SFB 823

Asymptotics of improved generalized moments estimators for spatial autoregressive error models

Carsten Drinkuth, Matthias Arnold

Discussion Paper

Nr. 22/2013

Asymptotics of Improved Generalized Moments Estimators for Spatial Autoregressive Error Models

by

Carsten Drinkuth

Fakultät Statistik, TU Dortmund

D-44221 Dortmund, Germany

carsten.drinkuth@tu-dortmund.de

and

Matthias Arnold*

Fakultät Statistik, TU Dortmund

D-44221 Dortmund, Germany

arnold@statistik.tu-dortmund.de

Abstract

This paper considers linear models with a spatial autoregressive error structure. Extending Arnold and Wied (2010), who develop an improved GMM estimator for the parameters of the disturbance process to reduce the bias of existing estimation approaches, we establish the asymptotic normality of a new weighted version of this improved estimator and derive the efficient weighting matrix. We also show that this efficiently weighted GMM estimator is feasible as long as the regression matrix of the underlying linear model is non-stochastic and illustrate the performance of the new estimator by a Monte Carlo simulation and an application to real data.

JEL Classification: C 13, C 21

Keywords: GMM estimation, Spatial autoregression, Regression residuals, Asymptotic normality

*Corresponding author. Phone: +49/231/755 5203, Fax: +49/231/755 5284.

1. Introduction and Summary

We consider data observed on spatial units like regions or districts, where dependencies between units induced by spatial closeness should be taken care of in statistical model building. In particular, we consider linear models with a spatially autoregressive error structure along the lines of Cliff and Ord (1973). There are two competing estimation approaches for the corresponding parameters. They can either be estimated by ML, see Anselin (1988), or, computationally more efficient, by the Kelejian and Prucha (1999) generalized method of moments (GMM) approach. This procedure bases on a system of three moment conditions, which can be expressed as means of quadratic forms depending on the innovation process. To create an empirical counterpart, Kelejian and Prucha (1999) replace the unobservable disturbances by regression residuals and optimize a quadratic objective function in terms of both the unknown autoregressive parameter and the unknown variance of the innovation process. The resulting estimator is consistent but suffers from a considerable bias if the sample size is small. To improve upon small-sample properties, Arnold and Wied (2010) develop a modified GMM estimator for the spatial autoregressive parameter by formulating the theoretical moment conditions in terms of residuals. This approach substantially improves bias and MSE in small samples, while the large sample properties agree with Kelejian and Prucha (1999).

Arnold and Wied (2012) exploit this idea to linear panel data models with spatial error terms of Kapoor et al. (2007) and Baltagi and Liu (2011) transfer the idea to the spatial moving average error process in Fingleton (2008).

This paper analyzes the asymptotic properties of the Arnold and Wied (2010) estimator and gives conditions for asymptotic normality. For the sake of generality the regressors of our linear models are allowed to be stochastic. It is well known that the statistical properties of GMM estimators can be improved by efficiently weighting the corresponding moment conditions. We therefore consider a weighted version of the estimator of Arnold and Wied (2010). The asymptotic framework of Kelejian and Prucha (2010) enables a quite general result on asymptotic normality which covers different model classes. We present the special form of the efficient weighting matrix as well as a consistent estimator for it. For non-stochastic regressors, the efficiently weighted estimator is shown to be feasible in the sense that the unknown parameters enter the efficient weighting matrix only as scalar factors, so that they are unnecessary to calculate the minimum. As an additional contribution, we generalize the moment conditions of Arnold and Wied (2010) to the case of a nonsymmetric projection matrix which maps the disturbances to the residuals.

The rest of the paper is organized as follows: Section 2 introduces the linear model with spatially autoregressive errors and the weighted version of the residual based GMM estimator. Section 3 presents the asymptotic results and the efficient weighting matrix. We develop a consistent estimation of the weighting matrix and show that the efficiently weighted estimator is feasible. Proofs are deferred to the appendix. Section 4 conducts a Monte Carlo simulation to examine the small sample performance and a real world data example is analyzed in Section 5. The paper ends with a short summary and suggestions for further research.

2. Model and estimator

This paper considers a linear regression model with n observations units as follows:

$$y_n = X_n\beta + u_n, \tag{1}$$

where y_n denotes the $(n \times 1)$-vector of observations on the dependent variable, X_n is the $(n \times k)$-matrix on the explanatory variables and β stands for the $(k \times 1)$-vector of regression coefficients. The $(n \times 1)$ disturbance vector u_n is generated as

$$u_n = \varrho W_n u_n + \epsilon_n \tag{2}$$

with an $(n \times n)$-matrix W_n of known constants, a scalar parameter ϱ and an $(n \times 1)$ innovation vector ϵ_n. We impose the following assumptions.

Assumption 1. *For the innovation process* $\{\epsilon_{i,n} : 1 \leq i \leq n, n \geq 1\}$*, it holds that*

$$\begin{aligned} E(\epsilon_{i,n}) &= 0 \\ E(\epsilon_{i,n}^2) &= \sigma^2 \quad \textit{with } 0 \leq \sigma^2 \leq b < \infty \\ E(|\epsilon_{i,n}|^{4+\eta}) &< \infty \quad \textit{for some } \eta > 0. \end{aligned}$$

Furthermore, $\epsilon_{1,n}, \ldots, \epsilon_{n,n}$ *are independent for all* $n \geq 1$*.*

Assumption 2. *a) The diagonal elements of* W_n *are zero for all* $n \geq 1$*. b) The row sums of* W_n *are equal to one for all* $n \geq 1$*. c)* $|\varrho| < 1$*.*

Assumptions 1 and 2 imply that $u_n = (I_n - \varrho W_N)^{-1}\epsilon_n$ such that

$$\text{Cov}(u_n) = \sigma^2(I_n - \varrho W_n)^{-1}(I_n - \varrho W_n^{'})^{-1} =: \Omega_{u,n}, \tag{3}$$

where I_n denotes the $(n \times n)$-identity matrix and A^{-1} and $A^{'}$ stand for the inverse and the transpose of a matrix A. Feasible generalized least squares estimation of β requires estimates of the unknown scalar parameters ϱ and σ^2.

