} \mathbf{M}_{XZ,n} = \frac{1}{n}\sum_{i=1}^n \mathbf{M}_{XZ,ni} \ . \end{aligned} \tag{58}$$

Remark 7. 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 $\frac{1}{T}\sum_{t=1}^T \mathbf{X}_{it,2:d_x} u_{it} = \frac{1}{T}\sum_{t=1}^T \mathbf{x}_{it,1:d_x} u_{it}$ the $T \to \infty$ 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 $\frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{it,1:d_z}u_{it}$ converges in distribution to

$$\mathbf{m}_{niZu} = \sqrt{\Omega_{uu,i}}\left(\sum_{j=1}^{n} W_{ij}^{\tau_1}\int \tilde{\mathcal{B}}_{vj1} d\mathcal{W}_{ui}, \ldots, \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}}\int \tilde{\mathcal{B}}_{vjq_\rho} d\mathcal{W}_{ui}, \left(\int \tilde{\mathcal{B}}_{vi} d\mathcal{W}_{ui}\right)'\right)' . \tag{59}$$

To obtain the first term of $\mathbf{m}_n$ we have to combine $\mathbf{M}_{ZZ,n}$ provided by (50), $\mathbf{M}_{XZ,ni}$ given by (58) and $\mathbf{m}_{niZU}$. Then the continuous mapping theorem yields

$$\mathbf{m}_{nT} \xrightarrow{d} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sqrt{\Omega_{uu,i}}\left(\mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}\begin{pmatrix}\sum_{j=1}^{n} W_{ij}^{\tau_1}\int \tilde{\mathcal{B}}_{vj1} d\mathcal{W}_{ui}\\ \vdots \\ \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}}\int \tilde{\mathcal{B}}_{vjq_\rho} d\mathcal{W}_{ui}\\ \int \tilde{\mathcal{B}}_{vi} d\mathcal{W}_{ui}\end{pmatrix}\right) . \tag{60}$$

$\mathbf{M}_{ZZ,n}$ 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 (41)), the rows $2:k+1$ have to be equal to the limit of $\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_{it,2:d_x}u_{it} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it,1:d_x}u_{it}$. This yields

$$\mathbf{m}_{nT} \xrightarrow{d} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sqrt{\Omega_{uu,i}}\begin{pmatrix}\left[\mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}\begin{pmatrix}\sum_{j=1}^{n} W_{ij}^{\tau_1}\int \tilde{\mathcal{B}}_{vj1} d\mathcal{W}_{ui}\\ \vdots \\ \sum_{j=1}^{n} W_{ij}^{\tau_{q_\rho}}\int \tilde{\mathcal{B}}_{vjq_\rho} d\mathcal{W}_{ui}\\ \int \tilde{\mathcal{B}}_{vi} d\mathcal{W}_{ui}\end{pmatrix}\right]_{1,1}\\ \hline \int \tilde{\mathcal{B}}_{vi} d\mathcal{W}_{ui}\end{pmatrix} = \mathbf{m}_n . \tag{61}$$

It remains to calculate the limit distribution of (42). 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 $1/T$ and $1/\sqrt{n}$). Using the above results and the continuous mapping theorem we get

$$\begin{aligned} \mathbf{M}_{ni} &= \mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}\mathbf{M}'_{XZ,ni}, \qquad (62)\\ \mathbf{M}_{n} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}_{ni} = \frac{1}{n}\mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}\mathbf{M}'_{XZ,n}\ . \qquad (63) \end{aligned}$$

This yields the *second result*: $(\gamma', \delta')'$ can be consistently estimated. $\sqrt{n}T\left(\hat{\gamma}_{D2SLS} - \gamma\right)$ 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 (61) and (62), 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) and Johansen (1995) 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 B_{vi} fixed for all $i = 1, \dots, n$: (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[7]

$$\begin{aligned} \mathbf{V}_n &= \frac{1}{n}\sum_{i=1} V_{ni} \text{ where } \mathbf{V}_{ni} = \Omega_{uu,i}\cdot\tilde{\Upsilon}_{ni}\ . \qquad (64)\\ &\tilde{\Upsilon}_{ni} = \int\left[\begin{pmatrix}\left[\mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}m_{Z,n}\right]_{(1,1)}\\ \tilde{\mathcal{B}}_{vi}\end{pmatrix}\begin{pmatrix}\left[\mathbf{M}_{XZ,n}\mathbf{M}_{ZZ,n}^{-1}m_{Z,n}\right]_{(1,1)}\\ \tilde{\mathcal{B}}_{vi}\end{pmatrix}'\right],\\ &m_{Z,n} = \left(\sum_{j=1}^{n}W_{ij}^{\tau_1}\tilde{\mathcal{B}}_{vj1}, \dots, \sum_{j=1}^{n}W_{ij}^{\tau_{q_\rho}}\tilde{\mathcal{B}}_{vjq_\rho}, \tilde{\mathcal{B}}'_{vi}\right)' \qquad (65) \end{aligned}$$

