SFB 823

Improved GMM estimation of panel data models with spatially correlated error components

Matthias Arnold, Dominik Wied

Discussion Paper

Nr. 29/2011

SFB 823

Improved GMM estimation of panel data models with spatially correlated error components

by

Matthias Arnold
Fakultät Statistik, TU Dortmund
D-44221 Dortmund, Germany
arnold@statistik.tu-dortmund.de

and

Dominik Wied *
Fakultät Statistik, TU Dortmund
D-44221 Dortmund, Germany
wied@statistik.tu-dortmund.de

This version: August 30, 2011

Abstract

We modify a previously suggested GMM estimator in a spatial panel regression model by taking into account the difference between disturbances and regression residuals and derive its asymptotic properties. Simulation results and an empirical application to Indonesian rice data illustrate the improvement in finite samples.

JEL Classification: C13, C21

Keywords: GMM estimation, panel regression, regression residuals, spatial autoregression

*Corresponding author. Phone: +49/231/755 3869, Fax: +49/231/755 5284. Financial support by Deutsche Forschungsgemeinschaft (SFB 823, project A1) is gratefully acknowledged.

1. Introduction and Summary

In this paper we consider a panel regression model where the disturbances are correlated both spatially and time-wise. To estimate the parameters of this correlation structure, Kapoor et al. (2007) suggest a GMM estimator which is a generalization of the estimator suggested by Kelejian and Prucha (1999) for the cross-sectional case. It has been used in empirical applications by many authors. Applications include multinational enterprise activity (Badinger and Egger, 2010b), export performance of Mexican states (Gamboa, 2010), effects of active labor market policies in Germany (Hujer et al., 2009) and the impact of knowledge capital stocks on total factor productivity in Europe (Fischer et al., 2009).

The statistical properties of the GMM estimator proposed by Kapoor et al. (2007) have been investigated by Larch and Walde (2009), who run a simulation study to compare the GMM estimator with a comparable ML estimator. Under normality, the GMM estimator is competitive with respect to ML. For non-normally distributed errors, the GMM estimator outperforms the quasi-ML estimator.

This paper follows up on the work on finite sample properties, i.e. we generalize an idea of Arnold and Wied (2010) for cross-sectional data to the panel case in order to improve the estimator in small and moderate samples. Our main point is the following: When calculating the GMM estimator, the unobservable disturbances of the regression model have to be replaced by the regression residuals. But then one should also calculate the theoretical moment conditions in terms of the residuals, not in terms of the disturbances. In doing so, the bias of the estimators can be essentially reduced. We point this out by some Monte Carlo evidence as well as by an analytical illustration.

As a second contribution, we derive asymptotic normality of the GMM estimators, an issue that several authors worked on in other contexts, see e.g. Lee (2004) for (quasi) ML estimation of spatial autoregressive models, Lee and Yu (2010) for ML estimation of spatial autoregressive panel data models with fixed effects and Kelejian and Prucha (2010), Badinger and Egger (2010a) and Lee and Liu (2010) for GMM estimation of spatial autoregressive models with autoregressive and heteroscedastic disturbances. Due to the nonlinear structure of the estimators, the exact finite sample distribution is unknown so that inference on the parameters has to depend on asymptotic approximations. However, the asymptotic distribution provides a good approximation to the finite sample distribution even for small sample sizes.

The remainder of the paper is organized as follows: Section 2 presents the spatial model, the estimation procedure and the analytic illustration, Section 3 provides the asymptotic results, Section 4 gives some Monte Carlo evidence and Section 5 presents an empirical application to Indonesian rice farming data which reveals the importance of our approach. Proofs are deferred to the Appendix.

2. The Model and the estimator

This paper considers a panel regression model with spatially correlated disturbances as follows:

$$
\begin{aligned}
y_N &= X_N\beta + u_N, \\
u_N &= \rho(I_T \otimes W_n)u_N + \varepsilon_N, \\
\varepsilon_N &= (e_T \otimes I_n)\mu_n + \nu_N, \\
\nu_N &= [\nu_n(1)', \ldots, \nu_n(T)']', \\
y_N &= [y_n(1)', \ldots, y_n(T)']', \\
X_N &= [X_n(1)', \ldots, X_n(T)']', \\
u_N &= [u_n(1)', \ldots, u_n(T)']', \\
\varepsilon_N &= [\varepsilon_n(1)', \ldots, \varepsilon_n(T)']',
\end{aligned}
$$

where for each time period $t = 1, \ldots, T$, $y_n(t)$ is the $n \times 1$ vector of observations on the dependent variable, $X_n(t)$ is the $n \times k$ matrix of observations on the exogenous regressors and $u_n(t)$ is the $n \times 1$ vector of spatially correlated disturbances. The serial dependence is captured by an error component structure for the innovation vector ε_N, where e_T is a $T \times 1$ vector of ones, I_T is the $T \times T$ identity matrix, μ_n is the $n \times 1$ vector of individual effects and the $N \times 1$ vector ν_N captures the remainder error terms which vary over both the cross-sectional units and the time periods.

We impose the following assumptions:

Assumption 1. *a) For all $i \in \{1, \dots, n\}$, $n \geq 1$, the $\mu_{i,n}$ are independent identically distributed with zero mean, variance σ_μ^2, $0 < \sigma_\mu^2 < b_\mu < \infty$ and finite fourth moments.*
b) For all $i \in \{1, \dots, n\}$, $n \geq 1$, $t \in \{1, \dots, T\}$, the $\nu_{it,n}$ are independent identically distributed with zero mean, variance σ_ν^2, $0 < \sigma_\nu^2 < b_\nu < \infty$ and finite fourth moments.
c) For all $i \in \{1, \dots, n\}$, $n \geq 1$, $t \in \{1, \dots, T\}$, the $\nu_{it,n}$ and $\mu_{i,n}$ are independent.

Assumption 2. *a) For all $i \in \{1, \dots, n\}$, $n \geq 1$, $w_{ii,n} = 0$ and $\sum_{j=1}^{n} w_{ij,n} = 1$.*
b) $|\rho| < 1$.

Assumption 2 ensures that the matrix $I_n - \rho W_n$ is nonsingular so that

$$\mathrm{Cov}(u_N) = \Omega_{u,N} = \left[I_T \otimes (I_n - \rho W_n)^{-1}\right] \Omega_{\varepsilon,N} \left[I_T \otimes (I_n - \rho W_n^{'})^{-1}\right], \tag{1}$$

with $\Omega_{\varepsilon,N} = \sigma_\mu^2 (J_T \otimes I_n) + \sigma_\nu^2 I_N$, where $J_T = e_T e_T^{'}$ is a $T \times T$ matrix with all elements equal to one. Kapoor et al. (2007) decompose $\Omega_{\varepsilon,N}$ as

$$\Omega_{\varepsilon,N} = \sigma_\nu^2 Q_{0,N} + \sigma_1^2 Q_{1,N},$$

where

$$\begin{aligned} Q_{0,N} &= \left(I_T - \frac{J_T}{T}\right) \otimes I_n, \\ Q_{1,N} &= \frac{J_T}{T} \otimes I_n, \end{aligned}$$

