

Jungyoon Lee and [Peter Robinson](#)  
Series estimation under cross-sectional  
dependence

Article (Published version)  
(Refereed)

**Original citation:** Lee, Jungyoon and Robinson, Peter (2016) *Series estimation under cross-sectional dependence*. [Journal of Econometrics](#), 190 (1). pp. 1-17. ISSN 0304-4076

DOI: [10.1016/j.jeconom.2015.08.001](https://doi.org/10.1016/j.jeconom.2015.08.001)

Reuse of this item is permitted through licensing under the Creative Commons:

© 2015 The Authors  
CC-BY 4.0

This version available at: <http://eprints.lse.ac.uk/63380/>  
Available in LSE Research Online: May 2017

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.



# Series estimation under cross-sectional dependence



Jungyoon Lee<sup>a</sup>, Peter M. Robinson<sup>b,\*</sup>

<sup>a</sup> Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK

<sup>b</sup> London School of Economics, Houghton Street, London WC2A 2AE, UK

## ARTICLE INFO

### Article history:

Received 19 June 2013

Received in revised form

2 August 2015

Accepted 3 August 2015

Available online 11 September 2015

### JEL classification:

C12

C13

C14

C21

### Keywords:

Series estimation

Nonparametric regression

Semiparametric regression

Spatial data

Cross-sectional dependence

Mean square rate of convergence

Functional central limit theorem

Data-driven studentization

## ABSTRACT

An asymptotic theory is developed for series estimation of nonparametric and semiparametric regression models for cross-sectional data under conditions on disturbances that allow for forms of cross-sectional dependence and heterogeneity, including conditional and unconditional heteroscedasticity, along with conditions on regressors that allow dependence and do not require existence of a density. The conditions aim to accommodate various settings plausible in economic applications, and can apply also to panel, spatial and time series data. A mean square rate of convergence of nonparametric regression estimates is established followed by asymptotic normality of a quite general statistic. Data-driven studentizations that rely on single or double indices to order the data are justified. In a partially linear model setting, Monte Carlo investigation of finite sample properties and two empirical applications are carried out.

© 2015 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Economic agents are typically interdependent, due for example to externalities, spill-overs or the presence of common shocks. Such dependence is often overlooked or ignored in cross-sectional or panel data analysis. In order to account for possible cross-sectional dependence, one needs first to establish a framework under which its structure can be suitably formalized, and which permits an asymptotic statistical theory that is useful in statistical inference, in particular a central limit theorem for estimates of functions or parameters of interest. Several approaches to modelling cross-sectional dependence prominent in recent literature can accomplish this. One class of models postulates unobserved common factors that affect some or all individual units, see e.g. Andrews (2005), Pesaran (2006) and Bai (2009), and entail persistent cross-sectional dependence. Two other classes involve a concept of “economic location” or “economic distance”. In economic data,

cross-sectional units correspond to economic agents such as individuals or firms, envisaged as positioned in some socio-economic (even geographical) space, whereby their relative locations underpin the strength of dependence, see e.g. Conley (1999) and Pinkse et al. (2002). The spatial autoregressive (SAR) model of Cliff and Ord (1968, 1981), see e.g. Arbia (2006), Lee (2002, 2004) and Kelejian and Prucha (1998, 1999), employs spatial weight matrices whose elements consist of inverse pairwise economic distances between agents, whence the dependent variable or disturbance for a given unit is affected by a weighted average of the other sampled units' variables. The weights are presumed known and reflect the proximity between agents, leaving a small number of parameters to be estimated. Alternatively, mixing conditions extending ones familiar from the time series literature, have been employed. Conley (1999) and Jenish and Prucha (2012), for example, develop spatial mixing and functions-of-mixing conditions in terms of economic distance between agents, under a suitable stationarity assumption, while an alternative type of condition was proposed by Pinkse et al. (2007). Another approach, of Robinson (2011), employs a possibly non-stationary, linear process for disturbances, with dependence in regressors expressed in terms of the departure of joint densities from the product of marginals; a degree of heterogeneity across

\* Corresponding author. Tel.: +44 20 7955 7516; fax: +44 20 7955 6592.

E-mail address: [p.m.robinson@lse.ac.uk](mailto:p.m.robinson@lse.ac.uk) (P.M. Robinson).

units is permitted, as well as strong dependence analogous to long memory in time series, which is ruled out by mixing conditions, as well as weak dependence, and the model can also accommodate economic distances, as well as lattice or irregularly-spaced data.

On the other hand, nonparametric and semiparametric estimation has become well established in econometric analysis, enabling assumptions of a known parametric functional form, that are frequently not warranted by economic theory, to be dropped or relaxed. There are many theoretical results on nonparametric kernel estimation under temporal dependence, see e.g. [Robinson \(1983\)](#), [Jenish \(2012\)](#), [Robinson \(2011\)](#) and [Robinson and Thawornkaiwong \(2012\)](#) have considered kernel estimation in nonparametric regression and partially linear regression, under forms of cross-sectional dependence. The asymptotic behaviour of series estimation under independence has been studied in [Andrews \(1991\)](#) and [Newey \(1997\)](#). For weakly dependent time series data, [Chen and Shen \(1998\)](#) and [Chen et al. \(2012\)](#) offer a rather complete treatment of asymptotic theory and robust inference of general sieve  $M$ -estimation, while [Chen and Christensen \(2015\)](#) shows that spline and wavelet series regression estimation obtains the optimal uniform convergence rate of [Stone \(1982\)](#).

This paper presents an asymptotic theory for series estimation of nonparametric and semiparametric regression models that covers fairly general cross-sectional heterogeneity and dependence, mainly of a weak form. The conditions of the paper may be relevant to cross-sectional, spatial, time series and panel data, and follow the framework of [Robinson \(2011\)](#), with modifications necessitated by the nature of series estimates relative to kernel ones. Our asymptotic results can easily be modified to cover linear and nonlinear parametric regression. Our other main contribution is establishing a theoretical background for a studentization method that offers an alternative to the existing variance estimation literature. In the spatial context, an extension of heteroscedasticity and autocorrelation consistent (HAC) estimation of the covariance matrix in the limiting normal distribution familiar from the time series literature, see e.g. [Hannan \(1957\)](#), is possible if additional information is available, such as the locations or geographical or economic distances between units. [Conley \(1999\)](#) considered HAC estimation under a stationary random field with measurement error in distances, as did [Kelejian and Prucha \(2007\)](#) for SAR-type models, and [Robinson and Thawornkaiwong \(2012\)](#) in a semiparametric regression set-up. However, the small sample performance of HAC estimation can be poor and an alternative studentization that can produce more accurately sized tests was suggested by [Kiefer et al. \(2000\)](#) in a time series setting. The present paper provides theoretical justification for employing a version of such a studentization in spatial or spatio-temporal data, though it critically relies on an assumption that the practitioner can suitably order the data across one or two dimensions, as may be relevant when geographical locations are known, or there are one or two characteristics believed to be strongly associated with dependence between individuals.

The paper is structured as follows. In Section 2, the model setting is outlined. In Section 3, series estimation is introduced and a mean square rate of convergence for the nonparametric component is established. Section 4 contains asymptotic normality results, covering slower-than- $\sqrt{n}$ , as well as  $\sqrt{n}$ , rate of convergence. In the latter setting, Section 5 presents data-driven studentizations in one and two dimensions. Using the semiparametric partially linear regression model, Section 6 presents a Monte Carlo study of finite sample performance and two empirical examples. Section 7 concludes. Two appendices contain proofs.

## 2. Model setting

The paper commences from the nonparametric regression model

$$Y_i = m(X_i) + U_i, \quad i = 1, 2, \dots, \quad (1)$$

relating observable random variables  $(X_i, Y_i) \in \mathcal{X} \times \mathbb{R}$ , for some set  $\mathcal{X} \subset \mathbb{R}^d$ , where  $m : \mathcal{X} \rightarrow \mathbb{R}$  and  $U_i$  satisfies

$$U_i = \sigma(X_i)e_i, \quad (2)$$

where  $\sigma : \mathcal{X} \rightarrow \mathbb{R}$  and  $e_i \in \mathbb{R}$  are unobservable random variables with zero mean and finite variance, independent of  $\{X_i\}_{i=1}^\infty$  and  $\sigma : \mathcal{X} \rightarrow \mathbb{R}$ . We regard  $m$  and  $\sigma$  as nonparametric functions. The factor  $\sigma(X_i)$  allows for conditional heteroscedasticity in  $Y_i$ , and the factor  $e_i$  for dependence and unconditional heteroscedasticity. We observe  $(X_i, Y_i)$  for  $i = 1, \dots, n$ , while the  $e_i$  (and thence the  $U_i$  and  $Y_i$ ) can form triangular arrays, so  $e_i = e_{in}$ , etc., but this feature will be suppressed in our notation; triangular arrays enable coverage of a wide range of models for spatial dependence, including SAR models with normalized weight matrices, and stationary models for panel data or multi-dimensional lattice or irregularly-spaced data where the single index  $i$  in (1) and (2) requires a re-labelling of multiple indices which is liable to change as  $n$  increases, as discussed by [Robinson \(2011\)](#), who considered kernel estimation of  $m$ . In some of our work  $e_i$  can have semi-strong dependence analogous to that found in long memory time series models. In Sections 3 and 4 we qualify (1) and (2) by detailed regularity conditions, including also restrictions on the dependence of  $X_i$ .

Under the preceding conditions  $m(x) = E(Y_i|X_i = x)$  for  $x \in \mathcal{X}$ . We will estimate  $m$  by a series nonparametric regression estimate  $\hat{m}$ , constructed as a linear combination of pre-specified approximating functions. More generally, we are interested in estimating a  $d \times 1$  vector functional  $a(m)$  of  $m$ , as in [Andrews \(1991\)](#) and [Newey \(1997\)](#), where  $a(m)$  can be estimated by  $a(\hat{m})$ . Simple nonparametric examples of  $a(m)$  include the value of  $m$ ,  $a(m) = (m(x_1), \dots, m(x_d))'$ , and the value of the partial derivative,  $a(m) = \left( \partial/\partial x_\ell m(x)|_{x_1}, \dots, \partial/\partial x_\ell m(x)|_{x_d} \right)'$ , at multiple fixed points  $(x_1, \dots, x_d) \in \mathcal{X}^d$ , where  $x_\ell$  is the  $\ell$ th element of  $x$ . Of semiparametric examples of  $a(\cdot)$ , the partially linear regression model, will be discussed in detail in Section 5. Other  $a(\cdot)$ , including nonlinear functionals, can be found in [Newey \(1997\)](#). [Andrews \(1991\)](#) established asymptotic normality for series estimates of a vector-valued linear  $a(\hat{m})$ , with  $X_i$  and  $U_i$  independent and non-identically distributed, and indicated that his proof can be extended to cover strong mixing time series regressors. [Newey \(1997\)](#) established uniform and integrated mean square rates of consistency for  $\hat{m}(x)$  and asymptotic normality of  $a(\hat{m})$  when  $X_i$  and  $U_i$  are independent and identically distributed (iid) and  $a(m)$  is a possibly nonlinear scalar functional, also describing conditions under which  $a(\hat{m})$  converges to  $a(m)$  at parametric rate. [Chen and Shen \(1998\)](#) and [Chen et al. \(2012\)](#) considered these issues for sieve extreme estimates with weakly dependent time series, with rules of inference that are robust to weak dependence. [Chen et al. \(2012\)](#) also indicated that for certain cases of slower-than- $\sqrt{n}$  rate of convergence such as when  $a(m) = m$ , the asymptotic variance of the estimate  $a(\hat{m})$  coincides with that obtained under independence, as found for kernel estimation by [Robinson \(1983\)](#), for example.

## 3. Estimation of $m$ and mean square convergence rate

The estimation of  $m$  is based on user-chosen approximating functions  $p_s(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ ,  $s = 1, 2, \dots$ , and a data-free integer  $\kappa$  denoting the number of  $p_s(\cdot)$  employed. Denote

$$p^k(\cdot) = (p_1(\cdot), \dots, p_k(\cdot))', \quad k \geq 1; \quad (3)$$

$$P = [p^\kappa(X_1), \dots, p^\kappa(X_n)]', \quad \kappa \geq 1;$$

$$Y = (Y_1, \dots, Y_n)'; \quad \hat{\beta} = (P'P)^{-1}P'Y, \quad (4)$$

where  $A^-$  denotes the Moore–Penrose pseudo-inverse of a matrix  $A$ , and a series estimate of  $m(x)$  by

$$\hat{m}(x) = p^\kappa(x)' \hat{\beta}. \quad (5)$$

As usual, the choice of  $\kappa$  entails a bias/variance trade-off, where increasing  $\kappa$  increases variance while reducing bias, and *vice versa*. In this connection, our first assumption restricts smoothness of the unknown function  $m$ . We first introduce a weighted sup norm for functions in order to allow for unbounded support  $\mathcal{X} = \mathbb{R}^q$  of  $X_i$ :

$$|g|_{\infty,w} := \sup_{x \in \mathcal{X}} |g(x)|(1 + \|x\|^2)^{-w/2},$$

for some  $w \geq 0$ , see e.g. [Chen et al. \(2005\)](#) and [Chen \(2007\)](#). The weighted sup norm becomes the more familiar sup norm  $|g|_{\infty} = \sup_{x \in \mathcal{X}} |g(x)|$  when  $w = 0$ , and this suffices in place of  $|g|_{\infty,w}$  in the following assumption when the  $X_i$  are uniformly bounded.

**Assumption A1.** Let  $w \geq 0$  be the largest value such that  $\max_{i \geq 1} E \|X_i\|^{2w} < \infty$ . There exists a sequence of vectors  $\beta = \beta_{\kappa}$  and a number  $\alpha > 0$  satisfying,

$$|m - p^{\kappa'} \beta|_{\infty,w} = O(\kappa^{-\alpha}), \quad \text{as } \kappa \rightarrow \infty.$$

Conditions similar to [Assumption A1](#), with  $|\cdot|_{\infty}$  in place of  $|\cdot|_{\infty,w}$ , were used in earlier series estimation literature, see e.g. [Andrews \(1991\)](#) and [Newey \(1997\)](#). Further insights into those conditions for various  $p_s(\cdot)$ , including polynomials, trigonometric polynomials, splines and orthogonal wavelets, can be found in [Chen \(2007, p. 5573\)](#). [Chen et al. \(2005\)](#) noted that the weight function  $(1 + \|x\|^2)^{-w/2}$  used here could be regarded as an alternative to the trimming used in kernel estimation when support is unbounded,  $\mathcal{X} = \mathbb{R}^q$ . Conditions imposing an upper bound on  $\kappa$  may necessitate stronger smoothness of  $m$ .

For any non-negative definite matrix  $B$  denote by  $\underline{\lambda}(B)$  and  $\bar{\lambda}(B)$  the smallest and largest eigenvalues respectively, and for any real matrix  $B$  denote the spectral norm  $\|B\| = \bar{\lambda}^{1/2}(B'B)$ . As in [Andrews \(1991\)](#) and [Newey \(1997\)](#), for example, introduce the sequence

$$\xi = \xi_{\kappa} = \sup_{x \in \mathcal{X}} \|p^{\kappa}(x)\|. \tag{6}$$

If  $m$  is believed bounded, one might choose bounded  $p_s(\cdot)$ , whence  $\xi \leq C\sqrt{\kappa}$ , where  $C$  throughout denotes a generic arbitrarily large positive constant. It may be possible to obtain the rate of  $\xi$  exactly: [Newey \(1997\)](#) showed that under suitable conditions,  $\xi \sim \kappa$  when the  $p_s(\cdot)$  are orthogonal polynomials, and  $\xi \sim \kappa^{1/2}$  when they are B-splines,  $a \sim b$  meaning that  $a/b$  tends to a positive finite limit. Define

$$Q = Q_n = n^{-1}P'P. \tag{7}$$

Denote by  $c$  an arbitrarily small generic positive constant. The following assumption presupposes elimination of any redundant  $p_s(\cdot)$ :

**Assumption A2.** For all sufficiently large  $n$ ,  $E(p_s(X_i)^2) \leq \infty$  for  $1 \leq s \leq \kappa$ ,  $1 \leq i \leq n$  and  $\underline{\lambda}(E(Q)) \geq c$ .

The first inequality of [A2](#) implies  $E \|p^{\kappa}(X_i)\|^2 \leq C\kappa$  which is equivalent to  $\text{tr}(E(Q)) \leq C\kappa$  if the  $X_i$  are identically distributed, whereby  $E(Q) = E(p^{\kappa}(X_1)p^{\kappa}(X_1)')$ .

**Assumption A3.** The model [\(1\)](#), [\(2\)](#) holds, where the sequences  $\{X_i\}_{i=1}^{\infty}$  and  $\{e_i\}_{i=1}^{\infty}$  are mutually independent,  $\max_{i \geq 1} E\sigma^4(X_i) < \infty$ , and the  $e_i$  have zero means and finite variances.

To describe dependence in  $\{X_i\}$  introduce

$$\zeta_{ij} = \sup_{f \in F_{ij}} \text{Cov}(f(X_i), f(X_j)), \tag{8}$$

$$F_{ij} = \{f : Ef^2(X_i) = Ef^2(X_j) = 1\},$$

$$\Delta = \Delta_n = \frac{1}{n(n-1)} \sum_{i,j=1, i \neq j}^n \zeta_{ij}. \tag{9}$$

The quantity  $\zeta_{ij}$  is bounded above by the maximal correlation coefficient of [Rozanov \(1963\)](#), for example. For Gaussian  $X_i$ , it follows from Lemma 10.2 of [Rozanov \(1963\)](#) that  $\zeta_{ij} \leq |\text{Corr}(X_i, X_j)|$ . Further, for a stationary Gaussian process  $X_i$ ,  $\zeta_{ij}$  is upper-bounded by the  $\alpha$ -mixing coefficient with index  $|i - j|$ . More generally, the  $X_i$  need not have absolutely continuous distribution functions. The quantity  $\Delta$  is an overall measure of dependence; the property  $\Delta \sim n^{-1}$  is analogous to weak dependence of time series, while a rate between that and the upper bound  $\Delta \leq 1$  is analogous to strong dependence.

**Assumption A4.** As  $n \rightarrow \infty$ ,  $\kappa^{-1} + \kappa \xi^2(n^{-1} + \Delta) \rightarrow 0$ .

The  $\Delta$  component is vacuous under independence of  $X_i$  and otherwise indicates that the stronger the dependence the smaller  $\kappa$  needs be. In light of [A1](#), this necessitates greater smoothness of  $m$ . The  $\kappa \xi^2/n \rightarrow 0$  component reduces under suitable conditions to  $\kappa = o(n^{1/2})$  for B-splines and  $\kappa = o(n^{1/3})$  for orthonormal polynomials.

