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

Harry H. Kelejian and Ingmar R. Prucha*

February 1999

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 statistic. We then apply this result to derive the large sample distribution of the Moran $I$ test statistic for a variety of important models for which general spatial correlation testing procedures are not available. In order to establish these results we also give a new central limit theorem for linear-quadratic forms.

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

*Department of Economics, University of Maryland, College Park, MD 20852
1 Introduction

Spatial models have a long history in the regional science and geography literature. In recent years issues concerning spatial dependence between cross sectional units have also received increasing attention in the economics literature. 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. The matrix appearing in the quadratic form is typically referred to as the spatial weights 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 original specification and not Cliff and Ord’s generalized specification.

The conditions maintained by the above results concerning the large sam-

---

1We would like to thank Dilip Madan and Benedikt Pötscher for helpful comments.
4The generalization contains Moran’s original specification as a special case corresponding to a model where the only regressor relates to the intercept.
ple distribution of the Moran $I$ test statistic are restrictive and rule out many potential applications. In particular, the results do not allow for heteroskedasticity in the innovations. As will become evident, this rules out many applications which relate to qualitative and limited dependent variable models, even though such models have become of increasing interest in the analysis of spatial data sets. Furthermore the above results do not cover 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. Available results concerning the large sample distribution of the Moran $I$ statistic also do not cover nonlinear models. Furthermore, they have only been derived under the null hypothesis of zero spatial correlation, but not under the alternative hypothesis, which would be needed in order to be able to assess the power of the test. Also, the available results do not account for the triangular nature of the data even though, as illustrated below, this is implied by the specifications of many models considered in practice.

One interpretation of the Moran $I$ test statistic was given by Burridge (1980) who demonstrated its equivalence with the Lagrange multiplier test statistic derived from a model without a spatial lag. Of course, other tests for spatial correlation have been considered. Most of these tests are close cousins to the Moran $I$ test and are 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).

The purpose of this paper is two-fold. First, on a theoretical level, we derive a set of general results concerning the asymptotic distribution of quadratic forms based on estimated residuals. Our 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 quadratic forms. Second, we apply our results and formally establish the limiting behavior of the Moran $I$ test statistic for spatial correlation in a wide variety of models. These models include a variety of limited dependent variable models, a sample section model, and linear cross sectional models which contain a spatially lagged dependent variable. Our results are such that applications to still other models of interest will become evident. Our
formal demonstrations are based on sets of low level assumptions.\footnote{The general results could also be used to establish the limiting distribution of the above mentioned variants of Largange multiplier tests for spatial correlation. We note, however, that the general results may also be of interest in other, not necessarily spatial, contexts.}

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

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 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 $n \in \mathbb{N}$ 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 $P_n$ are 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 $|P_n| = [\text{tr}(P_n^* P_n)]^{1/2}$ where $\text{tr}(\cdot)$ denoted 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 \( P_n \). Furthermore, we say the row and column sums of the (sequence of \( n \times n \)) matrices \( P_n \) are bounded in absolute value if

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

As a point of interest, we note that the above condition is identical to the condition that the sequences of the maximum column sum matrix norms and maximum row sum matrix norms of \( P_n \) are bounded; cp. Horn and Johnson (1985, pp.294-5). Finally, throughout the paper all random variables are assumed to be defined on some probability space \((\Omega, \mathcal{F}, P)\).

## 2 Testing for Spatial Autocorrelation: A Preliminary Discussion

As remarked in the introduction, we will provide general results concerning the limiting distribution of linear-quadratic forms, which will be useful in establishing the limiting distribution of the Moran \( I \), as well as various Lagrangian multiplier test statistics for spatial autocorrelation for a wide class of models. To motivate and put our results into context we start with a preliminary discussion of the Moran \( I \) test statistic within the context of the following spatial model \((n \in \mathbb{N})\):\(^7\)

\[
y_n = \lambda_0 M_n y_n + X_n \beta_0 + u_n, \quad u_n = \rho_0 M_n u_n + \varepsilon_n,
\]

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 matrices of known constants, \( \beta_0 \) is the \( k \times 1 \) vector of regression parameters, \( \lambda_0 \) and \( \rho_0 \) are scalar autoregressive parameters with \(|\lambda_0| < 1\) and \(|\rho_0| < 1\), \( u_n \) is the \( n \times 1 \) vector of regression disturbances, and \( \varepsilon_n \) is an \( n \times 1 \) vector of innovations. The variables \( M_n y_n \) and \( M_n u_n \) are typically

\(^7\)The model in (2.1) is of interest in and of itself, and hence our “motivational” discussion should also be of interest in and of themselves. It is for this reason that we specify the model in a complete manner. The results on the asymptotic properties of the Moran \( I \) test statistic for this model are summarized in Section 4.2.
referred to as spatial lags of $y_n$ and $u_n$, respectively.\(^8\) In the terminology of, e.g., Anselin (1988) the above model represents a first order autoregressive spatial model with first order autoregressive disturbances – for short a spatial ARAR(1,1) model. Since spatial weights 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 $\varepsilon_n$ to depend on $n$. We condition our analysis on the realized values of the exogenous variables and so, henceforth, the matrices $X_n$ will be viewed as matrices of constants. 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.

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 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.\(^9\) This estimator is based on a generalized moments estimator for the spatial autoregressive parameter introduced earlier in Kelejian and Prucha (1995).

Assume the following conditions concerning the spatial model (2.1):\(^{10}\) All diagonal elements of the spatial weighting matrices $M_n$ are zero. The matrices $(I - \lambda_0 M_n)^{-1}$ and $(I - \rho_0 M_n)^{-1}$ 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 bounded in absolute value. The regressor matrices $X_n$ have full column rank (for $n$ large enough), and the elements of $X_n$ are 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.

---

\(^8\)We note that the subsequent discussion trivially extends to the case where we allow for different spatial weights matrices for $y_n$ and $u_n$. Our specification of a common spatial weights matrix was solely adopted for reasons of notational simplicity.

\(^9\)We note that since $E(M_n y_n u_n') \neq 0$ the model in (2.1) cannot be estimated consistently by ordinary least squares.

\(^{10}\)For a discussion of these assumptions see, e.g., Kelejian and Prucha (1995, 1998).
with $E(\varepsilon_{i,n}) = 0$, $E(\varepsilon_{i,n}^2) = \sigma^2$ and finite $4 + \eta$ moments (for some $\eta > 0$). Now consider, for example, the following instrumental variable estimator for $\theta_0 = (\lambda_0, \beta_0)'$:

$$\hat{\theta}_n = (\hat{\lambda}_n, \hat{\beta}_n)' = (\tilde{D}_n')^{-1}\tilde{D}_ny_n$$  \hspace{1cm} (2.2)

with $\tilde{D}_n = H_n(H_n' H_n)^{-1}H_n' D_n$ 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 bounded in absolute value. It then follows from that paper that

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

where $P_n = (n^{-1}\tilde{D}_n' \tilde{D}_n)^{-1}n^{-1}D_n' H_n (n^{-1}H_n' H_n)^{-1}$ with $p \lim_{n \to \infty} P_n = P$ finite, and $F_n' = H_n'(I - \rho_0 M_n)^{-1}$ with $n^{-1/2}F_n' \varepsilon_n \overset{D}{\to} N(0, \sigma^2 \Phi_F')$ with $\Phi_F = \lim_{n \to \infty} n^{-1}F_n' F_n$. 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 weights matrix $W_n$ is given by

$$I_n = \frac{Q_n^*}{\sigma_{Q_n}^*}, \hspace{1cm} Q_n^* = \hat{u}_n' W_n \hat{u}_n$$  \hspace{1cm} (2.4)

where $\sigma_{Q_n}$ is some normalizing factor. We allow for $W_n \neq M_n$ to allow for misspecification of the spatial weights 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

$$n^{-1/2}\hat{u}_n' W_n \hat{u}_n = n^{-1/2}u_n' W_n u_n$$  \hspace{1cm} (2.5)

$$+ n^{-1}u_n'(W_n + W_n')D_n n^{1/2}(\hat{\theta}_n - \theta_0) + o_p(1).$$

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 bounded in absolute value. Consequently we have from (2.3), observing that $P_n = P + o_p(1)$ and $n^{-1/2}F_n' \varepsilon_n = O_p(1)$:

$$n^{-1}u_n'(W_n + W_n')D_n n^{1/2}(\hat{\theta}_n - \theta_0)$$  \hspace{1cm} (2.6)

$$= n^{-1}u_n'(W_n + W_n')D_n \left[ n^{-1/2}P_n F_n' \varepsilon_n + o_p(1) \right]$$

$$= n^{-1/2}b_n' \varepsilon_n + o_p(1)$$

7
with \( b_n' = d_n'PF_n' \). Since, from (2.1), \( u_n = (I - \rho_0 M_n)^{-1} \varepsilon_n \) it follows that
\[
u_n'W_nu_n = \varepsilon_n'A_n\varepsilon_n
\]
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
\[
n^{-1/2}\tilde{u}_n'W_n\tilde{u}_n = n^{-1/2} [\varepsilon_n'A_n\varepsilon_n + b_n'\varepsilon_n] + o_p(1). \tag{2.7}
\]
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 bounded in absolute value, and that row and column sums of the \( n \times n \) matrices \( A_n \) are 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 then \( d_n = 0 \). 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 element 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 qualitative and 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:
\[
Q_n = \varepsilon_n'A_n\varepsilon_n + b_n'\varepsilon_n \tag{3.1}
\]
\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j,n} \varepsilon_{i,n} \varepsilon_{j,n} + \sum_{i=1}^{n} b_{i,n} \varepsilon_{i,n},
\]
8
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 \( \varepsilon_{1,n}, \ldots, \varepsilon_{n,n} \) are mutually stochastically 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 \( \sup_{1 \leq j \leq n, n \geq 1} \sum_{i=1}^{n} |a_{ij,n}| < \infty \). The elements of the array of real numbers \( \{ b_{i,n} : 1 \leq i \leq n, \ n \geq 1 \} \) satisfy \( \sup_{n} \sum_{i=1}^{n} \sum_{i=1}^{n} |b_{i,n}|^{2+\eta_1} < \infty \) 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 \( \mu_{Q_n} \), and the variance, say \( \sigma^2_{Q_n} \), of \( Q_n \) is given by:

