CONTENTS

On spatial processes and asymptotic inference under near-epoch dependence

Nazgul Jenish [a,1], Ingmar R. Prucha [b,*]

[a] *Department of Economics, New York University, 19 West 4th Street, New York, NY 10012, United States*
[b] *Department of Economics, University of Maryland, College Park, MD 20742, United States*

ARTICLE INFO

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 α-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 α-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 spatial dependence.

1. Introduction

Models with spatially dependent data have recently attracted considerable attention in various fields of economics including labor and public economics, IO, political economy, international and urban economics. In these models, strategic interaction, neighborhood effects, shared resources and common shocks lead to interdependences in the dependent and/or explanatory variables, with the variables indexed by their location in some socioeconomic space.[2] Insofar as these locations are deterministic, observations can be modeled as a realization of a dependent heterogeneous process indexed by a point in $\mathbb{R}^d$, $d > 1$, i.e., as a *random field*.

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.

To date, linear spatial autoregressive models, also known as Cliff and Ord (1981) type models,[3] have arguably been one of the most popular approaches to modeling spatial dependence in the econometrics literature. The asymptotic theory in these models is facilitated, loosely speaking, by imposing specific structural conditions on the data generating process, and by exploiting some underlying independence assumptions. Another popular 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 α-mixing random fields on unevenly spaced lattices that allow for nonstationary processes with trending moments.

---

* Corresponding author. Tel.: +1 3014053499; fax: +1 3014053542.
E-mail addresses: nazgul.jenish@nyu.edu (N. Jenish), prucha@econ.umd.edu (I.R. Prucha).
[1] Tel.: +1 212 998 3891.
[2] The space and metric are not restricted to physical space and distance.

---

[3] For recent contributions see, e.g., Robinson (2010, 2009), Yu et al. (2008), Kelejian and Prucha (2010, 2007, 2004), Lee (2007, 2004), and Chen and Conley (2001).

However, some important classes of dependent processes are not necessarily mixing, including linear autoregressive (AR) and infinite moving average (MA(∞)) processes. Sufficient conditions for the α-mixing property of linear processes[4] are fairly stringent, and involve three types of restrictions (i) smoothness of the density functions of the innovations, (ii) sufficiently fast rates of decay of the coefficients, and (iii) invertibility of the linear process. There are examples demonstrating that the mixing property can fail for any of these reasons. In particular, Andrews (1984) showed that a simple AR(1) process of independent Bernoulli innovations is not α-mixing. Similar examples have been constructed for random fields, see, e.g. Doukhan and Lang (2002). Thus, mixing may break down in the case of discrete innovations. Further, Gorodetskii (1977) showed that the strong mixing property may fail even in the case of continuously distributed (normal) innovations when the coefficients of the linear process do not decline sufficiently fast. As these examples suggest, the mixing property is generally not preserved under infinite transformations of mixing processes. Yet stochastic processes generated as functionals of some underlying process arise in a wide range of models, with autoregressive models being the leading example. 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 (1975), Bierens (1981), Wooldridge (1986), Gallant and White (1988), Andrews (1988), Pötscher and Prucha (1997), Davidson (1992, 1993, 1994) and de Jong (1997), among others. Doukhan and Louhichi (1999) introduced an alternative class of dependent processes called "θ-weakly dependent".

In deriving our limit theorems we then only assume that the process is NED on a mixing input process, i.e., that the process can be approximated by a mixing input process in the NED sense, rather than to assume that the process itself is mixing. Of course, every mixing process is trivially also NED on itself, and thus the class of processes that are NED on a mixing input process includes the class of mixing processes. There are several advantages to working with the enlarged class of process that are NED on a mixing process. First, *linear* processes with discrete innovations, which results in the process to not satisfy the strong mixing property, will still be NED on the mixing input process of innovations, provided the latter are mixing. We note that, in particular, the NED property holds in both examples of Andrews (1984) and Gorodetskii (1977), by Proposition 1 of this paper. Second, as shown in this paper, nonlinear MA(∞) random fields are also NED under some mild conditions, while such conditions are not readily available for mixing. Third, the NED property is often easy to verify. For instance, the sufficient conditions for MA(∞) random fields involve only smoothness conditions on the functional form and absolute summability of the coefficients, which are not difficult to check, while verification of mixing is usually more difficult.

The paper derives a CLT and an LLN for spatial processes that are near epoch dependent on an α-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 a 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 α-mixing.[5]

The rest of the paper is organized as follows. Section 2 introduces the concept of NED spatial processes and gives of examples random fields satisfying this condition. Section 3 contains the LLN and CLT for NED spatial processes. Section 4 establishes the asymptotic properties of spatial GMM estimators. All proofs are relegated to the Appendices.

2. NED spatial processes

Let $D \subset \mathbb{R}^d$, $d \geq 1$, be a lattice of (possibly) unevenly placed 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 $(\Omega, \mathfrak{F}, P)$ with $D_n \subseteq T_n \subseteq D$. The space $\mathbb{R}^d$ is equipped with the metric $\rho(i, j) = \max_{1 \leq l \leq d} |j_l - i_l|$, 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 $\mathbb{R}^{p_\varepsilon}$, respectively. We assume that $\mathbb{R}^{p_z}$ and $\mathbb{R}^{p_\varepsilon}$ are normed metric spaces equipped with the Euclidean norm, which we denote (in an obvious misuse of notation) as $|\cdot|$. For any random vector Y, let $\|Y\|_p = \left[E |Y|^p \right]^{1/p}$, $p \geq 1$, denote its L_p-norm. Finally, let $\mathfrak{F}_{i,n}(s) = \sigma(\varepsilon_{j,n}; j \in T_n : \rho(i, j) \leq s)$ be the σ-field generated by the random vectors $\varepsilon_{j,n}$ located in the s-neighborhood of location i.

Throughout the paper, we maintain these notational conventions and the following assumption concerning D.

Assumption 1. The lattice $D \subset \mathbb{R}^d$, $d \geq 1$, is infinitely 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.

---

[4] These conditions for linear processes with general independent innovations were first established by Gorodetskii (1977). Doukhan and Guyon (1991) generalized them to random fields.

[5] This important early contribution employs Bolthausen's (1982) CLT for stationary α-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.

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 ε if*

$$\left\| Z_{i,n} - E(Z_{i,n}|\mathfrak{F}_{i,n}(s)) \right\|_p \leq d_{i,n}\psi(s) \tag{1}$$

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 ε 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 ε.

Recall that $D_n \subseteq T_n$. Typically, T_n will be an infinite subset of D, and often $T_n = D$. However, as discussed in more detail in Jenish and Prucha (2011), to cover Cliff–Ord type processes 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

$$\begin{aligned} \left\| Z_{i,n} - E(Z_{i,n}|\mathfrak{F}_{i,n}(s)) \right\|_p &\leq \left\| Z_{i,n} \right\|_p + \left\| E(Z_{i,n}|\mathfrak{F}_{i,n}(s)) \right\|_p \\ &\leq 2 \left\| Z_{i,n} \right\|_p \end{aligned}$$

by the Minkowski and the conditional Jensen inequalities. Given this, we may choose $d_{i,n} \leq 2 \left\| Z_{i,n} \right\|_p$, 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 not use the present term. The concept of time series NED processes was later formalized by McLeish (1975), Wooldridge (1986), 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.

We note that aside from the NED condition, a number of different notions of dependence have been used in the time series literature. For instance, Pötscher and Prucha (1997) considered a more general dependence condition (called L_p-approximability). They use more general approximating functions than the conditional mean in Definition 1 to describe the dependence structure of a process. Similar conditions are also used by Lu (2001), Lu and Linton (2007), among others. These conditions allow for more general choices of approximating functions than the conditional expectation. One of the main results in this paper is a central limit theorem, which requires the existence of second moments. Since for $p = 2$ the conditional mean is the best approximator in the sense of minimizing the mean squared error, our use of the conditional mean as an approximating function is not restrictive. Still, in particular applications it may be convenient to work with some other $\mathfrak{F}_{i,n}(s)$-measurable approximating function, say $h_{i,s,n}$.

Of course, if one can show that $\left\| Z_{i,n} - h_{i,s,n} \right\|_2 \leq d_{i,n}\psi(s)$, then this also established (1) for $p = 2$.

In the spatial literature, NED processes were considered in the special context of density estimation by Hallin et al. (2001, 2004), albeit they did not use this term. The first paper proves asymptotic normality of the kernel density estimator for linear random fields, the second paper shows L_1-consistency of the kernel density estimator for nonlinear functionals of i.i.d. random fields. We note that neither of these papers establishes a central limit theorem for nonlinear NED random fields.

As discussed earlier, an important motivation for considering NED processes is that mixing is generally not preserved under transformations involving infinitely many arguments. However, as illustrated below, the output process is generated as a function of infinitely many input variables in a wide range of models. 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 approximation error declines "sufficiently fast" as the conditioning set of input variables expands.

