Series Estimation under Cross-sectional Dependence

Jungyoon Lee* and Peter M. Robinson†‡

June 6, 2013

Abstract

An asymptotic theory is developed for nonparametric and semiparametric series estimation under general cross-sectional dependence and heterogeneity. A uniform rate of consistency, asymptotic normality, and sufficient conditions for $\sqrt{n}$ convergence, are established, and a data-driven studentization new to cross-sectional data is justified. The conditions accommodate various cross-sectional settings plausible in economic applications, and apply also to panel and time series data. Strong, as well as weak dependence are covered, and conditional heteroscedasticity is allowed.

JEL classifications: C12; C13; C14; C21

Keywords: Series estimation; Nonparametric regression; Spatial data; Cross-sectional dependence; Uniform rate of consistency; Functional central limit theorem; Data-driven studentization

....

*New College of the Humanities, 19 Bedford Square, London WC1B 3HH, UK

†London School of Economics, Houghton Street, London WC2A 2AE, UK.
Research supported by ESRC Grant ES/J007242/1

‡Corresponding author. Tel. 44 20 7955 7516. fax 44 20 7955 6592

E-mail address: p.m.robinson@lse.ac.uk

1. Introduction

Economic agents are typically interdependent, due for example to externalities, spillovers or the presence of common shocks. Such dependence is often overlooked in cross-sectional or panel data analysis, in part due to a lack of relevant econometric literature. The implications of dependence for econometric analysis have long been studied in the context of time series data, where dependence is naturally modeled in terms of temporal distance between observations, but the nature of cross-sectional dependence hinders a simple multi-dimensional extension of time series methods to spatial econometric data, due to the typical lack of a natural ordering in cross-sectional data. In order to account for possible cross-sectional dependence, one needs first to establish a framework under which its structure can be suitably formalised, 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 that describe dependence and other features. Several approaches to modeling cross-sectional dependence prominent in recent literature can accomplish this.

One class of models postulates unobserved common factors that affect some/all of individual units, see Andrews (2005), Pesaran (2006) and Bai (2009), and can give rise to 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 between them. For a detailed discussion and examples of such proximity, see e.g. Conley (1999) and Pinkse, Slade and Brett (2002). Another class of models stems from the spatial autoregressive (SAR) model of Cliff and Ord (1968, 1981), see e.g. Lee (2002, 2004), Kelejian and Prucha (1998, 1999), and employs spatial weight matrices whose elements consist of inverse pairwise economic distances between agents, whence the dependent variable or disturbance of a given unit is assumed to be affected by a weighted average of the dependent variable or disturbance of the other sampled units. The weights are presumed known and reflect the proximity between agents, leaving a small number of parameters to be estimated. The SAR model has gained popularity in empirical works, see e.g. Arbia (2006). Alternatively, mixing conditions, familiar from the time series literature, have been employed. Conley (1999), 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, Shen and Slade (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 units is permitted, as well as strong dependence analogous to long memory in time series, which is ruled out by mixing conditions and can also accommodate economic distances, as well as lattice or irregularly-spaced data. This approach's ability to cover both weak and strong dependence in the disturbances and regressors allows the development of a fairly general of theory.

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 partly 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), while for weakly dependent time series data, Chen and Shen (1998) and Chen, Liao and Sun (2011) offer a rather complete treatment of asymptotic theory and robust inference of general sieve M-estimation.

The present paper presents an asymptotic theory for nonparametric and semiparametric series estimation that covers quite general cross-sectional heterogeneity and dependence, including weak and strong dependence. The conditions of the paper, while designed for spatial settings, lend themselves also to time series, spatio-temporal data, 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 the use of a studentization method that offers an alternative to the existing variance estimation literature in spatial settings. In the spatial context, an extension of HAC (heteroscedasticity and autocorrelation consistent) estimation familiar from the time series literature, see e.g. Hannan (1957), is possible if additional information is available, such as the economic distances between units. Conley (1999) considered HAC estimation under a stationary random field with measurement error in distances, Kelejian and Prucha (2007) for SAR-type models, and Robinson and Thawornkaiwong (2010) in a semiparametric regression set-up. However, in time series settings small sample performance of HAC estimation can be poor and an alternative studentization that can produce more accurately sized tests was suggested by Kiefer, Vogelsang and Bunzel (2000). The present paper provides theoretical justification for employing such a studentization in spatial or spatio-temporal data.

The paper is structured as follows. In Section 2, the model setting is outlined. In Section 3, series estimation is introduced and a uniform rate of convergence for the nonparametric component is established. Section 4 contains asymptotic normality results. Section 5 presents sufficient conditions for $\sqrt{n}$ convergence of certain semiparametric estimators, with data-driven studentization. Using the semiparametric partly linear regression model, Section 6 presents a Monte Carlo study of finite sample performance and two empirical examples. Section 7 concludes. The Appendix contains the proofs.

2. Model setting

This paper commences from the nonparametric regression model

$$Y_i = m(X_i) + U_i, \quad i = 1, 2, \cdots, n, \tag{1}$$

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

$$U_i = \sigma(X_i) e_i, \quad e_i = \sum_{j=1}^{\infty} b_{ij} \varepsilon_j, \quad \sum_{j=1}^{\infty} b_{ij}^2 < \infty \quad i = 1, 2, \cdots, n, \tag{2}$$

where $\sigma : \mathcal{X} \to \mathbb{R}$, the b_{ij} are real constants, and $\{\varepsilon_j, j \geq 1\}$ is a sequence of independent random variables with zero mean and unit variance, independent of $\{X_j, j \geq 1\}$. We regard m and σ as nonparametric functions, and the b_{ij} as unknown. The dependence on both i and j of b_{ij}, rather than just their difference $i - j$, distinguishes (2) from representations for stationary time series, and we also allow $b_{ij} = b_{ijn}$ to depend on n. Both these aspects are important in enabling 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. The e_i, and thence U_i and Y_i, can thus form triangular arrays (and our proofs allow this of X_i and ε_j also), but we suppress the additional n subscript for ease of notation. In SAR models, where the b_{ij} tend to reflect economic distances between agents, we have, for all i, $b_{ij} = 0$ for $j > n$, but in models featured in the spatial statistics literature for stationary observations on $\mathbb{Z}^D$ or $\mathbb{R}^D$, with D denoting the spatial dimension, it is frequently natural to allow b_{ij} to be non-zero for all j, analogously to autoregressive time series models; indeed the square-summability condition in (2) only ensures e_i has finite variance, so long range dependence in e_i is potentially permitted. The factor $\sigma(X_i)$ in e_i allows for a degree of conditional or unconditional heteroscedasticity in U_i. In Section 3 we qualify (1) and (2) by detailed regularity conditions, including restrictions on the dependence and heterogeneity of X_i.

Under (1) and (2) , $m(x) = E(Y_i|X_i = x)$ for $x \in \mathcal{X}$. We will thence 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), Newey (1997), where $a(m)$ can be estimated by $a(\hat{m})$. There are many applications in which a known functional $a(m)$ is of interest. Simple examples include the value of m at multiple fixed points $(x_1, \cdots, x_d) \in \mathcal{X}^d$, $a(m) = (m(x_1), \cdots, m(x_d))'$, and the value of the partial derivative at fixed points,

$$a\,(m) = \Big(\frac{\partial m(\mathrm{x})}{\partial \mathrm{x}_\ell}\Big|_{x_1}, \cdots, \frac{\partial m(\mathrm{x})}{\partial \mathrm{x}_\ell}\Big|_{x_d}\Big)',$$

where x_ℓ the ℓ^{th} element of x. As an example of a nonlinear functional a, Newey (1997) took Y_i to be log consumption and $X_i = (\log p_i, \log I_i)'$, a 2×1 vector of log price and log income, with the demand function at a fixed point $X_i = x$ given by $\exp(m(x))$, whereas approximate consumer surplus is the integral of the demand function over a range of prices,

$$a(m) = \int_{\underline{p}}^{\bar{p}} \exp\big(m(\log t, \log \bar{I})\big)\,dt,$$

for a fixed income $\bar{I}$. For approximate consumer surplus at multiple fixed values of income, $a(m)$ would take a vector form. Another example of $a(\cdot)$, in the partly linear regression model, will be discussed in detail in Section 5.

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 regressors without too much difficulty. Newey (1997) established a uniform rate of consistency for $\hat{m}(x) - m(x)$ and asymptotic normality of $a(\hat{m}) - a(m)$ when X_i and U_i are independent and identically distributed (iid) and $a(g)$ 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), Chen, Liao and Sun (2011) considered these issues for sieve extreme estimates with weakly dependent time series, with rules of inference that are robust to weak dependence. They also indicated that for certain cases of slower-than-$\sqrt{n}$ rate of convergence such as when $a(m) = m$, the asymptotic variance of coincides with that obtained under independence, as found for kernel estimation by Robinson (1983), for example.

3. Estimation of m and uniform consistency rate

The estimation of m is based on approximating functions $p_s(\cdot) : \mathcal{X} \to \mathbb{R}$, $s = 1, 2, \cdots$. A deterministic sequence of positive integers $K = K_n$, nondecreasing in n, denotes the number of $p_s(\cdot)$ used. Its choice gives rise to a bias/variance trade-off, for a given choice of the $p_s(\cdot)$, increasing K reduces bias while increasing variance. of the estimate $\hat{m}$. Denoting $p^K(\cdot) = (p_1(\cdot), \cdots, p_K(\cdot))'$, $p = p_n = [p^K(X_1), \cdots, p^K(X_n)]'$, let

$$\hat{\beta} = (p'p)^{-}p'Y$$

where $Y = (Y_1, \cdots, Y_n)'$ and A^{-} denotes the Moore-Penrose pseudo-inverse of a matrix A.

Denote a series estimate of $m(x)$by

$$\hat{m}(x) = p^K(x)'\hat{\beta}. \tag{3}$$

In order to establish a uniform consistency rate of $\hat{m}(x)$. we introduce the following conditions.

Assumption A1. For $n = 1, 2, \cdots$, *the* X_i $i = 1, ..., n$, *are iid with probability density function* $f(x)$, $x \in X$, *and for* $i \neq j$, X_i *and* X_j *have joint probability density function* $f_{ij}(x, y)$, $x, y \in X$, *and* $\{X_i\}_{i=1}^n$, *is independent of* $\{\varepsilon_j\}_{j=1}^\infty$.

Assumption A2. U_i *satisfies (2) for bounded positive* $\sigma(x)$, *and independent* ε_j *satisfying* $E(\varepsilon_j^2) = 1$ *and* $\max_{j \geq 1} E|\varepsilon_j|^{2+\nu} < \infty$, *for some* $\nu > 0$.

For $k \geq 1$, define the $k \times k$ matrix

$$B_k = E(p^k(X_i)p^k(X_i)'), \quad k = 1, 2, \cdots . \tag{4}$$

Let $\underline{\lambda}(A)$ and $\bar{\lambda}(A)$ denote the minimal and maximal eigenvalues of a non-negative definite real matrix A, and for any real matrix A define the spectral norm$\|A\| = \bar{\lambda}^{1/2}(A'A)$. For a function $g : x \in \mathcal{X}$, define the uniform norm $|g|_\infty = \sup_{x \in \mathcal{X}} |g(x)|$. As in Andrews (1991) and Newey (1997), define

$$\xi(k) = \sup_{x \in \mathcal{X}} \|p^k(x)\|, \quad k = 1, 2, \cdots .$$

If m is known to be bounded, one may choose bounded and non-vanishing $p_s(\cdot)$, whence $\xi(k)$ increases at rate $\sqrt{k}$: $\sup_{x\in\mathcal{X}} \|p^k(x)\| = \sup_{x\in\mathcal{X}} \big(\sum_{i=1}^{k} p_i^2(x)\big)^{1/2} \le C\sqrt{k}$. It is sometimes possible to obtain the rate of $\xi(k)$ explicitly in terms of k. Newey (1997) provides examples where under suitable conditions, $\xi(k) = k$ when the $p_s(\cdot)$ are orthogonal polynomials, and $\xi(k) = k^{1/2}$ when they are B-splines.

Assumption A3. *(i) There exists $c > 0$ such that $\underline{\lambda}(B_k) \ge c, \quad \forall k \ge 1$.*

(ii) K and $p^K(\cdot)$ are such that $K^2\xi^4(K) = o(n)$.

Assumption A3(i) requires B_k to be nonsingular for all k and was also assumed by Andrews (1991) and Newey (1997); it requires any redundant $p_s(\cdot)$ to be eliminated. Assumption 3(ii) imposes an upper bound on the rate of increase of $\xi(K)$ as $K \to \infty$. Using the explicit bounds $\xi(K)$ mentioned before the assumption, A3 (ii) reduces to $K = o(n^{1/4})$ for B-splines and $K = o(n^{1/6})$ for orthonormal polynomials, under suitable conditions.

Assumption A4. *There exist a sequence of vectors β_K and a number $\alpha > 0$ satisfying*

$$|m - p^{K\prime}\beta_K|_\infty = O(K^{-\alpha}), \ \text{as } K \to \infty,$$

Assumption A4, which is standard in the series estimation literature, see e.g. Andrews (1991) and Newey (1997), can be seen as a smoothness condition on $m(\cdot)$, if the $p_s(\cdot)$ are ordered so that higher values of s correspond to less smooth functions, whence, the smoother $m(\cdot)$, the faster the decay of the coefficients of the vector β_K. Some further insights into Assumption A4 for certain choices of the $p_s(\cdot)$, including polynomials, trigonometric polynomials, splines and orthogonal wavelets, can be found in Chen (2007, pp. 5573). Assumption A4 controls the bias of $\hat{m}$, and α is also related to the number of the regressors. Newey (1997) pointed out that for splines and power series, Assumption A4 is satisfied with $\alpha = s/q$ where s is the number of continuous derivatives of m.. Conditions imposing an upper bound on the rate of increase in K, such as A3 (ii), may necessitate stronger smoothness of m.

Next we will introduce an assumption that is required to control the strength of dependence in the X_i across i. Define

$$\triangle_n = \sum_{i,j=1, i\neq j}^{n} \int_{\mathcal{X}^2} |f_{ij}(x,y) - f(x)f(y)| dx dy. \tag{5}$$

The rate of growth of $\triangle_n$ is a measure of bivariate dependence across the X_i, and has upper bound $2n^2$, whereas $\triangle_n = 0$ in case of independence across i. We may view the condition $\triangle_n = O(n)$ as an analogue to the concept of weak dependence in the time series literature. Quantities of a similar nature were used by Robinson (2011) and Robinson and Thawornkaiwong (2012). For Gaussian X_i, $\triangle_n$ has a simple upper bound. Denoting $\sigma_{ij}^{(X)} = Cov(X_i, X_j)$, and assuming for simplicity that $\sigma_{ii}^{(X)} = \sigma_i^{(X)} = 1$, if , for some $c_0 < 1$, $|\sigma_{ik}^{(X)}| \le c_0$, $i, k = 1, \cdots, n$; $i \ne k$, $n \ge 1$, then

$$\triangle_n \le C \sum_{i,k=1, i\ne k}^{n} |\sigma_{ik}^{(X)}|, \ n \ge 1, \tag{6}$$

see Proposition 1 in Appendix B. Clearly, if $\max_{1\le k\le n} \sum_{i=1}^{n} |\sigma_{ik}^{(X)}| \le Cn$, then $\triangle_n = O(n)$.

Assumption A5. *As* $n \to \infty$, $n^{-2}K^2\xi^4(K)\triangle_n = o(1)$.

Assumption A5 indicates that the stronger the dependence in X_i the smaller is the required K. In light of Assumption A4, this necessitates a stronger assumption on the smoothness of m. When $\triangle_n = O(n)$, A5 reduces to $K^2\xi^4(K) = o(n)$, which is stated in A3(ii). Otherwise, A5 is a stronger condition on the upper bound of the growth in K and $\xi(K)$ than A3(ii).

To state our first theorem, further notation is needed. Define normalised functions $P^k(x) = B_k^{-1/2} p^k(x)$ with B_k as in (4) such that $E(P^k(X_i)P^k(X_i)') = I_k$. We write $P(x) = P^K(x)$, suppressing the superscript K in the sequel for ease of notation. Note that $P(\cdot) = [P_{1K}(\cdot), \cdots, P_{KK}(\cdot)]'$, with the double subscripts in $P_{sK}(\cdot)$ arising from the definition $P(\cdot) = B_K^{-1/2} p^K(\cdot)$. Such normalised functions were also used in Newey (1997). Let $P = P_n = (P(X_1), \cdots, P(X_n))' \in \mathbb{R}^{n\times K}$. For a given sequence $K = K_n$, define the following $K \times K$ variance-covariance matrix Σ_n of the $K \times 1$ vector sum $\sum_{i=1}^{n} P(X_i)U_i/\sqrt{n}$:

$$\begin{aligned}
\Sigma_n = \quad & E(P'UU'P/n) = Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} P(X_i)U_i\right) \\
= \quad & \frac{1}{n}\sum_{i,k=1}^{n} E\big(P(X_i)U_iU_kP'(X_k)\big) \\
= \quad & \frac{1}{n}\sum_{i,k=1}^{n} \gamma_{ik} E(\sigma(X_i)\sigma(X_k)P(X_i)P'(X_k)),
\end{aligned} \tag{7}$$

where

$$\gamma_{ik} := Cov(\sum_{j=1}^{\infty} b_{ij}\varepsilon_j, \sum_{j=1}^{\infty} b_{kj}\varepsilon_j) = \sum_{j=1}^{\infty} b_{ij}b_{kj}.$$

The following theorem obtains a uniform rate of convergence.

Theorem 1 *Under Assumptions A1-A5,*

$$\sup_{x \in \mathcal{X}} |\hat{m}(x) - m(x)| = O_p\left(\xi(K)\left[\sqrt{\frac{tr(\Sigma_n)}{n}} + K^{-\alpha}\right]\right), \quad as \quad n \to \infty.$$

The rate obtained coincides with that of Newey (1997) for iid X_i and U_i, in which case $\Sigma_n = \sigma^2 E(P(X_i)P(X_i)') = \sigma^2 I_K$, leading to $tr(\Sigma_n) = O(K)$. The first term in square brackets reflects the contribution of the variance of $\hat{m}$, while the second term reflects the bias.

