On Spatial Processes and Asymptotic Inference under Near-Epoch Dependence

Nazgul Jenish\textsuperscript{1} and Ingmar R. Prucha\textsuperscript{2}

April 28, 2010

\textsuperscript{1}Department of Economics, New York University, 19 West 4th Street, New York, NY 10012. Tel.: 212-998-3891, Email: nazgul.jenish@nyu.edu

\textsuperscript{2}Department of Economics, University of Maryland, College Park, MD 20742. Tel.: 301-405-3499, Email: prucha@econ.umd.edu
Abstract

The development of a general inferential theory for nonlinear models with cross-sectionally or spatially dependent data has been hampered by a lack of appropriate limit theorems. To facilitate a general asymptotic inference theory relevant to economic applications, this paper first extends the notion of near-epoch dependent (NED) processes used in the time series literature to random fields. The class of processes that is NED on, say, an $\alpha$-mixing process, is shown to be closed under infinite transformations, and thus accommodates models with spatial dynamics. This would generally not be the case for the smaller class of $\alpha$-mixing processes. The paper then derives a central limit theorem and law of large numbers for NED random fields. These limit theorems allow for fairly general forms of heterogeneity including asymptotically unbounded moments, and accommodate arrays of random fields on unevenly spaced lattices. The limit theorems are employed to establish consistency and asymptotic normality of GMM estimators. These results provide a basis for inference in a wide range of models with cross-sectional or spatial dependence.

JEL Classification: C10, C21, C31

Key words: Random fields, near-epoch dependent processes, central limit theorem, law of large numbers, GMM estimator
1 Introduction

Models with cross-sectionally or spatially dependent data have recently attracted considerable attention in various fields of economics including labor and public economics, IO, political economy, macroeconomics, international economics, regional science and urban economics. In these models cross sectional dependence typically stems from agents’ strategic interaction, neighborhood effects, shared resources and common shocks. Furthermore, economic agents are typically heterogeneous, e.g., vary in size. Mathematically, agents’ observations may thus be viewed as a realization of a dependent heterogeneous process indexed by a point in $\mathbb{R}^d$, $d > 1$, i.e., a random field.\(^2\)

The aim of this paper is to define a class of random fields that is sufficiently general to accommodate many applications of interest, and to establish corresponding limit theorems that can be used for asymptotic inference. In particular, we apply these limit theorems to prove consistency and asymptotic normality of generalized method of moments (GMM) estimators for a general class of nonlinear spatial models.

There have been different approaches to modeling cross-sectional dependence in the econometrics literature. Two large strands of the literature have focused on (i) linear spatial autoregressive models, also known as Cliff-Ord (1981) type models, and (ii) common factor models. In both cases, progress was facilitated, loosely speaking, by imposing specific structural conditions on the data generating process, and/or by exploiting some underlying (conditional) independence assumptions. Another common approach to model dependence is through mixing conditions. Various mixing concepts developed for time series processes have been extended to random fields. However, the respective limit theorems for random fields have not been sufficiently general to accommodate many of the processes encountered in economics. This hampered the development of a general asymptotic inference theory for nonlinear models with cross-sectional dependence. Towards filling this gap, Jenish and Prucha (2009) have recently introduced a set of limit theorems (CLT, ULLN, LLN) for $\alpha$-mixing random fields on unevenly spaced lattices that allow for nonstationary processes with trending moments.

Yet some dependent spatial processes do not satisfy mixing conditions. As with time series, spatial processes may be generated as functions of infinitely many elements of an input process, with a spatial moving average process being

\(^1\)We thank the participants of the Cowles Foundation Conference, Yale, June 2009, as well as the seminar participants at the Columbia University for helpful discussions. This research benefitted from a University of Maryland Ann G. Wylie Dissertation Fellowship for the first author and from financial support from the National Institute of Health through SBIR grant 1 R43 AG027622 for the second author.

\(^2\)The space and metric are not restricted to physical space and distance.

\(^3\)For recent contributions see, e.g., Robinson (2009, 2007), Yu, de Jong and Lee (2008), Kelejian and Prucha (2007a,b,2004), Lee (2007a,b, 2004), and Chen and Conley (2001).

\(^4\)For recent contributions, see, e.g., Andrews (2005), Bai and Ng (2004), Bai (2003), Phillips and Sul (2003).
a leading example. As is well-known, the mixing property is not necessarily preserved under transformations involving infinite number of lags. For instance, Andrews (1984) shows that a simple time series AR(1) process fails to be $\alpha$-mixing though its input process is independent. Thus, it is important to develop an asymptotic theory for a generalized class of random fields that is “closed with respect to infinite transformations.”

To tackle this problem, the paper first extends the concept of near-epoch dependent (NED) processes used in the time series literature to spatial processes. The notion dates back to Ibragimov (1962), and Billingsley (1968). The NED concept, or variants thereof, have been used extensively in the time series literature by McLeish (1975a,b), Bierens (1981), Wooldridge (1986), Gallant (1987), Gallant and White (1988), Andrews (1988), Pötscher and Prucha (1991,1997), Davidson (1992, 1993, 1994) and de Jong (1997), among others. Doukhan and Louhichi (1999), Ango Nze and Doukhan (2004) introduced an alternative class of dependent processes called “$\theta$-weakly dependent.” We provide various examples suggesting that the class of NED spatial processes, which subsumes mixing processes, is sufficiently broad to cover many applications of interest. For instance, we show that it covers Cliff-Ord type processes, autoregressive and moving average random fields, and, more generally, nonlinear spatial Bernoulli shifts.

The paper then derives a CLT and an LLN for spatial processes that are near epoch dependent on an $\alpha$-mixing input process. These limit theorems allow for fairly general forms of heterogeneity including asymptotically unbounded moments, and accommodate arrays of random fields on unevenly spaced lattices. The LLN can be combined with the generic ULLN in Jenish and Prucha (2009) to obtain an ULLN for NED spatial processes. In the time series literature, CLTs for NED processes were derived by Wooldridge (1986), Davidson (1992, 1993), and de Jong (1997). Interestingly, our CLT contains as a special case the CLT of Wooldridge (1986), Theorem 3.13 and Corollary 4.4.

In addition, we give conditions under which the NED property is preserved under transformations. These results play a key role in verifying the NED property in applications. Thus, the NED property is compatible with considerable heterogeneity and dependence, invariant under transformations, and leads to a CLT and LLN under fairly general conditions. All these features make it a convenient tool for modeling spatial dependence.

As an application, we establish consistency and asymptotic normality of spatial GMM estimators. These results provide a fundamental basis for constructing confidence intervals and testing hypothesis for GMM estimators in nonlinear spatial models. Our results also expand on Conley (1999), who established the asymptotic properties of GMM estimators assuming that the data generating process and the moment functions are stationary and $\alpha$-mixing.\footnote{This important early contribution employs Bolthausen’s (1982) CLT for stationary $\alpha$-mixing random fields on the regular lattice $\mathbb{Z}^2$. However, the mixing and stationarity assumptions may not hold in many applications. The present paper relaxes these critical assumptions.}

The rest of the paper is organized as follows. Section 2 introduces the concept of NED spatial processes and gives some mild conditions that ensure preserv-
tion of the NED property under transformations. Section 3 provides examples of NED spatial processes. Section 4 contains the CLT and LLN for NED spatial processes. Section 5 establishes the asymptotic properties of spatial GMM estimators. All proofs are collected in the appendices.

2 Definition of NED Spatial Processes

Let $D \subset \mathbb{R}^d$, $d \geq 1$, be a (possibly) unevenly spaced lattice of locations in $\mathbb{R}^d$, and let $Z = \{ Z_{i,n}, i \in D_n, n \geq 1 \}$ and $\varepsilon = \{ \varepsilon_{i,n}, i \in T_n, n \geq 1 \}$ be triangular arrays of random fields defined on a probability space $(\Omega, \mathcal{F}, P)$ with $D_n \subseteq T_n \subseteq D$. The space $\mathbb{R}^d$ is equipped with the metric $\rho(i, j) = \max_{1 \leq l \leq d} |i_l - j_l|$, where $i_l$ is the $l$-th component of $i$. The distance between any subsets $U, V \subseteq D$ is defined as $\rho(U, V) = \inf \{ \rho(i, j) : i \in U \text{ and } j \in V \}$. Furthermore, let $|U|$ denote the cardinality of a finite subset $U \subset D$.

The random variables $Z_{i,n}$ and $\varepsilon_{i,n}$ are possibly vector-valued taking their values in $\mathbb{R}^p_z$ and $\mathbb{R}^p_z$, respectively. We assume that $\mathbb{R}^p_z$ and $\mathbb{R}^p_z$ are normed metric spaces equipped with the Euclidean norm, which we denote (in an obvious misuse of notation) as $|.|$. For any random vector $Y$, let $\| Y \|_p = [E|Y|^p]^{1/p}$, $p \geq 1$, denote its $L_p$-norm. Finally, let $\mathcal{F}_{i,n}(s) = \sigma(\varepsilon_{j,n}; j \in T_n : \rho(i, j) \leq s)$ be the $\sigma$-field generated by the random vectors $\varepsilon_{j,n}$ located in the $s$-neighborhood of location $i$.

Throughout the paper, we maintain the above notational conventions as well as the following assumption concerning $D$.

**Assumption 1** The lattice $D \subset \mathbb{R}^d$, $d \geq 1$, is infinite countable. All elements in $D$ are located at distances of at least $\rho_0 > 0$ from each other, i.e., for all $i, j \in D : \rho(i, j) \geq \rho_0$; w.l.o.g. we assume that $\rho_0 > 1$.

The assumption of a minimum distance has also been used by Conley (1999) and Jenish and Prucha (2009). It ensures the growth of the sample size as the sample regions $D_n$ and $T_n$ expand. The setup is thus geared towards what is referred to in the spatial literature as increasing domain asymptotics.

We now introduce the notion of near-epoch dependent (NED) random fields.

**Definition 1** Let $Z = \{ Z_{i,n}, i \in D_n, n \geq 1 \}$ be a random field with $\| Z_{i,n} \|_p < \infty$, $p \geq 1$, let $\varepsilon = \{ \varepsilon_{i,n}, i \in T_n, n \geq 1 \}$ be a random field, where $|T_n| \to \infty$ as $n \to \infty$, and let $d = \{ d_{i,n}, i \in D_n, n \geq 1 \}$ be an array of finite positive constants. Then the random field $Z$ is said to be $L_p(d)$-near-epoch dependent on the random field $\varepsilon$ if

$$\| Z_{i,n} - E(Z_{i,n} | \mathcal{F}_{i,n}(s)) \|_p \leq d_{i,n} \psi(s) \quad (1)$$
for some sequence \( \psi(s) \geq 0 \) with \( \lim_{s \to \infty} \psi(s) = 0 \). The \( \psi(s) \), which are w.l.o.g. assumed to be non-increasing, are called the NED coefficients, and the \( d_{i,n} \) are called NED scaling factors. \( Z \) is said to be \( L_p \)-NED on \( \varepsilon \) of size \(-\lambda \) if \( \psi(s) = O(s^{-\mu}) \) for some \( \mu > \lambda > 0 \). Furthermore, if \( \sup_n \sup_{i \in D_n} d_{i,n} < \infty \), then \( Z \) is said to be uniformly \( L_p \)-NED on \( \varepsilon \).

Recall that \( D_n \subseteq T_n \). Typically, \( T_n \) will be an infinite subset of \( D \) and often \( T_n = D \). However, to cover Cliff-Ord type processes, see Section 3.3, \( T_n \) is allowed to depend on \( n \) and to be finite provided that it increases in size with \( n \).

The role of the scaling factors \( \{d_{i,n}\} \) is to allow for the possibility of “unbounded moments”, i.e., \( \sup_n \sup_{i \in D_n} d_{i,n} = \infty \). Unbounded moments may reflect trends in the moments in certain directions, in which case we may also use, as in the time series literature, the terminology of “trending moments”. The NED property is thus compatible with a considerable amount of heterogeneity.

In establishing limit theorems for NED processes, we will have to impose restrictions on the scaling factors \( d_{i,n} \). In this respect, observe that

\[
\|Z_{i,n} - E(Z_{i,n}|\mathcal{F}_{i,n}(s))\|_p \leq \|Z_{i,n}\|_p + \|E(Z_{i,n}|\mathcal{F}_{i,n}(s))\|_p \leq 2 \|Z_{i,n}\|_p
\]

by the Minkowski and the conditional Jensen inequalities. Given this, we may choose \( d_{i,n} \leq 2 \|Z_{i,n}\|_p \), and consequently w.l.o.g. \( 0 \leq \psi(s) \leq 1 \); see, e.g., Davidson (1994), p. 262, for a corresponding discussion within the context of time series processes. Note that by the Lyapunov inequality, if \( Z_{i,n} \) is \( L_p \)-NED, then it is also \( L_q \)-NED with the same coefficients \( \{d_{i,n}\} \) and \( \{\psi(s)\} \) for any \( q \leq p \).

Our definition of NED for spatial processes is adapted from the definition of NED for time series processes. In the time series literature, the NED concept first appeared in the works of Ibragimov (1962) and Billingsley (1968), although they did use the present term. The concept of time series NED processes was later formalized by McLeish (1975a,b), Wooldridge (1986), Gallant (1987), Gallant and White (1988). These authors considered only \( L_2 \)-NED processes. Andrews (1988) generalized it to \( L_p \)-NED processes for \( p \geq 1 \). Davidson (1992, 1993, 1994) and de Jong (1997) further extended it to allow for trending time series processes.

An important motivation for considering NED processes is that mixing is not necessarily preserved under transformations involving infinitely many arguments; see, e.g., Andrews (1984) for simple examples in a time series context. However, as illustrated below, for a wide range of models the output process is obtained as a function of infinitely many input variables. In those situations mixing of the input process does not necessarily carry over to the output process, and thus limit theorems for averages of the output process cannot simply be established from limit theorems for mixing processes. Nevertheless, as with time series processes, we show below that limit theorems can be extended to spatial processes that are NED on a mixing input process, provided the approxima-
tion error declines “sufficiently fast” as the conditioning set of input variables expands.

A further attractive feature of NED processes is that the NED property is preserved under transformations. Econometric estimators are usually defined either explicitly as functions of some underlying data generating processes or implicitly as optimizers of a function of the data generating process. Thus, if the data generating process is NED on some input process, the question arises under what conditions functions of random fields are also NED on the same input process.

Various conditions that ensure preservation of the NED property under transformations have been established in the time series literature by Gallant and White (1988), and Davidson (1994). In fact, these results extend to random fields. In particular, the NED property is preserved under summation and multiplication, and carries over from a random vector to its components and vice versa. For future reference, we now state some results for generalized classes of nonlinear functions. Their proofs are analogous to those in the time series literature, and therefore omitted.

Consider transformations of \( Z_{i,n} \) given by a family of functions

\[ g_{i,n} : \mathbb{R}^{p_x} \to \mathbb{R}. \]

The functions \( g_{i,n} \) are assumed Borel-measurable for all \( n \) and \( i \in D \). They are furthermore assumed to satisfy the following Lipschitz condition: For all \((z, z^*) \in \mathbb{R}^{p_x} \times \mathbb{R}^{p_x}\) and all \( i \in D_n \) and \( n \geq 1 \):

\[ |g_{i,n}(z) - g_{i,n}(z^*)| \leq B_{i,n}(z, z^*)|z - z^*| \tag{2} \]

where

\[ B_{i,n}(z, z^*) : \mathbb{R}^{p_x} \times \mathbb{R}^{p_x} \to \mathbb{R}_+, \]

is Borel-measurable. Of course, this condition would be devoid of meaning without further restrictions on \( B_{i,n}(z, z^*) \), which are given in the next propositions.

