Improved GMM estimation of random effects 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 29, 2012

Abstract

We modify a previously suggested GMM estimator in a spatial panel regression model, which has recently received considerable interest in empirical applications, by taking into account the difference between disturbances and regression residuals. Consistency and asymptotic normality of the estimator are derived. Analytic results, simulation evidence and an empirical application to Indonesian rice data illustrate the improvement in finite samples.

JEL Classification: C13, C21
Keywords: Panel regression; Regression residuals; Spatial autoregression; Spatial econometrics

*Corresponding author. Phone: +49/231/755 3869, Fax: +49/231/755 5284. Financial support by Deutsche Forschungsgemeinschaft (SFB 823, project A1) is gratefully acknowledged. We are grateful to the editor James LeSage and two anonymous referees who provided numerous helpful comments which improved the paper.

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 the 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 the spatial autoregressive error model for cross-sectional data to the panel case in order to improve the estimator in small and moderate samples. Baltagi and Liu (2011) carry over this idea to the spatial moving average error model in recent work. The 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. Although this remains true for all sample sizes, the effect is especially relevant for moderate sample sizes. In return, this helps to improve significance tests for the regression coefficients in the sense that actual rejection probabilities are closer to the nominal level. 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 n observation units and T time points and 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 non-stochastic $n \times k$ matrix of observations on the exogenous regressors with the corresponding $k \times 1$ vector β of regression coefficients and $u_n(t)$ is the $n \times 1$ vector of spatially correlated disturbances with spatial correlation parameter ρ and spatial weighting matrix $W_n = (w_{ij,n})_{1 \leq i,j \leq n}$. 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 and I_n are identity matrices of the respective dimension, the $n \times 1$ vector of individual effects $\mu_n = (\mu_{1,n}, \ldots, \mu_{n,n})$ is constant for all time periods and the $N \times 1$ vector ν_N with $\nu_n(t) = (\nu_{1t,n}, \ldots, \nu_{nt,n}), t = 1, \ldots, T$, captures the remainder error terms which vary over both the cross-sectional units and the time periods. Note that $N = n \cdot T$ and that the quantities form triangular arrays. The model is similar to the model used in Kapoor et al. (2007). Mutl and Pfaffermayr (2011) consider a refinement by e.g. including an additional spatial autoregressive term $\lambda W_N y_N$ in the equation for y_N. For ease of exposition, we just consider the case $\lambda = 0$, i.e. we assume that there is only a spatial error component, no spatial lag.
We impose the following assumptions:

Assumption 1. *a) For all $i \in \{1, \ldots, n\}$, $n \geq 1$, 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 $i \in \{1, \ldots, n\}$, $n \geq 1$, $t \in \{1, \ldots, T\}$, 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 $i \in \{1, \ldots, n\}$, $n \geq 1$, $t \in \{1, \ldots, T\}$, the $\nu_{it,n}$ and $\mu_{i,n}$ are independent.

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

Assumption 2 restricts the degree of cross-sectional correlation between the model disturbances and serves for the next lemma.

Lemma 1. *Under Assumption 2, the matrix $I_n - \rho W_n$ is nonsingular.*

With Lemma 1,

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

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:

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

where $\hat{\Omega}_{u,N}$ is an estimator for $\Omega_{u,N}$, typically a plug-in estimator in which the true parameter values ρ, σ_μ^2 and σ_ν^2 are replaced by consistent estimates.
The corresponding regression residuals $\tilde{u}_N$ are given by

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

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 $M_N = I_N - X_N(X_N' \hat{\Omega}_{u,N}^{-1} X_N)^{-1} X_N' \hat{\Omega}_{u,N}^{-1}$. Whereas efficient GLS estimation would require knowledge of the parameters, our residual based approach exploits the difference between unobservable disturbances and observable regression residuals. This difference can be characterized by M_N, respectively, and is always known in applications because it only depends on the choice of estimator for β. In practice, we typically perform a two-stage estimation procedure in which we first use the OLS residuals to obtain initial estimates and the FGLS-estimator after this. The procedure is described in detail below.

Let

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

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:

$$
\begin{aligned}
\mathrm{E}\left(\frac{1}{n(T-1)}\tilde{\varepsilon}_N' Q_{0,N}\tilde{\varepsilon}_N\right) &= \frac{\sigma_\mu^2}{n(T-1)}\mathrm{tr}(M_N' Q_{0,N} M_N (J_T \otimes I_n)) \\
&\quad + \frac{\sigma_\nu^2}{n(T-1)}\mathrm{tr}(M_N' Q_{0,N} M_N) =: c_{1,N}^* \qquad (2) \\
\mathrm{E}\left(\frac{1}{n(T-1)}\bar{\tilde{\varepsilon}}_N' Q_{0,N}\bar{\tilde{\varepsilon}}_N\right) &= \frac{\sigma_\mu^2}{n(T-1)}\mathrm{tr}[M_N'(I_T \otimes W_n')Q_{0,N}(I_T \otimes W_n)M_N(J_T \otimes I_n)] \\
&\quad + \frac{\sigma_\nu^2}{n(T-1)}\mathrm{tr}[M_N'(I_T \otimes W_n')Q_{0,N}(I_T \otimes W_n)M_N] \\
&=: c_{2,N}^* \qquad (3) \\
\mathrm{E}\left(\frac{1}{n(T-1)}\bar{\tilde{\varepsilon}}_N' Q_{0,N}\tilde{\varepsilon}_N\right) &= \frac{\sigma_\mu^2}{n(T-1)}\mathrm{tr}[M_N'(I_T \otimes W_n')Q_{0,N}M_N(J_T \otimes I_n)] \\
&\quad + \frac{\sigma_\nu^2}{n(T-1)}\mathrm{tr}[M_N'(I_T \otimes W_n')Q_{0,N}M_N] =: c_{3,N}^* \qquad (4) \\
\mathrm{E}\left(\frac{1}{n}\tilde{\varepsilon}_N' Q_{1,N}\tilde{\varepsilon}_N\right) &= \frac{\sigma_\mu^2}{n}\mathrm{tr}(M_N' Q_{1,N} M_N (J_T \otimes I_n)) \\
&\quad + \frac{\sigma_\nu^2}{n}\mathrm{tr}(M_N' Q_{1,N} M_N) =: c_{4,N}^* \qquad (5) \\
\mathrm{E}\left(\frac{1}{n}\bar{\tilde{\varepsilon}}_N' Q_{1,N}\bar{\tilde{\varepsilon}}_N\right) &= \frac{\sigma_\mu^2}{n}\mathrm{tr}[M_N'(I_T \otimes W_n')Q_{1,N}(I_T \otimes W_n)M_N(J_T \otimes I_n)] \\
&\quad + \frac{\sigma_\nu^2}{n}\mathrm{tr}[M_N'(I_T \otimes W_n')Q_{1,N}(I_T \otimes W_n)M_N] =: c_{5,N}^* \qquad (6) \\
\mathrm{E}\left(\frac{1}{n}\bar{\tilde{\varepsilon}}_N' Q_{1,N}\tilde{\varepsilon}_N\right) &= \frac{\sigma_\mu^2}{n}\mathrm{tr}[M_N'(I_T \otimes W_n')Q_{1,N}M_N(J_T \otimes I_n)] \\
&\quad + \frac{\sigma_\nu^2}{n}\mathrm{tr}[M_N'(I_T \otimes W_n')Q_{1,N}M_N] =: c_{6,N}^*. \qquad (7)
\end{aligned}
$$

Let

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

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

$$
\Gamma_N \cdot (\rho, \rho^2, \sigma_\mu^2, \sigma_\nu^2)' - \gamma_N = 0, \tag{8}
$$

where

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

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 (8). Since Γ_N and γ_N are not observable, (8) 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