The rates obtained in Theorem 1 need to be verified to be $o_p(1)$ to establish uniform consistency of $\hat{m}$. The requirement $\xi(K)K^{-\alpha} = o(1)$ of negligible bias suggests that it may be desirable to choose bounded $p_s(\cdot)$. To evaluate the variance contribution, suppose for the time being that the original $p_s(\cdot)$ and thus the normalized functions $P_{1K}, \cdots, P_{KK}$, are uniformly bounded. Then, $tr(\Sigma_n) = K \cdot \sum_{i,k=1}^{n} \gamma_{ik}/n$, making the variance contribution $\xi(K)\sqrt{K}\big(\sum_{i,k=1}^{n} \gamma_{ik}/n^2\big)^{1/2}$. Under weak dependence of the e_i, $\sum_{i,k=1}^{n} \gamma_{ik} = O(n)$, meaning the rate becomes $\xi(K)\sqrt{K/n} = K/\sqrt{n}$ which is $o(1)$ by Assumption A3 (ii). Under strong dependence of the e_i, the rate is slower and further conditions restricting the increase of K and $\xi(K)$ may be needed to show uniform consistency. In the iid setting, the uniform rate obtained by Newey (1997) was improved by de Jong (2002), under the additional assumption of compact $\mathcal{X}$.

4. Asymptotic normality

Our ultimate interest lies in inference on the functional $a(m)$, for which a central limit theorem is the first step. Denoting $\theta_0 = a(m)$ and $\hat{\theta} = a(\hat{m})$, we obtain in this section the asymptotic distribution of $\hat{\theta} - \theta_0$, which requires further assumptions. Recall that $a(\cdot)$ is a vector-valued functional operator.

Assumption B1. *One of the following two assumptions holds.*

(i) *$a(g)$ is a linear operator in g.*

(ii) *For some $\epsilon > 0$, there exists a linear operator $D(g)$ and a constant $C = C_\epsilon < \infty$ such that $\|a(g) - a(m) - D(g - m)\| \le C(|g - m|_\infty)^2$, if $|g - m|_\infty \le \epsilon$.*

Assumption B2. *For some $C < \infty$, $D(\cdot)$ of Assumption B1 satisfies $\|D(g)\| \le C|g|_\infty$.*

Assumptions B1 and B2 are taken from Newey (1997). Assumption B2 requires $D(\cdot)$ to be continuous, which follows from the fact that $D(\cdot)$ is the Frechet-differential of $a(\cdot)$ at m : a functional $a(\cdot)$ is Frechet-differentiable at m if there exists a bounded linear operator $D(\cdot)$ such that, for any $\delta > 0$, there exists $\epsilon > 0$ such that $\|a(g) - a(m) - D(g-m)\| \leq \delta |g-m|_\infty$ if $|g-m|_\infty \leq \epsilon$. Assumption B1(ii) imposes a stronger smoothness condition on $a(\cdot)$ at m than Frechet differentiability. It is not restrictive, see e.g. its verification for some $a(\cdot)$ in Newey (1997, pp. 153). When $a(\cdot)$ is a linear operator, its Frechet-derivative is itself, $D(g) = a(g)$.

With $D(\cdot)$ as in Assumption B1, and the $K \times 1$ vector of normalised functions $P(\cdot)$ as defined above, define the $K \times d$ matrix

$$A = (D(P_{1K}), \cdots, D(P_{KK}))' \in \mathbb{R}^{K \times d}.$$

Consider a linear operator $a(g) = (g(x_1), \cdots, g(x_d))'$, for some $(x_1, \cdots, x_d) \in \mathcal{X}^d$. The linearity of $a(g)$ yields $a(P_{sK}) = D(P_{sK}) = \big(P_{sK}(x_1), \cdots, P_{sK}(x_d)\big)'$, $s = 1, \cdots, K$.

Denote by $\bar{V}_n$ the $d \times d$ conditional variance matrix of the sum $\sum_{i=1}^{n} A' P(X_i) U_i / \sqrt{n}$,

$$\bar{V}_n = Var\left(\sum_{i=1}^{n} A' P(X_i) U_i / \sqrt{n} | X_1, \cdots, X_n \right) = \frac{1}{n} \sum_{i,k=1}^{n} \gamma_{ik} \sigma(X_i) \sigma(X_k) A' P(X_i) P'(X_k) A.$$

To gain insight into $\bar{V}_n$ and its role in our statement of the asymptotic distribution, note that one may alternatively write

$$\bar{V}_n = A^{*\prime} B_K^{-1} \left[\frac{1}{n} \sum_{i,k=1}^{n} \gamma_{ik} \sigma(X_i) \sigma(X_k) p^K(X_i) p^{K\prime}(X_k) \right] B_K^{-1} A^*,$$

where $A^* = (D(p_1), \cdots, D(p_K))' = B_K^{1/2} A \in \mathbb{R}^{K \times d}$, the matrix of Frechet-derivatives of the original series functions. One sees that the matrix $\bar{V}_n$ takes the form of the conditional variance matrix of a nonlinear function of least squares estimates, where their conditional variance matrix (for a possibly misspecified model) is sandwiched between $A^{*'}$ and A^*. Assumption B3 below specifies conditions under which $\bar{V}_n$ is the correct normalising matrix to be used in Theorem 2 below. Two alternative representations of $\bar{V}_n$ in terms of $P(\cdot)$ or $p^K(\cdot)$ have been given; in statements of assumptions and theorems, quantities will be written in terms of $P(\cdot)$ to facilitate discussion of $\bar{V}_n$.

Assumption B3. *As* $n \to \infty$*,*

$$\text{(i)} \quad \xi^2(K)tr(\Sigma_n) = o(n^{1/2}).$$

$$\text{(ii)} \quad K^3\xi^6(K)tr(\Sigma_n)\left(\frac{1}{n} + \frac{\triangle_n}{n^2}\right) = o(1).$$

$$\text{(iii)} \quad n\xi^2(K)K^{-2\alpha+1} = o(1).$$

Assumption B3 combines various conditions on the rate of increase of K, $\xi(K)$, $tr(\Sigma_n)$ and $\triangle_n$ as $n \to \infty$. The rate of increase of $tr(\Sigma_n)$ depends on that of K and the strength of dependence in U_i and X_i. Theorem 1 required a smoothness condition $\xi(K)K^{-\alpha} = o(1)$ on m, while Theorem 2 will need the stronger smoothness condition B3 (iii). Revisiting the case of bounded functions $P_{sK}(\cdot)$ and weakly dependent e_i leading to $tr(\Sigma_n) = O(K)$, note that B3 (i) is implied by A3 (ii), while B3 (ii) becomes $K^4\xi^6(K) = o(n)$, which implies A3 (ii).

Assumption B4. *The bandwidth* K *and series functions* $p^K(\cdot)$ *are such that, as* $n \to \infty$*,*

$$\frac{\xi^2(K)}{\sqrt{n}} \max_{1\le j\le n}\left\{\sum_{i=1}^{n}|b_{ij}|\right\} = o(1).$$

Assumption B4 requires the influence of ε_j for any j on $U_i, i = 1, 2, \cdots$ to die off more quickly if $\xi(K)$ grows faster.

Assumption B5. *As* $n \to \infty$*,* $\|\bar{V}_n^{-1}\| = O_p(1)$*.*

Assumption B5 trivially holds when the random matrix $\bar{V}_n$ converges in probability to a finite nonsingular matrix, as considered in the following section's discussion of $\sqrt{n}$ -convergence, whose validation requires stronger restrictions both on $a(\cdot)$ and the strength of dependence in the X_i and U_i. Theorem 3 allows $\bar{V}_n$ to diverge with n as long as $\|\hat{V}_n - V_n\| = o(1)$ for some sequence of deterministic nonsingular matrices V_n, which still requires, albeit weaker, dependence restrictions. We present Theorem 2 separately from Theorem 3, to separate assumptions yielding asymptotic normality from those required for $\|\hat{V}_n - V_n\| = o_p(1)$. Assumption B5 assumes the derivative matrix A to have rank d for all $K \ge d$. Throughout we denote by $B^{1/2}$ the unique positive definite square root of a positive definite matrix B.

Theorem 2 *Under assumptions A1-A5 and B1-B5,*

$$\sqrt{n}\bar{V}_n^{-1/2}(\hat{\theta} - \theta_0) \to_d N(0, I_d), \quad as \quad n \to \infty. \tag{8}$$

The proof of Theorem 2 is given in Appendix A.

Defining

$$V_n = E(\bar{V}_n) = Var\left(\sum_{i=1}^{n} A'P(X_i)U_i/\sqrt{n}\right) = \frac{1}{n}\sum_{i,k=1}^{n}\gamma_{ik}E\big[\sigma(X_i)\sigma(X_k)A'P(X_i)P'(X_k)A\big],$$

we study conditions under which $\|\bar{V}_n - V_n\|$ converges in probability to zero, allowing $\bar{V}_n$ to be replaced by V_n in (8). In Theorem 3 below, the i^{th} element of $\hat{\theta}$ is shown to be $\sqrt{n}(V_n^{-1/2})_{ii}$-consistent, where $(V_n^{-1/2})_{ii}$ denotes the i^{th} diagonal element of $V_n^{-1/2}$. To gain some intuition of implications of this rate, we focus in this paragraph on scalar $a(\cdot)$. We rule out the possibility of shrinking V_n, which corresponds to presence of negative dependence in X_i or U_i, which is relatively unlikely for real data. The above expression for V_n suggests that $V_n = O(1)$ corresponds to short range dependence in $A'P(X_i)U_i$ if K were fixed. This may still allow for the possibility of long range dependence in $A'P(X_i)$ or U_i to a certain degree. With increasing K, V_n may be increasing even under short range dependence of $A'P(X_i)U_i$. The main contribution of this paper is developing inference procedures when V_n is unknown and deriving asymptotic distribution results under additional generality in the strength of dependence in both X_i and U_i.

The following two conditions restrict the strength of dependence in the X_i and U_i across i. Again, an upper bound is imposed on $\triangle_n$.

Assumption B6. *As $n \to \infty$,*

$$\frac{\xi^8(K)(n+\triangle_n)}{n^2}\Big(\max_{1\le j\le n}\sum_{i=1}^{n}\left|\gamma_{ij}\right|\Big)^2 = o(1).$$

Assumption B6 indicates how the dependence in the data restricts the choice of K and the $p_s(.)$. The stronger the dependence, the slower the rate of increase in K and $\xi(K)$, leading to further repercussions on the smoothness condition in Assumption B3 (iii), where a larger value of α would be needed to compensate for slower rate of growth in K.

Next we state an assumption on the strength of dependence in X_i across i in terms of 4th joint cumulants. Recalling that $A = (A_1, \cdots, A_K)' \in \mathbb{R}^{K\times d}$, introduce the following notations:

$$\begin{aligned} h_i^{(\ell)} &= \sigma(X_i)A_\ell' P(X_i), \\ \bar{h}_i^{(\ell)} &= \sigma(X_i)A_\ell' P(X_i) - E\left(\sigma(X_i)A_\ell' P(X_i)\right), \quad 1\le i\le n, \quad 1\le \ell \le d, \end{aligned} \tag{9}$$

so $\bar{h}_i^{(\ell)}$ is a de-meaned version of $h_i^{(\ell)}$.

Assumption B7. *$E\big[(\bar{h}_i^{(\ell)})^4\big] < \infty$ and $\kappa(\bar{h}_{i_1}^{(\ell)}, \bar{h}_{i_2}^{(p)}, \bar{h}_{i_3}^{(\ell)}, \bar{h}_{i_4}^{(p)})$ are such that*

$$\max_{1\le \ell,p\le d}\frac{1}{n^2}\left|\sum_{i_1,i_2,i_3,i_4=1}^{n}\gamma_{i_1i_2}\gamma_{i_3i_4}\kappa(\bar{h}_{i_1}^{(\ell)}, \bar{h}_{i_2}^{(p)}, \bar{h}_{i_3}^{(\ell)}, \bar{h}_{i_4}^{(p)})\right| = o(1), \tag{10a}$$

where $\kappa(.,.,.,.)$ denotes the fourth joint cumulant of its arguments.

This assumption is not restrictive and may allow strong dependence in both X_i and U_i. One can have arbitrarily strong dependence in U_i if the $\bar{h}_i^{(\ell)}$ are weakly dependent in fourth cumulants, upper bounding the left side of (10a) by

$$C\max_{1\le \ell,p\le d}\frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n}|\kappa(\bar{h}_{i_1}^{(\ell)}, \bar{h}_{i_2}^{(p)}, \bar{h}_{i_3}^{(\ell)}, \bar{h}_{i_4}^{(p)})|,$$

since $|\gamma_{ik}| \le C < \infty$.

Assumption B8. *As $n \to \infty$, $\|V_n^{-1}\| = O(1)$.*

The following theorem may be compared with Theorem 2.

Theorem 3 *Under Assumptions B7-B8,*

$$\|\bar{V}_n^{-1}\| = O_p(1), \tag{11}$$

$$\|\bar{V}_n - V_n\| = o_p(1). \tag{12}$$

Consequently, under Assumptions A1-A5 and B1-B7,

$$\sqrt{n}V_n^{-1/2}(\hat{\theta}-\theta_0) \to_d N(0, I_d). \tag{13}$$

The following section develops statistical inference on the basis of Theorem 3.

5. $\sqrt{n}-$ inference

We establish sufficient conditions for V_n to converge to a finite limit V, as $n \to \infty$, whence Theorem 3 implies $\sqrt{n}-$convergence of $\hat{\theta}$ to θ_0. Attainment of this parametric rate by semiparametric estimates has received wide interest in the econometric literature, following Robinson (1988) and Powell, Stock and Stoker (1989), who used kernel estimation. Newey (1997) developed $\sqrt{n}$-convergence of a general semiparametric estimate, using series estimation, while Chen and Shen (1998) considering extensions for weakly dependent time series. It is of interest to develop results in a general cross-sectional dependence setting, since semiparametric estimates, such as for the partly linear regression model, are widely used in empirical work. For statistical inference this requires a studentization that is robust against general spatial dependence and heterogeneity. The present section discusses $\sqrt{n}$-convergence and studentization.

5. 1 $\sqrt{n}$ - convergence

The following assumption, from Newey (1997), states the key condition for $\sqrt{n}-$convergence.

Assumption C1. *There exists a* $d\times 1$ *vector-valued function* $w(x) = (w_1(x), \cdots, w_d(x))'$ *with the following properties.*

(i) $E[w(X_i)w'(X_i)]$ *is finite and nonsingular,*

(ii) $D(m) = E[w(X_i)m(X_i)]$, $D(P_{sK}) = E[w(X_i)P_{sK}(X_i)], 1 \leq s \leq K$ *for all* K,

(iii) $E[\|w(X_i) - \delta_K P(X_i)\|^2] \to 0$ *for some sequence of fixed* $d \times K$ *matrices* δ_K.

Sufficient conditions for Assumption C1 can be found in Newey (1997, pp. 155). The vector-valued function $w(\cdot)$ is the element of the domain of $D(\cdot)$ used in the Riesz representation of $D(\cdot)$. Assumption C1 (iii) requires such $w(\cdot)$ to lie in the linear span of the series functions. Newey (1997) verified that Assumption C1 holds for the semiparametric estimands in the partly linear and single index models and also for the case of average consumer surplus estimation, where the quantity of interest is the approximate consumer surplus integrated over a range of income (as discussed in Section 2).

By Assumption C1, $D(P_{sK}) = E[w(X_i)P_{sK}(X_i)]$, $1 \leq s \leq K$, so one can write $A = E[P(X_i)w'(X_i)]$. Since the $K \times 1$ vector of normalized functions $P(\cdot)$ satisfies $E[P(X_i)P'(X_i)] = I_K$, $A'P(x)$ can be written as the mean square projection of $w(x)$ on the $K \times 1$ vector $P(\cdot)$ of approximating functions:

$$A'P(x) = A'I_K^{-1}P(x) = E[w(X_i)P'(X_i)]E[P(X_i)P'(X_i)]^{-1}P(x).$$

Denote $d \times 1$ vector $A'P(x) =: v_K(x) = (v_{1K}(x), \cdots, v_{dK}(x))'$, with the subscript K indicating that v_K is a mean-square projection of w onto the linear space spanned by K series functions. Then V_n can be written as

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

Next, replace $v_K(\cdot)$ by the function $w(\cdot)$ to give the matrix

$$W_n = \frac{1}{n}\sum_{i,k=1}^{n} \gamma_{ik} E[\sigma(X_i)\sigma(X_k)w(X_i)w'(X_k)].$$

The following assumption provides sufficient conditions for $\sqrt{n}$ -convergence of $a(\hat{m})$ to $a(m)$.

Assumption C2. (i) $V = \lim_{n\to\infty} W_n$ *exists;* (ii) $\sum_{i,k=1}^{n} |\gamma_{ik}| = O(n)$.

Existence of the limit V is a condition imposed on the collective strength of dependence in U_i and X_i, comparable to Assumption A4 of Robinson and Thawornkaiwong (2012). Assumption C2 (ii) is a weak dependence restriction on e_i.

Theorem 4. *Under Assumptions C1 and C2,*

$$V_n \to V < \infty, \quad as \quad n \to \infty.$$

Consequently, under Assumptions A1-A5, B1-B7, and C1-C2,

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, V), \quad as \quad n \to \infty.$$

Theorem 4 obtains $\sqrt{n}$ - convergence for certain semiparametric estimates under weak dependence. The following sub-section discusses estimation of V, needed for interval estimation or hypothesis testing.

