$k$-Nearest Neighbour Estimation of Inverse-Density-Weighted Expectations with Dependent Data*

Ba Chu† Carleton University
David T. Jacho-Chávez‡ Indiana University

Abstract

This paper considers the problem of estimating expected values of functions that are inversely weighted by an unknown density using the $k$-Nearest Neighbour method. It establishes the $\sqrt{T}$-consistency and asymptotic normality of an estimator that allows for time series data. In the random sampling scheme, the proposed estimator is also shown to be asymptotically semiparametric efficient. Monte Carlo experiments show that the proposed estimator performs as good as alternative methods in finite sample applications.

*EEL codes: C14, C22
Keywords: Semiparametric estimation, nearest neighbour density estimation.

1 Introduction

This paper addresses the problem of estimating definite integrals of conditional expectations such as

$$\theta_0 = \int_{x \in \mathcal{X}} E(W|X = x)dx = E\left[\frac{Y}{f(X)}\right],$$

(1.1)

*The authors would like to thank the co-editor and three referees for their valuable comments that led to corrections and various improvements in the paper. Francesco Bravo and Juan Carlos Escanciano also provided many helpful comments and suggestions on an earlier draft of this paper. The usual disclaimer applies.

†Department of Economics, Carleton University, B-857 Loeb Building, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada. Phone: +1 (613) 520 2600 (ext. 1546). Fax: +1 (613) 520 3906. E-mail: ba_chu@carleton.ca. Web Page: http://http-server.carleton.ca/~bchu/

‡Department of Economics, Indiana University, Wylie Hall 251, 100 South Woodlawn Avenue, Bloomington, IN 47405–7104, USA. Phone: +1 (812) 855 7928. Fax: +1 (812) 855 3736. E-mail: djachoch@indiana.edu. Web Page: http://mypage.iu.edu/~djachoch/
for some random variable $W$, where $X \subset \text{Supp}(X)$. The last equality follows from rewriting $Y = W\mathbb{1}(X \in \mathcal{X})$, where $\mathbb{1}(\cdot)$ is the indicator function that equals one if its argument is true and zero otherwise. $f(\cdot)$ represents the unknown marginal density of the continuous scalar random variable $X$, and $E[\cdot]$ is the expectation taken with respect to the joint distribution of $(Y, X)$. Consistent estimation of objects like (1.1) are important in economics because, as pointed out by Lewbel and Schennach (2007), numerous existing semiparametric estimators make direct or indirect use of (1.1), see e.g. Härdle and Stoker (1989), Hausman and Newey (1995), Newey (1997), McFadden (1999), Lewbel (1997, 1998, 2000, 2007), Honoré and Lewbel (2002), Newey and Ruud (2005), Hong and White (2005), Hall and Yatchew (2005), Khan and Lewbel (2007) and Lewbel, Linton, and McFadden (in press). Furthermore, its analysis with dependent data is also important because objects like (1.1) often represent consumer surpluses that need to be calculated with repeated observations of quantities demanded, and prices in time, see e.g. Newey and McFadden (1994, Section 8, p. 2195). In particular, if one has data $\{Y_t, X_t\}_{t=1}^T$, a natural estimator of (1.1) is

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^T \frac{Y_t}{f(X_t)},$$

(1.2)

where $\hat{f}(\cdot)$ denotes a consistent nonparametric density estimator. Various authors have established the $\sqrt{T}$-consistency and asymptotic normality of (1.2) using Rosenblatt’s (1956b) kernel estimator in place of $\hat{f}(\cdot)$ for example, see e.g. Lewbel (1997, 1998, 2000, 2007), Honoré and Lewbel (2002) among others. However, their results often require rather strong assumptions on $f(\cdot)$ such as smoothness conditions and compact support upon which it is bounded away from zero, see e.g. Khan and Lewbel (2007) and Jacho-Chávez (2010). Alternatively, this paper establishes the weak convergence of a version of (1.2) that imposes no explicit restrictions on the support of $X$, apart from general moment conditions on the joint distribution of $(Y, X)$. Furthermore, this is achieved while relaxing the independent and identically distributed (i.i.d) assumption prevalent in this literature to allow for strictly stationary and ergodic data $\{Y_t, X_t\}_{t=1}^T$.

Specifically, this paper proposes the usage of the (uniform) $k$-Nearest Neighbour ($k$-NN) density estimator of $f(\cdot)$ in (1.2) instead, i.e.,

$$\hat{f}_T(X_t) = \frac{k}{2TR_T(X_t, k)},$$

(1.3)

where

$$R_T(X_t, k) = \text{the Euclidean distance between } X_t \text{ and the } k\text{-th nearest neighbour of } X_t \text{ among all the } X_s\text{'s for } s \neq t$$

for $s = 1, \ldots, T$. Estimator (1.3) corresponds to Mack and Rosenblatt’s (1979) version with
uniform weights, as originally proposed by Loftsgaarden and Quesenberry (1965). In particular, using a score (a.k.a. weighting) function the estimator (1.2) can then be reformulated as

$$\hat{\theta} = \frac{2}{k} \sum_{t=1}^{T} J(t/T) Y_{[t]} R_T \left(X_{(t)}, k\right),$$

(1.4)

where $X_{(t)}$ is the $t$-th order statistic; $Y_{[t]}$ is the concomitant associated with $X_{(t)}$; and $J(\cdot)$ is a bounded, positive and smooth score function satisfying $\int_{0}^{1} J(s) ds = 1$, $\int_{0}^{1} J^2(s) \leq 1$, and $J'(s) < \infty$ for every $s \in [0, 1]$. Thus, this score function is essentially the trimming function $I[a \leq s \leq b]$ for some $0 \leq a, b \leq 1$.\(^2\) The purpose of the introduction of this weighting function in (1.4) is twofold: Methodologically, the proof of this paper first generalizes results in Yang (1981) to dependent data and applies them to obtain the main result of the paper. Since Yang (1981) utilizes a similar but more general weighting scheme, the introduction of this weighting function at this point therefore guarantees that results in Section 2.2 are effectively a generalization of those in Yang (1981) applicable to a wider set of problems beyond the scope of this paper. Empirically, since this weighting function can play the role of a trimming function, i.e. $a > 0$ and $b < 1$, its introduction can help to attenuate the impact of outliers in the regressor process $X_t$, possibly causing the $k$-NN density $\hat{f}_T(X_t)$ to take near-zero values, or to have undersmooth tails.

The usage of nonparametric $k$-NN estimator of $f(\cdot)$, in place of a kernel estimator for example is particularly helpful in (1.2), because (1.4) is theoretically easier to handle than (1.2), since it does not involve the ratio of two random quantities. Another important advantage of the $k$-NN approach is its local adaptation, a property that is not enjoyed by any kernel method for example. For these reasons the $k$-NN approach has already been used in other semiparametric settings assuming i.i.d. sampling, see e.g. Robinson (1987, 1995), Newey (1990), Delgado (1992), Delgado and Stengos (1994), Liu and Lu (1997), Heckman, Ichimura, and Todd (1998), Abadie and Imbens (2006; in press), Pinkse (2006), Li (2006), Jacho-Chávez (2008), and Jun and Pinkse (2009a,b,c). Although results are available for the nonparametric $k$-NN estimation of densities and regression functions with time series, see e.g. Yakowitz (1987), Boente and Fraiman (1988, 1990), Tran and Yakowitz (1993), and Lu and Cheng (1998) among many others, we have not seen any previous use of the $k$-NN method in semiparametric inference with dependent data as in this paper.

An estimator of (1.1) closely related to ours is Lewbel and Schennach’s (2007) ‘ordered data’

---

\(^1\)Although other weighting schemes can be chosen as suggested in Mack and Rosenblatt, the usage of (1.3) in (1.2) lead to various simplifications in the statistical theory.

\(^2\)Notice that setting $a = 0$ and $b = 1$ will amount to using the entire sample in (1.4), i.e. no trimming or uniform weights.
estimator based on nearest neighbour *spaces* defined as

\[ \tilde{\theta} = \frac{1}{k} \sum_{1}^{T-k} Y[t] \left( X(t+k) - X(t) \right). \]  

(1.5)

Lewbel and Schennach (2007) showed that if \( \{Y_t, X_t\}_{t=1}^{T} \) is a sequence of i.i.d. random variables and under other regularity conditions, \( \sqrt{T}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, E[\text{var}(Y|X)/f^2(X)]) \) when \( k = o(\ln T) \) as \( k \to \infty \). They derived the semiparametric efficiency bound for regular estimators of \( \theta_0 \) and proved that \( \tilde{\theta} \) achieves it.

Although similar in nature, estimators (1.4) and (1.5) are fundamentally different. In particular, \( k \) in (1.4) refers to the \( k \)-th order statistic from the (conditionally on \( X_t \)) sample \( \{\|X_t - X_s\|\}_{s=1}^{T-1} \) with \( t \neq s \), while \( k \) in (1.5) refers to the \( k \)-th order statistic from the original sample \( \{X_t\}_{t=1}^{T} \). These differences also make Lewbel and Schennach’s (2007) limiting distribution theory not applicable to (1.4) for either fixed or increasing \( k \). Although, (1.4) admittedly requires \( T \)-times more operations than (1.5), the asymptotic properties of (1.4) are established here under generally weaker conditions that do not explicitly require \( X_t \) to have bounded support. Furthermore, the i.i.d. sampling assumption is also relaxed allowing for dependent data. It is also shown that a version of the proposed estimator that utilizes uniform weights, i.e. \( J(s) = I[0 \leq s \leq 1] \), also achieves the semiparametric efficiency bound in the i.i.d. case.

The rest of the paper is organized as follows: Section 2 introduces some notation used throughout. A general central limit theorem for linear functionals of concomitants of order statistics is then provided that could potentially be useful in other settings, for instance, in the nonparametric estimation of a regression function (see e.g. Yang, 1981). The \( \sqrt{T} \)-consistency and asymptotical normality of (1.4) is thereafter established using mild moment conditions and the slowest possible rate of convergence for \( k \). Section 3 provides some Monte Carlo evidence of the proposed estimator under various data generating processes and compares it to the ordered data estimator. Finally, Section 4 concludes by pointing out directions for future research. All proofs are gathered in the appendices.

## 2 Main Results

### 2.1 Basic Notations

We start by introducing some notation: \( F(x) \) is the CDF of \( X \); \( f(x) \) is the PDF of \( X \); \( F(y|x) \) is the CDF of \( Y \) conditional on \( X = x \) (namely, the conditional cumulative probability distribution function of \( Y \)); \( f(y|x) \) is the probability density function of \( Y \) conditional on \( X = x \); \( f(x, y) \)

\(^3\)The study of concomitants as an important class of statistics with many potential applications has been more recently reviewed by David and Nagaraja (1998).
is the joint probability density of $X$ and $Y$; $F_T(x)$ is the empirical CDF of $\{X\}_{t=1}^T$; $F_T(x, y)$ is the empirical joint CDF of $\{X, Y\}_{t=1}^T$; $\|A\|_p$ is the $L_p$-norm of $A$, $\{E[|A|^p]\}^{1/p}$; $\|A\|_{p, \mathcal{I}}$ is the $L_p$-norm of $A$ conditional on $\mathcal{I}$, $\{E[|A|^p|\mathcal{I}]\}^{1/p}$; $\|f\|_{p, \mathcal{I}}$ is the $\ell_p$-norm of $f(x)$, $\{\int_\mathbb{R} |f(x)|^p dx\}^{1/p}$; $\Rightarrow$ means weak convergence (or distributional convergence); $\Omega$ is the sample space; $\omega$ is an element in $\Omega$; $\mathcal{F}_t$ is the Borel algebra generated by $\{(X_0, Y_0), (X_1, Y_1), \ldots, (X_t, Y_t)\}$; $\mathcal{F}^t$ is the Borel algebra generated by $\{(X_t, Y_t), (X_{t+1}, Y_{t+1}), \ldots\}$; $\mathcal{F}_{X,t}$ is the Borel algebra generated by $\{X_0, X_1, \ldots, X_t\}$; $\mathcal{F}_{X,t}$ is the Borel algebra generated by $\{X_t, X_{t+1}, \ldots\}$.

Let $\{Y_t, X_t\}_{t=1}^T$ denote a strictly stationary and ergodic series of bivariate random variables on a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, with a bounded continuous density, $f(x, y)$. Given a sample of data, $\{Y_t, X_t\}_{t=1}^T$, let $Y[t]$ denote the concomitant (or the induced order statistics) of the $t$-th order statistics $X(t)$; and the smoothing parameter $k_T \equiv k$ is a sequence of positive integers that depends on $T$ such that $k/T \to 0$ as $T \to \infty$. Throughout the paper, we shall also assume that the support of $f(\cdot)$, $\text{Supp}f = \{x : f(x) \geq \eta, \text{ for some } \eta > 0\}$, is the entire real line $\mathbb{R}$.4

To formulate our theory in Section 2.3, we shall make use of an assumption of conditional independence, which is often employed in stochastic analysis (see e.g. Clauser and Horne, 1974), between observables. In particular, let $\alpha^*(t) = \alpha(\mathcal{F}_{Y,t}, \mathcal{F}_{Y,0}, \mathcal{F}_{X,t}, \mathcal{F}_{X,0})$ denote the conditional independence coefficient of $Y_t$ such that

$$|P(F \cap G|K \cap H) - P(F|K)P(G|H)| \leq \alpha^*(t),$$

where

$$\alpha^*(t) = \sup_{K \in \mathcal{F}_{X,0}} \sup_{H \in \mathcal{F}'_X} \sup_{F \in \mathcal{F}_{Y,0}, G \in \mathcal{F}'_Y} |P(F \cap G|K \cap H) - P(F|K)P(G|H)|.$$

This coefficient essentially measures the degree of conditional independence between two flows of Borel algebras generated by the process $Y_t$ when sub-Borel algebras generated by the process $X_t$ are given. In particular, suppose that Assumption (A*1) below holds, then $P(F \cap G|K \cap H) - P(F|K)P(G|H) = P(F \cap G|K \cap H) - P(F|K \cap H)P(G|K \cap H)$ – thus, the conditional independence coefficient $\alpha^*(t)$ is a natural congruence of the conditional strong mixing concept, proposed by Rao (2009).

### 2.2 Asymptotic Normality of Linear Functionals of Concomitants of Order Statistics

In this section, we shall consider the following statistics:

$$T(F_T) = \frac{1}{T} \sum_{t=1}^T J^C(t/T) h(X(t), Y[t]), \quad (2.1)$$

---

4This allow for probability density functions such as the standard normal and Student’s $t$ commonly found in practice.
which can also be written as a functional of empirical distributions,

\[ T(F_T) = \int_{\mathbb{R}^2} J^\circ(F_T(x)) h(x, y) dF_T(x, y), \]

where \( h(x, y) \) is some real-valued function of \((x, y)\); and \( J^\circ(\cdot) \) is an arbitrary bounded and smooth score function, i.e., \( J^\circ(s) < \infty \) and \( J^\circ(\cdot) < \infty \) for every \( s \in [0, 1] \). We shall note at this point that this score function is a generalization of the weighting function \( J(s) \) introduced earlier in that it does not need to integrate to one, and it is also used by Yang (1981) whose results are extended to the non-i.i.d. case in this section. Typical examples of this score function include the Gaussian-type score function, \( J^\circ_G(s) = \exp(-s^2) \), or the quartic score function \( J^\circ_Q(s) = (1 - s^2)^2 \mathbb{I}(s^2 \leq 1) \) among others.

The following set of assumptions are needed to establish the main result in this section. To avoid any confusion, we shall note at the outset that henceforth generic constants, \( p, q, q^*, \delta, \) and \( \ell \), are mutually independent from one condition to another.

**Assumption A:**

(A1) **Moment Bounds:**

(a) Given some integer, \( p > 1 \), \( \| h(X_0, Y_0) - m_h(X_0) \|_p < \infty \), where \( m_h(x) = \int_{\mathbb{R}} h(x, y) f(y|x) dy \).

(b) \( \| [h(X_0, Y_0) - m_h(X_0)]^2 \|_{2p/(p-1)} < \infty \).

(c) \( \| q^2(F(X_0)) J^\circ_2(F(X_0)) \|_{2p} < \infty \) with \( q(F) = \{F(1-F)\}^{1/2-\delta} \) for a generic constant, \( \delta > 1/2 \).