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

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

$$
G_N \cdot (\tilde{\rho}, \tilde{\rho}^2, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2)' - g_N = \delta_N(\tilde{\rho}, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2), \tag{9}
$$

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,

$$
\varepsilon'_N C_{j,N} \varepsilon_N,
$$

where

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

Let $\tilde{C}_{j,N} = \Omega_{\varepsilon,N}^{\frac{1}{2}} C_{j,N} \Omega_{\varepsilon,N}^{\frac{1}{2}}$ where $\Omega_{\varepsilon,N}^{\frac{1}{2}}$ is the square root of the matrix $\Omega_{\varepsilon,N}$ with $\Omega_{\varepsilon,N}^{\frac{1}{2}} \cdot \Omega_{\varepsilon,N}^{\frac{1}{2}} = \Omega_{\varepsilon,N}$. Using a spectral decomposition of $\tilde{C}_{j,N}$, we have

$$\varepsilon_N^{'} C_{j,N} \varepsilon_N = \xi_N^{'} \tilde{C}_{j,N} \xi_N = \sum_{i=1}^{N} \lambda_{ji,N} \zeta_{i,N}^2$$

and

$$\varepsilon_N^{'} C_{j,N} \varepsilon_N - E(\varepsilon_N^{'} C_{j,N} \varepsilon_N) = \xi_N^{'} \tilde{C}_{j,N} \xi_N - E(\xi_N^{'} \tilde{C}_{j,N} \xi_N) = \sum_{i=1}^{N} \lambda_{ji,N} (\zeta_{i,N}^2 - 1), \tag{10}$$

where $\xi_N = \Omega_{\varepsilon,N}^{-\frac{1}{2}} \varepsilon_N$, the $\lambda_{ji,N}$ are the eigenvalues of $\tilde{C}_{j,N}$ and the $\zeta_{i,N}^2$ are independent random variables with expectation 1, see e.g. Rotar (1973), de Jong (1987) or Mikosch (1991) and the references therein. Note that $E(\varepsilon_N^{'} C_{j,N} \varepsilon_N) = c_{j,N}^*$ with the $c_{j,N}^*$ from equations (2) - (7).
Let S_N be the corresponding covariance matrix of our properly scaled empirical moment conditions which depends on the distribution of the ε_N. For normally distributed ε_N, for $i, j = 1, \ldots, 6$ the elements of S_N are given by

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

As discussed in Kapoor et al. (2007), p.108, in the absence of normality, this matrix will not be strictly optimal, but it has the advantage of simplicity and can be viewed as an approximation to the more complex true covariance matrix. Furthermore, note that the asymptotic results below do not depend on the normality assumption.
We define our weighted GMM estimator for $\theta := (\rho, \sigma_\mu^2, \sigma_\nu^2)$ as

$$\hat{\theta} := (\hat{\rho}, \hat{\sigma}_\mu^2, \hat{\sigma}_\nu^2) = \mathrm{argmin} \left\{ R_N(\tilde{\theta}) : \tilde{\rho} \in [-1, 1], \tilde{\sigma}_\mu^2 \in [0, b_\mu], \tilde{\sigma}_\nu^2 \in [0, b_\nu] \right\} \tag{11}$$

with $\tilde{\theta} = (\tilde{\rho}, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2)$ and $R_N(\tilde{\theta}) := \delta_N \left(\tilde{\rho}, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2\right)^{'} S_{W,N} \delta_N \left(\tilde{\rho}, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2\right)$.

For the weighting matrix $S_{W,N}$, one can choose any matrix which converges against a symmetric positive definite matrix S_W for $n \to \infty$. Given the explanations from above and assuming the invertibility of S_N^{-1}, it would be efficient to use $S_{W,N} = S_N^{-1}$ which would require knowledge of the true parameter values contained in S_N^{-1}. However, a two-stage approach is possible, i.e. we first use GMM estimation with $S_{W,N}$ calculated from initial values for the parameters. To be more precise, we suggest to take $\sigma_\nu^2 = 1$ and $\sigma_\mu^2 = 0$ which corresponds to no panel structure. This yields consistent GMM estimates which do not require any a priori information. For calculating G_N and g_N, in this stage we use the matrix M_N which corresponds to OLS. After that, we use the initial estimates to obtain better estimates with the estimated matrix S_N^{-1} and use the matrix M_N which corresponds to FGLS for obtaining G_N and g_N in the second stage. 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:

$$G_{jk}^i := E(g_{jk,N}^i), j = 1,2,3, k = 1,2,3,4, i = 0,1 \ g_j^i := E(g_{j,N}^i), j = 1,2,3, i = 0,1.$$

Afterwards, we 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

$$\begin{aligned} \mathrm{E}\left(\hat{\sigma}_\mu^2\right) &\approx \frac{\mathrm{E}(g_j^0) - \sigma_\nu^2 G_{j4}^0}{G_{j3}^0} \\ &= \frac{\mathrm{tr}\left(M_N' Q_{0,N} M_N[\sigma_\mu^2(J_T \otimes I_n) + \sigma_\nu^2 I_N]\right) - \sigma_\nu^2 \mathrm{tr}(M_N' Q_{0,N} M_N)}{\mathrm{tr}(M_N' Q_{0,N} M_N (J_T \otimes I_n))} \\ &= \sigma_\mu^2. \end{aligned} \tag{12}$$

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

$$\begin{aligned} \mathrm{E}\left(\hat{\sigma}_\mu^2\right) &\approx \frac{\sigma_\mu^2}{n(T-1)} \mathrm{tr}\left[(T-1) M_N Q_{1,N} M_N Q_{1,N} - M_N Q_{0,N} M_N Q_{1,N}\right] \\ &\quad + \frac{\sigma_\nu^2}{nT(T-1)} \mathrm{tr}\left[(T-1) M_N Q_{1,N} M_N - M_N Q_{0,N} M_N\right] \end{aligned} \tag{13}$$

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 n tends to infinity and T remains fixed. Remember that $N = n \cdot T$. To derive the asymptotic results, some additional assumptions will be imposed, at first some conditions on the regressor matrix.

Assumption 3. *a) For the entries* $(x_{ij,N}), i = 1, \dots, N,\ j = 1, \dots, k,$ *of* X_N *it holds* $|x_{ij,N}| < k_X$*, where* k_X *does not depend on* N.
b) $\lim_{n\to\infty} \frac{1}{n} X_N' \Omega_{u,N} X_N =: Q_{X'\Omega X}$ *and* $\lim_{n\to\infty} \frac{1}{n} X_N' X_N =: Q_{X'X}$*, where* $Q_{X'\Omega X}$ *and* $Q_{X'X}$ *are positive definite matrices.*
c) $\lim_{n\to\infty} \Gamma_N =: \Gamma_0$*, where* Γ_0 *is a constant* (6×4)*-matrix.*
d) $\lim_{n\to\infty} \gamma_N =: \gamma_0$*, where* γ_0 *is a constant* (6×1)*-vector.*

Assumptions 3 a) and b) are standard in the spatial econometrics literature and correspond to Assumption 3 in Kapoor et al. (2007). Among other reasons, these assumptions are needed to control the difference between residuals and error terms, see the proofs of Lemma 2. Assumptions 3 c) and d) ensure that the expressions in the theoretical system of equations have a well-defined limit. This is needed to derive the asymptotic covariance matrix of the estimator $\hat{\theta}$. These are no strong assumptions because it follows from construction that e.g. Γ_N is $O(1)$. Note that, with Assumptions 3 a) and b), Γ_0 and γ_0 do not depend on the concrete choice of M_N (OLS- or FGLS), see Lemma 2. This fact ensures the validity of our two-stage approach, namely that we can already obtain consistent initial estimates for θ with M_N corresponding to the OLS estimator which in the second step are refined by M_N corresponding to the FGLS estimator.
Next, we impose an identifiability condition which is crucial for consistency and asymptotic normality.

