

On the Asymptotic Distribution of the Moran I Test Statistic with Applications

Harry H. Kelejian and Ingmar R. Prucha*

Department of Economics

University of Maryland

College Park, MD 20852

October 2000

Abstract

By far, the most popular test for spatial correlation is the one based on Moran's (1950) I test statistic. Despite this, the available results in the literature concerning the large sample distribution of this statistic are limited and have been derived under assumptions that do not cover many applications of interest. In this paper we first give a general result concerning the large sample distribution of Moran I type test statistics. We then apply this result to derive the large sample distribution of the Moran I test statistic for a variety of important models. In order to establish these results we also give a new central limit theorem for linear-quadratic forms.

JEL Classification: C12, C21

Key words: Moran I test, spatial autocorrelation, asymptotic distribution, central limit theorem

*Corresponding author: Ingmar Prucha, Department of Economics, University of Maryland, College Park, MD 20742, Tel.: 301-405-3499, Fax: 301-405-3542, Email: prucha@econ.umd.edu

1 Introduction*

Spatial models have a long history in the regional science and geography literature.¹ In recent years issues concerning spatial dependence between cross sectional units have also received increasing attention in the economics literature.² One form of such dependence relates to spill-overs between the dependent variables; another relates to spill-overs between the disturbance terms, which leads to, among other things, the problem of spatial correlation. The efficiency and properties of estimators as well as the properties of other statistics will in general depend upon whether or not a model's disturbance terms are indeed spatially correlated. As a result, it is important to test for the presence of spatial correlation.

By far, the most popular test for spatial correlation is the one based on the Moran (1950) I test statistic. In essence this test statistic is formulated as a (properly normalized) quadratic form in terms of the variables that are being tested for spatial correlation. Moran's original specification standardizes the variables by subtracting the sample mean, and then deflating by an appropriate factor. Cliff and Ord (1972, 1973, 1981) generalized Moran's I statistic in order to derive a test for spatial correlation in a linear regression model. The Cliff and Ord generalization is formulated in terms of a quadratic form of estimated residuals.³ The matrix appearing in the quadratic form is typically referred to as the spatial weighting matrix. Assuming that the innovations are i.i.d. normal Cliff and Ord derive the large sample distribution of the Moran I test statistic, and also derive its small sample moments. Sen (1976) derives the large sample distribution of the Moran I test statistic under the weaker assumption that the variables are i.i.d. distributed, given that certain moment conditions hold. However, his analysis only covers Moran's

*This research was supported in part by the National Science Foundation through grant SES-0001780. We would like to thank Dilip Madan, Benedikt Pötscher, three anonymous referees and the Editor for helpful comments.

¹See, e.g., Anselin (1988), Bennett and Hordijk (1986), Cliff and Ord (1973, 1981), and Cressie (1993), and references cited therein.

²Some recent applied and theoretical studies include Dubin (1988), Moulton (1990), Delong and Summers (1991), Case (1991, 1993), Krugman (1991, 1995), Case, Hines, and Rosen (1993), Holtz-Eakin (1994), Besley and Case (1995), Shroder (1995), Conley (1996), Topa (1996), Audretsch and Feldman (1996), Ausubel et al. (1997), Kelejian and Prucha (1998, 1999), Pinkse and Slade (1998), Pinkse (1999), and Lee (1999a,b,c).

³The generalization contains Moran's original specification as a special case corresponding to a model where the only regressor relates to the intercept.

original specification and not Cliff and Ord's generalized specification.

The conditions maintained by the above results concerning the large sample distribution of the Moran I test statistic are restrictive and rule out many potential applications. For instance, the results do not allow for heteroskedasticity in the innovations. As will become evident, this rules out many applications which relate to limited dependent variable models, even though such models have become of increasing interest in the analysis of spatial data sets. Furthermore there are no formal results in the literature which cover spatial regression models where one of the regressors is a spatial lag of the dependent variable, even though such models are widely used in applied work and, indeed, primary interest often focuses on the coefficient of the spatially lagged dependent variable. Also, to the best of our knowledge, the results that are available in the literature concerning the large sample distribution of the Moran I statistic have only been derived under the null hypothesis of zero spatial correlation, but not under the alternative hypothesis, which is needed in order to assess the power of the test. Furthermore, the available results do not account for the triangular nature of the data even though, as illustrated below, this is implied by the specification of many models considered in practice.⁴

For purposes of interpretation we note that the Moran I test statistic is equivalent to the Lagrange multiplier test statistic derived from a linear regression model without a spatial lag, as was demonstrated by Burridge (1980). Other tests for spatial correlation have also been considered. We note that most of these tests are again formulated as Lagrangian multipliers tests, where the unrestricted model is specified at different degrees of generality. An inspection of those test statistics shows that they are essentially quadratic forms, or in some cases linear-quadratic forms of estimated residuals; see, e.g., Anselin and Florax (1995) and Anselin et al. (1996). Therefore, large sample issues concerning these Lagrangian multiplier test statistics are essentially identical to those concerning the Moran I test statistic, and so the gaps in the literature concerning the asymptotic properties of Lagrangian multipliers tests are analogous to those for the Moran I test statistic.

⁴An exception is a recent paper by Pinkse (1999). That paper, which also provides novel results concerning the large sample distribution of Moran I type test statistics, seems to have been written independently at approximately the same time as this paper. In contrast to Pinkse we allow for heteroskedastic innovations and for models with spatial lags in the dependent and independent variables. We note that the two papers also differ in various other aspects, and that neither paper contains the other as a special case.

The purpose of this paper is to fill in these gaps in the literature, with a particular focus on establishing the limiting distribution of the Moran I test statistic for various limited dependent variable models (the Tobit, dichotomous, polychotomous and a sample selection model) and for linear cross sectional models which contain a spatially lagged dependent variable. Instead of developing asymptotic results for each model separately, we first derive a set of general results concerning the asymptotic distribution of quadratic forms based on estimated residuals, and then apply those results to formally establish the asymptotic distribution of the Moran I test statistic for spatial correlation in the models of interest. Our general results allow for heteroskedastic innovations, for spatial lags in the dependent variable, for estimated disturbances which depend nonlinearly on the data and estimated parameters, and for the possible triangular nature of the data. We also give general results concerning the consistent estimation of the asymptotic variance of linear-quadratic forms. The results are such that applications to still other models of interest, as well as to Lagrangian multiplier test statistics in a spatial context, will become evident. Finally, our formal demonstrations are based on sets of low level assumptions.

Under the adopted setup the quadratic forms based on estimated residuals are seen to be asymptotically equivalent to quadratic forms (or in certain cases, to linear-quadratic forms) in the unobserved model innovations. As a technical point of interest we note that the elements of the matrix of the latter quadratic form will frequently depend on the sample size, i.e., form triangular arrays. In deriving our asymptotic results we therefore needed a central limit theorem for linear-quadratic forms that allows for heteroskedastic (possibly nonnormally distributed) innovations, and that allows for the weights in the linear-quadratic form to depend on the sample size. Since, to the best of our knowledge, such a central limit theorem is not directly available in the literature, we establish in this paper such a central limit theorem.⁵

The paper is organized as follows. A linear model containing a spatially lagged dependent variable is specified in Section 2 which serves to motivate certain essential issues. Our central limit theorem for linear-quadratic forms in model innovations is given in Section 3. This section also contains results relating to quadratic forms in estimated model disturbances, as well as

⁵Of course, various central limit theorems for quadratic forms have been considered in the literature; see, e.g., Whittle (1964), Beran (1972), Sen (1976), Girraitis and Taqqu (1998) and Pinkse (1999). Unfortunately, the assumptions maintained by those theorems are not satisfied within the context of the present paper.

results relating to the estimation of the variance of linear-quadratic forms. Applications of our theoretical results to linear models containing spatially lagged dependent variables, as well as to various limited dependent variable models are given in Section 4. Suggestions for further work are given in Section 5. Proofs are relegated to the appendix.

We adopt the following notation and conventions: Let P_n with $n \in \mathbf{N}$ be some matrix; we then denote the (i, j) -th element of P_n as $p_{ij,n}$. Similarly, if r_n is a vector, then $r_{i,n}$ denotes the i -th element of r_n . An analogous convention is adopted for matrices and vectors that do not depend on the index n , in which case the index n is suppressed on the elements. We say the elements of the (sequence of) matrices P_n are uniformly bounded in absolute value if $\sup_{1 \leq i, j \leq n, n \geq 1} |p_{ij,n}| < \infty$. As a matrix norm for P_n we choose $|P_n| = [tr(P_n' P_n)]^{1/2}$ where $tr(\cdot)$ denotes the trace operator. If P_n is a square matrix, then P_n^{-1} denotes the inverse of P_n . If P_n is singular, then P_n^{-1} should be interpreted as the generalized inverse of P_n . We say the row and column sums of the (sequence of $n \times n$) matrices P_n are uniformly bounded in absolute value if

$$\sup_{1 \leq i \leq n, n \geq 1} \sum_{j=1}^n |p_{ij,n}| < \infty \text{ and } \sup_{1 \leq j \leq n, n \geq 1} \sum_{i=1}^n |p_{ij,n}| < \infty.$$

Furthermore we adopt the conventions of Dhrymes (1978), p. 523, concerning vector differentiation. In particular, if $\psi(\alpha)$ is a scalar function and $\alpha = [\alpha_1, \dots, \alpha_m]'$, then $\partial\psi/\partial\alpha = [\partial\psi/\partial\alpha_1, \dots, \partial\psi/\partial\alpha_m]$. Finally, throughout the paper all random variables are assumed to be defined on some probability space $(\Omega, \mathfrak{F}, P)$.

2 Testing for Spatial Autocorrelation: A Preliminary Discussion

As remarked in the introduction, an aim of this paper is to establish the limiting behavior of the Moran I test statistic for a variety of limited dependent variable models, and the linear spatial first order autoregressive model with first order autoregressive disturbances - for short the spatial ARAR(1,1) model, adopting the terminology of, e.g., Anselin (1988). Towards deriving the limiting behavior of the Moran I test statistic for these particular models we first develop a set of general results concerning the limiting distribution

of linear-quadratic forms that can be used to establishing the limiting distribution of the Moran I test statistic not only for these particular models but for a wide class of models. The results can also be used to establish the limiting distribution of various Lagrangian multiplier test statistics for spatial autocorrelation. To motivate and put the general results into context we start with a preliminary discussion of the Moran I test statistic within the context of the widely used spatial ARAR(1,1) model ($n \in \mathbf{N}$):⁶

$$\begin{aligned} y_n &= \lambda_0 M_n y_n + X_n \beta_0 + u_n, \\ u_n &= \rho_0 M_n u_n + \varepsilon_n, \end{aligned} \tag{2.1}$$

where y_n is the $n \times 1$ vector of observations on the dependent variable, X_n is the $n \times k$ matrix of observations on k exogenous variables, M_n is an $n \times n$ spatial weighting matrix of known constants, β_0 is the $k \times 1$ vector of regression parameters, λ_0 and ρ_0 are scalar autoregressive parameters with $|\lambda_0| < 1$ and $|\rho_0| < 1$, u_n is the $n \times 1$ vector of regression disturbances, and ε_n is an $n \times 1$ vector of innovations. The variables $M_n y_n$ and $M_n u_n$ are typically referred to as spatial lags of y_n and u_n , respectively.⁷ Since spatial weighting matrices are often row normalized we allow for the elements of M_n to depend on n , i.e., to form triangular arrays. For reasons of generality we also permit the elements of X_n and ε_n to depend on n . We condition our analysis on the realized values of the exogenous variables and so, henceforth, X_n will be viewed as a matrix of constants. We note that this specification is consistent with the exogenous variables being generated by a process that entails both spatial correlation and heteroskedasticity. We also note that X_n may contain spatial lags of exogenous variables. The spatial weights $m_{ij,n}$ will typically be specified to be nonzero if cross sectional unit j relates to i in a meaningful way. In such cases, units i and j are said to be neighbors. Usually neighboring units are taken to be those units which are close in some dimension – e.g., geographic, technological, etc. Clearly the weighting matrix will reflect the ordering of the data, which can be arbitrary.

⁶This model will be used throughout the paper to motivate and to give practical meaning to some of the general results and corresponding catalogues of assumptions. Since this model is also a focal point of the paper we specify in the following the model in a complete manner. The specific results on the asymptotic properties of the Moran I test statistic for this model will be summarized in Section 4.2.

⁷We note that the subsequent discussion trivially extends to the case where we allow for different spatial weighting matrices for y_n and u_n . Our specification of a common spatial weighting matrix was solely adopted for reasons of notational simplicity. As a point of interest we note that this specification is typically assumed in practice.

The above spatial ARAR(1,1) model has been applied widely in the literature. It is an extension of the spatial model put forth by Cliff and Ord (1973, 1981), and a variant of the model considered by Whittle (1954). One approach of estimating a spatial ARAR(1,1) model is by maximum likelihood. However the computation of the maximum likelihood estimator may be difficult even for moderate sample sizes. In a recent paper Kelejian and Prucha (1998) introduced an instrumental variable estimator for the above model that is simple to compute and demonstrated its asymptotic normality under a set of “low level” conditions.⁸ This estimator is based on a generalized moments estimator for the spatial autoregressive parameter. The latter estimator was introduced in Kelejian and Prucha (1999).

Assume the following conditions concerning the spatial model (2.1):⁹ All diagonal elements of the spatial weighting matrix M_n are zero, and the matrices $I - \lambda_0 M_n$ and $I - \rho_0 M_n$ are nonsingular. The row and column sums of the matrices M_n , $(I - \lambda_0 M_n)^{-1}$, and $(I - \rho_0 M_n)^{-1}$ are uniformly bounded in absolute value. The regressor matrices X_n have full column rank (for n large enough), and the elements of X_n are uniformly bounded in absolute value. The innovations $\{\varepsilon_{i,n} : 1 \leq i \leq n, n \geq 1\}$ are distributed identically, and for each n the innovations $\{\varepsilon_{i,n} : 1 \leq i \leq n\}$ are distributed (jointly) independently with $E(\varepsilon_{i,n}) = 0$, $E(\varepsilon_{i,n}^2) = \sigma^2$. The innovations are furthermore assumed to have finite $(4 + \eta)$ -th moments (for some $\eta > 0$). Now consider the following instrumental variable estimator for $\theta_0 = (\lambda_0, \beta'_0)'$:

$$\hat{\theta}_n = (\hat{\lambda}_n, \hat{\beta}'_n)' = (\hat{D}'_n \hat{D}_n)^{-1} \hat{D}'_n y_n \quad (2.2)$$

with $\hat{D}_n = H_n(H'_n H_n)^{-1} H'_n D_n$ and $D_n = [M_n y_n, X_n]$, and where H_n denotes some $n \times p$ matrix of instruments ($p \geq k+1$). This estimator was considered in Kelejian and Prucha (1998) at the first stage of their two stage procedure. In the following we assume that all additional regularity conditions maintained in that paper also hold here, including that the elements of H_n are uniformly bounded in absolute value. It then follows from that paper that

$$n^{1/2}(\hat{\theta}_n - \theta_0) = P_n[n^{-1/2} F'_n \varepsilon_n] + o_p(1) \quad (2.3)$$

⁸The model in (2.1) cannot be estimated consistently by ordinary least squares since, in general, the disturbances are correlated with the spatial lag variable $M_n y_n$. Lee (1999b) demonstrates, however, that for an interesting subclass of models, corresponding to additional assumptions on the weighting matrix, ordinary least squares can be consistent.

⁹For a discussion of these assumptions see, e.g., Kelejian and Prucha (1998, 1999).

where $P_n = (n^{-1}\hat{D}'_n\hat{D}_n)^{-1}n^{-1}D'_nH_n(n^{-1}H'_nH_n)^{-1}$ with $p\lim_{n\rightarrow\infty}P_n = P$ a finite matrix, and $F'_n = H'_n(I - \rho_0M_n)^{-1}$ with $n^{-1/2}F'_n\varepsilon_n \xrightarrow{D} N(0, \sigma^2\Phi_F)$ where $\Phi_F = \lim_{n\rightarrow\infty}n^{-1}F'_nF_n$.

Let $\hat{u}_n = y_n - D_n\hat{\theta}_n$ denote the vector of estimated disturbances based on (2.1). Then the Moran I test statistic based on these estimated disturbances and on some spatial weighting matrix W_n is given by

$$I_n = \frac{Q_n^*}{\tilde{\sigma}_{Q_n^*}}, \quad Q_n^* = \hat{u}'_n W_n \hat{u}_n \quad (2.4)$$

where $\tilde{\sigma}_{Q_n^*}$ is some normalizing factor. We allow for $W_n \neq M_n$ to allow for misspecification of the spatial weighting matrix by the researcher, but otherwise postulate that W_n satisfies the same assumptions as M_n .