(A2) **Conditional Moment:**

(a) \( \| \partial m_h(x)/\partial x \|_{q^*, \ell} < \infty \) for some integer, \( q^* > 1 \).

(b) \( \| m_h(X_0) \|_p < \infty \).

(A3) **Conditional Joint Moments:**

(a) \( \sum_{r=1}^\infty \| E[h(X_r, Y_r)|X_r, \mathcal{F}_0] - E[m_h(X_r)|\mathcal{F}_0] \|_{2, \mathcal{F}_0}^{1/2} \|_{p/(p-1)} < \infty \).

(b) \( \sum_{r=1}^\infty \| E[h(X_0, Y_0)h(X_r, Y_r)|X_0, X_r] - m_h(X_0)m_h(X_r) \|_{p/(p-1)} < \infty \).

In the i.i.d. framework, these assumptions imply conditions usually needed to establish asymptotic results for concomitants of order statistics, see e.g. Watterson (1959), David (1973), David and Galambos (1974), Yang (1977), Yang (1981), Stute (1993) and Khaledi and Kochar (2000) in order settings. To the best of our knowledge, we have not seen similar results for non-i.i.d. data. Thus, the result established in this section can be viewed as an extension of Yang’s (1981) result to the non-i.i.d. case.

The following theorem will be used in Section 2.3:
Theorem 1 Suppose Assumptions (A1), (A2), and (A3) hold. Then

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ J^\circ(t/T)h(X(t), Y(t)) - \int_{\mathbb{R}} J^\circ(F(X))h(X, Y)dF(X, Y) \right\} \stackrel{W}{\rightarrow} N(0, \sigma^2_W),
\]

where \( \sigma^2_W = E[W^2_0] + 2 \sum_{\tau=1}^{\infty} E[W_0 W_\tau] \) and

\[
W_\tau = J^\circ(F(X_\tau))[h(X_\tau, Y_\tau) - m_h(X_\tau)] - \int_{\mathbb{R}} J^\circ(F(x))[\mathbb{I}(X_\tau \leq x) - F(x)]dm_h(x). \tag{2.2}
\]

Remark 1 In the i.i.d. case, Theorem 1 is equivalent to the seminal result obtained by Yang (1981). Assumption (A2) just implies that \( m_h(x) \) is of bounded variation in any finite interval. This is directly assumed in Theorem 1 in Yang (1981) for example. Similarly, Assumption (A1) collapses to the existence of finite second moments of \( Y \) as required in Yang (1981).

Remark 2 Assumption (A3) effectively imposes a minimum amount of cross-serial correlation for \( Y_t \) and \( X_t \) via the summability conditions on the \( L_p \)-norms of conditional moment differences of \( h(X_t, Y_t) \). In particular, when \( \{X_t, Y_t\} \) are i.i.d., then one can immediately verify that the left hand sides of Conditions (A3a) and (A3b) are zero – thus, Assumption (A3) is satisfied. This assumption has two advantages over conventional mixing assumptions, which are often employed in time series analysis. In particular, they are easier to verify than mixing conditions in a given application, and they are also weaker, see, e.g., Lemma 4 in Appendix.

Finally, as pointed out by Yang (1981), Theorem 1 can be used to construct consistent estimators of various conditional quantities such as \( E[Y|X = x] \), \( \Pr(Y \in A|X = x) \) and \( \text{var}(Y|X = x) \) with dependent data. The specific details of such extensions are beyond the scope of this paper, but interested readers can find more information in Yang (1981).

2.3 Asymptotic Normality of the \( k \)-NN Estimator of Inverse-Density Weighted Expectations

We now introduced some assumptions before stating the main result of the paper. As before, we shall note at the outset that henceforth generic constants, \( p, q, q^*, \delta, \) and \( \ell \), are mutually independent from one condition to another.

Assumption A*:

(A*1) Data Generating Process: The regressand process \( Y_t \) depends on the backward Borel algebra \( \mathcal{F}_{X_t} \) or the forward Borel algebra \( \mathcal{F}_X^\circ \) via \( X_t \). That is, \( E[Y_t|\mathcal{F}_{X_t}] = g(X_t) \) and \( E[Y_t|\mathcal{F}_X^\circ] = g(X_t) \).

(A*2) Moment Conditions: Let \( q(F) = \{F(1 - F)\}^{1/2 - \delta} \) for some generic constant, \( \delta > 1/2 \), and
(a) \( E[g^2(X_0)] < \infty \).

(b) \( E \left[ \left( \frac{\partial g(x)}{\partial x} f(x) \right)_{x=X_t}^2 \right] < \infty \).

(c) \( \| Y_0 \|_{q, F_{X,0}} \|_{2^p} < \infty \), where \( p \) and \( q \) are some positive integers such that \( 1/p + 1/q < 1 \), \( p > 1 \) and \( q \geq p \). We shall note at this point that any \( p > 1 \) can be used, depending on what is needed.

(d) \( E[(Y_0 - g(X_0))^4] < \infty \).

(e) \( \| Y_0 - g(X_0) \|_{f(X_0)} < \infty \).

(f) \( \| q^2(F(X_0))J^2(F(X_0)) \|_{2p} < \infty \).

(A*3) Strong Mixing:

(a) \( Y_t \) (conditioning on \( X_t \)) is a strong mixing sequence such that \( \sum_{t=1}^{\infty} \{ \alpha^*(t) \}^{\frac{p-1}{p}} \frac{1}{q} < \infty \) for some integers, \( p \) and \( q \), such that \( 0 < \frac{p-1}{p} - \frac{1}{q} < 1 \).

(b) \( X_t \) is a strong mixing sequence such that \( \sum_{t=1}^{\infty} t^{\ell-1} \alpha^{1-k^2} (F_{X,0}, F_X^k) < \infty \) for some integer \( p > 2 \), where \( \alpha(F_{X,0}, F_X^k) \) is the standard strong mixing coefficient proposed by Rosenblatt (1956a); and \( \ell \) is some integer such that \( \ell > 1 \).

(A*4) Smoothing Parameter: \( k_T = O(T^{\frac{\ell - 1}{2\ell}}) \) such that \( \lim_{T \to \infty} k_T/T = 0 \) for some integer \( \ell > 1 \).

Assumption (A*1) is satisfied, for example, in the nonlinear regression model, \( Y_t = g(X_t) + \epsilon_t \), where the disturbances, \( \epsilon_t \), are serially dependent in such a way that they are uncorrelated with \( X_t \) for all \( t \in [1, T] \). In this sense, one can allow for various degrees of nonlinear dependence between \( Y_t \) and \( X_t \), as well as some forms of heteroskedasticity in \( \epsilon_t \), i.e. \( \epsilon_t = \sigma^2(X_{t-1}) \zeta_t \), where \( \sigma(X_t) \) is a function of \( X_t \) and \( \{ \zeta_t \}_{t=1}^{T} \) are i.i.d. random variables with zero mean and finite variance.

In the same spirit as Robinson (1987) or Delgado (1992), the usage of the \( k \)-NN estimator allows us to relax Lewbel and Schennach’s (2007) assumptions to general moment conditions for certain functionals of \( \{ X_0, Y_0 \} \) as described in Assumption (A*2), whilst allowing for possibly unbounded support of \( f(x) \). However, for some forms of \( f(x) \) the validity of Assumption (A*2) implicitly requires bounded support conditions on \( X_t \). For example, if \( X_t \sim N(0,1) \), where \( X_t \) is independent of \( Y_t - g(X_t) \), then, Assumption (A*2e) is satisfied only if \( E[\exp\{pX_0^2\}] \propto \)

\(^5\)However, they admittedly preclude popular models such as ARMAX, or general regression models with lagged regressors such as \( Y_t = g(X_t, X_{t-1}) + \epsilon_t \). Unfortunately, we have been unable to extend our proofs to allow for this type of dependence.
\[ \int_{\mathbb{R}} \exp\{(p - 1/2)x^2\}dx < \infty. \] The latter is not satisfied unless \( p < 1/2 \) or \( X \) has bounded support.\(^6\)

Moreover, Lewbel and Schennach (2007, Conditions 3-4, p. 192) also require \( f(x) \) and \( g(x) \) to be uniformly Hölder continuous. In contrast, Assumption (A*2b) merely requires the finite second moment of the derivative of the ratio \( g(x)/f(x) \) to be finite, i.e. Lewbel and Schennach (2007, Conditions 3-4, p. 192) imply that the moduli of continuity of \( f(x) \) and \( g(x) \) are finite, making Assumption (A*2b) automatically hold.

Since the dependence structure of \( \{Y_t, X_t\} \) is governed by both cross-sectional and serial dependence, Assumption (A*3a) serves as a summability condition on the maximum degree of mixing of the regressand process \( Y_t \) for every possible mixing pattern in \( X_t \). Assumption (A*3b) basically imposes a rather mild restriction on the mixing coefficients of \( X_t \). This assumption is slightly stronger than other mixing conditions often found in time series analysis, see e.g. Phillips (1987). Nevertheless, owing to this assumption, one can derive a slower rate of divergence for \( k_T \) – which is discussed below. An example of this type of mixing is the nonlinear regression model \( Y_t = g(X_t) + \epsilon_t \), as described above. If the regressor process \( X_t \) satisfies Assumption (A*3b) and the disturbance \( \epsilon_t \) is mixing in such a way that \( \sum_{t=1}^{\infty} \alpha_{\epsilon}^{(p-1)/p-1/q}(t) < \infty \), where \( \alpha_{\epsilon}(t) \) is the conventional mixing coefficient of \( \epsilon_t \), then Assumption (A*3a) is satisfied.

In addition, notice that the conditional strong mixing coefficient, \( \alpha^*(t) \), relates to the standard strong mixing coefficients, \( \alpha(\mathcal{F}_{X,0}, \mathcal{F}_X^t) \) and \( \alpha(\mathcal{F}_0, \mathcal{F}_t) \) as follows: First, recall the relation\(^7\)
\[
\begin{align*}
P(F \cap G|K \cap H) - P(F|K)P(G|H) &= \frac{P(F \cap K \cap G \cap H) - P(F \cap K)P(G \cap H)}{P(K \cap H)} \\
&+ P(F \cap K)P(G \cap H) \left[ \frac{P(K)P(H) - P(K \cap H)}{P(K \cap H)P(K)P(H)} \right].
\end{align*}
\]

It then follows that
\[
\alpha^*(t) \leq \left\{ \inf_{K \in \mathcal{F}_{X,0}, H \in \mathcal{F}_X^t} \frac{1}{P(K \cap H)} \right\}^{-1} \sup_{A \in \mathcal{F}_0, B \in \mathcal{F}^t} | P(A \cap B) - P(A)P(B) | \\
+ \sup_{K \in \mathcal{F}_{X,0}, H \in \mathcal{F}_X^t} \frac{P(F|K)P(G|H)}{P(K \cap H)} \sup_{K \in \mathcal{F}_{X,0}, H \in \mathcal{F}_X^t} | P(K \cap H) - P(K)P(H) | \\
= C_1 \alpha(\mathcal{F}_0, \mathcal{F}^t) + C_2 \alpha(\mathcal{F}_{X,0}, \mathcal{F}_X^t),
\]

where \( C_1 = \left\{ \inf_{K \in \mathcal{F}_{X,0}, H \in \mathcal{F}_X^t} \frac{1}{P(K \cap H)} \right\}^{-1} < \infty \) and \( C_2 = \sup_{K \in \mathcal{F}_{X,0}, H \in \mathcal{F}_X^t} \frac{P(F|K)P(G|H)}{P(K \cap H)} < \infty. \)

\(^6\)We are indebted to an anonymous referee for pointing this out, and providing this example.

\(^7\)We thank and anonymous referee for pointing this out.
Applying the elementary inequality $|x + y|^{1/r} \leq |x|^{1/r} + |y|^{1/r}$ for some integer $r \geq 1$, one can strengthen Assumption (A*3) as

(A*3’) Strong Mixing:

(a) $(Y_t, X_t)$ is a strong mixing sequence of bivariate random variables such that

$$\sum_{t=1}^{\infty} \alpha^{\frac{p-1}{p} - \frac{1}{q}} (F_0, F_t) < \infty$$

for some integers, $p$ and $q$, such that $p > 2$ and $0 < \frac{p-1}{p} - \frac{1}{q} < 1$.

(b) $X_t$ is a strong mixing sequence such that

$$\max \left\{ \sum_{t=1}^{\infty} \alpha^{\frac{p-1}{p} - \frac{1}{q}} (F_{X,0}, F_{X,t}), \sum_{t=1}^{\infty} \ell^{-1} \alpha^{\frac{1}{p} - \frac{1}{q}} (F_{X,0}, F_{X,t}) \right\} < \infty,$$

where $\ell$ is some integer such that $\ell > 1$.

The condition $\lim_{T \to \infty} k_T/T = 0$ in (A*4) is standard in the pointwise asymptotic theory for $k$-NN, see e.g. Bhattacharya and Mack (1987). However, $k_T = O(T^{4/5})$ implies that $k_T$ diverges at a rate,\(^8\) $T^{\frac{3}{2}}$, which is much slower than many rates in the current literature. For example, Bhattacharya and Mack (1987) showed that the weak convergence of $k$-NN density holds for $k_T = O(T^{4/5})$. The slow rate of divergence obtained in the present paper is due to an application of Cox and Kim’s (1995) inequality in the proof of Theorem 2.

**Theorem 2** Let Assumptions (A*1), (A*2), (A*3) and (A*4) hold. Then,

$$\sqrt{T} (\hat{\theta}_{k_T} - \theta_0) \overset{W}{\to} N(0, \sigma_{W*}^2),$$

where $\sigma_{W*}^2 = E[W_0^2] + 2 \sum_{t=1}^{\infty} E[W_0^* W_t^*]$ with $W_t^* = J(F(X_t)) \frac{Y_t - g(X_t)}{f(X_t)}$.

A few remarks are in order:

- **Remark 3** In the i.i.d. case, the only valid continuous score function in the unit interval $[0, 1]$ is the uniform weight, $J(s) = \mathbb{I}[0 \leq s \leq 1]$. Indeed, $\text{var}(\sqrt{T} (\hat{\theta}_{k_T} - \theta_0)) = [\text{var}(Y | X)/f^2(X)]$ which corresponds to the semiparametric efficiency bound derived in Severini and Tripathi (2001, Section 7, p. 41) and Lewbel and Schennach (2007, footnote 4, p. 193). The choice of any other weighting function would therefore imply a larger variance. This should not preclude practitioners from using a non-uniform (or discontinuous) weighting function in situations where outliers are present.

\(^8\)If $\ell = 2$, then $k = O(T^{3/4})$. 

10
Remark 4 The asymptotic variance \( \sigma^2_{W^*} \), defined in Theorem 2 can be estimated using either \( s^2_{W^*,1} \) or \( s^2_{W^*,2} \) as follows:

\[
s^2_{W^*,1} = \frac{1}{T} \sum_{t=1}^{T} \hat{W}_t^* + 2 \frac{1}{T} \sum_{t=1}^{m} \sum_{\tau=t+1}^{T} \hat{W}_t^* \hat{W}_{t-\tau}^*,
\]

where \( \hat{W}_t^* = J(F_T(X_t))2T(Y_t - \hat{g}_T(X_t))R(X_t, k)/kT, \) \( F_T(\cdot) \) is the empirical CDF, \( \hat{g}_T(\cdot) \) is a consistent estimator of the conditional expectation functional \( g(\cdot) \) if \( g(\cdot) \) is unknown, and \( m \) is a bandwidth that grows slowly with \( T \), see e.g. White (1984, Chap. 6). However, \( s^2_{W^*,1} \) is not constrained to be positive, i.e. when there are large negative sample serial covariances, \( s^2_{W^*,1} \) can take on negative values. Therefore, in view of Newey and West (1987) one can also use

\[
\begin{align*}
\hat{W}_t^* & = J(F_T(X_t))2T(Y_t - \hat{g}_T(X_t))R(X_t, k)/kT, \\
\hat{g}_T(\cdot) & = \frac{1}{T} \sum_{t=1}^{T} \hat{W}_t^* \\
\hat{f}(X_t) & = \frac{1}{T} \sum_{t=1}^{m} \sum_{\tau=t+1}^{T} \hat{W}_t^* \hat{W}_{t-\tau}^*,
\end{align*}
\]

where \( w_{\tau m} = 1 - \tau/(m + 1) \). The consistency of \( s^2_{W^*,1} \) and \( s^2_{W^*,2} \) can then be established by non-trivial extensions of asymptotic arguments in Newey and West (1987), Hansen (1992) and Davidson and de Jong (2000).\(^9\) This is left for future research.