5.2 Studentization

The Introduction discussed difficulties with HAC estimation in dealing with cross-sectional dependence. It is possible, however to studentize $a(\hat{m}) - a(m)$ without employing any particular dependence structure, following the approach of Kiefer, Vogelsgag and Bunzel (2000)

Recalling the definitions $B_K = E(p^K(X_i)p^K(X_i)')$, $P(x) = B_K^{-1/2}p^K(x)$, $A = (D(P_{1K}), \cdots, D(P_{KK}))' \in \mathbb{R}^{K \times d}$, $A^* = (D(p_1), \cdots, D(p_K))' = B_K^{1/2}A \in \mathbb{R}^{K \times d}$ with $D(\cdot)$ given in Assumption B1(i), introduce the estimates of A^* and $B_{K:}$.

$$\hat{A}^* = \frac{\partial a(p^{K\prime}\beta)}{\partial \beta}\Big|_{\beta=\hat{\beta}}, \quad \hat{B}_K = p'p/n = \sum_{i=1}^{n} p^K(X_i)p^K(X_i)'/n. \tag{14}$$

We thence construct the sample analogue $\hat{A}^{*\prime}\hat{B}_K^{-1}p'\hat{U}/\sqrt{n}$ of $A'P'U/\sqrt{n}$, where, for $\hat{M} = \big(\hat{m}(X_1), \cdots, \hat{m}(X_n)\big)$, $\hat{U} = Y - \hat{M}$. Next set

$$\hat{S}^*_{n,m} = \sum_{i=1}^{m} \hat{A}^{*\prime}\hat{B}_K^{-1}p^K(X_i)\hat{U}_i/\sqrt{n}, \quad 1 \le m \le n,$$

and define

$$\hat{C}_n = \frac{1}{n}\sum_{m=1}^{n} \hat{S}^*_{n,m}\hat{S}^{*\prime}_{n,m}, \quad \Psi_d = \int_0^1 [W_d(r) - rW_d(1)][W_d(r) - rW_d(1)]'dr,$$

where $W_d(\cdot)$ denotes a d-dimensional vector of independent Brownian motions, so Ψ_d is the integral of the outer product of the d-dimensional multivariate Brownian bridge. Recall that $EW_d(r)W_d(u)' = rI, \quad 0 \le r \le u \le 1$.

Assumption C3. *(i)* $\sum_{i=1}^{[rn]} \sum_{k=[rn]+1}^{n} |\gamma_{ik}| = o(n)$ *uniformly in* $r \in [0,1]$*;*

$$\text{(ii) } \max_{1 \le i \le n} \sum_{k=1}^{n} |\gamma_{ik}| = O(1).$$

Assumption C2 (ii) used in Theorem 4 required e_i to be weakly dependent. Assumption C3 (ii) further rules out the presence of any "dominant" unit whose error covariances with new units added to the sample are persistently significant. Assumption C3 (i) requires some falling-off of dependence as $|i-k|$ increases, which inevitably necessitates the ordering of the data to carry at least some information of the structure of dependence, albeit with a significant relaxation from the time series setting where dependence is a function of $|i-k|$. Both C3 (i) and (ii) are natural implications of weak dependence in the time series context where dependence is a fast-decreasing function of time lag. Our setting differs in allowing $\gamma_{ik} = \gamma_{ikn}$ to admit a triangular array structure, and in relaxing the link between γ_{ik} and $|i-k|$. For example, Assumption C3(i) is satisfied if there exists a positive function $\eta(\cdot)$ such that $|\gamma_{ik}| \le \eta(i-k), i, k = 1, 2, \cdots$, and $\sum_{j=-\infty}^{\infty} \eta(j) < \infty$ (see Proposition 2 in Appendix B). If γ_{ik} has triangular array structure, as in the pure SAR model for example, then Assumption C3 (i) is potentially more restrictive. In this setting, Assumption C3 (ii) allows a unit i to interact with infinitely many others as the sample increases, so long as the bilateral interaction $\gamma_{ikn}, k = 1, 2, \cdots$, decays suitably fast as n increases, whereas C3 (i) requires a faster uniform-in-n rate of reduction in γ_{ikn} as $|i-k|$ increases. Therefore, our assumptions require that the ordering of data carries some meaning. This rules out random data collection, without any record of how units are related. Another issue is that in spatial settings units cannot be unambiguously ordered, for example if they are observed on a plane. Nonetheless, there are many economic applications where data can be ordered to satisfy Assumption C3. For example, with firm data, one may expect that firms using similar inputs or producing similar outputs would exhibit high correlation in disturbances, the knowledge of which can help order the data. These considerations are pursued in the Monte Carlo study in the following section.

Assumption C4. (i) $\triangle_n = O(n)$; (ii) $tr(\Sigma_n) = O(K)$; (iii) $\bar{\lambda}(B_K) = O(1)$; (iv) $\sqrt{n}\xi^3(K)K^{-\alpha} = o(1)$.

Assumption C4 (i) can be seen as weak dependence condition on the X_i, whereas Assumption C4 (iii) restricts the choice of the $p_s(.)$, requiring second moments to be bounded. Assumption C4 (ii) is a condition on the strength of dependence across i in the combined quantity $P(X_i)U_i$. Assumption C4 (iv) strengthens the smoothness condition of Assumption B3 (iii).

Assumption C5. $E(\varepsilon_j^4) = \kappa < \infty$, $j = 1, 2, \cdots$.

Recall that the functional derivative $D(\cdot)$ from Assumptions B1 and B2 is the Frechet differential of the functional $a(\cdot)$, evaluated at m. Now, let $D(\cdot; g)$ denote the functional derivative of $a(\cdot)$ evaluated at g. Let $D(\cdot; g) = \big(D_1(\cdot; g), \cdots, D_d(\cdot; g)\big)'$.

Assumption C6. *For some* $0 < C, \epsilon < \infty$ *and all* $\tilde{g}, \bar{g}$ *such that* $|\tilde{g} - m|_\infty \leq \epsilon$ *and* $|\bar{g} - m|_\infty \leq \epsilon$, $\|D_i(g; \tilde{g}) - D_i(g; \bar{g})\| \leq C|g|_\infty |\tilde{g} - \bar{g}|_\infty$, $i = 1, \cdots, d$.

Assumption C6 is from Newey (1997) and requires the $D_i(\cdot; g)$ to exhibit continuity over g, where the derivative is taken.

The following theorem finds that the asymptotic distribution of $\hat{\theta} - \theta_0$, when studentized by $\hat{C}_n$, is free from the unknown variance matrix V and only depends on d.

Theorem 5. *Under the assumptions of Theorem 4, and Assumptions C1-C6,*

$$\hat{C}_n^{-1/2} \sqrt{n}(\hat{\theta}_n - \theta_0) \to_d \Psi_d^{-1/2} W_d(1).$$

Now, suppose we are interested in testing the hypothesis $H_0 : a(m) = r$ against the alternative $H_1 : a(m) \neq r$ for a $d \times 1$ fixed vector r. Then we can construct the statistic $t_n^* = n(\hat{\theta} - r)' \hat{C}_n^{-1} (\hat{\theta} - r)$. Since $t_n^* = \|\sqrt{n}(\hat{\theta} - r)' \hat{C}_n^{-1/2}\|^2$, Theorem 5 implies the following result.

Theorem 6. *Under the assumptions of Theorem 5,*

$$\begin{aligned} t_n^* &\Rightarrow W_d(1)' \Psi_d^{-1} W_d(1), \quad \text{under} \quad H_0, \\ t_n^* &\Rightarrow \infty, \quad \text{under} \quad H_1. \end{aligned}$$

The critical values c_α satisfying $Pr(t_n \leq c_\alpha) \to 1 - \alpha$, required to carry out hypothesis tests can be obtained from Table 2 of Kiefer, Vogelsgag and Bunzel (2000) for $d = 1, \cdots, 30$.

6. Numerical results for the partly linear model

The present section contains Monte Carlo investigation of finite-sample performance of our estimates and proposed rules of statistical inference in simulated cross-sectional settings, and applies them to two empirical data sets, in both cases for the partly linear regression, which is discussed in the following sube-section.

6.1. Partly linear regression model

Partly linear regression restricts $m(\cdot)$ in (1) by requiring $m(\cdot)$ that a d-dimensional proper subvector, denoted Z_i, of X_i enters linearly. Denoting by W_i the vector consisting of the remaining elements of X_i, the model can written as

$$Y_i = Z_i'\delta_0 + h_0(W_i) + U_i, \tag{15}$$

where $h_0(\cdot)$ is a nonparametric function and δ_0 is an unknown parameter vector. The 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. This model has received much attention in kernel estimation, see e.g. Robinson (1988) and Fan and Li (1999), where δ_0 can be estimated at $\sqrt{n}$ rate despite first stage nonparametric estimation having a slower-than-$\sqrt{n}$ rate.

Series estimation of (15) had been considered in Chamberlain (1986), where the choice of series functions takes into account the partly linear form. The first d are functions of Z_i only, while the remaining $K-d$ are functions of W_i only. The series estimate of δ_0 is then the first d elements of $\hat{\beta}$, and $\hat{h}(x) = \hat{m}(z,w) - z'\hat{\delta}$. At first glance, the series estimation of δ_0 may seem very different in form from the kernel estimate, where first-stage non-parametric regression estimates of Y_is and Z_i in terms of W_i are required, but they are in fact very similar, as explained below.

As in Robinson (1988), subtracting $E(Y_i|W_i) = E(Z_i|W_i)'\delta_0 + h_0(W_i)$ from (15) yields

$$Y_i - E(Y_i|W_i) = [Z_i - E(Z_i|W_i)]'\delta_0 + U_i.$$

Robinson (1988) thence replaced the nonparametric conditional expectations by kernel estimates, and obtained the least squares estimate

$$\tilde{\delta} = [(Z - \tilde{E}(Z|W))'(Z - \tilde{E}(Z|W))]^{-1}(Z - \tilde{E}(Z|W))'(y - \tilde{E}(y|Z)),$$

with $Z = (Z_1, \cdots, Z_n)'$, $W = (W_1, \cdots, W_n)'$, and $\tilde{E}(Z|W)$ and $\tilde{E}(y|W)$ denoting kernel estimates of the $n \times d$ matrix of conditional expectations $E(Z|W)$ and the $n \times 1$ vector $E(y|W)$. In series estimation, the same operation is implemented, albeit implicitly. Recall that $\hat{\delta}$ is the first d elements of $\hat{\beta}$ such that $\hat{\beta} = (\hat{\delta}', \hat{\lambda}')' = (p'p)^{-}p'Y \in \mathbb{R}^K$, with $p^K(Z_i, W_i) = (Z_i', q(W_i)')'$, where $q(\cdot)$ is the vector of $K-d$ series functions in terms of W_i. Define the $n \times n$ "hat" matrix $M := I - \mathcal{P}(\mathcal{P}'\mathcal{P})^{-}\mathcal{P}'$, for the $n \times (K-d)$ matrix $\mathcal{P} = (q(W_1), \cdots, q(W_n))'$. Then by partitioned regression

$$\hat{\delta} = (Z'MZ)^{-}Z'My.$$

The projections $\mathcal{P}(\mathcal{P}'\mathcal{P})^{-1}\mathcal{P}'Z$ and $\mathcal{P}(\mathcal{P}'\mathcal{P})^{-1}\mathcal{P}'y$ are series estimates of $E(Z|X)$ and $E(y|X)$. Therefore, the series estimate $\hat{\delta}$ of δ_0 effectively takes the same form as the kernel estimate $\tilde{\delta}$ of Robinson (1988), with series estimates of $E(Z|W)$ and $E(y|W)$ replacing corresponding kernel estimates.

Next, we clarify the functional $a(\cdot)$ used to represent δ_0. There is more than one $a(\cdot)$ that yields $a(m) = \delta_0$. Andrews (1991) notes that one could write $a(m) = \partial m(w,z)/\partial z = \delta_0$ for any w, z. Here we use the functional of Newey (1997), since this leads to $\sqrt{n}$-consistency. Denote $\mathtt{Z}^* = \mathtt{Z} - E(\mathtt{Z}|\mathtt{W})$, where $\mathtt{Z}$ and $\mathtt{W}$ are random variables independent of the data used to construct $\hat{\delta}$. Suppose $E(\mathtt{Z}^*\mathtt{Z}^{*\prime})$ is non-singular, which is an identification condition for δ_0, and consider

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

Now

$$\begin{aligned} E(\mathtt{Z}^*\mathtt{Z}^{*\prime}) &= E(\mathtt{ZZ}') - E[E(\mathtt{Z}|\mathtt{X})\mathtt{Z}'] - E[\mathtt{Z}E(\mathtt{Z}'|\mathtt{W})] + E[E(\mathtt{Z}|\mathtt{W})E(\mathtt{Z}'|\mathtt{W})] \\ &= E(\mathtt{ZZ}') - E[E(\mathtt{Z}|\mathtt{W})\mathtt{Z}'] = E(\mathtt{Z}^*\mathtt{Z}'), \end{aligned}$$

since $E[\mathtt{Z}E(\mathtt{Z}'|\mathtt{W})] = E[E(\mathtt{Z}|\mathtt{W})E(\mathtt{Z}'|\mathtt{W})]$ by the law of iterative expectations, and

$$E[\mathtt{Z}^* h_0(\mathtt{W})] = E[\mathtt{Z}h_0(\mathtt{W})] - E[E(\mathtt{Z}|\mathtt{W})h_0(\mathtt{W})] = 0.$$

Thus

$$a(m) = [E(\mathtt{Z}^*\mathtt{Z}^{*\prime})]^{-1}\left\{E(\mathtt{Z}^*\mathtt{Z}')\delta_0 + E[\mathtt{Z}^* h_0(\mathtt{W})]\right\} = \delta_0.$$

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

$$\begin{aligned} a(\hat{m}) &= E\left\{[E(\mathtt{Z}^*\mathtt{Z}^{*\prime})]^{-1}\mathtt{Z}^*\hat{m}(\mathtt{X},\mathtt{W})\right\} \\ &= [E(\mathtt{Z}^*\mathtt{Z}^{*\prime})]^{-1}\left[E(\mathtt{Z}^*\mathtt{Z}')\hat{\delta} + E[\mathtt{Z}^* q(\mathtt{W})']\hat{\lambda}\right] = \hat{\delta}. \end{aligned}$$

6.2 Monte Carlo study of finite sample performance

Our simulations take both W_i and Z_i in (15) to be one-dimensional, and throughout set $\delta_0 = 0.3$ and $h(w) = \log(1 + w^2)$. Two issues we address relate to the difficulty of ordering the data in line with the requirements of Assumption C3. First, there may be noise in our information about the ordering, for example even if one knows which characteristic of individual units underpins the structure of spatial dependence, it may be observed with error. Second, it may not be straightforward to order the data when the underlying dependence structure is complex, for example when units reside on a plane. Our experiments entail two designs that cover the two issues separately.

In the first set of simulations, 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 $s = (s_1, \cdots, s_n)'$, 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 $Y_i = \log(1 + X_i^2) + 0.3Z_i + U_i$.

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

$$\epsilon' = (\epsilon'_1, \cdots, \epsilon'_n)' \sim N(0, 4I_n), \quad \epsilon'' = (\epsilon''_1, \cdots, \epsilon''_n)' \sim N(0, 25I_n),$$
$$\epsilon''' = (\epsilon'''_1, \cdots, \epsilon'''_n)' \sim N(0, 100I_n), \quad \epsilon'''' = (\epsilon''''_1, \cdots, \epsilon''''_n)' \sim N(0, 400I_n),$$

we take

$$s'_i = s_i + \epsilon'_i, \quad s''_i = s_i + \epsilon''_i, \quad s'''_i = s_i + \epsilon'''_i, \quad s''''_i = s_i + \epsilon''''_i, i = 1, ..., n,$$

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, $K = 4, 6, 9$ were tried, and for the series functions of W_i, the first $K - 1$ 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 1 the Monte Carlo MSE, bias and variance of $\widehat{m}$ at a fixed point $(w, z) = (0.5, 0.5)$, the Monte Carlo integrated MSE (MISE) of $\widehat{m}$ to convey global performance, and the MSE of $\hat{\delta}$. The bias and variance of $\widehat{m}(0.5, 0.5)$ are in line with the prediction that larger K reduce bias while increasing variance, while under all values of ρ, $K = 4$ or $K = 6$ led to the smallest MSE for $n = 100$, while $K = 6$ did so for $n = 400$. For the MISE, $K = 4$ was best when $n = 100$, and $K = 6$ was best for $n = 400$. The Monte Carlo MSE of estimate $\hat{\delta}$ was relatively insensitive to K across all 8 settings, which is important as the optimal choice of K for semiparametric estimation is often more difficult than for nonparametric estimation.

Our next objective is to investigate performance of the studentization in Section

5.3. Theorem 5 implies in this setting,

$$n(\hat{\delta}-\delta_0)'\hat{C}_n^{-1}(\hat{\delta}-\delta_0) \to_d W_1^2(1)\sqrt{\Psi_1}, \sqrt{n/\hat{C}_n}(\hat{\delta}-\delta_0) \to_d W_1(1)/\sqrt{\Psi_1}.$$

Kiefer, Vogelsang and Bunzel,2000, Table 2) gave simulated values of the percentiles of $W_1^2(1)/\Psi_1$, from which we derive the 99.5^{th}, 97.5^{th} and 95^{th} percentiles of $W_1(1)/\sqrt{\Psi_1}$ as $\sqrt{101.2}$, $\sqrt{46.39}$ and $\sqrt{28.88}$, respectively. We thence construct the asymptotic 95% confidence interval for δ_0, $\left[\hat{\delta}-\sqrt{46.39\hat{C}_n/n}, \hat{\delta}+\sqrt{46.39\hat{C}_n/n}\right]$. Table 2 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 ρ and is fairly insensitive to K. Similar patterns are observed under perturbed ordering.

Table 3 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 $Var\left(\epsilon_i'''\right) = 100$ and $Var\left(\epsilon_i''''\right) = 400$, the results are encouraging.

Table 4 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 improve with increasing perturbations.