We now give examples of NED spatial processes. First, spatial Cliff and Ord (1981) type autoregressive processes are NED under some weak conditions on the spatial weight coefficients. These models have been used widely in applications. For recent contributions on estimation strategies for these models see, e.g., Robinson (2010, 2009), Kelejian and Prucha (2010, 2007, 2004) and Lee (2007, 2004). The second example is linear infinite moving average (MA(∞)) random fields. In preparation of the example, we first give a more general result, which shows that the NED property is satisfied by random fields generated from nonlinear Lipschitz type functionals of some $\mathbb{R}^{p_\varepsilon}$-valued random field $\varepsilon = \{\varepsilon_{in}, i \in D\}$:

$$Z_{in} = H_{in}((\varepsilon_{jn})_{j \in D}) \tag{2}$$

where $H_{in} : \mathcal{E}^D \to \mathbb{R}^{p_z}, \mathcal{E} \subseteq \mathbb{R}^{p_\varepsilon}$, are measurable functions satisfying for all $e, e' \in \mathcal{E}^D$

$$\left| H_{in}(e) - H_{in}(e') \right| \leq \sum_{j \in D} w_{ijn} \left| e_j - e'_j \right| \quad \text{with } w_{ijn} \geq 0 \tag{3}$$

with

$$\lim_{s \to \infty} \sup_{n,i \in D} \sum_{j \in D: \rho(i,j) > s} w_{ijn} = 0, \quad \text{and} \quad \|\varepsilon\|_2 = \sup_{n,i \in D} \|\varepsilon_{in}\|_2 < \infty. \tag{4}$$

Proposition 1. *Under conditions* (3)–(4), $Z = \{Z_{in}, i \in D_n, n \geq 1\}$ *given by* (2) *is well-defined, and is L_2-NED on ε with $\psi(s) = \|\varepsilon\|_2 \sup_{n,i \in D} \sum_{j \in D: \rho(i,j) > s} w_{ijn}$.*

We now use the above proposition to establish the NED property for linear MA (∞) random fields. Linear MA (∞) random fields may arise as solutions of autoregressive models. For any $k \in \mathbb{N}$ and fixed vectors $v_l \in \mathbb{Z}^d, l = 1, \ldots, k$, consider the following autoregressive random field:

$$Z_i = \sum_{l=1}^{k} a_l Z_{i-v_l} + \varepsilon_i \tag{5}$$

where $a = \sum_{l=1}^{k} |a_l| < 1$, $\{\varepsilon_i, i \in \mathbb{Z}^d\}$ are i.i.d. with $\|\varepsilon_i\|_q < \infty, q \geq 1$. Model (5) is also known as a k-nearest-neighbor or interaction model with the radius of interaction $r = \max_{1 \leq l \leq k} |v_l|$.

As shown by Doukhan and Lang (2002), there exists a stationary solution of (5) given by:

$$Z_i = \sum_{m=0}^{\infty} \sum_{m_1 + \cdots + m_k = m} \frac{m!}{m_1! \ldots m_k!} a_1^{m_1} \ldots a_k^{m_k} \varepsilon_{i-(m_1 v_1 + \cdots + m_k v_k)}$$

with $m_i \in \mathbb{N}$. Thus, (5) can be represented as a linear random field $Z_i = \sum_{j \in \mathbb{Z}^d} w_j \varepsilon_{i-j}$, with

$$w_j = \sum_{m \geq |j|/r} \sum_{V(j,m)} \frac{m!}{m_1! \ldots m_k!} a_1^{m_1} \ldots a_k^{m_k},$$

where $V(j, m) = \{(m_1, \ldots, m_k) \in \mathbb{N}^k : m_1 + \cdots + m_k = m, m_1 v_1 + \cdots + m_k v_k = j\}$, observing that $V(j, m)$ is empty if $m < |j|/r$. Observing further that

$$\sum_{m_1 + \cdots + m_k = m} \frac{m!}{m_1! \ldots m_k!} \left| a_1^{m_1} \ldots a_k^{m_k} \right| = a^m$$

the coefficients w_j can be bounded as

$$\left| w_j \right| \leq \sum_{m \geq |j|/r} \sum_{m_1 + \cdots + m_k = m} \frac{m!}{m_1! \ldots m_k!} \left| a_1^{m_1} \ldots a_k^{m_k} \right|$$

$$= \sum_{m \geq |j|/r}^{\infty} a^m = (1-a)^{-1} a^{|j|/r}.$$

Rewriting the process Z_i as $Z_i = \sum_{j \in \mathbb{Z}^d} w_{ij} \varepsilon_j$ with $w_{ij} = w_{i-j}$ it follows from Proposition 1 that the random field (5) is L_p-NED on ε with the NED coefficients $\psi(s) = \|\varepsilon\|_q (1-a)^{-1} (1-a^{1/r})^{-1} a^{s/r}$.

The asymptotic theory of AR and MA (∞), satisfying the NED condition, can be useful in a variety of empirical applications where the data are cross sectionally correlated. For instance, Pinkse et al.'s (2002) study of spatial price competition among firms that produce differentiated products in one example of an empirical application with cross sectional dependence. They model the price charged by firm at location i in the geographic (or product characteristic) space as a linear spatial autoregressive process. Another example is Fogli and Veldkamp (2011) who investigate spatial correlation in the female labor force participation (LFP). In particular, they consider a spatial autoregression of county i's LFP rate on LFP rates of its neighbors. Dell (2010) examines the impact of *mita*, the forced mining labor system in colonial Peru and Bolivia, on household consumption and child growth across different regions. Although her model is not spatially autoregressive, the regressors and errors exhibit persistent spatial correlation, which can be modeled as a spatial NED process.

As discussed, an 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 $g_{i,n} : \mathbb{R}^{p_z} \to \mathbb{R}$. 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 $(z, z^\bullet) \in \mathbb{R}^{p_z} \times \mathbb{R}^{p_z}$ and all $i \in D_n$ and $n \geq 1$:

$$\left| g_{i,n}(z) - g_{i,n}(z^\bullet) \right| \leq B_{i,n}(z, z^\bullet) \left| z - z^\bullet \right| \tag{6}$$

where $B_{i,n}(z, z^\bullet) : \mathbb{R}^{p_z} \times \mathbb{R}^{p_z} \to \mathbb{R}_+$ is Borel-measurable. Of course, this condition would be devoid of meaning without further restrictions on $B_{i,n}(z, z^\bullet)$, which are given in the next propositions.

Proposition 2. *Suppose $g_{i,n}(\cdot)$ satisfies Lipschitz condition (6) with $\left| B_{i,n}(z, z^\bullet) \right| \leq C < \infty$, for all $(z, z^\bullet) \in \mathbb{R}^{p_z} \times \mathbb{R}^{p_z}$ 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 $\{2Cd_{i,n}\}$.*[6]

Proposition 3. *Suppose $g_{i,n}(\cdot)$ satisfies Lipschitz condition (6) with*

$$\sup_s \left\| B_{i,n}^{(s)} \right\|_2 < \infty \quad and \quad \sup_s \left\| B_{i,n}^{(s)} \left| Z_{i,n} - \widetilde{Z}_{i,n}^s \right| \right\|_r < \infty \tag{7}$$

for some $r > 2$, where $B_{i,n}^{(s)} = B_{i,n}(Z_{i,n}, \widetilde{Z}_{i,n}^s)$ and $\widetilde{Z}_{i,n}^s = E\left[Z_{i,n} | \mathfrak{F}_{i,n}(s) \right]$. If $\left\| g_{i,n}(Z_{i,n}) \right\|_2 < \infty$ and $Z_{i,n}$ is L_2-NED of size $-\lambda$ on $\{\varepsilon_{i,n}\}$ 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 $\{\varepsilon_{i,n}\}$ with scaling factors

$$d'_{i,n} = d_{i,n}^{(r-2)/(2r-2)} \sup_s \left\| B_{i,n}^{(s)} \right\|_2^{(r-2)/(2r-2)}$$

$$\times \left\| B_{i,n}^{(s)} \left| Z_{i,n} - \widetilde{Z}_{i,n}^s \right| \right\|_r^{r/(2r-2)}.$$

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 4.

3. Limit theorems

3.1. Law of large numbers

In this section, we present a LLN for real valued random fields $Z = \{Z_{i,n}, i \in D_n, n \geq 1\}$ that are L_1-NED on some vector-valued α-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. For ease of reference, we state below the definition of the α-mixing coefficients employed in the paper.

Definition 2. Let $\mathfrak{A}$ and $\mathfrak{B}$ be two σ-algebras of $\mathfrak{F}$, and let

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

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 α-mixing coefficients for the random field ε are defined as:

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

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

$$\overline{\alpha}(u, v, r) \leq \varphi(u, v)\widehat{\alpha}(r) \tag{8}$$

where the function $\varphi(u, v)$ is nondecreasing in each argument, and $\widehat{\alpha}(r) \to 0$ as $r \to \infty$. The idea is to account separately for the two

---

[6] The proof of the proposition shows that $\left\| g_{i,n}(Z_{i,n}) - E\left[g_{i,n}(Z_{i,n}) | \mathfrak{F}_{i,n}(s) \right] \right\|_p \leq 2C \left\| Z_{i,n} - \widetilde{Z}_{i,n}^s \right\|_p$, which explains the 2 in the scaling factor for $g_{i,n}(Z_{i,n})$.