Remark 5 For the i.i.d. data case, estimators \( s^2_{W^*,1} \) and \( s^2_{W^*,2} \) then become

\[
s^2_{W^*,3} = \frac{1}{T} \sum_{t=1}^{T} \left[ J(F_T(X_t)) \frac{Y_t - \hat{g}_T(X_t)}{f(X_t)} \right]^2.
\]

The strong consistency of this estimator in the i.i.d. case is established in the following Theorem:

**Theorem 3** Suppose that \( \{Y_t, X_t\}_{t=1}^{T} \) is a sequence of i.i.d. data and that \( \hat{g}_T(\cdot) \) is an uniformly strong consistent estimator for \( g(\cdot) \), i.e. \( \sup_{x \in \mathbb{R}} |\hat{g}_T(x) - g(x)| = o_{a.s.}(1) \). Let Assumptions (A*2d) and (A*2e) hold and let \( k_T \) increase with \( T \) in such a way that \( \lim_{T \to \infty} \frac{k_T}{\log T} = \infty \) and \( \lim_{T \to \infty} \frac{k_T}{T} = 0 \). Then, it follows that

\[
s^2_{W^*,3} \Rightarrow \sigma^2_{W^*} \equiv E[W^*_0^2] \ a.s.
\]

where \( W^*_t = J(F(X_t)) \frac{Y_t - g(X_t)}{f(X_t)} \).

\(^9\)Notice that Davidson and de Jong (2000) assume that a stochastic process, \( W_t \), is a random function with parameters, \( \beta \). In our case, \( W_t \) is a pseudo-observation with an unknown function.
3 Monte Carlo Experiments

In this Section, we provide Monte Carlo evidence of the small sample performance of the proposed estimator in comparison with Lewbel and Schennach’s (2007) estimator. Although the consistency of the latter was only proven under the i.i.d. assumption, our numerical results below show it also seems to work for dependent data.

We generate sequences of pseudo-random numbers \( \{Y_t, X_t\}_{t=1}^T \) where \( Y_t = 2X_t(1 + e_t)I(0 < X_t < 1) \), and \( \{X_t\}_{t=1}^T \) follows the following Data Generating Processes (DGP) where \( \{e_t\}_{t=1}^T \) and \( \{Z_t\}_{t=1}^T \) denote independent sequences of i.i.d. \( N(0,1) \) random variables:

DGP 1. Independent and Identically Distributed Data (IID), \( X_t = Z_t \).

DGP 2. Threshold Autoregressive Model (TAR),
\[
X_t = \begin{cases} 
0.6X_{t-1} + Z_t & ; \ X_{t-1} < 1, \\
-0.5X_{t-1} + Z_t & ; \ X_{t-1} \geq 1.
\end{cases}
\]

DGP 3. Sign autoregressive model (SIGN), \( X_t = \text{sign}(X_{t-1}) + 0.43Z_t \), where \( \text{sign}(x) = I(x > 0) - I(x < 0) \).

DGP 4. Tem Map model (TEM MAP),
\[
X_t = \begin{cases} 
X_{t-1}/\alpha & ; \ 0 < X_{t-1} < \alpha, \\
(1 - X_{t-1})/(1 - \alpha) & ; \ \alpha \leq X_{t-1} \leq 1,
\end{cases}
\]

where \( \alpha = 0.49999 \) and \( Y_0 \) is generated from \( U[0,1] \).

DGP 5. Markov Regime-Switching model (MARKOV RS),
\[
X_t = \begin{cases} 
0.6X_{t-1} + Z_t & ; \ S_t = 0, \\
-0.5X_{t-1} + Z_t & ; \ S_t = 1,
\end{cases}
\]

where \( S_t \) is a latent state variable that follows a 2-state Markov chain with transition probabilities \( P(S_t = 1|S_{t-1} = 0) = P(S_t = 0|S_{t-1} = 1) = 0.3 \).

DGP 6. Exponential AR(1) model (EXP-AR(1)), \( X_t = 0.5X_{t-1} + 10X_{t-1} \exp(-X_{t-1}^2) + Z_t \).

DGP 1 corresponds to Lewbel and Schennach’s (2007) original design, while DGP 2–6 are taken from Escanciano and Jacho-Chávez (2010). In each of 1000 replications we generate sets of \( T = 100, 200 \) and 400 pseudo-random numbers following these DGPs. Estimators (1.4), using uniform weights, i.e. \( J(s) = 1 \) for all \( s \in [0,1] \), and (1.5) are calculated for different values of \( k = 1, 2, \ldots, 30 \). Figures 1 and 2 display the finite sample behaviour of the simulated
\(\hat{\theta}(s) - \theta_0\) and \(\tilde{\theta}(s) - \theta_0\) as functions of \(k\) respectively in the form of box plots for each design. These figures contrast the estimators’ different bias and variance behaviour as \(k\) changes. Overall, the results are qualitative similar between both sets of estimators for most DGPs at large \(k\) and \(T\). However, the performance of the proposed estimator seems to be quite robust to the choice of smoothing parameter \(k\), i.e. bias and the box plot’s ‘whiskers’ are somewhat invariant to the choice of \(k\). This could be a consequence of the slow divergence rate implied by Assumption \((A^*4)\). On the other hand, the ordered-data estimator’s simulated bias and variance display the standard tradeoff usually found in nonparametric smoothers, see \(T = 100\) in Figure 1. The simulated bias and variability of both estimators decreases as sample size \(T\) increases.

Finally, Table 1 shows the performance in terms of Monte Carlo bias (Bias), standard deviation (Std. Dev.), root mean squared error (RMSE), mean absolute error (MAE), and the interquartile range (IQR) of each estimator at the value of \(k\) that minimized the simulated RMSE. As displayed in Figures 1 and 2, the results are qualitative similar for both estimators, but the proposed estimator displays less variation for DGP 2 and 3. Furthermore, as predicted by our theory, both estimators have numerically similar variances in DGP 1.

### 4 Summary and Extensions

In this paper, we have established the asymptotic normality of the \(k\)-NN estimators of inverse-density-weighted expectations involving strictly stationary, ergodic data. This is achieved under weaker conditions than previously required by other authors. In particular, no explicit restrictions on the support of the probability density function \(f(X_t)\) in the denominator are needed apart from mild and apparently minimum moment conditions. A slow rate of divergence for the ‘smoothing’ parameter \(k_T\) is also obtained via some conditions on the mixing coefficients of the bivariate process \((Y_t, X_t)\). The asymptotic theory of our \(k\)-NN estimator is then developed from a general asymptotic result of linear functionals of concomitants of order statistics for dependent data, which could be potentially useful in other problems.

Since the Euclidean distance is immediately adaptable to vector-valued random variables, extensions to the multivariate case, i.e. \(X_t \in \mathbb{R}^d\) where \(d \geq 2\), are possible. This would ultimately require conditions on the unconditional/conditional cumulants of \((X_t, Y_t)\) for example. Similarly, as in the current literature of semiparametric inference using the \(k\)-NN method, the question on how one chooses the number of nearest neighbours, \(k\), in a given application remains an open question. Although our estimator is asymptotically semiparametric efficient for i.i.d. data, the bound of semiparametric estimator of inverse-density-weighted expectations is yet to be derived formally in the time series case. Similarly, a self-normalized (a.k.a. studentized) central limit theorem in cases where \(E[W_t^2] = \infty, W_t = Z_t - E[E[Y_t|X_t]/f(X_t)]\), is also possible as in Khan.
and Tamer (in press) but beyond the scope of this paper. Finally, relaxation of Assumption (A*1) to allow for a larger class of dependent processes is also worth exploring, specially for possibly non-stationary processes.

References


Appendix A: Main Proofs

In this section, we use the notation \( k_T = k \), and \( \sum_{t=1}^{T_1} = \sum_{t=1}^{T} \) interchangeably throughout.

**Proof of Theorem 1:** The proof of this Theorem follows 3 steps:

**Step 1:** Firstly, we use the Gâteaux differentiation, see e.g. Koroljuk and Borovskich (1994), and the Taylor formula to decompose the functional \( T(F_T) - T(F) \). Let \( F_\epsilon = F + \epsilon [F_T - F] \) be some \( \epsilon \)-perturbation of a distribution function, \( F \), the Gâteaux derivative of \( T(F) \) in the direction of \( F_T - F \) is given by

\[
T'(F_T - F) = \lim_{\epsilon \downarrow 0} \frac{T(F_\epsilon) - T(F)}{\epsilon}
\]

\[
= \lim_{\epsilon \downarrow 0} \epsilon^{-1} \left[ \int_{\mathbb{R}^2} J^\circ(F_\epsilon(x)) h(x,y) dF(x,y) + \epsilon \int_{\mathbb{R}^2} J^\circ(F_\epsilon(x)) h(x,y) d(F_T(x,y) - F(x,y)) 
- \int_{\mathbb{R}^2} J^\circ(F(x)) h(x,y) dF(x,y) \right]
\]

\[
= \lim_{\epsilon \downarrow 0} \int_{\mathbb{R}^2} J^\circ(F_\epsilon(x)) h(x,y) d(F_T(x,y) - F(x,y)) + \lim_{\epsilon \downarrow 0} \frac{\int_{\mathbb{R}^2} h(x,y) [J^\circ(F_\epsilon(x)) - J^\circ(F(x))] dF(x,y)}{\epsilon}
\]

\[
= \int_{\mathbb{R}^2} J^\circ(F(x)) h(x,y) d(F_T(x,y) - F(x,y)) + \int_{\mathbb{R}^2} h(x,y) J^\circ(F(x)) [F_T(x,y) - F(x,y)] dF(x,y),
\]
where the last term follows from $J^s(F)(F_T - F) = \lim_{\epsilon \to 0} \frac{J^s(F_\epsilon) - J^s(F)}{\epsilon}$. Moreover, since

$$\int_{\mathbb{R}^2} h(x, y)J^s(F(x))[F_T(x) - F(x)]dF(x, y) = \int_{\mathbb{R}^2} J^s(F(x))[F_T(x) - F(x)]h(x, y)f(y|x)dydF(x) = \int_{\mathbb{R}} J^s(F(x))[F_T(x) - F(x)]m_h(x)dF(x) = -\int_{\mathbb{R}} J^s(F(x))[F_T(x) - F(x)]dm_h(x) - \int_{\mathbb{R}} J^s(F(x))m_h(x)d[F_T(x) - F(x)],$$

because $\int_{\mathbb{R}} m_h(x)[F_T(x) - F(x)]dJ^s(F(x))|^{+\infty}_{-\infty} = 0$, we have

$$T'(F_T - F) = \int_{\mathbb{R}^2} J^s(F(x))h(x, y)d(F_T(x, y) - F(x, y)) - \int_{\mathbb{R}} J^s(F(x))[F_T(x) - F(x)]dm_h(x)$$

$$- \int_{\mathbb{R}} J^s(F(x))m_h(x)d(F_T(x) - F(x))$$

$$= \frac{1}{T} \sum_{t=1}^{T} J^s(F(X_t))[h(X_t, Y_t) - m_h(X_t)] - \frac{1}{T} \sum_{t=1}^{T} \int_{\mathbb{R}} J^s(F(x))\mathbb{1}(X_t \leq x) - F(x)]dm_h(x)$$

$$= \frac{1}{T} \sum_{t=1}^{T} \left\{ J^s(F(X_t))[h(X_t, Y_t) - m_h(X_t)] - \int_{\mathbb{R}} J^s(F(x))\mathbb{1}(X_t \leq x) - F(x)]dm_h(x) \right\}.$$

In view of the above equation, we obtain the following decomposition:

$$T(F_T) - T(F) = T'(F_T - F) + R_T$$

$$= \frac{1}{T} \sum_{t=1}^{T} \left\{ J^s(F(X_t))[h(X_t, Y_t) - m_h(X_t)] - \int_{\mathbb{R}} J^s(F(x))\mathbb{1}(X_t \leq x) - F(x)]dm_h(x) \right\} \text{ (A-1)}$$

$$+ R_T, \text{ where } R_T := T(F_T) - T(F) - T'(F_T - F).$$

**Step 2**: Secondly, we prove the weak convergence of (A-1). An application of a martingale approximation, see e.g. Wu and Woodroofe (2004) for the process $\sum_{t=1}^{T} W_t$, where $W_t$ is defined in Theorem 1, yields

$$\sum_{t=1}^{T} W_t = \sum_{t=1}^{T} W_t^* + \tilde{W}_1 - \tilde{W}_{T+1},$$

where $W_t^*$ is a martingale difference sequence defined in Lemma 3; and \( \tilde{W}_t = \sum_{s=1}^{\infty} E[W_s | \mathcal{F}_{t-1}] \). Whence, we can apply the Lindeberg-Feller CLT for martingale difference sequences to $T^{-1/2} \sum_{t=1}^{T} W_t^*$ because under assumptions (A1a), (A2a) and (A3a) the variance of $T^{-1/2} \sum_{t=1}^{T} W_t^*$ is $E[W_0^{'*2}]$ which can easily be derived by a sequential application of Lemmas 1-3. The Lindeberg condition is satisfied as follows: For an arbitrarily small constant, $\delta$, it follows that

$$\frac{1}{T} \sum_{t=1}^{T} E\left[ E[W_t^{'*2}|W_t^{'*} > \sqrt{\delta}\mathcal{F}_{t-1}] \right] = E[W_0^{'*2}|W_0^{'*} > \sqrt{\delta}] \leq \|W_0^{'*2}\|_p \times P(W_0^*' > \sqrt{\delta}) \leq \|W_0^2\|_p \times \sqrt{\delta^2 T}.$$

Assumptions (A1a), (A2a) and Hausdorff-Young’s inequality imply that $\leq \|W_0^2\|_p < \infty$, while Chebyshev’s inequality and Lemma 3 imply that $P(W_0^*' > \sqrt{\delta}) \leq E[W_0^{'*2}]/\delta^2 T \to 0$ as
\[ T \rightarrow \infty. \] Therefore,
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} W_t^* \xrightarrow{w} N(0, \sigma_W^2). \tag{A-2}
\]

To derive the asymptotic behavior of \( T^{-1/2}(\bar{W}_1 - \bar{W}_{T+1}) \), first notice that \( W_t \) is \( \mathcal{F}_t \)-measurable, and therefore given some small number, \( \delta > 0 \), we have
\[
P \left( \left| \frac{W_1 - W_{T+1}}{\sqrt{T}} \right| \geq \delta \right) \leq 2P \left( \frac{|W_1|}{\sqrt{T}} \geq \frac{\epsilon}{2} \right) \rightarrow 0 \text{ as } T \rightarrow \infty. \tag{A-3}
\]

Therefore, (A-2) and (A-3) imply that (A-1) is such that
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} W_t \xrightarrow{w} N(0, \sigma_W^2). \tag{A-4}
\]

**Step 3:** We conclude the proof by showing that \( \mathcal{R}_T = o_p(T^{-1/2}) \), a somehow very lengthy step. We first define \( K(t) = \int_0^t J^o(v)dv \) and rewrite
\[
\mathcal{R}_T = \int_{\mathbb{R}^2} [J^o(F_T(x)) - J^o(F(x))]h(x, y)dF_T(x, y) + \int_{\mathbb{R}} m_h(x)d[K(F_T(x)) - K(F(x))]
- \int_{\mathbb{R}} [J^o(F_T(x)) - J^o(F(x))]m_h(x)dF_T(x) + \int_{\mathbb{R}} [F_T(x) - F(x)]J^o(F(x))dm_h(x).
\]

Since \( \int_{\mathbb{R}} m_h(x)d[K(F_T(x)) - K(F(x))] = -\int_{\mathbb{R}} [K(F_T(x)) - K(F(x))]dm_h(x) \), we obtain the following decomposition:

\[
\sqrt{T}\mathcal{R}_T = \int_{\mathbb{R}^2} \left[ \frac{J^o(F_T(x)) - J^o(F(x))}{F_T(x) - F(x)} - J^o(F(x)) \right] [\sqrt{T}(F_T(x) - F(x))]h(x, y)dF_T(x, y)
- \int_{\mathbb{R}} \left[ \frac{K(F_T(x)) - K(F(x))}{F_T(x) - F(x)} - J^o(F(x)) \right] [\sqrt{T}(F_T(x) - F(x))]dm_h(x)
- \int_{\mathbb{R}} \left[ \frac{J^o(F_T(x)) - J^o(F(x))}{F_T(x) - F(x)} - J^o(F(x)) \right] [\sqrt{T}(F_T(x) - F(x))]m_h(x)dF_T(x)
+ \int_{\mathbb{R}^2} J^o(F(x)) [\sqrt{T}(F_T(x) - F(x))]h(x, y)dF_T(x, y)
- \int_{\mathbb{R}} J^o(F(x)) [\sqrt{T}(F_T(x) - F(x))]m_h(x)dF_T(x) \]

\[ = \mathcal{T}_{T;1} - \mathcal{T}_{T;2} - \mathcal{T}_{T;3} + \mathcal{T}_{T;4}, \] where \( J^o(F) = \partial J^o(G) / \partial G|_{G=F} \).