and $\sigma_1^2 = \sigma_\nu^2 + T\sigma_\mu^2$. They provide GMM estimators for ρ, σ_ν^2 and σ_1^2. Basically, we build on this approach, but with two modifications. First, we do not follow their reparameterization but estimate ρ, σ_ν^2 and σ_μ^2 directly. Of course, our estimators for σ_ν^2 and σ_μ^2 provide an estimator for σ_1^2 just as well as the estimators of Kapoor et al. (2007) for σ_ν^2 and σ_1^2 can be used to estimate σ_μ^2. The second modification exploits the difference between unobservable disturbances and observable regression residuals. For the cross-sectional case, this idea was introduced by Arnold and Wied (2010), and it also applies to the panel case considered here. The main idea is as follows: Since the disturbance vector u_N is typically not observable, estimation has to rely on the residual vector

$$\tilde{u}_N = y_N - X_N \tilde{\beta}_N,$$

where $\tilde{\beta}_N$ is an estimator of β. Typical examples for $\tilde{\beta}_N$ are the OLS estimator and the feasible GLS estimator:

$$\begin{aligned}
\hat{\beta}_{OLS} &= (X_N' X_N)^{-1} X_N' y_N \\
\hat{\beta}_{FGLS} &= (X_N' \hat{\Omega}_{u,N}^{-1} X_N)^{-1} X_N' \hat{\Omega}_{u,N}^{-1} y_N.
\end{aligned}$$

The corresponding regression residuals $\tilde{u}_N$ are given by

$$\tilde{u}_N = M_N u_N,$$

where M_N depends on $\tilde{\beta}_N$. For example, OLS corresponds to $M_N = I_N - X_N(X_N' X_N)^{-1} X_N'$ and FGLS corresponds to $M_N = I_N - X_N(X_N' \hat{\Omega}_{u,N}^{-1} X_N)^{-1} X_N' \hat{\Omega}_{u,N}^{-1}$. Note that M_N is always known in applications because it only depends on the choice of estimator for β. Let

$$\begin{aligned}
\tilde{\varepsilon}_N &= M_N \varepsilon_N \\
\tilde{\bar{\varepsilon}}_N &= (I_T \otimes W_N) \tilde{\varepsilon}_N = (I_T \otimes W_N) M_N \varepsilon_N
\end{aligned}$$

Since the unobservable disturbances of the model have to be replaced by the regression residuals, we suggest to also calculate the theoretical moment conditions in terms of the residuals.

Consequently, we use the following six moment conditions:

$$
\begin{aligned}
\mathrm{E}\left(\frac{1}{n(T-1)}\tilde{\varepsilon}_N' Q_{0,N}\tilde{\varepsilon}_N\right) &= \frac{\sigma_\mu^2}{n(T-1)}\mathrm{tr}(M_N' Q_{0,N} M_N(J_T\otimes I_n)) \\
&\quad +\frac{\sigma_\nu^2}{n(T-1)}\mathrm{tr}(M_N' Q_{0,N} M_N) \\
\mathrm{E}\left(\frac{1}{n(T-1)}\bar{\tilde{\varepsilon}}_N' Q_{0,N}\bar{\tilde{\varepsilon}}_N\right) &= \frac{\sigma_\mu^2}{n(T-1)}\mathrm{tr}[M_N'(I_T\otimes W_n')Q_{0,N}(I_T\otimes W_n)M_N(J_T\otimes I_n)] \\
&\quad +\frac{\sigma_\nu^2}{n(T-1)}\mathrm{tr}[M_N'(I_T\otimes W_n')Q_{0,N}(I_T\otimes W_n)M_N] \\
\mathrm{E}\left(\frac{1}{n(T-1)}\bar{\tilde{\varepsilon}}_N' Q_{0,N}\tilde{\varepsilon}_N\right) &= \frac{\sigma_\mu^2}{n(T-1)}\mathrm{tr}[M_N'(I_T\otimes W_n')Q_{0,N}M_N(J_T\otimes I_n)] \\
&\quad +\frac{\sigma_\nu^2}{n(T-1)}\mathrm{tr}[M_N'(I_T\otimes W_n')Q_{0,N}M_N] \\
\mathrm{E}\left(\frac{1}{n}\tilde{\varepsilon}_N' Q_{1,N}\tilde{\varepsilon}_N\right) &= \frac{\sigma_\mu^2}{n}\mathrm{tr}(M_N' Q_{1,N} M_N(J_T\otimes I_n)) \\
&\quad +\frac{\sigma_\nu^2}{n}\mathrm{tr}(M_N' Q_{1,N} M_N) \\
\mathrm{E}\left(\frac{1}{n}\bar{\tilde{\varepsilon}}_N' Q_{1,N}\bar{\tilde{\varepsilon}}_N\right) &= \frac{\sigma_\mu^2}{n}\mathrm{tr}[M_N'(I_T\otimes W_n')Q_{1,N}(I_T\otimes W_n)M_N(J_T\otimes I_n)] \\
&\quad +\frac{\sigma_\nu^2}{n}\mathrm{tr}[M_N'(I_T\otimes W_n')Q_{1,N}(I_T\otimes W_n)M_N] \\
\mathrm{E}\left(\frac{1}{n}\bar{\tilde{\varepsilon}}_N' Q_{1,N}\tilde{\varepsilon}_N\right) &= \frac{\sigma_\mu^2}{n}\mathrm{tr}[M_N'(I_T\otimes W_n')Q_{1,N}M_N(J_T\otimes I_n)] \\
&\quad +\frac{\sigma_\nu^2}{n}\mathrm{tr}[M_N'(I_T\otimes W_n')Q_{1,N}M_N].
\end{aligned}
$$

Let

$$
\begin{aligned}
\tilde{u}_N &= M_N u_N, \\
\bar{\tilde{u}}_N &= (I_T\otimes W_n)M_N u_N, \\
\tilde{\bar{u}}_N &= M_N(I_T\otimes W_n)u_N, \\
\bar{\tilde{\bar{u}}}_N &= (I_T\otimes W_n)M_N(I_T\otimes W_n)u_N.
\end{aligned}
$$

Substituting $\tilde{\varepsilon}_N$ and $\bar{\tilde{\varepsilon}}_N$ by

$$\begin{aligned}
\tilde{\varepsilon}_N &= M_N \varepsilon_N = M_N u_N - \rho M_N (I_T \otimes W_n) u_N, \\
&= \tilde{u}_N - \rho \bar{\tilde{u}}_N, \\
\bar{\tilde{\varepsilon}}_N &= (I_T \otimes W_n) M_N \varepsilon_N = (I_T \otimes W_n) M_N u_N - \rho (I_T \otimes W_n) M_N (I_T \otimes W_n) u_N, \\
&= \bar{\tilde{u}}_N - \rho \bar{\bar{\tilde{u}}}_N,
\end{aligned}$$

expanding and collecting terms, our residual based theoretical system of equations is given by

$$\Gamma_N(\rho, \rho^2, \sigma_\mu^2, \sigma_\nu^2)' - \gamma_N = 0, \tag{2}$$

where

$$\Gamma_N = \begin{pmatrix} \gamma^0_{11,N} & \gamma^0_{12,N} & \gamma^0_{13,N} & \gamma^0_{14,N} \\ \gamma^0_{21,N} & \gamma^0_{22,N} & \gamma^0_{23,N} & \gamma^0_{24,N} \\ \gamma^0_{31,N} & \gamma^0_{32,N} & \gamma^0_{33,N} & \gamma^0_{34,N} \\ \gamma^1_{11,N} & \gamma^1_{12,N} & \gamma^1_{13,N} & \gamma^1_{14,N} \\ \gamma^1_{21,N} & \gamma^1_{22,N} & \gamma^1_{23,N} & \gamma^1_{24,N} \\ \gamma^1_{31,N} & \gamma^1_{32,N} & \gamma^1_{33,N} & \gamma^1_{34,N} \end{pmatrix}, \quad \gamma_N = \begin{pmatrix} \gamma^0_{1,N} \\ \gamma^0_{2,N} \\ \gamma^0_{3,N} \\ \gamma^1_{1,N} \\ \gamma^1_{2,N} \\ \gamma^1_{3,N} \end{pmatrix}.$$