**Proposition 1** Suppose \( g_{i,n}(Z_{i,n}) \) satisfies Lipschitz condition (2) with

\[ |B_{i,n}(z, z^*)| \leq C < \infty \]

for all \((z, z^*) \in \mathbb{R}^{p_x} \times \mathbb{R}^{p_x}\) and all \( i \) and \( n \). If for \( p \geq 1 \) the \( \{Z_{i,n}\} \) are \( L_p \)-NED of size \(-\lambda\) on \( \{\varepsilon_{i,n}\} \) with scaling factors \( \{d_{i,n}\} \), then \( g_{i,n}(Z_{i,n}) \) is also \( L_p \)-NED of size \(-\lambda\) on \( \{\varepsilon_{i,n}\} \) with scaling factors \( \{2Cd_{i,n}\} \).

**Proposition 2** Suppose \( g_{i,n}(Z_{i,n}) \) satisfies Lipschitz condition (2) with

\[ \sup_s \left\| B_{i,n}^{(s)} \right\|_2 < \infty \] and \( \sup_s \left\| B_{i,n}^{(s)} \right\| \left\| Z_{i,n} - \tilde{Z}_{i,n} \right\|_r < \infty \tag{3} \]
for some $r > 2$, where $B_{i,n}^{(s)} = B_{i,n}(Z_{i,n}, \tilde{Z}_{i,n})$ and $\tilde{Z}_{i,n} = E[Z_{i,n} \mid \tilde{\delta}_{i,n}(s)]$. If $\|g_{i,n}(Z_{i,n})\|_2 < \infty$ and $Z_{i,n}$ is $L_2$-NED of size $-\lambda$ on $\{\tilde{\varepsilon}_{i,n}\}$ with scaling factors $\{d_{i,n}\}$, then $g_{i,n}(Z_{i,n})$ is $L_2$-NED of size $-\lambda(r-2)/(2r-2)$ on $\{\tilde{\varepsilon}_{i,n}\}$ with scaling factors $d'_{i,n} = d_{i,n}^{(r-2)/(2r-2)} \sup_s \|B_{i,n}^{(s)}\|_2^{(r-2)/(2r-2)} \|B_{i,n}^{(s)}[Z_{i,n} - \tilde{Z}_{i,n}])\|_r^{r/(2r-2)}$.

Thus, the NED property is hereditary under reasonably weak conditions. These conditions facilitate verification of the NED property in practical application. In particular, we will use them in the proof of asymptotic normality of spatial GMM estimators in Section 5.

### 3 Examples of NED Spatial Processes

In this section, we give examples of some important classes of spatial processes, which are widely used in applications, and show that they satisfy the NED property. Of course, to establish limit theorems the NED property would be accompanied by mixing assumptions on the input process.

#### 3.1 Moving Average Random Fields

Consider the infinite moving average (or linear) random field on $\mathbb{Z}^d$, say, $Y = \{Y_i, i \in \mathbb{Z}^d\}$ defined as

$$Y_i = \sum_{j \in \mathbb{Z}^d} g_{ij} \varepsilon_j,$$

where $\varepsilon = \{\varepsilon_i, i \in \mathbb{Z}^d\}$ is a zero mean vector-valued random field on $\mathbb{Z}^d$ and the $g_{ij}$ are some real numbers. We assume furthermore that the coefficients $g_{ij}$ and the innovations $\varepsilon_i$ satisfy, respectively,

$$\lim_{s \to \infty} \sup_{i \in \mathbb{Z}^d} \sum_{j \in \mathbb{Z}^d, \rho(i,j) > s} |g_{ij}| = 0,$$

and

$$\sup_{i \in \mathbb{Z}^d} \|\varepsilon_i\|_p < \infty \text{ where } p \geq 1.$$

Linear stationary random fields on $\mathbb{Z}^2$ have been explored by Whittle (1954) in a paper that significantly influenced the subsequent modeling of spatial dependencies. More recently, linear random fields on $\mathbb{Z}^d, d \geq 1$, have been studied by Doukhan and Guyon (1991), see also Doukhan (1994). Analogous to the time series literature, $Y_i$ is defined as the limit in $L_p$ as $s \to \infty$ of the partial sums $Y_i^s = \sum_{j \in \mathbb{Z}^d, \rho(i,j) \leq s} g_{ij} \varepsilon_j$. The following existence results is proven by Doukhan (1994), pp. 75-81.\(^6\)

\(^6\)In fact, Doukhan (1994) proves a more general result that also covers the case $p < 1$. 

6
Proposition 3 (Doukhan, 1994) Under conditions (5)-(6) the distribution of linear random field \( Y \) in (4) is well defined. In particular, the finite dimensional distributions of \( Y \) are limits in \( L_p \) as \( s \to \infty \) of the those of the random fields \( Y^s = \{ Y^s_i, i \in \mathbb{Z}^d \} \).

We next give a result that shows that under the same conditions, the linear field \( Y \) is \( L_p \)-NED on the field \( \varepsilon \).

Proposition 4 Under conditions (5)-(6) the linear random field \( Y \) in (4) is uniformly \( L_p \)-NED on \( \varepsilon \).

The nice feature of this result is that it does not impose any additional conditions over and above those employed in Proposition 3 to establish the existence of the linear field.

3.2 Autoregressive Random Fields

The linear random fields of the previous example may arise as solutions of autoregressive spatial models. For example, Whittle (1954) considered the following simple autoregressive model on \( \mathbb{Z}^2 \):

\[
Y_{i_1,i_2} = \alpha Y_{i_1+1,i_2} + \beta Y_{i_1,i_2+1} + \varepsilon_{i_1,i_2}
\]

where \( |\alpha| + |\beta| < 1 \) and \( \varepsilon = \{ \varepsilon_{i_1,i_2}, (i_1,i_2) \in \mathbb{Z}^2 \} \) are i.i.d. random variables with zero mean and finite variance. Whittle (1954), p. 447, gives the following stationary solution of (7):

\[
Y_{i_1,i_2} = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \binom{j_1 + j_2}{j_1} \alpha^{j_1} \beta^{j_2} \varepsilon_{i_1+j_1,i_2+j_2}
\]

Thus, (8) is a linear random field with

\[
g_{ij} = \binom{j_1 - i_1 + j_2 - i_2}{j_1 - i_1} \alpha^{j_1 - i_1} \beta^{j_2 - i_2},
\]

for \( j = (j_1,j_2) \geq i = (i_1,i_2) \) and zero, otherwise. Furthermore, note that the assumptions of Proposition 4 are satisfied for \( p = 2 \), since \( \sum_{j \geq i} |g_{ij}| = \sum_{k=0}^{\infty} \sum_{l=0}^{k} \binom{k}{l} |\alpha|^l |\beta|^{k-l} = \sum_{k=0}^{\infty} (|\alpha| + |\beta|)^k < \infty \), and hence

\[
0 \leq \lim_{s \to \infty} \sup_{i \in \mathbb{Z}^2} \sum_{j \geq i, \rho(i,j) > s} |g_{ij}| \leq \lim_{s \to \infty} \sup_{k=s} \sum_{k=0}^{\infty} (|\alpha| + |\beta|)^k = 0,
\]

observing that \( |\alpha| + |\beta| < 1 \).
3.3 Cliff-Ord Type Spatial Processes

Consider the following Cliff-Ord type model:

\[ Y_n = \lambda W_n Y_n + X_n \beta + u_n \]
\[ u_n = \rho M_n u_n + v_n \]

where \( Y_n = (Y_{1,n}, \ldots, Y_{n,n})' \) is an \( n \times 1 \) vector of endogenous variables, \( X_n = (X_{ij,n}) \) is \( n \times k \) matrix of regressors, \( W_n = (w_{ij,n}) \) and \( M_n = (m_{ij,n}) \) are \( n \times n \) nonstochastic weight matrices, \( v_n = (v_{1,n}, \ldots, v_{n,n})' \) is an \( n \times 1 \) vector of disturbances, \( \lambda \) and \( \rho \) are scalar parameters typically referred to as spatial autoregressive parameters, and \( \beta \) is a \( k \times 1 \) parameter vector. To simplify presentation, we assume in the following that \( k = 1 \) and use the notation \( X_n = (X_{1,n}, \ldots, X_{n,n})' \).

The above model is frequently referred to as a spatial ARAR(1,1) model, and estimation strategies for this model see, e.g., Robinson (2009, 2007), Kelejian and Prucha (2007, 2004, 2007a,b), and Lee (2007a,b, 2004).

The spatial weights \( m_{ij,n} \) and \( w_{ij,n} \) are typically thought of as to depend on some measure of distance, say \( d_{ij} \), between units \( i \) and \( j \), where the weights decline as the distance increases. Thus although in Cliff-Ord models observations are indexed by natural numbers, i.e., \( i = 1, \ldots, n \), a leading case would be where \( i \) correspond to some location, say \( s_i \in \mathbb{R}^d \), and the distance between \( i \) and \( j \) is given by \( d_{ij} = \rho(s_i, s_j) \). In the following, we assume that \( D_n = \{s_1, \ldots, s_n\} \subset D \), where the lattice \( D \) satisfies Assumption 1. However, to simplify notation we will use \( \rho(i, j) \) instead of \( \rho(s_i, s_j) \). We assume furthermore that for all \( n \)

\[ I_n - \lambda W_n \text{ and } I_n - \rho M_n \text{ are nonsingular.} \] (10)

The reduced form of the model is then given by \( Y_n = A_n X_n \beta + B_n v_n \), with \( A_n = (a_{ij,n}) = (I_n - \lambda W_n)^{-1} \) and \( B_n = (b_{ij,n}) = (I_n - \lambda W_n)^{-1} (I_n - \rho M_n)^{-1} \). In scalar notation we have for \( i = 1, \ldots, n \):

\[ Y_{i,n} = \beta \sum_{j=1}^{n} a_{ij,n} X_{j,n} + \sum_{j=1}^{n} b_{ij,n} v_{j,n} \]

Although for fixed \( n \), the output process \( Y_{i,n} \) depends on only a finite number of elements of the input process \( \varepsilon_{i,n} = (X_{i,n}, v_{i,n})' \), the mixing property of \( \varepsilon_{i,n} \) may not carry over to \( Y_{i,n} \). The reason is that the number of elements composing the spatial lags grows unboundedly with the sample size so that the mixing property can break down in the limit. This is especially important when analyzing the asymptotic properties of Cliff-Ord type processes.

Towards establishing that \( Y = \left\{ Y_{i,n}, s_i \in D_n, n \geq 1 \right\} \) is NED on \( \varepsilon = \left\{ \varepsilon_{i,n}, s_i \in D_n, n \geq 1 \right\} \), we maintain the following assumptions:

\[ \lim_{n \to \infty} \sup_{1 \leq i \leq n} \sum_{1 \leq j \leq n, \rho(i,j) > s} |a_{ij,n}| + |b_{ij,n}| = 0 \] (11)
and
\[ \sup_n \sup_{1 \leq i \leq n} \| \varepsilon_{i,n} \|_p < \infty \text{ for some } p \geq 1. \] (12)

We have the following result.

**Proposition 5** Under conditions (10)-(12), the process \( Y = \{Y_{i,n}, s_i \in D_n, n \geq 1\} \), where \( Y_{i,n} \) is defined by (9), is uniformly \( L_p\)-NED on the process \( \varepsilon = \{\varepsilon_{i,n}, s_i \in D_n, n \geq 1\} \), where \( \varepsilon_{i,n} = (X_{i,n}, v_{i,n})' \).

It is readily seen that a sufficient condition for (11) is that for some \( \gamma > 0 \),
\[ \sup_n \sup_{1 \leq i \leq n} \sum_{j=1}^n (|a_{ij,n}| + |b_{ij,n}|) \rho(i,j) \gamma < \infty \] (13)
since
\[
\begin{align*}
\sup_n \sup_{1 \leq i \leq n} \sum_{1 \leq j \leq n, \rho(i,j) > s} |a_{ij,n}| + |b_{ij,n}| & \leq \sup_n \sup_{1 \leq i \leq n} \sum_{1 \leq j \leq n, \rho(i,j) > s} (|a_{ij,n}| + |b_{ij,n}|) \frac{\rho(i,j) \gamma}{s^\gamma} \\
& \leq \frac{1}{s^\gamma} \sup_n \sup_{1 \leq i \leq n} \sum_{j=1}^n (|a_{ij,n}| + |b_{ij,n}|) \rho(i,j) \gamma \to 0 \text{ as } s \to \infty.
\end{align*}
\]

A condition analogous to (13) has been used recently by Kelejian and Prucha (2007a), and should be satisfied in a wide range of applications. It is slightly stronger than the typical assumption in the Cliff-Ord literature which maintains assumptions that imply that the row and column sums of the absolute elements of the matrices \( A_n \) and \( B_n \) are uniformly bounded; compare, e.g., Kelejian and Prucha (2007b) and Lee (2007a,b, 2004).

### 3.4 Spatial Bernoulli Shifts

Consider a real-valued random field \( Y = \{Y_i, i \in \mathbb{Z}^d\} \) defined as:
\[ Y_i = f_i(\varepsilon_{i+j}, j \in \mathbb{Z}^d) \] (14)
where \( \varepsilon = \{\varepsilon_i, i \in \mathbb{Z}^d\} \) is a real-valued random field and the \( f_i : \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{R} \) are measurable functions. The process (14) is a generalization of one-parameter Bernoulli shifts considered by Doukhan and Louhichi (1999). A simple example of a spatial Bernoulli shift is the infinite moving average random field of the first example. More generally, the \( f_i \) may be nonlinear functions. We assume that the \( f_i \) satisfy the following Lipschitz-type regularity condition:
\[ |f_i(x_j, j \in \mathbb{Z}^d) - f_i(y_j, j \in \mathbb{Z}^d)| \leq \sum_{j \in \mathbb{Z}^d} w_{ij} |x_j - y_j| \] (15)
for some nonnegative constants $w_{ij}$ such that

$$\lim_{s \to \infty} \sup_{i \in \mathbb{Z}^d} \sum_{j \in \mathbb{Z}^d; \rho(i, j) > s} w_{ij} = 0. \quad (16)$$

Finally, we assume that the input process $\varepsilon = \{\varepsilon_i, i \in \mathbb{Z}^d\}$ has uniformly bounded second moments, i.e.,

$$\sup_{i \in \mathbb{Z}^d} \|\varepsilon_i\|_2 < \infty \quad (17)$$

Then, one can establish the following result.

**Proposition 6** Under conditions (15)-(17), the random field $Y = \{Y_i, i \in \mathbb{Z}^d\}$ given by (14) is well-defined and uniformly $L_2$-NED on the random field $\varepsilon = \{\varepsilon_i, i \in \mathbb{Z}^d\}$.

Conditions (15)-(17) are analogous to those used by Doukhan and Louhichi (1999) for time-series Bernoulli shifts. Condition (15) is fulfilled if the functions $f_i$ have bounded partial derivatives, i.e., $|\partial f_i / \partial x_j| \leq w_{ij}$.

Thus, the class of NED spatial processes covers fairly large classes of linear and nonlinear transformations of random fields.

## 4 Limit Theorems

### 4.1 Central Limit Theorem

In the following, we introduce our CLT for real valued random fields $Z = \{Z_{i,n}, i \in D_n, n \geq 1\}$, where $Z$ is assumed to be $L_2$-NED on some vector-valued $\alpha$-mixing random field $\varepsilon = \{\varepsilon_{i,n}, i \in T_n, n \geq 1\}$ with the NED coefficients $\{\psi(s)\}$ and scaling factors $\{d_{i,n}\}$, where $D_n \subseteq T_n \subseteq D$ and the lattice $D$ satisfies Assumption 1. In the sequel, we will use the following notation:

$$S_n = \sum_{i \in D_n} Z_{i,n}; \sigma_n^2 = \text{var}(S_n).$$

For ease of reference, we state below the definition of the $\alpha$-mixing coefficients employed in the paper.