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

$$\varphi(u, v) = (u + v)^\tau, \quad \tau \geq 0, \tag{9}$$

$$\varphi(u, v) = \min\{u, v\}. \tag{10}$$

The above mixing conditions have been used extensively in the random fields literature including Takahata (1983), Nahapetian (1987), Bulinskii (1989), Bulinskii and Doukhan (1990) and Bradley (1993). They are satisfied by fairly large classes of random fields. Bradley (1993) provides examples of random fields satisfying conditions (8)–(9) with $u = v$ and $\tau = 1$. Furthermore, Bulinskii (1989) constructs moving average random fields satisfying the same conditions with $\tau = 1$ for any given decay rate of coefficients $\widehat{\alpha}(r)$. Clearly, standard mixing coefficients in the time series literature are covered by conditions (8)–(9) when $\tau = 0$.

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

Assumption 2. (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., $\sup_n \sup_{i \in D_n} E |Z_{i,n}/c_{i,n}|^p < \infty$.
(b) The α-mixing coefficients of the input field ε satisfy (8) for some function $\varphi(u, v)$ which is nondecreasing in each argument, and some $\widehat{\alpha}(r)$ such that $\sum_{r=1}^\infty r^{d-1}\widehat{\alpha}(r) < \infty$.

Theorem 1. *Let $\{D_n\}$ be a sequence of arbitrary finite subsets 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, 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 ε satisfy Assumption 2, then*

$$\frac{1}{M_n |D_n|} \sum_{i \in D_n} (Z_{i,n} - EZ_{i,n}) \xrightarrow{L_1} 0,$$

where $M_n = \max_{i \in D_n} \max(c_{i,n}, d_{i,n})$.

This LLN can 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.

Assumption 2(a) is a standard moment condition employed in weak LLNs for dependent processes. It requires the existence of moments of order slightly greater than 1. As in Theorem 2, $c_{i,n}$ 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 $c_{i,n} = d_{i,n} = 1$. 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.

3.2. Central limit theorem

In this section, we present a CLT for real valued random fields $Z = \{Z_{i,n}, i \in D_n, n \geq 1\}$ that are L_2-NED on some vector-valued α-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 following, we will

use the following notation:

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

The CLT relies on the following assumptions.

Assumption 3. The α-mixing coefficients of ε satisfy (8) and (9) for some $\tau \geq 0$ and $\widehat{\alpha}(r)$, such that for some $\delta > 0$

$$\sum_{r=1}^\infty r^{d(\tau_*+1)-1}\widehat{\alpha}^{\frac{\delta}{2(2+\delta)}}(r) < \infty, \tag{11}$$

where $\tau_* = \delta\tau/(2 + \delta)$.

Assumption 3 restricts the dependence structure of the input process ε. Note that if $\tau = 1$ this assumption also covers the case where $\varphi(u, v)$ is given by (10).

Assumption 4 (*Uniform $L_{2+\delta}$ Integrability*).

(a) 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 3.*
(b) $\inf_n |D_n|^{-1}M_n^{-2}\sigma_n^2 > 0$, where $M_n = \max_{i\in D_n} c_{i,n}$.
(c) NED coefficients satisfy $\sum_{r=1}^\infty r^{d-1}\psi(r) < \infty$.
(d) NED scaling factors satisfy $\sup_n \sup_{i\in D_n} c_{i,n}^{-1}d_{i,n} \leq C < \infty$.

Assumptions 4(a), (b) are standard in the limit theory of mixing processes, e.g., Wooldridge (1986), Davidson (1992), de Jong (1997) and Jenish and Prucha (2009). Assumption 4(a) is satisfied if $Z_{i,n}/c_{i,n}$ are uniformly L_p-bounded for $p > 2 + \delta$, i.e., $\sup_{n,i\in D_n} \|Z_{i,n}/c_{i,n}\|_p < \infty$.

Assumption 4(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, Assumption 4(b) reduces to $\liminf_{n\to\infty} |D_n|^{-1}\sigma_n^2 > 0$, as is, e.g., maintained in Bolthausen (1982). Assumption 4(c) controls the size of the NED coefficients which measure the error in the approximation of $Z_{i,n}$ by ε. Intuitively, the approximation errors have to decline sufficiently fast with each successive approximation. Assumption 4(c) is satisfied if $\psi(r) = O(r^{-d-\gamma})$ for some $\gamma > 0$, i.e., $\psi(r)$ is of size $-d$. Finally, Assumption 4(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 4(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 2. *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 α-mixing random field $\varepsilon = \{\varepsilon_{i,n}, i \in T_n, n \geq 1\}$. Suppose Assumptions 3 and 4 hold, then*

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

Theorem 2 contains the CLT for α-mixing random fields given in Jenish and Prucha (2009) as a special case. It also contains as a special case the CLT for time series NED processes of Wooldridge (1986, see Theorem 3.13 and Corollary 4.4).

Theorem 2 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 α-mixing random field $\varepsilon = \{\varepsilon_{i,n}, i \in T_n, n \geq 1\}$. Suppose Assumptions 3 and 4 hold with $|Z_{i,n}|$ denoting the Euclidean norm of $Z_{i,n}$ and σ_n^2 replaced by $\lambda_{\min}(\Sigma_n)$, where $\Sigma_n = \mathrm{Var}(S_n)$ and $\lambda_{\min}(\cdot)$ is the smallest eigenvalue, then*

$$\Sigma_n^{-1/2} S_n \Longrightarrow N(0, I_k).$$

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

4. Large sample properties of spatial GMM estimators

We now apply the limit theorems of the previous section 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 α-mixing basis process to accommodate spatial lags in the data process as discussed above, (ii) allow for unevenly placed locations, 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.[7]

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

$$Eq_{i,n}(Z_{i,n}, \theta_{0n}) = 0. \tag{12}$$

Then, the corresponding spatial GMM estimator is defined as

$$\widehat{\theta}_n = \arg\min_{\theta \in \Theta} Q_n(\omega, \theta), \tag{13}$$

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

$$Q_n(\omega, \theta) = R_n(\theta)' P_n R_n(\theta),$$

with $R_n(\theta) = |D_n|^{-1} \sum_{i \in D_n} q_{i,n}(Z_{i,n}, \theta)$, and where the P_n are some positive semidefinite weighting matrices. To show consistency, consider the following non-stochastic analogue of Q_n, say

$$\overline{Q}_n(\theta) = [ER_n(\theta)]' P [ER_n(\theta)], \tag{14}$$

where P denotes the probability limit of P_n. Given the moment condition (12), $E[R_n(\theta_{0n})] = 0$, the functions $\overline{Q}_n$ are minimized at θ_{0n}. 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 θ_{0n} we establish, loosely speaking, convergence of the minimizers $\widehat{\theta}_n$ to the minimizers θ_{0n} 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 θ_{0n}.

Assumption 5. (a) The parameter space Θ is a compact metric space with metric ν.
(b) The functions $q_{i,n} : \mathbb{R}^{p_z} \times \Theta \to \mathbb{R}^{p_q}$ are $\mathfrak{B}^{p_z}/\mathfrak{B}^{p_q}$-measurable for each $\theta \in \Theta$, and continuous on Θ for each $z \in \mathbb{R}^{p_z}$.
(c) The elements of the $p_q \times p_q$ real matrices P_n are $\mathfrak{B}$-measurable, and P_n is positive semidefinite. Furthermore $P = p \lim P_n$ exists and P is positive definite.
(d) The minimizers θ_{0n} are identifiably unique in the sense that for every $\varepsilon > 0$, $\liminf_{n \to \infty} \left[\inf_{\theta \in \Theta : \nu(\theta, \theta_{0n}) \geq \varepsilon} [ER_n(\theta)]' [ER_n(\theta)] \right] > 0$.

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 Θ. Given those assumptions the existence of measurable functions $\widehat{\theta}_n$ that solves (13) 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 $\varepsilon > 0$:

$$\liminf_{n \to \infty} \left[\inf_{\theta \in \Theta : \nu(\theta, \theta_{0n}) \geq \varepsilon} \left| \overline{Q}_n(\theta) - \overline{Q}_n(\theta_{0n}) \right| \right] > 0,$$

observing that $\overline{Q}_n(\theta_{0n}) = 0$. Thus, under Assumption 5(d) the minimizers θ_{0n} are identifiably unique; compare, e.g., Gallant and White (1988, p. 19). For interpretation, consider the important special case where $\theta_{0n} = \theta_0$, $ER_n(\theta)$ does not depend on n, and is continuous in θ. In this case, identifiable uniqueness of θ_0 is equivalent to the assumption that θ_0 is the unique solution of the moment conditions, i.e., $E[R_n(\theta)] \neq 0$ for all $\theta \neq \theta_0$; compare, e.g., Pötscher and Prucha (1997, p. 16).