Given the maintained assumptions we demonstrate in the appendix that

$$\begin{aligned} n^{-1/2}\hat{u}'_n W_n \hat{u}_n &= n^{-1/2}u'_n W_n u_n \\ &\quad - n^{-1}u'_n(W_n + W'_n)D_n n^{1/2}(\hat{\theta}_n - \theta_0) + o_p(1). \end{aligned} \quad (2.5)$$

We also show in the appendix that $n^{-1}u'_n(W_n + W'_n)D_n - d'_n = o_p(1)$, and that the elements of $d'_n = E[n^{-1}u'_n(W_n + W'_n)D_n]$ are uniformly bounded in absolute value. Consequently we have from (2.3), observing that $P_n = P + o_p(1)$ and $n^{-1/2}F'_n\varepsilon_n = O_p(1)$:

$$\begin{aligned} &n^{-1}u'_n(W_n + W'_n)D_n n^{1/2}(\hat{\theta}_n - \theta_0) \\ &= n^{-1}u'_n(W_n + W'_n)D_n [n^{-1/2}P_n F'_n \varepsilon_n + o_p(1)] \\ &= n^{-1/2}b'_n \varepsilon_n + o_p(1) \end{aligned} \quad (2.6)$$

with $b'_n = -d'_n P F'_n$. Note that P is the probability limit of P_n and thus b_n is nonstochastic. Since, from (2.1), $u_n = (I - \rho_0 M_n)^{-1} \varepsilon_n$ it follows that $u'_n W_n u_n = \varepsilon'_n A_n \varepsilon_n$ where $A_n = \frac{1}{2}(I_n - \rho_0 M'_n)^{-1}(W_n + W'_n)(I_n - \rho_0 M_n)^{-1}$. It then follows from (2.5) and (2.6) that

$$n^{-1/2}\hat{u}'_n W_n \hat{u}_n = n^{-1/2}[\varepsilon'_n A_n \varepsilon_n + b'_n \varepsilon_n] + o_p(1). \quad (2.7)$$

Thus, the large sample distribution of Moran's I statistic based on estimated disturbances involves the large sample distribution of a linear-quadratic form in innovations. In the appendix we also show that the elements of the $n \times 1$

vectors b_n are uniformly bounded in absolute value, and that row and column sums of the $n \times n$ matrices A_n are uniformly bounded in absolute value.

Some comments are in order. First, as a preview, under the null hypothesis of zero spatial autocorrelation $\rho_0 = 0$, and so $A_n = \frac{1}{2}(W_n + W_n')$, which has zeroes as its diagonal elements. If $\rho_0 \neq 0$, the diagonal elements of A_n will generally not be zero. It will become evident that this distinction concerning the diagonal elements of A_n plays a crucial role in determining the power of our proposed test. Second, the linear term in (2.6) is a consequence of using estimated disturbances and the presence of a spatial lag in y . In the absence of a spatial lag in y , i.e., $\lambda_0 = 0$, we have $b_n = 0$, since $d_n = 0$. Third, we note that the elements of $(I_n - \rho_0 M_n)^{-1}$ and hence those of A_n , will in general depend on n , even if the elements of W_n and M_n do not depend on the sample size. As a consequence we allow for triangular arrays in the central limit theorem for linear-quadratic forms given below. We also allow for the innovations to be heteroskedastic, so that the theorem can be applied for testing purposes in a wide variety of limited dependent variable models, as was remarked above.

3 Some Useful Large Sample Results for Linear-Quadratic Forms

3.1 Central Limit Theorem for Linear-Quadratic Forms

Consider the following linear-quadratic form:

$$\begin{aligned} Q_n &= \varepsilon_n' A_n \varepsilon_n + b_n' \varepsilon_n \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{ij,n} \varepsilon_{i,n} \varepsilon_{j,n} + \sum_{i=1}^n b_{i,n} \varepsilon_{i,n}, \end{aligned} \tag{3.1}$$

where the $\varepsilon_{i,n}$ are (real valued) random variables, and the $a_{ij,n}$ and $b_{i,n}$ denote the (real valued) coefficients of the quadratic and linear form. We will make use of the following set of assumptions:

Assumption 1 *The real valued random variables of the array $\{\varepsilon_{i,n} : 1 \leq i \leq n, n \geq 1\}$ satisfy $E\varepsilon_{i,n} = 0$. Furthermore, for each $n \geq 1$ the random variables $\varepsilon_{1,n}, \dots, \varepsilon_{n,n}$ are totally independent.*

Assumption 2 The elements of the array of real numbers $\{a_{ij,n} : 1 \leq i, j \leq n, n \geq 1\}$ satisfy $a_{ij,n} = a_{ji,n}$ and $\sup_{1 \leq j \leq n, n \geq 1} \sum_{i=1}^n |a_{ij,n}| < \infty$.¹⁰ The elements of the array of real numbers $\{b_{i,n} : 1 \leq i \leq n, n \geq 1\}$ satisfy $\sup_n n^{-1} \sum_{i=1}^n |b_{i,n}|^{2+\eta_1} < \infty$ for some $\eta_1 > 0$.

Assumption 3 We assume that one of the following two conditions holds.

- (a) $\sup_{1 \leq i \leq n, n \geq 1} E |\varepsilon_{i,n}|^{2+\eta_2} < \infty$ for some $\eta_2 > 0$ and $a_{ii,n} = 0$.
- (b) $\sup_{1 \leq i \leq n, n \geq 1} E |\varepsilon_{i,n}|^{4+\eta_2} < \infty$ for some $\eta_2 > 0$ (but possibly $a_{ii,n} \neq 0$).

Given the above assumptions the mean, say μ_{Q_n} , and the variance, say $\sigma_{Q_n}^2$, of Q_n is given by:¹¹

$$\begin{aligned} \mu_{Q_n} &= \sum_{i=1}^n a_{ii,n} \sigma_{i,n}^2, \\ \sigma_{Q_n}^2 &= 4 \sum_{i=1}^n \sum_{j=1}^{i-1} a_{ij,n}^2 \sigma_{i,n}^2 \sigma_{j,n}^2 + \sum_{i=1}^n b_{i,n}^2 \sigma_{i,n}^2 \\ &\quad + \sum_{i=1}^n \left\{ a_{ii,n}^2 \left[\mu_{i,n}^{(4)} - \sigma_{i,n}^4 \right] + 2b_{i,n} a_{ii,n} \mu_{i,n}^{(3)} \right\}, \end{aligned} \tag{3.2}$$

with $\sigma_{i,n}^2 = E(\varepsilon_{i,n}^2)$ and $\mu_{i,n}^{(s)} = E(\varepsilon_{i,n}^s)$ for $s = 3, 4$. In the case where $a_{ii,n} = 0$ we have $\mu_{Q_n} = 0$, and the last two terms in the expression for $\sigma_{Q_n}^2$ are zero.¹² We now have the following central limit theorem for linear-quadratic forms:

Theorem 1 : Suppose Assumptions 1-3 hold and $n^{-1} \sigma_{Q_n}^2 \geq c$ for some $c > 0$. Then

$$\frac{Q_n - \mu_{Q_n}}{\sigma_{Q_n}} \xrightarrow{D} N(0, 1). \tag{3.3}$$

¹⁰The assumption of symmetry of the elements of A_n is maintained w.l.o.g. since $\varepsilon_n' A_n \varepsilon_n = \varepsilon_n' [(A_n + A_n')/2] \varepsilon_n$.

¹¹For a derivation of the subsequent expressions, compare, e.g., the proof of Theorem 1 below.

¹²We note that the variance of Q_n can also be expressed as $\sigma_{Q_n}^2 = 2 \sum_{i=1}^n \sum_{j=1}^n a_{ij,n}^2 \sigma_{i,n}^2 \sigma_{j,n}^2 + \sum_{i=1}^n b_{i,n}^2 \sigma_{i,n}^2 + \sum_{i=1}^n \left\{ a_{ii,n}^2 \left[\mu_{i,n}^{(4)} - 3\sigma_{i,n}^4 \right] + 2b_{i,n} a_{ii,n} \mu_{i,n}^{(3)} \right\}$. Under normality the last two terms are zero even if $a_{ii,n} \neq 0$. Given the last two terms are zero we can also write $\sigma_{Q_n}^2 = 2 \text{tr}(A_n' \Sigma_n A_n \Sigma_n) + b_n' \Sigma_n b_n$ with $\Sigma_n = \text{diag}(\sigma_{i,n}^2)$.

Note that the theorem allows the elements of A_n to depend on n , and does not restrict its diagonal elements to be zero. Furthermore, the theorem allows the innovations $\varepsilon_{i,n}$ to be heteroskedastic. As will be demonstrated in Section 4, this permits the derivation of the limiting distribution of the Moran I test statistic in a variety of limited dependent variable models. Finally, we note that Assumptions 1-3 are such that they do not depend upon a particular ordering of the data. This is important in a spatial setting because, unlike for time series, there is typically no natural ordering of spatial data.¹³

Existing results on the limiting distribution of quadratic forms include publications by Whittle (1964), Beran (1972), Sen (1976) and Giraitis and Taqqu (1998). In contrasting those results with the above theorem we note that the papers by Whittle, Beran, and Giraitis and Taqqu do not allow the weights in the quadratic form to depend on the sample size. The latter two papers actually assume that $a_{ij,n} = a_{|i-j|}$. Sen allows for the weights to depend on the sample size n , but his proof refers to a central limit theorem for martingale difference sequence. His proof is hence not completely formal (unless we assume additionally that weights do not depend on n .) Beran and Sen assume $a_{ii,n} = 0$, and also assume that the $\varepsilon_{i,n} = \varepsilon_i$ are i.i.d. Giraitis and Taqqu assume that the $\varepsilon_{i,n} = \varepsilon_i$ are products of linear process with i.i.d. innovations. The recent central limit theorem given in Pinkse (1999) also postulates that the random processes are i.i.d.

3.2 Asymptotic Distribution of Quadratic Forms Based on Estimated Disturbances

In the following we develop results on the limiting behavior of quadratic forms based on estimated disturbances, utilizing the above derived central limit theorem for linear-quadratic forms. The underlying model specification will be kept fairly general to cover, in particular, the various limited dependent variable models considered in Section 4 as well as the spatial ARAR(1,1) model. More specifically, for the subsequent discussion we consider the following (possibly nonlinear) model:

$$g_{i,n}(z_n, \theta_0) = u_{i,n}, \quad i = 1, \dots, n, \quad (3.4)$$

¹³We note that Theorem 1 also holds if the sample size is taken to be k_n rather than n (with $k_n \uparrow \infty$ as $n \rightarrow \infty$).

with $z_n = (z_{1,n}, \dots, z_{n,n})$ and where $z_{i,n}$ denotes the $1 \times p_z$ vector of endogenous and exogenous variables corresponding to the i -th unit, $u_{i,n}$ denotes the disturbance term corresponding to the i -th unit, and θ_0 denotes the $p_\theta \times 1$ true parameter vector.¹⁴ We emphasize that in allowing for the $g_{i,n}$ to depend on $z_{1,n}, \dots, z_{n,n}$ the specification allows for spatial lags. In the following we will write (3.4) more compactly as $g_n(z_n, \theta) = u_n$ with $g_n(z_n, \theta) = (g_{1,n}(z_n, \theta), \dots, g_{n,n}(z_n, \theta))'$ and $u_n = (u_{1,n}, \dots, u_{n,n})$.

The Moran I test statistic corresponding to model (3.4) would be based on the quadratic form $\hat{u}_n' W_n \hat{u}_n$, where $\hat{u}_n = g_n(z_n, \hat{\theta}_n)$ denotes the vector of estimated residuals based on some estimator $\hat{\theta}_n$. We maintain the following assumptions:

Assumption 4 *The elements of the row and column sums of the real $n \times n$ matrices $W_n = (w_{ij,n})$ are uniformly bounded in absolute value.*¹⁵

Assumption 5 *The true parameter vector θ_0 is an interior element of $\Theta \subseteq \mathbb{R}^{p_\theta}$.*

Assumption 6

(a) *The random variables of the array $\{z_{i,n} : 1 \leq i \leq n, n \geq 1\}$ take their values in $Z \subseteq \mathbb{R}^{p_z}$. The functions*

$$g_{i,n} : Z^n \times \Theta_* \rightarrow \mathbb{R}, \quad 1 \leq i \leq n, n \geq 1,$$

are Borel measurable for each $\theta \in \Theta_$, where $\Theta \subseteq \Theta_*$ and Θ_* is an open subset. Furthermore the functions $g_{i,n}(\underline{z}_1, \dots, \underline{z}_n, \cdot)$ are twice continuously partially differentiable on Θ_* for each $(\underline{z}_1, \dots, \underline{z}_n) \in Z^n$.*

(b) *If $g_{i,n}(z_n, \theta)$ is linear in θ , let*

$$\Delta_{i,n}(z_n) = \left| \frac{\partial g_{i,n}(z_n, \theta)}{\partial \theta} \right|,$$

otherwise let

$$\Delta_{i,n}(z_n) = \sup \left\{ |g_{i,n}(z_n, \theta)|, \left| \frac{\partial g_{i,n}(z_n, \theta)}{\partial \theta} \right|, \left| \frac{\partial^2 g_{i,n}(z_n, \theta)}{\partial \theta \partial \theta'} \right| : \theta \in \Theta \right\};$$

then $\sup_{1 \leq i \leq n, n \geq 1} E(\Delta_{i,n}(z_n))^2 < \infty$.

¹⁴For the spatial ARAR(1,1) model we have $z_{i,n} = (y_{i,n}, x_{i1,n}, \dots, x_{iK,n})$ and $\theta_0 = (\lambda_0, \beta_0')'$.

¹⁵We note that in this section we do not assume that the diagonal elements of W_n are zero.

We now have the following lemma concerning the asymptotic relationship between $n^{-1/2}\hat{u}'_n W_n \hat{u}_n$ and $n^{-1/2}u'_n W_n u_n$:

Lemma 1 : Suppose Assumptions 4, 5 and 6 hold. Let $\hat{\theta}_n$ be some estimator for θ_0 with $n^{1/2}(\hat{\theta}_n - \theta_0) = O_p(1)$. Then (abusing notation in an obvious way)

$$\begin{aligned} n^{-1/2}\hat{u}'_n W_n \hat{u}_n &= n^{-1/2}u'_n W_n u_n \\ &\quad + n^{-1}u'_n(W_n + W'_n)\frac{\partial g_n(z_n, \theta_0)}{\partial \theta}n^{1/2}(\hat{\theta}_n - \theta_0) + o_p(1). \end{aligned} \quad (3.5)$$

Furthermore

$$n^{-1}\hat{u}'_n(W_n + W'_n)\frac{\partial g_n(z_n, \hat{\theta}_n)}{\partial \theta} - n^{-1}u'_n(W_n + W'_n)\frac{\partial g_n(z_n, \theta_0)}{\partial \theta} = o_p(1). \quad (3.6)$$

The above lemma shows that in general the limiting distribution of $n^{-1/2}\hat{u}'_n W_n \hat{u}_n$ will differ from that of $n^{-1/2}u'_n W_n u_n$, unless the probability limit of the term $n^{-1}u'_n(W_n + W'_n)[\partial g_n(z_n, \theta_0)/\partial \theta]$ is zero. We note that this limit will (given typical assumptions) be zero if $\partial g_n(z_n, \theta_0)/\partial \theta$ only depends on exogenous variables. This limit will also typically be zero if $w_{ii,n} = 0$, $Eu_{i,n} = 0$ and $(u_{i,n}, \partial g_{i,n}(z_n, \theta_0)/\partial \theta)$ is distributed independently over i . For the spatial ARAR(1,1) model we have $n^{-1}u'_n(W_n + W'_n)[\partial g_n(z_n, \theta_0)/\partial \theta] = -n^{-1}u'_n(W_n + W'_n)D_n$, and equation (2.5) is seen to be a special case of (3.5). We note that for this model the probability limit of this term is non-zero, due to the presence of a spatial lag in the dependent variable. The last part of the lemma shows how the term can be estimated consistently.

In many cases u_n will be linear in ε_n , where ε_n is the basic vector of innovations. For example, for the spatial ARAR(1,1) model, $u_n = (I_n - \rho_0 M_n)^{-1}\varepsilon_n$. Furthermore in many cases, and again as illustrated by the spatial ARAR(1,1) model, $n^{1/2}(\hat{\theta}_n - \theta_0)$ will be (asymptotically) linear in ε_n . This motivates the following assumption.

Assumption 7 There exists a linear-quadratic form $Q_n = \varepsilon'_n A_n \varepsilon_n + b'_n \varepsilon_n$ such that

$$\begin{aligned} &n^{-1/2}u'_n W_n u_n + n^{-1}u'_n(W_n + W'_n)\frac{\partial g_n(z_n, \theta_0)}{\partial \theta}n^{1/2}(\hat{\theta}_n - \theta_0) \\ &= n^{-1/2}Q_n + o_p(1), \end{aligned} \quad (3.7)$$

and where the elements of ε_n , A_n and b_n satisfy Assumptions 1-3.

For the spatial ARAR(1,1) model the above assumption is clearly satisfied in light of (2.6). If the probability limit of the term $n^{-1}u'_n(W_n + W'_n)[\partial g_n(z_n, \theta_0)/\partial \theta]$ is zero, then we can take $b_n = 0$. We note that for the spatial ARAR(1,1) model $b_n = 0$ in the absence of a spatial lag of y_n (since in that case $d_n = 0$).

