*Econometric Theory*, **28**, 2012, 769–803.
doi:10.1017/S026646661100079X

# *k*-NEAREST NEIGHBOR ESTIMATION OF INVERSE-DENSITY-WEIGHTED EXPECTATIONS WITH DEPENDENT DATA

BA CHU
*Carleton University*

DAVID T. JACHO-CHÁVEZ
*Emory University*

This paper considers the problem of estimating expected values of functions that are inversely weighted by an unknown density using the $k$-nearest neighbor ($k$-NN) method. It establishes the $\sqrt{T}$-consistency and the asymptotic normality of an estimator that allows for strictly stationary time-series data. The consistency of the Bartlett estimator of the derived asymptotic variance is also established. The proposed estimator is also shown to be asymptotically semiparametric efficient in the independent random sampling scheme. Monte Carlo experiments show that the proposed estimator performs well in finite sample applications.

## 1. INTRODUCTION

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

$$\theta_0 = \int_{x \in \mathcal{X}} \mathrm{E}(W|X = x)dx = \mathrm{E}\left[\frac{Y}{f(X)}\right], \tag{1.1}$$

for some random variable $W$, where $\mathcal{X} \subset \mathrm{Supp}(X)$ and Supp stands for support. The last equality follows from rewriting $Y = W\mathbb{I}(X \in \mathcal{X})$, where $\mathbb{I}(\cdot)$ is the indicator function that equals one if its argument is true and zero otherwise. Here $f(\cdot)$ represents the unknown marginal density of the continuous scalar random variable $X$, and $\mathrm{E}[\cdot]$ is the expectation taken with respect the joint distribution of $(Y, X)$. Consistent estimation of objects like (1.1) are important in economics

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, suggestions, and corrections. The authors acknowledge funding from the Social Science and Humanities Research Council of Canada (MBF Grant 410-2011-1700). The usual disclaimer applies. Address correspondence to David T. Jacho-Chávez, Department of Economics, Emory University, Rich Building 306, 1602 Fishburne Dr., Atlanta, GA 30322-2240, USA; e-mail: djachocha@emory.edu.

© Cambridge University Press 2012 

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

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 (2011). 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, Sect. 8, p. 2195). In particular, if one has data $\{Y_t, X_t\}_{t=1}^T$, a natural estimator of (1.1) is

$$\widehat{\theta} = \frac{1}{T}\sum_{1}^{T}\frac{Y_t}{\widehat{f}(X_t)}, \tag{1.2}$$

where $\widehat{f}(\cdot)$ 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 $\widehat{f}(\cdot)$, for example; see, e.g., Lewbel (1997, 1998, 2000, 2007), Honoré and Lewbel, among others. However, their results often require smoothness and support conditions on $f(\cdot)$; see, e.g., Khan and Lewbel and Jacho-Chávez (2010). Alternatively, this paper establishes the weak convergence of a version of (1.2) under 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 neighbor ($k$-NN) density estimator of $f(\cdot)$ in (1.2) instead; i.e.,

$$\widehat{f}_T(X_t) = \frac{k}{2TR_T(X_t, k)}, \tag{1.3}$$

where $R_T(X_t, k) \doteq$ the Euclidean distance ($|\cdot|$) between $X_t$ and the $k$th nearest neighbor of $X_t$ among all the $X_s$'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).[1] It then follows that estimator (1.2) can be reformulated as

$$\widehat{\theta} = \frac{2}{k}\sum_{1}^{T} Y_t R_T(X_t, k). \tag{1.4}$$

The usage of the 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*

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

*adaptation*, a property that is not enjoyed by most kernel methods, 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, 2011), Pinkse (2006), Li (2006), Jacho-Chávez (2008), and Jun and Pinkse (2009a, 2009b, 2009c). 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, 1993; 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 problems involving the estimation of finite-dimensional parameters with dependent data as in this paper.

