Econometric Theory, **32**, 2016, 714–739.
doi:10.1017/S0266466614000905

SPATIAL SEMIPARAMETRIC MODEL WITH ENDOGENOUS REGRESSORS

NAZGUL JENISH

This paper proposes a semiparametric generalized method of moments estimator (GMM) estimator for a partially parametric spatial model with endogenous spatially dependent regressors. The finite-dimensional estimator is shown to be consistent and root-n asymptotically normal under some reasonable conditions. A spatial heteroscedasticity and autocorrelation consistent covariance estimator is constructed for the GMM estimator. The leading application is nonlinear spatial autoregressions, which arise in a wide range of strategic interaction models. To derive the asymptotic properties of the estimator, the paper also establishes a stochastic equicontinuity criterion and functional central limit theorem for near-epoch dependent random fields.

1. INTRODUCTION

Strategic behavior of agents is one of the defining features of modern economic models. Such behavior often involves interaction among economic agents and thus leads to interdependence of their choices. For instance, a monopolistically competitive firm takes into account prices charged by other neighboring firms in its price-setting decision. From an econometric point of view, this implies a model in which other observations on the response variable enter the regression as endogenous regressors, i.e., a spatially autoregressive model.

The existing estimation theory of spatially dependent models has mainly focused on fully parametric or nonparametric specifications. Recent contributions include Lee (2007), Robinson (2011), Jenish and Prucha (2012), and Jenish (2012), among others. However, nonparametric estimation suffers from the curse of dimensionality, while parametric estimation is susceptible to serious misspecification problems. Semiparametric estimation is the standard compromise between the two approaches. Yet there have been few contributions to semiparametric estimation of spatially dependent models except for Gao, Lu and Tjøstheim (2006), Su (2012), Robinson and Thawornkaiwong (2012).

In this paper, we propose a semiparametric GMM estimator of a spatial model with endogenous regressors which may include spatial lags of the response variable. This model can arise as an equilibrium of economic games, and thus have a wide range of applications. It has two key characteristics: nonlinear spatial interactions or spillovers in the dependent variable, and flexible functional

The author is grateful to Co-Editor Oliver Linton and two anonymous referees for their constructive comments which led to a considerable improvement of the paper. Address correspondence to Nazgul Jenish, Department of Economics, New York University, 19 West 4th Street, New York, NY 10012. Tel.: 212-998-3891, email: nazgul.jenish@nyu.edu

forms for the effect of the exogenous variables. Given spatial dependence in the response variable, the data-generating process is assumed to be spatially near-epoch dependent (NED) on some mixing process. As shown in Jenish (2012), this weak dependence condition is less restrictive than mixing and satisfied in many important applications including nonlinear spatial autoregressive models. Under this dependence condition, we prove consistency and root-n asymptotic normality of the finite-dimensional estimator. To derive the asymptotic results, the paper also establishes a stochastic equicontinuity criterion and functional central limit theorem for near-epoch dependent random fields, which may be also useful in other semiparametric and nonparametric settings. We also construct a heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimator for our semiparametric estimator.

To our knowledge, the proposed semiparametric GMM estimator has not been considered in the existing literature. Gao, Lu and Tjøstheim (2006) suggest a two-step semiparametric estimator of a partially linear regression for stationary α-mixing random fields. The α-mixing concept was originally introduced by Rosenblatt (1956). Rosenblatt (1985) further extends this dependence concept to random fields, and establishes a number of important asymptotic results on the parameter and spectral density estimation of α-mixing random fields. Our model relies on a weaker dependence condition than mixing. It also generalizes Gao et al. (2006) by allowing endogenous regressors and nonlinear parametric functions. Su (2012) studies a linear spatial autoregressive model with independent innovations, in which spatial dependence is modeled by means of a known spatial weight matrix and an unknown scalar parameter, also known as a Cliff–Ord spatial autoregressive model. Our model is different from Cliff–Ord type specifications in important ways. First, it does not assume any known spatial weight matrix, and hence allows for a more robust estimation of both the autoregressive parameters and the covariance matrix of their estimators. Second, the proposed asymptotic theory employs the machinery of random fields, while the limit theory of Cliff–Ord type models often exploits a parametric dependence structure represented by a linear-quadratic form of the independent innovation process. However, neither of these asymptotic approaches dominates the other. In the version of the Cliff–Ord model analyzed by Su (2012), the spatial lags enter the model linearly. In contrast to Su (2012), our semiparametric estimator allows for nonlinear autoregressions. This feature renders the asymptotic theory more intricate, and in particular requires establishing a stochastic equicontinuity property of the empirical process, which is not needed in Su (2012). Recently, Robinson and Thawornkaiwong (2012) have established root-n asymptotic normality of an IV estimator of a partially linear regression. Their asymptotic theory exploits a specific cross-sectional dependence structure in the error process generated as a linear process of independent innovations, which precludes some of autoregressive random fields considered in this paper.

The proposed estimator also differs from semiparametric GMM estimators in the time series literature, which typically relies on the i.i.d. assumption. It generalizes the seminal contribution of Robinson (1988) on semiparametric estimation of a partially linear regression with exogenous regressors to the spatial endogenous setting. Andrews (1994) provides a general framework for proving consistency and root-n asymptotic normality of semiparametric estimators that minimize a criterion function that depends on a preliminary infinite-dimensional nuisance parameter estimator. Our GMM estimator fits this framework. However, verification of the high-level assumption of that paper is nontrivial, and requires a series of new results on stochastic equicontinuity, uniform convergence rates, and asymptotic normality for weakly dependent spatial processes established in this paper.

The structure of the paper is as follows. Section 2 describes the estimation procedure. Section 3 establishes rates of uniform consistency of the first-step nonparametric estimator as well as consistency of the finite-dimensional parameter estimator. Section 4 proves root-n asymptotic normality of the finite-dimensional estimator and constructs a consistent estimator of its variance matrix. Section 5 gives a stochastic equicontinuity criterion and functional central limit theorem for near-epoch dependent random fields. Section 6 contains a Monte Carlo study. All proofs are collected in the appendices.

2. MODEL AND ESTIMATION PROCEDURE

We begin by introducing some basic notation. All random processes are defined on a common probability space $(\Omega, \mathfrak{F}, P)$, and take their values in $\mathbb{R}^p$, which is equipped with the Euclidean norm denoted by $|\cdot|$. Furthermore, let $\|X\|_q = [E|X|^q]^{1/q}$ denote the L_q-norm of a random variable X, and $|A| = trace(A'A)^{1/2}$ denote the norm of a nonrandom matrix A.

We consider double arrays, $\{W_{in}, i \in \mathbb{Z}^d, n \geq 1\}$, of random fields, i.e., stochastic processes indexed by $i = (i_1, \ldots, i_d) \in \mathbb{Z}^d$. In this paper, we are concerned with estimation of the semiparametric model:

$$Y_{1in} = h(Y_{2in}, \theta_0) + g(X_{in}) + U_{in} \text{ for } X_{in} \in \mathcal{X}, \quad (1)$$

$$E(U_{in}|X_{in}) = 0 \text{ a.s.} \quad (2)$$

$$E(Z_{in}U_{in}) = 0, \quad (3)$$

where Y_{1in} is the scalar response variable, Y_{2in} is the vector of (possibly) endogenous variables, X_{in} is the vector of exogenous variables that takes values in some set $\mathcal{X} \subseteq \mathbb{R}^k$, Z_{in} is the vector of instruments, θ_0 is the finite-dimensional parameter vector, $h(\cdot,\cdot)$ is a known function, and $g(\cdot)$ is an unknown function. Let $W_{in} = \left(X_{in}', Y_{1in}, Y_{2in}', Z_{in}'\right)'$ take values in the Euclidean space, $\mathbb{R}^{p_w}$.

In the absence of cross-sectional dependence, (1)–(3) reduces to some well-known models in the econometrics literature. For example, if Y_{2in} is exogenous and $h(Y_{2in}; \theta) = Y_{2in}'\theta$, then (1)–(3) is the partially linear regression studied by Robinson (1988), and for nonlinear $h(\cdot,\cdot)$, the model is the partially parametric regression analyzed by Andrews (1994).

The vector of endogenous variables Y_{2in} may contain spatial lags of Y_{1in}. This feature is critical for a broad array of applications ranging from IO and international economics to social interactions and networks. For instance, consider the following spatial autoregressive model:

$$Y_i = h_1\left(\left(Y_{i-j}\right)_{j\in\mathbb{Z}^d, 0<|j|\le r}, \theta_1\right) + h_2\left(X_{2i}, \theta_2\right) + g\left(X_{1i}\right) + U_i, \tag{4}$$

where $r > 0$ is the fixed radius of interaction or neighborhood, $X_i = \left(X'_{1i}, X'_{2i}\right)'$ are exogenous variables, $h_2(\cdot)$ and $g(\cdot)$ are, respectively, known and unknown functions, and $h_1 : \mathbb{R}^{|N(r)|} \times \mathbb{R}^{p_\theta} \to \mathbb{R}$ is a known function of spatial lags with $N(r) = \left\{ j \in \mathbb{Z}^d : 0 < |j| \le r\right\}$. If $\{X_i\}$ and $\{U_i\}$ are independent, model (4) satisfies conditions (1)-(3) with the instruments $Z_i = \left(X_{i-j}\right)_{j\in N(r)}$.

This model may arise in a number of applications. For example, in a model of spatial price competition between firms, the dependent variable would be the price charged by the firm at location i that depends on the prices of its neighbors, see Pinkse et al. (2002). Another example is a housing demand model in which the demand at location i is correlated with the demand at the neighboring locations, see Ioannides and Zabel (2003).

The data-generating process $\{W_{in}\}$ in model (4) exhibits intrinsic spatial dependence. For estimators to have desirable statistical properties, some restrictions need to be placed on the dependence structure of $\{W_{in}\}$. Specifically, $\{W_{in}\}$ has to satisfy some weak dependence property, e.g., mixing. As discussed in Jenish and Prucha (2012), the mixing property may fail in autoregressive models for a number of reasons, including the discrete nature of the innovations, slow decay of the coefficients etc. At the same time, many autoregressive processes satisfy a less restrictive dependence condition called near-epoch-dependence (NED), for examples see Jenish (2012).

In the following, we therefore assume that $\{W_{in}\}$ is spatially NED on some α-mixing random field $\{V_{in}\}$. For the definition of α-mixing random fields, see, e.g., Jenish (2012). As for the NED concept, we state its definition below:

DEFINITION 1. *The random field* $\left\{W_{in}, i \in \mathbb{Z}^d, n \ge 1\right\}$, $\|W_{in}\|_q < \infty$, $q \ge 1$, *is* L_q*-NED on* $\{V_{in}, i \in \mathbb{Z}^d, n \ge 1\}$ *iff*

$$\sup_{n, i\in\Lambda_n} \|W_{in} - E(W_{in}|\mathfrak{F}_{in}(s))\|_q \le \psi(s)$$

for some sequence $\psi(s) \to 0$ *as* $s \to \infty$, *where* $\mathfrak{F}_{in}(s) = \sigma(V_{jn}; j \in \mathbb{Z}^d : |i - j| \le s)$. $\{W_{in}\}$ *is* L_q*-NED of size* $-\eta$ *if* $\psi(s) = O(s^{-\eta-\varepsilon})$ *for some* $\varepsilon > 0$.

The NED concept was introduced by Ibragimov (1962) and Billingsley (1968), and has since been used extensively in the time series literature.

For example, the NED condition is satisfied in model (4). Specifically, if for some $a_j \ge 0$ and any $y, y' \in \mathbb{R}^{|N(r)|}$

$$\left|h_1(y) - h_1(y')\right| \le \sum_{j\in N(r)} a_j \left|y_j - y'_j\right|, \; 0 < a = \sum_{j\in N(r)} a_j < 1, \tag{5}$$

and $\|V_i\|_2 < \infty$, $V_i = h_2(X_{2i}, \theta_2) + g(X_{1i}) + U_i$, then by Proposition 1 of Jenish (2012), there exists a unique solution of (4) that is L_2-NED on $\{V_i\}$ with the NED coefficients $\psi(s) = 2\|Y_0\|_2 a^{-1} a^{[s/r]}$, where $\|Y_0\|_2 = \sup_i \|Y_i\|_2$ and $[\cdot]$ is the integer part function.