Given the above assumption and given the assumptions of Lemma 1 hold, it follows from that lemma that $n^{-1/2}\hat{u}'_n W_n \hat{u}_n = n^{-1/2}Q_n + o_p(1)$, and consequently the limiting behavior of $n^{-1/2}\hat{u}'_n W_n \hat{u}_n$ is identical with that of $n^{-1/2}Q_n$. For the spatial ARAR(1,1) model this was established in (2.7).

As before, let μ_{Q_n} and $\sigma_{Q_n}^2$ denote the mean and variance of the linear-quadratic form Q_n , and recall the explicit expressions for μ_{Q_n} and $\sigma_{Q_n}^2$ given in (3.2). We next give a result concerning the asymptotic behavior of $\hat{u}'_n W_n \hat{u}_n / \hat{\sigma}_{Q_n}$, where $\hat{\sigma}_{Q_n}^2$ denotes some estimator for $\sigma_{Q_n}^2$. We note that in particular applications $\hat{\sigma}_{Q_n}^2$ may have been chosen such that the estimator is, e.g., consistent under the null hypothesis of zero spatial autocorrelation, but inconsistent under the alternative hypothesis. To cover such situations we allow for the estimator $\hat{\sigma}_{Q_n}^2$ to be (possibly) inconsistent.

Theorem 2 : Suppose Assumption 4-7 hold and $n^{-1}\sigma_{Q_n}^2 \geq c$ for some $c > 0$. Let $\bar{\sigma}_{Q_n}^2$ be a sequence of real numbers such that $n^{-1}\hat{\sigma}_{Q_n}^2 - n^{-1}\bar{\sigma}_{Q_n}^2 = o_p(1)$ and $\lim_{n \rightarrow \infty} \sigma_{Q_n}^2 / \bar{\sigma}_{Q_n}^2 = \Lambda^2 > 0$. Then

$$\frac{\hat{u}'_n W_n \hat{u}_n}{\hat{\sigma}_{Q_n}} = \Lambda \left(\frac{Q_n - \mu_{Q_n}}{\sigma_{Q_n}} \right) + \frac{\mu_{Q_n}}{\bar{\sigma}_{Q_n}} + o_p(1). \quad (3.8)$$

Consequently:

(a) If $\mu_{Q_n} = 0$ (which is the case if $a_{ii,n} = 0$), then

$$\frac{\hat{u}'_n W_n \hat{u}_n}{\hat{\sigma}_{Q_n}} \xrightarrow{D} N(0, \Lambda^2).$$

(b) If $|n^{-1/2}\mu_{Q_n}| \rightarrow \infty$ as $n \rightarrow \infty$, then for every constant $\delta > 0$

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{\hat{u}'_n W_n \hat{u}_n}{\hat{\sigma}_{Q_n}} \right| > \delta \right) = 1.$$

We note that for the Moran I test statistic we have $A_n = \frac{1}{2}(W_n + W'_n)$, and hence $a_{ii,n} = 0$, under the null hypothesis of zero spatial autocorrelation. Result (a) in the above theorem thus provides a general result for the limiting distribution of the Moran I test statistic under the null hypothesis. Clearly, if $\hat{\sigma}_{Q_n}^2$ is a consistent estimator (in the sense that $n^{-1}\hat{\sigma}_{Q_n}^2 - n^{-1}\sigma_{Q_n}^2 = o_p(1)$), then $\Lambda = 1$. Under the alternative hypothesis we will typically have $a_{ii,n} \neq 0$ - e.g., for the spatial ARAR(1,1) model $A = \frac{1}{2}(I - \rho_0 M')^{-1}(W + W'_n)(I - \rho_0 M)^{-1}$. Result (b) in the above theorem provides a general result for the consistency of the Moran I test statistic, i.e., for the power of the test to approach unity as $n \rightarrow \infty$, for an arbitrary significance level of the test. The condition $|n^{-1/2}\mu_{Q_n}| \rightarrow \infty$ is, e.g., satisfied if $0 < \text{const} \leq \sigma_{i,n}^2$ and $0 < \text{const} \leq a_{ii,n}$.

3.3 Estimation of the Variance of the Linear-Quadratic Form

In Theorem 2 we employed an estimator for the variance $\sigma_{Q_n}^2$ of the linear-quadratic form $Q_n = \varepsilon'_n A_n \varepsilon_n + b'_n \varepsilon_n$. In many situations, including the case of the spatial ARAR(1,1) model and the models considered in the next section, we have $A_n = (W_n + W'_n)/2$ under the null hypothesis of zero spatial autocorrelation. Let $\hat{\sigma}_{i,n}^2$ and $\hat{\mu}_{i,n}^{(s)}$ denote estimators for $\sigma_{i,n}^2$ and $\mu_{i,n}^{(s)}$ with $s = 3, 4$, and let $\hat{b}_{i,n}$ denote estimators for the $b_{i,n}$ (since in practice $b_{i,n}$ will typically be unobserved, unless $b_{i,n} = 0$). Recalling the expression for $\sigma_{Q_n}^2$ given in (3.2), it is natural to consider the following estimator:

$$\begin{aligned} \hat{\sigma}_{Q_n}^2 &= \sum_{i=1}^n \sum_{j=1}^{i-1} (w_{ij,n} + w_{ji,n})^2 \hat{\sigma}_{i,n}^2 \hat{\sigma}_{j,n}^2 + \sum_{i=1}^n \hat{b}_{i,n}^2 \hat{\sigma}_{i,n}^2 \\ &\quad + \sum_{i=1}^n \left\{ w_{ii,n}^2 \left[\hat{\mu}_{i,n}^{(4)} - \hat{\sigma}_{i,n}^4 \right] + 2\hat{b}_{i,n} w_{ii,n} \hat{\mu}_{i,n}^{(3)} \right\}. \end{aligned} \quad (3.9)$$

Of course, in many contexts we will have $w_{ii,n} = 0$, but we do not postulate this here for reasons of generality.¹⁶

¹⁶We note that if $w_{ii,n} = 0$ the third and fourth moments of the innovations (and estimators thereof) drop from the respective expressions considered in this section, and then any assumptions maintained with respect to those moments become mute.

To accommodate the analysis of the test statistic both under the null and under the alternative hypothesis we allow for the estimators $\hat{\sigma}_{i,n}^2$, $\hat{\mu}_{i,n}^{(s)}$ ($s = 3, 4$) and $\hat{b}_{i,n}$ to be possibly inconsistent. In the following assumption, let $\bar{\sigma}_{i,n}^2$, $\bar{\mu}_{i,n}^{(s)}$ ($s = 3, 4$) and $\bar{b}_{i,n}$ denote nonstochastic sequences that the assumption requires to exist. These could be the expected values of the estimators, or alternatively, if the estimators are defined as functions of, say $\hat{\theta}_n$, then the nonstochastic sequences may correspond to the value of those functions at the expected value or probability limit of $\hat{\theta}_n$; for further interpretations see also the discussion after Theorem 3 below. Also let

$$\begin{aligned} \bar{\sigma}_{Q_n}^2 = & \sum_{i=1}^n \sum_{j=1}^{i-1} (w_{ij,n} + w_{ji,n})^2 \bar{\sigma}_{i,n}^2 \bar{\sigma}_{j,n}^2 + \sum_{i=1}^n \bar{b}_{i,n}^2 \bar{\sigma}_{i,n}^2 \\ & + \sum_{i=1}^n \left\{ w_{ii,n}^2 \left[\bar{\mu}_{i,n}^{(4)} - \bar{\sigma}_{i,n}^4 \right] + 2\bar{b}_{i,n} w_{ii,n} \bar{\mu}_{i,n}^{(3)} \right\}, \end{aligned} \quad (3.10)$$

denote the corresponding nonstochastic analogue of $\hat{\sigma}_{Q_n}^2$. We maintain the following assumption.

Assumption 8 •

(a) The estimators $\hat{\sigma}_{i,n}^2$ and $\hat{\mu}_{i,n}^{(s)}$ satisfy

$$|\hat{\sigma}_{i,n}^2 - \bar{\sigma}_{i,n}^2| \leq \phi_n \text{ and } |\hat{\mu}_{i,n}^{(s)} - \bar{\mu}_{i,n}^{(s)}| \leq \phi_n, \quad (3.11)$$

where $\phi_n = o_p(1)$, and $\sup_{1 \leq i \leq n, n \geq 1} \bar{\sigma}_{i,n}^2 < \infty$ and $\sup_{1 \leq i \leq n, n \geq 1} |\bar{\mu}_{i,n}^{(s)}| < \infty$ ($s = 3, 4$).

(b) The estimators $\hat{b}_{i,n}$ satisfy

$$|\hat{b}_{i,n} - \bar{b}_{i,n}| \leq \Psi_{i,n} \psi_n, \quad (3.12)$$

where $\psi_n = o_p(1)$, the $\Psi_{i,n} \in \mathbb{R}_+$ satisfy $\sup_n n^{-1} \sum_{i=1}^n E \Psi_{i,n}^2 < \infty$ and $\sup_n n^{-1} \sum_{i=1}^n |\bar{b}_{i,n}|^2 < \infty$.

We now have the following theorem concerning the stochastic convergence of $\hat{\sigma}_{Q_n}^2$.

Theorem 3 : Suppose Assumptions 4 and 8 hold. Then $n^{-1}(\hat{\sigma}_{Q_n}^2 - \bar{\sigma}_{Q_n}^2) = o_p(1)$ with $n^{-1}\bar{\sigma}_{Q_n}^2 \leq \text{const} < \infty$.

Of course, if $(w_{ij,n} + w_{ji,n})/2 = a_{ij,n}$ and $\bar{b}_{i,n} = b_{i,n}$, and if $\bar{\sigma}_{i,n}^2 = \sigma_{i,n}^2$, $\bar{\mu}_{i,n}^{(s)} = \mu_{i,n}^{(s)}$ ($s = 3, 4$), then $\bar{\sigma}_{Q_n}^2 = \sigma_{Q_n}^2$ and the above theorem establishes the consistency of $\hat{\sigma}_{Q_n}^2$ for $\sigma_{Q_n}^2$. We note that in this leading case the boundedness conditions maintained in Assumption 8 concerning $\bar{b}_{i,n}$ and $\bar{\sigma}_{i,n}^2, \bar{\mu}_{i,n}^{(s)}$ ($s = 3, 4$) are then automatically satisfied under Assumptions 2 and 3.

For a further interpretation of the assumptions of the theorem consider again the above leading case. We note that in this case condition (3.11) maintained in Assumption 8(a) can, e.g., be implied – as is easily verified – from either one of the following conditions.

Condition A: $\sigma_{i,n}^2 = h_{i,n}(\theta_0)$ where the functions $h_{i,n} : \Theta_* \rightarrow \mathbb{R}_+$ are continuously partially differentiable,

$$\sup_{1 \leq i \leq n, n \geq 1} \sup_{\theta \in \Theta} \left| \frac{\partial h_{i,n}(\theta)}{\partial \theta} \right| < \infty,$$

and $\hat{\sigma}_{i,n}^2 = h_{i,n}(\hat{\theta}_n)$ with $n^{1/2}(\hat{\theta}_n - \theta_0) = O_p(1)$. (If $w_{ii,n} \neq 0$ analogous conditions are also assumed to hold for the estimators of $\mu_{i,n}^{(s)}$, $s = 3, 4$.)

Condition B: Assumptions 1, 3, 5 and 6 hold. Furthermore $\sigma_{i,n}^2 = \sigma^2$ with $g_{i,n}(z_n, \theta_0) = u_{i,n} = \varepsilon_{i,n}$, and $\hat{\sigma}_{i,n}^2 = \hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n \hat{u}_{i,n}^2$ with $g_{i,n}(z_n, \hat{\theta}_n) = \hat{u}_{i,n}$ and $n^{1/2}(\hat{\theta}_n - \theta_0) = O_p(1)$. (If $w_{ii,n} \neq 0$ analogous conditions are also assumed to hold for the estimators of $\mu_{i,n}^{(s)}$, $s = 3, 4$.)

For an interpretation of Assumption 8(b) consider again the spatial ARAR(1,1). From the discussion surrounding (2.5) - (2.7) we see that in this case we have $b_n = -F_n P' d_n$ with $d'_n = E(n^{-1} u'_n (W_n + W_n) D_n)$ and $F_n = (I_n - \rho_0 M'_n)^{-1} H_n$. (Recall that P is the probability limit of P_n , and hence b_n is nonstochastic.) Let $\bar{b}_n = -H_n P' d_n$. Then $\bar{b}_n = b_n$ under the null hypothesis of zero spatial correlation. It seems natural to estimate \bar{b}_n by $\hat{b}_n = -H_n P'_n \hat{d}_n$ with $\hat{d}'_n = n^{-1} \hat{u}'_n (W_n + W_n) D_n$. Let $\hat{\gamma}_n = P'_n \hat{d}_n$ and $\gamma_n = P' d_n$, then $\hat{b}_{i,n} - \bar{b}_{i,n} = -\sum_{r=1}^p h_{ir,n}(\hat{\gamma}_{r,n} - \gamma_{r,n})$. Taking $\Psi_{i,n} = \sum_{r=1}^p |h_{ir,n}|$ and $\psi_n =$

$\sum_{r=1}^p |\hat{\gamma}_{r,n} - \gamma_{r,n}|$ we have $|\hat{b}_{i,n} - \bar{b}_{i,n}| \leq \Psi_{i,n} \psi_n$. Since the $h_{ir,n}$ are uniformly bounded in absolute value it follows that $\sup_n n^{-1} \sum_{i=1}^n \Psi_{i,n}^2 < \infty$. As a by-product of the proofs of the claims made in the discussion surrounding (2.5) - (2.7) we have $\hat{d}_n - d_n = o_p(1)$. Consequently $\hat{\gamma}_n - \gamma_n = o_p(1)$ and $\psi_n = o_p(1)$, and thus Assumption 8(b) holds for this setting.

4 Testing for Spatial Autocorrelation: Limited Dependent Variable Models and the ARAR(1,1) Model

In the following we utilize the above general results to formally establish the limiting behavior of the Moran I test statistic for spatial correlation for a variety of limited dependent variable models, and for the spatial ARAR(1,1) model. In particular, we will establish for each of these models the following asymptotic normality result for the Moran I test statistic under the null hypothesis of zero spatial autocorrelation:

$$I_n = \frac{Q_n^*}{\tilde{\sigma}_{Q_n^*}} \xrightarrow{D} N(0, 1) \quad (4.1)$$

where $Q_n^* = \hat{u}_n' W_n \hat{u}_n$ and $\tilde{\sigma}_{Q_n^*}$ is a normalization factor. The specification of the vector of estimated disturbances \hat{u}_n and the normalization factor $\tilde{\sigma}_{Q_n^*}$ will be model specific. Recall that for the spatial ARAR(1,1) model we have $u_n = (I_n - \rho_0 M_n)^{-1} \varepsilon_n$ and thus $u_n = \varepsilon_n$ under the null hypothesis $\rho_0 = 0$. More generally we may think of the disturbance process to be of the form $u_n = \Xi_n \varepsilon_n$, where Ξ_n is a $n \times n$ nonstochastic matrix. Under the null hypothesis of zero spatial correlation we have $\Xi_n = I_n$ and $u_n = \varepsilon_n$. In the following we will hence also write $Q_n^* = \hat{\varepsilon}_n' W_n \hat{\varepsilon}_n$ under the null hypothesis of zero spatial autocorrelation.

Throughout this section we will assume that the spatial weighting matrix W_n satisfies Assumption 4 and that all of its diagonal elements are zero. We also assume that for some constant c_w and for all $n > 1$:

$$0 < c_w \leq n^{-1} \sum_{i=1}^n \sum_{j=1}^n (w_{ij,n} + w_{ji,n})^2. \quad (4.2)$$

Clearly, a sufficient condition for (4.2) is that the elements of W_n are non-negative, each row has at least one nonzero element, and the nonzero elements

of W_n are uniformly bounded away from zero. In the following we will also assume that all non-stochastic regressors are uniformly bounded in absolute value. Furthermore, we maintain Assumption 5 concerning the parameter space Θ and the true parameter vector θ_0 . We also assume that the parameter space for the regression parameters is a subset of a compact set, except in the case of the ARAR(1,1) model.¹⁷

As will become clear, for the limited dependent variable models specified below the innovations entering the Moran I test statistic will typically be heteroskedastic. As was discussed in the Introduction, results available in the literature concerning the limiting distribution of the Moran I test statistic assume that the innovations are homoskedastic, and hence do not cover these models.¹⁸

4.1 Limited Dependent Variable Models

For each of the limited dependent variable models considered below the normalization factor in the Moran I test statistic will be of the following general form:

$$\begin{aligned}\tilde{\sigma}_{Q_n^*}^2 &= (1/2) \sum_{i=1}^n \sum_{j=1}^n (w_{ij,n} + w_{ji,n})^2 \hat{\sigma}_{i,n}^2 \hat{\sigma}_{j,n}^2 \\ &= \text{tr} \left(W_n \hat{\Sigma}_n W_n \hat{\Sigma}_n + W_n' \hat{\Sigma}_n W_n \hat{\Sigma}_n \right)\end{aligned}\tag{4.3}$$

where $\hat{\Sigma}_n = \text{diag}(\hat{\sigma}_{i,n}^2)$. The variance estimators $\hat{\sigma}_{i,n}^2$ will be specified explicitly for each of the models considered.