For $i = 0, 1$, the elements of Γ_N and γ_N are

$$
\begin{aligned}
\gamma_{11,N}^{i} &= \frac{2}{n(T-1)^{1-i}} \mathrm{E}\left[\tilde{u}_N' Q_{i,N} \bar{\tilde{u}}_N\right], \quad \gamma_{21,N}^{i} = \frac{2}{n(T-1)^{1-i}} \mathrm{E}\left[\bar{\tilde{u}}_N' Q_{i,N} \bar{\bar{\tilde{u}}}_N\right], \\
\gamma_{31,N}^{i} &= \frac{2}{n(T-1)^{1-i}} \mathrm{E}\left[\bar{\tilde{u}}_N' Q_{i,N} \bar{\tilde{u}}_N + \bar{\bar{\tilde{u}}}_N' Q_{i,N} \tilde{u}_N\right], \\
\gamma_{12,N}^{i} &= \frac{-1}{n(T-1)^{1-i}} \mathrm{E}\left[\bar{\tilde{u}}_N' Q_{i,N} \bar{\tilde{u}}_N\right], \\
\gamma_{22,N}^{i} &= \frac{-1}{n(T-1)^{1-i}} \mathrm{E}\left[\bar{\bar{\tilde{u}}}_N' Q_{i,N} \bar{\bar{\tilde{u}}}_N\right], \quad \gamma_{32,N}^{i} = \frac{-1}{n(T-1)^{1-i}} \mathrm{E}\left[\bar{\bar{\tilde{u}}}_N' Q_{i,N} \bar{\tilde{u}}_N\right], \\
\gamma_{13,N}^{i} &= \frac{1}{n(T-1)^{1-i}} \mathrm{tr}\left[M_N' Q_{i,N} M_N (J_T \otimes I_n)\right], \quad \gamma_{14,N}^{i} = \frac{1}{n(T-1)^{1-i}} \mathrm{tr}\left[M_N' Q_{i,N} M_N\right], \\
\gamma_{23,N}^{i} &= \frac{1}{n(T-1)^{1-i}} \mathrm{tr}\left[M_N' (I_T \otimes W_n') Q_{i,N} (I_T \otimes W_n) M_N (J_T \otimes I_n)\right], \\
\gamma_{24,N}^{i} &= \frac{1}{n(T-1)^{1-i}} \mathrm{tr}\left[M_N' (I_T \otimes W_n') Q_{i,N} (I_T \otimes W_n) M_N\right], \\
\gamma_{33,N}^{i} &= \frac{1}{n(T-1)^{1-i}} \mathrm{tr}\left[M_N' (I_T \otimes W_n') Q_{i,N} M_N (J_T \otimes I_n)\right], \\
\gamma_{34,N}^{i} &= \frac{1}{n(T-1)^{1-i}} \mathrm{tr}\left[M_N' (I_T \otimes W_n') Q_{i,N} M_N\right], \quad \gamma_{1,N}^{i} = \frac{1}{n(T-1)^{1-i}} \mathrm{E}\left[\tilde{u}_N' Q_{i,N} \tilde{u}_N\right], \\
\gamma_{2,N}^{i} &= \frac{1}{n(T-1)^{1-i}} \mathrm{E}\left[\bar{\tilde{u}}_N' Q_{i,N} \bar{\tilde{u}}_N\right], \quad \gamma_{3,N}^{i} = \frac{1}{n(T-1)^{1-i}} \mathrm{E}\left[\bar{\tilde{u}}_N' Q_{i,N} \tilde{u}_N\right].
\end{aligned}
$$

The true parameter values provide the unique solution of the theoretical system of equations (2). Since Γ_N and γ_N are not observable, (2) is replaced by an empirical counterpart. To that purpose, we leave out the expectation operator and replace $\bar{\tilde{u}}_N$ and $\bar{\bar{\tilde{u}}}_N$, which are not observable, by

$$
\begin{aligned}
\tilde{\bar{\tilde{u}}}_N &= M_N (I_T \otimes W_n) M_N u_N, \\
\tilde{\bar{\bar{\tilde{u}}}}_N &= (I_T \otimes W_n) M_N (I_T \otimes W_n) M_N u_N,
\end{aligned}
$$

respectively. The corresponding empirical system of equations can then be written as

$$
G_N(\rho, \rho^2, \sigma_\mu^2, \sigma_\nu^2)' - g_N = \delta_N(\rho, \sigma_\mu^2, \sigma_\nu^2), \tag{3}
$$

where

$$
G_N = \begin{pmatrix} g^0_{11,N} & g^0_{12,N} & g^0_{13,N} & g^0_{14,N} \\ g^0_{21,N} & g^0_{22,N} & g^0_{23,N} & g^0_{24,N} \\ g^0_{31,N} & g^0_{32,N} & g^0_{33,N} & g^0_{34,N} \\ g^1_{11,N} & g^1_{12,N} & g^1_{13,N} & g^1_{14,N} \\ g^1_{21,N} & g^1_{22,N} & g^1_{23,N} & g^1_{24,N} \\ g^1_{31,N} & g^1_{32,N} & g^1_{33,N} & g^1_{34,N} \end{pmatrix}, \quad g_N = \begin{pmatrix} g^0_{1,N} \\ g^0_{2,N} \\ g^0_{3,N} \\ g^1_{1,N} \\ g^1_{2,N} \\ g^1_{3,N} \end{pmatrix},
$$