Assumption 4. *For the probability limit* S_W *of* $S_{W,N}$*, the matrix* $\Gamma_0' S_W \Gamma_0$ *is positive definite and* S_W *is symmetric.*

If we denote $R_0(\tilde{\theta}) = (\Gamma_0(\tilde{\rho}, \tilde{\rho}^2, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2)' - \gamma_0)' S_W (\Gamma_0(\tilde{\rho}, \tilde{\rho}^2, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2)' - \gamma_0)$, 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). Assumption 4 e.g. rules out the case that X_N contains only a constant such that $X_N = e_N$. In this case, OLS would lead to $M_N = I_N - \frac{1}{N} J_N$ and the first three moment conditions would collapse because $Q_{0,N} M_N = 0$. Consequently, the matrix Γ_0 would contain rows in which all elements are equal to 0 so that $\Gamma_0' S_W \Gamma_0$ would not have full rank.
We also need an eigenvalue condition which corresponds to the expression in equation (10). This condition ensures that not some eigenvalues of $\tilde{C}_{j,N}$ are too large compared to the others.

Assumption 5. *For $j = 1, \ldots, 6$, the random variables from equation* (10) *fulfill the Ljapunov condition, i.e., for some $\delta > 0$ it holds*

$$\lim_{n\to\infty} n^{1+\frac{\delta}{2}} \sum_{i=1}^{N} |\lambda_{ji,N}|^{2+\delta} E|\zeta_{i,N}^2 - 1|^{2+\delta} = 0.$$

The following lemma gives information about the existence of certain limits.

Lemma 2. *Let Assumption 1 - 5 be true.*
a) The matrices Γ_0 and γ_0 do not depend on the concrete choice of the sequence $(M_N : N = 1, 2, \ldots)$.
b) For $n \to \infty$, $G_N - \Gamma_N = O_{\mathsf{P}}(n^{-1})$.
c) For $n \to \infty$, $g_N - \gamma_N = O_{\mathsf{P}}(n^{-1})$.

Now, we can derive consistency.

Theorem 1. *Under Assumptions 1 - 5, for $n \to \infty$ and for any sequence of matrices $(S_{W,N} : N = 1, 2, \ldots)$ with $S_{W,N} \to_p S_W$,*

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

Again, the fact that $S_{W,N}$ is arbitrary justifies the validity of our two stage approach because it ensures that we can obtain initial consistent estimates for θ.
Finally, we prove asymptotic normality with an additional lemma.

Lemma 3. *Under Assumptions 1 - 5, for $n \to \infty$, $\sqrt{n}(G_N \cdot (\rho, \rho^2, \sigma_\mu^2, \sigma_\nu^2)' - g_N) \to_d N(0, S_0)$, where $S_0 := \lim_{n\to\infty} S_N$ is a constant symmetric (6×6)-matrix.*

Theorem 2. *Under Assumptions 1 - 5, for $n \to \infty$ and for any sequence of matrices $(S_{W,N} : N = 1, 2, \ldots)$ with $S_{W,N} \to S_W$, the asymptotic distribution of $(\hat{\rho}, \hat{\sigma}_\mu^2, \hat{\sigma}_\nu^2)$ as $n \to \infty$ is given by*

$$\sqrt{n} \begin{pmatrix} \hat{\rho} - \rho \\ \hat{\sigma}_\mu^2 - \sigma_\mu^2 \\ \hat{\sigma}_\nu^2 - \sigma_\nu^2 \end{pmatrix} \to_d N\left(0, \left(Q'\Gamma_0' S_W \Gamma_0 Q\right)^{-1} Q'\Gamma_0' S_W S_0 S_W \Gamma_0 Q \left(Q'\Gamma_0' S_W \Gamma_0 Q\right)^{-1}\right),$$

where

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

By choosing $S_{W,N} = S_N^{-1}$, we would get $S_W = S_0^{-1}$ and the asymptotic covariance matrix would simplify to $\left(Q'\Gamma_0' S_0^{-1} \Gamma_0 Q\right)^{-1}$.
However, S_N^{-1} depends on the unknown parameters. So, in applications, Γ_0 can be replaced by G_N, whereas Q 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 the often used Columbus data set which contains information about crime rates from 49 neighborhoods in Columbus, Ohio, 1980, see Anselin (1988). The weighting matrix W is specified such that entry (i, j) is nonzero if neighborhoods i and j share a common border and the row sums are standardized to one. With $n = 49$, we analyze values of $T \in \{2, 5, 10\}$ (leading to $N \in \{49, 245, 490\}$), $\rho \in \{-0.5, 0, 0.5\}$ and $\sigma_\mu^2 = \sigma_\nu^2 \in \{1, 2\}$.

We use eight regressors $x_1, \ldots, x_8$ from the original data set and stack the data from 1980 to get a panel structure. x_1 is the intercept, x_2 describes the housing value in \$1,000, x_3 describes the household income in \$1,000, x_4 describes the percentage of housing units without plumbing, x_5 describes the distance to the Central Business District, x_6 is a north-south dummy (north=1), x_7 is an east-west dummy (east=1) and x_8 is a core-periphery dummy (core=1), see the documentation of the function columbus in the R-package *spdep*, R Development Core Team (2012). For each of the corresponding settings (varying T, ρ, σ_μ^2 and σ_ν^2), 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 (11). In both cases, we perform a two stage estimation procedure.

To keep the simulation setting as realistic as possible, the true parameter values are not used in the estimation procedure. In the first stage, OLS-residuals are used to estimate ρ, σ_μ^2 and σ_ν^2, i.e., $M_N = I_N - X_N(X_N' X_N)^{-1} X_N'$. The optimal weighting matrices for the moment conditions also depend on the true parameter values so we take starting values which correspond to a scalar covariance matrix for the disturbances ($\rho = 0$, $\sigma_\mu^2 = 1$, $\sigma_\nu^2 = 1$). This yields two sets of initial parameter estimates, one for the approach of Kapoor et al. (2007) and one for our approach.

In the second stage, FGLS-residuals are used to improve the estimates, i.e., now we take $M_N = I_N - X_N(X_N' \hat{\Omega}_{u,N}^{-1} X_N)^{-1} X_N' \hat{\Omega}_{u,N}^{-1}$. Both $\hat{\Omega}_{u,N}$ and the respective optimal weighting matrices are calculated by plugging in the parameter estimates of the first stage. Tables 1 and 2 give the resulting biases and mean square errors of the estimators, multiplied by 100.

- Table 1 here -

- Table 2 here -

Table 1 reveals the following aspects for the bias:

i) In almost all cases, the bias of our modified estimator is substantially reduced by some $80 - 90\%$ compared to the KKP estimator.

ii) In the second stage, the bias is generally smaller for the KKP estimator and our modified estimator, except for σ_ν^2 in small samples and for zero correlation. Note that the weighting matrix in the first stage is calculated with the starting value $\rho = 0$ which coincides with the true value in the case of zero correlation. Moreover, the bias of $\hat{\rho}$, $\hat{\sigma}_\nu^2$ nearly vanishes. For $\hat{\sigma}_\mu^2$, the bias vanishes rather fast for our estimator, but not for the KKP estimator.

iii) The bias of $\hat{\sigma}_\mu^2$ and of $\hat{\sigma}_\nu^2$ increases with the true variances by a factor of 2.

Calculating the analytical expressions in (13) for the true parameter values essentially yields the same results.