The second set of simulations aims to investigate the implications of ordering spatial data with a single index when the underlying structure is more complex. We generate random locations on a plane, then order based on an ascending distance from the origin (0,0). To generate the data, we follow the random location setting of Robinson and Thawornkaiwong (2012), where the vector of locations of the observations, denoted $s_1, \cdots, s_n$, 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 K , 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 ρ, but also by the distances between locations. The fact that we are considering locations on a plane rather than along a line implies that ρ 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 ρ led this quantity to be of of similar magnitude to that in the AR(1) model with lag-1 autocorrelation $\rho = 0, 0.2, 0.4, 0.6$: for $n = 100$, in our spatial setting it took values $100, 152, 255, 384$ for $\rho = 0, 0.2, 0.4, 0.52$, respectively, which are comparable to $100, 150, 232, 396$ in the AR(1) with $\rho = 0, 0.2, 0.4, 0.6$; for $n = 400$, it took values $400, 611, 949, 1602$ for $\rho = 0, 0.2, 0.4, 0.52$, respectively, which are comparable to $400, 599, 930, 1590$ in the AR(1) with $\rho = 0, 0.2, 0.4, 0.6$.

We report in Table 5 the Monte Carlo MSE, bias and variance of $\widehat{m}(0.5, 0.5)$, MISE of $\widehat{m}$, and MSE of $\hat{\delta}_0$. Again, patterns of bias and variance of $\widehat{m}$ with changing K is in line with predictions, and $K = 4$, 6 generated the lowest MSE for all combinations for $n = 100$, 400 respectively.

As mentioned before, we ordered the data in ascending Euclidean distance from the origin for the purpose of studentization. Table 6 reports Monte Carlo average length of 95% confidence intervals. As before, it decreases with n, increases with ρ and shows little variation across K. Table 7 reports the empirical coverage probabilities for the 99%, 95% and 90% confidence intervals, which seem despite the issues of ordering and dependence. Table 8 reports empirical power of testing $H_0 : \delta_0 = \delta$ against $H_1 : \delta_0 \neq \delta$, for $\delta = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 1$ with 5% significance level. The asymptotic distribution of the test statistic is symmetric, and as expected, powers reported for $\delta = 0.1$ and 0.2 are similar to those reported for $\delta = 0.5$ and 0.4, respectively.

Table 1: Monte Carlo MSE, Variance and Bias

ρ	n	K	$MSE(\hat{m}_x)$	$Var(\hat{m}_x)$	$Bias(\hat{m}_x)$	$MISE(\hat{m})$	$MSE(\hat{\delta})$
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

Table 2: Monte Carlo average 95 % CI length

n	K	$\rho = 0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$
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 3: Coverage Probabilities

			s			s’			s”’			s”’			s””		
ρ	n	K	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 4: Empirical power of 95% test, $K = 6$

		s				s'				s"				s"'				s""			
	$\rho\backslash\delta$	**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

Table 5: Monte Carlo MSE, Variance and Bias

ρ	n	K	$MSE(\hat{m}_x)$	$Var(\hat{m}_x)$	$Bias(\hat{m}_x)$	$MISE(\hat{m})$	$MSE(\hat{\delta})$
0	100	4	0.1884	0.024	0.4054	0.1965	0.0149
		6	0.0315	0.0258	0.0752	0.0587	0.0131
		9	0.0384	0.0384	-0.0029	0.0808	0.0136
	400	4	0.1717	0.006	0.407	0.1785	0.0037
		6	0.016	0.0064	0.098	0.0264	0.0031
		9	0.0081	0.0077	0.0211	0.0213	0.0031
0.2	100	4	0.1891	0.0316	0.3969	0.2009	0.0191
		6	0.0394	0.0334	0.0775	0.0676	0.017
		9	0.0433	0.0432	0.0097	0.0873	0.017
	400	4	0.1707	0.0083	0.403	0.1811	0.004
		6	0.0168	0.0083	0.0924	0.028	0.0035
		9	0.01	0.0099	0.0138	0.0233	0.0034
0.4	100	4	0.198	0.0473	0.3881	0.2107	0.0184
		6	0.0529	0.0475	0.0734	0.0815	0.0179
		9	0.0578	0.0578	0.0028	0.1009	0.018
0.35	400	4	0.177	0.0118	0.4064	0.1827	0.0046
		6	0.0214	0.0115	0.0996	0.0321	0.0042
		9	0.013	0.0126	0.0205	0.0272	0.0042
0.52	100	4	0.2089	0.0558	0.3913	0.2186	0.0225
		6	0.0614	0.0542	0.085	0.0942	0.021
		9	0.0654	0.065	0.0195	0.1146	0.0212
0.5	400	4	0.1778	0.0164	0.4018	0.1878	0.0062
		6	0.0264	0.017	0.0968	0.0387	0.0058
		9	0.0183	0.018	0.0171	0.0343	0.0058

Table 6: Monte Carlo average 95 % CI length

n	K	$\rho = 0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.52$
100	4	0.6241	0.6653	0.6816	0.717
	6	0.5823	0.628	0.6471	0.6776
	9	0.5812	0.624	0.6484	0.6747
n	K	$\rho = 0$	$\rho = 0.2$	$\rho = 0.35$	$\rho = 0.5$
400	4	0.3099	0.3234	0.3483	0.3793
	6	0.2869	0.3026	0.3288	0.361
	9	0.2862	0.3016	0.3271	0.3584

Table 7: Coverage Probabilities

		$\rho = 0$			$\rho = 0.2$			$\rho = 0.4$			$\rho = 0.52$		
n	K	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
100	4	0.901	0.946	0.99	0.881	0.935	0.984	0.892	0.939	0.991	0.876	0.939	0.981
	6	0.904	0.951	0.987	0.889	0.932	0.985	0.874	0.938	0.986	0.879	0.93	0.989
	9	0.888	0.946	0.989	0.885	0.932	0.989	0.884	0.936	0.988	0.877	0.925	0.982
		$\rho = 0$			$\rho = 0.2$			$\rho = 0.35$			$\rho = 0.5$		
n	K	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
400	4	0.9	0.943	0.985	0.905	0.945	0.987	0.904	0.956	0.993	0.881	0.932	0.976
	6	0.895	0.947	0.985	0.912	0.957	0.992	0.896	0.944	0.988	0.884	0.934	0.975
	9	0.901	0.952	0.99	0.911	0.959	0.987	0.892	0.944	0.989	0.882	0.923	0.975

Table 8: Empirical power of 95% test

$n = 100$	$\rho\backslash\delta$	0	0.1	0.2	**0.3**	0.4	0.5	0.7	1
100	0	0.538	0.296	0.13	0.049	0.123	0.299	0.759	0.992
	0.2	0.481	0.291	0.111	0.068	0.108	0.293	0.716	0.966
	0.4	0.481	0.263	0.111	0.062	0.124	0.291	0.686	0.969
	0.52	0.446	0.267	0.115	0.07	0.12	0.273	0.649	0.941
400	0	0.964	0.789	0.284	0.3	0.4	0.775	0.995	1
	0.2	0.951	0.738	0.276	0.043	0.287	0.753	0.993	1
	0.35	0.912	0.676	0.259	0.056	0.248	0.664	0.988	1
	0.5	0.868	0.602	0.235	0.066	0.225	0.613	0.963	1

6.3 Empirical examples

We apply our methodology in two illustrative empirical examples, using (15) with data from Yatchew (2003). Series estimation yields similar estimates of δ_0 to the kernel ones in Yatchew (2003). To test the hypothesis $H_0 : \delta_{0\ell} = 0$ against $H_1 : \delta_{0\ell} \neq 0$, $\ell = 1, \cdots, d$, the test using the usual t-statistic derived under independence contrasted with that based on the test statistic $t_n^* = n(\hat{\theta} - r)'\hat{C}_n^{-1}(\hat{\theta} - r), r = 0$ of our Theorem 6, which allows for spatial dependence.

The first example involves hedonic pricing of housing attributes. The data consist of 92 detached homes in Ottawa that were 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 $h(.)$ has two arguments, being location coordinates s =south and w = west):

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

The first set of columns of Table 9 recalls the kernel estimates reported in Yatchew (2003) based on methods and theory of Robinson (1988), and the second set reports the corresponding series estimation results, based on series functions $(1, s, w, sw)$. 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 the previous section, we ordered the data in ascending distance from the geographical coordinate $(s, w) = (0, 0)$, expecting spatial dependence in the error terms of neighbouring houses. In Table 9, SE refers to standard error computed under the assumption of independence, and TS^* is the test statistic t_n^* of Section 4.5 with critical values $46.39, 28.88$ at sizes 5% and 10%, respectively. Test statistics labelled * are significant at 5% level and those with $\triangle$ at 10% level. Our tests, which attempt to account for dependence, find that the presence of fireplace and luxury bathroom are 5% significant, while square footage, presence of fireplace, luxury bathroom, and garage are 10% significant; the latter findings may be more informative, bearing in mind the small sample size of 92. The contrasting conclusions on the significance of δ estimates between the t-test under independent errors and 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.

Table 9: Hedonic House Pricing

	kernel			**series**			
Variable	Coef	SE	t-stat	Coef	SE	t-stat	TS^*
frplc	12.6	5.8	2.17*	12.7	5.62	2.26*	126.23*
grge	12.9	4.9	2.63*	12.8	4.31	2.97*	29.98$^\triangle$
lux	57.6	10.6	5.43*	58.2	11.3	5.15*	177.10*
acginc	0.6	0.23	2.61*	0.61	0.2	3.08*	22.06
dhwy	1.5	21.4	0.07	-9.2	5.86	-1.57	10.38
lotarea	3.1	2.2	1.41	3.8	1.85	2.03*	22.12
nrbed	6.4	4.8	1.33	7.8	4.2	1.85$^\triangle$	14.57
usespc	24.7	10.6	2.33*	23.6	11.6	2.04*	37.67$^\triangle$

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

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

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 kilometers of distribution wire per customer). The non-parametrically involved regressor is $cust$ (log of the number of customers). 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., For the purpose of the present paper, we are interested in estimating the δ_i and testing their significance, $H_0 : \delta_l = 0, \quad$ versus $H_1 : \delta_l \neq 0$ for $l = 1, \cdots, d$, 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 repeat the kernel estimates of the δ_i and their standard errors assuming uncorrelatedness of error terms, taken from Yatchew (2003). The second set of columns report the $\widehat{\delta}_i$, using the first three Legendre polynomials in the series estimaton. Again, test statistics labelled $*$ are 5% significant, while those labelled △ are 10% significant. In order to apply the studentization of Section 5.2, 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 TS_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 TS_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% rsignificant using 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 $wage \cdot pcap$ were 10% significant and PUC, which was 5% significant under uncorrelatedness, was 10% significant based on ordering according to number of employees.

Table 10: Cost function in Electricity Distribution

	kernel			**series**				
	Coef	SE	t-stat	Coef	SE	t-stat	TS_w^*	TS_e^*
wage	-6.298	12.453	-0.506	-6.002	15.736	-0.381	0.426	0.261
pcap	-1.393	1.6	-0.872	-2.531	1.846	-1.371	$44.08^{\triangle}$	$35.433^{\triangle}$
$\frac{1}{2}wage^2$	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
$wage \cdot pcap$	0.534	0.599	0.891	2.044	1.553	1.317	$43.155^{\triangle}$	$40.27^{\triangle}$
PUC	-0.086	0.039	-2.205^*	-0.043	0.017	-2.6^*	11.042	$28.893^{\triangle}$
kwh	0.033	0.086	0.384	0.0828	0.102	0.8085	8.208	9.486
life	-0.634	0.115	-5.513^*	-0.613	0.124	-4.935^*	104.6^*	92.7^*
lf	1.249	0.436	2.865^*	0.746	0.486	1.535	$39.669^{\triangle}$	$36.587^{\triangle}$
kmwire	0.399	0.087	4.586^*	0.442	0.088	5.012^*	202.65^*	151.02^*

7. Conclusion

The 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, including a uniform rate of consistency, asymptotic normality and sufficient conditions for $\sqrt{n}$ - convergence, for which case a robust data-driven studentization method that offers an alternative to existing methods of inference was introduced. The framework of cross-sectional dependence and heterogeneity of this paper and its technical arguments, may be used to establish asymptotic theory for other estimation methods.

Appendix A. Proofs of Theorems 1-5.

In addition to the spectral norm introduced in Section 2, three other matrix norms appear in the proofs. Let $\|\cdot\|_E$ denote Euclidean norm, $\|\cdot\|_C$ maximum column absolute sum norm, and $\|\cdot\|_R$ maximum row absolute sum norm, so when $A = (a_{ij})$ is a $q \times q$ matrix

$$\|A\|_E^2 = \Big(\sum_{i,j=1}^{q} a_{ij}^2\Big), \quad \|A\|_C = \max_{1 \le j \le q} \Big(\sum_{i=1}^{q} |a_{ij}|\Big), \quad \|A\|_R = \max_{1 \le i \le q} \Big(\sum_{j=1}^{q} |a_{ij}|\Big).$$

The following inequalities will be useful:

$$\|A\|^2 \le \|A\|_R \|A\|_C, \quad |tr(AB)| \le \|A\|_E \|B\|_E, \|AB\|_E \le \|A\|_E \|B\|, \tag{16}$$

see e.g. Searle (1982), Horn and Johnson (1990).

In Section 3, we introduced the $K \times 1$ vector of normalised functions $P(x) = P^K(x) = B_K^{-1/2} p^K(x)$ satisfying $E(P(X_i)P(X_i)') = I_K$. Given that the series estimate $\hat{m}(\cdot)$ projects onto the linear space spanned by $p_1(\cdot), \cdots, p_K(\cdot)$, $\hat{m}(\cdot)$ is invariant to any nonsingular linear transformation of the $p_s(.)$. Hence,

$$\hat{m}(x) = p^K(x)'\hat{\beta} = P(x)'\hat{\gamma}, \tag{17}$$

where $\hat{\beta} = (p'p)^- p'Y \in \mathbb{R}^K$ with

$$p = p_n = [p^K(X_1), \cdots, p^K(X_n)]' \in \mathbb{R}^{n \times K}, \qquad Y = Y_n = (Y_1, \cdots, Y_n)' \in \mathbb{R}^n$$

and $\hat{\gamma} = (P'P)^- P'Y \in \mathbb{R}^K$, where $P = P_n = [P(X_1), \cdots, P(X_n)]' \in \mathbb{R}^{n \times K}$. To show such invariance, one can use the equality $P = pB_K^{-1/2}$ to establish that

$$\hat{\gamma} = (P'P)^- P'Y = B_K^{1/2}(p'p)^- B_K^{1/2} B_K^{-1/2} p'Y = B_K^{1/2}\hat{\beta}.$$

because $(p'p)^-$ being a Moore-Penrose inverse implies $(P'P)^- = (B_K^{-1/2} p'p B_K^{-1/2})^- = B_K^{1/2}(p'p)^- B_K^{1/2}$.

The proofs of Theorems 1-5 benefit from the algebraic convenience of the representation $\hat{m}(x) = P(x)'\hat{\gamma}$. Assumptions imposed on quantities involving $p^K(\cdot)$ such as $\xi(K)$ will continue to hold for their counterparts defined in terms of $P^K(\cdot)$. To show this fact for Assumption A4, note that $p^K(x)'\beta_K = P(x)'\gamma_K$, where $\gamma_K = B_K^{1/2}\beta_K$, so Assumption A4 implies

$$|m - P'\gamma_K|_\infty = O(K^{-\alpha}), \quad \text{as} \quad K \to \infty.$$

To verify that assumptions involving the upper bound $\xi(K)$ continue to hold for the corresponding quantity based on $P(\cdot)$, define:

$$\zeta(K) = \sup_{x \in \mathcal{X}} \|P^k(x)\|.$$

Then, for some $C < \infty$, $\zeta(k) \leq C\xi(k)$ for all $k \geq 1$, because

$$\zeta(k) = \sup_{x \in \mathcal{X}} \|B_k^{-1/2} p^k(x)\| \leq \|B_k^{-1/2}\| \sup_{x \in \mathcal{X}} \|p^k(x)\| \leq C\xi(k),$$

noting that by Assumption A3(i) and symmetry and positive semi-definiteness of B_K,

$$\|B_K^{-1/2}\| = \|B_K^{-1}\|^{1/2} = (\bar{\lambda}(B_K^{-1}))^{1/2} = (\underline{\lambda}(B_K))^{-1/2} \leq C.$$

The bound indicates that assumptions involving the upper bound $\xi(K)$ continue to hold also for $\zeta(K)$. The proofs will use $\hat{m}(x) = P(x)'\hat{\gamma}$, but wherever needed, the translation between the two alternative representations of $\hat{m}$ given in (17) is clarified.

Proof of Theorem 1. Let $M = M_n = (m(X_1), \cdots, m(X_n))' \in \mathbb{R}^n$ and $\hat{Q} = \hat{Q}_n = P'P/n \in \mathbb{R}^{K \times K}$. We decompose $\hat{m}(x) - m(x)$ into bias and stochastic terms. Let $\gamma_K = B_K^{1/2}\beta_K$ for β_K of Assumption A4. Write:

$$\hat{m}(x) - m(x) = [P(x)'(\hat{\gamma} - \gamma_K)] + [P(x)'\gamma_K - m(x)],$$

where $\hat{\gamma} = (P'P)^- P'Y = (\hat{Q})^- P'Y/n$. Recall $\Sigma_n = E(P'UU'P/n)$, the $K \times K$ variance matrix of $\sum_{i=1}^{n} P(X_i)U_i/\sqrt{n}$. We shall show below that

$$\|\hat{\gamma} - \gamma_K\| = O_p\left(\frac{tr(\Sigma_n)^{1/2}}{n^{1/2}} + K^{-\alpha}\right). \tag{18}$$

Then, by the definition of $\zeta(K)$ and Assumption A4,

$$\begin{aligned} |\hat{m} - m|_\infty &\leq |P'(\hat{\gamma}_K - \gamma_K)|_\infty + |P'\gamma_K - m|_\infty \\ &\leq \zeta(K)\|\hat{\gamma}_K - \gamma_K\| + O(K^{-\alpha}) \\ &= O_p\left(\zeta(K)\left[\frac{tr(\Sigma_n)^{1/2}}{n^{1/2}} + K^{-\alpha}\right]\right), \end{aligned}$$

as claimed.