**Definition 2** Let $\mathcal{A}$ and $\mathcal{B}$ be two $\sigma$-algebras of $\mathcal{F}$, and let

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup(\|P(AB) - P(A)P(B)\|, A \in \mathcal{A}, B \in \mathcal{B}),$$

For $U \subseteq D_n$ and $V \subseteq D_n$, let $\sigma_n(U) = \sigma(\varepsilon_{i,n}; i \in U)$ and $\alpha_n(U,V) = \alpha(\sigma_n(U), \sigma_n(V))$. Then, the $\alpha$-mixing coefficients for the random field $\varepsilon$ are defined as:

$$\overline{\alpha}(u, v, r) = \sup_n \sup_{U,V}(\alpha_n(U,V), |U| \leq u, |V| \leq v, \rho(U,V) \geq r).$$

10
Dobrushin (1968) showed that weak dependence conditions based on the above mixing coefficients are satisfied by broad classes of random fields including Markov fields. In contrast to standard mixing numbers for time-series processes, the mixing coefficients for random fields depend not only on the distance between two datasets but also their sizes. To explicitly account for such dependence, it is furthermore assumed that

$$\exists(u, v, r) \leq \varphi(u, v)\tilde{a}(r)$$

(18)

where the function $\varphi(u, v)$ is nondecreasing in each argument, and $\tilde{a}(r) \to 0$ as $r \to \infty$. The idea is to account separately for the two different aspects of dependence: (i) decay of dependence with the distance, and (ii) accumulation of dependence as the sample region expands. The two common choices of $\varphi(u, v)$ in the random fields literature are

$$\varphi(u, v) = (u + v)^\tau, \tau \geq 0,$$

(19)

and

$$\varphi(u, v) = \min\{u, v\}.$$  

(20)

The above mixing conditions have been used extensively in the random fields literature including Neaderhouser (1978), Takahata (1983), Nahapetian (1987), Bulinskii (1989), Bulinskii and Doukhan (1990), Tran (1990) and Bradley (1993). They are satisfied by fairly large classes of random fields. Bradley (1993) provides examples of random fields satisfying conditions (18)-(19) with $u = v$ and $\tau = 1$. Furthermore, Bulinskii (1989) constructs infinite moving average random fields satisfying the same conditions with $\tau = 1$ for any given decay rate of coefficients $\tilde{a}(r)$. Clearly, standard mixing coefficients in the time series literature are covered by conditions (18)-(19) when $\tau = 0$.

Following the literature, we employ the above mixing conditions for the input random field, and impose the following restriction on the decay rates of the mixing coefficients.

**Assumption 2** The $\alpha$-mixing coefficients of $\varepsilon$ satisfy (18) and (19) for some $\tau \geq 0$ and $\tilde{a}(r)$, such that for some $\delta > 0$

$$\sum_{r=1}^{\infty} r^{d(\tau + \delta - 1)} \frac{1}{\tilde{a}^2(2\tau + \delta)} (r) < \infty,$$

(21)

where $\tau = \delta/2 + \delta$.

Note that if $\tau = 1$ this assumption also covers the case where $\varphi(u, v)$ is given by (20). Furthermore, the CLT relies on the following conditions regarding moments, the NED coefficients and the NED scaling factors.

**Assumption 3 (a) (Uniform $L_{2+\delta}$ integrability)** There exists an array of positive constants $\{c_{i,n}\}$ such that

$$\lim_{k \to \infty} \sup_{n} \sup_{i \in D_n} E[|Z_{i,n}/c_{i,n}|^{2+\delta} \mathbf{1}(|Z_{i,n}/c_{i,n}| > k)] = 0,$$

where $\mathbf{1}(\cdot)$ is the indicator function and $\delta > 0$ is as in Assumption 2.
(b) $\inf_n |D_n|^{-1}M_n^{-2}\sigma_n^2 > 0$, where $M_n = \max_{i \in D_n} c_{i,n}$ and $\{D_n\}$ is a sequence of finite subsets of $D$ such that $|D_n| \to \infty$ as $n \to \infty$.

(c) NED coefficients satisfy $\sum_{r=1}^{\infty} r^{d-1}\psi(r) < \infty$.

(d) NED scaling factors satisfy $\sup_n \sup_{i \in D_n} c_{i,n}^{-1}d_{i,n} \leq C < \infty$.

We can now state the following CLT for $L^2$-NED random fields.

**Theorem 1** Suppose $\{D_n\}$ is a sequence of finite subsets such that $|D_n| \to \infty$ as $n \to \infty$ and $\{T_n\}$ is a sequence of subsets such that $D_n \subseteq T_n \subseteq D$ of the lattice $D$ satisfying Assumption 1. Let $Z = \{Z_{i,n}, i \in D_n, n \geq 1\}$ be a real valued zero-mean random field that is $L^2$-NED on a vector-valued $\alpha$-mixing random field $\varepsilon = \{\varepsilon_{i,n}, i \in T_n, n \geq 1\}$. Suppose Assumptions 2 and 3 hold, then

$$\sigma_n^{-1}S_n \implies N(0,1).$$

We note that Theorem 1 contains as a special case the CLT for time series NED processes of Wooldridge (1986), see Theorem 3.13 and Corollary 4.4.

Assumption 2 restricts the dependence structure of the input process $\varepsilon$.

Assumptions 3(a),(b) are standard in the limit theory of mixing processes, e.g., Wooldridge (1986), Davidson (1992), and de Jong (1997), and Jenish and Prucha (2009).

Assumption 3(a) is satisfied if the $Z_{i,n}/c_{i,n}$ are uniformly $L_p$-bounded for some $p > 2 + \delta$, i.e., $\sup_n \sup_{i \in D_n} \|Z_{i,n}/c_{i,n}\|_p < \infty$. The nonrandom constants $c_{i,n}$ allow for processes with asymptotically unbounded moments. Such processes are encountered frequently in applications. For instance, Bera and Simlai (2005) report on sharp spikes in the variances of housing prices in Boston. The constants $c_{i,n}$ can be thought of as upper bounds on the moments of the individual terms, i.e., $\|Z_{i,n}\|_p \leq c_{i,n} < \infty$. If $\sup_n \sup_{i \in D_n} \|Z_{i,n}\|_p \leq M < \infty$, the constants $c_{i,n}$ can be set to 1.

Assumption 3(b) is an asymptotic negligibility condition that ensures that no single summand influences disproportionately the entire sum. In the case of uniformly $L_{2+\delta}$-bounded fields, 3(b) reduces to $\liminf_{n \to \infty} |D_n|^{-1}\sigma_n^2 > 0$, as is, e.g., maintained in Bolthausen (1982).

Assumption 3(c) controls the size of the NED coefficients which measure the error in the approximation of $Z_{i,n}$ by $\varepsilon$. Intuitively, the approximation errors have to decline sufficiently fast with each successive approximation. Assumption 3(c) is satisfied if $\psi(r) = O(r^{d-\gamma})$ for some $\gamma > 0$, i.e., $\psi(r)$ is of size $-d$.

Finally, Assumption 3(d) is a technical condition, which ensures that the order of magnitude of the NED scaling factors does not exceed that of the $2+\delta$ moments. For instance, suppose the constant $c_{i,n}$ can be chosen as $c_{i,n} = \|Z_{i,n}\|_{2+\delta}$, and the NED scaling numbers as $d_{i,n} \leq 2\|Z_{i,n}\|_2$. Then Assumption 3(d) is satisfied, since by Lyapunov’s inequality, $\|Z_{i,n}\|_2 \leq \|Z_{i,n}\|_{2+\delta}$. This condition has also been used by de Jong (1997) and Davidson (1992).
Theorem 1 can be easily extended to vector-valued fields using the standard Cramér-Wold device.

**Corollary 1** Suppose \( \{D_n\} \) is a sequence of finite subsets such that \( |D_n| \to \infty \) as \( n \to \infty \) and \( \{T_n\} \) is a sequence of subsets such that \( D_n \subseteq T_n \subseteq D \) of the lattice \( D \) satisfying Assumption 1. Let \( Z = \{Z_{i,n}, i \in D_n, n \geq 1\} \) with \( Z_{i,n} \in \mathbb{R}^k \) be a zero-mean random field that is \( L_2 \)-NED on a vector-valued \( \alpha \)-mixing random field \( \varepsilon = \{\varepsilon_{i,n}, i \in T_n, n \geq 1\} \). Suppose Assumptions 2 and 3 hold with \( |Z_{i,n}| \) denoting the Euclidean norm of \( Z_{i,n} \) and \( \sigma_n^2 \) replaced by \( \lambda_{\min}(\Sigma_n) \), where \( \Sigma_n = \text{Var}(S_n) \) and \( \lambda_{\min}(\cdot) \) is the smallest eigenvalue, then

\[
\Sigma_n^{-1/2}S_n \Rightarrow N(0, I_k).
\]

Furthermore, \( \sup_n |D_n|^{-1} \lambda_{\max}(\Sigma_n) < \infty \), where \( \lambda_{\max}(\cdot) \) denotes the largest eigenvalue.

### 4.2 Law of Large Numbers

To prove consistency of spatial estimators, we also derive a weak LLN for random fields which are for \( L_1 \)-NED on some mixing field. This LLN can then be used to establish uniform convergence of random functions by combining it with the generic ULLN given in Jenish and Prucha (2009), which transforms pointwise LLNs (at a given parameter value) into ULLNs. The LLN maintains the following moment and mixing assumptions.

**Assumption 4** (a) There exist nonrandom positive constants \( \{c_{i,n}, i \in D_n, n \geq 1\} \) such that \( Z_{i,n}/c_{i,n} \) is uniformly \( L_p \)-bounded for some \( p > 1 \), i.e., \( \sup_n \sup_{i \in D_n} \mathbb{E} |Z_{i,n}/c_{i,n}|^p < \infty \).

(b) The \( \alpha \)-mixing coefficients of the input field \( \varepsilon \) satisfy (18) for some function \( \varphi(u,v) \) which is nondecreasing in each argument, and some \( \hat{\alpha}(r) \) such that \( \sum_{r=1}^{\infty} r^{d-1} \hat{\alpha}(r) < \infty \).

We now have the following weak LLN.

**Theorem 2** Let \( \{D_n\} \) be a sequence of arbitrary finite subsets of \( D \) such that \( |D_n| \to \infty \) as \( n \to \infty \), where \( D \subseteq \mathbb{R}^d \), \( d \geq 1 \) is as in Assumption 1, and let \( T_n \) be a sequence of subsets of \( D \) such that \( D_n \subseteq T_n \). Suppose further that \( Z = \{Z_{i,n}, i \in D_n, n \geq 1\} \) is \( L_1 \)-NED on \( \varepsilon = \{\varepsilon_{i,n}, i \in T_n, n \geq 1\} \) with the scaling factors \( d_{i,n} \). If \( Z \) and \( \varepsilon \) satisfy Assumption 4, then

\[
\frac{1}{M_n|D_n|} \sum_{i \in D_n} (Z_{i,n} - EZ_{i,n}) \overset{L_1}{\to} 0,
\]

where \( M_n = \max_{i \in D_n} \max(c_{i,n}, d_{i,n}) \).
We note that $Z$ and $\varepsilon$ can be vector-valued random fields. In this case, $|Z_{i,n}|$ should be understood as the Euclidean norm in the respective vector-space.

Assumption 4(a) is a standard moment condition employed in weak LLNs for dependent processes. It requires existence of moments of order slightly greater than 1. As in Theorem 1, $c_{i,n}$ and $d_{i,n}$ are the scaling factors that reflect the magnitudes of potentially trending moments. The case of variables with uniformly bounded moments is covered by setting $c_{i,n} = d_{i,n} = 1$. Assumption 4(b) is weaker than the mixing conditions used for the CLT. Furthermore, the LLN does not require any restrictions on the NED coefficients.

In the time series literature, weak LLNs for NED processes have been obtained by Andrews (1988) and Davidson (1993), among others. Andrews (1988) derives an $L_1$-law for triangular arrays of $L_1$-mixingales. He then shows that NED processes are $L_1$-mixingales, and hence, satisfy his LLN. Davidson (1993) extends the latter result to processes with trending moments.

5 Large Sample Properties of Spatial GMM Estimators

In this section, we apply the above developed limit theorems to establish the large sample properties of spatial GMM estimators under a reasonably general set of assumptions that should cover a wide range of empirical problems. More specifically, our consistency and asymptotic normality results (i) maintain only that the spatial data process is NED on an $\alpha$-mixing basis process to accommodate spatial lags in the data process as discussed above, (ii) allow for the data process to be located on an unevenly spaced grid, and (iii) allow for the data process to be non-stationary, which will frequently be the case in empirical applications. We also give our results under a set of primitive sufficient conditions for easier interpretation by the applied researcher.\footnote{In an important contribution, Conley (1999) gives a first set of results regarding the asymptotic properties of GMM estimators under the assumption that the data process is stationary and $\alpha$-mixing. Conley also maintains some high level assumption such as first moment continuity of the moment function, which in turn immediately implies uniform convergence - see, e.g., Pötscher and Prucha (1989) for a discussion. Our results extend Conley (1999) in several important directions, as indicated above. We establish uniform convergence from primitive sufficient conditions via the generic uniform law of large numbers given in Jenish and Prucha (2009) and the law of large numbers given as Theorem 2 above.}

We continue with the basic set-up of Section 2. Consider the moment function $q_{i,n} : \mathbb{R}^p \times \Theta \to \mathbb{R}^p$, where $\Theta$ denotes the parameter space, and let $\theta^0_n \in \Theta$ denote the parameter vector of interest (which we allow to depend on $n$ for reasons of generality). Suppose the following moment conditions hold

$$E q_{i,n}(Z_{i,n}, \theta^0_n) = 0,$$

then the corresponding spatial GMM estimator is defined as
\( \hat{\theta}_n = \arg\min_{\theta \in \Theta} Q_n(\omega, \theta), \)  

(23)

where \( Q_n : \Omega \times \Theta \to \mathbb{R}, \)

\[
Q_n(\omega, \theta) = R_n(\theta)^t P_n R_n(\theta), \\
R_n(\theta) = |D_n|^{-1} \sum_{i \in D_n} q_{i,n}(Z_{i,n}, \theta),
\]

where the \( P_n \) are a positive definite weighting matrices. To prove that \( \hat{\theta}_n \) is a consistent estimator for \( \theta_n^0 \) consider the following non-stochastic analogue of \( Q_n \), say

\[
\overline{Q}_n(\theta) = [ER_n(\theta)]^t P [ER_n(\theta)],
\]

(24)

where \( P \) denotes the probability limit of \( P_n \). Then given the moment condition (22) clearly \( R_n(\theta_n^0) = 0 \), and the functions \( \overline{Q}_n \) are minimized at \( \theta_n^0 \) with \( \overline{Q}_n(\theta_n^0) = 0 \). In proving consistency, we follow the classical approach; see, e.g., Gallant and White (1988) or Pötscher and Prucha (1997) for more recent expositions. In particular, given identifiable uniqueness of \( \theta_n^0 \) we establish, loosely speaking, convergence of the minimizers \( \hat{\theta}_n \) to the minimizers \( \theta_n^0 \) by establishing convergence of the objective function \( Q_n(\omega, \theta) \) to its non-stochastic analogue \( \overline{Q}_n(\theta) \) uniformly over the parameter space.

Throughout the sequel, we maintain the following assumptions regarding the parameter space, the GMM objective function and the unknown parameters \( \theta_n^0 \).

**Assumption 5 (a)** The parameter space \( \Theta \) is a compact metric space with metric \( \nu \).

(b) The functions \( q_{i,n} : \mathbb{R}^p \times \Theta \to \mathbb{R}^p \) are \( \mathcal{B}^p / \mathcal{B}^p \)-measurable for each \( \theta \in \Theta \), and continuous on \( \Theta \) for each \( z \in \mathbb{R}^p \).