Table 2 reveals the following aspects for the MSE:
i) In basically all cases except for positive correlation ρ in the first stage, the MSE of our modified estimator is reduced compared to the KKP estimator although the improvement is not as large as for the bias. For σ_ν^2, the MSE's are more or less equal.
ii) The second stage reduces the MSE for both estimators, especially for $\hat{\rho}$.
iii) The MSE of $\hat{\sigma}_\mu^2$ and of $\hat{\sigma}_\nu^2$ increases with the true variances by a factor of 4.

The reduction of bias and MSE in the second stage, respectively, is partly caused by the fact that FGLS-residuals are a better replacement for the unobservable disturbances as compared to the OLS-residuals of the first stage. The second and more important reason are the optimal weighting matrices for the moment conditions, which in the second stage can be consistently estimated, whereas in the first stage, we only use starting values for the parameters.
Basically, one could try to further improve the estimates in a third stage with an updated optimal weighting matrix, calculated from the estimates of the second stage. To assess the room for improvement of additional iterations, we also ran simulations with the optimal weighting matrices which would not be known in practical applications. These simulations revealed that further MSE reduction is limited to about 5% for $T = 2$, 4% for $T = 5$ and 3% for $T = 10$ in most situations (for positive correlation the effect is partially larger) so that more than one iteration seems to be more or less superfluous.
We do not report detailed results of these simulations here, but they are available from the authors upon request.
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. By our finite sample adjustment, these significance tests, which are performed by plugging the parameter estimates into (1), can be improved. Table 3 compares the performance of the estimation approaches with respect to empirical rejection probabilities of the F-tests for statistical significance of all parameters, where the nominal level is $\alpha = 0.05$. For OLS regression in the first stage, we use

$$\widehat{\mathrm{Cov}}(\hat{\beta}) = \left(X_N^T X_N\right)^{-1} X_N^T \hat{\Omega}_{u,N} X_N \left(X_N^T X_N\right)^{-1};$$

for the second stage, the usual FGLS standard errors are computed.

- Table 3 here -

We can see that the empirical rejection probabilities exceed the nominal level of 0.05, especially for positive correlation ρ, and also for $N = 490$ (further simulations with subsets of the data indicate that for given N the overrejection probabilities are the smaller the larger n and the smaller the amount of regressors is). For our modified estimator, these overrejection probabilities are always smaller, and this is true for both OLS regression in the first stage and FGLS regression in the second stage. We conclude that our small sample adjustment helps to avoid false rejections.
In addition to the first simulation example, we have performed simulations for the data example from Kapoor et al. (2007). Here, we keep $T = 5$ and $\sigma_\mu^2 = \sigma_\nu^2 = 1$ fixed and let n and ρ vary. 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 $u_{1,n}$ 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.
The results are reported in Table 4, Table 5 and Table 6.

- Table 4 here -

- Table 5 here -

- Table 6 here -

Basically, the results are more or less similar to the other setting although the bias improvement is not that large and although the MSE of the KKP estimator is slightly smaller as compared to our estimator. We conjecture that the amount of regressors plays an important role. In contrast to the other setting, the empirical rejection probabilities are close to the nominal level for $N = 200$. Tables 4 and 5 suggest that for both estimators, bias and MSE are of order $1/n$ which is in line with the analytical illustration (13). The relative improvement caused by the modification does not seem to depend on the sample size, and this holds true not only for the bias reduction but also for the overrejection probabilities presented in Table 6. In contrast to this, in the Columbus example where n is fixed, the biases of all estimators and the MSE of $\hat{\rho}$ do not decrease with T, while the MSE of $\hat{\sigma}_\mu^2$ and of $\hat{\sigma}_\nu^2$ do decrease.

5. Application to Indonesian rice farming

We illustrate our results with an empirical analysis of Indonesian rice farming data which is used in several contexts in the econometric literature, see e.g. Horrace and Schmidt (2000), Druska and Horrace (2004), Feng and Horrace (2007) or Arnold and Wied (2010). [2] 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). With this, we have $N = n \cdot T = 171 \cdot 3 = 513$. 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 $w_{ij,n}$ of the spatial weighting matrix W_n is positive if observations i and j belong to (a) farms located in the same village. The row sums of W_n 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

[2] The data set is available in the data archive of the *Journal of Applied Econometrics* corresponding to the article by Horrace and Schmidt (2000).

approach. Initial estimates are obtained from OLS. In a second step, these estimates are used to perform FGLS regression with updated GMM estimates for ρ, σ_μ^2 and σ_ν^2, where the optimal weighting matrices are estimated by plugging in the estimates of the first stage. 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.

6. Summary and conclusions

This paper provides a finite sample adjustment for a GMM estimator in a spatial panel regression model suggested by Kapoor et al. (2007). The main idea is to explicitly take into account that observable regression residuals are different from the true but unobservable disturbances. The resulting modified moment conditions improve the finite sample properties of the GMM estimator in the sense that the bias of the estimators is largely reduced.

We illustrate the effect of this improvement in three ways. First, an analytical illustration shows that in contrast to the estimator of Kapoor et al. (2007), the modified expected system of equations is in fact solved by the true parameter values. Second, this finding is confirmed by simulation results. Finally, an empirical application to Indonesian rice data indicates that optimization problems regarding solutions which are outside the parameter space might be circumvented with the improvement.

As a second contribution, we derive asymptotic normality for the case that the number of observation units tends to infinity and provide a consistent estimator for the asymptotic covariance matrix of the estimators. This allows for asymptotically valid tests.

Consequently, our results should be useful for practitioners working in spatial econometrics.

References

Amann, H. and J. Escher (2008): *Analysis II*, Birkhäuser Verlag.

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

Anselin, L. (1988): *Spatial econometrics*, Dordrecht.

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.

Baltagi, B. and L. Liu (2011): "An improved generalized moments estimator for a spatial moving average error model," *Economics Letters*, 113(3), 282–284.

Davidson, J. (1994): *Stochastic limit theory*, Oxford University Press.

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.

Feng, Q. and W. Horrace (2007): "Fixed-effect estimation of technical efficiency with time-invariant dummies," *Economics Letters*, 95, 247–252.

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.

Horrace, W. and P. Schmidt (2000): "Multiple Comparisons with the Best, with Economic Applications," *Journal of Applied Econometrics*, 15(1), 1–26.

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.

Mutl, J. and M. Pfaffermayr (2011): "The Hausman Test in a Cliff and Ord Panel Model," *Econometrics Journal*, 14, 48–76.

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.

R Development Core Team (2012): *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

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

7. Appendix section

Proof of Lemma 1
It suffices to show that the eigenvalues of $I_n - \rho W_n$ are different from zero. Let λ be an arbitrary eigenvalue and $v \neq 0$ the corresponding eigenvector. Denote the row sum norm of a matrix with $||\cdot||_R$ and the maximum norm of a vector with $||\cdot||_\infty$, then

$$\begin{aligned}(I_n - \rho W_n)v &= \lambda v \\ \Rightarrow |1-\lambda|||v||_\infty &= ||\rho W_n v||_\infty \leq |\rho|||W_n v||_\infty < ||W_n v||_\infty \leq ||W_n||_R||v||_\infty = ||v||_\infty \\ \Rightarrow |1-\lambda| < 1, & \end{aligned}$$

