# Series Estimation under Cross-sectional Dependence

Jungyoon Lee *

London School of Economics

*Job market paper*

November 14, 2011

**Abstract**

This paper develops an asymptotic theory of the series estimation under general cross-sectional dependence and heterogeneity, covering both nonparametric and semiparametric estimates. A uniform rate of consistency, asymptotic normality and sufficient conditions for the $\sqrt{n}$ rate of convergence are established. A new data-driven studentization that dispenses with the need for "distance measures", as required by the spatial HAC estimation, is introduced and leads to asymptotically correct inference. Conditions imposed on dependence are carefully formulated to accommodate various cross-sectional settings plausible in economic applications and readily apply to panel and time series data. Strong, as well as weak dependence are covered, filling the current gap in the theoretical literature, and conditional heteroscedasticity is allowed. A finite sample Monte Carlo study indicates a highly satisfactory performance of the estimation and testing procedures. Applying the methods of the paper to two illustrative examples reveals some subtle yet interesting differences in results, compared to that under the assumption of independence.

*JEL Classifications*: C13; C14; C21

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

*I am deeply indebted to my advisor Prof. Peter M. Robinson for his generous guidance and encouragement. I am most grateful to Donald Andrews, Xiahong Chen, Liudas Giraitis, Javier Hidalgo, Tatiana Komarova, Dennis Kristensen, Oliver Linton, Whitney Newey, Marcia Schafgans, Myung Hwan Seo and Martin Weidner for their insightful comments. I also thank A. Yatchew for the kind permission to use his datasets. Financial support of Samsung Scholarship Foundation is gratefully acknowledged. E-mail: j.lee2@lse.ac.uk

# 1 Introduction

This paper presents an asymptotic theory of the series estimation under general cross-sectional dependence and heterogeneity and proposes a data-driven studentization that is designed to circumvent particular challenges faced by the so-called heteroscedasticity, autocorrelation consistent (HAC) inference with cross-sectional data. In this section, a brief discussion of the motivation is given with some relevant literature, followed by a summary of the contribution of the paper.

## 1.1 Background

Economic agents are typically interdependent, due for example to competition, externalities and spill-overs. Such dependence is often overlooked in cross-sectional or panel data analysis, in part due to a lack of econometric literature that deals with the issue at hand. Implications of dependence on econometric analysis have long been studied in the context of time series data, where the temporal dependence is naturally modeled in terms of the distance between observations along the time axis. Unfortunately, the nature of cross-sectional dependence observed in economic data hinders a simple multi-dimensional extension of time series literature to spatial data. For example, the index of observations in economic cross-sectional data cannot be used to describe the dependence between units in the way that the time index can be. This is because there is often no natural ordering of cross-sectional data and the indices do not represent relative positioning of the units sampled.

In order to start accounting for possible cross-sectional dependence, one needs first to establish a framework under which the structure of such dependence can be suitably formalised. Two classes of models of spatial dependence have been prominent in recent literature the set-up of which involves a concept of "economic location". In economic data, cross-sectional units correspond to economic agents such as individuals or firms. One could envisage that these agents are positioned in some socio-economic (even geographical) space, whereby their relative locations in this space 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).

The first class of models that has attracted much recent theoretical research is the Spatial Autoregressive (SAR) model of Cliff and Ord (1981), see e.g. Lee (2002, 2004), Kelejian and Prucha (1998, 1999), Robinson (2010), Rossi (2010). In this approach, the dependent variable (or disturbance) of a given unit is assumed to be affected by a weighted average of the dependent variables (or disturbances) of the other sampled units. The weights used in the averaging are presumed to be known and reflect the degree of proximity between agents, leaving a finite number of parameters (often scalar) to be estimated to explain the spatial dependence. The SAR model has gained popularity in empirical works, see e.g. Arbia (2006).

The second class of models involve the use of mixing coefficients familiar from time series literature. Conley (1999) and the related papers assume that the relevant space and locations of the agents corresponding to each sample observations are known, possibly with some measurement error. Mixing conditions are developed in terms of the distance between agents' locations, under a suitable stationarity assumption. An alternative mixing condition in spatial setting was proposed in Pinkse, Shen and Slade (2007).

Robinson (2009) has offered a new way of modeling cross-sectional dependence, which does *not* hinge on the idea of economic distance although can certainly accommodate it. A general, possibly non-stationary, linear process is assumed for disturbances, which, unlike a mixing framework, allows possible strong dependence. The dependence in the regressors is phrased in terms of the departure of joint densities from the product of marginals, allowing possible heterogeneity across units. The model's ability to cover both weak and strong dependence in the error term and regressors allows the development of a general set of theory. While the model accommodates many spatial settings plausible in economic data, no new concepts, other than those familiar from standard econometric literature, need to be introduced.

On the other hand, nonparametric and semiparametric estimation have become an established method in econometric analysis. Such methods allow researchers to drop the assumption of known parametric functional form that is often not warranted by economic theory. There are many theoretical results on nonparametric kernel estimation under temporal dependence, see e.g. Robinson (1983) and Hidalgo (1997). Robinson (2009) and Robinson and Thawornkaiwong (2010) have considered kernel estimation in the nonparametric regression model and the partly linear regression model, respectively, under cross-sectional dependence.

Series estimation is an alternative method of nonparametric estimation, whose main advantage over kernel estimation is three-fold. When economic theory generates certain restrictions on the nonparametric function of interest, such as monotonicity, convexity and additive separability, series estimation offers a more natural way of using such information in estimation by reflecting them in the choice of series functions. Secondly, it is computationally convenient, because the data is summarized by a relatively few estimated coefficients. Thirdly, from a theoretical point of view, theories can be developed in a unified way to include both nonparametric regression and general semiparametric quantities, as will be made clear in the later sections of this paper. This is in contrast to kernel estimation where an asymptotic theory for each semiparametric model needs to be developed separately.

The asymptotic behaviour of the series estimation under independence has been studied in Andrews (1991) and Newey (1997). For weakly dependent time series data, Chen and Shen (1998) and Chen, Liao and Sun (2011) offer a rather complete treatment of asymptotic theory and robust inference of the general sieve M estimation, which includes series estimation as a special case. Contributions of these papers will be discussed in more detail in Section 2, once the model and the quantities of interest have been properly introduced.

## 1.2 Contribution of the paper

The two contributions of this paper are the development of an *asymptotic theory of series estimation* under cross-sectional dependence and heterogeneity and a *studentization method* that is new to the spatial setting and semi-parametric estimation.

This paper produces an asymptotic theory that covers general cross-sectional heterogeneity and dependence, including weak and strong dependence. The conditions of the paper, while designed for spatial setting, readily lend themselves to time series and panel data, expanding the applicability of the results to those settings. They follow the framework of Robinson (2009), however the nature of series estimation necessitated some modifications. This paper offers alternative conditions in terms of the $4^{th}$ order cumulants familiar from time series literature, enabling to avoid conditions on joint densities which may be difficult to verify for some processes. Due to a number of similarities of series estimation to OLS in linear regression, asymptotic results derived here easily extend to the linear regression.

The other main contribution of this paper is the development of a studentization method that overcomes, in part, the limitation of the existing variance estimation literature in spatial setting which requires additional distance measures. As mentioned above, a key difference of spatial data from time series is that the index of observations does not necessarily carry any information on the strength of dependence between units. In a stationary time series, it is natural to model the dependence between two observations in terms of the distance in time. If such dependence decreases to zero fast enough with increasing time lag between observations, the variance of a sample mean can be consistently estimated nonparametrically by the so-called HAC estimate, see e.g. Hannan (1957), Newey and West (1987). Let $U_1, U_2, \cdots, U_T$ be a sample of mean zero random variables. The HAC estimate of the variance $Var(\bar{U}) = E(\sum_{t,s=1}^{T} U_t U_s)/T^2$ of the sample mean $\bar{U}$ is defined as $\widehat{Var}(\bar{U}) = \sum_{t,s=1}^{T} U_t U_s w(|t-s|/m)/T^2$, with some weight function $w(\tau)$ decreasing in $\tau$ and a bandwidth parameter $m = m_T$ that increases slowly to infinity with T. In the simple case $w(\tau) = I(\tau \in [0,1])$, the above estimator effectively truncates out observation pairs $(U_t, U_s)$ with $|t-s| > m$, whose contribution to the quantity of interest is negligible. Such truncation prevents the variance of the estimator from exploding with $T$ and allows a consistent estimation of the variance.

In the spatial context, an extension of the HAC estimation is feasible if additional information which may take the role of the time indices in the afore-mentioned truncation is available, e.g. the socio-economic or geographical distance between units which underpin the structure of the spatial dependence. Conley (1999) has considered HAC estimation under a stationary random field with measurement error in distance measures, Kelejian and Prucha (2007) for models of Cliff and Ord (1981) and Robinson and Thawornkaiwong (2010) for a more general set-up than Cliff-Ord type models. Bester, Conley, Hansen and Vogelsang (2008) consider the asymptotic theory of the HAC estimation when a fixed, rather than a vanishing, proportion of the sample is used in the variance estimation.

In practice, the availability of such additional information on distances cannot be taken for granted. In a typical cross-sectional or panel data, location or distance observations are not readily available. More importantly, the knowledge of the correct socio-economic "distance" that underline the dependence structure is at best an educated guess and such distances may be inherently unobservable in some cases. Hence, practitioners often face ambiguity in their choices of distance measures, which may not generate much confidence. Studentization introduced in Section 5 of this paper offers an alternative that does not require truncation, in a manner similar to Kiefer, Vogelsang and Bunzel (2000) that deals with the linear regression in time series. Such studentization may be of great benefit to both cases, with and without additional information on distance measures, enabling inference for the latter, and offering a tool for robustifying results in the former.

The paper is structured as follows. In Section 2, the setting of the model is outlined. In Section 3, the 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 the $\sqrt{n}$ rate of convergence of certain semiparametric estimators, with data-driven studentization. Section 6 reports a Monte

Carlo study of the finite sample properties. Section 7 discusses some empirical examples and Section 8 concludes. The Appendix contains the proofs.

## 2 Setting of the model

This paper discusses inference on the following nonparametric regression model,

$$Y_i = g_0(X_i) + U_i, \quad i = 1, 2, \cdots, n, \tag{2.1}$$

where $Y_i, U_i \in \mathbb{R}$ and $X_i \in \mathscr{X} \subset \mathbb{R}^q$ are random variables and $g_0 : \mathscr{X} \to \mathbb{R}$ is an unknown function of interest. The error term $U_i$ of the model is assumed to follow

$$U_i = \sigma(X_i)e_i, \quad e_i = \sum_{j=1}^{\infty} b_{ij}\varepsilon_j, \quad \sum_{j=1}^{\infty} b_{ij}^2 < \infty \quad i = 1, 2, \cdots, n, \tag{2.2}$$

where $\sigma(\cdot)$ is a real function, $\{b_{ij}, i, j \geq 1\}$ are unknown constants and $\{\varepsilon_j, j \geq 1\}$ are independent random variables with zero mean and unit variance. Processes $\{X_i\}$ and $\{\varepsilon_j\}$ are assumed to be independent of each other. The linear process $e_i$ in $U_i$ was also used in Robinson (2009) and Robinson and Thawornkaiwong (2010). Quantities $Y_i, X_i, U_i, e_i, \varepsilon_j, b_{ij}$ are allowed to admit a triangular structure throughout this work, accommodated by the proofs of later theorems. The additional $n$ subscript in e.g. $b_{ijn} = b_{ij}$ is suppressed for the ease of notation. Triangular array structure takes into account the possible need to re-label observations as $n$ increases in panel-data or multi-dimensional lattice data as this work uses a single index $i$ for observations, see Robinson (2009) for discussion. Allowing coefficients $b_{ij} = b_{ijn}$ to vary with $n$ is important in making (2.2) to cover the popular SAR model, whereby $e_i = \sum_{j=1}^{n} b_{ijn}\varepsilon_j$ will include summation only up to $n$.

The structure (2.2) is designed to encompass various forms of spatial dependence and heterogeneity in the unobserved errors $U_i$, that could arise in economic applications. Conditional and unconditional heteroscedasticity of the errors $U_i$ is allowed, while the restrictions later imposed on $b_{ij}$'s are rather mild, affording an ample scope for possible non-stationarity/heterogeneity across $i$. For example, $b_{ij}$'s need not exhibit any form of conformity across $i$ and $j$, in particular, be affected by $|i - j|$. The specification (2.2) also accommodates the idea of "economic distance", in which case $b_{ij}$ will be determined by distances between units. Restrictions on dependence and heterogeneity in $X_i$ are stated and discussed in Section 3.

Errors (2.2) obviously cover equally-spaced stationary time series, where $b_{ij}$ is of the form $b_{|i-j|}$. An alternative to (2.2) is a mixing framework, which would allow us to relax the condition of independence between $\{X_i\}_{i=1}^n$ and $\{e_i\}_{i=1}^n$. However, a mixing framework necessitates the introduction of some distance measures and the notion of stationarity, which are not always justifiable in economic applications. More importantly, long-range dependence is not covered by a mixing-framework.

Regarding the function $g_0(x) = E(Y_i|X_i = x)$ in (2.1), it denotes the conditional expectation of $Y_i$ at $X_i = x$. For any given function $g(\cdot) : \mathscr{X} \to \mathbb{R}$, let $a(g)$ denote a $d \times 1$ vector-valued functional of $g(\cdot)$, i.e. a mapping from a possible conditional expectation to a real vector. There are many applications where a (known) functional $a(g_0)$ of the conditional expectation $g_0$ is of interest. It can be estimated by $a(\hat{g})$, where $\hat{g}(\cdot)$ denotes a series estimator of $g_0(\cdot)$, constructed as a linear combination of pre-specified approximating functions. Simple examples of $a(g)$ include the value of the function at multiple fixed points, $a(g) = (g(x_1), \cdots, g(x_d))'$, $(x_1, \cdots, x_d) \in \mathscr{X}^d$, which is of interest in nonparametric regression estimation, and the value of partial derivative of the function with respect to the $\ell^{th}$ argument at fixed points,

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

which is of interest in the case of nonparametric derivative estimation. An example of nonlinear functional $a(\cdot)$ given in Newey (1997) is the consumer surplus. Letting $Y_i$ be the log consumption and $X_i = (\log p_i, \log I_i)'$, a $2 \times 1$ vector of log price and log income, the estimated demand function at a fixed point $X_i = x$ is given by $\exp(\hat{g}(x))$, whereas the approximate consumer surplus is equal to the integral of the demand function over a range of prices. For a fixed income $\bar{I}$, an estimator of this functional, when $\underline{p}$ and $\bar{p}$ represent the lower and upper bounds on the price, is

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

If one was interested in the approximate consumer surplus at multiple fixed values of income, $a(\hat{g})$ would take a vector form. Further example of $a(\cdot)$ arises in the context of the partly linear regression model and will be discussed in detail in Section 5.

Previously, Andrews (1991) showed asymptotic normality for a vector-valued linear functional $a(\hat{g})$, using for $\{X_i\}_{i=1}^n$ and $\{U_i\}_{i=1}^n$ independent and non-identically distributed ($i.n.i.d$) setting and, in addition, indicating that "the proof can be extended to cover strong mixing regressors without too much difficulty." Newey (1997) has established uniform rate of convergence for $|\hat{g}(x) - g_0(x)|$ and asymptotic normality result for

$a(\hat{g}) - a(g_0)$ when $\{X_i\}_{i=1}^n$ and $\{U_i\}_{i=1}^n$ are both *i.i.d.* and $a(g)$ is a general (possibly nonlinear) scalar functional. Newey (1997) has also offered a set of conditions on the functional $a(\cdot)$, under which $a(\hat{g})$ converges to $a(g_0)$ at the parametric rate. Chen and Shen (1998) and Chen, Liao and Sun (2011) consider the problem of sieve extreme estimation for weakly dependent time series setting. In the context of series estimation, Chen and Shen (1998)'s results yield a convergence rate for the nonparametric regression estimate $\hat{g}(\cdot)$ and asymptotic normality for $a(\hat{g})$ in the case of $\sqrt{n}$-rate of convergece. Chen, Liao and Sun (2011) offer an asymptotic normality result that also covers the case of slower-than-$\sqrt{n}$ rate of convergence and provide methods of inference robust to time series weak dependence. They also unveil a rather striking fact that for certain cases of slower-than-$\sqrt{n}$ rate of convergence, the asymptotic variance of the estimate $a(\hat{g})$ coincides with that obtained under independence. An important example is the case of nonparametric regression function evaluated at a finite number of fixed points, for which a similar observation was made by Robinson (1983) for kernel estimation.

# 3 Estimation of $g_0$ and uniform consistency rate

Estimation of $g_0$ is based on the use of approximating functions. Denote by $p_s(\cdot), s = 1, 2, \cdots$ a set of approximating functions from $\mathscr{X}$ to $\mathbb{R}$:

$$p^k(\cdot) = (p_1(\cdot), \cdots, p_k(\cdot))'.$$

Next, introduce a deterministic sequence of positive integers $K = K_n$, nondecreasing in $n$, which denotes the number of approximating functions used in the series estimation where $n$ stands for the sample size. The integer $K$ can be regarded as a bandwidth parameter, analogous to the window length in kernel estimation, and its choice gives rise to a bias/variance trade-off as seen below. Under a suitable choice of approximating functions, larger values of $K$ will reduce the bias while increasing the variance of the estimate $\hat{g}$. A number of assumptions introduced in the following two sections reflect the reliance of the theory on a suitable choice of $K$.

Let $\hat{\beta} = (\mathbf{p}'\mathbf{p})^- \mathbf{p}'Y \in \mathbb{R}^K$, where $\mathbf{p} = \mathbf{p_n} = [p^K(X_1), ..., p^K(X_n)]' \in \mathbb{R}^{n \times K}$, $Y = (Y_1, ..., Y_n)' \in \mathbb{R}^n$, and $A^-$ denotes the Moore-Penrose inverse for a matrix $A$.

**Definition 1.** A series estimator of $g_0(x)$, at a fixed point $x \in \mathscr{X}$, based on $K$ approximating functions, $p^K(\cdot) = (p_1(\cdot), \cdots, p_K(\cdot))'$, is given by

$$\hat{g}(x) = p^K(x)'\hat{\beta}. \tag{3.1}$$

In the remainder of this section, we establish a uniform consistency rate of the estimate $\hat{g}(x)$.

**Assumption A1.** *The random variables* $\{X_i\}_{i=1}^n$, $n = 1, 2, \cdots$, *are independent of* $\{\varepsilon_j\}_{j=1}^\infty$ *and identically distributed with the probability density function* $f(x)$, $x \in \mathscr{X}$. *The joint density of* $X_i$ *and* $X_j$, $f_{ij}(x, y)$, $x, y \in \mathscr{X}$, *exists for all* $i$ *and* $j$.