4.1. Consistency

Given the minimizers θ_{0n} are identifiably unique, $\widehat{\theta}_n$ is a consistent estimator for θ_{0n} if Q_n converges uniformly to $\overline{Q}_n$, i.e., if $\sup_{\theta \in \Theta} \left| Q_n(\omega, \theta) - \overline{Q}_n(\theta) \right| \overset{p}{\to} 0$ 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 $\mathfrak{B}^{p_z}/\mathfrak{B}^{p_q}$-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 Θ for some $p > 1$ if $\sup_n \sup_{i \in D_n} E \sup_{\theta \in \Theta} \left| f_{i,n}(Z_{i,n}; \theta) \right|^p < \infty$.
(b) The random functions $f_{i,n}(Z_{i,n}; \theta)$ are said to be Lipschitz in the parameter θ on Θ if

$$\left| f_{i,n}(Z_{i,n}, \theta) - f_{i,n}(Z_{i,n}, \theta^\bullet) \right| \leq L_{i,n}(Z_{i,n}) h(\nu(\theta, \theta^\bullet)) \quad \text{a.s.,} \tag{15}$$

for all $\theta, \theta^\bullet \in \Theta$ 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 $\limsup_{n \to \infty} |D_n|^{-1} \sum_{i \in D_n} EL_{i,n}^\eta < \infty$ for some $\eta > 0$.

Towards establishing consistency of $\widehat{\theta}_n$ 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 Θ 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 ε is α-mixing with α-mixing coefficients the conditions stated in Assumption 2(b).
(c) They are Lipschitz in the parameter θ on Θ.

---

[7] 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 α-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 1.

Assumption 6(a) implies that $\sup_{n,i\in D_n} E\left|q_{i,n}(Z_{i,n};\theta)\right|^p < \infty$ for each $\theta \in \Theta$. Assumption 6(b) then allow us to apply the LLN given as Theorem 1 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 2 or Proposition 3 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 2. 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. θ. 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. θ if $\theta \in \mathbb{R}^k$.

Our consistency results for the spatial GMM estimator given by (13) 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 $D \subset \mathbb{R}^d$, $d \geq 1$ is as in Assumption 1. Suppose further that Assumptions 5 and 6 hold. Then*

$$\nu(\widehat{\theta}_n, \theta_{0n}) \xrightarrow{p} 0 \quad as\ n \to \infty,$$

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

4.2. Asymptotic normality

We next establish that the spatial GMM estimators defined by (13) is asymptotically normally distributed. For that purpose, we need a stronger set of assumptions than for consistency, including differentiability of the moment functions in θ. It proves helpful to adopt the notation ∇_θ in place of $\partial/\partial\theta$.[8]

Assumption 7. (a) The minimizers θ_{0n} lie uniformly in the interior of Θ with $\Theta \subseteq \mathbb{R}^k$. Furthermore $E[R_n(\theta_{0n})] = 0$.
(b) The functions $q_{i,n} : \mathbb{R}^{p_z} \times \Theta \to \mathbb{R}^{p_q}$ are continuously differentiable w.r.t. θ for each $z \in \mathbb{R}^{p_z}$.
(c) The functions $q_{i,n}(Z_{i,n};\theta_{0n})$ are uniformly L_2-NED on ε of size $-d$, and for some $\delta' > 0 \sup_{n,i\in D_n} E\left|q_{i,n}(Z_{i,n};\theta_{0n})\right|^{2+\delta'} < \infty$. The functions $\nabla_\theta q_{i,n}(Z_{i,n};\theta)$ are uniformly L_1-NED on ε.
(d) The input process $\varepsilon = \{\varepsilon_{i,n}, i \in T_n, n \geq 1\}$, where $D_n \subseteq T_n \subseteq D$, is α-mixing and the mixing coefficients satisfy Assumption 3 for some $\delta < \delta'$, where δ' is the same as in Assumption 7(c).
(e) The functions $\nabla_\theta q_{i,n}$ are p-dominated on Θ for some $p > 1$.
(f) The functions $\nabla_\theta q_{i,n}$ are Lipschitz in θ on Θ.
(g) $\inf_n \lambda_{\min}(|D_n|^{-1}\Sigma_n) > 0$ where $\Sigma_n = \text{Var}\left[\sum_{i\in D_n} q_{i,n}(Z_{i,n}, \theta_{0n})\right]$.
(h) $\inf_n \lambda_{\min}\left[E\nabla_\theta R_n(\theta_{0n})'\nabla_\theta R_n(\theta_{0n})\right] > 0$.

The first part of Assumption 7(a) is needed to ensure that the estimator $\widehat{\theta}_n$ lies in the interior of Θ with probability tending to one, and facilitates the application of the mean value theorem to $R_n(\widehat{\theta}_n)$ around θ_{0n}. 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 $R_n(\theta_{0n})$. 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 Θ, which is implied via Assumptions 7(c), (d), (e), (f). Finally, Assumption 7(h) ensures positive-definiteness of the variance–covariance matrix of the GMM estimator.

Given the above assumptions, we have the following asymptotic normality result for the spatial GMM estimator defined by (13).

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*

$$\left(A_n^{-1}B_nB_n'A_n^{-1'}\right)^{-1/2}|D_n|^{1/2}\left(\widehat{\theta}_n - \theta_{0n}\right) \Longrightarrow N(0, I_k),$$

where

$$A_n = [E\nabla_\theta R_n(\theta_{0n})]'\, P\, [E\nabla_\theta R_n(\theta_{0n})] \quad and$$
$$B_n = [E\nabla_\theta R_n(\theta_{0n})]'\, P\, \left[|D_n|^{-1}\,\Sigma_n\right]^{1/2}.$$

Moreover, $|A_n| = O(1)$; $\left|A_n^{-1}\right| = O(1)$; $|B_n| = O(1)$; $\left|\left(B_nB_n'\right)^{-1}\right| = O(1)$ and hence, $\widehat{\theta}_n$ is $|D_n|^{1/2}$-consistent for θ_{0n}.

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 α-mixing input process, rather than postulating that $q_{i,n}$ and $\nabla_\theta q_{i,n}$ are α-mixing. As such, Theorem 4 should provide a basis for constructing confidence intervals and hypothesis testing in a wider range of spatial models.

Using Proposition 3, 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 $\{\theta_{0n}\}$ on Θ, the functions $q_{i,n}(Z_{i,n};\theta_{0n})$ and $\nabla_\theta q_{i,n}(Z_{i,n};\theta_{0n})$ satisfy Lipschitz condition (6) in z, that is, for $g_{i,n} = q_{i,n}$ or $\nabla_\theta q_{i,n}$:

$$\left|g_{i,n}(z;\theta_n) - g_{i,n}(z^\bullet;\theta_n)\right| \leq B_{i,n}(z, z^\bullet)\,|z - z^\bullet|.$$

Furthermore, for *the $r > 2$* as specified in Assumption 8,

$$\sup_{n,i\in D_n} \sup_s \left\|B_{i,n}^{(s)}\right\|_2 < \infty \quad and$$

$$\sup_{n,i\in D_n} \sup_s \left\|B_{i,n}^{(s)}\left|Z_{i,n} - \widetilde{Z}_{i,n}^s\right|\right\|_r < \infty$$

where $B_{i,n}^{(s)} = B_{i,n}(Z_{i,n}, \widetilde{Z}_{i,n}^s)$ with $\widetilde{Z}_{i,n}^s = E\left[Z_{i,n}|\mathfrak{F}_{i,n}(s)\right]$.

5. 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 spatial dependence. More specifically, the paper extends the notion of near-epoch dependent (NED) processes used in the time series literature to spatial processes. This allows us 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 functionals of mixing processes. 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 α-mixing process. Apart from covering a larger class of

---

[8] To ensure that the derivatives are defined on the border of Θ, we assume in the following that the moment functions are defined on an open set containing Θ, and that the $q_{i,n}$ and $\nabla_\theta q_{i,n}$ are restrictions to Θ.

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.[9] One possible approach may be to augment the contributions of the present paper with theory from point processes.

Acknowledgments

Appendix A. Proofs for Sections 2 and 3

Throughout, let $\mathfrak{F}_{i,n}(s) = \sigma(\varepsilon_{j,n}; j \in T_n : \rho(i,j) \leq s)$ be the σ-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.

Proof of Proposition 1. The proof is available online on the authors' webpages. □

Proof of Theorem 1. Define $Y_{i,n} = Z_{i,n}/M_n$, then to prove the theorem, it suffices to show that $|D_n|^{-1} \sum_{i \in D_n} (Y_{i,n} - EY_{i,n}) \xrightarrow{L_1} 0$. 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 2(a)

$$\sup_{n,i \in D_n} E\left|Y_{i,n}\right|^p \leq \sup_{n,i \in D_n} E\left|Z_{i,n}/c_{i,n}\right|^p < \infty. \tag{A.1}$$

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

$$\sup_{n,i \in D_n} \left\|Y_{i,n} - E(Y_{i,n}|\mathfrak{F}_{i,n}(s))\right\|_1 \leq \sup_{n,i \in D_n} M_n^{-1} d_{i,n} \psi(s) \leq \psi(s), \tag{A.2}$$

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 ε.