Proof of (18). The matrix $\hat{Q}$ in $\hat{\gamma} = (\hat{Q})^{-}P'Y/n$ depends on $(X_1, \cdots, X_n)$, so invertibility of $\hat{Q}$ for any given sample cannot be taken for granted. Let $1_n = I(\underline{\lambda}(\hat{Q}) \geq a)$, where $I(.)$ is the indicator function and $a \in (0, 1)$. Then the inverse of $\hat{Q}$ exists when $1_n = 1$. It will be shown that

$$P(1_n = 1) \to 1 \text{ as } n \to \infty, \tag{19}$$

so that $\hat{Q}^{-1}$ exists with probability approaching 1. First write

$$1_n(\hat{\gamma} - \gamma_K) = 1_n \left[\hat{Q}^{-1}P'(Y - M)/n + \hat{Q}^{-1}P'(M - P\gamma_K)/n\right]. \tag{20}$$

By elementary inequalities

$$\begin{aligned} \|1_n(\hat{\gamma} - \gamma_K)\| &\leq \|1_n\hat{Q}^{-1}P'U/n\| + \|1_n\hat{Q}^{-1}P'(M - P\gamma_K)/n\| \\ &\leq \|1_n\hat{Q}^{-1}\|\|P'U/n\| + \|1_n\hat{Q}^{-1}P'/\sqrt{n}\|\|(M - P\gamma_K)/\sqrt{n}\|. \end{aligned} \tag{21}$$

We prove below that

$$\|1_n\hat{Q}^{-1}P'/\sqrt{n}\| = O_p(1), \tag{22}$$

$$\|P'U/n\| = O_p\left(\frac{tr(\Sigma_n)^{1/2}}{\sqrt{n}}\right), \tag{23}$$

$$\|(M - P\gamma_K)/\sqrt{n}\| = O_p(K^{-\alpha}). \tag{24}$$

which lead to

$$\|1_n(\hat{\gamma} - \gamma_K)\| = O_p\left(\frac{tr(\Sigma_n)^{1/2}}{n^{1/2}} + K^{-\alpha}\right),$$

and in turn to $\|\hat{\gamma}_K - \gamma_K\| = O_p\left(tr(\Sigma_n)^{1/2}/n^{1/2} + K^{-\alpha}\right)$. To see this, use $1 - 1_n = o_p(1)$ and the triangle inequality to obtain

$$\begin{aligned} \|\hat{\gamma} - \gamma_K\| &\leq \|1_n(\hat{\gamma} - \gamma_K)\| + \|(1 - 1_n)(\hat{\gamma} - \gamma_K)\| \\ &\leq \|1_n(\hat{\gamma} - \gamma_K)\| + o_p(1)\|\hat{\gamma} - \gamma_K\|. \end{aligned} \tag{25}$$

Thus

$$\begin{aligned} &\|\hat{\gamma} - \gamma_K\|(1 + o_p(1)) \leq \|1_n(\hat{\gamma} - \gamma_K)\|, \\ &\|\hat{\gamma} - \gamma_K\| \leq \|1_n(\hat{\gamma} - \gamma_K)\|/(1 + o_p(1)) = O_p\left(tr(\Sigma_n)^{1/2}/n^{1/2} + K^{-\alpha}\right). \end{aligned} \tag{26}$$

*Proof of (*19). It suffices to show that $\underline{\lambda}(\hat{Q}) \to_p 1$, as $n \to \infty$. Recalling that $P(x) = B_K^{-1/2}p^K(x) = [P_{1K}(x), \cdots, P_{KK}(x)]$,

$$\begin{aligned} E\left[tr\left\{(\hat{Q} - I)^2\right\}\right] &= \sum_{p,\ell=1}^{K} E[\{n^{-1}\sum_{i=1}^{n} P_{pK}(X_i)P_{\ell K}(X_i) - 1(\ell = p)\}^2] \\ &= n^{-2}\sum_{p,\ell=1}^{K} Var\left(\sum_{i=1}^{n} P_{pK}(X_i)P_{\ell K}(X_i)\right), \end{aligned}$$

noting that $E(\hat{Q}) = n^{-1}\sum_{i=1}^{n} E(P(X_i)P'(X_i)) = I$. For any pair $p, \ell = 1, \cdots, k$,

$$
\begin{aligned}
& Var\left(\sum_{i=1}^{n} P_{pK}(X_i)P_{\ell K}(X_i)\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov\left\{P_{pK}(X_i)P_{\ell K}(X_i), P_{pK}(X_j)P_{\ell K}(X_j)\right\} \\
& = \sum_{i=1}^{n} Var\left(P_{pK}(X_i)P_{\ell K}(X_i)\right) + \sum_{i,j=1, j\neq i}^{n} Cov\left\{P_{pK}(X_i)P_{\ell K}(X_i), P_{pK}(X_j)P_{\ell K}(X_j)\right\} \\
& =: V_{n,1}^{(p,\ell)} + V_{n,2}^{(p,\ell)}.
\end{aligned}
$$

so $E[\|\hat{Q} - I\|^2] \leq n^{-2}\sum_{p,\ell=1}^{K}(V_{n,1}^{(p,\ell)} + V_{n,2}^{(p,\ell)})$. One has

$$
\frac{1}{n^2}V_{n,1}^{(p,\ell)} = \frac{1}{n^2}\sum_{i=1}^{n} Var\big(P_{pK}(X_i)P_{\ell K}(X_i)\big) \leq \frac{\zeta^4(K)}{n}.
$$

To bound $V_{n,2}^{(p,\ell)}$ we use Assumption A5:

$$
\begin{aligned}
\frac{1}{n^2}|V_{n,2}^{(p,\ell)}| &= \Big| \int P_{pK}(x)P_{\ell K}(x)P_{pK}(y)P_{\ell K}(y)\Big(\frac{1}{n^2}\sum_{i,j=1, j\neq i}^{n}\{f_{ij}(x,y) - f(x)f(y)\}\Big)dxdy\Big| \\
&\leq \zeta^4(K)\Big(\frac{1}{n^2}\sum_{i,j=1, i\neq j}\int |f_{ij}(x,y) - f(x)f(y)|dxdy\Big) = \zeta^4(K)n^{-2}\triangle_n.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E\left[tr\left\{(\hat{Q} - I)^2\right\}\right] &= \sum_{p,\ell=1}^{K}(V_{n,1}^{(p,\ell)} + V_{n,2}^{(p,\ell)}) \\
&\leq \frac{K^2\zeta^4(K)}{n} + \frac{K^2\zeta^4(K)\triangle_n}{n^2} \\
&= K^2\zeta^4(K)\left(\frac{1}{n} + \frac{\triangle_n}{n^2}\right) = o(1), \qquad (27)
\end{aligned}
$$

by Assumptions A3(ii), and A5. Hence it remains to verify that $|\underline{\lambda}(\hat{Q}) - \underline{\lambda}(I)| \leq \left[tr\left\{(\hat{Q} - I)^2\right\}\right]^{1/2}$. The symmetric matrix $\hat{Q} - I$ can be written as $C(\hat{\Lambda} - I)C'$, where $C = (c_{ij}) \in \mathbb{R}^{K\times K}$ is an orthonormal eigenvector matrix such that $C'C = I$ and $\hat{\Lambda}$ is a diagonal matrix consisting of eigenvalues of $\hat{Q}$. Consequently, $(\hat{Q} - I)^2 = C(\hat{\Lambda} - I)^2C'$. Now $tr\{(\hat{Q} - I)^2\} = tr\left(C(\hat{\Lambda} - I)^2C'\right) = \sum_{\ell=1}^{K}(\lambda_\ell(\hat{Q}) - 1)^2$, because

$$
tr\left(C(\hat{\Lambda} - I)^2C'\right) = \sum_{i=1}^{K}\sum_{j=1}^{K} c_{ij}^2(\hat{\lambda}_j - 1)^2 = \sum_{j=1}^{K}(\hat{\lambda}_j - 1)^2\left(\sum_{i=1}^{K} c_{ij}^2\right) = \sum_{j=1}^{K}(\hat{\lambda}_j - 1)^2,
$$

because the columns of C are orthonormal. Therefore,

$$(\underline{\lambda}(\hat{Q})-1)^2 \le tr\{(\hat{Q}-I)^2\}, \qquad |\underline{\lambda}(\hat{Q})-1| \le [tr\{(\hat{Q}-I)^2\}]^{1/2} = o_p(1),$$

as was concluded in (27). This completes the proof of (19).

Proof of (22). Since $\hat{Q}$ is symmetric and non-negative definite,

$$\|1_n\hat{Q}^{-1}\| = 1_n\bar{\lambda}(\hat{Q}^{-1}) = 1_n(\underline{\lambda}(\hat{Q}))^{-1}.$$

The facts $1_n \to_p 1$ and $\underline{\lambda}(\hat{Q}) \to_p 1$ established above imply $1_n(\underline{\lambda}(\hat{Q}))^{-1} \to_p 1$. Hence, by Slutsky's theorem, $\|1_n\hat{Q}^{-1}\| = O_p(1)$, and thence

$$\|1_n\hat{Q}^{-1}P'/\sqrt{n}\|^2 = \|1_n\hat{Q}^{-1}P'P\hat{Q}^{-1}/n\| = \|1_n\hat{Q}^{-1}\| = O_p(1).$$

Proof of (23). We have

$$\|P'U/n\| = \frac{1}{\sqrt{n}}\|P'U/\sqrt{n}\| = \frac{1}{\sqrt{n}}\left[\bar{\lambda}\left(\frac{P'UU'P}{n}\right)\right]^{1/2} = O_p\left(\frac{tr\,(\Sigma_n)^{1/2}}{n^{1/2}}\right).$$

Proof of (24). We have

$$\begin{aligned}\|(M-P\gamma_K)/\sqrt{n}\|^2 &= (M-P\gamma_K)'(M-P\gamma_K)/n \\ &= \frac{1}{n}\sum_{i=1}^{n}(g(X_i)-P(X_i)\gamma_K)^2 = O_p(K^{-2\alpha}),\end{aligned}$$

by Assumption 4.

This completes the proof of (18) and thus of the theorem. ■

Proof of Theorem 2. Let $T_n = A'P'U/n$, where $P = p^K B_K^{-1/2} \in \mathbb{R}^n$, $A = \big(D(P_{1K}), D(P_{2K}), \cdots, D(P_{KK})\big)' \in \mathbb{R}^{K\times d}$ and $U = (U_1, \cdots, U_n)' \in \mathbb{R}^n$. Write

$$\hat{\theta}_n - \theta_0 = T_n + r_n, \quad r_n := \hat{\theta}_n - \theta_0 - T_n.$$

We shall show that

$$\sqrt{n}\bar{V}_n^{-1/2} r_n = o_p(1), \tag{28}$$

$$\sqrt{n}\bar{V}_n^{-1/2} T_n \to_d N(0, I_d), \tag{29}$$

which implies (8).

Proof of (28). By the same argument as in the proof of Theorem 1, (28) follows if we show that

$$1_n\sqrt{n}\bar{V}_n^{-1/2} r_n = o_p(1).$$

We shall use the bound $\|1_n\sqrt{n}\bar{V}_n^{-1/2}r_n\| \leq \sqrt{n}\|\bar{V}_n^{-1/2}\|\|1_n r_n\|$. To evaluate $\|1_n r_n\|$, recall $\bar{m} = P'\gamma_K$ and write

$$\begin{aligned} r_n = \hat{\theta}_n - \theta_0 - T_n = \{&a(\hat{m}) - a(m) - D(\hat{m}) + D(m)\} \\ &+\{D(\hat{m}) - D(\bar{m}) - T_n\} + \{D(\bar{m}) - D(m)\}. \end{aligned}$$

Then

$$\begin{aligned} \|r_n\| &\leq \|a(\hat{m}) - a(m) - D(\hat{m}) + D(m)\| \\ &+ \|D(\hat{m}) - D(\bar{m}) - T_n\| + \|D(\bar{m}) - D(m)\| \\ =: & \quad \|r_{n,1}\| + \|r_{n,2}\| + \|r_{n,3}\|. \end{aligned}$$

To show (28), note that by the assumptions of the theorem $\|\bar{V}_n^{-1/2}\| = \|\bar{V}_n^{-1}\|^{1/2} = O_p(1)$. Thus, it suffices to prove that

$$1_n\sqrt{n}\|r_{n,i}\| = o_p(1), \quad i = 1, 2, 3.$$

For $i = 1$, by Assumption B1, $\|r_{n,1}\| = O_p(|\hat{m} - m|_\infty^2)$. Thus by Theorem 1 and Assumption B3(i), (iii)

$$\sqrt{n}\|r_{n,1}\| = O_p\left(\sqrt{n}\zeta(K)^2\left(\frac{tr(\Sigma_n)}{n} + K^{-2\alpha}\right)\right) = o_p(1).$$

For $i = 2$, to bound $\|r_{n,2}\|$ recall the notation $\hat{\gamma} = (P'P)^- P'Y = \hat{Q}^- P'Y/n$, $Y = M + U$ and $A = (D(P_{1K}), \cdots, D(P_{KK}))'$. Then

$$D(\hat{m}) = D(P'\hat{\gamma}) = A'\hat{\gamma} = A'\hat{Q}^- P'(M + U)/n, \tag{30}$$
$$D(\bar{m}) = D(P'\gamma_K) = A'\gamma_K. \tag{31}$$

As in the proof of Theorem 1, one can replace $1_n\hat{Q}^-$ with $1_n\hat{Q}^{-1}$. Hence

$$\begin{aligned} \|1_n r_{n,2}\| &= \|1_n(A'\hat{Q}^{-1}P'Y/n - A'\gamma_K - A'P'U/n)\| \\ &= \|1_n A'\hat{Q}^{-1}P'(M + U)/n - A'\gamma_K - A'P'U/n\| \\ &= \|1_n A'(\hat{Q}^{-1} - I)P'U/n + A'\hat{Q}^{-1}P'(M - P\gamma_K)/n\| \\ &\leq \|1_n A'(\hat{Q}^{-1} - I)P'U/n\| + \|A'\hat{Q}^{-1}P'(M - P\gamma_K)/n\| \\ &\leq \|A'\|\|1_n(\hat{Q}^{-1} - I)\|\|P'U/n\| + \|A'\|\|1_n\hat{Q}^{-1}P'/\sqrt{n}\|\|(M - P\gamma_K)/\sqrt{n}\|. \end{aligned}$$

Note that $\|A\|^2 \leq \zeta^2(K)$, $\|1_n\hat{Q}^{-1}\| = O_p(1)$, and by (22)- (24),

$$\|1_n\hat{Q}^{-1}P'/\sqrt{n}\| = O_p(1), \quad \|(M - P\gamma_K)/\sqrt{n}\| = O_p(K^{-\alpha}), \quad \|P'U/n\| = O_p\left((tr(\Sigma_n)/n)^{1/2}\right).$$

Next, $\|1_n(\hat{Q}^{-1} - I)\| = \|1_n\hat{Q}^{-1}(I - \hat{Q})\| \leq \|1_n\hat{Q}^{-1}\|\|I - \hat{Q}\| = O_p(\|I - \hat{Q}\|)$, so

$$\|r_{n,2}\| = O_p(1)\sqrt{K}\zeta(K)\left(\|I - \hat{Q}\|(tr(\Sigma_n)/n)^{1/2} + K^{-\alpha}\right).$$

To bound $\|I-\hat{Q}\|$ note that $E[\|\hat{Q}-I\|^2] \le E\left[tr\left\{(\hat{Q}-I)^2\right\}\right]$. From (27),

$$\sqrt{n}\|r_{n,2}\| \le \left(nK\zeta^2(K)\right)^{1/2}\left\{\left[K^2\zeta^4(K)\left(\frac{1}{n}+\frac{\triangle_n}{n^2}\right)\frac{tr(\Sigma_n)}{n}\right]^{1/2}+K^{-\alpha}\right\} = o_p(1)$$

by Assumptions B3(ii) and (iii).

For $i=3$, by linearity of $D(\cdot)$ and Assumption B2 and A4, $\|r_{n,3}\| = O(|\bar{m}-m|_\infty) = O(K^{-\alpha})$, $\sqrt{n}\|r_{n,3}\| = O_p(\sqrt{n}K^{-\alpha}) = o_p(1)$,by Assumptions B3(iii), which implies $nK^{-2\alpha} = o(1)$.

Proof of (29). To show asymptotic normality of the main term $\sqrt{n}\bar{V}_n^{-1/2}T_n$, introduce the representation

$$\begin{aligned}
\sqrt{n}\bar{V}_n^{-1/2}T_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{V}_n^{-1/2}A'P(X_i)U_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{V}_n^{-1/2}A'P(X_i)\sigma(X_i)\sum_{j=1}^{\infty}b_{ij}\varepsilon_j \\
&= \sum_{j=1}^{\infty}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{V}_n^{-1/2}A'P(X_i)\sigma(X_i)b_{ij}\right)\varepsilon_j = \sum_{j=1}^{\infty}w_{jn}\varepsilon_j,
\end{aligned}$$

letting

$$w_{jn} = \sum_{i=1}^{n}\bar{V}_n^{-1/2}A'P(X_i)\sigma(X_i)b_{ij}/\sqrt{n}. \tag{32}$$

Noting that w_{jn} is a function of $\{X_i\}_{i=1}^n$, we show asymptotic normality conditional on $\|\bar{V}_n^{-1}\| \le C$ and $\{X_i\}_{i=1}^n$, treating w_{jn} as non-random. The key point here is to obtain the conditional asymptotic distribution to be $N(0, I_d)$, which is independent of $\{X_i\}_{i=1}^n$, whence . the required unconditional result follows.