\[
\begin{align*}
\mu_{Q_n} &= \sum_{i=1}^{n} a_{ii,n} \sigma^2_{i,n}, \\
\sigma^2_{Q_n} &= 4 \sum_{i=1}^{n} \sum_{j=1}^{i-1} a^2_{ij,n} \sigma^2_{i,n} \sigma^2_{j,n} + \sum_{i=1}^{n} b^2_{i,n} \sigma^2_{i,n} \\
&\quad + \sum_{i=1}^{n} \left\{ a^2_{ii,n} \left[ \mu_{i,n}^{(4)} - \sigma^4_{i,n} \right] + 2 b_{i,n} a_{ii,n} \mu_{i,n}^{(3)} \right\},
\end{align*}
\]

with \( \sigma^2_{i,n} = E(\varepsilon^2_{i,n}) \) and \( \mu_{i,n}^{(s)} = E(\varepsilon^s_{i,n}) \) 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 \( \sigma^2_{Q_n} \) are zero.

We now have the following central limit theorem for linear–quadratic forms:

\[\text{proof of Theorem 1 below.}\]

\[\text{We note that the variance of } Q_n \text{ can also be expressed as }\]

\[
\sigma^2_{Q_n} = 2 \sum_{i=1}^{n} \sum_{j=1}^{n} a^2_{ij,n} \sigma^2_{i,n} \sigma^2_{j,n} + \sum_{i=1}^{n} b^2_{i,n} \sigma^2_{i,n}
\]
**Theorem 1**: Suppose Assumptions 1-3 hold and \( n^{-1} \sigma_{Q_n}^2 \geq c \) for some \( c > 0 \). Then

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

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 qualitative and limited dependent variable models. Finally, we note that all of the maintained assumptions 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.\(^{14}\)

The existing literature on the limiting distribution of quadratic forms includes papers by Whittle (1964), Beran (1972), Sen (1976) and Giraitis and Taqq (1998). In contrasting those results with the above theorem we note that the papers by Whittle, Beran, and Giraitis and Taqq do not allow the weights in the quadratic form to depend on the sample size. The latter two papers actually assume that \( a_{ij,n} = a_{j-i} \). 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 Taqq assume that the \( \varepsilon_{i,n} = \varepsilon_i \) are products of linear process with i.i.d. innovations.

\[
\sigma_{Q_n}^2 = 2\text{tr}(A_n' \Sigma_n A_n \Sigma_n) + b_n' \Sigma_n b_n \text{ with } \Sigma_n = \text{diag}(\sigma_{i,n}^2).
\]

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 \( \Sigma_n = \text{diag}(\sigma_{i,n}^2) \).

\(^{14}\)We 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 \( n \to \infty \)).
3.2 Asymptotic Distribution of Quadratic Forms Based on Estimated Disturbances

The above illustrative example was given in terms of a spatial ARAR(1,1) model. For the subsequent discussion we consider more generally the following (possibly nonlinear) model:

\[ g_{i,n}(z_n, \theta_0) = u_{i,n}, \quad i = 1, \ldots, n, \]

with \( z_n = (z_{1,n}, \ldots, z_{n,n}) \) and where \( z_{i,n} \) denotes the 1× \( p_z \) 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 \( \theta_0 \) denotes the \( p_\theta \times 1 \) true parameter vector.\(^{15}\) We emphasize that in allowing for the \( g_{i,n} \) to depend on \( z_{1,n}, \ldots, z_{n,n} \) the specification allows for spatial lags. In the following we will write (3.4) more compactly as \( g_n(z_n, \theta_0) = u_n \) with \( g_n(z_n, \theta) = (g_{1,n}(z_{1,n}, \theta), \ldots, g_{n,n}(z_{n,n}, \theta))^\prime \) and \( u_n = (u_{1,n}, \ldots, u_{n,n}) \).

In this case, the Moran I test statistic would be based on the quadratic form \( \tilde{u}_n^\prime W_n \tilde{u}_n \) where \( \tilde{u}_n = g_n(z_n, \hat{\theta}_n) \) denotes the vector of estimated residuals based on some estimator \( \hat{\theta}_n \). In the following we provide results concerning the asymptotic distribution of quadratic forms based on estimated disturbances. 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 bounded in absolute value.\(^{16}\)

**Assumption 5** The true parameter vector \( \theta_0 \) 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

\[ g_{i,n} : Z^n \times \Theta_* \to \mathbb{R}, \quad 1 \leq i \leq n, n \geq 1, \]

are Borel measurable for each \( \theta \in \Theta_* \), where \( \Theta \subseteq \Theta_* \) and \( \Theta_* \) is an open subset. Furthermore the functions \( g_{i,n}(z_1, \ldots, z_n) \) are twice continuously partially differentiable on \( \Theta_* \) for each \( (z_1, \ldots, z_n) \in Z^n \).

\(^{15}\)For the spatial ARAR(1,1) model we have \( z_{i,n} = (y_{i,n}, x_{i1,n}, \ldots, x_{iK,n}) \) and \( \theta_0 = (\lambda_0, \beta_0^\prime)^\prime \).

\(^{16}\)We note that in this section we do not assume that the diagonal elements of \( W_n \) are zero.
(b) If \( g_{i,n}(z_n, \theta) \) is linear in \( \theta \), let

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

otherwise let

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

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

We now have the following lemma concerning the asymptotic relationship between \( n^{-1/2}\tilde{u}_n W_n \hat{u}_n \) 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 \( \theta_0 \) with \( n^{1/2}(\hat{\theta}_n - \theta_0) = O_p(1) \). Then (abusing notation in an obvious way)

\[
n^{-1/2}\tilde{u}_n' W_n \hat{u}_n = n^{-1/2}u_n' W_n u_n \tag{3.5}
\]

\[
+ n^{-1}u_n'(W_n + W_n') \frac{\partial g_n(z_n, \theta_0)}{\partial \theta} n^{1/2} (\hat{\theta}_n - \theta_0) + o_p(1).
\]

Furthermore

\[
n^{-1} \tilde{u}_n'(W_n + W_n') \frac{\partial g_n(z_n, \hat{\theta}_n)}{\partial \theta} - n^{-1}u_n'(W_n + W_n') \frac{\partial g_n(z_n, \theta_0)}{\partial \theta} = o_p(1). \tag{3.6}
\]

The above lemma shows that in general the limiting distribution of \( n^{-1/2}\tilde{u}_n' W_n \hat{u}_n \) 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 \( (u_{i,n}, \theta g_{i,n}(z_n, \theta_0)/\partial \theta) \) is distributed independently over \( i \). In terms of 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 in this illustration 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 $\varepsilon_n$, where $\varepsilon_n$ is the basic vector of innovations. For example, in the illustrative 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 above spatial ARAR(1,1) model, $n^{1/2}(\hat{\theta}_n - \theta_0)$ will be (asymptotically) linear in $\varepsilon_n$. This motivates the following assumption.

**Assumption 7** There exists a linear-quadratic form $Q_n = \varepsilon_n' A_n \varepsilon_n + \varepsilon_n' b_n \varepsilon_n$ such that

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

and where the elements of $\varepsilon_n$, $A_n$ and $b_n$ satisfy Assumptions 1-3.

For the illustrative spatial ARAR(1,1) model the above assumption is clearly satisfied in light of (2.6). If the probability limit of the term $n^{-1} u_n' (W_n + W'_n) \frac{\partial g_n(z_n, \theta_0)}{\partial \theta}$ 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 illustrative spatial ARAR(1,1) model this was established in (2.7).

As before, let $\mu_{Q_n}$ and $\sigma^2_{Q_n}$ denote the mean and variance of the linear-quadratic form $Q_n$, and recall the explicit expressions for $\mu_{Q_n}$ and $\sigma^2_{Q_n}$ given in (3.2). We next give a result concerning the asymptotic behavior of $\hat{u}_n' W_n \hat{u}_n/\sigma_{Q_n}$, where $\hat{\sigma}^2_{Q_n}$ denotes some estimator for $\sigma^2_{Q_n}$. We note that in particular applications $\hat{\sigma}^2_{Q_n}$ 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 $\hat{\sigma}^2_{Q_n}$ to be (possibly) inconsistent.