To describe dependence in  $e_i$  define

$$\gamma_{ik} = E(e_i e_k); \quad \omega = \omega_n = \max_{1 \leq i \leq n} \frac{1}{n} \sum_{k=1}^n |\gamma_{ik}|. \tag{10}$$

The property

$$\omega \sim n^{-\zeta}, \quad \zeta \in (0, 1], \tag{11}$$

can be said to cover weak dependence in  $e_i$  when  $\zeta = 1$ , and “semi-strong dependence” when  $\zeta \in (0, 1)$ , the latter case corresponding to long memory in stationary time series. Define

$$\chi = \omega\kappa \quad \text{if } \sup_x \sigma^2(x) < \infty; = \omega\xi\kappa^{1/2}, \quad \text{otherwise.} \tag{12}$$

Both versions of  $\chi$  are identical when  $\xi \sim \kappa^{1/2}$ . The following theorem obtains an integrated mean square rate of convergence.

**Theorem 1.** Under [Assumptions A1–A4](#), as  $n \rightarrow \infty$ ,

$$\max_{1 \leq i \leq n} E|\hat{m}(X_i) - m(X_i)|^2 = O(\chi + \kappa^{-2\alpha}). \tag{13}$$

Thus if  $(\chi + \kappa^{-2\alpha}) \rightarrow 0$ ,  $\hat{m}$  is consistent. For example if  $\xi \sim \kappa^{\rho}$  [\(13\)](#) becomes  $O(\kappa^{1/2+\rho}n^{-\zeta} + \kappa^{-2\alpha})$  in view of [\(11\)](#), leading to the optimal rate  $\kappa \sim n^{\zeta/(1/2+\rho+2\alpha)}$ . The rate in [\(13\)](#) coincides with [Newey’s \(1997\)](#) for iid  $X_i$  and  $e_i$ , and bounded  $\sigma$ , when  $\chi \sim \kappa/n$ , and weak dependence in  $e_i$  ( $\zeta = 1$  in [\(11\)](#)) leaves the rate unchanged. On the other hand, more generally under only  $E\sigma^4(X_i) \leq C$  and [\(11\)](#), we have  $\chi \sim \xi\kappa^{1/2}n^{-\zeta}$ , entailing stronger restrictions on  $\kappa$  and/or  $\xi$  to achieve consistency. [Newey \(1997\)](#) also covered a general concept of uniform convergence (where his uniform rate was improved by [de Jong \(2002\)](#), for compact  $\mathcal{X}$ ). A uniform rate of convergence in our setting may be shown to be  $O_p(\xi(\chi^{1/2} + \kappa^{-\alpha}))$  under a modification of [A1](#) stated in terms of sup norm, using the proof of [Theorem 1](#). More recently, [Chen and Christensen \(2015\)](#) verified that spline and wavelet series regression estimators for weakly dependent regressors attain the optimal uniform convergence rate of [Stone \(1982\)](#).

#### 4. Asymptotic normality

Our interest now lies in inference on the  $d \times 1$  deterministic functional  $a(m)$ , for which a central limit theorem is the first step. For this purpose, we denote

$$\theta_0 = a(m), \quad \hat{\theta} = \hat{\theta}_n = a(\hat{m}),$$

and introduce the following assumptions from [Newey \(1997\)](#), modified here with the weighted sup norm:

**Assumption B1.** Either: (i)  $a(g)$  is a linear operator in  $g$ , or: (ii) For any  $\epsilon > 0$ , there exists a linear operator  $D(g)$  and constant  $C_\epsilon < \infty$  such that  $\|a(g) - a(m) - D(g - m)\| \leq C_\epsilon \|g - m\|_{\infty, w}^2$ , if  $\|g - m\|_{\infty, w} \leq \epsilon$ .

**Assumption B2.**  $\|D(g)\| \leq C \|g\|_{\infty, w}$ .

B2 implies continuity of  $D(\cdot)$ , the Frechet-differential of  $a(\cdot)$  at  $m$ , while B1(ii) imposes a stronger smoothness condition on  $a(\cdot)$  at  $m$  and is not restrictive, see e.g. Newey (1997, p. 153). When  $a(\cdot)$  is a linear operator, its Frechet-derivative is itself,  $D(g) = a(g)$ . For example, for the linear operator  $a(g) = (g(x_1), \dots, g(x_d))'$ ,  $a(p_s) = D(p_s) = (p_s(x_1), \dots, p_s(x_d))'$ ,  $s = 1, \dots, \kappa$ .

Define the  $\kappa \times d$  matrix

$$A = A_n = (D(p_1), \dots, D(p_\kappa))'.$$

Also, using (3), let

$$v = v_n = n^{-1/2} A' P' U, \quad U = (U_1, \dots, U_n)', \quad (14)$$

and thence the conditional and unconditional variance matrices

$$\begin{aligned} \bar{V} &= \bar{V}_n = E(vv' | X_i, i = 1, \dots, n) \\ &= \frac{1}{n} \sum_{i,k=1}^n \gamma_{ik} \sigma(X_i) \sigma(X_k) A' p^\kappa(X_i) p^\kappa(X_k)' A, \end{aligned}$$

$$V = V_n = E(vv') = \frac{1}{n} \sum_{i,k=1}^n \gamma_{ik} E[\sigma(X_i) \sigma(X_k) A' p^\kappa(X_i) p^\kappa(X_k)' A].$$

**Assumption B3.** For a positive non-increasing scalar sequence  $\tau = \tau_n$ ,  $\tau \lambda(V) \geq c$  for all sufficiently large  $n$ .

B3 requires  $A$  to have rank  $d$  for all  $\kappa \geq d$ . The case  $\tau \rightarrow 0$  as  $n \rightarrow \infty$  is relevant when  $a$  is a nonparametric functional, converging at rate  $\sqrt{n\tau}$ , while the case  $\tau \equiv 1$  is relevant when  $a$  is a semiparametric functional, converging at rate  $\sqrt{n}$ , the latter situation being discussed in detail in the following section. B3 does not satisfactorily allow both nonparametric and semiparametric components in  $a$ , which will have different rates; theory suggests they should use different  $\kappa$ , but our statistic is insufficiently general to permit this. We strengthen the rate condition A4 to:

**Assumption B4.** As  $n \rightarrow \infty$ ,

$$\xi^4 \tau \kappa (1 + n\Delta) + \xi^4 n \tau (\chi^2 + \kappa^{-2\alpha}) \rightarrow 0.$$

To prove asymptotic normality we introduce a more detailed specification of  $e_i$ :

**Assumption B5.** (2) holds with

$$e_i = \sum_{j=1}^{\infty} b_{ij} \varepsilon_j, \quad i = 1, 2, \dots, \quad (15)$$

where the  $\varepsilon_j, j \geq 1$ , are independent random variables with zero mean and unit variance, the  $\varepsilon_j^2$  are uniformly integrable, and the  $b_{ij}$  are real constants such that for some  $\eta > 0$  and all sufficiently large  $n$ ,

$$\sum_{i=1}^n |b_{ij}| + \sum_{j=1}^{\infty} |j^\eta b_{ij}| \leq C, \quad i = 1, 2, \dots; j = 1, 2, \dots \quad (16)$$

The linear process (15) was employed by Robinson (2011) and the dependence on both  $i$  and  $j$  of  $b_{ij}$  (like  $\gamma_{ij}$ ), rather than just their

difference  $i - j$ , distinguishes (2) from representations for stationary time series, as does our allowance for them to be triangular arrays, thereby covering models described after (1) and (2) in Section 2. In SAR models, where the  $b_{ij}$  tend to reflect economic distances between agents, we have, for all  $i, b_{ij} = 0$  for  $j > n$ , and  $\sum_{j=1}^n |b_{ij}| \leq C$  is commonly assumed, see e.g. Kelejian and Prucha (1998) and Lee (2004). In models featured in the spatial statistics literature it is frequently natural to allow  $b_{ij}$  to be non-zero for all  $j$ , analogously to autoregressive time series models. Condition (16) implies weak dependence in  $e_i$ : noting that

$$\gamma_{ik} = E(e_i e_k) = \sum_{j=1}^{\infty} b_{ij} b_{kj}, \quad (17)$$

it is easily seen that  $\sum_{k=1}^n |\gamma_{ik}| \leq C$  and thus  $\omega = O(n^{-1})$  in (10).

Finally, some falling off of fourth cumulants is required in the assumption below.

**Assumption B6.** Defining  $w_i^{(h)} := p_h(X_i) \sigma(X_i)$  for  $i \geq 1, 1 \leq h \leq \kappa$ , as  $n \rightarrow \infty$

$$\frac{\xi^2}{n^2} \sum_{i_2, i_4=1}^n \max_{i_1, i_3} \left( \sum_{h_1, \dots, h_4=1}^{\kappa} \left( \text{cum}(w_{i_1}^{(h_1)}, \dots, w_{i_4}^{(h_4)}) \right)^2 \right)^{1/2} \rightarrow 0. \quad (18)$$

Under independence of the  $X_i$ , or more generally finite dependence (similar to  $m$ -dependence in time series), this condition entails  $\kappa^2 \xi^2 / n \rightarrow 0$  as  $n \rightarrow \infty$ .

Denote by  $B^{1/2}$  the unique positive definite square root of a positive definite matrix  $B$ .

**Theorem 2.** Under Assumptions A1–A4 and B1–B6, as  $n \rightarrow \infty$ ,

$$\sqrt{n} \bar{V}^{-1/2} (\hat{\theta} - \theta_0) \rightarrow_d N(0, I_d), \quad (19)$$

$$\sqrt{n} V^{-1/2} (\hat{\theta} - \theta_0) \rightarrow_d N(0, I_d). \quad (20)$$

Replacing  $V$  by a data dependent quantity without affecting the limit distribution in (20) typically requires additional information, as discussed in Section 1. In the weakly dependent time series context, Chen et al. (2012) found that under conditions on  $a(\cdot)$  that preclude  $\sqrt{n}$ -convergence of  $a(\hat{m})$ ,  $V$  reduces asymptotically to the same matrix as under the assumption of independence (cf Robinson (1983) for kernel estimates), which in our setting is  $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \gamma_{ii} E[\sigma(X_i)^2 A' p^\kappa(X_i) p^\kappa(X_i)' A]$ . This prompts interest in developing, for spatial settings, inference under  $\sqrt{n}$ -convergence, when such a simplification does not result. As indicated in Section 1, one example of  $a(m)$  is the vector  $(m(x_1), \dots, m(x_d))'$  whence, using Theorem 2 and after a relatively simple studentization a test for a specific functional form of  $m(x)$  analogous to that of Robinson (1983) can be developed, though such a test lacks satisfactory consistency properties and much more work would be required to justify a consistent test. Of course in case of a nonparametric  $a(m)$  the rate of convergence in Theorem 2 is slower than  $\sqrt{n}$ , whereas in the sequel we focus on the, semiparametric,  $\sqrt{n}$  case.

### 5. Studentization

We present a studentization for a  $\sqrt{n}$ -convergent semiparametric estimate  $\hat{\theta}$ . Attainment of the parametric  $\sqrt{n}$  rate by semiparametric estimates has attracted wide interest in the econometric literature, following Robinson (1988) and Powell et al. (1989), who used kernel estimation. Newey (1997) developed  $\sqrt{n}$ -convergence of a series semiparametric estimate, while Chen

and Shen (1998) establish extensions for weakly dependent time series. In our spatial setting statistical inference requires a studentization that is robust to dependence and heterogeneity across  $i$ . Section 1 discussed difficulties with HAC estimation in dealing with cross-sectional dependence. Given a suitable ordering of data, however,  $a(\hat{m})$  can be studentized without assuming any particular dependence structure, following the time series approach of Kiefer et al. (2000). We consider two different versions of such studentization, depending on whether data are to be ordered according to one or two indices. Below, we first present some common assumptions that are required for both versions of the studentization, then each version is considered in detail in the following subsections.

Define, for a fixed  $r \in [0, 1]$ :

$$V^{(r)} := \frac{1}{[rn]} \sum_{i,k=1}^{[rn]} \gamma_{ik} E[\sigma(X_i)\sigma(X_k)A'p^k(X_i)p^k(X_k)'A].$$

Throughout this section, we impose the following unprimitive assumption.

**Assumption C1.** There exists a finite positive definite matrix  $\Omega$  such that  $\lim_{n \rightarrow \infty} V^{(r)} = \Omega$  for all  $r \in [0, 1]$ .

Sufficient conditions for C1 involve technical restrictions on the functional derivative  $D(\cdot)$  of  $a(\cdot)$  that can be found in Assumption 7 of Newey (1997), where they are verified to hold for the estimands in partially linear and single index models and also when the quantity of interest is the approximate consumer surplus. In a non-iid setting, we also require weak dependence in  $e_i$  for C1 to hold, which is implied by B5. Existence of  $\Omega$  is a condition on the collective strength of dependence in  $U_i$  and  $X_i$ , comparable to Assumption A4 of Robinson and Thawornkaiwong (2012). C1 also requires a degree of homogeneity across units, although somewhat less stringent than identity of distribution.

**Assumption C2.** As  $n \rightarrow \infty$ : (i)  $\Delta = O(n^{-1})$ ; (ii)  $\chi = O(\kappa/n)$ ; (iii)  $\sqrt{n}\xi^3\kappa^{\frac{1}{2}-\alpha} = o(1)$ ; (iv)  $\kappa^2\xi^4 = o(n)$ .

Assumption C2(i) implies weak dependence of  $X_i$ , C2(ii) restricts the dependence across  $i$  of  $p^k(X_i)U_i$ , C2(iii) strengthens the smoothness condition of B3(iii) and C2(iv) further restricts the growth of  $\kappa$  and  $\xi$ , but is satisfied by, for example, the rate mentioned after Assumption B6.

**Assumption C3.**  $\max_{j \geq 1} E(\varepsilon_j^4) < \infty$ , and  $\sup_{x \in \mathcal{X}} |\sigma(x)| < \infty$ .

Let  $D(\cdot; g)$  denote the functional derivative of  $a(\cdot)$  evaluated at  $g$ , and  $D(\cdot; g) = (D_1(\cdot; g), \dots, D_d(\cdot; g))'$ . The following assumption from Newey (1997) requires  $D_i(\cdot; g)$  to exhibit continuity over  $g$ .

**Assumption C4.** For any  $\epsilon > 0$  and all  $\tilde{g}, \bar{g}$  such that  $|\tilde{g} - m|_{\infty, w} \leq \epsilon$  and  $|\bar{g} - m|_{\infty, w} \leq \epsilon$ , there exists  $C_\epsilon < \infty$  such that  $\|D_i(g; \tilde{g}) - D_i(g; \bar{g})\| \leq C_\epsilon |g|_{\infty} |\tilde{g} - \bar{g}|_{\infty, w}$ ,  $i = 1, \dots, d$ .

### 5.1. Studentization for one dimensional ordering

In this subsection, we consider the case when one has information to suitably order the data with a single index. We first introduce the quantities used in the studentization and discuss in detail the requisite assumption on the ordering.

We estimate  $A$  by

$$\hat{A} = \frac{\partial a(p^{k'}\beta)}{\partial \beta} \Big|_{\beta = \hat{\beta}} \quad (21)$$

and with  $\hat{M} = (\hat{m}(X_1), \dots, \hat{m}(X_n))$ ,  $\hat{U} = Y - \hat{M}$ , set

$$\hat{S}_{n,m}^* = \sum_{i=1}^m \hat{A}'Q^{-1}p^k(X_i)\hat{U}_i/\sqrt{n}, \quad 1 \leq m \leq n,$$

$$\hat{B}_n = \frac{1}{n} \sum_{m=1}^n \hat{S}_{n,m}^* \hat{S}_{n,m}^{*'},$$

$$\Psi_d = \int_0^1 [W_d(r) - rW_d(1)][W_d(r) - rW_d(1)]'dr,$$

where  $W_d(\cdot)$  denotes a  $d$ -dimensional vector of independent Brownian motions, so  $\Psi_d$  is the integral of the outer product of the  $d$ -dimensional multivariate Brownian bridge, and we note that  $EW_d(r)W_d(u)' = rI_d$ ,  $0 \leq r \leq u \leq 1$ .

**Assumption C5.** As  $n \rightarrow \infty$ : (i)  $\sum_{i=1}^{[rn]} \sum_{k=[rn]+1}^n |\gamma_{ik}| = o(n)$  uniformly in  $r \in [0, 1]$ ; (ii)  $\max_{1 \leq i \leq n} \sum_{k=1}^n |\gamma_{ik}| = O(1)$ .

Assumption C5(ii) further rules out the presence of any “dominant” unit whose error covariances with new units added to the sample are persistently significant. In our other assumptions, the ordering of the  $n$  observations is arbitrary, but Assumption C5(i), required to justify our studentization, is unfortunately highly sensitive to ordering, requiring some falling-off of dependence as  $|i-k|$  increases (Lemma 2 in Appendix B shows that it is satisfied if there exists a positive function  $\eta(\cdot)$  such that  $|\gamma_{ik}| \leq \eta(i-k)$ ,  $i, k = 1, 2, \dots$ , and  $\sum_{j=-\infty}^{\infty} \eta(j) < \infty$ ). This is often reasonable with time series, but for spatial data there is generally no natural ordering and the bulk of the  $n!$  possible orderings will not satisfy C5(i), so considerable care would be required in ordering the data. In some economic applications data might be ordered with respect to some relevant (explanatory) variable. For example, with firm data, firms using similar inputs or producing similar outputs might be expected to exhibit high correlation in disturbances. It may be that some economic distances are available, as in SAR models, in which case they can facilitate ordering by indicating neighbouring units although there would still remain considerable arbitrariness in the ordering. Such considerations are pursued in the Monte Carlo study in the following section. Other approaches to the modelling of cross sectional data that would justify alternative studentizations can also be challenged. Mixing assumptions, for example, are, like ours, nonparametric, but they generally presuppose that observations are recorded on a Euclidean space and distances between units are known to the practitioner for inference robust to dependence, using HAC variance estimation, to be carried out. Parametric models, such as SAR and factor models, with the former requiring the practitioner to specify one or more  $n \times n$  spatial weight matrices, run obvious risks of misspecification. If the practitioner has no information that suggests a plausible ordering of data, e.g. in the case one has a random permutation of observations, one cannot use our studentization to carry out inference that is robust to cross-sectional dependence, just as one cannot with any other alternative method available.

**Theorem 3.** Under the assumptions of Theorem 2, and Assumptions C1–C5, as  $n \rightarrow \infty$

$$n(\hat{\theta} - \theta_0)' \hat{B}_n^{-1} (\hat{\theta} - \theta_0) \rightarrow_d W_d(1)' \Psi_d^{-1} W_d(1).$$

Theorem 3 may be employed in approximate hypothesis tests and interval estimation, with critical values given in Table 2 of Kiefer et al. (2000).