To estimate model (1)–(3), we proceed as in Robinson (1988). Taking expectation of (1) conditional on X_{in} and subtracting it from (1) gives

$$\varrho(Y_{1in}, Y_{2in}, X_{in}, \theta_0) \equiv U_{in} = Y_{1in} - E(Y_{1in}|X_{in}) - h(Y_{2in}, \theta_0) + E(h(Y_{2in}; \theta_0)|X_{in}).$$

Then, moment conditions imply

$$E\left[\mathbf{1}_{\mathcal{X}}(X_{in})(Z_{in} - E(Z_{in}|X_{in}))\varrho(Y_{1in}, Y_{2in}, X_{in}, \theta_0)\right] = 0, \quad \textbf{(6)}$$

where $\mathbf{1}_{\mathcal{X}}(\cdot)$ is the indicator of $\mathcal{X}$. Thus, the parameter θ_0 can be identified from moment condition (6) provided that the latter has a unique solution $\theta = \theta_0$. Define

$$\tau_{10}(x) = E(Z_{in}|X_{in} = x), \ \tau_{20}(x) = E(Y_{1in}|X_{in} = x),$$
$$\tau_{30}(x, \theta_0) = E(h(Y_{2in}; \theta_0)|X_{in} = x), \ \tau_0 = (\tau_{10}, \tau_{20}, \tau_{30}),$$

and re-write moment condition (6) as

$$E[m(W_{in}, \theta_0, \tau_0)] = 0, \quad \textbf{(7)}$$

where

$$m(W_{in}, \theta, \tau) = \mathbf{1}_{\mathcal{X}}(X_{in}) m^*(W_{in}, \theta, \tau), \quad \textbf{(8)}$$
$$m^*(W_{in}, \theta, \tau) = (Z_{in} - \tau_1(X_{in}))(Y_{1in} - \tau_2(X_{in}) - h(Y_{2in}, \theta) + \tau_3(X_{in}, \theta)). \quad \textbf{(9)}$$

Clearly, the GMM estimator of θ based on (7) is infeasible since τ_0 is unknown. Nevertheless, we can replace τ_0 by a consistent nonparametric estimator, $\widehat{\tau}$, and obtain a feasible GMM estimator of θ, $\widehat{\theta}$, based on these approximate moment conditions. Under some regularity conditions that guarantee asymptotic independence of $\widehat{\tau}$ and $\widehat{\theta}$, the feasible estimator $\widehat{\theta}$ will be asymptotically equivalent to the infeasible estimator. This observation suggests the following estimation strategy. We first estimate τ_0 by the Nadaraya–Watson kernel estimator:

$$\widehat{\tau}_1(x) = \left(\overline{n} b_{1n}^k\right)^{-1} \sum_{i \in \Lambda_n} Z_{in} K_1((x - X_{in})/b_{1n}) / \widehat{f}_1(x), \quad \textbf{(10)}$$
$$\widehat{\tau}_2(x) = \left(\overline{n} b_{2n}^k\right)^{-1} \sum_{i \in \Lambda_n} Y_{1in} K_2((x - X_{in})/b_{2n}) / \widehat{f}_2(x),$$
$$\widehat{\tau}_3(x, \theta) = \left(\overline{n} b_{3n}^k\right)^{-1} \sum_{i \in \Lambda_n} h(Y_{2in}, \theta) K_3((x - X_{in})/b_{3n}) / \widehat{f}_3(x),$$
$$\widehat{f}_l(x) = \left(\overline{n} b_{ln}^k\right)^{-1} \sum_{i \in \Lambda_n} K_l\left(\frac{x - X_{in}}{b_{ln}}\right), l = 1, 2, 3,$$

where $\overline{n} = |\Lambda_n|$ is the sample size, $k = \dim X_{in}$, and $\widehat{\tau}_1(x)$ is the $p_z \times 1$ ($p_z = \dim Z_{in}$) vector-function whose components are element-by-element kernel estimators of Z_{in}. We then obtain a GMM estimator $\widehat{\theta}$ by minimizing

$$Q_n(\theta) = \overline{m}_n'(\theta, \widehat{\tau}) \Sigma_n \overline{m}_n(\theta, \widehat{\tau}),$$

where Σ_n is some weighting matrix and $\overline{m}_n(\theta, \tau) = \overline{n}^{-1} \sum_{i \in \Lambda_n} m_{\varepsilon_n}(W_{in}, \theta, \tau)$ is the sample analog of moment function (7). The sample moment function is constructed according to the formula

$$m_\varepsilon(W_{in}, \theta, \tau) = \zeta_\varepsilon(X_{in}) m^*(W_{in}, \theta, \tau),$$

where $m^*(\cdot)$ is as in (9), and $\zeta_\varepsilon(\cdot)$ is the smoothed version of the indicator trimming function, $\mathbf{1}_{\mathcal{X}}(\cdot)$, given by the formula:

$$\zeta_\varepsilon(x) = \int_{\mathbb{R}^k} 1_{\mathcal{X}_\varepsilon}(x - \varepsilon z) \phi(z) dz = \frac{1}{\varepsilon^k} \int_{\mathcal{X}_\varepsilon} \phi\left(\frac{x-z}{\varepsilon}\right) dz, \tag{11}$$

$$\phi(x) = \begin{cases} c_k \exp\left(1/\left(|x|^2 - 1\right)\right), & |x| < 1 \\ 0, & |x| \geq 1 \end{cases}$$

$$\mathcal{X}_\varepsilon = \{u : |x - u| < \varepsilon, \varepsilon > 0, \text{ for some } x \in \mathcal{X}\},$$

$$\varepsilon_n = \overline{n}^{-1/2-\epsilon} \varepsilon^*/3, \text{ for some } \epsilon > 0 \text{ and } \varepsilon^* > 0.$$

In general, semiparametric estimators based on a first-step nonparametric kernel estimator require trimming to guarantee that the denominator of the kernel estimator is bounded away from zero on some bounded set. The latter is critical for uniform consistency of the kernel estimator, and consequently, for consistency of the second-step estimator. The standard choice of the trimming function is the indicator function of the set $\mathcal{X}$ on which the density of X_{in} is bounded away from zero. For instance, $\mathcal{X}$ could be chosen as $\mathcal{X} = \left\{x : \inf_{n \geq 1} \inf_{\mathcal{X}} \overline{n}^{-1} \sum_{i \in \Lambda_n} f_{in}(x) \geq L\right\}$ for some $L > 0$. In our case, we replace indicator function $\mathbf{1}_{\mathcal{X}}(\cdot)$ with its smoothed version $\zeta_\varepsilon(\cdot)$ to ensure smoothness of the sample moment function. The latter is needed to verify the NED property of the sample moment function from that of W_{in}.

Mathematically, the smoothed trimming function $\zeta_\varepsilon(x)$ is the convolution of the indicator function with a mollifier function $\phi(\cdot)$. This operation transforms the discontinuous indicator function into an infinitely differentiable function $\zeta_\varepsilon(x)$. When $\varepsilon \to 0$, the smoothed trimming function, $\zeta_\varepsilon(\cdot)$, converges in L_1-norm to $\mathbf{1}_{\mathcal{X}}(\cdot)$. The constant c_k is a normalizing constant such that $\int \phi(x) dx = 1$. The properties of $\zeta_\varepsilon(\cdot)$ are collected in Lemma B.2 of the Appendix.

3. CONSISTENCY

Throughout the sequel, we consider q-times continuously differentiable functions with finite Sobolev norm:

$$\|f\|_{q,r,\mathcal{U}} = \sum_{|\mu| \leq q} \left(\int_{\mathcal{U}} \left|D^\mu f(u)\right|^r du\right)^{1/r}, \tag{12}$$

where $D^\mu f(u) = \frac{\partial^{|\mu|}}{\partial u_1^{\mu_1} \ldots \partial u_p^{\mu_p}} f(u)$, $u = (u_1, \ldots, u_p)$ is a p-vector of nonnegative integers with $|\mu| = \sum_{j=1}^{p} \mu_j$. Let $\mathcal{S}^{q,r}(\mathcal{U})$, $1 \le r < \infty$, denote the Sobolev space endowed with the above norm, and let $\mathcal{C}^\omega(\mathcal{U})$ denote the space of ω-times continuously differentiable functions on $\mathcal{U}$.

Consistency of the finite-dimensional parameter estimator relies heavily on uniform consistency of nonparametric estimators $\widehat{\tau}$. Uniform consistency of $\widehat{\tau}$ is in turn ensured by the following conditions:

Assumption 1.

(a) $\Theta \subset \mathbb{R}^{p_\theta}$ is compact, $\theta_0 \in \Theta$, $\mathcal{X} \subset \mathbb{R}^k$ is open bounded. Θ^* and $\mathcal{X}^*$ are open ε^*-neighborhoods of Θ and $\mathcal{X}$, respectively, for small $\varepsilon^* > 0$, $\mathcal{U} = \mathcal{X}^* \times \Theta^*$ is a Lipschitz domain [1].

(b) For some large $B < \infty$ and $q_1, q_2, q_3 > 0$, $\tau_0 = (\tau_{10}, \tau_{20}, \tau_{30}) \in \mathcal{T}$, where

$$\mathcal{T} = \Big\{(\tau_1, \tau_2, \tau_3) \colon \|\tau_1(\cdot)\|_{q_1,2,\mathcal{X}^*} \le B,\ \|\tau_2(\cdot)\|_{q_2,2,\mathcal{X}^*} \le B,\ \|\tau_3(\cdot,\cdot)\|_{q_3,2,\mathcal{X}^* \times \Theta^*} \le B\Big\}.$$

Assumption 2.

(a) $\mathcal{W} = \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Z}$ is an open bounded subset of $\mathbb{R}^{p_w}$.

(b) $\{W_{in} = \left(X_{in}', Y_{1in}, Y_{2in}', Z_{in}'\right)', i \in \Lambda_n\}$ is L_2-NED on $\left\{V_{in}, i \in \mathbb{Z}^d\right\}$ of size $-\eta$, $\eta > d$. $\left(Y_{1in}, Y_{2in}', Z_{in}'\right)'$ lie in $\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Z}$.

(c) Mixing numbers of $\{V_{in}\}$ satisfy $\alpha(k,l,r) \le (k+l)^\varsigma \widehat{\alpha}(r)$, $\varsigma \ge 0$, $\widehat{\alpha}(r)$ s.t. $\sum_{r=1}^\infty r^{d(\varsigma+1)-1} \widehat{\alpha}(r) < \infty$.[2]

Assumption 3.

(a) X_{in} is continuous with the density $f_{in}(x) \in \mathcal{C}^\omega(\mathbb{R}^k)$ for $\omega \ge \max\{q_1, q_2, q_3\} + 1$.

(b) $\sup_{n\ge1} \sup_{\mathbb{R}^k} \left|\overline{n}^{-1} \sum_{i\in\Lambda_n} D^\mu f_{in}(x)\right| < \infty$ for $|\mu| \le \omega$.

(c) $\inf_{n\ge1} \inf_{\mathcal{X}^*} \overline{n}^{-1} \sum_{i\in\Lambda_n} f_{in}(x) \ge L > 0$, where $\mathcal{X}^*$ is bounded.

Assumption 4. For ω defined in Assumption 3(a):

(a) $\tau_{l0}(x) f_{in}(x) \in \mathcal{C}^\omega(\mathbb{R}^k)$ and $\sup_{n\ge1} \sup_{\mathbb{R}^k} \left|\overline{n}^{-1} \sum_{i\in\Lambda_n} D^\mu \left[\tau_{l0}(x) f_{in}(x)\right]\right| < \infty$, $l = 1, 2$, for $|\mu| \le \omega$.

(b) $\tau_{30}(x,\theta) f_{in}(x) \in \mathcal{C}^\omega(\mathbb{R}^k \times \Theta^*)$ and $\sup_{n\ge1} \sup_{\mathbb{R}^k \times \Theta^*} \left|\overline{n}^{-1} \sum_{i\in\Lambda_n} D^{\mu_x} D^{\mu_\theta} \left[\tau_{30}(x,\theta) f_{in}(x)\right]\right| < \infty$ for $|\mu_x| + |\mu_\theta| \le \omega$.

(c) $\sup_{\Theta^*} \sup_{\mathcal{Y}_2} |D^{\mu_y} D^{\mu_\theta} h(y;\theta)| < \infty$ for $\left|\mu_y\right| + |\mu_\theta| \le q_3$.

Assumption 5. The kernels $K_l : \mathbb{R}^k \to \mathbb{R}$, $l = 1,2,3$ satisfy:

(a) $\int K_l(x)dx = 1$, $\int x^\mu K_l(x)dx = 0$ for $1 \le |\mu| \le \omega - 1$, $\int |x^\mu K_l(x)|\, dx < \infty$ for $|\mu| = \omega$.

https://doi.org/10.1017/S0266466614000905 Published online by Cambridge University Press

(b) For $|\mu| \leq q_l$, $l = 1,2,3$, $D^{\mu} K_l(x) \to 0$ as $|x| \to \infty$, and $D^{\mu} K_l(x)$ are absolutely integrable and have Fourier transforms $\Psi_{l,\mu}(\lambda) = (2\pi)^k \int \exp(\mathbf{i}\lambda' x) D^{\mu} K_l(x) dx$ satisfying $\int (1 + |\lambda|)|\Psi_{l,\mu}(\lambda)| d\lambda < \infty$, where $\mathbf{i} = \sqrt{-1}$.

Assumption 1 defines the set of nonparametric functions $\tau(\cdot)$. This set needs to be restricted to obtain stochastic equicontinuity of the moment functions in τ. Assumption 2 specifies the dependence structure of the data-generating process. Assumption 3 summarizes the properties of the density function of X_{in}. Assumption 3(a) can be relaxed to allow for mixed discrete-continuous regressors X_{in} as in Bierens (1983). Assumption 3(b) is a boundedness condition on the density of X_{in}. Assumption 3(c) is critical as uniform consistency holds only over bounded sets on which the densities are bounded away from zero. Assumption 4 contains standard smoothness and boundedness conditions required to quantify the rates of convergence. Assumption 5(a) requires bias-reducing kernels, which are needed to obtain faster rates of convergence. Assumption 5(b) is a technical condition used in the proof of Theorem 1, which relies on the Fourier transform of the kernel.