An estimator of (1.1) closely related to ours is Lewbel and Schennach's (2007) "ordered data" estimator based on nearest-neighbor *spacings* defined as

$$\widetilde{\theta} = \frac{1}{k} \sum_{1}^{T-k} Y_{[t]} \left( X_{(t+k)} - X_{(t)} \right), \tag{1.5}$$

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

Although similar in nature, estimators (1.4) and (1.5) are fundamentally different. In particular, for the pair $(Y_{[t]}, X_{(t)})$, where $X_{(t)}$ is in the support of $\{X_t\}_{t=1}^{T}$, its contribution to the sum in (1.5) is $Y_{[t]} \times |X_{(t+k)} - X_{(t)}|$, whereas, in (1.4), its contribution can be written as $Y_{[t]} \times \min\left(|X_{(t)} - X_{(t-k)}|, |X_{(t+k)} - X_{(t)}|\right)$. That is, Lewbel and Schennach's (2007) estimator can be understood as a functional of "one-sided" Euclidean distances between the $X_t$'s and their $k$th "right-hand side" nearest neighbor, while the proposed estimator (1.4) is a functional of "two-sided" Euclidean distances between the $X_t$'s and their $k$th nearest neighbor (either to their left or right).[2] These differences also make Lewbel and Schennach's 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 moment conditions with dependent data. Furthermore, the two-sided nature of the proposed estimator enables it to use more local information than the ordered-data estimator. This in turn may partly explain the numerical "robustness" of (1.4) to the choice of $k$ shown in our Monte Carlo exercise below. Finally, one can also verify that, in the i.i.d. case, (1.4) also achieves the semiparametric efficiency bound.

The rest of the paper is organized as follows: Section 2 introduces some notations used throughout. The $\sqrt{T}$-consistency and the asymptotic normality of (1.4) are then established using a set of mild moment conditions and the slowest

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

possible rate of divergence for the bandwidth $k_T$ under the strong-mixing dependence conditions. The consistency of the Bartlett estimator of the asymptotic variance of (1.4) is also proved under general conditions in Section 2.3. 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 of the proofs are gathered in the Appendixes.

## 2. MAIN RESULTS

### 2.1. Basic Notations

Let $\{Y_t, X_t\}_{t=1}^T$ be a strictly stationary and ergodic sequence of random variables defined on a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, and valued in $\mathbb{R}^2$. The smoothing parameter $k_T \doteq k$ is a sequence of positive integers that depends on $T$ such that $k/T \longrightarrow 0$ as $T \longrightarrow \infty$. $F(x)$ is the cumulative distribution function (CDF) of $X$, and $f(x)$ is the probability density function (PDF) of $X$. Throughout this paper, we shall also assume that the support of $f(\cdot)$, $\text{Supp}(f) = \{x : \ f(x) \geq \eta, \text{ for some } \eta > 0\}$, could be the entire real line $\mathbb{R}$.[3] $F(y|x)$ is the CDF of $Y$ conditional on $X = x$; $f(y|x)$ is the PDF of $Y$ conditional on $X = x$; $f(x, y)$ is the bounded and continuous joint PDF of $X$ and $Y$; $F_T(x)$ is the empirical CDF of $\{X_t\}_{t=1}^T$; $F_T(x, y)$ is the empirical joint CDF of $\{X_t, Y_t\}_{t=1}^T$; $\|A\|_p$ is the $L_p$-norm of $A$, $\{\mathrm{E}[|A|^p]\}^{1/p}$; $\|A\|_{p,\mathcal{I}}$ is the $L_p$-norm of $A$ conditional on $\mathcal{I}$, $\{\mathrm{E}[|A|^p|\mathcal{I}]\}^{1/p}$; $\|f\|_{p,\ell}$ is the $\ell_p$-norm of $f$, $\{\int_{\mathbb{R}} |f(x)|^p dx\}^{1/p}$; $\overset{W}{\Longrightarrow}$ 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 $X_t$; and $\mathcal{F}_X^t$ is the Borel algebra generated by $\{X_t, X_{t+1}, \ldots\}$.

To formulate our theory in Section 2.2, we shall make use of an assumption of conditional independence, which has been employed in stochastic analysis (see, e.g., Clauser and Horne, 1974), between the observables. In particular, let $\alpha^*(t) =: \alpha(\mathcal{F}_{Y,0}, \mathcal{F}_Y^t | \mathcal{F}_{X,0}, \mathcal{F}_X^t)$ denote the *conditional independence* coefficient of $Y_t$ where

$$\alpha^*(t) := \sup_{K \in \mathcal{F}_{X,0}, H \in \mathcal{F}_X^t} \sup_{F \in \mathcal{F}_{Y,0}, G \in \mathcal{F}_Y^t} |P(F \cap G | K \cap H) - P(F|K)P(G|H)|.$$

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 A1 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).

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

### 2.2. Asymptotic Normality of the *k*-NN Estimator of Inverse- Density-Weighted Expectations

We now introduce some assumptions before stating the main result of this section. We shall note at the outset that henceforth generic constants, $p$, $q$, and $\ell$, are mutually independent from one condition to another unless otherwise stated.