(c) The elements of the \( p_q \times p_q \) real matrices \( P_n \) are \( \mathcal{B} \)-measurable, and \( P_n \) is positive definite a.s. Furthermore \( P = \text{p} \lim P_n \) exists and \( P \) is positive definite.

(d) The minimizers \( \theta_n^0 \) are identifiably unique in the sense that for every \( \varepsilon > 0 \),

\[
\liminf_{n \to \infty} [\inf_{\theta \in \Theta: \varepsilon(\theta, \theta_n^0) \geq \varepsilon} [ER_n(\theta)]^t [ER_n(\theta)]] > 0.
\]

Compactness of the parameter space as maintained in Assumption 5(a) is typical for the GMM literature. Assumptions 5(b),(c) imply that \( Q_n(\cdot, \theta) \) is measurable for all \( \theta \in \Theta \), and \( Q_n(\omega, \cdot) \) is continuous on \( \Theta \). Given those assumptions the existence of measurable functions \( \hat{\theta}_n \) that solves (23) follows, e.g., from Lemma A3 of Pötscher and Prucha (1997).

Since \( P \) is positive definite, it is readily seen that Assumption 5(d) implies that for every \( \varepsilon > 0 \):

\[
\lim \inf_{n \to \infty} \left[ \inf_{\theta \in \Theta: \varepsilon(\theta, \theta_n^0) \geq \varepsilon} [\overline{Q}_n(\theta) - \overline{Q}_n(\theta_n^0)] \right] > 0,
\]
observing that \( \overline{Q}_n(\theta_n^0) = 0 \). Thus under Assumption 5(d) the minimizers \( \theta_n^0 \) are identifiably unique; compare, e.g., Gallant and White (1988), p.19. For interpretation consider the important special case where \( \theta_n^0 = \theta^0 \), \( R_n(\theta) = R(\theta) \) and \( R(\theta) \) is continuous. In this case identifiability of \( \theta^0 \) is equivalent to the assumption that \( \theta^0 \) is the unique solution of the moment conditions, i.e., \( R_n(\theta) \neq 0 \) for all \( \theta \neq \theta^0 \); compare, e.g., Pötscher and Prucha (1997), p. 16.

5.1 Consistency

Given the minimizers \( \theta_n^0 \) are identifiably unique, \( \hat{\theta}_n \) is a consistent estimator for \( \theta_n^0 \) if \( Q_n \) converges uniformly to \( \overline{Q}_n \), i.e., if \( \sup_{\theta \in \Theta} |Q_n(\omega, \theta) - \overline{Q}_n(\theta)| \overset{p}{\to} 0 \) as \( n \to \infty \); this follows immediately from, e.g., Pötscher and Prucha (1997), Lemma 3.1.

We now proceed by giving a set of primitive domination and Lipschitz type conditions for the moment functions that ensure uniform convergence of \( Q_n \) to \( \overline{Q}_n \). The conditions are in line with those maintained in the general literature on M-estimation, e.g., Andrews (1987), Gallant and White (1988), and Pötscher and Prucha (1989, 1994).

**Definition 3** Let \( f_{i,n} : \mathbb{R}^{p_z} \times \Theta \to \mathbb{R}^{p_q} \) be \( \mathcal{B}^{p_z}/\mathcal{B}^{p_q} \)-measurable functions for each \( \theta \in \Theta \), then:

(a) The random functions \( f_{i,n}(Z_{i,n}; \theta) \) are said to be \( p \)-dominated on \( \Theta \) for some \( p > 1 \) if \( \sup_n \sup_{i \in D_n} E \sup_{\theta \in \Theta} |f_{i,n}(Z_{i,n}; \theta)|^p < \infty \).

(b) The random functions \( f_{i,n}(Z_{i,n}; \theta) \) are said to be Lipschitz in the parameter \( \theta \) on \( \Theta \) if

\[
|f_{i,n}(Z_{i,n}; \theta) - f_{i,n}(Z_{i,n}; \theta^*)| \leq L_{i,n}(Z_{i,n}) h(\nu(\theta, \theta^*)) \text{ a.s.,}
\]

for all \( \theta, \theta^* \in \Theta \) and \( i \in D_n \), \( n \geq 1 \), where \( h \) is a nonrandom function with \( h(x) \downarrow 0 \) as \( x \downarrow 0 \), and \( L_{i,n} \) are random variables with

\[
\limsup_{n \to \infty} |D_n|^{-1} \sum_{i \in D_n} EL_{i,n}^\eta < \infty \text{ for some } \eta > 0.
\]

Towards establishing consistency of \( \hat{\theta}_n \) we furthermore maintain the following moment and mixing assumptions.

**Assumption 6** The moment functions \( q_{i,n}(Z_{i,n}; \theta) \) have the following properties:

(a) They are \( p \)-dominated on \( \Theta \) for \( p = 2 \).

(b) They are uniformly \( L_1 \)-NED on \( \varepsilon = \{\varepsilon_{i,n}, i \in T_n, n \geq 1\} \), where \( D_n \subseteq T_n \subseteq D \), and \( \varepsilon \) is \( \alpha \)-mixing with \( \alpha \)-mixing coefficients the conditions stated in Assumption 4(b).

16
Assumption 6(a) implies that \( \sup_{n,i \in D_n} E |q_{i,n}(Z_{i,n}; \theta)|^p < \infty \) for each \( \theta \in \Theta \). Assumptions 6(b) then allow us to apply the LLN given as Theorem 2 above the sample moments \( R_n(\theta) = [D_n]^{-1} \sum_{i \in D_n} q_{i,n}(Z_{i,n}, \theta) \).

To verify Assumption 6(b) one can use either Proposition 1 or Proposition 2 to imply this condition from the lower level assumption that the data \( Z_{i,n} \) are \( L_1 \)-NED. For example, the \( q_{i,n} \) are \( L_1 \)-NED, if the \( Z_{i,n} \) are \( L_1 \)-NED and satisfy the Lipschitz condition of Proposition 1. Note that no restrictions on the sizes of the NED coefficients are required.

Assumption 6(c) ensures stochastic equicontinuity of \( q_{i,n} \) w.r.t. \( \theta \). Stochastic equicontinuity jointly with Assumption 6(a) and the pointwise LLN enable us to invoke the ULLN of Jenish and Prucha (2009) to prove uniform convergence of the sample moments, which in turn is used to establish that \( Q_n \) converges uniformly to \( \overline{Q}_n \). A sufficient condition for Assumption 6(c) is existence of integrable partial derivatives of \( q_{i,n} \) w.r.t. \( \theta \) if \( \theta \in \mathbb{R}^k \).

Our consistency results for the spatial GMM estimator given by (23) is summarized by the next theorem.

**Theorem 3** (Consistency) Suppose \( \{D_n\} \) is a sequence of finite sets of \( D \) such that \( |D_n| \to \infty \) as \( n \to \infty \), where \( D \subset \mathbb{R}^d, d \geq 1 \) as in Assumption 1. Suppose further that Assumptions 5 and 6 hold. Then

\[
\nu(\hat{\theta}_n, \theta^*) \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty,
\]

and \( \overline{Q}_n(\theta) \) is uniformly equicontinuous on \( \Theta \).

### 5.2 Asymptotic Normality

We next establish that the spatial GMM estimators defined by (23) is asymptotically normally distributed. For that purpose, we need a stronger set of assumptions than for consistency, including differentiability of the moment functions in \( \theta \). It proofs helpful to adopt the notation \( \nabla_\theta \) in place of \( \partial/\partial \theta \).

**Assumption 7 (a)** The minimizers \( \theta^*_n \) lie uniformly in the interior of \( \Theta \) with \( \Theta \subseteq \mathbb{R}^k \). Furthermore \( E [R_n(\theta^*_n)] = 0 \).

(b) The functions \( q_{i,n} : \mathbb{R}^p \times \Theta \to \mathbb{R}^p \) are continuously differentiable w.r.t. \( \theta \) in the interior of \( \Theta \) for each \( z \in \mathbb{R}^p \).

---

8To ensure that the derivatives are defined on the border of \( \Theta \), we assume in the following that the moment functions are defined on an open set containing \( \Theta \), and that the \( q_{i,n} \) and \( \nabla_\theta q_{i,n} \) are restrictions to \( \Theta \).
(c) The functions $q_{i,n}(Z_{i,n}; \theta_n^o)$ are uniformly $L_2$-NED on $\varepsilon$ of size $-d$, and $\sup_{n,i\in D} E |q_{i,n}(Z_{i,n}; \theta_n^o)|^{2+\delta'} < \infty$ for some $\delta' > 0$. The functions $\nabla_{\theta} q_{i,n}(Z_{i,n}; \theta)$ are uniformly $L_1$-NED on $\varepsilon$.

(d) The input process $\varepsilon = \{\varepsilon_{i,n}, i \in T_n, n \geq 1\}$, where $D_n \subseteq T_n \subseteq D$, is $\alpha$-mixing and the mixing coefficients satisfy Assumption 2 for some $\delta < \delta'$, where $\delta'$ is the same as in Assumption 7(c).

(e) The functions $\nabla_{\theta} q_{i,n}$ are $p$-dominated on $\Theta$ for some $p > 1$.

(f) The functions $\nabla_{\theta} q_{i,n}$ are Lipschitz in $\theta$ on $\Theta$.

(g) $\inf_n \lambda_{\min} \left( [D_n^{-1} \Sigma_n] \right) > 0$ where $\Sigma_n = \text{Var} \left[ \sum_{i\in D_n} q_{i,n}(Z_{i,n}; \theta_n^o) \right]$.

(h) $\inf_n \lambda_{\min} \left[ E \nabla_{\theta} R_n(\theta_n^o) \nabla_{\theta} R_n(\theta_n^o) \right] > 0$.

The first part of Assumption 7(a) is needed to ensure that the estimator $\hat{\theta}_n$ lies in the interior of $\Theta$ with probability tending to one, and facilitates the application of the mean value theorem to $R_n(\hat{\theta}_n)$ around $\theta_n^o$. The second part states in essence that the moment conditions are correctly specified. Its violation will generally invalidate the limiting distribution result.

Assumptions 7(c),(d),(g) enable us to apply the CLT for vector-valued NED processes given above as Corollary 1 to $R_n(\theta_n^o)$. Some low level sufficient conditions for Assumption 7(c) are given below. To establish asymptotic normality, we also need uniform convergence of $\nabla_{\theta} R_n$ on $\Theta$, which is implied via Assumptions 7(c),(d),(e),(f).

Given the above assumptions, we have the following asymptotic normality result for the spatial GMM estimator defined by (23).

**Theorem 4** Suppose $\{D_n\}$ is a sequence of finite sets of $D$ such that $|D_n| \to \infty$ as $n \to \infty$, where $D \subset \mathbb{R}^d, d \geq 1$ is as in Assumption 1. Suppose further that Assumptions 5-7 hold. Then

\[(A_n^{-1} B_n B_n' A_n^{-1})^{-1/2} |D_n|^{1/2} \left( \hat{\theta}_n - \theta_n^o \right) \Rightarrow N(0, I_k),\]

where

\[A_n = \left[ E \nabla_{\theta} R_n(\theta_n^o) \right]' P \left[ E \nabla_{\theta} R_n(\theta_n^o) \right] \quad \text{and} \quad B_n = \left[ E \nabla_{\theta} R_n(\theta_n^o) \right]' P \left( [D_n^{-1} \Sigma_n] \right)^{1/2}.

Moreover, $|A_n| = O(1)$; $|A_n^{-1}| = O(1)$; $|B_n| = O(1)$; $\left| (B_n B_n')^{-1} \right| = O(1)$ and hence, $\hat{\theta}_n$ is $|D_n|^{1/2}$-consistent for $\theta_n^o$.

As remarked above, relative to the existing literature Theorem 4 allows for nonstationary processes and only assumes that $q_{i,n}$ and $\nabla_{\theta} q_{i,n}$ are NED on an $\alpha$-mixing input process, rather than postulating that $q_{i,n}$ and $\nabla_{\theta} q_{i,n}$ are $\alpha$-mixing. The latter assumption is restrictive in that the $\alpha$-mixing property is not
preserved under transformation. Since the NED property is preserved under a wide range of transformations it accommodates a much larger class of dependent spatial processes, including infinite moving average and spatial autoregressive models. As such, Theorem 4 should provide a basis for constructing confidence intervals and hypothesis testing in a wider range of spatial models.

Using Proposition 2, we now give some sufficient conditions for Assumption 7(c).

**Assumption 8** The process \( \{Z_{i,n}, i \in D_n \subset T_n, n \geq 1\} \) is uniformly \( L_2\)-NED on \( \{\varepsilon_{i,n}, i \in T_n, n \geq 1\} \) of size \(-2d(r-1)/(r-2)\) for some \( r > 2 \).

**Assumption 9** For every sequence \( \{\theta_n\} \) on \( \Theta \), the functions \( q_{i,n}(Z_{i,n}; \theta_n) \) and \( \nabla_{\theta} q_{i,n}(Z_{i,n}; \theta_n) \) satisfy Lipschitz condition (2) in \( z \), that is, for \( g_{i,n} = q_{i,n} \) or \( \nabla_{\theta} q_{i,n} \):

\[
|g_{i,n}(z; \theta_n) - g_{i,n}(z^*; \theta_n)| \leq B_{i,n}(z, z^*) |z - z^*|.
\]

Furthermore, for the \( r > 2 \) as specified in Assumption 8,

\[
\sup_{n,i \in D_n} \sup_s \|B_{i,n}^{(s)}\|_2 < \infty \quad \text{and} \quad \sup_{n,i \in D_n} \sup_s \left\| B_{i,n}^{(s)}(Z_{i,n} - \tilde{Z}_{i,n}^s) \right\|_r < \infty
\]

where \( B_{i,n}^{(s)} = B_{i,n}(Z_{i,n}, \tilde{Z}_{i,n}^s) \) with \( \tilde{Z}_{i,n}^s = E[Z_{i,n}|\mathcal{F}_i(s)] \).

6 Conclusion

The paper develops an asymptotic inference theory for a class of dependent nonstationary random fields that could be used in a wide range of econometric models with cross-sectional or spatial dependence, e.g., social interaction models. More specifically, the paper extends the notion of near-epoch dependent (NED) processes used in the time series literature to spatial processes. This allows to accommodate larger classes of dependent processes than mixing random fields. The class of NED random fields is “closed with respect to infinite transformations” and thus should be sufficiently broad for many applications of interest. In particular, it covers autoregressive and infinite moving average random fields as well as nonlinear spatial Bernoulli shifts. The NED property is also compatible with considerable heterogeneity and preserved under transformations under fairly mild conditions. Furthermore, a CLT and an LLN are derived for spatial processes that are NED on an \( \alpha \)-mixing process. Apart from covering a larger class of dependent processes, these limit theorems also allow for arrays of nonstationary random fields on unevenly spaced lattices. Building on these limit results, the paper develops an asymptotic theory of spatial GMM estimators, which provides a basis for inference in a broad range of models with cross-sectional or spatial dependence.
Much of the random fields literature assumes that the process resides on an equally spaced grid. In contrast, and as in Jenish and Prucha (2009), we allow for locations to be unequally spaced. The implicit assumption of fixed locations seems reasonable for a large class of applications, especially in the short run. Still, an important direction for future work would be to extend the asymptotic theory to spatial processes with endogenous locations, while maintaining a set of assumptions that are reasonably easy to interpret. One possible approach may be to augment the contributions of the present paper with theory from point processes.