The assumption of identity of distribution on $\{X_i\}_{i=1}^n$ later facilitates proofs of theorems by offering some algebraic simplification, allowing us to afford more heterogeneity in the unobserved $\{U_i\}_{i=1}^n$, which was deemed more crucial than allowing for non-identical distribution of the observed $\{X_i\}_{i=1}^n$. Heterogeneity/non-stationarity in $X_i$ across $i$ is still allowed as the dependence in $X_i$ may vary across the index $i$.

**Assumption A2.** *The random variables* $\{U_i\}_{i=1}^n$, $n = 1, 2, \cdots$, *follows the linear specification (2.2) with some bounded positive function* $\sigma(x) : \mathscr{X} \to \mathbb{R}$, *and innovations* $\{\varepsilon_j\}$ *are independent across* $j$, *satisfying* $E(\varepsilon_j^2) = 1$ *and* $\max_{j \geq 1} E|\varepsilon_j|^{2+\nu} < \infty$, *for some* $\nu > 0$.

For any $k \geq 1$, define a $k \times k$ matrix

$$B_k := E(p^k(X_i)p^k(X_i)'), \quad k = 1, 2, \cdots . \tag{3.2}$$

Let $\underline{\lambda}(A)$ and $\bar{\lambda}(A)$ denote the minimal and maximal eigenvalues of a square matrix $A$. In this work, Euclidean norm is used for vectors: $\|a\|^2 = a'a$. For matrices, we use spectral norm, induced by Euclidean vector norm: $\|A\| = \max_{\|x\|=1} \|Ax\| = \bar{\lambda}^{1/2}(A'A)$. For functions, the uniform norm $|g|_\infty = \sup_{x \in \mathscr{X}} |g(x)|$ is used.

Define a sequence of scalar constants $\xi(k)$ as

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

Quantities similar $\xi(k)$ were also used in Andrews (1991) and Newey (1997). If it is known that $g_0$ is a bounded function, one may choose bounded and non-vanishing series functions, in which case $\xi(k)$ increases at the rate of $\sqrt{k}$: $\sup_{x \in \mathscr{X}} \|p^k(x)\| = \sup_{x \in \mathscr{X}} \big(\sum_{i=1}^k p_i^2(x)\big)^{1/2} \leq C\sqrt{k}$. 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 series functions are orthogonal polynomials, and $\xi(K) = K^{1/2}$ when they are B-splines.

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

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

Condition $\underline{\lambda}(B_k) \geq c$ of Assumption A3(i) requires $B_k$ to be nonsingular for all values of $k$ and is also assumed in Andrews (1991) and Newey (1997). When this assumption fails, some series functions in $p_s(\cdot), s \geq 1$ may be redundant and need to be eliminated to make it hold. 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 in the previous paragraph, A3 (ii) boils down to $K = o(n^{1/4})$ for the case of B-splines and $K = o(n^{1/6})$ in the case of orthonormal polynomials under the suitable conditions required for those expressions of $\xi(K)$.

**Assumption A4.** *The function $g_0(\cdot) : \mathscr{X} \to \mathbb{R}$ and series functions $p_s(\cdot), s \geq 1$, are such that there exist a sequence of vectors $\beta_K$ and a number $\alpha > 0$ satisfying, as $K \to \infty$,*

$$|g_0 - p^{K\prime}\beta_K|_\infty = O(K^{-\alpha}).$$

Assumption A4 is a standard condition used in the series estimation literature, and appears in Andrews (1991) and Newey (1997). It requires the uniform approximation error of $g_0(\cdot)$ by a linear combination of the chosen set of series functions to diminish fast enough. It can be seen as a smoothness condition imposed on $g_0(\cdot)$, if the functions $p_s(\cdot), s = 1, 2, \cdots$ are ordered so that higher values of $s$ correspond to less smooth functions. In such case, the smoother the function $g_0(\cdot)$ is, the faster is the rate of decay in the coefficients of the vector $\beta_K$ in the series expansion $p^{K\prime}\beta_K$ of $g_0(\cdot)$. Some further insights into Assumption A4 for certain choices of the approximating functions, including polynomials, trigonometric polynomials, splines and orthogonal wavelets, can be found in Chen (2007), pp. 5573. Assumption A4 will control the bias term of our estimate $\hat{g}$, and $\alpha$ is also related to the number of the regressors. Newey (1997) points out that for splines and power series, Assumption A4 is satisfied with $\alpha = s/q$ where $s$ is the number of continuous derivatives of $g_0$ and $q$ is the dimension of $x$. Conditions imposing an upper bound on the rate of increase in $K$, such as A3 (ii), may necessitate a stronger assumption on the smoothness of the unknown $g_0$.

Now, we will state an assumption that is required to control the strength of dependence in $X_i$'s across $i$. Introduce the quantity:

$$\triangle_n := \sum_{i,j=1, i\neq j}^{n} \int_{\mathscr{X}^2} |f_{ij}(x, y) - f(x)f(y)|dxdy. \tag{3.3}$$

The rate of growth of $\triangle_n$ is a measure of bi-variate dependence in $X_i$'s and has an upper bound of $2n^2$, a useful property used in the proofs. The quantity $\triangle_n$ is zero in case of independence across $i$ and we may view the condition $\triangle_n = O(n)$ as an analogue to the concept of short-range/weak dependence in time series literature. Quantities of similar nature were used in Robinson (2009) and Robinson and Thawornkaiwong (2010).

In the case that $X_i$'s are Gaussian random variables, $\triangle_n$ satisfies the following simple bound. Let $\sigma_{ij}^{(X)} := Cov(X_i, X_j)$. Assume for simplicity that $\sigma_{ii}^{(X)} = \sigma_i^{(X)} = 1$. If for some $c_0 < 1$, one has $|\sigma_{ik}^{(X)}| \le c_0, \forall i, k = 1, \cdots, n;\ i \neq k,\ n \ge 1$, then

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

see Proposition 1 in Appendix B. Clearly, if $\max_{1\le k\le n} \sum_{i=1}^{n} |\sigma_{ik}^{(X)}| \le Cn$, then $\triangle_n = O(n)$, whereas $\triangle_n = o(n^2)$ holds for a large class of covariances. Thus, in the Gaussian case, $\triangle_n$ can be replaced by the sum $\sum_{i,k=1:i\neq k}^{n} |\sigma_{ik}^{(X)}|$.

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

Assumption A5 indicates that stronger dependence in $X_i$ will require the use of smaller $K$. In the light of Assumption A4, this will necessitate a stronger assumption on the smoothness of the unknown function $g_0$. Under weak dependence, i.e. $\triangle_n = O(n)$, A5 reduces to $K^2\xi^4(K) = o(n)$ which is stated in A3(ii). Otherwise, A5 is a stronger condition than A3(ii) imposed on the upper bound of the growth in $K$ and $\xi(K)$.

To state our first theorem, it is necessary to introduce some notation. Define normalised functions $P^k(x) := B_k^{-1/2}p^k(x)$ with $B_k$ as in (3.2) and such that $E(P^k(X_i)P^k(X_i)') = I_k$. We shall write $P(x) = P^K(x)$ with $K = K_n$, suppressing the superscript $K$ for the rest of the paper for the ease of notation. Note that $P(\cdot) = [P_{1K}(\cdot), ..., P_{KK}(\cdot)]'$, with the double subscripts at $P_{sK}(\cdot)$ arising from the definition $P(\cdot) = B_K^{-1/2}p^K(\cdot)$. Such normalised functions were also used in Newey (1997). Let $\mathbf{P} = \mathbf{P}_n = (P(X_1), \cdots, P(X_n))' \in \mathbb{R}^{n\times K}$. For a given sequence $K = K_n$, define the following $K \times K$ variance-covariance matrix $\Sigma_n$ of the $K \times 1$ vector sum $\sum_{i=1}^{n} P(X_i)U_i/\sqrt{n}$:

$$\begin{aligned}\Sigma_n \quad &:= \quad E(\mathbf{P}'UU'\mathbf{P}/n) = Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} P(X_i)U_i\right) &(3.4)\\ &= \quad \frac{1}{n}\sum_{i,k=1}^{n} E\big(P(X_i)U_iU_kP'(X_k)\big)\\ &= \quad \frac{1}{n}\sum_{i,k=1}^{n} \gamma_{ik}E(\sigma(X_i)\sigma(X_k)P(X_i)P'(X_k)),\end{aligned}$$

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 the uniform rate of convergence of the estimator $\hat{g}(x)$.

**Theorem 1 (Uniform Rate of Convergence).** *Under Assumptions A1-A5,*

$$\sup_{x\in\mathscr{X}}|\hat{g}(x)-g_0(x)| = O_p\left(\xi(K)\left[\sqrt{\frac{tr(\Sigma_n)}{n}}+K^{-\alpha}\right]\right), \quad as \quad n\to\infty.$$

This result coincides with the rate obtained by Newey (1997) for *i.i.d.* $\{X_i\}$ and $\{U_i\}$. In the latter case $\Sigma_n = \sigma^2 E(P(X_i)P(X_i)') = \sigma^2 I_K$ leading to $tr(\Sigma_n) = O(K)$. The proof of the above Theorem is given in the Appendix A. The first term in the rate of Theorem 1 reflects the contribution of the variance of $\hat{g}$, while the second term arises from the bias component. The uniform rate of consistency highlights the bias/variance trade-off in the selection of $K$: the use of larger $K$ reduces the bias and increases the variance.

The rates obtained in Theorem 1 need to be verified to be $o_p(1)$, to establish uniform consistency of the series estimate $\hat{g}$. The requirement $\xi(K)K^{-\alpha} = o(1)$ of negligible bias suggests that it may be favourable to choose series functions which are bounded. To evaluate the contribution of the variance, suppose for now that the original series functions 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 on $e_i$'s, $\sum_{i,k=1}^{n}\gamma_{ik} = O(n)$, meaning the rate becomes $\xi(K)\sqrt{K/n} = K/\sqrt{n}$ which is $o(1)$ by Assumption A3 (ii). Under strong dependence of $e_i$'s, the rate is slower and further conditions restricting the increase of $K$ and $\xi(K)$ may be needed to show uniform consistency.

# 4 Asymptotic normality

The previous section established the uniform rate of convergence for $\hat{g}-g_0$, whilst our ultimate interest lies in inference on the functional $a(g_0)$. Denote $\theta_0 = a(g_0)$ and $\hat{\theta} = a(\hat{g})$. In this section, we study the asymptotic distribution of $\hat{\theta}-\theta_0$. First, we provide some technical assumptions needed for establishing asymptotic normality. 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 $\|a(g)-a(g_0)-D(g-g_0)\|\le C(|g-g_0|_\infty)^2$, if $|g-g_0|_\infty\le\epsilon$.*

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

Assumptions B1 and B2 are the same as in Newey (1997). Assumption B2 requires the linear functional $D(\cdot)$ to be continuous, which follows from the fact that $D(\cdot)$ is the Frechet-differential of $a(\cdot)$ at $g_0$. A functional $a(\cdot)$ is said to be Frechet-differentiable at $g_0$ if there exists a bounded linear operator $D(\cdot)$ satisfying the following property: $\forall\delta > 0, \quad \exists\epsilon > 0$ such that $\|a(g) - a(g_0) - D(g - g_0)\| \leq \delta|g - g_0|_\infty$ if $|g - g_0|_\infty \leq \epsilon$. Assumption B1(ii) imposes a stronger smoothness condition on $a(\cdot)$ at $g_0$ 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, $D(g) = a(g)$.

Define a $K \times d$ matrix $A$, with $D(\cdot)$ as in Assumption B1 and the $K \times 1$ vector of normalised functions $P(\cdot)$ as defined above, setting

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

Consider a linear operator $a(g) = (g(x_1), \cdots, g(x_d))'$, for some $(x_1, \cdots, x_d) \in \mathscr{X}^d$. 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-covariance 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 an insight into the the matrix $\bar{V}_n$ and its role in the statement of the asymptotic distribution, note that one may alternatively write

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

where $A^* := (D(p_1), ..., D(p_K))' = B_K^{1/2}A \in \mathbb{R}^{K\times d}$, 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-covariance matrix of a nonlinear function of least squares estimates, where the matrix $A^*$ is the Jacobian term and $B_K^{-1}\left[\frac{1}{n}\sum_{i,k=1}^{n} \gamma_{ik}\sigma(X_i)\sigma(X_k)p^K(X_i)p^{K'}(X_k)\right]B_K^{-1}$ is the conditional variance-covariance matrix of LS estimates for a possibly misspecified model. Assumption B3 below specifies the conditions under which $\bar{V}_n$ is the correct normalising matrix to be used in the statement of the asymptotic result of Theorem 2 below.

Two alternative representations of $\bar{V}_n$ in terms of $P(\cdot)$ or $p^K(\cdot)$ were given above. In the statement of assumptions and theorems, quantities will be written in terms of the vector of normalized functions $P(\cdot)$ to facilitate discussion of the quantity $\bar{V}_n$ in a more tractable manner. We shall need the following assumptions.

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

$$\begin{aligned}
&\text{(i)} \quad \xi^2(K)tr(\Sigma_n) = o(n^{1/2}). \\
&\text{(ii)} \quad K^3\xi^6(K)tr(\Sigma_n)\left(\frac{1}{n} + \frac{\triangle_n}{n^2}\right) = o(1). \\
&\text{(iii)} \quad n\xi^2(K)K^{-2\alpha+1} = o(1).
\end{aligned}$$

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 $X_i$. Uniform consistency of Theorem 1 required smoothness condition $\xi(K)K^{-\alpha} = o(1)$ on the unknown $g_0$, while deriving asymptotic normality in Theorem 2 needs a stronger smoothness condition of B3 (iii). Revisiting the case of bounded functions $P_{sK}(\cdot)$'s and weakly dependent $e_i$'s 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.** *$K$ and 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 $\varepsilon_j$ of any particular $j$ 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 in the case when the random matrix $\bar{V}_n$ converges to a finite nonsingular matrix, considered in the next section of $\sqrt{n}$ rate of convergence. Validation of such convergence requires stronger restrictions both on the functional $a(\cdot)$ and the strength of dependence in $X_i$'s and $U_i$'s. Theorem 3 allows $\bar{V}_n$ to diverge with $n$ as long as approximation $\|\hat{V}_n - V_n\| = o(1)$ holds for some sequence of deterministic nonsingular matrices $V_n$. Such approximation still requires certain, although weaker, restrictions to be placed on the strength of dependence in $X_i$'s and $U_i$'s. 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 certainly assumes the derivative matrix $A$ to have rank $d$ for all $K \ge d$. Throughout this work, denote by $A^{1/2}$ the unique positive definite square root of a positive definite matrix $A$.

**Theorem 2 (Asymptotic Normality).** *Under assumptions A1-A5 and B1-B5,*

$$\sqrt{n}\bar{V}_n^{-1/2}(\hat{\theta} - \theta_0) \to_d N(0, I_d), \quad as \quad n \to \infty. \tag{4.1}$$

The proof of Theorem 2 is given in the Appendix A.

## 4.1 Properties of $\bar{V}_n$

The conditional covariance matrix $\bar{V}_n$ is a random quantity. In this section we study conditions, under which $\|\bar{V}_n - V_n\|$ converges to zero, where

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

This will allow us to present the asymptotic distribution result (4.1) for $(\hat{\theta} - \theta_0)$ with normalisation $V_n$. In Theorem 3 below, the $i^{th}$ element of the $d \times 1$ estimator $\hat{\theta}$ 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 an intuition of implications of this rate, let's focus on the case of scalar $a(\cdot)$ in this paragraph. We rule out the possibility of shrinking $V_n$ which corresponds to presence of negative dependence in $X_i$'s or $U_i$'s, as this is rather unlikely for real data. The above expression of $V_n$ indicates that $V_n = O(1)$ would correspond to the case of short range dependence in the combined quantity $A'P(X_i)U_i$ if $K$ were fixed. This may still allow for 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 state restrictions on the strength of dependence in $X_i$'s and $U_i$'s across $i$. Again, an upper bound is imposed on the rate of increase in the measure of bivariate dependence in $X_i$, $\triangle_n$.

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

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

Assumption B6 indicates how the dependence in the data restricts the choice of the bandwidth parameter $K$ and series functions. The stronger the dependence is, the slower the rate of increase in $K$ and $\xi(K)$ is required to be, leading to further repercussions on the smoothness in Assumption B3 (iii), where a larger value of $\alpha$ 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 their 4th order joint cumulant. The following definition is required to do this.

**Definition 2.** Let $Z_1, Z_2, Z_3, Z_4$ be zero-mean random variables with finite fourth moments. Then, the joint cumulant of these four random variables is defined as

$$\begin{aligned}\kappa(Z_1, Z_2, Z_3, Z_4) &:= E(Z_1Z_2Z_3Z_4) - E(Z_1Z_2)E(Z_3Z_4) - E(Z_1Z_3)E(Z_2Z_4) - E(Z_1Z_4)E(Z_2Z_3) \\ &= Cov(Z_1Z_2, Z_3Z_4) - Cov(Z_1, Z_3)Cov(Z_2, Z_4) - Cov(Z_1, Z_4)Cov(Z_2, Z_3).\end{aligned}$$

Recalling $A = (A_1, \cdots, A_K)' \in \mathbb{R}^{K\times d}$, introduce the following notations:

$$\begin{aligned} h_i^{(\ell)} &:= \sigma(X_i)A_\ell' P(X_i), \qquad (4.2)\\ \bar{h}_i^{(\ell)} &:= \sigma(X_i)A_\ell' P(X_i) - E\left(\sigma(X_i)A_\ell' P(X_i)\right), \quad 1 \le i \le n, \quad 1 \le \ell \le d.\end{aligned}$$

The latter term is a de-meaned version of the former, introduced here so that we can make use of the definition of joint cumulant for mean-zero random variables.

In the time series literature, see e.g. Brillinger (1968), the weak dependence characterization in terms of cumulants typically implies that the $4^{th}$ order cumulant satisfies,

$$\left|\sum_{i_1,i_2,i_3,i_4=1}^{n} \kappa(Z_{i_1}, Z_{i_2}, Z_{i_3}, Z_{i_4})\right| = O(n). \qquad (4.3)$$

**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

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

Comparing Assumption B7 to (4.3), one observes that Assumption B7 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 $\{\bar{h}_i^{(\ell)}\}$ are weakly dependent, c.f. (4.3):

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

noting $|\gamma_{ik}| \le \sqrt{\gamma_{ii}\gamma_{kk}} \le C < \infty$, $i, k = 1, \cdots, n$, $n \ge 1$.