which proves the lemma. ■

Proof of Lemma 2
a) Let Γ_N^* be the matrix Γ_N with all matrices M_N in Γ_N replaced by I_N and γ_N^* be the vector γ_N with all matrices M_N in γ_N replaced by I_N. To prove the result, we show that $\Gamma_N - \Gamma_N^* \to 0$ and $\gamma_N - \gamma_N^* \to 0$. Then, for each chosen M_N the resulting matrices Γ_N and Γ_N^* have the same limit with the following formal argument: Define Γ_N^1 based on matrices M_N^1 and Γ_N^2 based on matrices M_N^2, then $\Gamma_N^1 - \Gamma_N^2 = (\Gamma_N^1 - \Gamma_N^*) + (\Gamma_N^* - \Gamma_N^2)$. Both summands converge to the same limit and thus Γ_N^1 has the same limit as Γ_N^2. Consequently, the limit of Γ_N does not depend on M_N.
It remains to prove $\Gamma_N - \Gamma_N^* \to 0$ and $\gamma_N - \gamma_N^* \to 0$. Since for an arbitrary matrix A, $E(u_N' A u_N) = tr(A\Omega_{u,N})$, each entry in each of these differences can be written as

$$c_1 \cdot \frac{1}{n}\sum_{d=1}^{m} tr(D_{d,N}),$$

where c_1 is a constant, m is a finite number and $D_{d,N}$ is a matrix product, where at least one factor is equal to $I_N - M_N$ and the other factors are bounded.
The result then follows from $tr(D_{d,N}) \leq rank(I_N - M_N) \cdot c_2 = k \cdot c_2 < \infty$, where c_2 is a constant, because with this, each entry in each of the differences $\Gamma_N - \Gamma_N^*$ (and $\gamma_N - \gamma_N^*$) converges against 0. ■

b) and c) The proof follows up on ideas of the proofs of Lemma A1, A2 and A3 in Kapoor et al. (2007). We show the result exemplarily for G_N as the proof can simply be transferred to g_N. Note that

$$G_N - \Gamma_N = (G_N - G_N^*) + (G_N^* - \Gamma_N) := A_N + B_N,$$

where G_N^* is the matrix G_N with $M_N u_N$ replaced by u_N. We show that each component of A_N and B_N converges to 0 in probability.

Consider first A_N. Each component can be written as

$$\frac{1}{n} \cdot c_3 \cdot (u_N' M_N' E_N M_N u_N - u_N' E_N u_N) = \frac{1}{n} \cdot c_3 \cdot (u_N'(M_N' E_N M_N - E_N) u_N)$$
$$= \frac{1}{n} \cdot c_3 \cdot (u_N'(M_N' - I_N) E_N M_N u_N) + \frac{1}{n} \cdot c_3 \cdot (u_N' E_N (M_N - I_N) u_N),$$

where the entries of E_N are $O_P(1)$ and c_3 is a constant. The result then follows from the fact that the entries of $M_N - I_N$ are $O_P(n^{-1})$ and that the entries of u_N, E_N and M_N are $O_P(1)$. The convergence of B_N follows from the fact that $E(B_N) = 0$ by construction and that $Var(B_N) = \frac{1}{n^2} tr(F_N)$, where F_N is a $N \times N$ matrix whose entries are $O_P(1)$, converges against 0. ■

Remark to the proof of Lemma 2.a) An example for $D_{d,N}$ can be obtained by considering $\gamma_{11,N}^i$,

$$tr\left[M_N Q_{i,N} M_N (I_T \otimes W_n) \Omega_{u,N}\right] - tr\left[I_N Q_{i,N} I_N (I_T \otimes W_n) \Omega_{u,N}\right]$$
$$= \{tr\left[M_N Q_{i,N} M_N (I_T \otimes W_n) \Omega_{u,N}\right] - tr\left[M_N Q_{i,N} I_N (I_T \otimes W_n) \Omega_{u,N}\right]\}$$
$$+ \{tr\left[I_N Q_{i,N} M_N (I_T \otimes W_n) \Omega_{u,N}\right] - tr\left[I_N Q_{i,N} I_N (I_T \otimes W_n) \Omega_{u,N}\right]\}$$
$$= tr\left[M_N Q_{i,N} (M_N - I_N)(I_T \otimes W_n) \Omega_{u,N}\right] + tr\left[I_N Q_{i,N} (M_N - I_N)(I_T \otimes W_n) \Omega_{u,N}\right],$$

so that in this case $D_{1,N} = M_N Q_{i,N}(M_N - I_N)(I_T \otimes W_n)\Omega_{u,N}$ and $D_{2,N} = I_N Q_{i,N}(M_N - I_N)(I_T \otimes W_n)\Omega_{u,N}$. □

Remark to the proof of Lemma 2.b) and c) In the entry in the first row and first column (corresponding to $g_{11,N}^0$), $E_N = Q_{0,N} M_N (I_T \otimes W_n)$ and $c_3 = \frac{2}{T-1}$ (which does not change with n). □

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 identifiability condition. ■

Remark to the proof of Theorem 1 The basic idea of the proof is that the uniform convergence of $R_N(\tilde{\theta})$ to $R_0(\tilde{\theta})$ allows for applying an "argmin-theorem" yielding the convergence of $\operatorname{argmin}_{\tilde{\theta} \in S} R_N(\tilde{\theta})$ to $\operatorname{argmin}_{\tilde{\theta} \in S} R_0(\tilde{\theta})$, where the latter one is well-defined due to the identifiability condition. □

Proof of Lemma 3
For $j = 1, \ldots, 6$, the j-th row of $\sqrt{n}(G_N \cdot (\rho, \rho^2, \sigma_\mu^2, \sigma_\nu^2)' - g_N)$ is given by

$$\sqrt{n}(\varepsilon_N' C_{j,N} \varepsilon_N - c_{j,N}^*)$$

with $c^*_{j,N} = E(\varepsilon'_N C_{j,N}\varepsilon_N)$ from equations (2) - (7). With equation (10), every linear combination $\sum_{j=1}^6 c_j \tilde{C}_{j,N}$ with $\sum_{j=1}^6 c_j^2 = 1$ of this vector can be written as

$$\sqrt{n}\sum_{i=1}^{N}\sum_{j=1}^{6} c_j \lambda_{ji,N}(\zeta_{i,N}^2 - 1).$$

With Assumption 5, this linear combination fulfills the Ljapunov condition, i.e., for some $\delta > 0$ it holds

$$\begin{aligned}
&\lim_{n\to\infty}\sum_{i=1}^{N} E\left|\sqrt{n}\sum_{j=1}^{6} c_j\lambda_{ji,N}(\zeta_{i,N}^2 - 1)\right|^{2+\delta} \\
&\leq \lim_{n\to\infty} n^{1+\frac{\delta}{2}}\sum_{i=1}^{N}\sum_{j=1}^{6}|\lambda_{ji,N}|^{2+\delta}E|\zeta_{i,N}^2 - 1|^{2+\delta} = 0.
\end{aligned}$$

This directly allows for applying the central limit theorem from Davidson (1994), Theorem 23.11. Consequently, the linear combinations are asymptotically normal so that multivariate normality follows by the Cramér-Wold device. ■

Proof of Theorem 2
Due to the smoothness of the target function the estimators are the zeros of the derivative

$$\Psi(\tilde{\rho}, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2) := 2 \cdot Q'[(\tilde{\rho}, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2)] \cdot G_N' \cdot S_{W,N} \cdot (G_N \cdot (\tilde{\rho}, \tilde{\rho}^2, \tilde{\sigma}_\mu^2, \tilde{\sigma}_\nu^2)' - g_N).$$