To this end, Kelejian and Prucha (1999) suggest a GMM approach as an alternative to (quasi) maximum likelihood estimation. The corresponding moment conditions can be expressed as quadratic forms in the innovation vector ϵ_n. Arnold and Wied (2010) improve the finite sample properties by explicitly taking into account the difference between unobservable disturbances and observable regression residuals $\hat{u}_n$, where the latter are given by

$$\hat{u}_n = M_n u_n = M_n y_n,$$

and the projection matrix M_n depends on the estimation approach for β. For example, OLS leads to $M_n = I_n - X_n(X_n^{'}X_n)^{-1}X_n^{'}$ and FGLS gives $M_n = I_n - X_n(X_n^{'}\hat{\Omega}_{u,n}^{-1}X_n)^{-1}X_n^{'}\hat{\Omega}_{u,n}^{-1}$. The difference between unobservable disturbances and observable regression residuals can be characterized by M_n, respectively, and M_n is always available in applications because it depends only on the choice of estimator for β.

The main idea of Arnold and Wied (2010) is to calculate the theoretical moment conditions in terms of the residuals, since the empirical counterpart has to rely on the residuals anyway. The resulting theoretical moment conditions can be expressed as

$$E\begin{pmatrix} n^{-1}\epsilon_n' A_{1,n}\epsilon_n \\ n^{-1}\epsilon_n' A_{2,n}\epsilon_n \\ n^{-1}\epsilon_n' A_{3,n}\epsilon_n \end{pmatrix} = 0, \tag{4}$$

where

$$\begin{aligned} A_{1,n} &= M_n' M_n - \mathrm{diag}(M_n' M_n) \\ A_{2,n} &= M_n' W_n' W_n M_n - \mathrm{diag}(M_n' W_n' W_n M_n) \\ A_{3,n} &= M_n' W_n' M_n - \mathrm{diag}(M_n' W_n' M_n) \end{aligned}$$

and $\mathrm{diag}(A)$ stands for a diagonal matrix with the same main diagonal elements as A. Making use of

$$M_n\epsilon_n = M_n u_n - \varrho M_n W_n u_n \quad \text{and} \quad W_n M_n \epsilon_n = W_n M_n u_n - \varrho W_n M_n W_n u_n,$$

the theoretical system of equations can be written as

$$\Gamma_n \cdot \begin{pmatrix} \varrho \\ \varrho^2 \\ \sigma^2 \end{pmatrix} - \gamma_n = 0.$$

The (3×3)-matrix Γ_n is given by $\Gamma_n =$

$$\begin{pmatrix} \frac{2}{n}E\left(u_n' M_n' M_n W_n u_n\right) & -\frac{1}{n}E\left(u_n' W_n' M_n' M_n W_n u_n\right) & \frac{1}{n}\mathrm{tr}\left(M_n' M_n\right) \\ \frac{2}{n}E\left(u_n' M_n' \left[W_n + W_n'\right] M_n W_n u_n\right) & -\frac{1}{n}E\left(u_n' W_n' M_n' M_n W_n' W_n M_n W_n u_n\right) & \frac{1}{n}\mathrm{tr}\left(M_n' W_n' W_n M_n\right) \\ \frac{1}{n}E\left(u_n' M_n' \left[W_n + W_n'\right] M_n W_n u_n\right) & -\frac{1}{n}E\left(u_n' W_n' M_n' W_n M_n W_n u_n\right) & \frac{1}{n}\mathrm{tr}\left(M_n' W_n M_n\right) \end{pmatrix},$$

and the (3×1)-vector γ_n by

$$\gamma_n = \left(\frac{1}{n} E\left(u_n' M_n' M_n u_n\right) \ , \ \frac{1}{n} E\left(u_n' M_n' W_n' W_n M_n u_n\right) \ , \ \frac{1}{n} E\left(u_n' M_n' W_n M_n u_n\right) \right)',$$

where tr(.) stands for the trace of a matrix. The corresponding empirical counterpart is

$$H_n \cdot \begin{pmatrix} \varrho \\ \varrho^2 \\ \sigma^2 \end{pmatrix} - h_n =: v_n(\varrho, \sigma^2), \tag{5}$$

where

$$H_n = \begin{pmatrix} \frac{2}{n}\hat{u}_n' M_n W_n \hat{u}_n & -\frac{1}{n}\hat{u}_n' W_n' M_n' M_n W_n \hat{u}_n & \frac{1}{n}\mathrm{tr}\left(M_n' M_n\right) \\ \frac{2}{n}\hat{u}_n' \left[W_n + W_n'\right] M_n W_n \hat{u}_n & -\frac{1}{n}\hat{u}_n W_n' M_n' W_n' W_n M_n W_n \hat{u}_n & \frac{1}{n}\mathrm{tr}\left(M_n' W_n' W_n M_n\right) \\ \frac{1}{n}\hat{u}_n' \left[W_n + W_n'\right] M_n W_n \hat{u}_n & -\frac{1}{n}\hat{u}_n' W_n' M_n' W_n M_n W_n \hat{u}_n & \frac{1}{n}\mathrm{tr}\left(M_n' W_n M_n\right) \end{pmatrix},$$

$$h_n = \left(\frac{1}{n}\hat{u}_n' \hat{u}_n \ , \frac{1}{n}\hat{u}_n' W_n' W_n \hat{u}_n \ , \frac{1}{n}\hat{u}_n' W_n \hat{u}_n\right)'.$$

We slightly refine the moment conditions of Arnold and Wied (2010), because we allow for a nonsymmetric projection matrix M_n such that GLS regression is covered. Now we can formally define the weighted residual based GMM estimator for ϱ and σ^2.

Definition 1. *In the spatial error model (1) and (2), let Ψ_n be a sequence of stochastic (3×3) weighting matrices, which for $n \to \infty$ converges against a positive definite deterministic matrix Ψ. The weighted residual based GMM estimator $(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})$ is given by*

$$(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n}) = \underset{(\varrho,\sigma^2)\in[-1,1]\times[0,b]}{\operatorname{argmin}} \ v_n(\varrho, \sigma^2)' \Psi_n v_n(\varrho, \sigma^2).$$

Definition 1 contains the estimator of Arnold and Wied (2010) as a special case for $\Psi_n = I_3$ for all n.

3. Asymptotic results

To develop the asymptotic results, we need some further assumptions, which base on the assumption sets of Kelejian and Prucha (2010) and Arnold and Wied (2010).

