Functionals of order statistics and their multivariate concomitants with application to semiparametric estimation by nearest neighbours

Chu, Ba and Huynh, Kim and Jacho-Chavez, David

Carleton University, Bank of Canada, Emory University

2013

Functionals of Order Statistics and their Multivariate Concomitants with Application to Semiparametric Estimation by Nearest Neighbors

Ba M. Chu[*] Carleton University

Kim P. Huynh[†] Bank of Canada

David T. Jacho-Chávez[‡] Emory University

Abstract

This paper studies the limiting behavior of general functionals of order statistics and their multivariate concomitants for weakly dependent data. The asymptotic analysis is performed under a *conditional moment*-based notion of dependence for vector-valued time series. It is argued, through analysis of various examples, that the dependence conditions of this type can be effectively implied by other dependence formations recently proposed in time-series analysis, thus it may cover many existing linear and nonlinear processes. The utility of this result is then illustrated in deriving the asymptotic properties of a semiparametric estimator that uses the k-Nearest Neighbor estimator of the inverse of a multivariate unknown density. This estimator is then used to calculate consumer surpluses for electricity demand in Ontario for the period 1971 to 1994. A Monte Carlo experiment also assesses the efficacy of the derived limiting behavior in finite samples for both these general functionals and the proposed estimator.

Keywords: Order statistics; multivariate concomitant; k-nearest neighbor; semiparametric estimation; consumer surplus calculation.

AMS 2000 subject classification: 62G30; 62H12; 62G05; 62G07; 62E20

[*]Department of Economics, Carleton University, B-857 Loeb Building, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada. E-mail: ba_chu@carleton.ca. Web Page: http://http-server.carleton.ca/~bchu/

[†]Bank of Canada, 234 Wellington Street, Ottawa, Ontario K1A 0G9, Canada. E-mail: kim@huynh.tv. Web Page: http://www.huynh.tv/

[‡]Corresponding Author: Department of Economics, Emory University, Rich Building 306, 1602 Fishburne Dr., Atlanta, GA 30322-2240, USA. E-mail: djachocha@emory.edu. Web Page: http://userwww.service.emory.edu/~djachoc/

1 Introduction

Let $(\boldsymbol{X}_t^\top, Y_t)^\top$ be a $\mathbb{R}^{N+1}$-valued time series process on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $Y_{(1)} < \cdots < Y_{(t)} < \cdots < Y_{(T)}$ be the order statistics; and $\boldsymbol{X}_{[t]}$ paired with $Y_{(t)}$ is called the concomitant of the t-th order statistics in the sample $\{\boldsymbol{X}_t^\top, Y_t\}_{t=1}^T$.

The use of order statistics and their multivariate concomitants often arises in various statistical problems. For example, selection procedures dictates that s-observations ($< T$) are chosen on the basis of their Y-values. Then the corresponding $\boldsymbol{X}$-values represent their associated characteristics. Alternatively Y might represent the score on a screening test and $\boldsymbol{X}$ can represent the score of a later test. Concomitants have also proven useful in the estimation of parameters using doubly censored samples, i.e. Watterson (1958) in estimating means, Barnett et al. (1976) in estimating correlations, and Stokes (1977) in ranked set. The properties of concomitants have been studied extensively by many authors, and important contributions include but are not limited to David and Galambos (1974), Yang (1977), Nagaraja and David (1994), Khaledi and Kochar (2000) and Arnold et al. (2009), inter alios. The study of concomitants as an important class of statistics has been more recently reviewed by David and Nagaraja (1998).

A significant line of work has focused on the asymptotic distribution of general functions of concomitants. For example, Yang (1981a,b) proved the asymptotic normality, under mild regularity conditions, of functionals of the form

$$\frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) W_{[i]}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) h(Z_{(i)}, W_{[i]}),$$

where $J(\cdot)$ is a bounded, smooth score function which may depend on n, and $h(z, w)$ is a known $\mathbb{R}-$valued function. Stute (1993) established a functional central limit theorem for U-functions of concomitants defined as

$$\frac{1}{n(n-1)}\sum_{1\le i\neq j\le [nt]} K(W_{[i]}, W_{[j]}),$$

where $K(\cdot,\cdot)$ is any symmetric second-order U-kernel. Previous research have used the independent and identically distributed (i.i.d.) assumption in their execution. The notable exceptions are Puri and Tran (1980), Wu (1988), and Tran and Wu (1993). Specifically, Wu (1988) and Tran and Wu (1993) established the asymptotic theory for linear combinations of functions of order statistics (or *L-estimates*, as termed by Serfling, 1980, Chapter 8) for nonstationary time series. This assumption is often too strong in applications where data is collected sequentially over time.

Therefore, this paper studies the limiting behavior of general functionals of ordered statistics and their multivariate concomitants with *dependent* data. In particular, we prove the $\sqrt{T}$-asymptotic normality, under fairly mild regularity conditions, of functionals such as

$$\mathfrak{T}_T = \frac{1}{T}\sum_{t=1}^{T} J(t/T) h(\boldsymbol{X}_{[t]}, Y_{(t)}), \tag{1.1}$$

where $J(\cdot)$ is a bounded smooth score function and $h(\boldsymbol{x}, y)$ is some $\mathbb{R}$-valued known function of $(\boldsymbol{x}^\top, y)^\top \in \mathbb{R}^{N+1}$.

Studying the limiting properties of statistics such as (1.1) is important, because its usage in semi-parametric estimation can avoid the presence of random denominators and the usage of trimming functions altogether. For example, consider the *Single Index* model which is widely studied in the Statistics literature, see, e.g., Ichimura (1993), Härdle et al. (1993), Carroll et al. (1997), inter alia. Let $\{(Y_t^*, \boldsymbol{Z}_t^{*\top}, \boldsymbol{X}_t^{*\top})\}_{t=1}^T$ denote a vector-valued time series with the contemporaneous dependence generated by the partially linear single-index model

$$Y_t^* = g(\boldsymbol{Z}_t^{*\top}\boldsymbol{\alpha}) + \boldsymbol{X}_t^{*\top}\boldsymbol{\beta} + \epsilon_t, \tag{1.2}$$

where $\boldsymbol{Z}_t^*$ and $\boldsymbol{X}_t^*$ are random covariate vectors; $g(\cdot)$ represents an unknown, possibly non-differentiable, function; ϵ denote i.i.d. mean-zero random errors, which are independent of $\left(\boldsymbol{Z}_t^{*\top}, \boldsymbol{X}_t^{*\top}\right)^\top$; and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the unknown finite-dimensional parameters to be estimated. The Semiparametric Least Squares (SLS) estimator minimizes $Q_T(\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top) \doteq \sum_{t=1}^T \{Y_t^* - \widehat{g}(\boldsymbol{Z}_t^{*\top}\boldsymbol{\alpha}; \boldsymbol{\beta}) - \boldsymbol{X}_t^{*\top}\boldsymbol{\beta}\}^2$, where $\widehat{g}(\boldsymbol{Z}_t^{*\top}\boldsymbol{\alpha}; \boldsymbol{\beta})$ represents the Nearest Neighbor (NN) regression function estimator of $g(\boldsymbol{Z}_t^{*\top}\boldsymbol{\alpha})$, which can then be written in a form congruent with Eq. (1.1) as follows: Let $Y_t(\boldsymbol{\alpha}) \doteq \boldsymbol{Z}_t^{*\top}\boldsymbol{\alpha}$ and $W_t(\boldsymbol{\beta}) \doteq \boldsymbol{X}_t^{*\top}\boldsymbol{\beta}$,

$$\widehat{g}(Y_t(\boldsymbol{\alpha}); \boldsymbol{\beta}) \doteq \frac{1}{(T-1)h_T} \sum_{s=1}^{T-1} [Y_{[s]}^* - W_{[s]}(\boldsymbol{\beta})] K\left(\frac{F_T(Y_t(\boldsymbol{\alpha})) - s/(T-1)}{h_T}\right),$$

where $F_T(\cdot)$ is the empirical distribution function; $K(\cdot)$ is a kernel (weight) function; and $\boldsymbol{X}_{[s]}(\boldsymbol{\beta}) \doteq (Y_{[s]}^*, W_{[s]}(\boldsymbol{\beta}))$ denotes a vector of the concomitants of the order statistics $Y_{(s)}(\boldsymbol{\alpha})$ in the sample $\{(Y_1^*, W_1(\boldsymbol{\beta}), Y_1(\boldsymbol{\alpha})), \dots, (Y_{t-1}^*, W_{t-1}(\boldsymbol{\beta}), Y_{t-1}(\boldsymbol{\alpha})), (Y_{t+1}^*, W_{t+1}(\boldsymbol{\beta}), Y_{t+1}(\boldsymbol{\alpha})), \dots, (Y_T^*, W_T(\boldsymbol{\beta}), Y_T(\boldsymbol{\alpha}))\}$. Stute (1984) shows that in the i.i.d. case, the asymptotic behavior of $\widehat{g}(y; \boldsymbol{\beta})$ is the same as that of

$$\widehat{g}^*(y; \boldsymbol{\beta}) = \frac{1}{T-1} \sum_{s=1}^{T-1} J_T(s/(T-1))\{Y_{[s]}^* - W_{[s]}(\boldsymbol{\beta})\},$$

where $J_T\left(s/(T-1)\right) \doteq h_T^{-1} K\left((F(y) - s/(T-1))/h_T\right)$. Therefore the statistics $\widehat{g}^*(y; \boldsymbol{\beta})$ becomes a special case of (1.1) with a sample size-varying score function, $J_T(\cdot)$. Unlike Ichimura's (1993) SLS estimator, the NN regression function estimator contains no random denominator, and no trimming function is needed. However, the non-differentiability of the objective function $Q_T(\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)$ poses many technical challenges to the derivation of the asymptotic properties of the implied SLS estimator of (1.2) which are beyond the scope of this paper.

Instead, the results of this paper are applied here to study the asymptotic properties of a semiparametric k-Nearest Neighbor (k-NN) based estimators of objects such as:

$$\theta_0 = \int_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} E[Y|\boldsymbol{X} = \boldsymbol{x}] d\boldsymbol{x} = E\left[\frac{Y}{f(\boldsymbol{X})}\right], \tag{1.3}$$

where $\boldsymbol{\mathcal{X}}$ is some subset in the support of the multivariate density $f(\boldsymbol{X})$. Note that, without any loss of generality, assume that $\boldsymbol{\mathcal{X}}$ equals the whole support of $f(\boldsymbol{X})$. Let $\mathrm{Supp}(f) = \{\boldsymbol{x} \in \mathbb{R}^N : f(\boldsymbol{x}) \geq \epsilon \text{ for some } \epsilon > 0\}$, where $f(\boldsymbol{x})$ is assumed to be continuous and bounded, denote the support of $f(\boldsymbol{x})$.

In Economics, the object $E[Y|\boldsymbol{X} = \boldsymbol{x}]$ could represent nonparametric demand or supply functions for a product. In which case quantities such as θ_0 can be used to calculate consumer or producer surplus. The latter are paramount in Microeconomic theory. Recently, the asymptotic properties of various estimators of (1.3) when $N \geq 1$ has been studied by Lewbel and Schennach (2007), Jacho-Chávez (2008), Chu and Jacho-Chávez (2012), and Lu et al. (2012) under a variety of sampling schemes. A semiparametric estimator that utilizes the k-NN multivariate density estimator of $f(\cdot)$ is discussed in this paper. This density estimator was first proposed by Loftsgaarden and Quesenberry (1965). Pointwise consistency and asymptotic normality of the k-NN density estimator have been established under various data generating processes: see Moore and Yackel (1977a,b) for i.i.d. samples, Boente and Fraiman (1988, 1990) for mixing processes and Tran and Yakowitz (1993) and Li and Tran (2009) for mixing random fields.

The object of interest is the k-NN density in the aforementioned papers, but in the proposed semiparametric estimator the inverse of the k-NN multivariate density estimator is used instead. The proof of asymptotic normality for this semiparametric estimator, therefore, requires strong consistency of the inverse k-NN multivariate density estimator, which is established here under a fairly mild regularity condition involving a random dependence coefficient. Other estimators using the nearest-neighbors technique in semiparametric problems include Robinson (1987, 1995) and references therein.

Our method of proof can be viewed as a combination of the Gâteaux differential (see, e.g., Koroljuk and Borovskich, 1994, p. 48) and the martingale approximation approach developed in Gordin (1969), Philipp and Stout (1975) and Wu and Woodroofe (2004). These are commonly used methods to establish central limit theorems involving stationary data. Theoretically, another contribution is the introduction of a new method of quantifying the notion of contemporaneous dependence for $\mathbb{R}^{N+1}$-valued processes. In particular, it can be shown that the proposed dependence conditions are related to conditional mixing concepts, as introduced by Rao (2009), and random dependence coefficients, as independently proposed by Bickel and Bühlmann (1999) and Dedecker and Prieur (2005).

The paper is organized as follows: Section 2 summarizes some mathematical notations and definitions, then introduces three popular time series processes discussed in the text. Section 3 presents the conditions under which the asymptotic normality of functionals of order statistics and their multivariate concomitants is established. This section also demonstrate that our conditions hold for the examples introduced in Section 2. Section 3.1 provides sufficient conditions for our theoretical results to hold for stationary causal processes while Section 3.2 established the asymptotic normality of the new proposed semiparametric estimator of (1.3). Throughout, a discussion on how our conditions of weak dependence compare to other existing ones is presented. Section 4 presents some Monte Carlo evidence of the small-sample performance of the proposed asymptotic approximations, as well as an empirical application to the calculation of consumer surplus in consumer electricity demand in Ontario for the period 1971 to 1994. Proofs of the main theorems and results of technical flavor are gathered in Section 5 and the appendices.

2 Basic Notations, Definitions, and Examples

2.1 Basic Notations and Definitions

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, which is sufficiently rich to accommodate $(\boldsymbol{X}^\top, Y)$ and $\mathcal{T}$ be a measure-preserving, bijective and bimeasurable mapping from Ω onto itself. Let $\mathcal{I}$ denote the Borel algebra of invariant sets $A \in \mathcal{A}$ such that $\mathcal{T}^{-1}A = A$. If all the elements of $\mathcal{I}$ are of measure 0 or 1, then a sequence of random variables, $(\boldsymbol{X}^\top, Y)$, defined on $(\Omega, \mathcal{A}, \mathbb{P})$ is said to be ergodic. Define a strictly stationary vector-valued sequence of random variables, $(\boldsymbol{X}_t^\top, Y_t)$, which can be represented as $\boldsymbol{X}_t = \boldsymbol{K} \circ \mathcal{T}^t = \boldsymbol{K}(\mathcal{T}^t\omega)$ and $Y_t = H \circ \mathcal{T}^t = H(\mathcal{T}^t\omega)$ for all $\omega \in \Omega$, where $\boldsymbol{K}$ is a N-dimensional vector of Borel functions and H is a scalar Borel function. This formulation allows stationary causal processes. Let $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{N,t}, \epsilon_{N+1,t})$ denote a [N+1]-dimensional vector of i.i.d. random noises, $X_{it} = K_i(\ldots, \epsilon_{i,t-1}, \epsilon_{i,t})$ for $i = 1, \ldots, N$, and $Y_t = H(\ldots, \epsilon_{N+1,t-1}, \epsilon_{N+1,t})$. Then $(\boldsymbol{X}_t^\top, Y_t)$ is a causal process, thus naturally falls into the framework; and indeed Y_t depends on the filtration of $(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_T)$ via the filtration of $\boldsymbol{X}_t$. We emphasize that the class of causal processes is rather vast because all time series models used in practice (scalar, vector, or functional) have this representation (cf. Tong, 1990).

For every $i \in 1, \ldots, N$, let $\mathcal{F}_{t,X_i} = \sigma(X_{i,s},\ s \le t) = \mathcal{T}^{-t}\mathcal{F}_{0,X_i}$ and $\mathcal{F}^t_{X_i} = \sigma(X_{i,s},\ s \ge t) = \mathcal{T}^t\mathcal{F}_{0,X_i}$ be Borel algebras generated by $(X_{i,0}, \ldots, X_{i,t})$ and $(X_{i,t}, \ldots, X_{i,T})$ respectively. Let $\mathcal{F}_t = \sigma\big((\boldsymbol{X}_s, Y_s),\ s \le t\big) = \mathcal{T}^{-t}\mathcal{F}_0$ be he smallest σ-algebra in the product Borel algebra, $\mathcal{F}_{t,X_1} \otimes \mathcal{F}_{t,X_2} \otimes \cdots \otimes \mathcal{F}_{t,X_N} \otimes \mathcal{F}_{t,Y}$, generated by $\{(\boldsymbol{X}_0^\top, Y_0)^\top, \ldots, (\boldsymbol{X}_t^\top, Y_t)^\top\}$; $\mathcal{F}^t = \sigma\big((\boldsymbol{X}_s, Y_s),\ s \ge t\big) = \mathcal{T}^t\mathcal{F}_0$ represents the smallest σ-algebra in the product Borel algebra, $\mathcal{F}^t_{X_1} \otimes \mathcal{F}^t_{X_2} \otimes \cdots \otimes \mathcal{F}^t_{X_N} \otimes \mathcal{F}^t_Y$, generated by $\{(\boldsymbol{X}_t^\top, Y_t)^\top, \ldots, (\boldsymbol{X}_T^\top, Y_T)^\top\}$; $\mathcal{F}_{t,\boldsymbol{X}} = \sigma(\boldsymbol{X}_s,\ s \le t) = \mathcal{T}^{-t}\mathcal{F}_{0,\boldsymbol{X}}$ is the smallest σ-algebra in the product Borel algebra, $\mathcal{F}_{t,X_1} \otimes \mathcal{F}_{t,X_2} \otimes \cdots \otimes \mathcal{F}_{t,X_N}$, generated by $(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_t)$; $\mathcal{F}^t_{\boldsymbol{X}} = \sigma(\boldsymbol{X}_s,\ s \ge t) = \mathcal{T}^t\mathcal{F}_{0,\boldsymbol{X}}$ represent the smallest σ-algebra in the product Borel algebra, $\mathcal{F}^t_{X_1} \otimes \mathcal{F}^t_{X_2} \otimes \cdots \otimes \mathcal{F}^t_{X_N}$, generated by $(\boldsymbol{X}_t, \ldots, \boldsymbol{X}_T)$; $\mathcal{F}_{t,Y} = \sigma(Y_s,\ s \le t) = \mathcal{T}^{-t}\mathcal{F}_{0,Y}$ is a Borel algebra generated by $(Y_0, \ldots, Y_t)$; $\mathcal{F}^t_Y = \sigma(Y_s,\ s \ge t) = \mathcal{T}^t\mathcal{F}_{0,Y}$ is a Borel algebra generated by $(Y_t, \ldots, Y_T)$; $\mathcal{F}_{\boldsymbol{X}_t} = \sigma(\boldsymbol{X}_t)$ represents the smallest σ-algebra in the product Borel algebra, $\mathcal{F}_{X_{1,t}} \otimes \mathcal{F}_{X_{2,t}} \otimes \cdots \otimes \mathcal{F}_{X_{N,t}}$, generated by $\boldsymbol{X}_t$; $\mathcal{F}_{Y_t} = \sigma(Y_t)$ is a Borel algebra generated by Y_t, while $F(y|\mathcal{I})$ is the invariant distribution of Y_t, $\lim_{\tau \longrightarrow \infty} P(Y_\tau \le y | Y_0 \in \mathcal{I})$, where $\mathcal{I}$ represents the invariant sets with Borel algebra consisting of probability measures 0 or 1. The continuity of the following probability distribution functions: $F(\boldsymbol{x})$, $F(y)$, and $F(\boldsymbol{x}, y)$ will be assumed throughout this paper – so that ties among the $\boldsymbol{X}$ and Y-variates can be neglected in probability. Finally the quantity $\|A\|_p$ is the L_p-norm of A, i.e. $\{E[|A|^p]\}^{1/p}$; $\|A\|_{p,\mathcal{I}}$ is the L_p-norm of A conditional on $\mathcal{I}$, i.e. $\{E[|A|^p|\mathcal{I}]\}^{1/p}$. Note that it is possible to simplify the reading by assuming that $\mathcal{I} = \{\Omega, \emptyset\}$; in this case, any $\mathcal{I}$-measurable random variable will become a constant, i.e. $E[A|\mathcal{I}] = E[A]$ and $E[A|Y = y, \mathcal{I}] = E[A|Y = y]$.

2.2 Examples

For simplicity, we set $N = 1$ and let ξ_t denote an i.i.d. mean-zero random variable in the following data generating processes (d.g.p.'s):

- Moving Average (MA) model, $X_t = \sum_{i=0}^{\infty} \theta_i \xi_{t-i}$, where θ_i are MA coefficients.

- Bilinear (BILINEAR) model, $X_t = aX_{t-1} + \xi_t + bX_{t-1}\xi_t$, where a and b take their values on $\mathbb{R}$, see, e.g., Tong (1990).

- Generalized Autoregressive Conditional Heteroskedasticity (1,1) (GARCH) model, $X_t = \sigma_t \xi_t$ with $\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2$, where ω, α and β take their values on $\mathbb{R}$.

We now introduce three examples that will serve as illustrations on how a wide range of popular d.g.p.'s satisfy the main assumptions stated below in Section 3.

Example 1: $Y_t = X_t \epsilon_t$, where ϵ_t are i.i.d. mean-zero random variables independent of X_t, and X_t can admit one of the d.g.p.'s above.

Example 2: $Y_t = X_t Z_t$, where Z_t is a mean-zero stochastic process that can also follow one of the above d.g.p.'s, where Z_t is independent of X_t.

Example 3: $Y_t = X_t + Z_t$, where Z_t is as in the Example 2 above.

It will be shown in the next section that the required conditions in the paper are satisfied in these 3 examples. Example 1 is the base of our numerical experimentation in Section 4.1.

3 Assumptions and Main Results

Expressing (1.1) as a functional of empirical distribution functions F_T, yields

$$
\begin{aligned}
\mathfrak{T}(F_T) &= \frac{1}{T}\sum_{t=1}^{T} J(t/T) h(\boldsymbol{X}_{[t]}, Y_{(t)}) \\
&= \sum_{t=1}^{T} J(F_T(Y_{(t)})) h(\boldsymbol{X}_{[t]}, Y_{(t)}) \left[F_T(\boldsymbol{X}_{[t+1]}, Y_{(t+1)}) - F_T(\boldsymbol{X}_{[t]}, Y_{(t)})\right] \\
&= \int_{\mathbb{R}^{N+1}} J\left(F_T(y)\right) h(\boldsymbol{x}, y) dF_T(\boldsymbol{x}, y),
\end{aligned}
$$

where $F_T(y)$ is the empirical distribution of y and $F_T(\boldsymbol{x}, y)$ is the joint empirical distribution of $(\boldsymbol{X}^\top, Y)^\top$. Similarly, let $m_h(y;\mathcal{I}) \doteq E[h(\boldsymbol{X}, Y)|Y = y, \mathcal{I}]$ and $m_h(y) \doteq E[h(\boldsymbol{X}, Y)|Y = y]$.

The following regularity conditions are introduced to facilitate our theoretical development:

A1 Moment Bounds:

For a given integer, $p > 1$, $\max\{\|h(\boldsymbol{X}_0, Y_0)\|_{p,\mathcal{I}}, \|h(\boldsymbol{X}_0, Y_0)\|_{2p/(p-1),\mathcal{I}}\} < \infty$.

A2 Conditional Moments:

(a) $\left\|m_h^{'}(Y_0;\mathcal{I})\right\|_{p^*,\mathcal{I}} < \infty$, where $m_h^{'}(\cdot;\mathcal{I})$ is the first derivative of $m_h(\cdot;\mathcal{I})$.

(b) $\lim_{\tau\longrightarrow\infty} \left\|\sup_y |P(Y_\tau \leq y|\mathcal{F}_{Y_0}) - F(y|\mathcal{I})|\right\|_{q^*,\mathcal{I}} = 0$, where p^* and q^* are such integers that $1/p^* + 1/q^* = 1 + (p-1)/p$.

A3 Conditional Joint Moments:

(a) $\sum_{\tau=1}^{\infty} \left\| \|m_h(Y_\tau;\mathcal{F}_0) - m_h(Y_\tau;\mathcal{I})\|_{2,\mathcal{F}_0} \right\|_{p/(p-1),\mathcal{I}} < \infty$.

(b) $\sum_{\tau=1}^{\infty} \|E[h(\boldsymbol{X}_\tau, Y_\tau)h(\boldsymbol{X}_0, Y_0)|\mathcal{F}_{\tau,Y}, \mathcal{I}] - m_h(Y_\tau;\mathcal{I})m_h(Y_0;\mathcal{I})\|_{p/(p-1),\mathcal{I}} < \infty$.