### 5.2. Studentization for two dimensional ordering

We can partially relax the ordering required in the previous subsection. When one has geographical locations of individuals

or spatio-temporal data, it is natural to consider two dimensional ordering of the data. For a given number  $n$  of observations, suppose we can choose  $s$  and  $t$  such that  $n = st$ . In the case of geographical locations, this setting does not necessarily entail a grid or regular spacing. Suppose each location has two (real valued) coordinates, north and east. One may divide the south–north axis into  $s$  intervals, possibly of different length so that each has  $t$  observations. In each such interval one then orders the  $t$  observations along the west–east axis, and thence finally gives each observation two subscripts,  $j = 1, \dots, s$  and  $k = 1, \dots, t$ . The choice of  $s$  and  $t$  introduces arbitrariness. Moreover, having gone through such an indexing procedure, so e.g.  $X_{jk}$  corresponds to  $x(u_i, v_i)$ , we can have for observation points  $X_{j_1 k_1} = x(u_{i_1}, v_{i_1})$ ,  $X_{j_2 k_2} = x(u_{i_2}, v_{i_2})$ , the outcome  $v_{i_1} < v_{i_2}$  but  $k_1 > k_2$ , say, so while the procedure is appropriate for (possibly irregular) lattice data, for example, it entails an element of arbitrariness in relation to general geographical data sets on two dimensions. There is also an underlying difference in our attitudes to the one- and two-dimensional orderings of this and the previous section. In the former case, our attitude is that the ordering may be possible (e.g. it will be in the time series case) but typically more or less problematic, whereas in the latter we take for granted that we know locations, at least approximately. It should also be added that if in fact we know actual locations in either case, our procedures do not in general use that information, only orderings, so will not be ideal.

Though we adopt the above simplified method of ordering in establishing our results, they will also hold if the  $s$  intervals contain differing numbers  $t_j$  of observations, where  $\sum_{j=1}^s t_j = n$ . It would seem also to be possible to extend our approach and results to three or more dimensions.

Some of the previously defined quantities need to be rewritten to incorporate two dimensional ordering of data. Note that quantities that do not depend on the ordering of data such as  $\hat{\theta}$  do not need to be restated. The model can be rewritten as  $Y_{jk} = m(X_{jk}) + U_{jk}$ . Denote  $\gamma_{jkh} = Cov(U_{jk}, U_{th})$ , where  $U_{jk} = \sum_{i=1}^{\infty} \sum_{h=1}^{\infty} b_{jkih} \varepsilon_{ih}$  and let

$$V = \frac{1}{n} \sum_{j,t=1}^s \sum_{k,h=1}^t \gamma_{jkh} E[\sigma(X_{jk})\sigma(X_{th})A'p^k(X_{jk})p^k(X_{th})'A],$$

$$\hat{S}_{n,m,\ell}^* = \sum_{j=1}^m \sum_{k=1}^{\ell} \hat{A}'Q^{-1}p^k(X_{jk})\hat{U}_{jk}/\sqrt{n},$$

$$\hat{C}_n = \frac{1}{n} \sum_{m=1}^s \sum_{\ell=1}^t \hat{S}_{n,m,\ell}^* \hat{S}_{n,m,\ell}^{*'},$$

$$\Phi_d = \int_0^1 \int_0^1 [W_d(r, u) - ruW_d(1, 1)] \times [W_d(r, u) - ruW_d(1, 1)]' drdu,$$

where  $W_d(\cdot, \cdot)$  denotes a  $d$ -dimensional vector of independent Brownian sheets such that  $W_d(0, 0) = 0$  and  $Cov(W_d(r, u), W_d(r', u')) = (r \wedge r')(u \wedge u')I_d$  for  $r, u, r', u' \in [0, 1]$ .

**Assumption C5'.** As  $n \rightarrow \infty$ : (i)  $\sum_{k,h=1}^t \sum_{j=1}^{[rs]} \sum_{i=[rs]+1}^s |\gamma_{jkh}| = o(n)$  uniformly in  $r \in [0, 1]$  and  $\sum_{j,t=1}^s \sum_{k=1}^{[ut]} \sum_{h=[ut]+1}^t |\gamma_{jkh}| = o(n)$  uniformly in  $u \in [0, 1]$ ; (ii)  $\max_{1 \leq i \leq s, 1 \leq h \leq t} \sum_{j=1}^s \sum_{k=1}^t |\gamma_{jkh}| = O(1)$ .

**Assumption C5'(ii)** is identical to **C5(ii)**, just rewritten with the two subscripts. **Assumption C5'(i)** requires a suitability of ordering in both directions.

**Theorem 4.** Under the assumptions of **Theorem 2**, and **Assumptions C1–C4, C5'**, as  $n \rightarrow \infty$

$$n(\hat{\theta} - \theta_0)' \hat{C}_n^{-1} (\hat{\theta} - \theta_0) \rightarrow_d W_d(1, 1)' \Phi_d^{-1} W_d(1, 1).$$

**Table 1** provides critical values for  $W_d(1, 1)' \Phi_d^{-1} W_d(1, 1)$  simulated from 50,000 iterations, where  $W_d(1, 1)$  was approximated by normalized sums of iid Gaussian random variables in 1000 steps in both directions.

### 6. Numerical results for the partially linear model

The present section contains Monte Carlo investigation of finite-sample performance of our estimates and two empirical applications, in all cases for the partially linear regression model, which is discussed in the following sub-section.

#### 6.1. Partially linear regression model

We partition  $X_i = (W_i', Z_i)'$  and specialize (1) to

$$Y_i = Z_i' \delta_0 + h(W_i) + U_i, \tag{22}$$

where  $h(\cdot)$  is a nonparametric function and  $\delta_0$  is an unknown parameter vector, so  $m(x) = m(w, z) = z' \delta_0 + h(w)$ . This model is particularly suitable when  $Z_i$  contains categorical variables, and is often used when the overall number of regressors is large, when a fully nonparametric specification suffers the curse of dimensionality. Kernel estimation of this model has received much attention see e.g. **Robinson (1988)** and **Fan and Li (1999)**, where  $\delta_0$  was shown to be estimable at  $\sqrt{n}$  rate despite first stage nonparametric estimation having a slower-than- $\sqrt{n}$  rate. Series estimation of (22) was considered in **Chamberlain (1986)** where the choice of series functions takes into account the partially linear form, the first  $d$  functions being  $Z_i$ , and the remaining  $\kappa - d$  are functions of  $W_i$  only. The series estimate of  $\delta_0$  is then the first  $d$  elements of  $\hat{\beta}$ , and  $\hat{h}(x) = \hat{m}(w, z) - z' \hat{\delta}$ . There is more than one  $a(\cdot)$  that yields  $a(m) = \delta_0$ . **Andrews (1991)** considered  $a(m) = \partial m(w, z) / \partial z = \delta_0$  for any  $w, z$ , earlier mentioned by **Robinson (1988)**, in a kernel context. Here we use the functional of **Newey (1997)**, which leads to  $\sqrt{n}$ -consistency. Denote  $Z^* = Z - E(Z|W)$ , where  $Z$  and  $W$  are random variables independent of the data used to construct  $\hat{\delta}$ . Suppose  $E(Z^* Z^{*'})$  is non-singular, an identification condition, and consider

$$a(m) = E \{ [E(Z^* Z^{*'})]^{-1} Z^* m(W, Z) \}.$$

Now  $E(Z^* Z^{*'}) = E[Z^* Z' - Z^* E(Z'|W)] = E(Z^* Z')$ , since  $E[Z^* E(Z'|W)] = E[ZE(Z'|W)] - E[E(Z|W)E(Z'|W)] = 0$ , and  $E[Z^* m(W, Z)] = E[Z^* (Z' \delta_0 + h(W))] = E(Z^* Z') \delta_0$  since  $E[Z^* h(W)] = E[Z h(W)] - E[E(Z|W)h(W)] = 0$ . Thus

$$a(m) = \delta_0.$$

Likewise, since  $\hat{\beta} = (\hat{\delta}', \hat{\lambda}')'$ ,  $\hat{m}(w, z) = z' \hat{\delta} + q(w)' \hat{\lambda}$ ,

$$a(\hat{m}) = E \{ [E(Z^* Z^{*'})]^{-1} Z^* \hat{m}(X, W) \} = [E(Z^* Z^{*'})]^{-1} [E(Z^* Z') \hat{\delta} + E[Z^* q(W)'] \hat{\lambda}] = \hat{\delta}.$$

#### 6.2. Monte Carlo study of finite sample performance

Our simulations take both  $W_i$  and  $Z_i$  in (22) to be one-dimensional, and throughout set  $\delta_0 = 0.3$  and  $h(w) = \log(1+w^2)$ . Our main focus is to investigate performance of studentizations of Section 5 when the ordering requirements of **C5** and **C5'** may be problematic due to noise in our information about the ordering. For example, even if one knows which characteristic(s) of individual units underpins the structure of spatial dependence, it may be observed with error. For both settings of Sections 5.1 and 5.2, we introduce varying degrees of perturbations to the correct

**Table 1**

Asymptotic critical values of  $W_d(1, 1)' \Phi_d^{-1} W_d(1, 1)$ .

%	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$	$d = 7$	$d = 8$	$d = 9$	$d = 10$
90	26.45	56.12	89.06	125.13	164.52	205.41	251.52	299.46	349.92	404.22
92.5	31.94	65.19	101.88	140.16	182.74	225.43	275.62	327.14	379.69	438.89
95	40.13	78.37	119.49	161.47	208.18	254.78	310.96	364.84	422.63	485.4
97.5	56.94	102.44	149.76	200.09	253.48	306.56	366.9	428.62	496.25	562.6
99	80.51	137.37	192.4	253.15	313.18	376.38	450.14	516.98	596.31	667.22
%	$d = 11$	$d = 12$	$d = 13$	$d = 14$	$d = 15$	$d = 16$	$d = 17$	$d = 18$	$d = 19$	$d = 20$
90	462.01	518.89	579.63	642.53	705.94	772.63	843.9	918.99	988.69	1066.3
92.5	497.43	556.68	620.25	688.05	754.72	824.05	897.5	975.44	1048.85	1127.71
95	548.63	609.57	677.77	750.63	820.3	897.52	973.74	1052.87	1126.82	1213.57
97.5	631.45	697.63	777.26	855.09	927.58	1013.52	1096.11	1180.25	1271.74	1353.66
99	749.3	816.72	901.14	990.75	1065.59	1162.26	1245.07	1345.38	1446.13	1535.03
%	$d = 21$	$d = 22$	$d = 23$	$d = 24$	$d = 25$	$d = 26$	$d = 27$	$d = 28$	$d = 29$	$d = 30$
90	1143.37	1228.08	1301.17	1384.64	1473.12	1556.33	1651.56	1745.87	1838.19	1931.44
92.5	1209.83	1298.81	1373.56	1460.09	1555.47	1636.52	1736.99	1836.96	1928.30	2034.14
95	1296.34	1391.67	1471.74	1567.32	1660.19	1748.68	1854.77	1959.23	2047.54	2161.92
97.5	1447.76	1544.98	1642.89	1742.46	1840.65	1944.36	2048.66	2145.17	2247.40	2367.51
99	1634.47	1752.84	1834.49	1954.47	2068.41	2180.41	2277.93	2406.17	2492.32	2630.90

ordering of data and report their effects on Monte Carlo coverage probabilities and power of test.

In the first set of simulations which is relevant to the studentization of Section 5.1, we generate random locations for individual units along a line, which determines the underlying dependence structure. We then compare the performance of our studentization under the correct ordering of data with that under a perturbed ordering, where locations are observed subject to error, but used to order the data. To be specific, the locations of the observations, denoted  $s = (s_1, \dots, s_n)'$ , were generated by a random draw from the uniform distribution over  $[0, n]$ . Keeping them fixed across replications,  $U_i$  and  $Z_i$  were generated independently as scalar normal variables with mean zero and covariances  $Cov(U_i, U_j) = Cov(Z_i, Z_j) = \rho^{|s_i - s_j|}$ , using various  $\rho \in (0, 1)$ . To construct  $W_i$ , we generate another scalar normal random variable  $V_i$  in the same way as  $U_i$  and  $Z_i$  and let  $W_i = 1 + V_i + 0.5Z_i$ . The dependent variable is then  $Y_i = \log(1 + W_i^2) + 0.3Z_i + U_i$ .

For the studentization, we add noise to the locations, to generate four sets of “perturbed” locations: defining

$$\begin{aligned} \epsilon' &= (\epsilon'_1, \dots, \epsilon'_n)' \sim N(0, 4I_n), \\ \epsilon'' &= (\epsilon''_1, \dots, \epsilon''_n)' \sim N(0, 25I_n), \\ \epsilon''' &= (\epsilon'''_1, \dots, \epsilon'''_n)' \sim N(0, 100I_n), \\ \epsilon'''' &= (\epsilon''''_1, \dots, \epsilon''''_n)' \sim N(0, 400I_n), \end{aligned}$$

we take

$$\begin{aligned} s'_i &= s_i + \epsilon'_i, & s''_i &= s_i + \epsilon''_i, & s'''_i &= s_i + \epsilon'''_i, \\ s''''_i &= s_i + \epsilon''''_i, & i &= 1, \dots, n. \end{aligned}$$

We base the studentization on 5 different orderings of the data, according to the five sets of locations  $s, s', s'', s''', s''''$ .

We consider two sample sizes,  $n = 100, 400$  and 4 levels of dependence,  $\rho = 0, 0.2, 0.4, 0.6$ . For each of the 8 combinations, three bandwidth values,  $\kappa = 4, 6, 9$  were tried, and for the series functions of  $W_i$ , the first  $\kappa - 1$  orthonormal Legendre polynomials were used. The results are based on 1000 replications.

We first analyse performance of the estimates of both the nonparametric function  $m$  and semiparametric quantity  $a(m)$ . We report in Table 2 the Monte Carlo mean squared error (MSE), biases and variances of  $\hat{m}$  at a fixed point  $x = (w, z) = (0.5, 0.5)$ , the Monte Carlo integrated MSE (MISE) of  $\hat{m}$  to convey global performance, and the MSE of  $\hat{\delta}$ . The bias and variance of  $\hat{m}(0.5, 0.5)$  are in line with the prediction that larger  $\kappa$  reduce bias while increasing variance, while under all values of  $\rho, \kappa = 4$  or  $\kappa = 6$  led to the smallest MSE for  $n = 100$ , while  $\kappa = 6$  did

**Table 2**

Monte Carlo MSE, variance and bias, design 1.

$\rho$	$n$	$\kappa$	$MSE(\hat{m}_x)$	$Var(\hat{m}_x)$	$Bias(\hat{m}_x)$	$MISE(\hat{m})$	$MSE(\hat{\delta})$	
0	100	4	0.0353	0.0283	0.0842	0.0595	0.0126	
		6	0.035	0.0347	0.017	0.0701	0.0125	
		9	0.0463	0.0463	0.0039	0.0989	0.0132	
	400	4	0.0162	0.0071	0.0956	0.0265	0.0033	
		6	0.0082	0.0079	0.0174	0.0199	0.0033	
		9	0.0098	0.0098	-0.0024	0.025	0.0034	
	0.2	100	4	0.0526	0.0453	0.0855	0.0863	0.0216
			6	0.055	0.0546	0.0201	0.0992	0.022
			9	0.0671	0.067	0.0066	0.1261	0.0229
400		4	0.0219	0.0121	0.099	0.033	0.005	
		6	0.0141	0.0135	0.0254	0.0278	0.0051	
		9	0.0151	0.0151	0.0041	0.0334	0.0051	
0.4		100	4	0.0693	0.0647	0.0674	0.106	0.0268
			6	0.0757	0.0756	0.005	0.1207	0.0273
			9	0.0915	0.0915	-0.002	0.1493	0.0278
	400	4	0.025	0.0148	0.1014	0.0394	0.0065	
		6	0.0175	0.0168	0.0265	0.0347	0.0065	
		9	0.0193	0.0192	0.0058	0.0404	0.0065	
	0.6	100	4	0.0863	0.0809	0.0738	0.1326	0.0341
			6	0.0861	0.0859	0.0112	0.1465	0.0348
			9	0.1028	0.1028	-0.0013	0.1739	0.0358
400		4	0.034	0.0253	0.0931	0.0517	0.0107	
		6	0.0272	0.0267	0.0222	0.0481	0.0107	
		9	0.0301	0.0301	-0.0006	0.0542	0.0107	

so for  $n = 400$ . For the MISE,  $\kappa = 4$  was best when  $n = 100$ , and  $\kappa = 6$  was best for  $n = 400$ . The Monte Carlo MSE of  $\hat{\delta}$  was relatively insensitive to  $\kappa$  across all 8 settings, which is important as optimal choice of  $\kappa$  for semiparametric estimation is often more ambiguous than for nonparametric estimation.

Our next objective is to investigate performance of the studentization of Section 5.1. Theorem 3 implies in this setting,

$$\sqrt{n/\hat{B}_n}(\hat{\delta} - \delta_0) \rightarrow_d W_1(1)/\sqrt{\Psi_1}, \quad \text{as } n \rightarrow \infty.$$

Kiefer et al. (2000)'s Table 2 gave simulated values of the percentiles of  $W_1^2(1)/\Psi_1$ , from which we derive the 99.5th, 97.5th and 95th percentiles of  $W_1(1)/\sqrt{\Psi_1}$  as  $\sqrt{101.2}, \sqrt{46.39}$  and  $\sqrt{28.88}$ , respectively. We thence construct asymptotic 95% confidence intervals for  $\delta_0$ ,  $\left[ \hat{\delta} - \sqrt{46.39\hat{C}_n/n}, \hat{\delta} + \sqrt{46.39\hat{C}_n/n} \right]$ .

Table 3 reports the Monte Carlo average length of this interval based on correctly ordering the data according to  $s$ . The length decreases with  $n$  and increases with  $\rho$  and is fairly insensitive to  $\kappa$ . Similar patterns are observed under perturbed ordering.

**Table 3**  
Monte Carlo average 95% CI length for  $\delta_0$ , design 1.

$n$	$\kappa$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$
100	4	0.5605	0.6746	0.7447	0.8328
	6	0.5608	0.6701	0.7401	0.8276
	9	0.5608	0.6736	0.7353	0.8224
400	4	0.2955	0.3519	0.4043	0.4889
	6	0.2933	0.3501	0.4039	0.4874
	9	0.2922	0.3489	0.402	0.4869

Table 4 reports empirical coverage probabilities for the 99%, 95% and 90% confidence intervals under the five different orderings of data, based on locations  $s, s', s'', s'''$ , and  $s''''$ . When  $\rho = 0$ , studentizations with all orderings produce a rather precise coverage probabilities for both samples sizes. For  $\rho = 0.2, 0.4, 0.6$  and correct ordering based on  $s$ , the coverage probabilities suffer slightly in the smallest sample  $n = 100$ , while being fairly good for  $n = 400$ , at least for  $\rho = 0.2$  and  $0.4$ . The more we perturb the ordering, a gradual deterioration is reported. Nevertheless, even with the greatest perturbations, caused by substantial noises  $Var(\epsilon_i''''') = 100$  and  $Var(\epsilon_i''''') = 400$ , the results are encouraging.