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

The following theorem establishes asymptotic normality if $\hat{\theta}$.

**Theorem 3** *Under Assumptions B7-B8,*

$$\|\bar{V}_n^{-1}\| = O_p(1), \quad \text{and,} \tag{4.4}$$

$$\|\bar{V}_n - V_n\| = o_p(1). \tag{4.5}$$

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

$$\sqrt{n} V_n^{-1/2}(\hat{\theta} - \theta_0) \to_d N(0, I_d). \tag{4.6}$$

Consistent estimation of $V_n$ that facilitates inference based on Theorem 3 typically requires availability of some additional information in the cross-sectional setting, as discussed in Section 1. In the context of weakly dependent time series data, Chen, Liao and Sun (2011) found that under certain conditions on the functional $a(\cdot)$ that preclude the $\sqrt{n}$ rate of convergence of $a(\hat{g})$, $V_n$ reduces asymptotically to the same matrix as under the assumption of independence, which in our setting is equal to $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \gamma_{ii} E\big[\sigma(X_i)^2 A' P(X_i) P'(X_i) A\big]$. This engenders a particular interest in devising, for the spatial setting, a method of robust inference for the case of $\sqrt{n}$ rate of convergence, where no such simplification of the asymptotic variance holds. This problem is the focus of the next section.

# 5 $\sqrt{n}$ rate inference

Theorem 3 provides sufficient conditions for convergence $\sqrt{n} V_n^{-1/2}(a(\hat{g}) - a(g_0)) \to_d N(0, I_d)$, where $V_n$ is a $d \times d$ matrix that may grow with $n$. In this section, we establish sufficient conditions under which $V_n$ converges to a finite limit $V$, as $n \to \infty$, which in turn implies the parametric $\sqrt{n}$ rate convergence of $\hat{\theta}$ to $\theta_0$. Attainment of the parametric rate of convergence by some semiparametric estimates have received wide interest in econometric literature, starting from Robinson (1988) and Powell, Stock and Stocker (1989). This type of results are available for the two well-known semiparametric models: the single index model and partly linear model. While in the kernel estimation each semiparametric model needs to be considered separately, Newey (1997) has shown that series estimation allows introducing a general semiparametric estimate encompassing both afore-mentioned popular models, enabling attainment of a unified theory of $\sqrt{n}$-rate of convergence. Chen and Shen (1998) obtained similar results for weakly dependent time series case. It is of interest to extend these results to the setting of cross-sectional dependence, since semiparametric estimates, such as in the partly linear regression model, are widely used in empirical works, generating a need for a method of inference robust against general spatial dependence and heterogeneity. This section provides a data-driven studentization method that overcomes certain limitations of the existing alternatives.

### 5.1 Partly linear regression model

Before starting the formal statement of theory, we discuss the partly linear regression model in some detail, as the semiparametric estimate of this model satisfies the conditions of this section and will be used in the Monte Carlo study and empirical examples. This model is a popular alternative to the fully nonparametric regression model and imposes a restriction on the nonparametric function $g_0(\cdot)$ that a $d$-dimensional subset of the regressors enter $g_0(\cdot)$ linearly. For notational convenience, denote this subset by $Z_i$ and the remaining regressors by $X_i$. Then the model can written as

$$Y_i = Z_i'\delta_0 + h_0(X_i) + U_i, \tag{5.1}$$

where $h_0(\cdot)$ is a function of unknown non-parametric form. The model is particularly suitable when $Z_i$ are categorical variables, and is often used when the number of regressors is large since the fully nonparametric specification suffers from the curse of dimensionality. This model has received much attention in kernel estimation, see e.g. Robinson (1988) and Fan and Li (1999), where it has been noted that the parameter $\delta_0$ can be estimated at the $\sqrt{n}$ rate despite the first stage nonparametric estimate having a slower-than-$\sqrt{n}$ rate of convergence.

Series estimation of (5.1) had been considered in Chamberlain (1986), where the choice of the series functions takes into account the partly linear regression form. The first $d$ number of series functions are set to be $Z_i$, while the remaining $K - d$ number of series functions include only $X_i$ in their arguments. The series estimate of $\delta_0$ is then the first $d$ elements of $\hat{\beta}$, and $\hat{h}(x) = \hat{g}(z, x) - z'\hat{\delta}$. It is notable that the series estimation of $\delta_0$ is very different from that in kernel estimation literature, where the first step nonparametric regression estimates of $Y$ and $Z$ in terms of $X$ are required.

There is more than one functional $a(\cdot)$ that yields $a(g_0) = \delta_0$. Andrews (1991) notes $a(g_0) = \partial g_0(x, z)/\partial z = \delta_0$ for any value of $z$. In this work, we use the following functional as in Newey (1997), since this facilitates verification of conditions for $\sqrt{n}$-consistency. Denote $Z^* = Z - E(Z|X)$. Suppose $E(Z^*Z^{*\prime})$ is a non-singular matrix, which is an identification condition for $\delta_0$, and consider the following functional of $g_0$:

$$\begin{aligned} a(g_0) &:= E\left\{[E(Z^*Z^{*\prime})]^{-1}Z^*g_0(X,Z)\right\} \\ &= [E(Z^*Z^{*\prime})]^{-1}\left\{E(Z^*Z')\delta_0 + E[Z^*h_0(X)]\right\} = \delta_0. \end{aligned} \tag{5.2}$$

The last equality follows from

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

since $E[ZE(Z'|X)] = E[E(Z|X)E(Z'|X)]$ by the law of iterative expectation, and

$$E[Z^*h_0(X)] = E[Zh_0(X)] - E[E(Z|X)h_0(X)] = 0.$$

To see how this functional can be used to characterise the series estimate of $\delta_0$, denote $\hat{\beta} = (\hat{\delta}', \hat{\lambda}')'$ and $p(x,z) = (z', q(x)')'$, where $q(\cdot)$ is the vector of $K-d$ series functions in terms of $x$. Then, $\hat{g}(x,z) = z'\hat{\delta} + q(x)'\hat{\lambda}$ hence,

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

## 5.2 $\sqrt{n}$ rate of convergence

Returning to the discussion of the $\sqrt{n}$ rate of convergence, the following assumption states the key condition and is from Newey (1997).

**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) $D(g_0) = E[w(X_i)g_0(X_i)]$, $D(P_{sK}) = E[w(X_i)P_{sK}(X_i)], 1 \le s \le K$ *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 $\delta_K$.*

Discussion of 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 $D(\cdot)$ that is used in the Riesz representation of $D(\cdot)$. Assumption C1 (iii) requires such function $w(\cdot)$ to lie in the linear span of the series functions. Newey (1997) explicitly verifies 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. The verification for the partly linear regression case is straightforward in the view of (5.2). Interested readers are referred to pp. 155 of Newey (1997).

By Assumption C1, $D(P_{sK}) = E[w(X_i)P_{sK}(X_i)]$, $1 \leq s \leq K$. Thus, 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, define $d \times d$ matrix $W_n$ where $v_K(\cdot)$ is replaced by the function $w(\cdot)$:

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

The following assumption provides sufficient conditions for $\sqrt{n}$ rate of convergence of $a(\hat{g})$ to $a(g_0)$.

**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 (2010). Assumption C2 (ii) is a weak dependence restriction for $e_i$'s.

**Theorem 4. ($\sqrt{n}$ rate of convergence).** *Under assumptions C1 and C2,*

$$V_n \to V < \infty, \quad as \quad n \to \infty. \tag{5.3}$$

*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 has obtained the $\sqrt{n}$ rate of convergence for certain semiparametric estimates under weak dependence. The asymptotic variance-covariance matrix $V$ is unknown and needs to be estimated to construct a confidence interval or carry out hypothesis testing for the unknown $\theta_0$. The next subsection considers the issues related to this.

### 5.3 Studentization

In the earlier Section 1, possible problems of using the HAC estimator in the cross-sectional setting have been discussed. Under the conditions for $\sqrt{n}$ rate of convergence of $a(\hat{g})$ to $a(g_0)$ given in this section, it is possible to construct a new studentization for $a(\hat{g}) - a(g_0)$ that does not require availability of economic distances. Theorem 4 states $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, V)$, where the matrix

$$V = \lim_{n\to\infty} A'\mathbf{P}'E(UU'|X)\mathbf{P}A/n = \lim_{n\to\infty} A^{*\prime}B_K^{-1}\mathbf{p}'E(UU'|X)\mathbf{p}B_K^{-1}A^*/n$$

is unknown. It is not possible to consistently estimate $V$, unless one resorts to additional information of suitable distance measures, as considered in Conely (1999), Kelejian and Prucha (2007) and Robinson and Thawornkaiwong (2010). Instead, we devise a matrix $\hat{C}_n$, defined in the subsequent discussion, such that the limit of $\sqrt{n}\hat{C}_n^{-1/2}(\hat{\theta}_n - \theta_0)$ is free from unknown parameters. Similar idea was used in a setting of linear OLS estimation in Kiefer, Vogelsang and Bunzel (2000).

Recall some notations: $B_K = E(p^K(X_i)p^K(X_i)')$, $P(x) = B_K^{-1/2}p^K(x)$, $A = (D(P_{1K}), \cdots, D(P_{KK}))' \in \mathbb{R}^{K\times d}$, $A^* = (D(p_1), ..., D(p_K))' = B_K^{1/2}A \in \mathbb{R}^{K\times d}$ with $D(\cdot)$ from Assumption B1(i). Denote by $\hat{A}^*$ and $\hat{B}_K$ the estimates of the corresponding true values $A^*$ and $B_K$.

$$\hat{A}^* := \frac{\partial a(p^{K\prime}\beta)}{\partial \beta}\Big|_{\beta=\hat{\beta}}, \quad \hat{B}_K := \mathbf{p}'\mathbf{p}/n = \sum_{i=1}^{n} p^K(X_i)p^K(X_i)'/n. \tag{5.4}$$

Given $\hat{A}^*$ and $\hat{B}_K$, we can construct the sample analogue of $A'\mathbf{P}'U/\sqrt{n}$ by $\hat{A}^{*\prime}\hat{B}_K^{-1}\mathbf{p}'\hat{U}/\sqrt{n}$, where $\hat{U} = Y - \hat{G}$, with $\hat{G} = \big(\hat{g}(X_1), \cdots, \hat{g}(X_n)\big)$, is the $n \times 1$ vector of corresponding residuals. To introduce $\hat{C}_n$, set

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

Now, define

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

where $W_d(\cdot)$ denotes a $d$-dimensional vector of independent Brownian motions and $\Psi_d$ is the integral of the outer product of $d$-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 $r \in [0,1]$;

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

Previously, Assumption C2 (ii) of Theorem 4 required $e_i$'s to be weakly dependent. C3 (ii) further rules out presence of any "dominant" unit whose error covariances with new units added to the sample are persistently significant. To see implications of Assumption C3 (i), recall Assumption C2 (ii): $\sum_{i,k=1}^{n} |\gamma_{ik}| = \sum_{i=1}^{n} \gamma_{ii} + \sum_{i,k=1:i\neq k}^{n} |\gamma_{ik}| = O(n)$. The first summation is always $O(n)$, and there are two possibilities for the second summation, either $\sum_{i,k=1:i\neq k}^{n} |\gamma_{ik}| = o(n)$ or, the knife-edge case that $\sum_{i,k=1:i\neq k}^{n} |\gamma_{ik}|$ grows at the exact rate of $n$. In the former case, Assumption C3 (i) imposes no further restriction while in the latter case, some falling-off of dependence as $|i-k|$ increases is required. Both C3 (i) and (ii) are natural implications of weak dependence in the time series context where the dependence is a fast-decreasing function of the distance in time. The current setting differs from the time series in two ways; firstly it allows $\gamma_{ik} = \gamma_{ikn}$ to admit a triangular array structure, and secondly it relaxes the link between $\gamma_{ik}$ 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, Appendix B. If $\gamma_{ik}$ takes on a triangular array structure, as allowed in the pure SAR model, 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, as long as the bilateral interaction $\gamma_{ikn}, k = 1,2,\cdots$, falls suitably fast in $n$, whereas C3 (i) requires a faster uniform-in-$n$ rate of reduction in $\gamma_{ikn}$ as $|i-k|$ increases.

**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 $X_i$'s, whereas Assumption C4 (iii) is a restriction on the choice of the approximating functions, requiring their 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 C (iv) strengthens the smoothness condition of Assumption B3 (iii).

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

Recall the functional derivative $D(\cdot)$ from Assumptions B1 and B2. It is Frechet differential of the functional $a(\cdot)$, evaluated at $g_0$. 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* $|\tilde{g} - g_0|_\infty \leq \epsilon$ *and* $|\bar{g} - g_0|_\infty \leq \epsilon$, $\|D_i(g;\tilde{g}) - D_i(g;\bar{g})\| \leq C|g|_\infty|\tilde{g} - \bar{g}|_\infty$, $i = 1, \cdots, d$.

Assumption C5 is the same as in Newey (1997) and requires the functional derivatives $D_i(\cdot;g)$ to exhibit continuity over $g$, the point at which the derivative is taken.

The following theorem shows that the asymptotic distribution for the estimation error $\hat{\theta} - \theta_0$, when studentized by the matrix $\hat{C_n}$, is free from the unknown variance matrix $V$ and only depends on $d$, and is non-Gaussian.

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

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

Now, suppose we are interested in testing the hypothesis $H_0 : a(g_0) = r$ against the alternative $H_1 : a(g_0) \neq r$ for a $d \times 1$ fixed vector $r$. Then the test statistic can be constructed as $t_n^* := n(\hat{\theta} - r)'\hat{C}_n^{-1}(\hat{\theta} - r)$. Since $t_n^* = \|\sqrt{n}(\hat{\theta} - r)'\hat{C}_n^{-1/2}\|^2$, Theorem 5 implies the following result.

**Theorem 6.** Under 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 $c_\alpha$, satisfying $Pr(t_n \leq c_\alpha) \rightarrow 1 - \alpha$, required to carry out hypothesis tests can be obtained from Table 2 of Kiefer *et al.* (2000) for $d = 1, \cdots, 30$. In particular, for $d = 1$, $c_{5\%} = 46.39$, and $c_{10\%} = 28.88$. Correspondingly, the $97.5^{th}$ and $95^{th}$ percentiles for $\Psi_1^{-1/2}W_1(1)$ in Theorem 5 are $\sqrt{46.39}$ and $\sqrt{28.88}$.

## 6 Monte Carlo Study of Finite-Sample Performance

In this section, we focus on the partly linear model of (5.1) where regressors $X_i$ and $Z_i$ are both one-dimensonal:

$$Y_i = \delta_0 Z_i + h_0(X_i) + U_i.$$

It was noted in Section 5 that the functional $a(g_0) = \delta_0$ of (5.2) satisfies the conditions of Theorem 4. Therefore the studentization devised in Theorem 5 and 6 applies. We set the true model at $\delta_0 = 0.3$ and $h_0(x) = \log(1 + x^2)$.

To generate the data, we follow the random location setting used in the Monte Carlo study of Robinson and Thawornkaiwong (2010), where the locations of the observations, denoted $s_1, \cdots, s_n$, were generated by a random draw from the uniform distribution over $[0, 4n^{1/2}] \times [0, 4n^{1/2}]$. This setting was chosen over more tractable alternatives such as SAR or lattice setting since it offers a model closer to what may be envisaged in many economic applications. Keeping these locations fixed across replications, $U_i$ and $Z_i$ were generated independently as scalar normal random variables with mean zero and covariances $Cov(U_i, U_j) = Cov(Z_i, Z_j) = \rho^{\|s_i - s_j\|}$. To construct $X_i$, we generate another scalar normal random variable $V_i$ in the same way as $U_i$ and $Z_i$ and let $X_i = 1 + V_i + 0.5Z_i$. The dependent variable is then formed as $Y_i = \log(1 + X_i^2) + 0.3Z_i + U_i$.

It is important to note that the dependence parameter $\rho$ does *not* carry the same meaning as the AR coefficient of time series, and $|\rho| < 1$ is not a sufficient condition for weak dependence. It has not been possible to ascertain the exact value of the boundary between weak and strong dependence in terms of $\rho$, however, it is clear that it lies significantly below 1. To see this, consider, as the corresponding deterministic location setting, a two-dimensional lattice data with each axis of $\sqrt{n}$-length. There are $n$ number of points on this plane and the summation of all pair-wise distances is a lot smaller than the summation of such distances when the $n$ points are lined up along a line of length $n$ instead, as in stationary AR (1) time series. Recall weak dependence is characterised by $\sum_{i,k=1}^{n} \gamma_{ik} = O(n)$, the LHS of which in the lattice setting is

$$\sum_{i,k=1}^{n} \rho^{\|s_i - s_k\|}, \quad \text{with} \quad s_i, s_k \in \{(1,1)', (1,2)', \cdots, (1, [\sqrt{n}])', (2,1)', \cdots, ([\sqrt{n}], [\sqrt{n}])'\},$$

recalling that location $s_i$, of the $i^{th}$ observation, is the vector of coordinates in the two dimensional lattice. On the other hand, the LHS is $\sum_{i,k=1}^{n} \rho^{|i-k|}$ for the AR(1) time series. It is obvious that for any value of $0 < \rho$, the former summation is larger than the latter. Therefore, values of $0 < \rho < 1$ close to 1 yield strong dependence in the two-dimensional lattice setting under consideration.

To cover a wide-range of different settings, 16 combinations of $n = 100, 200, 400, 1200$ and $\rho = 0, 0.2, 0.4, 0.8$ were considered. The value $\rho = 0.8$ was chosen, having in mind that this case may correspond to a strong dependence case. For each combination, three values of $K = 3, 5, 8$ were tried and 2000 iterations carried out. For the series functions of $X_i$, the first $K - 1$ orthonormal Legendre polynomials were used.

The first objective of this simulation study is to analyse the finite sample performance of the series estimation for both the nonparametric function $g$ and semiparametric quantity $a(g)$ under differing sample sizes, degrees of dependence and choices of K. We report in Table 1 the Monte Carlo MSE, bias and variance of the nonparametric regression estimate at a fixed point $(x, z) = (0.5, 0.5)$, and the Monte Carlo integrated MSE, defined as $E[(\hat{g}(X_i) - g_0(X_i))^2]$ conveying how the nonparametric estimation performs globally. Table 1 also contains the Monte Carlo MSE of the estimate $\hat{\delta}$ of $\delta_0$. The Monte Carlo variance and bias of the nonparametric estimate at a fixed point are in line with the prediction that larger values of $K$ reduce the bias while increasing variance. As for the Monte Carlo MSE for $\hat{g}(0.5, 0.5)$, under all four values of $\rho$, $K = 5$ led to the smallest MSE for $n = 100$ while $K = 8$ did so for $n = 200, 400, 1200$. For $n = 200, 400, 1200$, the larger the sample size, the more dramatic was the improvement in the MSE from using a larger $K$, as the reduction in the bias was offset by a relatively small increase in variance. The Monte Carlo MISE on the other hand suggested $K = 5$ to be the best out of three for $n = 100, 200$ and $K = 8$ for $n = 400, 1200$. The Monte Carlo MSE of the semiparametric estimate $\hat{\delta}$ shows remarkable invariance to the choice of $K$ across all of the 16 settings, which is especially important as the optimal choice of the bandwidth parameter $K$ for semiparametric estimate is often more difficult than in the case of nonparametric estimation. See Robinson and Thawornkaiwong (2010) for a discussion.