---

[7]Note that as with m_{nT} the term $\mathbf{M}_{XZ,Tn}\mathbf{M}_{ZZ,Tn}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{j=1}^{n}W_{ij}^{\tau_1}Z_{it,1},\dots,\sum_{j=1}^{n}W_{ij}^{\tau_{q_\rho}}Z_{it,q_\rho},x'_{it}\right)'$ is equal to $\left(\left[\mathbf{M}_{XZ,Tn}\mathbf{M}_{ZZ,Tn}^{-1}\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{j=1}^{n}W_{ij}^{\tau_1}Z_{it,1},\dots,\sum_{j=1}^{n}W_{ij}^{\tau_{q_\rho}}Z_{it,q_\rho},x'_{it}\right)'\right]_{1,1},x'_{it}\right)'$ by the fact the projection P_W applied to x_{it} is x_{it}.

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

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

An estimate $\mathbf{V}_{nT}$ of $\mathbf{V}_n$ is derived by means of

$$\begin{aligned}\mathbf{V}_{nT} &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\Omega}_{uu,i}\frac{1}{T^2}\sum_{t=1}^{T}\Upsilon_{nTi}\Upsilon_{nTi}' \\ \Upsilon_{nTi} &= \begin{pmatrix}\left[\left(\sum_{l=1}^{n}\sum_{t=1}^{T}\mathbf{X}_{lt,1:d_x}\mathbf{Z}_{lt,1:d_z}'\right)\left(\sum_{j=1}^{n}\sum_{t=1}^{T}\mathbf{Z}_{jt,1:d_z}\mathbf{Z}_{jt,1:d_z}'\right)^{-1}\mathbf{Z}_{it,1:d_z}\right]_{(1,1)} \\ \mathbf{x}_{it}\end{pmatrix} \ .\end{aligned} \tag{67}$$

Combining (67) and $\mathbf{M}_{nT}$, which is an estimate of $\mathbf{M}_n$, we arrive at an estimate of the covariance matrix

$$\mathbf{D}_{nT} = \mathbf{M}_{nT}^{-1}\mathbf{V}_{nT}\mathbf{M}_{nT}^{-1} \ . \tag{68}$$

Equipped with these terms we obtain

$$\begin{aligned} S_{\gamma,nT} &= \left(\sqrt{n}T\mathbf{R}\left(\widehat{\gamma}_{D2SLS}-\gamma\right)\right)'\left(\mathbf{R}\mathbf{D}_{nT}\mathbf{R}'\right)^{-1}\left(\sqrt{n}T\mathbf{R}\left(\widehat{\gamma}_{D2SLS}-\gamma\right)\right) \\ S_{\gamma,nT} &\overset{d}{\to} S_{\gamma,n} = \left(\sqrt{n}T\mathbf{R}\left(\widehat{\gamma}_{D2SLS}-\gamma\right)\right)'\left(\mathbf{R}\mathbf{D}_{n}\mathbf{R}'\right)^{-1}\left(\sqrt{n}T\mathbf{R}\left(\widehat{\gamma}_{D2SLS}-\gamma\right)\right) \ .\end{aligned} \tag{69}$$

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 $\mathbf{D}_n$.

B The Rank Condition and the Instruments $\sum_{j=1}^{n} W_{ij}\tilde{x}_{jtv}$

Lemma 1. Given the model assumptions of Section 2. Suppose that the $\tau_v = 1$ 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. The rank of the $k+1 \times q_\rho + k$ matrices $M_{XZ,ni}$ and $M_{XZ,n}$ is $k+1$ almost surely.