¹⁷We note that for nonlinear estimation problems the parameter space is typically taken to be compact; cp., e.g., Pötscher and Prucha (1997).

¹⁸Anselin and Kelejian (1997) give a result concerning the large sample distribution of the Moran I test statistic for a spatial ARAR(1,1) model. However, their result is based on a central limit theorem for m -dependent variables and the assumption that the spatial weighting matrix is band diagonal. Clearly, this assumption is restrictive and mostly covers situations where the data expand “bandwise” in one direction, and is thus in essence geared more towards a time series context.

4.1.1 The Tobit Model

Consider the Tobit model ($i = 1, \dots, n$)

$$\begin{aligned} y_i^* &= x_i \beta_0 + \nu_i, \\ y_i &= \begin{cases} y_i^* & \text{if } y_i^* \geq 0 \\ 0 & \text{otherwise} \end{cases}, \end{aligned} \quad (4.4)$$

where y_i^* is a latent variable, y_i is the observed dependent variable, x_i is a non-stochastic $1 \times k$ vector of regressors, β_0 is a $k \times 1$ vector of regression parameters, and the ν_i 's are i.i.d. $N(0, \sigma_{v0}^2)$ where σ_{v0}^2 is bounded by some constants c_σ and C_σ : $0 < c_\sigma < \sigma_{v0} < C_\sigma < \infty$. The model is furthermore assumed to satisfy sufficient regularity conditions such that the maximum likelihood estimators of $\theta_0 = (\beta_0', \sigma_{v0}^2)'$, say $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_{v0}^2)'$, is $n^{1/2}$ -consistent.¹⁹

Let $\Phi_N(\cdot)$ and $\phi_N(\cdot)$ denote the c.d.f. and p.d.f. of the standardized normal distribution. Then, results given in Amemiya (1985, pp. 370-71) imply

$$y_i = f(x_i, \theta_0) + \varepsilon_i \quad (4.5)$$

with

$$f(x_i, \theta_0) = \sigma_{v0} \Phi_N \left(\frac{x_i \beta_0}{\sigma_{v0}} \right) \left[\frac{x_i \beta_0}{\sigma_{v0}} + \frac{\phi_N(x_i \beta_0 / \sigma_{v0})}{\Phi_N(x_i \beta_0 / \sigma_{v0})} \right], \quad (4.6)$$

and where $E(\varepsilon_i) = 0$ and $\sigma_i^2 = E(\varepsilon_i^2) = h_i(\theta_0)$ with

$$\begin{aligned} h_i(\theta_0) &= \sigma_{v0}^2 \Phi_N \left(\frac{x_i \beta_0}{\sigma_{v0}} \right) \left[\left(\frac{x_i \beta_0}{\sigma_{v0}} \right)^2 + 1 \right. \\ &\quad \left. + \frac{x_i \beta_0}{\sigma_{v0}} \frac{\phi_N(x_i \beta_0 / \sigma_{v0})}{\Phi_N(x_i \beta_0 / \sigma_{v0})} \right] - f(x_i, \theta_0)^2. \end{aligned} \quad (4.7)$$

Assume that $\sigma_i^2 = h_i(\theta_0) \geq \text{const} > 0$ and let $\hat{\varepsilon}_{i,n} = y_i - f(x_i, \hat{\theta}_n)$ and $\hat{\sigma}_{i,n}^2 = h_i(\hat{\theta}_n)$. In the appendix we demonstrate that under the above assumptions for the Tobit model the Moran I test statistic converges in distribution to a standardized normal, i.e., (4.1) holds. The result in (4.1) can now be used in the usual way to test for the presence of spatial autocorrelation.

¹⁹ Actually, for the subsequent result $\hat{\theta}_n$ could be any $n^{1/2}$ -consistent estimators.

4.1.2 A Dichotomous Model

Consider the model ($i = 1, \dots, n$)

$$\begin{aligned} y_i^* &= x_i\beta_0 + \nu_i, \\ y_i &= \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{otherwise} \end{cases}, \end{aligned} \quad (4.8)$$

where y_i^* is a latent variable and y_i is the observed binary dependent variable. The vectors x_i and β_0 are as in Section 4.1.1, but ν_i is only assumed to be i.i.d. with zero mean and c.d.f. $\Phi(\cdot)$. (Note that we are distinguishing between Φ_N defined above and Φ .) We assume furthermore that the distribution has continuous second order derivatives and is symmetric around zero²⁰, and thus $\Pr(y_i = 1) = \Phi(x_i\beta_0)$ and $\Pr(y_i = 0) = 1 - \Phi(x_i\beta_0)$. Among others, this specification covers the probit as well as the logit model. We assume again that the model satisfies sufficient regularity conditions such that the maximum likelihood estimator of β_0 , say $\hat{\beta}_n$, is $n^{1/2}$ -consistent. (Here $\theta_0 = \beta_0$ and $\hat{\theta}_n = \hat{\beta}_n$.)

The above specification clearly implies that

$$y_i = \Phi(x_i\beta_0) + \varepsilon_i \quad (4.9)$$

where $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = \Phi(x_i\beta_0)[1 - \Phi(x_i\beta_0)]$. Assume that $E(\varepsilon_i^2) \geq \text{const} > 0$ and let $\hat{\varepsilon}_i = y_i - \Phi(x_i\hat{\beta}_n)$ and $\hat{\sigma}_{i,n}^2 = \Phi(x_i\hat{\beta}_n)[1 - \Phi(x_i\hat{\beta}_n)]$. In the appendix we demonstrate that under the above assumptions the Moran I test statistic converges in distribution to a standardized normal, i.e., (4.1) holds. The result in (4.1) can now be used in the usual way to test for the presence of spatial autocorrelation.

4.1.3 A Sample Selection Model

Consider the sample selection model described in Amemiya (1985, pp. 385-387):

$$\begin{aligned} y_{1i}^* &= x_{1i}\beta_{10} + \nu_{1i} \\ y_{2i}^* &= x_{2i}\beta_{20} + \nu_{2i} \\ y_{2i} &= \begin{cases} y_{2i}^* & \text{if } y_{1i}^* > 0 \\ 0 & \text{if } y_{1i}^* \leq 0 \end{cases} \end{aligned} \quad (4.10)$$

²⁰Inspection of the appendix shows that this symmetry assumption could be dropped. The assumption is only made to highlight the standard logit and probit cases.

where y_{1i}^* and y_{2i}^* are latent variables, y_{2i} is the observed dependent variable, x_{1i} and x_{2i} are nonstochastic $1 \times k_1$ and $1 \times k_2$ vectors of regressors, β_{10} and β_{20} are $k_1 \times 1$ and $k_2 \times 1$ parameter vectors, and (ν_{1i}, ν_{2i}) is i.i.d. as $N(0, \Sigma_0)$ where $\Sigma_0 = (\sigma_{ij0})$ with $0 < c_\sigma < |\Sigma_0| < C_\sigma < \infty$. We assume again that the model satisfies sufficient regularity conditions such that the maximum likelihood estimator of $\theta_0 = (\beta'_{10}, \beta'_{20}, \sigma_{220}, \sigma_{120})'$, say $\hat{\theta}_n = (\hat{\beta}'_{1n}, \hat{\beta}'_{2n}, \hat{\sigma}_{22n}, \hat{\sigma}_{12n})'$, is $n^{1/2}$ -consistent.²¹

Results given in Amemiya imply that for those i for which $y_{2i} \neq 0$:

$$y_{2i} = f(x_{1i}, x_{2i}, \theta_0) + \varepsilon_i; \quad (4.11)$$

with

$$f(x_{1i}, x_{2i}, \theta_0) = x_{2i}\beta_{20} + \sigma_{120} \frac{\phi_N(x_{1i}\beta_{10})}{\Phi_N(x_{1i}\beta_{10})} \quad (4.12)$$

and where $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = h_i(\theta_0)$ with

$$h_i(\theta_0) = \sigma_{220} - \sigma_{120}^2 \left[x_{1i}\beta_{10} \frac{\phi_N(x_{1i}\beta_{10})}{\Phi_N(x_{1i}\beta_{10})} + \left(\frac{\phi_N(x_{1i}\beta_{10})}{\Phi_N(x_{1i}\beta_{10})} \right)^2 \right]. \quad (4.13)$$

Assume that $\sigma_i^2 = h_i(\theta_0) \geq \text{const} > 0$ and let $\hat{\varepsilon}_{i,n} = y_{2i} - f(x_{1i}, x_{2i}, \hat{\theta}_n)$ and $\hat{\sigma}_{i,n}^2 = h_i(\hat{\theta}_n)$. In the appendix we demonstrate that under the above assumptions for the sample selection model the Moran I test statistic converges in distribution to a standardized normal, i.e., (4.1) holds. The result in (4.1) can now be used in the usual way to test for the presence of spatial autocorrelation.

4.1.4 A Polychotomous Model

Consider the polychotomous model which has m categories ($i = 1, \dots, n$; $j = 1, \dots, m$):

$$\Pr(y_i = j) = P_j(x_i, \beta_0) \quad (4.14)$$

where y_i denotes the observed endogenous variable, x_i is a non-stochastic $1 \times k$ vector of exogenous variables and β_0 is a $k \times 1$ vector of parameters.

²¹As demonstrated by Amemiya (1985, p. 386) the likelihood function for this model depends on σ_{110} only through β_{110}/σ_{110} and $\sigma_{120}/\sigma_{110}$. In the absence of further information σ_{110} is not identified and we hence put $\sigma_{110} = 1$. We note that the analysis can be readily extended to the case where σ_{110} can be identified.

We assume that the functions $P_j(\cdot)$ possess bounded first and second order derivatives. We also assume that for all x_i and all vectors β in the parameter space $0 < P_j(x_i, \beta) < 1$ ($j = 1, \dots, m$). Among others, these specifications are consistent with many ordered, as well as unordered polychotomous dependent variable models which are based on either the normal or logit specifications - see, e.g., Maddala (1983, Chapter 2), and Greene (1997, Chapter 19). We assume again that the model satisfies sufficient regularity conditions such that the maximum likelihood estimator of β_0 , say $\hat{\beta}_n$, is $n^{1/2}$ -consistent. (Here $\theta_0 = \beta_0$ and $\hat{\theta}_n = \hat{\beta}_n$.)

Given the above specification we have

$$y_i = f(x_i, \beta_0) + \varepsilon_i, \quad i = 1, \dots, n \quad (4.15)$$

with

$$f(x_i, \beta_0) = \sum_{j=1}^m j P_j(x_i, \beta_0) \quad (4.16)$$

and where $E(\varepsilon_i) = 0$ and $\sigma_i^2 = E(\varepsilon_i^2) = h_i(\beta_0)$ with

$$h_i(\beta_0) = \sum_{j=1}^m j^2 P_j(x_i, \beta_0) - \left[\sum_{j=1}^m j P_j(x_i, \beta_0) \right]^2 \quad (4.17)$$

Let $\hat{\varepsilon}_{i,n} = y_i - f(x_i, \hat{\beta}_n)$ and $\hat{\sigma}_{i,n}^2 = h_i(\hat{\beta}_n)$. In the appendix we demonstrate that under the above assumptions the Moran I test statistic converges in distribution to a standardized normal, i.e., (4.1) holds. The result in (4.1) can now be used in the usual way to test for the presence of spatial autocorrelation.

4.2 Spatial ARAR(1,1) Model

Next consider again the spatial ARAR(1,1) model specified in Section 2. As remarked, although this model was discussed in Section 2 in part to motivate the theoretical developments in Section 3, this model is also of interest in and of itself because it is widely used.

We continue to maintain all assumptions postulated in Section 2. In testing for spatial correlation in the disturbances, i.e., in testing $H_0 : \rho_0 = 0$ against $H_1 : \rho_0 \neq 0$, we can again utilize the Moran I test statistic given in

(4.1). Motivated by the results in Section 2 we now specify the normalization factor in the Moran I test as:

$$\begin{aligned}\tilde{\sigma}_{Q_n^*}^2 &= (1/2)\hat{\sigma}_n^4 \sum_{i=1}^n \sum_{j=1}^n (w_{ij,n} + w_{ji,n})^2 + \hat{\sigma}_n^2 \sum_{i=1}^n \hat{b}_{i,n}^2 \\ &= \hat{\sigma}_n^4 \text{tr}(W_n W_n + W_n' W_n) + \hat{\sigma}_n^2 \hat{b}_n' \hat{b}_n\end{aligned}\quad (4.18)$$

where $\hat{\sigma}_n^2 = n^{-1} \hat{u}_n' \hat{u}_n$ and $\hat{b}_n = -H_n P_n' \hat{d}_n$ with $\hat{d}_n' = n^{-1} \hat{u}_n' (W_n + W_n') D_n$. In the appendix we demonstrate that the asymptotic normality result (4.1) for the Moran I test statistic holds under the maintained assumptions if H_0 is true. This result can again be used in the usual way to test for the presence of spatial autocorrelation. In the appendix we demonstrate also that under H_1 , and some additional mild assumptions, we have $\lim_{n \rightarrow \infty} P(|I_n| > \tau) = 1$ for all $\tau > 0$. That is the test is consistent, in that under the alternative hypothesis the power of the test tends to unity.²²

5 Concluding Remarks

In this paper we formally derive the large sample distribution of Moran I (type) test statistics for spatial correlation in the context of limited dependent variable models (the Tobit, dichotomous, polychotomous and a sample selection model), as well as for the widely used spatial ARAR(1,1) model. Towards establishing these results we first derive a set of general results concerning the asymptotic distribution of quadratic forms based on estimated residuals, and then apply those results to derive the asymptotic distribution of the Moran I test statistic for the models of interest. The results are such that applications to still other models are evident. The analysis also necessitated the derivation of a new central limit theorem for linear-quadratic forms that allows for heteroskedastic and non-normally distributed innovations (and for the elements of the linear-quadratic forms to constitute triangular arrays).

One suggestion for further research is to study, via Monte Carlo methods, the actual distribution of the Moran I test statistic in finite samples under the

²²Since $d_n' = E[n^{-1} u_n' (W_n + W_n') D_n] = \left(E[n^{-1} u_n' (W_n + W_n') M_n y_n], 0_{1 \times k} \right)$ we may alternatively specify $\hat{d}_n' = \left(n^{-1} \hat{u}_n' (W_n + W_n') M_n y_n, 0_{1 \times k} \right)$ without changing the result on the limiting distribution of the Moran I test statistic.

conditions considered in this paper, and to compare the actual distribution with the large sample distribution. In this regard, it would be especially interesting to focus on limited dependent variable models, since there are no Monte Carlo results in the literature relating to the Moran I test statistic for such models.²³ Another suggestion for further research relates to the need of developing appropriate estimation methods for (some of) these models if the model's disturbance terms are found to be spatially correlated.

²³For Monte Carlo results relating to the small sample distribution of the Moran I test statistic in the context of a linear cross sectional regression model see Anselin and Florax (1995), and Anselin and Kelejian (1997), as well as the references cited there-in.

A Appendix²⁴

Let $\{X_{i,n}, 1 \leq i \leq k_n, n \geq 1\}$ be an array of random variables defined on a probability space $(\Omega, \mathfrak{F}, P)$ with $E|X_{i,n}| < \infty$. Let $\{\mathfrak{F}_{i,n}, 1 \leq i \leq k_n, n \geq 1\}$ be an array of sub-sigma fields with $\mathfrak{F}_{i-1,n} \subseteq \mathfrak{F}_{i,n}$. We then call $\{X_{i,n}, \mathfrak{F}_{i,n}, 1 \leq i \leq k_n, n \geq 1\}$ a martingale difference array if $X_{i,n}$ is $\mathfrak{F}_{i,n}$ -measurable and $E(X_{i,n} | \mathfrak{F}_{i-1,n}) = 0$ (with $\mathfrak{F}_{0,n} = \{\emptyset, \Omega\}$). In the following we shall use the following central limit theorem for martingale difference arrays (with $k_n \uparrow \infty$ as $n \rightarrow \infty$).

Theorem A.1 *Let $\{X_{i,n}, \mathfrak{F}_{i,n}, 1 \leq i \leq k_n, n \geq 1\}$ be a square integrable martingale difference array. Suppose that for all $\epsilon > 0$*

$$\sum_{i=1}^{k_n} E [X_{i,n}^2 \mathbf{I}(|X_{i,n}| > \epsilon) | \mathfrak{F}_{i-1,n}] \xrightarrow{p} 0, \quad (\text{A.1})$$

and

$$\sum_{i=1}^{k_n} E [X_{i,n}^2 | \mathfrak{F}_{i-1,n}] \xrightarrow{p} 1. \quad (\text{A.2})$$

Then $\sum_{i=1}^{k_n} X_{i,n} \xrightarrow{D} N(0, 1)$.