Assumption 3. *a) For all $n \geq 1$, with probability one, the matrix X_n of (1) is of full column rank and the absolute entries of X_n are bounded, $|x_{ij,n}| < c_X < \infty$. b) $Q := \lim_{n\to\infty} n^{-1}(X_n^{'}X_n)$ is a finite regular matrix with probability one. c) The row and column sums of the absolute elements of W_n, $P_n := (I_n - \varrho W_n)^{-1}$ and M_n are bounded by c_W, c_P and c_M, respectively, i.e., for $i, j = 1, \ldots, n$, $n \geq 1$, with probability one it holds*

$$\sum_{i=1}^{n} |w_{ij,n}| < c_W \quad , \quad \sum_{j=1}^{n} |w_{ij,n}| < c_W$$
$$\sum_{i=1}^{n} |p_{ij,n}| < c_P \quad , \quad \sum_{j=1}^{n} |p_{ij,n}| < c_P$$
$$\sum_{i=1}^{n} |m_{ij,n}| < c_M \quad , \quad \sum_{j=1}^{n} |m_{ij,n}| < c_M.$$

Assumption 4. *For all $n \geq 1$, there are random vectors $d_{i,n} \sim (1, p)$ and $\Delta_n \sim (p, 1)$ with $E(|d_{ij,n}|^{2+\delta}) \leq c_d < \infty$ for some $\delta > 0$ and $\sqrt{n}\|\Delta_n\| = O_P(1)$, such that*

$$u_{i,n} - \hat{u}_{i,n} = d_{i,n}\Delta_n.$$

Assumption 5. *For each $n \geq 1$, the smallest eigenvalue $\lambda_{\min}\left(\Gamma_n^{'}\Gamma_n\right)$ of the matrix $\Gamma_n^{'}\Gamma_n$ is bounded away from zero:*

$$0 < \lambda_{\maltese} < \lambda_{\min}(\Gamma_n^{'}\Gamma_n).$$

Assumption 6. *a) For the sequence Ψ_n of stochastic weighting matrices and the deterministic matrix Ψ it holds that $\Psi_n - \Psi = o_P(1)$. b) The smallest and largest eigenvalues $\lambda_{\min}(\Psi)$ and $\lambda_{\max}(\Psi)$ of Ψ fulfill*

$$0 < \lambda_* < \lambda_{\min}(\Psi) \quad \textit{and} \quad \lambda_{\max}(\Psi) < \lambda^* < \infty.$$

Assumption 7. *Let $D_n = (d_{1,n}^{'}, \ldots, d_{n,n}^{'})^{'}$ with $d_{i,n}$ from Assumption 4. For each real* *matrix A with bounded row sums and column sums, it holds*

$$\frac{1}{n}D_n^{'}Au_n - \frac{1}{n}E(D_n^{'}Au_n) = o_P(1).$$

Assumption 8. *For Δ_n from 4, there is a deterministic matrix T_n with $|t_{ij,n}| < c_T < \infty$, so that*

$$\sqrt{n}\Delta_n = \frac{1}{\sqrt{n}}T_n^{'}\epsilon_n + o_P(1).$$

These Assumptions allow for different asymptotic results. First, we consider consistency of $(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})$. Arnold and Wied (2010) show that their unweighted version $(\hat{\varrho}_{\mathrm{RB},n}, \hat{\sigma}^2_{\mathrm{RB},n})$ is asymptotically equivalent to the estimator $(\hat{\varrho}_{\mathrm{KP},n}, \hat{\sigma}^2_{\mathrm{KP},n})$ of Kelejian and Prucha (1999), i.e.

$$(\hat{\varrho}_{\mathrm{RB},n}, \hat{\sigma}^2_{\mathrm{RB},n}) \xrightarrow{P} (\hat{\varrho}_{\mathrm{KP},n}, \hat{\sigma}^2_{\mathrm{KP},n}).$$

Hence, the consistency of $(\hat{\varrho}_{\mathrm{RB},n}, \hat{\sigma}^2_{\mathrm{RB},n})$ follows directly from the consistency of $(\hat{\varrho}_{\mathrm{KP},n}, \hat{\sigma}^2_{\mathrm{KP},n})$ (see Theorem 1 of Kelejian and Prucha (1999)). Further, the consistency remains valid if the estimator is constructed from a weighted objective function as long as the sequence of weighting matrices converges to a regular and deterministic matrix. Therefore, under Assumptions 1 to 5, which essentially coincide with the assumptions of Arnold and Wied (2010) or Kelejian and Prucha (1999), and additionally under Assumption 6, we have

$$(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n}) \xrightarrow{P} (\varrho, \sigma^2). \tag{6}$$

A detailed proof of this result can be constructed following the arguments of the proof of Theorem 1 of Kelejian and Prucha (2010).

For the asymptotic normality of the weighted residual based GMM estimator $(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})$ we start with some preliminary considerations. The estimator is build by replacing an unobservable vector of quadratic forms in ϵ_n by an observable vector of quadratic forms in $\hat{u}_n$, i.e. we can explain $v_n(\varrho, \sigma^2)$ in (5) by

$$v_n(\varrho, \sigma^2) = \begin{pmatrix} n^{-1}\hat{u}_n^{'}C_{1,n}\hat{u}_n \\ n^{-1}\hat{u}_n^{'}C_{2,n}\hat{u}_n \\ n^{-1}\hat{u}_n^{'}C_{3,n}\hat{u}_n \end{pmatrix} \tag{7}$$

with $C_{i,n} := \frac{1}{2}(P_n^{-1})'(A_{i,n}+A_{i,n}^{'})P_n^{-1}$, $i \in \{1,2,3\}$. In a second step, (7) can be rewritten in terms of a sum of quadratic plus linear forms in ϵ_n. Consequently, a central limit theorem (Theorem A1 of Kelejian and Prucha (2010)) applies and yields the following main result:

Theorem 1. *Under Assumptions 1 to 8, for $n \to \infty$ it holds that*

$$\sqrt{n}\left(\begin{pmatrix} \hat{\varrho}_{\Psi,n} \\ \hat{\sigma}^2_{\Psi,n} \end{pmatrix} - \begin{pmatrix} \varrho \\ \sigma^2 \end{pmatrix}\right) \xrightarrow{d} N\left(0, (G_n^{'}\Psi G_n)^{-1}G_n^{'}\Psi S_n \Psi G_n (G_n^{'}\Psi G_n)^{-1}\right)$$

where

$$\begin{aligned} G_n &= \Gamma_n \begin{pmatrix} 1 & 0 \\ 2\varrho & 0 \\ 0 & 1 \end{pmatrix} \\ S_n &= Cov\left(\frac{1}{\sqrt{n}}\begin{pmatrix} \frac{1}{2}\epsilon_n^{'}(A_{1,n}+A_{1,n}^{'})\epsilon_n + a_{1,n}^{'}\epsilon_n \\ \frac{1}{2}\epsilon_n^{'}(A_{2,n}+A_{2,n}^{'})\epsilon_n + a_{2,n}^{'}\epsilon_n \\ \frac{1}{2}\epsilon_n^{'}(A_{3,n}+A_{3,n}^{'})\epsilon_n + a_{3,n}^{'}\epsilon_n \end{pmatrix}\right), \end{aligned}$$

$a_{i,n} = T_n\alpha_{i,n}$ with T_n of Assumption 8 and $\alpha_{i,n} = 2n^{-1}E(D_n^{'}C_{i,n}u_n)$, $i \in \{1,2,3\}$.

The proof of Theorem 1 is given in the appendix. For the entries $s_{k\ell,n}$ of S_n it holds

$$s_{k\ell,n} = \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij,k,n}+a_{ji,k,n})(a_{ij,\ell,n}+a_{ji,\ell,n})\sigma^4 + \frac{1}{n}\sum_{i=1}^{n} a_{i,k,n}a_{i,\ell,n}\sigma^2, \tag{8}$$

where $a_{ij,h,n}$ denotes the element in row i and column j of the matrix $A_{h,n}$ and $a_{i,h,n}$ stands for the i^{th} entry of the vector $a_{h,n}$. This result is a direct consequence of Lemma A.1 of Kelejian and Prucha (2010).

Since the regression matrix X_n is allowed to be stochastic, Theorem 1 covers more complex model classes like the SARAR(1,1) model with spatial dependencies in both the response and error terms. On the other hand, in some applications all regressors $x_{i,n}$ are deterministic. Therefore, we state a special case of Theorem 1 for spatial error models with deterministic regressors.

Corollary 1. *Suppose that Assumptions 1 to 8 hold. Assume further that the matrix X_n and hence the matrix D_n of Assumption 4 are non-stochastic. Then it holds that*

$$\sqrt{n}\left(\begin{pmatrix}\hat{\varrho}_{\Psi,n}\\ \hat{\sigma}^2_{\Psi,n}\end{pmatrix}-\begin{pmatrix}\varrho\\ \sigma^2\end{pmatrix}\right)\xrightarrow{d} N\left(0,(G_n'\Psi G_n)^{-1}G_n'\Psi S_n^*\Psi G(G_n'\Psi G_n)^{-1}\right),$$

with G_n as in Theorem 1. In this case the entries $s^*_{kl,n}$ of the matrix

$$S_n^* = Cov\left(\frac{1}{\sqrt{n}}\begin{pmatrix}\frac{1}{2}\epsilon_n'(A_{1,n}+A_{1,n}')\epsilon_n\\ \frac{1}{2}\epsilon_n'(A_{2,n}+A_{2,n}')\epsilon_n\\ \frac{1}{2}\epsilon_n'(A_{3,n}+A_{3,n}')\epsilon_n\end{pmatrix}\right)$$

are of the form

$$s^*_{kl,n}=\frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij,k,n}+a_{ji,k,n})(a_{ij,\ell,n}+a_{ji,\ell,n})\sigma^4. \tag{9}$$

Corollary 1 immediately follows from Theorem 1. For deterministic regressors, S_n^* equals the covariance matrix of the original moment conditions. Furthermore, S_n^* does not depend on ϱ but only on σ^2, since the matrices $A_{i,n}$, $i\in\{1,2,3\}$, are fully known. This reduces the complexity of S_n^* and the asymptotic covariance matrix of $(\hat{\varrho}_{\Psi,n},\hat{\sigma}^2_{\Psi,n})$ considerably. Therefore, we now restrict ourselves to deterministic regression matrices X_n and develop further results for this model class.

The covariance matrix S_n^* depends on σ^2 only through a scalar parameter. Consequently, with the continuous mapping theorem we receive a consistent estimator $\hat{S}_n^*$ simply by plugging in a consistent estimator of σ^2, for example $\hat{\sigma}_{\mathrm{RB},n}$. We now turn to the matrix G_n. Consistency of $(\hat{\varrho}_{\Psi,n},\hat{\sigma}^2_{\Psi,n})$ implies $H_n-\Gamma_n=o_P(1)$. Again, plugging in a consistent estimator for ϱ gives a consistent estimator $\hat{G}_n$ of G_n by

$$\hat{G}_n=H_n\begin{pmatrix}1&0\\ 2\hat{\varrho}_{\mathrm{RB},n}&0\\ 0&1\end{pmatrix}.$$

We now combine these results with the help of the continuous mapping theorem and

Slutzky's theorem to yield a consistent estimator for the asymptotic covariance matrix given in Corollary 1.

Lemma 1. *In the spatial error model 1 and 2 with a deterministic regression matrix* X_n *and known weighting matrix* Ψ*, it holds under Assumptions 1 to 8 that*

$$(\hat{G}_n^{'}\Psi\hat{G}_n)^{-1}\hat{G}_n^{'}\Psi\hat{S}_n^{*}\Psi\hat{G}_n(\hat{G}_n^{'}\Psi\hat{G}_n)^{-1} \xrightarrow{p} (G_n^{'}\Psi G_n)^{-1}G_n^{'}\Psi S_n^{*}\Psi G_n(G_n^{'}\Psi G_n)^{-1}. \qquad (10)$$

Next we consider the weighting matrix Ψ. An efficient choice of weighting matrix improves the statistical properties of GMM estimators. The optimal weighting matrix is the inverse of S_n^* so that we define the efficiently weighted residual based GMM estimator as

$$(\hat{\varrho}_{\text{eff},n}, \hat{\sigma}^2_{\text{eff},n}) = \underset{(\varrho,\sigma^2)\in[-1,1]\times[0,b]}{\text{argmin}} \; v_n(\varrho,\sigma^2)'(S_n^*)^{-1}v_n(\varrho,\sigma^2).$$

The estimator is not operational for applications because $(S_n^*)^{-1}$ depends on σ^2. However, we can first estimate σ^2 and S_n^* and then use this estimates for the weighting matrix in a second step. This leads to the following two-stage estimation procedure:

1. Estimate ϱ and σ^2 with the unweighted estimator $(\hat{\varrho}_{\text{RB},n}, \hat{\sigma}^2_{\text{RB},n})$ of Arnold and Wied (2010), i.e. set $\Psi_n = I$.