THEOREM 1. *Let $\{\Lambda_n\}$ be a sequence of finite sets of $\mathbb{Z}^d$ s.t. $\overline{n} = |\Lambda_n| \to \infty$. Under Assumptions 1-5,*

A. $\sup_{\mathcal{X}^*} |D^{\mu_x} \widehat{\tau}_l(x) - D^{\mu_x} \tau_{l0}(x)| = O_p(\overline{n}^{-\eta/(2\eta+d)} b_{ln}^{-k-|\mu_x|-d/(2\eta+d)}) + O_p(b_{ln}^{\omega-|\mu_x|})$, $|\mu_x| \leq q_l$, $l = 1,2$;

B. $\sup_{\mathcal{X}^* \times \Theta^*} |D^{\mu_x} D^{\mu_\theta} \widehat{\tau}_3(x,\theta) - D^{\mu_x} D^{\mu_\theta} \tau_{30}(x,\theta)| = O_p(\overline{n}^{-\eta/(2\eta+d)} b_{3n}^{-k-|\mu_x|-d/(2\eta+d)}) + O_p(b_{3n}^{\omega-|\mu_\theta|-|\mu_x|})$, *for* $|\mu_\theta| + |\mu_x| \leq q_3$, *provided that the right-hand sides of both equalities are* $o_p(1)$.

Part A of Theorem 1 extends results of Bierens (1983) and Andrews (1995) to NED random fields. For the case one-dimensional lattice ($d = 1$), the rates of convergence reduce to those in Theorem 1(b) of Andrews (1995). Part B of the theorem establishes uniform convergence rates over parameter θ, in addition to x, for functions that also depend on the finite-dimensional parameter.

The uniform convergence rates in Theorem 1 are suboptimal, see Stone (1982) and Lu and Linton (2007). Relative to the i.i.d. case, there is a loss in the speed of uniform convergence. The magnitude of this loss depends on the rate of decay of the NED coefficients, η, and the dimension of lattice, d: convergence is slower, the stronger the dependence and the higher the dimension of the lattice.

To ensure that the rates obtained in Theorem 1 are $o_p(1)$, we use the following

Assumption 6. $b_{ln} = c_l \overline{n}^{-\gamma_l}$, with $c_l > 0$ and $0 < \gamma_l < \eta/[(2\eta+d)(k+q_l)+d]$, $l = 1,2,3$.

This bandwidth condition is not restrictive for most applications. For instance, it is automatically satisfied in model (4). In this and many other models (see Jenish, 2012), the NED coefficients decay at an exponential rate, i.e., $\eta = \infty$, and Assumption 6 reduces to $0 < \gamma < 1/[2(k+q)]$, where k is the dimension of x and q is the degree of smoothness of $\tau(\cdot)$. The latter condition is reasonable for practical values of k and q, e.g., if $k = 1$ and $q = 3$, then γ must satisfy $0 < \gamma < 1/8$.

Based on Theorem 1, we can now establish consistency of the parametric estimator, which requires additionally the following assumption:

Assumption 7.

(a) $E\left[m(W_{in},\theta_0,\tau_0)\right] = 0$. $p\lim_{n\to\infty}\Sigma_n = \Sigma$, Σ is positive definite, and There exists a function $\overline{f}(x)$ such that $\lim_{n\to\infty}\sup_{x\in\mathcal{X}}\left|\overline{n}^{-1}\sum_{i\in\Lambda_n} f_{in}(x) - \overline{f}(x)\right| = 0$.

(b) $\tau_{40}(x,\theta) = E[Z_{in}h(Y_{2in},\theta)|X_{in} = x]$ and $\tau_{50}(x) = E[Z_{in}Y_{1in}|X_{in} = x]$ do not depend on i or n, $\sup_{\mathcal{X}\times\Theta}|\tau_{40}(x,\theta)| < \infty$ and $\sup_{\mathcal{X}}|\tau_{50}(x)| < \infty$.

(c) $m(\theta,\tau_0)'\Sigma m(\theta,\tau_0)$ is uniquely minimized on Θ at θ_0, where $m(\theta,\tau)$ is a nonrandom function defined:

$$m(\theta,\tau) = \int_{\mathcal{X}}[\Upsilon(\theta,\tau(x)) + \Upsilon_1(\theta,x)]\overline{f}(x)dx, \text{ where:}$$

$$\Upsilon(\theta,\tau(x)) = [\tau_{10}(x) - \tau_1(x)][\tau_{20}(x) - \tau_2(x) - \tau_{30}(x,\theta) + \tau_3(x,\theta)],$$

$$\Upsilon_1(\theta,x) = \tau_{10}(x)[\tau_{30}(x,\theta) - \tau_{30}(x,\theta_0)] + \tau_{40}(x,\theta_0) - \tau_{40}(x,\theta).$$

(d) $\varepsilon_n = \overline{n}^{-1/2-\epsilon}\varepsilon^*/3$ for some $\epsilon > 0$ and ε^* defined in Assumption 1.

Assumption 7 is the identification condition: parts (a), (b), and (d) ensure convergence of the objective function to a finite nonrandom function in part (c) that is uniquely minimized at the true parameter. In particular, part (d) guarantees convergence of the smooth trimming function to $\mathbf{1}_{\mathcal{X}}(\cdot)$. The parameter ε_n is chosen such that the trimming set, $\mathcal{X}_{2\varepsilon_n}$, lies inside the set $\mathcal{X}^*$, and ε_n decays at a rate faster than $\overline{n}^{-1/2}$, not to affect the asymptotic distribution of the finite-dimensional estimator.

THEOREM 2. *Let* $\{\Lambda_n\}$ *be a sequence of finite sets of* $\mathbb{Z}^d$ *s.t.* $\overline{n}=|\Lambda_n| \to \infty$. *Under Assumptions 1–7,* $\widehat{\theta}_n \xrightarrow{p} \theta_0$.

4. ASYMPTOTIC NORMALITY AND COVARIANCE MATRIX ESTIMATION

In this section, we derive the limiting distribution of the parametric estimator and construct its covariance matrix estimator. We maintain Assumptions 1–5 and 7, which imply consistency. For asymptotic normality, we need to strengthen the bandwidth condition as follows:

Assumption 8. $b_{ln} = c_l\overline{n}^{-\gamma_l}$, with $c_l > 0$, $l = 1,2,3$

$$[4\omega]^{-1} < \gamma_l < \min\{\eta/[(2\eta+d)(k+q_l)+d],\ (2\eta-d)/[4k(2\eta+d)+4d]\},$$

https://doi.org/10.1017/S0266466614000905 Published online by Cambridge University Press

where ω is as in Assumption 3(a), q_1, q_2, q_3 are defined in Assumption 1 and, in addition, satisfy $\min(q_1, q_2, q_3) > (p_w + 1)/2$ for $p_w = \dim W_{in}$.

These stronger bandwidth conditions are needed for asymptotic independence of $\widehat{\tau}$ and $\widehat{\theta}$. Assumption 8 ensures $\overline{n}^{1/4} \sup_{\mathcal{X}\times\Theta} |\widehat{\tau}(x,\theta) - \tau_0(x,\theta)| \xrightarrow{p} 0$. This condition is standard in the semiparametric literature, see, e.g., Andrews (1994). Assumption 8 is not void. For example, if $q > 3k + 2/3$ and $\eta > d$, then $\eta/[(2\eta+d)(k+q_l)+d] < (2\eta-d)/[4(2\eta+d)k+4d]$, and the set of γ satisfying $[4\omega]^{-1} < \gamma < \eta/[(2\eta+d)(k+q)+d]$ is nonempty for any $\omega > q > k(2\eta+d)/(2\eta-d)-2$. For exponential decay rates of the NED coefficients, Assumption 8 reduces to $1/(4\omega) < \gamma < \min\{1/(2k+2q),\ 1/(4k)\}$, which is compatible for all $k \le q < \omega$.

Assumption 9.

(a) θ_0 is in the interior of Θ, and $\mathcal{W}$ is an open bounded convex subset of $\mathbb{R}^{p_w}$. Assumption 2(c) holds with $\widehat{\alpha}(r)$ s.t. $\sum_{r=1}^{\infty} r^{d(\varsigma+1)-1}\widehat{\alpha}^{1/2}(r) < \infty$.

(b) There exists a nonrandom matrix $M(\theta,\tau)$ such that $M = M(\theta_0,\tau_0)$ is of full column rank and

$$\sup_{\Theta\times\mathcal{T}} \left| \overline{n}^{-1} \sum_{i\in\Lambda_n} E\left[\frac{\partial}{\partial\theta'} m(W_{in},\theta,\tau)\right] - M(\theta,\tau)\right| \to 0.$$

(c) $S(\theta_0,\tau_0) = \lim_{n\to\infty} Var\left[\overline{n}^{1/2}\overline{m}_n(\theta_0,\tau_0)\right]$ exists and is positive definite.

Convexity of $\mathcal{W}$ in Assumption 9(a) is needed to verify the stochastic equicontinuity criterion of Section 5. Assumptions 9(b)–(c) are standard conditions that ensure convergence of the Jacobian and covariance matrices of the sample moments.

THEOREM 3. *Let $\{\Lambda_n\}$ be a sequence of finite sets of $\mathbb{Z}^d$ s.t. $\overline{n}=|\Lambda_n| \to \infty$. Under Assumptions 1–5 and 7–9, $\overline{n}^{1/2}\left(\widehat{\theta}_n - \theta_0\right) \Longrightarrow N(0,V)$, where $V = ASA'$, $S = S(\theta_0,\tau_0)$, $A = \left(M'\Sigma M\right)^{-1} M'\Sigma$.*

Let $\widehat{S}$ be a consistent estimator of S. Then, the covariance matrix V is consistently estimated by $\widehat{V} = \widehat{A}\widehat{S}\widehat{A}'$, where $\widehat{A} = \left(\widehat{M}'\Sigma_n\widehat{M}\right)^{-1}\widehat{M}'\Sigma_n$ and $\widehat{M} = \frac{1}{\overline{n}}\sum_{i\in\Lambda_n}\frac{\partial}{\partial\theta'} m_{\varepsilon_n}(W_{in},\widehat{\theta},\widehat{\tau})$.

To consistently estimate $S = S(\theta_0,\tau_0)$, we construct the following spatial HAC estimator:

$$\widehat{S}\left(\widehat{\theta},\widehat{\tau}\right) = \frac{1}{\overline{n}} \sum_{i\in\Lambda_n} \sum_{j\in\Lambda_n:|i-j|\le\beta_n} K((i-j)/\beta_n) m_{\varepsilon_n}\left(W_{in},\widehat{\theta},\widehat{\tau}\right) m_{\varepsilon_n}\left(W_{jn},\widehat{\theta},\widehat{\tau}\right)', \qquad \textbf{(13)}$$

where $K((i-j)/\beta_n) = K((i_1-j_1)/\beta_n, \ldots, (i_d-j_d)/\beta_n)$ is a d-dimensional kernel, and β_n is a bandwidth parameter. For ease of exposition, we assume that the sample grows at the same rate in all d-dimensions so that we can use the same bandwidth window β_n. The consistency result below remains valid for bandwidths varying with the direction. We maintain the following assumption on the kernel function.

Assumption 10. $K:\mathbb{R}^d \to [-1,1]$, $K(0)=1$, $K(x)=0$ for $|x|>1$, $\int_{\mathbb{R}^d}|K(x)|\,dx<\infty$, $K(\cdot)$ is symmetric and continuous at 0 and at all but finite number of points.

This assumption is satisfied by many standard kernels including the rectangular, Bartlett, Parzen, and Tuckey–Hanning kernels. Among these kernels, the Bartlett and Parzen kernels produce a positive semidefinite covariance matrix estimator. For instance, one can use the following product Bartlett kernel:

$$K((i-j)/\beta_n) = \begin{cases} \Pi_{k=1}^{d}(1-|i_k-j_k|/\beta_n), & \text{if } |i_k-j_k|<\beta_n, \quad 1\le k\le d \\ 0, & \text{else.} \end{cases}$$

THEOREM 4. *Let* $\{\Lambda_n\}$ *be a sequence of finite sets of* Λ*, where* $\Lambda\subset\mathbb{R}^d$*,* $d\ge 1$ *is a discrete lattice satisfying the minimum distance assumption, such that* $\overline{n}=|\Lambda_n|\to\infty$*, and let* $\beta_n^d = O(\overline{n}^{1/4})$*. Under Assumptions 1, 2-10,* $\widehat{S}(\widehat{\theta},\widehat{\tau}) \xrightarrow{p} S(\theta_0,\tau_0)$.

This covariance estimator extends that of Conley (1999) in two directions: (i) from mixing to NED random fields, and (ii) from the parametric to semiparametric setting. In the fully parametric case, the bandwidth assumption can be relaxed to $\beta_n^d = O(\overline{n}^{1/3})$. In the semiparametric case, the bandwidth condition is more restrictive: the bandwidth parameter must increase at a slower rate, $\beta_n^d = O(\overline{n}^{1/4})$ to account for the first-step nonparametric kernel estimator, which is $o_p(\overline{n}^{-1/4})$.

5. STOCHASTIC EQUICONTINUITY CRITERION AND FUNCTIONAL CENTRAL LIMIT THEOREM (CLT)

To prove asymptotic normality of our semiparametric estimator, we will need a stochastic equicontinuity criterion for empirical processes. To our knowledge, no such results have been derived for NED random fields. Andrews (1991), Theorems 2 and 4, obtains a stochastic equicontinuity criterion and empirical CLT for time series NED processes. In this section, we extend these results to NED random fields. The stochastic equicontinuity criterion and functional CLT below can be also used to establish asymptotic properties of various semiparametric and seminonparametric estimators for heterogenous spatially dependent data, and may therefore be interesting in their own right.