A Appendix: Proofs for Section 3

Proof of Proposition 4: Let \( \mathfrak{S}_j(s) = \sigma(\varepsilon_j; j \in \mathbb{Z}^d; \rho(i,j) \leq s) \). By the Minkowski inequality for infinite sums, the conditional Jensen inequality, and (6), we have

\[
\|Y_i - E[Y_i|\mathfrak{S}_i(s)]\|^p_p = \left\| \sum_{j \in \mathbb{Z}^d, \rho(i,j) > s} g_{ij} \{\varepsilon_j - E[\varepsilon_j|\mathfrak{S}_i(s)]\} \right\|^p_p
\leq \sum_{j \in \mathbb{Z}^d, \rho(i,j) > s} |g_{ij}| \|\varepsilon_j - E[\varepsilon_j|\mathfrak{S}_i(s)]\|^p_p \leq K \psi(s).
\]

with \( K = 2 \sup_{j \in \mathbb{Z}^d} \|\varepsilon_j\|^p_p < \infty \) and \( \psi(s) = \sup_{l \in \mathbb{Z}^d} \sum_{j \in \mathbb{Z}^d, \rho(i,j) > s} |g_{ij}| \). Since \( \psi(s) \to 0 \) as \( s \to \infty \) by condition (5) we have \( \|Y_i - E[Y_i|\mathfrak{S}_i(s)]\|^p_p \to 0 \) as \( s \to \infty \), which proves the claim. Q.E.D.

Proof of Proposition 5: Let \( \mathfrak{S}_{i,n}(s) = \sigma(\varepsilon_{j,n}; 1 \leq j \leq n; \rho(i,j) \leq s) \). By the Minkowski inequality, the conditional Jensen inequality, and (12), we have

\[
\|Y_{i,n} - E[Y_{i,n}|\mathfrak{S}_{i,n}(s)]\|^p_p \leq \beta \sum_{1 \leq j \leq n, \rho(i,j) > s} |a_{ij,n}| \|X_{j,n} - E[X_{j,n}|\mathfrak{S}_{i,n}(s)]\|^p_p
+ \sum_{1 \leq j \leq n, \rho(i,j) > s} |b_{ij,n}| \|v_{j,n} - E[v_{j,n}|\mathfrak{S}_{i,n}(s)]\|^p_p \leq K \psi(s)
\]

with \( K = \max\{2\beta \|X_{j,n}\|^p_p, 2 \|v_{j,n}\|^p_p\} \times \infty \) and \( \psi(s) = \sup_{n,1 \leq i \leq n} \sum_{1 \leq j \leq n, \rho(i,j) > s} (|a_{ij,n}| + |b_{ij,n}|) \). Since \( \psi(s) \to 0 \) as \( s \to \infty \) by condition (11) we have \( \|Y_{i,n} - E[Y_{i,n}|\mathfrak{S}_{i,n}(s)]\|^p_p \to 0 \) as \( s \to \infty \), which proves the claim. Q.E.D.

Proof of Proposition 6: We first show that \( Y_i \) is well defined as the limit in \( L_2 \) as \( s \to \infty \) of the random variables

\[
Y_i^s = f_i(\varepsilon_{i+j}; j \in \mathbb{Z}^d),
\]

\footnote{Pinkse et al. (2007) made an interesting contribution in this direction. Their catalogue of assumptions is at the level of Bernstein blocks. Without further sufficient conditions, verification of those assumptions would typically be challenging in practical situations.}
where

\[ \varepsilon_{i+j}^{(s)} = \begin{cases} \varepsilon_{i+j} & \text{for } \rho(i, j) \leq s \\ 0 & \text{for } \rho(i, j) > s \end{cases} \]

In light of condition (15) we have for any \( s, m \in \mathbb{N} \):

\[
\|Y_i^{s+m} - Y_i^s\|_2 = \left\| f_i(\varepsilon_{i+j}^{(s+m)}) - f_i(\varepsilon_{i+j}^{(s)}) \right\|_2 \leq \sum_{j \in \mathbb{Z}^d: s < \rho(i, j) \leq s+m} w_{ij} \|\varepsilon_{i+j}\|_2
\]

with \( K = \sup_{i \in \mathbb{Z}^d} \|\varepsilon_i\|_2 \). In light of (16)-(17) the r.h.s. converges monotonically to zero as \( s \to \infty \). Thus clearly \( Y_i^s \) is a Cauchy series in \( L_2 \), and hence \( Y_i \) is well defined since the \( L_2 \) space is a Banach space and as such complete.

Now, let \( \tilde{G}_i(s) = \sigma(\varepsilon_{i+j}; j \in \mathbb{Z}^d: \rho(i, j) \leq s) \). By the minimum mean-squared error property of the conditional expectation, we have

\[
\|Y_i - E[Y_i|\tilde{G}_i(s)]\|_2 \leq \left\| f_i(\varepsilon_{i+j}, \mathbb{Z}^d) - f_i(\varepsilon_{i+j}, j \in \mathbb{Z}^d) \right\|_2 \leq K \sum_{j \in \mathbb{Z}^d: \rho(i, j) > s} w_{ij} \leq K \psi(s),
\]

with

\[ \psi(s) = \sup_{i \in \mathbb{Z}^d \rho(i, j) > s} \sum_{j \in \mathbb{Z}^d: \rho(i, j) > s} w_{ij} \to 0 \]

as \( s \to \infty \) by condition (16). This completes the proof of the lemma. \( \Box \)

**B Appendix: Proofs for Section 4**

Throughout this appendix let \( \tilde{G}_{i,n}(s) = \sigma(\varepsilon_{j,n}; j \in T_n : \rho(i, j) \leq s) \) be the \( \sigma \)-field generated by the random vectors \( \varepsilon_{j,n} \) located in the \( s \)-neighborhood of location \( i \). Furthermore, \( C \) denotes a generic constant that does not depend on \( n \) and may be different from line to line.

The proof of the CLT builds on Ibragimov and Linnik (1971), pp. 352-355, and makes use of the following lemmata:

**Lemma B.1** (Brockwell and Davis (1991), Proposition 6.3.9). Let \( Y_n, n = 1, 2, ... \) and \( V_{ns}, s = 1, 2, ...; n = 1, 2, ... \), be random vectors such that

(i) \( V_{ns} \Rightarrow V_s \) as \( n \to \infty \) for each \( s = 1, 2, ... \)

(ii) \( V_s \Rightarrow V \) as \( s \to \infty \), and

(iii) \( \lim_{s \to \infty} \lim \sup_{n \to \infty} P(|Y_n - V_{ns}| > \epsilon) = 0 \) for every \( \epsilon > 0 \).

Then \( Y_n \Rightarrow V \) as \( n \to \infty \).
Lemma B.2 (Ibragimov and Linnik (1971)) Let $L_p(\mathfrak{F}_1)$ and $L_p(\mathfrak{F}_2)$ denote, respectively, the class of $\mathfrak{F}_1$-measurable and $\mathfrak{F}_2$-measurable random variables $\xi$ satisfying $\|\xi\|_p < \infty$. Let $X \in L_p(\mathfrak{F}_1)$ and $Y \in L_q(\mathfrak{F}_2)$. Then, for any $1 \leq p, q, r < \infty$ such that $p^{-1} + q^{-1} + r^{-1} = 1$, 

$$|\text{Cov}(X, Y)| < 4\alpha^{1/r}(\mathfrak{F}_1, \mathfrak{F}_2) \|X\|_p \|Y\|_q$$

where $\alpha(\mathfrak{F}_1, \mathfrak{F}_2) = \sup_{A \in \mathfrak{F}_1, B \in \mathfrak{F}_2}(|P(AB) - P(A)P(B)|)$.

To prove the CLT for NED random fields, we first establish some moment inequalities and a slightly modified version of the CLT for mixing fields developed in Jenish and Prucha (2009). It is helpful to introduce the following notation. Let $X = \{X_{i,n}, i \in D_n, n \geq 1\}$ be a random field, then $\|X\|_q := \sup_{n,i \in D_n} \|X_{i,n}\|_q$ for $q \geq 1$.

Lemma B.3 Let $\{X_{i,n}\}$ be uniformly $L_2$-NED on a random field $\{\xi_{i,n}\}$ with $\alpha$-mixing coefficients $\mathfrak{m}(u, v, r) \leq (u + v)^{r}\hat{\alpha}(r)$, $\tau \geq 0$. Let $S_n = \sum_{i \in D_n} X_{i,n}$ and suppose that the NED coefficients of $\{X_{i,n}\}$ satisfy $\sum_{r=1}^\infty r^{d-1}\psi(r) < \infty$ and $\|X\|_{2+\delta} < \infty$ for some $\delta > 0$. Then,

(a)

$$|\text{Cov}(X_{i,n}, X_{j,n})| \leq \|X\|_{2+\delta} \left\{ C_1 \|X\|_{2+\delta} [h/3]^{d\tau^* \cdot \alpha^{\delta/(2+\delta)}(\lfloor h/3 \rfloor)} + C_2 \psi(\lfloor h/3 \rfloor) \right\}$$

where $h = \rho(i, j)$ and $\tau^* = \delta \tau/(2+\delta)$. If, in addition, $\sum_{r=1}^\infty r^{d-1}\psi(r) < \infty$, then for some constant $C < \infty$, not depending on $n$

$$\text{Var}(S_n) \leq C \|X\|_2 |D_n|.$$

(b)

$$|\text{Cov}(X_{i,n}, X_{j,n})| \leq \|X\|_2 \left\{ C_3 \|X\|_2 [h/3]^{d\tau^* \cdot \alpha^{\delta/(4+2\delta)}(\lfloor h/3 \rfloor)} + C_4 \psi(\lfloor h/3 \rfloor) \right\}$$

where $h = \rho(i, j)$ and $\tau^* = \delta \tau/(4+2\delta)$. If, in addition, $\sum_{r=1}^\infty r^{d-1}\psi(r) < \infty$ where $\tau^* = \delta \tau/(4+2\delta)$, then for some constant $C < \infty$, not depending on $n$

$$\text{Var}(S_n) \leq C \|X\|_2 |D_n|.$$

Proof of Lemma B.3: (a) For any $i \in D_n$ and $m > 0$, let $$\xi_{i,n}^m = E(X_{i,n}|\mathfrak{F}_{i,n}(m)), \quad \eta_{i,n}^m = X_{i,n} - \xi_{i,n}^m$$

By the Jensen and Lyapunov inequalities, we have for all $i \in D_n, n, m \in \mathbb{N}$ and any $1 \leq q \leq 2 + \delta$

$$E|\xi_{i,n}^m|^q = E\{|E(X_{i,n}|\mathfrak{F}_{i,n}(m))|^q\} \leq E\{|E(X_{i,n}|\mathfrak{F}_{i,n}(m))|^q\} = E|\xi_{i,n}^m|^q$$
and thus
\[ \|\xi_{i,n}^m\|_q \leq \|X_{i,n}\|_q \leq \|X\|_{2+\delta}, \quad \|\eta_{i,n}^m\|_q \leq 2 \|X_{i,n}\|_q \leq 2 \|X\|_{2+\delta}. \]

Thus, both \(\xi_{i,n}^m\) and \(\eta_{i,n}^m\) are uniformly \(L_{2+\delta}\) bounded. Also, note that
\[
\sup_{n,i \in D_n} \|\eta_{i,n}^m\|_2 \leq \psi(m),
\]
given that the \(\{X_{i,n}\}\) is uniformly \(L_2\)-NED on \(\{\varepsilon_{i,n}\}\) and thus the NED-scaling factors can be chosen w.l.g. to be one. Furthermore, let \(\sigma(\xi_{i,n}^m)\) denote the \(\sigma\)-field generated by \(\xi_{i,n}^m\). Since \(\sigma(\xi_{i,n}^m) \subseteq 3_{i,n}(m)\), the mixing coefficients of \(\xi_{i,n}^m\) satisfy
\[
\pi_1(1, 1, r) \leq \left\{ \begin{array}{ll}
1, & r \leq 2m \\
\pi(M^{d}, M^{d}, r - 2m), & r > 2m
\end{array} \right.
\]
where \(\pi(u, v, r)\) are the mixing coefficients of the input process \(\varepsilon\), since the \(m\)-neighborhood of any point on \(D\) contains at most \(M^{d}\) points of \(D\) for some \(M\) that does not depend on \(m\), see the proof of Lemma A.1 of Jenish and Prucha (2009).

Now, decompose \(X_{i,n}\) and \(X_{j,n}\) as
\[
X_{i,n} = \xi_{i,n}^{[h/3]} + \eta_{i,n}^{[h/3]}, \quad \text{and} \quad X_{j,n} = \xi_{j,n}^{[h/3]} + \eta_{j,n}^{[h/3]},
\]
where \(h = \rho(i, j)\). Then,
\[
|Cov(X_{i,n}X_{j,n})| = \left|Cov\left(\xi_{i,n}^{[h/3]} + \eta_{i,n}^{[h/3]} ; \xi_{j,n}^{[h/3]} + \eta_{j,n}^{[h/3]}\right)\right| \leq \left|Cov\left(\xi_{i,n}^{[h/3]} ; \xi_{j,n}^{[h/3]}\right)\right| + \left|Cov\left(\xi_{i,n}^{[h/3]} ; \eta_{j,n}^{[h/3]}\right)\right| + \left|Cov\left(\eta_{i,n}^{[h/3]} ; \xi_{j,n}^{[h/3]}\right)\right| + \left|Cov\left(\eta_{i,n}^{[h/3]} ; \eta_{j,n}^{[h/3]}\right)\right|. \tag{B.1}
\]

We will now bound separately each term on the r.h.s. of the last inequality.

First, using Lemma B.2 with \(p = q = 2 + \delta\), and \(r = (2 + \delta)/\delta\) yields the following bound on the first term:
\[
\left|Cov\left(\xi_{i,n}^{[h/3]} ; \xi_{j,n}^{[h/3]}\right)\right| \leq 4 \left\|\xi_{i,n}^{[h/3]}\right\|_{2+\delta} \left\|\xi_{j,n}^{[h/3]}\right\|_{2+\delta} \pi^{\delta/(2+\delta)}(1, 1, [h/3]) \tag{B.2}
\]
\[
\leq 4 \left\|X\right\|_{2+\delta} \pi^{\delta/(2+\delta)} \left(M [h/3]^d, M [h/3]^d, h - 2 [h/3]\right) \leq C_1 \left\|X\right\|_{2+\delta} [h/3]^d \pi^{\delta/(2+\delta)} ([h/3])
\]
where \(\tau_* = \delta \tau/(2 + \delta)\).

Second, the Cauchy-Schwartz inequality gives the following bound on the second and third terms:
\[
\left|Cov\left(\xi_{i,n}^{[h/3]} ; \eta_{j,n}^{[h/3]}\right)\right| \leq 4 \left\|\xi_{i,n}^{[h/3]}\right\|_2 \left\|\eta_{j,n}^{[h/3]}\right\|_2 \leq 4 \left\|X\right\|_2 \psi([h/3]) \tag{B.3}
\]
Similarly, the fourth term can be bounded as:

\[ \left| \text{Cov} \left( \eta_i^{[h/3]}, \eta_j^{[h/3]} \right) \right| \leq 8 \|X\|_2 \psi \left( [h/3] \right) \]  \hspace{1cm} (B.4)

Collecting (B.2)-(B.4), we have

\[ |\text{Cov} (X_{i,n}, X_{j,n})| \leq \|X\|_{2+\delta} \left\{ C_1 \|X\|_{2+\delta} [h/3]^{d\tau r} \alpha^{\delta/(2+\delta)} ([h/3]) + C_2 \psi ([h/3]) \right\} \]

which proves the first inequality.

Using this inequality as well as the bounds on the sizes of the sets given in Lemma A.1 of Jenish and Prucha (2009), we have

\[ \text{Var} (S_n) \leq \sum_{i \in D_n} \text{Var} (X_{i,n}) + \sum_{i,j \in D_n, i \neq j} \left| \text{Cov} (X_{i,n}, X_{j,n}) \right| \]

\[ \leq 2 |D_n| \|X\|_{2+\delta}^2 + C |D_n| \|X\|_{2+\delta} \left[ \sum_{r=1}^\infty \|X\|_{2+\delta} \psi ([\rho(i,j)/3]) \right] \]

\[ \leq 2 |D_n| \|X\|_{2+\delta} + C |D_n| \left\{ \sum_{r=1}^\infty r^{d(r+1)-1} \alpha^{\delta/(2+\delta)} (r) + \sum_{r=1}^\infty r^{d-1} \psi (r) \right\} \leq C |D_n| \]

for some constant \( C < \infty \), not depending on \( n \).