It is helpful to note at this point that, for a scalar-valued X_t, the statistics defined by Eq. (1.1) can also be represented as $\mathfrak{T}(F_T) = \frac{1}{T}\sum_{t=1}^{T} J(t/T)h(X_{(t)}, Y_{[t]})$. Accordingly, we shall replace $m_h(Y;\mathcal{I})$ and $m_h(Y;\mathcal{F}_0)$ in Assumptions A1, A2 and A3 with $m_h(X;\mathcal{I})$ and $m_h(X;\mathcal{F}_0)$ respectively.

Remark 3.1. Assumption A1 entails that moments of the function $h(\cdot,\cdot)$ are bounded up to a certain order, e.g. any $p > 1$ can be used depending on what is needed. This mild moment-type condition is often employed to obtain many central limit theorems and invariance principles, i.e. Lyapunov's central limit theorem. As for Assumption A2, condition A2a is automatically fulfilled by any Lipschitz continuous function, $m_h(\cdot)$, though this condition does not imply Lipschitz continuity. Condition A2b entails asymptotic weak independence (in the ergodic sense) of the random process Y_t. This condition is quite close in spirit to the mixing characteristic introduced by Rinott and Rotar (1999, p. 613); rather than conditioning the sum of random elements belonging to a 'past' Borel algebra on a 'future' Borel algebra, we condition a 'future' random element on a 'past' Borel algebra. Condition A2b may be weaker than the usual mixing conditions (see, e.g., Bradley, 1986 for definition of various mixing concepts). For example, by virtue of the covariance inequality for strong mixing random variables (Ibragimov, 1962), one can verify that, for stationary ergodic processes, $\left\|\sup_y |P(Y_\tau \le y|\mathcal{F}_0) - F(y)|\right\|_{q^*} \le 2(2^{1/q^*} + 1)\alpha_\tau^{1/q^*-1/r^*}$ for some $r^* \ge q^* \ge 1$, where α_τ represents Rosenblatt's (1956a) strong mixing coefficient. Hence, strong mixing implies Condition A2b. Indeed, many causal processes used in practice, e.g. stationary and ergodic Markov chains, have been shown to satisfy this condition (see, e.g., Pham and Tran, 1985, Pham, 1986, among many others).

Over the past decades, many approaches have been proposed to formalize weak dependence. In this context, we now discuss how the notion of weak dependence introduced here compares to other existing ones. Perhaps, the most popular are the strong mixing property and its variants like β, ϕ, ρ and ψ mixing coefficients, which were developed in the seminal papers of Rosenblatt (1956a) and Ibragimov (1962). The general idea is to measure the maximal dependence between two events pertaining to the backward $\sigma-$ algebra $\mathcal{F}_t$ and the forward $\sigma-$ algebra $\mathcal{F}^{t+m}$, respectively. The memory is fading as this maximal dependence decays to zero, as m increases to infinity. For example, the strong mixing dependence is formalized by

$$\alpha_m = \sup_{\substack{A\in\mathcal{F}_t \\ B\in\mathcal{F}^{t+m}}} |P(A\cap B) - P(A)P(B)|.$$

A sequence is $\alpha-$mixing if α_m tends to zero for a sufficiently large m. Recently, Rao (2009) has introduced the concept of *conditional strong mixing*, i.e. let $\mathcal{M}$ be a $\sigma-$algebra of $\mathcal{A}$, a sequence is said to be conditionally strong mixing if there exists a nonnegative $\mathcal{M}-$ measurable random variable $\alpha_m^*(\mathcal{M})$ converging to zero a.s. as m goes to infinity, such that

$$|P(A\cap B|\mathcal{M}) - P(A|\mathcal{M})P(B|\mathcal{M})| \le \alpha_m^*(\mathcal{M}) \text{ a.s.}$$

In Section 3.3 below we shall show that, for stationary and ergodic processes, $\{\boldsymbol{X}_t^\top, Y_t\}$, the Cesàro summability of the *conditional strong mixing* coefficient $\alpha_m^*(\mathcal{M})$ effectively implies Condition A3b, but not vice versa.

Although many results have been established for strongly mixing sequences, see e.g. Bradley (2007) and Rio (2000), many classes of time series have been shown not to satisfy these conditions, i.e. Andrews (1984). Therefore, Bickel and Bühlmann (1999) and Dedecker and Prieur (2005) independently introduced a new concept of weak dependence. Their notion of weak dependence makes explicitly the asymptotic independence between 'past' and 'future'. Roughly speaking, the covariance between measurable functions of the 'past' and 'future' becomes small as the distance between the 'past' and the 'future' is large. The decay rate of this covariance is measured through the L_p-distance between the conditional expectation of a Lipschitz function, $g(\cdot)$, of a L_p-integrable random variable, X, given $\mathcal{M}$ and the expectation of $g(X)$. Thus the $\mathcal{M}-$measurable random $\theta-$ coefficient is defined as

$$\theta_p(\mathcal{M}) = \sup\left\{\|E[g(X)|\mathcal{M}] - E[g(X)]\|_p,\ \text{for some function } g \in \Lambda^{(1)}\right\},$$

where $\Lambda^{(1)}$ denotes the class of Lipschitz functions with the Lipschitz coefficient at most equal to one.

The regularity conditions A3a and A3b essentially imply that the dependence coefficient of θ-type is Cesàro summable. Here the σ-algebra $\mathcal{M}$ contains two sub σ-algebras, $\mathcal{F}_{Y_\tau}$ and $\mathcal{F}_0$, and an important difference being that the functions $m_h(\cdot)$ and $h(\cdot)$ in the above conditions may not need to be Lipschitzian.

An alternative approach to define weak dependence is based on a martingale projection, $\mathcal{P}_t(X) = E[X|\mathcal{F}_t] - E[X|\mathcal{F}_{t-1}]$. In the context where sequences, X_t, are stationary and ergodic Markov chains, Wu (2005, 2007) has employed some regularity conditions regarding the Cesàro summability of the L_p-norm of $\mathcal{P}_t(X)$ and successfully proved strong invariance principles with nearly optimal bounds. Wu (2007) also validates that these regularity conditions can be directly inferred from the Cesàro summability of an input/output dependence measure, which is defined as the L_p distance between the conditional expectation of the Markov chain $X_t = g(\dots, \epsilon_{t-1}, \epsilon_t)$ for some i.i.d. $(\epsilon_t)_{t\in\mathbb{Z}}$ and the conditional expectation of a decoupled sequence of X_t. In Section 3.1, we demonstrate that by taking into account the Cesàro summability of the L_p-norm of a *conditional* input/output dependence measure the asymptotic normality of the statistics (1.1) can be established. As a result, Corollary 3.1 is a special case of Theorem 3.1 because Assumption B3 below explicitly implies Assumption A3.

We shall now demonstrate that the d.g.p.'s provided in Section 2.2 can fulfill Assumptions A1-A3.
Example 1 (continued): For ease of derivation, we take the function $h(x, y) = (xy)^2$. Since $\mathcal{F}_t = \mathcal{F}_{t,Y} \doteq \sigma\left((X_s, \epsilon_s) : s \le t\right)$, Assumption A1 then becomes

$$\max\left\{\|\epsilon_0^2\|_p \|X_0^4\|_p, \|\epsilon_0^2\|_{2p/(p-1)} \|X_0^4\|_{2p/(p-1)}\right\} < \infty. \tag{3.1}$$

Condition A2a is equivalent to

$$\left\|\epsilon_0^2\right\|_{p^*} \left\|X_0^3\right\|_{p^*} < \infty. \tag{3.2}$$

It is straightforward to check that $\|m_h(X_\tau; \mathcal{F}_0) - m_h(X_\tau)\|_{2,\mathcal{F}_0} = 0$ and $E[h(X_0, Y_0)h(X_\tau, Y_\tau)|\mathcal{F}_{\tau,X}] - m_h(X_\tau)m_h(X_0) = 0$, which then imply Assumption A3.

- MA: The MA coefficients θ_i must be chosen so as to verify (3.1) and (3.2). In addition, if $\lim_{\tau\to\infty}\sum_{t=\tau}^{\infty}\left(\sum_{i=t}^{\infty}|\theta_i|\right)^{\delta/(1+\delta)}=0$, where δ is some positive generic constant, then X_t satisfies the absolutely-regular mixing condition (Pham and Tran, 1985). The absolute regularity then implies Condition A2b.

- BILINEAR: The stationarity condition is $a^2+b^2<1$ (see Tong (1990, p. 159)). Therefore the existence of higher-order moments of ϵ_0 will validate (3.1) and (3.2). Pham (1986) shows that, under some regularity conditions, X_t is geometrically ergodic. This automatically implies Condition A2b.

- GARCH: If $E[\log(\alpha\xi_t^2+\beta)]<0$, then this GARCH(1,1) model has a non-anticipative strictly stationary solution, which also satisfies the geometrically absolutely-regular mixing condition (see Francq and Zakoïan, 2010, p. 71). Since absolute regularity is stronger than strong mixing, Condition A2b is validated.

Example 2 (continued): Define $h(x,y)=xy$. Assumption A1 then becomes

$$\max\left\{\|X_0^2Z_0\|_p,\|X_0^2Z_0\|_{2p/(p-1)}\right\}<\infty. \tag{3.3}$$

Meanwhile, Condition A2a is fulfilled because of $m_h(X_0)=0$. Also, in view of Example 1, the various representations of X_t satisfy Condition A2b. We shall now verify Assumption A3: Since $\mathcal{F}_\tau=\mathcal{F}_{\tau,Y}\doteq\sigma\left((X_s,Z_s):s\leq\tau\right)$, it then follows that $m_h(X_\tau;\mathcal{F}_0)=X_\tau^2E[Z_\tau|X_\tau,\mathcal{F}_0]=X_\tau^2E[Z_\tau|Z_0]$ because Z is independent of X. Therefore, an application of Hölder's inequality yields

$$\begin{aligned}\left\|\|m_h(X_\tau;\mathcal{F}_0)-m_h(X_\tau)\|_{2,\mathcal{F}_0}\right\|_{p/(p-1)} &= \left\|\|X_\tau^2\|_{2,\mathcal{F}_0}E[Z_\tau|Z_0]\right\|_{p/(p-1)}\\ &\leq \left\|\|X_\tau^2\|_{2,\mathcal{F}_0}\right\|_{p_1^*}\|E[Z_\tau|Z_0]\|_{q_1^*}\\ &\leq \left\|X_\tau^4\right\|_{p_1^*/2}^{1/2}\|E[Z_\tau|Z_0]\|_{q_1^*},\end{aligned}$$

where $p_1^*>2$ and $q_1^*>1$ such that $1/p_1^*+1/q_1^*=(p-1)/p$; and

$$\begin{aligned}\|E[h(X_0,Y_0)h(X_\tau,Y_\tau)|\mathcal{F}_{\tau,X}-m_h(X_\tau)m_h(X_0)\|_{p/(p-1)} &\leq \left\|X_0^2X_\tau^2\right\|_{p/(p-1)}E[Z_0Z_\tau]\\ &\leq \left\|X_0^2\right\|_{2p/(p-1)}\left\|X_\tau^2\right\|_{2p/(p-1)}|E[Z_0Z_\tau]|.\end{aligned}$$

Assuming that the coefficients in the time-series models, defined in Example 1, of X_t ensure that X_t is strictly stationary and $\max\left\{\|X_0^2\|_p,\left\|X_0^2\right\|_{2p/(p-1)},\left\|X_\tau^4\right\|_{p_1^*/2}\right\}<\infty$, Assumption A3 is thus essentially weaker than

$$\max\left\{\sum_{\tau=1}^{\infty}\|E[Z_\tau|Z_0]\|_{q_1^*},\sum_{\tau=1}^{\infty}|E[Z_0Z_\tau]|\right\}<\infty. \tag{3.4}$$

An application of covariance inequalities for mixing random variables (see, e.g., Truong and Stone 1992 and Ibragimov 1962) yields

$$\begin{aligned}\sum_{\tau=1}^{\infty}\|E[Z_\tau|Z_0]\|_{q_1^*} &\leq 2(2^{1/q_1^*}+1)\|Z_0\|_{r_1^*}\sum_{\tau=1}^{\infty}\alpha_\tau^{1/q_1^*-1/r_1^*},\ \text{where } r_1^*\geq q_1^*\geq 1,\\ \sum_{\tau=1}^{\infty}|E[Z_\tau|Z_0]| &\leq 8\|Z_0\|_{p_2^*}\|Z_\tau\|_{q_2^*}\sum_{\tau=1}^{\infty}\alpha_\tau^{1-1/p_2^*-1/q_2^*},\ \text{where } 1/p_2^*+1/q_2^*<1.\end{aligned}$$

We shall now establish that Conditions (3.3) and (3.4) indeed holds under the various d.g.p. of Z_t.

- MA: Z_t is strong mixing with $\alpha_\tau = O\left(\sum_{t=\tau+1}^{\infty} G_t(r)^{1/(1+r)}\right)$, where, for either an even positive integer, r, or $0 < r \leq 2$, $G_t(r) \doteq \begin{cases} 2\sum_{i=t}^{\infty} |\theta_i|^r, & r\leq 2, \\ 2^{r-1}\left(\sum_{i=t}^{\infty} \theta_i^2\right)^{r/2}, & r\geq 2. \end{cases}$ (see Davidson, 1994, Theorem 14.9, for sufficient conditions pertaining to this result). Therefore, the MA coefficients θ_i must be chosen so as to warrant the existence of the higher-order moments of Z_t such that $\max\left\{\sum_{\tau=1}^{\infty}\left(\sum_{t=\tau+1}^{\infty} G_t(r)^{1/(1+r)}\right)^{1/q_1^*-1/r_1^*}, \sum_{\tau=1}^{\infty}\left(\sum_{t=\tau+1}^{\infty} G_t(r)^{1/(1+r)}\right)^{1-1/p_2^*-1/q_2^*}\right\} < \infty$, which then validates Conditions (3.3) and (3.4).

- BILINEAR: Geometric ergodicity implies absolute regularity with a geometric convergence rate, which is in turn stronger than strong mixing. Therefore, in view of Example 1, we obtain $\alpha_\tau = O\left(\ell^\tau\right)$ for some $0 < \ell < 1$. It then follows that, if q_1^*, r_1^*, p_2^* and q_2^* are chosen in such a way that $\max\left\{\ell^{1/q_1^*-1/r_1^*}, \ell^{1-1/p_2^*-1/q_2^*}\right\} < 1$ so conditions (3.3) and (3.4) are validated.

- GARCH: The GARCH(1,1) process also satisfies the absolutely-regular mixing condition with a geometric convergence rate. The verification of Conditions (3.3) and (3.4) can be done in exactly the same way as for the bilinear model.

Example 3 (continued): Define $h(x,y) = xy$, it is immediate to obtain via Hölder's inequality that:

$$\begin{aligned} \left\| \|m_h(X_\tau;\mathcal{F}_0) - m_h(X_\tau)\|_{2,\mathcal{F}_0} \right\|_{p/(p-1)} &\leq \left\|X_\tau^2\right\|_{p_1^*/2}^{1/2} \|E[Z_\tau|Z_0]\|_{q_1^*}, \\ \|E[h(X_0,Y_0)h(X_\tau,Y_\tau)|\mathcal{F}_{\tau,X} - m_h(X_\tau)m_h(X_0)\|_{p/(p-1)} &\leq \|X_0\|_{2p/(p-1)} \|X_\tau\|_{2p/(p-1)} |E[Z_0 Z_\tau]|. \end{aligned}$$

Therefore, Assumptions A1, A2, and A3 can be easily verified for all d.g.p. as in Example 2.

Prior to stating the first theorem of this paper, it is necessary to define the following martingale difference sequence: $W_t^* \doteq \sum_{s=t}^{\infty}\{E[W_s|\mathcal{F}_t] - E[W_s|\mathcal{F}_{t-1}]\}$, where $W_t = J(F(Y_t|\mathcal{I}))\{h(\boldsymbol{X}_t, Y_t) - m_h(Y_t|\mathcal{I})\} - \int_{\mathbb{R}} J(F(y|\mathcal{I}))\{\mathbb{I}(Y_t \leq y) - F(y|\mathcal{I})\}dm_h(y|\mathcal{I})$, and $\mathbb{I}(\cdot)$ is the standard indicator function. The following theorem states the main result of the paper.

Theorem 3.1. *Suppose that Assumptions A1, A2, and A3 hold. Then*

$$\sqrt{T}(\mathfrak{T}(F_T) - \mathfrak{T}(F)) \overset{W}{\Longrightarrow} \mathcal{N}\left(0, \sigma_W^2(\mathcal{I})\right), \tag{3.5}$$

where $\sigma_W^2(\mathcal{I}) = E[W_1^{*2}|\mathcal{I}] = E[W_0^2|\mathcal{I}] + 2\sum_{s=1}^{\infty} E[W_0 W_s|\mathcal{I}]$.

3.1 Stationary Causal Processes

Suppose that $(\boldsymbol{X}_t^\top, Y_t)^\top$ is a $\mathbb{R}^{N+1}$-valued stationary causal process on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In other words, $X_{i,t} = g_i(\xi_{i,t})$, $\forall\, i \in \{1,\ldots,N\}$, and $Y_t = g_{N+1}(\xi_{N+1,t})$ with $\xi_{j,t} = (\ldots, \epsilon_{j,0}, \ldots, \epsilon_{j,t-1}, \epsilon_{j,t})$ $\forall\, j \in \{1,\ldots,N+1\}$, for some measurable functions $\boldsymbol{g}(\cdot) = (g_1(\cdot), \ldots, g_{N+1}(\cdot))^\top$.

Let's us define $\widetilde{g}(\boldsymbol{X}_t, Y_t) = \widetilde{g} \circ \boldsymbol{g}(\boldsymbol{\xi}_t) = h(\boldsymbol{X}_t, Y_t) - m_h(Y_t)$, where $\boldsymbol{\xi}_t = (\xi_{1,t}, \ldots, \xi_{N,t}, \xi_{N+1,t}) = (\boldsymbol{\epsilon}_1^\top, \ldots, \boldsymbol{\epsilon}_t^\top)$ with $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \ldots, \epsilon_{N+1,t})^\top$, as a joint stochastic process of $\boldsymbol{X}_t$ and Y_t; and define $\widetilde{h}_\ell(\boldsymbol{\xi}_s) =$

$E[\widetilde{g} \circ \boldsymbol{g}(\boldsymbol{\xi}_{s+\ell})|\boldsymbol{\xi}_s]$ for some $\ell \geq 1$ as the ℓ-step conditional expectation of $\widetilde{g} \circ \boldsymbol{g}(\boldsymbol{\xi}_t)$. The input/output dependence measures $\boldsymbol{\alpha}^*_{t-1}$, as proposed by Wu (2007), is given by

$$\alpha^*_{t-1} = \left\| \widetilde{h}_1(\boldsymbol{\xi}_{t-1}) - \widetilde{h}_1(\boldsymbol{\xi}^*_{t-1}) \right\|_{p/(p-1)},$$

where $\boldsymbol{\xi}^*_t = (\xi^*_{1,t}, \ldots, \xi^*_{i,t}, \ldots, \xi^*_{N+1,t})^\top$ with $\xi^*_{i,t} = (\xi'_{i,0}, \epsilon_{i,1}, \ldots, \epsilon_{i,t-1}, \epsilon_{i,t})$ (the sequence $\xi'_{i,0}$ denotes an i.i.d. copy of $\xi_{i,0}$ for every $i \in \{1, \ldots, N+1\}$).

The following assumptions guarantee that Theorem 3.1 holds for vector-valued stationary, ergodic Markov chains.

B1 Moments Bounds: $\|\widetilde{g} \circ \boldsymbol{g}(\boldsymbol{\xi}_0)\|_p < \infty$ for some integer, p, satisfying $p > \frac{2p}{p-1}$.

B2 Conditional Moments: $\|m'_h(\xi_{N+1,0})\|_{p^*} < \infty$ for some integer, $p^* > 1$.

B3 Input/Output Dependence: (a) $\sum_{\tau=1}^{\infty} \alpha^*_{\tau-1} < \infty$; (b) $\sum_{\tau=1}^{\infty} \left\| \alpha^*_{\tau-1}(\xi_{N+1,\tau}) \right\|_{p/(p-1)} < \infty$, where $\alpha^*_{\tau-1}(\xi_{N+1,\tau}) = \|\widetilde{h}_1(\boldsymbol{\xi}^\circ_{\tau-1}|\xi_{N+1,\tau}) - \widetilde{h}_1(\boldsymbol{\xi}^{\circ*}_{\tau-1}|\xi_{N+1,\tau})\|_{2,\mathcal{F}_{\tau,Y}}$, $\boldsymbol{\xi}^\circ_t = (\xi_{1,t}, \ldots, \xi_{N,t})$ and $\widetilde{h}_\ell(\boldsymbol{\xi}^\circ_s|\xi_{N+1,s+\ell}) = E\left[\widetilde{g} \circ \boldsymbol{g}(\boldsymbol{\xi}_{s+\ell})|\boldsymbol{\xi}^\circ_s, \xi_{N+1,s+\ell}\right]$.

Corollary 3.1. *Suppose that Assumptions B1, B2, and B3 hold. Then the asymptotic normality stated in Theorem 3.1 follows.*

3.2 Semiparametric Estimation by k-Nearest Neighbor

This section illustrates the usage of the above result for deriving the asymptotic properties of the k-NN semiparametric estimator of objects like (1.3). Throughout this section, Supp(f) is assumed to be a compact subspace in the N-dimensional real space $(\mathbb{R}^N, \|\cdot\|, \mathcal{L})$ equipped with the Euclidean distance $\|\cdot\|$ and the Lebesgue measure $\mathcal{L}$.

As mentioned in the Introduction, consistent estimation of objects like (1.3) are important because numerous existing semiparametric estimators in Economics and Statistics make direct or indirect use of quantities such as (1.3), see e.g. Härdle and Stoker (1989), Hausman and Newey (1995), Lewbel (1998), Hong and White (2005), Hall and Yatchew (2005), Jacho-Chávez (2008), Chu and Jacho-Chávez (2012), inter alia. More recently, using Rosenblatt's (1956b) kernel density estimator, Lu et al. (2012) have studied the estimation of (1.3) in an i.i.d. setting with possibly missing-at-random Y_t.

In this section, the proposed estimator uses the k-NN multivariate density estimator instead. In particular, if the stationary density, $f(\boldsymbol{x})$, is unknown and one has data $\{\boldsymbol{X}_t^\top, Y_t\}_{t=1}^T$, then it can be estimated by Loftsgaarden and Quesenberry's (1965) k-NN multivariate density estimator, i.e.

$$\widehat{f}(\boldsymbol{x}) = \frac{k_T}{T\mathcal{L}(V_{R_T(\boldsymbol{x},k_T)})}, \tag{3.6}$$

where $R_T(\boldsymbol{x}, k_T)$ denotes the Euclidean distance between $\boldsymbol{x}$ and its k-th nearest neighbor among the $\{\boldsymbol{X}_t\}_{t=1}^T$. The quantity $V_{R_T(\boldsymbol{x},k_T)}$ is the volume of a ball with the radius $R_T(\boldsymbol{x}, k_T)$, since $\mathcal{L}(V_{R_T(\boldsymbol{x},k_T)}) = R_T^N(\boldsymbol{x}, k_T)\mathcal{L}(V_1)$ with $\mathcal{L}(V_1) = \frac{\pi^{N/2}}{\Gamma((p+2)/2)}$ is the volume of the unit sphere in $\mathbb{R}^N$. Note that (3.6) corresponds to Mack and Rosenblatt's (1979) version of the k-NN with uniform weights. Hereafter, the notation k is used to refer to k_T unless confusion is likely.

In view of Eq. (3.6), an estimator of θ_0 in Eq. (1.3) is then given by

$$\widehat{\theta} = \frac{\mathcal{L}(V_1)}{k} \sum_{t=1}^{T} J(t/T) Y_{(t)} R_T^N(\boldsymbol{X}_{[t]}, k). \tag{3.7}$$

Here the weight function $J(\cdot)$ is assumed known and satisfies, apart from the smoothness and boundedness condition as in (1.1), the relation $\int_0^1 |J(\tau) - 1| \leq 1$.

Define the following *pseudo-metric* on $\mathbb{Z}^+$:

$$\rho(t,s) = \left[E\|\boldsymbol{X}_t - \boldsymbol{X}_s\|^2\right]^{1/2}.$$

This pseudo-metric is called the deviation generated by the vector-valued random function $\boldsymbol{X}$. Let $B_\epsilon(t) = \{s \in \mathbb{Z}^+ : \ \rho(t,s) < \epsilon\}$ denote the $\rho-$ball of radius $\epsilon > 0$ with the center point $t \in \mathbb{Z}^+$. Given that there exists a finite covering of $\mathbb{Z}^+$ by $\rho-$balls, we denote by $N(\epsilon, \mathbb{Z}^+, \rho)$ the number of elements in the least $\epsilon-$covering of $\mathbb{Z}^+$. The quantity $H(\epsilon, \mathbb{Z}^+, \rho) = \log N(\epsilon, \mathbb{Z}^+, \rho)$ is then the *entropy* of $\mathbb{Z}^+$ with respect to the pseudo-metric ρ.