Let $m(\cdot,\cdot):\mathcal{W}\times\mathcal{T}\to\mathbb{R}$ be a real function indexed by infinite-dimensional metric space $\mathcal{T}$. We assume that for each $\tau\in\mathcal{T}$, $m(w,\tau)$ is Borel measurable in w and the family $\{m(\cdot,\tau)\}$ belongs to the Sobolev space, $\mathcal{S}^{q,2}(\mathcal{W})$, equipped with

https://doi.org/10.1017/S0266466614000905 Published online by Cambridge University Press

norm (12). Following Andrews (1991), we take the pseudometric ρ on $\mathcal{T}$ to be

$$\rho_{\mathcal{T}}(\tau_1,\tau_2) = \|m(\cdot,\tau_1) - m(\cdot,\tau_2)\|_{\mathcal{W}} = \left(\int_{\mathcal{W}} |m(w,\tau_1) - m(w,\tau_2)|^2 dw\right)^{1/2}$$

and consider empirical processes $\nu_n(\cdot)$ defined as

$$\nu_n(\tau) = \frac{1}{\sqrt{|\Lambda_n|}} \sum_{i\in\Lambda_n} [m(W_{in},\tau) - Em(W_{in},\tau)].$$

DEFINITION 2. $\{\nu_n(\cdot), n \geq 1\}$ *is uniformly stochastically equicontinuous iff for every* $\varepsilon > 0$ *and* $\epsilon > 0$, *there exists* $\delta > 0$ *such that* $\limsup_{n\to\infty} P^*$ $(\sup_{\tau_1,\tau_2\in\mathcal{T}:\rho_{\mathcal{T}}(\tau_1,\tau_2)<\delta} |\nu_n(\tau_1) - \nu_n(\tau_2)| > \epsilon) < \varepsilon$, *where* P^* *denotes* P*-outer probability.*

Assumption 11.

(a) $\mathcal{W}$ is an open bounded subset of $\mathbb{R}^p$ with minimally smooth boundary.
(b) $\sup_{\tau\in\mathcal{T}} \|m(\cdot,\tau)\|_{q,2,\mathcal{W}} < \infty$ for some integer $q > (p+1)/2$.
(c) $\{W_{in}\}$ is a $\mathcal{W}$-valued random field that is L_2-NED of size $-d$ on $\{V_{in}, i \in \Lambda\}$, where $\Lambda \subset \mathbb{R}^d$ is a discrete lattice satisfying the minimum distance assumption. The mixing coefficients of $\{V_{in}\}$ satisfy $\alpha(k,l,r) \leq (k+l)^{\varsigma}\widehat{\alpha}(r)$, $\varsigma \geq 0$, and $\widehat{\alpha}(r)$ s.t. $\sum_{r=1}^{\infty} r^{d(\varsigma+1)-1}\widehat{\alpha}^{1/2}(r) < \infty$.
(d) $m(w,\tau)$ satisfies for any w, $w^\bullet \in \mathcal{W}$: $|m(w,\tau) - m(w^\bullet,\tau)| \leq C|w - w^\bullet|$ for some $C < \infty$.
(e) For any $\tau = (\tau_1,\ldots,\tau_p)' \in \mathcal{T}^p$ and $p \geq 1$, $S_p(\tau) = \lim_{n\to\infty} E(\nu_n(\tau)$ $\nu_n(\tau)')$ exists and is positive definite, where $\nu_n(\tau) = (\nu_n(\tau_1),\ldots,\nu_n(\tau_p))'$.

Assumptions 11(a)–(b) are identical to Assumptions D(i)–(ii) of Andrews (1991) who provides definition and examples of sets with minimally smooth boundaries. Assumption 11(c) restricts the dependence structure of the random field, and for $d = 1$, $\varsigma = 0$, reduces to Assumption C(iii) of Andrews (1991). Assumptions 11(a)–(c) jointly imply stochastic equicontinuity. To obtain a functional central limit theorem (FCLT), one additionally needs convergence of finite-dimensional distributions. The last two assumptions serve this purpose: Assumption 11(d) allows to establish the NED property of $\{m(W_{in})\}$ from that of $\{W_{in}\}$, and Assumption 11(e) guarantees convergence of the covariance matrices.

THEOREM 5.

A. *Under Assumptions 11(a)–(c),* $\{\nu_n(\cdot), n \geq 1\}$ *is uniformly stochastically equicontinuous and* $(\mathcal{T},\rho)$ *is totally bounded.*
B. *Under Assumptions 11(a)–(e),* $\nu_n(\cdot)$ *converges weakly to a zero-mean Gaussian process with covariance function* $S(\cdot,\cdot)$ *whose sample paths are uniformly continuous on* $(\mathcal{T},\rho)$ *a.s.*

6. NUMERICAL RESULTS

In this section, we examine the finite sample performance of our GMM and HAC covariance estimators. We consider the following specification:

$$Y_{i,j} = \theta \arctan\left(Y_{i-1,j} + Y_{i,j-1} + Y_{i+1,j} + Y_{i,j+1}\right) + \sin\left(X_{i,j}\right) + U_{i,j}, \tag{14}$$

where $(i,j) \in \mathbb{Z}^2$, $X_{i,j}$ is a scalar random variable generated according to

$$X_{i,j} = 0.2495\left(X_{i-1,j} + X_{i,j-1} + X_{i+1,j} + X_{i,j+1}\right) + \xi_{i,j}, \tag{15}$$

$\{U_{i,j}\}$ and $\{\xi_{i,j}\}$ are independent and i.i.d. $N(0,1)$. To check sensitivity of our results to the degree of persistence, we furthermore consider two different values of θ: $\theta = 0.15$ and $\theta = 0.2$.

Autoregressive processes (14)–(15) are defined implicitly as solutions to spatial difference equations. Both equations satisfy contraction mapping conditions of Jenish (2012), and consequently, there exist unique stationary solutions of these equations. The data are simulated on a rectangular grid Λ_n of $(m_1+300)\times(m_2+300)$ locations. To control for boundary effects, we discard the 300 outer boundary points along each of the axes and use the sample of size $\overline{n} = m_1 m_2$ for estimation.

By Proposition 1 of Jenish (2012), $\{Y_{i,j}\}$ is L_2-NED on $\{V_{i,j}\}$, $V_{i,j} = \sin(X_{i,j}) + U_{i,j}$, and $\{X_{i,j}\}$ is L_2-NED on $\{\xi_{i,j}\}$, and consequently, $\{(Y_{i,j}, X_{i,j})'\}$ is L_2-NED on $\{(U_{i,j}, \xi_{i,j})'\}$ with the NED coefficients decaying at a geometric rate.

To construct a trimmed GMM estimator in Model 2, we take $\mathcal{X} = \{x \in \mathbb{R} : |x| < 30\}$, $\mathcal{X}^* = \{x \in \mathbb{R} : |x| < 31\}$, $\varepsilon^* = 1$, $\varepsilon_n = \frac{1}{3}\overline{n}^{-0.51}$. We use a bias-reducing normal kernel of order 9 and bandwidth parameters $\gamma_l = \overline{n}^{-1/21}$, $l = 1,2,3$, consistent with Assumption 8.

The instruments are $Z_{i,j} \equiv (X_{i-1,j-1}, X_{i-1,j}, X_{i-1,j+1}, X_{i,j-1}, X_{i,j+1}, X_{i+1,j-1}, X_{i+1,j}, X_{i+1,j+1})$. Finally, we use the Bartlett kernel given in Section 4 and the bandwidth $\beta_n = \overline{n}^{1/8}$ to construct the HAC covariance estimator.

The results of simulations based on 1000 Monte-Carlo repetitions are reported in Table 1.

The finite-sample biases are sizeable in smaller samples, e.g., for $\overline{n} = 200$ the bias is about 6% when $\theta = 0.2$, and about 10% when $\theta = 0.15$. Nevertheless, the finite sample bias declines as the sample size increases, consistent with our asymptotic theory. Specifically, for the sample size of 800 and larger, the reported biases are in the range of 3–5%. Thus, the results suggest that a five-fold increase of the sample size leads to a two-fold decrease in biases. The results are sensitive to the autoregressive parameter θ: the larger θ, the smaller the finite sample bias. This is not surprising since θ determines the strength of the signal relative to the noise.

Generally, such finite-sample biases are not uncommon in the classical (non-spatial) semiparametric literature. There are some Monte Carlo studies for semiparametric estimators that suggest quite large finite-sample biases for the sample sizes used in our simulations. For example, Chen and Khan (2001) report biases

TABLE 1. Simulation results for semiparametric model: $Y_{ij} = \theta \text{ atan } (Y_{i-1,j} + Y_{i+1,j} + Y_{i,j-1} + Y_{i,j+1}) + \sin (X_{ij}) + U_{ij}$

					$\theta = 0.15$					
Sample Size	Mean	Bias (%)	SD	RMSE	MAD	25^{th} Pct.	50^{th} Pct.	75^{th} Pct.	C.R. (95%)	C.R. (90%)
$N = 200$	0.165	10.201	0.16	0.156	0.045	0.128	0.159	0.191	95.0	90.7
$N = 400$	0.162	7.711	0.06	0.065	0.040	0.129	0.161	0.187	96.3	92.2
$N = 600$	0.161	7.201	0.11	0.111	0.041	0.133	0.160	0.189	97.1	92.0
$N = 800$	0.158	5.238	0.05	0.051	0.036	0.131	0.158	0.184	96.7	92.0
$N = 1000$	0.157	4.639	0.05	0.055	0.036	0.131	0.160	0.184	97.6	93.2
					$\theta = 0.20$					
$N = 200$	0.213	6.440	0.10	0.098	0.041	0.179	0.208	0.239	94.8	90.7
$N = 400$	0.211	5.334	0.06	0.061	0.038	0.180	0.210	0.235	96.3	91.8
$N = 600$	0.210	5.225	0.10	0.103	0.039	0.184	0.210	0.237	96.7	91.7
$N = 800$	0.207	3.742	0.05	0.047	0.034	0.182	0.208	0.234	96.9	92.0
$N = 1000$	0.207	3.261	0.05	0.053	0.034	0.183	0.208	0.233	97.2	92.6

in the range of 7–15% for the sample sizes of 200–800 in their semiparametric estimator of the partially linear censored model. In our paper, the problem is further exacerbated by (i) presence of endogenous regressors (spatial lags) which enter the regression in a highly nonlinear way; (ii) the Lipschitz condition which shrinks the range (variability) of the dependent variable, thereby lowering the signal-to-noise ratio, and (iii) dimensionality of the index space, which is now a two-dimensional lattice.

Finally, we test the performance of the HAC covariance estimator by computing coverage rates for the 95% and 90% confidence intervals. The actual coverage rates are within the range of 81–91% for the 90% nominal interval, and within 88–95% for the 95% nominal interval. Overall, the simulations results are consistent with our asymptotic theory of the previous sections: the finite sample bias in the GMM estimator decays and the coverage rates of the HAC estimator improve as the sample size increases.

NOTES

1. For formal definition, see Dacorogna (2004). Loosely, a domain is Lipschitz, if its boundary is Lipschitz-continuous.
2. For the definition, see Jenish (2012).

REFERENCES

Andrews, D.W.K. (1991) An empirical process central limit theorem for dependent non-identically distributed random variables. *Journal of Multivariate Analysis* 38, 187–203.

Andrews, D.W.K. (1994) Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* 62, 43–72.

Andrews, D.W.K. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560–596.

Bierens, H.J. (1983) Uniform consistency of kernel estimators of a regression function under generalized conditions. *Journal of the American Statistical Association* 78, 699–707.

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: John Wiley and Sons.

Chen, S. & S. Khan (2001) Semiparametric estimation of a partially linear censored regression model. *Econometric Theory* 17, 567–590.

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

Dacorogna, B. (2004) *Introduction to the Calculus of Variations*. Imperial College Press, London.

Gao, J., Lu, Z. & D. Tjøstheim (2006) Estimation in semiparametric spatial regression. *Annals of Statistics* 34, 1395–1435.

Hörmander, L. (1976) *Linear Partial Differential Operators.* Vol. I, Springer-Verlag, Berlin Heidelberg New York.

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

Ioannides Y. & J. Zabel (2003) Neighborhood effects and housing demand. *Journal of Applied Econometrics* 18, 563–584.

Jenish, N. (2012) Nonparametric spatial regression under near-epoch dependence. *Journal of Econometrics* 167, 224–239.

Jenish, N. & I.R. Prucha (2009) Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics* 150, 86–98.
Jenish, N. & I.R. Prucha (2012) On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics* 170, 2012, 178–190.
Lee, L.-F. (2007) GMM and 2SLS for mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489–514.
Lu, Z. & O. Linton (2007) Local linear fitting under near epoch dependence. *Econometric Theory* 23, 37–70.
Pinkse, J., M. Slade & C. Brett (2002) Spatial price competition: a semiparametric approach. *Econometrica* 70, 1111–1153.
Robinson, P.M. (1988) Root-n-consistent semiparametric regression. *Econometrica* 56, 931–954.
Robinson, P.M. (2011) Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics* 165*,* 5–19.
Robinson, P.M. & S. Thawornkaiwong (2012) Statistical inference on regression with spatial dependence. *Journal of Econometrics* 167*,* 521–542.
Rosenblatt, M. (1956). A central limit theorem and strong mixing condition. *Proceedings of the National Academy of Sciences of the U.S.A.* 42, 43–47.
Rosenblatt, M. (1985) *Stationary Sequences and Random Fields*. Birkhäuser, Boston Basel Stuttgart.
Stone S.J. (1982) Optimal global rates of convergence for nonparametric regression. *Annals of Statistics* 10, 1040–1053.
Su, L. (2012) A semiparametric GMM estimation of spatial autoregressive models. *Journal of Econometrics* 167, 543–560.