$$
\begin{aligned}
g^i_{11,N} &= \frac{2}{n(T-1)^{1-i}}\left[\tilde{u}'_N Q_{i,N} \tilde{\bar{\tilde{u}}}_N\right], \quad g^i_{21,N} = \frac{2}{n(T-1)^{1-i}}\left[\bar{\tilde{u}}'_N Q_{i,N} \bar{\tilde{\bar{\tilde{u}}}}_N\right], \\
g^i_{31,N} &= \frac{1}{n(T-1)^{1-i}}\left[\bar{\tilde{u}}'_N Q_{i,N} \tilde{\bar{\tilde{u}}}_N + \bar{\tilde{\bar{\tilde{u}}}}'_N Q_{i,N} \tilde{u}_N\right], \\
g^i_{12,N} &= \frac{-1}{n(T-1)^{1-i}}\left[\tilde{\bar{\tilde{u}}}'_N Q_{i,N} \tilde{\bar{\tilde{u}}}_N\right], \quad g^i_{22,N} = \frac{-1}{n(T-1)^{1-i}}\left[\bar{\tilde{\bar{\tilde{u}}}}'_N Q_{i,N} \bar{\tilde{\bar{\tilde{u}}}}_N\right], \\
g^i_{32,N} &= \frac{-1}{n(T-1)^{1-i}}\left[\bar{\tilde{\bar{\tilde{u}}}}'_N Q_{i,N} \tilde{\bar{\tilde{u}}}_N\right], \quad g^i_{1,N} = \frac{1}{n(T-1)^{1-i}}\left[\tilde{u}'_N Q_{i,N} \tilde{u}_N\right], \\
g^i_{2,N} &= \frac{1}{n(T-1)^{1-i}}\left[\bar{\tilde{u}}'_N Q_{i,N} \bar{\tilde{u}}_N\right], \quad g^i_{3,N} = \frac{1}{n(T-1)^{1-i}}\left[\bar{\tilde{u}}'_N Q_{i,N} \tilde{u}_N\right].
\end{aligned}
$$

For the third and fourth columns of G_N, we simply take the corresponding elements of Γ_N because they are observable.

It is well known that GMM estimators can be improved by a suitable weighting of the moment conditions. The optimal weighting matrix is given by the inverse of the covariance matrix of the moment conditions. Therefore, we proceed by calculating the covariance matrix of our empirical moment conditions. Since $\tilde{\varepsilon}_N = M_N \varepsilon_N$, $\bar{\tilde{\varepsilon}}_N = (I_T \otimes W_N) M_N \varepsilon_N$, the random variates on the left hand side of our moment conditions can be written as quadratic forms in ε_N,

$$
\varepsilon'_N C_{N,i} \varepsilon'_N,
$$

where

$$
\begin{aligned}
C_{N,1} &= \frac{1}{n(T-1)} M_N' Q_{0,N} M_N, \\
C_{N,2} &= \frac{1}{n(T-1)} M_N' (I_T \otimes W_N') Q_{0,N} (I_T \otimes W_N) M_N, \\
C_{N,3} &= \frac{1}{n(T-1)} M_N' (I_T \otimes W_N') Q_{0,N} M_N, \\
C_{N,4} &= \frac{1}{n} M_N' Q_{1,N} M_N, \\
C_{N,5} &= \frac{1}{n} M_N' (I_T \otimes W_N') Q_{1,N} (I_T \otimes W_N) M_N, \\
C_{N,6} &= \frac{1}{n} M_N' (I_T \otimes W_N') Q_{1,N} M_N.
\end{aligned}
$$

Let $\tilde{C}_{j,N} = \Omega_{\varepsilon,N}^{\frac{1}{2}} C_{j,N} \Omega_{\varepsilon,N}^{\frac{1}{2}}$. Using a spectral decomposition of $\tilde{C}_{j,N}$, we have

$$
\varepsilon_N' C_{j,N} \varepsilon_N = \tilde{\varepsilon}_N' \tilde{C}_{j,N} \tilde{\varepsilon}_N = \sum_{i=1}^{n} \lambda_{ji,N} \zeta_{i,N}^2, \tag{4}
$$

where $\tilde{\varepsilon}_N = \Omega_{\varepsilon,N}^{-\frac{1}{2}} \varepsilon_N$, the $\lambda_{ji,N}$ are the eigenvalues of $\tilde{C}_{j,N}$ and the $\zeta_{i,N}^2$ are independent χ_1^2-distributed random variables, see e.g. Rotar (1973), de Jong (1987) or Mikosch (1991) and the references therein.

Let S_N be the corresponding covariance matrix of our empirical moment conditions. Assuming normality, for $i, j = 1, \ldots, 6$ the covariances between the moment conditions are given by

$$
S_{N,ij} = \mathrm{Cov}(\varepsilon_N' C_{N,i} \varepsilon_N, \varepsilon_N' C_{N,j} \varepsilon_N) = 2\mathrm{tr}(C_{N,i} \Omega_{\varepsilon,N} C_{N,j} \Omega_{\varepsilon,N}).
$$

We define our weighted GMM estimator for $\theta := (\rho, \sigma_\mu^2, \sigma_\nu^2)$ as

$$
(\hat{\rho}, \hat{\sigma}_\mu^2, \hat{\sigma}_\nu^2) = \mathrm{argmin} \left\{ R_N(\tilde{\theta}) : \tilde{\rho} \in [-1,1], \tilde{\sigma}_\mu^2 \in [0, b_\mu], \tilde{\sigma}_\nu^2 \in [0, b_\nu] \right\} \tag{5}
$$

with $\tilde{\theta} = (\tilde{\rho}, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2)$ and $R_N(\tilde{\theta}) := \delta_N \left(\tilde{\rho}, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2\right)' \frac{1}{n} S_N^{-1} \delta_N \left(\tilde{\rho}, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2\right)$.

As we will prove in Section 3, our GMM approach provides consistent estimates, a feature it shares with the approach by Kapoor et al. (2007). The main advantage of the residual based approach presented here is a bias reduction for finite samples. To shed light on this, we give a small analytical illustration. To this purpose, we replace the elements of G_N and g_N in our empirical moment conditions by their respective expectations and calculate the minimizing values for ρ, σ_μ^2 and σ_ν^2 in this "expected" empirical system of equations. Although explicit formulas for these minimizing values could in principle be derived, these formulas are more or less useless because they are very intricate. We can nonetheless get some insight by considering the special case of $\rho = 0$. The j^{th} row of the empirical system of equations ($j = 1, 2, 3$) is then given by

$$\sigma_\mu^2 G_{j3}^0 + \sigma_\nu^2 G_{j4}^0 = g_j^0 \Leftrightarrow \sigma_\mu^2 = \frac{g_j^0 - \sigma_\nu^2 G_{j4}^0}{G_{j3}^0},$$

so e.g. the first row yields

$$\begin{aligned} \mathrm{E}\left(\hat{\sigma}_\mu^2\right) &\approx \frac{\mathrm{E}(g_j^0) - \sigma_\nu^2 G_{j4}^0}{G_{j3}^0} \\ &= \frac{\mathrm{tr}\left(M_N' Q_{0,N} M_N [\sigma_\mu^2 (J_T \otimes I_n) + \sigma_\nu^2 I_N]\right) - \sigma_\nu^2 \mathrm{tr}(M_N' Q_{0,N} M_N)}{\mathrm{tr}(M_N' Q_{0,N} M_N (J_T \otimes I_n))} \\ &= \sigma_\mu^2. \end{aligned} \tag{6}$$

Similar calculations for the other five rows yield the same result so that we can expect the bias of the estimator to be small. For the purpose of comparison, we perform the corresponding calculations for the first and fourth moment conditions of Kapoor et al. (2007). Here, we find that

$$\begin{aligned} \mathrm{E}\left(\hat{\sigma}_\mu^2\right) &\approx \frac{\sigma_\mu^2}{n(T-1)} \mathrm{tr}\left[(T-1) M_N Q_{1,N} M_N Q_{1,N} - M_N Q_{0,N} M_N Q_{1,N}\right] \\ &\quad + \frac{\sigma_\nu^2}{nT(T-1)} \mathrm{tr}\left[(T-1) M_N Q_{1,N} M_N - M_N Q_{0,N} M_N\right] \end{aligned} \tag{7}$$

so that we can expect this estimator to be biased in finite samples.