Some further regularity conditions are:

C1 Data Generating Processes: Let $(\boldsymbol{X}_t^\top, Y_t)^\top$ be a $\mathbb{R}^{N+1}$-valued stationary, ergodic process on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The scalar process Y_t depends on the backward Borel algebra $\mathcal{F}_{t,\boldsymbol{X}}$ or the forward Borel algebra $\mathcal{F}_{\boldsymbol{X}}^t$ via $\boldsymbol{X}_t$, i.e., $E[Y_t|\mathcal{F}_{t,\boldsymbol{X}}] = g(\boldsymbol{X}_t)$ and $E[Y_t|\mathcal{F}_{\boldsymbol{X}}^t] = g(\boldsymbol{X}_t)$.

C2 Moment Conditions:

(a) Let $h(\boldsymbol{X}_t, Y_t) = \frac{Y_t - g(\boldsymbol{X}_t)}{f(\boldsymbol{X}_t)}$, then $\|h(\boldsymbol{X}_0, Y_0)\|_p < \infty$ for some $p \geq 4$.

(b) $\|Y\|_4 < \infty$.

(c) $\left\{\int_{\mathcal{X}} g^p(\boldsymbol{x}) d\boldsymbol{x}\right\}^{1/p} < \infty$ for some $p \geq 1$.

(d) $\left\|m_h'(Y_0)\right\|_{p^*} < \infty$. for some $p^* > 1$.

C3 Conditional Probabilities:

(a) $\lim_{T \longrightarrow \infty} \sup_{1 \leq t \leq T} \sum_{s=1}^{T} s^{-1/2} \|P(\boldsymbol{X}_s \in A_t) - P(\boldsymbol{X}_s \in A_t|\mathcal{F}_0)\|_{2\ell} < \infty$ for some integer, $\ell \geq 1$, where $\{A_t\}_{t=1}^T$ is a collection of some disjoint random sets containing the sequence $\{\boldsymbol{X}_t\}_{t=1}^T$ such that $\mathcal{X} \subset \bigcup_{t=1}^T A_t$.

(b) $\lim_{\tau \longrightarrow \infty} \sup_{y \in \mathbb{R}} \|P(Y_\tau \leq y|\mathcal{F}_{Y_0}) - F(y)\|_{q^*} = 0$ for some $q^* \geq 1$ such that $1/p^* + 1/q^* = 1 - (p-1)/p$.

C4 Conditional Moments:

(a) $\sum_{\tau=1}^\infty \left\| \|m_h(Y_\tau; \mathcal{F}_0) - m_h(Y_\tau)\|_{2,\mathcal{F}_0} \right\|_{p/(p-1)} < \infty$.

(b) $\sum_{\tau=1}^\infty \|E\left[h(\boldsymbol{X}_\tau, Y_\tau) h(\boldsymbol{X}_0, Y_0)|\mathcal{F}_{\tau,Y}\right] - m_h(Y_\tau) m_h(Y_0)\|_{p/(p-1)} < \infty$.

(c) $\sum_{\tau=1}^{\infty} \|E[Y_0 Y_\tau | \mathcal{F}_{\tau,\boldsymbol{X}}] - g(\boldsymbol{X}_0) g(\boldsymbol{X}_\tau)\|_{p/(p-1)} < \infty$.

C5 Structure of the Space $\mathcal{X}$:

(a) $\mathcal{X}$ is a compact space.

(b) $\int_0^1 \exp\left(\frac{1}{2} H(x, \mathbb{Z}^+, \rho)\right) dx < \infty$.

C6 Equicontinuity: the p.d.f. $f(\boldsymbol{x})$ is bounded and equicontinuous; and the conditional expectation $g(\boldsymbol{x})$ is equicontinuous.

C7 Bandwidth: $k_T = O\left(T^{\frac{\ell+2}{2\ell}}\right)$ for some integer, $\ell \geq 3$.

Remark 3.2. Assumption C1 says that information is accumulated over time such that $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$. This assumption is satisfied, for example, in the nonlinear regression model, $Y_t = g(\boldsymbol{X}_t) + \epsilon_t$, where the disturbance, ϵ_t, is serially correlated in such a way that it is independent of $\boldsymbol{X}_t$ for all $t \in [1, T]$. Hence, it allows various degrees of nonlinear dependence between Y_t and $\boldsymbol{X}_t$. This formulation also constitutes a variety of stationary causal processes used in many realistic applications (see, e.g., Priestley, 1988 and Tong, 1990). Assumption C2 is a collection of some basic moment conditions. Assumption C3 states that the L_p distances between probability distributions and their corresponding conditional distributions should be minimum as the process moves further from its past. This is a natural extension of the dependence coefficient, $\alpha(\mathcal{M}, X)$ (based on the L_1 distance), which was introduced in Rio (2000, Eq. 1.10c).

Remark 3.3. Assumption C5 requires the subspace $\mathcal{X}$ to be compact in $\mathbb{R}^N$; and the $\mathbb{R}^N-$valued random process, $\boldsymbol{X}_t$, indexed by the set of nonnegative integers, $\mathbb{Z}^+$, satisfies a standard metric entropy condition . Hence, the convergence of the integral merely depends on the size of the covering numbers $N(\epsilon, \mathbb{Z}^+, \rho)$ for $\epsilon \longrightarrow 0$. Since $\int_0^1 \epsilon^{-k} d\epsilon < \infty$ for some $k < 1$, the integral condition C5b roughly entails that the entropy grows at a slower order than $-\log(\epsilon)$.

Remark 3.4. Having $\lim_{T \longrightarrow \infty} k_T / T = 0$ is standard in the pointwise asymptotic theory for k-NN, see e.g. Bhattacharya and Mack (1987). However, $k_T = O(T^{\frac{\ell+2}{2\ell}})$ implies that k_T diverges at a rate $T^{\frac{\ell+2}{2\ell}}$, which is much slower than other rates currently found in the literature, i.e. If $\ell = 3$, then $k = O(T^{5/6})$. For example, unlike Bhattacharya and Mack (1987) who showed that the weak convergence of k-NN density holds for $k_T = O(T^{4/5})$, the slow rate of divergence obtained here is because of an application of Peligrad et al.'s (2007) inequality in the proof of Theorem 3.2 below.

The following theorem states the asymptotic behavior of estimator (1.3):

Theorem 3.2. *Suppose Assumptions C1-C7 hold. Then, we have*

$$\sqrt{T}(\widehat{\theta} - \theta_0) \overset{W}{\Longrightarrow} \mathcal{N}(0, \sigma^2_{W^*}),$$

where $\sigma^2_{W^*} = E[W_0^{*2}] + 2\sum_{\tau=1}^{\infty} E[W_0^* W_\tau^*]$ *with* $W_t^* = J(F(Y_t))\{h(\boldsymbol{X}_t, Y_t) - m_h(Y_t)\} - \int_{\mathbb{R}} J(F(y))\{\mathbb{I}(Y_t \leq y) - F(y)\} dm_h(y)$.

3.3 Conditions for Mixing Processes

This section restates Assumption C4a, C4b and C4c in terms of the conditional dependence coefficients (see e.g. Dedecker and Prieur, 2007) and the conditional mixing coefficients (see e.g. Rao, 2009, for definition of this concept) of mixing processes.

Let $\Omega = \boldsymbol{\Omega}_{\boldsymbol{X}} \times \Omega_Y$ denote a 'sufficiently rich' probability space; and $(\boldsymbol{\omega}_1, \omega_2)$ denote elementary events in this probability space. First, we define marginal random sets: $A^*_\tau(y) = \{\boldsymbol{\omega}_1 \in \boldsymbol{\Omega}_{\boldsymbol{X}} : (\boldsymbol{X}_\tau(\boldsymbol{\omega}_1), Y_\tau(\omega_2)) \in A_\tau,$ where A_τ is some set in $\mathcal{F}_\tau$ and $Y_\tau(\omega_2) = y$ for a given y in $\mathbb{R}\}$, and $B^*_\tau(\boldsymbol{x}) = \{\omega_2 \in \Omega_Y : (\boldsymbol{X}_\tau(\boldsymbol{\omega}_1), Y_\tau(\omega_2)) \in A_\tau,$ where A_τ is some set in $\mathcal{F}_\tau$ and $\boldsymbol{X}_\tau(\boldsymbol{\omega}_1) = \boldsymbol{x}$ for a given $\boldsymbol{x}$ in $\mathcal{X}\}$. The *conditional* strong mixing coefficients are defined as follows:

$$
\begin{aligned}
\alpha^*_\tau(Y_0 = y_1, Y_\tau = y_2) &= \sup_{\substack{A^*_0(y_1) \in \mathcal{F}_{0,\boldsymbol{X}} \\ A^*_\tau(y_2) \in \mathcal{F}_{\tau,\boldsymbol{X}}}} \left| P(A^*_0(y_0) A^*_\tau(y_2)) - P(A^*_0(y_0)) P(A^*_\tau(y_2)) \right|, \\
\alpha^*_\tau(\boldsymbol{X}_0 = \boldsymbol{x}_1, \boldsymbol{X}_\tau = \boldsymbol{x}_2) &= \sup_{\substack{B^*_0(\boldsymbol{x}_1) \in \mathcal{F}_{0,Y} \\ B^*_\tau(\boldsymbol{x}_2) \in \mathcal{F}_{\tau,Y}}} \left| P(B^*_0(\boldsymbol{x}_1) B^*_\tau(\boldsymbol{x}_2)) - P(B^*_0(\boldsymbol{x}_1) B^*_\tau(\boldsymbol{x}_2)) \right|,
\end{aligned}
$$

for some y_1 and y_2 in $\mathbb{R}$ and some $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in $\mathcal{X}$, respectively. Note that the $\alpha^*_\tau(\cdot)$ used here and that used in Section 3.1 are different. In what follows, this conflict in notation will cause no difficulty because their meaning will be clear from the context they are employed.

Let us define the following conditional dependence coefficients:

$$
\begin{aligned}
\boldsymbol{\tau}(\mathcal{F}_{\boldsymbol{X}_0}, \boldsymbol{X}_\tau(\boldsymbol{\omega}_1) | Y_\tau(\omega_2) = y) &= \int_{\mathcal{X}} \left| F(\boldsymbol{x}|y, \boldsymbol{X}_0) - F(\boldsymbol{x}|y) \right| d\boldsymbol{x}, \\
\alpha(\mathcal{F}_{\boldsymbol{X}_0}, \boldsymbol{X}_\tau(\boldsymbol{\omega}_1) | Y_\tau(\omega_2) = y) &= \sup_{\boldsymbol{x} \in \mathcal{X}} \left| F(\boldsymbol{x}|y, \boldsymbol{X}_0) - F(\boldsymbol{x}|y) \right|,
\end{aligned}
$$

where $F(\boldsymbol{x}|y, \boldsymbol{X}_0) = P(\boldsymbol{X}_\tau(\boldsymbol{\omega}_1) \leq \boldsymbol{x} | Y_\tau(\omega_2) = y, \boldsymbol{X}_0)$ and $F(\boldsymbol{x}|y) = P(\boldsymbol{X}_\tau(\boldsymbol{\omega}_1) \leq \boldsymbol{x} | Y_\tau(\omega_2) = y)$. The above coefficients are the conditional analogues of the random dependence coefficients: $\alpha(\mathcal{M}, X)$ introduced by Rio (2000, Eq. 1.10c), and $\tau(\mathcal{M}, X)$, introduced by Dedecker and Prieur (2005). These coefficients are weaker than the corresponding mixing coefficients and can be computed in many situations.

Lemma 3.3. *Let q, r, p_1, p_2, and p_3 denote generic constants that may differ from one context to another.*

(1) Suppose that

$$
\max\left\{ \left\| \left\| \sup_{\boldsymbol{x} \in \mathcal{X}} \frac{\partial^N}{\partial x_1 \dots \partial x_N} h(\boldsymbol{x}, Y_\tau) \right\|_{4,\mathcal{F}_0} \right\|_{p_1}, \sum_{d=0}^{N-1} \sum_{j_1 + \dots + j_N = d} \left\| \left\| \sup_{\boldsymbol{x} \in \mathcal{X}} \frac{\partial^d}{\partial x_1^{j_1} \dots \partial x_N^{j_N}} h(\boldsymbol{x}, Y_\tau) \right\|_{4,\mathcal{F}_0} \right\|_{p_2} \right\} < \infty,
$$

$$
\sum_{\tau=1}^{\infty} \left\| \left\| \alpha\left(\mathcal{F}_{\boldsymbol{X}_0}, \boldsymbol{X}_\tau | Y_\tau\right) \right\|_{4,\mathcal{F}_0} \right\|_{p_1} < \infty,
$$

and

$$
\sum_{\tau=1}^{\infty} \left\| \left\| \boldsymbol{\tau}\left(\mathcal{F}_{\boldsymbol{X}_0}, \boldsymbol{X}_\tau | Y_\tau\right) \right\|_{4,\mathcal{F}_0} \right\|_{p_2} < \infty,
$$

where $p_1,\ p_2 > 1$ such that $1/p_1 + 1/p_2 = (p-1)/p$. Then Assumption C4a holds.

(2) Suppose that

$$\max\left\{\left\|\|h(\boldsymbol{X}_0,Y_0)\|_{q,\mathcal{F}_{0,Y}}\right\|_{p_1},\left\|\|h(\boldsymbol{X}_0,Y_0)\|_{r,\mathcal{F}_{0,Y}}\right\|_{p_2}\right\}<\infty$$

and

$$\sum_{\tau=1}^{\infty}\left\|\{\alpha_\tau^*(Y_0,Y_\tau)\}^{1-q^{-1}-r^{-1}}\right\|_{p_3}<\infty,$$

where q, r, p_1, p_2, and p_3 are some integers satisfying $1/q+1/r<1$ and $1/p_3=(p-1)/p-1/p_1-1/p_2$. Then Assumption C4b holds.

(3) Suppose that

$$\max\left\{\left\|\|Y_0\|_{q,\mathcal{F}_{0,\boldsymbol{X}}}\right\|_{p_1},\left\|\|Y_0\|_{r,\mathcal{F}_{0,\boldsymbol{X}}}\right\|_{p_2}\right\}<\infty$$

and

$$\sum_{\tau=1}^{\infty}\left\|\{\alpha_\tau^*(\boldsymbol{X}_0,\boldsymbol{X}_\tau)\}^{1-q^{-1}-r^{-1}}\right\|_{p_3}<\infty,$$

where q, r, p_1, p_2, and p_3 are some integers satisfying $1/q+1/r<1$ and $1/p_3=(p-1)/p-1/p_1-1/p_2$. Then Assumption C4c holds.

4 Numerical Results

4.1 Monte Carlo Simulation Study

This section examines how well the asymptotic approximations established in Theorems 3.1 and 3.2 perform in small samples. We use the d.g.p. of Example 1 in Section 2.2, i.e. $Y_t = X_t\epsilon_t$, in each of 2000 replications. The $\{\epsilon_t\}_{t=1}^T$ are i.i.d. standardize samples (multiplied by -1) from a gamma distribution with shape parameter equal to 12 and scale parameter equal to 1/2. The $\{X_t\}_{t=1}^T$ are generated from Moving Average (MA), Bilinear (BILINEAR) and Generalized Autoregressive Conditional Heteroskedasticity (1,1) (GARCH) models where $\{\xi_t\}_{t=1}^T$ are i.i.d. samples from a mixtures of skewed normal distributions as follows: With probability $1/(1+4\sqrt{2/17})$ they come from a skewed normal with location parameter equals 0, scale parameter equals 1, and shape parameter equals -4, and with probability $4\sqrt{2/17}/(1+4\sqrt{2/17})$ they come from another independent skewed normal with location parameter equals 0, scale parameter equals 1, and shape parameter equals -1. The other parameters are chosen as: (MA) $\theta_0 = 0.7$, $\theta_1 = 0.3$, and $\theta_j = 0$ for $j = 2, 3, \ldots$; (BILINEAR) $a = -0.5$, and $b = 0.5$; (GARCH) $\omega = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$. We set $T \in \{100, 200, 400\}$.

As for Theorem 3.1, we set $h(x, y) = xy$ and $J(u) = \phi(u)$, where $\phi(\cdot)$ represents the probability density function of a standard normal random variable. We call this Case 1 and it is such that $\mathfrak{T}(F) = 0$. Similarly, for Theorem 3.2, we set $J(u) = 1$ and choose k in (3.7) by standard cross-validation, see e.g. Hart and Vieu (1990). This is called Case 2 and θ_0 in (1.3) equals 0 as well.

Table 1, Figures 1 and 2 summarize the results. The figures show the QQ-Plots of the simulated samples for Cases 1 and 2 standardize by their Monte Carlos mean and standard deviations. As to assess the quality of the asymptotic normal approximation in small samples, the 45-degrees line is also included. For the (MA) process, the asymptotic normal approximation is very accurate for all sample sizes and in all cases. For the nonlinear processes (BILINEAR) and (GARCH) the normal approximation at the tails is better for a sample size of 400 than it is for 100 observations. In all models, as suggested by our results, the small sample tail behavior of the estimators becomes closer to that of a normal as the sample sizes increases.

From the results of Table 1, one observes that the proposed estimators work very well in terms of Monte Carlo bias (Bias), standard deviation (Std. Dev.) and inter-quartile range (IQR). In general, the proposed estimator of $\mathfrak{T}(F) = 0$ in Case 1 is generally unbiased for the (MA) and (GARCH) processes, and in Case 2, the bias tends to decrease with the sample size quite rapidly. In all the cases, as predicted by Theorems 3.1 and 3.2, the simulated standard deviations and inter-quartile ranges decrease when the sample size increases.

4.2 Empirical Example: Consumer Surplus Estimation

In a classic paper, Engle et al. (1986) used a partial linear model to study the impact of weather and other variables on electricity demand. Since fully nonparametric models are becoming popular in Economics, see e.g. Huynh and Jacho-Chávez (2009), we use this framework to illustrate the utility of our estimator by applying our methodology to estimate consumer surpluses based on a fully nonparametrically estimated demand function for electricity in the Canadian province of Ontario. In particular, we calculate monetary gains obtained by consumers when facing lower prices than the one they are willing to pay - this is known as consumer surplus in Economics.

The data for this analysis come from the Ontario Hydro Corporation and is made publicly available by Yatchew (2003). The data consist of 288 quarterly observations in Ontario for the period 1971 to 1994 of the following variables: elec_t - log of monthly electricity sales in millions of Canadian dollars, temp_t - heating and cooling degree days relative to 68°F, relprice_t - log of ratio of price of electricity to the price of natural gas, and gdp_t - log of Ontario gross domestic product in millions of Canadian dollars. Notice that temp_t is the difference between the number of days the temperature is below 68°F (20°C or room temperature) and number of days the temperature is above 68°F. If the net cooling days is negative it implies that the monthly temperature is colder than 68°F while positive is that it is positive (more hotter) days.

We set $Y_t := \text{elec}_t - \text{gdp}_t$ - this normalization is suggested by Yatchew (2003, Chapter 4) to enforce a cointegration relationship, and $\boldsymbol{X}_t := [\text{temp}_t, \text{relprice}_t]$ in (1.3). We then estimate a version of θ_0 at different levels of temp_t. In particular, our calculation asks what would happen if the relative prices of electricity versus natural gas in 1994 (fourth quarter) were set to 1989 prices (fourth quarter). During that period, electricity prices increased by 22.6 percent while gas increased by 9.2 percent for a net increase of 13.4 percent. For illustrative purposes, we compute the consumer surplus for the three levels of temp_t: 1) median cold month (-397), 2) zero, and 3) median hot month (94). Our estimates utilize $k = [83, 2]$ found via cross-validation as suggested by Hart and Vieu (1990). Overall, the largest consumer surplus gain is for cold months with 62.7 million normalized dollars. When temp_t tends to be hotter (94) the consumer surplus is 57.8 million. If temp_t equals zero, the consumer surplus is 53.9

million. Overall, these estimates are reasonable, since consumer surplus gains are expected to be higher for households when the temperature is extremely cold or hot.

Acknowledgements

We would like to thank the Editor, Associate Editor and two anonymous referees for many helpful comments and suggestions. We also thank Juan Carlos Escanciano, Ingrid Van Keilegom, and Lanh Tran for their insightful and useful comments and suggestions in earlier drafts of this paper. We acknowledge the usage of the 'np' package by Hayfield and Racine (2008), and the assistance of Takuya Noguchi in implementing the Monte Carlo experiments in the Quarry High Performance Clusters at Indiana University. We gratefully acknowledge support by the Social Sciences and Humanities Research Council of Canada grant (SSHRC Grant 410-2011-1700). The views expressed in this article are those of the authors. No responsibility for them should be attributed to the Bank of Canada. All remaining errors are the responsibility of the authors.

5 Proofs of Theorems

For brevity the proofs of the theorems are presented in concise format while all auxiliary results are available in Appendix A. Similarly, Appendix B restates various known results in the literature that are used in our proofs for the paper to be self-contained. The notation 'Const.' refers to any generic positive constant that may take different values for each appearance.

5.1 Proof of Theorem 3.1

The proof proceeds in three steps:

Step 1: Let $F_\epsilon(\cdot|\mathcal{I}) = F(\cdot|\mathcal{I}) + \epsilon(F_T - F(\cdot|\mathcal{I}))$ denote a ϵ-perturbation of $F(\cdot|\mathcal{I})$ in the direction $(F_T - F(\cdot|\mathcal{I}))$. Then,

$$\phi(\epsilon) = \mathfrak{T}(F_\epsilon) = \int_{\mathbb{R}^{N+1}} J(F_\epsilon(y))h(\boldsymbol{x}, y)dF_\epsilon(\boldsymbol{x}, y).$$

Since the Gâteau derivative of $\mathfrak{T}(F_\epsilon)$ in the direction $F_T - F$ is defined as the right derivative of $\phi(\epsilon)$ at 0, we obtain

$$\begin{aligned}
& \mathfrak{T}'(F_T - F(\cdot|\mathcal{I})) = \phi'(0^+), \\
= & \int_{\mathbb{R}^{N+1}} \{F_T(y) - F(y|\mathcal{I})\}J'(F(y|\mathcal{I}))h(\boldsymbol{x}, y)dF(\boldsymbol{x}, y|\mathcal{I}) + \int_{\mathbb{R}^{N+1}} J(F(y|\mathcal{I}))h(\boldsymbol{x}, y)d\{F_T(\boldsymbol{x}, y) - F(\boldsymbol{x}, y|\mathcal{I})\} \\
= & \int_{\mathbb{R}^{N+1}} \{F_T(y) - F(y|\mathcal{I})\}J'(F(y|\mathcal{I}))h(\boldsymbol{x}, y)f(\boldsymbol{x}|Y = y, \mathcal{I})f(y|\mathcal{I})d\boldsymbol{x}dy \\
+ & \int_{\mathbb{R}^{N+1}} J(F(y|\mathcal{I}))h(\boldsymbol{x}, y)d\{F_T(\boldsymbol{x}, y) - F(\boldsymbol{x}, y|\mathcal{I})\}, \\
= & \int_{\mathbb{R}} \{F_T(y) - F(y|\mathcal{I})\}J'(F(y|\mathcal{I}))m_h(y; \mathcal{I})dF(y|\mathcal{I}) + \int_{\mathbb{R}^{N+1}} J(F(y|\mathcal{I}))h(\boldsymbol{x}, y)d\{F_T(\boldsymbol{x}, y) - F(\boldsymbol{x}, y|\mathcal{I})\}, \\
= & \int_{\mathbb{R}} \{F_T(y) - F(y|\mathcal{I})\}m_h(y; \mathcal{I})dJ(F(y|\mathcal{I})) + \int_{\mathbb{R}^{N+1}} J\left(F(y|\mathcal{I})\right)h(\boldsymbol{x}, y)d\{F_T(\boldsymbol{x}, y) - F(\boldsymbol{x}, y|\mathcal{I})\},
\end{aligned}$$

where $m_h(y;\mathcal{I}) = E[h(\boldsymbol{X},Y)|Y=y,\mathcal{I}]$. Integration by parts yields $\int_{\mathbb{R}}\{F_T(y) - F(y|\mathcal{I})\}m_h(y;\mathcal{I})dJ(F(y|\mathcal{I})) = -\int_{\mathbb{R}} J(F(y|\mathcal{I}))m_h(y;\mathcal{I})d\{F_T(y) - F(y|\mathcal{I})\} - \int_{\mathbb{R}} J(F(y|\mathcal{I}))\{F_T(y) - F(y|\mathcal{I})\}dm_h(y|\mathcal{I})$. It follows that

$$
\begin{aligned}
\mathfrak{T}^{'}(F_T - F(\cdot|\mathcal{I})) &= \int_{\mathbb{R}^{N+1}} J(F(y|\mathcal{I}))h(\boldsymbol{x},y)dF_T(\boldsymbol{x},y) - \int_{\mathbb{R}} J(F(y|\mathcal{I}))m_h(y;\mathcal{I})dF_T(y) \\
&- \int_{\mathbb{R}} J(F(y|\mathcal{I}))\{F_T(y) - F(y|\mathcal{I})\}dm_h(y|\mathcal{I}), \\
&= \frac{1}{T}\left\{\sum_{t=1}^{T} J(F(Y_t|\mathcal{I}))\{h(\boldsymbol{X}_t,Y_t) - m_h(Y_t|\mathcal{I})\}\right. \\
&- \left.\sum_{t=1}^{T}\int_{\mathbb{R}} J(F(y|\mathcal{I}))\{\mathbb{I}(Y_t \leq y) - F(y|\mathcal{I})\}dm_h(y|\mathcal{I})\right\}.
\end{aligned}
$$

Hence, by the Taylor formula: $\phi(1) = \phi(0) + \sum_{k=1}^{\ell-1} k!^{-1}\phi^{(k)}(u)\big\|_{u=0^+} + \ell!^{-1}\phi^{(\ell-1)}(v)$, where $v \in [0,1]$, we obtain:

$$
\mathfrak{T}(F_T) - \mathfrak{T}(F(\cdot|\mathcal{I})) = \mathfrak{T}^{'}(F_T - F(\cdot|\mathcal{I})) + \mathcal{R}_T, \tag{5.1}
$$

where $\mathcal{R}_T$ is the remainder of the above expansion.