Proof. We follow Phillips and Hansen (1990)[Lemma 3] to show that the rows of the corresponding matrices are independent. Thus, consider the vectors

$$\mathbf{Z}_{it,1:d_z} = \begin{pmatrix} \sum_{j=1}^{n} W_{ij}x_{jt1} \\ \vdots \\ \sum_{j=1}^{n} W_{ij}x_{jtq_\rho} \\ x_{it1} \\ \vdots \\ x_{itk} \end{pmatrix} \text{ and } \mathbf{X}_{it,1:d_x} = \begin{pmatrix} \sum_{j=1}^{n} W_{ij}y_{jt} \\ x_{it1} \\ \vdots \\ x_{itk} \end{pmatrix}.$$

of dimension $q_\rho + k$ and $1 + k$, respectively. The corresponding transpose vectors are:

$$\begin{aligned} \mathbf{Z}'_{it,1:d_z} &= \left(\sum_{j=1}^{n} W_{ij}x_{jt1}, \ldots, \sum_{j=1}^{n} W_{ij}x_{jtq_\rho}, x_{it1}, \ldots, x_{itk}\right) \\ \mathbf{X}'_{it,1:d_x} &= \left(\sum_{j=1}^{n} W_{ij}y_{jt}, x_{it1}, \ldots, x_{itk}\right) . \end{aligned}$$

In the following we calculate the limits of $\frac{1}{T^2}\sum_{t=1}^{T}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}$, $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}$ and $\frac{1}{T^2}\sum_{t=1}^{T}\mathbf{X}_{it,1:d_x}\mathbf{Z}'_{it,1:d_z}$.

Let us start with $\sum_{t=1}^{T} \mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}$ where we get

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

$$\sum_{t=1}^{T}\left(\begin{array}{ccc|ccc}
\sum_{j=1}^{n} W_{ij}x_{jt1}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & \sum_{j=1}^{n} W_{ij}x_{jt1}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho} & \sum_{j=1}^{n} W_{ij}x_{jt1}x_{it1} & \cdots & \sum_{j=1}^{n} W_{ij}x_{jt1}x_{itk} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\sum_{j=1}^{n} W_{ij}x_{jtq_\rho}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & \sum_{j=1}^{n} W_{ij}x_{jtq_\rho}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho} & \sum_{j=1}^{n} W_{ij}x_{jtq_\rho}x_{it1} & \cdots & \sum_{j=1}^{n} W_{ij}x_{jtq_\rho}x_{itk} \\
\hline
x_{it1}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & x_{it1}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho} & x_{it1}x_{it1} & \cdots & x_{it1}x_{itk} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
x_{itk}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & x_{itk}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho} & x_{itk}x_{it1} & \cdots & x_{itk}x_{itk}
\end{array}\right).$$

Since $\sum_{j=1}^{n} W_{ij}x_{jtv}\sum_{j=1}^{n} W_{ij}x_{jtw} = \sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}W_{il}x_{jtv}x_{ltw}$ the matrix (70) can be written as:

$$\sum_{t=1}^{T} \mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z} = \tag{71}$$

$$\sum_{t=1}^{T}\left(\begin{array}{ccc|ccc}
\sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}W_{il}x_{jt1}x_{lt1} & \cdots & \sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}W_{il}x_{jt1}x_{ltq_\rho} & \sum_{j=1}^{n} W_{ij}x_{jt1}x_{it1} & \cdots & \sum_{j=1}^{n} W_{ij}x_{jt1}x_{itk} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}W_{il}x_{jtq_\rho}x_{lt1} & \cdots & \sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}W_{il}x_{jtq_\rho}x_{ltq_\rho} & \sum_{j=1}^{n} W_{ij}x_{jtq_\rho}x_{it1} & \cdots & \sum_{j=1}^{n} W_{ij}x_{jtq_\rho}x_{itk} \\
\hline
x_{it1}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & x_{it1}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho} & x_{it1}x_{it1} & \cdots & x_{it1}x_{itk} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
x_{itk}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & x_{itk}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho} & x_{itk}x_{it1} & \cdots & x_{itk}x_{itk}
\end{array}\right).$$

The limit of (71) divided by T^2 was abbreviated by $\mathbf{M}_{ZZ,ni}$. By the functional central limit theorem this expression converges in distribution to

$$\mathbf{M}_{ZZ,ni} = \lim_{t\to\infty}\frac{1}{T^2}\sum_{t=1}^{T} \mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z} = \tag{72}$$

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

When we consider (71) 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 (72), 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. $\sum_{i=1}^{n} W_{ij}\tilde{\mathcal{B}}_{vj1}$ for the first row, $\sum_{i=1}^{n} W_{ij}\tilde{\mathcal{B}}_{vjq_\rho}$ for row q_ρ, $\tilde{\mathcal{B}}_{vi1}$ for row $q_\rho + 1$, …and $\tilde{\mathcal{B}}_{vik}$ for row $d_z = q_\rho + k$. These "factors" are independent for