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

\[
\frac{\hat{u}_n' W_n \hat{u}_n}{\sigma_{Q_n}} = \frac{Q_n - \mu_{Q_n}}{\sigma_{Q_n}} + \frac{\mu_{Q_n}}{\sigma_{Q_n}} + o_p(1).
\]

(3.8)

Consequently:

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

\[
\frac{\hat{u}_n' W_n \hat{u}_n}{\sigma_{Q_n}} \xrightarrow{D} N(0, \Lambda^2).
\]

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

\[
\lim_{n \to \infty} \mathbb{P} \left( \left| \frac{\hat{u}_n' W_n \hat{u}_n}{\sigma_{Q_n}} \right| > \delta \right) = 1.
\]

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 \( \hat{\sigma}^2_{Q_n} \) is a consistent estimator then \( \Lambda = 1 \). Under the alternative 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 \( n \to \infty \), for an arbitrary significance level of the test. The condition \( |n^{-1/2} \mu_{Q_n}| \to \infty \) is, e.g., satisfied if \( 0 < \text{const} \leq \sigma^2_{i,n} \) 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 \( \sigma^2_{Q_n} \) 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 cases considered in the next section, we have \( A_n = (W_n + W'_n)/2 \) under the null hypothesis of zero spatial autocorrelation. Let \( \hat{\sigma}^2_{i,n} \) and \( \hat{\mu}_{i,n} \) denote estimators for \( \sigma^2_{i,n} \) and \( \mu_{i,n} \) 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 $\sigma^2_{Q_n}$ given in (3.2), it is natural to consider the following estimator:

$$
\hat{\sigma}^2_{Q_n} = \sum_{i=1}^{n} \sum_{j=1}^{n-1} (w_{ij,n} + w_{ji,n})^2 \hat{\sigma}_{i,j,n}^2 + \sum_{i=1}^{n} \hat{b}_{i,n} \hat{\sigma}_{i,n}^2 
$$

$$
+ \sum_{i=1}^{n} \left\{ w_{ii,n}^2 \left[ \hat{\mu}_{i,n}^{(4)} - \hat{\sigma}_{i,n}^4 \right] + 2\hat{b}_{i,n} w_{ii,n} \hat{\mu}_{i,n}^{(3)} \right\}.
$$

Of course, in many contexts we will have $w_{ii,n} = 0$, but we do not postulate this here for reasons of generality.\footnote{We 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. Let $\sigma_{i,n}^2$, $\mu_{i,n}^{(s)}$ ($s = 3, 4$) and $\hat{b}_{i,n}$ denote the “nonstochastic analogues” of those estimators, and let

$$
\sigma^2_{Q_n} = \sum_{i=1}^{n} \sum_{j=1}^{n-1} (w_{ij,n} + w_{ji,n})^2 \sigma_{i,j,n}^2 + \sum_{i=1}^{n} \hat{b}_{i,n} \sigma_{i,n}^2 
$$

$$
+ \sum_{i=1}^{n} \left\{ w_{ii,n}^2 \left[ \mu_{i,n}^{(4)} - \sigma_{i,n}^4 \right] + 2\hat{b}_{i,n} w_{ii,n} \mu_{i,n}^{(3)} \right\},
$$

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

**Assumption 8**

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

$$
\left| \hat{\sigma}_{i,n}^2 - \sigma_{i,n}^2 \right| \leq \phi_n \quad \text{and} \quad \left| \hat{\mu}_{i,n}^{(s)} - \mu_{i,n}^{(s)} \right| \leq \phi_n,
$$

where $\phi_n = o_p(1)$, and $\sup_{1 \leq i \leq n, n \geq 1} \sigma_{i,n}^2 < \infty$ and $\sup_{1 \leq i \leq n, n \geq 1} \left| \mu_{i,n}^{(s)} \right| < \infty$ ($s = 3, 4$).

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

$$
\left| \hat{b}_{i,n} - \hat{b}_{i,n} \right| \leq \Psi_{i,n} \psi_n,
$$

where $\psi_n = o_p(1)$, the $\Psi_{i,n} \in \mathbb{R}$ satisfy $\sup_{n \geq 1} n^{-1} \sum_{i=1}^{n} E \Psi_{i,n}^2 < \infty$ and $\sup_{n \geq 1} n^{-1} \sum_{i=1}^{n} \left| \hat{b}_{i,n} \right|^2 < \infty$.\footnote{The use of the notation $o_p(1)$ indicates that the term converges in probability to zero as $n \to \infty$.}
We now have the following theorem concerning the stochastic convergence of $\hat{\sigma}^2_{Q_n}$.

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

Of course, if $(w_{ij,n} + w_{ji,n})/2 = a_{ij,n}$ and $\bar{b}_{i,n} = b_{i,n}$, and if $\sigma^2_{i,n} = \sigma^2_{i,n}$, $\bar{\mu}^{(s)}_{i,n} = \mu^{(s)}_{i,n}$ ($s = 3, 4$), then $\sigma^2_{Q_n} = \sigma^2_{Q_n}$ and the above theorem establishes the consistency of $\hat{\sigma}^2_{Q_n}$ for $\sigma^2_{Q_n}$. We note that in this leading case the boundedness conditions maintained in Assumption 8 concerning $\bar{b}_{i,n}$ and $\sigma^2_{i,n}$, $\bar{\mu}^{(s)}_{i,n}$ ($s = 3, 4$) 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**: $\sigma^2_{i,n} = h_{i,n}(\theta_0)$ where the functions $h_{i,n} : \Theta_n \to \mathbb{R}_+$ are continuously partially differentiable,

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

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

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

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 $F_n = (I_n - \rho_0 M'_n)^{-1} H_n$. 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 \( \gamma_n = P'_d d_n \), then \( \hat{b}_{i,n} - \bar{b}_{i,n} = \sum_{r=1}^{p} h_{i,r,n} (\hat{\gamma}_{r,n} - \gamma_{r,n}) \). Taking \( \Psi_{i,n} = \sum_{r=1}^{p} |h_{i,r,n}| \) and \( \psi_n = \sum_{r=1}^{n} |\hat{\gamma}_{r,n} - \gamma_{r,n}| \), we have \( |\hat{b}_{i,n} - \bar{b}_{i,n}| \leq \Psi_{i,n} \psi_n \). Since the \( h_{i,r,n} \) are 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 Applications: Tests for Spatial Autocorrelation

In the following we apply 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, a sample selection model, and for the spatial ARAR(1,1) model. In particular, we will establish for each of these models that under the null hypothesis of zero spatial autocorrelation

\[ I_n = \frac{Q^*_n}{\sigma Q^*_n} \overset{D}{\longrightarrow} N(0,1) \]  \hspace{1cm} (4.1)

where \( Q^*_n = \hat{u}'_n W_n \hat{u}_n \) and \( \sigma Q^*_n \) is a normalization factor. The specification of the vector of estimated disturbances \( \hat{u}_n \) and the normalization factor \( \sigma Q^*_n \) will be model specific. Under the null hypothesis of zero spatial autocorrelation we have \( u_n = \varepsilon_n \) and hence we will also write \( Q^*_n = \varepsilon'_n W_n \varepsilon_n \).

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

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

Clearly, a sufficient condition for (4.2) is that the elements of \( W_n \) are nonnegative, each row has at least one nonzero element, and the nonzero elements of \( W_n \) are bounded away from zero. In the following we will also assume that all non-stochastic regressors are bounded in absolute value. Furthermore, we maintain Assumption 5 concerning the parameter space \( \Theta \) and the true
parameter vector \( \theta_0 \). 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.\(^{18}\)

As will become clear, for limited dependent variable models and sample selection models the innovations entering the Moran \( I \) test statistic will typically be heteroskedastic. We note that, as was discussed in the Introduction, results available in the literature concerning the limiting distribution of this test statistic assume that the innovations are homoskedastic, and hence do not cover these models.\(^{19}\)

### 4.1 Limited Dependent Variable Models

For each of the limited dependent variable models and the sample selection model considered below the normalization factor in the Moran \( I \) test statistic will be of the following general form:

\[
\hat{\sigma}_{\hat{Q}_n}^2 = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{ij,n} + w_{ji,n})^2 \hat{\sigma}_{i,n}^2 \hat{\sigma}_{j,n}^2
\]

\[= \text{tr} \left( W_n \tilde{\Sigma}_n W_n \tilde{\Sigma}_n + W_n \tilde{\Sigma}_n W_n \tilde{\Sigma}_n \right)\]

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

#### 4.1.1 The Tobit Model

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

\[
y_i^* = x_i \beta_0 + \nu_i,
\]

\[
y_i = \begin{cases} y_i^* & \text{if } y_i^* \geq 0 \\ 0 & \text{otherwise} \end{cases}
\]

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, \( \beta_0 \) is a \( k \times 1 \) vector of regression

\(^{18}\)We note that for nonlinear estimation problems the parameter space is typically taken to be compact; cp., e.g., Pötscher and Prucha (1997).

\(^{19}\)Anselin 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 weights matrix is band diagonal. Clearly, this assumption is restrictive and mostly covers situations where the data expand “bandwise” in one direction.
parameters, and the $\nu_i$'s are i.i.d. $N(0, \sigma_{\nu_i}^2)$ with $0 < c_\sigma < \sigma_{\nu_i} < C_\sigma < \infty$. The model is furthermore assumed to satisfy sufficient regularity conditions such that the maximum likelihood estimators of $\theta_0 = (\beta_0^*, \sigma_{\nu_i}^2)^\prime$, say $\hat{\theta}_n = (\beta_n^*, \sigma_{\nu_i}^2)^\prime$, is $n^{1/2}$-consistent.\(^{20}\)

Let $\Phi_N(.)$ and $\phi_N(.)$ 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

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

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

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

Assume that $\sigma_i^2 = h_i(\theta_0) \geq \text{const} > 0$ and let $\bar{\varepsilon}_{i, n} = y_i - f(x_i, \hat{\theta}_n)$ 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 (4.1) holds. The result in (4.1) can now be used in the usual way to test for the presence of spatial autocorrelation.