The second objective is to investigate how the studentization of Section 4 performs in finite samples. Theorem 5 implies in this setting,

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

$$\frac{\sqrt{n}(\hat{\delta} - \delta_0)}{\sqrt{\hat{C}_n}} \to_d \frac{W_1(1)}{\sqrt{\Psi_1}}.$$

Kiefer *et al.* (2000, Table 2) give simulated values of the percentiles of $W_1^2(1)/\Psi_1$, from which the corresponding percentiles of the square-rooted quantity $W_1(1)/\sqrt{\Psi_1}$ can be easily derived. The $99.5^{th}, 97.5^{th}$ and $95^{th}$ percentile of $W_1(1)/\sqrt{\Psi_1}$ are $\sqrt{101.2}$, $\sqrt{46.39}$ and $\sqrt{28.88}$, respectively. Based on this, we construct the asymptotic 95% confidence interval for $\delta_0$:

$$\Pr\left(\delta_0 \in \left[\hat{\delta} - \sqrt{46.39\frac{\hat{C}_n}{n}}, \hat{\delta} + \sqrt{46.39\frac{\hat{C}_n}{n}}\right]\right) \to 0.95.$$

Table 2 reports the Monte Carlo average length of the 95% confidence intervals. The length of confidence intervals decreases with the sample size, and does not report much variation over the choice of $K$ or $\rho$.

Table 3 reports the empirical coverage probabilities for the 99%, 95% and 90% asymptotic confidence intervals. For $\rho = 0, 0.2, 0.4$, the reported coverage probabilities are remarkably precise even for small sample sizes of 100 and 200. For $\rho = 0.8$, the coverage probabilities are poor, however this is not surprising given that in this random location setting $\rho = 0.8$ is thought to result in strong dependence, therefore the studentization is not applicable here as conditions of Theorems 5 and 6 are violated.

In Table 4, we examine finite sample performance of the studentization when the data exhibit weak dependence near the boundary of strong dependence. This is of particular interest because finite sample behaviour of the HAC-type estimates in such case are known to be poor. For illustration, we briefly divert to standard time series Gaussian AR(1) setting, where $|\rho| < 1$ is the condition for weak dependence. Table 4 reports the coverage probabilities for $\rho = 0.6, 0.7, 0.8, 0.9$. The results show that although the coverage probabilities are not strikingly satisfactory for the small sample size of 100, they improve noticeably with increasing sample size. The reported coverage probabilities become rather precise for the sample size of 200 even for the value $\rho = 0.8$, which represents the case of high persistence in such a small sample. Even for $\rho = 0.9$, the coverage probabilities are rather precise for sample size of 1200.

Table 5 reports empirical power of testing $H_0 : \delta_0 = \delta$ against $H_1 : \delta_0 \neq \delta$, for $\delta = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 1$. Since the true $\delta_0$ used in data generation is 0.3, the fourth column reports empirical size of the test. The results are in line with the prediction that the power improves with increasing sample size. For $\delta = 0.2, 0.4$, stronger dependence led to better power.

Figures 1-8 plot the histograms of the standardized quantity $(\hat{\delta} - \bar{\delta}_{MC})/SD_{MC}(\hat{\delta})$, for $\rho = 0.2, 0.8$ where $\bar{\delta}_{MC}$ is the average of $\hat{\delta}$ from 2000 iterations and $SD_{MC}(\hat{\delta})$ is the Monte Carlo standard deviation. The probability density function of a standard normal random variable is superimposed for the ease of comparison. The histograms show that the normal approximation appears good even for the sample size $n = 100$.

Figures 7-16 plot the histograms of the studentized quantity from Theorem 5: $\hat{C}_n^{-1/2}\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d \Psi_d^{-1/2} W_d(1)$. There is a clear difference between $\rho = 0.2$ and $\rho = 0.8$, with the latter leading to fatter non-Gaussian tails.

Table 1: Monte Carlo MSE, Variance and Bias

| | | $\rho = \mathbf{0}$ | | | | | $\rho = \mathbf{0.2}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $K$ | $MSE(\hat{g}_x)$* | $Var(\hat{g}_x)$ | $Bias(\hat{g}_x)$ | $MISE(\hat{g})$ | $MSE(\hat{\delta})$ | $MSE(\hat{g}_x)$ | $Var(\hat{g}_x)$ | $Bias(\hat{g}_x)$ | $MISE(\hat{g})$ | $MSE(\hat{\delta})$ |
| 100 | 3 | 0.1852 | 0.0238 | 0.4018 | 0.194 | 0.0147 | 0.1879 | 0.0279 | 0.4001 | 0.1959 | 0.0153 |
| | 5 | 0.032 | 0.0261 | 0.077 | 0.0592 | 0.0131 | 0.036 | 0.0301 | 0.0766 | 0.0632 | 0.0134 |
| | 8 | 0.0375 | 0.0375 | 0.0045 | 0.081 | 0.0135 | 0.0416 | 0.0416 | 0.005 | 0.0848 | 0.0136 |
| 200 | 3 | 0.1805 | 0.0128 | 0.4096 | 0.1853 | 0.0074 | 0.1821 | 0.0139 | 0.4102 | 0.1842 | 0.0081 |
| | 5 | 0.0219 | 0.0132 | 0.0934 | 0.0368 | 0.0066 | 0.0235 | 0.0149 | 0.093 | 0.0381 | 0.0069 |
| | 8 | 0.0174 | 0.017 | 0.021 | 0.041 | 0.0067 | 0.0187 | 0.0184 | 0.0182 | 0.0423 | 0.0069 |
| 400 | 3 | 0.1724 | 0.0067 | 0.407 | 0.1791 | 0.0037 | 0.174 | 0.0069 | 0.4087 | 0.18 | 0.004 |
| | 5 | 0.0168 | 0.0071 | 0.0985 | 0.0267 | 0.0032 | 0.0171 | 0.0075 | 0.0979 | 0.027 | 0.0035 |
| | 8 | 0.0085 | 0.0081 | 0.021 | 0.0218 | 0.0033 | 0.0093 | 0.0088 | 0.0214 | 0.022 | 0.0035 |
| 1200 | 3 | 0.1713 | 0.0021 | 0.4114 | 0.1764 | 0.0012 | 0.1724 | 0.0022 | 0.4125 | 0.1774 | 0.0013 |
| | 5 | 0.0139 | 0.0024 | 0.1072 | 0.0202 | 0.001 | 0.014 | 0.0025 | 0.1074 | 0.0203 | 0.0011 |
| | 8 | 0.0034 | 0.0026 | 0.0289 | 0.009 | 0.001 | 0.0036 | 0.0027 | 0.0304 | 0.0092 | 0.0011 |
| | | $\rho = \mathbf{0.4}$ | | | | | $\rho = \mathbf{0.8}$ | | | | |
| $n$ | $K$ | $MSE(\hat{g}_x)$ | $Var(\hat{g}_x)$ | $Bias(\hat{g}_x)$ | $MISE(\hat{g})$ | $MSE(\hat{\delta})$ | $MSE(\hat{g}_x)$ | $Var(\hat{g}_x)$ | $Bias(\hat{g}_x)$ | $MISE(\hat{g})$ | $MSE(\hat{\delta})$ |
| 100 | 3 | 0.1978 | 0.0312 | 0.4081 | 0.2006 | 0.0167 | 0.252 | 0.1054 | 0.3829 | 0.2645 | 0.0323 |
| | 5 | 0.0411 | 0.0334 | 0.0874 | 0.0669 | 0.0149 | 0.1115 | 0.1055 | 0.0779 | 0.1482 | 0.0298 |
| | 8 | 0.0443 | 0.0441 | 0.0171 | 0.088 | 0.0154 | 0.1172 | 0.1171 | 0.0106 | 0.1663 | 0.0301 |
| 200 | 3 | 0.1799 | 0.0139 | 0.4074 | 0.1859 | 0.008 | 0.2007 | 0.0536 | 0.3834 | 0.2204 | 0.0173 |
| | 5 | 0.0237 | 0.0154 | 0.0907 | 0.039 | 0.0071 | 0.0597 | 0.0524 | 0.085 | 0.0866 | 0.016 |
| | 8 | 0.02 | 0.0199 | 0.0124 | 0.0431 | 0.0072 | 0.0567 | 0.0566 | 0.0121 | 0.0909 | 0.0161 |
| 400 | 3 | 0.1747 | 0.0079 | 0.4084 | 0.1816 | 0.0042 | 0.1958 | 0.0289 | 0.4085 | 0.2003 | 0.0096 |
| | 5 | 0.0185 | 0.0088 | 0.0984 | 0.0283 | 0.0036 | 0.038 | 0.0279 | 0.1005 | 0.0535 | 0.0086 |
| | 8 | 0.0102 | 0.0098 | 0.0201 | 0.0233 | 0.0036 | 0.0299 | 0.0293 | 0.0235 | 0.0492 | 0.0085 |
| 1200 | 3 | 0.1724 | 0.0026 | 0.412 | 0.1767 | 0.0013 | 0.1786 | 0.0101 | 0.4105 | 0.1844 | 0.0032 |
| | 5 | 0.0145 | 0.0029 | 0.1076 | 0.0205 | 0.0012 | 0.0212 | 0.01 | 0.1061 | 0.0297 | 0.0029 |
| | 8 | 0.0040 | 0.0031 | 0.0302 | 0.0096 | 0.0012 | 0.0111 | 0.0103 | 0.0285 | 0.0192 | 0.0029 |

* $\hat{g}_x$ denotes point-wise nonparametric estimate $\hat{g}(0.5, 0.5)$.

Table 2: Monte Carlo average 95 % CI length

| $n$ | $K$ | $\rho = 0$ | $\rho = 0.2$ | $\rho = 0.4$ | $\rho = 0.8$ |
|---|---|---|---|---|---|
| 100 | 3 | 0.6259 | 0.6146 | 0.6238 | 0.6014 |
| | 5 | 0.5868 | 0.5721 | 0.5833 | 0.5595 |
| | 8 | 0.5883 | 0.5714 | 0.5808 | 0.5595 |
| 200 | 3 | 0.4325 | 0.4387 | 0.4394 | 0.4358 |
| | 5 | 0.4044 | 0.4074 | 0.4076 | 0.4047 |
| | 8 | 0.4043 | 0.4054 | 0.4054 | 0.4044 |
| 400 | 3 | 0.3052 | 0.3088 | 0.3035 | 0.3132 |
| | 5 | 0.2858 | 0.2872 | 0.2847 | 0.2903 |
| | 8 | 0.2854 | 0.2855 | 0.284 | 0.2906 |
| 1200 | 3 | 0.1783 | 0.1808 | 0.1782 | 0.1785 |
| | 5 | 0.1663 | 0.1688 | 0.1667 | 0.1679 |
| | 8 | 0.1657 | 0.168 | 0.1656 | 0.1667 |

Table 3: Coverage Probabilities

| | | $\rho = \mathbf{0}$ | | | $\rho = \mathbf{0.2}$ | | |
|---|---|---|---|---|---|---|---|
| $n$ | $K\backslash 1-\alpha$ | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 |
| 100 | 3 | 0.9015 | 0.948 | 0.991 | 0.8905 | 0.9455 | 0.989 |
| | 5 | 0.9025 | 0.948 | 0.988 | 0.896 | 0.944 | 0.987 |
| | 8 | 0.8935 | 0.9405 | 0.9885 | 0.8905 | 0.9365 | 0.985 |
| 200 | 3 | 0.8965 | 0.9435 | 0.9895 | 0.892 | 0.942 | 0.987 |
| | 5 | 0.9025 | 0.9455 | 0.9895 | 0.8985 | 0.9465 | 0.991 |
| | 8 | 0.906 | 0.95 | 0.9905 | 0.899 | 0.9475 | 0.991 |
| 400 | 3 | 0.9 | 0.9435 | 0.987 | 0.901 | 0.948 | 0.99 |
| | 5 | 0.8975 | 0.9495 | 0.9925 | 0.895 | 0.9455 | 0.9895 |
| | 8 | 0.8945 | 0.947 | 0.992 | 0.8945 | 0.948 | 0.9875 |
| 1200 | 3 | 0.9135 | 0.9565 | 0.9885 | 0.9065 | 0.948 | 0.9895 |
| | 5 | 0.918 | 0.9565 | 0.9875 | 0.9005 | 0.9475 | 0.9905 |
| | 8 | 0.914 | 0.958 | 0.9895 | 0.9045 | 0.9475 | 0.988 |
| | | $\rho = \mathbf{0.4}$ | | | $\rho = \mathbf{0.8}$ | | |
| $n$ | $K\backslash 1-\alpha$ | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 |
| 100 | 3 | 0.9015 | 0.935 | 0.9845 | 0.7685 | 0.847 | 0.9385 |
| | 5 | 0.882 | 0.9385 | 0.9855 | 0.7455 | 0.8275 | 0.939 |
| | 8 | 0.877 | 0.933 | 0.9815 | 0.75 | 0.8295 | 0.933 |
| 200 | 3 | 0.9065 | 0.947 | 0.989 | 0.7685 | 0.847 | 0.94 |
| | 5 | 0.905 | 0.949 | 0.986 | 0.7475 | 0.826 | 0.9365 |
| | 8 | 0.894 | 0.9435 | 0.987 | 0.751 | 0.83 | 0.931 |
| 400 | 3 | 0.8785 | 0.937 | 0.9895 | 0.749 | 0.8285 | 0.938 |
| | 5 | 0.884 | 0.937 | 0.9885 | 0.7475 | 0.8215 | 0.9365 |
| | 8 | 0.88 | 0.9355 | 0.988 | 0.7435 | 0.83 | 0.9415 |
| 1200 | 3 | 0.8925 | 0.9465 | 0.991 | 0.734 | 0.8205 | 0.9365 |
| | 5 | 0.8875 | 0.9485 | 0.9865 | 0.724 | 0.819 | 0.9275 |
| | 8 | 0.8905 | 0.9455 | 0.987 | 0.7235 | 0.8125 | 0.9325 |

$(1-\alpha)$ is the confidence level.

Table 4: Coverage Probabilities for Time Series AR(1)

| | | $\rho = \mathbf{0.6}$ | | | $\rho = \mathbf{0.7}$ | | |
|---|---|---|---|---|---|---|---|
| $n$ | $K\backslash 1-\alpha$ | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 |
| 100 | 3 | 0.8575 | 0.9195 | 0.978 | 0.862 | 0.913 | 0.979 |
| | 5 | 0.8575 | 0.918 | 0.9755 | 0.849 | 0.9065 | 0.975 |
| | 8 | 0.858 | 0.9125 | 0.974 | 0.8495 | 0.906 | 0.9725 |
| 200 | 3 | 0.9 | 0.9415 | 0.986 | 0.881 | 0.935 | 0.985 |
| | 5 | 0.8925 | 0.943 | 0.985 | 0.8765 | 0.9355 | 0.986 |
| | 8 | 0.8925 | 0.943 | 0.984 | 0.8735 | 0.9335 | 0.985 |
| 400 | 3 | 0.905 | 0.9465 | 0.9875 | 0.8915 | 0.941 | 0.984 |
| | 5 | 0.903 | 0.9495 | 0.986 | 0.887 | 0.939 | 0.984 |
| | 8 | 0.904 | 0.95 | 0.987 | 0.8855 | 0.939 | 0.984 |
| 1200 | 3 | 0.893 | 0.9405 | 0.988 | 0.894 | 0.9465 | 0.988 |
| | 5 | 0.898 | 0.9465 | 0.987 | 0.8955 | 0.9415 | 0.9895 |
| | 8 | 0.8955 | 0.945 | 0.9895 | 0.8965 | 0.9405 | 0.989 |
| | | $\rho = \mathbf{0.8}$ | | | $\rho = \mathbf{0.9}$ | | |
| $n$ | $K\backslash 1-\alpha$ | 0.9 | 0.95 | 0.99 | 0.9 | 0.95 | 0.99 |
| 100 | 3 | 0.828 | 0.891 | 0.9585 | 0.7615 | 0.8375 | 0.935 |
| | 5 | 0.818 | 0.881 | 0.955 | 0.75 | 0.8175 | 0.927 |
| | 8 | 0.818 | 0.877 | 0.9505 | 0.7355 | 0.8145 | 0.927 |
| 200 | 3 | 0.8665 | 0.927 | 0.9795 | 0.824 | 0.893 | 0.962 |
| | 5 | 0.8645 | 0.921 | 0.979 | 0.8205 | 0.8875 | 0.958 |
| | 8 | 0.8585 | 0.9155 | 0.981 | 0.813 | 0.88 | 0.9585 |
| 400 | 3 | 0.8865 | 0.9345 | 0.987 | 0.8535 | 0.9085 | 0.9685 |
| | 5 | 0.8795 | 0.9335 | 0.988 | 0.85 | 0.9075 | 0.9685 |
| | 8 | 0.878 | 0.9345 | 0.9875 | 0.8495 | 0.9045 | 0.9685 |
| 1200 | 3 | 0.8905 | 0.938 | 0.984 | 0.8865 | 0.938 | 0.9885 |
| | 5 | 0.8925 | 0.937 | 0.985 | 0.888 | 0.942 | 0.989 |
| | 8 | 0.893 | 0.9365 | 0.9845 | 0.887 | 0.9415 | 0.989 |

$(1-\alpha)$ is the confidence level.