By the Cramer-Wold device, to derive asymptotic normality of the vector $\sqrt{n}\bar{V}_n^{-1/2}T_n$, we consider the scalar $\sum_{j=1}^{\infty}c'w_{jn}\varepsilon_j$, for any fixed vector $c \in \mathbb{R}^d$ such that $c'c=1$. Write

$$\sqrt{n}c'\bar{V}_n^{-1/2}T_n = \sum_{j=1}^{N(n)}c'w_{jn}\varepsilon_j + \sum_{j=N(n)+1}^{\infty}c'w_{jn}\varepsilon_j, \tag{33}$$

where the integer $N(n)$ is chosen to be the smallest satisfying $\sum_{j=N(n)+1}^{\infty}(c'w_{jn})^2 \le 1/\log n$. Then

$$E\Big(\sum_{j=N(n)+1}^{\infty}c'w_{jn}\varepsilon_j\Big)^2 = O\Big(\sum_{j=N(n)+1}^{\infty}(c'w_{jn})^2\Big) = o(1),$$

so the second sum on the right side of (33) is $o_p(1)$. Since the $c'w_j\varepsilon_j$ are martingale differences under assumption A2, asymptotic normality of the first sum on the right side of (33) is established by verifying the following two sufficient conditions of Scott (1973), adapted for our setting.

$$\sum_{j=1}^{N(n)} E\big((c'w_j\varepsilon_j)^2\big) \to_p 1, \tag{34}$$

$$\sum_{j=1}^{N(n)} E\big((c'w_{jn}\varepsilon_j)^2 1(|c'w_{jn}\varepsilon_j| > \delta)\big) \to_p 0, \quad \forall\delta > 0. \tag{35}$$

By Assumption A2, we have

$$\sum_{j=1}^{N(n)} E\big((c'w_j\varepsilon_j)^2\big) = \sum_{j=1}^{N(n)} (c'w_{jn})^2.$$

By the choice of $N(n)$,

$$\sum_{j=1}^{N(n)} (c'w_{jn})^2 = \sum_{j=1}^{\infty} (c'w_{jn})^2 - \sum_{j=N(n)+1}^{\infty} (c'w_{jn})^2 = 1 + o(1).$$

Next let ν be as in Assumption A2. Then,

$$\begin{aligned}
\sum_{j=1}^{N(n)} E[(c'w_{jn}\varepsilon_j)^2 1(|c'w_{jn}\varepsilon_j| > \delta)] &= \sum_{j=1}^{N(n)} (c'w_{jn})^2 E[\varepsilon_j^2 1(|c'w_{jn}\varepsilon_j| > \delta)] \\
\le \sum_{j=1}^{N(n)} (c'w_{jn})^2 \left(\frac{|c'w_{jn}|}{\delta}\right)^{\nu} E|\varepsilon_j|^{2+\nu} &= \delta^{-\nu} \sum_{j=1}^{N(n)} |c'w_{jn}|^{2+\nu} E|\varepsilon_j|^{2+\nu} \\
\le C\delta^{-\nu} \sum_{j=1}^{N(n)} |c'w_{jn}|^{2+\nu} &\le C\delta^{-\nu} \max_{1\le j\le n} |c'w_{jn}|^{\nu} \sum_{j=1}^{N(n)} (c'w_{jn})^2.
\end{aligned}$$

The first inequality follows from $1(|c'w_{jn}\varepsilon_j| > \delta) \le (|c'w_{jn}\varepsilon_j|/\delta)^{\nu}$. With $\sum_{j=1}^{N(n)} (c'w_{jn})^2 \to 1$, (35) is verified once we show that $\max_{j\ge 1} |c'w_{jn}|^{\nu} \to 0$. Conditionally on $X_1, \cdots, X_n$, the following holds for any $j \ge 1$:

$$\begin{aligned}
|c'w_{jn}| &= \left| \frac{c'}{\sqrt{n}} \bar{V}_n^{-1/2} \sum_{i=1}^{n} A'P(X_i)\sigma(X_i)b_{ij} \right| \\
&\le \|c\| \|\bar{V}_n^{-1/2}\| \frac{1}{\sqrt{n}} \max_{1\le j\le n} \sum_{i=1}^{n} |b_{ij}| \|A'P(X_i)\sigma(X_i)\| \\
&= O\left(\frac{\zeta(K)^2}{\sqrt{n}} \max_{1\le j\le n} \sum_{i=1}^{n} |b_{ij}| \right) = o(1), \qquad (36)
\end{aligned}$$

by Assumption B4 and the bound $\|A'P(X_i)\sigma(X_i)\| \le C\|A\|\|P(X_i)\| \le C\zeta^2(K)$. ■

Proof of Theorem 3. We will prove later that

$$\|\bar{V}_n - V_n\| = o_p(1), \tag{37}$$

implying that $V_n^{-1}\bar{V}_n \to_p I$ since $\|V_n^{-1}\bar{V}_n - I\| \le \|V_n^{-1}\|\|\bar{V}_n - V_n\| = o_p(1)$. It follows that $\|\bar{V}_n^{-1}\| \le \|V_n^{-1}\|\|V_n\bar{V}_n^{-1}\| = O_p(1)$.To show the final statement, (13), of Theorem 3, write:

$$\sqrt{n}V_n^{-1/2}(\hat{\theta} - \theta_0) = \sqrt{n}\bar{V}_n^{-1/2}(\hat{\theta} - \theta_0) + \sqrt{n}\big(V_n^{-1/2} - \bar{V}_n^{-1/2}\big)(\hat{\theta} - \theta_0).$$

The first term was shown to converge in distribution to $N(0, I_p)$ in Theorem 2, while the second term is negligible:

$$\|\sqrt{n}\big(V_n^{-1/2} - \bar{V}_n^{-1/2}\big)(\hat{\theta} - \theta_0)\| \le \|\big(V_n^{-1/2}\bar{V}_n^{1/2} - I\big)\|\|\sqrt{n}\bar{V}_n^{-1/2}(\hat{\theta} - \theta_0)\| = o_p(1),$$

since $V_n^{-1/2}\bar{V}_n^{1/2} \to_p I$ from $V_n^{-1}\bar{V}_n \to_p I$, and thus $\|V_n^{-1/2}\bar{V}_n^{1/2} - I\| = o_p(1)$.

Proof of (37). The result follows if $|(\bar{V}_n - V_n)_{\ell p}| = o_p(1)$, for all $\ell, p = 1, \cdots, d$, where $(B)_{\ell p}$ denotes the $(\ell, p)^{th}$ element of a matrix B. Using the notation in (9),

$$\begin{aligned}(\bar{V}_n - V_n)_{\ell p} &= \frac{1}{n}\sum_{i,j=1}^{n} \gamma_{ij}\left\{\sigma(X_i)A'_\ell P(X_i)\sigma(X_j)P'(X_j)A_p - E(\sigma(X_i)A'_\ell P(X_i)\sigma(X_j)P'(X_j)A_p)\right\} \\ &= \frac{1}{n}\sum_{i,j=1}^{n} \gamma_{ij}\left\{h_i^{(\ell)}h_j^{(p)} - E(h_i^{(\ell)}h_j^{(p)})\right\}.\end{aligned}$$

Since

$$h_i^{(\ell)}h_j^{(p)} - E(h_i^{(\ell)}h_j^{(p)}) = \left\{\bar{h}_i^{(\ell)}\bar{h}_j^{(p)} - E(\bar{h}_i^{(\ell)}\bar{h}_j^{(p)})\right\} + \bar{h}_j^{(p)}E(h_i^{(\ell)}) + \bar{h}_i^{(\ell)}E(h_j^{(p)}),$$

we obtain that

$$\begin{aligned}(\bar{V}_n - V_n)_{\ell p} &= \frac{1}{n}\sum_{i,j=1}^{n} \gamma_{ij}\left\{\bar{h}_i^{(\ell)}\bar{h}_j^{(p)} - E(\bar{h}_i^{(\ell)}\bar{h}_j^{(p)})\right\} \\ &+ \frac{1}{n}\sum_{i,j=1}^{n} \gamma_{ij}\bar{h}_j^{(p)}E(h_i^{(\ell)}) + \frac{1}{n}\sum_{i,j=1}^{n} \gamma_{ij}\bar{h}_i^{(\ell)}E(h_j^{(p)}) \\ =: &\quad S_{1,n} + S_{2,n} + S_{3,n}.\end{aligned}$$

We shall show that

$$Var(S_{k,n}) = o(1), \quad k = 1, 2, 3, \tag{38}$$

which proves (37).

Proof of (38) for k=1. We have

$$Var(S_{1,n}) = \frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n} \gamma_{i_1i_2}\gamma_{i_3i_4} Cov\left(\bar{h}_{i_1}^{(\ell)}\bar{h}_{i_2}^{(p)}, \bar{h}_{i_3}^{(\ell)}\bar{h}_{i_4}^{(p)}\right).$$

Denote $\phi_{ij}^{(\ell,p)} = Cov(\bar{h}_i^{(\ell)}, \bar{h}_j^{(p)})$ and by $\Phi^{(\ell,p)}$ the $n\times n$ matrix whose $(i,j)^{th}$ element is $\phi_{ij}^{(\ell,p)}$. One has

$$\begin{aligned} Var(S_{1,n}) &= \frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n} \gamma_{i_1i_2}\gamma_{i_3i_4}\kappa(\bar{h}_{i_1}^{(\ell)}, \bar{h}_{i_2}^{(p)}, \bar{h}_{i_3}^{(\ell)}, \bar{h}_{i_4}^{(p)}) && (39)\\ &+ \frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n} \gamma_{i_1i_2}\gamma_{i_3i_4}\phi_{i_1i_3}^{(\ell,\ell)}\phi_{i_2i_4}^{(p,p)} && (40)\\ &+ \frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n} \gamma_{i_1i_2}\gamma_{i_3i_4}\phi_{i_1i_4}^{(\ell,p)}\phi_{i_2i_3}^{(p,\ell)}. && (41)\end{aligned}$$

Denote by $\Gamma = \Gamma_n$ the $n\times n$ matrix whose $(i,j)^{th}$ element is γ_{ij}. By Assumption B7, the right hand side of (39) is $o(1)$. To bound (40) and (41), write

$$\begin{aligned} \frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n} \gamma_{i_1i_2}\gamma_{i_3i_4}\phi_{i_1i_3}^{(\ell,\ell)}\phi_{i_2i_4}^{(p,p)} &= \frac{1}{n^2}tr\left(\Gamma\Phi^{(p,p)}\Gamma\Phi^{(\ell,\ell)}\right),\\ \frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n} \gamma_{i_1i_2}\gamma_{i_3i_4}\phi_{i_1i_4}^{(\ell,p)}\phi_{i_2i_3}^{(p,\ell)} &= \frac{1}{n^2}tr\left(\Gamma\Phi^{(p,\ell)}\Gamma\Phi^{(p,\ell)}\right).\end{aligned}$$

By properties of matrix norms, we see that

$$\left|tr\left(\Gamma\Phi^{(p,p)}\Gamma\Phi^{(\ell,\ell)}\right)\right| \le \|\Gamma\Phi^{(p,p)}\|_E\|\Gamma\Phi^{(\ell,\ell)}\|_E \le \|\Gamma\|^2\|\Phi^{(p,p)}\|_E\|\Phi^{(\ell,\ell)}\|_E. \quad (42)$$

Now write $\|\Phi^{(p,p)}\|_E^2 = \sum_{i,j=1}^{n}(\phi_{ij}^{(p,p)})^2 = \sum_{i=1,i=j}^{n}(\phi_{ii}^{(p,p)})^2 + \sum_{i,j=1,i\neq j}^{n}(\phi_{ij}^{(p,p)})^2$. For $i = j$, $|\phi_{ii}^{(p,p)}| = Var(\bar{h}_i^{(p)}) \le \zeta^4(K)$. For $i \neq j$, $|\phi_{ij}^{(p,p)}| \le C\zeta^4(K)\int_{\mathcal{X}^2}|f_{ij}(x,y) - f(x)f(y)|dxdy$, since $|\sigma(X_i)A_p'P(X_i)| \le C\zeta^2(K)$. Therefore,

$$\|\Phi^{(p,p)}\|_E^2 \le Cn\zeta^8(K) + C\zeta^8(K)\sum_{i,j=1,i\neq j}^{n}\left(\int|f_{ij}(x,y) - f(x)f(y)|dxdy\right)^2.$$

Clearly $\int|f_{ij}(x,y) - f(x)f(y)|dxdy \le 2$ for all i and j. Hence

$$\sum_{i,j=1,i\neq j}^{n}\left(\int|f_{ij}(x,y) - f(x)f(y)|dxdy\right)^2 \le 2\sum_{i,j=1,i\neq j}^{n}\int|f_{ij}(x,y)-f(x)f(y)|dxdy = 2\triangle_n.$$

Thus, for any $p = 1, \cdots, d$,

$$\|\Phi^{(p,p)}\|_E^2 = \sum_{i,j=1}^{n} (\phi_{ij}^{(p,p)})^2 \leq C\zeta^8(K)(n + \triangle_n). \tag{43}$$

Hence, by (42) and Assumption B6,

$$\frac{1}{n^2}\|\Gamma\|^2\|\Phi^{(p,p)}\|_E\|\Phi^{(\ell,\ell)}\|_E \leq \frac{1}{n^2}\left(\max_{j\geq 1}\sum_{i=1}^{n}\left|\gamma_{ij}\right|\right)^2 \zeta^8(K)(n+\triangle_n) = o(1),$$

by (16) , and by symmetry of Γ

$$\|\Gamma\|^2 \leq \|\Gamma\|_C^2 = \left(\max_{j\geq 1}\sum_{i=1}^{n}\left|\gamma_{ij}\right|\right)^2.$$

Similarly, it follows that $n^{-2}tr\left(\Gamma\Phi^{(p,\ell)}\Gamma\Phi^{(p,\ell)}\right) = o(1)$, which completes the proof of (38) when $k = 1$.

Proof of (38) for k=2,3. We have

$$\begin{aligned} Var(S_{2,n}) &= \frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n} \gamma_{i_1i_2}\gamma_{i_3i_4}E(h_{i_1}^{(\ell)})E(h_{i_3}^{(\ell)})E(\bar{h}_{i_2}^{(p)}\bar{h}_{i_4}^{(p)}) \\ &= \frac{1}{n^2}\sum_{i_2,i_4=1}^{n}\left(\sum_{i_1=1}^{n}\gamma_{i_1i_2}E(h_{i_1}^{(\ell)})\right)\left(\sum_{i_3=1}^{n}\gamma_{i_3i_4}E(h_{i_3}^{(\ell)})\right)\phi_{i_2i_4}^{(p,p)} \\ &\leq \frac{1}{n^2}\left(\zeta^2(K)\left|\max_{1\leq j\leq n}\sum_{i=1}^{n}\gamma_{ij}\right|\right)^2\sum_{i,j=1}^{n}|\phi_{ij}^{(p,p)}| \\ &\leq \frac{1}{n^2}\left(\zeta^2(K)\max_{1\leq j\leq n}\sum_{i=1}^{n}\left|\gamma_{ij}\right|\right)^2\sum_{i,j=1}^{n}|\phi_{ij}^{(p,p)}| \end{aligned}$$

using the bound $E|h_i^{(\ell)}| \leq C\zeta^2(K)$. By the same steps in the two lines prior to (43),

$$\sum_{i,j=1}^{n}|\phi_{ij}^{(p,p)}| \leq C\zeta^4(K)(n+\triangle_n).$$

This, together with Assumption B6 yields

$$Var(S_{n,2}) \leq \frac{C\zeta^8(K)(n+\triangle_n)}{n^2}\left(\max_{j\geq 1}\sum_{i=1}^{n}\left|\gamma_{ij}\right|\right)^2 = o(1).$$

■

Proof of Theorem 4. By the triangle inequality, $\|V_n - V\| \leq \|V_n - W_n\| + \|W_n - V\|$,where $\|W_n - V\| = o(1)$ holds by Assumption C2 (i). To bound $\|V_n - W_n\|$ note that

$$V_n - W_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\gamma_{ik}E[\sigma(X_i)\sigma(X_k)\{v_K(X_i)v_K'(X_k) - w(X_i)w'(X_k)\}].$$

We shall establish $\|V_n - W_n\| = o(1)$ by showing that all elements $(V_n - W_n)_{\ell,p}$, $1 \leq \ell, p \leq d$, of $V_n - W_n$ converge to zero. We have

$$\begin{aligned} |(V_n - W_n)_{\ell p}| &= \left|\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\gamma_{ik}E[\sigma(X_i)\sigma(X_k)(v_{\ell K}(X_i)v_{pK}(X_k) - w_\ell(X_i)w_p(X_k))]\right| \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}|\gamma_{ik}|E[|\sigma(X_i)\sigma(X_k)\{v_{\ell K}(X_i)v_{pK}(X_k) - w_\ell(X_i)w_p(X_k)\}|]. \end{aligned}$$

Notice that

$$\begin{aligned} & E[|\sigma(X_i)\sigma(X_k)\{v_{\ell K}(X_i)v_{pK}(X_k) - w_\ell(X_i)w_p(X_k)\}|] \\ & \leq CE[|v_{\ell K}(X_i)\{v_{pK}(X_k) - w_p(X_k)\}|] + CE[|\{v_{\ell K}(X_i) - w_\ell(X_i)\}w_p(X_k)|] \\ & \leq C\left(E[v_{\ell K}^2(X_i)]\right)^{1/2}\left(E[\{v_{pK}(X_k) - w_p(X_k)\}^2]\right)^{1/2} \\ & +C\left(E[\{v_{\ell K}(X_i) - w_\ell(X_i)\}^2]\right)^{1/2}\left(E[w_p^2(X_k)]\right)^{1/2} = o(1), \end{aligned}$$

because for any $p = 1, \cdots, d$, $E[w_p^2(X_i)] < \infty$ by Assumption C1 (i), $E[\{v_{pK}(X_i) - w_p(X_i)\}^2] = o(1)$ by Assumption C1 (iii) and $E[v_{pK}^2(X_i)] < \infty$. The latter follows from

$$E[v_{pK}^2(X_i)] \leq 2E[\{v_{pK}(X_i) - w_p(X_i)\}^2] + 2E[w_p^2(X_i)] < \infty.$$