### 4.1.2 A Dichotomous Model

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

$$y_i^* = x_i \beta_0 + \nu_i, \quad (4.8)$$

$$y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{otherwise} \end{cases},$$

where $y_i^*$ is a latent variable and $y_i$ is the observed binary dependent variable. The vectors $x_i$ and $\beta_0$ are as in Section 4.1, but $\nu_i$ is only assumed to be i.i.d. with zero mean and c.d.f. $\Phi(.)$. We assume furthermore that the distribution is symmetric around zero, and thus $\Pr(y_i = 1) = \Phi(x_i \beta_0)$ and

\(^{20}\)Actually, for the subsequent result $\hat{\theta}_n$ could be any $n^{1/2}$-consistent estimators.
Pr(y_i = 0) = 1 − Φ(x_iβ_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

\[
y_i = \Phi(x_i\beta_0) + \varepsilon_i
\]

where \( E(\varepsilon_i) = 0 \) and \( E(\varepsilon_i^2) = \Phi(x_i\beta_0)[1 - \Phi(x_i\beta_0)] \). Assume that \( E(\varepsilon_i^2) \geq const > 0 \) and let \( \hat{\varepsilon}_i = y_i - \Phi(x_i\hat{\beta}_n) \) and \( \hat{\sigma}^2_{i,n} = \Phi(x_i\hat{\beta}_n)[1 - \Phi(x_i\hat{\beta}_n)] \). In the appendix we demonstrate that under the above assumptions (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):

\[
\begin{align*}
y^*_1 &= x_i\beta_{10} + \nu_{1i} \quad (4.10) \\
y^*_2 &= x_i\beta_{20} + \nu_{2i} \\
y_{2i} &= \begin{cases} y^*_2 & \text{if } y^*_1 > 0 \\ 0 & \text{if } y^*_1 \leq 0 \end{cases}
\end{align*}
\]

where \( y^*_1 \) and \( y^*_2 \) 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, \( \beta_{10} \) and \( \beta_{20} \) are \( k_1 \times 1 \) and \( k_2 \times 1 \) parameter vectors, and \( (\nu_{1i}, \nu_{2i}) \) is i.i.d. as \( N(0, \Sigma_0) \) where \( \Sigma_0 = (\sigma_{ij0}) \) with \( 0 < c_0 < |\Sigma_0| < C_0 < \infty \). 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 \( y_{2i} \neq 0 \):

\[
y_{2i} = f(x_{1i}, x_{2i}, \theta_0) + \varepsilon_i
\]

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

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

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

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

Assume that \( \sigma^2_i = h_i(\theta_0) \geq \text{const} > 0 \) and let \( \hat{\varepsilon}_{i,n} = y_{2i} - f(x_{1i}, x_{2i}, \hat{\theta}_n) \) and \( \hat{\sigma}^2_{i,n} = h_i(\hat{\theta}_n) \). In the appendix we demonstrate that (4.1) holds under the above assumptions for the sample selection model. 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 \((i = 1, \ldots, n; j = 1, \ldots, m)\):

\[ \Pr(y_i = j) = P_j(x_i \beta_0) \]  (4.14)

where \( y_i \) denotes the observed endogenous variable, \( x_i \) is a non-stochastic \( 1 \times k \) vector of exogenous variables and \( \beta_0 \) is a \( k \times 1 \) vector of parameters. We assume that the functions \( P_j(.) \) posses bounded first and second order derivatives. We also assume that for all \( x_i \) and all vectors \( \beta \) in the parameter space \( 0 < P_j(x_i \beta) < 1 \) (\( j = 1, \ldots, m \)). 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 \( \beta_0 \), 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

\[ y_i = f(x_i \beta_0) + \varepsilon_i, \ i = 1, \ldots, n \]  (4.15)

with

\[ f(x_i \beta_0) = \sum_{j=1}^{m} j P_j(x_i \beta_0) \]  (4.16)

21
and where $E(\varepsilon_i) = 0$ and $\sigma_i^2 = E(\varepsilon_i^2) = h_i(\beta_0)$ with

$$h_i(\beta_0) = \sum_{j=1}^{m} j^2 P_j(x_i \beta_0) - \left[ \sum_{j=1}^{m} j P_j(x_i \beta_0) \right]^2$$ (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 (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. 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 that 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:

$$\hat{\sigma}_{Q_n}^2 = (1/2) \hat{\sigma}_n^4 \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{ij,n} + w_{ji,n})^2 + \hat{\sigma}_n^2 \sum_{i=1}^{n} \hat{b}_{i,n}^2$$ (4.18)

$$= \hat{\sigma}_n^4 \text{tr} (W_n W_n' + W_n' W_n) + \hat{\sigma}_n^2 \hat{\nu}_n \hat{\nu}_n$$

where $\hat{\sigma}_n^2 = n^{-1} \hat{\omega}_n \hat{\omega}_n$ and $\hat{b}_n = H_n P_n' \hat{d}_n$ with $\hat{d}_n = n^{-1} \hat{\omega}_n (W_n + W_n') D_n$. In the appendix we demonstrate that (4.1) 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 $\lim_{n \to \infty} P(|I_n| > \tau) = 1$ for all $\tau > 0$. That is the test is consistent, in that under the alternative hypothesis the power of the test tends to unity.\(^{22}\)

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

In this paper we give a general results concerning the large sample distribution of Moran $I$ (type) test statistics. The result allows for heteroskedastic and non-normally distributed innovations, spatial lags of the dependent variable, nonlinearities, and the possibly triangular nature of the data. We then apply this result to formally establish the asymptotic distribution of the Moran $I$ test statistic for spatial correlation in a variety of models. These include qualitative and limited dependent variable models, a sample selection model, and a linear cross sectional model which contains a spatially lagged dependent variable. Other applications of our results are evident. In order to establish these results it was necessary to derive a (to the best of our knowledge) new central limit theorem for linear-quadratic forms.

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 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 qualitative and limited dependent variable models and sample selection models, since there are no Monte Carlo results in the literature relating to the Moran $I$ test statistic for such models.\footnote{For 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.} 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.
A Appendix\textsuperscript{24}

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 \( E(X_{i,n} \mid \mathfrak{F}_{i-1,n}) = 0 \) (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 \( \epsilon > 0 \)

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

and

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

Then \( \sum_{i=1}^{k_n} X_{i,n}^2 \xrightarrow{D} N(0, 1) \).

** Remark.** The above theorem is given in Gänssler 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

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

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

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

\textsuperscript{24}The 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:

\[
E |y_{t,n}|^3 \leq \text{const} < \infty, \quad (A.4)
\]

\[
n^{-1}u'_n D_n = O_p(1),
\]

\[
n^{-1}u'_n(W_n + W'_n)D_n = O_p(1),
\]

\[
n^{-1}D'_n D_n = O_p(1),
\]

\[
n^{-1}D'_n W_n D_n = O_p(1),
\]

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

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

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

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 (1995): First, if $S_n$ and $R_n$ are two $n \times n$ matrices whose row and column sums are bounded in absolute value, then so are the row and column sums of $S_nR_n$. Second, maintaining this notation, if $T_n$ is a conformable matrix whose elements are bounded in absolute value, then so are the elements of $T_nR_n$. Now recall that $D_n = [M_n y_n, X_n]$ and observe that because of (2.1) we have

\[
y_n = (I_n - \lambda_0 M_n)^{-1} X_n \beta_0 + (I_n - \lambda_0 M_n)^{-1}(I_n - \rho_0 M_n)^{-1} \varepsilon_n, \tag{A.8}
\]

\[
u_n = (I_n - \rho_0 M_n)^{-1} \varepsilon_n.
\]
In light of this and given the maintained assumptions it is readily seen that all components of \( n^{-1}u'(W_n + W_n')D_n \) are composed of terms of the form \( n^{-1}B_n^*\varepsilon_n \) and \( n^{-1}C_n^*\varepsilon_n \) where \( B_n^* \) is a non-stochastic vector whose elements are bounded in absolute value, and where \( C_n^* \) is a non-stochastic matrix whose row and column sums are bounded in absolute value. The claim in (A.6) now follows since \( E(n^{-1}C_n^*\varepsilon_n) = \sigma^2 n^{-1}tr(C_n^*) \). Next observe that

\[
\text{var}(n^{-1}B_n^*\varepsilon_n) = \sigma^2 n^{-2}B_n^*C_n^*B_n^* = o(1). 
\]

Using the results of Section 3.1 we have

\[
\text{var}(n^{-1}C_n^*\varepsilon_n) = \sigma^4 n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} (c_{ij,n} + c_{ji,n})^2 + (\mu^4 - \sigma^4) n^{-2} \sum_{i=1}^{n} c_{ii,n}. \tag{A.9}
\]

The expressions on the r.h.s. are readily seen to be \( o(1) \) - cp. Kelejian and Prucha (1995) - and thus \( \text{var}(n^{-1}u'(W_n + W_n')D_n) = o(1) \). 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 bounded in absolute value.