Table 5 reports empirical power of testing  $H_0 : \delta_0 = \delta$  against  $H_1 : \delta_0 \neq \delta$ , for  $\delta = 0.3, 0.4, 0.5, 0.7$ ; of course columns corresponding to  $\delta = 0.3$  report empirical size. Not surprisingly, for  $\rho = 0$ , powers across different orderings are similar, while for  $\rho = 0.2, 0.4$  and  $0.6$ , power tends to improve with increasing perturbations.

**Table 4**  
Coverage probabilities for  $\delta_0$ , design 1.

$\rho$	$n$	$\kappa$	$s$			$s'$			$s''$			$s'''$			$s''''$			
			0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	
0	100	4	0.894	0.951	0.987	0.891	0.942	0.987	0.897	0.936	0.991	0.889	0.937	0.989	0.892	0.94	0.987	
		6	0.897	0.944	0.99	0.891	0.947	0.989	0.892	0.942	0.99	0.891	0.937	0.99	0.895	0.942	0.985	
		9	0.896	0.938	0.989	0.891	0.94	0.989	0.889	0.937	0.984	0.878	0.933	0.985	0.882	0.94	0.984	
	400	4	0.908	0.947	0.988	0.91	0.947	0.989	0.909	0.948	0.989	0.906	0.95	0.989	0.907	0.95	0.991	
		6	0.897	0.947	0.991	0.897	0.947	0.991	0.891	0.944	0.992	0.897	0.955	0.99	0.902	0.951	0.992	
		9	0.89	0.952	0.991	0.889	0.951	0.988	0.891	0.944	0.99	0.891	0.947	0.989	0.897	0.943	0.992	
	0.2	100	4	0.865	0.939	0.979	0.863	0.927	0.98	0.859	0.922	0.975	0.836	0.915	0.972	0.829	0.898	0.971
			6	0.862	0.926	0.984	0.855	0.919	0.984	0.849	0.916	0.979	0.837	0.913	0.978	0.828	0.899	0.97
			9	0.852	0.917	0.984	0.853	0.908	0.981	0.848	0.906	0.978	0.845	0.902	0.972	0.827	0.892	0.973
400		4	0.89	0.951	0.989	0.889	0.951	0.988	0.881	0.946	0.986	0.876	0.941	0.983	0.868	0.936	0.986	
		6	0.889	0.948	0.988	0.889	0.946	0.987	0.881	0.941	0.988	0.877	0.938	0.985	0.866	0.934	0.983	
		9	0.883	0.943	0.986	0.885	0.94	0.986	0.876	0.934	0.986	0.872	0.934	0.984	0.863	0.929	0.983	
0.4		100	4	0.863	0.918	0.976	0.858	0.916	0.97	0.837	0.904	0.971	0.825	0.898	0.968	0.807	0.881	0.959
			6	0.866	0.916	0.971	0.858	0.909	0.969	0.846	0.91	0.966	0.827	0.891	0.964	0.797	0.877	0.962
			9	0.864	0.913	0.973	0.86	0.907	0.972	0.841	0.901	0.96	0.84	0.901	0.969	0.796	0.868	0.967
	400	4	0.887	0.941	0.992	0.885	0.938	0.99	0.88	0.933	0.986	0.878	0.931	0.986	0.866	0.916	0.977	
		6	0.893	0.935	0.992	0.889	0.933	0.99	0.885	0.93	0.99	0.88	0.929	0.991	0.868	0.912	0.975	
		9	0.891	0.939	0.992	0.891	0.936	0.991	0.887	0.931	0.99	0.879	0.928	0.99	0.865	0.91	0.975	
	0.6	100	4	0.863	0.93	0.974	0.855	0.923	0.974	0.846	0.921	0.963	0.823	0.894	0.961	0.804	0.872	0.951
			6	0.865	0.919	0.978	0.864	0.915	0.976	0.84	0.912	0.967	0.821	0.893	0.956	0.799	0.874	0.948
			9	0.86	0.914	0.979	0.856	0.907	0.977	0.846	0.902	0.971	0.816	0.88	0.958	0.796	0.862	0.945
400		4	0.866	0.923	0.977	0.861	0.921	0.977	0.858	0.915	0.975	0.846	0.913	0.973	0.835	0.905	0.96	
		6	0.872	0.927	0.98	0.869	0.923	0.981	0.864	0.919	0.978	0.861	0.912	0.975	0.837	0.897	0.96	
		9	0.869	0.93	0.981	0.867	0.928	0.978	0.864	0.922	0.977	0.856	0.913	0.975	0.84	0.901	0.966	

**Table 5**  
Empirical power of 5% test on  $\delta_0$ ,  $\kappa = 6$ , design 1.

	$\rho \setminus \delta$	$s$				$s'$				$s''$				$s'''$				$s''''$			
		0.3	0.4	0.5	0.7	0.3	0.4	0.5	0.7	0.3	0.4	0.5	0.7	0.3	0.4	0.5	0.7	0.3	0.4	0.5	0.7
$n = 100$	0	0.049	0.113	0.319	0.778	0.058	0.112	0.32	0.781	0.064	0.119	0.314	0.779	0.063	0.121	0.32	0.781	0.06	0.12	0.31	0.779
	0.2	0.061	0.131	0.275	0.656	0.073	0.141	0.28	0.667	0.078	0.142	0.287	0.679	0.085	0.139	0.314	0.695	0.102	0.163	0.313	0.726
	0.4	0.082	0.127	0.245	0.597	0.084	0.131	0.246	0.607	0.096	0.139	0.273	0.635	0.102	0.167	0.28	0.642	0.119	0.174	0.331	0.683
	0.6	0.07	0.108	0.229	0.542	0.077	0.115	0.236	0.55	0.079	0.136	0.263	0.587	0.106	0.158	0.293	0.601	0.128	0.187	0.304	0.639
$n = 400$	0	0.053	0.287	0.751	0.994	0.053	0.287	0.749	0.994	0.052	0.283	0.751	0.994	0.05	0.288	0.754	0.995	0.05	0.29	0.756	0.994
	0.2	0.049	0.219	0.624	0.971	0.049	0.22	0.622	0.971	0.054	0.228	0.633	0.969	0.059	0.224	0.632	0.968	0.064	0.243	0.648	0.971
	0.4	0.059	0.177	0.508	0.942	0.062	0.183	0.507	0.943	0.067	0.188	0.517	0.946	0.069	0.193	0.531	0.946	0.084	0.209	0.546	0.952
	0.6	0.077	0.164	0.387	0.846	0.079	0.169	0.394	0.848	0.085	0.17	0.401	0.85	0.087	0.174	0.406	0.857	0.095	0.191	0.441	0.872

The second set of simulations considers the studentization of Section 5.2. To generate the data, we follow the random location setting of Robinson and Thawornkaiwong (2012), where the locations of the observations, denoted  $s_1, \dots, s_n$ , were generated by a random draw from the uniform distribution over  $[0, 2n^{1/2}] \times [0, 2n^{1/2}]$ . Keeping these locations fixed across replications,  $U_i$  and  $Z_i$  were generated independently as scalar normal random variables with mean zero and covariances  $Cov(U_i, U_j) = Cov(Z_i, Z_j) = \rho^{\|s_i - s_j\|}$ . We generated  $W_i$  and  $Y_i$ , and used the same  $\kappa$ , series functions and number of replications, as before, but took  $\rho = 0, 0.2, 0.4, 0.52$  for  $n = 100$  and  $\rho = 0, 0.2, 0.35, 0.5$  for  $n = 400$ . The random location setting implies the degree of dependence is determined not only by  $\rho$ , but also by the distances between locations. The fact that we are considering locations on a plane rather than along a line implies that  $\rho$  produces differing strengths of dependence compared to the familiar time series AR(1) model, making it difficult to get a sense of the degree of dependence in the data generated. One measure of dependence that might be used in comparisons is  $\sum_{i,j=1}^n |Cov(U_i, U_j)|$ . Our choices of  $\rho$  led this quantity to be of similar magnitude to that in the AR(1) model with lag-1 autocorrelation  $\varphi = 0, 0.2, 0.4, 0.6$ : for  $n = 100$ , in our spatial setting it took values 100, 156, 246, 401 for  $\rho = 0, 0.2, 0.4, 0.52$ , respectively, which are comparable to 100, 150, 232, 396 in the time series AR(1) with  $\varphi = 0, 0.2, 0.4, 0.6$ ; for  $n = 400$ , it took values 400, 618, 932, 1597 for  $\rho = 0, 0.2, 0.35, 0.5$ , respectively, which are comparable to 400, 599, 930, 1590 in the time series AR(1) with  $\varphi = 0, 0.2, 0.4, 0.6$ .

**Table 6**  
Monte Carlo MSE, variance and bias, design 2.

$\rho$	$n$	$\kappa$	MSE( $\hat{m}_x$ )	Var( $\hat{m}_x$ )	Bias( $\hat{m}_x$ )	MISE( $\hat{m}$ )	MSE( $\hat{\delta}$ )	
0	100	4	0.1851	0.0239	0.4016	0.1923	0.0151	
		6	0.0337	0.0273	0.08	83.4583	0.0133	
		9	0.037	0.0368	0.0158	83.4967	0.0137	
	400	4	0.173	0.0062	0.4084	0.1787	0.0038	
		6	0.017	0.0071	0.0998	365.4258	0.0032	
		9	0.0085	0.0081	0.0213	365.4463	0.0032	
	0.2	100	4	0.1936	0.0325	0.4014	0.1995	0.0165
			6	0.0406	0.0339	0.0815	83.4625	0.0151
			9	0.0441	0.044	0.0103	83.5007	0.0154
400		4	0.1741	0.0077	0.4078	0.1804	0.0039	
		6	0.0184	0.0087	0.0989	365.4674	0.0036	
		9	0.0113	0.0108	0.0226	365.487	0.0036	
0.4		100	4	0.1937	0.0423	0.3892	0.2094	0.019
			6	0.0511	0.0456	0.0745	83.4556	0.0173
			9	0.0549	0.0549	0.0029	83.4942	0.0176
0.35	400	4	0.1766	0.0109	0.4071	0.1815	0.0048	
		6	0.0215	0.0114	0.1003	365.3836	0.0041	
		9	0.0122	0.0117	0.0233	365.4038	0.004	
0.52	100	4	0.2195	0.0662	0.3916	0.2288	0.0245	
		6	0.0735	0.0666	0.0826	83.519	0.023	
		9	0.0766	0.0765	0.0109	83.5562	0.0231	
0.5	400	4	0.181	0.0164	0.4057	0.1858	0.0062	
		6	0.0265	0.0164	0.1007	365.3476	0.0053	
		9	0.0167	0.0161	0.0228	365.3677	0.0053	

**Table 7**  
Monte Carlo average 95% CI length for  $\delta_0$ , design 2.

$n$	$\kappa$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.4, 0.35$	$\rho = 0.52, 0.5$
100	4	0.5696	0.596	0.6135	0.6602
	6	0.5304	0.5582	0.5748	0.6193
	9	0.5348	0.5606	0.5785	0.6202
400	4	0.2853	0.2982	0.3108	0.3379
	6	0.2634	0.2765	0.2905	0.3186
	9	0.262	0.2754	0.2896	0.3178

**Table 8**  
Coverage probabilities for  $\delta_0$ , design 2.

$\rho$	$n$	$\kappa$	$s$			$s'$			$s''$			$s'''$			$s''''$			
			0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	
0	100	4	0.892	0.943	0.992	0.898	0.947	0.991	0.889	0.957	0.992	0.896	0.954	0.992	0.909	0.948	0.988	
		6	0.89	0.951	0.987	0.896	0.95	0.99	0.9	0.951	0.988	0.895	0.947	0.989	0.901	0.953	0.988	
		9	0.899	0.952	0.988	0.904	0.958	0.988	0.9	0.944	0.985	0.906	0.946	0.987	0.904	0.954	0.986	
	400	4	0.91	0.962	0.987	0.913	0.954	0.99	0.909	0.961	0.991	0.905	0.96	0.989	0.901	0.956	0.991	
		6	0.912	0.956	0.987	0.91	0.955	0.992	0.915	0.954	0.99	0.91	0.951	0.99	0.902	0.959	0.988	
		9	0.914	0.961	0.987	0.906	0.958	0.99	0.908	0.959	0.989	0.908	0.95	0.99	0.908	0.961	0.99	
	0.2	100	4	0.889	0.938	0.99	0.897	0.942	0.986	0.889	0.938	0.987	0.891	0.94	0.979	0.885	0.938	0.985
			6	0.876	0.937	0.99	0.882	0.933	0.988	0.873	0.932	0.985	0.875	0.929	0.981	0.883	0.94	0.983
			9	0.886	0.935	0.988	0.884	0.941	0.986	0.867	0.934	0.989	0.879	0.934	0.978	0.889	0.938	0.989
400		4	0.916	0.954	0.991	0.904	0.949	0.991	0.911	0.959	0.989	0.899	0.94	0.985	0.896	0.946	0.988	
		6	0.907	0.952	0.987	0.897	0.946	0.985	0.902	0.954	0.988	0.881	0.932	0.985	0.888	0.944	0.986	
		9	0.907	0.951	0.988	0.902	0.948	0.983	0.909	0.957	0.99	0.883	0.934	0.983	0.887	0.938	0.988	
0.4		100	4	0.878	0.937	0.985	0.864	0.938	0.987	0.878	0.931	0.982	0.872	0.929	0.985	0.861	0.927	0.983
			6	0.876	0.928	0.983	0.867	0.928	0.984	0.867	0.92	0.977	0.86	0.921	0.983	0.859	0.929	0.985
			9	0.881	0.937	0.983	0.882	0.93	0.98	0.867	0.927	0.973	0.863	0.919	0.975	0.857	0.92	0.981
0.35	400	4	0.893	0.953	0.99	0.893	0.944	0.988	0.876	0.943	0.989	0.879	0.934	0.986	0.86	0.929	0.982	
		6	0.888	0.95	0.991	0.885	0.94	0.993	0.875	0.935	0.991	0.869	0.928	0.984	0.866	0.934	0.983	
		9	0.893	0.953	0.99	0.89	0.939	0.991	0.876	0.935	0.992	0.872	0.927	0.985	0.877	0.93	0.981	
0.52	100	4	0.873	0.928	0.978	0.845	0.91	0.974	0.829	0.906	0.975	0.826	0.898	0.969	0.821	0.883	0.966	
		6	0.857	0.911	0.981	0.83	0.901	0.973	0.81	0.884	0.97	0.8	0.877	0.962	0.803	0.876	0.961	
		9	0.849	0.916	0.981	0.835	0.902	0.973	0.815	0.896	0.969	0.807	0.878	0.958	0.808	0.883	0.967	
0.5	400	4	0.876	0.938	0.989	0.86	0.924	0.985	0.853	0.915	0.978	0.841	0.904	0.977	0.829	0.903	0.974	
		6	0.877	0.932	0.985	0.865	0.921	0.987	0.844	0.912	0.976	0.831	0.903	0.973	0.835	0.904	0.973	
		9	0.875	0.933	0.986	0.862	0.922	0.987	0.839	0.911	0.977	0.833	0.904	0.97	0.835	0.905	0.971	

We used the two dimensional location coordinate  $s_i = (s_{1i}, s_{2i})$  to order the data with two indices. The first coordinate  $s_{1i}$  was used to divide the sample into  $n^{1/2}$  groups numbered by the first index and the second coordinate  $s_{2i}$  was then used to order within each group, generating the second index. For perturbations to the ordering, we added two dimensional error terms to  $s_i$ , leading to  $s'_i, s''_i, s'''_i, s''''_i$ , respectively:

$$\begin{aligned} \epsilon'_i &\sim N(0, 4I_2), & \epsilon''_i &\sim N(0, 25I_2), \\ \epsilon'''_i &\sim N(0, 100I_2), & \epsilon''''_i &\sim N(0, 400I_2), \end{aligned}$$

where  $\epsilon'_i, \epsilon''_i, \epsilon'''_i, \epsilon''''_i$  are independent across  $i$ . It is worth noting that due to the difference in the settings, above errors represent more significant perturbations than in the previous setting, since the variance of  $s_{1i}$  and  $s_{2i}$  in the current set-up is smaller than the variance of  $s_i$  used in the previous one.

We report in Table 6 the Monte Carlo MSE, bias and variance of  $\hat{m}(0.5, 0.5)$ , MISE of  $\hat{m}$ , and MSE of  $\hat{\delta}_0$ . Again, patterns of bias and variance of  $\hat{m}$  with changing  $\kappa$  are in line with predictions, and  $\kappa = 6, 9$  tended to generate the lowest MSE for settings with  $n = 100, 400$ , respectively. Table 7 reports Monte Carlo average length of 95% confidence intervals. As before, it decreases with  $n$ , increases with  $\rho$  and shows little variation across  $\kappa$ . Table 8 reports the empirical coverage probabilities for the 99%, 95% and 90% confidence intervals based on critical values of Table 1. As expected, adverse impacts of perturbed ordering increase with  $\rho$  and the magnitude of perturbations, although the results are again encouraging even for the greatest perturbations and  $\rho = 0.52, 0.5$ . Table 9 reports empirical power of testing  $H_0 : \delta_0 = \delta$  against  $H_1 : \delta_0 \neq \delta$ , for  $\delta = 0.3, 0.4, 0.5, 0.7$  with 5% significance level. Again, results are similar to those of the previous setting, with power improving with increasing perturbations for  $\rho \neq 0$ , and powers across different orderings are similar when  $\rho = 0$ .