Step 2: We are now ready to prove the weak convergence of $\mathfrak{T}^{'}(F_T - F(\cdot|\mathcal{I}))$. An application of the martingale approximation for the process $\sum_1^T W_t$ yields

$$
\sum_1^T W_t = \sum_1^T W_t^* + \widetilde{W}_1 - \widetilde{W}_{T+1}, \tag{5.2}
$$

where W_t^* is defined from Lemma A.4; and $\widetilde{W}_t = \sum_{s=1}^{\infty} E[W_s|\mathcal{F}_{t-1}]$. The asymptotic behavior of $T^{-1/2}\sum_1^T W_t$ is determined by those of two terms, $T^{-1/2}\sum_1^T W_t^*$ and $T^{-1/2}(\widetilde{W}_1 - \widetilde{W}_{T+1})$.

To derive the asymptotic behavior of $T^{-1/2}\sum_1^T W_t^*$, note that W_t^* is a martingale difference sequence. Thus a Central Limit Theorem for martingale difference sequences (see e.g. Chow and Teicher, 1978, p. 336) can be applied; and the variance of the above sum is just $E\left[W_1^{*2}|\mathcal{I}\right]$ given in Lemma A.4.

Under Assumptions A1, A2 and A3a, the variance $E[W_1^{*2}|\mathcal{I}]$ can easily be derived by sequentially applying Lemmas A.2-A.4. To finish with deriving the asymptotic normality of $T^{-1/2}\sum_1^T W_t^*$, we need to check the Lindeberg condition: For an arbitrarily small constant, δ, it follows that

$$
\begin{aligned}
&\frac{1}{T}\sum_1^T E\left[E\left[W_t^{*2}\mathbf{1}(\frac{W_t^*}{\sqrt{T}} > \delta)\middle|\mathcal{F}_{t-1}\right]\middle|\mathcal{I}\right] = E[W_t^{*2}\mathbf{1}(W_t^* > \delta\sqrt{T})|\mathcal{I}], \\
\leq\ & \|W_0^2\|_{p,\mathcal{I}}P(W_t^* \geq \delta\sqrt{T}|\mathcal{I}) \overset{p}{\Longrightarrow} 0,
\end{aligned}
$$

where the last inequality follows from Hölder's inequality; and the limit in probability follows from the Tchebyshev inequality and Assumptions A2a and A3a. Hence, we obtain

$$
\frac{1}{\sqrt{T}}\sum_1^T W_t^* \overset{W}{\Longrightarrow} N(0,\sigma_W^2). \tag{5.3}
$$

To derive the asymptotic behavior of $T^{-1/2}(\widetilde{W}_1 - \widetilde{W}_{T+1})$, note that W_t is $\mathcal{F}_t$-measurable. Given some generic constant, $\delta > 0$, under Assumptions A2 and A3a we obtain

$$P\left(\left|\frac{\widetilde{W}_1 - \widetilde{W}_{T+1}}{\sqrt{T}}\right| \geq \delta\right) \leq 2P\left(\frac{|\widetilde{W}_1|}{\sqrt{T}} \geq \frac{\delta}{2}\right) \leq \frac{\|\widetilde{W}_1\|_p}{\delta\sqrt{T}} \longrightarrow 0 \text{ as } T \longrightarrow \infty. \tag{5.4}$$

From Eqs. (5.3) and (5.4), we obtain

$$\sqrt{T}\mathfrak{T}^{'}\left(F_T - F(\cdot|\mathcal{I})\right) \overset{W}{\Longrightarrow} N(0, \sigma_W^2). \tag{5.5}$$

Step 3: We conclude the proof by studying the limiting behavior of the remainder term, $\mathcal{R}_T$, from the Gâteau expansion, i.e.

$$\begin{aligned}
\mathcal{R}_T &= \int_{\mathbb{R}^{N+1}} h(\boldsymbol{x}, y)\{J(F_T(y)) - J(F(y|\mathcal{I}))\}dF_T(\boldsymbol{x}, y) + \int_{\mathbb{R}} m_h(y;\mathcal{I})d\{K(F_T(y)) - K(F(y|\mathcal{I}))\} \\
&+ \int_{\mathbb{R}} J(F(y|\mathcal{I}))\{F_T(y) - F(y|\mathcal{I})\}dm_h(y|\mathcal{I}),
\end{aligned}$$

where $K(u) = \int_0^u J(v)dv$. Some basic algebra yield

$$\begin{aligned}
&\mathcal{R}_T \\
= &\int_{\mathbb{R}^{N+1}} h(\boldsymbol{x}, y)\left\{\frac{J(F_T(y)) - J(F(y|\mathcal{I}))}{F_T(y) - F(y|\mathcal{I})} - J^{'}\left(F(y|\mathcal{I})\right)\right\}\left\{F_T(y) - F(y|\mathcal{I})\right\}dF_T(\boldsymbol{x}, y) \\
- &\int_{\mathbb{R}} \left\{\frac{K(F_T(y)) - K(F(y|\mathcal{I}))}{F_T(y) - F(y|\mathcal{I})} - J(F(y|\mathcal{I}))\right\}\left\{F_T(y) - F(y|\mathcal{I})\right\}dm_h(y|\mathcal{I}) \\
- &\int_{\mathbb{R}} \left\{\frac{J(F_T(y)) - J(F(y|\mathcal{I}))}{F_T(y) - F(y|\mathcal{I})} - J^{'}\left(F(y|\mathcal{I})\right)\right\}\{F_T(y) - F(y|\mathcal{I})\}m_h(y;\mathcal{I})dF_T(y) \\
+ &\left\{\int_{\mathbb{R}^{N+1}} J^{'}\left(F(y|\mathcal{I})\right)\left\{F_T(y) - F(y|\mathcal{I})\right\}h(\boldsymbol{x}, y)dF_T(\boldsymbol{x}, y)\right. \\
- &\left.\int_{\mathbb{R}} J^{'}\left(F(y|\mathcal{I})\right)\left\{F_T(y) - F(y|\mathcal{I})\right\}m_h(y;\mathcal{I})dF_T(y)\right\}, \\
= &\mathcal{R}_{\rm a} - \mathcal{R}_{\rm b} - \mathcal{R}_{\rm c} + \mathcal{R}_{\rm d},
\end{aligned}$$

where the definitions of $\mathcal{R}_{\rm a}$, $\mathcal{R}_{\rm b}$, $\mathcal{R}_{\rm c}$ and $\mathcal{R}_{\rm d}$ should be apparent. We now bound each term in the last equality as follows:

Terms $\mathcal{R}_{\rm a}$, $\mathcal{R}_{\rm b}$, $\mathcal{R}_{\rm c}$: Using an absolute value inequality, we have

$$\begin{aligned}
\sqrt{T}\mathcal{R}_{\rm a} &\leq \sup_{1 \leq t \leq T} \sqrt{T}|F_T(Y_t) - F(Y_t|\mathcal{I})| \\
&\times \sup_{1 \leq t \leq T}\left|\frac{J(F_T(Y_t)) - J(F(Y_t|\mathcal{I}))}{F_T(Y_t) - F(Y_t|\mathcal{I})} - J^{'}\left(F(Y_t|\mathcal{I})\right)\right|\frac{1}{T}\sum_{t=1}^{T}|h(\boldsymbol{X}_t, Y_t)|.
\end{aligned}$$

The Birkhoff-Khintchine theorem (see e.g. Varadhan, 2001, p. 132) yields, under Assumption A1,

$$\frac{1}{T}\sum_{t=1}^{T}|h(\boldsymbol{X}_t, Y_t)| \Longrightarrow E[|h(\boldsymbol{X}_t, Y_t)|\,|\mathcal{I}] \text{ P-a.s.}$$

Moreover, by the same arguments used in the proof of Step 2, we can prove that Assumption A2b implies that $\overline{\lim}_{T\longrightarrow\infty}\sup_{y\in\mathbb{R}}\sum_{s=1}^{T}\|P(Y_s\le y|\mathcal{F}_{Y_0})-P(Y_s\le y|\mathcal{I})\|_{p/(p-1)}<\infty$. Hence, it immediately follows that $\sup_{1\le t\le T}\sqrt{T}|F_T(Y_t)-F(Y_t|\mathcal{I})|=O_p(1)$ by Skorokhod representation theorem. Since the Birkhoff-Khintchine theorem implies that $|F_T(Y_t)-F(Y_t|\mathcal{I})|=o_{a.s.}(1)$ for every $t\in\{1,\dots,T\}$, an application of the stochastic mean-value theorem (see, e.g., White and Domowitz (1984)) yields

$$\sup_{1\le t\le T}\left|\frac{J(F_T(Y_t))-J(F(Y_t|\mathcal{I}))}{F_T(Y_t)-F(Y_t|\mathcal{I})}-J^{'}(F(Y_t|\mathcal{I}))\right|=o_{a.s.}(1).$$

Therefore, we obtain $\sqrt{T}\mathcal{R}_{\rm a}=o_p(1)$. Using a similar argument and the Lebesgue dominated convergence theorem, we can verify that $\sqrt{T}\mathcal{R}_{\rm b}=o_p(1)$ and $\sqrt{T}\mathcal{R}_{\rm c}=o_p(1)$.

Term $\mathcal{R}_{\rm d}$: Firstly, notice that

$$\begin{aligned}
&E[\left(\sqrt{T}\mathcal{R}_{\rm d}\right)^2]=E\left[J^{'2}(F(Y_t|\mathcal{I}))\{F_T(Y_t)-F(Y_t|\mathcal{I})\}^2\{h(\boldsymbol{X}_t,Y_t)-m_h(Y_t|\mathcal{I})\}^2\right]\\
+\;&\frac{1}{T}\sum_{\substack{t=1\\s=1\\t\ne s}}^{T}E\left[J^{'}(F(Y_t|\mathcal{I}))J^{'}(F(Y_s|\mathcal{I}))\{F_T(Y_t)-F(Y_t|\mathcal{I})\}\{F_T(Y_s)-F(Y_s|\mathcal{I})\}\right.\\
&\left.\times\{h(\boldsymbol{X}_t,Y_t)-m_h(Y_t|\mathcal{I})\}\{h(\boldsymbol{X}_s,Y_s)-m_h(Y_s|\mathcal{I})\}\right]=\mathcal{R}_{\rm d;1}+\mathcal{R}_{\rm d;2}.
\end{aligned}$$

Using the fact that $J^{'}(F)$ is bounded and Hölder's inequality, we obtain $\mathcal{R}_{\rm d;1}\le$ Const. $\times\left\|\{F_T(Y_t)-F(Y_t|\mathcal{I})\}^2\right\|_{p/(p-1)}\left\|\{h(\boldsymbol{X}_t,Y_t)-m_h(Y_t|\mathcal{I})\}^2\right\|_p$. In view of the Birkhoff-Khintchine theorem and the dominated convergence theorem, we obtain $\left\|\{F_T(Y_t)-F(Y_t|\mathcal{I})\}^2\right\|_p=o_{a.s.}(1)$. Hence, under Assumption A1, we have

$$\left\|\{h(\boldsymbol{X}_t,Y_t)-m_h(Y_t|\mathcal{I})\}^2\right\|_p<\infty.$$

Then, $\mathcal{R}_{\rm d;1}=o_{a.s.}(1)$.

In addition, we have

$$\begin{aligned}
\mathcal{R}_{\rm d;2}\;\le\;&\text{Const.}\times\|\{F_T(Y_t)-F(Y_t|\mathcal{I})\}\|_{2p}^2\times\\
&\sum_{\tau=1}^{T}\left(1-\frac{\tau}{T}\right)\|E[h(\boldsymbol{X}_\tau,Y_\tau)h(\boldsymbol{X}_0,Y_0)|\mathcal{F}_{\tau,Y},\mathcal{I}]-m_h(Y_\tau)m_h(Y_0)\|_{p/(p-1)}\text{, where}
\end{aligned}$$

$\|\{F_T(Y_t)-F(Y_t|\mathcal{I})\}\|_{2p}^2=o_{a.s.}(1)$. Thus, under Assumption A3b, we can deduce that $\mathcal{R}_{\rm d;2}=o(1)$.

5.2 Proof of Corollary 3.1

Assumption B1 essentially implies Assumption A1. Assumption B2 implies Assumption A2a while Assumption A2b naturally follows from the ergodicity of the processes under study.

In light of Lemma A.5, Assumption B3a implies that $\sum_{t=1}^{\infty}\mathcal{K}_t(p)<\infty$, where the term $\mathcal{K}_t(p)$ first appeared in Lemma A.2. By working through *Step 2* in the proof of Theorem 3.1, in view of Assumption A1 one can derive the limit for the term $\mathfrak{T}^{'}(F_T-F)$. To bound the remaining term, $\mathcal{R}_T$, in *Step 1* in the proof of Theorem 3.1, one needs Assumptions A1 and A3b. Assumption A3b is verified as follows:

$$
\begin{aligned}
&E\left[h(\boldsymbol{X}_\tau, Y_\tau)h(\boldsymbol{X}_0, Y_0)|\mathcal{F}_{\tau,Y}\right] - m_h(Y_\tau)m_h(Y_0)\\
&\qquad = E\left[\{h(\boldsymbol{X}_0, Y_0) - m_h(Y_0)\}E\left[h(\boldsymbol{X}_\tau, Y_\tau) - m_h(Y_\tau)|\mathcal{F}_0, \mathcal{F}_{\tau,Y}\right]\middle|\mathcal{F}_{\tau,Y}\right]\\
&\qquad\qquad \leq \|h(\boldsymbol{X}_0, Y_0) - m_h(Y_0)\|_{2,\mathcal{F}_{0,Y}} \left\|E\left[h(\boldsymbol{X}_\tau, Y_\tau) - m_h(Y_\tau)\middle|\mathcal{F}_{0,\boldsymbol{X}}, \mathcal{F}_{\tau,Y}\right]\right\|_{2,\mathcal{F}_{\tau,Y}}.
\end{aligned}
$$

An application of Hölder's inequality yields

$$
\begin{aligned}
&\left\|E\left[h(\boldsymbol{X}_\tau, Y_\tau) \;-\; m_h(Y_\tau)\middle|\mathcal{F}_{0,\boldsymbol{X}}, \mathcal{F}_{\tau,Y}\right]\right\|_{2,\mathcal{F}_{\tau,Y}}\\
=\;& \left\|E\left[\widetilde{h}_1(\boldsymbol{\xi}^{\circ}_{\tau-1}|\xi_{N+1,\tau})|\boldsymbol{\xi}^{\circ}_0, \xi_{N+1,\tau}\right] - E\left[\widetilde{h}_1(\boldsymbol{\xi}^{\circ *}_{\tau-1}|\xi_{N+1,\tau})|\boldsymbol{\xi}^{\circ}_0, \xi_{N+1,\tau}\right]\right\|_{2,\mathcal{F}_{\tau,Y}}\\
\leq\;& \left\|\widetilde{h}_1(\boldsymbol{\xi}^{\circ}_{\tau-1}|\xi_{N+1,\tau}) - \widetilde{h}_1(\boldsymbol{\xi}^{\circ *}_{\tau-1}|\xi_{N+1,\tau})\right\|_{2,\mathcal{F}_{\tau,Y}}\\
=\;& \alpha^*_{\tau-1}(\xi_{N+1,\tau}),
\end{aligned}
$$

where $\boldsymbol{\xi}^{\circ *}_\tau$ is $\boldsymbol{\xi}^{\circ}_\tau$ with the whole past $\boldsymbol{\xi}^{\circ}_0$ replaced by an i.i.d. copy $\boldsymbol{\xi}^{\circ'}_0$. Another application of Hölder's inequality yields

$$
\begin{aligned}
&\left\|E\left[h(\boldsymbol{X}_\tau, Y_\tau)h(\boldsymbol{X}_0, Y_0)|\mathcal{F}_{\tau,Y}\right] - m_h(Y_\tau)m_h(Y_0)\right\|_{p/(p-1)}\\
\leq\;& \left\|\|h(\boldsymbol{X}_0, Y_0) - m_h(Y_0)\|_{2,\mathcal{F}_{0,Y}}\right\|_{p/(p-1)} \left\|\alpha^*_{\tau-1}(\xi_{N+1,\tau})\right\|_{p/(p-1)}.
\end{aligned}
$$

Hence, it immediately follows that $\sum_{\tau=1}^{\infty}\left\|\alpha^*_{\tau-1}(\xi_{N+1,\tau})\right\|_{p/(p-1)} < \infty$ implies Assumption A3b.

5.3 Proof of Theorem 3.2

Some algebra yields the following expansion:

$$
\begin{aligned}
\sqrt{T}(\widehat{\theta} - \theta_0) \;=\;& \frac{1}{\sqrt{T}}\sum_{t=1}^{T} J(t/T)\frac{Y_{(t)} - g(\boldsymbol{X}_{[t]})}{f(\boldsymbol{X}_{[t]})}\\
&+\; \frac{1}{\sqrt{T}}\sum_{t=1}^{T} J(t/T)\left(Y_{(t)} - g(\boldsymbol{X}_{[t]})\right)\left\{\frac{T\mathcal{L}(V_1)R^N_T(\boldsymbol{X}_{[t]}, k)}{k} - \frac{1}{f(\boldsymbol{X}_{[t]})}\right\}\\
&+\; \sqrt{T}\left\{\sum_{t=1}^{T} J(t/T)\frac{\mathcal{L}(V_1)g(\boldsymbol{X}_{[t]})R^N_T(\boldsymbol{X}_{[t]}, k)}{k} - \int_{\boldsymbol{\mathcal{X}}} g(\boldsymbol{x})d\boldsymbol{x}\right\},\\
=\;& \mathcal{T}_{1T} + \mathcal{T}_{2T} + \mathcal{T}_{3T},
\end{aligned}
$$

where the definitions of $\mathcal{T}_{lT}$ $(l = 1, 2, 3)$ should be apparent. These three components are analyzed in each of the following 3 steps:

Step 1: Since Assumptions C2, C3b, C4a, and C4b imply Assumptions A1, A2, and A3. An application of Theorem 3.1 yields

$$\mathcal{T}_{1T} \overset{W}{\Longrightarrow} \mathcal{N}(0, \sigma^2_{W^*}).$$

The proof concludes by showing that $\mathcal{T}_{2T} = o_p(1)$ and $\mathcal{T}_{3T} = o_p(1)$.

Step 2: Since $E[Y_{(t)} - g(\boldsymbol{X}_{[t]})|\mathcal{F}_{T,\boldsymbol{X}}] = 0$, we can show that

$$\begin{aligned}
E[\mathcal{T}_{2T}] &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} J(t/T)E\left[\left\{Y_{(t)} - g(\boldsymbol{X}_{[t]})\right\}\left\{\frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_{[t]}, k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_{[t]})}\right\}\right] \\
&= E\left[\left\{\frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_{[t]}, k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_{[t]})}\right\} E[Y_{(t)} - g(\boldsymbol{X}_{[t]})|\mathcal{F}_{T,\boldsymbol{X}}]\right] \frac{1}{\sqrt{T}}\sum_{t=1}^{T} J(t/T) \\
&= 0.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
E[\mathcal{T}_{2T}^2] &= \frac{1}{T}\sum_{t=1}^{T} J^2(t/T)E\left[\{Y_{(t)} - g(\boldsymbol{X}_{[t]})\}^2\left\{\frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_{[t]}, k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_{[t]})}\right\}^2\right] \\
&+ \frac{1}{T}\sum_{\substack{s=1\\t=1\\s\neq t}}^{T} J(t/T)J(s/T)E\left[\left\{Y_{(t)} - g(\boldsymbol{X}_{[t]})\right\}\left\{Y_{(s)} - g(\boldsymbol{X}_{[s]})\right\}\right. \\
&\quad \left.\times\left\{\frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_{[t]}, k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_{[t]})}\right\}\left\{\frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_{[s]}, k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_{[s]})}\right\}\right] \\
&= \mathcal{T}_{2T;1} + \mathcal{T}_{2T;2}.
\end{aligned}$$

Using the Cauchy–Schwarz inequality and the fact that the score function $J(\cdot)$ is bounded, we have

$$\mathcal{T}_{2T;1} \leq \text{Const.} \times \left\|\{Y_t - g(\boldsymbol{X}_t)\}^2\right\|_2 \left\|\left\{\frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_t, k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_t)}\right\}^2\right\|_2,$$

and, by virtue of the stationarity of $\{\boldsymbol{X}_t, Y_t\}$,

$$\begin{aligned}
\mathcal{T}_{2T;2} \leq\ & \text{Const.} \\
& \times\frac{1}{T}\sum_{\substack{s=1\\t=1\\s\neq t}}^{T} \left\|\left\{\frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_t, k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_t)}\right\}\left\{\frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_s, k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_s)}\right\}\right\|_p \\
& \|E[Y_tY_s|\boldsymbol{X}_t, \boldsymbol{X}_s] - g(\boldsymbol{X}_t)g(\boldsymbol{X}_s)\|_{p/(p-1)}
\end{aligned}$$

$$
\begin{aligned}
\le\ & \text{Const.} \times \sup_{1\le t\le T} \left\| \left\{ \frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_t,k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_t)} \right\} \left\{ \frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_s,k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_s)} \right\} \right\|_p \\
& \times \frac{1}{T} \sum_{\substack{s=1\\t=1\\s\neq t}}^{T} \| E[Y_tY_s|\boldsymbol{X}_t,\boldsymbol{X}_s] - g(\boldsymbol{X}_t)g(\boldsymbol{X}_s)\|_{p/(p-1)} \\
=\ & \text{Const.} \times \sup_{1\le t\le T} \left\| \left\{ \frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_t,k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_t)} \right\} \left\{ \frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_s,k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_s)} \right\} \right\|_p \\
& \times 2 \sum_{\tau=1}^{T} \left(1 - \frac{\tau}{T}\right) \| E[Y_\tau Y_0|\mathcal{F}_{\tau,\boldsymbol{X}}] - g(\boldsymbol{X}_\tau)g(\boldsymbol{X}_0)\|_{p/(p-1)} \\
\le\ & \text{Const.} \times \sup_{1\le t\le T} \left\| \left\{ \frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_t,k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_t)} \right\} \left\{ \frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_s,k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_s)} \right\} \right\|_p \\
& \times \sum_{\tau=1}^{T} \| E[Y_\tau Y_0|\mathcal{F}_{\tau,\boldsymbol{X}}] - g(\boldsymbol{X}_\tau)g(\boldsymbol{X}_0)\|_{p/(p-1)}
\end{aligned}
$$

Under Assumption C2b, we have $\|Y_t - g(\boldsymbol{X}_t)\|_4 \le \|Y\|_4 < \infty$. Moreover, under Assumptions C3 and C7, Lemma A.6 and the dominated convergence theorem yield

$$
\lim_{T\longrightarrow\infty} \left\| \left\{ \frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_t,k_T)}{k_T} - \frac{1}{f(\boldsymbol{X}_t)} \right\}^2 \right\|_2 = 0.
$$