3. Asymptotic results

This section proves the consistency and asymptotic normality of the GMM estimators as the number of observation units tends to infinity. For this, some additional assumptions will be imposed. First, we consider consistency.

Assumption 3. *a) For $n \to \infty$, $G_N \xrightarrow{p} G_0$ where G_0 is a constant (6×4)-matrix.*

b) For $n \to \infty$, $G_N - \Gamma_N = o_{\mathsf{P}}(1)$.

c) For $n \to \infty$, $g_N \xrightarrow{p} g_0$ where g_0 is a constant (6×1)-vector.

d) For $n \to \infty$, $g_N - \gamma_N = o_{\mathsf{P}}(1)$.

e) For $n \to \infty$, $nS_N \xrightarrow{p} S_0$ where G_0 is a constant (6×6)-matrix.

Assumption 4. *For the true parameter vector $(\rho, \sigma_\mu^2, \sigma_\nu^2)$ the matrix $G_0' S_0^{-1} G_0$ is positively definite.*

If we denote $R_0(\theta^*) = (G_0\theta^* - g_0)' S_0^{-1} (G_0\theta^* - g_0)$, then Assumption 4 yields, for arbitrary $\epsilon > 0$, the inequality

$$\inf_{\{\tilde{\theta}: |\tilde{\theta} - \theta| \geq \epsilon\}} \left| R_0(\tilde{\theta}) - R_0(\theta) \right| > 0$$

and thus guarantees the identifiability of θ, see also Kelejian and Prucha (1999).

Theorem 1. *Under Assumption 1 - 4, for $n \to \infty$,*

$$(\hat{\rho}, \hat{\sigma}_\mu^2, \hat{\sigma}_\nu^2) \xrightarrow{P} (\rho, \sigma_\mu^2, \sigma_\nu^2).$$

Whereas consistency requires the existence of certain limits and some identifiability condition, we additionally need some eigenvalue conditions for the proof of asymptotic normality.

Assumption 5. *(a) For $j = 1, \ldots, 6$, the eigenvalues $\lambda_{ji,N}$ of $\tilde{C}_{j,N}$ fulfill the Ljapunov* *condition, i.e., for some $\delta > 0$ it holds*

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \lambda_{ji,N}^{2+\delta}}{\left(\sum_{i=1}^{n} \lambda_{ji,N}^{2}\right)^{1+\delta}} = 0.$$

(b) The Ljapunov condition is fulfilled for the eigenvalues of any linear combination $\sum_{j=1}^{6} c_j \tilde{C}_{j,N}$ with $\sum_{j=1}^{6} c_j^2 = 1$.

Lemma 1. *Under Assumption 5, $\sqrt{n}(G_N\theta - g_N) \to N(0, S_0)$ as $n \to \infty$.*

Theorem 2. *Under Assumption 1-5, the asymptotic distribution of $(\hat{\rho}, \hat{\sigma}_\mu^2, \hat{\sigma}_\nu^2)$ as $n \to \infty$ is given by*

$$\sqrt{n}\begin{pmatrix} \hat{\rho} - \rho \\ \hat{\sigma}_\mu^2 - \sigma_\mu^2 \\ \hat{\sigma}_\nu^2 - \sigma_\nu^2 \end{pmatrix} \to N(0, (DG^{'} G_0^{'} S_0^{-1} G_0 DG)^{-1}),$$

where

$$(DG)[(\rho, \sigma_\mu^2, \sigma_\nu^2)] := DG := \begin{pmatrix} 1 & 0 & 0 \\ 2\rho & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In applications, G_0 can be replaced by G_N, whereas DG and S_0^{-1} can be estimated by a plug-in method in which the true parameter values are replaced by the GMM estimators for ρ, σ_μ^2 and σ_ν^2. This provides a consistent estimator for the asymptotic covariance matrix.

4. Finite sample Monte Carlo evidence

This section compares the finite sample properties of the GMM estimators for $N = 50$, 100, 200, $T = 5$, $\rho = -0.5, 0, 0.5$ and $\sigma_\mu^2 = \sigma_\nu^2 = 1$. We consider two different weighting matrices W_N. The first one is specified such that each element of u_n is directly related to the elements immediately after and immediately before it. For the first and the last elements of u_n, we imply a circular setting such that for example u_1 is directly related to the second and last element of u_n. This weighting matrix is marked by $J = 2$ since there are two nonzero elements in each row of W_N. The second weighting matrix is labeled by $J = 6$. Here, each element of u_n is directly related to the three elements immediately after and the three elements immediately before it. For both weighting matrices, the row sums are standardized to one. We use two regressors x_1 and x_2 which are the same as in Kapoor et al. (2007): x_1 is the intercept and x_2 is per capita income in contiguous counties in Virginia in the years 1996-2000. For each of the 18 corresponding settings (three different values of ρ, two different weighting matrices and three different sample sizes), we generate 1000 realizations of our regression model and calculate parameter estimates in two different ways, first as in Kapoor et al. (2007) and second as in (5). In both cases, we first use the known optimal weighting matrix (denoted as S_0 in our model) and second, we use the iterative procedure in which the optimal weighting matrix is estimated. Tables 1 and 2 give the resulting biases and mean square errors of the estimators.

- Table 1 here -

- Table 2 here -

Table 1 reveals that the biases of the estimators for ρ and σ_ν^2 are virtually zero, whereas the estimators for σ_μ^2 are both downwards biased. Our modified residual based estimator reduces the bias by up to 95%. Calculating the analytical expressions in (6) and (7) for the true parameter values essentially yields the same result.

Table 2 shows that the mean square errors of the estimators for ρ and σ_ν^2 are very close to each other. With respect to the estimators for σ_μ^2, the MSE of our modified version is slightly larger than the MSE of the estimator of Kapoor et al. (2007). We conclude that the effect of reduced bias comes at the expense of a larger MSE.

There may be situations in which the parameters ρ, σ_μ^2 and σ_ν^2 are of interest in their own right. However, in most applications one is interested in these parameters only because they are needed for significance tests for the regression coefficients contained in β. This is done by plugging the parameter estimates into (1). Consequently, Table 3 compares the performance of the two different estimation approaches with respect to empirical rejection probabilities where the nominal level is $\alpha = 0.05$. Again, we compare the case of known and unknown optimal weighting matrix.

- Table 3 here -

We can see that the empirical rejection probabilities exceed the nominal level of 0.05. For our modified estimator, these overrejection probabilities are smaller by up to $40 - 50\%$, whereby the improvement of our procedure is larger if the optimal weighting matrix is known.

5. Application to Indonesian rice farming

We illustrate our results with an empirical analysis of Indonesian rice farming data. We have data of 171 rice farms over six growing seasons. The farms are located in six different villages. We use a standard random effects model for the data related to the wet growing seasons to regress the output (ln(rice)) on the covariates seed, urea, phosphate (TSP), labor and land as well as dummies for pesticides (DP), high yield varieties (DV1) and mixed varieties (DV2). For a detailed description of the data see Erwidodo (1990). The disturbances are assumed to be spatially correlated across cross-sectional units where the typical element w_{ij} of the spatial weighting matrix W is positive if observations i and j belong to (a) farms located in the same village and (b) the same growing season. The row sums of W are standardized to one.