APPENDIX A: Proof of Theorem 5

More detailed proofs are available in the working paper on the author's webpage. Throughout appendices, C denotes a generic constant that does not depend on n and may vary from line to line.

To prove Theorem 5, we need the following lemma.

LEMMA A.1. *If* $\{Z_{in}, i \in \Lambda, n \geq 1\}$ *is* L_2*-NED with the NED numbers* $\{\psi(s)\}$ *on* $\{V_{in}, i \in \Lambda, n \geq 1\}$*, where* $\Lambda \subset \mathbb{R}^d$ *is a discrete lattice, which satisfies the minimum distance assumption. The* α*-mixing coefficients of* $\{V_{in}\}$ *satisfy* $\alpha(k,l,r) \leq (k+l)^{\varsigma}\widehat{\alpha}(r)$ *for* $\varsigma \geq 0$*, and* $\|Z\|_{2+\delta} = \sup_{i,n}\|Z_{in}\|_{2+\delta} < \infty$ *for* $\delta > 0$*. Then, for* $i \neq j$ *and* $\varsigma_* = \varsigma\delta/(2+\delta)$

$$\left|Cov\left(Z_{in}, Z_{jn}\right)\right| \leq C\left\{\|Z\|_{2+\delta}^2[|i-j|/3]^{d\varsigma_*}\widehat{\alpha}^{\delta/(2+\delta)}([|i-j|/3]) + \|Z\|_2\psi([|i-j|/3])\right\}.$$

The proof of this lemma is given in the online version of the paper available on the author's website.

Proof of Theorem 5. *Part A.* We verify assumptions of Theorem 1 of Andrews (1991), which provides a generic stochastic equicontinuity criterion for $m(x,\tau)$ that has for each $\tau \in \mathcal{T}$ a pointwise convergent series expansion of the form $m(w,\tau) = \sum_{j=1}^{\infty} c_j(\tau)h_j(w)$ for all $w \in \mathcal{W}$ with respect to the orthonormal Fourier basis $\left\{h_j(w) = (b-a)^{-p/2} e^{2\pi \mathbf{i}\kappa(j)'(w-a\mathbf{1})/(b-a)}\right\}$, where $\mathbf{1}$ is the p-vector of ones and $\kappa(j) = (\kappa_1, ..., \kappa_p)$ is a p-vector of integers. Verification of Assumptions A(i)-(ii) and the first part of A(iii) of Andrews (1991) does not rely on the spatial dependence structure of the process, and is the same as in the proof of Theorem 4 of that paper, see pp. 199–200. There, Andrews in

https://doi.org/10.1017/S0266466614000905 Published online by Cambridge University Press

particular shows that Assumptions 11(a)–(b) imply $\sup_{\tau\in\mathcal{T}}\sum_{j=J}^{\infty}|c_j(\tau)|^2/a_j\to 0$ as $J\to\infty$ for $a_j=j^{-2q/p+\varepsilon}$, $\varepsilon\in(0,-1+(2q-1)/p)$.

Thus, it remains only to verify the second part of Assumptions A(iii). Note that $\sup_{i,j,n}|h_j(W_{in})|\leq\overline{B}=(b-a)^{-p/2}$, and the functions $h_j(w)$ satisfy the Lipschitz condition $|h_j(w)-h_j(w^\bullet)|\leq B_j|w-w^\bullet|$ with the Lipschitz constants $B_j=2C\pi(b-a)^{-1-p/2}pj^{1/p}$. Then, by Proposition 1 of Jenish and Prucha, $\{h_j(W_{in})\}$ is L_2-NED on $\{V_{in}\}$ with the NED numbers $\{2B_j\psi(s)\}$. Furthermore, by Lemma A.1 of Jenish and Prucha (2009),

$$\sup_{i\in\Lambda} card\{j\in\Lambda: r\leq|i-j|<r+1\}\leq Cr^{d-1}. \tag{A.1}$$

Using this inequality and Lemma A.1 with $Z_{in}=h_j(W_{in})$ and $\delta=\infty$ gives

$$\begin{aligned}\gamma_j&=\lim\sup_{n\to\infty} Var\left(\overline{n}^{-1/2}\sum_{i\in\Lambda_n}h_j(W_{in})\right)\leq 2\overline{B}^2\\&\quad+C\lim\sup_{n\to\infty}\overline{n}^{-1}\sum_{i\in\Lambda_n}\sum_{r=1}^{\infty}\sum_{l\in\Lambda_n:|i-l|\in[r,r+1)}\left\{[|i-l|/3]^{d\varsigma}\widehat{\alpha}([|i-l|/3])+\psi([|i-l|/3])\right\}\\&\leq 2\overline{B}^2+C\left\{\sum_{r=1}^{\infty}r^{d-1}r^{d\varsigma}\widehat{\alpha}(r)+B_j\sum_{r=1}^{\infty}r^{d-1}\psi(r)\right\}=C_1+C_2B_j<\infty,\end{aligned}$$

by Assumption 11(c) since $\widehat{\alpha}(r)\leq 1$. Hence,

$$\sum_{j=1}^{\infty}a_j\gamma_j=C_1\sum_{j=1}^{\infty}j^{-2q/p+\varepsilon}+2C_2C\pi(b-a)^{-1-p/2}p\sum_{j=1}^{\infty}j^{1/p-2q/p+\varepsilon}<\infty,$$

which verifies Assumptions A(iii) of Andrews (1991) and thus completes the proof of part A.

Finally, part B, i.e., finite dimensional convergence, follows the CLT of Jenish and Prucha (2012). ■

APPENDIX B: Proofs for Section 3

In the following, w.p.1 denotes "with probability approaching 1". Proof of Theorem 1 relies on the following lemma. Let $\tau(x,\theta)\equiv E(\varphi(Y_{in},\theta)|X_{in}=x)$, and let $\widehat{\tau}(x,\theta)$ be the kernel estimator of $\tau(x,\theta)$ for the kernel $K(\cdot)$ and bandwidth b_n.

LEMMA B.1. *Let $\{(Y_{in},X_{in}')'\}$ be a $\mathcal{Y}\times\mathbb{R}^k$-valued random field that is L_2-NED with the NED coefficients $\psi(m)=O(m^{-\eta})$ on $\{V_{in}\}$, whose α-mixing coefficients satisfy $\alpha(k,l,r)\leq(k+l)^{\varsigma}\widehat{\alpha}(r)$ for $\varsigma\geq 0$ and $\widehat{\alpha}(r)$ s.t. $\sum_{r=1}^{\infty}r^{d(\varsigma_*+1)-1}\widehat{\alpha}^{\delta/(2+\delta)}(r)<\infty$, $\varsigma_*=\varsigma\delta/(2+\delta)$, $\delta>0$. Suppose further that*

(a) *The function $\varphi:\mathcal{Y}\times\Theta^*\to\mathbb{R}$ satisfies for any y, $y^\bullet\in\mathcal{Y}$: $|\varphi(y,\theta)-\varphi(y^\bullet,\theta)|\leq L(\theta)|y-y^\bullet|$ with $\sup_{\theta\in\Theta^*}L(\theta)<\infty$, and $\sup_{n,i}\sup_{\theta\in\Theta^*}\|\varphi(Y_{in},\theta)\|_{2+\delta}<\infty$ for the above δ.*

(b) *For some integer $\omega\geq 1$, $\tau(x,\theta)f_{in}(x)\in\mathcal{C}^{\omega}(\mathbb{R}^k)$ w.r.t. x, $\sup_{n,\mathcal{X}^*\times\Theta^*}\left|\overline{n}^{-1}\sum_{i\in\Lambda_n}D^{\mu_x}[\tau(x,\theta)f_{in}(x)]\right|<\infty$ for $|\mu_x|\leq\omega$, and Assumption 3 holds for this ω.*

(c) *Assumption 5(a) is satisfied, and Assumption 5(b) holds with* $q_l = \omega - 1$ *and* $l = 1$. *Then, for* $|\mu_x| \le \omega - 1$

$$\sup_{\mathcal{X}^* \times \Theta^*} \left| D^{\mu_x} \widehat{\tau}(x,\theta) - D^{\mu_x} \tau(x,\theta) \right| = O_p\left(\overline{n}^{-\eta/(2\eta+d)} b_n^{-k-|\mu_x|-d/(2\eta+d)} \right) + O_p\left(b_n^{\omega-|\mu_x|} \right),$$

provided that the right-hand side of this equality is $o_p(1)$.

The proof of the lemma is similar to that of Theorem 1(b) in Andrews (1995), who give similar results for NED time-series processes, and is therefore omitted.

Proof of Theorem 1. Since $|Y_{in}| < \infty$ and $|Z_{in}| < \infty$, we will use Lemma B.1 with $\delta = \infty$. Then, $\delta/(2+\delta) \to 1$ and $\varsigma_* \to \varsigma$ as $\delta \to \infty$, and hence the mixing coefficients must satisfy $\sum_{r=1}^{\infty} r^{d(\varsigma+1)-1} \widehat{\alpha}(r) < \infty$, which holds by Assumption 2(c). Part A follows immediately from Lemma B.1. To establish part B, observe that $D^{\mu_\theta} \widehat{\tau}_3(x,\theta)$ is the kernel estimator of $D^{\mu_\theta} h(Y_{2in},\theta)$. By Assumption 4(c), $\sup_{\mathcal{Y}_2 \times \Theta^*} \left| D^{\mu_\theta} h(Y_{2in},\theta) \right| < \infty$, which is a domination condition that allows interchanging differentiation and integration in $D^{\mu_\theta} \tau_{30}(x,\theta) = D^{\mu_\theta} E\left(h(Y_{2in};\theta)|X_{in} = x\right) = E\left(D^{\mu_\theta} h(Y_{2in};\theta)|X_{in} = x\right)$. Assumption (a) of Lemma B.1 holds by Assumption 4(c). Part B now follows Lemma B.1. ∎

The proof of Theorem 2 makes use of the following lemmata.

LEMMA B.2. *Let* $\zeta_\varepsilon(x) = \frac{1}{\varepsilon^k} \int_{\mathcal{X}_\varepsilon} \phi\left(\frac{x-z}{\varepsilon}\right) dz$, *where* $\mathcal{X}_\varepsilon$ *is the* ε*-neighborhood of* $\mathcal{X} \subset \mathbb{R}^k$, $0 < \varepsilon < \varepsilon^*/2$, ε^* *is as in Assumption 1 and* $\phi(x)$ *is defined as:*

$$\phi(x) = \begin{cases} c_k \exp\left(1/\left(|x|^2 - 1\right)\right), & |x| < 1, \\ 0, & |x| \ge 1, \end{cases}$$

where c_k *is a normalizing constant such that* $\int \phi(x) dx = 1$. *Then,*

(a) *supp*$[\zeta_\varepsilon(x)] \subset \mathcal{X}_{2\varepsilon}$, $\zeta_\varepsilon(x) = 0$ *for* $x \in \mathbb{R}^k \backslash \mathcal{X}_{2\varepsilon}$.

(b) $D^\mu \zeta_\varepsilon(x)$ *exists and continuous on* $\mathbb{R}^k$ *for all* $|\mu| \ge 0$.

(c) $\int_{\mathbb{R}^k} \left| \zeta_\varepsilon(x) - \mathbf{1}_{\mathcal{X}}(x) \right| dx \to 0$ *as* $\varepsilon \to 0$.

LEMMA B.3.

(a) *Under Assumptions 1, 2, 4(c), for* $0 < \varepsilon_n < \varepsilon^*/2$, $\{m_{\varepsilon_n}(W_{in},\theta,\tau)\}$ *and* $\left\{\frac{\partial}{\partial\theta'} m_{\varepsilon_n}(W_{in},\theta,\tau)\right\}$ *are* L_2*-NED on* $\{V_{in}\}$ *of size* $-\eta$.

(b) *Under Assumptions 1, 7(a)-(b),* $Em(W_{in},\theta,\tau) = \int_{\mathcal{X}} [\Upsilon(\theta,\tau(x)) + \Upsilon_1(\theta,x)] f_{in}(x) dx$, *and*

$$Em_{\varepsilon_n}(W_{in},\theta,\tau) = \int_{\mathcal{X}_{2\varepsilon_n}} \zeta_{\varepsilon_n}(x) \left[\Upsilon(\theta,\tau(x)) + \Upsilon_1(\theta,x)\right] f_{in}(x) dx + \int_{\mathcal{X}_{2\varepsilon_n} \backslash \mathcal{X}} \zeta_{\varepsilon_n}(x) \Upsilon_2(x) f_{in}(x) dx,$$

where $\Upsilon_2(x) = \tau_{50}(x) - \tau_{40}(x,\theta_0) + \tau_{10}(x)(\tau_{30}(x,\theta_0) - \tau_{20}(x))$, $\tau_{50}(x)$, $\tau_{40}(x,\theta_0)$, $\Upsilon(\theta,\tau(x))$ *and* $\Upsilon_1(\theta,x)$ *are defined in Assumption 7(b)–(c).*

https://doi.org/10.1017/S0266466614000905 Published online by Cambridge University Press

Proof of Lemma B.2. $\phi(x)$ is infinitely differentiable on $\mathbb{R}^k$ and $\text{supp}[\phi(x)] = \{x : |x| \le 1\}$. Thus, $\phi(x)$ is a *test function* in the terminology of Hörmander (1976), and the lemma follows from Theorems 1.2.1 and 1.6.3 of this monograph with $u(x) = 1_{\mathcal{X}_\varepsilon}(x)$. ■

Proof of Lemma B.3. *Part (a).* By Lemma B.2(c), Assumptions 1–2 and 4(c), $m_\varepsilon(w,\theta,\tau)$ is continuously differentiable in w, and hence, satisfies a Lipschitz condition in w with a bounded Lipschitz coefficient. Then, by Proposition 2 of Jenish and Prucha (2012), $\{m_\varepsilon(W_{in},\theta,\tau)\}$ is L_2-NED of the same size as $\{W_{in}\}$. Similarly, $\frac{\partial}{\partial\theta} m_\varepsilon(W_{in},\theta,\tau)$ is also L_2-NED of the same size as $\{W_{in}\}$.