Hence, we have $\mathcal{T}_{2T;1} \longrightarrow 0$. By the same argument, we can show that, under Assumption C4c, $\mathcal{T}_{2T;2} \longrightarrow 0$. Therefore, it follows that $\mathcal{T}_{2T} = o_p(1)$.
Step 3: Notice that

$$
\begin{aligned}
\mathcal{T}_{3T} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{J(1/T) - 1\} \frac{g(\boldsymbol{X}_{[t]})}{\widehat{f}(\boldsymbol{X}_{[t]})} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{g(\boldsymbol{X}_t)}{\widehat{f}(\boldsymbol{X}_t)} - \sqrt{T} \int_{\mathcal{X}} g(\boldsymbol{x})d\boldsymbol{x} \\
&= \mathcal{T}_{3T;1} + \mathcal{T}_{3T;2},
\end{aligned}
$$

where $\widehat{f}(\boldsymbol{x}) = k_T/T\mathcal{L}(V_1)R_T^N(\boldsymbol{x},k_T)$ as previously defined. We start with the first term $\mathcal{T}_{3T;1}$ equals

$$
\begin{aligned}
& \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{J(t/T) - 1\} g(\boldsymbol{X}_{[t]}) \left\{ \frac{1}{\widehat{f}(\boldsymbol{X}_{[t]})} - \frac{1}{f(\boldsymbol{X}_{[t]})} \right\} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{J(t/T) - 1\} \frac{g(\boldsymbol{X}_{[t]})}{f(\boldsymbol{X}_{[t]})} \\
=\ & \mathcal{T}_{3T;1a} + \mathcal{T}_{3T;1b}.
\end{aligned}
$$

An application of the Cauchy-Schwarz inequality and Hölder's inequality for mixed norms, i.e. Let

$0 < p_i \leq \infty$ and $1/r = \sum_{i=1}^k 1/p_i$, then $\| \prod_{i=1}^k u_i \|_r \leq \prod_{i=1}^k \|u_i\|_{p_i}$, yields

$$
\begin{aligned}
& E[|\mathcal{T}_{3T;1a}|^2] \leq \frac{1}{T}\sum_{t=1}^{T}\{J(t/T)-1\}^2 E\left[g^2(\boldsymbol{X}_t)\left|\frac{1}{\widehat{f}(\boldsymbol{X}_t)}-\frac{1}{f(\boldsymbol{X}_t)}\right|^2\right] \\
+ & \ 2\frac{1}{T}\sum_{\substack{t=1\\ s=1\\ t\neq s}}^{T} |\{J(t/T)-1\}\{J(s/T)-1\}|\, E\left[|g(\boldsymbol{X}_{[t]})g(\boldsymbol{X}_{[s]})|\left|\frac{1}{\widehat{f}(\boldsymbol{X}_{[t]})}-\frac{1}{f(\boldsymbol{X}_{[t]})}\right|\left|\frac{1}{\widehat{f}(\boldsymbol{X}_{[s]})}-\frac{1}{f(\boldsymbol{X}_{[s]})}\right|\right] \\
\leq & \ \|g^2(\boldsymbol{X}_0)\|_2 \left\|\left|\frac{1}{\widehat{f}(\boldsymbol{X}_t)}-\frac{1}{f(\boldsymbol{X}_t)}\right|^2\right\|_2 \frac{1}{T}\sum_{t=1}^{T}\{J(t/T)-1\}^2 \\
+ & \ 2\|g(\boldsymbol{X}_0)\|_4^2 \left\|\frac{1}{\widehat{f}(\boldsymbol{X}_t)}-\frac{1}{f(\boldsymbol{X}_t)}\right\|_4^2 \frac{1}{T}|J(0)-1|\sum_{\tau=1}^{T}|J(\tau/T)-1|.
\end{aligned}
$$

Under Assumptions C3 and C7, Lemma A.6 holds. The score function $J(\cdot)$ is bounded. Assumption C2b implies that $E[g^4(\boldsymbol{X}_0)] < \infty$. It then follows that $\lim_{T\longrightarrow\infty} E[|\mathcal{T}_{3T;1a}|^2] = 0$. With the same argument, we can also show that $\lim_{T\longrightarrow\infty} E[|\mathcal{T}_{3T;1b}|^2] = 0$. Hence, we obtain

$$\mathcal{T}_{3T;1} = o_p(1). \tag{5.6}$$

Step 3: To finish the present proof, we need to show that $\mathcal{T}_{3T;2} = o_p(1)$. Suppose that the distribution function $F(\boldsymbol{x})$ is invertible (i.e., given some $\boldsymbol{x} \in \mathcal{X}$ there exists a realization of the Uniform[0,1] random variable, U, such that $\boldsymbol{x} = F^{-1}(u)$, where $F^{-1}(u)$ is the inverse cdf). In view of Lemma B.5 and the definition of integrals along a curve in $\mathbb{R}^N$ (*or* line integrals), see e.g. Apostol (1969, Definition 10.1), one can derive the following approximation:

$$
\begin{aligned}
E\left[\frac{g(\boldsymbol{X})}{f(\boldsymbol{X})}\right] &= \int_{\mathcal{X}} \frac{g(\boldsymbol{x})}{f(\boldsymbol{x})}dF(\boldsymbol{x}) = \int_{[0,1]} \frac{g\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)}du = \sum_{s=0}^{2^{\mathcal{N}}} \int_{s2^{-\mathcal{N}}}^{(s+1)2^{-\mathcal{N}}} \frac{g\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)}du \\
&\approx 2^{-\mathcal{N}} \sum_{s=0}^{2^{\mathcal{N}}} \frac{g\left(F^{-1}(u_s)\right)}{f\left(F^{-1}(u_s)\right)},
\end{aligned}
$$

where $u_s \in [s2^{-\mathcal{N}}, (s+1)2^{-\mathcal{N}})$. Without any loss of generality let us choose $\mathcal{N} = \log_2 T$. It then follows that:

$$\mathcal{T}_{3T;2} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\frac{g(\boldsymbol{X}_t)}{\widehat{f}(\boldsymbol{X}_t)} - \frac{g(F^{-1}(u_t))}{f\left(F^{-1}(u_t)\right)}\right\} + o_p(1).$$

Recalling Assumption C5a, the set $\mathcal{X}$ is closed and bounded. As T becomes sufficiently large, it is possible to find a $\rho-$ball of radius $\delta_{\mathfrak{N}(T)} = \epsilon 2^{-\mathfrak{N}(T)}$, which centers at τ_t such that $F^{-1}(u_t) \in B_{\delta_{\mathfrak{N}}}(\tau_t)$, where $\mathfrak{N}(T)$ increases with T, thus implicitly depends on T, though in what follows we will suppress the dependence of $\mathfrak{N}$ on T unless confusion is likely. Hence, we have

$$\mathcal{T}_{3T;2} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\frac{g(\boldsymbol{X}_t)}{\widehat{f}(\boldsymbol{X}_t)} - \frac{g(\boldsymbol{X}_{\tau_t})}{f(\boldsymbol{X}_{\tau_t})}\right\} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\frac{g(\boldsymbol{X}_{\tau_t})}{f(\boldsymbol{X}_{\tau_t})} - \frac{g(F^{-1}(u_t))}{f(F^{-1}(u_t))}\right\} + o_p(1).$$

To this end we take center points, $\{\tau_t\}_{t=1}^T$, in the finite subset of $\mathbb{Z}^+$, $\Pi_{\mathfrak{N}} \doteq \{s \in \mathbb{Z}^+ : \ \sup_{t\in\mathbb{Z}^+} \inf_{s\in\Pi_{\mathfrak{N}}} \rho(t,s) \le \delta_{\mathfrak{N}}\}$. Therefore, for a given $t \in \mathbb{Z}^+$, there exists a point, $\tau_t \in \Pi_{\mathfrak{N}}$, such that $\rho(t,\tau_t) \le \delta_{\mathfrak{N}}$. Next, let us define a mapping, $\pi_{\mathfrak{N}} : \ \mathbb{Z}^+ \longrightarrow \Pi_{\mathfrak{N}}$ such that $\log|\Pi_{\mathfrak{N}}| \le H(\delta_{\mathfrak{N}}, \mathbb{Z}^+, \rho)$, where $|.|$ denotes the cardinality of a set, and $\rho(t, \pi_{\mathfrak{N}} t) \le \delta_{\mathfrak{N}}$. It immediately follows that

$$\begin{aligned} \mathcal{T}_{3T;2} &\le \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{g(\boldsymbol{X}_t)}{\widehat{f}(\boldsymbol{X}_t)} - \frac{g(\boldsymbol{X}_{\pi_{\mathfrak{N}}t})}{f(\boldsymbol{X}_{\pi_{\mathfrak{N}}t})} \right| + \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{g(\boldsymbol{X}_{\pi_{\mathfrak{N}}t})}{f(\boldsymbol{X}_{\pi_{\mathfrak{N}}t})} - \frac{g\left(F^{-1}(u_t)\right)}{f\left(F^{-1}(u_t)\right)} \right| \\ &= \mathcal{T}_{3T;2;A} + \mathcal{T}_{3T;2;B}. \end{aligned}$$

In the sequel, we shall bound the terms $\mathcal{T}_{3T;2;A}$ and $\mathcal{T}_{3T;2;B}$. The triangular inequality yields

$$\begin{aligned} \mathcal{T}_{3T;2;A} &\le \frac{1}{\sqrt{T}} \left| \sum_{t=1}^T g(\boldsymbol{X}_t) \left\{ \frac{1}{\widehat{f}(\boldsymbol{X}_t)} - \frac{1}{f(\boldsymbol{X}_t)} \right\} \right| + \frac{1}{\sqrt{T}} \left| \sum_{t=1}^T \frac{1}{f(\boldsymbol{X}_t)} \left(g(\boldsymbol{X}_t) - g(\boldsymbol{X}_{\pi_{\mathfrak{N}}t}) \right) \right| \\ &+ \frac{1}{\sqrt{T}} \left| \sum_{t=1}^T g(\boldsymbol{X}_{\pi_{\mathfrak{N}}t}) \left\{ \frac{1}{f(\boldsymbol{X}_t)} - \frac{1}{f(\boldsymbol{X}_{\pi_{\mathfrak{N}}t})} \right\} \right| \\ &= \mathcal{T}_{3T;2;A;i} + \mathcal{T}_{3T;2;A;ii} + \mathcal{T}_{3T;2;A;iii}. \end{aligned}$$

Now, to show that $\mathcal{T}_{3T;2;A;i} = o_p(1)$, we study the second moment

$$\begin{aligned} E|\mathcal{T}_{3T;2;A;i}|^2 &= E\left[g^2(\boldsymbol{X}_t) \left| \frac{1}{\widehat{f}(\boldsymbol{X}_t)} - \frac{1}{f(\boldsymbol{X}_t)} \right|^2 \right] \\ &+ 2\frac{1}{T} \sum_{\substack{t=1 \\ s=1 \\ s<t}} E\left[g(\boldsymbol{X}_t) g(\boldsymbol{X}_s) \left(\frac{1}{\widehat{f}(\boldsymbol{X}_t)} - \frac{1}{f(\boldsymbol{X}_t)} \right) \left(\frac{1}{\widehat{f}(\boldsymbol{X}_s)} - \frac{1}{f(\boldsymbol{X}_s)} \right) \right] \\ &\le \left\| \frac{1}{\widehat{f}(\boldsymbol{X}_t)} - \frac{1}{f(\boldsymbol{X}_t)} \right\|_4^2 \left\{ \|g(\boldsymbol{X}_t)\|_4^2 + 2 \sum_{\tau=1}^T \|g(\boldsymbol{X}_0) g(\boldsymbol{X}_\tau)\|_2| \right\} \\ &\le 3\,\|g(\boldsymbol{X}_0)\|_4^2 \left\{ T^{1/2} \left\| \frac{1}{\widehat{f}(\boldsymbol{X}_t)} - \frac{1}{f(\boldsymbol{X}_t)} \right\|_4 \right\}^2. \end{aligned}$$

Assumptions C3a, C7, together with Lemma A.6 yield $\sup_{1\le t\le T} T^{1/2} |\widehat{f}^{-1}(\boldsymbol{X}_t) - f^{-1}(\boldsymbol{X}_t)|^2 = o_{a.s.}(1)$. Consequently, by virtue of Assumption C2b and the dominated convergence theorem, we prove that $E|\mathcal{T}_{3T;2;A;i}|^2 = o(1)$. Next, to show that $\mathcal{T}_{3T;2;A;ii} = o_p(1)$. Since the density function $f(\boldsymbol{x})$ is bounded, the equicontinuity condition (cf. Assumption C6) implies that

$$\begin{aligned} E[\mathcal{T}_{3T;2;A;ii}] &\le \text{Const.} \times \frac{1}{\sqrt{T}} \sum_{t=1}^T E^{1/2}\left[\|\boldsymbol{X}_t - \boldsymbol{X}_{\pi_{\mathfrak{N}}t}\|^2 \right] \\ &= \text{Const.} \times \frac{1}{\sqrt{T}} \sum_{t=1}^T \rho(t, \pi_{\mathfrak{N}} t) \le \text{Const.} \times \frac{1}{\sqrt{T}} \sum_{t=1}^T E^{1/2}[\|\boldsymbol{X}_t - \boldsymbol{X}_{\pi_{\mathfrak{N}}t}\|^2] \\ &\le \text{Const.} \times \frac{1}{\sqrt{T}} \sum_{\ell=\mathfrak{N}}^\infty \sum_{t=1}^T E^{1/2}[\|\boldsymbol{X}_{\pi_\ell t} - \boldsymbol{X}_{\pi_{\ell-1}t}\|^2]. \end{aligned}$$

To show that the quantity $E[\mathcal{T}_{3T;2;A;ii}]$ converges to zero, in view of the Kronecker lemma, it is sufficient to show that $\sum_{\ell=\mathfrak{N}}^{\infty}\sum_{t=1}^{T}\frac{1}{\sqrt{t}}E^{1/2}[\|\boldsymbol{X}_{\pi_\ell t}-\boldsymbol{X}_{\pi_{\ell-1}t}\|^2]<\infty$. Using the formula $\sum_{u=1}^{T}\frac{1}{\sqrt{u}}\leq 2\sqrt{T}$ and the relation: $\rho(\pi_{\ell-1}t,\pi_\ell t)\leq 3\delta_\ell$, one obtains:

$$
\begin{aligned}
\sum_{\ell=\mathfrak{N}}^{\infty}\sum_{t=1}^{T}\frac{1}{\sqrt{t}}E^{1/2}[\|\boldsymbol{X}_{\pi_\ell t}-\boldsymbol{X}_{\pi_{\ell-1}t}\|^2] &\leq \sum_{\ell=\mathfrak{N}}^{\infty}\delta_\ell \exp\{\frac{1}{2}H(\delta_\ell,\mathbb{Z}^+,\rho)\},\\
&\approx \sum_{\ell=\mathfrak{N}}^{\infty}\int_{\epsilon 2^{-\ell}}^{\epsilon 2^{-\ell+1}}\exp\{\frac{1}{2}H(\delta_\ell,\mathbb{Z}^+,\rho)\}\Delta\delta_\ell \leq \int_0^{\epsilon}\exp\{\frac{1}{2}H(x,\mathbb{Z}^+,\rho)\}dx.
\end{aligned}
$$

Invoking Assumption C5b, we obtain $\mathcal{T}_{3T;2;A;ii}=o_p(1)$. An analogous argument together with Assumptions C2b and C6 yield $\mathcal{T}_{3T;2;A;iii}=o_p(1)$.

To this end, to bound the term $\mathcal{T}_{3T;2;B}$, the triangular inequality yields

$$
\begin{aligned}
\mathcal{T}_{3T;2;B} &\leq \frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\frac{1}{f(\boldsymbol{X}_{\pi_{\mathfrak{N}}t})}\left(g(\boldsymbol{X}_{\pi_{\mathfrak{N}}t})-g(F^{-1}(u_t))\right)\right|\\
&\quad+\frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}g(F^{-1}(u_t))\left(\frac{1}{f(\boldsymbol{X}_{\pi_{\mathfrak{N}}t})}-\frac{1}{f(F^{-1}(u_t))}\right)\right| = \mathcal{T}_{3T;2;B;i}+\mathcal{T}_{3T;2;B;ii}.
\end{aligned}
$$

Next, we prove that $\mathcal{T}_{3T;2;B;i}=o_p(1)$, because the relation $\mathcal{T}_{3T;2;B;ii}=o_p(1)$ can be proved by the same argument. The boundedness of the density function, $f(\boldsymbol{x})$, and the equicontinuity condition of $g(\boldsymbol{x})$ (cf. Assumption C6) results in:

$$
\begin{aligned}
E[|\mathcal{T}_{3T;2;B;i}|] &\leq \text{Const.}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}E^{1/2}[\|\boldsymbol{X}_{\pi_{\mathfrak{N}}t}-F^{-1}(u_t)\|^2]\\
&\leq \text{Const.}\left\{\epsilon\sqrt{T}2^{-\mathfrak{N}}+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}E^{1/2}\left[\|\boldsymbol{X}_{\pi_{\mathfrak{N}}t}-\boldsymbol{X}_t\|^2\right]\right\}\\
&\leq \text{Const.}\left\{\epsilon\sqrt{T}2^{-\mathfrak{N}}+\frac{1}{\sqrt{T}}\sum_{\ell=\mathfrak{N}}^{\infty}\sum_{t=1}^{T}E^{1/2}\left[\|\boldsymbol{X}_{\pi_\ell t}-\boldsymbol{X}_{\pi_{\ell-1}t}\|\right]\right\}.
\end{aligned}
$$

Take $\mathfrak{N}(T)$ such that $2^{\mathfrak{N}}=\sqrt{T}/a_T$ for some sequence a_T decreasing to zero. Then the quantity $\sqrt{T}2^{-\mathfrak{N}}$ tends to 0 as T tends to infinity. Furthermore, to prove that the quantity $\frac{1}{\sqrt{T}}\sum_{\ell=\mathfrak{N}}^{\infty}\sum_{t=1}^{T}E^{1/2}\left[\|\boldsymbol{X}_{\pi_\ell t}-\boldsymbol{X}_{\pi_{\ell-1}t}\|\right]$ converges to zero, in view of the Kronecker lemma, it is sufficient to show that $\sum_{\ell=\mathfrak{N}}^{\infty}\sum_{t=1}^{T}\frac{1}{\sqrt{t}}E^{1/2}\left[\|\boldsymbol{X}_{\pi_\ell t}-\boldsymbol{X}_{\pi_{\ell-1}t}\|\right]<\infty$. Using the argument analogous as before, we obtain

$$
\sum_{\ell=\mathfrak{N}}^{\infty}\sum_{t=1}^{T}\frac{1}{\sqrt{t}}E^{1/2}\left[\|\boldsymbol{X}_{\pi_\ell t}-\boldsymbol{X}_{\pi_{\ell-1}t}\|\right]\leq\int_0^{\epsilon}\exp\{\frac{1}{2}H(x,\mathbb{Z}^+,\rho)\}dx.
$$

Invoking Assumption C5b, the desired result is obtained.

Finally, collecting all the relevant terms, we obtain $\mathcal{T}_{3T}=o_p(1)$.

5.4 Proof of Lemma 3.3

To prove (1), first note that $\mathbb{I}(Y_\tau(\omega_2)=y)\{m_h(Y_\tau|\mathcal{F}_0)-m_h(Y_\tau)\}=\int_{\mathcal{X}} h(\boldsymbol{x},y)d\{F(\boldsymbol{x}|y,\boldsymbol{X}_0)-F(\boldsymbol{x}|y)\}$, Applying the following inequality (see Csörgö, 1981):

$$\left|\int_{\mathcal{X}} f(\boldsymbol{x})dg(\boldsymbol{x})\right| \le \left|\int_{\mathcal{X}} g(\boldsymbol{x})\frac{\partial^N}{\partial x_1\dots\partial x_N}f(\boldsymbol{x})d\boldsymbol{x}\right| + 2N(2M)^{N-1}\sup_{\boldsymbol{x}\in\mathcal{X}}|g(\boldsymbol{x})|\sum_{d=0}^{N-1}\sum_{j_1+\cdots+j_N=d}\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\frac{\partial^d}{\partial x_1^{j_1}\dots\partial x_N^{j_N}}f(\boldsymbol{x})\right|,$$

where, without any loss of generality, we may take $\mathcal{X}=\prod_{i=1}^N[-M,M]$ as a compact subspace of $\mathbb{R}^N$ and obtain:

$$\begin{aligned}
&|\mathbb{I}(Y_\tau(\omega_2)=y)\{m_h(Y_\tau;\mathcal{F}_0)-m_h(Y_\tau)\}|\\
\le\ & \left|\int_{\mathcal{X}}\frac{\partial^N}{\partial x_1\dots\partial x_N}h(\boldsymbol{x},y)[F(\boldsymbol{x}|y,\boldsymbol{X}_0)-F(\boldsymbol{x}|y)]d\boldsymbol{x}\right|\\
+\ & 2N(2M)^{N-1}\sup_{\boldsymbol{x}\in\mathcal{X}}|F(\boldsymbol{x}|y,\boldsymbol{X}_0)-F(\boldsymbol{x}|y)|\sum_{d=0}^{N-1}\sum_{j_1+\cdots+j_N=d}\left|\frac{\partial^d}{\partial x_1^{j_1}\dots\partial x_N^{j_N}}h(\boldsymbol{x},y)\right|\\
\le\ & \sup_{\boldsymbol{x}\in\mathcal{X}}\frac{\partial^N}{\partial x_1\dots\partial x_N}h(\boldsymbol{x},y)\int_{\mathcal{X}}|F(\boldsymbol{x}|y,\boldsymbol{X}_0)-F(\boldsymbol{x}|y)|\,d\boldsymbol{x}\\
+\ & 2N(2M)^{N-1}\sup_{\boldsymbol{x}\in\mathcal{X}}|F(\boldsymbol{x}|y,\boldsymbol{X}_0)-F(\boldsymbol{x}|y)|\sum_{d=0}^{N-1}\sum_{j_1+\cdots+j_N=d}\left|\sup_{\boldsymbol{x}\in\mathcal{X}}\frac{\partial^d}{\partial x_1^{j_1}\dots\partial x_N^{j_N}}h(\boldsymbol{x},y)\right|.
\end{aligned}$$

An application of Hölder's inequality and Minkowski's inequality yields

$$\begin{aligned}
\|m_h(Y_\tau;\mathcal{F}_0)-m_h(Y_\tau)\|_{2,\mathcal{F}_0} \le\ & \left\|\sup_{\boldsymbol{x}\in\mathcal{X}}\frac{\partial^N}{\partial x_1\dots\partial x_N}h(\boldsymbol{x},Y_\tau)\right\|_{4,\mathcal{F}_0}\|\boldsymbol{\tau}\left(\mathcal{F}_{\boldsymbol{X}_0},\boldsymbol{X}_\tau|Y_\tau\right)\|_{4,\mathcal{F}_0}\\
+\ & 2N(2M)^{N-1}\|\alpha\left(\mathcal{F}_{\boldsymbol{X}_0},\boldsymbol{X}_\tau|Y_\tau\right)\|_{4,\mathcal{F}_0}\\
& \sum_{d=0}^{N-1}\sum_{j_1+\cdots+j_N=d}\left\|\sup_{\boldsymbol{x}\in\mathcal{X}}\frac{\partial^d}{\partial x_1^{j_1}\dots\partial x_N^{j_N}}h(\boldsymbol{x},Y_\tau)\right\|_{4,\mathcal{F}_0}.
\end{aligned}$$

Applying Minkowski's inequality and Hölder's inequality for mixed norms, we obtain:

$$\begin{aligned}
\left\|\|m_h(Y_\tau;\mathcal{F}_0)-m_h(Y_\tau)\|_{2,\mathcal{F}_0}\right\|_{p/(p-1)} \le\ & \left\|\left\|\sup_{\boldsymbol{x}\in\mathcal{X}}\frac{\partial^N}{\partial x_1\dots\partial x_N}h(\boldsymbol{x},Y_\tau)\right\|_{4,\mathcal{F}_0}\right\|_{p_1}\left\|\|\boldsymbol{\tau}\left(\mathcal{F}_{\boldsymbol{X}_0},\boldsymbol{X}_\tau|Y_\tau\right)\|_{4,\mathcal{F}_0}\right\|_{p_2}\\
+\ & 2N(2M)^{N-1}\left\|\|\alpha\left(\mathcal{F}_{\boldsymbol{X}_0},\boldsymbol{X}_\tau|Y_\tau\right)\|_{4,\mathcal{F}_0}\right\|_{p_1}\\
& \times\sum_{d=0}^{N-1}\sum_{j_1+\cdots+j_N=d}\left\|\left\|\sup_{\boldsymbol{x}\in\mathcal{X}}\frac{\partial^d}{\partial x_1^{j_1}\dots\partial x_N^{j_N}}h(\boldsymbol{x},Y_\tau)\right\|_{4,\mathcal{F}_0}\right\|_{p_2}.
\end{aligned}$$

Therefore, the first part of the lemma immediately follows.