Next we show that for each given $s > 0$, the conditional mean $V_{i,n}^s = E(Y_{i,n}|\mathfrak{F}_{i,n}(s))$ 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 $s > 0$, $i \in D_n$, $n \geq 1$:

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

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,

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

and thus in light of Assumption 2(b)

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

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

$$\left\||D_n|^{-1} \sum_{i \in D_n} \left[E(Y_{i,n}|\mathfrak{F}_{i,n}(s)) - EY_{i,n}\right]\right\|_1 \to 0 \quad \text{as } n \to \infty. \tag{A.3}$$

Furthermore observe that from (A.2) and the Minkowski inequality

$$\left\||D_n|^{-1} \sum_{i \in D_n} \left(Y_{i,n} - E(Y_{i,n}|\mathfrak{F}_{i,n}(s))\right)\right\|_1 \leq \psi(s). \tag{A.4}$$

Given (A.3) and (A.4), and observing that $\lim_{s \to \infty} \psi(s) = 0$ it now follows that

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

This completes the proof of the LLN. □

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

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

  (i) $V_{ns} \Longrightarrow V_s$ *as $n \to \infty$ for each $s = 1, 2, \ldots$*
 (ii) $V_s \Longrightarrow V$ *as $s \to \infty$, and*
(iii) $\lim_{s \to \infty} \limsup_{n \to \infty} P(|Y_n - V_{ns}| > \epsilon) = 0$ *for every $\epsilon > 0$.*

Then $Y_n \Longrightarrow V$ as $n \to \infty$.

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

---

[9] Pinkse et al. (2007) made an interesting contribution in this direction. Their catalog of assumption is at the level of Bernstein blocks. Without further sufficient conditions, verification of those assumptions would typically be challenging in practical situations.

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 A.3. *Let $\{X_{i,n}\}$ be uniformly L_2-NED on a random field $\{\varepsilon_{i,n}\}$ with α-mixing coefficients $\overline{\alpha}(u, v, r) \leq (u + v)^\tau \widehat{\alpha}(r)$, $\tau \geq 0$. 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) $\left| \text{Cov}\left(X_{i,n} X_{j,n}\right) \right| \leq \|X\|_{2+\delta} \{C_1 \|X\|_{2+\delta} [h/3]^{d\tau_*} \widehat{\alpha}^{\delta/(2+\delta)} ([h/3]) + C_2 \psi ([h/3])\}$, *where $h = \rho(i, j)$ and $\tau_* = \delta\tau/(2 + \delta)$. If, $\sum_{r=1}^\infty r^{d(\tau_*+1)-1} \widehat{\alpha}^{\delta/(2+\delta)}(r) < \infty$, then for some $C < \infty$, not depending on n*

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

(b) $\left| \text{Cov}\left(X_{i,n} X_{j,n}\right) \right| \leq \|X\|_2 \{C_3 \|X\|_{2+\delta} [h/3]^{d\tau^*} \widehat{\alpha}^{\delta/(4+2\delta)} ([h/3]) + C_4 \psi ([h/3])\}$, *where $h = \rho(i, j)$ and $\tau^* = \delta\tau/(4 + 2\delta)$. If, $\sum_{r=1}^\infty r^{d(\tau^*+1)-1} \widehat{\alpha}^{\delta/(4+2\delta)}(r) < \infty$ where $\tau^* = \delta\tau/(4 + 2\delta)$, then for some $C < \infty$, not depending on n*

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

Proof of Lemma A.3. The proof is available online on the authors' webpages. □

Theorem A.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 $\{\varepsilon_{i,n}; i \in D_n, n \in \mathbb{N}\}$ is an array of zero-mean random variables with α-coefficients $\overline{\alpha}(u, v, r) \leq C(u + v)^\tau \widehat{\alpha}(r)$ for some constants $C < \infty$ and $\tau \geq 0$. Suppose for some $\delta > 0$ and $\gamma > 0$*

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

and

$$\widehat{\alpha}(r) = O(r^{-d(2\mu+1)-\gamma})$$

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

$$\sigma_n^{-1} \sum_{i \in D_n} \varepsilon_{i,n} \Longrightarrow N(0, 1)$$

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

Proof of Theorem A.1. The proof is available online on the authors' webpages. □

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

Proof of Theorem 2. 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 $M_n = \max_{i \in D_n} c_{i,n}$ and $Y_{i,n} = Z_{i,n}/M_n$. Also, let $\sigma_{Z,n}^2 = \text{Var}\left[\sum Z_{i,n}\right]$ and $\sigma_{Y,n}^2 = \text{Var}\left[\sum Y_{i,n}\right] = M_n^{-2} \sigma_{Z,n}^2$. Since

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

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

$$S_n = \sum_{i \in D_n} Y_{i,n}, \qquad \sigma_n^2 = \text{Var}(S_n).$$

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 σ_n^2 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

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

and hence

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

so that Assumption 4(a) implies that

$$\lim_{k \to \infty} \sup_{n, i \in D_n} E[|Y_{i,n}|^{2+\delta} \mathbf{1}(|Y_{i,n}| > k)] = 0. \tag{A.5}$$

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

$$\begin{aligned} \left\|Y_{i,n} - E(Y_{i,n}|\mathfrak{F}_{i,n}(s))\right\|_2 &= M_n^{-1} \left\|Z_{i,n} - E(Z_{i,n}|\mathfrak{F}_{i,n}(s))\right\|_2 \\ &\leq c_{i,n}^{-1} d_{i,n} \psi(s) \leq C \psi(s) \end{aligned} \tag{A.6}$$

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 ε with the NED coefficients $\psi(m)$. Finally, observe that by Assumption 4(b):

$$\inf_n |D_n|^{-1} \sigma_n^2 > 0. \tag{A.7}$$

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

$$B|D_n| \leq \sigma_n^2. \tag{A.8}$$

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

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

$$Y_{i,n} = \xi_{i,n}^s + \eta_{i,n}^s$$

where $\xi_{i,n}^s = E(Y_{i,n}|\mathfrak{F}_{i,n}(s))$, $\eta_{i,n}^s = Y_{i,n} - \xi_{i,n}^s$. Let

$$S_{n,s} = \sum_{i \in D_n} \xi_{i,n}^s; \qquad \widetilde{S}_{n,s} = \sum_{i \in D_n} \eta_{i,n}^s$$

$$\sigma_{n,s}^2 = \text{Var}\left[S_{n,s}\right], \qquad \widetilde{\sigma}_{n,s}^2 = \text{Var}\left[\widetilde{S}_{n,s}\right].$$

Repeated use of the Minkowski inequality yields:

$$\left|\sigma_n - \sigma_{n,s}\right| \leq \widetilde{\sigma}_{n,s}, \qquad \left|\sigma_n - \widetilde{\sigma}_{n,s}\right| \leq \sigma_{n,s}. \tag{A.9}$$

Observe that

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

and hence

$$\begin{aligned} &\left\|\eta_{i,n}^s - E(\eta_{i,n}^s|\mathfrak{F}_{i,n}(m))\right\|_2 \\ &= \|Y_{i,n} - E[Y_{i,n}|\mathfrak{F}_{i,n}(s)] - E[Y_{i,n}|\mathfrak{F}_{i,n}(m)] \\ &\quad + E[(Y_{i,n}|\mathfrak{F}_{i,n}(s))|\mathfrak{F}_{i,n}(m)]\|_2 \\ &= \begin{cases} \left\|Y_{i,n} - E(Y_{i,n}|\mathfrak{F}_{i,n}(m))\right\|_2 \leq C\psi(m), & \text{if } m \geq s, \\ \left\|Y_{i,n} - E(Y_{i,n}|\mathfrak{F}_{i,n}(s))\right\|_2 \leq C\psi(s) \leq C\psi(m), & \text{if } m < s \end{cases} \end{aligned}$$

since by definition the sequence $\psi(m)$ is non-increasing. Thus, for any fixed $s > 0$, $\{\eta_{i,n}^s\}$ is uniformly L_2-NED on ε with the same NED coefficients $\psi(m)$ as the random field $\{Y_{i,n}\}$. Furthermore, as shown in the proof of Lemma A.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 3, and observing that $\tau^* = \delta\tau/(4+2\delta) \leq \tau_* = \delta\tau/(2+\delta)$ and $\widehat{\alpha}^{\delta/(2+\delta)}(r) \leq \widehat{\alpha}^{\frac{\delta}{2(2+\delta)}}(r)$ we have

$$\sum_{r=1}^{\infty} r^{d(\tau_*+1)-1}\widehat{\alpha}^{\delta/(2+\delta)}(r) \leq \sum_{r=1}^{\infty} r^{d(\tau_*+1)-1}\widehat{\alpha}^{\frac{\delta}{2(2+\delta)}}(r) < \infty,$$

$$\sum_{r=1}^{\infty} r^{d(\tau^*+1)-1}\widehat{\alpha}^{\delta/(4+2\delta)}(r) \leq \sum_{r=1}^{\infty} r^{d(\tau_*+1)-1}\widehat{\alpha}^{\frac{\delta}{2(2+\delta)}}(r) < \infty.$$