$t = 1, \dots, T$ for $\mathbf{Z}_{it,1:d_z}$. The same property is carried over to the Brownian motions where we have $\sum_{i=1}^{n} W_{ij}\tilde{\mathcal{B}}_{vjq_1}(r), \dots \tilde{\mathcal{B}}_{vik}(r)$ with $r \in [0,1]$. (Note that $\tilde{\mathcal{B}}_{vi\lambda}(r)$ and $\tilde{\mathcal{B}}_{vj\iota}(r)$ are independent (for all $\lambda \neq \iota$ for $i = j$ and all λ, ι if $i \neq j$) by assumption. In addition for each $\tilde{\mathcal{B}}_{vi\lambda}(r_1) - \tilde{\mathcal{B}}_{vi\lambda}(r_2)$ is independent from $\tilde{\mathcal{B}}_{vi\lambda}(r_2) - \tilde{\mathcal{B}}_{vi\lambda}(r_3)$, with $0 \leq r_1 < r_2 < r_3 \leq 1$ by the independent increment property of the Brownian motion.) Therefore, each row of the matrix $\mathbf{M}_{ZZ,ni}$ arises from a mixture with the independent "mixture weights" $\sum_{i=1}^{n} W_{ij}\tilde{\mathcal{B}}_{vj1}, \dots, \tilde{\mathcal{B}}_{vik}$ for row $1, \dots, q_{\rho+k}$, respectively. Each of these row vectors has dimension $q_\rho + k$. We consider $q_\rho + k$ such mixtures. Therefore, the matrix (72) has rank $q_\rho + k$ almost surely.

Remark 8. Note that $\mathbf{M}_{ZZ,nTi}$, $\mathbf{M}_{ZZ,nT}$, $\mathbf{M}_{ZZ,ni}$ are $\mathbf{M}_{ZZ,n}$ symmetric matrices.

Next we show that $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}$ still has rank $q_\rho + k$. To do this we jump back to (71) and take sums over the index i. This yields

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z} =$$

$$\left(\begin{array}{ccc|ccc}
\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{l=1}^{n} W_{ij}W_{il}x_{jt1}x_{lt1} & \cdots & \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}W_{il}x_{ltq_\rho}x_{lt1} & \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n} W_{ij}x_{jt1}x_{it1} & \cdots & \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}W_{il}x_{jtq_\rho}x_{lt1} & \cdots & \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{l=1}^{n} W_{ij}W_{il}x_{ltq_\rho}x_{ltq_\rho} & \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho}x_{it1} & \cdots & \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n} \\
\hline
\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it1}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & \sum_{i=1}^{n}\sum_{t=1}^{T} x_{it1}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho} & \sum_{i=1}^{n}\sum_{t=1}^{T} x_{it1}x_{it1} & \cdots & \sum_{i=1}^{n}\sum_{t=1}^{T} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{n}\sum_{t=1}^{T} x_{itk}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & \sum_{i=1}^{n}\sum_{t=1}^{T} x_{itk}\sum_{j=1}^{n}\sum_{t=1}^{T} W_{ij}x_{jtq_\rho} & \sum_{i=1}^{n}\sum_{t=1}^{T} x_{itk}x_{it1} & \cdots & \sum_{i=1}^{n}\sum_{t=1}^{T}
\end{array}\right.$$

Note that row v of (73) is a linear combination of $\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{Z}'_{it,1:d_z}$ with the element $\left[\mathbf{Z}_{it,1:d_z}\right]_v$. As in the case of (71) this carries over to the limit such that the rank of the $q_\rho + k \times q_\rho + k$ matrix $\mathbf{M}_{ZZ,n} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\mathbf{Z}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z}$ is $q_\rho + k$ almost surely.

In the next step we investigate the rank of $\frac{1}{T^2}\sum_{t=1}^{T}\mathbf{X}_{it,1:d_x}\mathbf{Z}'_{it,1:d_z}$. We have to show that the rank of this term is $k+1$ (a.s.). For the just identified case we meet $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\mathbf{X}_{it,1:d_x}\mathbf{Z}'_{it,1:d_z}$, we show that also the limit of this term has rank $k+1$ (a.s.). Consider