We estimate ρ, σ_μ^2 and σ_ν^2 in two ways, once following Kapoor et al. (2007) and once by our residual based approach. As to the regression coefficients, the results of the random effects specification mostly agree with the results of a fixed effects model like in Druska and Horrace (2004) or Arnold and Wied (2010). However, there is a considerable discrepancy in the estimates for ρ. Whereas the residual based approach produces an estimate of 0.78, which is very much in line with previous studies of these data, the approach of Kapoor et al. (2007) yields an estimate of 1.23, which is not only far away from previous results but also outside the parameter space. To illustrate this, Figure 1 presents "profile" target functions R_N for both estimators for different values of ρ, where the variance parameters are replaced by their respective estimates ($\hat{\sigma}_\nu^2 = 0.066$ and $\hat{\sigma}_1^2 = 0.102$ for Kapoor et al. (2007), $\hat{\sigma}_\mu^2 = 0.012$ and $\hat{\sigma}_\nu^2 = 0.065$ for the residual based approach).

- Figure 1 here -

For Kapoor et al. (2007), the minimizing value ($\rho = 1.23$) is not included in the parameter space. If the search is restricted on the parameter space, the optimum would be the boundary ($\rho = 1$) which is not a good choice either because $\hat{\Omega}_{u,N}$ would then be singular. For the residual based approach, such problems do not occur. Although there is a local minimum about 1.23, the global minimum is $\rho = 0.78$. We conclude that the residual based modification of the GMM estimators can also circumvent optimization problems.

References

Amemiya, T. (1973): "Regression Analysis when the Dependent Variable Is Truncated Normal," *Econometrica*, 41(6), 997–1016.

Arnold, M. and D. Wied (2010): "Improved GMM estimation of the spatial autoregressive error model," *Economics Letters*, 108, 65–68.

Badinger, H. and P. Egger (2010a): "Estimation of higher-order spatial autoregressive cross-section models with heteroskedastic disturbances," *Papers in Regional Science*, 90, 213–235.

——— (2010b): "Horizontal vs. Vertical Interdependence in Multinational Activity," *Oxford Bulletin for Economics and Statistics*, 72(6), 744–768.

Davidson, J. (1994): *Stochastic Limit Theory*, Oxford Univerity Press, New York.

de Jong, P. (1987): "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75, 261–277.

Druska, V. and W. C. Horrace (2004): "Generalized Moments Estimation for Spatial Panel Data: Indonesian Rice Farming," *American Journal of Agricultural Economics*, 86(1), 185–198.

Erwidodo (1990): *Panel data analysis on farm-level efficiency, input demand and output supply of rice farming in West Java, Indonesia*, Ph.D. dissertation, Department of Agricultural Economics, Michigan State University.

Fischer, M. M., T. Scherngell, and M. Reismann (2009): "Knowledge Spillovers and Total Factor Productivity: Evidence Using a Spatial Panel Data Model," *Geographical Analysis*, 41(2), 204–220.

Gamboa, O. R. E. (2010): "The (un)lucky neighbour: Differences in export performance across Mexico's states," *Papers in Regional Science*, 89(4), 777–799.

Hujer, R., P. J. M. Rodrigues, and K. Wolf (2009): "Estimating the macroeconomic effects of active labour market policies using spatial econometric methods," *International Journal of Manpower*, 30(7), 648–671.

Jennrich, R. (1969): "Asymptotic Properties of Non-Linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40, 633–643.

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

Kelejian, H. and I. Prucha (1999): "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40, 509–533.

——— (2010): "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances," *Journal of Econometrics*, 157, 53–67.

Larch, M. and J. Walde (2009): "Finite sample properties of alternative GMM estimators for random effects models with spatially correlated errors," *Annals of Regional Science*, 43, 473–490.

Lee, L. (2004): "Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 72, 1899–1925.

Lee, L. and X. Liu (2010): "Efficient GMM Estimation of High Order Spatial Autoregressive Models with Autoregressive Disturbances," *Econometric Theory*, 26, 187–230.

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

Mikosch, T. (1991): "Functional Limit Theorems for Random Quadratic Forms," *Stochastic Processes and their Applications*, 37, 81–98.

Poetscher, B. and K. Prucha (1991): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part I: Consistency and approximation concepts," *Econometric Reviews*, 10, 125–216.

Rotar, V. I. (1973): "Some Limit Theorems for Polynomials of Second Degree," *Theory of Probability and its Applications*, 18, 499–507.

6. Appendix section

Proof of Theorem 1

This follows by standard arguments as e.g. presented in Poetscher and Prucha (1991), Amemiya (1973) or Jennrich (1969), using the uniform convergence of $R_N(\tilde{\theta})$ to $R_0(\tilde{\theta})$ and the identificability condition. ■

Proof of Lemma 1

Equation (4) in combination with Assumption 5(a) implies asymptotic normality for each moment condition by Theorems 23.6 and 23.11 of Davidson (1994). Since every linear combination of the moment conditions can be written as

$$\sum_{j=1}^{6} c_j \tilde{\varepsilon}_N' C_{ji,N} \tilde{\varepsilon}_N = \tilde{\varepsilon}_N' \left(\sum_{j=1}^{6} c_j C_{ji,N} \right) \tilde{\varepsilon}_N,$$

these linear combinations are also asymptotically normal by Assumption 5(b) so that multivariate normality follows by the Cramér-Wold device which proves the Lemma. ■

Proof of Theorem 2

Due to the smoothness of the target function the estimators are the zeros of the derivative

$$\Psi(\tilde{\rho}, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2) := 2DGG_N' S_N^{-1}(G_N\theta - g_N).$$

With the multivariate mean value theorem it holds that

$$
\begin{aligned}
\Psi\begin{pmatrix}\hat{\rho}\\ \hat{\sigma}_\mu^2\\ \hat{\sigma}_\nu^2\end{pmatrix} = 0 &= \Psi\begin{pmatrix}\rho\\ \sigma_\mu^2\\ \sigma_\nu^2\end{pmatrix} + \left[D\Psi\begin{pmatrix}\bar{\rho}\\ \bar{\sigma}_\mu^2\\ \bar{\sigma}_\nu^2\end{pmatrix}\right]\begin{pmatrix}\hat{\rho}-\rho\\ \hat{\sigma}_\mu^2-\sigma_\mu^2\\ \hat{\sigma}_\nu^2-\sigma_\nu^2\end{pmatrix}\\
\Leftrightarrow \begin{pmatrix}\hat{\rho}-\rho\\ \hat{\sigma}_\mu^2-\sigma_\mu^2\\ \hat{\sigma}_\nu^2-\sigma_\nu^2\end{pmatrix} &= \left[D\Psi\begin{pmatrix}\bar{\rho}\\ \bar{\sigma}_\mu^2\\ \bar{\sigma}_\nu^2\end{pmatrix}\right]^{-1}\Psi\begin{pmatrix}\rho\\ \sigma_\mu^2\\ \sigma_\nu^2\end{pmatrix},
\end{aligned}
$$