Using part (a) of Lemma A.3 with $X_{i,n} = Y_{i,n}$ and recalling (A.8), we have

$$B|D_n| \leq \sigma_n^2 = \text{Var}\,(S_n) \leq C\,|D_n|\,.$$

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

$$\widetilde{\sigma}_{n,s}^2 = \text{Var}\,(\widetilde{S}_{n,s}) \leq C\,|D_n|\,\|\eta_{i,n}^s\|_2 = C|D_n|\psi(s) \qquad (A.10)$$

in light of (A.6). Hence,

$$\lim_{s\to\infty} \limsup_{n\to\infty} \frac{\widetilde{\sigma}_{n,s}^2}{\sigma_n^2} \leq C\lim_{s\to\infty}\psi(s) = 0. \qquad (A.11)$$

Furthermore, by (A.9) we have

$$\lim_{s\to\infty} \limsup_{n\to\infty} \left|1 - \frac{\sigma_{n,s}}{\sigma_n}\right| \leq \lim_{s\to\infty} \limsup_{n\to\infty} \frac{\widetilde{\sigma}_{n,s}}{\sigma_n} = 0 \qquad (A.12)$$

and hence for all $s \geq 1$ and $n \geq 1$

$$\frac{\sigma_{n,s}}{\sigma_n} \leq C < \infty. \qquad (A.13)$$

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 Theorem A.1.

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

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

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

$$\overline{\alpha}_\xi(u, v, r) \leq \overline{\alpha}(uMs^d, vMs^d, r - 2s) \leq C\,(u+v)^\tau\,\widehat{\alpha}\,(r - 2s)\,.$$

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

$$\sum_{r=1}^{\infty} r^{d(\tau_*+1)-1}\widehat{\alpha}^{\frac{\delta}{2(2+\delta)}}(r) < \infty,$$

where $\tau_* = \delta\tau/(2+\delta)$, which implies

$$\widehat{\alpha}\,(r) = o(r^{-2d(2+\delta)(\tau_*+1)/\delta}) = o(r^{-d[2(\tau+2/\delta)+1]-d})$$
$$= o(r^{-d[2\mu+1]-d})$$

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

We next show that for sufficiently large s,

$$0 < \liminf_{n\to\infty} |D_n|^{-1}\sigma_{n,s}^2.$$

By (A.8),

$$B^{1/2} \leq \inf |D_n|^{-1/2}\sigma_n.$$

Since $\lim_{s\to\infty}\psi(s) = 0$, there exists s_* such that in light of (A.10) for all $s \geq s_*$,

$$|D_n|^{-1/2}\widetilde{\sigma}_{n,s} \leq C\psi^{1/2}(s) \leq B^{1/2}/2. \qquad (A.14)$$

Hence by (A.9) for all $s \geq s_*$, $|D_n|^{-1/2}(\sigma_n - \widetilde{\sigma}_{n,s}) \leq |D_n|^{-1/2}\sigma_{n,s}$, and thus $\inf_n |D_n|^{-1/2}\sigma_{n,s} \geq \inf_n |D_n|^{-1/2}\sigma_n - \sup_n |D_n|^{-1/2}\widetilde{\sigma}_{n,s}$. Using (A.7) and (A.14), we have

$$\liminf_{n\to\infty} |D_n|^{-1/2}\sigma_{n,s} \geq B^{1/2} - \frac{B^{1/2}}{2} = \frac{B^{1/2}}{2} > 0.$$

Thus, for all $s \geq s_*$,

$$\sigma_{n,s}^{-1}\sum_{i\in D_n} \xi_{i,n}^s \Longrightarrow N(0, 1) \quad \text{as } n \to \infty. \qquad (A.15)$$

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

5. *CLT for* $\sigma_n^{-1}\sum_{i\in D_n} Y_{i,n}$.

Finally, using Lemma A.1 we now show that, given the maintained NED assumption, the just established CLT in (A.15) for the approximators $\xi_{i,n}^s$ can be carried over to the $Y_{i,n}$. Define

$$W_n = \sigma_n^{-1}\sum_{i\in D_n} Y_{i,n}, \qquad V_{ns} = \sigma_n^{-1}\sum_{i\in D_n} \xi_{i,n}^s,$$

$$W_n - V_{ns} = \sigma_n^{-1}\sum_{i\in D_n} \eta_{i,n}^s$$

so that we can exploit Lemma A.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 A.1. By Markov's inequality and (A.11), for every $\epsilon > 0$ we have

$$\lim_{s\to\infty} \limsup_{n\to\infty} P(|W_n - V_{ns}| > \epsilon)$$

$$= \lim_{s\to\infty} \limsup_{n\to\infty} P\left(\left|\sigma_n^{-1}\sum_{i\in D_n} \eta_{i,n}^s\right| > \epsilon\right)$$

$$\leq \lim_{s\to\infty} \limsup_{n\to\infty} \frac{\widetilde{\sigma}_{n,s}^2}{\epsilon^2\sigma_n^2} = 0.$$

Next observe that $V_{ns} = \frac{\sigma_{n,s}}{\sigma_n}\left[\sigma_{n,s}^{-1}\sum_{i\in D_n} \xi_{i,n}^s\right]$. We proceed to show $W_n \Longrightarrow V$ by contradiction. For that purpose let $\mathcal{M}$ be the set of all probability measures on $(\mathbb{R}, \mathfrak{B})$, and observe that we can metrize $\mathcal{M}$ by, e.g., the Prokhorov distance $d(\cdot, \cdot)$. Let μ_n and μ be the probability measures corresponding to W_n and V, respectively, then $W_n \Longrightarrow V$, or $\mu_n \Longrightarrow \mu$, iff $d(\mu_n, \mu) \to 0$ as $n \to \infty$. Now suppose μ_n does not converge to μ. Then for some $\epsilon > 0$ there exists a subsequence $\{n(m)\}$ such that $d(\mu_{n(m)}, \mu) > \epsilon$ for all $n(m)$. By (A.13), we have $0 \leq \sigma_{n,s}/\sigma_n \leq C < \infty$ for all $s, n \geq 1$. Hence, $0 \leq \sigma_{n(m),s}/\sigma_{n(m)} \leq C < \infty$ for all $n(m)$. Consequently, for $s = 1$ there exists a subsubsequence $\{n(m(l_1))\}$ 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 subsubsubsequence $\{n(m(l_1(l_2)))\}$ 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 $\{n_l\}$ 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

$$\lim_{l\to\infty} \frac{\sigma_{n_l,s}}{\sigma_{n_l}} = p(s) \qquad (A.16)$$

for $s = 1, 2, \ldots$ Given (A.15), it follows that as $l \to \infty$

$$V_{n_l s} \Longrightarrow V_s \sim N(0, p^2(s)).$$

Then, it follows from (A.12) that

$$\lim_{s \to \infty} |p(s) - 1| \leq \lim_{s \to \infty} \lim_{l \to \infty} \left| p(s) - \frac{\sigma_{n_l, s}}{\sigma_{n_l}} \right|$$
$$+ \lim_{s \to \infty} \sup_{n \geq 1} \left| \frac{\sigma_{n, s}}{\sigma_n} - 1 \right| = 0.$$

Thus $V_s \Longrightarrow V$ and thus by Lemma A.1 $W_{n_l} \Longrightarrow V \sim N(0, 1)$ as $l \to \infty$. Since $\{n_l\} \subseteq \{n(m)\}$ 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. The proof is available online on the authors' webpages. □

Appendix B. Proofs for Section 4

Proof of Theorem 3. We show that

$$\sup_{\theta \in \Theta} |Q_n(\theta) - \overline{Q}_n(\theta)| \xrightarrow{p} 0 \tag{B.1}$$

as $n \to \infty$. As discussed in the text, given that the θ_{0n} are identifiably unique it then follows immediately from, e.g., Pötscher and Prucha (1997, Lemma 3.1), that $\nu(\widehat{\theta}_n, \theta_{0n}) \xrightarrow{p} 0$ as $n \to \infty$ as claimed.

We start by proving that

$$|D_n|^{-1} \sum_{i \in D_n} \left[q_{i,n}(Z_{i,n}, \theta) - E q_{i,n}(Z_{i,n}, \theta) \right] \xrightarrow{p} 0 \tag{B.2}$$

for each $\theta \in \Theta$, by applying the LLN given as Theorem 1 in the text to $q_{i,n}(Z_{i,n}, \theta)$. By Assumption 6(a), we have $\sup_{n, i \in D_n} E \left| q_{i,n}(Z_{i,n}, \theta) \right|^p < \infty$ for each $\theta \in \Theta$ and $p = 2$, which verifies Assumption 2(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 ε, and hence w.o.l.g. we can take $d_{i,n} = 1$. Furthermore, by Assumption 6(b) the input process ε is α-mixing, and the α-mixing coefficients satisfy Assumption 2(b). Consequently (B.2) follows directly from Theorem 1 applied to $q_{i,n}(Z_{i,n}, \theta)$.