Table 5: Empirical power of 95% test

| | $\rho$ | $K\backslash\delta$ | 0 | 0.1 | 0.2 | **0.3***| 0.4 | 0.5 | 0.7 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|
| n=100 | 0 | 3 | 0.5005 | 0.2775 | 0.106 | 0.052 | 0.1055 | 0.2675 | 0.714 | 0.9725 |
| | | 5 | 0.5385 | 0.3045 | 0.1275 | 0.052 | 0.123 | 0.2985 | 0.76 | 0.9875 |
| | | 8 | 0.533 | 0.3125 | 0.1265 | 0.0595 | 0.1235 | 0.2875 | 0.756 | 0.981 |
| | 0.2 | 3 | 0.5105 | 0.2845 | 0.1125 | 0.0545 | 0.123 | 0.289 | 0.722 | 0.9735 |
| | | 5 | 0.5645 | 0.322 | 0.121 | 0.056 | 0.1295 | 0.328 | 0.7735 | 0.9845 |
| | | 8 | 0.573 | 0.3245 | 0.1255 | 0.0635 | 0.1245 | 0.327 | 0.7755 | 0.982 |
| | 0.4 | 3 | 0.506 | 0.306 | 0.116 | 0.065 | 0.119 | 0.2935 | 0.689 | 0.9725 |
| | | 5 | 0.5485 | 0.3265 | 0.134 | 0.0615 | 0.1275 | 0.303 | 0.742 | 0.986 |
| | | 8 | 0.5545 | 0.3325 | 0.1345 | 0.067 | 0.1355 | 0.303 | 0.7465 | 0.987 |
| | 0.8 | 3 | 0.5225 | 0.3455 | 0.2085 | 0.153 | 0.2055 | 0.3465 | 0.682 | 0.949 |
| | | 5 | 0.559 | 0.374 | 0.226 | 0.1725 | 0.2225 | 0.3705 | 0.731 | 0.9675 |
| | | 8 | 0.555 | 0.3715 | 0.2235 | 0.1705 | 0.2225 | 0.3735 | 0.7245 | 0.966 |
| 200 | 0 | 3 | 0.784 | 0.4735 | 0.1615 | 0.0565 | 0.1695 | 0.4815 | 0.912 | 0.9985 |
| | | 5 | 0.825 | 0.527 | 0.1825 | 0.0545 | 0.187 | 0.517 | 0.9295 | 1 |
| | | 8 | 0.8245 | 0.5285 | 0.1925 | 0.05 | 0.1925 | 0.522 | 0.933 | 0.9995 |
| | 0.2 | 3 | 0.765 | 0.474 | 0.1815 | 0.058 | 0.17 | 0.46 | 0.9055 | 0.9995 |
| | | 5 | 0.8145 | 0.5195 | 0.1985 | 0.0535 | 0.1905 | 0.511 | 0.937 | 0.9995 |
| | | 8 | 0.813 | 0.5245 | 0.2055 | 0.0525 | 0.19 | 0.508 | 0.9345 | 0.9995 |
| | 0.4 | 3 | 0.7565 | 0.465 | 0.169 | 0.053 | 0.1705 | 0.486 | 0.908 | 0.9985 |
| | | 5 | 0.799 | 0.521 | 0.192 | 0.051 | 0.19 | 0.532 | 0.9415 | 0.9995 |
| | | 8 | 0.7965 | 0.5255 | 0.198 | 0.0565 | 0.193 | 0.529 | 0.9465 | 1 |
| | 0.8 | 3 | 0.6985 | 0.455 | 0.2385 | 0.153 | 0.258 | 0.481 | 0.8755 | 0.996 |
| | | 5 | 0.7425 | 0.4975 | 0.261 | 0.174 | 0.2815 | 0.5175 | 0.8975 | 0.998 |
| | | 8 | 0.75 | 0.4955 | 0.264 | 0.17 | 0.282 | 0.514 | 0.9 | 0.9985 |
| 400 | 0 | 3 | 0.938 | 0.736 | 0.2865 | 0.0565 | 0.269 | 0.7345 | 0.9895 | 1 |
| | | 5 | 0.964 | 0.7715 | 0.3075 | 0.0505 | 0.3145 | 0.7725 | 0.994 | 1 |
| | | 8 | 0.964 | 0.7725 | 0.3045 | 0.053 | 0.3155 | 0.777 | 0.992 | 1 |
| | 0.2 | 3 | 0.9445 | 0.7185 | 0.28 | 0.052 | 0.2765 | 0.7085 | 0.9905 | 1 |
| | | 5 | 0.965 | 0.7765 | 0.3155 | 0.0545 | 0.305 | 0.759 | 0.9955 | 1 |
| | | 8 | 0.966 | 0.776 | 0.317 | 0.052 | 0.2995 | 0.761 | 0.9955 | 1 |
| | 0.4 | 3 | 0.9385 | 0.7195 | 0.2935 | 0.063 | 0.292 | 0.7445 | 0.9945 | 1 |
| | | 5 | 0.956 | 0.77 | 0.318 | 0.063 | 0.304 | 0.783 | 0.995 | 1 |
| | | 8 | 0.9575 | 0.779 | 0.313 | 0.0645 | 0.3065 | 0.7785 | 0.996 | 1 |
| | 0.8 | 3 | 0.898 | 0.674 | 0.35 | 0.1715 | 0.3155 | 0.6495 | 0.9745 | 1 |
| | | 5 | 0.919 | 0.7225 | 0.371 | 0.1785 | 0.3385 | 0.699 | 0.983 | 1 |
| | | 8 | 0.9225 | 0.7265 | 0.369 | 0.17 | 0.3405 | 0.7 | 0.984 | 1 |

$^*H_0 : \delta_0 = \delta, H_1 : \delta_0 \neq \delta$. Data was generated from $\delta_0 = 0.3$.

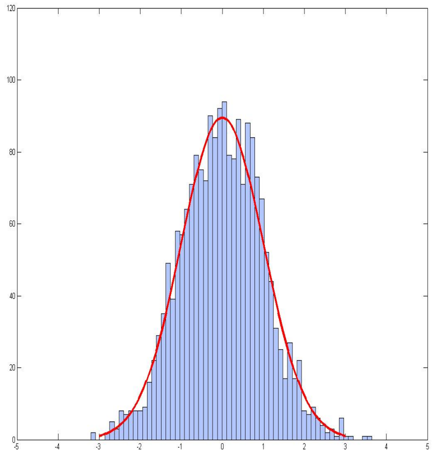


Figure 1: Histogram of $\hat{\delta}_{std}$, $\rho = 0.2, n = 100$

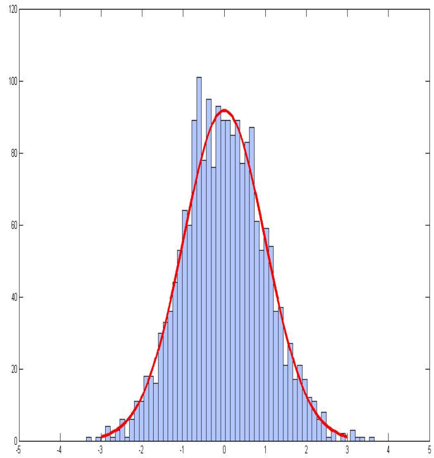


Figure 2: Histogram of $\hat{\delta}_{std}$, $\rho = 0.2, n = 200$

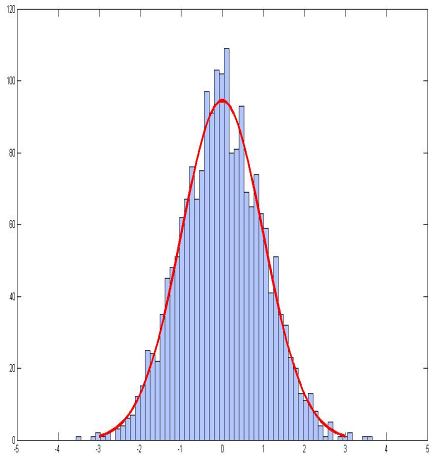


Figure 3: Histogram of $\hat{\delta}_{std}$, $\rho = 0.2, n = 400$

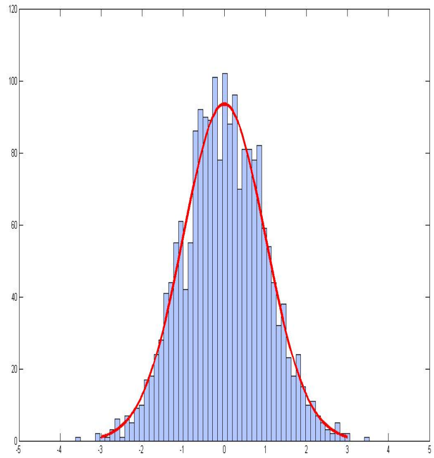


Figure 4: Histogram of $\hat{\delta}_{std}$, $\rho = 0.2, n = 1200$

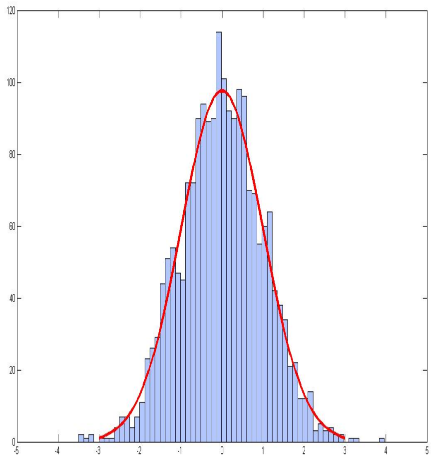

Figure 5: Histogram of $\hat{\delta}_{std}$, $\rho = 0.8, n = 100$

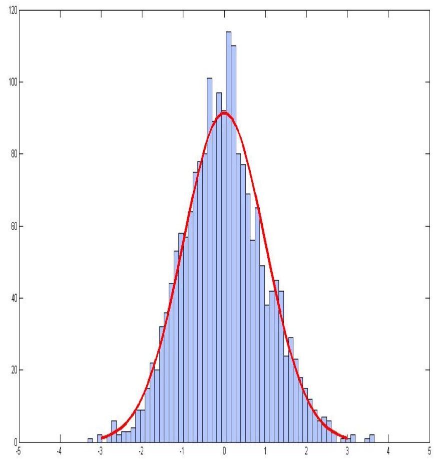

Figure 6: Histogram of $\hat{\delta}_{std}$, $\rho = 0.8, n = 200$

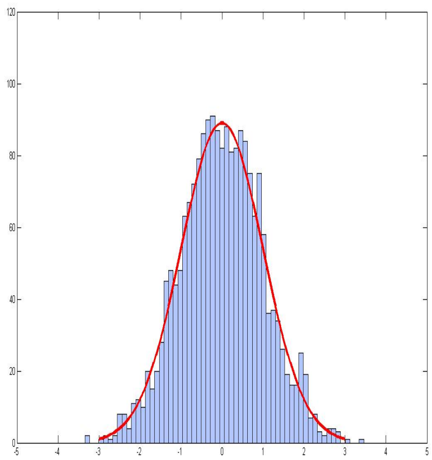

Figure 7: Histogram of $\hat{\delta}_{std}$, $\rho = 0.8, n = 400$

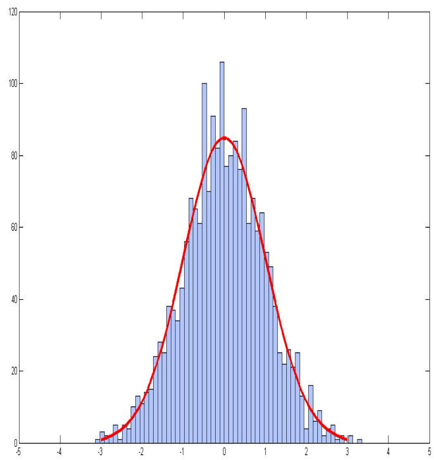

Figure 8: Histogram of $\hat{\delta}_{std}$, $\rho = 0.8, n = 1200$

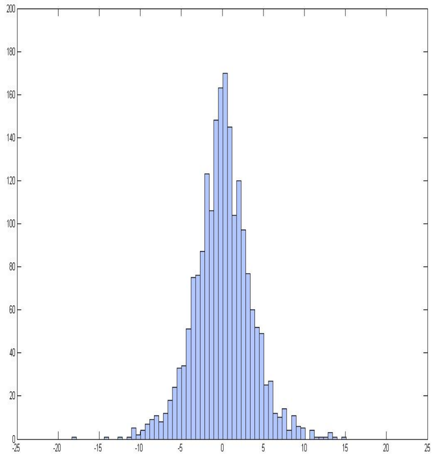

Figure 9: Histogram of $\hat{\delta}_{stu}$, $\rho = 0.2, n = 100$

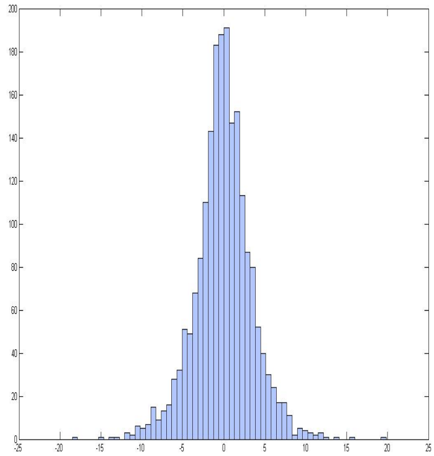

Figure 10: Histogram of $\hat{\delta}_{stu}$, $\rho = 0.2, n = 200$

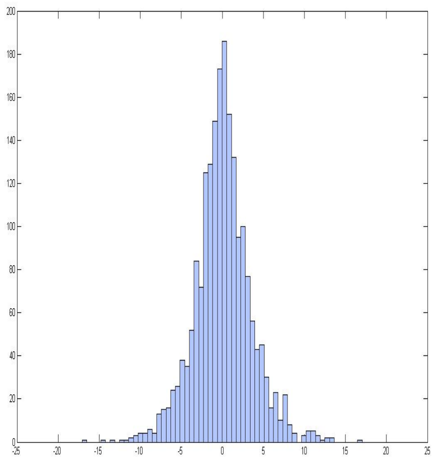

Figure 11: Histogram of $\hat{\delta}_{stu}$, $\rho = 0.2, n = 400$

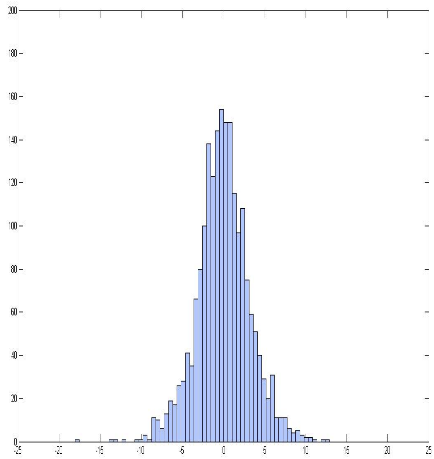

Figure 12: Histogram of $\hat{\delta}_{stu}$, $\rho = 0.2, n = 1200$

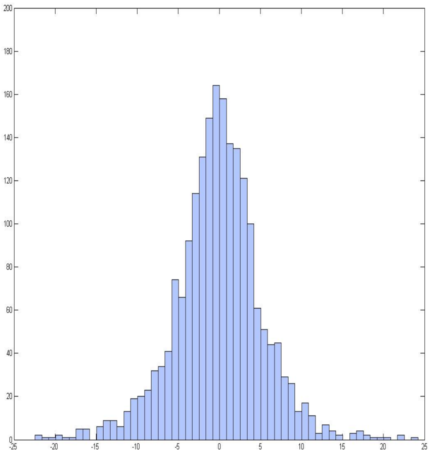

Figure 13: Histogram of $\hat{\delta}_{stu}$, $\rho = 0.8, n = 100$

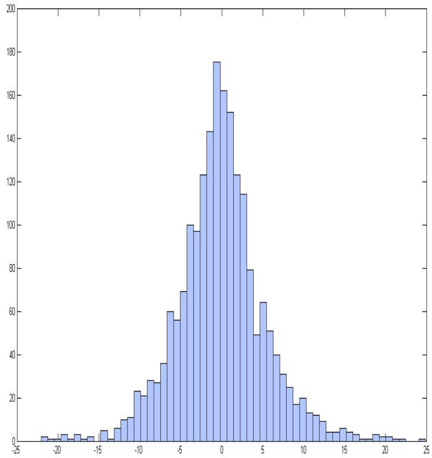

Figure 14: Histogram of $\hat{\delta}_{stu}$, $\rho = 0.8, n = 200$

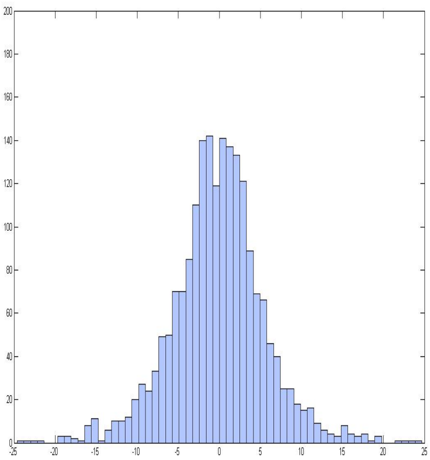

Figure 15: Histogram of $\hat{\delta}_{stu}$, $\rho = 0.8, n = 400$

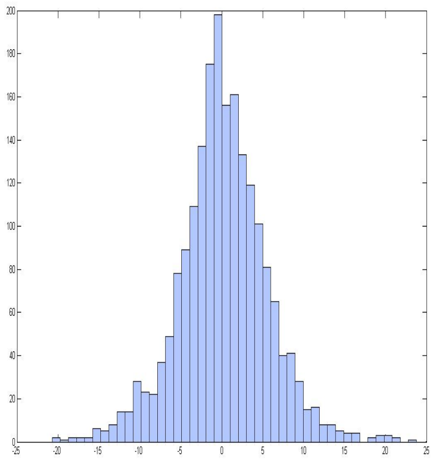

Figure 16: Histogram of $\hat{\delta}_{stu}$, $\rho = 0.8, n = 1200$

# 7 Empirical examples

This section presents two illustrative empirical examples in which the series estimation and studentization method of this paper are applied. The examples are from Yatchew (2003) and are analysed by fitting the partly linear specification of

$$Y_i = Z_i'\delta_0 + h_0(X_i) + U_i.$$

The series estimation yields similar values of $\hat{\delta}$ to the kernel estimates reported 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 assumption is contrasted with that based on the test statistic $t_n^* := n(\hat{\theta} - r)'\hat{C}_n^{-1}(\hat{\theta} - r), r = 0$ of Theorem 6 of this paper, which allows for spatial dependence.

The first example involves a hedonic pricing of housing attributes. The data consists of a relatively small sample 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 regressors contain various attributes of the house including the 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*). In the nonparametric function enter two location coordinates denoted $s$ and $w$ (south and west) of the house:

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

The first set of columns of Table 6 recalls the results of kernel semiparametric estimation reported in Yatchew (2003) based on the work of Robinson (1988). The second set of columns reports the corresponding results from series estimation. The estimates of coefficients, their standard errors and the t-statistics are broadly similar, reporting significance of many of the regressors at the 5% level. However, the test statistic $t_n^*$ of this work, which accounts for dependence, reports that the presence of fire place and luxury bathroom are significant at the 5% significance level and square footage, presence of fire place, luxury bathroom, and garage at 10% level, which may be more informative, bearing in mind the small sample size of 92.