2. Use $\hat{\sigma}^2_{\text{RB},n}$ to consistently estimate S_n^* and finally calculate the weighted estimator $(\hat{\varrho}_{\text{RBW},n}, \hat{\sigma}^2_{\text{RBW},n})$ by

$$(\hat{\varrho}_{\text{RBW},n}, \hat{\sigma}^2_{\text{RBW},n}) = \underset{(\varrho,\sigma^2)\in[-1,1]\times[0,b]}{\text{argmin}} \; v_n(\varrho,\sigma^2)'\left(\hat{S}_n^*\right)^{-1}v_n(\varrho,\sigma^2).$$

Note that any procedure that leads to a consistent estimation of σ^2 could be used in the first step. We still recommend the estimator of Arnold and Wied (2010) for the sake bias reduction. This two-stage estimator leads to an operational alternative to the efficiently weighted estimator. For the case of non-stochastic regressors, this two-step procedure is superfluous since the unknown parameter enters S_n^* only as a scalar factor such that the efficient GMM estimator $(\hat{\varrho}_{\text{eff},n}, \hat{\sigma}^2_{\text{eff},n})$ is feasible. This remarkable property of $(\hat{\varrho}_{\text{RBW},n}, \hat{\sigma}^2_{\text{RBW},n})$ and $(\hat{\varrho}_{\text{eff},n}, \hat{\sigma}^2_{\text{eff},n})$ is formulated in the next theorem.

Theorem 2. *In the spatial error model (1) and (2) with a non-stochastic regression matrix X_n, the efficiently weighted residual based GMM estimator $(\hat{\varrho}_{\text{eff},n}, \hat{\sigma}^2_{\text{eff},n})$ is feasible in the sense that it equals any estimator weighted by the inverse of a consistent estimation $\hat{S}^*_n$ of S^*_n, i.e.*

$$\underset{(\varrho,\sigma^2)\in[-1,1]\times[0,b]}{\operatorname{argmin}} \; v_n(\varrho,\sigma^2)'(S^*_n)^{-1}v_n(\varrho,\sigma^2) = \underset{(\varrho,\sigma^2)\in[-1,1]\times[0,b]}{\operatorname{argmin}} \; v_n(\varrho,\sigma^2)'(\hat{S}^*_n)^{-1}v_n(\varrho,\sigma^2).$$

The proof is straightforward since both objective functions are quadratic forms with regular weighting matrices that differ only with regard to a scalar parameter (σ^4 for the elements of S^*_n, $\hat{\sigma}^4$ for the elements of $\hat{S}^*_n$, compare (9)). Because this parameter does not influence the minima of the objective functions, both minima are equal.

4. Monte Carlo Simulation

The weighted version $(\hat{\varrho}_{\text{RBW},n}, \hat{\sigma}^2_{\text{RBW},n})$ improves the estimator of Arnold and Wied (2010), which itself has been developed to improve the estimator of Kelejian and Prucha (1999) in the first place. We investigate the small sample properties with the help of a Monte Carlo simulation. Our specifications follow those of Arnold and Wied (2010), i.e. we consider the model $y_n = X_n\beta + u_n$ with $u_n = \varrho W_n u_n + \epsilon_n$ for $n = 20, 100, 400$, $\varrho = -0.5, 0, 0.5$ and $\sigma^2 = 1$. The matrix W_n relates every $u_{i,n}$ to the three elements immediately preceding and succeeding it, each with the value $\frac{1}{6}$. The matrix X_n contains an intercept and two binary regressors and we simulate 10,000 observations of ϵ_n with $\epsilon_{i,n} \sim$ i.i.d. $N(0,1)$. We compare our new weighted estimator $(\hat{\varrho}_{\text{RBW},n}, \hat{\sigma}^2_{\text{RBW},n})$ with the two unweighted versions $(\hat{\varrho}_{\text{RB},n}, \hat{\sigma}^2_{\text{RB},n})$ and $(\hat{\varrho}_{\text{KP},n}, \hat{\sigma}^2_{\text{KP},n})$ with respect to the bias and the MSE. Table 1 shows the results.

Our new weighted residual based estimator for the autoregressive parameter ϱ reduces the bias up to 98% as compared to the estimator of Kelejian and Prucha (1999) and up to 92% when compared to the unweighted version of Arnold and Wied (2010) for $n = 20$. Especially for this small sample size, the MSE of our estimator rises up to 60% compared to the unweighted version, but it still deceeds the MSE of the Kelejian and Prucha (1999) estimator. The rising variance might result from the higher complexity of the two-stage estimation procedure. The results for the estimators of σ^2 show essentially the same