With the mean value theorem for vector valued functions in integral form (see Amann and Escher, 2008, Theorem 3.10) it holds

$$\begin{aligned}
\Psi\begin{pmatrix}\hat{\rho}\\ \hat{\sigma}_\mu^2\\ \hat{\sigma}_\nu^2\end{pmatrix} &= 0 = \Psi\begin{pmatrix}\rho\\ \sigma_\mu^2\\ \sigma_\nu^2\end{pmatrix} + \int_0^1\left[D\Psi\begin{pmatrix}\rho + s(\hat{\rho}-\rho)\\ \sigma_\mu^2 + s(\hat{\sigma}_\mu^2 - \sigma_\mu^2)\\ \sigma_\nu^2 + s(\hat{\sigma}_\nu^2 - \sigma_\nu^2)\end{pmatrix}\right]ds \cdot \begin{pmatrix}\hat{\rho}-\rho\\ \hat{\sigma}_\mu^2 - \sigma_\mu^2\\ \hat{\sigma}_\nu^2 - \sigma_\nu^2\end{pmatrix} \\
\Leftrightarrow \begin{pmatrix}\hat{\rho}-\rho\\ \hat{\sigma}_\mu^2 - \sigma_\mu^2\\ \hat{\sigma}_\nu^2 - \sigma_\nu^2\end{pmatrix} &= -\left[\int_0^1\left[D\Psi\begin{pmatrix}\rho + s(\hat{\rho}-\rho)\\ \sigma_\mu^2 + s(\hat{\sigma}_\mu^2 - \sigma_\mu^2)\\ \sigma_\nu^2 + s(\hat{\sigma}_\nu^2 - \sigma_\nu^2)\end{pmatrix}\right]ds\right]^{-1}\Psi\begin{pmatrix}\rho\\ \sigma_\mu^2\\ \sigma_\nu^2\end{pmatrix},
\end{aligned}$$

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

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

Due to the consistency of $(\hat{\rho}, \hat{\sigma}_\mu^2, \hat{\sigma}_\nu^2)$, the process $(P_n(s) : s \in [0,1])$, defined by

$$P_n(s) := \begin{pmatrix} \rho + s(\hat{\rho} - \rho) \\ \sigma_\mu^2 + s(\hat{\sigma}_\mu^2 - \sigma_\mu^2) \\ \sigma_\nu^2 + s(\hat{\sigma}_\nu^2 - \sigma_\nu^2) \end{pmatrix}, s \in [0,1],$$

converges to $(\rho, \sigma_\mu^2, \sigma_\nu^2)$. Then, with the extended continuous mapping theorem, the process $(D\Psi(P_n(s)), s \in [0,1])$ converges to $2Q^{'}\Gamma_0^{'}S_W\Gamma_0 Q$. It follows

$$\begin{aligned}\operatorname*{plim}_{n\to\infty} \int_0^1 \left[D\Psi \begin{pmatrix} \rho + s(\hat{\rho} - \rho) \\ \sigma_\mu^2 + s(\hat{\sigma}_\mu^2 - \sigma_\mu^2) \\ \sigma_\nu^2 + s(\hat{\sigma}_\nu^2 - \sigma_\nu^2) \end{pmatrix} \right] ds &= \int_0^1 \operatorname*{plim}_{n\to\infty} \left[D\Psi \begin{pmatrix} \rho + s(\hat{\rho} - \rho) \\ \sigma_\mu^2 + s(\hat{\sigma}_\mu^2 - \sigma_\mu^2) \\ \sigma_\nu^2 + s(\hat{\sigma}_\nu^2 - \sigma_\nu^2) \end{pmatrix} \right] ds \\ &= \int_0^1 2Q^{'}\Gamma_0^{'}S_W\Gamma_0 Q ds = 2Q^{'}\Gamma_0^{'}S_W\Gamma_0 Q \end{aligned} \tag{14}$$

With Lemma 3, $\sqrt{n}(G_N \cdot (\rho, \rho^2, \sigma_\mu^2, \sigma_\nu^2)' - g_N)$ converges to $N(0, S_0)$ so that $\sqrt{n}\Psi(\rho, \sigma_\mu^2, \sigma_\nu^2)$ converges to $2Q^{'}\Gamma_0^{'}S_W N(0, S_0)$.
With this and (14),

$$\sqrt{n} \begin{pmatrix} \hat{\rho} - \rho \\ \hat{\sigma}_\mu^2 - \sigma_\mu^2 \\ \hat{\sigma}_\nu^2 - \sigma_\nu^2 \end{pmatrix} \to_d -(Q^{'}\Gamma_0^{'}S_W\Gamma_0 Q)^{-1} 2Q\Gamma_0 S_W N(0, S_0).$$

This completes the proof. ■

Figure 1: Profile target functions for ρ

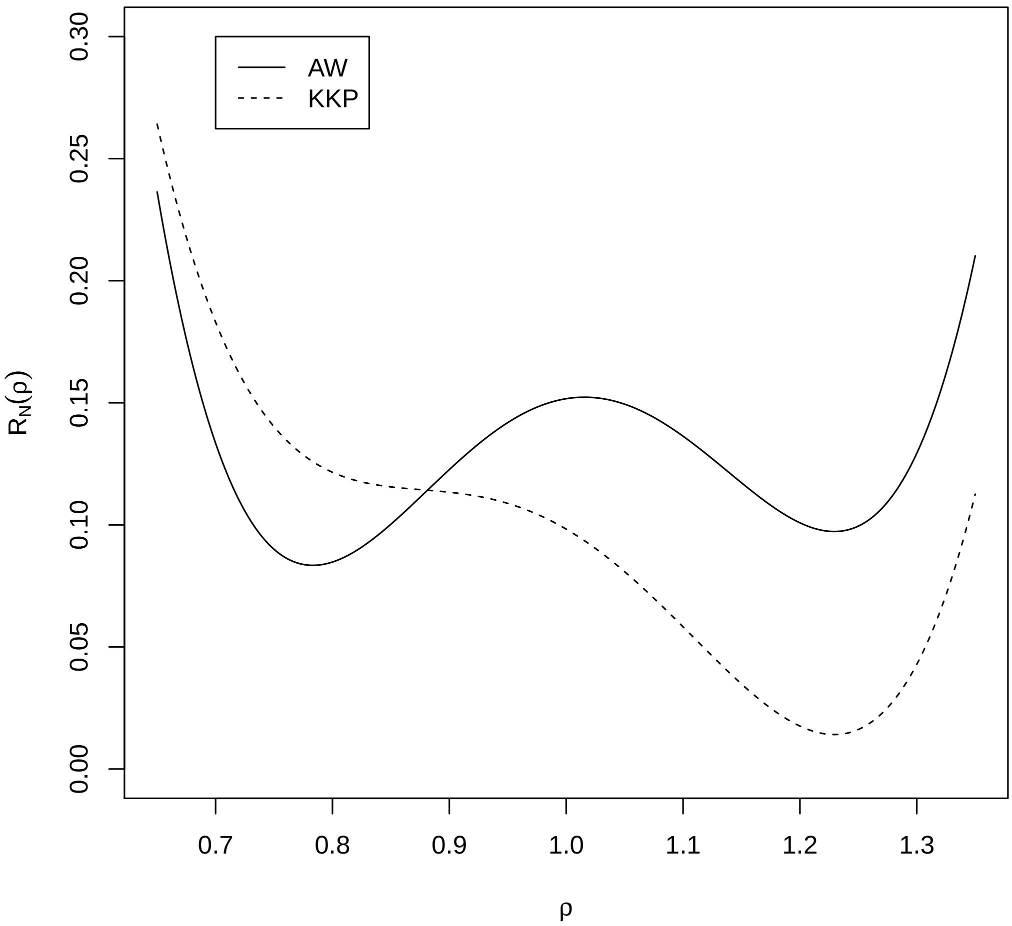

Table 1: Bias of the (Columbus) estimators for first stage (OLS-residuals, upper line) and for the iterative procedure (FGLS-residuals, lower line), multiplied by 100