### 6.3. Empirical examples

We apply our methodology in two illustrative empirical examples, using (22) with data from Yatchew (2003). Series

**Table 9**  
Empirical power of 5% test on  $\delta_0$ ,  $\kappa = 6$ , design 2.

	$\rho \setminus \delta$	$s$				$s'$				$s''$				$s'''$				$s''''$			
		0.3	0.4	0.5	0.7	0.3	0.4	0.5	0.7	0.3	0.4	0.5	0.7	0.3	0.4	0.5	0.7	0.3	0.4	0.5	0.7
$n = 100$	0	0.049	0.117	0.333	0.829	0.05	0.117	0.327	0.833	0.049	0.114	0.327	0.83	0.053	0.124	0.322	0.836	0.3	0.4	0.5	0.7
	0.2	0.063	0.123	0.295	0.779	0.067	0.134	0.314	0.79	0.068	0.127	0.32	0.803	0.071	0.148	0.344	0.818	0.06	0.132	0.339	0.799
	0.4	0.072	0.132	0.315	0.769	0.072	0.137	0.301	0.783	0.08	0.137	0.339	0.79	0.079	0.15	0.334	0.794	0.071	0.153	0.358	0.812
	0.52	0.089	0.124	0.267	0.678	0.099	0.144	0.283	0.714	0.116	0.158	0.327	0.733	0.123	0.175	0.333	0.75	0.124	0.187	0.361	0.76
$n = 400$	0	0.044	0.32	0.846	0.999	0.045	0.325	0.852	0.999	0.046	0.331	0.838	0.999	0.049	0.348	0.864	0.998	0.041	0.355	0.842	0.998
	0.2	0.048	0.323	0.823	1	0.054	0.342	0.81	1	0.046	0.342	0.823	1	0.068	0.361	0.821	1	0.056	0.346	0.845	1
	0.35	0.05	0.287	0.779	1	0.06	0.295	0.791	1	0.065	0.318	0.789	1	0.072	0.365	0.808	0.999	0.066	0.347	0.81	1
	0.5	0.068	0.249	0.692	0.998	0.079	0.266	0.714	1	0.088	0.293	0.726	0.999	0.097	0.342	0.773	0.999	0.096	0.36	0.783	0.999

**Table 10**  
Cost function in electricity distribution.

	Kernel			Series			$t_{n,w}^*$	$t_{n,e}^*$
	Coef	SE	t-stat	Coef	SE	t-stat		
<i>wage</i>	-6.298	12.453	-0.506	-6.002	15.736	-0.381	0.426	0.261
<i>pcap</i>	-1.393	1.6	-0.872	-2.531	1.846	-1.371	44.08 $\Delta$	35.433 $\Delta$
$\frac{1}{2}wage^2$	0.72	2.13	0.3388	1.731	12.837	0.135	0.061	0.036
$\frac{1}{2}pcap^2$	0.032	0.066	0.485	0.148	0.318	0.466	1.593	1.491
<i>wage</i> · <i>pcap</i>	0.534	0.599	0.891	2.044	1.553	1.317	43.155 $\Delta$	40.27 $\Delta$
<i>PUC</i>	-0.086	0.039	-2.205*	-0.043	0.017	-2.6*	11.042	28.893 $\Delta$
<i>kwh</i>	0.033	0.086	0.384	0.0828	0.102	0.8085	8.208	9.486
<i>life</i>	-0.634	0.115	-5.513*	-0.613	0.124	-4.935*	104.6*	92.7*
<i>lf</i>	1.249	0.436	2.865*	0.746	0.486	1.535	39.669 $\Delta$	36.587 $\Delta$
<i>kmwire</i>	0.399	0.087	4.586*	0.442	0.088	5.012*	202.65*	151.02*

estimation yields similar estimates of  $\delta_0$  to the kernel ones in Yatchew (2003). For the hypothesis  $H_0 : \delta_{0\ell} = 0$  against  $H_1 : \delta_{0\ell} \neq 0$ ,  $\ell = 1, \dots, d$ , the test that assumes independence of disturbances is contrasted with that based on an extension of Theorem 3 or 4, which allow for spatial dependence. Denoting by  $\hat{\delta}_\ell$  the  $\ell$ th element of  $\hat{\delta}$  and  $\hat{B}_{\ell n}, \hat{C}_{\ell n}$  the  $(\ell, \ell)$ th element of  $\hat{B}_n, \hat{C}_n$ , respectively, the statistic  $t_n^* = n\hat{\delta}_\ell^2 / \hat{B}_{\ell n}$  and  $t'_n = n\hat{\delta}_\ell^2 / \hat{C}_{\ell n}$  satisfy:

$$t_n^* \rightarrow_d W_1^2(1) / \Psi_1 \text{ under } H_0; t_n^* \rightarrow_p \infty \text{ under } H_1,$$

$$t'_n \rightarrow_d W_1^2(1, 1) / \Phi_1 \text{ under } H_0; t'_n \rightarrow_p \infty \text{ under } H_1$$

(cf Kiefer et al. (2000, pp. 712–713)).

The first empirical example concerns the following cost function of distributing electricity, also from Yatchew (2003):

$$tc = \delta_1 wage + \delta_2 pcap + \frac{\delta_3}{2} wage^2 + \frac{\delta_4}{2} pcap^2 + \delta_5 wage \cdot pcap + \delta_6 PUC + \delta_7 kwh + \delta_8 life + \delta_9 lf + \delta_{10} kmwire + h(cust) + u.$$

The dependent variable,  $tc$ , is the log total cost per customer. The parametrically involved regressors are *wage* (log wage rate), *pcap* (log price of capital), *PUC* (a dummy for public utility commissions that deliver additional services, and therefore may benefit from economies of scale), *life* (log of the remaining life of distribution assets), *lf* (log of the load factor, measuring capacity utilization relative to peak usage), and *kmwire* (log of kilometres of distribution wire per customer). The nonparametrically involved regressor is *cust* (log of the number of customers). The disturbance is now denoted  $u$ . Yatchew (2003) was interested in estimating the conditional expectation of  $tc$  given  $cust$ , holding the other regressors fixed, as the shape of this curve reveals whether there are increasing/decreasing returns to scale in electricity distribution. We are interested in estimating  $\delta$  and testing their significance,  $H_0 : \delta_l = 0$ , versus  $H_1 : \delta_l \neq 0$  for  $l = 1, \dots, d$ , when allowing for dependence in the disturbance  $u$ . The data consists of 81 municipal distributors in Ontario, Canada in 1993.

The first set of columns of Table 10 repeats the kernel estimates of  $\delta$  and their standard errors, assuming uncorrelatedness of error terms, from Yatchew (2003) based on methods and theory of

Robinson (1988). The second set of columns reports the  $\hat{\delta}_i$ , using the first three Legendre polynomials in series estimation. Test statistics labelled \* are 5% significant, while those labelled  $\Delta$  are 10% significant. To apply the studentization of Section 5.1, two different orderings were tried. First, the data were ordered in the ascending wage rate faced by the firm, with the rationale that firms may be subject to input shocks, and those with similar wage rate may use similar inputs, leading to dependence in disturbances. Test statistics based on this studentization are denoted  $t_{n,w}^*$ . Second, the data were ordered according to the number of employees of the firm, which is a measure of size, noting that firms of similar size may be subject to similar shocks, or alternatively, may be dependent due to competition. Test statistics based on this studentization are denoted  $t_{n,e}^*$ . Inference based on the assumption of uncorrelated disturbances found *PUC*, *life*, *lf* and *kmwire* to be 5% significant using kernel estimation, while *PUC*, *life* and *kmwire* are 5% significant with series estimation. When allowing for dependence in disturbances, and with both orderings, *life* and *kmwire* were still found to be 5% significant, while *lf*, *pcap* and *wage* · *pcap* were 10% significant and *PUC*, which was 5% significant under uncorrelatedness, was 10% significant based on ordering according to number of employees.

The second example involves hedonic pricing of housing attributes, the data concerning 92 detached homes in Ottawa sold during 1987. The dependent variable is the sale price of a given house (*price*), while the parametrically involved regressors consist of attributes of the house, including lot size (*lotarea*), square footage of housing (*usespc*), number of bedrooms (*nrbed*), average neighbourhood income (*acginc*), distance to highway (*dhwy*), presence of garage (*grge*), fireplace (*frplc*), and luxury bathroom (*lux*). The nonparametric function  $h(\cdot)$  has two arguments, being location coordinates  $s = \text{south}$  and  $w = \text{west}$ :

$$price = h(s, w) + \delta_1 frplc + \delta_2 grge + \delta_3 lux + \delta_4 acginc + \delta_5 dhwy + \delta_6 lotarea + \delta_7 nrbed + \delta_8 usespc + u.$$

In series estimation of  $h(s, w)$ , we used approximating functions  $(1, s, w, sw)$ . The first set of columns of Table 11 recalls the kernel estimates reported in Yatchew (2003), and the second set

**Table 11**  
Hedonic house pricing.

Variable	Kernel			Series			
	Coef	SE	t-stat	Coef	SE	t-stat	$t'_n$
frplc	12.6	5.8	2.17*	12.7	5.62	2.26*	42.40*
grge	12.9	4.9	2.63*	12.8	4.31	2.97*	80.88*
lux	57.6	10.6	5.43*	58.2	11.3	5.15*	260.74*
acginc	0.6	0.23	2.61*	0.61	0.2	3.08*	29.64 $\Delta$
dhwy	1.5	21.4	0.07	-9.2	5.86	-1.57	21.32
lotarea	3.1	2.2	1.41	3.8	1.85	2.03*	31.56 $\Delta$
nrbed	6.4	4.8	1.33	7.8	4.2	1.85 $\Delta$	19.51
usespc	24.7	10.6	2.33*	23.6	11.6	2.04*	53.13*

reports the corresponding series estimation results. The estimates of coefficients, their standard errors and the  $t$ -statistics are broadly similar, revealing significance of many of the regressors at the 5% level.

In applying the studentization of Section 5.2, we divided the north axis to 9 intervals, the first two containing 11 units each and the rest 10 units, anticipating possible spatial dependence in disturbances of neighbouring houses. In Table 11, SE refers to standard error computed under the assumption of independence, and the test statistic  $t'_n$  defined above has critical values 40.3, 26.45 at sizes 5% and 10%, respectively. Again, test statistics labelled \* are significant at 5% level and those with  $\Delta$  at 10% level. Our tests, which attempt to account for dependence, find that the presence of fireplace, garage, luxury bathroom and square footage are 5% significant, while average neighbourhood income and lot size are 10% significant. The contrasting conclusions on the significance of  $\delta$  estimates between the test under independent errors (denoted  $t$ -stat) and the test  $t'_n$  allowing for dependence may be due to cross-sectional dependence in the data, as seems natural, given that prices of houses of the same type, sold in the same year and city, would have been subject to an overlapping set of demand and supply side factors, driven by the same macroeconomic fundamentals.

### 7. Conclusion

This paper has established a theoretical background for series estimation of a vector-valued functional of the nonparametric regression function under cross-sectional dependence and heterogeneity. Theoretical results include a mean square rate of convergence and asymptotic normality. Robust data-driven studentizations that require a suitable ordering of observations across either one or two dimensions offer an alternative to existing methods of inference. The framework of cross-sectional dependence and heterogeneity of this paper and its technical arguments may be used to establish asymptotic theory in some other settings.

### Acknowledgements

We are extremely grateful to two referees for numerous penetrating and constructive comments, which have led to many corrections and improvements. Second author's research was supported by ESRC Grant ES/J007242/1.

### Appendix A. Proofs of Theorems 1–4

**Proof of Theorem 1.** This follows that of Theorem 1 in Newey (1997), with modifications arising from relaxing his iid setting. Throughout the proofs write  $p(\cdot) = p^\kappa(\cdot)$ . We have  $E(\hat{m}(X_i) - m(X_i))^2 \leq 2[E(p(X_i)'(\hat{\beta} - \beta))^2 + E(p(X_i)'\beta - m(X_i))^2]$ , whence (13) follows from Lemma 1 and A2 on showing that

$$\|\hat{\beta} - \beta\| = O_p(\chi^{1/2} + \kappa^{-\alpha}). \quad (23)$$

As in Newey (1997), if we replace  $p(\cdot)$  by the normalized  $E(Q)^{-1/2}p$ , we can take  $E(Q) = I_\kappa$ , where  $I_r$  denotes the  $r$ -rowed identity matrix. Though this transformation affects  $\beta$  and  $\hat{\beta}$ , we leave the notation unchanged because  $\hat{m}(x)$  is unchanged. To prove (23), given (7) and (3), put  $1_n = l(\hat{\lambda}(Q) \geq 1/2)$ , where  $l(\cdot)$  is the indicator function. It will be shown that

$$1 - 1_n = o_p(1) \text{ as } n \rightarrow \infty, \quad (24)$$

whence  $Q^{-1}$  exists with probability approaching 1. Writing  $M = (m(X_1), \dots, m(X_n))'$ ,  $1_n(\hat{\beta} - \beta) = 1_n[Q^{-1}P'(Y - M)/n + Q^{-1}P'(M - P\beta)/n]$ , by elementary inequalities

$$\begin{aligned} \|1_n(\hat{\beta} - \beta)\| &\leq \|1_n Q^{-1}\| \|P'U/n\| \\ &\quad + \|1_n Q^{-1}P'/\sqrt{n}\| \|(M - P\beta)/\sqrt{n}\| = O_p(\chi^{1/2} + \kappa^{-\alpha}), \end{aligned} \quad (25)$$

since  $\|1_n Q^{-1}\| \leq \bar{\lambda}(Q^{-1}) \rightarrow_p 1$ , and thus  $\|1_n Q^{-1}P'/\sqrt{n}\|^2 = \|1_n Q^{-1}P'^{-1}/n\| = \|1_n Q^{-1}\| \leq \bar{\lambda}(Q^{-1}) = O_p(1)$ ;  $\|(M - P\beta)/\sqrt{n}\|^2 = n^{-1} \sum_{i=1}^n (m(X_i) - p(X_i)'\beta)^2 = O_p(\kappa^{-2\alpha})$  by Lemma 1 and  $\|P'U/n\| = O_p(\chi^{1/2})$ , the final statement following from

$$\begin{aligned} E\|P'U/n\|_2^2 &= \frac{1}{n^2} \sum_{i,j=1}^n \gamma_{ij} E(\sigma(X_i)\sigma(X_j)p(X_i)p'(X_j)) \\ &\leq \frac{1}{n^2} \sum_{i,j=1}^n |\gamma_{ij}| \left( E\sigma^4(X_i) E\sigma^4(X_j) E\|p(X_i)\|^4 E\|p(X_j)\|^4 \right)^{1/4} \\ &\leq \frac{C}{n^2} \sum_{i,j=1}^n |\gamma_{ij}| \left( E\|p(X_i)\|^4 \right)^{1/2} \\ &\leq C\xi\omega (Etr(Q))^{1/2} \leq C\xi\kappa^{1/2}\omega, \end{aligned} \quad (26)$$

which can be improved to

$$E\|P'U/n\|_2^2 \leq \frac{C}{n^2} \sum_{i,j=1}^n |\gamma_{ij}| E\|p(X_j)\|^2 \leq C\kappa\omega$$

when  $\sigma$  is bounded. Thus (23) follows from (24). To prove (24), for any  $\varepsilon > 0$ ,

$$\begin{aligned} P(1 - 1_n > \varepsilon) &\leq P(\hat{\lambda}(Q) < 1/2) \leq P(|\hat{\lambda}(Q) - 1| > 1/2) \\ &\leq 4E(\hat{\lambda}(Q) - 1)^2 \leq 4E\|Q - I_\kappa\|_2^2, \end{aligned}$$

since  $\hat{\lambda}(Q) - 1 = \hat{\lambda}(Q - I_\kappa)$  and where  $\|A\|_2 = tr^{1/2}(A'A)$  denotes the Euclidean norm of a matrix  $A$ . We show that

$$E\|Q - I_\kappa\|_2^2 \leq C\kappa\xi^2(n^{-1} + \Delta) \rightarrow 0. \quad (27)$$

Denoting  $\psi_{p\ell}^{(i,j)} = Cov\{p_p(X_i)p_\ell(X_i), p_p(X_j)p_\ell(X_j)\}$ , the left side of (27) is  $a_1 + a_2$  where

$$a_1 = \frac{1}{n^2} \sum_{p,\ell=1}^{\kappa} \sum_{i=1}^n \psi_{p\ell}^{(i,i)}, \quad a_2 = \frac{1}{n^2} \sum_{p,\ell=1}^{\kappa} \sum_{i,j=1, j \neq i}^n \psi_{p\ell}^{(i,j)}.$$

By elementary inequalities

$$\begin{aligned} |\psi_{p\ell}^{(i,i)}| &\leq \left( \sup_x p_p^2(x) \sup_x p_\ell^2(x) E p_p^2(X_i) E p_\ell^2(X_i) \right)^{1/2} \\ &\leq C \left( \sup_x p_p^2(x) E p_\ell^2(X_i) + \sup_x p_\ell^2(x) E p_p^2(X_i) \right) \end{aligned}$$

and thus, applying A4,

$$a_1 \leq \frac{C}{n^2} \sup_x \|p^\kappa(x)\|^2 \sum_{i=1}^n E\|p(X_i)\|^2 \leq C\kappa\xi^2/n = o(1),$$

while

$$\begin{aligned} \psi_{p\ell}^{(i,j)} &= \sqrt{\psi_{p\ell}^{(i,i)}} \sqrt{\psi_{p\ell}^{(j,j)}} \text{Cov} \left\{ p_p(X_i) p_\ell(X_i) / \sqrt{\psi_{p\ell}^{(i,i)}}, \right. \\ &\quad \left. p_p(X_j) p_\ell(X_j) / \sqrt{\psi_{p\ell}^{(j,j)}} \right\} \\ &\leq C \zeta_{ij} \left( \sup_x p_\ell^2(x) \sup_x p_p^2(x) \right)^{1/2} \\ &\quad \times \left( E p_p^2(X_i) E p_\ell^2(X_i) E p_p^2(X_j) E p_\ell^2(X_j) \right)^{1/4}, \end{aligned}$$

and thus

$$\begin{aligned} a_2 &\leq \frac{C}{n^2} \left( \sup_x \|p^k(x)\|^2 \right) \sum_{i,j=1, j \neq i}^n \zeta_{ij} \left( E \|p(X_i)\|^2 + E \|p(X_j)\|^2 \right) \\ &\leq C \kappa \xi^2 \Delta = o(1), \end{aligned}$$

by A4, which completes the proof of (27), and thus of the theorem. ■

**Proof of Theorem 2.** We will prove later that

$$\|\bar{V} - V\|_2 = o_p(\tau^{-1}), \tag{28}$$

implying with B3 that  $\|\bar{V}V^{-1} - I_d\| \leq \|\bar{V} - V\| \|V^{-1}\| \rightarrow_p 0$  and thus

$$V\bar{V}^{-1} \rightarrow_p I_d, \quad \|\bar{V}^{-1}\| \leq \|V^{-1}\| \|V\bar{V}^{-1}\| = O_p(\tau). \tag{29}$$