Remark. The above theorem is given in Gansler and Stute (1977), pp. 365, and represents a special case of Corollary 3.1 in Hall and Heyde (1980). The theorem is restated here for the convenience of the reader. It is a variant of a central limit theorem by McLeish (1974); compare also Brown (1971). It is readily seen that a sufficient condition for (A.1) is that

$$\sum_{i=1}^{k_n} E [|X_{i,n}|^{2+\delta} | \mathfrak{F}_{i-1,n}] \xrightarrow{p} 0$$

for some $\delta > 0$; cp., e.g., Potscher and Prucha (1997), p. 235. In turn, applying Chebychev's inequality, it follows that a sufficient condition for the latter condition is that

$$\sum_{i=1}^{k_n} E \left\{ E [|X_{i,n}|^{2+\delta} | \mathfrak{F}_{i-1,n}] \right\} \rightarrow 0 \quad (\text{A.3})$$

²⁴The proofs given in this appendix are, at this point, written in a detailed fashion to make it easier for the reader to check them.

for some $\delta > 0$.

Proof of some assertions made in Section 2. We first note that the following results have been established in Kelejian and Prucha (1998) in the proof of their Theorems 2 and 3:

$$\begin{aligned} E |\bar{y}_{i,n}|^3 &\leq \text{const} < \infty, \\ n^{-1} u'_n D_n &= O_p(1), \\ n^{-1} u'_n (W_n + W'_n) D_n &= O_p(1), \\ n^{-1} D'_n D_n &= O_p(1), \\ n^{-1} D'_n W_n D_n &= O_p(1), \end{aligned} \tag{A.4}$$

where $\bar{y}_n = M_n y_n$.

We now demonstrate that (2.5) holds. From (2.1) we have $\hat{u}_n = u_n - D_n(\hat{\theta}_n - \theta_0)$ and so

$$\begin{aligned} n^{-1/2} \hat{u}'_n W_n \hat{u}_n &= n^{-1/2} u'_n W_n u_n \\ &\quad - n^{-1} u'_n (W_n + W'_n) D_n n^{1/2} (\hat{\theta}_n - \theta_0) \\ &\quad + (\hat{\theta}_n - \theta_0)' [n^{-1} D'_n W_n D_n] [n^{1/2} (\hat{\theta}_n - \theta_0)]. \end{aligned} \tag{A.5}$$

In light of (2.3) and its discussion we have $n^{1/2}(\hat{\theta}_n - \theta_0) = O_p(1)$. Utilizing (A.4) it then follows that the third term on the r.h.s. of (A.5) is $o_p(1)$, which proves the claim.

We next show that

$$d'_n = E [n^{-1} u'_n (W_n + W'_n) D_n] = O(1) \tag{A.6}$$

and that

$$n^{-1} u'_n (W_n + W'_n) D_n - d'_n = o_p(1). \tag{A.7}$$

To demonstrate this we use two evident results (see, e.g., Kelejian and Prucha (1999): First, if \mathbf{S}_n and \mathbf{R}_n are two $n \times n$ matrices whose row and column sums are uniformly bounded in absolute value, then so are the row and column sums of $\mathbf{S}_n \mathbf{R}_n$. Second, maintaining this notation, if \mathbf{T}_n is a conformable matrix whose elements are uniformly bounded in absolute value, then so are the elements of $\mathbf{T}_n \mathbf{R}_n$. Now recall that $D_n = [M_n y_n, X_n]$ and observe that because of (2.1) we have

$$\begin{aligned} y_n &= (I_n - \lambda_0 M_n)^{-1} X_n \beta_0 + (I_n - \lambda_0 M_n)^{-1} (I_n - \rho_0 M_n)^{-1} \varepsilon_n, \\ u_n &= (I_n - \rho_0 M_n)^{-1} \varepsilon_n. \end{aligned} \tag{A.8}$$

In light of this and given the maintained assumptions it is readily seen that all components of $n^{-1}u'_n(W_n + W'_n)D_n$ are composed of terms of the form $n^{-1}B_n^{*'}\varepsilon_n$ and $n^{-1}\varepsilon'_nC_n^*\varepsilon_n$ where B_n^* is a non-stochastic vector whose elements are uniformly bounded in absolute value, and where C_n^* is a non-stochastic matrix whose row and column sums are uniformly bounded in absolute value. The claim in (A.6) now follows since $E(n^{-1}\varepsilon'_nC_n^*\varepsilon_n) = \sigma^2 n^{-1}tr(C_n^*)$. Next observe that $var(n^{-1}B_n^{*'}\varepsilon_n) = \sigma^2 n^{-2}B_n^{*'}B_n^* = o(1)$. Using the results of Section 3.1 we have

$$var(n^{-1}\varepsilon'_nC_n^*\varepsilon_n) = \sigma^4 n^{-2} \sum_{i=1}^n \sum_{j=1}^{i-1} (c_{ij,n}^* + c_{ji,n}^*)^2 + (\mu^{(4)} - \sigma^4) n^{-2} \sum_{i=1}^n c_{ii,n}^*. \quad (\text{A.9})$$

The expressions on the r.h.s. are readily seen to be $o(1)$ - cp. Kelejian and Prucha (1999) - and thus $var(n^{-1}u'_n(W_n + W'_n)D_n) = o(1)$. The claim in (A.7) now follows from Chebychev's inequality.

Consider the vectors $b_n = -F_n P' d_n = -(I_n - \rho_0 M'_n)^{-1} H_n P' d_n$ and the matrices $A_n = \frac{1}{2}(I_n - \rho_0 M'_n)^{-1}(W_n + W'_n)(I_n - \rho_0 M_n)^{-1}$ and observe that the matrix P and the vectors d_n are of finite dimension. Given (A.6) and the assumptions maintained with respect to $(I_n - \rho_0 M_n)^{-1}$, W_n and H_n it follows from the results given after (A.7) that the elements of b_n and that the row and column sums of the matrices A_n are uniformly bounded in absolute value.

For later use we establish furthermore that that

$$n^{-1}\hat{u}'_n(W_n + W'_n)D_n - d'_n = o_p(1). \quad (\text{A.10})$$

Observe that $n^{-1}\hat{u}'_n(W_n + W'_n)D_n = n^{-1}u'_n(W_n + W'_n)D_n - (\hat{\theta}_n - \theta_0)' n^{-1}D_n(W_n + W'_n)D_n$. The second term on the r.h.s. is $o_p(1)$ in light of the consistency of $\hat{\theta}_n$ and (A.4), and hence (A.10) follows from (A.7). ■

Proof of Theorem 1. Clearly $\mu_{Q_n} = \sum_{i=1}^n a_{ii,n} \sigma_{i,n}^2$ in light of the independence of the $\varepsilon_{i,n}$ and since $E\varepsilon_{i,n} = 0$. Next observe that

$$Q_n - \mu_{Q_n} = \sum_{i=1}^n Y_{i,n} \quad (\text{A.11})$$

with²⁵

$$Y_{i,n} = a_{ii,n}(\varepsilon_{i,n}^2 - \sigma_{i,n}^2) + 2\varepsilon_{i,n} \sum_{j=1}^{i-1} a_{ij,n}\varepsilon_{j,n} + b_{i,n}\varepsilon_{i,n}, \quad i = 1, \dots, n. \quad (\text{A.12})$$

Consider the σ -fields $\mathfrak{F}_{0,n} = \{\emptyset, \Omega\}$, $\mathfrak{F}_{i,n} = \sigma(\varepsilon_{1,n}, \dots, \varepsilon_{i,n})$, $1 \leq i \leq n$. By construction $\mathfrak{F}_{i-1,n} \subseteq \mathfrak{F}_{i,n}$, $Y_{i,n}$ is $\mathfrak{F}_{i,n}$ -measurable, and it is easily seen that $E(Y_{i,n} \mid \mathfrak{F}_{i-1,n}) = 0$. Thus $\{Y_{i,n}, \mathfrak{F}_{i,n}, 1 \leq i \leq n, n \geq 1\}$ forms a martingale difference array. Consequently $\sigma_{Q_n}^2 = \sum_{i=1}^n E(Y_{i,n}^2)$. The expression for the variance of Q_n given in the text before the theorem follows upon observing that

$$E(Y_{i,n}^2) = a_{ii,n}^2 \left[\mu_{i,n}^{(4)} - \sigma_{i,n}^4 \right] + 4 \sum_{j=1}^{i-1} a_{ij,n}^2 \sigma_{i,n}^2 \sigma_{j,n}^2 + b_{i,n}^2 \sigma_{i,n}^2 + 2b_{i,n} a_{ii,n} \mu_{i,n}^{(3)}. \quad (\text{A.13})$$

Let $X_{i,n} = Y_{i,n}/\sigma_{Q_n}$, then $\{X_{i,n}, \mathfrak{F}_{i,n}, 1 \leq i \leq n, n \geq 1\}$ also forms a martingale difference array. We now prove that

$$\frac{Q_n - \mu_{Q_n}}{\sigma_{Q_n}} = \sum_{i=1}^n X_{i,n} \xrightarrow{D} N(0, 1) \quad (\text{A.14})$$

by showing that the $X_{i,n}$ satisfies the remaining conditions of Theorem A.1. In particular, we demonstrate that the $X_{i,n}$ satisfy condition (A.3), which in turn is sufficient for (A.1), and that the $X_{i,n}$ satisfy condition (A.2).

We first consider the case where Assumption 3(b) holds (in conjunction with Assumptions 1 and 2) and take $0 < \delta \leq \min(\eta_1, \eta_2)/2$. We note that under the maintained moment assumptions on the $\varepsilon_{i,n}$ there exists then a finite constant, say K_e , such that $\sigma_{i,n}^2 \leq K_e$, $E|\varepsilon_{i,n}|^3 \leq K_e$, $E|\varepsilon_{i,n}|^s E|\varepsilon_{j,n}|^s \leq K_e$, and $E|\varepsilon_{i,n}^2 - \sigma_{i,n}^2|^s \leq K_e$ for $s \leq 2 + \delta$. We note further that under the maintained assumptions on the $a_{ij,n}$ and $b_{i,n}$ there exists a finite constant, say K_p , such that $\sum_{j=1}^n |a_{ij,n}| \leq K_p$ and $n^{-1} \sum_{j=1}^n |b_{i,n}|^s \leq K_p$ for $s \leq 2 + \delta$. Observe that $\sum_{j=1}^n |a_{ij,n}|^r \leq K_p^r$ for $r \geq 1$ and that $\sum_{k=1}^n |a_{ik,n}| |a_{rk,n}| \leq \sum_{k=1}^n |a_{ik,n}| \sum_{k=1}^n |a_{rk,n}| \leq K_p^2$.

²⁵We adopt the convention that any sum with an upper index of less than one is zero.

In the following let $q = 2 + \delta$ and let $1/q + 1/p = 1$. Using the triangle and Hölder's inequalities we then have

$$\begin{aligned}
|Y_{i,n}|^q &= \left| a_{ii,n}(\varepsilon_{i,n}^2 - \sigma_{i,n}^2) + 2\varepsilon_{i,n} \sum_{j=1}^{i-1} a_{ij,n}\varepsilon_{j,n} + b_{i,n}\varepsilon_{i,n} \right|^q \\
&\leq 2^q \left[|a_{ii,n}|^{\frac{1}{p}} |a_{ii,n}|^{\frac{1}{q}} |\varepsilon_{i,n}^2 - \sigma_{i,n}^2| + \sum_{j=1}^{i-1} |a_{ij,n}|^{\frac{1}{p}} |a_{ij,n}|^{\frac{1}{q}} 2 |\varepsilon_{i,n}| |\varepsilon_{j,n}| \right]^q \\
&\quad + 2^q |b_{i,n}|^q |\varepsilon_{i,n}|^q \\
&\leq 2^q \left[\sum_{j=1}^i |a_{ij,n}|^{\frac{p}{p}} \right]^{\frac{q}{p}} \left[|a_{ii,n}|^{\frac{q}{q}} |\varepsilon_{i,n}^2 - \sigma_{i,n}^2|^q + \sum_{j=1}^{i-1} |a_{ij,n}|^{\frac{q}{q}} 2^q |\varepsilon_{i,n}|^q |\varepsilon_{j,n}|^q \right] \\
&\quad + 2^q |b_{i,n}|^q |\varepsilon_{i,n}|^q \\
&\leq 2^{2q} (K_p)^{\frac{q}{p}} \left[|a_{ii,n}| |\varepsilon_{i,n}^2 - \sigma_{i,n}^2|^q + \sum_{j=1}^{i-1} |a_{ij,n}| |\varepsilon_{i,n}|^q |\varepsilon_{j,n}|^q \right] + 2^q |b_{i,n}|^q |\varepsilon_{i,n}|^q.
\end{aligned}$$

Consequently

$$\begin{aligned}
&\sum_{i=1}^n E \{ E[|Y_{i,n}|^q \mid \mathfrak{F}_{i-1,n}] \} \\
&\leq 2^{2q} (K_p)^{\frac{q}{p}} \sum_{i=1}^n \left[|a_{ii,n}| E |\varepsilon_{i,n}^2 - \sigma_{i,n}^2|^q + \sum_{j=1}^{i-1} |a_{ij,n}| E |\varepsilon_{i,n}|^q E |\varepsilon_{j,n}|^q \right] \\
&\quad + 2^q \sum_{i=1}^n |b_{i,n}|^q E |\varepsilon_{i,n}|^q \\
&\leq 2^{2q} (K_p)^{\frac{q}{p}} K_e \sum_{i=1}^n \sum_{j=1}^i |a_{ij,n}| + 2^q K_e \sum_{i=1}^n |b_{i,n}|^q \\
&\leq n \left[2^{2q} (K_p)^{\frac{q}{p}+1} K_e + 2^q K_p K_e \right]
\end{aligned}$$

and thus

$$\begin{aligned}
& \sum_{i=1}^n E \left\{ E \left[|X_{i,n}|^{2+\delta} \mid \mathfrak{F}_{i-1,n} \right] \right\} \\
&= \frac{1}{[n^{-1}\sigma_{Q_n}^2]^{1+\delta/2}} \frac{1}{n^{1+\delta/2}} \sum_{i=1}^n E \left\{ E \left[|Y_{i,n}|^{2+\delta} \mid \mathfrak{F}_{i-1,n} \right] \right\} \\
&\leq \frac{1}{[n^{-1}\sigma_{Q_n}^2]^{1+\delta/2}} \frac{1}{n^{\delta/2}} \left[2^{2q}(K_p)^{\frac{q}{p}+1} K_e + 2^q K_p K_e \right].
\end{aligned}$$

Since $0 < c \leq n^{-1}\sigma_{Q_n}^2$ the r.h.s. of the last inequality goes to zero as $n \rightarrow \infty$, which proves that condition (A.3) holds.

Utilizing that the $\varepsilon_{i,n}$ are independent with zero mean it follows that

$$\begin{aligned}
& E[Y_{i,n}^2 \mid \mathfrak{F}_{i-1,n}] \\
&= E \left\{ \left[a_{ii,n}(\varepsilon_{i,n}^2 - \sigma_{i,n}^2) + 2\varepsilon_{i,n} \sum_{j=1}^{i-1} a_{ij,n}\varepsilon_{j,n} + b_{i,n}\varepsilon_{i,n} \right]^2 \mid \varepsilon_{i-1,n}, \dots, \varepsilon_{1,n} \right\} \\
&= a_{ii,n}^2 \left(\mu_{i,n}^{(4)} - \sigma_{i,n}^4 \right) + 4\sigma_{i,n}^2 \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} a_{ij,n} a_{ik,n} \varepsilon_{j,n} \varepsilon_{k,n} \\
&\quad + 4 \left[a_{ii,n} \mu_{i,n}^{(3)} + b_{i,n} \sigma_{i,n}^2 \right] \sum_{j=1}^{i-1} a_{ij,n} \varepsilon_{j,n} + b_{i,n}^2 \sigma_{i,n}^2 + 2a_{ii,n} b_{i,n} \mu_{i,n}^{(3)}.
\end{aligned}$$

Recalling that $\sigma_{Q_n}^2 = \sum_{i=1}^n E(Y_{i,n}^2)$ and utilizing the expression for $E(Y_{i,n}^2)$ in (A.13) yields

$$\begin{aligned}
& \sum_{i=1}^n E[X_{i,n}^2 \mid \mathfrak{F}_{i-1,n}] - 1 = \frac{1}{\sigma_{Q_n}^2} \sum_{i=1}^n [E[Y_{i,n}^2 \mid \mathfrak{F}_{i-1,n}] - E(Y_{i,n}^2)] \\
&= \frac{1}{\sigma_{Q_n}^2} \left[8 \sum_{i=1}^n \sigma_{i,n}^2 \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} a_{ij,n} a_{ik,n} \varepsilon_{j,n} \varepsilon_{k,n} + 4 \sum_{i=1}^n \sigma_{i,n}^2 \sum_{j=1}^{i-1} a_{ij,n}^2 (\varepsilon_{j,n}^2 - \sigma_{j,n}^2) \right. \\
&\quad \left. + 4 \sum_{i=1}^n \left[a_{ii,n} \mu_{i,n}^{(3)} + b_{i,n} \sigma_{i,n}^2 \right] \sum_{j=1}^{i-1} a_{ij,n} \varepsilon_{j,n} \right] \\
&= \frac{1}{n^{-1}\sigma_{Q_n}^2} [8H_{1,n} + 4H_{2,n} + 4H_{3,n}]
\end{aligned}$$

with

$$\begin{aligned}
H_{1,n} &= n^{-1} \sum_{i=1}^n \sigma_{i,n}^2 \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} a_{ij,n} a_{ik,n} \varepsilon_{j,n} \varepsilon_{k,n}, \\
H_{2,n} &= n^{-1} \sum_{i=1}^n \sigma_{i,n}^2 \sum_{j=1}^{i-1} a_{ij,n}^2 (\varepsilon_{j,n}^2 - \sigma_{j,n}^2) \\
H_{3,n} &= n^{-1} \sum_{i=1}^n \left[a_{ii,n} \mu_{i,n}^{(3)} + b_{i,n} \sigma_{i,n}^2 \right] \sum_{j=1}^{i-1} a_{ij,n} \varepsilon_{j,n}.
\end{aligned}$$

Since $0 < c \leq n^{-1} \sigma_{Q_n}^2$ condition (A.2) holds if $H_{i,n} \xrightarrow{p} 0$ as $n \rightarrow \infty$ for $i = 1, 2, 3$. Clearly $EH_{1,n} = 0$. Utilizing that the $\varepsilon_{i,n}$ are independent with zero mean and variance $\sigma_{i,n}^2$ it is not difficult (although somewhat tedious) to see that

$$\begin{aligned}
EH_{1,n}^2 &\leq 2n^{-2} \sum_{i=1}^n \sum_{r=1}^{i-1} \sum_{j=1}^{r-1} |a_{ij,n}| |a_{rj,n}| \sum_{k=1}^{j-1} |a_{ik,n}| |a_{rk,n}| \sigma_{i,n}^2 \sigma_{r,n}^2 \sigma_{j,n}^2 \sigma_{k,n}^2 \\
&\leq 2K_e^4 n^{-2} \sum_{i=1}^n \sum_{j=1}^n |a_{ij,n}| \sum_{r=1}^n |a_{rj,n}| \sum_{k=1}^n |a_{ik,n}| |a_{rk,n}| \\
&\leq n^{-1} 2K_e^4 K_p^4 \rightarrow 0 \text{ as } n \rightarrow \infty.
\end{aligned}$$

This proves $H_{1,n} \xrightarrow{p} 0$.