To prove (2), first note that

$$
\begin{aligned}
E[h(\boldsymbol{X}_\tau, Y_\tau)h(\boldsymbol{X}_0, Y_0)|\mathcal{F}_{\tau,Y}] - m_h(Y_\tau)m_h(Y_0) &= E\left[(h(\boldsymbol{X}_\tau, Y_\tau) - m_h(Y_\tau))\,(h(\boldsymbol{X}_0, Y_0) - m_h(Y_0))\,\middle|\mathcal{F}_{\tau,Y}\right] \\
&= \operatorname{cov}(h(\boldsymbol{X}_\tau, Y_\tau), h(\boldsymbol{X}_0, Y_0)|\mathcal{F}_{\tau,Y}).
\end{aligned}
$$

Define a random variable, $H_\tau(y) = \mathbb{I}(Y_\tau(\omega_2) = y)h(\boldsymbol{X}_\tau(\boldsymbol{\omega}_1), Y_\tau(\omega_2))$ for some $y \in \mathbb{R}$. Using result B.4, we obtain:

$$
\begin{aligned}
&\operatorname{cov}[h(\boldsymbol{X}_0, Y_0), h(\boldsymbol{X}_\tau, Y_\tau)|Y_0 = y_1, Y_\tau = y_2] = \operatorname{cov}[H_0(y_1), H_\tau(y_2)] \\
&\quad \leq 6\|H_0(y_1)\|_q \|H_\tau(y_2)\|_r \Bigg\{ \sup_{\substack{A_0 \in \mathcal{F}_0 \\ A_\tau \in \mathcal{F}_\tau}} |P\left((\boldsymbol{X}_0, Y_0) \in A_0, (\boldsymbol{X}_\tau, Y_\tau) \in A_\tau | Y_0 = y_1, Y_\tau = y_2\right) \\
&\qquad - \ P\left((\boldsymbol{X}_0, Y_0) \in A_0 | Y_0 = y_1\right) P\left((\boldsymbol{X}_\tau, Y_\tau) \in A_\tau | Y_\tau = y_2\right)|^{1-1/q-1/r} \Bigg\} \\
&\qquad = 6\|H_0(y_1)\|_q \|H_\tau(y_2)\|_r \{\alpha_\tau^*(Y_0 = y_1, Y_\tau = y_2)\}^{1-1/q-1/r}.
\end{aligned}
$$

It immediately follows that:

$$
\begin{aligned}
& \|\operatorname{cov}(h(\boldsymbol{X}_\tau, Y_\tau), h(\boldsymbol{X}_0, Y_0)|\mathcal{F}_{\tau,Y})\|_{p/(p-1)} \\
\leq\ & 6 \left\| \|h(\boldsymbol{X}_0, Y_0)\|_{q,\mathcal{F}_{0,Y}} \|h(\boldsymbol{X}_0, Y_0)\|_{r,\mathcal{F}_{0,Y}} \{\alpha_\tau^*(Y_0, Y_\tau)\}^{1-1/q-1/r} \right\|_{p/(p-1)} \\
\leq\ & 6 \left\| h(\boldsymbol{X}_0, Y_0)\|_{q,\mathcal{F}_{0,Y}} \right\|_{p_1} \left\| \|h(\boldsymbol{X}_0, Y_0)\|_{r,\mathcal{F}_{0,Y}} \right\|_{p_2} \left\| \{\alpha_\tau^*(Y_0, Y_\tau)\}^{1-1/q-1/r} \right\|_{p_3},
\end{aligned}
$$

where the last inequality follows from an application of Hölder's inequality for mixed norms. Hence, the second part of the lemma has been proved. The third part can be proved in the same way.

A Auxiliary Results

We state and prove a generalized version of Young's inequality for the convolution of two functions of $\mathbb{R}^N$-valued random variables on the Lebesgue Space. The following result is an extension of Young's convolution theorem, see for example Wheeden and Zygmund (1977, p. 146).

Lemma A.1 (Generalized Young's Inequality)**.** *Let p and q satisfy $1 \leq p, q \leq \infty$ and $1/p + 1/q \geq 1$, and let r be defined by $1/r = 1/p + 1/q - 1$. If $\boldsymbol{x} \in \mathbb{R}^N$, $f \in L^p(\mathbb{R}^N, \mathcal{A}, \mu)$ and $g \in L^q(\mathbb{R}^N, \mathcal{A}, \mu)$, then $f * g \in L^r(\mathbb{R}^N, \mathcal{A}, \mu)$, where μ is the Lebesgue measure such that $\int_{\mathbb{R}} \mu(d\boldsymbol{x}) = 1$, and*

$$
\|f * g\|_r \leq \|f\|_p \|g\|_q.
$$

Proof. First, we write

$$
\begin{aligned}
f * g(\boldsymbol{x}) &= \int_{\mathbb{R}^N} f^{p/r}(\boldsymbol{t}) g^{q/r}(\boldsymbol{x} - \boldsymbol{t}) f^{p(1/p-1/r)}(\boldsymbol{t}) g^{q(1/q-1/r)}(\boldsymbol{x} - \boldsymbol{t}) \mu(d\boldsymbol{t}), \\
&= \int_{\mathbb{R}^N} f^{p/r}(\boldsymbol{t}) g^{q/r}(\boldsymbol{x} - \boldsymbol{t}) f^{p/p_1}(\boldsymbol{t}) g^{q/p_2}(\boldsymbol{x} - \boldsymbol{t}) \mu(d\boldsymbol{t}),
\end{aligned}
$$

where $1/p_1 = 1/p - 1/r$ and $1/p_2 = 1/q - 1/r$. An application of Hölder's inequality for mixed norms yields:

$$|f * g(\boldsymbol{x})| \leq \left\{\int_{\mathbb{R}^N} f^p(\boldsymbol{t}) g^q(\boldsymbol{x}-\boldsymbol{t})\mu(d\boldsymbol{t})\right\}^{1/r} \left\{\int_{\mathbb{R}^N} f^p(\boldsymbol{t})\mu(d\boldsymbol{t})\right\}^{1/p_1} \left\{\int_{\mathbb{R}^N} g^q(\boldsymbol{x}-\boldsymbol{t})\mu(d\boldsymbol{t})\right\}^{1/p_2}.$$

In view of the assumption that $\int_{\mathbb{R}} \mu(d\boldsymbol{x}) = 1$, an application of Hölder's inequality yields

$$\begin{aligned} |f * g(\boldsymbol{x})|_r &\leq \|f\|_p^p \|g\|_q^q \|f\|_p^{pr/p_1} \|g\|_q^{qr/p_2} \\ &= \|f\|_p^{p(1+r/p_1)} \|g\|_q^{q(1+r/p_2)}. \end{aligned}$$

It follows that $\|f * g(\boldsymbol{x})\|_r \leq \|f\|_p \|g\|_q$. ◻

The following Lemmas A.2-A.4 are needed for the proof of Theorem 3.1.

Lemma A.2. *Let $\|m_h^{'}(Y_0|\mathcal{I})\|_{p^*,\mathcal{I}} < \infty$, $\lim_{t\longrightarrow\infty} \|P(Y_t \leq y|\mathcal{F}_{Y_0}) - F(y|\mathcal{I})\|_{q^*,\mathcal{I}} = 0$, and $\sum_{\tau=1}^{\infty} \|\|m_h(Y_\tau;\mathcal{F}_0) - m_h(Y_\tau;\mathcal{I})\|_{2,\mathcal{F}_0}\|_{p/(p-1),\mathcal{I}} < \infty$ hold with $1/p^* + 1/q^* = 1 + (p-1)/p$. Set $\mathcal{K}_t(p) = \|E[W_t|\mathcal{F}_0]\|_{p/(p-1),\mathcal{I}}$. Then, $\sum_{\tau=1}^{\infty} \mathcal{K}_\tau(p) < \infty$.*

Proof. An application of Minkowski's inequality yields:

$$\begin{aligned} \mathcal{K}_p(t) &\leq \|E\left[J(F(Y_t|\mathcal{I}))\{h(\boldsymbol{X}_t, Y_t) - m_h(Y_t)\}|\mathcal{F}_0\right]\|_{p/(p-1)} \\ &+ \left\|\int_{\mathbb{R}} J(F(y|\mathcal{I}))\{F(y|\mathcal{F}_0) - F(y|\mathcal{I})\} m_h^{'}(y|\mathcal{I}) dy\right\|_{p/(p-1)} \\ &= \mathcal{K}_I + \mathcal{K}_{II}, \end{aligned}$$

where $F(y|\mathcal{F}_0) = P(Y_t \leq y|\mathcal{F}_0) = P(Y_t - y \leq 0|\mathcal{F}_0)$, and the definitions of $\mathcal{K}_I$and $\mathcal{K}_{II}$ should be apparent. By Hölder's inequality and the law of iterated expectations, we have:

$$\begin{aligned} \mathcal{K}_I &\leq \left\|E\left[[J(F(Y_t|\mathcal{I}))\{h(\boldsymbol{X}_t, Y_t) - m_h(Y_t|\mathcal{I})|Y_t, \mathcal{F}_0\}|Y_t, \mathcal{F}_0\right]\big|\mathcal{F}_0\right]\right\|_{p/(p-1),\mathcal{I}} \\ &\leq \left\|\|J(F(Y_t|\mathcal{I}))\|_{2,\mathcal{F}_0} \|E[h(\boldsymbol{X}_t, Y_t)|Y_t, \mathcal{F}_0] - m_h(Y_t|\mathcal{I})\|_{2,\mathcal{F}_0}\right\|_{p/(p-1),\mathcal{I}} \\ &\leq \text{Const.} \times \left\|\|E[h(\boldsymbol{X}_t, Y_t)|Y_t, \mathcal{F}_0] - m_h(Y_t|\mathcal{I})\|_{2,\mathcal{F}_0}\right\|_{p/(p-1),\mathcal{I}}, \end{aligned}$$

where the last inequality follows because the function $J(\cdot)$ is bounded. In addition, by applying the generalized Young inequality for variables on the Lebesgue spaces in Lemma A.1 above, i.e. $\|u * v\|_r \leq \|u\|_p \|v\|_q$, where $u * v(x) = \int_{\mathbb{R}^N} u(x-y)v(y)pdf(y)dy$ and $1/p + 1/q = 1 + 1/r$, we immediately obtain $\mathcal{K}_{II} = \|u * v(Y_t)\|_{p/(p-1)} \leq \|F(y|\mathcal{F}_0) - F(y|\mathcal{I})\|_{p^*} \|J(F(y|\mathcal{I})) m_h^{'}(y|\mathcal{I})\|_{q^*} \leq \text{Const.} \times \|F(y|\mathcal{F}_0) - F(y|\mathcal{I})\|_{p^*} \|m_h^{'}(y|\mathcal{I})\|_{q^*}$, using the fact that the function $J(\cdot)$ is bounded. ◻

Lemma A.3. *For some generic constants, $p \geq 2$ and $q^* \geq 1$, defined in Lemma A.2, let*

$$\|h(\boldsymbol{X}_0, Y_0)\|_{p,\mathcal{I}} < \infty \text{ and } \|m_h^{'}(Y_0|\mathcal{I})\|_{q^*,\mathcal{I}} < \infty. \tag{A.1}$$

If

$$\sum_1^{\infty} \mathcal{K}_t(p) < \infty, \tag{A.2}$$

then (a) $\lim_{s\longrightarrow-\infty} E[W_0|\mathcal{F}_{-s}] = 0$ *(P-a.s.)*, (b) $\lim_{t\longrightarrow\infty} t\mathcal{K}_t(p) = 0$, (c) $\sum_1^\infty E\,|E[W_0W_t|\mathcal{I}]| < \infty$, (d) $\sum_1^\infty |E[W_0W_t]| < \infty$, (e) $\sum_{t=s}^\infty E|E[W_t|\mathcal{F}_{s-1}]| < \infty$, (f) $\sum_{t=s}^\infty E|E[W_t|\mathcal{F}_s]| < \infty$.

Proof. Firstly, note that $\mathcal{I} \subset \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$. We shall now prove each claim as follows:
Result (a): Condition (A.2) implies that $\lim_{t\longrightarrow\infty} \mathcal{K}_t(p) = 0$ or $\lim_{t\longrightarrow\infty} \|E[W_t|\mathcal{F}_0]\|_{p/(p-1)} = 0$. By the strict stationarity and ergodicity of $(\boldsymbol{X}_t, Y_t)$, we have $E[|E[W_s|\mathcal{F}_0]|^{\frac{p}{p-1}}] = E[|E[\mathcal{T}^{-s}W_s|\mathcal{T}^s\mathcal{F}_0]|^{\frac{p}{p-1}}] = E[|E[W_0|\mathcal{F}_{-s}]|^{\frac{p}{p-1}}]$. Thus equation (a) follows.
Result (b): Consider the decomposition $\sum_1^\infty t\mathcal{K}_t(p) = \sum_1^{T_0} t\mathcal{K}_t(p) + \sum_{T_0+1}^{T_1} t \times \mathcal{K}_t(p) + \sum_{T_1+1}^{T_2} t\mathcal{K}_t(p) + \cdots + \sum_{T_n+1}^{\infty} t\mathcal{K}_t(p)$, as $n \longrightarrow \infty$, in such a way that $T_0 > T_1 - T_0 > T_2 - T_1 > \ldots$. It then follows that $\lim_{n\longrightarrow\infty}(T_n - T_{n-1}) = 0$. An application of the Kronecker lemma and (A.2) yield $\lim_{T_0\longrightarrow\infty} T_0^{-1}\sum_1^{T_0} t\mathcal{K}_t(p) = 0$, $\lim_{T_1-T_0\longrightarrow\infty}(T_1-T_0)^{-1}\sum_{T_0+1}^{T_1} t\mathcal{K}_t(p) = 0$, $\ldots$,$\lim_{T_n-T_{n-1}\longrightarrow\infty}(T_n - T_{n-1})^{-1}\sum_{T_{n-1}+1}^{T_n} t\mathcal{K}_t(p) = 0$. Hence, it follows that $\sum_1^{T_0} t\mathcal{K}_t(p) = o(T_0)$, $\sum_{T_0+1}^{T_1} t\mathcal{K}_t(p) = o(T_1 - T_0)$, $\ldots$, $\sum_{T_{n-1}}^{T_n} t\mathcal{K}_t(p) = o(T_n - T_{n-1})$, and the result follows after noticing that as n becomes sufficiently large, $\lim_{n\longrightarrow\infty}(T_n - T_{n-1}) = 0$.
Result (c): Hölder's and Jensen's inequalities yield $|E[W_0W_t|\mathcal{I}]| = |E[W_0 \times E[W_t|\mathcal{F}_0]|\mathcal{I}]| \leq \|W_0\|_{p,\mathcal{I}} \times \|E[W_t|\mathcal{F}_0]\|_{p/(p-1),\mathcal{I}}$. By Minkowski' inequality and Young's inequality, we can see that $\|W_0\|_{p,\mathcal{I}} \leq$ Const. $\times \{\|h(\boldsymbol{X}_0, Y_0) - m_h(Y_0;\mathcal{I})\|_{p,\mathcal{I}} + \|F(Y_0|\mathcal{I}) - F(Y_0)\|_{p^*,\mathcal{I}}\|m_h'(Y_0|\mathcal{I})\|_{q^*,\mathcal{I}}\}$, where p^* and q^* are defined in Lemma A.2. In view of Eq. (A.1) and the inequality $\|h(\boldsymbol{X}_0, Y_0) - m_h(Y_0;\mathcal{I})\|_{p,\mathcal{I}} \leq 2\|h(\boldsymbol{X}_0, Y_0)\|_{p,\mathcal{I}} < \infty$, it follows that $\|W_0\|_{p,\mathcal{I}} < \infty$. In addition, $\|E[W_t|\mathcal{F}_0]\|_{p/(p-1),\mathcal{I}} =$
$\left\|E[\mathcal{T}^{-t}W_t|\mathcal{T}^t\mathcal{F}_0]\right\|_{p/(p-1),\mathcal{I}} = \|E[W_0|\mathcal{F}_{-t}]\|_{p/(p-1),\mathcal{I}} \longrightarrow E[W_0] = 0$ as $t \longrightarrow \infty$. Thus the result follows.
Result (d): By the law of iterated expectation and the Jensen inequality, we obtain $\sum_1^\infty |E[W_0W_t]| = \sum_1^\infty |E[E[W_0W_t|\mathcal{I}]]| \leq \sum_1^\infty E|E[W_0W_t|\mathcal{I}]| < \infty$ in view of (c).
Result (e): An application of Hölder's inequality and Jensen's inequality yield
$\sum_s^\infty E|E[W_t|\mathcal{F}_{s-1}]| = \sum_s^\infty E|E[\mathcal{T}^{-(s-1)}W_t|\mathcal{T}^{s-1}\mathcal{F}_{s-1}]| \leq \sum_s^\infty \|E[W_{t-s+1}|\mathcal{F}_0]\|_{\frac{p}{p-1}} = \sum_s^\infty \mathcal{K}_{t-s+1}(p) = \sum_1^\infty \mathcal{K}_s(p) < \infty$.
Result (f): By Hölder inequality and Eq. (e), we have $\sum_s^\infty E|E[W_t|\mathcal{F}_s]| = \sum_s^\infty E|E[\mathcal{T}^{-s}W_t|\mathcal{T}^s\mathcal{F}_s]| = E|W_0| + \sum_{s+1}^\infty E|E[W_{t-s}|\mathcal{F}_0]| \leq \sum_1^\infty \mathcal{K}_t(p) + \|W_0\|_p < \infty$. □

Lemma A.4. *For some integer $p > 1$, let $\|W_0\|_{p,\mathcal{I}} < \infty$ and $\sum_1^\infty \mathcal{K}_t(p) < \infty$. Then,*

$$\sigma_W^2(\mathcal{I}) = E[W_1^{*2}|\mathcal{I}] = E[W_0^2|\mathcal{I}] + 2\sum_{s=1}^\infty E[W_0W_s|\mathcal{I}] < \infty. \tag{A.3}$$

Proof. Let's define $\xi_t = \mathbb{I}(\omega \in A)\{E[W_t|\mathcal{F}_0] - E[W_t|\mathcal{F}_{-1}]\}$, where A is a subset in $\mathcal{I}$. Hence, we obtain $\mathbb{I}(\omega \in A)W_1^* = \sum_1^\infty \xi_t$. Therefore, $E[W_1^{*2}|A \in \mathcal{I}] = E[\mathbb{I}(\omega \in A)W_1^{*2}] = \sum_{\substack{t=1\\s=1}}^\infty E[\xi_s\xi_t]$. We deduce that

$E[\xi_s \xi_t]$ equals

$$\begin{aligned}
& E\big[\mathbb{I}(\omega \in A)\{E[W_t|\mathcal{F}_0] - E[W_t|\mathcal{F}_{-1}]\}\{E[W_s|\mathcal{F}_0] - E[W_s|\mathcal{F}_{-1}]\}\big] \\
= \;& E\big[\mathbb{I}(\omega \in A)\{E[W_t|\mathcal{F}_0]E[W_s|\mathcal{F}_0] - E[W_t|\mathcal{F}_0]E[W_s|\mathcal{F}_{-1}] \\
- \;& E[W_t|\mathcal{F}_{-1}]E[W_s|\mathcal{F}_0] + E[W_t|\mathcal{F}_{-1}]E[W_s|\mathcal{F}_{-1}]\}\big], \\
= \;& E\big[\mathbb{I}(\omega \in A)E[W_tE[W_s|\mathcal{F}_0]\mathcal{F}_0]\big] - E\big[\mathbf{1}(\omega \in A)E[W_tE[W_s|\mathcal{F}_{-1}]|\mathcal{F}_0]\big] \\
- \;& E\big[\mathbb{I}(\omega \in A)E[W_sE[W_t|\mathcal{F}_{-1}]|\mathcal{F}_0]\big] + E\big[\mathbf{1}(\omega \in A)E[W_tE[W_s|\mathcal{F}_{-1}]|\mathcal{F}_0]\big], \\
= \;& E\big[\mathbb{I}(\omega \in A)W_tE[W_s|\mathcal{F}_0]\big] - E\big[\mathbf{1}(\omega \in A)W_tE[W_s|\mathcal{F}_{-1}]\big] \\
- \;& E\big[\mathbb{I}(\omega \in A)W_sE[W_t|\mathcal{F}_{-1}]\big] + E\big[\mathbf{1}(\omega \in A)W_tE[W_s|\mathcal{F}_{-1}]\big], \\
= \;& E\big[W_tE[W_s|\mathcal{F}_0]|A \in \mathcal{I}\big] - E\big[W_tE[W_s|\mathcal{F}_{-1}]|A \in \mathcal{I}\big].
\end{aligned}$$

Since $E\big[W_tE[W_s|\mathcal{F}_{-1}]|A \in \mathcal{I}\big] = E[\mathcal{T}^{s-t}W_tE[W_s|\mathcal{T}^{t-s}\mathcal{F}_{-1}]] = E[W_s \times E[W_t|\mathcal{F}_{-1}]|A \in \mathcal{I}]$ it then follows $E[\xi_s\xi_t]$ equals

$$\begin{aligned}
& E\big[W_tE[W_s|\mathcal{F}_0]|A \in \mathcal{I}\big] - E\big[W_{t+1}E[W_{s+1}|\mathcal{F}_0]|A \in \mathcal{I}\big] \\
= \;& E\big[W_tE[W_s|\mathcal{F}_0]|A \in \mathcal{I}\big] - E\big[W_{t+1}E[W_s|\mathcal{F}_0]|A \in \mathcal{I}\big] \\
+ \;& E\big[W_{t+1}E[W_s|\mathcal{F}_0]|A \in \mathcal{I}\big] - E\big[W_{t+1}E[W_{s+1}|\mathcal{F}_0]|A \in \mathcal{I}\big], \\
= \;& E\big[(W_t - W_{t+1})E[W_s|\mathcal{F}_0]|A \in \mathcal{I}\big] + E\big[W_{t+1}(E[W_s|\mathcal{F}_0] - E[W_{s+1}|\mathcal{F}_0])|A \in \mathcal{I}\big].
\end{aligned}$$

Hence, we obtain that $\sum_{t=1,s=1,t\neq s}^{T} \xi_s\xi_t$ equals

$$\begin{aligned}
- \;& \sum_{s=1}^{T} E\big[(W_{T+1} - W_1)E[W_s|\mathcal{F}_0]|A \in \mathcal{I}\big] - \sum_{t=1}^{T} E\big[W_{t+1}(E[W_{T+1}|\mathcal{F}_0] - E[W_1|\mathcal{F}_0])|A \in \mathcal{I}\big], \\
= \;& \sum_{t=1}^{T} \Big\{E\big[W_1E[W_t|\mathcal{F}_0]|A \in \mathcal{I}\big] + E\big[W_{t+1}E[W_1|\mathcal{F}_0]|A \in \mathcal{I}\big]\Big\} \\
- \;& \sum_{s=1}^{T} \Big\{E\big[W_{T+1}E[W_s|\mathcal{F}_0]|A \in \mathcal{I}\big] + E\big[W_{s+1}E[W_{T+1}|\mathcal{F}_0]|A \in \mathcal{I}\big]\Big\}, \\
= \;& \mathcal{A}_1 + \mathcal{A}_2.
\end{aligned}$$

Moreover, some basic algebra yield $\mathcal{A}_1$ equal to

$$\begin{aligned}
& \sum_{t=1}^{T} \Big\{E[\mathcal{T}^{-1}W_1E[W_t|\mathcal{T}^1\mathcal{F}_0]|A \in \mathcal{I}] + E[\mathcal{T}^{-1}W_{t+1}E[W_1|\mathcal{T}^1\mathcal{F}_0]|A \in \mathcal{I}]\Big\} \\
= \;& \sum_{t=1}^{T} \Big\{E[W_0E[W_{t-1}|\mathcal{F}_0]|A \in \mathcal{I}] + E[W_tW_0|A \in \mathcal{I}]\Big\}, \\
= \;& E\big[W_0^2|A \in \mathcal{I}\big] + \sum_{s=1}^{T-1} \Big\{E\big[W_0E[W_s|\mathcal{F}_0]|A \in \mathcal{I}\big] + E\big[W_sW_0|A \in \mathcal{I}\big]\Big\} + E\big[W_0W_T|A \in \mathcal{I}\big], \\
= \;& E\big[W_0^2|A \in \mathcal{I}\big] + 2\sum_{s=1}^{T-1} E\big[W_0W_s|A \in \mathcal{I}\big] + E[W_0W_T|A \in \mathcal{I}].
\end{aligned}$$

Since $E[W_0|\mathcal{F}_{-\infty}] = 0$, then $\lim_{T\longrightarrow\infty} E[W_0 W_T|A \in \mathcal{I}] = 0$.
Hence $\mathcal{A}_1 = E\big[W_0^2|A \in \mathcal{I}\big] + 2\sum_{s=1}^{T-1} E\big[W_0 W_s|A \in \mathcal{I}\big]$.
The law of iterated expectations and Hölder's inequality yields

$$\begin{aligned}
\mathcal{A}_2 &= \sum_{s=1}^{T}\Big\{E\big[E[W_{T+1}|\mathcal{F}_0]E[W_s|\mathcal{F}_0]|A \in \mathcal{I}\big] + E\big[W_{s+1}E[W_{T+1}|\mathcal{F}_0]|A \in \mathcal{I}\big]\Big\} \\
&\leq \sum_{s=1}^{T}\Big\{\|E[W_s|\mathcal{F}_0]\|_p \|E[W_{T+1}|\mathcal{F}_0]\|_{\frac{p}{p-1}} + \|W_{s+1}\|_p \|E[W_{T+1}|\mathcal{F}_0]\|_{\frac{p}{p-1}}\Big\} \\
&= \mathcal{K}_{T+1}(p)\sum_{s=1}^{T}\mathcal{K}_s(p) + T\mathcal{K}_{T+1}(p)\|W_0\|_{p,\mathcal{I}}.
\end{aligned}$$

By referring to Lemma A.3 which asserts the finiteness of the individual conditional/unconditional (joint) moments, we can immediately verify that, as T goes to infinity, the term $\mathcal{A}_1$ is finite and the term $\mathcal{A}_2$ vanishes. Now, by extending the set A to the whole algebra of invariant sets $\mathcal{I}$, we obtain Eq. (A.3). □

Lemma A.5. *Suppose that $\|m_h^{'}(Y_0)\|_{p^*} < \infty$ and $\sum_{t=1}^{\infty} \alpha_{t-1}^* < \infty$, where p^* is some integer greater than one. Then, it follows that $\sum_1^{\infty} \mathcal{K}_t(p) < \infty$.*

Proof. Note that for any measurable function $\widetilde{f}(\cdot)$, we have
$E[\widetilde{f}(Y_t)|\mathcal{F}_{t,\boldsymbol{X}}] = E[\widetilde{f}(Y_t)|\sigma(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, \boldsymbol{X}_0)] = E[\widetilde{f}(Y_t)|\sigma(\widetilde{\boldsymbol{\xi}}_t, \widetilde{\boldsymbol{\xi}}_{t-1}, \ldots, \widetilde{\boldsymbol{\xi}}_0)]$, where $\widetilde{\boldsymbol{\xi}}_t^\top = (\xi_{1,t}, \ldots, \xi_{N,t})$. Since $\sigma(\widetilde{\boldsymbol{\xi}}_{t-1}) \subset \sigma(\widetilde{\boldsymbol{\xi}}_t)$, it then follows that $E[\widetilde{f}(Y_t)|\mathcal{F}_{t,\boldsymbol{X}}] = E[Y_t|\mathcal{F}_{\boldsymbol{X}_t}]$.