For later use we establish furthermore that

\[
n^{-1}\hat{u}'(W_n + W_n')D_n - d'_n = o_p(1). \tag{A.10}
\]

Observe that \( n^{-1}\hat{u}'(W_n + W_n')D_n = n^{-1}u'(W_n + W_n')D_n - (\hat{\theta}_n - \theta_0)'n^{-1}D_n(W_n + W_n')D_n \). 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 \( \mu Q_n = \sum_{i=1}^{n} a_{i,i} \sigma^2 \) in light of the independence of the \( \varepsilon_{i,n} \) and since \( E\varepsilon_{i,n} = 0 \). Next observe that

\[
Q_n - \mu Q_n = \sum_{i=1}^{n} Y_{i,n} \tag{A.11}
\]

with

\[
Y_{i,n} = a_{i,i} (\varepsilon_{i,n}^2 - \sigma^2_{i,n}) + 2\varepsilon_{i,n} \sum_{j=1}^{i-1} a_{i,j} \varepsilon_{j,n} + b_{i,n} \varepsilon_{i,n}, \quad i = 1, \ldots, n. \tag{A.12}
\]

\(^{25}\)We adopt the convention that any sum with an upper index of less than one is zero.
Consider the σ-fields $\mathcal{F}_{0,n} = \{\emptyset, \Omega\}$, $\mathcal{F}_{i,n} = \sigma(\varepsilon_{1,n}, \ldots, \varepsilon_{i,n})$, $1 \leq i \leq n$. By construction $\mathcal{F}_{i-1,n} \subseteq \mathcal{F}_{i,n}$, $Y_{i,n}$ is $\mathcal{F}_{i,n}$-measurable, and it is easily seen that $E(Y_{i,n} \mid \mathcal{F}_{i-1,n}) = 0$. Thus $\{Y_{i,n}, \mathcal{F}_{i,n}, 1 \leq i \leq n, n \geq 1\}$ forms a martingale difference array. Consequently $\sigma_Q^2 = \sum_{i=1}^n E(Y_{i,n}^2)$. The expression for the variance of $Q_n$ given in the text before the theorem follows upon observing that

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

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

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

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 $K_\varepsilon$, such that $\sigma_{i,n}^2 \leq K_\varepsilon$, $E|\varepsilon_{i,n}|^3 \leq K_\varepsilon$, $E|\varepsilon_{i,n}|^s E|\varepsilon_{j,n}|^s \leq K_\varepsilon$, and $E|\varepsilon_{i,n}^2 - \sigma_{i,n}^2|^s \leq K_\varepsilon$ for $s \leq 2 + \delta$. We note further that under the maintained assumptions on the $a_{i,j,n}$ and $b_{i,n}$ there exists a finite constant, say $K_p$, such that $\sum_{j=1}^n |a_{i,j,n}| \leq K_p$ and $n^{-1} \sum_{j=1}^n |b_{i,n}|^s \leq K_p$ for $s \leq 2 + \delta$. Observe that $\sum_{j=1}^n |a_{i,j,n}|^r \leq K_p^r$ for $r \geq 1$ and that $\sum_{k=1}^n |a_{i,k,n}| \overset{P}{\leq} K_p^\prime$. In the following let $q = 2 + \delta$ and let $1/q + 1/p = 1$. Using the triangle and Hölder’s inequalities we then have

$$|Y_{i,n}|^q = \left|a_{i,n}(\varepsilon_{i,n}^2 - \sigma_{i,n}^2) + 2\varepsilon_{i,n} \sum_{j=1}^{i-1} a_{i,j,n}\varepsilon_{j,n} + b_{i,n}\varepsilon_{i,n}\right|^q \leq 2^q \left[|a_{i,n}|^\frac{q}{p} |a_{i,n}|^\frac{q}{q} |\varepsilon_{i,n}^2 - \sigma_{i,n}^2|^q + \sum_{j=1}^{i-1} |a_{i,j,n}|^\frac{q}{p} |a_{i,j,n}|^\frac{q}{q} 2^q |\varepsilon_{i,n}|^q |\varepsilon_{j,n}|^q \right]^q$$

$$\leq 2^q \left[\sum_{j=1}^i |a_{i,j,n}|^\frac{q}{p} \right]^\frac{q}{q} \left[|a_{i,n}|^\frac{q}{q} |\varepsilon_{i,n}^2 - \sigma_{i,n}^2|^q + \sum_{j=1}^{i-1} |a_{i,j,n}|^\frac{q}{p} 2^q |\varepsilon_{i,n}|^q |\varepsilon_{j,n}|^q \right]$$

27
\[ +2^q |b_{i,n}|^q |\varepsilon_{i,n}|^q \]

\[ \leq 2^{2q}(K_p)^{\theta} \left[ |a_{i,n}| |\varepsilon_{i,n}^2 - \sigma_{i,n}^2|^q + \sum_{j=1}^{i-1} |a_{i,j,n}| |\varepsilon_{i,n}|^q |\varepsilon_{j,n}|^q \right] + 2^q |b_{i,n}|^q |\varepsilon_{i,n}|^q. \]

Consequently

\[ \sum_{i=1}^{n} E \left\{ E \left[ |Y_{i,n}|^q \mid \tilde{\mathcal{F}}_{i-1,n} \right] \right\} \]

\[ \leq 2^{2q}(K_p)^{\theta} \sum_{i=1}^{n} \left[ |a_{i,n}| E |\varepsilon_{i,n}^2 - \sigma_{i,n}^2|^q + \sum_{j=1}^{i-1} |a_{i,j,n}| E |\varepsilon_{i,n}|^q E |\varepsilon_{j,n}|^q \right] \]

\[ + 2^q \sum_{i=1}^{n} |b_{i,n}|^q E |\varepsilon_{i,n}|^q \]

\[ \leq 2^{2q}(K_p)^{\theta} K_e \sum_{i=1}^{n} \sum_{j=1}^{i} |a_{i,j,n}| + 2^q K_e \sum_{i=1}^{n} |b_{i,n}|^q \]

\[ \leq n \left[ 2^{2q}(K_p)^{\theta + 1} K_e + 2^q K_p K_e \right] \]

and thus

\[ \sum_{i=1}^{n} E \left\{ E \left[ |X_{i,n}|^{2+\delta} \mid \tilde{\mathcal{F}}_{i-1,n} \right] \right\} \]

\[ = \frac{1}{[n^{-1}\sigma_{\tilde{\mathcal{F}}_{i-1,n}}^{2}]^{1+\delta/2}} \frac{1}{n^{1+\delta/2}} \sum_{i=1}^{n} E \left\{ E \left[ |Y_{i,n}|^{2+\delta} \mid \tilde{\mathcal{F}}_{i-1,n} \right] \right\} \]

\[ \leq \frac{1}{[n^{-1}\sigma_{\tilde{\mathcal{F}}_{i-1,n}}^{2}]^{1+\delta/2}} \frac{1}{n^{1+\delta/2}} \left[ 2^{2q}(K_p)^{\theta + 1} K_e + 2^q K_p K_e \right]. \]

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

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

\[ E \left[ Y_{i,n}^2 \mid \tilde{\mathcal{F}}_{i-1,n} \right] \]

\[ = E \left\{ \left[ a_{i,n}(\varepsilon_{i,n}^2 - \sigma_{i,n}^2) + 2\varepsilon_{i,n} \sum_{j=1}^{i-1} a_{i,j,n} \varepsilon_{j,n} + b_{i,n} \varepsilon_{i,n} \right] \right\}^{2} \varepsilon_{i-1,n}, \ldots, \varepsilon_{1,n} \]

28
\[ a_{ii,n}^2 \left( \mu_{i,n}^{(4)} - \sigma_{i,n}^4 \right) + 4 \sigma_{i,n}^2 \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} a_{ij,n} a_{ik,n} \varepsilon_{j,n} \varepsilon_{k,n} + 4 \left[ a_{ii,n} \mu_{i,n}^{(3)} + b_{i,n} \sigma_{i,n}^2 \right] \sum_{j=1}^{i-1} a_{ij,n} \varepsilon_{j,n} + b_{i,n}^2 \sigma_{i,n}^2 + 2 a_{ii,n} b_{i,n} \mu_{i,n}^{(3)}. \]

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

\[
\sum_{i=1}^{n} E \left[ X_{i,n}^2 \mid \mathcal{F}_{i-1,n} \right] - 1 = \frac{1}{\sigma_Q^2} \sum_{i=1}^{n} E \left[ Y_{i,n}^2 \mid \mathcal{F}_{i-1,n} \right] - E(Y_{i,n}^2) \]

\[
= \frac{1}{\sigma_Q^2} \left[ 8 \sum_{i=1}^{n} \sigma_{i,n}^2 \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} a_{ij,n} a_{ik,n} \varepsilon_{j,n} \varepsilon_{k,n} + 4 \sum_{i=1}^{n} \sigma_{i,n}^2 \sum_{j=1}^{i-1} a_{ij,n}^2 (\varepsilon_{j,n}^2 - \sigma_{j,n}^2) + 4 \sum_{i=1}^{n} \left[ a_{ii,n} \mu_{i,n}^{(3)} + b_{i,n} \sigma_{i,n}^2 \right] \sum_{j=1}^{i-1} a_{ij,n} \varepsilon_{j,n} \right] \]

\[
= \frac{1}{\sigma_Q^2} \left[ 8H_{1,n} + 4H_{2,n} + 4H_{3,n} \right]
\]

with

\[ H_{1,n} = n^{-1} \sum_{i=1}^{n} \sigma_{i,n}^2 \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} a_{ij,n} a_{ik,n} \varepsilon_{j,n} \varepsilon_{k,n}, \]

\[ H_{2,n} = n^{-1} \sum_{i=1}^{n} \sigma_{i,n}^2 \sum_{j=1}^{i-1} a_{ij,n}^2 (\varepsilon_{j,n}^2 - \sigma_{j,n}^2) \]