**Assumption A1** (Data generating process). The regressand process $Y_t$ depends on $\mathcal{F}_{X,t}$ via $X_t$; i.e., $\mathrm{E}[Y_t|\mathcal{F}_{X,t}] = g(X_t)$.

**Assumption A2** (Moment conditions).

(a) $\|g(X_0)\|_2 < \infty$.

(b) $\left\| \frac{\partial}{\partial x}\left(\frac{g(x)}{f(x)}\right)\Big|_{x=X_t} \right\|_2 < \infty$.

(c) $\left\| \|Y_0\|_{q,\mathcal{F}_{X,0}} \right\|_{2\frac{p-1}{p}} < \infty$, 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) $\|Y_0 - g(X_0)\|_4 < \infty$.

(e) $\left\| \frac{Y_0 - g(X_0)}{f(X_0)} \right\|_p < \infty$.

**Assumption A3** (Strong mixing).

(a) $Y_t$ (conditioning on $X_t$) is a strong-mixing sequence such that $\max\left\{ \sum_{t=1}^{\infty} \{\alpha^*(t)\}^{(p-1)/p-1/q}, \sum_{t=1}^{\infty} \alpha^{1-(2/p)}\left(\mathcal{F}_0, \mathcal{F}^t\right) \right\} < \infty$ for some integers, $p$ and $q$, such that $0 < \frac{p-1}{p} - \frac{1}{q} < 1$, where $\alpha(\cdot,\cdot)$ is the standard strong mixing coefficient in Rosenblatt (1956a).

(b) $X_t$ is a strong mixing sequence such that $\sum_{t=1}^{\infty} t^{\ell-1}\alpha^{1-(2/p)}(\mathcal{F}_{X,0}, \mathcal{F}_X^t) < \infty$ for some integers $p > 2$ and $\ell > 1$.

**Assumption A4** (Smoothing parameter). $k_T = O(T^{(\ell+1)/2\ell})$ such that $\lim_{T\to\infty} k_T/T = 0$ for some integer $\ell > 1$.

Assumption A1 is used in the conditioning arguments in the proofs below. It 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 $X_t$,[4] 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 A2, while allowing for possibly unbounded support of $f(x)$. For example, consider the following data generating process (similar to the one in the

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

Monte Carlo experiments below) where $Y_t = 2X_t(1+e_t)\phi(X_t)$, $X_t$ and $e_t$ are independent of each other standard normal random variables, and $\phi(\cdot)$ represents the probability density function of a standard normal. Notice that the support of $Y_t$ and $X_t$ is the entire real line. Furthermore, $f(X_t) \equiv \phi(X_t)$, $g(X_t) \equiv \mathrm{E}[Y_t|X_t] = 2X_t\phi(X_t)$, and $Y_t - g(X_t) = 2X_t\phi(X_t) \times e_t$. In this case, $\theta = \mathrm{E}[Y_0/\phi(X_0)] = 0$ and the integrability conditions in Assumption A2 are satisfied. On the other hand, for some forms of $f(x)$ the validity of Assumption A2 implicitly requires bounded support conditions on $X_t$. For example, if $X_t$ is again a standard normal random variable but independent of $Y_t - g(X_t)$, then Assumption A2e is only satisfied if $\mathrm{E}\left[\exp\{pX_0^2\}\right] \propto \int_{\mathbb{R}} \exp\{(p-1/2)x^2\}dx < \infty$. The latter is not satisfied unless $p < 1/2$ or $X_t$ has bounded support.[5]

Notice that Lewbel and Schennach (2007, Conds. 3-4, p. 192) require $f(x)$ and $g(x)$ to be uniformly Hölder continuous. In contrast, Assumption A2b merely requires the finite second moment of the derivative of the ratio $g(x)/f(x)$ to be finite. However, Conditions 3-4 in Lewbel and Schennach (p. 192) together with their Condition 2 (bounded support) make Assumption A2b above automatically hold.