The contrasting conclusions on the significance of $\delta$-coefficients between the t-test under independent errors and test $t_n^*$ allowing for dependence may be due to a presence

| Variable | **kernel** Coef | SE | t-stat | **series** 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 |
| *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 | 14.57 |
| *usespc* | 24.7 | 10.6 | 2.33* | 23.6 | 11.6 | 2.04* | 37.67 |

Table 6: Hedonic House Pricing. Series estimation based on (1,s,w,sw) as basis functions. SE reports standard error under assumption of independence, $TS^*$ is test statistic $t_n^*$ of Section 5 with critical values $46.39, 28.88$ at sizes 5% and 10%, respectively. Test statistics with * are significant at 5% level.

of cross-sectional dependence in the data. This seems to be natural as the dependent variable is price of houses of the same type, sold in the same year and city, which would have been subject to an overlapping set of demand and supply side factors, driven by the same macroeconomic fundamentals.

The second example concerns an analysis of household gasoline demand using a Canadian survey data, see also Yatchew and No (2001). The data are from the "National private vehicle use survey" between October 1994 and September 1996 and consist of 6490 households. The following specification is given in Yatchew (2003):

$$\begin{aligned} dist &= h(price) + \delta_1 income + \delta_2 drivers + \delta_3 hhsize + \delta_4 youngsingle + \delta_5 age \\ &\quad + \delta_6 retire + \delta_7 urban + monthlydummies + u, \end{aligned}$$

where $dist$ is the log distance traveled per month by the household, $price$ is the log price of a litre of gasoline, $drivers$ is the log number of licensed drivers in the household, $hhsize$ is the log size of the household, $youngsingle$ is a dummy for singles up to age of 35, $age$ is the log age of the head, $retire$ is a dummy for the head over the age of 65 and $urban$ is the dummy for urban dwellers. Table 7 reports the results of kernel estimation from Yatchew (2003) and the corresponding series estimation outcomes. The coefficient estimates and test statistics are broadly similar between the kernel and series estimation methods, with all coefficients apart from that of $age$ being highly significant, and t-statistics of $income, drivers, retire$ and $urban$ are indeed very large. The test $t_n^*$ that takes into account spatial dependence in the data still report very high significance of those four variables, while $hhsize$ is no longer significant at 10% level. This suggests that once the effect of the number of drivers in a household has been accounted for, the

| Variable | kernel Coef | SE | t-stat | series Coef | SE | t-stat | $TS^*$ |
|---|---|---|---|---|---|---|---|
| *income* | 0.3 | 0.02 | 15* | 0.286 | 0.0196 | 14.57* | 3202.3* |
| *drivers* | 0.565 | 0.033 | 17.121* | 0.558 | 0.031 | 17.993* | 1925.8* |
| *hhsize* | 0.094 | 0.026 | 3.615* | 0.098 | 0.0248 | 3.953* | 17.8 |
| *youngsingle* | 0.198 | 0.061 | 3.246* | 0.165 | 0.056 | 2.939* | 82.8* |
| *age* | -0.075 | 0.044 | -1.704 | -0.071 | 0.0408 | -1.745 | 45.4 |
| *retire* | -0.198 | 0.032 | -6.188* | -0.195 | 0.0292 | -6.682* | 271.9* |
| *urban* | -0.325 | 0.02 | -16.25* | -0.286 | 0.0164 | -17.461* | 2454.4* |

Table 7: Household Gasoline Demand. Series estimation based on (1, cos(x), cos(2x)) as basis functions. SE reports standard error under assumption of independence, $TS^*$ is test statistic $t_n^*$ of Section 5 with critical values $46.39, 28.88$ at sizes 5% and 10%, respectively. Test statistics with * are significant at 5% level. For the possible endogeneity of *price*, see Yatchew and No (2001, pp.1706-1707).

size of the household does not affect the distance traveled by the household.

## 8 Conclusion

This paper has established the theoretical background for the series estimation of a vector-valued functional of the nonparametric regression function under cross-sectional dependence and nonstationarity. A uniform rate of consistency, asymptotic normality and sufficient conditions for the $\sqrt{n}$ rate of convergence were provided. Importantly, a data-driven studentization method that dispenses with the need for additional distance measures was introduced for the $\sqrt{n}$-consistent semiparametric estimates. A Monte Carlo study reports a highly satisfactory finite sample performance of the proposed studentization. The problem of inference for nonparametric or semiparametric estimates that do not achieve the $\sqrt{n}$ rate of convergence remains open and calls for further research.

The framework of cross-sectional dependence and non-stationarity of this paper and its asymptotic arguments, e.g. application of the FCLT, may be used to establish asymptotic theory for other estimation methods under the cross-sectional setting. The robust inference offered by the studentization of this paper provides a new tool for inference with cross-sectional data and needs to be extended to other commonly used methods such as GMM estimation of parametric models.

# A Appendix A. Proofs of Theorems 1-5.

The main matrix norm used in this work is *spectral* norm, $\|A\|^2 := \bar{\lambda}(A'A)$, defined as the largest eigenvalue of the matrix $A'A$. It is submultiplicative, i.e. $\|AB\| \le \|A\|\|B\|$, and when $A$ is positive semi-definite and symmetric, it satisfies $\|A^{1/2}\|^2 = \|A\|$ and $\|A^{-1}\|^{-1} = \|A\|$. When $A$ is positive semi-definite, symmetric and random, one has that

$$\|A\| = O_p(E\|A\|) = O_p\big(E(\bar{\lambda}(A))\big) \le O_p\big(E(tr(A))\big) = O_p\big(tr(E(A))\big).$$

In addition, three other matrix norms appear in the proofs. Let $\|\cdot\|_E$ denote Euclidean norm for matrix, $\|\cdot\|_C$ maximum column sum norm and $\|\cdot\|_R$ maximum row sum norm. Let $A = (a_{ij})$ be a $q \times q$ matrix. Then,

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

The following inequalities hold:

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

$$\|AB\|_E \le \|A\|_E\|B\|, \quad \|AB\|_E \le \|A\|_E\|B\|_E.$$

The above facts can be found in Searle (1982), Horn and Johnson (1990) and the appendix of Davies (1973).

**Alternative representations of $\hat{g}$ in $p^K$ and $P$**

In Section 3, we introduced a $K \times 1$ 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 estimator $\hat{g}(\cdot)$ is a projection of the unknown function $g_0(\cdot)$ onto the linear space spanned by $p_1(\cdot), \cdots, p_K(\cdot)$, the estimate $\hat{g}(\cdot)$ is invariant to any nonsingular linear transformation of approximating functions. Hence,

$$\hat{g}(x) = p^K(x)'\hat{\beta} = P(x)'\hat{\gamma}, \tag{A.1}$$

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

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

and $\hat{\gamma} = (\mathbf{P}'\mathbf{P})^{-}\mathbf{P}'Y \in \mathbb{R}^K$, where

$$\mathbf{P} = \mathbf{P}_n = [P(X_1), \cdots, P(X_n)]' \in \mathbb{R}^{n\times K}.$$

To show such invariance, one can use the equality $\mathbf{P} = \mathbf{p}B_K^{-1/2}$ to establish the following relation between $\hat{\gamma}$ and $\hat{\beta}$:

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

Above equality holds, because the fact that $(\mathbf{p}'\mathbf{p})^{-}$ is the Moore-Penrose inverse of $\mathbf{p}'\mathbf{p}$

implies $(\mathbf{P}'\mathbf{P})^- = (B_K^{-1/2}\mathbf{p}'\mathbf{p}B_K^{-1/2})^- = B_K^{1/2}(\mathbf{p}'\mathbf{p})^- B_K^{1/2}$.[1]

Proof of Theorems 1-5 benefits from algebraic convenience of studying the representation $\hat{g}(x) = P(x)'\hat{\gamma}$ instead of $\hat{g}(x) = p^K(x)'\hat{\beta}$. 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

$$p^K(x)'\beta_K = P(x)'\gamma_K, \quad \text{where} \quad \gamma_K = B_K^{1/2}\beta_K.$$

Therefore, Assumption A4 implies

$$|g_0 - P'\gamma_K|_\infty = O(K^{-\alpha}), \quad \text{as} \quad K \to \infty.$$

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

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

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

$$\zeta(k) = \sup_{x\in\mathscr{X}} \|B_k^{-1/2}p^k(x)\| \le \|B_k^{-1/2}\| \sup_{x\in\mathscr{X}} \|p^k(x)\| \le C\xi(k), \tag{A.2}$$

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

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

The bound indicates that assumptions involving the upper bound $\xi(K)$ continue to hold also for $\zeta(K)$. The rest of the proof will be completed using $\hat{g}(x) = P(x)'\hat{\gamma}$. Wherever needed, translation to and from the two alternative representations of $\hat{g}$ given in (A.1) is clarified.

**Proof of Theorem 1.** Let $G := G_n = (g_0(X_1), \cdots, g_0(X_n))' \in \mathbb{R}^n$ and $\hat{Q} := \hat{Q}_n = \mathbf{P}'\mathbf{P}/n \in \mathbb{R}^{K\times K}$. We shall use these notations for the rest of the proof. To study the order of $|\hat{g} - g_0|_\infty$, we decompose the quantity $\hat{g}(x) - g_0(x)$ into the bias and stochastic terms. Let $\gamma_K = B_K^{1/2}\beta_K$ for $\beta_K$ of Assumption A4. Write:

$$\hat{g}(x) - g_0(x) = [P(x)'(\hat{\gamma} - \gamma_K)] + [P(x)'\gamma_K - g_0(x)], \tag{A.3}$$

where $\hat{\gamma} = (\mathbf{P}'\mathbf{P})^-\mathbf{P}'Y = (\hat{Q})^-\mathbf{P}'Y/n$. Recall

$$\Sigma_n = E(\mathbf{P}'UU'\mathbf{P}/n)$$

the $K \times K$ variance-covariance matrix of the vector $\sum_{i=1}^n P(X_i)U_i/\sqrt{n}$. We shall show

[1] Existence and uniqueness of Moore-Penrose inverse were established in Penrose (1955). The four Penrose conditions can be found in Searle (1982), pp. 212.

below that

$$\|\hat{\gamma} - \gamma_K\| = O_p\left(\frac{tr(\Sigma_n)^{1/2}}{n^{1/2}} + K^{-\alpha}\right). \tag{A.4}$$

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

$$\begin{aligned}
|\hat{g} - g_0|_\infty &\leq |P'(\hat{\gamma}_K - \gamma_K)|_\infty + |P'\gamma_K - g_0|_\infty \\
&\leq \zeta(K)\|\hat{\gamma}_K - \gamma_K\| + O(K^{-\alpha}) \\
&= O_p\left(\zeta(K)\left[\frac{tr(\Sigma_n)^{1/2}}{n^{1/2}} + K^{-\alpha}\right]\right).
\end{aligned}$$

Therefore, we obtain the statement of Theorem 1:

$$|\hat{g} - g_0|_\infty = O_p\left(\xi(K)\left[\frac{tr(\Sigma_n)^{1/2}}{n^{1/2}} + K^{-\alpha}\right]\right).$$

*Proof of (A.4).* Observe that the matrix $\hat{Q}$ in $\hat{\gamma} = (\hat{Q})^{-}\mathbf{P}'Y/n$ depends on the sample $(X_1, \cdots, X_n)$ of random variables. Thus invertibility of $\hat{Q}$ for any given sample cannot be taken for granted. Let $1_n := I(\underline{\lambda}(\hat{Q}) \geq a)$ be the indicator function for the smallest eigenvalue of $\hat{Q}$, $\underline{\lambda}(\hat{Q})$, to be greater than some positive number $a < 1$. Then the inverse of $\hat{Q}$ exists when $1_n = 1$. It will be shown that $Pr(1_n = 1) \to 1$ as $n \to \infty$, so that $\hat{Q}^{-1}$ exists with probability tending to 1. First we study the quantity $1_n(\hat{\gamma} - \gamma_K)$, subsequently used to get the required result. Decompose $1_n(\hat{\gamma} - \gamma_K)$ as follows:

$$1_n(\hat{\gamma} - \gamma_K) = 1_n\left[\hat{Q}^{-1}\mathbf{P}'(Y - G)/n + \hat{Q}^{-1}\mathbf{P}'(G - \mathbf{P}\gamma_K)/n\right]. \tag{A.5}$$

Applying triangle inequality to (A.5) and the property $\|AB\| \leq \|A\|\|B\|$ of the spectral norm gives

$$\begin{aligned}
\|1_n(\hat{\gamma} - \gamma_K)\| &\leq \|1_n\hat{Q}^{-1}\mathbf{P}'U/n\| + \|1_n\hat{Q}^{-1}\mathbf{P}'(G - \mathbf{P}\gamma_K)/n\| \\
&\leq \|1_n\hat{Q}^{-1}\|\|\mathbf{P}'U/n\| + \|1_n\hat{Q}^{-1}\mathbf{P}'/\sqrt{n}\|\|(G - \mathbf{P}\gamma_K)/\sqrt{n}\|. \quad (A.6)
\end{aligned}$$

Below we shall prove that

$$\begin{aligned}
\|1_n\hat{Q}^{-1}\mathbf{P}'/\sqrt{n}\| &= O_p(1), & (A.7)\\
\|\mathbf{P}'U/n\| &= O_p\left(\frac{tr(\Sigma_n)^{1/2}}{\sqrt{n}}\right), & (A.8)\\
\|(G - \mathbf{P}\gamma_K)/\sqrt{n}\| &= O_p(K^{-\alpha}). & (A.9)
\end{aligned}$$

These lead to

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

which gives $\|\hat{\gamma}_K - \gamma_K\| = O_p\left(tr(\Sigma_n)^{1/2}/n^{1/2} + K^{-\alpha}\right)$. To see this, use the fact that

$1 - 1_n = o_p(1)$ and the triangle inequality, to obtain

$$\begin{aligned} \|\hat{\gamma} - \gamma_K\| &\leq \|1_n(\hat{\gamma} - \gamma_K)\| + \|(1 - 1_n)(\hat{\gamma} - \gamma_K)\| \\ &\leq \|1_n(\hat{\gamma} - \gamma_K)\| + o_p(1)\|\hat{\gamma} - \gamma_K\|. \end{aligned} \tag{A.10}$$

Thus

$$\begin{aligned} &\|\hat{\gamma} - \gamma_K\|(1 + o_p(1)) \leq \|1_n(\hat{\gamma} - \gamma_K)\|, \\ &\|\hat{\gamma} - \gamma_K\| \leq \|1_n(\hat{\gamma} - \gamma_K)\|/(1 + o_p(1)) = O_p\left(tr(\Sigma_n)^{1/2}/n^{1/2} + K^{-\alpha}\right). \end{aligned} \tag{A.11}$$

*Proof of* $1_n \to_p 1$. It suffices to show that $\underline{\lambda}(\hat{Q}) \to_p 1$, as $n \to \infty$ .

First we derive $tr\left\{(\hat{Q} - I)^2\right\} = o_p(1)$. Recall the definition $P(x) := B_K^{-1/2}p^K(x) = [P_{1K}(x), ..., P_{KK}(x)]$. Observe that

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

noting that $E(\hat{Q}) = I$: $E(\hat{Q}) = n^{-1}\sum_{i=1}^{n} E(P(X_i)P'(X_i))$ where $E(P(X_i)P'(X_i)) = B_K^{-1/2}E[p^K(X_i)p^K(X_i)']B_K^{-1/2} = B_K^{-1/2}B_K B_K^{-1/2} = I$. For any pair $p, \ell = 1, \cdots, k$,

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

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

$$\begin{aligned} E\left[tr\left\{(\hat{Q}-I)^2\right\}\right] &= \sum_{p,\ell=1}^{K}(V_{n,1}^{(p,\ell)}+V_{n,2}^{(p,\ell)}) \\ &\leq \frac{K^2\zeta^4(K)}{n}+\frac{K^2\zeta^4(K)\triangle_n}{n^2} \\ &= K^2\zeta^4(K)\left(\frac{1}{n}+\frac{\triangle_n}{n^2}\right)=o(1), \end{aligned} \tag{A.12}$$

by Assumptions A3(ii), and A5.

Hence to show $\underline{\lambda}(\hat{Q}) \to_p 1$, it suffices 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 $\hat{Q}-I=C(\hat{\Lambda}-I)C'$, where $C=(c_{ij})\in\mathbb{R}^{K\times K}$ is 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 columns of $C$ are orthonormal. Therefore,

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

as was concluded in (A.12). This completes the proof of $Pr(1_n=1)\to 1$ as $n\to\infty$.

Now we prove (A.7) -(A.9).

*Proof of (A.7).* Note that $\hat{Q}$ is symmetric and nonnegative definite. Thus, by the properties of the spectral norm,

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

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 theorem, $\|1_n\hat{Q}^{-1}\|=O_p(1)$. Therefore,

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

*Proof of (A.8).* One has

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

*Proof of (A.9).* We have,

$$\|(G-\mathbf{P}\gamma_K)/\sqrt{n}\|^2=(G-\mathbf{P}\gamma_K)'(G-\mathbf{P}\gamma_K)/n=\frac{1}{n}\sum_{i=1}^{n}(g(X_i)-P(X_i)\gamma_K)^2=O_p(K^{-2\alpha}),$$

by Assumption 4, which completes the proof of (A.4) and of the theorem. ■

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

$$\sqrt{n}\bar V_n^{-1/2} r_n = o_p(1), \tag{A.13}$$
$$\sqrt{n}\bar V_n^{-1/2} T_n \to_d N(0, I_d), \tag{A.14}$$

which implies convergence (4.1) of Theorem 2.