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

Since \( 0 < c \leq n^{-1} \sigma_Q^2 \), condition (A.2) holds if \( H_{i,n} \xrightarrow{p} 0 \) as \( n \to \infty \) for \( i = 1, 2, 3 \). Clearly \( E H_{i,n} = 0 \). 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

\[ E H_{1,n}^2 \leq 2n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} |a_{ij,n}| |a_{rj,n}| \sum_{k=1}^{j-1} |a_{ik,n}| |a_{rk,n}| \sigma_{i,n}^2 \sigma_{r,n}^2 \sigma_{j,n}^2 \sigma_{k,n}^2 \]

\[ \leq 2K_e^4 n^{-2} \sum_{i=1}^{n} |a_{ij,n}| \sum_{r=1}^{n} |a_{rj,n}| \sum_{k=1}^{n} |a_{ik,n}| |a_{rk,n}| \]

\[ \leq n^{-1} 2K_e^4 K_p^4 \to 0 \text{ as } n \to \infty. \]

29
This proves $H_{1,n} \xrightarrow{p} 0$.

Next observe that

$$H_{2,n} = \sum_{i=1}^{n-1} \phi_{i,n}(\varepsilon_{i,n}^2 - \sigma_{i,n}^2),$$

$$\phi_{i,n} = n^{-1} \sum_{j=i+1}^{n} \sigma_{j,n}^2 a_{ji,n}^2.$$  

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

$$\limsup_{n \to \infty} \sum_{i=1}^{n-1} \phi_{i,n} = \limsup_{n \to \infty} n^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sigma_{j,n}^2 a_{ji,n}^2$$

$$\leq K_\varepsilon \limsup_{n \to \infty} n^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ji,n}^2 \leq K_\varepsilon K_p^2 < \infty$$

and

$$\lim_{n \to \infty} \sum_{i=1}^{n-1} \phi_{i,n}^2 = \lim_{n \to \infty} n^{-2} \sum_{i=1}^{n-1} \left[ \sum_{j=i+1}^{n} \sigma_{j,n}^2 a_{ji,n}^2 \right]^2$$

$$\leq K_\varepsilon^2 \lim_{n \to \infty} n^{-2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ji,n}^2 \sum_{k=1}^{n} a_{ki,n}^2 \leq \lim_{n \to \infty} \sum_{i=1}^{n-1} K_\varepsilon^2 K_p^2 = 0.$$  

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

$$H_{3,n} = \sum_{i=1}^{n-1} \varphi_{i,n} \varepsilon_{i,n},$$

$$\varphi_{i,n} = n^{-1} \sum_{j=i+1}^{n} \left[ a_{jj,n} + b_{j,n} \sigma_{j,n}^2 \right] a_{ji,n}.$$  

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

$$\limsup_{n \to \infty} \sum_{i=1}^{n-1} |\varphi_{i,n}| \leq \limsup_{n \to \infty} n^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[ |a_{jj,n}| + |b_{j,n}| \sigma_{j,n}^2 \right] |a_{ji,n}|$$
\[
\leq K_e \limsup_{n \to \infty} n^{-1} \sum_{j=1}^{n} |a_{j,j,n}| + |b_{j,n}| \sum_{i=1}^{n-1} |a_{j,i,n}|
\]

\[
\leq 2K_e K_p^2 < \infty
\]

and

\[
\lim_{n \to \infty} \sum_{i=1}^{n-1} \varphi_{i,n}^2 \leq \lim_{n \to \infty} n^{-2} \sum_{i=1}^{n-1} \left[ \sum_{j=i+1}^{n} \left( |a_{j,j,n}| \mu_{j,n}^{(3)} + |b_{j,n}| \sigma_{j,n}^2 \right) |a_{j,i,n}| \right]^2
\]

\[
\leq K_e^2 \lim_{n \to \infty} n^{-2} \sum_{i=1}^{n-1} \left( \sum_{j=i+1}^{n} \left( |a_{j,j,n}| + |b_{j,n}| \right) |a_{j,i,n}| \right)^2
\]

\[
\leq K_e^2 \lim_{n \to \infty} n^{-2} \sum_{i=1}^{n-1} \left( K_p^2 + \sum_{j=1}^{n} |a_{j,n}| |a_{j,i,n}| \right)^2
\]

Using Hölder’s inequality we have for \( q = 2 + \delta \) and \( 1/q + 1/p = 1 \)

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

Consequently

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

Applying again the weak law of large numbers (for martingale difference arrays) in Davidson (1994), p. 299, gives \( H_{3,n} \stackrel{p}{\to} 0 \). 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 \( a_{i,i,n} = 0 \).

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

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

with

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

$$= n^{-1/2} \sum_{r=1}^{p_0} \sum_{s=1}^{p_0} \frac{\partial^2 g_n(z_n, \tilde{\theta}_n)'W_n g_n(z_n, \tilde{\tilde{\theta}}_n)}{\partial \theta_r \partial \theta_s} (\tilde{\theta}_n - \theta_0)(\tilde{\theta}_s - \theta_0)$$

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

$$\left| \frac{\partial^2 g_{i,n}(z_n, \tilde{\theta}_n)g_{j,n}(z_n, \tilde{\tilde{\theta}}_n)}{\partial \theta_r \partial \theta_s} \right|$$

$$\leq \left| \frac{\partial^2 g_{i,n}(z_n, \tilde{\theta}_n)}{\partial \theta_r \partial \theta_s} \right| \left| g_{j,n}(z_n, \tilde{\tilde{\theta}}_n) \right| + \left| \frac{\partial g_{i,n}(z_n, \tilde{\theta}_n)}{\partial \theta_r} \right| \left| g_{j,n}(z_n, \tilde{\tilde{\theta}}_n) \right|$$

$$+ \left| \frac{\partial g_{i,n}(z_n, \tilde{\theta}_n)}{\partial \theta_s} \right| \left| g_{j,n}(z_n, \tilde{\tilde{\theta}}_n) \right| + \left| g_{i,n}(z_n, \tilde{\tilde{\theta}}_n) \right| \left| \frac{\partial^2 g_{j,n}(z_n, \tilde{\tilde{\theta}}_n)}{\partial \theta_r \partial \theta_s} \right|$$

$$\leq 4\Delta_{i,n}\Delta_{j,n}.$$ 

Therefore

$$|R_{1n}| \leq 4 \sum_{r=1}^{p_0} \sum_{s=1}^{p_0} \varphi_n n^{1/2} \left| \tilde{\theta}_r - \theta_0 \right| \left| \tilde{\theta}_s - \theta_0 \right|.$$ 

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

$$E\varphi_n \leq n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |w_{ij,n}| \left[ E\Delta_{i,n}^2 \right]^{1/2} \left[ E\Delta_{j,n}^2 \right]^{1/2}$$

$$\leq K_\delta n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |w_{ij,n}| \leq K_\delta K_w < \infty,$$
where $K_\delta$ and $K_w$ are the suprema for $E\Delta^2_{i,n}$ and $\sum_{j=1}^n |w_{i,j,n}|$. Consequently $\varphi_n = O_p(1)$. Since $n^{1/2} |\hat{\theta}_{r,n} - \theta,0| = O_p(1)$ and $|\hat{\theta}_{s,n} - \theta,0| = o_p(1)$ by assumption it follows that $R_{1n} = o_p(1)$. This proves the first part of the theorem.

Applying a first order Taylor expansion yields

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

with

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

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

$$r_{2s,n} = n^{-1} \sum_{r=1}^{p_g} \frac{\partial^2 g_{i,n}(z_n, \hat{\theta}_n) g_{j,n}(z_n, \hat{\theta}_n)}{\partial \theta_r \partial \theta_s} (\hat{\theta}_{r,n} - \theta_{r,0})$$

$$= n^{-1} \sum_{r=1}^{p_g} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j,n} \frac{\partial^2 g_{i,n}(z_n, \hat{\theta}_n) g_{j,n}(z_n, \hat{\theta}_n)}{\partial \theta_r \partial \theta_s} (\hat{\theta}_{r,n} - \theta_{r,0}).$$

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

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

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

**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

$$n^{-1/2} \hat{\varepsilon}_n W_n \hat{\varepsilon}_n = n^{-1/2} Q_n + o_p(1)$$

and consequently

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

$$= n^{-1/2} \left[ \frac{\hat{\sigma}^2_{Q_n}}{\hat{\sigma}^2_{Q_n}} \right]^{1/2} + n^{-1/2} \mu_{Q_n} + o_p(1)$$

$$= n^{-1/2} \left[ \frac{\hat{\sigma}^2_{Q_n}}{\hat{\sigma}^2_{Q_n}} \right]^{1/2} + n^{-1/2} \mu_{Q_n} + o_p(1).$$

33
By assumption \( n^{-1} \sigma_{Q_n}^2 - n^{-1} \sigma_{Q_n}^2 = o_p(1) \) and \( \lim_{n \to \infty} \sigma_{Q_n}^2 / \sigma_{Q_n}^2 = \Lambda^2 > 0 \).