n	ϱ		Bias	MSE		Bias	MSE
20	-0.5	$\hat{\varrho}_{\mathrm{RBW}}$	-0.0127	0.8031	$\hat{\sigma}^2_{\mathrm{RBW}}$	-0.0583	0.1426
20	-0.5	$\hat{\varrho}_{\mathrm{RB}}$	-0.1428	0.5677	$\hat{\sigma}^2_{\mathrm{RB}}$	-0.0926	0.1353
20	-0.5	$\hat{\varrho}_{\mathrm{KP}}$	-0.5996	1.0271	$\hat{\sigma}^2_{\mathrm{KP}}$	-0.2751	0.1556
20	0	$\hat{\varrho}_{\mathrm{RBW}}$	-0.0173	0.9288	$\hat{\sigma}^2_{\mathrm{RBW}}$	-0.0630	0.1332
20	0	$\hat{\varrho}_{\mathrm{RB}}$	-0.1519	0.5796	$\hat{\sigma}^2_{\mathrm{RB}}$	-0.0923	0.1258
20	0	$\hat{\varrho}_{\mathrm{KP}}$	-0.6610	1.0921	$\hat{\sigma}^2_{\mathrm{KP}}$	-0.2667	0.1494
20	0.5	$\hat{\varrho}_{\mathrm{RBW}}$	-0.0148	0.8683	$\hat{\sigma}^2_{\mathrm{RBW}}$	-0.0527	0.1400
20	0.5	$\hat{\varrho}_{\mathrm{RB}}$	-0.1621	0.5471	$\hat{\sigma}^2_{\mathrm{RB}}$	-0.0803	0.1264
20	0.5	$\hat{\varrho}_{\mathrm{KP}}$	-0.6667	0.9960	$\hat{\sigma}^2_{\mathrm{KP}}$	-0.2334	0.1384
100	-0.5	$\hat{\varrho}_{\mathrm{RBW}}$	0.0018	0.0455	$\hat{\sigma}^2_{\mathrm{RBW}}$	-0.0135	0.0222
100	-0.5	$\hat{\varrho}_{\mathrm{RB}}$	-0.0281	0.0524	$\hat{\sigma}^2_{\mathrm{RB}}$	-0.0184	0.0225
100	-0.5	$\hat{\varrho}_{\mathrm{KP}}$	-0.0991	0.0630	$\hat{\sigma}^2_{\mathrm{KP}}$	-0.0591	0.0241
100	0	$\hat{\varrho}_{\mathrm{RBW}}$	-0.0096	0.0359	$\hat{\sigma}^2_{\mathrm{RBW}}$	-0.0154	0.0202
100	0	$\hat{\varrho}_{\mathrm{RB}}$	-0.0285	0.0390	$\hat{\sigma}^2_{\mathrm{RB}}$	-0.0167	0.0202
100	0	$\hat{\varrho}_{\mathrm{KP}}$	-0.0934	0.0493	$\hat{\sigma}^2_{\mathrm{KP}}$	-0.0498	0.0211
100	0.5	$\hat{\varrho}_{\mathrm{RBW}}$	-0.0192	0.0203	$\hat{\sigma}^2_{\mathrm{RBW}}$	-0.0092	0.0213
100	0.5	$\hat{\varrho}_{\mathrm{RB}}$	-0.0262	0.0192	$\hat{\sigma}^2_{\mathrm{RB}}$	-0.0090	0.0214
100	0.5	$\hat{\varrho}_{\mathrm{KP}}$	-0.0730	0.0252	$\hat{\sigma}^2_{\mathrm{KP}}$	-0.0315	0.0211
400	-0.5	$\hat{\varrho}_{\mathrm{RBW}}$	-0.0007	0.0103	$\hat{\sigma}^2_{\mathrm{RBW}}$	-0.0040	0.0053
400	-0.5	$\hat{\varrho}_{\mathrm{RB}}$	-0.0074	0.0116	$\hat{\sigma}^2_{\mathrm{RB}}$	-0.0050	0.0054
400	-0.5	$\hat{\varrho}_{\mathrm{KP}}$	-0.0249	0.0124	$\hat{\sigma}^2_{\mathrm{KP}}$	-0.0154	0.0055
400	0	$\hat{\varrho}_{\mathrm{RBW}}$	-0.0036	0.0078	$\hat{\sigma}^2_{\mathrm{RBW}}$	-0.0048	0.0051
400	0	$\hat{\varrho}_{\mathrm{RB}}$	-0.0076	0.0081	$\hat{\sigma}^2_{\mathrm{RB}}$	-0.0049	0.0051
400	0	$\hat{\varrho}_{\mathrm{KP}}$	-0.0228	0.0087	$\hat{\sigma}^2_{\mathrm{KP}}$	-0.0128	0.0052
400	0.5	$\hat{\varrho}_{\mathrm{RBW}}$	-0.0048	0.0035	$\hat{\sigma}^2_{\mathrm{RBW}}$	-0.0024	0.0052
400	0.5	$\hat{\varrho}_{\mathrm{RB}}$	-0.0057	0.0035	$\hat{\sigma}^2_{\mathrm{RB}}$	-0.0023	0.0052
400	0.5	$\hat{\varrho}_{\mathrm{KP}}$	-0.0158	0.0038	$\hat{\sigma}^2_{\mathrm{KP}}$	-0.0076	0.0052

Table 1: Results of Monte Carlo simulation

	estimate	standard error
intercept	59.96	5.77
income	-0.92	0.35
housing value	-0.31	0.09
ρ	0.59	0.16
σ^2	104.59	7.07

Table 2: Estimation results and estimated standard errors

expansion, but to a lower degree. For larger sample sizes, the bias of our new estimator still considerably deceeds the bias of the two competitors, while the MSE's of the three estimators hardly differ for growing sample sizes.

5. Application to Columbus

As a real data example, we consider the Columbus data of Cliff and Ord (1973), which is often used in spatial econometrics. It contains observations on 49 districts of Columbus, Ohio, for the variables *CRIME* (residential burglaries and vehicle thefts per thousand households), *INC* (household income in \$1000) and *HOVAL* (housing value in \$1000). A spatial weighting matrix is also available. Cliff and Ord (1973) suggest the following regression relationship

$$CRIME = \beta_0 + \beta_1 INC + \beta_2 HOVAL + u$$

The disturbance vector u is assumed to follow a spatial autoregressive process, i.e., $u = \varrho W u + \epsilon$.

We estimate the regression coefficients by feasible generalized least squares, where we plug in the weighted GMM estimates into the disturbance covariance matrix (3).

Table 2 shows the results. Household income and housing values both affect crime rates negatively. There is pronounced spatial dependence within the disturbances, and Theorem 1 allows for asymptotic standard errors of $\hat{\rho}$. The corresponding asymptotic 95% confidence interval for ρ is $[0.27, 0.91]$ so that the amount of spatial dependence is significantly different from zero.

6. Summary and conclusions

This article analyzes the asymptotic and finite properties of a residual based GMM estimator which simultaneously estimates the autoregressive parameter and the unknown error variance in linear models with spatially autoregressive error terms. For the sake of generality, the regression matrix X_n is assumed to be stochastic. A weighted version of an improved estimator of Arnold and Wied (2010). is shown to be consistent and asymptotically normal. The limit distribution provides the special form of the efficient weighting matrix, i.e the inverse of the covariance matrix. For deterministic regressors this matrix has a very simple form. It does not depend on the autoregressive parameter and the unknown error variance enters the matrix only as a scalar parameter. Therefore, it can easily be estimated. Moreover, in this case the efficiently weighted GMM estimator is even feasible. Additionally, a Monte-Carlo simulation shows that the small sample performance of our new efficiently weighted GMM estimator dominates the performance of both the unweighted version and the GMM estimator of Kelejian and Prucha (1999) concerning the bias and the MSE. Based on the asymptotic results we can additionally construct confidence sets and calculate standard errors for our estimator which is applied to the Columbus data.