*Proof of (A.13).* Again, let $1_n = I(\underline{\lambda}(\hat Q) \geq a)$ for some positive number $a < 1$ as in the proof of Theorem 1, hence $1_n = 1 + o_p(1)$. By the same argument as in proof of Theorem 1, (A.13) follows if we show that

$$1_n\sqrt{n}\bar V_n^{-1/2} r_n = o_p(1). \tag{A.15}$$

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 $\bar g = P'\gamma_K$. Write

$$r_n = \hat\theta_n - \theta_0 - T_n = \{a(\hat g) - a(g_0) - D(\hat g) + D(g_0)\} + \{D(\hat g) - D(\bar g) - T_n\} + \{D(\bar g) - D(g_0)\}.$$

Then

$$\begin{aligned}
\|r_n\| &\leq \|a(\hat g) - a(g_0) - D(\hat g) + D(g_0)\| \\
&+ \|D(\hat g) - D(\bar g) - T_n\| + \|D(\bar g) - D(g_0)\| \\
&=: \|r_{n,1}\| + \|r_{n,2}\| + \|r_{n,3}\|.
\end{aligned}$$

To show (A.13), note that by assumption of the theorem, $\|\bar V_n^{-1/2}\| = \|\bar V_n^{-1}\|^{1/2} = O_p(1)$. Thus, it suffices to prove that

$$1_n\sqrt{n}\|r_{n,i}\| = o_p(1), \quad i = 1, 2, 3. \tag{A.16}$$

For $i = 1$, by Assumption B1, $\|r_{n,1}\| = O_p(|\hat g - g_0|_\infty^2)$. Thus by Theorem 1 and Assumption B3(i), (iii)

$$\sqrt{n}\|r_{n,1}\| = O_p\left(\sqrt{n}\zeta(K)^2\big(\frac{tr(\Sigma_n)}{n} + K^{-2\alpha}\big)\right) = o_p(1). \tag{A.17}$$

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

$$D(\hat g) = D(P'\hat\gamma) = A'\hat\gamma = A'\hat Q^{-}\mathbf{P}'(G+U)/n, \tag{A.18}$$
$$D(\bar g) = D(P'\gamma_K) = A'\gamma_K. \tag{A.19}$$

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

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

Note that $\|A\|^2 \leq \zeta^2(K)$, $\|1_n\hat{Q}^{-1}\| = O_p(1)$, and by (A.7)- (A.9),

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

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

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

To bound $\|I - \hat{Q}\|$ note that $E[\|\hat{Q} - I\|^2] = E\left[\bar{\lambda}\left\{(\hat{Q} - I)^2\right\}\right] \leq E\left[tr\left\{(\hat{Q} - I)^2\right\}\right]$. From (A.12),

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

by Assumptions B3(ii) and (iii).

For $i = 3$, by linearity of $D(\cdot)$ and Assumption B2 and A4, $\|r_{n,3}\| = O(|\bar{g} - g_0|_\infty) = O(K^{-\alpha})$,

$$\sqrt{n}\|r_{n,3}\| = O_p(\sqrt{n}K^{-\alpha}) = o_p(1), \tag{A.20}$$

by Assumptions B3(iii), which implies $nK^{-2\alpha} = o(1)$.

*Proof of (A.14).* To show asymptotic normality of the main term $\sqrt{n}\bar{V}_n^{-1/2}T_n$, introduce the following 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

$$w_{jn} := \sum_{i=1}^n \bar{V}_n^{-1/2}A'P(X_i)\sigma(X_i)b_{ij}/\sqrt{n}. \tag{A.21}$$

Noting that $w_{jn}$ is a function of $\{X_i\}_{i=1}^n$, we show asymptotic normality conditional on $\|\bar{V}_n^{-1}\| \leq C$ 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 $\{X_i\}_{i=1}^n$.

This yields the required unconditional asymptotic normality result of Theorem 2. Such line of reasoning was used in Robinson (2009).

By Cramer-Wold device, to derive asymptotic normality of the vector $\sqrt{n}\bar{V}_n^{-1/2}T_n$, we focus on a scalar summation $\sum_{j=1}^{\infty} c' w_{jn}\varepsilon_j$ with any fixed vector $c \in \mathbb{R}^d$ such that $c'c = 1$. Consider splitting $\sqrt{n}c'\bar{V}_n^{-1/2}T_n$ into two sums,

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

where the integer $N(n)$ is chosen to be the smallest satisfying

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

The choice of $N(n)$ is deterministic once we condition on $\{X_i\}_{i=1}^n$. The purpose of this truncation is to make the contribution from the second summation negligible:

$$\begin{aligned}\big(\sum_{j=N(n)+1}^{\infty} c' w_{jn}\varepsilon_j\big)^2 &= O_p\big(E\big(\sum_{j=N(n)+1}^{\infty} c' w_{jn}\varepsilon_j\big)^2\big) \\ &= O_p\big(\sum_{j=N(n)+1}^{\infty} (c' w_{jn})^2\big) = O_p\left(\frac{1}{\log n}\right) = o_p(1).\end{aligned}$$

Since $\{c' w_j \varepsilon_j\}$ are martingale differences under assumption A2, asymptotic normality of the first summation is established by verifying the following two sufficient conditions for asymptotic normality from Scott (1973), adapted for our setting.

$$\sum_{j=1}^{N(n)} E\big((c' w_j \varepsilon_j)^2\big) \to_p 1, \tag{A.22}$$

$$\sum_{j=1}^{N(n)} E\big((c' w_{jn}\varepsilon_j)^2 1(|c' w_{jn}\varepsilon_j| > \delta)\big) \to_p 0, \quad \forall \delta > 0. \tag{A.23}$$

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 $\nu$ be as in Assumption A2. Then,

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

The first inequality follows from $1(|c'w_{jn}\varepsilon_j| > \delta) \leq (|c'w_{jn}\varepsilon_j|/\delta)^\nu$. With $\sum_{j=1}^{N(n)} (c'w_{jn})^2 \to 1$, (A.23) is verified once we show that $\max_{j\geq 1} |c'w_{jn}|^\nu \to 0$. Conditionally on $X_1, \cdots, X_n$, the following holds for any $j \geq 1$:

$$\begin{aligned} |c'w_{jn}| &= \left| \frac{c'}{\sqrt{n}} \bar{V}_n^{-1/2} \sum_{i=1}^n A'P(X_i)\sigma(X_i)b_{ij} \right| \\ &\leq \|c\| \|\bar{V}_n^{-1/2}\| \frac{1}{\sqrt{n}} \max_{1\leq j\leq n} \sum_{i=1}^n |b_{ij}| \|A'P(X_i)\sigma(X_i)\| \\ &= O\left( \frac{\zeta(K)^2}{\sqrt{n}} \max_{1\leq j\leq n} \sum_{i=1}^n |b_{ij}| \right) = o(1), \end{aligned} \tag{A.24}$$

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

**Proof of Theorem 3.** We will prove later that $\|\bar{V}_n - V_n\| = o_p(1)$. Then, $V_n^{-1}\bar{V}_n \to_p I$ since $\|V_n^{-1}\bar{V}_n - I\| \leq \|V_n^{-1}\| \|\bar{V}_n - V_n\| = o_p(1)$, which in turn gives $V_n\bar{V}_n^{-1} \to_p I$. It follows that

$$\|\bar{V}_n^{-1}\| \leq \|V_n^{-1}\| \|V_n\bar{V}_n^{-1}\| = O_p(1).$$

Now, to show the final statement, (4.6), of the 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:

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

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* $\|\bar{V}_n - V_n\| = o_p(1)$. By definition of the spectral norm, $\|\bar{V}_n - V_n\| = o_p(1)$ 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 $B$. Then, using notation (4.2),

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

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

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

We shall show that

$$Var(S_{k,n}) = o(1), \quad k = 1, 2, 3, \tag{A.25}$$

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

*Proof of (A.25), k=1.* We have

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

Introduce the notation, $\phi_{ij}^{(\ell,p)} := Cov(\bar{h}_i^{(\ell)}, \bar{h}_j^{(p)})$ and denote by $\Phi^{(\ell,p)}$ the $n \times n$ matrix whose $(i,j)^{th}$ element is $\phi_{ij}^{(\ell,p)}$. Recall that by the Definition 2 of joint $4^{th}$ order cumulant,

$$Cov(Z_1Z_2, Z_3Z_4) = \kappa(Z_1, Z_2, Z_3, Z_4) + Cov(Z_1, Z_3)Cov(Z_2, Z_4) + Cov(Z_1, Z_4)Cov(Z_2, Z_3).$$

One has

$$\begin{aligned}
Var(S_{1,n}) &= \frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n} \gamma_{i_1i_2}\gamma_{i_3i_4}\kappa(\bar{h}_{i_1}^{(\ell)}, \bar{h}_{i_2}^{(p)}, \bar{h}_{i_3}^{(\ell)}, \bar{h}_{i_4}^{(p)}) && \text{(A.26)} \\
&+ \frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n} \gamma_{i_1i_2}\gamma_{i_3i_4}\phi_{i_1i_3}^{(\ell,\ell)}\phi_{i_2i_4}^{(p,p)} && \text{(A.27)} \\
&+ \frac{1}{n^2}\sum_{i_1,i_2,i_3,i_4=1}^{n} \gamma_{i_1i_2}\gamma_{i_3i_4}\phi_{i_1i_4}^{(\ell,p)}\phi_{i_2i_3}^{(p,\ell)}. && \text{(A.28)}
\end{aligned}$$

Denote by $\Gamma = \Gamma_n$ the $n \times n$ matrix whose $(i,j)^{th}$ element is $\gamma_{ij}$. Firstly, by Assumption

B7, the RHS of (A.26) is $o(1)$. To bound (A.27) and (A.28), write

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

By the properties of matrix norms given earlier, we see that

$$\left|tr\left(\Gamma\Phi^{(p,p)}\Gamma\Phi^{(\ell,\ell)}\right)\right| \leq \|\Gamma\Phi^{(p,p)}\|_E\|\Gamma\Phi^{(\ell,\ell)}\|_E \leq \|\Gamma\|^2\|\Phi^{(p,p)}\|_E\|\Phi^{(\ell,\ell)}\|_E. \tag{A.29}$$

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

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

It is clear that $\int|f_{ij}(x,y) - f(x)f(y)|dxdy \leq 2$ for all $i$ and $j$. Hence,

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

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

$$\|\Phi^{(p,p)}\|_E^2 = \sum_{i,j=1}^{n}(\phi_{ij}^{(p,p)})^2 \leq C\zeta^8(K)(n+\triangle_n). \tag{A.30}$$

Hence, by (A.29) and Assumption B6,

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

since by the property of spectral norm $\|A\|^2 \leq \|A\|_C\|A\|_R$, and by the symmetry of $\Gamma$,

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

Similarly, it follows that $n^{-2}tr\left(\Gamma\Phi^{(p,\ell)}\Gamma\Phi^{(p,\ell)}\right) = o(1)$, which completes the proof of (A.25) when $k=1$.

*Proof of (A.25), k=2,3.* Recall, $S_{n,2} = n^{-2} \sum_{i,j=1}^{n} \gamma_{ij} \bar{h}_j^{(p)} E(h_i^{(\ell)})$. Therefore,

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

using the bound $E|h_i^{(\ell)}| \leq C\zeta^2(K)$. By the same steps taken in two lines prior to (A.30),

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

This, together with Assumption B6 yields

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

■

**Proof of Theorem 4.** We need to show $\|V_n - V\| = o(1)$, as $n \to \infty$. 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 elements $(V_n - W_n)_{\ell,p}$, $1 \leq \ell, p \leq d$, of $V_n - W_n$ converges to zero. We have that

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

Notice that

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

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

$$
E[v_{pK}^2(X_i)] \le 2E[\{v_{pK}(X_i) - w_p(X_i)\}^2] + 2E[w_p^2(X_i)] < \infty. \tag{A.31}
$$

Hence,

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

by Assumption C2 (ii). This completes the proof of the Theorem. ■

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

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

where $[rn]$ denotes the integer part of $rn$. Based on the statement of 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 $\hat{C}_n = \frac{1}{n}\sum_{m=1}^{n} S_n^*(\frac{m}{n})S_n^*(\frac{m}{n})' \sim \int_0^1 S_n^*(r)S_n^*(r)'dr$. Therefore, continuous mapping theorem gives

$$
V^{-1/2}\hat{C}_nV^{-1/2} \Rightarrow \Psi_d. \tag{A.32}
$$

Write

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

By Lemma 1-3, $\hat{C}_n^{-1/2}V^{1/2} \Rightarrow \Psi_d^{-1/2}$, and by Theorem 4, $\sqrt{n}V^{-1/2}(\hat{\theta}_n - \theta_0) \to_d N(0, I_d)$, where convergence of the two terms is joint, completing the proof. ■

# B Appendix B. Lemmas 1-3. Propositions 1-2.

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. Skorohod metric $d(\cdot,\cdot)$ in $D[0,1]$ is given by:

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

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

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

Denote by

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

the $d\times 1$ vector-valued summations.

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 $X(\cdot) = (X_1(\cdot),\cdots,X_d(\cdot))' \in D[0,1]^d$ and $Y(\cdot) = (Y_1(\cdot),\cdots,Y_d(\cdot))' \in D[0,1]^d$, define the metric:

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

Lemma 1 states functional central limit theorem (FCLT) for $S_n(r)$ in $D[0,1]^d$ equipped with the metric $d'(\cdot,\cdot)$. The notation $\Rightarrow_{D[0,1]^d}$ signifies weak convergence of the associated probability measures in $D[0,1]^d$.

In the following proofs we will need some notations. For $j\ge 1$, introduce a $j\times K$ random matrix $\mathbf{P}_j = (P(X_1),\cdots,P(X_j))'$ and $j\times 1$ random vectors, $G_j = (g_0(X_1),\cdots,g_0(X_j))'$ and $\hat{G}_j = (\hat{g}(X_1),\cdots,\hat{g}(X_j))'$.

**Lemma 1.** Under Assumptions of Theorem 5,

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

**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) had established two sufficient conditions for weak convergence of probability measures in this multi-dimensional product space. These two conditions, adapted here for (B.1), 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 $S_n(\cdot)$.

We first establish the following two statements, which will be subsequently used to obtain the above two facts: for any $0\le r\le u\le 1$,

$$ES_n(r)S_n(u)' \to r\cdot V, \tag{B.2}$$

$$E|S_{n\ell}(u)-S_{n\ell}(r)|^2 \le C\Big|\frac{[un]-[rn]}{n}\Big|, \quad \ell=1,\cdots,d \tag{B.3}$$

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, $E(S_nS_n') = V_n \to V$. Therefore,

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

Hence (B.2) follows if we show that $E\big(S_n(r)(S_n(u)' - S_n(r)')\big) \to 0$. This is done by showing that the corresponding limit of each element of the vector is zero. For $\ell, p = 1, \cdots, d$,

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

by Assumption C3 (i), and because

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

as shown in the proof of Theorem 4. This completes the proof of (B.2).

To prove (B.3), observe that

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

by Assumption C3 (ii), which proves (B.3).

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

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

Using Cramer-Wold device, it suffices to show that for any $d \times 1$ vectors $c_1', \cdots, c_k'$, the scalar random variable

$$Q_n := \sum_{l=1}^{k} c_l' S_n(r_l) \to_d \sum_{l=1}^{k} c_l' V^{1/2} W_d(r_l) =: Q. \tag{B.4}$$

Write $S_n(r) = \sum_{j=1}^{\infty} w_{j,[rn]}\varepsilon_j$ with $w_{j,[rn]}$ as in (A.21), with $\bar{V}_n$ replaced by $V$. Then,

$Q_n = \sum_{j=1}^{\infty} w_{jn}^* \varepsilon_j$ with $w_{jn}^* = \sum_{l=1}^{k} c_l' w_{j,[r_l n]}$. Then by (B.2),

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

By (A.24), 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 convergence (B.4) follows by the same argument as in the proof of asymptotic normality (A.14).

Finally, we establish tightness for individual component of the vector $S_n(r)$, which completes the proof of the lemma. Noting $S_{n\ell}(\cdot) \in D[0,1]$, $\ell = 1, \cdots, d$, 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 $C > 0$,

$$E[|S_{n\ell}(s) - S_{n\ell}(r)|^{2\beta} |S_{n\ell}(t) - S_{n\ell}(s)|^{2\beta}] \leq C\big|t - r\big|^{2\alpha}, \quad \ell = 1, \cdots, d. \tag{B.5}$$

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

$$E|S_{n\ell}(u) - S_{n\ell}(r)|^4 \leq C\Big|\frac{[un] - [rn]}{n}\Big|^2. \tag{B.6}$$

To see (B.6) implies (B.5), note that for $\beta = 1$, the LHS of (B.5) is

$$\begin{aligned} E[|S_{n\ell}(s) - S_{n\ell}(r)|^2 |S_{n\ell}(t) - S_{n\ell}(s)|^2] &\leq \big\{E[|S_{n\ell}(s) - S_{n\ell}(r)|^4] E[|S_{n\ell}(t) - S_{n\ell}(s)|^4]\big\}^{1/2} \\ &\leq C\Big(\Big|\frac{[sn] - [rn]}{n}\Big|^2 \Big|\frac{[tn] - [sn]}{n}\Big|^2\Big)^{1/2} \\ &= C\Big|\frac{[sn] - [rn]}{n}\Big| \Big|\frac{[tn] - [sn]}{n}\Big| \\ &\leq C\Big|\frac{[tn] - [rn]}{n}\Big|^2, \end{aligned} \tag{B.7}$$

where the first step uses the Cauchy-Schwarz inequality, the second inequality follows from (B.6) 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 (B.7) implies (B.5) 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 LHS's of (B.5) and (B.7) vanish, and thus (B.5) holds.

To verify (B.6), denote by $e_\ell$, a $d$-dimensional vector, whose $\ell^{th}$ element is 1 and the 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 LHS of (B.6) with the new notation and noting $E(\varepsilon_j^4) = \kappa < \infty$, $\forall j$ by

Assumption C5, we obtain

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

where the last step follows from (B.3). This completes the proof of the lemma. ■

**Lemma 2.** Under Assumptions of Theorem 5,

$$\Big(\hat{S}_n(r)\Big)_{0\leq r\leq 1} \Rightarrow_{D[0,1]^d} \big(V^{1/2}\{W_d(r)-rW_d(1)\}\big)_{0\leq r\leq 1}. \tag{B.8}$$

**Proof of Lemma 2.** Since $\hat{U}_i - U_i = g_0(X_i) - \hat{g}(X_i)$,

$$L_n(r) := \hat{S}_n(r) - S_n(r) = \sum_{i=1}^{[rn]} A'P(X_i)\{g_0(X_i)-\hat{g}(X_i)\}/\sqrt{n}.$$

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

$$\begin{aligned}
L_n(r) &= \sum_{i=1}^{[rn]} A'P(X_i)\{g_0(X_i)-P'(X_i)\gamma_K\}/\sqrt{n} + \sum_{i=1}^{[rn]} A'P(X_i)P'(X_i)(\gamma_K-\hat{\gamma})/\sqrt{n} \\
&= A'\mathbf{P}'_{[rn]}(G_{[rn]}-\mathbf{P}_{[rn]}\gamma_K)/\sqrt{n} + A'\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}(\gamma_K-\hat{\gamma})/\sqrt{n},
\end{aligned}$$

leading to

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

We shall show that

$$\sup_{r\in[0,1]}\|a_n(r)\| = o_p(1), \tag{B.9}$$

$$\ell_n(r) \Rightarrow_{D[0,1]^d} rV^{1/2}W_d(1), \tag{B.10}$$

which, together with Lemma 1, prove (B.8).