We only prove that \( \mathcal{T}_{T;1} = o_p(1) \) and \( \mathcal{T}_{T;4} = o_p(1) \), since the proof of \( \mathcal{T}_{T;2} = o_p(1) \) and \( \mathcal{T}_{T;3} = o_p(1) \) follow similar steps.

Firstly, an application of the absolute-value inequalities yields
\[ |T_{T;1}| \leq \sup_{x \in \mathbb{R}} \left| \frac{1}{T} \sum_{t=1}^{T} \left( \frac{J^o(F_T(X_t)) - J^o(F(X_t))}{F_T(X_t) - F(x)} - J^o'(F(X_t)) \right) h(X_t, Y_t) \right| \]

\[ \leq \sup_{x \in \mathbb{R}} \left| \frac{1}{T} \sum_{t=1}^{T} \left( \frac{J^o(F_T(x)) - J^o(F(x))}{F_T(x) - F(x)} - J^o'(F(x)) \right) \times \frac{1}{T} \sum_{t=1}^{T} |h(X_t, Y_t)| \right| \]

\[ = T_{T;1a} \times T_{T;1b} \times T_{T;1c} \]

The strict stationarity assumption implies the existence of an invariant probability measure, \( P(\cdot) \), such that \( \sum_{t=1}^{\infty} \| P(X_t|X_0) - P(X_t) \|_{\mathcal{E}_T} < \infty \). Therefore, by using the same arguments as in Step 2 above without any further condition to be imposed, one can show that \( T_{T;1a} = O_p(1) \). Since \( |F_T(x) - F(x)| = o_{a.s.}(1) \) in view of the Birkhoff-Khintchine ergodic theorem, see e.g. Skorokhod (2005, Theorem 5.2.4, p. 130), an application of the mean-value theorem for random functions, see e.g. White and Domowitz (1984), yields \( T_{T;1b} = o_p(1) \). Assumptions (A1a), (A2b) and Minkowski’s inequality yield \( E|h(X_0,Y_0)| < \infty \). Thus, an application of the Birkhoff-Khintchine ergodic theorem yields \( \lim_{T \to \infty} T_{T;1c} = E[|h(X_0,Y_0)||\mathcal{G}] < \infty \) P-a.s., which implies that \( T_{T;1c} = O_{a.s.}(1) \). Therefore, it follows that \( T_{T;1c} = o_p(1) \), and conclude that \( T_{T;1} = o_p(1) \).

From Tchebyshev’s inequality, in order to prove that \( T_{T;4} = o_p(1) \) we only need to show that \( \lim_{T \to \infty} E[T_{T;4}^2] = 0 \) where

\[ E[T_{T;4}^2] = \frac{1}{T} \left\{ \sum_{t=1}^{T} E \left[ J^{o^2}(F(X_t))[F_T(X_t) - F(X_t)]^2[h(X_t, Y_t) - m_h(X_t)]^2 \right] \right. \]

\[ + \sum_{s \neq t} E \left[ J^{o^2}(F(X_s))[F_T(X_s) - F(X_s)][F_T(X_t) - F(X_t)] \times \right. \]

\[ \left. [h(X_s, Y_s) - m_h(X_s)][h(X_t, Y_t) - m_h(X_t)] \right\} \]

\[ = T^{-1} \{ \overline{T}_{T;4a} + \overline{T}_{T;4b} \}. \]

Term \( \overline{T}_{T;4a} \) in the last inequality is bounded above

\[ \left| \overline{T}_{T;4a} \right| \leq \left\| \sup_{x \in \mathbb{R}} \frac{|F_T(x) - F(x)|}{q^2(F(x))} \right\|_p \left\| q^2(F(X_0)) J^{o^2}(F(X_0)) \right\|_{2p} \left\| |h(X_0,Y_0) - m_h(X_0)|^2 \right\|_{2p}^{\frac{2p}{p-1}} \]

\[ \leq \overline{T}_{T;4a;I} \times \overline{T}_{T;4a;II} \times \overline{T}_{T;4a;III}. \]

Next, using the law of iterated expectations, \( \overline{T}_{T;4b} \) can be rewritten as
An application of Hölder’s inequality yields

\[
\left| \mathcal{T}_{T:4b} \right| = \sum_{s=1}^{T} E \left[ E \left[ J^\prime\prime(F(X_s)) J^\prime(F(X_t)) [F_T(X_s) - F(X_s)][F_T(X_t) - F(X_t)] \times \right. \right. \\
\left. \left. [h(X_s, Y_s) - m_h(X_s)][h(X_t, Y_t) - m_h(X_t)] \mathcal{F}_{X,T} \right] \right] \\
= \sum_{s=1}^{T} E \left[ J^\prime\prime(F(X_s)) J^\prime(F(X_t)) [F_T(X_s) - F(X_s)][F_T(X_t) - F(X_t)] \times \right. \\
\left. \left. E \left[ [h(X_s, Y_s) - m_h(X_s)][h(X_t, Y_t) - m_h(X_t)] \mathcal{F}_{X,T} \right] \right] \\
= \sum_{s=1}^{T} E \left[ J^\prime\prime(F(X_s)) J^\prime(F(X_t)) [F_T(X_s) - F(X_s)][F_T(X_t) - F(X_t)] \times \right. \\
\left. \left. \left\{ E \left[ h(X_s, Y_s)h(X_t, Y_t)|X_s, X_t| - m_h(X_s)m_h(X_t) \right] \right\} \right] \\
\leq \sum_{s=1}^{T} \left\{ \left\| J^\prime\prime(F(X_s)) J^\prime(F(X_t)) [F_T(X_s) - F(X_s)][F_T(X_t) - F(X_t)] \right\|_p \right. \\
\left. \left\| E \left[ h(X_s, Y_s)h(X_t, Y_t)|X_s, X_t| - m_h(X_s)m_h(X_t) \right] \right\|_p^{-1} \right\}. \\
\]

An application of Hölder’s inequality yields

\[
\left| \mathcal{T}_{T:4b} \right| \leq \left\| J^\prime\prime(F(X_s)) J^\prime(F(X_t)) [F_T(X_s) - F(X_s)][F_T(X_t) - F(X_t)] \right\|_p \times \\
\sum_{s=1}^{T} \left\| E \left[ h(X_s, Y_s)h(X_t, Y_t)|X_s, X_t| - m_h(X_s)m_h(X_t) \right] \right\|_p^{-1} \\
\leq 2 \left\| J^\prime\prime(F(X_s)) J^\prime(F(X_t)) q(F(X_s))q(F(X_t)) \right\|_{2p} \left\| \sup_{x \in \mathbb{R}} \left| \frac{F_T(x) - F(x)}{q(F(X_s))} \right| \right\|_{2p} \times \\
\sum_{t=1}^{T-1} \left\| E \left[ h(X_0, Y_0)h(X_\tau, Y_\tau)|X_0, X_\tau| - m_h(X_0)m_h(X_\tau) \right] \right\|_p^{-1} \\
= \mathcal{T}_{T:4b,I} \times \mathcal{T}_{T:4b,II} \times \mathcal{T}_{T:4b,III}. \\
\]
Now, using the following indicator function inequality, see e.g. Tran and Wu (1993, p. 669)

\[ |I(t - s \geq 0) - t| \leq C \left( \frac{1}{s(1 - s)} \right)^{\frac{1}{2}} \left( \frac{1}{t(1 - t)} \right)^{\frac{\delta - 1}{2}}, \]

where \( C \) is a generic constant, we obtain \(|(F_T(x) - F(x))/q(F(x))| \leq CT^{-1} \sum t q^*(F(X_t))\), where \( q^*(F) = \{F(1 - F)\}^{\delta - 1/2} \) for some generic constant, \( \delta > 1/2 \). Since \( q^*(F) \) is a bounded function vanishing at its domain boundaries, it follows that \( \sum t^{-3/2} \sum q^*(F(X_t)) |F_0|_2 < \infty \). Therefore, Maxwell and Woodroofe (2000, p. 128) implies that \( T^{-1} \sum q^*(F(X_t)) = O_{a.s.}(\sqrt{2T^{-1}\log(\log(T))}) \) which is \( o_{a.s.}(1) \) as \( T \to \infty \); and we immediately see that

\[ \sup_{x \in \mathbb{R}} \left| \frac{F_T(x) - F(x)}{q(F(x))} \right| = o_{a.s.}(1), \]

and by the dominated convergence theorem, we conclude that \( \overline{T}_{T,4a,l} = o_{a.s.}(1) \) and \( \overline{T}_{T,4b,l} = o_{a.s.}(1) \). Therefore, along with Assumptions (A1b) and (A1c), an application of Hölder’s inequality (i.e. \( \|X\|_r \leq \|X\|_s \) for any \( r \leq s \)) yields

\[ \frac{1}{T} \overline{T}_{T,4a} = o_{a.s.}(1). \]

(A-5)

Similarly, it then follows from Assumptions (A1c) and (A3b) that

\[ \frac{1}{T} \overline{T}_{T,4b} = o_{a.s.}(1), \]

(A-6)

and we conclude that \( \lim_{T \to \infty} E[\overline{T}_{T,4}] = 0 \) and \( \sqrt{T} R_T = o_p(1) \). The result of the Theorem then follows as needed. \( \blacksquare \)

**Proof of Theorem 2:** The proof of this Theorem follows 4 steps:

**Step 1:** Firstly, notice that

\[
\hat{\theta}_{kT} - \theta_0 = \frac{2}{kT} \sum_{t=1}^{T} J(t/T) Y[t] R(X(t), kT) - \int_{\mathbb{R}} g(x) dx \\
= \frac{1}{T} \sum_{t=1}^{T} J(t/T) Y[t] - g(X(t)) f(X(t)) + \sum_{t=1}^{T} \left\{ J(t/T) \frac{2(Y[t] - g(X(t))) R(X(t), kT)}{kT} - \frac{1}{T} \frac{Y[t] - g(X(t))}{f(X(t))} \right\} \\
+ \sum_{t=1}^{T} J(t/T) \frac{2g(X(t)) R(X(t), kT)}{kT} - \int_{\mathbb{R}} g(x) dx \\
= \mathcal{L}_{T,1} + \mathcal{L}_{T,2} + \mathcal{L}_{T,3},
\]

where \( g(x) \doteq E[Y|x] \). Hence, the asymptotic behavior of \( \hat{\theta}_{kT} - \theta_0 \) is derived from the asymptotic behaviors of terms \( \mathcal{L}_{T,1}, \mathcal{L}_{T,2} \) and \( \mathcal{L}_{T,3} \); these are analyzed in the remaining steps.

**Step 2:** Let \( h(X, Y) = (Y - g(X))/f(X) \), and \( m_h(X_t) = E[h(X_t, Y_t)|X_t] = 0 \). Then Assumption (A2) in Theorem 1 is satisfied. Furthermore, Assumption (A*2) implies (A1), while Assumptions
(A*1) and (A*3) imply Assumption (A3), in view of Lemma 4. Therefore, an application of Theorem 1 gives us
\[
\sqrt{T} \mathcal{L}_{T;1} \xrightarrow{W} N(0, \sigma^2_{W^*}). \tag{A-7}
\]

**Step 3:** We now verify that \( \sqrt{T} \mathcal{L}_{T;2} = o_p(1) \), by first noticing that

\[
\mathcal{L}_{T;2} = \frac{1}{T} \sum_{t=1}^{T} [Y_t - g(X(t))] \left\{ J(t/T) \frac{2TR(X(t), k)}{k} - \frac{1}{f(X(t))} \right\}
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} [Y_t - g(X(t))] \left\{ \frac{2TR(X(t), k)}{k} - \frac{1}{f(X(t))} \right\}
\]

\[
+ \frac{1}{T} \sum_{s=1}^{T} \left\{ [Y_t - g(X(t))] [J(t/T) - 1] \frac{2TR(X(t), k)}{k} \right\}
\]

\[
= \mathcal{L}_{T;2a} + \mathcal{L}_{T;2b}.
\]

Assumption (A*1) implicitly implies that the law of iterated expectation yields \( E[\sqrt{T} \mathcal{L}_{T;2a}] = T^{-1/2} \sum_{t=1}^{T} E \left[ \left\{ 2TR(X(t), k)/k - 1/f(X(t)) \right\} E[Y_t - g(X(t)|X_t)] \right] = 0 \), and we only need to show that \( \lim_{T \to \infty} E[(\sqrt{T} \mathcal{L}_{T;2a})^2] = 0 \). We proceed to write

\[
E[(\sqrt{T} \mathcal{L}_{T;2a})^2] = \frac{1}{T} \sum_{t=1}^{T} E \left[ [Y_t - g(X_t)]^2 \left\{ \frac{2TR(X_t, k)}{k} - \frac{1}{f(X_t)} \right\}^2 \right]
\]

\[
+ \frac{1}{T} \sum_{t=1}^{T} E \left[ [Y_t - g(X_t)][Y_s - g(X_s)] \left\{ \frac{2TR(X_t, k)}{k} - \frac{1}{f(X_t)} \right\} \left\{ \frac{2TR(X_s, k)}{k} - \frac{1}{f(X_s)} \right\} \right]
\]

\[
= \mathcal{L}_{T;2a;I} + \mathcal{L}_{T;2a;II}.
\]

Then by the Cauchy-Schwarz inequality, we have

\[
\mathcal{L}_{T;2a;I} = E \left[ [Y_t - g(X_t)]^2 \left\{ \frac{2TR(X_t, k)}{k} - \frac{1}{f(X_t)} \right\}^2 \right]
\]

\[
\leq E^{1/2} \left[ \sup_{1 \leq t \leq T} \left| \frac{2TR(X_t, k)}{k} - \frac{1}{f(X_t)} \right|^4 \right] E^{1/2} \left[ [Y_t - g(X_t)]^4 \right];
\]

and by Assumption (A*1), the law of the iterated expectation, the Jensen inequality, and the Cauchy-Schwarz inequality, we have
\[ |\mathcal{L}_{T;2a;II}| \]
\[ = \frac{1}{T} \sum_{s=1}^{T} E \left( \left[ \frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)} \right] \left( \frac{2TR(X_s,k)}{k} - \frac{1}{f(X_s)} \right) E \left[ |Y_t - g(X_t)||Y_s - g(X_s)||X_s, X_t| \right] \right) \]
\[ \leq \frac{1}{T} \sum_{s=1}^{T} E \left[ \left( \sup_{1 \leq t \leq T} \left| \frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)} \right| \right)^2 \left( \sup_{1 \leq s \leq T} \left| \frac{2TR(X_s,k)}{k} - \frac{1}{f(X_s)} \right| \right) \times \right.
\[ \left. \left| E \left[ |Y_t - g(X_t)||Y_s - g(X_s)||X_s, X_t| \right] \right| \right] \]
\[ \leq \left\| \sup_{1 \leq t \leq T} \left| \frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)} \right| \right\|^2 2^p \frac{1}{T} \sum_{s=1}^{T-1} \left\| E \left[ Y_\tau Y_0 |X_\tau, X_0 \right] - g(X_\tau)g(X_0) \right\|_{p+1}. \]