for some $(\bar{\rho}, \bar{\sigma}_\mu^2, \bar{\sigma}_\nu^2)$ between $(\rho, \sigma_\mu^2, \sigma_\nu^2)$ and $(\hat{\rho}, \hat{\sigma}_\mu^2, \hat{\sigma}_\mu^2)$. $D\Psi$ is given by

$$
\begin{aligned}
D\Psi\begin{pmatrix}\bar{\rho}\\ \bar{\sigma}_\mu^2\\ \bar{\sigma}_\nu^2\end{pmatrix} &= 2(DG)[(\bar{\rho}, \bar{\sigma}_\mu^2, \bar{\sigma}_\nu)]G_N' S_N^{-1} G_N (DG)'[(\bar{\rho}, \bar{\sigma}_\mu^2, \bar{\sigma}_\nu)]\\
&\quad +2\left[\left(G_N[(\bar{\rho}, \bar{\sigma}_\mu^2, \bar{\sigma}_\nu)] - g_N\right)' S_N^{-1} G_N \begin{pmatrix}0&0&0\\2&0&0\\0&0&0\\0&0&0\end{pmatrix}\right] \otimes \begin{pmatrix}1\\0\\0\end{pmatrix}.
\end{aligned}
$$

It follows

$$
\begin{aligned}
\sqrt{n}\begin{pmatrix}\hat{\rho}-\rho\\ \hat{\sigma}_\mu^2-\sigma_\mu^2\\ \hat{\sigma}_\nu^2-\sigma_\nu^2\end{pmatrix} &= \left((DG)[(\bar{\rho}, \bar{\sigma}_\mu^2, \bar{\sigma}_\nu)]G_N' \frac{1}{n} S_N^{-1} G_N (DG)'[(\bar{\rho}, \bar{\sigma}_\mu^2, \bar{\sigma}_\nu)] + o_{\mathsf{P}}(1)\right)^{-1}\cdot\\
&\quad DG' G_N' \frac{1}{n} S_N^{-1} \sqrt{n}(G_N\theta - g_N).
\end{aligned}
$$

With Lemma 1, $\sqrt{n}(G_N\theta - g_N)$ converges to $N(0, S_0)$ whereas the preceding term converges by the consistency of Theorem 1, the Continuous Mapping Theorem and Slutzky's Theorem to

$$
(DG' G_0' S_0^{-1} G_0 DG)^{-1} DG' G_0' S_0^{-1}
$$

with Assumption 3. The theorem then follows by Slutzky's Theorem. ■

Figure 1: Profile target functions for ρ

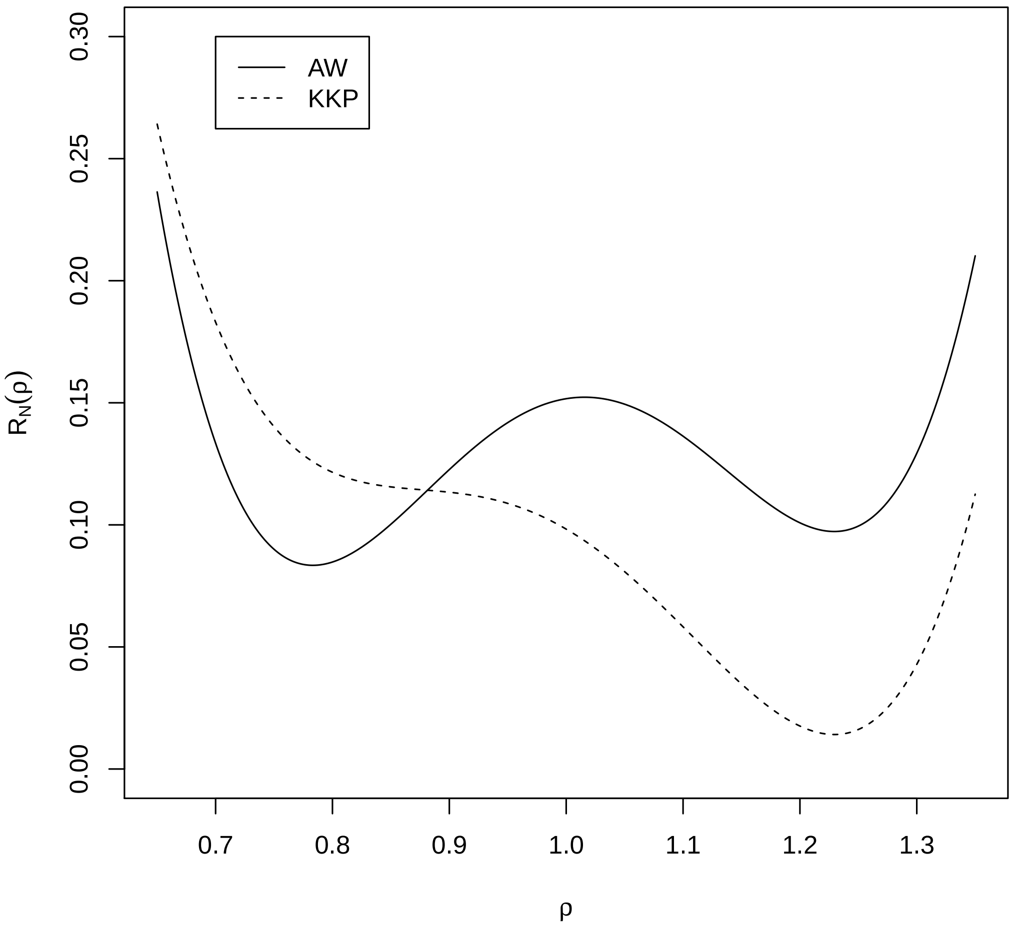

Table 1: Bias of the estimators for known optimal weighting matrix (upper line) and for the iterative procedure (lower line)