Next, by Proposition 1 of Jenish and Prucha (2009), Assumption 6(c) implies that $q_{i,n}$ is L_0 stochastically equicontinuous on Θ, i.e., for every $\varepsilon > 0$

$$\limsup_{n \to \infty} \frac{1}{|D_n|} \sum_{i \in D_n} P \left(\sup_{\nu(\theta, \theta^\bullet) \leq \delta} \left| q_{i,n}(Z_{i,n}, \theta) - q_{i,n}(Z_{i,n}, \theta^\bullet) \right| > \varepsilon \right)$$
$$\to 0 \quad \text{as } \delta \to 0.$$

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 (B.2) it now follows directly from that theorem that

$$\sup_{\theta \in \Theta} |R_n(\theta) - E R_n(\theta)| \xrightarrow{p} 0 \tag{B.3}$$

with $R_n(\theta) = |D_n|^{-1} \sum_{i \in D_n} q_{i,n}(Z_{i,n}, \theta)$, and that the $E R_n(\theta)$ are uniformly equicontinuous on Θ in the sense that

$$\limsup_{n \to \infty} \sup_{\theta^\bullet \in \Theta} \sup_{\nu(\theta, \theta^\bullet) \leq \delta} |E R_n(\theta) - E R_n(\theta^\bullet)| \to 0 \quad \text{as } \delta \to 0.$$

To prove (B.1), observe that

$$\sup_{\theta \in \Theta} |Q_n(\theta) - \overline{Q}_n(\theta)|$$
$$\leq \sup_{\theta \in \Theta} |R_n(\theta)' P R_n(\theta) - E R_n(\theta) P E R_n(\theta)|$$
$$+ \sup_{\theta \in \Theta} |R_n(\theta)'(P_n - P) R_n(\theta)|$$
$$\leq \sup_{\theta \in \Theta} |R_n(\theta)' P R_n(\theta) - E R_n(\theta) P E R_n(\theta)|$$
$$+ 2 \sup_{\theta \in \Theta} |R_n(\theta)|^2 |P_n - P|. \tag{B.4}$$

Furthermore observe that Assumption 6(a) we have $E[\sup_{\theta \in \Theta} |q_{i,n}(Z_{i,n}, \theta)|] \leq K$ and $E \left[\sup_{\theta \in \Theta} |q_{i,n}(Z_{i,n}, \theta)| \right]^2 \leq K$ for some finite constant K. Thus

$$\sup_{\theta \in \Theta} E |R_n(\theta)| \leq E \sup_{\theta \in \Theta} |R_n(\theta)| \leq |D_n|^{-1}$$
$$\times \sum_{i \in D_n} E \sup_{\theta \in \Theta} |q_{i,n}(Z_{i,n}, \theta)| \leq K \tag{B.5}$$

$$E \sup_{\theta \in \Theta} |R_n(\theta)|^2$$
$$\leq |D_n|^{-2} \sum_{i,j \in D_n} E \left[\sup_{\theta \in \Theta} |q_{i,n}(Z_{i,n}, \theta)| \sup_{\theta \in \Theta} |q_{j,n}(Z_{j,n}, \theta)| \right]$$
$$\leq |D_n|^{-2} \sum_{i,j \in D_n} \left[E \left(\sup_{\theta \in \Theta} |q_{i,n}(Z_{i,n}, \theta)| \right)^2 \right]^{1/2}$$
$$\times \left[E \left(\sup_{\theta \in \Theta} |q_{j,n}(Z_{j,n}, \theta)| \right)^2 \right]^{1/2} \leq K. \tag{B.6}$$

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

$$\sup_{\theta \in \Theta} |R_n(\theta)' P R_n(\theta) - E R_n(\theta) P E R_n(\theta)| \xrightarrow{p} 0. \tag{B.7}$$

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

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

Proof of Theorem 4. Clearly by Theorem 3 we have $\widehat{\theta}_n - \theta_{0n} = o_p(1)$.

Step 1. The estimators $\widehat{\theta}_n$ corresponding to the objective function (13) satisfy the following first order conditions:

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

The $o_p(1)$ term on the r.h.s. reflects that the first order conditions may not hold if $\widehat{\theta}_n$ falls onto the boundary of Θ, and that the probability of that event goes to zero as $n \to \infty$, since the θ_{0n} are uniformly in the in the interior of Θ by Assumption 7(a). If $\widehat{\theta}_n$ is in the interior of Θ, then the l.h.s. of (B.8) is zero.

Taking the mean value expansion of $R_n(\widehat{\theta}_n)$ about θ_{0n} yields

$$R_n(\widehat{\theta}_n) = R_n(\theta_{0n}) + \nabla_\theta R_n(\widetilde{\theta}_n)(\widehat{\theta}_n - \theta_{0n}) \tag{B.9}$$

where $\widetilde{\theta}_n \in \Theta$ is between $\widehat{\theta}_n$ and θ_{0n} (component-by-component). Let

$$\widehat{A}_n = \nabla_\theta R_n(\widehat{\theta}_n)' P_n \nabla_\theta R_n(\widetilde{\theta}_n) \quad \text{and}$$

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

then combining (B.8) and (B.9) gives

$$
\begin{aligned}
|D_n|^{1/2} \left(\widehat{\theta}_n - \theta_{0n}\right) &= \left[I - \widehat{A}_n^+ \widehat{A}_n\right] |D_n|^{1/2} \left(\widehat{\theta}_n - \theta_{0n}\right) \\
&\quad - \widehat{A}_n^+ \nabla_\theta R_n(\widehat{\theta}_n)' P_n \left[|D_n|^{1/2} R_n(\theta_{0n})\right] \\
&\quad + \widehat{A}_n^+ o_p(1) \\
&= \left[I - \widehat{A}_n^+ \widehat{A}_n\right] |D_n|^{1/2} \left(\widehat{\theta}_n - \theta_{0n}\right) \\
&\quad - \widehat{A}_n^+ \widehat{B}_n \left[\Sigma_n^{-1/2} |D_n| R_n(\theta_{0n})\right] \\
&\quad + \widehat{A}_n^+ o_p(1),
\end{aligned}
\tag{B.10}
$$

where $\widehat{A}_n^+$ denotes the generalized inverse of $\widehat{A}_n$.

Step 2. By Assumptions 7(c) the $q_{i,n}(Z_{i,n}, \theta_{0n})$ 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 $\{q_{i,n}(Z_{i,n}, \theta_{0n}), i \in D_n\}$ satisfies all assumptions of the CLT for vector-valued NED processes, given as Corollary 1 in the text, with $c_{in} = 1$. (Note that Assumption 4(d) is satisfied automatically since the $q_{i,n}(Z_{i,n}, \theta_{0n})$ are uniformly L_2-NED.) Hence,

$$
\begin{aligned}
\Sigma_n^{-1/2} |D_n| R_n(\theta_{0n}) &= \Sigma_n^{-1/2} \sum_{i \in D_n} q_{i,n}(Z_{i,n}, \theta_{0n}) \\
&\Longrightarrow N(0, I_{p_q}),
\end{aligned}
\tag{B.11}
$$

with $\Sigma_n = \operatorname{Var}\left[\sum_{i \in D_n} q_{i,n}(Z_{i,n}, \theta_{0n})\right]$ 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 1 in the text with $c_{i,n} = 1$, observing that Assumption 2(b) is implied by 3. By argumentation analogous as used in the proof of consistency we have

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

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_0-equicontinuous on Θ. Given L_0-equicontinuity and Assumption 7(e), we have by the ULLN of Jenish and Prucha (2009):

$$\sup_{\theta \in \Theta} |\nabla_\theta R_n(\theta) - E\nabla_\theta R_n(\theta)| \xrightarrow{p} 0. \tag{B.12}$$

and furthermore, the $E\nabla_\theta R_n(\theta)$ are uniformly equicontinuous on Θ in the sense:

$$\limsup_{n \to \infty} \sup_{\theta' \in \Theta} \sup_{|\theta - \theta'| < \delta} |E\nabla_\theta R_n(\theta) - E\nabla_\theta R_n(\theta)| \to 0 \tag{B.13}$$

as $\delta \to 0$. In light of (B.12) and (B.13), and given that $\widehat{\theta}_n - \theta_{0n} = o_p(1)$ and hence $\widetilde{\theta}_n - \theta_{0n} = o_p(1)$, if follows further that

$$\nabla_\theta R_n(\widehat{\theta}_n) - E\nabla_\theta R_n(\theta_{0n}) \xrightarrow{p} 0, \quad \text{and}$$

$$\nabla_\theta R_n(\widetilde{\theta}_n) - E\nabla_\theta R_n(\theta_{0n}) \xrightarrow{p} 0.$$