$$\sum_{t=1}^{T}\mathbf{X}_{it,1:d_z}\mathbf{Z}'_{it,1:d_z} = \tag{74}$$

$$\sum_{t=1}^{T}\left(\begin{array}{ccc|ccc}
\sum_{j=1}^{n} W_{ij}y_{jt}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & \sum_{j=1}^{n} W_{ij}y_{jt}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho} & \sum_{j=1}^{n} W_{ij}y_{jt}x_{it1} & \cdots & \sum_{j=1}^{n} W_{ij}y_{jt}x_{itk} \\
\hline
x_{it1}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & x_{it1}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho} & x_{it1}x_{it1} & \cdots & x_{it1}x_{itk} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
x_{itk}\sum_{j=1}^{n} W_{ij}x_{jt1} & \cdots & x_{itk}\sum_{j=1}^{n} W_{ij}x_{jtq_\rho} & x_{itk}x_{it1} & \cdots & x_{itk}x_{itk}
\end{array}\right).$$

Since $\sum_{j=1}^{n} W_{ij} y_{jt} \sum_{j=1}^{n} W_{ij} x_{jtw} = \sum_{j=1}^{n} \sum_{l=1}^{n} W_{ij} W_{il} y_{jt} x_{ltw}$ the matrix (74) can be written as:

$$\sum_{t=1}^{T} \mathbf{X}_{it,1:d_z} \mathbf{Z}'_{it,1:d_z} = \tag{75}$$
$$\sum_{t=1}^{T} \left(\begin{array}{ccc|ccc} \sum_{j=1}^{n} \sum_{l=1}^{n} W_{ij} W_{il} y_{jt} x_{lt1} & \cdots & \sum_{j=1}^{n} \sum_{l=1}^{n} W_{ij} W_{il} y_{jt} x_{ltq_\rho} & \sum_{j=1}^{n} W_{ij} x_{jt} x_{it1} & \cdots & \sum_{j=1}^{n} W_{ij} y_{jt} x_{itk} \\ \hline x_{it1} \sum_{j=1}^{n} W_{ij} x_{jt1} & \cdots & x_{it1} \sum_{j=1}^{n} W_{ij} x_{jt,q_\rho} & x_{it1} x_{it1} & \cdots & x_{it1} x_{itk} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{itk} \sum_{j=1}^{n} W_{ij} x_{jt1} & \cdots & x_{itk} \sum_{j=1}^{n} W_{ij} x_{jt,q_\rho} & x_{itk} x_{it1} & \cdots & x_{itk} x_{itk} \\ & & & & & \end{array} \right).$$

The limit of the sum of (75) divided by T^2 provides us with $\mathbf{M}_{XZ,ni}$. To derive this limit we meet the terms discussed in (44). Therefore by the functional central limit theorem the expression (75) converges in distribution to

$$\mathbf{M}_{XZ,ni} = \lim_{t\to\infty} \frac{1}{T^2} \sum_{t=1}^{T} \mathbf{X}_{it,1:d_x} \mathbf{Z}'_{it,1:d_z} = \tag{76}$$
$$\left(\begin{array}{ccc|ccc} \int \sum_{j=1}^{n} \sum_{l=1}^{n} K_{ij} W_{il} \beta' \tilde{\mathcal{B}}_{vj} \tilde{\mathcal{B}}_{vl1} & \cdots & \int \sum_{j=1}^{n} \sum_{l=1}^{n} K_{ij} W_{il} \beta' \tilde{\mathcal{B}}_{vj} \tilde{\mathcal{B}}_{vlq_\rho} & \int \sum_{j=1}^{n} K_{ij} \beta' \mathbf{B}_{vj} \tilde{\mathcal{B}}_{vi1} & \cdots & \int \sum_{j=1}^{n} K_{ij} \beta' \tilde{\mathcal{B}}_{vj} \tilde{\mathcal{B}}_{vik} \\ \hline \int \tilde{\mathcal{B}}_{vi1} \sum_{j=1}^{n} W_{ij} \tilde{\mathcal{B}}_{vj1} & \cdots & \int \tilde{\mathcal{B}}_{vi1} \sum_{j=1}^{n} W_{ij} \tilde{\mathcal{B}}_{vjq_\rho} & \int \mathbf{B}_{vi1} \tilde{\mathcal{B}}_{vi1} & \cdots & \int \tilde{\mathcal{B}}_{vi1} \tilde{\mathcal{B}}_{vik} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \int \tilde{\mathcal{B}}_{vjk} \sum_{j=1}^{n} W_{ij} \tilde{\mathcal{B}}_{vj1} & \cdots & \int \tilde{\mathcal{B}}_{vik} \sum_{j=1}^{n} W_{ij} \tilde{\mathcal{B}}_{vjq_\rho} & \int \mathbf{B}_{vik} \tilde{\mathcal{B}}_{vi1} & \cdots & \int \tilde{\mathcal{B}}_{vik} \tilde{\mathcal{B}}_{vik} \end{array} \right).$$