Writing  $r = \hat{\theta} - \theta_0 - n^{-1/2}v$ , with  $v$  defined in (14), (19) is implied if

$$\sqrt{n}\bar{V}^{-1/2}r = o_p(1), \tag{30}$$

$$\bar{V}^{-1/2}v \rightarrow_d N(0, I_d). \tag{31}$$

*Proof of (30).* By the same argument as in the proof of Theorem 1, (30) follows if we show that  $\|1_n \sqrt{n}\bar{V}^{-1/2}r\| \leq \sqrt{n}\|\bar{V}^{-1/2}\| \|1_n r\| = o_p(1)$ . Now denoting  $\hat{m}(\cdot) = p(\cdot)'\beta$ ,

$$\begin{aligned} \|r\| &\leq \|a(\hat{m}) - a(m) - D(\hat{m}) + D(m)\| \\ &\quad + \|D(\hat{m}) - D(m) - n^{-1/2}v\| + \|D(\hat{m}) - D(m)\| \\ &=: \|r_1\| + \|r_2\| + \|r_3\|. \end{aligned}$$

Since  $\|\bar{V}^{-1/2}\| = O_p(\tau^{1/2})$  it suffices to prove that  $1_n \sqrt{n\tau} \|r_i\| = o_p(1)$ ,  $i = 1, 2, 3$ . By B1,  $\|r_1\| = O_p(\|\hat{m} - m\|_{\infty, w}^2)$ . By the triangle inequality, A1 and (23),

$$\begin{aligned} \|\hat{m} - m\|_{\infty, w}^2 &\leq 2|p'(\hat{\beta} - \beta)|_{\infty, w}^2 + 2|p'\beta - m|_{\infty, w}^2 \\ &\leq 2\xi^2 \|\hat{\beta} - \beta\|^2 + O(\kappa^{-2\alpha}) \\ &= O_p(\xi^2(\chi + \kappa^{-2\alpha})), \end{aligned} \tag{32}$$

and by B4,

$$\sqrt{n\tau} \|r_1\| = O_p(\sqrt{n\tau} \xi^2(\chi + \kappa^{-2\alpha})) = o_p(1).$$

Next, since

$$\begin{aligned} D(\hat{m}) &= D(P'\hat{\beta}) = A'\hat{\beta} = A'Q^{-1}P'(M + U)/n, \\ D(\bar{m}) &= D(P'\beta) = A'\beta, \end{aligned} \tag{33}$$

we have

$$\begin{aligned} \|1_n r_2\| &= \|1_n A'Q^{-1}P'(M + U)/n - 1_n A'\beta - 1_n A'P'U/n\| \\ &\leq \|1_n A'(Q^{-1} - I)P'U/n\| + \|1_n A'Q^{-1}P'(M - P\beta)/n\| \\ &\leq \|A\| \|1_n(Q^{-1} - I)\| \|P'U/n\| \\ &\quad + \|A\| \|1_n Q^{-1}P'/\sqrt{n}\| \|(M - P\beta)/\sqrt{n}\|. \end{aligned}$$

Now  $\|A\| \leq C\xi$ ,  $\|1_n^{-1}Q\| = O_p(1)$ , and by (12), Lemma 1 and (26),  $\|1_n^{-1}Q^{-1}P'/\sqrt{n}\| = O_p(1)$ ,  $\|(M - P\beta)/\sqrt{n}\| = O_p(\kappa^{-\alpha})$ ,  $\|P'U/n\| = O_p(\chi^{1/2})$ , while  $\|1_n(Q^{-1} - I)\|^2 = O_p(\|I - Q\|_2^2) = O(\kappa\xi^2(n^{-1} + \Delta))$  from (27). Thus

$$\begin{aligned} \sqrt{n\tau} \|1_n r_2\| &= O_p(\xi(n\tau)^{1/2}(\xi\kappa^{1/2}(n^{-1/2} + \Delta^{1/2}) + \kappa^{-\alpha})) \\ &= o_p(1) \end{aligned}$$

by B4. Finally, by linearity of  $D(\cdot)$ , B2 and A1,  $\sqrt{n\tau} \|r_3\| = O(\sqrt{n\tau} |\bar{m} - m|_{\infty, w}) = O(\sqrt{n\tau} \kappa^{-\alpha}) = o_p(1)$ , by B4.

*Proof of (31).* With  $z_i = A'p^k(X_i)\sigma(X_i)$ , write

$$v = \sum_{j=1}^{\infty} t_j \varepsilon_j, \quad t_j = t_{jn} = \sum_{i=1}^n z_i b_{ij} / \sqrt{n}. \tag{34}$$

For  $N = N_n$ , write  $v_1 = \sum_{j=1}^N t_j \varepsilon_j$ ,  $v_2 = v - v_1$ . Since

$$E \|z_i\|^2 \leq C\xi^2 (E\sigma^4(X_i) E \|p^k(X_i)\|^4)^{1/2} \leq C\xi^3 \kappa^{1/2}, \tag{35}$$

we have

$$\begin{aligned} E \|t_j\|^2 &\leq \frac{C}{n} \sum_{i=1}^n \sum_{k=1}^n (E \|z_i\|^2 + E \|z_k\|^2) |b_{ij} b_{kj}| \\ &\leq \frac{C\xi^3 \kappa^{1/2}}{n} \left( \sum_{i=1}^n |b_{ij}| \right)^2 \end{aligned} \tag{36}$$

and thus

$$\begin{aligned} E \|v_2\|^2 &= \sum_{j=N+1}^{\infty} E \|t_j\|^2 \\ &\leq \frac{C\xi^3 \kappa^{1/2} N^{-\eta}}{n} \left( \max_j \sum_{k=1}^n |b_{kj}| \right) \sum_{i=1}^n \sum_{j=N+1}^{\infty} j^\eta |b_{ij}| \\ &\leq C\xi^3 \kappa^{1/2} N^{-\eta} \max_{1 \leq i \leq n} \sum_{j=1}^{\infty} j^\eta |b_{ij}| = o(\tau^{-1}), \end{aligned}$$

by B5, on choosing  $N$  such that  $(\xi^3 \kappa^{1/2} \tau)^{1/\eta} / N \rightarrow 0$ . Thus from (31) it remains to consider  $\bar{V}^{-1/2}v_1$ . Defining

$$\tilde{V} = \tilde{V}_n = \sum_{j=1}^N t_j t_j',$$

in view of (17), (36)  $\|\bar{V} - \tilde{V}\|_2 = \sum_{j=N+1}^{\infty} \|t_j\|^2 = o_p(\tau^{-1})$ , so arguing as in (29),

$$\tilde{V}^{-1}\bar{V} \rightarrow_p I_d, \quad \|\tilde{V}^{-1}\| = O_p(\tau). \tag{37}$$

We now show that for any  $c \in \mathbb{R}^d$  such that  $c'c = 1$

$$c'\tilde{V}^{-1/2}v_1 \rightarrow_d N(0, 1), \tag{38}$$

conditionally on  $\{X_i\}_{i=1}^{\infty}$  whence (38) holds unconditionally also, and

$$\tilde{V}^{-1/2}v_1 \rightarrow_d N(0, I_d)$$

by the Cramer–Wold device. Writing  $w_j = w_{jn} = c'\tilde{V}^{-1/2}t_j$ , since the  $w_j \varepsilon_j$  are martingale differences under A3 it suffices to check the following two sufficient conditions of Scott (1973):

$$\sum_{j=1}^N E(w_j \varepsilon_j)^2 \rightarrow_p 1, \tag{39}$$

$$\sum_{j=1}^N E((w_j \varepsilon_j)^2 1(|w_j \varepsilon_j| > \delta) | \{X_i\}_{i=1}^{\infty}) \rightarrow_p 0, \quad \forall \delta > 0. \tag{40}$$

By independence of  $\{\varepsilon_i\}$  and  $\{X_i\}$ , and  $\sum_{j=1}^N w_j^2 \equiv 1$ , the left side of (39) is  $E\left(\sum_{j=1}^N w_j^2\right) \equiv E(1) = 1$ , so (39) holds. As in the proof of Theorem 4 of Robinson (2011), again using the independence, the term of (40) is bounded by

$$\max_{j \geq 1} E(\varepsilon_j^2 1(\varepsilon_j^2 > \delta/r)) + \frac{1}{r\delta} \sum_{j=1}^N w_j^4,$$

for any  $r > 0$ . By B5, for any  $\delta$  the first term can be made arbitrarily small by choosing  $r$  small enough, while from (36)

$$\begin{aligned} \sum_{j=1}^N w_j^4 &\leq \frac{1}{n^2} \|\tilde{V}^{-1/2}\|^4 \sum_{j=1}^N \|t_j\|^4 \\ &= O_p\left(\frac{\xi^6 \kappa \tau^2}{n^2} \sum_{j=1}^N \left(\sum_{i=1}^n |b_{ij}|\right)^4\right) \\ &= O_p\left(\frac{\xi^6 \kappa \tau^2}{n} \left(\max_{j \geq 1} \sum_{i=1}^n |b_{ij}|\right)^3 \left(\max_{1 \leq i \leq n} \sum_{j=1}^n |b_{ij}|\right)\right) \\ &= O_p\left(\frac{\xi^6 \kappa \tau^2}{n}\right), \end{aligned} \tag{41}$$

which is  $o_p(1)$ , to verify (40) in view of B4. It follows from the above calculations that  $\tilde{V}^{-1/2}v = (\tilde{V}^{-1/2}\tilde{V}^{1/2})\tilde{V}^{-1/2}v_1 + \tilde{V}^{-1/2}v_2 = \tilde{V}^{-1/2}v_1 + o_p(1)$ , to prove (31). This proves (19). Finally, writing  $\sqrt{n}\tilde{V}^{-1/2}(\hat{\theta} - \theta_0) = \sqrt{n}\tilde{V}^{-1/2}(\hat{\theta} - \theta_0) + \sqrt{n}(V^{-1/2} - \tilde{V}^{-1/2})(\hat{\theta} - \theta_0)$ , (20) follows from (19) and (29).

*Proof of (28).* Let  $z_i^{(\ell)}$  denote the  $\ell$ th element of  $z_i$ . The  $(\ell, p)$ th element of  $\tilde{V} - V$  is  $n^{-1} \sum_{i,j=1}^n \gamma_{ij} \left\{ z_i^{(\ell)} z_j^{(p)} - E(z_i^{(\ell)} z_j^{(p)}) \right\}$ , which, from Isserlis' formula, has variance

$$\begin{aligned} \frac{1}{n^2} \sum_{i_1, i_2, i_3, i_4=1}^n \gamma_{i_1 i_2} \gamma_{i_3 i_4} \text{cum}(z_{i_1}^{(\ell)}, z_{i_2}^{(p)}, z_{i_3}^{(\ell)}, z_{i_4}^{(p)}) \\ + \frac{1}{n^2} \sum_{i_1, i_2, i_3, i_4=1}^n \gamma_{i_1 i_2} \gamma_{i_3 i_4} \phi_{i_1 i_3}^{(\ell, \ell)} \phi_{i_2 i_4}^{(p, p)} \\ + \frac{1}{n^2} \sum_{i_1, i_2, i_3, i_4=1}^n \gamma_{i_1 i_2} \gamma_{i_3 i_4} \phi_{i_1 i_4}^{(\ell, p)} \phi_{i_2 i_3}^{(p, \ell)}, \end{aligned} \tag{42}$$

where  $\phi_{ij}^{(\ell, p)} = \text{Cov}(z_i^{(\ell)}, z_j^{(p)})$ .

For each  $\ell, p = 1, \dots, d$ , the left side of (42) is bounded by

$$\begin{aligned} \frac{1}{n^2} \sum_{i_2, i_4=1}^n \sum_{i_1=1}^n |\gamma_{i_1 i_2}| \sum_{i_3=1}^n |\gamma_{i_3 i_4}| \left| \text{cum}(z_{i_1}^{(\ell)}, z_{i_2}^{(p)}, z_{i_3}^{(\ell)}, z_{i_4}^{(p)}) \right| \\ \leq \frac{1}{n^2} \sum_{i_2, i_4=1}^n \max_{i_1, i_3} \left| \text{cum}(z_{i_1}^{(\ell)}, z_{i_2}^{(p)}, z_{i_3}^{(\ell)}, z_{i_4}^{(p)}) \right| \\ \times \max_{i_2} \sum_{i_1=1}^n |\gamma_{i_1 i_2}| \max_{i_4} \sum_{i_3=1}^n |\gamma_{i_3 i_4}| \\ \leq \frac{C}{n^2} \sum_{i_2, i_4=1}^n \max_{i_1, i_3} \left| \text{cum}(z_{i_1}^{(\ell)}, z_{i_2}^{(p)}, z_{i_3}^{(\ell)}, z_{i_4}^{(p)}) \right| \\ \leq \frac{C}{n^2} \sum_{i_2, i_4=1}^n \max_{i_1, i_3} \sum_{h_1, \dots, h_4=1}^{\kappa} |A_{h_1 \ell} A_{h_2 p} A_{h_3 \ell} A_{h_4 p}| \\ \times |\text{cum}(w_{i_1}^{(h_1)}, \dots, w_{i_4}^{(h_4)})| \end{aligned}$$

$$\begin{aligned} \leq \frac{C}{n^2} \sum_{h=1}^{\kappa} A_{h\ell}^2 \sum_{h=1}^{\kappa} A_{hp}^2 \sum_{i_2, i_4=1}^n \max_{i_1, i_3} \\ \times \left( \sum_{h_1, \dots, h_4=1}^{\kappa} \left( \text{cum}(w_{i_1}^{(h_1)}, \dots, w_{i_4}^{(h_4)}) \right)^2 \right)^{1/2} \\ \leq \frac{C\xi^2}{n^2} \sum_{i_2, i_4=1}^n \max_{i_1, i_3} \\ \times \left( \sum_{h_1, \dots, h_4=1}^{\kappa} \left( \text{cum}(w_{i_1}^{(h_1)}, \dots, w_{i_4}^{(h_4)}) \right)^2 \right)^{1/2} = o(1), \end{aligned}$$

by B6, noting

$$\begin{aligned} \sum_{h_1, \dots, h_4=1}^{\kappa} |A_{h_1 \ell} A_{h_2 p} A_{h_3 \ell} A_{h_4 p}| &\leq \left( \sum_{h_1, \dots, h_4=1}^{\kappa} (A_{h_1 \ell} A_{h_2 p} A_{h_3 \ell} A_{h_4 p})^2 \right)^{1/2} \kappa^2 \\ &\leq \|A\|^2 \kappa^2 \leq C\xi^2 \kappa^2 \end{aligned}$$

with  $\|A\| \leq C\xi$  from B2.

From (35),  $|\phi_{ii}^{(\ell, p)}| \leq C\xi^3 \kappa^{1/2}$ , and the first term in (43) is bounded by

$$\begin{aligned} \frac{1}{n^2} \sum_{i_1, i_2, i_3, i_4=1}^n \left| \gamma_{i_1 i_2} \gamma_{i_3 i_4} \phi_{i_1 i_3}^{(\ell, \ell)} \right| \left( \phi_{i_2 i_2}^{(p, p)} \phi_{i_4 i_4}^{(p, p)} \right)^{1/2} \\ \leq \frac{C}{n^2} \xi^3 \kappa^{1/2} \sum_{i_1, i_2, i_3, i_4=1}^n \left| \gamma_{i_1 i_2} \gamma_{i_3 i_4} \phi_{i_1 i_3}^{(\ell, \ell)} \right| \\ \leq \frac{C}{n^2} \xi^3 \kappa^{1/2} \left( \sum_{i=1}^n |\phi_{ii}^{(\ell, \ell)}| + \sum_{\substack{i, j=1 \\ i \neq j}}^n |\phi_{ij}^{(\ell, \ell)}| \right) \\ \leq C\xi^6 \kappa (n^{-1} + \Delta) = o_p(\tau^{-2}) \end{aligned} \tag{44}$$

by A4 and B4, since

$$\begin{aligned} |\phi_{ij}^{(\ell, \ell)}| &= \sqrt{\phi_{ii}^{(\ell, \ell)}} \sqrt{\phi_{jj}^{(\ell, \ell)}} \left| \text{Cov}\left(z_i^{(\ell)} / \sqrt{\phi_{ii}^{(\ell, \ell)}}, z_j^{(\ell)} / \sqrt{\phi_{jj}^{(\ell, \ell)}}\right) \right| \\ &\leq C\xi^3 \kappa^{1/2} \zeta_{ij}, \quad i \neq j. \end{aligned}$$

By symmetry the second term in (43) is also of the order (44). ■

**Proof of Theorem 3.** The proof of Theorem 3 is similar to, and simpler than the proof of Theorem 4 and is thus omitted here, but can be found in the online appendix of the paper on the author's website.<sup>1</sup>

**Proof of Theorem 4.** Define the  $d \times 1$  vector summation

$$\hat{S}_n^*(r, u) = \sum_{j=1}^{[rs]} \sum_{k=1}^{[ut]} \hat{A}^{*'} Q^{-1} p(X_{jk}) \hat{U}_{jk} / \sqrt{n}, \quad 0 \leq r, u \leq 1,$$

where  $[\cdot]$  denotes integer part. From Lemmas 3–5, one has weak convergence  $(\hat{S}_n^*(r, u))_{r, u \in [0, 1]} \Rightarrow (\Omega^{1/2} \{W_d(r, u) - ruW_d(1, 1)\})_{r, u \in [0, 1]}$  in the space  $\{D([0, 1]^2)\}^d$ . Observe that  $\hat{C}_n = n^{-1} \sum_{m=1}^s \sum_{\ell=1}^t S_n^*(m/s, \ell/t) S_n^*(m/s, \ell/t)' \sim \int_0^1 \int_0^1 S_n^*(r, u) S_n^*(r, u)' dr du$ , so the continuous mapping theorem gives  $\Omega^{-1/2} \hat{C}_n \Omega^{-1/2} \Rightarrow \Phi_d$ . Writing

$$\begin{aligned} n(\hat{\theta}_n - \theta_0)' \hat{C}_n^{-1} (\hat{\theta}_n - \theta_0) \\ = \left( \sqrt{n} \Omega^{-1/2} (\hat{\theta}_n - \theta_0) \right)' \left( \Omega^{1/2} \hat{C}_n^{-1} \Omega^{1/2} \right) \left( \sqrt{n} \Omega^{-1/2} (\hat{\theta}_n - \theta_0) \right). \end{aligned}$$

<sup>1</sup> <https://sites.google.com/site/jungyoonleeecon>.