(b) To prove the second part of the lemma, apply Lemma B.2 with \( p = 2+\delta \), \( q = 2 \), and \( r = 2(2+\delta)/\delta \) to obtain the following bound on the first term:

\[ |\text{Cov} (\xi_i^{[h/3]}, \xi_j^{[h/3]})| \leq C_5 \|X\|_{2+\delta} \|X\|_{2} [h/3]^{d\tau r^*} \alpha^{\delta/(4+2\delta)} ([h/3]) , \]  \hspace{1cm} (B.5)

where \( \tau^* = \delta \tau/(4+2\delta) \). The other terms on the r.h.s. of (B.1) are bounded as in part (a). Collecting (B.3)-(B.5) gives

\[ |\text{Cov} (X_{i,n}X_{j,n})| \leq \|X\|_2 \left\{ C_5 \|X\|_{2+\delta} [h/3]^{d\tau r^*} \alpha^{\delta/(4+2\delta)} ([h/3]) + C_4 \psi ([h/3]) \right\} \]

as required. Finally, using similar arguments as in the proof of part (a), we can bound \( \text{Var} (S_n) \) as

\[ \text{Var} (S_n) \leq C \|X\|_2 |D_n| \]

for some constant \( C < \infty \), not depending on \( n \). \hspace{2cm} Q.E.D.

**Theorem B.1** Suppose \( \{D_n\} \) is a sequence of finite subsets of \( D \), satisfying Assumption 1, with \( |D_n| \to \infty \) as \( n \to \infty \). Suppose further that \( \{\xi_{i,n}; i \in D_n, n \in \mathbb{N}\} \) is an array of zero-mean random variables with \( \alpha \)-coefficients \( \overline{\alpha}(u,v,r) \leq \)
\(C(u + v)^\gamma \hat{\alpha}(r)\) for some constants \(C < \infty\) and \(\tau \geq 0\). Suppose for some \(\delta > 0\) and \(\gamma > 0\)

\[
\lim_{k \to \infty} \sup_{n, i \in D_n} E[|\varepsilon_{i,n}|^{2+\delta} 1(|\varepsilon_{i,n}| > k)] = 0
\]

and

\[
\hat{\alpha}(r) = O(r^{-d(2\mu+1)-\gamma})
\]

with \(\mu = \max\{\tau, 1/\delta\}\), and suppose \(\liminf_{n \to \infty} |D_n|^{-1} \sigma_n^2 > 0\), then

\[
\sigma_n^{-1} \sum_{i \in D_n} \varepsilon_{i,n} \rightarrow N(0,1).
\]

where \(\sigma_n^2 = \text{Var} \left( \sum_{i \in D_n} \varepsilon_{i,n} \right) \).

The above CLT is in essence a variant of CLT for \(\alpha\)-mixing random fields given as Corollary 1 of Theorem 1 in Jenish and Prucha (2009), applied to mixing coefficients of the type \(\pi(u, v, r) \leq C(u + v)^\gamma \hat{\alpha}(r), \tau \geq 0\).

**Proof of Theorem B.1:** The proof of the theorem is largely the same as the proof of Theorem 1 in Jenish and Prucha (2009). We will first show that all assumptions of that theorem except Assumption 3(c) are satisfied. We will then show that the entire proof goes through if Assumption 3(c) is replaced by the condition \(\hat{\alpha}(r) = O(r^{-d(2\mu+1)-\gamma})\) with \(\mu = \max\{\tau, 1/\delta\}\).

Clearly, Assumptions 1, 2 and 5 of that theorem with \(c_{i,n} = 1\) are satisfied. To verify Assumptions 3(a) note that

\[
\sum_{r=1}^{\infty} \pi(1, 1, r) r^{d(2+\delta)/\delta - 1} \leq C \sum_{r=1}^{\infty} r^{-2d(\mu-1/\delta)-1-\gamma} \leq C \sum_{r=1}^{\infty} r^{-1-\gamma} < \infty,
\]

since \(\mu = \max\{\tau, 1/\delta\}\). As shown in the proof of Corollary 1 in Jenish and Prucha (2009), the latter condition implies Assumption 3(a) of Theorem 1 in Jenish and Prucha (2009). Furthermore, it is easy to see that Assumption 3(b) is also satisfied. Indeed, for any \(u + v \leq 4\), we have

\[
\sum_{r=1}^{\infty} \pi(u, v, r) r^{d-1} \leq 4^d C \sum_{r=1}^{\infty} r^{d-1-\hat{\alpha}(r)} \leq C \sum_{r=1}^{\infty} r^{d-1} r^{-d(2\mu+1)-\gamma} < \infty.
\]

Thus, all assumptions, except Assumption 3(c), of Theorem 1 in Jenish and Prucha (2009) hold, and hence, all steps of its proof which do not rely on that assumptions remain valid in our case. Assumption 3(c) is only used in step 5 of that proof. Specifically, all arguments in that step continue to hold given we show that there exists sequence \(m_n\) such that

\[
m_n^d |D_n|^{-1/2} \rightarrow 0 \text{ as } n \rightarrow \infty \quad (B.6)
\]

and

\[
\pi(1, |D_n|, m_n)|D_n|^{1/2} \rightarrow 0 \text{ as } n \rightarrow \infty \quad (B.7)
\]
Though $\pi_1$ is used instead of $\pi_1[D_n]$ in the proof of Theorem 1 of Jenish and Prucha (2009), in fact the proof only relies on the coefficient $\pi_1[D_n]$; see Step 9 ($|E_{3,n} \rightarrow 0$) of the proof in the working version of the paper. The desired sequence $m_n$ can be chosen as

$$m_n = \left(|D_n|^{1/2} / \log |D_n|\right)^{1/d}.$$ 

It is immediate that (B.6) holds,

$$m_n|D_n|^{-1/2} = (\log |D_n|)^{-1} \rightarrow 0.$$

To verify (B.7), observe

$$C[1, |D_n|, m_n]|D_n|^{1/2} \leq C[D_n]\left(\gamma + 1/2\right)m_n^{-d(\gamma + 1)-\gamma}$$

$$\leq C[D_n]\left(\gamma + 1/2\right)|D_n|^{-\gamma - 1/2} |D_n|^{-\gamma/(2d)} |\log |D_n||^{(2\mu + 1)+\gamma/d}$$

$$\leq C[D_n]^{-\gamma/(2d)} |\log |D_n||^{(2\mu + 1)+\gamma/d} \rightarrow 0.$$

The rest of the proof is the same, word-by-word, as the proof of Theorem 1 of Jenish and Prucha (2009).

Proof of Theorem 1: Since the proof is lengthy it is broken into steps.

1. Transition from $Z_{i,n}$ to $Y_{i,n} = Z_{i,n}/M_n$

   Let $M_n = \max_{i \in D_n} c_i$ and $Y_{i,n} = Z_{i,n}/M_n$. Also, let $\sigma^2_{Z,n} = Var[\sum Y_{i,n}]$ and $\sigma^2_{Z,n} = Var[\sum Z_{i,n}] = M_n^{-2}\sigma^2_{Z,n}$. Since

   $$\sigma^{-1}_{Y,n} \sum_{i \in D_n} Y_{i,n} = \sigma^{-1}_{Z,n} \sum_{i \in D_n} Z_{i,n},$$

   to prove the theorem, it suffices to show that $\sigma^{-1}_{Y,n} \sum_{i \in D_n} Y_{i,n} \Rightarrow N(0,1)$. Therefore, it proves convenient to switch notation from the text and to define

   $$S_n = \sum_{i \in D_n} Y_{i,n}, \quad \sigma^2_n = Var(S_n).$$

That is, in the following, $S_n$ denotes $\sum_{i \in D_n} Y_{i,n}$ rather than $\sum_{i \in D_n} Z_{i,n}$, and $\sigma^2_n$ denotes the variance of $\sum_{i \in D_n} Y_{i,n}$ rather than of $\sum_{i \in D_n} Z_{i,n}$. We now establish moment and mixing conditions for $Y_{i,n}$ from the assumptions of the theorem. Observe that by definition of $M_n$

$$1(|Y_{i,n}| > k) = 1(|Z_{i,n}/M_n| > k) \leq 1(|Z_{i,n}/c_i| > k),$$

and hence

$$E(|Y_{i,n}|^{2+\delta} 1(|Y_{i,n}| > k)) \leq E(|Z_{i,n}/c_i|^{2+\delta} 1(|Z_{i,n}/c_i| > k))$$

so that Assumption 3(a) implies that

$$\lim_{k \rightarrow \infty} \sup_{n, i \in D_n} E(|Y_{i,n}|^{2+\delta} 1(|Y_{i,n}| > k)) = 0. \quad (B.8)$$
Hence, $Y_{i,n}$ is also uniformly $L_{2+\delta}$ bounded. Let $\|Y\|_{2+\delta} = \sup_{n,i \in D_n} \|Y_{i,n}\|_{2+\delta}$.

Further, note that

$$\|Y_{i,n} - E(Y_{i,n}|\mathfrak{F}_{i,n}(s))\|_2 = M_n^{-1} \|Z_{i,n} - E(Z_{i,n}|\mathfrak{F}_{i,n}(s))\|_2$$ (B.9)

since $\sup_{n,i \in D_n} c_{i,n}^{-1}d_{i,n} \leq C < \infty$, by assumption. Thus, $Y_{i,n}$ is uniformly $L_2$-NED on $\varepsilon$ with the NED coefficients $\psi(m)$. Finally, observe that by Assumption 3(b):

$$\inf_n |D_n|^{-1}\sigma_n^2 > 0.$$ (B.10)

Hence, there exists $0 < B < \infty$ such that for all $n$

$$B|D_n| \leq \sigma_n^2.$$ (B.11)

2. Decomposition of $Y_{i,n}$

For any fixed $s > 0$, decompose $X_{i,n}$ as

$$Y_{i,n} = \xi^s_{i,n} + \eta^s_{i,n}$$

where

$$\xi^s_{i,n} = E(Y_{i,n}|\mathfrak{F}_{i,n}(s)), \quad \eta^s_{i,n} = Y_{i,n} - \xi^s_{i,n}$$

Let

$$S_{n,s} = \sum_{i \in D_n} \xi^s_{i,n}; \quad \bar{S}_{n,s} = \sum_{i \in D_n} \eta^s_{i,n}$$

$$\sigma^2_{n,s} = \text{Var}[S_{n,s}], \quad \bar{\sigma}^2_{n,s} = \text{Var}[\bar{S}_{n,s}]$$

Repeated use of the Minkowski inequality yields:

$$|\sigma_n - \sigma_{n,s}| \leq \bar{\sigma}_{n,s}, \quad |\sigma_n - \bar{\sigma}_{n,s}| \leq \sigma_{n,s}.$$ (B.12)

Observe that

$$E[E(Y_{i,n}|\mathfrak{F}_{i,n}(s))|\mathfrak{F}_{i,n}(m))] = \begin{cases} E(Y_{i,n}|\mathfrak{F}_{i,n}(s)), & m \geq s, \\ E(Y_{i,n}|\mathfrak{F}_{i,n}(m)), & m < s. \end{cases}$$

and hence

$$\|\eta^s_{i,n} - E(\eta^s_{i,n}|\mathfrak{F}_{i,n}(m))\|_2$$

$$= \|Y_{i,n} - E[Y_{i,n}|\mathfrak{F}_{i,n}(s)] - E[Y_{i,n}|\mathfrak{F}_{i,n}(m)] + E[(Y_{i,n}|\mathfrak{F}_{i,n}(s))|\mathfrak{F}_{i,n}(m)]\|_2$$

$$= \begin{cases} \|Y_{i,n} - E[Y_{i,n}|\mathfrak{F}_{i,n}(s)]\|_2 \leq C\psi(m), & m \geq s, \\ \|Y_{i,n} - E[Y_{i,n}|\mathfrak{F}_{i,n}(s)]\|_2 \leq C\psi(s) \leq C\psi(m), & m < s. \end{cases}$$

since by definition the sequence $\psi(m)$ is non-increasing. Thus, for any fixed $s > 0$, $\{\eta^s_{i,n}\}$ is uniformly $L_2$-NED on $\varepsilon$ with the same NED coefficients $\psi(m)$.
as the random field \( \{Y_{i,n}\} \). Furthermore, as shown in the proof of Lemma B.3, \( \{\eta_{i,n}^s\} \) is also uniformly \( L_{2+\delta} \) bounded.

3. **Bounds for the Variances of \( \sum Y_{i,n} \) and \( \sum \eta_{i,n}^s \)**

First note that in light of Assumption 2, and observing that \( \tau^* = \delta \tau/(4 + 2\delta) \leq \tau_* = \delta \tau/(2 + \delta) \) and \( \hat{\alpha}^{\delta/(2+\delta)}(r) \leq \hat{\alpha}^{\delta/(2+\delta)}(r) \) we have

\[
\sum_{r=1}^{\infty} r^d(\tau^*+1)\hat{\alpha}^{\delta/(2+\delta)}(r) \leq \sum_{r=1}^{\infty} r^d(\tau_*+1)\hat{\alpha}^{\delta/(2+\delta)}(r) < \infty,
\]

\[
\sum_{r=1}^{\infty} r^d(\tau^*+1)\hat{\alpha}^{\delta/(4+2\delta)}(r) \leq \sum_{r=1}^{\infty} r^d(\tau_*+1)\hat{\alpha}^{\delta/(4+2\delta)}(r) < \infty.
\]

Using part (a) of Lemma B.3 with \( X_{i,n} = Y_{i,n} \) and recalling (B.11), we have

\[
B|D_n| \leq \sigma_{n}^2 = Var(S_n) \leq C|D_n|.
\]

for some \( B > 0 \). Using part (b) of Lemma B.3 with \( X_{i,n} = \eta_{i,n}^s \) we have

\[
\sigma_{n,s}^2 = Var(\tilde{S}_{n,s}) \leq C|D_n| \|\eta_{i,n}^s\|_2 = C|D_n|\psi(s) \quad (B.13)
\]

in light of (B.9). Hence,

\[
\lim_{s \to \infty} \lim_{n \to \infty} \sup_{n;i} \frac{\sigma_{n,s}^2}{\sigma_{n}^2} \leq C \lim_{s \to \infty} \psi(s) = 0. \quad (B.14)
\]

Furthermore, by (B.12) we have

\[
\lim_{s \to \infty} \lim_{n \to \infty} \sup_{n;i} \left| 1 - \frac{\sigma_{n,s}}{\sigma_{n}} \right| \leq \lim_{s \to \infty} \lim_{n \to \infty} \sup_{n;i} \frac{\sigma_{n,s}}{\sigma_{n}} = 0 \quad (B.15)
\]

and hence for all \( s \geq 1 \) and \( n \geq 1 \)

\[
\frac{\sigma_{n,s}}{\sigma_{n}} \leq C < \infty. \quad (B.16)
\]

4. **CLT for \( \sum_{i \in D_n} \xi_{i,n}^s \)**

We now show that for any fixed \( s > 0 \), \( \xi_{i,n}^s \) satisfies the CLT given in Theorem B.1.