Next observe that

$$\begin{aligned}
H_{2,n} &= \sum_{i=1}^{n-1} \phi_{i,n} (\varepsilon_{i,n}^2 - \sigma_{i,n}^2), \\
\phi_{i,n} &= n^{-1} \sum_{j=i+1}^n \sigma_{j,n}^2 a_{ji,n}^2.
\end{aligned}$$

Given our assumptions the $\varepsilon_{i,n}^2 - \sigma_{i,n}^2$ are independent with mean zero. Since also $E |\varepsilon_{i,n}^2 - \sigma_{i,n}^2|^{1+\delta} \leq K_e$ (with $\delta > 0$) it follows that the $\varepsilon_{i,n}^2 - \sigma_{i,n}^2$ are

uniformly integrable. Furthermore

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sum_{i=1}^{n-1} \phi_{i,n} &= \limsup_{n \rightarrow \infty} n^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sigma_{j,n}^2 a_{ji,n}^2 \\ &\leq K_e \limsup_{n \rightarrow \infty} n^{-1} \sum_{i=1}^{n-1} \sum_{j=1}^n a_{ji,n}^2 \leq K_e K_p^2 < \infty \end{aligned}$$

and

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} \phi_{i,n}^2 &= \lim_{n \rightarrow \infty} n^{-2} \sum_{i=1}^{n-1} \left[\sum_{j=i+1}^n \sigma_{j,n}^2 a_{ji,n}^2 \right]^2 \\ &\leq K_e^2 \lim_{n \rightarrow \infty} n^{-2} \sum_{i=1}^{n-1} \sum_{j=1}^n a_{ji,n}^2 \sum_{k=1}^n a_{ki,n}^2 \leq \lim_{n \rightarrow \infty} n^{-1} K_e^2 K_p^2 = 0. \end{aligned}$$

It now follows from the weak law of large numbers (for martingale difference arrays) in Davidson (1994), p. 299, that $H_{2,n} \xrightarrow{p} 0$.

The proof that $H_{3,n} \xrightarrow{p} 0$ is analogous. Observe that

$$\begin{aligned} H_{3,n} &= \sum_{i=1}^{n-1} \varphi_{i,n} \varepsilon_{i,n}, \\ \varphi_{i,n} &= n^{-1} \sum_{j=i+1}^n \left[a_{jj,n} \mu_{j,n}^{(3)} + b_{j,n} \sigma_{j,n}^2 \right] a_{ji,n}. \end{aligned}$$

The $\varepsilon_{i,n} \varphi_{i,n} / |\varphi_{i,n}|$ are then independent with zero mean and uniformly integrable given our moment assumptions. Furthermore

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sum_{i=1}^{n-1} |\varphi_{i,n}| &\leq \limsup_{n \rightarrow \infty} n^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left[|a_{jj,n}| |\mu_{j,n}^{(3)}| + |b_{j,n}| \sigma_{j,n}^2 \right] |a_{ji,n}| \\ &\leq K_e \limsup_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n [|a_{jj,n}| + |b_{j,n}|] \sum_{i=1}^{n-1} |a_{ji,n}| \\ &\leq 2K_e K_p^2 < \infty \end{aligned}$$

and

$$\begin{aligned}
\lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} \varphi_{i,n}^2 &\leq \lim_{n \rightarrow \infty} n^{-2} \sum_{i=1}^{n-1} \left[\sum_{j=i+1}^n \left[|a_{jj,n}| \left| \mu_{j,n}^{(3)} \right| + |b_{j,n}| \sigma_{j,n}^2 \right] |a_{ji,n}| \right]^2 \\
&\leq K_e^2 \lim_{n \rightarrow \infty} n^{-2} \sum_{i=1}^{n-1} \left[\sum_{j=1}^n [|a_{jj,n}| + |b_{j,n}|] |a_{ji,n}| \right]^2 \\
&\leq K_e^2 \lim_{n \rightarrow \infty} n^{-2} \sum_{i=1}^{n-1} \left[K_p^2 + \sum_{j=1}^n |a_{j,n}| |a_{ji,n}| \right]^2
\end{aligned}$$

Using Hölder's inequality we have for $q = 2 + \delta$ and $1/q + 1/p = 1$

$$\sum_{j=1}^n |b_{j,n}| |a_{ji,n}| \leq n^{\frac{1}{q}} \left[n^{-1} \sum_{j=1}^n |b_{j,n}|^q \right]^{\frac{1}{q}} \left[\sum_{j=1}^n |a_{ji,n}|^p \right]^{\frac{1}{p}} \leq n^{\frac{1}{q}} K_p.$$

Consequently

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} \varphi_{i,n}^2 \leq K_e^2 \left[\lim_{n \rightarrow \infty} n^{-1/2} (K_p^2 + n^{\frac{1}{2+\delta}} K_p) \right]^2 = 0.$$

Applying again the weak law of large numbers (for martingale difference arrays) in Davidson (1994), p. 299, gives $H_{3,n} \xrightarrow{p} 0$. We have thus demonstrated that also condition (A.2) holds, which completes the proof under Assumption 3(b).

The proof under Assumption 3(a) is identical to that under Assumption 3(b), except that under Assumption 3(a) all terms involving moments of order higher than $2 + \delta$ drop out from the expressions since in this case $a_{ii,n} = 0$. ■

Proof of Lemma 1. Since $\widehat{u}_n' W_n \widehat{u}_n = \widehat{u}_n' [(W_n + W_n')/2] \widehat{u}_n$ we prove the result w.l.o.g. under the assumption that W_n is symmetric. By Assumption 6 the functions $g_{i,n}(z_n, \theta)$, and hence $g_n(z_n, \theta)' W_n g_n(z_n, \theta)$, are twice continuously differentiable in θ . Applying a second order Taylor expansion to $\widehat{u}_n' W_n \widehat{u}_n = g_n(z_n, \widehat{\theta}_n)' W_n g_n(z_n, \widehat{\theta}_n)$ yields (abusing notation in an obvious way)

$$n^{-1/2} \widehat{u}_n' W_n \widehat{u}_n = n^{-1/2} u_n' W_n u_n + 2n^{-1} u_n' W_n \frac{\partial g_n(z_n, \theta_0)}{\partial \theta} n^{1/2} (\widehat{\theta}_n - \theta_0) + \frac{1}{2} R_{1n}$$

with

$$\begin{aligned}
R_{1n} &= n^{-1/2} (\widehat{\theta}_n - \theta_0)' \frac{\partial^2 g_n(z_n, \widetilde{\theta}_n)' W_n g_n(z_n, \widetilde{\theta}_n)}{\partial \theta \partial \theta'} (\widehat{\theta}_n - \theta_0) \\
&= n^{-1/2} \sum_{r=1}^{p_\theta} \sum_{s=1}^{p_\theta} \frac{\partial^2 g_n(z_n, \widetilde{\theta}_n)' W_n g_n(z_n, \widetilde{\theta}_n)}{\partial \theta_r \partial \theta_s} (\widehat{\theta}_{r,n} - \theta_{r,0}) (\widehat{\theta}_{s,n} - \theta_{s,0}) \\
&= n^{-1/2} \sum_{r=1}^{p_\theta} \sum_{s=1}^{p_\theta} \sum_{i=1}^n \sum_{j=1}^n w_{ij,n} \frac{\partial^2 g_{i,n}(z_n, \widetilde{\theta}_n) g_{j,n}(z_n, \widetilde{\theta}_n)}{\partial \theta_r \partial \theta_s} (\widehat{\theta}_{r,n} - \theta_{r,0}) (\widehat{\theta}_{s,n} - \theta_{s,0})
\end{aligned}$$

and where $\widetilde{\theta}_n$ is a vector of between values. Since $\widetilde{\theta}_n \rightarrow \theta_0$ we have $\widetilde{\theta}_n \in \Theta$ for large n in light of Assumption 5, and consequently by Assumption 6

$$\begin{aligned}
& \left| \frac{\partial^2 g_{i,n}(z_n, \widetilde{\theta}_n) g_{j,n}(z_n, \widetilde{\theta}_n)}{\partial \theta_r \partial \theta_s} \right| \\
& \leq \left| \frac{\partial^2 g_{i,n}(z_n, \widetilde{\theta}_n)}{\partial \theta_r \partial \theta_s} \right| |g_{j,n}(z_n, \widetilde{\theta}_n)| + \left| \frac{\partial g_{i,n}(z_n, \widetilde{\theta}_n)}{\partial \theta_r} \right| \left| \frac{\partial g_{j,n}(z_n, \widetilde{\theta}_n)}{\partial \theta_s} \right| \\
& \leq \left| \frac{\partial g_{i,n}(z_n, \widetilde{\theta}_n)}{\partial \theta_s} \right| \left| \frac{\partial g_{j,n}(z_n, \widetilde{\theta}_n)}{\partial \theta_r} \right| + |g_{i,n}(z_n, \widetilde{\theta}_n)| \left| \frac{\partial^2 g_{j,n}(z_n, \widetilde{\theta}_n)}{\partial \theta_r \partial \theta_s} \right| \\
& \leq 4\Delta_{i,n} \Delta_{j,n}.
\end{aligned}$$

Therefore

$$|R_{1n}| \leq 4 \sum_{r=1}^{p_\theta} \sum_{s=1}^{p_\theta} \varphi_n n^{1/2} \left| \widehat{\theta}_{r,n} - \theta_{r,0} \right| \left| \widehat{\theta}_{s,n} - \theta_{s,0} \right|.$$

with $\varphi_n = n^{-1} \sum_{i=1}^n \sum_{j=1}^n |w_{ij,n}| \Delta_{i,n} \Delta_{j,n}$. By Assumption 4, 6 and Hölder's inequality

$$\begin{aligned}
E\varphi_n &\leq n^{-1} \sum_{i=1}^n \sum_{j=1}^n |w_{ij,n}| [E\Delta_{i,n}^2]^{1/2} [E\Delta_{j,n}^2]^{1/2} \\
&\leq K_\delta n^{-1} \sum_{i=1}^n \sum_{j=1}^n |w_{ij,n}| \leq K_\delta K_w < \infty,
\end{aligned}$$

where K_δ and K_w are the suprema for $E\Delta_{i,n}^2$ and $\sum_{j=1}^n |w_{ij,n}|$. Consequently $\varphi_n = O_p(1)$. Since $n^{1/2} \left| \widehat{\theta}_{r,n} - \theta_{r,0} \right| = O_p(1)$ and $\left| \widehat{\theta}_{s,n} - \theta_{s,0} \right| = o_p(1)$ by

assumption it follows that $R_{1n} = o_p(1)$. This proves the first part of the lemma.

Applying a first order Taylor expansion yields

$$n^{-1}\widehat{u}'_n W_n \frac{\partial g_n(z_n, \widehat{\theta}_n)}{\partial \theta} = n^{-1}u'_n W_n \frac{\partial g_n(z_n, \theta_0)}{\partial \theta} + \frac{1}{2}R_{2n}$$

with

$$R_{2n} = (\widehat{\theta}_n - \theta_0)' n^{-1} \frac{\partial^2 g_n(z_n, \widetilde{\theta}_n)' W_n g_n(z_n, \widetilde{\theta}_n)}{\partial \theta \partial \theta'}$$

and where $\widetilde{\theta}_n$ is a vector of between values. Let $r_{2s,n}$ denote the s -th element of R_{2n} , then

$$\begin{aligned} r_{2s,n} &= n^{-1} \sum_{r=1}^{p_\theta} \frac{\partial^2 g_{i,n}(z_n, \widetilde{\theta}_n) g_{j,n}(z_n, \widetilde{\theta}_n)}{\partial \theta_r \partial \theta_s} (\widehat{\theta}_{r,n} - \theta_{r,0}) \\ &= n^{-1} \sum_{r=1}^{p_\theta} \sum_{i=1}^n \sum_{j=1}^n w_{ij,n} \frac{\partial^2 g_{i,n}(z_n, \widetilde{\theta}_n) g_{j,n}(z_n, \widetilde{\theta}_n)}{\partial \theta_r \partial \theta_s} (\widehat{\theta}_{r,n} - \theta_{r,0}). \end{aligned}$$

By arguments analogous to those used in the first part of the proof we have

$$|r_{2s,n}| \leq 4 \sum_{r=1}^{p_\theta} \varphi_n \left| \widehat{\theta}_{r,n} - \theta_{r,0} \right|.$$

Since φ_n was shown to be $O_p(1)$ and $\left| \widehat{\theta}_{r,n} - \theta_{r,0} \right|$ is $o_p(1)$ it follows that $r_{2s,n} = o_p(1)$. This proves the second part of the lemma. \blacksquare

Proof of Theorem 2. By Lemma 2, the assumptions of the theorem, and recalling that $Q_n = \varepsilon'_n A_n \varepsilon_n + b'_n \varepsilon_n$ we have

$$n^{-1/2} \widehat{u}'_n W_n \widehat{u}_n = n^{-1/2} Q_n + o_p(1)$$

and consequently

$$\frac{\widehat{u}'_n W_n \widehat{u}_n}{\widehat{\sigma}_{Q_n}} = \frac{n^{-1/2} \widehat{u}'_n W_n \widehat{u}_n}{[n^{-1} \widehat{\sigma}_{Q_n}^2]^{1/2}} = \frac{n^{-1/2} [Q_n - \mu_{Q_n}]}{[n^{-1} \widehat{\sigma}_{Q_n}^2]^{1/2}} + \frac{n^{-1/2} \mu_{Q_n}}{[n^{-1} \widehat{\sigma}_{Q_n}^2]^{1/2}} + \frac{o_p(1)}{[n^{-1} \widehat{\sigma}_{Q_n}^2]^{1/2}}.$$

By assumption $n^{-1} \widehat{\sigma}_{Q_n}^2 - n^{-1} \bar{\sigma}_{Q_n}^2 = o_p(1)$ and $\lim_{n \rightarrow \infty} \sigma_{Q_n}^2 / \bar{\sigma}_{Q_n}^2 = \Lambda^2 > 0$.

Since $n^{-1} \sigma_{Q_n}^2$ is additionally bounded away from zero we have $[n^{-1} \sigma_{Q_n}^2]^{1/2} / [n^{-1} \widehat{\sigma}_{Q_n}^2]^{1/2} =$

$\Lambda + o_p(1)$ and $o_p(1)/[n^{-1}\hat{\sigma}_{Q_n}^2]^{1/2} = o_p(1)$. Observing furthermore that $(Q_n - \mu_{Q_n})/\sigma_{Q_n} = O_p(1)$ yields

$$\frac{\hat{u}_n' W_n \hat{u}_n}{\hat{\sigma}_{Q_n}} = \Lambda \frac{Q_n - \mu_{Q_n}}{\sigma_{Q_n}} + \frac{n^{-1/2} \mu_{Q_n}}{[n^{-1}\hat{\sigma}_{Q_n}^2]^{1/2}} + o_p(1).$$

Result (a) of the theorem now follows immediately from Theorem 1, after setting $\mu_{Q_n} = 0$ in the above expression.