Up till now, we have not transferred our results to the task of testing for spatial dependence. Simple tests can be derived from our asymptotic results given in this paper. Further research concerning this aspect is recommended.

Acknowledgements. Financial support by Deutsche Forschungsgemeinschaft (SFB 823, project A1) is gratefully acknowledged. We are grateful to Walter Krämer for useful discussions and comments.

References

Anselin, L. (1988): *Spatial Econometrics*, Kluwer Academic Publishing, Dordrecht.

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

——— (2012): "Improved GMM estimation of random effects panel data models with spatially correlated error components," *Papers in Regional Science*, to appear.

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

Cliff, A. and J. Ord (1973): *Spatial autocorrelation*, Pion, London.

Fingleton, B. (2008): "A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices," *Empirical Economics*, 34, 35–57.

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 heteroscedastic disturbances," *Journal of Econometrics*, 157, 53–67.

Magnus, J. and H. Neudecker (1999): *Matrix Differential Calculus with Applications in Statistics and Econometrics, 2nd Edition*, Wiley, New York.

Poetscher, D. and I. Prucha (1997): *Dynamic Nonlinear Econometric Models: Asymptotic Theory*, Springer, New York.

7. Appendix section

Proof of Theorem 1

We consider the objective function

$$Q_n(\varrho, \sigma^2) := \upsilon_n(\varrho, \sigma^2)' \Psi_n \upsilon_n(\varrho, \sigma^2)$$

of $(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})$. From the consistency it follows

$$\left(\frac{\partial v_n(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})}{\partial(\varrho, \sigma^2)'}\right)' \Psi_n v_n(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n}) = 0. \tag{11}$$

As a function of ϱ and σ^2 $v_n(\varrho, \sigma^2)$ maps from $\mathbb{R}^2$ to $\mathbb{R}^3$ and thus is vector valued. Therefore, we need to apply a multivariate mean value theorem to develop a linear approximation of $v_n(\varrho, \sigma^2)$ (see Magnus and Neudecker (1999)). Let $\bar{\varrho}_n$ and $\bar{\sigma}^2_n$ lie on the real lines between ϱ and $\hat{\varrho}_{\Psi,n}$ and between σ^2 and $\hat{\sigma}^2_{\Psi,n}$ respectively. Then it holds

$$v_n(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n}) = v_n(\varrho, \sigma^2) + \frac{\partial v_n(\bar{\varrho}_n, \bar{\sigma}^2_n)}{\partial(\varrho, \sigma^2)'} \left(\begin{pmatrix} \hat{\varrho}_{\Psi,n} \\ \hat{\sigma}^2_{\Psi,n} \end{pmatrix} - \begin{pmatrix} \varrho \\ \sigma^2 \end{pmatrix} \right). \tag{12}$$

Plugging (12) in (11) yields

$$\begin{aligned}\left(\frac{\partial v_n(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})}{\partial(\varrho, \sigma^2)'}\right)' \Psi_n \frac{\partial v_n(\bar{\varrho}_n, \bar{\sigma}^2_n)}{\partial(\varrho, \sigma^2)'} \sqrt{n} \left(\begin{pmatrix} \hat{\varrho}_{\Psi,n} \\ \hat{\sigma}^2_{\Psi,n} \end{pmatrix} - \begin{pmatrix} \varrho \\ \sigma^2 \end{pmatrix} \right) \\ = - \left(\frac{\partial v_n(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})}{\partial(\varrho, \sigma^2)'}\right)' \Psi_n \sqrt{n} v_n(\varrho, \sigma^2).\end{aligned} \tag{13}$$

To eliminate the quadratic form on the left-hand side of equation (13) we first consider the form of the Jacobian matrix. It holds

$$\frac{\partial v_n(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})}{\partial(\varrho, \sigma^2)'} = H_n \begin{pmatrix} 1 & 0 \\ 2\hat{\varrho}_{\Psi,n} & 0 \\ 0 & 1 \end{pmatrix},$$

which leads to

$$\left(\frac{\partial v_n(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})}{\partial(\varrho, \sigma^2)'}\right)' \Psi_n \frac{\partial v_n(\bar{\varrho}_n, \bar{\sigma}^2_n)}{\partial(\varrho, \sigma^2)'} = \begin{pmatrix} 1 & 0 \\ 2\hat{\varrho}_{\Psi,n} & 0 \\ 0 & 1 \end{pmatrix}' H_n' \Psi_n H_n \begin{pmatrix} 1 & 0 \\ 2\bar{\varrho}_n & 0 \\ 0 & 1 \end{pmatrix} =: \hat{\Phi}_n$$

The theoretical counterpart Φ_n of $\hat{\Phi}_n$ is given by

$$\Phi_n := \begin{pmatrix} 1 & 0 \\ 2\varrho & 0 \\ 0 & 1 \end{pmatrix}' \Gamma_n' \Psi \Gamma_n \begin{pmatrix} 1 & 0 \\ 2\varrho & 0 \\ 0 & 1 \end{pmatrix}.$$

Since $\Psi_n \xrightarrow{p} \Psi$ by Assumption 6 and $H_n \xrightarrow{p} \Gamma_n$, it follows from the continuous mapping theorem, that $\hat{\Phi}_n \xrightarrow{p} \Phi_n$. In the next step, we consider the inverse Φ_n^{-1} of Φ_n and the generalized inverse $\hat{\Phi}_n^+$ of $\hat{\Phi}_n$. As explained by Kelejian and Prucha (2010), we use the generalized inverse of $\hat{\Phi}_n$ to include the case that $(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})$ does not lie within the parameter space. Because of the consistency, this only occurs on a set with measure zero. With Lemma F1 of Poetscher and Prucha (1997) it follows $\hat{\Phi}_n^+ \xrightarrow{p} \Phi_n^{-1}$, which leads to

$$\sqrt{n}\left(\begin{pmatrix} \hat{\varrho}_{\Psi,n} \\ \hat{\sigma}^2_{\Psi,n} \end{pmatrix} - \begin{pmatrix} \varrho \\ \sigma^2 \end{pmatrix}\right) = -\hat{\Phi}_n^+ \left(\frac{\partial v_n(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})}{\partial(\varrho, \sigma^2)'}\right)' \Psi_n \sqrt{n} v_n(\varrho, \sigma^2) + o_P(1). \quad (14)$$