*Proof of (B.9).* One has

$$\begin{aligned}
\sup_{r\in[0,1]}\|a_n(r)\| &\leq \|A'\|\sup_{r\in[0,1]}\|\mathbf{P}'_{[rn]}\|\sup_{r\in[0,1]}\|(G_{[rn]}-\mathbf{P}_{[rn]}\gamma_K)/\sqrt{n}\| && \text{(B.11)} \\
&= O_p(\sqrt{n}\xi^2(K)K^{-\alpha}) && \text{(B.12)}
\end{aligned}$$

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

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

by Assumption A4. Then (B.9) follows by Assumption C4.

*Proof of (B.10).* Recall $Y = G + U$, and one has $\hat{\gamma} = (\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'Y = (\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'(G - \mathbf{P}\gamma_K) + (\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'(\mathbf{P}\gamma_K + U)$. Therefore,

$$\sqrt{n}(\hat{\gamma} - \gamma_K) = \left(\frac{\mathbf{P}'\mathbf{P}}{n}\right)^{-1} \frac{\mathbf{P}'(G - \mathbf{P}\gamma_K)}{\sqrt{n}} + \left(\frac{\mathbf{P}'\mathbf{P}}{n}\right)^{-1} \frac{\mathbf{P}'U}{\sqrt{n}}.$$

Hence,

$$\begin{aligned} \ell_n(r) &= A'\left(\frac{\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}}{n}\right)\left(\frac{\mathbf{P}'\mathbf{P}}{n}\right)^{-1} \frac{\mathbf{P}'(G - \mathbf{P}\gamma_K)}{\sqrt{n}} \\ &+ A'\left(\frac{\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}}{n}\right)\left(\frac{\mathbf{P}'\mathbf{P}}{n}\right)^{-1} \frac{\mathbf{P}'U}{\sqrt{n}} =: \ell_{1,n}(r) + \ell_{2,n}(r). \end{aligned} \tag{B.13}$$

We shall show the following two results which constitute the proof of (B.10):

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

Noting that $(\mathbf{P}'\mathbf{P}/n)^{-1} = O_p(1)$ and $\sup_{r\in[0,1]} \left\|\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}/n\right\| = O_p(\xi^2(K))$, since $\|\sum_{i=1}^{[rn]} P(X_i)P'(X_i)/n\| \leq \xi^2(K)$, we obtain

$$\begin{aligned} \|\ell_{1,n}(r)\| &\leq \|A'\| \sup_{r\in[0,1]} \left\|\frac{\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}}{n}\right\| \left\|(\frac{\mathbf{P}'\mathbf{P}}{n})^{-1}\right\| \left\|\frac{\mathbf{P}'(G - \mathbf{P}\gamma_K)}{\sqrt{n}}\right\| \\ &\leq \|A'\| O_p(\xi^2(K)) \|\mathbf{P}'\| \left\|\frac{(G - \mathbf{P}\gamma_K)}{\sqrt{n}}\right\| = O_p(\sqrt{n}\xi^3(K)K^{-\alpha}) = o_p(1), \end{aligned}$$

by Assumption C4(iv). Next, write

$$\ell_{2,n}(r) = rA'\mathbf{P}'U/\sqrt{n} + A'\left(\left(\frac{\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}}{n}\right)\left(\frac{\mathbf{P}'\mathbf{P}}{n}\right)^{-1} - rI\right)\frac{\mathbf{P}'U}{\sqrt{n}}.$$

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

$$\sup_{r\in[0,1]} \|A'\left(\left(\frac{\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}}{n}\right)\left(\frac{\mathbf{P}'\mathbf{P}}{n}\right)^{-1} - rI\right)\frac{\mathbf{P}'U}{\sqrt{n}}\| = o_p(1).$$

One has $\|A\| = O(\xi(K))$ and $\left\|\mathbf{P}'U/\sqrt{n}\right\| = O(\sqrt{K})$ by Assumption C4 (ii). Next, we have

$$\begin{aligned}
& \sup_{r\in[0,1]} \left\| (\frac{[rn]}{n})(\frac{\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}}{[rn]})(\frac{\mathbf{P}'\mathbf{P}}{n})^{-1} - rI \right\| \\
\leq & \sup_{r\in[0,1]} \frac{[rn]}{n} \left\| \frac{\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}}{[rn]} - I \right\| \left\| (\frac{\mathbf{P}'\mathbf{P}}{n})^{-1} - I \right\| \\
+ & \sup_{r\in[0,1]} \frac{[rn]}{n} \left\| \frac{\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}}{[rn]} - I \right\| + \sup_{r\in[0,1]} \frac{[rn]}{n} \left\| (\frac{\mathbf{P}'\mathbf{P}}{n})^{-1} - I \right\| + o(1/n).
\end{aligned}$$

From the proof of Theorem 1, (A.12), we have

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

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

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

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

$$\begin{aligned}
\sup_{r\in[0,1]} (\frac{[rn]}{n})^2 \left\| \frac{\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}}{[rn]} - I \right\|^2 &= \sup_{r\in[0,1]} (\frac{[rn]}{n})^2 O_p\left( K^2\xi^4(K)(\frac{1}{[rn]} + \frac{\triangle_{[rn]}}{[rn]^2}) \right) \\
= \sup_{r\in[0,1]} \frac{[rn]}{n} O_p\left( K^2\xi^4(K)(\frac{1}{n} + \frac{\triangle_{[rn]}}{n[rn]}) \right) &= O_p\left( K^2\xi^4(K)/n \right), \qquad \text{(B.14)}
\end{aligned}$$

by Assumption A4 (i). Therefore,

$$\begin{aligned}
& \sup_{r\in[0,1]} \|A' \left( \left( \frac{\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}}{n} \right) \left( \frac{\mathbf{P}'\mathbf{P}}{n} \right)^{-1} - rI \right) \frac{\mathbf{P}'U}{\sqrt{n}} \| \\
& = O_p(K\sqrt{K}\xi^3(K)/\sqrt{n}) = o_p(1), \qquad \text{(B.15)}
\end{aligned}$$

with the last step following from Assumption A3 (ii). This completes the proof of Lemma 2. ■

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

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

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

Therefore,

$$\begin{aligned}
\sup_{r\in[0,1]} \|\hat{S}_n^*(r) - \hat{S}_n(r)\| &\leq \|\hat{A}^{*\prime}\hat{B}_K^{-1}B_K^{1/2} - A^{*\prime}B_K^{-1/2}\| \\
& \quad \cdot \sup_{r\in[0,1]} \|\mathbf{P}'_{[rn]}\hat{U}_{[rn]}/\sqrt{n}\| =: d_{n,1}d_{n,2}. \qquad \text{(B.16)}
\end{aligned}$$

We shall show that

$$d_{n,1} = O_p(K\xi^2(K)/\sqrt{n}) + O_p\big(\xi^2(K)\big(\sqrt{\frac{tr(\Sigma)}{n}} + K^{-\alpha}\big)\big), \tag{B.17}$$

$$d_{n,2} = O_p(K^{1/2} + K^{-\alpha}\sqrt{n}). \tag{B.18}$$

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).

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

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

$$\|\hat{B}_K^{-1}B_K^{1/2} - B_K^{-1/2}\| \leq \|\hat{B}_K^{-1} - B_K^{-1}\|\|B_K^{1/2}\| = O_p\left(K\xi^2(K)\frac{1}{\sqrt{n}}\right), \tag{B.19}$$

since by Assumption C4 (iii), $\|B_K\| = O(1)$, whereas $\|\hat{B}_K - B_K\|^2 = O_p\big(K^2\xi^4(K)/n\big)$ which can be shown using the same argument shown in obtaining an order of magnitude (A.12) for $\|\hat{Q} - I\|$ and applying Assumption C4 (i). Then, $\|\hat{B}_K^{-1} - B_K^{-1}\|^2 = O_p\big(K^2\xi^4(K)/n\big)$ 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 (B.17), it remains to evaluate the term $\|\hat{A}^* - A^*\|$. Newey (1997) showed that the estimate $\hat{A}^* = (\hat{A}_1^*, \cdots, \hat{A}_d^*)$ is equal to the quantity $\big(D(p_1;\hat{g}), \cdots, D(p_K;\hat{g})\big)'$ with probability approaching one. Recalling $D(\cdot;\hat{g}) = \big(D_1(\cdot;\hat{g}), \cdots, D_d(\cdot;\hat{g})\big)$, the $i^{th}$ column of $\hat{A}^* - A^*$ can be written as

$$\hat{A}_i^* - A_i^* = \big(D_i(p_1;\hat{g}) - D_i(p_1;g_0), \cdots, D_i(p_K;\hat{g}) - D_i(p_K;g_0)\big)', \quad i = 1, \cdots, d.$$

Using linearity of $D_i(g;\hat{g})$ in $g$, one writes

$$\begin{aligned} \|\hat{A}_i^* - A_i^*\|^2 &= (\hat{A}_i^* - A_i^*)'(\hat{A}_i^* - A_i^*) = |D_i\big((\hat{A}_i^* - A_i^*)'p^K;\hat{g}\big) - D_i\big((\hat{A}_i^* - A_i^*)'p^K;g_0\big)| \\ &\leq C|(\hat{A}_i^* - A_i^*)'p^K|_\infty|\hat{g} - g_0|_\infty \leq C\|\hat{A}_i^* - A_i^*\|\xi(K)|\hat{g} - g_0|_\infty, \end{aligned}$$

with the first inequality following from Assumption C6. Therefore, $\|\hat{A}_i^* - A_i^*\| = O_p(\xi(K)|\hat{g} - g_0|_\infty)$, for $i = 1, \cdots, p$. This allows us to bound

$$\begin{aligned} \|\hat{A}^* - A^*\|^2 &\leq tr\big((\hat{A}^* - A^*)'(\hat{A}^* - A^*)\big) = \sum_{i=1}^{p}(\hat{A}_i^* - A_i^*)'(\hat{A}_i^* - A_i^*) \\ &= \big(\sum_{i=1}^{p}\|\hat{A}_i^* - A_i^*\|^2\big) \leq C\xi^2(K)|\hat{g} - g_0|_\infty^2. \end{aligned}$$

Therefore, applying for $|\hat{g} - g_0|_\infty$ 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 (B.17).

Next, decompose $d_{n,2}$ as follows.

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

As in the proof of Lemma 2, one can bound

$$d_{n,21} \leq \sup_{r\in[0,1]} \|\mathbf{P}'_{[rn]}(G_{[rn]} - \mathbf{P}_{[rn]}\gamma_K)/\sqrt{n}\| + \sup_{r\in[0,1]} \|\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}/n\|\|\sqrt{n}(\hat{\gamma} - \gamma_K)\|.$$

From (B.12) it is seen that the first term on the RHS is $O_p(\sqrt{n}\xi(K)K^{-\alpha}) = o_p(1)$, by Assumption C4 (iv). By (A.11), $\|(\hat{\gamma} - \gamma_K)\sqrt{n}\| = O_p(tr(\Sigma_n)^{1/2} + K^{-\alpha}\sqrt{n}) = O_p(K^{1/2} + K^{-\alpha}\sqrt{n})$ from Assumption C4 (ii), whereas by (B.14), $\sup_{r\in[0,1]} \|\mathbf{P}'_{[rn]}\mathbf{P}_{[rn]}/[rn]\| = O_p(1) + O_p(K\zeta^2(K)/\sqrt{n}) = O_p(1)$ by Assumption A3 (ii). Thus, $d_{n,21} = O_p(K^{1/2} + K^{-\alpha}\sqrt{n})$.

Finally, one has

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

by Assumption C4(ii). Hence, $d_{n,2} = O_p(K^{1/2})$, which proves (B.18) and completes the proof of the lemma. ■

In the following proposition we provide the upper bound for $\triangle_n$ in (3.3) in the case of Gaussian random variables $X_i$.

**Proposition 1.** Let $X_i \sim N(0,1), i = 1, 2, \cdots$, be Gaussian variables with $\sigma_{ij}^{(X)} = Cov(X_i, X_j)$. If for some $c_0 < 1$, one has $|\sigma_{ik}^{(X)}| \leq c_0, \forall i, k = 1, 2, \cdots; i \neq k$, then,

$$\triangle_n \leq C \sum_{i,k=1,i\neq k}^{n} |\sigma_{ik}^{(X)}|, \quad n \geq 1. \tag{B.20}$$

**Proof of Proposition 1.** Recall that the bivariate density of $X \sim N(0,1), Y \sim N(0,1), Cov(X,Y) = \rho$ is

$$\begin{aligned} \mathrm{f}_\rho(\mathrm{x}, \mathrm{y}) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\big(-\mathrm{m}_\rho(\mathrm{x}, \mathrm{y})\big), \\ m_\rho(x, y) &:= \frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)}, \quad x, y \in \mathbb{R}. \end{aligned}$$

Then $\mathrm{f}_0(\mathrm{x}, \mathrm{y}) = \mathrm{f}(\mathrm{x})\mathrm{f}(\mathrm{y})$, where $f(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. We shall show that for all $|\rho| \leq c_0 < 1$,

$$|\mathrm{f}_\rho(\mathrm{x}, \mathrm{y}) - \mathrm{f}_0(\mathrm{x}, \mathrm{y})| \leq \mathrm{C}\rho \exp\big(\frac{-(\mathrm{x}^2 + \mathrm{y}^2)}{8}\big), \ \mathrm{x}, \mathrm{y} \in \mathbb{R}, \tag{B.21}$$

where $C$ does not depend on $\rho$. Since $f_{ik}(x,y) = \mathrm{f}_{\sigma_{\mathrm{ik}}}(\mathrm{x},\mathrm{y})$ is the bivariate density of $X_i$, $X_k$, the following holds by (B.21),

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

which proves (B.20).

*Proof of (B.21)* By the mean value theorem, applied in $|\rho| \leq c_0$,

$$|\mathrm{f}_\rho(\mathrm{x},\mathrm{y}) - \mathrm{f}_0(\mathrm{x},\mathrm{y})| \leq |\rho| \sup_{|\rho|\leq \mathrm{c}_0} |\mathrm{f}'_\rho(\mathrm{x},\mathrm{y})|. \tag{B.22}$$

Note that

$$\mathrm{f}'_\rho(\mathrm{x},\mathrm{y}) = \mathrm{f}_\rho(\mathrm{x},\mathrm{y}) \left( \frac{\rho}{1-\rho^2} - \frac{\partial \mathrm{m}_\rho(\mathrm{x},\mathrm{y})}{\partial \rho} \right). \tag{B.23}$$

One has

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

We shall show that

$$\mathrm{f}_\rho(\mathrm{x},\mathrm{y}) \leq \mathrm{c}\exp\big(-(\mathrm{x}^2+\mathrm{y}^2)/4\big) \tag{B.24}$$

$$\Big|\frac{\partial m_\rho(x,y)}{\partial \rho}\Big| \leq c(x^2+y^2),\ x,y \in \mathbb{R}, \tag{B.25}$$

where $c$ does not depend on $\rho$ and $x, y$, which together with (B.22) and (B.23) implies (B.21).

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 (B.24):

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

Next,

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

since $2|xy| \le x^2+y^2$, which proves (B.25) and completes the proof of the proposition. ■

**Proposition 2** Assume that there exists $\eta(j) \ge 0$, $j \in \mathbb{Z}$ such that $\sum_{j=-\infty}^{\infty} \eta(j) < \infty$ and $|\gamma_{ikn}| \le \eta(i-k), \forall i,k = 1,2,\cdots$. 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 $\tau_n := \sum_{|j|\ge \log n} \eta(j) \to 0$ as $n \to \infty$, and $\max_j \eta(j) \le C < \infty$. One has

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

This completes the proof of the proposition. ■

## References


Andrews, D.W.K. (1991). Asymptotic normality of series estimators for nonparametric and semiparametric models. *Econometrica*, **59**, 307-345.

Arbia, G. (2006). *Spatial Econometrics: Statistical Foundation and Applications to Regional Analysis.* Springer-Verlag, Berlin.

Bester, C. A., Conley, T. G., Hansen, C. B. and Vogelsang, T. J. (2008). Fixed-b asymptotics for spatially dependent robust nonparametric covariance matrix estimators. *Working paper.*

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

Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory.* Holden Day, Inc., San Francisco.

Chamberlain, G. (1986). Notes on Semiparametric Regression. *Unpublished manuscript*, Department of Economics, Harvard University.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. eds. Heckman. J.J. and Leamer, E.E. In: *Handbook of Econometrics.* Vol. 6B, Chapter 76, North-Holland.

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

Chen, X., Liao, Z. and Sun, Y. (2011). On inference of sieve M-estimation of functionals of semi-nonparametric time series models. *Preprint.*

Chen, X. and Shen, X. (1998). Sieve extreme estimates for weakly dependent data. *Econometrica*, **66**, 289-314.

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

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

Davies, R.B. (1973). Asymptotic inference in stationary Gaussian time-series. *Advances in Applied Probability*, **5**, 469-497.

Fan, Y., and 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.

Hidalgo, J. (1997). Nonparametric estimation with strongly dependent multivariate time series. *Journal of Time Series Analysis*, **18**, 97-122.

Horn, R.A. and Johnson, C.R. (1990). *Matrix Analysis.* Cambridge University Press.

Kelejian, H.H. and 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. and 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. and Prucha, I.R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, **140**, 131-154.

Kiefer, M., Vogelsang T.J. and 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 distributions of quasi-maximum likelihood estimators 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.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703-708.

Penrose, R. (1955). A generalized inverse for matrices. *Proc. Cambridge Phil. Soc.*, **51**, 406-413.

Phillips, P.C.B. and Durlauf, S. (1986). Multiple time series regression with integrated processes. *Review of Economic Studies*, **53**, 473-496.

Pinkse, J., Shen, L. and 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. and Brett, B. (2002). Spatial price competition: a semiparametric approach. *Econometrica*, **70**, 1111-1153.

Powell, J. L., Stock, J. H. and Stocker, T. M. (1989). Semiparametric estimation of the 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. (2009). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics*, forthcoming.

Robinson, P.M. (2010). Efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics*, **157**, 6-17.

Robinson, P.M. and Thawornkaiwong, S. (2010). Statistical inference on regression with spatial dependence. *Journal of Econometrics*, forthcoming.

Rossi, F. (2010). Improved test statistics for pure spatial autoregressive models. *Preprint.*

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 and Sons, Inc., New York.

Yatchew, Y. (2003). *Semiparametric Regression for the Applied Econometrician.* Cambridge University Press.

Yatchew, Y. and No, J. A. (2003). Household gasoline demand in Canada. *Econometrica*, **69**, 1697-1710.