When we consider (75) we observe that row v is a linear combination of $\mathbf{Z}'_{it,1:d_z}$ with the element $\left[\mathbf{X}_{it,1:d_x}\right]_v$. This property also translates to the limit (76). Each row of the matrix $\mathbf{M}_{XZ,ni} = \lim_{t\to\infty} \frac{1}{T^2} \sum_{t=1}^{T} \mathbf{X}_{it,1:d_x} \mathbf{Z}'_{it,1:d_z}$ is a vector of dimension $q_\rho + k$. Each row is mixed with a different and independent element, e.g. $\sum_{i=1}^{n} W_{ij} \beta' \tilde{\mathcal{B}}_{vj}$ in the first row, $\tilde{\mathcal{B}}_{vi1}$ in the second row, ... and $\tilde{\mathcal{B}}_{vik}$ in row $k+1$. Therefore, the matrix (76) has rank $k+1$ almost surely.

Last but not least we show that $\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} \mathbf{X}_{it,1:d_z} \mathbf{Z}'_{it,1:d_z}$ has rank $k+1$ (a.s.). This is done by using (75) and taking sums over the index i. Then each row v is a linear combination of $\sum_{i=1}^{n} \sum_{i=1}^{T} \mathbf{Z}'_{it,1:d_z}$ with the element $\left[\mathbf{X}_{it,1:d_x}\right]_v$. As already observed with $\mathbf{M}_{ZZ,n}$ when taking limits each row is an independent mixture, such that the rank of $\mathbf{M}_{XZ,n}$ is $k+1$ almost surely.

Remark 9. 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 $\mathbf{M}_{ZX,n} = \mathbf{M}'_{XZ,n}$.

□

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 structural 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 $NumS_t$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_t 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_t) and long term debt (LD_t), 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 for the maturities m = 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 BLS Employment Projection Program (http://www.bls.gov/emp/ep_data_input_output_matrix.htm). We excluded firms where we either detected problems in data (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:

$$DVR_{it} = \left[\frac{D_{it}}{S_{it} + D_{it}}\right] \cdot 100 \ , \tag{77}$$

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. 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 an 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

$$DD_{it} = \frac{VA_{it} - DP_{it}}{VA_{it}\sigma_{Ait}} \ . \tag{78}$$

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_E is a measure of the equity volatility. Based on Crosbie and Bohn (2003)

$$\begin{aligned}
VA_{it} &= VE_{it}\mathcal{N}(d_{1i}) + \exp(-y_{tm}M)(SD_{it} + LD_{it})\mathcal{N}(d_{2i}) \\
\sigma_{Ait} &= \sigma_{Ei}\frac{VE_{it}}{VA_{it}} \\
d_{1i} &= \frac{\log\left(VA_{it}/(SD_{it} + LD_{it})\right) + \left(y_{tm} + \frac{1}{2}\sigma^2_{Ait}\right)M}{\sqrt{\sigma^2_{Ait}M}} \\
d_{2i} &= d_{1i} - \sqrt{\sigma^2_{Ait}M} \ .
\end{aligned} \tag{79}$$

Following applied literature, the standard deviation of the firm value, σ_{Ait}, is derived by an implicit estimation from the Black/Scholes formula. Following industry praxis and finance literature 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 $\sigma_{Ei}VE_{it}$, 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_{Ei}^2 is estimated from log asset returns. Here e.g. the sample variance $\hat{\sigma}_{iE}^2$ (resulting in a constant equity volatility) can be used. In this article we follow Ericsson et al. (2009) and approximate the equity volatility by means of exponential smoothing:

$$\sigma_{E,it}^2 = \lambda\sigma_{E,it-1}^2 + (1-\lambda)(\Delta \log p_{it})^2 \tag{80}$$

with $\lambda = 0.94$. $\sigma_{E,it}^2$ 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.

Brockwell, P. J. and Davis, R. A. (2006). *Time Series: Theory and Methods*. Springer Series and Statistics. Springer, New York, 2 edition.

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.

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.

Hannan, E. J. and Deistler, M. (2012). *The Statistical Theory of Linear Systems E. J. Hannan and Manfred Deistler*. SIAM, Classics and Applied Mathematics, No. 70, Philadelphia.

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.

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.

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.

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

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.

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.

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

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.

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.

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