By assumption $c \leq n^{-1}\sigma_{Q_n}^2$ for some $c > 0$. In light of Assumption 1-3 it is not difficult to see that $n^{-1}\sigma_{Q_n}^2 \leq c_*$ for some $c_* < \infty$. Since $\lim_{n \rightarrow \infty} \sigma_{Q_n}^2/\bar{\sigma}_{Q_n}^2 = \Lambda^2 > 0$ it follows that there exist constants \bar{c} and \bar{c}_* such that for n large enough $0 < \bar{c} \leq n^{-1}\bar{\sigma}_{Q_n}^2 \leq \bar{c}_* < \infty$. Result (b) of the theorem now follows immediately from the lemma below with $\eta_n = \Lambda(Q_n - \mu_{Q_n})/\sigma_{Q_n} + o_p(1)$, $\varphi_n = [n^{-1}\hat{\sigma}_{Q_n}^2]^{1/2}$, $f_n = [n^{-1}\bar{\sigma}_{Q_n}^2]^{1/2}$ for $\bar{c} \leq n^{-1}\bar{\sigma}_{Q_n}^2 \leq \bar{c}_*$ and $f_n = (\bar{c}^{1/2} + \bar{c}_*^{1/2})/2$ otherwise, and $d_n = n^{-1/2}\mu_{Q_n}$. ■

Lemma A.1 *Let η_n and φ_n be a sequence of real valued random variables, and let d_n and f_n be sequences of real numbers with $0 < K_1 \leq f_n \leq K_2 < \infty$. Suppose $\eta_n = O_p(1)$, $\varphi_n - f_n = o_p(1)$ and $|d_n| \rightarrow \infty$ as $n \rightarrow \infty$. Then for every constant $\gamma > 0$ we have*

$$\lim_{n \rightarrow \infty} P(|\eta_n + d_n/\varphi_n| \leq \gamma) = 0.$$

Proof of Lemma. Choose some $\varepsilon > 0$ and some $K_* > K_2$. Since $\eta_n = O_p(1)$ it follows that there exists a positive constant M_ε such that $P(|\eta_n| > M_\varepsilon) \leq \varepsilon/2$ for all n . Since $\varphi_n - f_n = o_p(1)$ and $f_n \leq K_2 < \infty$ it follows further that there exists an index n_ε such that $P(|\varphi_n| \geq K_*) \leq \varepsilon/2$ for all $n \geq n_\varepsilon$. Next observe that

$$\begin{aligned} & P(|\eta_n + d_n/\varphi_n| \leq \gamma) \\ &= P(|\eta_n + d_n/\varphi_n| \leq \gamma, |\eta_n| > M_\varepsilon) + P(|\eta_n + d_n/\varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon) \\ &\leq P(|\eta_n| > M_\varepsilon) + P(|\eta_n + d_n/\varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon) \\ &\leq \varepsilon/2 + P(|\eta_n + d_n/\varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon, |\varphi_n| \geq K_*) \\ &\quad + P(|\eta_n + d_n/\varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon, |\varphi_n| < K_*) \\ &\leq \varepsilon/2 + P(\varphi_n \geq K_*) + P(|\eta_n + d_n/\varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon, |\varphi_n| < K_*). \end{aligned}$$

For $n \geq n_\varepsilon$ we thus have

$$P(|\eta_n + d_n/\varphi_n| \leq \gamma) \leq \varepsilon + P(|\eta_n + d_n/\varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon, |\varphi_n| < K_*).$$

Now let $\Omega_n = \{\omega \in \Omega : |\eta_n(\omega)| \leq M_\varepsilon, |\varphi_n(\omega)| < K_*\}$. Then $|d_n|/|\varphi_n(\omega)| \geq |d_n|/K_*$ for all $\omega \in \Omega_n$, $n \in \mathbb{N}$. Since $|d_n| \rightarrow \infty$ as $n \rightarrow \infty$ there exists an index n_ε^* such that $|d_n|/K_* > M_\varepsilon + 2\gamma$ for all $n \geq n_\varepsilon^*$. Consequently, provided that $n \geq n_\varepsilon^*$ we have for all $\omega \in \Omega_n$

$$|\eta_n(\omega) + d_n/\varphi_n(\omega)| \geq |d_n|/|\varphi_n(\omega)| - |\eta_n(\omega)| \geq 2\gamma.$$

Thus $P(|\eta_n + d_n/\varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon, |\varphi_n| < K_*) = 0$ for $n \geq n_\varepsilon^*$, and consequently $P(|\eta_n + d_n/\varphi_n| \leq \gamma) \leq \varepsilon$ for all $n \geq \max\{n_\varepsilon, n_\varepsilon^*\}$, which proves the claim. \blacksquare

Proof of Theorem 3. In light of Assumption 4 there exists a finite constant, say K_w , such that $n^{-1} \sum_{i=1}^n \sum_{j=1}^{i-1} (w_{ij,n} + w_{ji,n})^2 \leq K_w^2$ and $w_{ii,n}^2 \leq K_w^2$; compare the analogous discussion concerning the $a_{ij,n}$'s in the proof of Theorem 1. Furthermore, in light of Assumption 8 there exists a finite constant, say K_r , such that $n^{-1} \sum_{i=1}^n |\bar{b}_{i,n}|^\eta \leq K_r$ for $\eta \leq 2$, and $\bar{\sigma}_{i,n}^2 \leq K_r$ and $|\bar{\mu}_{i,n}^{(s)}| \leq K_r$ ($s = 3, 4$). Now observe that

$$\hat{\sigma}_{i,n}^2 \hat{\sigma}_{j,n}^2 - \bar{\sigma}_{i,n}^2 \bar{\sigma}_{j,n}^2 = (\hat{\sigma}_{i,n}^2 - \bar{\sigma}_{i,n}^2) \bar{\sigma}_{j,n}^2 + (\hat{\sigma}_{i,n}^2 - \bar{\sigma}_{i,n}^2)(\hat{\sigma}_{j,n}^2 - \bar{\sigma}_{j,n}^2) + \bar{\sigma}_{i,n}^2 (\hat{\sigma}_{j,n}^2 - \bar{\sigma}_{j,n}^2).$$

Assumption 8 then implies

$$|\hat{\sigma}_{i,n}^2 \hat{\sigma}_{j,n}^2 - \bar{\sigma}_{i,n}^2 \bar{\sigma}_{j,n}^2| \leq 2K_r \phi_n + \phi_n^2.$$

Consequently

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{i-1} (w_{ij,n} + w_{ji,n})^2 [\hat{\sigma}_{i,n}^2 \hat{\sigma}_{j,n}^2 - \bar{\sigma}_{i,n}^2 \bar{\sigma}_{j,n}^2] \right| &\leq K_w^2 (2K_r \phi_n + \phi_n^2), \\ \left| \frac{1}{n} \sum_{i=1}^n w_{ii,n}^2 [\hat{\sigma}_{i,n}^4 - \bar{\sigma}_{i,n}^4] \right| &\leq K_w^2 (2K_r \phi_n + \phi_n^2), \\ \left| \frac{1}{n} \sum_{i=1}^n w_{ii,n}^2 [\hat{\mu}_{i,n}^{(4)} - \bar{\mu}_{i,n}^{(4)}] \right| &\leq K_w^2 \phi_n. \end{aligned} \tag{A.15}$$

Similarly

$$\begin{aligned}
\left| \widehat{b}_{i,n}^2 - \bar{b}_{i,n}^2 \right| &\leq 2 |\bar{b}_{i,n}| \Psi_{i,n} \psi_n + \Psi_{i,n}^2 \psi_n^2, \\
\left| \widehat{b}_{i,n}^2 \widehat{\sigma}_{i,n}^2 - \bar{b}_{i,n}^2 \bar{\sigma}_{i,n}^2 \right| &\leq 2K_r |\bar{b}_{i,n}| \Psi_{i,n} \psi_n + K_r \Psi_{i,n}^2 \psi_n^2 \\
&\quad + 2 |\bar{b}_{i,n}| \Psi_{i,n} \psi_n \phi_n + \Psi_{i,n}^2 \psi_n^2 \phi_n + |\bar{b}_{i,n}|^2 \phi_n, \\
\left| \widehat{b}_{i,n} \widehat{\mu}_{i,n}^{(3)} - \bar{b}_{i,n} \mu_{i,n}^{(3)} \right| &\leq K_r \Psi_{i,n} \psi_n + \Psi_{i,n} \psi_n \phi_n + |\bar{b}_{i,n}| \phi_n.
\end{aligned}$$

Observing that by Hölder's and Lyapunov's inequalities $n^{-1} \sum_{i=1}^n |\bar{b}_{i,n}| \Psi_{i,n} \leq K_r^{1/2} B_n^{1/2}$ and $n^{-1} \sum_{i=1}^n \Psi_{i,n} \leq B_n^{1/2}$ with $B_n = n^{-1} \sum_{i=1}^n \Psi_{i,n}^2$ it follows from the above inequalities that

$$\begin{aligned}
\left| \frac{1}{n} \sum_{i=1}^n \left[\widehat{b}_{i,n}^2 \widehat{\sigma}_{i,n}^2 - b_{i,n}^2 \sigma_{i,n}^2 \right] \right| &\leq 2K_r^{3/2} B_n^{1/2} \psi_n + K_r B_n \psi_n^2 \quad (\text{A.16}) \\
&\quad + 2K_r^{1/2} B_n \psi_n \phi_n + B_n \psi_n^2 \phi_n + K_r \phi_n, \\
\left| \frac{1}{n} \sum_{i=1}^n \left[\widehat{b}_{i,n} \widehat{\mu}_{i,n}^{(3)} - b_{i,n} \mu_{i,n}^{(3)} \right] \right| &\leq K_r B_n \psi_n + B_n \psi_n \phi_n + K_r \phi_n.
\end{aligned}$$

In light of Assumption 8 we have $B_n = O_p(1)$. Since $\psi_n = o_p(1)$ and $\phi_n = o_p(1)$ it is now evident from (A.15) and (A.16) and the triangle inequality that $n^{-1}(\widehat{\sigma}_{Q_n}^2 - \bar{\sigma}_{Q_n}^2) = o_p(1)$. The claim that $n^{-1} \bar{\sigma}_{Q_n}^2 \leq \text{const} < \infty$ is easily checked in light of Assumptions 4 and 8. \blacksquare

Proof of Claims in Section 4.1. We first prove the claim (4.1) for the Tobit model. The proof utilizes Theorems 2(a) and 3. To cast the Tobit model (4.5)-(4.7) into the notation of those theorems let $z_n = (z_{1,n}, \dots, z_{n,n})$ with $z_{i,n} = (y_i, x_i)$. The elements of x_i are assumed to be bounded in absolute value by some finite constant, say C_x . In the following let $Z = \mathbb{R} \times \bar{X}$ with $\bar{X} = \prod_{i=1}^k [-C_x, C_x]$. The parameter space for β_0 , say Θ_β , was assumed to be an open subset of \mathbb{R}^k . Furthermore Θ_β was assumed to be contained in a compact set. Hence there exists a finite constant, say C_β , such $\Theta_\beta \subseteq \prod_{i=1}^k [-C_\beta, C_\beta]$. Now let $\Theta = \Theta_\beta \times (c_\sigma, C_\sigma)$ denote the parameter space, and define $\bar{\Theta} = \prod_{i=1}^k [-C_\beta, C_\beta] \times [c_\sigma, C_\sigma]$ and $\Theta_* = \prod_{i=1}^k (-C_{\beta*}, C_{\beta*}) \times (c_{\sigma*}, C_{\sigma*})$ with $C_\beta < C_{\beta*}$, $C_\sigma < C_{\sigma*}$, and $0 < c_{\sigma*} < c_\sigma$. Observe that the sets Θ , $\bar{\Theta}$,

and Θ_* are open, compact and open, respectively, and satisfy $\Theta \subseteq \overline{\Theta} \subseteq \Theta_*$ by construction. Next define

$$g_{i,n}(\underline{z}_1, \dots, \underline{z}_n, \theta) = \underline{y}_i - f(\underline{x}_i, \theta) \quad (\text{A.17})$$

for all $\underline{z}_j = (\underline{y}_j, \underline{x}_j) \in Z$, $j = 1, \dots, n$, and $\theta = (\beta, \sigma_v^2) \in \Theta_*$. Then (4.5) can be written as

$$u_{i,n} = \varepsilon_i = g_{i,n}(z_n, \theta_0) = y_i - f(x_i, \theta_0). \quad (\text{A.18})$$

Furthermore we have

$$\widehat{u}_{i,n} = \widehat{\varepsilon}_i = g_{i,n}(z_n, \widehat{\theta}_n) = y_i - f(x_i, \widehat{\theta}_n). \quad (\text{A.19})$$

We now check Assumptions 4-6. Assumption 4 is assumed to hold. Assumption 5 holds evidently since Θ is open. We next verify Assumption 6. Since $\phi_N(\cdot)$ and $\Phi_N(\cdot)$ are twice continuously differentiable, and since $\sup_{\underline{x} \in \overline{X}} \sup_{\theta \in \Theta_*} |\underline{x}\beta/\sigma_v| < \infty$, it is readily seen that $f(\underline{x}, \theta)$, $\partial f(\underline{x}, \theta)/\partial \theta$ and $\partial^2 f(\underline{x}, \theta)/\partial \theta \partial \theta'$ are continuous on $Z \times \Theta_*$. (Note that the derivatives are well defined since Θ_* is open.) Assumption 6(a) is hence clearly satisfied. Since \overline{X} and $\overline{\Theta}$ are compact, and continuous functions are bounded in absolute value on compact sets, it follows furthermore that here exists some finite constant K such that

$$\begin{aligned} \sup_{\theta \in \Theta} |x_i \beta| &\leq \sup_{\underline{x} \in \overline{X}} \sup_{\theta \in \overline{\Theta}} |\underline{x}\beta| < K, \\ \sup_{\theta \in \Theta} \left| \frac{x_i \beta}{\sigma_v} \right| &\leq \sup_{\underline{x} \in \overline{X}} \sup_{\theta \in \overline{\Theta}} \left| \frac{\underline{x}\beta}{\sigma_v} \right| < K, \\ \sup_{\theta \in \Theta} |f(x_i, \theta)| &\leq \sup_{\underline{x} \in \overline{X}} \sup_{\theta \in \overline{\Theta}} |f(\underline{x}, \theta)| < K, \\ \sup_{\theta \in \Theta} \left| \frac{\partial g_{i,n}(z_n, \theta)}{\partial \theta} \right| &= \sup_{\theta \in \Theta} \left| \frac{\partial f(x_i, \theta)}{\partial \theta} \right| \leq \sup_{\underline{x} \in \overline{X}} \sup_{\theta \in \overline{\Theta}} \left| \frac{\partial f(\underline{x}, \theta)}{\partial \theta} \right| < K, \\ \sup_{\theta \in \Theta} \left| \frac{\partial^2 g_{i,n}(z_n, \theta)}{\partial \theta \partial \theta'} \right| &= \sup_{\theta \in \Theta} \left| \frac{\partial^2 f(x_i, \theta)}{\partial \theta \partial \theta'} \right| \leq \sup_{\underline{x} \in \overline{X}} \sup_{\theta \in \overline{\Theta}} \left| \frac{\partial^2 f(\underline{x}, \theta)}{\partial \theta \partial \theta'} \right| < K. \end{aligned} \quad (\text{A.20})$$

Observing that $|y_i| \leq |y_i^*| \leq |x_i \beta_0| + |v_i| \leq K + |v_i|$ it follows further that

$$\sup_{\theta \in \Theta} |g_{i,n}(z_n, \theta)| \leq |y_i| + \sup_{\theta \in \Theta} |f(x_i, \theta)| \leq |v_i| + 2K. \quad (\text{A.21})$$

Since $Ev_i^2 < C_\sigma < \infty$ it follows immediately from (A.20) and (A.21) that also Assumption 6(b) holds.

We next verify Assumption 7 with $u_n = \varepsilon_n$, $Q_n = \varepsilon_n' A_n \varepsilon_n$, $A_n = (W_n + W_n')/2$ and $b_n = 0$. We first verify that the elements of ε_n , A_n and b_n satisfy Assumptions 1-3. The conditions postulated in Assumption 2 for the elements of A_n and b_n are trivially satisfied, given the elements of W_n are assumed to satisfy Assumption 4 and the elements of b_n are taken to be zero. Given our discussion in Section 4.1.1 the innovations $\varepsilon_{i,n} = \varepsilon_i$ clearly satisfy Assumption 1. We next show that the innovations satisfy Assumption 3. More specifically, since $a_{ii,n} = w_{ii,n} = 0$ it suffices to demonstrate that Assumption 3(a) holds for, say, $\eta_2 = 1$. Now

$$\begin{aligned}
E |\varepsilon_i|^3 &= E |y_i - f(x_i, \theta_0)|^3 \leq 4 [E |y_i|^3 + |f(x_i, \theta_0)|^3] \quad (\text{A.22}) \\
&\leq 4 \left\{ E |y_i|^3 + \left[\sup_{\underline{x} \in \bar{X}} \sup_{\theta \in \bar{\Theta}} |f(\underline{x}, \theta)| \right]^3 \right\} \\
&\leq 4 \{ E |y_i|^3 + K^3 \} \\
&\leq 4 \{ E |v_i|^3 + 3K E |v_i|^2 + 3K^2 E |v_i| + 2K^3 \}
\end{aligned}$$

where the last inequality utilizes that $|y_i| \leq K + |v_i|$. Since v_i is i.i.d. normal the expression on the r.h.s. of the last inequality equals a finite constant (that does not depend on i), which proves that Assumption 3(a) holds.