Hence,

$$|(V_n - W_n)_{\ell p}| \leq \left[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}|\gamma_{ik}|\right]\cdot o(1) = o(1),$$

by Assumption C2 (ii). ■

Proof of Theorem 5. This is based on Lemmas 1, 2 and 3, stated and proved in Appendix B. Define the $d \times 1$ vector summation

$$\hat{S}_n^*(r) = \sum_{i=1}^{[rn]}\hat{A}^{*\prime}\hat{B}_K^{-1}p^K(X_i)\hat{U}_i/\sqrt{n}, \quad 0 \leq r \leq 1,$$

where $[rn]$ denotes integer part. Based on Lemma 2 and 3, one has weak convergence $\left(\hat{S}_n^*(r)\right)_{r\in[0,1]} \Rightarrow \left(V^{1/2}\{W_d(r) - rW_d(1)\}\right)_{r\in[0,1]}$ in the space $D[0,1]^d$. Observe that $\hat{C}_n = \frac{1}{n}\sum_{m=1}^{n}S_n^*(m/n)S_n^*(m/n)' \sim \int_0^1 S_n^*(r)S_n^*(r)'dr$. Therefore, the continuous mapping theorem gives

$$V^{-1/2}\hat{C}_nV^{-1/2} \Rightarrow \Psi_d.$$

Write

$$\hat{C}_n^{-1/2}\sqrt{n}(\hat{\theta}_n - \theta_0) = (\hat{C}_n^{-1/2}V^{1/2})\big(\sqrt{n}V^{-1/2}(\hat{\theta}_n - \theta_0)\big).$$

By Lemmas 1-3, $\hat{C}_n^{-1/2}V^{1/2} \Rightarrow \Psi_d^{-1/2}$, and by Theorem 4, $\sqrt{n}V^{-1/2}(\hat{\theta}_n - \theta_0) \rightarrow_d N(0, I_d)$, where the two terms converge jointly. ■

Appendix B. Lemmas and propositions

Let $X(\cdot), Y(\cdot) \in D[0,1]$, the space of all real valued functions on $[0,1]$ that are right-continuous with finite left limits. The Skorohod metric $d(\cdot,\cdot)$ in $D[0,1]$ is given by:

$$d(X,Y) = \inf_{\varepsilon>0}\{\varepsilon : \|\lambda\| \leq \varepsilon, \sup_{r\in[0,1]} |X(r) - Y(\lambda(r))| \leq \varepsilon\}$$

where λ is any continuous mapping of $[0,1]$ onto itself with $\lambda(0) = 0$, $\lambda(1) = 1$ and

$$\|\lambda\| = \sup_{r,u\in[0,1]:r\neq u} \Big|\log\frac{\lambda(u)-\lambda(r)}{u-r}\Big|, \quad 0 \leq r < u \leq 1.$$

Introduce

$$S_n(r) = \sum_{i=1}^{[rn]} A'P(X_i)U_i/\sqrt{n}, \quad \hat{S}_n(r) = \sum_{i=1}^{[rn]} A'P(X_i)\hat{U}_i/\sqrt{n}, \quad r \in [0,1] \tag{44}$$

Note that $S_n(\cdot) \in D[0,1]^d = D[0,1] \times \cdots \times D[0,1]$, where $D[0,1]^d$ is the product space. Endowing each component space $D[0,1]$ with the well-known Skorohod metric $d(\cdot,\cdot)$, stated above, we assign the following metric to the product space $D[0,1]^p$ as was done in Phillips and Durlauf (1986), for example. For $X(\cdot) = (X_1(\cdot),\cdots,X_d(\cdot))' \in D[0,1]^d$ and $Y(\cdot) = (Y_1(\cdot),\cdots,Y_d(\cdot))' \in D[0,1]^d$, define the metric:

$$d'(X,Y) = \max_{1\leq\ell\leq d}\{d(X_\ell, Y_\ell) : X_\ell, Y_\ell \in D[0,1]\}.$$

Lemma 1 below states a functional central limit theorem (FCLT) for $S_n(r)$ in $D[0,1]^d$ equipped with the metric $d'(\cdot,\cdot)$, for which we use $\Rightarrow_{D[0,1]^d}$to signify weak convergence of the associated probability measures in $D[0,1]^d$. Note that unlike in much of the FCLT literature, $S_n(r)$ is not a partial sum because its summands form a triangular array. structure. Since K is a function of n and recalling $A = \big(D(P_{1K}), D(P_{2K}),\cdots, D(P_{KK})\big)' \in \mathbb{R}^{K\times d}$, the summand $A'P(X_i)U_i/\sqrt{n}$ of $S_n(r)$ can be written

$$A'P(X_i)U_i/\sqrt{n} = \sum_{l=1}^{K(n)} D(P_{lK})P_{lK}(X_i)\sum_{j=1}^{\infty} b_{ij}\varepsilon_j/\sqrt{n}.$$

Now consider the following representation of $S_n(r)$, as a weighted summation of the ε_j over $j = 1, ...,$ with triangular array weights:

$$\begin{aligned} S_n(r) &= \sum_{i=1}^{[rn]} \Big[\sum_{l=1}^{K(n)} D(P_{lK}) P_{lK}(X_i) \sum_{j=1}^{\infty} b_{ij}\varepsilon_j/\sqrt{n} \Big] \\ &= \sum_{j=1}^{\infty} \Big[\sum_{i=1}^{[rn]} \sum_{l=1}^{K(n)} D(P_{lK}) P_{lK}(X_i) b_{ij}/\sqrt{n} \Big] \cdot \varepsilon_j = \sum_{j=1}^{\infty} c_j(n;r)\varepsilon_j, \end{aligned} \tag{45}$$

where we denote

$$c_j(n;r) := \Big[\sum_{i=1}^{[rn]} \sum_{l=1}^{K(n)} D(P_{lK}) P_{lK}(X_i) b_{ij}/\sqrt{n} \Big], \quad r \in [0,1], \quad n \geq 1.$$

The specification $S_n(r) = \sum_{j=1}^{\infty} c_j(n;r)\varepsilon_j$ was previously considered in Kasahara and Maejima (1986) in a general FCLT for infinite weighted sums. It goes without saying that the alternative representations of $S_n(r)$ given by (44) and (45) are of course equivalent. For the rest of the proof, we find it more convenient to use the form in (44) instead of (45).

We need further notations. For $j \geq 1$, introduce a $j \times K$ random matrix $P_j = (P(X_1), \cdots, P(X_j))'$ and $j \times 1$ random vectors, $M_j = (m(X_1), \cdots, m(X_j))'$ and $\hat{M}_j = (\hat{m}(X_1), \cdots, \hat{m}(X_j))'$.

Lemma 1. *Under the assumptions of Theorem 5,*

$$\Big(S_n(r)\Big)_{0 \leq r \leq 1} \Rightarrow_{D[0,1]^d} \big(V^{1/2} W_d(r)\big)_{0 \leq r \leq 1}. \tag{46}$$

Proof of Lemma 1. Lemma 1 states weak convergence in the d-dimensional product space $D[0,1]^d$. Phillips and Durlauf (1986, pp. 487-489) established two sufficient conditions for weak convergence of probability measures in this multi-dimensional product space. These two conditions, adapted here for (46), are; convergence of finite dimensional distributions of $S_n(\cdot)$ to those of $V^{1/2}W_d(\cdot)$; and tightness of each component of the vector $S_n(\cdot)$. These properties will be established using the the following results: for any $0 \leq r \leq u \leq 1$,

$$ES_n(r)S_n(u)' \to r \cdot V, \tag{47}$$

$$E|S_{n\ell}(u) - S_{n\ell}(r)|^2 \leq C\Big|\frac{[un] - [rn]}{n}\Big|, \quad \ell = 1, \cdots, d, \tag{48}$$

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

$$ES_n(r)S_n(u)' = ES_n(r)S_n(r)' + E\big(S_n(r)(S_n(u)' - S_n(r)')\big).$$

By Theorem 4, $E(S_n S_n') = V_n \to V$, and therefore

$$ES_n(r)S_n(r)' = \frac{[rn]}{n}\frac{1}{[rn]}E(A'P_{[rn]}'U_{[rn]}U_{[rn]}'P_{[rn]}A) \to rV.$$

Hence (47) follows if we show that $E\big(S_n(r)(S_n(u)' - S_n(r)')\big) \to 0$. This is achieved by showing that the limit of each element of the vector is zero. For $\ell, p = 1, \cdots, d$,

$$\begin{aligned} |E\big[S_n(r)(S_n(u)' - S_n(r)')\big]_{\ell p}| &\le \frac{C}{n}\sum_{i=1}^{[rn]}\sum_{k=[rn]+1}^{[un]} |\gamma_{ik}|E|v_{\ell K}(X_i)v_{pK}(X_k)| \\ &\le \frac{C}{n}\sum_{i=1}^{[rn]}\sum_{k=[rn]+1}^{[un]} |\gamma_{ik}| = o(1), \end{aligned}$$

by Assumption C3 (i), and because

$$E|v_{\ell K}(X_i)v_{pK}(X_k)| \le \big(Ev_{\ell K}^2(X_i)Ev_{pK}^2(X_k)\big)^{1/2} < \infty,$$

as shown in the proof of Theorem 4. This completes the proof of (47).

To prove (48), observe that

$$\begin{aligned} E|S_{n\ell}(u) - S_{n\ell}(r)|^2 &= E\Big|\frac{1}{\sqrt{n}}\sum_{i=[rn]+1}^{[un]} A_\ell' P(X_i)U_i\Big|^2 \\ &\le \frac{1}{n}\sum_{i,k=[rn]+1}^{[un]} |\gamma_{ik}|E|\sigma(X_i)\sigma(X_k)v_{\ell K}(X_i)v_{\ell K}'(X_k)| \\ &\le \frac{C}{n}\sum_{i,k=[rn]+1}^{[un]} |\gamma_{ik}| \le \frac{C}{n}\sum_{i=[rn]+1}^{[un]} \big[\max_{1\le i\le n}\sum_{k=1}^{n}|\gamma_{ik}|\big] \\ &\le C\Big|\frac{[un]-[rn]}{n}\Big|, \end{aligned}$$

by Assumption C3 (ii), which proves (48).

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

$$\big(S_n(r_1), \cdots, S_n(r_k)\big) \to_d \big(V^{1/2}W_d(r_1), \cdots, V^{1/2}W_d(r_k)\big).$$

By the Cramer-Wold device, it suffices to show that for any $d \times 1$ vectors $c_1', \cdots, c_k'$, $Q_n \to_d Q$, where

$$Q_n = \sum_{l=1}^{k} c_l' S_n(r_l), \ \ Q = \sum_{l=1}^{k} c_l' V^{1/2} W_d(r_l).$$

Write $S_n(r) = \sum_{j=1}^{\infty} w_{j,[rn]}\varepsilon_j$ with $w_{j,[rn]}$ as in (32), with $\bar{V}_n$ replaced by V. Then $Q_n = \sum_{j=1}^{\infty} w_{jn}^* \varepsilon_j$ with $w_{jn}^* = \sum_{l=1}^{k} c_l' w_{j,[r_l n]}$. By (47),

$$Var(Q_n) = \sum_{j=1}^{\infty} (w_{jn}^*)^2 \to Var(Q) = \sum_{l,t=1}^{k} c_l' V c_t \cdot \min\{r_l, r_t\} < \infty.$$

By (36), which holds for all $c_l' w_{j,[r_l n]}$, $l = 1, \cdots, k$, we have $\max_{j \geq 1} |w_{jn}^*| = o(1)$, and $Q_n \to_d Q$ follows by the same argument as in the proof of asymptotic normality (29).

Finally, we establish tightness for individual component of the vector $S_n(r)$, which completes the proof of the lemma. Noting that $S_{n\ell}(\cdot) \in D[0,1]$, $\ell = 1, \cdots, d$, we verify the following sufficient condition for tightness given in Billingsley (1968, Theorem 15.6, pp.128): for any $0 \leq r \leq s \leq t \leq 1$, and some $\beta \geq 0$, $\alpha > \frac{1}{2}$ and $C > 0$,

$$E[|S_{n\ell}(s) - S_{n\ell}(r)|^{2\beta} |S_{n\ell}(t) - S_{n\ell}(s)|^{2\beta}] \leq C \big|t - r\big|^{2\alpha}, \quad \ell = 1, \cdots, d. \tag{49}$$

This is in turn derived by showing that for any $0 \leq r \leq u \leq 1$,

$$E|S_{n\ell}(u) - S_{n\ell}(r)|^4 \leq C \Big|\frac{[un] - [rn]}{n}\Big|^2. \tag{50}$$

To see that (50) implies (49), note that for $\beta = 1$, the left sideof 49) is

$$\begin{aligned} E[|S_{n\ell}(s) - S_{n\ell}(r)|^2 |S_{n\ell}(t) - S_{n\ell}(s)|^2] &\leq \big\{E[|S_{n\ell}(s) - S_{n\ell}(r)|^4] E[|S_{n\ell}(t) - S_{n\ell}(s)|^4]\big\}^{1/2} \\ &\leq C\Big(\Big|\frac{[sn] - [rn]}{n}\Big|^2 \Big|\frac{[tn] - [sn]}{n}\Big|^2\Big)^{1/2} \\ &= C\Big|\frac{[sn] - [rn]}{n}\Big| \Big|\frac{[tn] - [sn]}{n}\Big| \\ &\leq C\Big|\frac{[tn] - [rn]}{n}\Big|^2, \end{aligned} \tag{51}$$

where the first step uses the Schwarz inequality, the second inequality follows from (50), and the last inequality from $0 \leq r \leq s \leq t \leq 1$. As explained on pp.138 of Billingsley (1968), if $t - r \geq 1/n$, then (51) implies (49) with $\alpha = 1$: since $[nt_2] \leq nt_2$ and $[nt_1] \geq nt_1 - 1$,

$$\frac{[nt_2] - [nt_1]}{n} \leq \frac{nt_2 - nt_1 + 1}{n} = t_2 - t_1 + \frac{1}{n} \leq 2(t_2 - t_1).$$

On the other hand, if $t - r < 1/n$, then at least one of $[sn] - [rn] = 0$ or $[tn] - [sn] = 0$ holds. Then the left sidess of (49) and (51) vanish, and thus (49) holds.

To verify (50), denote by e_ℓ, a d-dimensional vector, whose ℓ^{th} element is 1 and other elements 0. Then one can write

$$S_{n\ell}(u) - S_{n\ell}(r) = \sum_{j=1}^{\infty} e'_\ell(w_{j,[un]} - w_{j,[rn]})\varepsilon_j =: \sum_{j=1}^{\infty} \lambda_{jn}\varepsilon_j.$$

Rewriting the left side of (50) with the new notation and noting Assumption C5, we obtain

$$\begin{aligned} E(\sum_{j=1}^{\infty} \lambda_{jn}\varepsilon_j)^4 &= \sum_{j_1,\cdots,j_4=1}^{\infty} \lambda_{j_1n}\lambda_{j_2n}\lambda_{j_3n}\lambda_{j_4n}E(\varepsilon_{j_1}\varepsilon_{j_2}\varepsilon_{j_3}\varepsilon_{j_4}) \\ &= 3[\sum_{j,j'=1:j\neq j'}^{\infty} \lambda_{jn}^2\lambda_{j'n}^2] + \kappa\sum_{j=1}^{\infty}\lambda_{jn}^4 \le C[\sum_{j=1}^{\infty}\lambda_{jn}^2]^2 \\ &= C\big(E|S_{n\ell}(u) - S_{n\ell}(r)|^2\big)^2 \le C\Big|\frac{[un]-[rn]}{n}\Big|^2, \end{aligned}$$

where the last step follows from (48). ■

Lemma 2. Under the assumptions of Theorem 5,

$$\Big(\hat{S}_n(r)\Big)_{0\le r\le 1} \Rightarrow_{D[0,1]^d} \Big(V^{1/2}\{W_d(r) - rW_d(1)\}\Big)_{0\le r\le 1}. \tag{52}$$

Proof of Lemma 2. Since $\hat{U}_i - U_i = m(X_i) - \hat{m}(X_i)$,

$$L_n(r) := \hat{S}_n(r) - S_n(r) = \sum_{i=1}^{[rn]} A'P(X_i)\{m(X_i) - \hat{m}(X_i)\}/\sqrt{n}.$$

We can write, using $\hat{m}(X_i) = P'(X_i)\hat{\gamma}$,

$$\begin{aligned} L_n(r) &= \sum_{i=1}^{[rn]} A'P(X_i)\{m(X_i) - P'(X_i)\gamma_K\}/\sqrt{n} + \sum_{i=1}^{[rn]} A'P(X_i)P'(X_i)(\gamma_K - \hat{\gamma})/\sqrt{n} \\ &= A'P'_{[rn]}(M_{[rn]} - P_{[rn]}\gamma_K)/\sqrt{n} + A'P'_{[rn]}P_{[rn]}(\gamma_K - \hat{\gamma})/\sqrt{n}, \end{aligned}$$

leading to

$$\begin{aligned} \hat{S}_n(r) &= S_n(r) + \frac{A'P'_{[rn]}(M_{[rn]} - P_{[rn]}\gamma_K)}{\sqrt{n}} - \frac{A'P'_{[rn]}P_{[rn]}(\hat{\gamma} - \gamma_K)}{\sqrt{n}} \\ =: &\quad S_n(r) + a_n(r) - \ell_n(r). \end{aligned}$$

We shall show that

$$\sup_{r\in[0,1]} \|a_n(r)\| = o_p(1), \tag{53}$$

$$\ell_n(r) \Rightarrow_{D[0,1]^d} rV^{1/2}W_d(1), \tag{54}$$

which, together with Lemma 1, prove (52).

Proof of (53). One has

$$\begin{aligned}\sup_{r\in[0,1]}\|a_n(r)\| &\leq \|A'\|\sup_{r\in[0,1]}\|P'_{[rn]}\|\sup_{r\in[0,1]}\|(M_{[rn]}-P_{[rn]}\gamma_K)/\sqrt{n}\| && (55)\\ &= O_p(\sqrt{n}\xi^2(K)K^{-\alpha}) && (56)\end{aligned}$$

because $\|A'\| \leq \zeta(K) \leq \xi(K)$, whereas

$$\sup_{r\in[0,1]}\|P'_{[rn]}\| = O_p(\sqrt{n}\xi(K)), \quad \sup_{r\in[0,1]}\|(M_{[rn]}-P_{[rn]}\gamma_K)/\sqrt{n}\| = O(K^{-\alpha}),$$

by Assumption A4. Then (53) follows by Assumption C4.