Since \( n^{-1} \sigma_{Q_n}^2 \) is additionally bounded away from zero we have \( \left[ n^{-1} \sigma_{Q_n}^2 \right]^{1/2} / \left[ n^{-1} \sigma_{Q_n}^2 \right]^{1/2} = \Lambda + o_p(1) \) and \( o_p(1) / \left[ n^{-1} \sigma_{Q_n}^2 \right]^{1/2} = o_p(1) \). Observing furthermore that \( (Q_n - \mu_{Q_n}) / \sigma_{Q_n} = O_p(1) \) yields

\[
\frac{\hat{u}_n W_n \hat{u}_n}{\sigma_{Q_n}} = \frac{Q_n - \mu_{Q_n}}{\sigma_{Q_n}} + \frac{n^{-1/2} \mu_{Q_n}}{\left[ n^{-1} \sigma_{Q_n}^2 \right]^{1/2}} + o_p(1).
\]

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 \( \lim_{n \to \infty} \sigma_{Q_n}^2 / \sigma_{Q_n}^2 = \Lambda^2 > 0 \) it follows that ther exist constants \( \sigma \) and \( \sigma_* \) such that for \( n \) large enough \( 0 < \sigma \leq n^{-1} \sigma_{Q_n}^2 \leq \sigma_\ast < \infty \). Result (b) of the theorem now follows immediately from the lemma below with \( \eta_n = \Lambda(Q_n - \mu_{Q_n}) / \sigma_{Q_n} + o_p(1) \), \( \varphi_n = \left[ n^{-1} \sigma_{Q_n}^2 \right]^{1/2} \), \( f_n = \left[ n^{-1} \sigma_{Q_n}^2 \right]^{1/2} \) for \( \sigma \leq n^{-1} \sigma_{Q_n}^2 \leq \sigma_\ast \), and \( f_n = (\sigma^{1/2} + \sigma_\ast^{1/2}) / 2 \) otherwise, and \( d_n = n^{-1/2} \mu_{Q_n} \).

**Lemma A.1** Let \( \eta_n \) and \( \varphi_n \) 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 \( |d_n| \to \infty \) as \( n \to \infty \). Then for every constant \( \gamma > 0 \) we have

\[
\lim_{n \to \infty} P \left( |\eta_n + d_n / \varphi_n| \leq \gamma \right) = 0.
\]

**Proof of Lemma.** Choose some \( \varepsilon > 0 \) and some \( K_\ast > K_2 \). Since \( \eta_n = O_p(1) \) it follows that there exists a positive constant \( M_\varepsilon \) 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 \( n_\varepsilon \) such that \( P(|\varphi_n| \geq K_\ast) \leq \varepsilon / 2 \) for all \( n \geq n_\varepsilon \). Next observe that

\[
P \left( |\eta_n + d_n / \varphi_n| \leq \gamma \right) = P \left( |\eta_n + d_n / \varphi_n| \leq \gamma, |\eta_n| > M_\varepsilon \right) + P \left( |\eta_n + d_n / \varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon \right)
\]
\[
\leq P \left( |\eta_n| > M_\varepsilon \right) + P \left( |\eta_n + d_n / \varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon \right)
\]
\[
\leq \varepsilon / 2 + P \left( |\eta_n + d_n / \varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon, |\varphi_n| \geq K_\ast \right)
\]
\[
+ P \left( |\eta_n + d_n / \varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon, |\varphi_n| < K_\ast \right)
\]
\[
\leq \varepsilon / 2 + P \left( \varphi_n \geq K_\ast \right) + P \left( |\eta_n + d_n / \varphi_n| \leq \gamma, |\eta_n| \leq M_\varepsilon, |\varphi_n| < K_\ast \right).
\]
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 $\Omega_n = \{ \omega \in \Omega : |\eta_n(\omega)| \leq M_\varepsilon, |\varphi_n(\omega)| < K_*, |d_n| \geq |d_n|/K_* \}$ for all $\omega \in \Omega_n$, $n \in \mathbb{N}$. Since $|d_n| \to \infty$ as $n \to \infty$ there exists an index $n^*_\varepsilon$ 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. ■

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

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

Assumption 8 then implies

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

Consequently

$$\left| \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{i-1} (w_{i,j,n} + w_{j,i,n})^2 \left[ \hat{\sigma}_{i,n}^2 - \sigma_{i,n}^2 \sigma_{j,n}^2 \right] \right| \leq K_w^2 (2K_r \phi_n + \phi_n^2) \quad \text{(1A.15)}$$

$$\left| \frac{1}{n} \sum_{i=1}^{n} w_{i,i,n}^2 \left[ \hat{\sigma}_{i,n}^4 - \sigma_{i,n}^4 \right] \right| \leq K_w^2 (2K_r \phi_n + \phi_n^2),$$

$$\left| \frac{1}{n} \sum_{i=1}^{n} w_{i,i,n}^2 \left[ \mu_{i,n}^{(4)} - \bar{\mu}_{i,n}^{(4)} \right] \right| \leq K_w^2 \phi_n .$$
Similarly
\[
\begin{align*}
|\hat{b}_{i,n}^2 - \overline{b}_{i,n}^2| & \leq 2 |\overline{b}_{i,n}| \Psi_{i,n} \psi_n + \Psi_{i,n}^2 \psi_n^2, \\
|\hat{b}_{i,n} \hat{\sigma}_{i,n}^2 - \overline{b}_{i,n} \overline{\sigma}_{i,n}^2| & \leq 2 K_r |\overline{b}_{i,n}| \Psi_{i,n} \psi_n + K_r \Psi_{i,n}^2 \psi_n^2 \\
& \quad + 2 |\overline{b}_{i,n}| \Psi_{i,n} \psi_n \phi_n + \Psi_{i,n}^2 \phi_n^2 + \overline{b}_{i,n}^2 \phi_n^2, \\
|\hat{b}_{i,n} \hat{\mu}_{i,n}^{(3)} - \overline{b}_{i,n} \overline{\mu}_{i,n}^{(3)}| & \leq K_r \Psi_{i,n} \psi_n + \Psi_{i,n} \phi_n + |\overline{b}_{i,n}| \phi_n.
\end{align*}
\]

Observing that by Hölder’s and Lyapunov’s inequalities \( n^{-1} \sum_{i=1}^n |\overline{b}_{i,n}| \Psi_{i,n} \leq K_r^{1/2} B_n^{1/2} \) 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
\[
\begin{align*}
\left| \frac{1}{n} \sum_{i=1}^n \left[ \hat{b}_{i,n}^2 \hat{\sigma}_{i,n}^2 - \overline{b}_{i,n}^2 \overline{\sigma}_{i,n}^2 \right] \right| & \leq 2 K_r^{3/2} B_n^{1/2} \psi_n + K_r B_n \psi_n^2 \\
& \quad + 2 K_r^{1/2} B_n \psi_n \psi_n \phi_n + B_n \psi_n \phi_n^2 + K_r \phi_n, \\
\left| \frac{1}{n} \sum_{i=1}^n \left[ \hat{b}_{i,n} \hat{\mu}_{i,n}^{(3)} - \overline{b}_{i,n} \overline{\mu}_{i,n}^{(3)} \right] \right| & \leq K_r B_n \psi_n + B_n \psi_n \phi_n + K_r \phi_n.
\end{align*}
\]

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 \( n^{-1} (\overline{\sigma}_{i,n}^2 - \overline{\sigma}_{i,n}^2) = o_p(1) \). The claim that \( n^{-1} \overline{\sigma}_{i,n}^2 \leq \text{const} < \infty \) is easily checked in light of assumptions 4 and 8.

\[\square\]

**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 \( z_n = (z_{1,n}, \ldots, z_{n,n}) \) 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 \( Z = \mathbb{R} \times \mathbf{X} \) with \( \mathbf{X} = \prod_{i=1}^k [-C_x, C_x] \). The parameter space for \( \beta_0 \), say \( \Theta_\beta \), was assumed to be an open subset of \( \mathbb{R}^k \). Furthermore \( \Theta_\beta \) was assumed to be contained in a compact set. Hence there exists a finite constant, say \( C_\beta \), 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 \( \overline{\Theta} = \prod_{i=1}^k [-C_\beta, C_\beta] \times [c_\sigma, C_\sigma] \) 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 \( \Theta, \overline{\Theta} \), and \( \Theta_* \) are open, compact and open, respectively, and satisfy \( \Theta \subseteq \overline{\Theta} \subseteq \Theta_* \) by construction. Next define
\[
g_{i,n}(\hat{z}_1, \ldots, \hat{z}_n, \theta) = y_i - f(x_i, \theta) \tag{A.17}
\]
for all \( z_j = (y_j, x_j) \in Z, j = 1, \ldots, n \), and \( \theta = (\beta, \sigma^2_v) \in \Theta_* \). Then (4.5) can be written as
\[
  u_{i,n} = \varepsilon_i = g_{i,n}(z_n, \theta_0) = y_i - f(x_i, \theta_0).
\]  

Furthermore we have
\[
  \tilde{u}_{i,n} = \tilde{\varepsilon}_i = g_{i,n}(z_n, \hat{\theta}_n) = y_i - f(x_i, \hat{\theta}_n). 
\]  