We observe by [Lemmas 3–5](#) that  $\Omega^{-1/2}\hat{C}_n\Omega^{-1/2} \Rightarrow \Phi_d$ , and by [Theorem 2](#) and [Assumption C1](#) that  $\sqrt{n}\Omega^{-1/2}(\hat{\theta}_n - \theta_0) \rightarrow_d N(0, I_d)$ , where the two terms converge jointly. ■

**Appendix B. Lemmas**

**Lemma 1.** Under [Assumption A1](#), as  $n \rightarrow \infty$ ,

$$\max_{1 \leq i \leq n} E|m(X_i) - p'(X_i)\beta|^2 = O(\kappa^{-2\alpha}).$$

**Proof of Lemma 1.** By [A1](#), for  $1 \leq i \leq n$ ,

$$\begin{aligned} E(m(X_i) - p'(X_i)\beta)^2 &= E\left(\left(m(X_i) - p'(X_i)\beta\right)^2 (1 + \|X_i\|^2)^{-w} (1 + \|X_i\|^2)^w\right) \\ &\leq |m - p'\beta|_{\infty, w}^2 E\left((1 + \|X_i\|^2)^w\right) = O(\kappa^{-2\alpha}). \quad \blacksquare \end{aligned}$$

**Lemma 2.** Assume there exists  $\eta(j) \geq 0, j \in Z$  such that  $\sum_{j=-\infty}^{\infty} \eta(j) < \infty$  and  $|\gamma_{ik}| \leq \eta(i - k), i, k = 1, 2, \dots$ . Then for any  $r \in [0, 1]$ ,

$$\sum_{i=1}^{[m]} \sum_{k=[rn]+1}^n |\gamma_{ik}| = o(n).$$

**Proof of Lemma 2.** Note that  $\tau_n = \sum_{|j| \geq \log n} \eta(j) \rightarrow 0$  as  $n \rightarrow \infty$ , and  $\max_j \eta(j) \leq C < \infty$ . We have

$$\begin{aligned} \sum_{i=1}^{[m]} \sum_{k=[rn]+1}^n |\gamma_{ik}| &\leq \sum_{i=1}^{[m]} \sum_{k=[rn]+1}^n \eta(i - k) \leq \sum_{i=1}^{[m]} \sum_{k=[rn]+\log n}^n \eta(i - k) \\ &\quad + \sum_{k=[rn]+1}^n \sum_{i=1}^{[m]-\log n} \eta(i - k) + C \sum_{i=[m]-\log n}^{[m]} \sum_{k=[rn]+1}^{[m]+\log n} 1 \\ &\leq \tau_n \sum_{i=1}^{[m]} 1 + \tau_n \sum_{k=[rn]+1}^n 1 + C(\log n)^2 \\ &\leq 2\tau_n n + C(\log n)^2 = o(n). \quad \blacksquare \end{aligned}$$

The remaining lemmas are needed for establishing the studentization of [Section 5.2](#), where we need to verify weak convergence of  $d$ -dimensional processes. We first introduce a metric for the coordinate space for each element of  $d$ -dimensional processes, then consider a metric for the  $d$ -dimensional processes which lie in a product space. Corresponding definitions and lemmas required for studentization of [Section 5.3](#) can be found in the online supplementary appendix.

Let  $X(\cdot, \cdot), Y(\cdot, \cdot) \in D_2([0, 1]^2) = D_2$  where  $D_2$  is the space of all real valued functions on  $[0, 1]^2$  that are right-continuous with finite left limits with respect to each of the two arguments  $(r, u) \in [0, 1]^2$ . Let  $\Lambda$  be the group of all transformations  $\lambda : [0, 1]^2 \rightarrow [0, 1]^2$  of the form  $\lambda(r, u) = (\lambda_1(r), \lambda_2(u))'$ , where  $\lambda_1, \lambda_2 : [0, 1] \rightarrow [0, 1]$  are continuous, strictly increasing functions satisfying  $\lambda_1(0) = \lambda_2(0) = 0$  and  $\lambda_1(1) = \lambda_2(1) = 1$ .

The ‘‘Skorohod’’ metric  $d(\cdot, \cdot)$  in  $D_2$  is defined as

$$d(X, Y) = \inf_{\lambda \in \Lambda} (\min\{\|X - Y\lambda\|, \|\lambda\|\}),$$

where  $\|X - Y\lambda\| = \sup_{r, u \in [0, 1]} |X(r, u) - Y(\lambda(r, u))|$  and  $\|\lambda\| = \sup_{r, u \in [0, 1]} |\lambda(r, u) - (r, u)|'$ . The space  $D_2$  equipped with the metric  $d(\cdot, \cdot)$  is separable and complete (cf. [Bickel and Wichura \(1971, p. 1662\)](#)), which is crucial to our need when dealing with the product space  $D_2^d = D_2 \times \dots \times D_2$ .

Introduce the  $d$ -dimensional processes

$$S_n(r, u) = \sum_{j=1}^{[rs]} \sum_{k=1}^{[ut]} A'p(X_{jk})U_{jk}/\sqrt{n}, \tag{45}$$

$$\hat{S}_n(r, u) = \sum_{j=1}^{[rs]} \sum_{k=1}^{[ut]} A'p(X_{jk})\hat{U}_{jk}/\sqrt{n}, \quad r, u \in [0, 1].$$

Note that  $S_n(\cdot, \cdot) \in D_2^d$ . For  $X(\cdot, \cdot) = (X_1(\cdot, \cdot), \dots, X_d(\cdot, \cdot))' \in D_2^d$  and  $Y(\cdot, \cdot) = (Y_1(\cdot, \cdot), \dots, Y_d(\cdot, \cdot))' \in D_2^d$ , define the metric in  $D_2^d$  as

$$d'(X, Y) = \max_{1 \leq \ell \leq d} \{d(X_\ell, Y_\ell) : X_\ell, Y_\ell \in D_2\}.$$

Then the  $\sigma$ -algebra generated by open sets of  $D_2^d$  is equivalent to the product of the  $\sigma$ -algebras generated by open sets in the component space  $D_2$  ([Billingsley \(1968\)](#), p. 225).

[Lemma 3](#) states a functional central limit theorem (FCLT) for  $S_n(r, u)$  in  $D_2^d$  equipped with the metric  $d'(\cdot, \cdot)$ , where  $\Rightarrow_{D_2^d}$  denotes weak convergence of the associated probability measures in  $D_2^d$ .

Note that  $S_n(r, u)$  can also be written as a sum with triangular array weights

$$S_n(r, u) = \sum_{i, h=1}^{\infty} t_{ih}(n : r, u)\varepsilon_{ih}, \tag{46}$$

where, with  $z_{jk} = A'p(X_{jk})\sigma(X_{jk})$  and

$$t_{ih}(n; r, u) := \sum_{j=1}^{[rs]} \sum_{k=1}^{[ut]} z_{jk}b_{jkh}/\sqrt{n}, \quad r, u \in [0, 1], n \geq 1,$$

with  $b_{jkh}$  as in  $U_{jk} = \sum_{i=1}^{\infty} \sum_{h=1}^{\infty} b_{jkh}\varepsilon_{ih}$ . In the proof, we find it more convenient to use the representation [\(45\)](#) for establishing tightness while [\(46\)](#) is used for deriving asymptotic normality.

**Lemma 3.** Under the assumptions of [Theorem 4](#) as  $n \rightarrow \infty$ ,

$$(S_n(r, u))_{0 \leq r, u \leq 1} \Rightarrow_{D_2^d} (\Omega^{1/2}W_d(r, u))_{0 \leq r, u \leq 1}. \tag{47}$$

**Proof of Lemma 3.** We use two sufficient conditions for weak convergence of probability measures in the product space  $D_2^d$  under the metric  $d'(\cdot, \cdot)$ : convergence of finite dimensional distributions of  $S_n(\cdot, \cdot)$  to those of  $\Omega^{1/2}W_d(\cdot, \cdot)$ , and tightness condition for each component of the vector  $S_n(\cdot, \cdot)$ . This is justified by using the same arguments as in [Phillips and Durlauf \(1986, pp. 487–489, Lemma A.1, A.3\)](#). While [Lemma A.3 of Phillips and Durlauf \(1986\)](#) on the tightness condition is a general result that applies also to the product space  $D_2^d$ , we need to show that their [Lemma A.1](#) on convergence of finite dimensional distributions shown for  $\{D[0, 1]\}^d$  remains valid also for  $D_2^d = \{D([0, 1]^2)\}^d$ . The proof of [Lemma A.1 of Phillips and Durlauf \(1986\)](#) rests on [Theorem 14.5 of Billingsley \(1968\)](#) derived for the space  $D[0, 1]$ . [Neuhaus \(1971, p. 1290\)](#) extended [Billingsley \(1968\)](#)'s result to the space  $D([0, 1]^2)$ . Hence the same arguments as those in [Phillips and Durlauf \(1986\)](#) imply validity of their [Lemma A.1](#) for the product space  $D_2^d$ .

Convergence of finite dimensional distributions and tightness for  $S_n(\cdot, \cdot)$  will be established below after showing: for any  $0 \leq r \leq r' \leq 1$  and  $0 \leq u \leq u' \leq 1$ ,

$$ES_n(r, u)S_n(r', u')' \rightarrow ru\Omega, \tag{48}$$

where  $S_n(r, u) = (S_{n1}(r, u), \dots, S_{nd}(r, u))'$ . Write

$$\begin{aligned} ES_n(r, u)S_n(r', u')' &= ES_n(r, u)S_n(r, u)' \\ &\quad + E(S_n(r, u)(S_n(r', u')' - S_n(r, u)' )'). \end{aligned}$$

Denote by  $P_{ru} = (p(X_{11}), p(X_{12}), \dots, p(X_{1[ut]}), p(X_{21}), p(X_{22}), \dots, p(X_{[rs][ut]}))'$  a  $[rs][ut] \times \kappa$  matrix and similarly by  $U_{ru}$  the  $[rs][ut] \times 1$  vector. By Assumption C5',  $E(S_n S_n') = V_n \rightarrow \Omega$ , and therefore

$$ES_n(r, u)S_n(r, u)' = \frac{[rs]}{s} \frac{[ut]}{t} \frac{1}{[rs]} \frac{1}{[ut]} E(A'P'_{ru}U_{ru}U'_{ru}P_{ru}A) \rightarrow ru\Omega.$$

Convergence (48) follows if we show that  $E(S_n(r, u)(S_n(r', u') - S_n(r, u))) \rightarrow 0$ . Denote  $d \times 1$  vector  $A'p(x) =: w_\kappa = (w_{1\kappa}(x), \dots, w_{d\kappa}(x))'$ . A precondition for C1 given by Assumption 7 of Newey (1997), along with weak dependence in  $e_{jk}$  implied by B5, lead to  $Ew_{p\kappa}^2(X_{jk}) < \infty$  for  $1 \leq p \leq \kappa, 1 \leq j \leq s, 1 \leq k \leq t$ . For  $\ell, p = 1, \dots, d$ ,

$$\begin{aligned} &|E[S_n(r, u)(S_n(r', u') - S_n(r, u))]'_{\ell p}] \\ &\leq \frac{C}{n} \sum_{j=1}^{[rs]} \sum_{\ell=[rs]+1}^{[r's]} \sum_{k=1}^{[ut]} \sum_{h=[ut]+1}^{[u't]} |\gamma_{jkh}| E|w_{\ell\kappa}(X_{jk})w_{p\kappa}(X_{th})| \\ &\leq \frac{C}{n} \left( \sum_{k,h=1}^t \sum_{j=1}^{[rs]} \sum_{\ell=[rs]+1}^{[r's]} |\gamma_{jkh}| + \sum_{j,\ell=1}^s \sum_{k=1}^{[ut]} \sum_{h=[ut]+1}^{[u't]} |\gamma_{jkh}| \right) = o(1), \end{aligned}$$

by C5'(i), and because

$$E|w_{\ell\kappa}(X_i)w_{p\kappa}(X_k)| \leq (Ew_{\ell\kappa}^2(X_i)Ew_{p\kappa}^2(X_k))^{1/2} < \infty.$$

This completes the proof of (48).

Next we show that finite dimensional distributions of  $S_n(\cdot)$  converge to those of  $\Omega^{1/2}W_d(\cdot)$ , that is, for an arbitrary integer  $k$ , and any distinct points  $r_1, \dots, r_k$  and  $u_1, \dots, u_k$  in  $[0, 1]$ ,

$$\begin{aligned} &(S_n(r_1, u_1), \dots, S_n(r_k, u_k)) \\ &\rightarrow_d (\Omega^{1/2}W_d(r_1, u_1), \dots, \Omega^{1/2}W_d(r_k, u_k)). \end{aligned}$$

By the Cramer-Wold device, it suffices to show that for any  $d \times 1$  vectors  $c'_1, \dots, c'_k, Q_n \rightarrow_d Q$ , where

$$Q_n = \sum_{l=1}^k c'_l S_n(r_l, u_l), \quad Q = \sum_{l=1}^k c'_l \Omega^{1/2}W_d(r_l, u_l).$$

From (46),  $Q_n = \sum_{i,h=1}^\infty t_{ih}^*(n; r, u)\varepsilon_{ih}$  with  $t_{ih}^*(n; r, u) = \sum_{l=1}^k c'_l t_{ih}(n; r_l, u_l)$ . By (48),

$$\text{Var}(Q_n) = \sum_{i,h=1}^\infty (t_{ih}^*)^2 \rightarrow \text{Var}(Q) = \sum_{l,t=1}^k c'_l \Omega c_t \cdot \min\{r_l, r_t\} < \infty.$$

By (41), which holds for all  $c'_l t_{ih}(n; r_l, u_l), l = 1, \dots, k$ , we have  $\sum_{i=1}^S \sum_{h=1}^T t_{ih}^*(n; r, u)^4 = o_p(1)$  for  $S$  and  $T$  similar to  $N$  mentioned in the proof of (31), and  $Q_n \rightarrow_d Q$  follows by the same argument as in the proof of asymptotic normality (31).

To verify tightness of  $S_n(r, u)$ , we define the increment of  $S_n(r, u)$ . For  $0 \leq r \leq r' \leq 1$  and  $0 \leq u \leq u' \leq 1$ , set

$$\begin{aligned} \delta(rur'u') &= (\delta_1(rur'u'), \dots, \delta_d(rur'u'))' \\ &:= S_n(r', u') + S_n(r, u) - S_n(r', u) - S_n(r, u') \\ &= \frac{1}{\sqrt{n}} \sum_{j=[rs]+1}^{[r's]} \sum_{k=[ut]+1}^{[u't]} A'p(X_{jk})U_{jk}. \end{aligned}$$

To establish tightness of joint probability measures on a product space such as  $D_2^d$ , a necessary and sufficient condition is that all of the marginal probability measures are tight on the coordinate space  $D_2$  (Phillips and Durlauf (1986, Lemma A.3), Billingsley (1968, p. 41)). A sufficient condition for tightness for each component in  $D_2$  established in (Bickel and Wichura (1971),

Theorem 3) is given by

$$\begin{aligned} E|\delta_\ell(rur'u')|^4 &\leq C \left( \frac{[r's] - [rs]}{s} \right)^2 \left( \frac{[u't] - [ut]}{t} \right)^2, \\ &\ell = 1, \dots, d, \end{aligned} \tag{49}$$

noting that the right hand side approaches the squared Lebesgue measure  $((r' - r)(u' - u))^2$  as  $s, t \rightarrow \infty$ . Note that

$$\begin{aligned} &E|\delta(rur'u')|^2 \\ &\leq \frac{1}{n} \sum_{j,\ell=[rs]+1}^{[r's]} \sum_{k,h=[ut]+1}^{[u't]} |\gamma_{jkh}| E|\sigma(X_{jk})\sigma(X_{th})w_{\ell\kappa}(X_{jk})w'_{\ell\kappa}(X_{th})| \\ &\leq \frac{C}{n} \sum_{j,\ell=[rs]+1}^{[r's]} \sum_{k,h=[ut]+1}^{[u't]} |\gamma_{jkh}| \\ &\leq \frac{C}{n} \sum_{j,\ell=[rs]+1}^{[r's]} \sum_{k,h=[ut]+1}^{[u't]} \left[ \max_{1 \leq j \leq s, 1 \leq k \leq t} \sum_{\ell=1}^s \sum_{h=1}^t |\gamma_{jkh}| \right] \\ &\leq C \left( \frac{[r's] - [rs]}{s} \right) \left( \frac{[u't] - [ut]}{t} \right) \end{aligned} \tag{50}$$

by Assumption C5'(ii).

To verify (49), denote by  $1_\ell$  a  $d$ -dimensional vector whose  $\ell$ th element is 1 and other elements 0, so

$$\begin{aligned} S_{n\ell}(u) - S_{n\ell}(r) &= \sum_{i,h=1}^\infty 1'_\ell(t_{ih}(n; r', u') - t_{ih}(n; r, u))\varepsilon_{ih} \\ &=: \sum_{i,h=1}^\infty \lambda_{ihn}\varepsilon_{ih}. \end{aligned} \tag{51}$$

By Assumption C3,

$$\begin{aligned} E \left( \sum_{i,h=1}^\infty \lambda_{ihn}\varepsilon_{ih} \right)^2 &= \sum_{i,h=1}^\infty \lambda_{ihn}^2, \tag{52} \\ E \left( \sum_{i,h=1}^\infty \lambda_{ihn}\varepsilon_{ih} \right)^4 &= \sum_{i_1, \dots, i_4=1}^\infty \sum_{h_1, \dots, h_4=1}^\infty \lambda_{i_1 h_1} \lambda_{i_2 h_2} \lambda_{i_3 h_3} \lambda_{i_4 h_4} \\ &\quad \times E(\varepsilon_{i_1 h_1} \varepsilon_{i_2 h_2} \varepsilon_{i_3 h_3} \varepsilon_{i_4 h_4}) \\ &= 3 \left[ \sum_{i,\ell'=1:i \neq \ell'}^\infty \sum_{h,h'=1:h \neq h'}^\infty \lambda_{i h}^2 \lambda_{\ell' h'}^2 \right] \\ &\quad + \kappa \sum_{i,h=1}^\infty \lambda_{i h}^4 \leq C \left[ \sum_{i,h=1}^\infty \lambda_{i h}^2 \right]^2. \end{aligned} \tag{53}$$

Thus from (50)–(53),

$$\begin{aligned} E|S_{n\ell}(r', u') - S_{n\ell}(r, u)|^4 &\leq C \left[ \sum_{i,h=1}^\infty \lambda_{i h}^2 \right]^2 \\ &\leq C (E|S_{n\ell}(r', u') - S_{n\ell}(r, u)|^2)^2 \\ &\leq C \left( \frac{[r's] - [rs]}{s} \right)^2 \left( \frac{[u't] - [ut]}{t} \right)^2 \end{aligned}$$

to prove (49). ■

**Lemma 4.** Under the assumptions of Theorem 4, as  $n \rightarrow \infty$ ,

$$\left( \hat{S}_n(r, u) \right)_{0 \leq r, u \leq 1} \Rightarrow_{D_2^d} \left( \Omega^{1/2} \{W_d(r, u) - ruW_d(1, 1)\} \right)_{0 \leq r, u \leq 1}. \tag{54}$$

**Proof of Lemma 4.** We have

$$\begin{aligned} \hat{S}_n(r, u) - S_n(r, u) &= \sum_{j=1}^{\lfloor rs \rfloor} \sum_{k=1}^{\lfloor ut \rfloor} A' p(X_{jk}) \{m(X_{jk}) - \hat{m}(X_{jk})\} / \sqrt{n} \\ &= \sum_{j=1}^{\lfloor rs \rfloor} \sum_{k=1}^{\lfloor ut \rfloor} A' p(X_{jk}) \{m(X_{jk}) - p'(X_{jk})\beta\} / \sqrt{n} \\ &\quad + \sum_{j=1}^{\lfloor rs \rfloor} \sum_{k=1}^{\lfloor ut \rfloor} A' p(X_{jk}) p'(X_{jk}) (\beta - \hat{\beta}) / \sqrt{n} \\ &= A' P'_{ru} (M_{ru} - P_{ru} \beta) / \sqrt{n} \\ &\quad + A' P'_{ru} P_{ru} (\beta - \hat{\beta}) / \sqrt{n} \\ &=: a_n(r, u) - \ell_n(r, u). \end{aligned}$$

We shall show that

$$\sup_{r, u \in [0, 1]} \|a_n(r, u)\| = o_p(1), \tag{55}$$

$$\ell_n(r, u) \Rightarrow_{D^d} ru \Omega^{1/2} W_d(1, 1), \tag{56}$$

which, together with Lemma 3, prove (54).