Part (b). Note that

$$Em(W_{in},\theta,\tau) = \int_{\mathcal{X}} \mathbf{1}_{\mathcal{X}}(x)[\tau_{10}(x)-\tau_1(x)][\tau_{20}(x)-\tau_2(x)-\tau_{30}(x,\theta)+\tau_3(x,\theta)]\, f_{in}(x)dx$$
$$+\int_{\mathcal{X}} \mathbf{1}_{\mathcal{X}}(x)\, E[(Z_{in}-\tau_{10}(X_{in}))(Y_{1in}-\tau_2(X_{in})-h(Y_{2in},\theta)$$
$$+\tau_3(X_{in},\theta))\,|X_{in}=x]\, f_{in}(x)dx.$$

The second term in the last expression can be further written as

$$\begin{aligned}Q_2 = \int_{\mathcal{X}} & E\left[(Z_{in}-\tau_{10}(X_{in}))(Y_{1in}-\tau_{20}(X_{in})\right.\\
&\left. -\ h(Y_{2in},\theta_0)+\tau_{30}(X_{in},\theta_0))\,|X_{in}=x\right] f_{in}(x)dx\\
&+\int_{\mathcal{X}} E\left[(Z_{in}-\tau_{10}(X_{in}))\,|X_{in}=x\right](\tau_{20}(x)-\tau_2(x)+\tau_3(x,\theta)\\
&-\ \tau_{30}(x,\theta_0))\, f_{in}(x)dx\\
&+\int_{\mathcal{X}} E\left[(Z_{in}-\tau_{10}(X_{in}))(h(Y_{2in},\theta_0)-h(Y_{2in},\theta))\,|X_{in}=x\right] f_{in}(x)dx\\
=&\ 0+0+\int_{\mathcal{X}}\left\{\tau_{40}(x,\theta_0)-\tau_{40}(x,\theta)+\tau_{10}(x)\left[\tau_{30}(x,\theta)-\tau_{30}(x,\theta_0)\right]\right\} f_{in}(x)dx,\end{aligned}$$

since the first term on the r.h.s. is the moment condition and is zero; the second term is zero by definition of τ_{10}. Similar arguments and Lemma B.2(b) yield the expression for $Em_\varepsilon(W_{in},\theta,\tau)$. ■

Proof of Theorem 2. To prove consistency of $\widehat{\theta}_n$, it suffices to verify Assumption C of Theorem A-1 of Andrews (1994).

Step 1. Verification of Assumption C(a). By Lemma B.3(a), $\{m_{\varepsilon_n}(W_{in},\theta,\tau)\}$ is L_2-NED of size $-\eta$, $\eta > d$. By Theorem 1 of Jenish and Prucha (2012), $\{m_{\varepsilon_n}(W_{in},\theta,\tau)\}$ satisfies an law of large numbers (LLN) for each (θ,τ):

$$\left|\overline{n}^{-1}\sum_{i\in\Lambda_n} m_{\varepsilon_n}(W_{in},\theta,\tau) - Em_{\varepsilon_n}(W_{in},\theta,\tau)\right| \xrightarrow{p} 0.$$

Next, we show that this convergence holds uniformly over $\Theta\times\mathcal{T}$. To this end, by the uniform law of large numbers (ULLN) of Jenish and Prucha (2009), it suffices to show that (i) $\{m_{\varepsilon_n}(W_{in},\theta,\tau)\}$ is stoch. equicontinuous on $\Theta\times\mathcal{T}$ w.r.t. pseudometric ρ_1, and (ii) $(\Theta\times\mathcal{T},\rho_1)$ is totally bounded, where $\rho_1((\theta,\tau),(\theta^\bullet,\tau^\bullet)) = |\theta-\theta^\bullet| + \sup_{\mathcal{X}\times\Theta}|\tau(x,\theta)-\tau^\bullet(x,\theta)|$.

First, note that Assumptions 1, 2, 4(c) and 7(e), for any $(\theta,\tau),(\theta^\bullet,\tau^\bullet)\in\Theta\times\mathcal{T}$:

$$\left|m_{\varepsilon_n}(W_{in},\theta,\tau)-m_{\varepsilon_n}(W_{in},\theta^\bullet,\tau^\bullet)\right| \le C\rho_1\left((\theta,\tau),(\theta^\bullet,\tau^\bullet)\right) \qquad \textbf{(B.1)}$$

https://doi.org/10.1017/S0266466614000905 Published online by Cambridge University Press

for some $C < \infty$, which proves stoch. ρ_1-equicontinuity of $\{m_{\varepsilon_n}(W_{in}, \theta, \tau)\}$ on $\Theta \times \mathcal{T}$. Moreover, $\mathcal{T}$ is uniformly ρ_u-equicontinuous on $\mathcal{X} \times \Theta$, where $\rho_u(\tau, \tau^\bullet) = \sup_{\mathcal{X}\times\Theta} |\tau(t) - \tau^\bullet(t)|$, and $\mathcal{T}$ is equibounded on $\mathcal{X} \times \Theta$. Then, by the Arzela-Ascoli theorem, $(\mathcal{T}, \rho_u)$ is totally bounded, and hence, $(\Theta \times \mathcal{T}, \rho_1)$ is also totally bounded. Thus, by the ULLN of Jenish and Prucha (2009)

$$\sup_{\Theta\times\mathcal{T}} \left| \overline{n}^{-1} \sum_{i\in\Lambda_n} m_{\varepsilon_n}(W_{in}, \theta, \tau) - E m_{\varepsilon_n}(W_{in}, \theta, \tau) \right| \xrightarrow{p} 0. \tag{B.2}$$

Next, by Lemmas B.2 and B.3(b), Assumptions 1, 3(b) and 7(b),(e)

$$\sup_{\Theta\times\mathcal{T}} \left| \overline{n}^{-1} \sum_{i\in\Lambda_n} E\left[m_{\varepsilon_n}(W_{in}, \theta, \tau) - m(W_{in}, \theta, \tau)\right] \right|$$
$$\leq C_1 \int_{\mathbb{R}^k} \left|\zeta_{\varepsilon_n}(x) - \mathbf{1}_{\mathcal{X}}(x)\right| dx + C_2 Leb\left(\mathcal{X}_{2\varepsilon_n} \backslash \mathcal{X}\right) \to 0,$$

where $Leb(A)$ denotes the Lebesgue measure of set A. Hence, $\{\overline{n}^{-1} \sum_{i\in\Lambda_n} Em(W_{in}, \theta, \tau)\})$ is uniformly ρ_1-equicontinuous on $\Theta \times \mathcal{T}$. By Lemma B.3(b) and Assumptions 1, 7(b)–(c),

$$\sup_{\Theta\times\mathcal{T}} \left| \overline{n}^{-1} \sum_{i\in\Lambda_n} Em(W_{in}, \theta, \tau) - m(\theta, \tau) \right| \leq C \sup_{x\in\mathcal{X}} \left| \overline{n}^{-1} \sum_{i\in\Lambda_n} f_{in}(x) - \overline{f}(x) \right| \to 0, \tag{B.3}$$

for the $m(\theta, \tau)$ is defined in Assumption 7(d). Now, collecting (B.2)–(B.3) gives

$$\sup_{\Theta\times\mathcal{T}} \left| \overline{n}^{-1} \sum_{i\in\Lambda_n} m_{\varepsilon_n}(W_{in}, \theta, \tau) - m(\theta, \tau) \right| \xrightarrow{p} 0. \tag{B.4}$$

It also follows that $m(\theta, \tau)$ is uniformly ρ_1-continuous on $\Theta \times \mathcal{T}$, and by total boundedness of $(\Theta \times \mathcal{T}, \rho_1)$, $\sup_{\Theta\times\mathcal{T}} |m(\theta, \tau)| < \infty$.

Step 2. Verification of Assumption C(b). We need to show $\sup_{\theta\in\Theta} |m(\theta, \widehat{\tau}) - m(\theta, \tau_0)| \xrightarrow{p} 0$. Note that $m(\theta, \tau_0) = \int_{\mathcal{X}} \Upsilon_1(\theta, x)\overline{f}(x)dx$ and hence $m(\theta, \widehat{\tau}) - m(\theta, \tau_0) = \int_{\mathcal{X}} \Upsilon(\theta, \widehat{\tau})\overline{f}(x)dx$. By Theorem 1,

$$\sup_{\Theta} |m(\theta, \widehat{\tau}) - m(\theta, \tau_0)| \leq C \sup_{\mathcal{X}^*} |\tau_{10}(x) - \widehat{\tau}_1(x)|$$
$$\times \sup_{\mathcal{X}^*\times\Theta^*} \left[|\tau_{20}(x) - \widehat{\tau}_2(x)| + |\tau_{30}(x, \theta) - \widehat{\tau}_3(x, \theta)| \right] \xrightarrow{p} 0.$$

We next show that $P(\widehat{\tau} \in \mathcal{T}) \to 1$. Observe that for $l = 1, 2$

$$\|\widehat{\tau}_l - \tau_{l0}\|_{q_l, \mathcal{X}^*} = \sum_{|\mu|\leq q_l} \left(\int_{\mathcal{X}^*} \left| D^\mu \widehat{\tau}_l(x) - D^\mu \tau_{l0}(x) \right|^2 dx \right)^{1/2}$$
$$\leq C \sum_{|\mu|\leq q_l} \sup_{\mathcal{X}^*} \left| D^\mu \widehat{\tau}_l(x) - D^\mu \tau_{l0}(x) \right|^2 \xrightarrow{p} 0,$$

https://doi.org/10.1017/S0266466614000905 Published online by Cambridge University Press

by part A of Theorem 1. It then follows $\|\widehat{\tau}_l\|_{q_l,\mathcal{X}^*} \le \|\widehat{\tau}_l - \tau_{l0}\|_{q_l,\mathcal{X}^*} + \|\tau_{l0}\|_{q_l,\mathcal{X}^*} \le o_p(1) + B$, $l = 1,2$. Similarly, part B of Theorem 1 implies $\|\widehat{\tau}_3\|_{q_2,\mathcal{X}^*\times\Theta^*} \le o_p(1) + B$.

Hence, $P\left(\|\widehat{\tau}_1\|_{q_1,\mathcal{X}^*} \le B, \|\widehat{\tau}_2\|_{q_2,\mathcal{X}^*} \le B, \|\widehat{\tau}_3\|_{q_3,\mathcal{X}^*\times\Theta^*} \le B\right) \to 1$, which verifies $P(\widehat{\tau} \in \mathcal{T}) \to 1$. Finally, the last condition of Assumption C(b) of Andrews (1994) is satisfied with $\widehat{\gamma} = \Sigma_n^{-1}$ in light of Assumption 7(a).

Step 3. Verification of Assumptions C(c) and C(d). Assumption C(b) is satisfied with $d(m,\gamma_0) = m'\Sigma m$ and $\gamma_0 = \Sigma$ since $\sup_{\Theta\times\mathcal{T}} |m(\theta,\tau)| < \infty$. Note that by uniform continuity of $m(\theta,\tau)$ in (θ,τ), $d(m(\theta,\tau_0),\gamma_0)$ is continuous on Θ. Assumption C(d) holds since (i) Θ is compact, (ii) $d(m(\theta,\tau_0),\gamma_0)$ is continuous on Θ, and (iii) $m(\theta,\tau_0)'\Sigma^{-1}m(\theta,\tau_0)$ is uniquely minimized at θ_0. By Theorem A-1 of Andrews (1994), $\widehat{\theta}_n \xrightarrow{p} \theta_0$. ■

APPENDIX C: Proofs for Section 4

Proof of Theorem 3.