We now check Assumptions 4-6. Assumption 4 is assumed to hold. Assumption 5 holds evidently since \( \Theta \) is open. We next verify Assumption 6. Since \( \phi_N(\cdot) \) and \( \Phi_N(\cdot) \) are twice continuously differentiable, and since \( \sup_{x \in X} \sup_{\theta \in \Theta} |x^2 \beta / \sigma_v| < \infty \), it is readily seen that \( f(x, \theta), \partial f(x, \theta) / \partial \theta \) and \( \partial^2 f(x, \theta) / \partial \theta \partial \theta' \) are continuous on \( Z \times \Theta_* \). (Note that the derivatives are well defined since \( \Theta_* \) is open.) Assumption 6(a) is hence clearly satisfied. Since \( X \) and \( \Theta \) 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
\[
  \sup_{\theta \in \Theta} |x_i \beta| \leq \sup_{x \in X} \sup_{\theta \in \Theta} |x^2 \beta| < K, 
\]
\[
  \sup_{\theta \in \Theta} \frac{|x_i \beta|}{\sigma_v} \leq \sup_{x \in X} \sup_{\theta \in \Theta} \frac{|x^2 \beta|}{\sigma_v} < K, 
\]
\[
  \sup_{\theta \in \Theta} |f(x_i, \theta)| \leq \sup_{x \in X} \sup_{\theta \in \Theta} |f(x, \theta)| < K, 
\]
\[
  \sup_{\theta \in \Theta} \left| \frac{\partial g_{i,n}(z_n, \theta)}{\partial \theta} \right| = \sup_{\theta \in \Theta} \left| \frac{\partial f(x_i, \theta)}{\partial \theta} \right| \leq \sup_{x \in X} \sup_{\theta \in \Theta} \left| \frac{\partial f(x, \theta)}{\partial \theta} \right| < K, 
\]
\[
  \sup_{\theta \in \Theta} \left| \frac{\partial^2 g_{i,n}(z_n, \theta)}{\partial \theta \partial \theta'} \right| = \sup_{\theta \in \Theta} \left| \frac{\partial^2 f(x_i, \theta)}{\partial \theta \partial \theta'} \right| \leq \sup_{x \in X} \sup_{\theta \in \Theta} \left| \frac{\partial^2 f(x, \theta)}{\partial \theta \partial \theta'} \right| < K. 
\]

Observing that \( |y_i| \leq |y^*_i| \leq |x_i \beta_0| + |v_i| \leq K + |v_i| \) it follows further that
\[
  \sup_{\theta \in \Theta} |g_{i,n}(z_n, \theta)| \leq |y_i| + \sup_{\theta \in \Theta} |f(x_i, \theta)| \leq |v_i| + 2K. 
\]  

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

We next verify Assumption 7 with \( u_n = \varepsilon_n, Q_n = \varepsilon_n^2 A_n, A_n = (W_n + W'_n) / 2 \) and \( b_n = 0 \). We first verify that the elements of \( \varepsilon_n, A_n \) 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

$$E|\varepsilon_i|^3 = E|y_i - f(x_i, \theta_0)|^3 \leq 4 \left[ E|y_i|^3 + |f(x_i, \theta_0)|^3 \right]$$

(A.22)

$$\leq 4 \left\{ E|y_i|^3 \sup_{x \in X} \sup_{\theta \in \Theta} |f(x, \theta)| \right\}^3$$

$$\leq 4 \left\{ E|y_i|^3 + K^3 \right\}$$

$$\leq 4 \left\{ E|v_i|^3 + 3K^2 |v_i| + 2K^3 \right\}$$

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

$$n^{-1}u_n'(W_n + W_n') \frac{\partial g_{i,n}(z_n, \theta_0)}{\partial \theta} = -2n^{-1} \varepsilon_n' A_n \frac{\partial f(x_i, \theta_0)}{\partial \theta} = o_p(1),$$

(A.23)

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 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 $\sigma_i^2$ are bounded by some finite constant. Consequently $n^{-1}w_n A_n \partial f(x_i, \theta_0)/\partial \theta$ has mean zero and its variance covariance matrix converges to zero; (A.23) now follows from Chebychev’s inequality and thus (3.7) holds. We have now verified all conditions of Assumption 7.

Now take $\sigma^2_{Q_n} = \sigma^2_{\hat{Q}_n}$, where the latter is defined in (4.1), and $\sigma^2_{Q_n} = \sigma^2_{\hat{Q}_n}$ with

$$\sigma^2_{Q_n} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij,n} + w_{ji,n})^2 \sigma_i^2 \sigma_j^2,$$

(A.24)
Then clearly $\Lambda = 1$. Furthermore, given (4.2) and since $\sigma_i^2 = h_i(\theta_0) \geq const > 0$ 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

$$n^{-1}\sigma_{Q_n}^2 - n^{-1}\sigma_{Q_n^*}^2 = o_p(1) \quad (A.25)$$

using Theorem 3. In the notation of that theorem we have $\sigma_{i,n}^2 = \sigma_{i,n}^2 = h_i(\theta_0)$, $\tilde{\sigma}_{i,n}^2 = \tilde{h}_i(\tilde{\theta}_n)$, and $\tilde{b}_{i,n} = \tilde{b}_{i,n} = b_{i,n} = 0$. To apply Theorem 3 we have to verify Assumptions 4 and 8. Assumption 4 is postulated. Next consider Assumption 8(a). Since $w_{i,n} = 0$ we can ignore all conditions involving third and fourth moments. To prove $|\tilde{\sigma}_{i,n}^2 - \sigma_{i,n}^2| \leq \phi_n$ 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 $\Theta_*$. Again, since continuous functions are bounded in absolute value on compact sets, there exists a finite constant, say $K_*$, such that

$$\sup_{1 \leq i \leq n, n \geq 1} \sup_{\theta \in \Theta} \left| \frac{\partial h_i(\theta)}{\partial \theta} \right| \leq \sup_{\bar{Y} \in \bar{Y}} \sup_{\theta \in \Theta} \left| \frac{\partial h(x_i, \theta)}{\partial \theta} \right| < K_*,$$

where $h(x_i, \theta) = h_i(\theta)$. Since $n^{1/2}(\tilde{\theta}_n - \theta_0) = O_p(1)$ 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 and sample selection models are analogous.

\[\square\]

**Proof of Claims in Section 4.2.** In the following let $\bar{y}_n = M_n y_n$. 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 $D_n = [\bar{y}_n, X_n]$, $\theta = (\lambda, \beta')'$, and $z_n = (z_{1,n}, \ldots, z_{n,n})$ with $z_{i,n} = (y_{i,n}, x_{1,n}, \ldots, x_{k,n})$. 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 $\Theta$ 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 $\theta$, respectively. To verify Assumption 6(b) observe that $\Delta_{i,n}^2 = |\partial g_{i,n}/\partial \theta|^2 = \bar{y}_{i,n}^2 + \sum_{j=1}^{k} x_{i,j,n}^2$. 

39
From the first result in (A.4) it follows that $E \bar{\sigma}^2_{Q_n} \leq \text{const} < \infty$. Since the $x_{ij,n}$ are assumed to be 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

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

(A.26)

where $\sigma^2_n = n^{-1} E(u'_n u_n) = n^{-1} E(\varepsilon'_n C_n^* \varepsilon_n) = \sigma^2 n^{-1} \text{tr}(C_n^*)$ with $C_n^* = (I_n - \rho_0 M_n)^{-1}(I_n - \rho_0 M_n')^{-1}$, and where $b_n = H_n P'd_n$. Furthermore let $\bar{\sigma}^2_{Q_n} = \bar{\sigma}^2_{Q_n}$. Clearly $\sigma^2_n = n^{-1} \varepsilon'_n u_n = n^{-1} \varepsilon'_n u_n + o_p(1)$, since $\varepsilon_n = u_n - D_n(\theta_n - \theta_0)$, $\theta_n$ 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 $n^{-1} u'_n u_n - \sigma^2_n = o_p(1)$ and consequently $\bar{\sigma}^2_{Q_n} = \sigma^2_{Q_n} = o_p(1)$. This verifies Assumption 8(a) for $\sigma^2_{Q_n}$ and $\bar{\sigma}^2_{Q_n}$ (observing that $\sigma^2_{i,n} = \sigma^2_n$ and $w_{ii,n} = 0$). Assumption 8(b) for $\bar{b}_{i,n}$ and $\bar{b}_{i,n}$ was verified at the end of Section 3.3. It now follows from Theorem 3 that $\bar{\sigma}^2_{Q_n} - \sigma^2_{Q_n} = o_p(1)$ as postulated in Theorem 2.

The expressions for the mean $\mu_{Q_n}$ and variance $\sigma^2_{Q_n}$ 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 $a_{i,n} = 0$, $b_n = \bar{b}_n = H_n P'd_n$ and $\sigma^2_{i,n} = \bar{\sigma}^2_{n} = \sigma^2$. Thus in this case we have $\mu_{Q_n} = 0$ and $\sigma^2_{Q_n} = \bar{\sigma}^2_{Q_n}$. In light of (4.2) we also have $n^{-1} \sigma^2_{Q_n} \geq c$ 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 $n^{-1} \sigma^2_{Q_n} > 0$, $\lim_{n \to \infty} n^{-1} \sigma^2_{Q_n} = s > 0$, $\lim_{n \to \infty} n^{-1} \sigma^2_{Q_n} = \sigma > 0$ and $\lim_{n \to \infty} n^{-1} \text{tr}(A_n) = a \neq 0$. Then clearly $n^{-1} \sigma^2_{Q_n} \geq c$ for some $c > 0$ and $\lim_{n \to \infty} \sigma^2_{Q_n} / \sigma^2_{Q_n} > 0$. Furthermore $n^{-1/2} |\mu_{Q_n}| = n^{1/2} |n^{-1} \text{tr}(A_n)| \to \infty$ for $n \to \infty$. The claim that $\lim_{n \to \infty} P(|I_n| > \tau)$ for arbitrary $\tau > 0$ now follows from part (b) of Theorem 2.

40
References