			$\hat{\rho}$		$\hat{\sigma}_\mu^2$		$\hat{\sigma}_\nu^2$	
N	J	ρ	KKP	AW	KKP	AW	KKP	AW
50	2	0.5	-0.0097	0.0020	-0.2443	-0.0533	-0.0219	-0.0252
			-0.0195	-0.0050	-0.2097	-0.0364	-0.0263	-0.0236
50	2	0	-0.0205	0.0015	-0.2473	-0.0479	-0.0394	-0.0331
			-0.0255	-0.0104	-0.1967	-0.0205	-0.0480	-0.0387
50	2	-0.5	-0.0343	-0.0108	-0.2656	-0.0822	-0.0474	-0.0299
			-0.0154	-0.0061	-0.2537	-0.0827	-0.0274	-0.0126
50	6	0.5	0.0062	0.0124	-0.2327	-0.0184	-0.0225	-0.0204
			0.0047	0.0297	-0.2125	-0.0323	-0.0244	-0.0132
50	6	0	-0.0376	-0.0057	-0.2308	-0.0216	-0.0320	-0.0251
			-0.0229	-0.0014	-0.2162	-0.0426	-0.0363	-0.0260
50	6	-0.5	-0.0378	0.0067	-0.2662	-0.0831	-0.0384	-0.0224
			-0.0169	0.0052	-0.2120	-0.0399	-0.0315	-0.0166
100	2	0.5	-0.0061	0.0001	-0.1289	-0.0340	-0.0077	-0.0102
			-0.0048	0.0030	-0.1135	-0.0250	-0.0101	-0.0121
100	2	0	-0.0103	0.0031	-0.1182	-0.0204	-0.0158	-0.0117
			-0.0105	0.0022	-0.1344	-0.0432	-0.0130	-0.0090
100	2	-0.5	-0.0104	0.0004	-0.1153	-0.0225	-0.0198	-0.0084
			-0.0121	-0.0025	-0.1289	-0.0386	-0.0229	-0.0119
100	6	0.5	-0.0225	-0.0018	-0.0997	0.0025	-0.0103	-0.0120
			-0.0159	-0.0021	-0.0847	-0.0018	-0.0121	-0.0127
100	6	0	-0.0243	0.0123	-0.1016	-0.0020	-0.0230	-0.0178
			-0.0268	0.0121	-0.1180	-0.0284	-0.0181	-0.0120
100	6	-0.5	-0.0248	0.0204	-0.1134	-0.0197	-0.0218	-0.0088
			-0.0141	0.0251	-0.1125	-0.0241	-0.0290	-0.0173
200	2	0.5	0.0000	0.0031	-0.0537	-0.0037	-0.0123	-0.0138
			-0.0002	0.0028	-0.0525	-0.0041	-0.0124	-0.0134
200	2	0	-0.0085	-0.0021	-0.0594	-0.0101	-0.0137	-0.0116
			-0.0007	0.0061	-0.0590	-0.0116	-0.0103	-0.0082
200	2	-0.5	-0.0019	0.0034	-0.0553	-0.0101	-0.0122	-0.0063
			-0.0080	-0.0022	-0.0560	-0.0107	-0.0132	-0.0069
200	6	0.5	-0.0012	0.0086	-0.0624	-0.0123	-0.0084	-0.0096
			-0.0099	0.0028	-0.0539	-0.0087	-0.0061	-0.0068
200	6	0	-0.0076	0.0093	-0.0662	-0.0177	-0.0101	-0.0077
			-0.0073	0.0107	-0.0522	-0.0062	-0.0051	-0.0023
200	6	-0.5	-0.0149	0.0061	-0.0491	-0.0035	-0.0110	-0.0048
			-0.0073	0.0134	-0.0421	0.0016	-0.0055	0.0008

Table 2: MSE of the estimators for known optimal weighting matrix (upper line) and for the iterative procedure (lower line)

			$\hat{\rho}$		$\hat{\sigma}_\mu^2$		$\hat{\sigma}_\nu^2$	
N	J	ρ	KKP	AW	KKP	AW	KKP	AW
50	2	0.5	0.019	0.021	0.259	0.297	0.051	0.052
			0.024	0.031	0.281	0.337	0.052	0.056
50	2	0	0.025	0.027	0.251	0.294	0.051	0.051
			0.026	0.032	0.280	0.341	0.051	0.051
50	2	-0.5	0.022	0.017	0.286	0.355	0.060	0.060
			0.018	0.023	0.270	0.316	0.053	0.053
50	6	0.5	0.110	0.112	0.269	0.325	0.050	0.051
			0.104	0.136	0.283	0.334	0.048	0.050
50	6	0	0.116	0.126	0.256	0.320	0.052	0.052
			0.106	0.148	0.264	0.310	0.054	0.055
50	6	-0.5	0.101	0.114	0.252	0.293	0.050	0.051
			0.098	0.143	0.291	0.366	0.053	0.055
100	2	0.5	0.007	0.007	0.139	0.149	0.027	0.027
			0.006	0.007	0.134	0.145	0.027	0.028
100	2	0	0.011	0.012	0.138	0.152	0.025	0.025
			0.011	0.012	0.138	0.147	0.026	0.026
100	2	-0.5	0.006	0.006	0.131	0.148	0.029	0.029
			0.007	0.007	0.131	0.144	0.030	0.031
100	6	0.5	0.022	0.021	0.141	0.159	0.025	0.025
			0.032	0.026	0.148	0.164	0.028	0.028
100	6	0	0.036	0.038	0.140	0.158	0.025	0.025
			0.037	0.040	0.133	0.145	0.026	0.026
100	6	-0.5	0.049	0.053	0.137	0.155	0.024	0.025
			0.046	0.053	0.139	0.155	0.025	0.025
200	2	0.5	0.003	0.003	0.074	0.078	0.014	0.014
			0.003	0.003	0.074	0.078	0.014	0.014
200	2	0	0.006	0.006	0.068	0.071	0.012	0.012
			0.006	0.006	0.073	0.076	0.012	0.012
200	2	-0.5	0.003	0.004	0.073	0.078	0.013	0.013
			0.003	0.004	0.071	0.076	0.013	0.013
200	6	0.5	0.007	0.007	0.070	0.073	0.013	0.013
			0.008	0.009	0.070	0.074	0.012	0.012
200	6	0	0.017	0.018	0.069	0.071	0.012	0.012
			0.016	0.017	0.075	0.079	0.013	0.013
200	6	-0.5	0.021	0.022	0.069	0.075	0.014	0.014
			0.020	0.022	0.073	0.080	0.012	0.012

Table 3: Empirical rejection probabilities of significance tests for the regression coefficients for known optimal weighting matrix (upper line) and for the iterative procedure (lower line); nominal level $\alpha = 0.05$

		$J = 2$				$J = 6$			
		β_1		β_2		β_1		β_2	
N	ρ	KKP	AW	KKP	AW	KKP	AW	KKP	AW
50	0.5	0.115	0.083	0.110	0.085	0.116	0.096	0.120	0.092
		0.155	0.126	0.222	0.194	0.134	0.109	0.178	0.145
50	0	0.117	0.089	0.107	0.083	0.113	0.087	0.111	0.090
		0.105	0.080	0.100	0.080	0.111	0.088	0.110	0.094
50	-0.5	0.113	0.090	0.099	0.081	0.093	0.076	0.102	0.081
		0.149	0.121	0.166	0.138	0.107	0.091	0.107	0.097
100	0.5	0.067	0.059	0.068	0.056	0.070	0.052	0.073	0.060
		0.151	0.139	0.178	0.168	0.111	0.091	0.143	0.131
100	0	0.080	0.064	0.086	0.079	0.064	0.058	0.069	0.055
		0.079	0.065	0.081	0.072	0.067	0.061	0.077	0.071
100	-0.5	0.084	0.077	0.072	0.063	0.077	0.068	0.073	0.066
		0.129	0.115	0.133	0.119	0.098	0.081	0.083	0.077
200	0.5	0.056	0.047	0.076	0.072	0.058	0.053	0.072	0.066
		0.119	0.113	0.138	0.132	0.114	0.111	0.145	0.135
200	0	0.052	0.045	0.052	0.050	0.060	0.057	0.063	0.055
		0.055	0.052	0.049	0.044	0.077	0.070	0.064	0.061
200	-0.5	0.056	0.054	0.048	0.047	0.056	0.054	0.059	0.055
		0.136	0.128	0.140	0.138	0.089	0.083	0.093	0.088