			$\hat{\rho}$		$\hat{\sigma}^2_\mu$		$\hat{\sigma}^2_\nu$	
N	$\sigma^2_\nu = \sigma^2_\mu$	ρ	KKP	AW	KKP	AW	KKP	AW
98	1	0.5	-7.3	-4.0	-24.9	-13.9	1.7	2.1
			-2.7	-0.8	-23.3	-2.6	-1.6	-1.8
98	1	0	-16.2	-3.5	-30.7	-6.3	0.2	0.2
			-4.8	0.6	-25.7	-3.3	-2.4	-1.9
98	1	-0.5	-22.2	-3.7	-35.3	-5.7	-3.9	-0.7
			-8.9	-1.6	-27.7	-3.4	-3.8	-2.2
98	2	0.5	-6.6	-3.0	-54.9	-30.2	5.9	6.3
			-1.0	-0.1	-50.1	-10.0	-1.3	-1.8
98	2	0	-17.4	-4.4	-59.8	-9.9	1.6	1.4
			-6.0	-0.3	-50.0	-2.3	-3.3	-2.5
98	2	-0.5	-20.8	-2.2	-68.3	-8.7	-7.9	-1.5
			-7.5	-0.5	-54.5	-5.0	-7.6	-5.5
245	1	0.5	-6.1	-3.3	-21.6	-11.9	3.6	3.5
			-0.5	-0.4	-20.2	-4.3	-0.3	-0.2
245	1	0	-17.0	-4.1	-24.2	-4.3	1.0	0.9
			-2.4	-0.4	-19.4	-0.4	-1.0	-0.9
245	1	-0.5	-20.6	-1.9	-28.2	-3.8	-2.7	0.6
			-2.8	0.1	-20.4	-1.3	-1.3	-0.7
245	2	0.5	-7.2	-3.3	-42.1	-21.6	8.9	7.7
			-1.4	-0.9	-39.6	-7.2	0.5	0.6
245	2	0	-17.3	-4.3	-48.8	-8.7	2.0	1.7
			-2.5	-0.4	-39.5	-1.4	-1.8	-1.7
245	2	-0.5	-19.1	-0.8	-56.0	-7.5	-5.4	1.2
			-3.0	-0.2	-41.9	-3.8	-2.8	-1.6
490	1	0.5	-7.6	-4.9	-18.8	-9.8	5.7	4.8
			-0.5	-0.3	-17.6	-2.7	-0.3	-0.2
490	1	0	-19.1	-5.2	-23.2	-4.9	2.1	1.7
			-1.2	-0.2	-18.0	-0.9	-0.1	-0.1
490	1	-0.5	-20.7	-1.2	-27.1	-4.6	-2.5	0.9
			-1.4	0.1	-19.1	-2.2	-0.8	-0.5
490	2	0.5	-6.3	-4.1	-35.7	-20.7	9.6	8.0
			0.0	0.0	-34.2	-4.8	-1.5	-1.6
490	2	0	-18.1	-3.9	-45.0	-8.3	3.3	2.8
			-1.0	0.1	-35.1	-0.5	-0.9	-0.9
490	2	-0.5	-21.1	-1.3	-54.2	-8.7	-5.3	1.4
			-1.2	0.2	-37.9	-4.2	-1.7	-1.1

Table 2: MSE of the (Columbus) estimators for first stage (OLS-residuals, upper line) and for the iterative procedure (FGLS-residuals, lower line), multiplied by 100

			$\hat{\rho}$		$\hat{\sigma}_\mu^2$		$\hat{\sigma}_\nu^2$	
N	$\sigma_\nu^2 = \sigma_\mu^2$	ρ	KKP	AW	KKP	AW	KKP	AW
98	1	0.5	3.3	5.8	15.1	18.7	4.1	4.5
			2.3	2.0	14.2	12.8	3.8	4.0
98	1	0	6.4	5.6	17.4	12.0	4.0	4.0
			3.8	3.1	14.3	11.9	3.8	3.9
98	1	-0.5	8.6	5.3	20.0	12.6	3.9	3.9
			4.3	3.3	15.1	12.8	3.9	3.8
98	2	0.5	3.2	5.7	65.9	79.0	18.2	19.9
			2.9	1.8	60.5	51.4	17.3	17.4
98	2	0	7.2	5.9	65.9	45.0	16.5	16.6
			4.0	3.3	54.0	44.7	15.9	16.4
98	2	-0.5	8.2	5.1	76.4	47.2	16.4	16.8
			4.0	3.1	57.7	46.3	16.8	16.5
245	1	0.5	3.8	4.8	9.9	10.8	1.8	1.8
			0.5	0.6	9.1	7.5	1.1	1.1
245	1	0	6.5	5.4	10.6	7.1	1.1	1.0
			1.1	1.0	8.7	7.2	1.0	1.0
245	1	-0.5	7.9	5.2	12.8	8.0	1.2	1.3
			1.2	1.2	9.2	7.6	1.1	1.1
245	2	0.5	4.4	5.2	37.5	43.6	8.6	8.7
			0.5	0.9	35.0	27.8	4.4	4.8
245	2	0	6.4	5.2	42.3	28.7	4.2	4.2
			1.0	0.9	35.2	28.7	3.9	4.0
245	2	-0.5	7.0	5.0	50.7	30.9	4.6	5.1
			1.1	1.1	36.3	28.9	4.3	4.2
490	1	0.5	7.1	6.7	7.8	9.4	3.0	1.8
			0.2	0.5	7.2	5.8	0.5	0.6
490	1	0	8.1	6.5	9.5	6.4	0.6	0.5
			0.5	0.5	7.6	6.2	0.4	0.4
490	1	-0.5	8.4	5.9	11.1	6.3	0.6	0.8
			0.5	0.5	7.6	5.8	0.5	0.4
490	2	0.5	6.7	6.5	31.7	41.0	8.1	6.4
			0.3	0.3	29.3	25.1	2.0	2.3
490	2	0	7.5	6.3	36.0	24.4	2.3	2.2
			0.5	0.4	29.3	24.3	1.9	1.9
490	2	-0.5	8.7	5.8	46.1	28.1	2.6	3.3
			0.5	0.5	30.9	24.4	2.1	2.1

Table 3: Empirical rejection probabilities of (Columbus) F-significance tests for the regression coefficients in percent for the first stage (OLS-residuals, upper line) and for the iterative procedure (FGLS-residuals, lower line), nominal level $\alpha = 5\%$

		$\sigma_\mu^2 = \sigma_\nu^2 = 1$		$\sigma_\mu^2 = \sigma_\nu^2 = 2$	
N	ρ	KKP	AW	KKP	AW
98	0.5	23.2	20.6	25.1	20.6
		18.7	14.5	20.7	14.9
98	0	23.4	11.4	23.8	11.0
		20.2	10.1	20.3	10.4
98	-0.5	18.3	8.7	19.7	8.7
		24.0	13.1	21.3	10.3
245	0.5	22.7	18.5	22.1	18.4
		16.7	13.0	15.9	12.5
245	0	23.0	10.5	25.2	12.1
		17.9	10.1	18.3	9.8
245	-0.5	19.7	9.3	17.7	7.7
		19.4	10.1	15.6	9.1
490	0.5	24.5	20.2	25.9	22.5
		17.0	12.4	16.4	13.7
490	0	25.4	10.8	26.5	11.1
		16.1	9.1	17.5	8.7
490	-0.5	19.2	9.6	19.5	8.8
		17.6	10.0	15.4	8.5

Table 4: Bias of the (Virginia) estimators for first stage (OLS-residuals, upper line) and for the iterative procedure (FGLS-residuals, lower line), multiplied by 100