Assumptions (A*2c), (A*2d), (A*3), and Lemma 4 imply that \( E^{1/2} \left[ |Y_t - g(X_t)|^4 \right] < \infty \) and \( \sum_{\tau=1}^\infty \left\| E \left[ Y_\tau Y_0 |X_\tau, X_0 \right] - g(X_\tau)g(X_0) \right\|_{p+1} < \infty \). Thus, in order to prove that \( \mathcal{L}_{T;2a;I} = o(1) \) and \( \mathcal{L}_{T;2a;II} = o(1) \), we need to show that
\[ E \left[ \sup_{1 \leq t \leq T} \left| \frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)} \right|^4 \right] = o(1), \quad (A-8) \]
\[ E \left[ \sup_{1 \leq t \leq T} \left| \frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)} \right|^{2p} \right] = o(1). \quad (A-9) \]

In fact, \( A-8 \) and \( A-9 \) immediately follow from the dominated convergence theorem if one can show that
\[ \lim_{T \to \infty} \sup_{1 \leq t \leq T} \left| \frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)} \right| = 0 \text{ P-a.s.} \quad (A-10) \]

Hence, it is necessary to demonstrate that
\[ \lim_{T \to \infty} \left| \frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)} \right| = 0 \quad (A-11) \]

almost everywhere (pointwise) for every \( X_t \) in the sample. First, we shall define a random event, \( A_t = \{ \omega \in \Omega : |2TR(X_t,k)/k - 1/f(X_t)| < \epsilon \} \), where \( \epsilon \) is some arbitrarily small positive generic constant. Thus, in light of the Borel-Cantelli lemma, Eq. (A-11) follows if
\[ \lim_{T \to \infty} \sum_{t=1}^{T} P(A_t) = \infty, \text{ which is analogous to } \lim_{T \to \infty} \sum_{t=1}^{T} P(A_t^c) < \infty. \quad (A-12) \]

In order to prove (A-12) we notice that that \( A_t^c \) can be rewritten as \( A_t^c = A_t^{c,1} \cup A_t^{c,2} \), where \( A_t^{c,1} = \{ \omega \in \Omega : 2TR(X_t,k)/k - 1/f(X_t) \geq \epsilon \} \) and \( A_t^{c,2} = \{ \omega \in \Omega : 2TR(X_t,k)/k - 1/f(X_t) \leq -\epsilon \}. \)
This implies that
\[ P(A_{t}^c) \leq P(A_{t,1}^c) + P(A_{t,2}^c). \]  
(A-13)

Now in order to bound the probabilities $P(A_{t,1}^c)$ and $P(A_{t,2}^c)$ in (A-13), we notice that \( \{2TR(X_t, k)/\kappa - 1/f(X_t) \geq \epsilon \} = \{ R(X_t, k) \geq k(2T)^{-1}(1 + \epsilon f(X_t))/f(X_t) \} \); and for some $s \neq t$ we define $q_{tT} = P\{\|X_s - X_t\| \leq \delta_{1T} \}$, where $\delta_{1T} = k(2T)^{-1}[(\epsilon f(X_t) + 1)/f(X_t)]$. Lemma 5 implies
\[ \lim_{T \to \infty} \frac{q_{tT}}{k} = f(X_t) \iff \lim_{T \to \infty} \frac{T q_{tT}}{k} = \epsilon f(X_t) + 1. \]  
(A-14)

In addition, after defining the Bernoulli random variable, $K_s = \mathbb{I}(\|X_s - X_t\| \geq \delta_{1T})$. It immediately follows that $P(A_{t,1}^c) = P\{R(X_t, k) \geq \delta_{1T} \} \leq P\{\sum_{s=1}^{T} K_s > T - k \} = P\{\sum_{s=1}^{T} (K_s - E[K_s]) > T - k - TE[K_s] \} = P\{\sum_{s=1}^{T} E[K_s] > T q_{tT} - k \}$. Equation (A-14) imply that $k/T q_{tT} \leq \epsilon \sup_{x \in C_{X_t}} f(x) + 1$, where $C_{X_t}$ is an open ball with center, $X_t$, such that the union $\bigcup_{t=1}^{T} C_{X_t}$ covers $\mathbb{R}$ for $T$ sufficiently large. Hence, after defining $a_{ct} = 1 + \epsilon \sup_{x \in C_{X_t}} f(x)$, we then obtain
\[ P(A_{t,1}^c) \leq P\left\{ \sum_{s=1}^{T} (K_s - E[K_s]) > T q_{tT}(1 - k/T q_{tT}) \right\} \leq P\left\{ \sum_{s=1}^{T} (K_s - E[K_s]) > T q_{tT} \frac{a_{ct} - 1}{a_{ct}} \right\} \leq \frac{E\left\{ \sum_{s=1}^{T} (K_s - E[K_s]) \right\}^{2\ell}}{(T q_{tT} a_{ct})^{2\ell}}, \]  
(A-15)

where the last inequality follows from Tchebychev's inequality with $\ell$ representing a positive integer. Now since $K_s$ is the indicator function of $X_s$, one can show that $K_s - E[K_s]$ is also mixing with the same coefficient, $\alpha(s)$, as $X_s$ by using the inequality in Truong and Stone (1992, Lemma 1, p. 82), i.e., \( |E[UV] - E[U]E[V]| \leq 4B_1B_2\alpha(t - s) \), where $U = u(X_t)$, $V = v(X_s)$, $|U| < B_1$, $|V| < B_2$ and $\alpha(\cdot)$ is the usual mixing coefficient. Therefore, Lemma 6 yields
\[ P(A_{t,1}^c) \leq C_\ell \left\{ \frac{T^{\ell}}{k^{2\ell}} \left( \sum_{p=1}^{m} \alpha^{p-1/2}(i) \right) \sum_{t=1}^{T} \frac{M_{t,pl}^2}{(T q_{tT} k)^{2\ell} a_{ct}^{2\ell}} \sum_{s=1}^{T} \frac{T^{s} P^{2\ell-s} \nu_{t}^{s}}{(T q_{tT} k)^{2\ell} a_{ct}^{2\ell}} \right\}^{1/p} \leq 1 \] and $\nu_t = q_{tT}^k(1 - q_{tT}) + (1 - q_{tT})^k q_{tT}$. Hence, we have
\[ \sum_{t=1}^{T} P(A_{t,1}^c) \leq C_\ell \left\{ \frac{T^{\ell+1}}{k^{2\ell}} \left( \sum_{p=1}^{m} \alpha^{p-1/2}(i) \right) \sum_{t=1}^{T} \frac{M_{t,pl}^2}{(T q_{tT} k)^{2\ell} a_{ct}^{2\ell}} \sum_{s=1}^{T} \frac{T^{s} P^{2\ell-s} \nu_{t}^{s}}{(T q_{tT} k)^{2\ell} a_{ct}^{2\ell}} \right\} \] \[ \leq C_\ell \left\{ \frac{T^{\ell+1}}{k^{2\ell}} \left( \sum_{p=1}^{m} \alpha^{p-1/2}(i) \right) \sup_{t} \frac{M_{t,pl}^2}{(T q_{tT} k)^{2\ell} a_{ct}^{2\ell}} \sum_{s=1}^{T} \frac{T^{s} P^{2\ell-s} \nu_{t}^{s}}{(T q_{tT} k)^{2\ell} a_{ct}^{2\ell}} \right\}. \]  
(A-16)

Assumptions (A*3b), (A*4) and the fact that $M_{t,pl}^2/(T q_{tT} k)^{2\ell} a_{ct}^{2\ell} \to 1/(\epsilon f(X_t) + 1)^{2\ell} a_{ct}^{2\ell} < \infty$ and $\nu_t/(T q_{tT} k)^{2\ell} a_{ct}^{2\ell} \to 1/(\epsilon f(X_t) + 1)^{2\ell} a_{ct}^{2\ell} < \infty$ as $T \to \infty$ then imply
Using the exact same arguments, we can also prove that
\[ \sum_{t=1}^{\infty} P(A_{c,t}^t) < \infty. \tag{A-17} \]
Hence, results (A-17) and (A-18) prove (A-12) in view of (A-13).

We now prove that \[ \sum_{t=1}^{\infty} P(A_{c,2,t}^t) < \infty. \tag{A-18} \]

Now we shall bound the terms \( \mathcal{L}_{T;2b} \) and \( \mathcal{L}_{T;2b;I} \). The Cauchy-Schwartz inequality yields
\[ E[(\sqrt{T} \mathcal{L}_{T;2b})^2] = \frac{1}{T} \sum_{t=1}^{T} E \left[ (J(t/T) - 1)^2 [Y_t - g(X_t)]^2 \left( \frac{2TR(X_t, k)}{k} \right)^2 \right] \]
\[ + \frac{1}{T} \sum_{t=1}^{T} \left\{ (J(t/T) - 1)(J(s/T) - 1) E \left[ [Y_t - g(X_t)][Y_s - g(X_s)] \frac{TR(X_t, k) TR(X_s, k)}{k} \right] \right\} \]
\[ = \mathcal{L}_{T;2b,I} + \mathcal{L}_{T;2b;II}. \]

Assumption (A*2d) implies that \{E[|Y_t - g(X_t)|^4]\}^{1/2} < \infty. Equation (A-11) and the dominated convergence theorem also imply that \{E[(2TR(X_t, k)/k)^4]\}^{1/2} \rightarrow E^{1/2} [1/f^4(X_t)]. Similarly, we obtain \( \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^{T} [J(t/T) - 1]^2 \rightarrow \int_0^1 J^2(s) ds - 1 \leq 0 \) because \(|J(s)| \leq 1\) for every \( s \in [0,1] \). Therefore, we conclude that
\[ \mathcal{L}_{T;2b,I} \rightarrow 0 \text{ as } T \rightarrow \infty. \tag{A-19} \]
Now Assumption (A*1) and the law of the iterated expectation yields

\[
|\mathcal{L}_{T;2b,II}| \leq \frac{1}{T} \sum_{t=1}^{T} \left\{ (J(t/T) - 1)(J(s/T) - 1) \right\}
\]

\[
E \left[ \frac{2TR(X_t, k)}{k} \frac{2TR(X_s, k)}{k} \left\{ E \left[ Y_t Y_s | X_t, X_s \right] - g(X_t)g(X_s) \right\} \right] \leq \frac{1}{T} \sup_{s \in [0,1]} \{(J(s) - 1)^2 \} \sum_{t=1}^{T} \left\{ 2TR(X_t, k) \frac{2TR(X_s, k)}{k} \left\{ E \left[ Y_t Y_s | X_t, X_s \right] - g(X_t)g(X_s) \right\} \right\} ;
\]

and an application of Hölder inequality gives

\[
|\mathcal{L}_{T;2b,II}| \leq \text{const.} \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{2TR(X_t, k)}{k} \frac{2TR(X_s, k)}{k} \left\{ E \left[ Y_t Y_s | X_t, X_s \right] - g(X_t)g(X_s) \right\} \right\} \|E \left[ Y_t Y_s | X_t, X_s \right] - g(X_t)g(X_s)\|_p^p \]

\[
\leq \left\| \frac{2TR(X_t, k)}{k} \right\|_{2p} \left\| \frac{2TR(X_s, k)}{k} \right\|_{2p} \frac{1}{T} \sum_{t=1}^{T-1} \left\| E \left[ Y_t Y_0 | X_t, X_0 \right] - g(X_t)g(X_0) \right\|_p^p .
\]

In view of equation (A-11) and the dominated convergence theorem, it follows that

\[
\|2TR(X_t, k)/k\|_{2p} \longrightarrow \|1/f(X_t)\|_{2p} < \infty,
\]

while the last term in the above equation converges to zero in view of (A*2c), (A*3a) and Lemma 4. We then also conclude that

\[
\mathcal{L}_{T;2b,II} \longrightarrow 0, \text{ as } T \longrightarrow \infty.
\]

Hence, Equations (A-19) and (A-20) yield

\[
\sqrt{T} \mathcal{L}_{T;2b} = o_p(1).
\]

(A-21)

**Step 4:** We conclude the proof by proving that \(\sqrt{T} \mathcal{L}_{T;3} = o_{a.s.}(1)\). Firstly, we rewrite

\[
\mathcal{L}_{T;3} = \frac{1}{T} \sum_{t=1}^{T} \left\{ J(t/T) - 1 \right\} \frac{2g(X_t)TR(X_t, k)}{k} + \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{2g(X_t)TR(X_t, k)}{k} - E \left[ g(X_t) \right] \right\} .
\]

Now using the variable transformation for integrals we obtain for finite \(T\), \(E \left[ g(X_t)/f(X_t) \right] \approx \sum_{t=1}^{T-1} \int_{F^{-1}(\tau^{t+1}/T)}^{F^{-1}(\tau^t/T)} g(u)/f(u)T^{-1} = T^{-1} \sum_{t=1}^{T-1} g(\xi(\tau))/f(\xi(\tau)) \),\(^{10}\) where \(F^{-1}(s) \equiv \inf\{x : F(x) \geq s\}\) and \(\xi(\tau)\) is some real number in the interval \([F^{-1}(\tau^{t+1}/T), F^{-1}(\tau^t/T))\), we further obtain\(^{11}\)

\(^{10}\)We obtain an exact equality as \(T\) diverges to infinity.

\(^{11}\)Note that it is straightforward to verify that the boundary term in this equation disappears at the order \(o_p(1)\).
\[ \mathcal{L}_{T;3} \approx \frac{1}{T} \sum_{t=1}^{T} \left[ J(t/T) - 1 \right] 2g(X(t))TR(X(t), k) + \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{2g(X(t))TR(X(t), k)}{k} - \frac{g(\xi(t))}{f(\xi(t))} \right\} \]

\[ = \frac{1}{T} \sum_{t=1}^{T} [J(t/T) - 1] 2g(X(t))TR(X(t), k) + \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{2g(X(t))TR(X_t, k)}{k} - \frac{g(X_t)}{f(X_t)} \right\} \]

\[ + \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{g(X(t))}{f(X(t))} - \frac{g(\xi(t))}{f(\xi(t))} \right\} = \mathcal{L}_{T;3a} + \mathcal{L}_{T;3b} + \mathcal{L}_{T;3c}. \]

In view of Assumption (A*2a) and equation (A-11), we can show that \( \sqrt{T} |\mathcal{L}_{T;3a}| = o_{a.s}(1) \) and \( \sqrt{T} |\mathcal{L}_{T;3b}| = o_{a.s}(1) \). However, term \( \mathcal{L}_{T;3c} \) requires some extra steps as follows: An application of the stochastic mean value theorem yields

\[ \sqrt{T} |\mathcal{L}_{T;3c}| = \frac{1}{\sqrt{T}} \left| \sum_{t=1}^{T} \frac{\partial}{\partial x} \left( \frac{g(x)}{f(x)} \right) \bigg|_{x=\eta(t)} (X(t) - \xi(t)) \right| \leq \left( \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{\partial}{\partial x} \left( \frac{g(x)}{f(x)} \right) \bigg|_{x=\eta(t)} \right\}^2 \right)^{1/2} \left( \sum_{t=1}^{T} (X(t) - \xi(t))^2 \right)^{1/2} = \tilde{\mathcal{L}}_{T;3c;1} \tilde{\mathcal{L}}_{T;3c;11}, \]

where the above inequality follows from the Cauchy-Schwarz inequality; and \( \eta(t) \) is a point on a sample path of \( X_t(\omega) \) such that \( \eta_t \in (\min(X_t(\xi(t)), \max(X_t(\xi(t)))) \) for each \( \omega \in \Omega \) – this point always exists in a large sample.