First, since \( \sup_{n,i \in D_n} E[|\xi_{i,n}^s|^{2+\delta}] < \infty \), the process \( \{\xi_{i,n}^s\} \) is uniformly \( L_{2+\delta'} \)-integrable for \( \delta' = \delta/2 \), i.e.,

\[
\lim_{k \to \infty} \sup_{n,i \in D_n} E[|\xi_{i,n}^s|^{2+\delta/2} 1(|\xi_{i,n}^s| > k)] = 0.
\]

Second, note that since \( \xi_{i,n}^s \) is a measurable function of \( \varepsilon_{i,n} \) for any \( u, v \in \mathbb{N} \) and \( r > 2s \)

\[
\overline{\sigma}_\xi(u, v, r) \leq \overline{\sigma}(u Ms^d, v Ms^d, r - 2s) \leq C (u + v)^{\gamma} \hat{\alpha}(r - 2s)
\]

28
We next to show that \( \tilde{\alpha}(r) = O(r^{-d(2\mu+1)\gamma}) \) for \( \mu = \max\{\tau, 2/\delta\} \) and some \( \gamma > 0 \). By assumption,

\[
\sum_{r=1}^{\infty} r^{d(\tau_r+1)^{-1} - \frac{d}{2(\tau_r+1)}} (r) < \infty,
\]

where \( \tau_r = \delta \tau_r/(2 + \delta) \), which implies

\[
\tilde{\alpha}(r) = o(r^{-d(2(\tau_r+1)/\delta)}) = o(r^{-d(2(\tau_r+1)/\delta) + 1/d}) = o(r^{-d(2\mu+1) - d})
\]

since \( \mu \leq \tau + 2/\delta \) for \( \mu = \max\{\tau, 2/\delta\} \). Thus, \( \tilde{\alpha}(r) = O(r^{-d(2\mu+1)\gamma}) \) for \( \gamma = d \).

We next show that for sufficiently large \( s \),

\[
0 < \lim \inf_{n \to \infty} |D_n|^{-1} \sigma_{n,s}^2.
\]

By (B.11),

\[
B^{1/2} \leq \inf |D_n|^{-1/2} \sigma_n
\]

Since \( \lim_{n \to \infty} \psi(s) = 0 \), there exists \( s_* \) such that in light of (B.13) for all \( s \geq s_* \),

\[
|D_n|^{-1/2} \sigma_{n,s} \leq C \psi^{1/2}(s) \leq B^{1/2}/2.
\]

Hence by (B.12) for all \( s \geq s_* \),

\[
|D_n|^{-1/2} (\sigma_n - \tilde{\sigma}_{n,s}) \leq |D_n|^{-1/2} \sigma_{n,s}
\]

and thus \( \inf_{n} |D_n|^{-1/2} \sigma_{n,s} \geq \inf_{n} |D_n|^{-1/2} \sigma_n - \sup_{n} |D_n|^{-1/2} \tilde{\sigma}_{n,s} \). Using (B.10) and (B.17), we have

\[
\lim \inf_{n \to \infty} |D_n|^{-1/2} \sigma_{n,s} \geq B^{1/2} - \frac{B^{1/2}}{2} = \frac{B^{1/2}}{2} > 0
\]

Thus, for all \( s \geq s_* \),

\[
\sigma_{n,s}^{-1} \sum_{i \in D_n} \xi_{i,n}^{s} \Longrightarrow N(0, 1) \text{ as } n \to \infty.
\]

Since the first \( s_* \) terms do not affect the analysis below we take in the following \( s_* = 1 \).

5. CLT for \( \sum_{i \in D_n} Y_{i,n} \)

Finally, using Lemma B.1 we now show that, given the maintained NED assumption, the just established CLT in (B.18) for the approximators \( \xi_{i,n}^{s} \) can be carried over to the \( \xi_{i,n}^{s} \) for the approximators \( \xi_{i,n}^{s} \) can be carried over to the \( Y_{i,n} \). Define

\[
W_n = \sigma_n^{-1} \sum_{D_n} Y_{i,n}, \quad V_{ns} = \sigma_n^{-1} \sum_{i \in D_n} \xi_{i,n}^{s}, \quad W_n - V_{ns} = \sigma_n^{-1} \sum_{i \in D_n} \eta_{i,n}^{s}
\]

so that we can exploit Lemma B.1 to prove that

\[
W_n = \sigma_n^{-1} \sum_{i \in D_n} Y_{i,n} \Longrightarrow V \sim N(0, 1).
\]
We first verify condition (iii) of Lemma B.1. By Markov’s inequality and (B.14), for every \( \epsilon > 0 \) we have

\[
\lim_{s \to \infty} \lim_{n \to \infty} \sup_{n} P(|W_{n} - V_{n,s}| > \epsilon) = \lim_{s \to \infty} \lim_{n \to \infty} P\left(\frac{\sigma_{n}^{-1} \sum_{i \in D_{n}} n_{i,n}^{s}}{\| \sigma_{n,s} \|} > \epsilon\right) \\
\leq \lim_{s \to \infty} \lim_{n \to \infty} \frac{\sigma_{n,s}^{2}}{\epsilon^{2} \sigma_{n}^{2}} = 0.
\]

Next observe that

\[
V_{n,s} = \frac{\sigma_{n,s}}{\sigma_{n}} \left[ \sigma_{n,s}^{-1} \sum_{i \in D_{n}} \xi_{i,n}^{s} \right].
\]

We proceed to show \( W_{n} \Rightarrow V \) by contradiction. For that purpose let \( M \) be the set of all probability measures on \((\mathbb{R}, \mathcal{B})\), and observe that we can metrize \( M \) by, e.g., the Prokhorov distance \( d(\cdot, \cdot) \). Let \( \mu_{n} \) and \( \mu \) be the probability measures corresponding to \( W_{n} \) and \( V \), respectively, then \( W_{n} \Rightarrow V \), or \( \mu_{n} \Rightarrow \mu \), if \( d(\mu_{n}, \mu) \to 0 \) as \( n \to \infty \). Now suppose \( \mu_{n} \) does not converge to \( \mu \). Then for some \( \epsilon > 0 \) there exists a subsequence \( \{n(m)\} \) such that \( d(\mu_{n(m)}, \mu) > \epsilon \) for all \( n(m) \). By (B.16), we have \( 0 \leq \sigma_{n,m,s}/\sigma_{n,m} \leq C < \infty \) for all \( s \geq 1 \). Hence, \( 0 \leq \sigma_{n,m,s}/\sigma_{n,m} \leq C < \infty \) for all \( n(m) \). Consequently, for \( s = 1 \) there exists a subsequence \( \{n(m(l_{1}))\} \) such that \( \sigma_{n(m(l_{1})),1}/\sigma_{n(m(l_{1}))} \to p(1) \) as \( l_{1} \to \infty \). For \( s = 2 \), there exists a subsubsequence \( \{n(m(l_{1}(l_{2})))\} \) such that \( \sigma_{n(m(l_{1}(l_{2}))),2}/\sigma_{n(m(l_{1}(l_{2})))} \to p(2) \) as \( l_{2} \to \infty \). The argument can be repeated for \( s = 3, 4, \ldots \). Now construct a subsequence \( \{n_{i}\} \) such that \( n_{1} \) corresponds to the first element of \( \{n(m(l_{1}))\} \), \( n_{2} \) corresponds to the second element of \( \{n(m(l_{1}(l_{2})))\} \), and so on, then

\[
\lim_{l \to \infty} \frac{\sigma_{n_{i},s}}{\sigma_{n_{i}}} = \rho(s) \quad (B.19)
\]

for \( s = 1, 2, \ldots \). Given (B.18), it follows that as \( l \to \infty \)

\[
V_{n,l} \Rightarrow V_{l} \sim N(0, \rho^{2}(s)).
\]

Then, it follows from (B.15) that

\[
\lim_{s \to \infty} |p(s) - 1| \leq \lim_{s \to \infty} \lim_{l \to \infty} \left| p(s) - \frac{\sigma_{n_{i},s}}{\sigma_{n_{i}}} \right| + \lim_{s \to \infty} \lim_{n \to \infty} \sup_{n \geq 1} \left| \frac{\sigma_{n_{i},s}}{\sigma_{n_{i}}} - 1 \right| = 0.
\]

Thus \( V_{l} \Rightarrow V \) and thus by Lemma B.1 \( W_{n,l} \Rightarrow V \sim N(0,1) \) as \( l \to \infty \). Since \( \{n_{i}\} \subset \{n(m)\} \) this contradicts the assumption that \( d(\mu_{n(m)}, \mu) > \epsilon \) for all \( n(m) \). This completes the proof of the CLT.

**Proof of Corollary 1**: To prove the theorem, we apply the Cramer-Wold device, and verify that for every \( \lambda \in \mathbb{R}^{k} \) with \( |\lambda| = 1 \), \( \sigma_{n}^{-1} \sum V_{i,n} \Rightarrow N(0,1) \), where \( V_{i,n} = \lambda^{T} Z_{i,n} \). Observe that using the properties of norms, we have

\[
|V_{i,n}|/c_{i,n} = |\lambda^{T} Z_{i,n}|/c_{i,n} \leq |\lambda| |Z_{i,n}|/c_{i,n} = |Z_{i,n}|/c_{i,n}
\]

30
and

\[ \mathbf{1}(|V_i,n|/c_i,n > k) = \mathbf{1}(|X_i,n|/c_i,n > k) \leq \mathbf{1}(|Z_i,n|/c_i,n > k), \]

and thus \( \lim_{k \to \infty} \sup_{n,j \in D_n} E[|V_i,n/c_i,n|^{2+\delta} \mathbf{1}(|V_i,n/c_i,n| > k)] = 0. \) Further-
more, observe that

\[ \|V_{i,n} - E(V_{i,n} | \mathcal{F}_{i,n}(s))\|_2 \leq |\lambda| \|Z_{i,n} - E(Z_{i,n} | \mathcal{F}_{i,n}(s))\|_2 \leq d_{i,n} \psi(s) \]

and that for \( \sigma_n^2 = \text{Var}(\sum_{t \in D_n} V_{i,n}) = \lambda' \Sigma_n \lambda \) we have

\[ \inf_n |D_n|^{-1} M_n^{-2} \sigma_n^2 = \inf_n |D_n|^{-1} M_n^{-2} \lambda' \Sigma_n \lambda \geq \inf_n |D_n|^{-1} M_n^{-2} \lambda \min(\Sigma_n) > 0. \]

From this we see that under the maintained assumptions \( V_{i,n} \) satisfies all as-
sumptions of the CLT for scalar-valued random \( \mathbf{f} \)elds (Theorem 1) and, there-
fore, \( \sigma_n^{-1} \sum V_{i,n} \Rightarrow N(0,1) \) as claimed.

Next define \( X_{i,n} = M_n^{-1} V_{i,n} \), then by analogous arguments as above

\[ |X_{i,n}| \leq |\lambda| |Z_{i,n}|/c_i,n = |Z_{i,n}|/c_i,n. \]

From the maintained uniform \( L_{2+\delta} \) integrability of \( |Z_{i,n}|/c_i,n \) it then follows
that \( \|X\|_{2+\delta} \leq \|Z\|_{2+\delta} \), which shows that the \( 2+\delta \) moments of \( X_{i,n} \) can be bounded by a constant that does not depend on \( \lambda \). Consequently if follows from
the last inequality in the proof of part (a) of Lemma B.3 that

\[ \text{Var}(\sum_{i \in D_n} X_{i,n}) = M_n^{-2} \lambda' \Sigma_n \lambda \leq C|D_n| \]

where \( C < \infty \) does not depend on \( n \) and \( \lambda \). Hence

\[ \sup_n |D_n|^{-1} M_n^{-2} \lambda \max(\Sigma_n) = \sup_n |D_n|^{-1} M_n^{-2} \sup_{|\lambda|=1} \lambda' \Sigma_n \lambda \leq C < \infty. \]

This proves the second claim of the lemma. \( \quad Q.E.D. \)

**Proof of Theorem 2:** Define \( Y_{i,n} = Z_{i,n}/M_n \), then to prove the theorem,
it suffices to show that \( |D_n|^{-1} \sum_{i \in D_n} (Y_{i,n} - EY_{i,n}) \xrightarrow{L_2} 0. \) We first establish moment and mixing conditions for \( Y_{i,n} \) from those for \( Z_{i,n} \). Observe that in
light of the definition of \( M_n \) and Assumption 4(a)

\[ \sup_{n,i \in D_n} E|Y_{i,n}|^p \leq \sup_{n,i \in D_n} E|Z_{i,n}/c_{i,n}|^p < \infty. \quad (B.20) \]

Thus, \( Y_{i,n} \) is uniformly \( L_p \)-bounded for \( p > 1 \). Let \( \mathcal{F}_{i,n}(s) = \sigma(\varepsilon_{j,n}; j \in T_n; \rho(i,j) \leq s) \). Since \( Z_{i,n} \) is \( L_1 \)-NED on \( \varepsilon = \{\varepsilon_{i,n}; i \in T_n, n \geq 1\} \):

\[ \sup_{n,i \in D_n} \|Y_{i,n} - E(Y_{i,n}|\mathcal{F}_{i,n}(s))\|_1 \leq \sup_{n,i \in D_n} M_n^{-1} d_{i,n} \psi(s) \leq \psi(s), \quad (B.21) \]

observing that \( M_n = \max_{i \in D_n} \max(c_{i,n}, d_{i,n}) \). Thus \( Y_{i,n} \) is also \( L_1 \)-NED on \( \varepsilon \).
Next we show that for each given $s > 0$, the conditional mean $V_{i,n}^s = E(Y_{i,n} | \mathcal{F}_{i,n}(s))$ satisfies the assumptions of the $L_1$-norm LLN of Jenish and Prucha (2009, Theorem 3). Using the Jensen and Lyapunov inequalities gives for all $s > 0$, $i \in D_n$:

$$E|V_{i,n}^s|^p \leq E\{E(|Y_{i,n}|^p | \mathcal{F}_{i,n}(s))\} \leq \sup_{n,i \in D_n} E|Y_{i,n}|^p < \infty.$$  

So, $V_{i,n}^s$ is uniformly $L_p$-bounded for $p > 1$ and hence uniformly integrable. For each fixed $s$, $V_{i,n}^s$ is a measurable function of $\{\varepsilon_{j,n}; j \in T_n : \rho(i,j) \leq s\}$.

Observe that under Assumption 1 there exists a finite constant $C$ such that the cardinality of the set $\{j \in T_n : \rho(i,j) \leq s\}$ is bounded by $Cs^d$; compare Lemma A.1 in Jenish and Prucha (2009). Hence,

$$\overline{\sigma}_{V^s,1,1,r} \leq \left\{ \begin{array}{ll} 1, & r \leq 2s \\overline{\sigma}(Cs^d, Cs^d, r - 2s), & r > 2s \end{array} \right.$$  

and thus in light of Assumption 6(b)

$$\sum_{r=1}^{\infty} r^{d-1}\overline{\sigma}_{V^s,1,1,r} \leq \sum_{r=1}^{2s} r^{d-1} + \varphi(Cs^d, Cs^d) \sum_{r=1}^{\infty} (r + 2s)^{d-1}\overline{\sigma}(r) < \infty.$$  

The above shows that indeed, for each fixed $s$, $V_{i,n}^s$ satisfies the assumptions of the $L_1$-norm LLN of Jenish and Prucha (2009, Theorem 3). Therefore, for each $s$, we have

$$\left\| D_n^{-1} \sum_{i \in D_n} [E(Y_{i,n} | \mathcal{F}_{i,n}(s)) - EY_{i,n}] \right\|_1 \to 0 \text{ as } n \to \infty. \quad \text{(B.22)}$$

Furthermore observe that from (B.21) and the Minkowski inequality

$$\left\| D_n^{-1} \sum_{i \in D_n} (Y_{i,n} - E(Y_{i,n} | \mathcal{F}_{i,n}(s))) \right\|_1 \leq \psi(s). \quad \text{(B.23)}$$

Given (B.22) and (B.23), and observing that $\lim_{s \to \infty} \psi(s) = 0$ it now follows that

$$\lim_{n \to \infty} \left\| D_n^{-1} \sum_{i \in D_n} (Y_{i,n} - EY_{i,n}) \right\|_1 = \lim_{s \to \infty} \lim_{n \to \infty} \left\| D_n^{-1} \sum_{i \in D_n} (Y_{i,n} - EY_{i,n}) \right\|_1 \leq \lim_{s \to \infty} \limsup_{n \to \infty} \left\| D_n^{-1} \sum_{i \in D_n} (Y_{i,n} - E(Y_{i,n} | \mathcal{F}_{i,n}(s))) \right\|_1 + \lim_{s \to \infty} \lim_{n \to \infty} \left\| D_n^{-1} \sum_{i \in D_n} (E(Y_{i,n} | \mathcal{F}_{i,n}(s)) - EY_{i,n}) \right\|_1 = 0.$$  

This completes the proof of the LLN. \quad Q.E.D.
C Appendix: Proofs for Section 5

Proof of Theorem 3: We show that
\[ \sup_{\theta \in \Theta} \left| Q_n(\theta) - \overline{Q}_n(\theta) \right| \xrightarrow{p} 0 \] (C.1)
as \( n \to \infty \). As discussed in the text, given that the \( \theta_n^o \) are identifiably unique it then follows immediately from, e.g., Pötscher and Prucha (1997), Lemma 3.1, that \( \nu(\theta_n^o, \theta_n^o) \xrightarrow{p} 0 \) as \( n \to \infty \) as claimed.