*Proof of (55).* One has

$$\begin{aligned} \sup_{r, u \in [0, 1]} \|a_n(r, u)\| &\leq \|A'\| \sup_{r, u \in [0, 1]} \|P'_{ru}\| \sup_{r, u \in [0, 1]} \|(M_{ru} - P_{ru} \beta) / \sqrt{n}\| \\ &= O_p(\sqrt{n} \xi^2 \kappa^{-\alpha}) \end{aligned} \tag{57}$$

$$= O_p(\sqrt{n} \xi^2 \kappa^{-\alpha}) \tag{58}$$

because  $\|A'\| \leq C\xi$ , whereas

$$\sup_{r, u \in [0, 1]} \|P'_{ru}\| = O_p(\sqrt{n} \xi),$$

$$\sup_{r, u \in [0, 1]} \|(M_{ru} - P_{ru} \beta) / \sqrt{n}\| = O(\kappa^{-\alpha}),$$

by Lemma 1. Then (55) follows by C2(iii).

*Proof of (56).* Since  $\sqrt{n}(\hat{\beta} - \beta) = Q^{-1} P'(M - P\beta) / \sqrt{n} + Q^{-1} P'U / \sqrt{n}$ , with  $\tilde{Q}_{ru} = P'_{ru} P_{ru} / n$ ,

$$\begin{aligned} \ell_n(r, u) &= A' \tilde{Q}_{ru} Q^{-1} P'(M - P\beta) / \sqrt{n} + A' \tilde{Q}_{ru} Q^{-1} \frac{P'U}{\sqrt{n}} \\ &=: \ell_{1,n}(r, u) + \ell_{2,n}(r, u). \end{aligned} \tag{59}$$

We shall show the following two results which constitute the proof of (56):

$$\sup_{r, u \in [0, 1]} \|\ell_{1,n}(r, u)\| = o_p(1), \quad \ell_{2,n}(r, u) \Rightarrow_{D[0, 1]^d} ru W_d(1, 1).$$

Noting that  $Q^{-1} = O_p(1)$  and  $\sup_{r, u \in [0, 1]} \|\tilde{Q}_r\| = O_p(\sqrt{\kappa})$ , since under A2

$$\left\| \sum_{j=1}^{\lfloor rs \rfloor} \sum_{k=1}^{\lfloor ut \rfloor} p(X_{jk}) p'(X_{jk}) / n \right\| = O_p(\sqrt{\kappa}),$$

we obtain

$$\begin{aligned} \|\ell_{1,n}(r)\| &\leq \|A'\| \sup_{r, u \in [0, 1]} \|\tilde{Q}_{ru}\| \|Q^{-1}\| \|P'(M - P\beta) / \sqrt{n}\| \\ &\leq \|A'\| O_p(\sqrt{\kappa}) \|P'\| \|(M - P\beta) / \sqrt{n}\| \\ &= O_p(\sqrt{n} \xi^2 \kappa^{\frac{1}{2} - \alpha}) = o_p(1), \end{aligned}$$

by C2(iii). Next, write

$$\ell_{2,n}(r, u) = rA' P'U / \sqrt{n} + A' (\tilde{Q}_{ru} Q^{-1} - ruI) P'U / \sqrt{n}.$$

Since the convergence  $ru(A' P'U / \sqrt{n}) \rightarrow_d ru \Omega^{1/2} W_d(1)$  follows from the proof of Theorem 2 and C1, it remains to verify that

$$\sup_{r \in [0, 1]} \|A' (\tilde{Q}_{ru} Q^{-1} - ruI) P'U / \sqrt{n}\| = o_p(1).$$

One has  $\|A'\| = O(\xi)$  and  $\|P'U / \sqrt{n}\| = O(\sqrt{\kappa})$  by Assumption C2(ii). Next, with  $\bar{Q}_{ru} = P'_{ru} P_{ru} / [rs][ut]$  we have

$$\begin{aligned} \sup_{r, u \in [0, 1]} \left\| \frac{[rs]}{s} \frac{[ut]}{t} \bar{Q}_{ru} Q^{-1} - ruI \right\| &\leq \sup_{r, u \in [0, 1]} \|\bar{Q}_{ru} - I\| \|Q^{-1} - I\| \\ &\quad + \sup_{r, u \in [0, 1]} \|\bar{Q}_{ru} - I\| + \|Q^{-1} - I\| + o(1/n). \end{aligned}$$

From (27) and Horn and Johnson (1990, pp. 335–336),

$$\|Q^{-1} - I_\kappa\|^2 = O_p(\kappa \xi^2 (n^{-1} + \Delta)) = O_p(\kappa \xi^2 / n),$$

with the last step following from Assumption C2(i). Similarly, one has

$$\begin{aligned} \sup_{r, u \in [0, 1]} \left( \frac{[rs]}{s} \frac{[ut]}{t} \right)^2 \|\bar{Q}_{ru} - I_\kappa\|^2 &= \sup_{r, u \in [0, 1]} \left( \frac{[rs]}{s} \frac{[ut]}{t} \right)^2 O_p \left( \frac{\kappa \xi^2}{[rs][ut]} + \frac{\kappa \xi^2 \Delta_{ru}}{[rs]^2 [ut]^2} \right) \\ &= O_p(\kappa \xi^2 / n). \end{aligned} \tag{60}$$

Therefore,

$$\sup_{r, u \in [0, 1]} \|A' (\tilde{Q}_{ru} Q^{-1} - ruI) P'U / \sqrt{n}\| = O_p \left( \frac{\kappa \xi^2}{\sqrt{n}} \right) = o_p(1), \tag{61}$$

with the last step following from Assumption C2(iv). ■

**Lemma 5.** Under the assumptions of Theorem 4, as  $n \rightarrow \infty$ ,  $\sup_{r \in [0, 1]} \|\hat{S}_n^*(r, u) - \hat{S}_n(r, u)\| = o_p(1)$ .

**Proof of Lemma 5.** We have

$$\|\hat{S}_n^*(r, u) - \hat{S}_n(r, u)\| = \|(\hat{A}' Q^{-1} - A') P'_{ru} \hat{U}_{ru} / \sqrt{n}\|,$$

and thus

$$\begin{aligned} \sup_{r, u \in [0, 1]} \|\hat{S}_n^*(r, u) - \hat{S}_n(r, u)\| &\leq \|\hat{A}' Q^{-1} - A'\| \sup_{r, u \in [0, 1]} \|P'_{ru} \hat{U}_{ru} / \sqrt{n}\| =: d_{n1} d_{n2}. \end{aligned} \tag{62}$$

We will show that

$$d_{n1} = O_p \left( \xi^2 \left( \sqrt{\frac{\kappa}{n}} + \kappa^{-\alpha} \right) \right), \tag{63}$$

$$d_{n2} = O_p(\kappa^{1/2} + \sqrt{n} \xi \kappa^{-\alpha}), \tag{64}$$

which imply

$$\begin{aligned} d_{n1} d_{n2} &= O_p(\xi^2 \sqrt{\kappa} / \sqrt{n} + \xi^2 \kappa^{-\alpha}) O_p(\sqrt{\kappa} + \xi \sqrt{n} \kappa^{-\alpha}) \\ &= O_p(\kappa \xi^2 / \sqrt{n} + \xi^3 \kappa^{1/2 - \alpha} + \xi^3 \kappa^{-2\alpha} \sqrt{n}) = o_p(1), \end{aligned}$$

by Assumptions C2(iii) and C2(iv). To see this, write

$$\begin{aligned} d_{n1} &= \|\hat{A}' Q^{-1} - A'\| \leq \|\hat{A} - A\| \|Q^{-1} - I_\kappa\| \\ &\quad + \|A\| \|Q^{-1} - I_\kappa\| + \|\hat{A} - A\|. \end{aligned}$$

Horn and Johnson (1990, pp. 335–336), (27) and C2(i) lead to  $\|Q^{-1} - I_\kappa\| = O_p(\xi^2 \kappa / n)$ .

To obtain (63), we need an upper bound on  $\|\hat{A} - A\|$ . Newey (1997) showed that  $\hat{A} = (\hat{A}_1, \dots, \hat{A}_d)$  equals  $(D(p_1; \hat{m}), \dots,$

$D(p_\kappa; \hat{m})'$  with probability approaching one as  $n \rightarrow \infty$ . Recalling  $D(\cdot; \hat{m}) = (D_1(\cdot; \hat{m}), \dots, D_d(\cdot; \hat{m}))$ , the  $i$ th column of  $\hat{A} - A$  can be written as

$$\hat{A}_i - A_i = (D_i(p_1; \hat{m}) - D_i(p_1; m), \dots, D_i(p_\kappa; \hat{m}) - D_i(p_\kappa; m))',$$

$$i = 1, \dots, d.$$

Using linearity of  $D_i(g; \hat{m})$  in  $g$ , write

$$\begin{aligned} \|\hat{A}_i - A_i\|^2 &= (\hat{A}_i - A_i)'(\hat{A}_i - A_i) \\ &= |D_i((\hat{A}_i - A_i)'p; \hat{m}) - D_i((\hat{A}_i - A_i)'p; m)| \\ &\leq C|(\hat{A}_i - A_i)'p|_\infty |\hat{m} - m|_{\infty, w} \\ &\leq C\|\hat{A}_i - A_i\|_\xi |\hat{m} - m|_{\infty, w}, \end{aligned}$$

with the first inequality following from Assumption C4. Therefore  $\|\hat{A}_i - A_i\| = O_p(\xi |\hat{m} - m|_{\infty, w})$ , for  $i = 1, \dots, d$ . This allows the bound

$$\begin{aligned} \|\hat{A} - A\|^2 &\leq \text{tr}((\hat{A} - A)'(\hat{A} - A)) \\ &= \sum_{i=1}^d \|\hat{A}_i - A_i\|^2 \leq C\xi^2 |\hat{m} - m|_{\infty, w}^2. \end{aligned}$$

Thus by (32),

$$\|\hat{A} - A\| = O_p(\xi^2 [\chi^{1/2} + \kappa^{-\alpha}]) = o_p(1),$$

by Assumption C2(ii)–(iv), completing the proof of (63).

Next note that

$$\begin{aligned} d_{n,2} &\leq \sup_{r,u \in [0,1]} \|P'_{ru}(\hat{U}_{ru} - U_{ru})/\sqrt{n}\| + \sup_{r,u \in [0,1]} \|P'_{ru}U_{ru}/\sqrt{n}\| \\ &= d_{n,21} + d_{n,22}. \end{aligned}$$

As in the proof of Lemma 4,

$$\begin{aligned} d_{n,21} &\leq \sup_{r,u \in [0,1]} \|P'_{ru}(M_{ru} - P_{ru}\beta)/\sqrt{n}\| \\ &\quad + \sup_{r,u \in [0,1]} \|\tilde{Q}_{ru}\| \|\sqrt{n}(\hat{\beta} - \beta)\|. \end{aligned}$$

From (58) it is seen that the first term on the right is  $O_p(\sqrt{n}\xi\kappa^{-\alpha})$ . By (23),  $\|(\hat{\beta} - \beta)\sqrt{n}\| = O_p(\sqrt{n}(\chi^{1/2} + \kappa^{-\alpha})) = O_p(\sqrt{\kappa} + \sqrt{n}\kappa^{-\alpha})$  from Assumption C2(ii), whereas by (60),  $\sup_{r,u \in [0,1]} \|\tilde{Q}_{ru}\| = O_p(1) + O_p(\kappa\xi^2/\sqrt{n}) = O_p(1)$  by Assumption C2(iv). Thus  $d_{n,21} = O_p(\kappa^{1/2} + \sqrt{n}\xi\kappa^{-\alpha})$ . Finally,

$$\begin{aligned} d_{n,22} &= \sup_{r,u \in [0,1]} \left( \frac{[rs][ut]}{st} \right)^{1/2} \|P'_{ru}U_{ru}/\sqrt{[rs][ut]}\| \\ &= O_p(\sqrt{n\chi}) = O_p(\sqrt{\kappa}), \end{aligned}$$

by Assumption C2(ii). Hence,  $d_{n,22} = O_p(\kappa^{1/2})$ , which proves (64). ■

### Appendix C. Supplementary data

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.jeconom.2015.08.001>.

### References

Andrews, D.W.K., 1991. Asymptotic normality of series estimators for nonparametric and semiparametric models. *Econometrica* 59, 307–345.  
 Andrews, D.W.K., 2005. Cross-section regression with common shocks. *Econometrica* 73, 1551–1585.

Arbia, G., 2006. *Spatial Econometrics: Statistical Foundation and Applications to Regional Analysis*. Springer-Verlag, Berlin.  
 Bai, J., 2009. Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.  
 Bickel, P.J., Wichura, M.J., 1971. Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* 42, 1656–1670.  
 Billingsley, P., 1968. *Convergence of Probability Measures*. John Wiley, New York.  
 Chamberlain, G., 1986. *Notes on Semiparametric Regression*. Department of Economics, Harvard University.  
 Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In: Heckman, J.J., Leamer, E.E. (Eds.), *Handbook of Econometrics*, vol. 6B. North-Holland, Amsterdam (Chapter 76).  
 Chen, X., Christensen, T.M., 2015. Optimal uniform convergence rates and asymptotic normality for series estimations under weak dependence and weak conditions. *J. Econometrics*. <http://dx.doi.org/10.1016/j.jeconom.2015.03.010>.  
 Chen, X., Hong, H., Tamer, E., 2005. Measurement error models with auxiliary data. *Rev. Econom. Stud.* 72, 343–366.  
 Chen, X., Liao, Z., Sun, Y., 2012. Sieve Inference on Semi-Nonparametric Time Series Models. Working Paper, Dept of Economics, Yale University.  
 Chen, X., Shen, X., 1998. Sieve extremum estimates for weakly dependent data. *Econometrica* 66, 289–314.  
 Cliff, A., Ord, J.K., 1968. *The Problem of Spatial Autocorrelation*. Joint Discussion Paper, University of Bristol: Department of Economics, vol. 26, Department of Geography, Series A, p. 15.  
 Cliff, A., Ord, J.K., 1981. *Spatial Processes, Models and Applications*. Pion, London.  
 Conley, T.G., 1999. GMM estimation with cross sectional dependence. *J. Econometrics* 92, 1–45.  
 de Jong, R.M., 2002. A note on “Convergence rates and asymptotic normality for series estimators”: uniform convergence rates. *J. Econometrics* 111, 1–9.  
 Fan, Y., Li, Q., 1999. Root-n-consistent estimation of partially linear time series models. *J. Nonparametr. Stat.* 11, 251–269.  
 Hannan, E.J., 1957. The variance of the mean of a stationary process. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 19, 282–285.  
 Horn, R.A., Johnson, C.R., 1990. *Matrix Analysis*. Cambridge University Press, New York.  
 Jenish, N., 2012. Nonparametric spatial regression under near-epoch dependence. *J. Econometrics* 167, 224–239.  
 Jenish, N., Prucha, I.R., 2012. On spatial processes and asymptotic inference under near-epoch dependence. *J. Econometrics* 170, 178–190.  
 Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J. Real Estate Finance Econ.* 17, 99–121.  
 Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *Internat. Econom. Rev.* 40, 509–533.  
 Kelejian, H.H., Prucha, I.R., 2007. HAC estimation in a spatial framework. *J. Econometrics* 140, 131–154.  
 Kiefer, M., Vogelsang, T.J., Bunzel, H., 2000. Simple robust testing of regression hypotheses. *Econometrica* 68, 695–714.  
 Lee, L-F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory* 18, 252–277.  
 Lee, L-F., 2004. Asymptotic distribution of quasi-maximum likelihood estimates for spatial autoregressive models. *Econometrica* 72, 1899–1925.  
 Neuhaus, G., 1971. On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* 42, 1285–1295.  
 Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. *J. Econometrics* 79, 147–168.  
 Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74, 967–1012.  
 Phillips, P.C.B., Durlauf, S., 1986. Multiple time series regression with integrated processes. *Rev. Econom. Stud.* 53, 473–496.  
 Pinkse, J., Shen, L., Slade, M.E., 2007. Central limit theorem for endogenous locations and complex spatial interactions. *J. Econometrics* 140, 215–225.  
 Pinkse, J., Slade, M.E., Brett, B., 2002. Spatial price competition: a semiparametric approach. *Econometrica* 70, 1111–1153.  
 Powell, J.L., Stock, J.H., Stoker, T.M., 1989. Semiparametric estimation of index coefficients. *Econometrica* 57, 1043–1430.  
 Robinson, P.M., 1983. Nonparametric estimators for time series. *J. Time Ser. Anal.* 4, 185–207.  
 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. *J. Econometrics* 165, 5–19.  
 Robinson, P.M., Thawornkaiwong, S., 2012. Statistical inference on regression with spatial dependence. *J. Econometrics* 167, 521–542.  
 Rozanov, Yu.A., 1963. *Stationary Random Processes*. Holden-Day, San Francisco.  
 Scott, D.J., 1973. Central limit theorems for martingales using a Skorokhod representation approach. *Adv. Appl. Probab.* 5, 119–137.  
 Stone, C.J., 1982. Optimal global rates of convergence for nonparametric regressions. *Ann. Statist.* 10, 1040–1053.  
 Yatchew, Y., 2003. *Semiparametric Regression for the Applied Econometrician*. Cambridge University Press, New York.