An application of the Birkhoff-Khintchine theorem together with Assumption (A*2b) gives

\[ \tilde{\mathcal{L}}_{T;3c;1} \rightarrow E^{1/2} \left[ \left\{ \frac{\partial}{\partial x} \left( \frac{g(x)}{f(x)} \right) \bigg|_{x=X_t} \right\}^2 \right] < \infty. \]

Moreover, note that

\[ \sum_{t=1}^{T} \left[ X(t) - F^{-1}(t/T) \right]^2 = \sum_{t=1}^{T} \left[ F_T^{-1}(t/T) - F^{-1}(t/T) \right]^2 \leq \sup_{u \in [0,1]} \left[ F_T^{-1}(u) - F^{-1}(u) \right] \sum_{t=1}^{T} \left| F_T^{-1}(t/T) - F^{-1}(t/T) \right|. \]

Hence, the strong law of large numbers for \( L \)-statistics for stationary and ergodic processes in Aaronson, Burton, Dehling, Gilat, Hill, and Weiss (1996, p. 2847), i.e., \( \lim_{T \to \infty} |F_T^{-1}(u) - F^{-1}(u)| = 0 \) P-a.s., for every \( u \in [0,1] \) implies that

\[ \lim_{T \to \infty} \sup_{u \in [0,1]} |F_T^{-1}(u) - F^{-1}(u)| = 0 \) P-a.s., and \( \sum_{t=1}^{T} \left| F_T^{-1}(t/T) - F^{-1}(t/T) \right| < \infty \) P-a.s.
Then in view of equation (A-22), we have
\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} [X(t) - F^{-1}(t/T)]^2 = 0 \text{ P-a.s.} \tag{A-23}
\]
Considering the fact that \(|F^{-1}(u) - F^{-1}(v)| \to 0\) as \(|u-v| \to 0\) and \(\xi(t) \in (F^{-1}(\frac{t}{T}), F^{-1}(\frac{t+1}{T}))\), putting \(u = \frac{t+1}{T}\) and \(v = \frac{t}{T}\), equation (A-23) yields
\[
\lim_{T \to \infty} \mathcal{L}_{T;3c,II} = 0 \text{ P-a.s.,} \tag{A-24}
\]
and it immediately follows that \(\sqrt{T} |\mathcal{L}_{T;3c}| = o_{a.s.}(1)\). Thus, we obtain
\[
\sqrt{T} |\mathcal{L}_{T;3c}| = o_{a.s.}(1). \tag{A-25}
\]
Therefore, Equations (A-7), (A-21) and (A-25) imply the main result of this theorem. ■

**Proof of Theorem 3:** This proof is performed using the following 3 steps:

**Step 1:** Firstly, an application of the triangle inequality yields
\[
\left| s_{W^*,3}^2 - \sigma_{W^*}^2 \right| = \frac{1}{T} \left| \sum_{t=1}^{T} \left( \frac{J^2(\widehat{F}_T(X_t)) \{Y_t - \widehat{g}_T(X_t)\}^2}{\widehat{f}_T^2(X_t)} - \frac{J^2(F(X_t)) \{Y_t - g(X_t)\}^2}{f^2(X_t)} \right) \right|
\leq \frac{1}{T} \left| \sum_{t=1}^{T} J^2(\widehat{F}_T(X_t)) \{Y_t - \widehat{g}_T(X_t)\}^2 \left( \frac{1}{\widehat{f}_T^2(X_t)} - \frac{1}{f^2(X_t)} \right) \right|
+ \frac{1}{T} \left| \sum_{t=1}^{T} \frac{1}{f^2(X_t)} \left( J^2(\widehat{F}_T(X_t)) \{Y_t - \widehat{g}_T(X_t)\}^2 - J^2(F(X_t)) \{Y_t - g(X_t)\}^2 \right) \right|
= \mathcal{M}_{T;1} + \mathcal{M}_{T;2}.
\]

**Step 2:** It is shown that \(\mathcal{M}_{T;1}\) vanishes almost surely. It is observed that
\[
\mathcal{M}_{T;1} \leq \sup_{x \in \mathbb{R}} \left| \frac{1}{\widehat{f}_T^2(x)} - \frac{1}{f^2(x)} \right| \left\{ \frac{1}{T} \sum_{t=1}^{T} J^2(\widehat{F}_T(X_t)) \{Y_t - \widehat{g}_T(X_t)\}^2 \right\}
= \mathcal{M}_{T;1a} \times \mathcal{M}_{T;1b}. \tag{A-26}
\]
To bound \(\mathcal{M}_{T;1a}\), one applies the following inequality: \(||x|^p - |y|^p| \leq p2^{p-1} \{ |y|^{p-1} |x-y| + |x-y|^p \}\) for some integer \(p > 1\), see Csörgő and Horváth (1997, p. 83), to derive
\[
\left| \frac{1}{\widehat{f}_T^2(x)} - \frac{1}{f^2(x)} \right| \leq 4 \left\{ \frac{1}{f^2(x)} \left| \frac{1}{\widehat{f}_T(x)} - \frac{1}{f(x)} \right| + \left| \frac{1}{\widehat{f}_T(x)} - \frac{1}{f(x)} \right|^2 \right\}.
\]
Therefore, reminiscent of the initial assumption that \(f(x)\) is bounded away from zero on its real-line support, to prove that \(\mathcal{M}_{T;1a} = o_{a.s.}(1)\) we need to verify that
\[
\sup_{x \in \mathbb{R}} \left| \frac{1}{\widehat{f}_T(x)} - \frac{1}{f(x)} \right| = o_{a.s.}(1). \tag{A-27}
\]
Using the same strategy as Devroye and Wagner (1977), let €\epsilon$ denote a positive generic constant, and choose another positive generic constant, $\delta$, such that $\left| \frac{1}{f(y)} - \frac{1}{f(x)} \right| < \epsilon/2$ whenever $x$ and $y$ are within $[-\delta/2, \delta/2]$. First, one needs to rewrite the probability of the event $\{ \omega \in \Omega : \sup_{x \in \mathbb{R}} \left| \frac{1}{f_T(x)} - \frac{1}{f(x)} \right| > \epsilon \}$ as

$$P \left( \sup_{x \in \mathbb{R}} \left| \frac{1}{f_T(x)} - \frac{1}{f(x)} \right| > \epsilon \right) = P (A_{T,1}(\omega)) + P (A_{T,2}(\omega)),$$

where $A_{T,1}(\omega) = \bigcup_{x \in \mathbb{R}} \left\{ R_T(x, k) < \frac{k}{2T} \frac{1-\epsilon f(x)}{f(x)} \right\}$ and $A_{T,2}(\omega) = \bigcup_{x \in \mathbb{R}} \left\{ R_T(x, k) > \frac{k}{2T} \frac{1+\epsilon f(x)}{f(x)} \right\}$. Suppose that the event $A_{T,1}$ occurs. Then, for every $x \in \mathbb{R}$, there exists an interval $S$, centered at $x$, with length, $|S|$, less than or equal to $\frac{k}{T} \frac{1-\epsilon f(x)}{f(x)}$. It follows that

$$\mu(S) \leq \frac{k}{T} \frac{1-\epsilon f(x)}{f(x)} = \frac{k}{T} \frac{1-\epsilon f(x)}{f(x)}$$

provided that $|S| \leq \frac{k}{T} \leq \delta$, where $\mu(S)$ is the theoretical distribution of $S$ and $\bar{f} = \inf_{x \in \mathbb{R}} f(x) > 0$. Note that the empirical distribution $\mu_T(S) = \frac{k}{T}$, we obtain

$$\mu_T(S) - \mu(S) \geq \frac{k}{T} \left\{ 1 - \frac{1-\epsilon f(x)}{1-\frac{\epsilon f(x)}{2}} \right\} = \frac{k}{T} \frac{\epsilon f(x)}{1-\frac{\epsilon f(x)}{2}} \geq \frac{k}{T} \frac{\epsilon f(x)}{1-\frac{\epsilon f(x)}{2}}. \tag{A-28}$$

Without any loss of generality, we may take $S_T$ as the class of all intervals with length less than or equal to $4\frac{k}{T} \leq \delta$. By virtue of Equation (A-28), we have

$$P (A_{T,1}(\omega)) \leq P \left( \sup_{S \in S_T} |\mu_T(S) - \mu(S)| \geq \frac{k}{T} \frac{\epsilon f(x)}{2} \right). \tag{A-29}$$

Now suppose that the event $A_{T,2}$ occurs. Then, for every $x \in \mathbb{R}$, there exists an interval $S$, centered at $x$, with length, $|S|$, such that $\frac{k}{T} \frac{1+\epsilon f(x)}{f(x)} < |S| \leq 4\frac{k}{T}$. It follows that

$$\mu(S) \geq \frac{k}{T} \left\{ \frac{4}{f(x)} - \frac{1}{f(x)} - \epsilon \right\} \frac{f(x)}{1+\frac{\epsilon f(x)}{2}}.$$

Hence, we can derive

$$\mu(S) - \mu_T(S) \geq \frac{k}{T} \left\{ \frac{4}{f(x)} - 1 \right\} \frac{f(x)}{1+\frac{\epsilon f(x)}{2}} \geq \frac{k}{T} \left\{ \frac{2}{f(x)} - \frac{\epsilon f(x)}{2} \right\} \geq \frac{k}{T} \left\{ \frac{2}{1+\frac{\epsilon f(x)}{2}} \right\}. \tag{A-30}$$

Equation (A-30) leads to

$$P (A_{T,2}(\omega)) \leq P \left( \sup_{S \in S_T} |\mu_T(S) - \mu(S)| \geq \frac{k}{T} \frac{2}{1+\frac{\epsilon f(x)}{2}} \right). \tag{A-31}$$
Combining Equations (A-29) and (A-31), we obtain

\[ P \left( \sup_{x \in \mathbb{R}} \left| \frac{1}{f_T(x)} - \frac{1}{f(x)} \right| > \varepsilon \right) \leq 2P \left( \sup_{S \in \mathbb{R}} |\mu_T(S) - \mu(S)| \geq \frac{k}{T+\varepsilon} \right). \]

The proof will be completed by an application of the Borel-Cantelli lemma if we show that

\[ \sum_{T=1}^{\infty} P \left( \sup_{S \in \mathbb{R}} |\mu_T(S) - \mu(S)| \geq \frac{k}{T+\varepsilon} \right) < \infty. \] (A-32)

Since

\[ P \left( \sup_{S \in \mathbb{R}} |\mu_T(S) - \mu(S)| \geq \frac{k}{T+\varepsilon} \right) \leq P \left( \sup \left\{ |\mu_T(S) - \mu(S)| : 0 < \mu(S) \leq \frac{4k}{T} \frac{\bar{f}}{(1-\frac{\varepsilon}{2}f)} \right\} \geq \frac{k}{T+\varepsilon} \right), \]

where \( \bar{f} = \sup_{x \in \mathbb{R}} f(x) \), then we shall verify that

\[ \mathcal{M}_{T;1a} = \sum_{T=1}^{\infty} P \left( \sup \left\{ |\mu_T(S) - \mu(S)| : 0 < \mu(S) \leq \frac{4k}{T} \frac{\bar{f}}{(1-\frac{\varepsilon}{2}f)} \right\} \geq \frac{k}{T+\varepsilon} \right) < \infty. \]

To do so, in view of Lemma 7, define \( b = \frac{4k}{T} \frac{\bar{f}}{(1-\frac{\varepsilon}{2}f)} \) and \( \zeta = \frac{k}{T+\varepsilon} \). First note that \( \varepsilon \), \( \bar{f} \), and \( \bar{f} \) are positive generic constants; and \( \lim_{T \to \infty} \frac{\bar{f}}{T} = 0 \). Then, for a given \( T \), one can always find a sufficiently large \( k > \max \left\{ \frac{f(1-\frac{\varepsilon}{2}f)}{4\zeta}, \frac{\bar{f}(2+\varepsilon)^2}{4\zeta} \right\} \) such that \( 0 < b \leq 1/4 \) and \( T > \max \{ 1/b, 8b/\zeta^2 \} \). An application of Lemma 7 yields

\[ \mathcal{M}_{T;1a} \leq \sum_{T=1}^{\infty} 16T^2 \exp \left\{ - \frac{T\zeta^2}{64b+4\zeta} \right\} + \sum_{T=1}^{\infty} 8T \exp \left\{ - \frac{Tb}{10} \right\} \]

\[ = \sum_{T=1}^{\infty} 16T^2 \left\{ \exp\{k\} \right\} - \frac{(\frac{\zeta^2}{2T})^2}{\sum_{T=1}^{\infty} \frac{1}{2^{0.25T}}} + \sum_{T=1}^{\infty} 8T \left\{ \exp\{k\} \right\} - \frac{T}{\sum_{T=1}^{\infty} \frac{1}{2^{0.25T}}} \]

\[ \leq \sum_{T=1}^{\infty} \left\{ 16T^2 \exp\{-\delta_1 k\} + 8T \exp\{-\delta_2 k\} \right\}, \]

where \( \delta_1 \) and \( \delta_2 \) are some generic constant such that \( 0 < \delta_1 < 1 \) and \( 0 < \delta_2 < 1 \). It immediately follows that, if \( \lim_{T \to \infty} \frac{k}{\log T} = \infty \), then \( \mathcal{M}_{T;1a} < \infty \). Therefore, Equation (A-32) follows. This implies Equation (A-27).

Next, to bound \( \mathcal{M}_{T;1b} \), the boundedness of the weight function \( J(\cdot) \) yields

\[ \mathcal{M}_{T;1b} \leq \text{Const.} \times \frac{1}{T} \sum_{t=1}^{T} \{ Y_t - \hat{g}_T(X_t) \}^2. \]
The triangle inequality is then applied to obtain

\[
\mathcal{M}_{T;1b} \leq \text{Const.} \times \left\{ \frac{1}{T} \sum_{t=1}^{T} \left( \{Y_t - \hat{g}_T(X_t)\}^2 - \{Y_t - g(X_t)\} \right) + \frac{1}{T} \sum_{t=1}^{T} \{Y_t - g(X_t)\}^2 \right\}
\]

\[
= \text{Const.} \times \{ \mathcal{M}_{T;1b;i} + \mathcal{M}_{T;1b;ii} \}
\]

In view of the inequality, \(||x||^p - |y|^p| \leq p2^{p-1} \{ |y|^{p-1} |x| + |x-y|^p \}\), we obtain

\[
\mathcal{M}_{T;1b;i} \leq 4 \left\{ \frac{1}{T} \sum_{t=1}^{T} \{Y_t - g(X_t)\} |\hat{g}_T(X_t) - g(X_t)| + |\hat{g}_T(X_t) - g(X_t)|^2 \right\}
\]

\[
\leq 4 \left\{ \sup_{x \in \mathbb{R}} |\hat{g}_T(x) - g(x)| \left\{ \frac{1}{T} \sum_{t=1}^{T} |Y_t - g(X_t)| \right\} + \sup_{x \in \mathbb{R}} |\hat{g}_T(x) - g(x)|^2 \right\}
\]

\[
= o_{a.s.}(1) \quad \text{(A-33)}
\]

by the assumption of the theorem, Assumption (A*2d), and the Khintchine Strong Law of Large Numbers, i.e. \(\frac{1}{T} \sum_{t=1}^{T} |Y_t - g(X_t)| \rightarrow E|Y_0 - g(X_0)| \text{ a.s.}\). Moreover, the Khintchine Strong Law of Large Numbers implies that, under Assumption (A*2d), \(\mathcal{M}_{T;1b;ii} \rightarrow E[(Y_0 - g(X_0))^2] < \infty \text{ a.s.}\). Therefore, in view of Equation (A-26) we can deduce that \(\mathcal{M}_{T;1} = o_{a.s.}(1)\).

**Step 3:** Finally, it is shown that the second term \(\mathcal{M}_{T;2}\) also vanishes asymptotically. Due to the boundedness of \(J(\cdot)\), an application of the triangle inequality yields

\[
\mathcal{M}_{T;2} \leq \sup_{x \in \mathbb{R}} \left| J^2(\hat{F}_T(x)) - J^2(F(x)) \right| \left\{ \frac{1}{T} \sum_{t=1}^{T} \frac{(Y_t - g(X_t))^2}{f^2(X_t)} \right\}
\]

\[
+ \text{Const.} \times \frac{1}{T} \sum_{t=1}^{T} \frac{|Y_t - \hat{g}_T(X_t)|^2 - |Y_t - g(X_t)|^2}{f^2(X_t)}
\]

\[
= \mathcal{M}_{T;2a} + \mathcal{M}_{T;2b}.
\]

The Glivenko-Cantelli theorem implies that \(\sup_{x \in \mathbb{R}} |\hat{F}_T(x) - F(x)| = o_{a.s.}(1)\). Since \(J(\cdot)\) is bounded and smooth, one can conclude that \(\sup_{x \in \mathbb{R}} \left| J^2(\hat{F}_T(x)) - J^2(F(x)) \right| = o_{a.s.}(1)\). In addition the Khintchine Strong Law of Large Numbers implies that, under Assumption (A*2e),

\[
\frac{1}{T} \sum_{t=1}^{T} \frac{(Y_t - g(X_t))^2}{f^2(X_t)} \rightarrow E \left[ \left| \frac{Y_0 - g(X_0)}{f(X_0)} \right|^2 \right] \text{ a.s.}
\]

It then follows that \(\mathcal{M}_{T;2a} = o_{a.s.}(1)\).