Step 1. We first show that

$$\sup_{\Theta\times\mathcal{T}} \left| \overline{n}^{-1} \sum_{i\in\Lambda_n} \frac{\partial}{\partial\theta'} m_{\varepsilon_n}(W_{in},\theta,\tau) - M(\theta,\tau) \right| \xrightarrow{p} 0, \tag{C.1}$$

for the $M(\theta,\tau)$ defined in Assumption 9(b). By Lemma B.3(a) and Theorem 1 of Jenish and Prucha (2012), $\left\{m_{\varepsilon_n}(W_{in},\theta,\tau)\right\}$ satisfies an LLN for each (θ,τ). Using arguments analogous to those in Step 1 of the proof of Theorem 2, one can strengthen this LLN to ULNN:

$$\sup_{(\theta,\tau)\in\Theta\times\mathcal{T}} \left| \overline{n}^{-1} \sum_{i\in\Lambda_n} \frac{\partial}{\partial\theta'} m_{\varepsilon_n}(W_{in},\theta,\tau) - E\frac{\partial}{\partial\theta'} m(W_{in},\theta,\tau) \right| \xrightarrow{p} 0. \tag{C.2}$$

Step 2. The estimator $\widehat{\theta}$ satisfies the following first order conditions:

$$\frac{\partial}{\partial\theta'} \overline{m}_n\left(\widehat{\theta},\widehat{\tau}\right) \Sigma_n \overline{m}_n\left(\widehat{\theta},\widehat{\tau}\right) = o_p(1). \tag{C.3}$$

Note that by assumption $m(W_{in},\theta,\tau)$ is continuously differentiable in the interior of Θ, and θ_0 is in the interior of Θ. Taking the mean value expansion of $\overline{m}_n\left(\widehat{\theta},\widehat{\tau}\right)$ about θ_0 yields

$$\overline{m}_n\left(\widehat{\theta},\widehat{\tau}\right) = \overline{m}_n\left(\theta_0,\widehat{\tau}\right) + \frac{\partial}{\partial\theta'}\overline{m}_n\left(\widetilde{\theta},\widehat{\tau}\right)\left(\widehat{\theta}-\theta_0\right), \tag{C.4}$$

where $\widetilde{\theta} \in \Theta$ is between $\widehat{\theta}$ and θ_0. By Theorem 2, $\widehat{\theta} \xrightarrow{p} \theta_0$ and hence $\widetilde{\theta} \xrightarrow{p} \theta_0$.

Plugging (C.4) into (C.3) gives

$$\overline{n}^{1/2}(\widehat{\theta}_n - \theta_0) = -\left[\frac{\partial}{\partial\theta}\overline{m}_n\left(\widehat{\theta},\widehat{\tau}\right)\Sigma_n \frac{\partial}{\partial\theta'}\overline{m}_n\left(\widetilde{\theta},\widehat{\tau}\right)\right]^{-1} \times \frac{\partial}{\partial\theta}\overline{m}_n\left(\widehat{\theta},\widehat{\tau}\right)\Sigma_n \overline{n}^{1/2}\overline{m}_n\left(\theta_0,\widehat{\tau}\right) + o_p(1).$$

https://doi.org/10.1017/S0266466614000905 Published online by Cambridge University Press

By (C.1), $\frac{\partial}{\partial\theta'}\overline{m}_n\left(\widehat{\theta},\widehat{\tau}\right) \xrightarrow{p} M = M(\theta_0,\tau_0)$ and $\frac{\partial}{\partial\theta'}\overline{m}_n\left(\widetilde{\theta},\widehat{\tau}\right) \xrightarrow{p} M = M(\theta_0,\tau_0)$. By Assumption 7(a), $\frac{\partial}{\partial\theta}\overline{m}_n\left(\widehat{\theta},\widehat{\tau}\right)\Sigma_n\frac{\partial}{\partial\theta'}\overline{m}_n\left(\widetilde{\theta},\widehat{\tau}\right) = M'\Sigma M + o_p(1)$ and $\frac{\partial}{\partial\theta}\overline{m}_n\left(\widehat{\theta},\widehat{\tau}\right)\Sigma_n = M'\Sigma + o_p(1)$. Since $M'\Sigma M$ is nonsingular, and $\overline{n}^{1/2}\overline{m}_n\left(\theta_0,\widehat{\tau}\right) = o_p(1)$ as shown in Step 3, we have

$$\overline{n}^{1/2}\left(\widehat{\theta}_n-\theta_0\right) = A\overline{n}^{1/2}\overline{m}_n\left(\theta_0,\widehat{\tau}\right)+o_p(1) \tag{C.5}$$

with $A = -\left(M'\Sigma M\right)^{-1}M'\Sigma$. Let

$$\widetilde{\nu}_n(\tau) = \overline{n}^{-1/2}\sum_{i\in\Lambda_n}\left\{m_{\varepsilon_n}(W_{in},\theta_0,\tau)-Em_{\varepsilon_n}(W_{in},\theta_0,\tau)\right\},\ \nu_n(\tau)$$
$$= \overline{n}^{-1/2}\sum_{i\in\Lambda_n}\left\{m(W_{in},\theta_0,\tau)-Em(W_{in},\theta_0,\tau)\right\}.$$

Then, $\overline{n}^{1/2}\left(\widehat{\theta}_n-\theta_0\right) = A\widetilde{\nu}_n(\widehat{\tau})+A\overline{n}^{1/2}E\overline{m}_n\left(\theta_0,\widehat{\tau}\right)+o_p(1)$.

Step 3. We now show that $\overline{n}^{1/2}E\overline{m}_n(\theta_0,\widehat{\tau}) = o_p(1)$. Note that

$$\left|\overline{n}^{1/2}E\overline{m}_n\left(\theta_0,\widehat{\tau}\right)\right| \le \left|\overline{n}^{-1/2}\sum_{i\in\Lambda_n}Em_{\varepsilon_n}\left(W_{in},\theta_0,\widehat{\tau}\right)-Em\left(W_{in},\theta_0,\widehat{\tau}\right)\right|$$
$$+\left|\overline{n}^{-1/2}\sum_{i\in\Lambda_n}Em\left(W_{in},\theta_0,\widehat{\tau}\right)\right|.$$

Using Lemma B.3(b), $\left|\overline{n}^{-1/2}\sum_{i\in\Lambda_n}E\left[m_{\varepsilon_n}\left(W_{in},\theta_0,\widehat{\tau}\right)-m\left(W_{in},\theta_0,\widehat{\tau}\right)\right]\right| = o_p(1)$. By the Cauchy–Schwartz inequality, $\left|\overline{n}^{-1/2}\sum_{i\in\Lambda_n}Em(W_{in},\theta_0,\widehat{\tau})\right|$ is less or equal to

$$\overline{n}^{-1/2}\left(\sum_{i\in\Lambda_n}\int_{\mathcal{X}}|\tau_{10}(x)-\widehat{\tau}_1(x)|^2\,dP_{in}(x)\right)^{1/2}$$
$$\times\left(\sum_{i\in\Lambda_n}\int_{\mathcal{X}}|\tau_{20}(x)-\widehat{\tau}_2(x)-\tau_{30}(x,\theta_0)+\widehat{\tau}_3(x,\theta_0)|^2\,dP_{in}(x)\right)^{1/2}$$
$$\le \overline{n}^{1/2}\sup_{\mathcal{X}^*}|\tau_{10}(x)-\widehat{\tau}_1(x)|\sup_{\mathcal{X}^*}|\widehat{\tau}_2(x)-\tau_{20}(x)+\widehat{\tau}_3(x,\theta_0)-\tau_{30}(x,\theta_0)| \xrightarrow{p} 0,$$

since $\sup_{\mathcal{X}^*}|\widehat{\tau}(x,\theta_0)-\tau_0(x,\theta_0)| = o_p(\overline{n}^{-1/4})$ by Assumption 8. Thus, $\overline{n}^{1/2}(\widehat{\theta}_n-\theta_0) = A\widetilde{\nu}_n(\widehat{\tau})+o_p(1)$.

Step 4. Using definitions of $\nu_n(\cdot)$ and $\widetilde{\nu}_n(\cdot)$ in Step 2 above, rewrite the last equality as

$$\overline{n}^{1/2}(\widehat{\theta}_n-\theta_0) = A\widetilde{\nu}_n(\tau_0)+A\left[\widetilde{\nu}_n(\widehat{\tau})-\nu_n(\widehat{\tau})\right]+A\left[\nu_n(\widehat{\tau})-\nu_n(\tau_0)\right]$$
$$+A\left[\nu_n(\tau_0)-\widetilde{\nu}_n(\tau_0)\right]+o_p(1).$$

We need to show that all terms in the squared brackets are $o_p(1)$. Note that for all $\tau \in \mathcal{T}$, $E\left|\widetilde{\nu}_n(\tau) - \nu_n(\tau)\right|$ is less or equal to

$$
\begin{aligned}
&2\overline{n}^{-1/2} \sum_{i \in \Lambda_n} E\left|\zeta_{\varepsilon_n}(X_{in}) - \mathbf{1}_{\mathcal{X}}(X_{in})\right| \\
&\quad \times \left|(Z_{in} - \tau_1(X_{in}))(Y_{1in} - \tau_2(X_{in}) - h(Y_{2in}, \theta_0) + \tau_3(X_{in}, \theta_0))\right| \\
&\quad \le 2 \sup_{\mathcal{X}^* \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Z}} |z - \tau_1(x)|\,|y_1 - \tau_2(x) - h(y_2, \theta_0) + \tau_3(x, \theta_0)| \times \\
&\qquad \times \left\{\sup_{\mathbb{R}^k} \overline{n}^{-1} \sum_{i \in \Lambda_n} f_{in}(x)\right\} \overline{n}^{1/2} \int_{\mathbb{R}^k} \left|\zeta_{\varepsilon_n}(x) - \mathbf{1}_{\mathcal{X}}(x)\right| dx \\
&\quad \le C\overline{n}^{1/2} Leb(\mathcal{X}_{2\varepsilon_n} \backslash \mathcal{X}) = o(1),
\end{aligned}
$$

since $Leb(\mathcal{X}_{2\varepsilon_n} \backslash \mathcal{X}) = o\left(\overline{n}^{-1/2}\right)$. Thus, for all $\tau \in \mathcal{T}$, $\widetilde{\nu}_n(\tau) - \nu_n(\tau) = o_p(1)$, and in particular, $\nu_n(\tau_0) - \widetilde{\nu}_n(\tau_0) = o_p(1)$. As established earlier, $P(\widehat{\tau} \in \mathcal{T}) \to 1$. By similar arguments, $\widetilde{\nu}_n(\widehat{\tau}) - \nu_n(\widehat{\tau}) = o_p(1)$.

We next prove that $\nu_n(\widehat{\tau}) - \nu_n(\tau_0) = o_p(1)$. To this end, it suffices to show: (i) $P(\widehat{\tau} \in \mathcal{T}) \to 1$ (already proven), (ii) $\rho_{\mathcal{T}}(\widehat{\tau}, \tau_0) \xrightarrow{p} 0$, and (iii) $\{\nu_n(\cdot), n \ge 1\}$ is stoch. equicontinuous at τ_0. Given (i)–(iii) and using standard arguments, we will then have

$$
\begin{aligned}
&\lim \sup_{n \to \infty} P\left(|\nu_n(\widehat{\tau}) - \nu_n(\tau_0)| > \epsilon\right) \\
&\quad \le \lim \sup_{n \to \infty} P^*\left(\sup_{\tau \in \mathcal{T}: \rho_{\mathcal{T}}(\tau, \tau_0) < \delta} |\nu_n(\tau) - \nu_n(\tau_0)| > \epsilon\right) < \varepsilon,
\end{aligned}
$$

where the last inequality is equivalent to (iii). To show (ii), observe that by Assumption 1, $m(\cdot, \theta_0, \tau)$ belongs to the Sobolev space, $\mathcal{S}^{q,2}(\mathcal{W})$, with the norm order $q > (p_w + 1)/2$ and by Theorem 1, $\rho_{\mathcal{T}}(\widehat{\tau}, \tau_0) \le Leb^{1/2}(\mathcal{W}) \sup_{w \in \mathcal{W}} |m(w, \theta_0, \widehat{\tau}) - m(w, \theta_0, \tau_0)| \to 0$. To establish (iii), we verify assumptions of part A of Theorem 5. Assumption 11(a) and 11(c) hold by Assumption 2 and 9. Finally, by Assumption 1,

$$
\sup_{\tau \in \mathcal{T}} \|m(\cdot, \theta_0, \tau)\|_{q,2,\mathcal{W}} = \sup_{\tau \in \mathcal{T}} \sum_{|\mu| \le q} \left\|D^{\mu_w} m(w, \theta_0, \tau)\right\|_{L_2(\mathcal{W})} < \infty. \qquad \textbf{(C.6)}
$$

Thus, $\{\nu_n(\cdot), n \ge 1\}$ is stoch. equicontinuous, and hence $\overline{n}^{1/2}(\widehat{\theta}_n - \theta_0) = A\widetilde{\nu}_n(\tau_0) + o_p(1)$.

Step 5. Finally, we show that $\widetilde{\nu}_n(\tau_0) \Rightarrow N(0, S)$ by verifying assumptions of Theorem 5. Assumptions 11(a)–(d) have been verified in Step 5 of this proof and also in Lemma B.3(a). Assumptions 11(e) for $p = 1$ holds by Assumption 9(c). Thus, $\nu_n(\tau_0) \Rightarrow N(0, S)$ and hence $\overline{n}^{1/2}\left(\widehat{\theta}_n - \theta_0\right) \Longrightarrow N(0, V)$. ■

Proof of Theorem 4.