			$\hat{\rho}$		$\hat{\sigma}_\mu^2$		$\hat{\sigma}_\nu^2$	
N	J	ρ	KKP	AW	KKP	AW	KKP	AW
50	2	0.5	-0.85	1.55	-33.60	-20.43	6.92	6.12
			-0.48	0.60	-26.09	-11.02	-2.38	-1.62
50	2	0	-8.24	-2.12	-32.60	-14.92	3.80	4.89
			-2.24	-1.21	-21.69	-4.90	-2.92	-1.64
50	2	-0.5	-17.45	-4.46	-34.15	-15.56	4.01	4.56
			-1.78	-1.21	-22.30	-6.99	-4.97	-3.13
50	6	0.5	-0.34	-2.18	-28.47	-11.79	6.09	7.35
			-0.39	-0.12	-21.71	-2.50	-2.55	-1.68
50	6	0	-8.44	-5.24	-29.75	-12.31	2.23	4.39
			-0.55	1.36	-21.08	-3.23	-3.45	-2.29
50	6	-0.5	-21.86	-12.55	-33.24	-16.43	0.93	3.76
			-2.74	-1.73	-22.06	-5.61	-3.71	-2.35
100	2	0.5	-1.76	-0.65	-17.15	-9.11	5.06	3.43
			-0.84	-0.07	-11.97	-3.46	-1.34	-1.41
100	2	0	-5.44	-1.43	-16.97	-7.71	1.83	2.15
			-1.96	-0.98	-11.13	-2.10	-1.17	-0.67
100	2	-0.5	-7.68	-1.04	-16.89	-6.82	1.41	2.35
			-1.25	-0.71	-10.21	-1.60	-1.92	-1.04
100	6	0.5	-3.71	1.82	-12.95	-3.76	5.29	3.56
			-2.39	0.33	-9.32	-0.12	-1.04	-1.06
100	6	0	-7.36	2.65	-15.90	-5.58	0.84	1.34
			-2.37	2.01	-10.35	-1.32	-2.50	-1.79
100	6	-0.5	-8.07	3.85	-18.04	-7.72	-0.30	2.69
			-2.24	2.44	-11.72	-2.87	-2.53	-1.11
200	2	0.5	-2.45	-1.25	-8.92	-4.57	2.47	1.79
			-0.43	-0.13	-6.70	-1.98	-0.48	-0.59
200	2	0	-2.82	-0.85	-8.67	-3.87	0.79	1.04
			-0.74	-0.27	-5.19	-0.52	-0.80	-0.55
200	2	-0.5	-1.51	0.67	-9.21	-4.10	0.62	1.92
			-0.57	-0.17	-5.88	-1.50	-1.13	-0.59
200	6	0.5	0.51	0.99	-7.90	-3.67	3.46	1.79
			-0.04	1.14	-5.90	-1.23	-0.13	-0.29
200	6	0	-5.73	-1.14	-9.48	-4.56	0.92	0.98
			-1.14	0.68	-6.24	-1.66	-0.87	-0.59
200	6	-0.5	-5.75	0.00	-9.17	-4.15	-0.31	0.94
			-0.54	1.42	-5.39	-1.21	-1.08	-0.44

Table 5: MSE of the (Virginia) estimators for for first stage (OLS-residuals, upper line) and for the iterative procedure (FGLS-residuals, lower line), multiplied by 100

			$\hat{\rho}$		$\hat{\sigma}_\mu^2$		$\hat{\sigma}_\nu^2$	
N	J	ρ	KKP	AW	KKP	AW	KKP	AW
50	2	0.5	11.27	11.04	33.12	34.15	10.89	9.06
			1.92	2.92	28.93	34.04	5.23	5.59
50	2	0	10.19	11.45	30.65	33.60	6.34	7.00
			2.82	3.20	27.43	34.95	4.99	5.33
50	2	-0.5	21.44	13.12	32.75	37.00	10.69	10.40
			2.30	2.92	26.13	35.26	5.79	6.46
50	6	0.5	32.80	38.74	32.38	33.52	6.65	7.27
			12.08	13.74	29.85	37.84	4.79	5.12
50	6	0	47.68	58.33	29.59	33.66	6.07	6.96
			12.11	13.70	28.61	34.71	5.06	5.22
50	6	-0.5	69.91	64.59	30.77	34.67	6.88	8.11
			11.35	15.56	26.43	34.90	4.95	5.06
100	2	0.5	5.12	4.08	15.06	15.27	4.69	4.19
			0.72	0.93	13.31	14.86	2.71	2.76
100	2	0	4.76	4.66	15.72	16.26	3.27	2.86
			1.16	1.29	14.65	16.72	2.53	2.59
100	2	-0.5	7.20	3.05	17.34	18.25	4.34	3.98
			0.63	0.74	14.88	17.60	2.90	2.90
100	6	0.5	16.76	14.69	14.63	15.92	4.46	4.24
			2.41	2.69	14.41	16.25	2.51	2.56
100	6	0	18.21	17.51	14.71	15.87	3.11	3.14
			3.57	4.03	14.16	15.93	2.47	2.50
100	6	-0.5	19.25	20.65	15.34	16.10	3.37	4.26
			5.15	5.46	14.24	15.89	2.77	2.81
200	2	0.5	1.39	1.42	7.90	7.95	2.04	2.01
			0.34	0.36	7.34	7.71	1.39	1.41
200	2	0	2.17	2.23	7.80	7.80	1.26	1.27
			0.53	0.54	7.32	7.90	1.15	1.16
200	2	-0.5	1.89	1.44	7.88	8.04	2.04	2.18
			0.33	0.33	6.97	7.51	1.45	1.45
200	6	0.5	8.34	4.17	8.24	8.34	2.01	1.64
			0.86	0.73	7.89	8.27	1.24	1.23
200	6	0	8.66	7.44	7.75	7.78	1.62	1.39
			1.80	1.87	7.39	7.76	1.28	1.29
200	6	-0.5	8.50	8.68	7.60	7.76	1.53	1.65
			2.20	2.33	6.86	7.35	1.38	1.38

Table 6: Empirical rejection probabilities of (Virginia) significance tests for the regression coefficients in percent for the first stage (OLS-residuals, upper line) and for the iterative procedure (FGLS-residuals, lower line), nominal level $\alpha = 5\%$

		$J = 2$				$J = 6$			
		β_1		β_2		β_1		β_2	
N	ρ	KKP	AW	KKP	AW	KKP	AW	KKP	AW
50	0.5	11.7	8.8	11.1	7.3	13.7	12.1	11.3	9.5
		12.5	10.5	14.1	12.3	13.3	11.6	14.7	12.8
50	0	12.9	9.2	12.1	9.4	9.4	7.3	10.2	7.7
		13.6	12.3	14.0	12.5	10.5	9.6	12.4	11.2
50	-0.5	8.5	6.9	8.6	6.7	10.2	8.3	10.3	8.3
		13.1	12.3	12.9	12.1	10.0	9.4	10.6	9.5
100	0.5	7.3	5.5	7.4	5.2	7.4	6.5	7.0	6.0
		7.6	6.8	7.2	7.0	7.9	6.9	9.2	8.2
100	0	8.3	6.9	8.1	7.1	8.0	6.5	8.0	6.8
		8.4	8.0	8.9	8.0	9.0	8.1	10.1	9.1
100	-0.5	7.0	5.8	6.6	5.7	6.3	5.0	6.2	5.6
		7.0	6.8	7.0	6.7	9.3	8.1	9.0	8.1
200	0.5	6.4	5.5	5.9	5.1	6.6	5.0	6.0	5.7
		6.7	6.1	7.2	7.0	7.1	6.0	7.3	6.7
200	0	5.7	5.4	5.9	5.1	6.1	5.7	5.9	5.7
		7.3	6.9	6.8	6.3	6.7	6.1	7.2	7.0
200	-0.5	6.4	5.9	6.6	6.0	5.7	5.3	5.6	4.8
		5.2	5.0	5.3	4.9	6.8	6.7	7.2	6.9