Using the inequality \(||x||^p - |y|^p| \leq p2^{p-1} \{ |y|^{p-1} |x| + |x-y|^p \}\), one can bound \(\mathcal{M}_{T;2b}\) as follows:

\[
\mathcal{M}_{T;2b} \leq \frac{1}{f^2} \left\{ \sup_{x \in \mathbb{R}} |\hat{g}_T(x) - g(x)| \frac{1}{T} \sum_{t=1}^{T} |Y_t - g(X_t)| + \sup_{x \in \mathbb{R}} |\hat{g}_T(x) - g(x)|^2 \right\} .
\]

It then follows from the assumption of the theorem and Assumption (A*2d) that, by the Khintchine Strong Law of Large Numbers, \(\mathcal{M}_{T;2b} = o_{a.s.}(1)\). Hence, it has been shown that \(\mathcal{M}_{T;2} = o_{a.s.}(1)\) and the result of the theorem follows. \(\blacksquare\)
Appendix B: Auxiliary Results

The following Lemmas are used to prove Theorem 1, 2 and 3.

**Lemma 1** Given two arbitrary integers, \( p \geq 2 \) and \( q^* \geq 1 \):

(a) Let \( \|\|E[h(X_t, Y_t) - m_h(X_t)]|X_t, \mathcal{F}_0]\|_{2, \mathcal{F}_0}^{1/2} < \infty \), and \( \|\partial m_h(x)/\partial x\|_{q^*, \ell} < \infty \), then \( \kappa_t(p) < \infty \);

(b) Let \( \sum_1^\infty ||E[h(X_t, Y_t) - m_h(X_t)]|X_t, \mathcal{F}_0]\|_{2, \mathcal{F}_0}^{1/2} < \infty \), then \( \sum_1^\infty \kappa_t(p) < \infty \);

where \( \kappa_t(p) = ||E[W_t|\mathcal{F}_0]\|_{p/(p-1)} \) represents the \( L_{p/(p-1)} \)-norm of the conditional expectation of \( W_t \) - defined in (2.2) - given \( \mathcal{F}_0 \) - the Borel algebra generated by \((X_0, Y_0)\).

**Proof.** Firstly, it follows from Minkowski’s inequality that

\[
\kappa_t(p) = \left\| E \left[ J^\circ(F(X_t))[h(X_t, Y_t) - m_h(X_t)] - \int_{\mathbb{R}} J^\circ(F(x)) [\mathbb{I}(X_t \leq x) - F(x)] dm_h(x) | \mathcal{F}_0 \right] \right\|_{p/(p-1)} \leq A_1 + A_2,
\]

where \( A_1 = ||E [J^\circ(F(X_t))[h(X_t, Y_t) - m_h(X_t)]| \mathcal{F}_0] \|_{p/(p-1)} \), and \( A_2 = ||E[\int_{\mathbb{R}} J^\circ(F(x)) [\mathbb{I}(X_t \leq x) - F(x)] dm_h(x)| \mathcal{F}_0] \|_{p/(p-1)} \). By Hölder’s inequality, the first term is bounded above by

\[
A_1 \leq ||E [J^\circ(F(X_t)) E[h(X_t, Y_t) - m_h(X_t)] | X_t, \mathcal{F}_0] | \mathcal{F}_0 \|_{p/(p-1)} \leq \left\| J^\circ(F(X_t)) \right\|_{2, \mathcal{F}_0} \left\| E[h(X_t, Y_t) - m_h(X_t)] | X_t, \mathcal{F}_0 \right\|_{2, \mathcal{F}_0} \leq C \left\| E[h(X_t, Y_t) - m_h(X_t)] | X_t, \mathcal{F}_0 \right\|_{2, \mathcal{F}_0}^{1/2}, \]

because \( J^\circ(\cdot) \leq C \) by assumption, \( \leq \infty \), by the assumptions of the Lemma.

Similarly, by redefining \( v(x) = J^\circ(F(x)) \partial m_h(x)/\partial x \) and \( u(X_t - x) = \mathbb{I}(X_t - x \leq 0) - F(x) \), we can write \( A_2 = ||E[\int_{\mathbb{R}} v(x) u(X_t - x) dx | \mathcal{F}_0] \|_{p/(p-1)} \). Hence, an application of Hausdorff-Young’s inequality yields

\[
A_2 \leq C_{p^*} C_{q^*}^{1/p^*} \left\| P(X_t \leq x | \mathcal{F}_0) - F(x) \right\|_{p^*} \left\| J^\circ(F(x)) \frac{\partial m_h(x)}{\partial x} \right\|_{q^*, \ell},
\]

where \( p^* \) and \( q^* \) are arbitrary integers such that \( p^* > 1, q^* > 1, 1/p^* + 1/q^* = 1 + (p - 1)/p \); and \( C_a = (a)^{1/a}(p')^{1/p'} \) with \( 1/a + 1/p' = 1 \). Notice that the invariant probability distribution, \( P(x) \), of the strictly stationary process \( X_t \) exists, then \( \lim_{t \to \infty} \|P(X_t < x | \mathcal{F}_0) - P(x)\|_{p^*} = 0 \) and \( A_2 \leq \infty \) from the assumptions of the Lemma. This concludes the proof. \( \blacksquare \)
Lemma 2 For two arbitrary integers, \( p \geq 2 \) and \( q^* \geq 1 \), let \( \|h(X_0,Y_0) - m_h(X_0)\|_p < \infty \), and \( \|\partial m_h(x)/\partial x\|_{q^*,t} < \infty \). If

\[
\sum_{i=1}^{\infty} \kappa_i(p) < \infty, \quad \text{(B-1)}
\]

then the following claims are true: (a) \( \lim_{s \to -\infty} E[W_0|F_s] = 0 \) \( P \)-a.s.; (b) \( \lim_{t \to -\infty} t\kappa_t(p) = 0 \); (c) \( \sum_{i=1}^{\infty} E[E[W_0W_i|G]] < \infty \); (d) \( \sum_{i=1}^{\infty} E[W_0W_i] < \infty \); (e) \( \sum_{t=s}^{\infty} E[E[W_t|F_{s+1}] < \infty \); (f) \( \sum_{t=s}^{\infty} E[E[W_t|F_s]] < \infty \), where \( W_t \) is defined in (2.2), and \( G \) is the Borel algebra of invariant sets such that \( G \subset F_0 \subset \cdots \subset F_\infty \).

Proof. We now prove each claim as follow:

Result (a):
Condition (B-1) implies that \( \lim_{t \to -\infty} \kappa_t(p) = 0 \). That is, \( \lim_{t \to -\infty} \|E[W_t|F_\infty]\|_{p/(p-1)} = 0 \). The strict stationarity of \( X_t \) & \( Y_t \) together with the \( F_\infty \)-measurability of the random function \( W_t \) imply that \( E[|E[W_0|F_\infty]|^{\frac{p}{p-1}}] = E[E[W_0|F_\infty]|^{\frac{p}{p-1}}] \). Then \( \lim_{s \to -\infty} E[W_0|F_s] = 0 \) \( P \)-a.s. Thus Equation (a) follows.

Result (b):
Consider the decomposition \( \sum_{i=1}^{\infty} t\kappa_i(p) = \sum_{i=1}^{T_0} t\kappa_i(p) + \sum_{T_0+1}^{T_1} t\kappa_i(p) + \sum_{T_1+1}^{T_2} t\kappa_i(p) + \cdots + \sum_{T_n+1}^{\infty} t\kappa_i(p), \) as \( n \to \infty \), in such a way that \( T_0 > T_1 - T_0 > T_2 - T_1 > \ldots \). It then follows from (B-1) and an application of the Kronecker lemma that \( \lim_{t \to -\infty} T_0^{-1} \sum_{i=1}^{T_0} t\kappa_i(p) = 0, \lim_{T_1 - T_0 \to -\infty} (T_1 - T_0)^{-1} \sum_{i=T_0+1}^{T_1} t\kappa_i(p) = 0, \ldots, \lim_{T_n - T_{n-1} \to -\infty} (T_n - T_{n-1})^{-1} \sum_{i=T_{n-1}+1}^{T_n} t\kappa_i(p) = 0. \) Therefore, \( \sum_{i=1}^{T_0} t\kappa_i(p) = o(T_0), \sum_{i=T_0+1}^{T_1} t\kappa_i(p) = o(T_1 - T_0), \ldots, \sum_{i=T_n+1}^{\infty} t\kappa_i(p) = o(T_n - T_{n-1}), \) and the result follows after noticing that as \( n \) becomes sufficiently large, \( \lim_{n \to \infty} (T_n - T_{n-1}) = 0 \).

Result (c):
It follows from Hörder’s inequality that \( \sum_{i=1}^{\infty} E[E[W_0W_i|G]] = \sum_{i=1}^{\infty} E[W_0W_i|G] \leq \sum_{i=1}^{\infty} E[\|W_0\|_{G,p}||W_i|\|_{G,p/(p-1)}] \leq \sum_{i=1}^{\infty} \|W_0\|_{G,p} \sum_{i=1}^{\infty} \kappa_i(p) \). If follows from the assumptions of the Lemma together with applications of Minkowski’s inequality and Hausdorff-Young’s inequality that \( \|W_0\|_{p} < \infty \). The result then follows from Assumption (B-1).

Result (d):
It follows from the law of iterated expectations and the Jensen inequality that \( \sum_{i=1}^{\infty} E[E[W_0W_i]] = \sum_{i=1}^{\infty} E[E[W_0W_i|G]] \leq \sum_{i=1}^{\infty} E[W_0W_i|G] < \infty \), where the last equality follows from result (c) in the Lemma.

Result (e):
As in the proof of result (a) Hörder’s inequality imply that \( \sum_{i=1}^{\infty} E[E[W_t|F_{s+1}]| F_0| \leq \sum_{i=1}^{\infty} E[E[W_t|F_{s+1}]| F_0|_{p/(p-1)} = \sum_{i=1}^{\infty} \kappa_{t+1}(p) = \sum_{i=1}^{\infty} \kappa_i(p) < \infty \).

Result (f):
Result (e) along with Hörder’s inequality yields \( \sum_{i=1}^{\infty} E[E[W_t|F_s]] = \sum_{i=1}^{\infty} E[E[W_{t-s}|F_0]] = E[W_0] + \sum_{i=1}^{\infty} E[E[W_{t-s}|F_0]] \leq \sum_{i=1}^{\infty} \kappa_i(p) + \|W_0\|_{p} < \infty \), because as in the proof of result (c), \( \|W_0\|_{p} < \infty \). 

36
Lemma 3 For some integer $p > 1$, let $\|W_0\|_p < \infty$ and $\sum_1^\infty \kappa_t(p) < \infty$. Furthermore, let us define the following martingale difference sequence: $W^*_t = \sum_1^\infty (E[W_s|F_t] - E[W_s|F_{t-1}]).$ Then,

$$\sigma^2_W = E[W^*_1^2] = E[W_0^2] + 2 \sum_{s=1}^\infty E[W_0W_s] < \infty. \quad (B-2)$$

Proof. It follows from the law of iterated expectations that it is sufficient to show that

$$E[W^*_1^2|\mathcal{G}] = E[W_0^2|\mathcal{G}] + 2 \sum_{s=1}^\infty E[W_0W_s|\mathcal{G}] < \infty. \quad (B-3)$$

Define $\xi_t = \mathbb{I}(\omega \in A) \{E[W_t|\mathcal{F}_0] - E[W_t|\mathcal{F}_{-1}]\}$, where $\omega$ is the probability element of the Borel subspace $(\Omega, \mathcal{G})$; and $A$ is any subset in $\mathcal{G}$. Hence, we obtain $\mathbb{I}(\omega \in A)W^*_1 = \sum_1^\infty \xi_t$ and $E[\mathbb{I}(\omega \in A)W^*_1^2] = \sum_{t=1,s=1}^\infty E[\xi_s\xi_t]$, where

$$E[\xi_s\xi_t] = E[\mathbb{I}(\omega \in A)\{E[W_s|\mathcal{F}_0] - E[W_s|\mathcal{F}_{-1}]\}\{E[W_s|\mathcal{F}_0] - E[W_s|\mathcal{F}_{-1}]\}]$$

$$= E[\mathbb{I}(\omega \in A)\{E[W_s|\mathcal{F}_0]E[W_s|\mathcal{F}_0] - E[W_s|\mathcal{F}_0]E[W_s|\mathcal{F}_{-1}] - E[W_s|\mathcal{F}_0]E[W_s|\mathcal{F}_{-1}]]] + E[W_s|\mathcal{F}_{-1}]E[W_s|\mathcal{F}_0] + E[W_s|\mathcal{F}_{-1}]E[W_s|\mathcal{F}_0]]$$

$$= E[\mathbb{I}(\omega \in A)\{E[W_s|\mathcal{F}_0] - E\mathbb{I}(\omega \in A)E[W_s|\mathcal{F}_{-1}]\}]$$

$$= E[\mathbb{I}(\omega \in A)\{E[W_s|\mathcal{F}_0] - E\mathbb{I}(\omega \in A)E[W_s|\mathcal{F}_{-1}]\}]$$

$$= E[\mathbb{I}(\omega \in A)W_sE[W_s|\mathcal{F}_0] - E\mathbb{I}(\omega \in A)W_sE[W_s|\mathcal{F}_{-1}]]$$

Since $E[\mathbb{I}(\omega \in A)W_tE[W_s|\mathcal{F}_{-1}]] = E[\mathbb{I}(\omega \in A)W_{t+s-1}E[W_s|\mathcal{F}_{-1+s-t}]] = E[\mathbb{I}(\omega \in A)W_sE[W_t|\mathcal{F}_{-1}]]$, we can write

$$E[\xi_s\xi_t] = E[\mathbb{I}(\omega \in A)W_tE[W_s|\mathcal{F}_0] - E\mathbb{I}(\omega \in A)W_sE[W_t|\mathcal{F}_0]]$$

$$= E[\mathbb{I}(\omega \in A)W_tE[W_s|\mathcal{F}_0] - E\mathbb{I}(\omega \in A)W_sE[W_t|\mathcal{F}_0]]$$

$$+ E[\mathbb{I}(\omega \in A)W_tE[W_s|\mathcal{F}_0] - E\mathbb{I}(\omega \in A)W_sE[W_t|\mathcal{F}_0]]$$

$$= E[\mathbb{I}(\omega \in A)(W_t - W_{t-1})E[W_s|\mathcal{F}_0]] + E[\mathbb{I}(\omega \in A)W_{t+1}(E[W_s|\mathcal{F}_0] - E[W_{s+1}|\mathcal{F}_0])].$$