For the first part of the right-hand side of equation (14) we have

$$\hat{\Phi}_n^+ \left(\frac{\partial v_n(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})}{\partial(\varrho, \sigma^2)'}\right)' \Psi_n - \Phi_n^{-1} \begin{pmatrix} 1 & 0 \\ 2\varrho & 0 \\ 0 & 1 \end{pmatrix} \Gamma_n' \Phi = o_P(1). \quad (15)$$

For the second part we use the asymptotic theory developed by Kelejian and Prucha (2010) again. With the alternative characterization of the empirical moment conditions given in (7) and Lemma C.1 of Kelejian and Prucha (2010) it holds

$$\sqrt{n} v_n(\varrho, \sigma^2) = \frac{1}{\sqrt{n}} \begin{pmatrix} \hat{u}_n' C_{1,n} \hat{u}_n \\ \hat{u}_n' C_{2,n} \hat{u}_n \\ \hat{u}_n' C_{3,n} \hat{u}_n \end{pmatrix} = \frac{1}{\sqrt{n}} \begin{pmatrix} u_n' C_{1,n} u_n \\ u_n' C_{2,n} u_n \\ u_n' C_{3,n} u_n \end{pmatrix} + \begin{pmatrix} \alpha_{1,n}' \sqrt{n} \Delta_n \\ \alpha_{2,n}' \sqrt{n} \Delta_n \\ \alpha_{3,n}' \sqrt{n} \Delta_n \end{pmatrix} + o_P(1),$$

where $\alpha_{i,n} = 2n^{-1} E(D_n' C_{i,n} u_n)$, $i \in \{1, 2, 3\}$. To get a representation depending on ϵ_n, we replace u_n by $P_n\epsilon_n$. Further, we make use of Assumption 8, which leads to

$$\sqrt{n}v_n(\varrho,\sigma^2) = \frac{1}{\sqrt{n}}\begin{pmatrix} \frac{1}{2}\epsilon_n'(A_{1,n}+A_{1,n}')\epsilon_n + a_{1,n}'\epsilon_n \\ \frac{1}{2}\epsilon_n'(A_{2,n}+A_{2,n}')\epsilon_n + a_{2,n}'\epsilon_n \\ \frac{1}{2}\epsilon_n'(A_{3,n}+A_{3,n}')\epsilon_n + a_{3,n}'\epsilon_n \end{pmatrix} + o_P(1) \tag{16}$$

with $a_{i,n} = T_n\alpha_{i,n}$, $i \in \{1,2,3\}$. The quadratic forms in (16) coincide with the original moment conditions and the linear forms are used to model the difference in distribution between $\hat{u}_n$ and u_n. In the next step we consider the expectation and the covariance matrix of the right-hand side of equation (15). From $E(\epsilon_n) = 0$ and $\text{tr}(A_{i,n}) = 0$ it follows

$$E\left(\frac{1}{2}\epsilon_n'(A_{i,n}+A_{i,n}')\epsilon_n + a_{i,n}'\epsilon_n\right) = 0\ , \quad i \in \{1,2,3\}.$$

Further we define

$$S_n := \text{Cov}\left(\frac{1}{\sqrt{n}}\begin{pmatrix} \frac{1}{2}\epsilon_n'(A_{1,n}+A_{1,n}')\epsilon_n + a_{1,n}'\epsilon_n \\ \frac{1}{2}\epsilon_n'(A_{2,n}+A_{2,n}')\epsilon_n + a_{2,n}'\epsilon_n \\ \frac{1}{2}\epsilon_n'(A_{3,n}+A_{3,n}')\epsilon_n + a_{3,n}'\epsilon_n \end{pmatrix}\right)$$

with entries given in (8). To finally develop the asymptotic distribution of $(\hat{\varrho}_{\Psi,n}, \hat{\sigma}^2_{\Psi,n})$, we use the central limit theorem (Theorem A.1 of Kelejian and Prucha (2010)) for a vector of quadratic forms and linear forms. Assuming that $\lambda_{\min}(S_n) \geq c_S > 0$, it leads to

$$\zeta_n := -S_n^{-1/2}\frac{1}{\sqrt{n}}\begin{pmatrix} \frac{1}{2}\epsilon_n'(A_{1,n}+A_{1,n}')\epsilon_n + a_{1,n}'\epsilon_n \\ \frac{1}{2}\epsilon_n'(A_{2,n}+A_{2,n}')\epsilon_n + a_{2,n}'\epsilon_n \\ \frac{1}{2}\epsilon_n'(A_{3,n}+A_{3,n}')\epsilon_n + a_{3,n}'\epsilon_n \end{pmatrix} \xrightarrow{d} N(0, I_3).$$

Putting $-S_n^{1/2}\zeta_n$ and (15) into (14) yields

$$\sqrt{n}\left(\begin{pmatrix} \hat{\varrho}_{\Psi,n} \\ \hat{\sigma}^2_{\Psi,n} \end{pmatrix} - \begin{pmatrix} \varrho \\ \sigma^2 \end{pmatrix}\right) = \Phi^{-1}\begin{pmatrix} 1 & 0 \\ 2\varrho & 0 \\ 0 & 1 \end{pmatrix}' \Gamma_n'\Psi S_n^{1/2}\zeta_n + o_P(1).$$

Finally, let the matrix G_n be defined as

$$G_n := \frac{\partial}{\partial(\varrho, \sigma^2)'} \Gamma_n \begin{pmatrix} \varrho \\ \varrho^2 \\ \sigma^2 \end{pmatrix} = \Gamma_n \begin{pmatrix} 1 & 0 \\ 2\varrho & 0 \\ 0 & 1 \end{pmatrix},$$

which leads to $\Phi_n = G_n' \Psi G_n$ and

$$\sqrt{n} \left(\begin{pmatrix} \hat{\varrho}_{\Psi,n} \\ \hat{\sigma}^2_{\Psi,n} \end{pmatrix} - \begin{pmatrix} \varrho \\ \sigma^2 \end{pmatrix} \right) = (G_n' \Psi G_n)^{-1} G_n \Psi S_n^{1/2} \zeta_n + o_P(1),$$

which completes the proof. ■