Since the dependence structure in (1.4) is governed by both cross-sectional and serial dependence, Assumption A3a to a certain extent serves as a *summability* condition on the maximum degree of mixing of the regressand process $Y_t$ for every possible mixing pattern in $X_t$. This assumption effectively implies the strong mixing condition, $\sum_{t=1}^{\infty} \alpha^{1-(2/p)}(\mathcal{F}_0, \mathcal{F}^t) < \infty$, required for a central limit theorem for stationary strong mixing sequences to be applied in the proof of Theorem 1 below. Assumption A3b in essence imposes rather mild restrictions on the mixing coefficient of $X_t$. This assumption is slightly stronger than other mixing conditions often found in the time-series analysis; see, e.g., Phillips (1987). Nevertheless, owing to this assumption, one could derive a slower rate of divergence for $k_T$, which is discussed below. This type of mixing is general enough to include standard models in nonlinear time series analysis such as those used in the Monte Carlo experiments below. Another example is the nonlinear regression model $Y_t = g(X_t) + \epsilon_t$, as described above. If the regressor process $X_t$ satisfies Assumption A3b 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 A3a 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[6]

$$
\begin{aligned}
&P(F \cap G|K \cap H) - P(F|K)P(G|H) \\
&\quad = \frac{P(F \cap K \cap G \cap H) - P(F \cap K)P(G \cap H)}{P(K \cap H)} \\
&\qquad + P(F \cap K)P(G \cap H)\left[\frac{P(K)P(H) - P(K \cap H)}{P(K \cap H)P(K)P(H)}\right].
\end{aligned}
$$

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

It then follows that

$$\alpha^*(t) \le \left\{ \inf_{\substack{K\in\mathcal{F}_{X,0}\\ H\in\mathcal{F}_X^t\\ K\cap H\neq\{\emptyset\}}} P(K\cap H) \right\}^{-1} \sup_{\substack{A\in\mathcal{F}_0\\ B\in\mathcal{F}^t}} |P(A\cap B)-P(A)P(B)|$$
$$+ \sup_{\substack{K\in\mathcal{F}_{X,0},\ H\in\mathcal{F}_X^t\\ F\in\mathcal{F}_{Y,0},\ G\in\mathcal{F}_Y^t\\ K\cap H\neq\{\emptyset\}}} \frac{P(F|K)P(G|H)}{P(K\cap H)} \sup_{\substack{K\in\mathcal{F}_{X,0}\\ H\in\mathcal{F}_X^t}} |P(K\cap H)-P(K)P(H)|$$
$$= \mathcal{C}_1\alpha\left(\mathcal{F}_0,\mathcal{F}^t\right)+\mathcal{C}_2\alpha\left(\mathcal{F}_{X,0},\mathcal{F}_X^t\right),$$

where $\mathcal{C}_1 = \left\{ \inf_{\substack{K\in\mathcal{F}_{X,0}\\ H\in\mathcal{F}_X^t\\ K\cap H\neq\{\emptyset\}}} P(K\cap H) \right\}^{-1} < \infty$

and $\mathcal{C}_2 = \sup_{\substack{K\in\mathcal{F}_{X,0},\ H\in\mathcal{F}_X^t\\ F\in\mathcal{F}_{Y,0},\ G\in\mathcal{F}_Y^t\\ K\cap H\neq\{\emptyset\}}} \frac{P(F|K)P(G|H)}{P(K\cap H)} < \infty.$

Applying the elementary inequality $|x+y|^{1/r} \le |x|^{1/r}+|y|^{1/r}$ for some integer $r \ge 1$, one can strengthen Assumption A3 to the following.

**Assumption A3′** (Strong mixing).

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

$$\max\left\{ \sum_{t=1}^{\infty} \alpha^{1-(2/p)}\left(\mathcal{F}_0,\mathcal{F}^t\right), \sum_{t=1}^{\infty} \alpha^{((p-1)/p)-(1/q)}\left(\mathcal{F}_0,\mathcal{F}^t\right) \right\} < \infty$$

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

$$\max\left\{ \sum_{t=1}^{\infty} \alpha^{((p-1)/p)-(1/q)}\left(\mathcal{F}_{X,0},\mathcal{F}_X^t\right), \sum_{t=1}^{\infty} t^{\ell-1}\alpha^{1-(2/p)}\left(\mathcal{F}_{X,0},\mathcal{F}_X^t\right) \right\} < \infty,$$

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

The condition $\lim_{T\to\infty} k_T/T = 0$ in Assumption A4 is standard in the pointwise asymptotic theory for $k$-NN; see, e.g., Bhattacharya and Mack (1987). However, $k_T = O(T^{(\ell+1)/2\ell})$ implies that $k_T$ diverges at a rate,[7] $T^{(\ell+1)/2\ell}$, which is much slower than many rates in the current literature. For example, Bhattacharya and Mack showed that the weak convergence of the $k$-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 the following theorem.