Since the processes $\boldsymbol{X}_t$ and Y_t are vector-valued stationary, ergodic Markov chains it follows that $F(y|\mathcal{I}) = \lim_{\tau\longrightarrow\infty} P(Y_\tau \leq y|Y_0 \in \mathcal{I}) = F(y)$ and thus $m_h(y;\mathcal{I}) = m_h(y)$. Hence,

$$\begin{aligned}
\mathcal{K}_t(p) &= \Bigg\|E\bigg[J(F(Y_t))\{h(\boldsymbol{X}_t, Y_t) - m_h(Y_t)\} \\
&\quad - \int_{\mathbb{R}} J(F(y))\{\mathbb{I}(Y_t \leq y) - F(y)\}dm_h(y)\big|\mathcal{F}_{\boldsymbol{X}_0}, \mathcal{F}_{Y_0}\bigg]\Bigg\|_{p/(p-1)} \\
&\leq \big\|E\big[J(F(Y_t))\{h(\boldsymbol{X}_t, Y_t) - m_h(Y_t)\}\big|\mathcal{F}_{\boldsymbol{X}_0}, \mathcal{F}_{Y_0}\big]\big\|_{p/(p-1)} \\
&\quad + \bigg\|\int_{\mathbb{R}} J(F(y))\{P(Y_t \leq y|\mathcal{F}_{Y_0}) - F(y)\}m_h^{'}(y)dy\bigg\|_{p/(p-1)} = \mathfrak{K}_{\mathrm{a}} + \mathfrak{K}_{\mathrm{b}}.
\end{aligned}$$

Since $J(F(Y_t))$ is bounded, $\mathfrak{K}_{\mathrm{a}} \leq \widetilde{h}_t(\boldsymbol{\xi}_0) = E[\widetilde{g} \circ \boldsymbol{g}(\boldsymbol{\xi}_t)|\boldsymbol{\xi}_0] = E\left[E\left[\widetilde{g} \circ \boldsymbol{g}(\boldsymbol{\xi}_t)|\boldsymbol{\xi}_{t-1}\right]|\boldsymbol{\xi}_0\right] = E[\widetilde{h}_1(\boldsymbol{\xi}_{t-1})|\boldsymbol{\xi}_0]$ and $E[\widetilde{h}_1(\boldsymbol{\xi}_{t-1}^*)|\boldsymbol{\xi}_0] = E[\widetilde{g} \circ \boldsymbol{g}(\boldsymbol{\xi}_t)] = 0$. Then, we have $\mathfrak{K}_{\mathrm{a}} = \|E[\widetilde{h}_1(\boldsymbol{\xi}_{t-1}) - \widetilde{h}_1(\boldsymbol{\xi}_{t-1}^*)|\boldsymbol{\xi}_0]\|_{p/(p-1)}$. An application of Jensen's inequality yields $\mathfrak{K}_{\mathrm{a}} \leq \|\widetilde{h}_1(\boldsymbol{\xi}_{t-1}) - \widetilde{h}_1(\boldsymbol{\xi}^*)\|_{p/(p-1)} = \alpha_{t-1}^*$.

Finally, $\boldsymbol{\xi}_t$ is a stationary, ergodic Markov chain so the $\lim_{t\longrightarrow\infty} \|P(Y_t \leq y|Y_0) - F(y)\|_{q^*} = 0$. Using the same argument as Lemma A.2, we can obtain $\mathfrak{K}_{\mathrm{b}} \longrightarrow 0$ as $t \longrightarrow \infty$. □

Lemma A.6. *Suppose that*

$$\overline{\lim_{T\longrightarrow\infty}} \sup_{1\leq t\leq T} \sum_{s=1}^{T} s^{-1/2} \|P(\boldsymbol{X}_s \in A_t) - P(\boldsymbol{X}_s \in A_t|\mathcal{F}_0)\|_{2\ell} < \infty, \tag{A.4}$$

where $\{A_t\}_{t=1}^{T}$ are disjoint sets containing the sequence $\{\boldsymbol{X}_t\}_{t=1}^{T}$ so that $\mathcal{X} \subseteq \bigcup_{t=1}^{T} A_t$.

(a) *Let*

$$k_T = O\left(T^{\frac{\ell+1}{2\ell}}\right) \text{ for some } \ell \geq 2, \tag{A.5}$$

then

$$\lim_{T\longrightarrow\infty} \sup_{1\leq t\leq T} \left| \frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_t)}{k_T} - \frac{1}{f(\boldsymbol{X}_t)} \right| = 0. \tag{A.6}$$

(b) *Let*

$$k_T = O\left(T^{\frac{\ell+2}{2\ell}}\right) \text{ for some } \ell \geq 3,$$

then

$$\lim_{T\longrightarrow\infty} \sup_{1\leq t\leq T} T^{1/2} \left| \frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_t)}{k_T} - \frac{1}{f(\boldsymbol{X}_t)} \right| = 0. \tag{A.7}$$

Proof. We prove Part (a) only since the proof of Part (b) essentially uses the same argument.

First, define the event

$$\begin{aligned}
\mathcal{A}_t(\omega) &= \left\{ \omega \in \Omega : \ \left| \frac{T\mathcal{L}(V_1)R_T^N(\boldsymbol{X}_t, k)}{k} - \frac{1}{f(\boldsymbol{X}_t)} \right| < \epsilon \right\} \\
&= \left\{ \omega \in \Omega : \ \frac{k}{T\mathcal{L}(V_1)} \left(\frac{1-\epsilon f(\boldsymbol{X}_t)}{f(\boldsymbol{X}_t)} \right) < R_T^N(\boldsymbol{X}_t, k) < \frac{k}{T\mathcal{L}(V_1)} \left(\frac{1+\epsilon f(\boldsymbol{X}_t)}{f(\boldsymbol{X}_t)} \right) \right\},
\end{aligned}$$

where ϵ is some arbitrarily small positive constant. To prove the *uniform* convergence defined in Eq. (A.6), it is necessary to show the *pointwise* convergence. In view of the Borel-Cantelli lemma, we need to show that

$$\lim_{T\longrightarrow\infty} \sum_{t=1}^{T} P(\mathcal{A}_t^c(\omega)) < \infty,$$

where

$$\begin{aligned}
\mathcal{A}_t^c(\omega) &= \left\{ \omega \in \Omega : \ R_T(\boldsymbol{X}_t, k) \geq \left(\frac{k}{T\mathcal{L}(V_1)} \left(\frac{1+\epsilon f(\boldsymbol{X}_t)}{f(\boldsymbol{X}_t)} \right) \right)^{1/N} \right\} \\
&\quad \bigcup \left\{ \omega \in \Omega : \ R_T(\boldsymbol{X}_t, k) \leq \left(\frac{k}{T\mathcal{L}(V_1)} \left(\frac{1-\epsilon f(\boldsymbol{X}_t)}{f(\boldsymbol{X}_t)} \right) \right)^{1/N} \right\} \\
&= \mathcal{A}_{1,t}^c(\omega) \cup \mathcal{A}_{2,t}^c(\omega).
\end{aligned}$$

Define $q_{1,T}(\boldsymbol{X}_t) = P\left(\|\boldsymbol{X}_s - \boldsymbol{X}_t\| \leq \delta_1(T)\right)$, where $\delta_1(T) = \left(\frac{k}{T\mathcal{L}(V_1)} \left(\frac{1+\epsilon f(\boldsymbol{X}_t)}{f(\boldsymbol{X}_t)} \right) \right)^{1/N}$, for some $s \neq t$; and $q_{2,T}(\boldsymbol{X}_t) = P\left(\|\boldsymbol{X}_s - \boldsymbol{X}_t\| \leq \delta_2(T)\right)$, where $\delta_2(T) = \left(\frac{k}{T\mathcal{L}(V_1)} \left(\frac{1-\epsilon f(\boldsymbol{X}_t)}{f(\boldsymbol{X}_t)} \right) \right)^{1/N}$. Using result B.1 and the fact that $\mathcal{L}(\|\boldsymbol{X}_s - \boldsymbol{X}_t\| \leq \delta_1(T)) = \frac{k}{T} \frac{1+\epsilon f(\boldsymbol{X}_t)}{f(\boldsymbol{X}_t)}$ together with $\mathcal{L}(\|\boldsymbol{X}_s - \boldsymbol{X}_t\| \leq \delta_2(T)) = \frac{k}{T} \frac{1-\epsilon f(\boldsymbol{X}_t)}{f(\boldsymbol{X}_t)}$, we obtain the following *pointwise* limits:

$$\limsup_{T\longrightarrow\infty} \frac{Tq_{1,T}(\boldsymbol{X}_t)}{k} = 1 + \epsilon f(\boldsymbol{X}_t), \tag{A.8}$$

$$\limsup_{T\longrightarrow\infty} \frac{Tq_{2,T}(\boldsymbol{X}_t)}{k} = 1 - \epsilon f(\boldsymbol{X}_t). \tag{A.9}$$

Next, let $K_{1,s}(\boldsymbol{X}_t) = \mathbb{I}(\|\boldsymbol{X}_s - \boldsymbol{X}_t\| \geq \delta_1(T))$ with $E[K_{1,s}(\boldsymbol{X}_t)|\mathcal{F}_{\boldsymbol{X}_t}] = 1 - q_{1,T}(\boldsymbol{X}_t)$ and $K_{2,s}(\boldsymbol{X}_t) = \mathbb{I}(\|\boldsymbol{X}_s - \boldsymbol{X}_t\| \leq \delta_2(T))$ with $E[K_{1,s}(\boldsymbol{X}_t)|\mathcal{F}_{\boldsymbol{X}_t}] = q_{2,T}(\boldsymbol{X}_t)$ denote conditional Bernoulli random variables. Then we have

$$\begin{aligned} P\left(\mathcal{A}_{1,t}^c(\omega)\right) &\leq P\left(\sum_{s=1}^{T} K_{1,s}(\boldsymbol{X}_t) \geq T - k \middle| \mathcal{F}_{\boldsymbol{X}_t}\right) \\ &= P\left(\sum_{s=1}^{T} (K_{1,s}(\boldsymbol{X}_t) - E[K_{1,s}(\boldsymbol{X}_t)|\mathcal{F}_{\boldsymbol{X}_t}]) \geq Tq_{1,T}(\boldsymbol{X}_t) - k \middle| \mathcal{F}_{\boldsymbol{X}_t}\right), \qquad \text{(A.10)} \\ P\left(\mathcal{A}_{2,t}^c(\omega)\right) &\leq P\left(\sum_{s=1}^{T} K_{2,s}(\boldsymbol{X}_t) \geq k \middle| \mathcal{F}_{\boldsymbol{X}_t}\right) \\ &= P\left(\sum_{s=1}^{T} (K_{2,s}(\boldsymbol{X}_t) - E[K_{2,s}(\boldsymbol{X}_t)|\mathcal{F}_{\boldsymbol{X}_t}]) \geq k - Tq_{2,T}(\boldsymbol{X}_t) \middle| \mathcal{F}_{\boldsymbol{X}_t}\right). \qquad \text{(A.11)} \end{aligned}$$

Since the subspace $\mathcal{X}$ is compact, without any loss of generality one may use $\{\boldsymbol{X}_t\}_{t=1}^T$, for a sufficiently large T, as a countable dense set in $\mathcal{X}$; and then find a sequence of positive constants, $\{\delta_t\}_{t=1}^T$, such that it is possible to construct balls, $B_t = \{\boldsymbol{x} \in \mathcal{X} : \|\boldsymbol{x} - \boldsymbol{X}_t\| < \delta_t\}$, with $\mathcal{L}(\partial B_t) = 0$ for each $1 \leq t \leq T$. Finally, set $A_1 = B_1$ and $A_{t+1} = B_{t+1} \setminus \bigcup_{s=1}^t B_s$ for $1 \leq t < T$. Then the space $\mathcal{X}$ can be covered by a finite sequence of disjoint sets, $\{A_t\}_{t=1}^T$. It immediately follows that

$$\begin{aligned} \frac{Tq_{1,T}(\boldsymbol{X}_t)}{k_T} &\leq 1 + \epsilon \sup_{\boldsymbol{x} \in B_t} f(\boldsymbol{x}) = 1 + a_t(\epsilon), \qquad \text{(A.12)} \\ \frac{Tq_{2,T}(\boldsymbol{X}_t)}{k_T} &\geq 1 - \epsilon \sup_{\boldsymbol{x} \in B_t} f(\boldsymbol{x}) = 1 - a_t(\epsilon). \qquad \text{(A.13)} \end{aligned}$$

We now proceed to bound the term defined in Eq. (A.10). An application of Tchebyshev's inequality yields for any integer $\ell > 1$:

$$\begin{aligned} & P\left(\sum_{s=1}^{T} (K_{1,s}(\boldsymbol{X}_t) - E[K_{1,s}(\boldsymbol{X}_t)|\mathcal{F}_{\boldsymbol{X}_t}]) \geq Tq_{1,T}(\boldsymbol{X}_t)\left(1 - \frac{k_T}{Tq_{1,T}(\boldsymbol{X}_t)}\right) \middle| \mathcal{F}_{\boldsymbol{X}_t}\right) \\ \leq\ & P\left(\max_{1 \leq \tau \leq T} \left\{\sum_{s=1}^{\tau} (K_{1,s}(\boldsymbol{X}_t) - E[K_{1,s}(\boldsymbol{X}_t)|\mathcal{F}_{\boldsymbol{X}_t}])\right\} \geq Tq_{1,T}(\boldsymbol{X}_t)\left(1 - \frac{k_T}{Tq_{1,T}(\boldsymbol{X}_t)}\right) \middle| \mathcal{F}_{\boldsymbol{X}_t}\right) \\ \leq\ & \frac{E\left[\left|\max_{1 \leq \tau \leq T}\left\{\sum_{s=1}^{\tau} (K_{1,s}(\boldsymbol{X}_t) - E[K_{1,s}(\boldsymbol{X}_t)|\mathcal{F}_{\boldsymbol{X}_t}])\right\}\right|^{2\ell} \middle| \mathcal{F}_{\boldsymbol{X}_t}\right]}{\left|Tq_{1,T}(\boldsymbol{X}_t)\frac{a_t(\epsilon)}{a_t(\epsilon)+1}\right|^{2\ell}} = \frac{\mathcal{B}_1}{\mathcal{B}_2}. \end{aligned}$$

Using Result B.3 and the inequality $\|E[g(\boldsymbol{X})|\mathcal{B}]\|_p \leq \|g(\boldsymbol{X})\|_p$ for any measurable function of $\boldsymbol{X}$ and a Borel algebra, $\mathcal{B}$, yields:

$$\begin{aligned} \mathcal{B}_1 &\leq C_{2\ell} T^{\ell} \Big\{ \|K_{1,1}(\boldsymbol{X}_t) - E[K_{1,1}(\boldsymbol{X}_t)|\mathcal{F}_{\boldsymbol{X}_t}]\|_{2\ell, \mathcal{F}_{\boldsymbol{X}_t}} \\ &+ 240 \sum_{s=1}^{T} s^{-1/2} \left\| E\left[K_{1,s}(\boldsymbol{X}_t) - E[K_{1,s}(\boldsymbol{X}_t)|\mathcal{F}_{\boldsymbol{X}_t}] \middle| \mathcal{F}_{\boldsymbol{X}_0}\right]\right\|_{2\ell} \Big\}^{2\ell} \\ &= C_{2\ell} T^{\ell} \{\mathcal{B}_{1\text{a}} + \mathcal{B}_{1\text{b}}\}, \end{aligned}$$

where

$$\begin{aligned}
\mathcal{B}_{1\mathrm{a}} &= \|K_{1,1}(\boldsymbol{X}_t) + q_{1,T}(\boldsymbol{X}_t) - 1\|_{2\ell,\mathcal{F}_{\boldsymbol{X}_t}} \\
&= \left\{(q_{1,T}(\boldsymbol{X}_t) - 1)^{2\ell} q_{1,T}(\boldsymbol{X}_t) + q_{1,T}^{2\ell}(\boldsymbol{X}_t)(1 - q_{1,T}(\boldsymbol{X}_t))\right\}^{1/2\ell}. \\
\mathcal{B}_{1\mathrm{b}} &= \left\| E\left[K_{1,s}(\boldsymbol{X}_t) + q_{1,T}(\boldsymbol{X}_t) - 1 \middle| \mathcal{F}_{\boldsymbol{X}_0}\right]\right\|_{2\ell} \\
&= \left\|q_{1,T}(\boldsymbol{X}_t) - P\left(\|\boldsymbol{X}_s - \boldsymbol{X}_t\| \leq \delta_1(T) \middle| \mathcal{F}_{\boldsymbol{X}_0}\right)\right\|_{2\ell} \\
&\leq \sup_{1\leq t\leq T} \left\|P\left(\boldsymbol{X}_s \in A_t\right) - P\left(\boldsymbol{X}_s \in A_t \middle| \mathcal{F}_{\boldsymbol{X}_0}\right)\right\|_{2\ell},
\end{aligned}$$

where the last inequality holds for a sufficiently large T. In view of the inequality $(a+b) \leq 2^{1-1/p}\left(a^p + b^p\right)^{1/p}$ for $a > 0$, $b > 0$, and $p \geq 1$, using Eq. (A.8) we obtain the ratio

$$\begin{aligned}
\frac{\mathcal{B}_1}{\mathcal{B}_2} &\leq 2^{2\ell-1} \frac{C_{2\ell}\frac{T^\ell}{k_T^{2\ell}}}{\left((1+\epsilon f(\boldsymbol{X}_t))\frac{a_t(\epsilon)-1}{a_t(\epsilon)}\right)^{2\ell}} \left\{(q_{1,T}(\boldsymbol{X}_t) - 1)^{2\ell} q_{1,T}(\boldsymbol{X}_t) + q_{1,T}^{2\ell}(\boldsymbol{X}_t)(1 - q_{1,T}(\boldsymbol{X}_t))\right. \\
&+ \left. 240\left(\sum_{s=1}^{T} s^{-1/2} \sup_{1\leq t\leq T}\left\|P\left(\boldsymbol{X}_s \in A_t\right) - P\left(\boldsymbol{X}_s \in A_t \middle| \mathcal{F}_{\boldsymbol{X}_0}\right)\right\|_{2\ell}\right)^{2\ell}\right\}.
\end{aligned}$$

It then follows that:

$$\begin{aligned}
\sum_{t=1}^{T} P\left(\mathcal{A}_{1,t}^c(\omega)\right) &\leq 2^{2\ell-1} C_{2\ell} \frac{T^{\ell+1}}{k_T^{2\ell}} \left\{\sup_{1\leq t\leq T} \frac{q_{1,T}(\boldsymbol{X}_t)(1 - q_{1,T}(\boldsymbol{X}_t))^2 + q_{1,T}^{2\ell}(1 - q_{1,T}(\boldsymbol{X}_t))}{\left((1+\epsilon f(\boldsymbol{X}_t))\frac{a_t(\epsilon)}{a_t(\epsilon)+1}\right)^{2\ell}}\right. \\
&+ \left. 240^{2\ell} \sup_{1\leq t\leq T}\left(\sum_{s=1}^{T} s^{-1/2}\left\|P\left(\boldsymbol{X}_s \in A_t\right) - P\left(\boldsymbol{X}_s \in A_t \middle| \mathcal{F}_{\boldsymbol{X}_0}\right)\right\|_{2\ell}\right)^{2\ell}\right\}.
\end{aligned}$$

Since the first term inside the brackets in the above equation is always finite, it follows from Eqs. (A.4) and (A.5) that

$$\overline{\lim_{T\longrightarrow\infty}} \sum_{t=1}^{T} P\left(\mathcal{A}_{1,t}^c(\omega)\right) \leq \infty. \tag{A.14}$$

Using the same argument results in:

$$\begin{aligned}
\sum_{t=1}^{T} P\left(\mathcal{A}_{2,t}^c(\omega)\right) &\leq 2^{2\ell-1} C_{2\ell} \frac{T^{\ell+1}}{k_T^{2\ell}} \left\{\sup_{1\leq t\leq T} \frac{q_{2,T}(\boldsymbol{X}_t)(1 - q_{2,T}(\boldsymbol{X}_t))^{2\ell} + q_{2,T}^{2\ell}(\boldsymbol{X}_t)(1 - q_{2,T}(\boldsymbol{X}_t))}{\left((1-\epsilon f(\boldsymbol{X}_t))\frac{a_t(\epsilon)}{1-a_t(\epsilon)}\right)^{2\ell}}\right. \\
&+ \left. 240^{2\ell} \sup_{1\leq t\leq T}\left(\sum_{s=1}^{T} s^{-1/2}\left\|P\left(\boldsymbol{X}_s \in A_t\right) - P\left(\boldsymbol{X}_s \in A_t \middle| \mathcal{F}_{\boldsymbol{X}_0}\right)\right\|_{2\ell}\right)^{2\ell}\right\}.
\end{aligned}$$

Hence, it follows that:

$$\overline{\lim_{T\longrightarrow\infty}} \sum_{t=1}^{T} P\left(\mathcal{A}_{2,t}^c(\omega)\right) \leq \infty. \tag{A.15}$$

Therefore, in view of Eqs. (A.14) and (A.15), the result follows. □

B Known Results

Lemma B.1. *(The 'Lebesgue Density Theorem') Let $\mathcal{Q}$ be a subclass of the Borel sets of $\mathbb{R}^N$ with the property that $\sup_{Q\in\mathcal{Q}} \frac{\mathcal{L}(Q^*)}{\mathcal{L}(Q)} \leq c < \infty$ for some constant c, where Q^* is the smallest cube centered at the origin that contains Q, and $\mathcal{L}(\cdot)$ denotes the volume of a set. Let $\mathcal{Q}_r$ be the subclass of $\mathcal{Q}$ containing only sets Q with $\mathcal{L}(Q) \leq r$. Let f be any density on $\mathbb{R}^N$. Let $\boldsymbol{z} + Q$ denote the translation of Q by $\boldsymbol{z}$. Then, for almost all $\boldsymbol{x}$,*

$$\lim_{r\downarrow 0} \sup_{Q\in\mathcal{Q}_r} \left| \frac{1}{\mathcal{L}(Q)} \int_{\boldsymbol{x}+Q} f(\boldsymbol{y})d\boldsymbol{y} - f(\boldsymbol{x}) \right| = 0.$$

Proof. See Devroye and Lugosi (2001, p. 42). □

The points $\boldsymbol{x}$ at which this convergence takes place are called Lebesgue points for f. Classes that satisfy the condition are the classes of all cubes, or all balls on $\mathbb{R}^N$.