We start by proving that
\[ |D_n|^{-1} \sum_{i \in D_n} [q_{i,n}(Z_{i,n}, \theta) - E_q(Z_{i,n}, \theta)] \xrightarrow{p} 0 \] (C.2)
for each \( \theta \in \Theta \), by applying the LLN given as Theorem 2 in the text to \( q_{i,n}(Z_{i,n}, \theta) \). First note that by Assumption 6(a), we have \( \sup_{n,i \in D_n} E |q_{i,n}(Z_{i,n}, \theta)|^p < \infty \) for each \( \theta \in \Theta \) and \( p = 2 \), which verifies Assumption 4(a) for \( q_{i,n}(Z_{i,n}, \theta) \) with \( c_{i,n} = 1 \). By Assumption 6(b), the \( q_{i,n}(Z_{i,n}, \theta) \) are uniformly \( L_1 \)-NED on \( \varepsilon \), and hence w.l.g. we can take \( d_{i,n} = 1 \). Furthermore, by Assumption 6(b) the input process \( \varepsilon \) is \( \alpha \)-mixing, and the \( \alpha \)-mixing coefficients satisfy Assumption 4(b). Consequently (C.2) follows directly from Theorem 2 applied to \( q_{i,n}(Z_{i,n}, \theta) \).

Next observe that by Proposition 1 of Jenish and Prucha (2009), Assumption 6(c) implies that \( q_{i,n} \) is \( L_0 \) stochastically equicontinuous on \( \Theta \), i.e., for every \( \varepsilon > 0 \)
\[ \limsup_{n \to \infty} \frac{1}{|D_n|} \sum_{i \in D_n} P \left( \sup_{\nu(\theta, \theta^*) \leq \delta} |q_{i,n}(Z_{i,n}, \theta) - q_{i,n}(Z_{i,n}, \theta^*)| > \varepsilon \right) \to 0 \text{ as } \delta \to 0. \]
Furthermore, in light of Assumption 6(a) the \( q_{i,n}(Z_{i,n}, \theta) \) clearly satisfy the domination condition postulated by the ULLN in Jenish and Prucha (2009), stated as Theorem 2 in that paper. Given that we have already verified the pointwise LLN in (C.2) it now follows directly from that theorem that
\[ \sup_{\theta \in \Theta} |R_n(\theta) - ER_n(\theta)| \xrightarrow{p} 0 \] (C.3)
with \( R_n(\theta) = |D_n|^{-1} \sum_{i \in D_n} q_{i,n}(Z_{i,n}, \theta) \), and that the \( ER_n(\theta) \) are uniformly equicontinuous on \( \Theta \) in the sense that
\[ \limsup_{n \to \infty} \sup_{\theta^* \in \Theta} \sup_{\nu(\theta, \theta^*) \leq \delta} |ER_n(\theta) - ER_n(\theta^*)| \to 0 \text{ as } \delta \to 0. \]

To prove (C.1) observe that
\[ \sup_{\theta \in \Theta} \left| Q_n(\theta) - \overline{Q}_n(\theta) \right| \geq \sup_{\theta \in \Theta} \left| R_n(\theta)' PR_n(\theta) - ER_n(\theta) PR_n(\theta) \right| + \sup_{\theta \in \Theta} \left| R_n(\theta)' (P_n - P) R_n(\theta) \right| \]
\[ \leq \sup_{\theta \in \Theta} \left| R_n(\theta)' PR_n(\theta) - ER_n(\theta) PR_n(\theta) \right| + 2 \sup_{\theta \in \Theta} \left| R_n(\theta)' (P_n - P) R_n(\theta) \right|. \]
Furthermore observe that Assumption 6(a) we have $E \left[ \sup_{\theta \in \Theta} |q_i, n(Z_i, n, \theta)| \right] \leq K$ and $E \left[ \sup_{\theta \in \Theta} |q_i, n(Z_i, n, \theta)|^2 \right] \leq K$ for some finite constant $K$. Thus

$$\sup_{\theta \in \Theta} E \left| R_n(\theta) \right| \leq E \sup_{\theta \in \Theta} \left| R_n(\theta) \right| \leq |D_n|^{-1} \sum_{i \in D_n} E \sup_{\theta \in \Theta} |q_i, n(Z_i, n, \theta)| \leq K \quad (C.5)$$

and

$$E \sup_{\theta \in \Theta} \left| R_n(\theta) \right|^2 \leq |D_n|^{-2} \sum_{i,j \in D_n} E \left[ \sup_{\theta \in \Theta} |q_i, n(Z_i, n, \theta)| \sup_{\theta \in \Theta} |q_j, n(Z_j, n, \theta)| \right] \leq |D_n|^{-2} \sum_{i,j \in D_n} \left[ E \left( \sup_{\theta \in \Theta} |q_i, n(Z_i, n, \theta)| \right)^2 \right]^{1/2} \left[ E \left( \sup_{\theta \in \Theta} |q_j, n(Z_j, n, \theta)| \right)^2 \right]^{1/2} \leq K. \quad (C.6)$$

Now consider the first terms on the r.h.s. of the last inequality of (C.4). From (C.5) we see that $E \left| R_n(\theta) \right|$ takes on its values in a compact set. Given (C.3) it now follows immediately from part (a.I) of Lemma 3.3 of Pötscher and Prucha (1997) that

$$\sup_{\theta \in \Theta} \left| R_n(\theta)' P R_n(\theta) - E R_n(\theta)' P R_n(\theta) \right| \xrightarrow{P} 0. \quad (C.7)$$

Next we show that also the second term on the r.h.s. of the last inequality of (C.4) converges in probability to zero. To see that this is indeed the case observe that $\sup_{\theta \in \Theta} \left| R_n(\theta) \right|^2 = O_p(1)$ in light of (C.6) and $|P_n - P| \xrightarrow{P} 0$ by assumption. This completes the proof of (C.1).

Having established that $E R_n(\theta)$ are uniformly equicontinuous on $\Theta$, the uniform equicontinuity of $\mathcal{Q}_n(\theta)$ on $\Theta$ follows immediately from Lemma 3.3(b) of Pötscher and Prucha (1997).

**Proof of Theorem 4:** Clearly by Theorem 3 we have $\hat{\theta}_n - \theta^o_n = o_p(1)$.

**Step 1.** The estimators $\hat{\theta}_n$ corresponding to the objective function (23) satisfy the following first order conditions:

$$\nabla_\theta R_n(\hat{\theta}_n)' P_n \left[ |D_n|^{1/2} R_n(\hat{\theta}_n) \right] = o_p(1). \quad (C.8)$$

The $o_p(1)$ term on the r.h.s. reflects that the first order conditions may not hold if $\hat{\theta}_n$ falls onto the boundary of $\Theta$, and that the probability of that event goes to zero as $n \to \infty$, since the $\theta^o_n$ are uniformly in the in the interior of $\Theta$ by Assumption 7(a). If $\hat{\theta}_n$ is in the interior of $\Theta$, then the l.h.s. of (C.8) is zero.

Taking the mean value expansion of $R_n(\theta_n)$ about $\theta^o_n$ yields

$$R_n(\theta_n) = R_n(\theta^o_n) + \nabla_\theta R_n(\bar{\theta}_n)(\theta_n - \theta^o_n) \quad (C.9)$$

where $\bar{\theta}_n \in \Theta$ is between $\hat{\theta}_n$ and $\theta^o_n$ (component-by-component). Let

$$\hat{A}_n = \nabla_\theta R_n(\hat{\theta}_n)' P_n \nabla_\theta R_n(\hat{\theta}_n) \quad \text{and} \quad \hat{B}_n = \nabla_\theta R_n(\hat{\theta}_n)' P_n \left[ |D_n|^{-1} \Sigma_n \right]^{1/2},$$

34
then combining (C.8) and (C.9) gives:

\[
\begin{align*}
|D_n|^{1/2} \left( \hat{\theta}_n - \theta_n^o \right) & \\
= & \left[ I - \hat{A}_n^+ \hat{A}_n \right] |D_n|^{1/2} \left( \hat{\theta}_n - \theta_n^o \right) - \hat{A}_n^+ \nabla_\theta R_n(\hat{\theta}_n) \lambda_n(1)
\end{align*}
\]

Step 2. By Assumptions 7(c) the \( q_{i,n}(Z_{i,n}, \theta_n^o) \) are uniformly \( L_2 \)-NED and uniformly \( L_{2+\delta} \)-integrable with \( c_{i,n} = 1 \). Given Assumptions 7(d),(g) it is now readily seen that the process \{\( q_{i,n}(Z_{i,n}, \theta_n^o) \), \( i \in D_n \)\} satisfies all assumptions of the CLT for vector-valued NED processes, given as Corollary 1 in the text, with \( c_{i,n} = 1 \). (Note that Assumption 3(d) is satisfied automatically since the \( q_{i,n}(Z_{i,n}, \theta_n^o) \) are uniformly \( L_2 \)-NED.) Hence,

\[
\Sigma_n^{-1/2} |D_n| R_n(\theta_n^o) = \Sigma_n^{-1/2} \sum_{i \in D_n} q_{i,n}(Z_{i,n}, \theta_n^o) \Rightarrow N(0, I_{p_q}), \quad (C.11)
\]

with \( \Sigma_n = \operatorname{Var} \left[ \sum_{i \in D_n} q_{i,n}(Z_{i,n}, \theta_n^o) \right] \) and \( \sup_n \lambda \max \left[ |D_n|^{-1} \Sigma_n \right] < \infty \).

Step 3. By Assumptions 7(c),(d),(e) the functions \( \nabla_\theta q_{i,n}(Z_{i,n}, \theta) \) satisfy for each \( \theta \in \Theta \) the LLN given as Theorem 2 in the text with \( c_{i,n} = 1 \), observing that Assumption 4(b) is implied by 2. By argumentation analogous as used in the proof of consistency we have

\[
|D_n|^{-1} \sum_{i \in D_n} \left( \nabla_\theta q_{i,n}(Z_{i,n}, \theta) - E \nabla_\theta q_{i,n}(Z_{i,n}, \theta) \right) \overset{p}{\rightarrow} 0.
\]

By Proposition 1 of Jenish and Prucha (2009), Assumption 7(f) implies that the \( \nabla_\theta q_{i,n}(Z_{i,n}, \theta) \) are uniformly \( L_q \)-equicontinuous on \( \Theta \). Given \( L_q \)-equicontinuity and Assumption 7(e), we have by the ULLN of Jenish and Prucha (2009), Theorem 2,

\[
\sup_{\theta \in \Theta} |\nabla_\theta R_n(\theta) - E \nabla_\theta R_n(\theta)| \overset{p}{\rightarrow} 0. \quad (C.12)
\]

and furthermore, the \( E \nabla_\theta R_n(\theta) \) are uniformly equicontinuous on \( \Theta \) in the sense that

\[
\lim_{n \to \infty} \sup_{\theta' \in \Theta} \sup_{|\theta - \theta'| < \delta} |E \nabla_\theta R_n(\theta) - E \nabla_\theta R_n(\theta)| \to 0 \quad (C.13)
\]
as \( \delta \to 0 \). In light of (C.12) and (C.13), and given that \( \tilde{\theta}_n - \theta_n^o = o_p(1) \) and hence \( \hat{\theta}_n - \theta_n^o = o_p(1) \), if follows further that

\[
\nabla_\theta R_n(\hat{\theta}_n) - E \nabla_\theta R_n(\theta_n^o) \overset{p}{\rightarrow} 0, \quad \text{and} \quad \nabla_\theta R_n(\tilde{\theta}_n) - E \nabla_\theta R_n(\theta_n^o) \overset{p}{\rightarrow} 0.
\]

35
Hence,
\[ \hat{A}_n - A_n \overset{P}{\rightarrow} 0 \quad \text{and} \quad \hat{B}_n - B_n \overset{P}{\rightarrow} 0, \]  
(C.14)
where \( \hat{A}_n \) and \( \hat{B}_n \) are as defined above, and
\[ A_n = \left[ E \nabla_\theta R_n(\theta^o_n) \right]' P \left[ E \nabla_\theta R_n(\theta^o_n) \right] \quad \text{and} \quad B_n = \left[ E \nabla_\theta R_n(\theta^o_n) \right]' P \left[ D_n^{-1} \Sigma_n \right]^{1/2}. \]

**Step 4.** Given Assumptions 7(e),(f), and since \( P \) is positive definite, we have \( |A_n| = O(1) \) and \( |A_n^{-1}| = O(1) \), respectively. Hence by, e.g., Lemma F1 in Pötscher and Prucha (1997) we have \( \hat{A}_n = O_p(1) \), \( \hat{A}_n^+ = O_p(1) \), \( \hat{A}_n \) is nonsingular with probability tending to one, and \( \hat{A}_n^+ - A_n^{-1} \overset{P}{\rightarrow} 0 \). In light of the above it follows from (C.10) that
\[ |D_n|^{1/2} \left( \hat{\theta}_n - \theta^o_n \right) = - \hat{A}_n^+ \hat{B}_n \left[ \Sigma_n^{-1/2} |D_n| R_n(\theta^o_n) \right] + o_p(1) \]
\[ = - A_n^{-1} B_n \left[ \Sigma_n^{-1/2} |D_n| R_n(\theta^o_n) \right] + o_p(1). \]
Recalling that \( \sup_\theta \lambda_{\max} \left[ |D_n|^{-1} \Sigma_n \right] < \infty \), Assumptions 7(e) implies that \( |B_n| = O_p(1) \). In light of Assumption 7(g),(h) \( B_n B'_n \) is invertible and furthermore \( (B_n B'_n)^{-1} = O(1) \). Thus \( \left( A_n^{-1} B_n B'_n A_n^{-1} \right)^{-1} \leq |A_n|^{-2} \left( B_n B'_n \right)^{-1} \) = \( O(1) \) and therefore
\[ \left( A_n^{-1} B_n B'_n A_n^{-1} \right)^{-1/2} |D_n|^{1/2} \left( \hat{\theta}_n - \theta^o_n \right) \]
\[ = - \left( A_n^{-1} B_n B'_n A_n^{-1} \right)^{-1/2} A_n^{-1} B_n \left[ \Sigma_n^{-1/2} |D_n| R_n(\theta^o_n) \right] + o_p(1). \]
The claim that \( \left( A_n^{-1} B_n B'_n A_n^{-1} \right)^{-1/2} |D_n|^{1/2} \left( \hat{\theta}_n - \theta^o_n \right) \Rightarrow N(0, I_k) \) now follows, e.g., from Corollary F4(b) in Pötscher and Prucha (1997).  

**References**