Next consider condition (3.7) of Assumption 7. Clearly to verify this condition it suffices to show that

$$\begin{aligned}
&n^{-1} u_n' (W_n + W_n') \frac{\partial g_{i,n}(z_n, \theta_0)}{\partial \theta} \quad (\text{A.23}) \\
&= -2n^{-1} \varepsilon_n' A_n \frac{\partial f(x_i, \theta_0)}{\partial \theta} = o_p(1),
\end{aligned}$$

since $n^{1/2}(\hat{\theta}_n - \theta_0) = O_p(1)$. The elements of $\partial f(x_i, \theta_0)/\partial \theta$ are nonstochastic and bounded in absolute value in light of (A.20). Since the elements of W_n are assumed to satisfy Assumption 4 it follows that the elements of $A_n \partial f(x_i, \theta_0)/\partial \theta$ are uniformly bounded in absolute value. From the discussion in Section 4.1.1 we have $E \varepsilon_i = 0$ and $\sigma_i^2 = E \varepsilon_i^2 = h_i(\theta_0)$ with $h_i(\theta_0)$ defined by (4.7). In light of in (A.22) the variances σ_i^2 are bounded by some finite constant. Consequently $n^{-1} \varepsilon_n' A_n \partial f(x_i, \theta_0)/\partial \theta$ has mean zero and its variance covariance matrix converges to zero; (A.23) now follows from Chebyshev's inequality and thus (3.7) holds. We have now verified all conditions of Assumption 7.

Now take $\widehat{\sigma}_{Q_n}^2 = \widetilde{\sigma}_{Q_n^*}^2$, where the latter is defined in (4.1), and $\overline{\sigma}_{Q_n}^2 = \sigma_{Q_n}^2$ with

$$\sigma_{Q_n}^2 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij,n} + w_{ji,n})^2 \sigma_i^2 \sigma_j^2. \quad (\text{A.24})$$

Then clearly $\Lambda = 1$. Furthermore, given (4.2) and since $\sigma_i^2 = h_i(\theta_0) \geq \text{const} > 0$ there exists some $c > 0$ such that $n^{-1}\sigma_{Q_n}^2 \geq c$.

The final step in our proof is now to demonstrate that

$$n^{-1}\widehat{\sigma}_{Q_n}^2 - n^{-1}\sigma_{Q_n}^2 = o_p(1) \quad (\text{A.25})$$

using Theorem 3. In the notation of that theorem we have $\sigma_{i,n}^2 = \overline{\sigma}_{i,n}^2 = \sigma_i^2 = h_i(\theta_0)$, $\widehat{\sigma}_{i,n}^2 = h_i(\widehat{\theta}_n)$, and $\widehat{b}_{i,n} = \overline{b}_{i,n} = b_{i,n} = 0$. To apply Theorem 3 we have to verify Assumptions 4 and 8. Assumption 4 is postulated. Next consider Assumption 8(a). Since $w_{ii,n} = 0$ we can ignore all conditions involving third and fourth moments. To prove $|\widehat{\sigma}_{i,n}^2 - \sigma_{i,n}^2| \leq \phi_n$ with $\phi_n = o_p(1)$ it suffices to prove Condition A, as was discussed after the theorem. Clearly $\partial h_i(\theta)/\partial \theta$ exists and is continuous on Θ_* . Again, since continuous functions are bounded in absolute value on compact sets, there exists a finite constant, say K_* , such that

$$\sup_{1 \leq i \leq n, n \geq 1} \sup_{\theta \in \Theta} \left| \frac{\partial h_i(\theta)}{\partial \theta} \right| \leq \sup_{\underline{x} \in \overline{X}} \sup_{\theta \in \Theta} \left| \frac{\partial h(\underline{x}, \theta)}{\partial \theta} \right| < K_*,$$

where $h(x_i, \theta) = h_i(\theta)$. Since $n^{1/2}(\widehat{\theta}_n - \theta_0) = O_p(1)$ this verifies Condition A. Assumption 8(b) holds trivially.

Having demonstrated that all conditions maintained by Theorem 2(a) are satisfied the claim (4.1) for the Tobit model follows directly from that theorem. The proofs of claim (4.1) for the dichotomous, sample selection, and polychotomous models are analogous. \blacksquare

Proof of Claims in Section 4.2. In the following let $\overline{y}_n = M_n y_n$. To prove the claims we use Theorems 2 and 3. To cast the spatial ARAR(1,1) into the notation of those theorems define $g_n(z_n, \theta) = y_n - D_n \theta$ where $D_n = [\overline{y}_n, X_n]$, $\theta = (\lambda, \beta')'$, and $z_n = (z_{1,n}, \dots, z_{n,n})$ with $z_{i,n} = (y_{i,n}, x_{i1,n}, \dots, x_{ik,n})$. The parameter space $\Theta = (-1, 1) \times \mathbb{R}^k$. Furthermore we take $Z = \mathbb{R}^{k+1}$ as the space for the $z_{i,n}$.

We now check the Assumptions 4-6. Assumption 4 is assumed to hold. Assumption 5 also holds since Θ is open. To verify Assumption 6(a) take $\Theta_* = \Theta$. The measurability and differentiability assumptions hold trivially since $g_{i,n}$ is continuous in all arguments and linear in θ , respectively. To verify Assumption 6(b) observe that $\Delta_{i,n}^2 = |\partial g_{i,n}/\partial \theta|^2 = \bar{y}_{i,n}^2 + \sum_{j=1}^k x_{ij,n}^2$. From the first result in (A.4) it follows that $E\bar{y}_{i,n}^2 \leq \text{const} < \infty$. Since the $x_{ij,n}$ are assumed to be uniformly bounded in absolute value Assumption 6(b) clearly holds.

From the discussion surrounding (2.5) - (2.7) it is readily seen that condition (3.7) of Assumption 7 holds with $A_n = \frac{1}{2}(I_n - \rho_0 M'_n)^{-1}(W_n + W'_n)(I_n - \rho_0 M_n)^{-1}$ and $b_n = -F_n P' d_n$ with $F_n = (I_n - \rho_0 M'_n)^{-1} H_n$. Furthermore it follows from that discussion that the elements of A_n and b_n satisfy Assumption 2. Assumption 1 and 3(b) are postulated in the catalogue of model assumptions. Thus also all conditions of Assumption 7 hold.

Next define

$$\bar{\sigma}_{Q_n}^2 = \frac{1}{2} \bar{\sigma}_n^4 \sum_{i=1}^n \sum_{j=1}^n (w_{ij,n} + w_{ji,n})^2 + \bar{\sigma}_n^2 \sum_{i=1}^n \bar{b}_{i,n}^2$$

where $\bar{\sigma}_n^2 = n^{-1} E(u'_n u_n) = n^{-1} E(\varepsilon'_n C_n^* \varepsilon_n) = \sigma^2 n^{-1} \text{tr}(C_n^*)$ with $C_n^* = (I_n - \rho_0 M_n)^{-1} (I_n - \rho_0 M'_n)^{-1}$, and where $\bar{b}_n = -H_n P' d_n$. Furthermore let $\hat{\sigma}_{Q_n}^2 = \hat{\sigma}_{Q_n^*}^2$. Clearly $\hat{\sigma}_n^2 = n^{-1} \hat{u}'_n \hat{u}_n = n^{-1} u'_n u_n + o_p(1)$, since $\hat{u}_n = u_n - D_n(\hat{\theta}_n - \theta_0)$, $\hat{\theta}_n$ is $n^{1/2}$ -consistent, and given the results in (A.4). By analogous arguments as those used to establish (A.9) we have $\text{var}(n^{-1} u'_n u_n) = \text{var}(n^{-1} \varepsilon'_n C_n^* \varepsilon_n) = o(1)$. Hence $n^{-1} u'_n u_n - \bar{\sigma}_n^2 = o_p(1)$ and consequently $\hat{\sigma}_n^2 - \bar{\sigma}_n^2 = o_p(1)$. This verifies Assumption 8(a) for $\hat{\sigma}_n^2$ and $\bar{\sigma}_n^2$ (observing that $\bar{\sigma}_{i,n}^2 = \bar{\sigma}_n^2$ and $w_{ii,n} = 0$). Assumption 8(b) for $\hat{b}_{i,n}$ and $\bar{b}_{i,n}$ was verified at the end of Section 3.3. It now follows from Theorem 3 that $\hat{\sigma}_{Q_n}^2 - \bar{\sigma}_{Q_n}^2 = o_p(1)$ as postulated in Theorem 2.

The expressions for the mean μ_{Q_n} and variance $\sigma_{Q_n}^2$ of the linear-quadratic form $Q_n = \varepsilon'_n A_n \varepsilon_n + b'_n \varepsilon_n$ are given in (3.2). Now assume that $H_0 : \rho_0 = 0$ holds. In this case we have $A_n = \frac{1}{2}(W_n + W'_n)$ and hence $a_{ii,n} = 0$, $b_n = \bar{b}_n = -H_n P' d_n$ and $\sigma_{i,n}^2 = \bar{\sigma}_n^2 = \sigma^2$. Thus in this case we have $\mu_{Q_n} = 0$ and $\sigma_{Q_n}^2 = \bar{\sigma}_{Q_n}^2$. In light of (4.2) we also have $n^{-1} \sigma_{Q_n}^2 \geq c$ for some $c > 0$. The claim that (4.1) holds under H_0 now follows directly from part (a) of Theorem 2.

Next assume that $H_1 : \rho_0 \neq 0$ holds. Assume furthermore that $n^{-1} \sigma_{Q_n}^2 > 0$, $\lim_{n \rightarrow \infty} n^{-1} \sigma_{Q_n}^2 = s > 0$, $\lim_{n \rightarrow \infty} n^{-1} \bar{\sigma}_{Q_n}^2 = \bar{s} > 0$ and $\lim_{n \rightarrow \infty} n^{-1} \text{tr}(A_n) =$

$a \neq 0$. Then clearly $n^{-1}\sigma_{Q_n}^2 \geq c$ for some $c > 0$ and $\lim_{n \rightarrow \infty} \sigma_{Q_n}^2 / \bar{\sigma}_{Q_n}^2 > 0$. Furthermore $|n^{-1/2}\mu_{Q_n}| = n^{1/2}|n^{-1}\text{tr}(A_n)| \rightarrow \infty$ for $n \rightarrow \infty$. The claim that $\lim_{n \rightarrow \infty} P(|I_n| > \tau) = 0$ for arbitrary $\tau > 0$ now follows from part (b) of Theorem 2. ■

References

- [1] Amemiya, T., 1985, *Advanced econometrics* (Harvard University Press, Cambridge).
- [2] Anselin, L., 1988, *Spatial econometrics: Methods and models* (Kluwer Academic Publishers, Boston).
- [3] Anselin, L., A. Bera, R. Florax, and M. Yoon, 1996, Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics* 26, 77-104.
- [4] Anselin, L. and R. Florax, 1995, Small sample properties of tests for spatial dependence in regression models: Some further results, in L. Anselin and R. Florax, eds., *New directions in spatial econometrics*, (Springer Verlag, New York) 21-74.
- [5] Anselin, A. and H. Kelejian, 1997, Testing for spatial autocorrelation in the presence of endogenous regressors, *International Regional Science Review* 20, 153-182.
- [6] Audretsch, D. and M. Feldmann, 1996, R&D spillovers and the geography of innovation and production, *American Economic Review* 86, 630-640.
- [7] Ausubel, L., P. Cramton, R. McAfee and J. McMillan, 1997, Synergies in wireless telephony: Evidence from broadcast PCS, *Journal of Economics and Management Strategy* 6, 497-527.
- [8] Bennett, R., and L. Hordijk, 1986, Regional econometric and dynamic models, in: P. Nijkamp, ed., *Handbook of regional and urban economics*, Vol. 1 (North Holland, Amsterdam) 407-441.
- [9] Beran, J., 1972, Rank spectral processes and tests for serial dependence, *Annals of Mathematical Statistics* 43, 1749-1766.
- [10] Besley, T. and A. Case, 1995, Incumbent behavior: Vote-Seeking, tax-setting, and yardstick competition, *American Economic Review* 85, 25-45.

- [11] Brown, B., 1971, Martingale central limit theorems, *Annals of Mathematical Statistics* 42, 59-66.
- [12] Burridge, P., 1980, On the Cliff-Ord test for spatial correlation, *Journal of the Royal Statistical Society B* 42, 107-108.
- [13] Case, A., 1991, Spatial patterns in household demand, *Econometrica* 59, 953-966.
- [14] Case, A., 1993, Interstate tax competition after TRA86, *Journal of Policy Analysis and Management* 12, 136-148.
- [15] Case, A., J. Hines Jr. and H. Rosen, 1993, Budget spillovers and fiscal policy independence: Evidence from the states, *Journal of Public Economics* 52, 285-307.
- [16] Cliff, A. and J. Ord., 1972, Testing for spatial autocorrelation among regression residuals, *Geographic Analysis* 4, 267-284.
- [17] Cliff, A. and J. Ord., 1973, *Spatial autocorrelation* (Pion, London).
- [18] Cliff, A. and J. Ord., 1981, *Spatial processes, models and applications* (Pion, London).
- [19] Conley, T., 1996, *Econometric modelling of cross sectional dependence*, Ph.D. Thesis (University of Chicago, Chicago)
- [20] Cressie, N., 1993, *Statistics of spatial data* (Wiley, New York).
- [21] Davidson, J., 1994, *Stochastic limit theory* (Oxford University Press, Oxford).
- [22] De Long, J. and L. Summers, 1991, Equipment investment and economic growth, *Quarterly Journal of Economics* 106, 445-502.
- [23] Dhrymes, P., 1978, *Introductory econometrics* (Springer Verlag, New York).
- [24] Dubin, R., 1988, Estimation of regression coefficients in the presence of spatially autocorrelated error terms, *Review of Economics and Statistics* 70, 466-474.

- [25] Gänslér, P. and W. Stute, 1977, *Wahrscheinlichkeitstheorie* (Springer Verlag, New York).
- [26] Giraitis, L. and M. Taqqu, 1998, Central limit theorems for quadratic forms with time-domain conditions, *Annals of Probability* 26, 377-398.
- [27] Green, W., 1997, *Econometric analysis* (Prentice Hall, Upper Saddle River).
- [28] Hall, P. and C. Heyde, 1980, *Martingale limit theory and its applications* (Academic Press, New York).
- [29] Holtz-Eakin, D., 1994, Public sector capital and the productivity puzzle, *Review of Economics and Statistics* 76, 12-21.
- [30] Kelejian, H. and I. Prucha, 1998, A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics* 17, 99-121.
- [31] Kelejian, H. and I. Prucha, 1999, A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* 40, 509-533.
- [32] Krugman, P., 1991, *Geography and trade* (MIT Press, Cambridge).
- [33] Krugman, P., 1995, *Development, geography, and economic theory* (MIT Press, Cambridge).
- [34] Lee, L., 1999a, Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances, *Department of Economics, Hong Kong University of Science and Technology*.
- [35] Lee, L., 1999b, Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models, *Department of Economics, Hong Kong University of Science and Technology*.
- [36] Lee, L., 1999c, Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models, *Department of Economics, Hong Kong University of Science and Technology*.

- [37] Maddala, G., 1983, *Limited-dependent and qualitative variables in econometrics* (Cambridge University Press, Cambridge).
- [38] McLeish, D., 1974, Dependent central limit theorems and invariance principles, *Annals of Probability* 2, 620-628.
- [39] Moran, P., 1950, Notes on continuous stochastic phenomena, *Biometrika* 37, 17-23.
- [40] Moulton, B., 1990, An illustration of a pitfall in estimating the effects of aggregate variables on micro units, *Review of Economics and Statistics* 72, 334-338.
- [41] Pinkse, J., 1999, Asymptotic properties of Moran and related tests and testing for spatial correlation in probit models, *Department of Economics, University of British Columbia and University College London*.
- [42] Pinkse, J., and M. Slade, 1998, Contracting in space: An application of spatial statistics to discrete-choice models, *Journal of Econometrics* 85, 125-154.
- [43] Pötscher, B. and I. Prucha, 1997, *Dynamic nonlinear econometric models, asymptotic theory* (Springer Verlag, New York).
- [44] Sen, A., 1976, Large sample-size distribution of statistics used in testing for spatial correlation, *Geographical Analysis* 9, 175-184.
- [45] Shroder, M., 1995, Games the states don't play: Welfare benefits and the theory of fiscal federalism, *Review of Economics and Statistics* 77, 183-191.
- [46] Topa, G., 1996, *Social interactions, local spillovers, and unemployment*, Ph.D. thesis (University of Chicago, Chicago).
- [47] Whittle, P., 1954, On stationary processes in the plane, *Biometrika* 41, 434-449.
- [48] Whittle, P., 1964, On the convergence to normality of quadratic forms of independent variables, *Theory of Probability and Its Applications* 9, 103-108.