Hence,

$$\widehat{A}_n - A_n \xrightarrow{p} 0 \quad \text{and} \quad \widehat{B}_n - B_n \xrightarrow{p} 0, \tag{B.14}$$

where $\widehat{A}_n$ and $\widehat{B}_n$ are as defined above, and

$$A_n = [E\nabla_\theta R_n(\theta_{0n})]' P [E\nabla_\theta R_n(\theta_{0n})] \quad \text{and}$$

$$B_n = [E\nabla_\theta R_n(\theta_{0n})]' P \left[|D_n|^{-1} \Sigma_n\right]^{1/2}.$$

Step 4. Given Assumptions 7(e), (f), and since P is positive definite, we have $|A_n| = O(1)$ and $\left|A_n^{-1}\right| = O(1)$, respectively. Hence by, e.g., Lemma F1 in Pötscher and Prucha (1997) we have $\widehat{A}_n = O_p(1), \widehat{A}_n^+ = O_p(1), \widehat{A}_n$ is nonsingular with probability tending to one, and $\widehat{A}_n^+ - A_n^{-1} \xrightarrow{p} 0$. In light of the above it follows from (B.10) that

$$
\begin{aligned}
|D_n|^{1/2} \left(\widehat{\theta}_n - \theta_{0n}\right) &= -\widehat{A}_n^+ \widehat{B}_n \left[\Sigma_n^{-1/2} |D_n| R_n(\theta_{0n})\right] + o_p(1) \\
&= -A_n^{-1} B_n \left[\Sigma_n^{-1/2} |D_n| R_n(\theta_{0n})\right] + o_p(1).
\end{aligned}
$$

Recalling that $\sup_n \lambda_{\max}\left[|D_n|^{-1} \Sigma_n\right] < \infty$, Assumptions 7(e) implies that $|B_n| = O_p(1)$. In light of Assumptions 7(g), (h) $B_n B_n'$ is invertible and furthermore $(B_n B_n')^{-1} = O(1)$. Thus $\left|\left(A_n^{-1} B_n B_n' A_n^{-1\prime}\right)^{-1}\right| \leq |A_n|^2 \left|(B_n B_n')^{-1}\right| = O(1)$ and therefore

$$
\begin{aligned}
&\left(A_n^{-1} B_n B_n' A_n^{-1\prime}\right)^{-1/2} |D_n|^{1/2} \left(\widehat{\theta}_n - \theta_{0n}\right) \\
&= -\left(A_n^{-1} B_n B_n' A_n^{-1\prime}\right)^{-1/2} A_n^{-1} B_n \left[\Sigma_n^{-1/2} |D_n| R_n(\theta_{0n})\right] + o_p(1).
\end{aligned}
$$

The claim that $\left(A_n^{-1} B_n B_n' A_n^{-1\prime}\right)^{-1/2} |D_n|^{1/2} \left(\widehat{\theta}_n - \theta_{0n}\right) \Longrightarrow N(0, I_k)$ now follows, e.g., from Corollary F4(b) in Pötscher and Prucha (1997). $\quad\square$

References

Andrews, D.W.K., 1984. Non-strong mixing autoregressive processes. Journal of Applied Probability 21, 930–934.

Andrews, D.W.K., 1987. Consistency in nonlinear econometric models: a generic uniform law of large numbers. Econometrica 55, 1465–1471.

Andrews, D.W.K., 1988. Laws of large numbers for dependent non-identically distributed variables. Econometric Theory 4, 458–467.

Bierens, H.J., 1981. Robust Methods and Asymptotic Theory. Springer Verlag, Berlin.

Billingsley, P., 1968. Convergence of Probability Measures. John Wiley and Sons, New York.

Bolthausen, E., 1982. On the central limit theorem for stationary mixing random fields. Annals of Probability 10, 1047–1050.

Bradley, R., 1993. Some examples of mixing random fields. The Rocky Mountain Journal of Mathematics 23 (2), 495–519.

Brockwell, P., Davis, R., 1991. Times Series: Theory and Methods. Springer Verlag.

Bulinskii, A.V., 1989. Limit Theorems Under Weak Dependence Conditions. Moscow University Press.

Bulinskii, A.V., Doukhan, P., 1990. Vitesse de convergence dans le theorem de limite centrale pour des champs melangeants des hypotheses de moment faibles. Comptes Rendus de l'Academie des Sciences, Serie I (Mathematique) 801–805.

Chen, X., Conley, T., 2001. A new semiparametric spatial model for panel time series. Journal of Econometrics 105, 59–83.

Cliff, A., Ord, J., 1981. Spatial Processes, Models and Applications. Pion, London.

Conley, T., 1999. GMM estimation with cross sectional dependence. Journal of Econometrics 92, 1–45.

Davidson, J., 1992. A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes. Econometric Theory 8, 313–329.

Davidson, J., 1993. An L_1 convergence theorem for heterogeneous mixingale arrays with trending moments. Statistics and Probability Letters 8, 313–329.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press.

de Jong, R.M., 1997. Central limit theorems for dependent heterogeneous random variables. Econometric Theory 13, 353–367.

Dell, M., 2010. The persistent effects of Peru's mining mita. Econometrica 78, 1863–1903.

Dobrushin, R., 1968. The description of a random field by its conditional distribution and its regularity condition. Theory of Probability and its Applications 13, 197–227.

Doukhan, P., Guyon, X., 1991. Mixing for linear random fields. Comptes Rendus de l'Academie des Sciences, Serie I (Mathematique) 313, 465–470.

Doukhan, P., Lang, G., 2002. Rates in the empirical central limit theorem for stationary weakly dependent random fields. Statistical Inference for Stochastic Processes 5, 199–228.

Doukhan, P., Louhichi, S., 1999. A new weak dependence condition and applications to moment inequalities. Stochastic Processes & their Applications 84, 313–342.

Fogli, A., Veldkamp, L., 2011. Nature or nurture? Learning and geography of female force participation. Econometrica 79, 1103–1138.

Gallant, A.R., White, H., 1988. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Basil Blackwell, New York.

Gorodetskii, V.V., 1977. On the strong mixing property for linear sequences. Theory of Probability and Its Applications 22, 411–413.

Hallin, M., Lu, Z., Tran, L.T., 2001. Density estimation for spatial linear processes. Bernoulli 7, 657–668.

Hallin, M., Lu, Z., Tran, L.T., 2004. Kernel density estimation for spatial processes: the L_1 theory. Journal of Multivariate Analysis 88, 61–75.

Ibragimov, I.A., 1962. Some limit theorems for stationary processes. Theory of Probability and its Applications 7, 349–382.

Ibragimov, I.A., Linnik, Y.V., 1971. Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen.

Jenish, N., Prucha, I.R., 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. Journal of Econometrics 150, 86–98.

Jenish, N., Prucha, I.R., 2011. On spatial processes and asymptotic inference under near epoch dependence. Working Paper. Version January.

Kelejian, H.H., Prucha, I.R., 2004. Estimation of simultaneous systems of spatially interrelated cross sectional equations. Journal of Econometrics 118, 27–50.

Kelejian, H.H., Prucha, I.R., 2007. HAC estimation in a spatial framework. Journal of Econometrics 140, 131–154.

Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53–67.

Lee, L.-F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.

Lee, L.-F., 2007. GMM and 2SLS for mixed regressive, spatial autoregressive models. Journal of Econometrics 137, 489–514.

Lu, Z., 2001. Asymptotic normality of kernel density estimators under dependence. Annals of the Institute of Statistical Mathematics 53, 447–468.

Lu, Z., Linton, O., 2007. Local linear fitting under near epoch dependence. Econometric Theory 23, 37–70.

McLeish, D.L., 1975. Invariance principles for dependent variables. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 32, 165–178.

Nahapetian, B., 1987. An approach to central limit theorems for dependent random variables. Theory of Probability and its Applications 32, 589–594.

Pinkse, J., Shen, L., Slade, M.E., 2007. A central limit theorem for endogenous locations and complex spatial interactions. Journal of Econometrics 140, 215–225.

Pinkse, J., Slade, L.M.E., Brett, , 2002. Spatial price competition: a semiparametric approach. Econometrica 70, 1111–1153.

Pötscher, B.M., Prucha, I.R., 1989. A uniform law of large numbers for dependent and heterogeneous data processes. Econometrica 57, 675–683.

Pötscher, B.M., Prucha, I.R., 1994. Generic uniform convergence and equicontinuity concepts for random functions. Journal of Econometrics 60, 23–63.

Pötscher, B.M., Prucha, I.R., 1997. Dynamic Nonlinear Econometric Models. Springer-Verlag, New York.

Robinson, P.M., 2009. Large-sample inference on spatial dependence. Journal of Econometrics 12, S68–S82. Tenth Anniversary Special Issue.

Robinson, P.M., 2010. Efficient estimation of the semiparametric spatial autoregressive model. Journal of Econometrics 157, 6–17.

Takahata, H., 1983. On the rates in the central limit theorem for weakly dependent random fields. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 64, 445–456.

Wooldridge, J., 1986. Asymptotic properties of econometric estimators. University of California San Diego, Department of Economics. Ph.D. Dissertation.

Yu, J., de Jong, R., Lee, L.-F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both N and T are large. Journal of Econometrics 146, 118–134.