Proof of (54). Recalling that $\hat{\gamma} = (P'P)^{-1}P'Y = (P'P)^{-1}P'(M-P\gamma_K)+(P'P)^{-1}P'(P\gamma_K+U)$,

$$\sqrt{n}(\hat{\gamma}-\gamma_K) = \widehat{Q}^{-1}P'(M-P\gamma_K)/\sqrt{n} + \widehat{Q}^{-1}P'U/\sqrt{n}.$$

Hence, with $\widetilde{Q}_r = P'_{[rn]}P_{[rn]}/n$,

$$\ell_n(r) = A'\widetilde{Q}_r\widehat{Q}^{-1}P'(M-P\gamma_K)/\sqrt{n} + A''\widetilde{Q}_r\widehat{Q}^{-1}\frac{P'U}{\sqrt{n}} =: \ell_{1,n}(r)+\ell_{2,n}(r). \qquad (57)$$

We shall show the following two results which constitute the proof of (54):

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

Noting that $\widehat{Q}^{-1} = O_p(1)$ and $\sup_{r\in[0,1]}\left\|\widetilde{Q}_r\right\| = O_p(\xi^2(K))$, since

$$\|\sum_{i=1}^{[rn]} P(X_i)P'(X_i)/n\| \leq \xi^2(K),$$

we obtain

$$\begin{aligned}\|\ell_{1,n}(r)\| &\leq \|A'\|\sup_{r\in[0,1]}\left\|\widetilde{Q}_r\right\|\left\|\widehat{Q}^{-1}\right\|\|P'(M-P\gamma_K)/\sqrt{n}\|\\ &\leq \|A'\|O_p\big(\xi^2(K)\big)\|P'\|\left\|(M-P\gamma_K)/\sqrt{n}\right\| = O_p(\sqrt{n}\xi^3(K)K^{-\alpha}) = o_p(1),\end{aligned}$$

by Assumption C4(iv). Next, write

$$\ell_{2,n}(r) = rA'P'U/\sqrt{n} + A'\left(\widetilde{Q}_r\widehat{Q}^{-1}-rI\right)P'U/\sqrt{n}.$$

Since the convergence $r(A'P'U/\sqrt{n}) \to_d rV^{1/2}W_d(1)$ was shown in the proofs of Theorems 2 and 4, it remains to verify that

$$\sup_{r\in[0,1]}\|A'\left(\widetilde{Q}_r\widehat{Q}^{-1}-rI\right)P'U/\sqrt{n}\| = o_p(1).$$

One has $\|A\| = O(\xi(K))$ and $\left\|P'U/\sqrt{n}\right\| = O(\sqrt{K})$ by Assumption C4 (ii). Next, with $\overline{Q}_r = P'_{[rn]}P_{[rn]}/[rn]$ we have

$$\sup_{r\in[0,1]}\left\|\frac{[rn]}{n}\overline{Q}_r\widehat{Q}^{-1} - rI\right\| \leq \sup_{r\in[0,1]}\left\|\overline{Q}_r - I\right\|\left\|\widehat{Q}^{-1} - I\right\| + \sup_{r\in[0,1]}\left\|\overline{Q}_r - I\right\| + \left\|\widehat{Q}^{-1} - I\right\| + o(1/n).$$

From the proof of Theorem 1, (27), we have

$$\|\hat{Q} - I\|^2 = O_p\left(K^2\xi^4(K)(\frac{1}{n} + \frac{\triangle_n}{n^2})\right).$$

This, by Horn and Johnson (1990, pp. 335-336), implies

$$\left\|\widehat{Q}^{-1} - I\right\|^2 = O_p\left(K^2\xi^4(K)(\frac{1}{n} + \frac{\triangle_n}{n^2})\right) = O_p\left(K^2\xi^4(K)/n\right),$$

with the last step following from Assumption C4(i). Similarly, one has that

$$\begin{aligned}\sup_{r\in[0,1]}(\frac{[rn]}{n})^2\left\|\overline{Q}_r - I\right\|^2 &= \sup_{r\in[0,1]}(\frac{[rn]}{n})^2 O_p\left(K^2\xi^4(K)(\frac{1}{[rn]} + \frac{\triangle_{[rn]}}{[rn]^2})\right)\\ &= \sup_{r\in[0,1]} O_p\left(K^2\xi^4(K)(\frac{1}{n} + \frac{\triangle_{[rn]}}{n[rn]})\right) = O_p\left(K^2\xi^4(K)/n\right),\end{aligned} \tag{58}$$

by Assumption A4 (i). Therefore,

$$\sup_{r\in[0,1]}\|A'\left(\widetilde{Q}_r\widehat{Q}^{-1} - rI\right)P'U/\sqrt{n}\| = O_p(K\sqrt{K}\xi^3(K)/\sqrt{n}) = o_p(1), \tag{59}$$

with the last step following from Assumption A3 (ii). ■

Lemma 3. Under the assumptions of Theorem 5, as $n \to \infty$, $\sup_{r\in[0,1]}\|\hat{S}_n^*(r) - \hat{S}_n(r)\| = o_p(1)$.

Proof of Lemma 3. Recall that $A' = A^{*\prime}B_K^{-1/2}$, $p'_{[rn]} = B_K^{1/2}P'_{[rn]}$. Thus,

$$\begin{aligned}\|\hat{S}_n^*(r) - \hat{S}_n(r)\| &= \|(\hat{A}^{*\prime}\hat{B}_K^{-1}p'_{[rn]}\hat{U}_{[rn]} - A'P'_{[rn]}\hat{U}_{[rn]})/\sqrt{n}\|\\ &= \|(\hat{A}^{*\prime}\hat{B}_K^{-1}B_K^{1/2} - A^{*\prime}B_K^{-1/2})P'_{[rn]}\hat{U}_{[rn]}/\sqrt{n}\|.\end{aligned}$$

Therefore,

$$\begin{aligned}\sup_{r\in[0,1]}\|\hat{S}_n^*(r) - \hat{S}_n(r)\| &\leq \|\hat{A}^{*\prime}\hat{B}_K^{-1}B_K^{1/2} - A^{*\prime}B_K^{-1/2}\|\\ &\quad\cdot\sup_{r\in[0,1]}\|P'_{[rn]}\hat{U}_{[rn]}/\sqrt{n}\| =: d_{n,1}d_{n,2}.\end{aligned} \tag{60}$$

We will show that

$$d_{n,1} = O_p(K\xi^2(K)/\sqrt{n}) + O_p\big(\xi^2(K)\big(\sqrt{\frac{tr(\Sigma)}{n}} + K^{-\alpha}\big)\big), \qquad (61)$$
$$d_{n,2} = O_p(K^{1/2} + K^{-\alpha}\sqrt{n}). \qquad (62)$$

Then, since $tr(\Sigma) = O_p(K)$ by Assumption C4 (ii),

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

by Assumption B3 (iii), C4 (iv) and B3 (ii), and

$$\begin{aligned} d_{n,1} &= \|\hat{A}^{*\prime}\hat{B}_K^{-1}B_K^{1/2} - A^{*\prime}B_K^{-1/2}\| \leq \|\hat{A}^{*\prime} - A^{*\prime}\|\|\hat{B}_K^{-1}B_K^{1/2} - B_K^{-1/2}\| \\ &\quad + \|A^{*\prime}\|\|\hat{B}_K^{-1}B_K^{1/2} - B_K^{-1/2}\| + \|\hat{A}^{*\prime} - A^{*\prime}\|\|B_K^{-1/2}\|. \end{aligned}$$

Note that $\|A^*\| \leq \xi(K)$, and by Assumption A3 (i), $\|B_K^{-1/2}\| = O_p(1)$. Now

$$\|\hat{B}_K^{-1}B_K^{1/2} - B_K^{-1/2}\| \leq \|\hat{B}_K^{-1} - B_K^{-1}\|\|B_K^{1/2}\| = O_p\left(K\xi^2(K)\frac{1}{\sqrt{n}}\right),$$

since by Assumption C4 (iii) $\|B_K\| = O(1)$, whereas $\|\hat{B}_K - B_K\|^2 = O_p\left(K^2\xi^4(K)/n\right)$, as can be shown using the argument used for the bound (27) for $\|\hat{Q} - I\|$ and applying Assumption C4 (i). Then $\|\hat{B}_K^{-1} - B_K^{-1}\|^2 = O_p\left(K^2\xi^4(K)/n\right)$ follows from Horn and Johnson (1990, pp 335-336), under Assumptions C4 (iii) and A3 (i), which imply $\|B_K\| = O(1)$ and $\|B_K^{-1}\| = O(1)$, as $n \to \infty$.

To obtain (61), it remains to evaluate the term $\|\hat{A}^* - A^*\|$. Newey (1997) showed that $\hat{A}^* = (\hat{A}_1^*, \cdots, \hat{A}_d^*)$ equals $\big(D(p_1;\hat{m}), \cdots, D(p_K;\hat{m})\big)'$ with probability approaching one. Recalling $D(\cdot;\hat{m}) = \big(D_1(\cdot;\hat{m}), \cdots, D_d(\cdot;\hat{m})\big)$, the i^{th} column of $\hat{A}^* - A^*$ can be written

$$\hat{A}_i^* - A_i^* = \big(D_i(p_1;\hat{m}) - D_i(p_1;m), \cdots, D_i(p_K;\hat{m}) - D_i(p_K;m)\big)', \quad i = 1, \cdots, d.$$

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

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

with the first inequality following from Assumption C6. Therefore $\|\hat{A}_i^* - A_i^*\| = O_p(\xi(K)|\hat{m} - m|_\infty)$, for $i = 1, \cdots, p$. This allows the bound

$$\begin{aligned} \|\hat{A}^* - A^*\|^2 &\leq tr\big((\hat{A}^* - A^*)'(\hat{A}^* - A^*)\big) = \sum_{i=1}^{p}(\hat{A}_i^* - A_i^*)'(\hat{A}_i^* - A_i^*) \\ &= \big(\sum_{i=1}^{p}\|\hat{A}_i^* - A_i^*\|^2\big) \leq C\xi^2(K)|\hat{m} - m|_\infty^2. \end{aligned}$$

Therefore, applying to $|\hat{m} - m|_\infty$ the bound of Theorem 1, we obtain

$$\|\hat{A}^* - A^*\| = O_p\left(\xi^2(K)\left[\sqrt{\frac{tr(\Sigma_n)}{n}} + K^{-\alpha}\right]\right)) = o_p(1),$$

by Assumption B3 (ii)-(iii), completing the proof of (61).

Next note that

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

As in the proof of Lemma 2,

$$d_{n,21} \le \sup_{r\in[0,1]} \|P'_{[rn]}(M_{[rn]} - P_{[rn]}\gamma_K)/\sqrt{n}\| + \sup_{r\in[0,1]} \|\widetilde{Q}_r\|\|\sqrt{n}(\hat{\gamma} - \gamma_K)\|.$$

From (56) it is seen that the first term on the right is $O_p(\sqrt{n}\xi(K)K^{-\alpha}) = o_p(1)$, by Assumption C4 (iv). By (26), $\|(\hat{\gamma} - \gamma_K)\sqrt{n}\| = O_p(tr(\Sigma_n)^{1/2} + K^{-\alpha}\sqrt{n}) = O_p(K^{1/2} + K^{-\alpha}\sqrt{n})$ from Assumption C4 (ii), whereas by (58), $\sup_{r\in[0,1]} \|\overline{Q}_r\| = O_p(1) + O_p(K\zeta^2(K)/\sqrt{n}) = O_p(1)$ by Assumption A3 (ii). Thus, $d_{n,21} = O_p(K^{1/2} + K^{-\alpha}\sqrt{n})$.

Finally,

$$d_{n,22} = \sup_{r\in[0,1]} (\frac{[rn]}{n})^{1/2}\|P'_{[rn]}U_{[rn]}/\sqrt{[rn]}\| = O_p(\sup_{r\in[0,1]} tr^{1/2}(\Sigma_{[rn]})) = O_p(\sqrt{K}),$$

by Assumption C4(ii). Hence, $d_{n,2} = O_p(K^{1/2})$, which proves (62). ∎

Te following proposition provides the upper bound for $\triangle_n$ in (5) in case of scalar Gaussian X_i.

Proposition 1. *Let $X_i \sim N(0,1), i = 1, 2, \cdots,$ and denote $\sigma_{ij}^{(X)} = Cov(X_i, X_j)$. If for some $c_0 < 1$, one has $|\sigma_{ik}^{(X)}| \le c_0$, $i, k = 1, 2, \cdots; i \ne k$, then* (6)

Proof of Proposition 1. Recall the standard bivariate normal density with correlation ρ is

$$f_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\big(-m_\rho(x, y)\big),$$

where

$$m_\rho(x, y) := \frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)}, \quad x, y \in \mathbb{R}.$$

Then $f_0(x, y) = f(x)f(y)$, where $f(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. We show that when $|\rho| \le c_0 < 1$,

$$|f_\rho(x, y) - f_0(x, y)| \le C\rho \exp\big(\frac{-(x^2 + y^2)}{8}\big), \ x, y \in \mathbb{R}, \tag{63}$$

where C does not depend on ρ. Since $f_{ik}(x,y) = f_{\sigma_{ik}}(x,y)$ is the bivariate density of X_i, X_k, it follows by (63) that

$$\begin{aligned}\triangle_n &= \sum_{i,k=1,i\neq k}^{n} \int |f_{ij}(x,y) - f(x)f(y)| dxdy \\ &\leq C \sum_{i,k=1,i\neq k}^{n} |\sigma_{ik}^{(X)}| \int \exp\big(-(x^2+y^2)/8\big) dxdy \\ &\leq C \sum_{i,k=1,i\neq k}^{n} |\sigma_{ik}^{(X)}|.\end{aligned}$$

Proof of (63) By the mean value theorem, applied for $|\rho| \leq c_0$,

$$|f_\rho(x,y) - f_0(x,y)| \leq |\rho| \sup_{|\rho|\leq c_0} |f'_\rho(x,y)|. \tag{64}$$

Note that

$$f'_\rho(x,y) = f_\rho(x,y)\left(\frac{\rho}{1-\rho^2} - \frac{\partial m_\rho(x,y)}{\partial \rho}\right). \tag{65}$$

where

$$\left|\frac{\rho}{1-\rho^2}\right| \leq \frac{c_0}{1-c_0^2}.$$

We show that

$$f_\rho(x,y) \leq c \exp\big(-(x^2+y^2)/4\big) \tag{66}$$

$$\Big|\frac{\partial m_\rho(x,y)}{\partial \rho}\Big| \leq c(x^2+y^2),\ x,y \in \mathbb{R}, \tag{67}$$

where c does not depend on ρ and x, y, which together with (64) and (65) implies (63).

Note that

$$\begin{aligned}m_\rho(x,y) &\geq \frac{x^2+y^2-2|\rho xy|}{2(1-|\rho|^2)} = \frac{|\rho|(x^2+y^2-2|xy|) + (1-|\rho|)(x^2+y^2)}{2(1-|\rho|^2)} \\ &\geq \frac{(1-|\rho|)(x^2+y^2)}{2(1-|\rho|^2)} \geq \frac{x^2+y^2}{2(1+|\rho|)} \geq \frac{x^2+y^2}{4},\end{aligned}$$

the second inequality following from $2|xy| \leq x^2+y^2$. This implies (66):

$$f_\rho(x,y) \leq \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\big(-(x^2+y^2)/4\big) \leq \frac{1}{2\pi\sqrt{1-c_0^2}} \exp\big(-(x^2+y^2)/4\big).$$

Next,

$$\begin{aligned}\left|\frac{\partial m_\rho(x,y)}{\partial \rho}\right| &= \left|\frac{-4(1-\rho^2)xy + 4\rho(x^2+y^2-2\rho xy)}{[2(1-\rho^2)]^2}\right| \\ &\leq \frac{|xy|}{(1-\rho^2)} + \frac{|x^2+y^2-2\rho xy|}{[(1-\rho^2)]^2} \leq \frac{|xy|}{(1-c_0^2)} + \frac{x^2+y^2+2|\rho xy|}{(1-c_0^2)^2} \\ &\leq c(x^2+y^2),\end{aligned}$$

which proves (67). ■

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

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

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

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

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.

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 Handbook of Econometrics. Vol. 6B, Chapter 76 (J. J. Heckman, E. E. Leamer, eds.) North-Holland, Amsterdam.

Chen, X., Liao, Z., Sun, Y., 2010. On inference of sieve M-estimation of functionals of semi-nonparametric time series models. Preprint.

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, 26, Department of Geography, Series A, 15.

Cliff, A., Ord, J.K., 1981. Spatial processes, models and applications. Pion, London.

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

de Jong, R.M., 2002. A note on "Convergence rates and asymptotic normality for series estimators": uniform convergence rates. Journal of Econometrics 111, 1-9.

Fan, Y., Li, Q., 1999, Root-n-consistent estimation of partially linear time series models. Journal of Nonparametric Statistics 11, 251-269.

Hannan, E. J., 1957. The variance of the mean of a stationary process. Journal of the Royal Statistical Society Series B 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. Journal of Econometrics 167, 224-239.

Jenish, N., Prucha, I.R., 2012. On spatial processes and asymptotic inference under near-epoch dependence. Journal of Econometrics 170, 178-190.

Kasahara, Y., Maejima, M., 1986. Functional limit theorems for weighted sums of I.I.D. random variables. Probability Theory and Related Fields 72, 161-183.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 99-121.

Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509-533.

Kelejian, H.H., Prucha, I.R., 2007. HAC estimation in a spatial framework. Journal of 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.

Newey, W. K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of 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. Review of Economic Studies 53, 473-496.

Pinkse, J., Shen, L., Slade, M. E., 2007. Central limit theorem for endogenous locations and complex spatial interactions. Journal of 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. Journal of Time Series Analysis 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. Journal of Econometrics 165, 5-19.

Robinson, P.M., Thawornkaiwong, S., 2012. Statistical inference on regression with spatial dependence. Journal of Econometrics 167, 521-542.

Scott, D.J., 1973. Central limit theorems for martingales using a Skorokhod representation approach. Advances in Applied Probability 5, 119-137.

Searle, S.R., 1982. Matrix algebra useful for statistics. John Wiley, New York.

Yatchew, Y., 2003. Semiparametric regression for the applied econometrician. Cambridge University Press, New York.