Let $\vartheta = (\theta, \tau)$, $m_{in}(\vartheta) = m_{\varepsilon_n}(W_{in}, \theta, \tau)$, and $S_n(\vartheta_0) = \overline{n}^{-1} E \sum_{i \in \Lambda_n} \sum_{j \in \Lambda_n} m_{in}(\vartheta_0) m_{in}'(\vartheta_0)$. Note that $S(\vartheta_0) = \lim_{n \to \infty} S_n(\vartheta_0)$. To prove the theorem, it suffices to show that

$$
\left|\widehat{S}_n(\widehat{\vartheta}) - \widehat{S}_n(\vartheta_0)\right| \xrightarrow{p} 0, \ \left|\widehat{S}_n(\vartheta_0) - E\widehat{S}_n(\vartheta_0)\right| \xrightarrow{p} 0, \text{ and } \left|E\widehat{S}_n(\vartheta_0) - S_n(\vartheta_0)\right| \to 0.
$$

Step 1. Proof of $\left|\widehat{S}_n(\vartheta_0) - E\widehat{S}_n(\vartheta_0)\right| \xrightarrow{p} 0$. For any $s > 0$, let $m_{in}^s = E(m_{in}|\mathfrak{F}_{in}(s))$, $\xi_{in}^s = m_{in} - m_{in}^s$ and $K_{ijn} = K((i-j)/\beta_n)$, where for simplicity we suppress dependence of $m_{in}(\vartheta_0)$ on ϑ_0. Using the decomposition $m_{in}m_{jn} = m_{in}^s m_{jn}^s + m_{in}^s \xi_{jn}^s + \xi_{in}^s m_{jn}^s + \xi_{in}^s \xi_{jn}^s$, write $\widehat{S}_n(\vartheta_0) - E\widehat{S}_n(\vartheta_0) = H_{1n} + H_{2n} + H_{3n} + H_{4n}$, where

$$H_{1n} = \overline{n}^{-1} \sum_{i,j\in\Lambda_n, |i-j|\le\beta_n} K_{ijn}\left(m_{in}^s m_{jn}^s - E m_{in}^s m_{jn}^s\right),$$

$$H_{2n} = \overline{n}^{-1} \sum_{i,j\in\Lambda_n, |i-j|\le\beta_n} K_{ijn}\left(m_{in}^s \xi_{jn}^s - E m_{in}^s \xi_{jn}^s\right),$$

$$H_{3n} = \overline{n}^{-1} \sum_{i,j\in\Lambda_n, |i-j|\le\beta_n} K_{ijn}\left(\xi_{in}^s m_{jn}^s - E \xi_{in}^s m_{jn}^s\right),$$

$$H_{4n} = \overline{n}^{-1} \sum_{i,j\in\Lambda_n, |i-j|\le\beta_n} K_{ijn}\left(\xi_{in}^s \xi_{jn}^s - E \xi_{in}^s \xi_{jn}^s\right).$$

Next, we show that H_{kn}, $k = 1, \ldots, 4$, converge to zero.

Set $s = \beta_n$, and let $\chi_{in}^s = \sum_{j\in\Lambda_n:|i-j|\le\beta_n} K_{ijn}\left(m_{in}^s m_{jn}^s - E m_{in}^s m_{jn}^s\right)$. Then,

$$Var(H_{1n}) = \overline{n}^{-2} \sum_{i\in\Lambda_n} \sum_{j\in\Lambda_n:|i-j|\le 3\beta_n} E\chi_{in}^s \chi_{jn}^s + \overline{n}^{-2} \sum_{i\in\Lambda_n} \sum_{j\in\Lambda_n:|i-j|> 3\beta_n} E\chi_{in}^s \chi_{jn}^s. \tag{C.7}$$

By Lemma A.1 of Jenish and Prucha (2009),

$$\left\|\chi_{in}^s\right\|_2 \le 2 \sum_{j\in\Lambda_n:|i-j|\le\beta_n} \left|K_{ijn}\right| \left\|m_{in}^s m_{jn}^s\right\|_2 \le C\beta_n^d \sup_{j\in\Lambda_n} \left\|m_{in}^s\right\|_4 \left\|m_{jn}^s\right\|_4$$
$$\le C\beta_n^d \sup_{n,i\in\Lambda_n} \left\|m_{in}\right\|_4^2,$$

since $\left\|m_{in}^s\right\|_p \le \left\|m_{in}\right\|_p$ for any $p \ge 1$. Using the Cauchy–Schwartz inequality yields the following bound on the first term in (C.7):

$$\overline{n}^{-2} \sum_{i\in\Lambda_n} \sum_{j\in\Lambda_n:|i-j|\le 3\beta_n} \left|E\chi_{in}^s \chi_{jn}^s\right| \le C\overline{n}^{-1}\beta_n^d \sup_{n,i\in\Lambda_n} \left\|\chi_{in}^s\right\|_2^2$$
$$\le C\overline{n}^{-1}\beta_n^d \sup_{n,i\in\Lambda_n} \left\|m_{in}\right\|_4^4 \beta_n^{2d} \le C\overline{n}^{-1}\beta_n^{3d} \to 0,$$

since $\beta_n^d = O(\overline{n}^{1/4})$. Using Lemma A.1 with $Z_{in} = \chi_{in}^s$, $\psi(h) = 0$ and $\delta = \infty$ implies for i, j s.t. $\left|i - j\right| > 3\beta_n$ that $\left|E\chi_{in}^s \chi_{jn}^s\right| \le C\beta_n^{2d} s^{d\varsigma} \widehat{\alpha}(|i-j|/3)$, since $\left\|\chi_{in}^s\right\|_p \le 2\sum_{j\in\Lambda_n:|i-j|\le\beta_n} \left|K_{ijn}\right| \left\|m_{in}^s\right\|_{2p} \left\|m_{jn}^s\right\|_{2p} \le C\beta_n^d$ for all $p \ge 1$.

Using the last inequality and inequality (A.1), we can now bound the second term in (C.7) for $s = \beta_n$

$$\overline{n}^{-2} \sum_{i \in \Lambda_n} \sum_{j \in \Lambda_n : |i-j| > 3\beta_n} \left| E\chi_{in}^s \chi_{jn}^s \right| \le C\overline{n}^{-2} \beta_n^{2d} \sum_{i \in \Lambda_n} \sum_{j \in \Lambda_n : |i-j| > 3\beta_n} s^{d\varsigma} \widehat{\alpha}([|i-j|/3])$$

$$\le C\overline{n}^{-2} \beta_n^{2d} \sum_{i \in \Lambda_n} \sum_{r=[\beta_n]}^{\infty} \sum_{j \in \Lambda_n : |i-j| \in [r,r+1)} s^{d\varsigma} \widehat{\alpha}(r)$$

$$\le C\overline{n}^{-1} \beta_n^{2d} \sum_{r=[\beta_n]}^{\infty} r^{d(\varsigma+1)-1} \widehat{\alpha}(r) \to 0,$$

by Assumption 9(c). Thus, $|H_{1n}| \xrightarrow{p} 0$. We now show that $|H_{2n}| \xrightarrow{p} 0$. By the Cauchy–Schwartz inequality,

$$E|H_{2n}| \le \overline{n}^{-1} \sum_{i,j \in \Lambda_n, |i-j| \le \beta_n} |K_{ijn}| E\left| m_{in}^s \xi_{jn}^s - E m_{in}^s \xi_{jn}^s \right|$$

$$\le 2\overline{n}^{-1} \sum_{i,j \in \Lambda_n, |i-j| \le \beta_n} \left\| m_{in}^s \right\|_2 \left\| \xi_{jn}^s \right\|_2 \le C\psi(\beta_n)\beta_n^d \to 0,$$

since $\psi(s) = o(s^{-d})$. Similarly, $|H_{3n}| \xrightarrow{p} 0$ and $|H_{4n}| \xrightarrow{p} 0$. Thus, $\left|\widehat{S}_n(\vartheta_0) - E\widehat{S}_n(\vartheta_0)\right| \xrightarrow{p} 0$, as required.

Step 2. Proof of $\left|E\widehat{S}_n(\vartheta_0) - S_n(\vartheta_0)\right| \to 0$. Write $E\widehat{S}_n(\vartheta_0) - S_n(\vartheta_0) = J_{1n} - J_{2n}$, where

$$J_{1n} = \overline{n}^{-1} \sum_{i \in \Lambda_n} \sum_{j \in \Lambda_n : |i-j| \le \beta_n} [K((i-j)/\beta_n) - 1] E m_{in} m_{jn} \text{ and } J_{2n}$$

$$= \overline{n}^{-1} \sum_{i \in \Lambda_n} \sum_{j \in \Lambda_n : |i-j| > \beta_n} E m_{in} m_{jn}.$$

We first show that J_{2n} converges to zero. Using Lemma A.1 with $Z_{in} = m_{in}$ and $\delta = \infty$ yields

$$\left| E m_{in} m_{jn} \right| \le C\left\{ [h/3]^{d\varsigma} \widehat{\alpha}([h/3]) + \psi([h/3]) \right\}. \quad \textbf{(C.8)}$$

Then, using the last inequality and (A.1) gives

$$|J_{2n}| \le C\overline{n}^{-1} \sum_{i \in \Lambda_n} \sum_{r=[\beta_n]}^{\infty} \sum_{j \in \Lambda_n : |i-j| \in [r,r+1)} \left\{ [|i-j|/3]^{d\varsigma} \widehat{\alpha}([|i-j|/3]) + \psi([|i-j|/3]) \right\}$$

$$\le C \sum_{r=[\beta_n/3]}^{\infty} r^{d-1} \left[r^{d\varsigma} \widehat{\alpha}(r) + \psi(r) \right] \to 0,$$

as the tail of a convergent series. Now, using (C.8) and similar arguments as for J_{2n} gives

$$|J_{1n}| \le C \sum_{r=1}^{\infty} \left| K(|a_{r,n}|/\beta_n) - 1 \right| r^{d-1} \left[[r/3]^{d\varsigma} \widehat{\alpha}([r/3]) + \psi([r/3]) \right] \to 0,$$

where $a_{r,n} = \arg\max_{r \le x \le r+1} |K(x/\beta_n) - 1|$. Since $\sum_{r=1}^{\infty} r^{d-1}\left[r^{d\varsigma}\widehat{\alpha}(r) + \psi([r])\right] < \infty$, $K(x)$ is continuous in a neighborhood of $x = 0$ and $\sup_{r,n}\left|a_{r,n}\right| < \infty$, the r.h.s of the last inequality converges to zero by the Dominated Convergence Theorem. Thus, $\left|E\widehat{S}_n(\vartheta_0) - S_n(\vartheta_0)\right| \to 0$, as required.

Step 3. Proof of $\left|\widehat{S}_n\left(\widehat{\vartheta}\right) - \widehat{S}_n(\vartheta_0)\right| \xrightarrow{p} 0$. Note that

$$\left|\widehat{S}_n\left(\widehat{\vartheta}\right) - \widehat{S}_n(\vartheta_0)\right| \le \overline{n}^{-1} \sum_{i\in\Lambda_n} \sum_{j\in\Lambda_n:|i-j|\le\beta_n} \left\{\left|m_{in}(\widehat{\vartheta})\right|\left|m_{jn}(\widehat{\vartheta}) - m_{jn}(\vartheta_0)\right| + \left|m_{jn}(\vartheta_0)\right|\left|m_{in}(\widehat{\vartheta}) - m_{in}(\vartheta_0)\right|\right\}.$$

We need to show that each of the two terms on the r.h.s. of the last inequality is $o_p(1)$. Since $P(\widehat{\tau} \in \mathcal{T}) \to 1$, for all i and n

$$\left|m_{in}(\widehat{\vartheta}) - m_{in}(\vartheta_0)\right| \le C_1\left|\widehat{\theta} - \theta_0\right| + C_2 \sup_{\mathcal{X}^*\times\Theta^*}\{|\widehat{\tau}_1(x) - \tau_{10}(x)| + |\widehat{\tau}_2(x,\theta) - \tau_{20}(x,\theta)| + |\widehat{\tau}_3(x) - \tau_{30}(x)|\} \text{ w.p.1.}$$

Since $\beta_n^d = O(\overline{n}^{1/4})$, $\sup_{\mathcal{X}^*\times\Theta^*}|\widehat{\tau} - \tau_0| = o_p(\overline{n}^{-1/4})$, $\left|\widehat{\theta} - \theta_0\right| = O_p(\overline{n}^{-1/2})$, $\sup_{n,i}\left|m_{in}(\widehat{\vartheta})\right| < \infty$ w.p.1,

$$\overline{n}^{-1}\sum_{i\in\Lambda_n}\sum_{j\in\Lambda_n:|i-j|\le\beta_n}\left|m_{in}(\widehat{\vartheta})\right|\left|m_{jn}(\widehat{\vartheta}) - m_{jn}(\vartheta_0)\right| \le C\sup_{n,i\in\Lambda_n}\left|m_{in}(\widehat{\vartheta})\right| \cdot \beta_n^d\left\{\left|\widehat{\theta} - \theta_0\right| + \sup_{\mathcal{X}^*\times\Theta^*}\{|\widehat{\tau}_1(x) - \tau_{10}(x)| + |\widehat{\tau}_2(x,\theta) - \tau_{20}(x,\theta)| + |\widehat{\tau}_3(x) - \tau_{30}(x)|\}\right\} \xrightarrow{p} 0,$$

Similarly, $\overline{n}^{-1}\sum_{i\in\Lambda_n}\sum_{j\in\Lambda_n:|i-j|\le\beta_n}\left|m_{jn}(\vartheta_0)\right|\left|m_{in}(\widehat{\vartheta}) - m_{in}(\vartheta_0)\right| \xrightarrow{p} 0$, which completes the proof. ■

https://doi.org/10.1017/S0266466614000905 Published online by Cambridge University Press