Hence, we obtain

$$\sum_{s=1, t=1}^T E[\xi_s\xi_t]$$

$$= - \sum_{s=1}^T E[\mathbb{I}(\omega \in A)(W_{T+1} - W_1)E[W_s|\mathcal{F}_0]] - \sum_{t=1}^T E[\mathbb{I}(\omega \in A)W_t+1(E[W_{T+1}|\mathcal{F}_0] - E[W_1|\mathcal{F}_0])]$$

$$= B_{T:1} + B_{T:2},$$

37
where \( \mathcal{B}_{T:1} = \sum_{t=1}^{T} \{ E[\mathbb{I}(\omega \in A) W_t E[W_t|\mathcal{F}_0]] + E[\mathbb{I}(\omega \in A) W_{t+1} E[W_t|\mathcal{F}_0]] \} \), and
\( \mathcal{B}_{T:2} = \sum_{s=1}^{T} \{ E[\mathbb{I}(\omega \in A) W_{T+1} E[W_s|\mathcal{F}_0]] + E[\mathbb{I}(\omega \in A) W_{s+1} E[W_{T+1}|\mathcal{F}_0]] \} \). It follows from the strict stationarity of \( \{Y_t, X_t\} \) and the measurability of \( W_t \) that
\[
\mathcal{B}_{T:1} = \sum_{t=1}^{T} \left\{ E[\mathbb{I}(\omega \in A) W_0 E[W_{t-1}|\mathcal{F}_0]] + E[\mathbb{I}(\omega \in A) W_t W_0] \right\} \\
= E[\mathbb{I}(\omega \in A) W_0^2] + \sum_{s=1}^{T-1} \left\{ E[\mathbb{I}(\omega \in A) W_0 E[W_s|\mathcal{F}_0]] + E[\mathbb{I}(\omega \in A) W_s W_0] \right\} + E[\mathbb{I}(\omega \in A) W_0 W_T] \\
= E[\mathbb{I}(\omega \in A) W_0^2] + 2 \sum_{s=1}^{T-1} E[\mathbb{I}(\omega \in A) W_0 W_s] + E[\mathbb{I}(\omega \in A) W_0 W_T],
\]
where \( E[\mathbb{I}(\omega \in A) W_0 W_T] = E[\mathbb{I}(\omega \in A) W_0 E[W_0|\mathcal{F}_{-T}]] \longrightarrow 0 \) as \( T \longrightarrow \infty \). Also, by Hölder’s inequality
\[
\mathcal{B}_{T:2} = \sum_{s=1}^{T} \left\{ E[\mathbb{I}(\omega \in A) E[W_s|\mathcal{F}_0] E[W_{T+1}|\mathcal{F}_0]] + E[\mathbb{I}(\omega \in A) W_{s+1} E[W_{T+1}|\mathcal{F}_0]] \right\} \\
\leq \sum_{s=1}^{T} \left\{ \|E[W_s|\mathcal{F}_0]\|_p \|E[W_{T+1}|\mathcal{F}_0]\|_{\frac{p}{q}} + \|W_{s+1}\|_p \|E[W_{T+1}|\mathcal{F}_0]\|_{\frac{p}{q}} \right\} \\
= \kappa_{T+1}(p) \sum_{s=1}^{T} \kappa_s(p) + T \kappa_{T+1}(p) \|W_0\|_p.
\]
The results follows after observing that, from Lemma 2, we have \( \lim_{T \to \infty} \mathcal{B}_{T:1} \leq \infty \), and \( \lim_{T \to \infty} \mathcal{B}_{T:2} = 0 \). ■

**Lemma 4** Given integers, \( p > 1 \) and \( q \geq p \), such that \( 1/p + 1/q < 1 \), let \( \|\|Y_0\||_{q,\mathcal{F}_{X,0}}\|_{2p/(p-1)} < \infty \) and \( \sum_{t=1}^{\infty} \{\alpha^*(t)\} \frac{e-1}{p} < \infty \). If Assumption (A*1) holds, then
\[
\sum_{\tau=1}^{\infty} \|E[X_\tau|Y_\tau, X_\tau] - g(X_0)g(X_\tau)\|_{\frac{p}{q}} < \infty, \text{ and} \tag{B-4}
\]
\[
\sum_{\tau=1}^{\infty} \left\|E[Y_\tau|X_\tau, \mathcal{F}_0] - g(X_\tau)\right\|_{2,\mathcal{F}_0} \frac{1}{p} < \infty. \tag{B-5}
\]

**Proof.** We only prove (B-4) as the proof of (B-5) follows similar steps (see e.g. McLeish, 1975). In order to establish (B-4), we follow the same strategy in Hall and Heyde (1980, p. 276) by first demonstrating that if \( |Y_0| \leq C_1 \) and \( |Y_t| \leq C_2 \) for some finite constants, \( C_1 \) and \( C_2 \), then
\[
|E[Y_0 Y_t|X_0, X_t] - g(X_0)g(X_t)| \leq 4C_1C_2\alpha^*(t). \tag{B-6}
\]
In the sequel, let us define truncated random variables, which is obvious in view of the first assumption and the inequality for some $|F|$. Hence, $(\text{B-7})$ follows from $(\text{B-7})$, $(\text{B-8})$ and $(\text{B-9})$.

Next, we shall establish that, if $Y_t \leq C$ for a generic constant, $C$, and $\|Y_t\|_{p,F_{X,t}} < \infty$ – which is obvious in view of the first assumption and the inequality $\|Y\|_{p,T} \leq \|Y\|_{q,T}$ if $p \leq q$ – for some $p > 1$, then

$$\|E[Y_0Y_t|X_0, X_t] - E[Y_0|X_0]E[Y_t|X_t]\|_{p,F_{X,t}} \leq 6C \|Y_0\|_{p,F_{X,0}} \|X_t\|_{q,F_T} \alpha^\ast(t) \frac{2^{t-1}}{p^{t-1}} \alpha^\ast(t).$$

(B-10)

In the sequel, let us define truncated random variables, $Y_{0,N} = Y_0 I(|Y_0| \leq N)$ and $Y_{0,N}' = Y_0 - Y_{0,N}$. We then obtain

$$|E[Y_0Y_t|X_0, X_t] - E[Y_0|X_0]E[Y_t|X_t]| \leq |E[Y_{0,N}Y_t|X_0, X_t] - E[Y_{0,N}|X_0]E[Y_t|X_t]| + |E[Y_{0,N}'Y_t|X_0, X_t] - E[Y_{0,N}'|X_0]E[Y_t|X_t]| \leq D_1 + D_2,$$
where $D_1 \leq 4CN\alpha^*(t)$ from (B-6). Moreover, Hölder’s inequality together with Tchebysev’s inequality yield $D_2 \leq 2CE||Y_{0,N}'||X_0| \leq 2CN^{1-p}E|Y_0|^p|X_0|$ for a generic constant $C$. The by setting $N = \|Y_0\|_{p,F_{X_0}}\{\alpha^*(t)\}^{-1/p}$, we have

$$|E[Y_0Y_t|X_0,X_t] - E[Y_0|X_0]E[Y_t|X_t]| \leq 6C\|Y_0\|_{p,F_{X_0}}\{\alpha^*(t)\}^{\frac{p-1}{p}}.$$  

Then (B-10) follows after taking $L_{p/(p-1)}$ norm on both sides of the above equation.

Finally, we now prove that (B-4) is implied by (B-6) and (B-10) under the assumptions of the Lemma: Firstly, define the truncated random variables, $Y_{t,M} = Y_tI(|Y_t| \leq M)$ and $Y_{t,M}' = Y_t - Y_{t,M}$, where $M = \{\alpha^*(t)\}^{-1/q}E[|Y_0|^q|X_0|]_2^{1/(q-1+p)}$ with $1/p + 1/q < 1$. Then we have

$$\|E[Y_0Y_t|X_0,X_t] - E[Y_0|X_0]E[Y_t|X_t]\|_{p^{-1}} \leq 6M\|Y_0\|_{p,F_{X_0}}\|Y_t,M\|_{p^{-1}} \{\alpha^*(t)\}^{\frac{p-1}{p}} + 2\|Y_0\|_{p,F_{X_0}}\|Y_t,M\|_{p^{-1}} \|F_{X_t}\|_{p^{-1}},$$

where the last inequality follows from (B-10) and Hölder’s inequality. Moreover, an application of Tchebysev’s inequality yields

$$\|Y_0\|_{p,F_{X_0}}\|Y_t,M\|_{p^{-1}} \|F_{X_t}\|_{p^{-1}} \leq 2\|E[|Y_0|^q|X_0|]_2^{\frac{1}{q-1+p}}\}^{\frac{p-1}{p}} \frac{M^{1-q}\|F_{X_t}\|_{p^{-1}}}{q}.$$

Therefore, by applying the inequality $\|Y\|_{p,I} \leq \|Y\|_{q,I}$ for $p \leq q$ we obtain

$$\|E[Y_0Y_t|X_0,X_t] - E[Y_0|X_0]E[Y_t|X_t]\|_{p^{-1}} \leq 6\|E[|Y_0|^q|X_0|]_2^{\frac{1}{q-1+p}} + 2\|E[|Y_0|^q|X_0|]_2\}^{\frac{p-1}{p}} \{\alpha^*(t)\}^{\frac{p-1}{p}} \{\alpha^*(t)\}^{\frac{p-1}{p}} \frac{M^{1-q}\|F_{X_t}\|_{p^{-1}}}{q},$$

and from the assumptions of the Lemma the right-hand side of the above inequality is finite and (B-4) follows immediately.

We shall now state implications of results in Wheeden and Zygmund (1977) and Cox and Kim (1995), as Lemmas 5 and 6 for convenience and completeness.

**Lemma 5** (Wheeden and Zygmund, 1977) Let $S_{x,h}$ denote a closed ball with center, $x$, and diameter, $h$. If $\int |f(x)|dx < \infty$, then we have $\lim_{h \to 0} \int_{S_{x,h}} |f(u)/L(S_{x,h})| du = f(x)$, where $L(\cdot)$ is the Lebesgue measure on $\mathbb{R}$.

**Proof.** See Wheeden and Zygmund (1977, p. 191).
Lemma 6 (Cox and Kim, 1995) Let $\xi_t$ be a strong mixing process. Let $\ell$ be a positive integer and assume that $E[\xi_t] = 0$, and that for some $p > 2$, $M_{p\ell} = \sup_t \{\|\xi_t\|_{p\ell}\} \leq 1$. Suppose further that there is a constant, $\nu$, not depending on $t$ such that $E[\|\xi_t^k\|] \leq \nu$, for some $2 \leq k \leq 2\ell$. Finally, assume that the mixing coefficients satisfy $\sum_{i=1}^{\infty} i^{\ell-1} \alpha^{1-2/p}(i) < \infty$. Then, there exists a generic constant, $C_\ell$, depending on $\ell$ but not on the probability distribution of $\xi_t$, $\nu$, $T$, and $P$ such that

$$E \left[ \left( \sum_{t=1}^{T} \xi_t \right)^{2\ell} \right] \leq C \left\{ T^{\ell} M_{p\ell}^{2\ell} \sum_{i=P}^{\infty} i^{\ell-1} \alpha^{1-2/p}(i) + \sum_{j=1}^{\ell} T^j P^{2\ell-j} \nu^j \right\}$$

for any integers, $T$ and $P \in (0,T)$.


Lemma 7 (Devroye and Wagner, 1980) Let $\mu_T$ and $\mu$ be 1-dimensional empirical distribution and theoretical distribution, respectively, and $Ia$ denote an interval with length $a > 0$. Then, for any $\zeta > 0$, $0 < b \leq 1/4$ and $T \geq \max\{1/b, 8b/\zeta^2\}$,

$$P \left( \sup \{|\mu_T(Ia) - \mu(Ia)| : 0 < \mu(Ia) \leq b\} \geq \zeta \right) \leq 16T^2 \exp \left\{ -T \zeta^2 / (64b + 4\zeta) \right\} + 8T \exp \left\{ -Tb / 10 \right\}.$$ 

Figure 1: Semiparametric Ordered-Data Estimator

Note: Box plots of 1000 simulated differences between the ordered-data estimator and $\theta_0$, where

$$\hat{\theta} = k^{-1} \sum_{t=1}^{T-k} Y_{[t]} (X_{t+k} - X_t)$$

for $k = 1, \ldots, 30$. 
Figure 2: Semiparametric $k$-Nearest Neighbour based Estimator

Note: Box plots of 1000 simulated differences between the $k$-Nearest Neighbour based Estimator and $\theta_0$, where $\hat{\theta} = (2/k) \sum_{1}^{T} Y_{[t]} R_{T} (X_{(t)}, k)$ for $k = 1, \ldots, 30$. 
<table>
<thead>
<tr>
<th>DGP</th>
<th>T</th>
<th>k</th>
<th>Bias</th>
<th>Std. Dev.</th>
<th>RMSE</th>
<th>MAE</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>IID</td>
<td>100</td>
<td>4</td>
<td>0.010</td>
<td>0.255</td>
<td>0.255</td>
<td>0.201</td>
<td>0.349</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.002</td>
<td>0.250</td>
<td>0.250</td>
<td>0.200</td>
<td>0.356</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>6</td>
<td>0.002</td>
<td>0.169</td>
<td>0.169</td>
<td>0.135</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.000</td>
<td>0.170</td>
<td>0.169</td>
<td>0.135</td>
<td>0.222</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>9</td>
<td>0.005</td>
<td>0.109</td>
<td>0.109</td>
<td>0.087</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.003</td>
<td>0.109</td>
<td>0.109</td>
<td>0.087</td>
<td>0.150</td>
<td></td>
</tr>
<tr>
<td>TAR</td>
<td>100</td>
<td>3</td>
<td>0.013</td>
<td>0.305</td>
<td>0.306</td>
<td>0.241</td>
<td>0.394</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>-0.010</td>
<td>0.301</td>
<td>0.301</td>
<td>0.238</td>
<td>0.399</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>3</td>
<td>-0.004</td>
<td>0.210</td>
<td>0.209</td>
<td>0.166</td>
<td>0.280</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>-0.010</td>
<td>0.210</td>
<td>0.210</td>
<td>0.167</td>
<td>0.286</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>5</td>
<td>-0.003</td>
<td>0.141</td>
<td>0.141</td>
<td>0.111</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>-0.004</td>
<td>0.140</td>
<td>0.140</td>
<td>0.110</td>
<td>0.180</td>
<td></td>
</tr>
<tr>
<td>SIGN</td>
<td>100</td>
<td>3</td>
<td>-0.064</td>
<td>0.452</td>
<td>0.490</td>
<td>0.341</td>
<td>0.449</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>-0.081</td>
<td>0.441</td>
<td>0.490</td>
<td>0.336</td>
<td>0.436</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>3</td>
<td>0.000</td>
<td>0.317</td>
<td>0.322</td>
<td>0.202</td>
<td>0.245</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>-0.035</td>
<td>0.313</td>
<td>0.325</td>
<td>0.201</td>
<td>0.225</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>4</td>
<td>0.011</td>
<td>0.173</td>
<td>0.173</td>
<td>0.114</td>
<td>0.175</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.000</td>
<td>0.174</td>
<td>0.175</td>
<td>0.113</td>
<td>0.165</td>
<td></td>
</tr>
<tr>
<td>TEM MAP</td>
<td>100</td>
<td>3</td>
<td>-0.039</td>
<td>0.120</td>
<td>0.125</td>
<td>0.101</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.041</td>
<td>0.142</td>
<td>0.148</td>
<td>0.119</td>
<td>0.196</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>5</td>
<td>-0.024</td>
<td>0.082</td>
<td>0.087</td>
<td>0.068</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.031</td>
<td>0.094</td>
<td>0.098</td>
<td>0.078</td>
<td>0.124</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>5</td>
<td>-0.011</td>
<td>0.059</td>
<td>0.061</td>
<td>0.048</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.015</td>
<td>0.065</td>
<td>0.067</td>
<td>0.053</td>
<td>0.086</td>
<td></td>
</tr>
<tr>
<td>MARKOV RS</td>
<td>100</td>
<td>5</td>
<td>0.011</td>
<td>0.254</td>
<td>0.254</td>
<td>0.199</td>
<td>0.322</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.004</td>
<td>0.253</td>
<td>0.253</td>
<td>0.198</td>
<td>0.327</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>8</td>
<td>0.014</td>
<td>0.169</td>
<td>0.170</td>
<td>0.135</td>
<td>0.220</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.010</td>
<td>0.171</td>
<td>0.172</td>
<td>0.136</td>
<td>0.223</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>9</td>
<td>-0.003</td>
<td>0.116</td>
<td>0.116</td>
<td>0.090</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>-0.002</td>
<td>0.119</td>
<td>0.119</td>
<td>0.093</td>
<td>0.157</td>
<td></td>
</tr>
<tr>
<td>EXP-AR(1)</td>
<td>100</td>
<td>3</td>
<td>-0.022</td>
<td>0.503</td>
<td>0.503</td>
<td>0.388</td>
<td>0.642</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>-0.086</td>
<td>0.469</td>
<td>0.476</td>
<td>0.379</td>
<td>0.610</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>6</td>
<td>0.007</td>
<td>0.308</td>
<td>0.308</td>
<td>0.239</td>
<td>0.379</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>-0.009</td>
<td>0.309</td>
<td>0.309</td>
<td>0.242</td>
<td>0.382</td>
<td></td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>6</td>
<td>0.004</td>
<td>0.196</td>
<td>0.195</td>
<td>0.152</td>
<td>0.252</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>-0.002</td>
<td>0.199</td>
<td>0.199</td>
<td>0.155</td>
<td>0.256</td>
<td></td>
</tr>
</tbody>
</table>

Note: Simulated bias (Bias), standard deviation (Std. Dev.), root mean squared error (RMSE), mean absolute error (MAE), and interquartile range (IQR) of 1000 replications for $k$ minimizing the Monte Carlo RMSE.