Lemma B.2. *(The 'Generalized Lebesgue Differentiation Theorem') Given an equicontinuous function, $g(\cdot)$, such that $\int_{\mathbb{R}^N} |g(\boldsymbol{x})|^p \mu(d\boldsymbol{x}) < \infty$ for some $p > 0$ and a Lebesgue measure, μ, then*

$$\lim_{h\longrightarrow 0} \frac{1}{\mu(Q_{\boldsymbol{x}}(h))} \int_{Q_{\boldsymbol{x}}(h)} |g(\boldsymbol{y}) - g(\boldsymbol{x})|^p \, d\mu(\boldsymbol{y}) = 0 \text{ a.e. } (\mu),$$

where $Q_{\boldsymbol{x}}(h)$ is a cube of center $\boldsymbol{x}$ with an edge length h.

Proof. The proof of this result immediately follows from the equicontinuity of $g(\boldsymbol{x})$ couple with the Lebesgue differentiation theorem stated in Wheeden and Zygmund (1977, p. 189). □

Lemma B.3. *Let $S_n = \sum_{i=1}^n X_i$, where $X_i = X_0 \circ \mathcal{T}^i$ is a stationary process, and $S_n^* = \max_{j\leq n} |S_j|$. Assume that $E[|X_1|^p] < \infty$, $p \geq 2$. Then*

$$\|S_n^*\|_p \leq C_p^{1/p} n^{1/2} \left[\|X_1\|_p + 240 \sum_{i=1}^n i^{-1/2} \|E[X_i|\mathcal{F}_0]\|_p \right],$$

where C_p is a generic constant that depends only on p.

Proof. See Peligrad et al. (2007). □

Lemma B.4. *Suppose that ξ is $\mathcal{F}_t$-measurable and η is $\mathcal{F}_{t+\tau}$-measurable for some $t, \tau \in \mathbb{N}^+$. If $E|\xi|^q < \infty$ and $E|\eta|^r < \infty$ for some q, $r > 1$ and $1/q + 1/r < 1$, then $\operatorname{cov}(\xi, \nu) \leq 6\|\xi\|_q \|\eta\|_r \alpha_\tau^{1-1/q-1/r}$, where α_τ is the conventional strong mixing coefficient.*

Proof. See Davydov (1968). □

Lemma B.5. *Let $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$, let $\mathcal{L}$ denote the Lebesgue measure on $(\Omega, \mathcal{F})$, and let $f = f(x) \in L_1$. Put*

$$f_n(x) = 2^n \int_{k2^{-n}}^{(k+1)2^{-n}} f(y)dy, \;\; k2^{-n} \leq x < (k + 1)2^{-n}.$$

Then $f_n(x) \longrightarrow f(x)$ ($\mathcal{L}$-a.s.).

Proof. See exercise 5 in Shiryaev (1996, p. 515). □

References

Andrews, D. W. K. (1984), "Nonstrong mixing autoregressive processes," *Journal of Applied Probability*, 21, 930–934.

Apostol, T. M. (1969), *Multi-Variable Calculus and Linear Algebra, with Applications to Differential Equations and Probability*, Wiley & Sons, 2nd ed.

Arnold, B. C., Castillo, E., and Sarabia, J. M. (2009), "Multivariate order statistics via multivariate concomitants," *Journal of Multivariate Analysis*, 100, 946–951.

Barnett, V., Green, P. J., and Robinson, A. (1976), "Concomitants and Correlation Estimates," *Biometrika*, 63, 323–328.

Bhattacharya, P. K. and Mack, Y. P. (1987), "Weak convergence of k-nn density and regression estimators with varying k and applications," *Annals of Statistics*, 15, 976–994.

Bickel, P. J. and Bühlmann, P. (1999), "A new mixing notion and functional central limit theorems for a sieve bootstrap in time series," *Bernoulli*, 5, 413–446.

Boente, G. and Fraiman, R. (1988), "Consistency of a nonparametric estimate of a density function for dependent variables," *Journal of Multivariate Analysis*, 25, 90–99.

— (1990), "Asymptotic distribution of robust estimators for nonparametric models from mixing processes," *Annals of Statistics*, 18, 891–906.

Bradley, R. (1986), "Basic properties of strong mixing conditions," in *Dependence in Probability and Statistics*, eds. Eberlein, E. and Taqqu, M., Boston: Birkhäuser, Progress in Probability and Statistics, pp. 165 – 192.

Bradley, R. C. (2007), *Introduction to Strong Mixing Conditions*, vol. I, II, III, Utah, USA: Kendrick Press.

Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997), "Generalized partially linear single-index models," *Journal of the American Statistical Association*, 92, 477 – 489.

Chow, Y. S. and Teicher, H. (1978), *Probability Theory: Independece, interchangeability, Martingales*, Springer-Verlag New York, Inc, 1st ed.

Chu, B. M. and Jacho-Chávez, D. T. (2012), "k-Nearest Neighbour Estimation of Inverse-Density-Weighted Expectations with Dependent Data," *Econometric Theory*, 28, 769 – 803.

Csörgö, S. (1981), *Empirical Characteristic Functions*, Carleton Mathematical Lecture Notes, Vol. 26. Carleton University, Ottawa.

David, H. A. and Galambos, J. (1974), "The asymptotic theory of concomitants of order statistics," *Journal of Applied Probability*, 11, 762–770.

David, H. A. and Nagaraja, H. N. (1998), “Concomitants of order statistics,” in *Handbook of Statistics, Vol. 16, Order Statistics: Theory and Methods*, eds. Balakrishnan, N. and Rao, C. R., Elsevier Science, pp. 487–513.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford, New York: Oxford University Press.

Davydov, Y. A. (1968), “Convergence of distributions generated by stationary stochastic processes,” *Theory of Probability and Its Applications*, 13, 691–696.

Dedecker, J. and Prieur, C. (2005), “New dependence coefficients. Examples and applications to statistics,” *Probability Theory and Related Fields*, 132, 203–236.

— (2007), “An empirical central limit theorem for dependent sequences,” *Stochastic Processes and Their Applications*, 117, 121–142.

Devroye, L. and Lugosi, G. (2001), *Combinatorial Methods in Density Estimation*, Springer-Verlag.

Engle, R. F., Granger, C. W. J., Rice, J., and Weiss, A. (1986), “Semiparametric Estimates of the Relation Between Weather and Electricity Sales,” *Journal of the American Statistical Association*, 81, 310 – 320.

Francq, C. and Zakoïan, J.-M. (2010), *GARCH Models: Structure, Statistical Inference and Financial Applications*, John Wiley & Sons.

Gordin, M. I. (1969), “The central limit theorem for stationary processes,” *Doklady Akademii Nauk SSSR*, 188, 739–741.

Hall, P. and Yatchew, A. (2005), “Unified approach to testing functional hypotheses in semiparametric contexts,” *Journal of Econometrics*, 127, 225–252.

Härdle, W., Hall, P., and Ichimura, H. (1993), “Optimal smoothing in single-index models,” *Annals of Statistics*, 21, 157 – 178.

Härdle, W. and Stoker, T. M. (1989), “Investigating Smooth Multiple Regression by the Method of Average Derivatives,” *Journal of the American Statistical Association*, 84, 986–995.

Hart, J. D. and Vieu, P. (1990), “Data-Driven Bandwidth Choice for Density Estimation Based on Dependent Data,” *The Annals of Statistics*, 18, 873–890.

Hausman, J. A. and Newey, W. K. (1995), “Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss,” *Econometrica*, 63, 1445–1476.

Hayfield, T. and Racine, J. S. (2008), “Nonparametric Econometrics: The np Package,” *Journal of Statistical Software*, 27, 1–32.

Hong, Y. and White, H. (2005), “Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence,” *Econometrica*, 73, 837–901.

Huynh, K. P. and Jacho-Chávez, D. T. (2009), “Growth and governance: A nonparametric analysis,” *Journal of Comparative Economics*, 37, 121–143.

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

Ichimura, H. (1993), “Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models,” *Journal of Econometrics*, 58, 71–120.

Jacho-Chávez, D. T. (2008), “k nearest-neighbor estimation of inverse density weighted expectations,” *Economics Bulletin*, 3, 1–6.

Khaledi, B.-E. and Kochar, S. (2000), “Stochastic comparisons and dependence among concomitants of order statistics,” *Journal of Multivariate Analysis*, 73, 262–281.

Koroljuk, V. S. and Borovskich, Y. V. (1994), *Theory of* U-*Statistics*, Dordrecht/Boston/London: Kluwer Academic Publishers.

Lewbel, A. (1998), “Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors,” *Econometrica*, 66, 105–121.

Lewbel, A. and Schennach, S. M. (2007), “A simple ordered data estimator for inverse density weighted expectations,” *Journal of Econometrics*, 136, 189–211.

Li, J. and Tran, L. T. (2009), “Nonparametric estimation of conditional expectation,” *Journal of Statistical Planning and Inference*, 139, 164–175.

Loftsgaarden, D. O. and Quesenberry, C. P. (1965), “A nonparametric estimate of a multivariate density function,” *Annals of Mathematical Statistics*, 36, 1049–1051.

Lu, X., Lian, H., and Liu, W. (2012), “Semiparametric Estimation for Inverse Density Weighted Expectations when Responses are Missing at Random,” *Journal of Nonparametric Statistics*, 24, 139 – 159.

Mack, Y. P. and Rosenblatt, M. (1979), “Multivariate k-Nearest Neighbor Density Estimates,” *Journal of Multivariate Analysis*, 9, 1–15.

Moore, D. S. and Yackel, J. W. (1977a), “Consistency properties of nearest neighbor density function estimators,” *Annals of Statistics*, 5, 143–154.

— (1977b), “Large sample properties of nearest neighbor density estimators,” in *Statistical Decision Theory and Related Topics II*, eds. Gupta, S. S. and Moore, D. S., New York: Academic Press.

Nagaraja, H. N. and David, H. A. (1994), “Distribution of the maximum of concomitants of selected order statistics,” *Annals of Statistics*, 22, 478–494.

Peligrad, M., Utev, S., and Wu, W. B. (2007), "A Maximal $\mathbb{L}_p$-inequality for stationary sequences and its applications," *Proceedings of The American Mathematical Society*, 135, 541–550.

Pham, T. D. (1986), "The mixing property of bilinear and generalized random coefficient autoregressive models," *Stochastic Processes and their Applications*, 23, 291 – 300.

Pham, T. D. and Tran, L. T. (1985), "Some mixing properties of time series models," *Stochastic Processes and their Applications*, 13, 297–303.

Philipp, W. and Stout, W. F. (1975), "Almost sure invariance principles for partial sums of weakly dependent random variables," *Memoirs of the American Mathematical Society*, 2.

Priestley, M. (1988), *Nonlinear and Nonstationary Time Series Analysis*, Academic Press.

Puri, M. L. and Tran, L. T. (1980), "Empirical Distribution Functions and Functions of Order Statistics for Mixing Random Variables," *Journal of Multivariate Analysis*, 10, 405–425.

Rao, B. L. S. P. (2009), "Conditional independence, conditional mixing and conditional association," *Annals of the Institute of Statistical Mathematics*, 61, 441–460.

Rinott, Y. and Rotar, V. (1999), "Some bounds on the rate of convergence in the CLT for martingales. I," *Theory of Probability and its Applications*, 43, 604 – 619.

Rio, E. (2000), *Théorie asymptotique des processus aléatoires faiblement dépendants*, vol. 31 of *Collection Mathématiques & Applications*, Berlin: Springer.

Robinson, P. M. (1987), "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 875–891.

— (1995), "Nearest-neighbour estimation of semiparametric regression models," *Journal of Nonparametric Statistics*, 5, 33–41.

Rosenblatt, M. (1956a), "A central limit theorem and a strong mixing condition," *Procedure of the National Academy of Science USA*, 42, 43–47.

— (1956b), "Remarks on some non-parametric estimates of a density function," *The Annals of Mathematical Statistics*, 27, 832–837.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons.

Shiryaev, A. N. (1996), *Probability*, Springer, 2nd ed.

Stokes, S. L. (1977), "Ranked set sampling with concomitant variables," *Communication in Statistics, Part A – Theory and Methods*, 6, 1207–1211.

Stute, W. (1984), "Asymptotic Normality of Nearest Neighbor Regression Function Estimates," *Annals of Statistics*, 12, 917–926.

— (1993), “U-functions of concomitants of order statistics,” *Probability and Mathematical Statistics*, 14, 143–155.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford University Press.

Tran, L. T. and Wu, B. (1993), “Order statistics for nonstationary time series,” *Annals of Institute of Statistical Mathematics*, 45, 665–686.

Tran, L. T. and Yakowitz, S. (1993), “Nearest neighbor estimators for random fields,” *Journal of Multivariate Analysis*, 44, 23–46.

Truong, Y. K. and Stone, C. J. (1992), “Nonparametric function estimation involving time series,” *Annals of Statistics*, 20, 77–97.

Varadhan, S. R. S. (2001), *Probability Theory*, American Mathematical Society, Courant Institute of Mathematical Sciences.

Watterson, G. A. (1958), “Linear estimation in censored samples from multivariate normal populations,” *Annals of Mathematical Statistics*, 30, 814–824.

Wheeden, R. L. and Zygmund, A. (1977), *Measure and Integral*, Dekker, New York.

White, H. and Domowitz, I. (1984), “Nonlinear Regression with Dependent Observations,” *Econometrica*, 52, 143–161.

Wu, B. (1988), “On Order Statistics in Time Series Analysis,” Ph.D. thesis, Indiana University.

Wu, W. B. (2005), “Nonlinear system theory: Another look at dependence,” *Proceedings of the National Academy of Sciences of the United States of America*, 102, 14150–14154.

— (2007), “Strong invariance principles for dependent random variables,” *Annals of Probability*, 35, 2294–2320.

Wu, W. B. and Woodroofe, M. (2004), “Martingale approximations for sums of stationary processes,” *Annals of Probability*, 32, 1674–1690.

Yang, S. S. (1977), “General distribution theory of the concomitants of order statistics,” *Annals of Statistics*, 5, 996–1002.

— (1981a), “Linear combinations of concomitants of order statistics with application to testing and estimation,” *Annals of the Institute of Statistical Mathematics*, 33, 463–470.

— (1981b), “Linear functions of concomitants of order statistics with application to nonparametric estimation of a regression function,” *Journal of the American Statistical Association*, 76, 658–662.

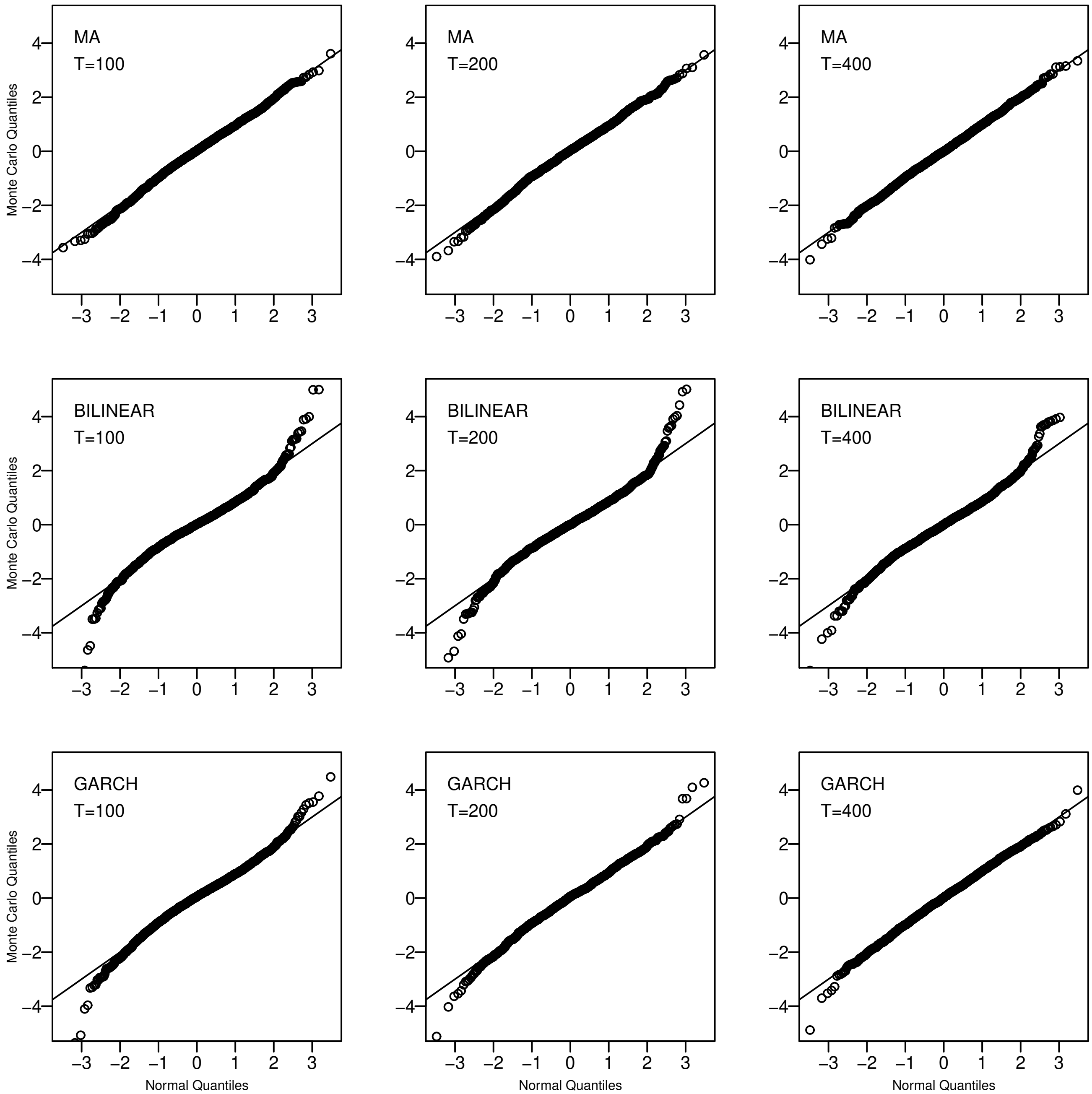
Yatchew, A. (2003), *Semiparametric Regression for the Applied Econometrician*, Themes in Modern Econometrics, Cambridge University Press, 1st ed.

Table 1: Monte Carlo Performance for Cases 1 & 2

		Case 1			Case 2		
Model	n	Bias	Std. Dev.	IQR	Bias	Std. Dev.	IQR
MA	100	0.000	0.034	0.043	-0.003	0.376	0.492
	200	0.000	0.025	0.032	-0.005	0.269	0.356
	400	0.000	0.017	0.023	-0.003	0.187	0.242
BILINEAR	100	-0.003	0.155	0.170	0.031	1.580	1.647
	200	-0.002	0.116	0.131	0.029	1.078	1.216
	400	0.001	0.080	0.095	0.007	0.753	0.919
GARCH	100	0.000	0.065	0.076	0.019	0.661	0.808
	200	0.000	0.046	0.058	0.015	0.461	0.604
	400	0.000	0.032	0.042	-0.001	0.327	0.431

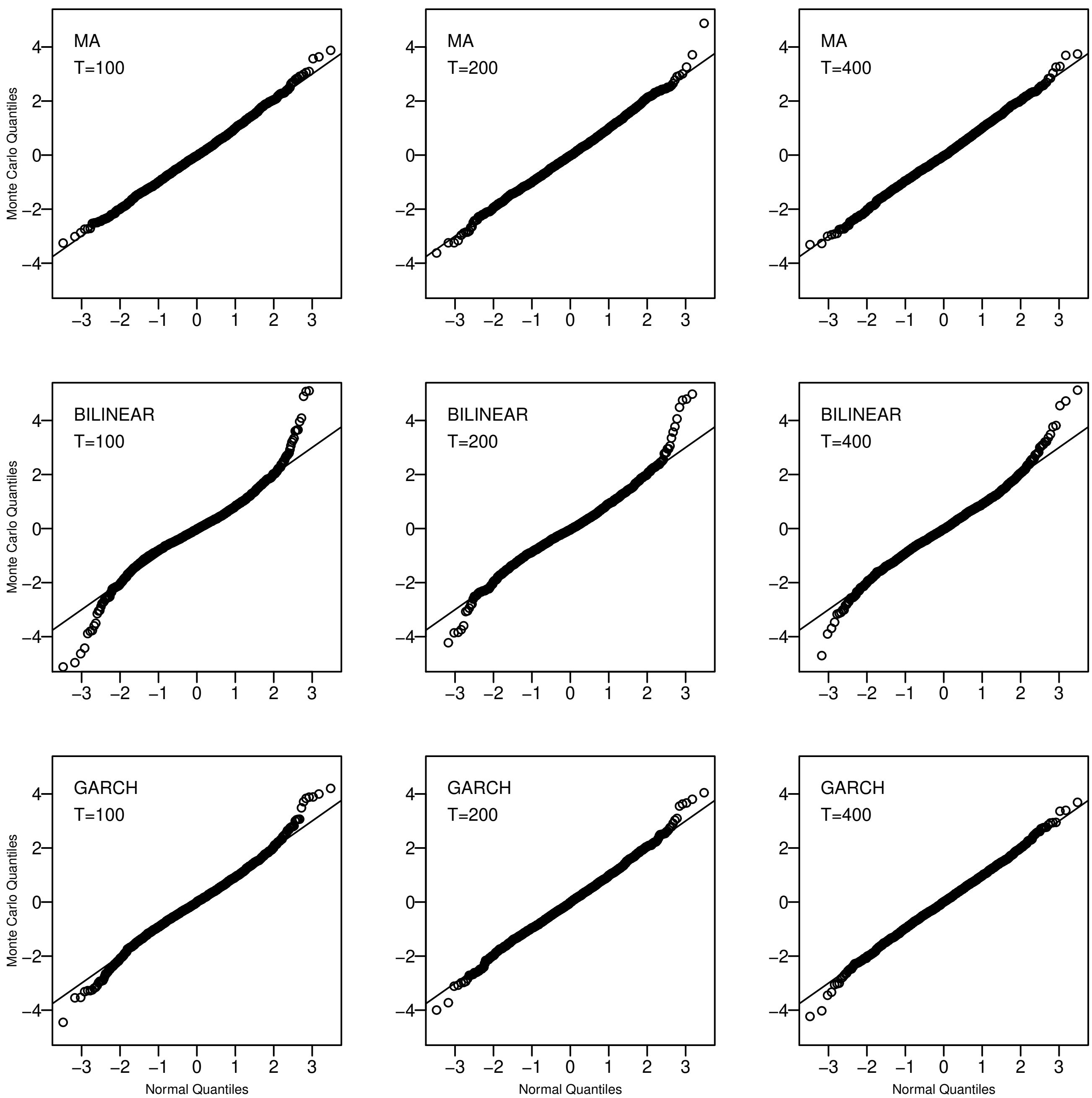
Note: This table reports Monte Carlo Bias (Bias), Standard Deviation (Std. Dev.) and Inter-Quartile Range (IQR) for Cases 1 and 2. Results based on 2000 Monte Carlo replications corresponding to the Moving Average (MA), Bilinear (BILINEAR) and Generalized Autoregressive Conditional Heteroskedasticity (1,1) (GARCH) models for X_t in Example 1.

Figure 1: Case 1



Note: QQ-Plots of the standardize Monte Carlo sample vs the theoretical quantiles of a standard normal distribution, with a 45-degrees line. Results based on 2000 Monte Carlo replications correspond to the Moving Average (MA), Bilinear (BILINEAR) and Generalized Autoregressive Conditional Heteroskedasticity (1,1) (GARCH) models for X_t in Example 1.

Figure 2: Case 2



Note: QQ-Plots of the standardize Monte Carlo sample vs the theoretical quantiles of a standard normal distribution, with a 45-degrees line. Results based on 2000 Monte Carlo replications correspond to the Moving Average (MA), Bilinear (BILINEAR) and Generalized Autoregressive Conditional Heteroskedasticity (1,1) (GARCH) models for X_t in Example 1.