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

THEOREM 1. *Let Assumptions A1, A2, A3, and A4 hold. Then*

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

*where* $\sigma^2_{W^*}=\mathrm{E}[W_0^{*2}]+2\sum_{\tau=1}^{\infty}\mathrm{E}[W_0^*W_\tau^*]$ *with* $W_t^*=\frac{Y_t-g(X_t)}{f(X_t)}$.

A few remarks are in order.

**Remark 1.** In the i.i.d. case, we can easily establish that Asy. $\mathrm{Var}(\sqrt{T}(\widehat{\theta}_{k_T}-\theta_0))=[\mathrm{var}(Y|X)/f^2(X)]$, which corresponds to the semiparametric efficiency bound derived in Severini and Tripathi (2001, Sect. 7, p. 41) and Lewbel and Schennach (2007, fn. 4, p. 193).

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

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

where $\widehat{W}_t^*=\frac{(Y_t-\widehat{g}_T(X_t))}{\widehat{f}_T(X_t)}$; $\widehat{g}_T(x)\doteq\widehat{g}(x,R_T(x,h_T))$ is the $h$-NN estimate of $g(x)$ defined in Lemma 6 in Appendix B below; $\widehat{f}_T(x)$ is the standard $k$-NN estimate of the density function $f(x)$; and $m\doteq m_T$ is a lag-truncation parameter that grows slowly with $T$; see, e.g., White (2001, Ch. 6). However, $s^2_{\widehat{W}^*,1}$ is not constrained to be positive; i.e., when there are large negative sample serial covariances, $s^2_{\widehat{W}^*,1}$ can take on negative values. Therefore, in view of Newey and West (1987), one may also use the following Bartlett estimator instead:[8]

$$s^2_{\widehat{W}^*,2}=\frac{1}{T}\sum_{1}^{T}\widehat{W}_t^{*2}+2\sum_{\tau=1}^{m}w_{\tau m_T}\left(\frac{1}{T-\tau}\sum_{t=1}^{T-\tau}\widehat{W}_t^*\widehat{W}_{t+\tau}^*\right), \tag{2.1}$$

where $w_{\tau m}=1-\tau/(m+1)$.

## 2.3. Asymptotic Variance Estimation

The *weak* consistency of the long-run variance estimator, (2.1), is established in this section. Toward that end, we now introduce further assumptions before stating the main result of this section. First, let $g^+(u)=\mathrm{E}[\max(Y,0)|X=u]$, $g^-=\mathrm{E}[\min(Y,0)|X=u]$, $\rho_\tau(u,v)=\mathrm{E}[Y_1Y_{\tau+1}|X_1=u,X_{\tau+1}=v]$, $\overline{\rho}_\tau(u,v)=\mathrm{E}[|Y_1Y_{\tau+1}||X_1=u,X_{\tau+1}=v]$, and $\sigma^2(u)=\mathrm{E}\left[(Y-g(X))^2|X=u\right]$.

**Assumption B1** (Moment conditions).

(a) $\max\left\{\left\|\frac{Y_t-g(X_t)}{f(X_t)}\right\|_8,\|Y\|_s\right\}<\infty$, for some $s>2$.

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

(b) $\max\left\{\left\|\left\|\frac{Y_1-g(X_1)}{f(X_1)}\right\|_{q,\mathcal{F}_{X_1}}\right\|_{10}, \left\|\left\|\frac{Y_1-g(X_1)}{f(X_1)}\right\|_{r,\mathcal{F}_{X_1}}\right\|_{10}, \left\|\left\|\frac{Y_1-g(X_1)}{f(X_1)}\right\|_{p_1,\mathcal{F}_{X_1}}\right\|_{20}, \left\|\left\|\frac{Y_1-g(X_1)}{f(X_1)}\right\|_{p_2,\mathcal{F}_{X_1}}\right\|_{20}, \left\|\left\|\frac{Y_1-g(X_1)}{f(X_1)}\right\|_{p_3,\mathcal{F}_{X_1}}\right\|_{20}, \left\|\left\|\frac{Y_1-g(X_1)}{f(X_1)}\right\|_{p_4,\mathcal{F}_{X_1}}\right\|_{20}\right\} < \infty,$
for some generic constants ($> 1$); $q, r, p_1, p_2, p_3$, and $p_4$ such that $1/q + 1/r < 1$ and $\sum_{i=1}^{4} 1/p_i < 1$.

(c) $\sigma^2(x) = \mathrm{E}\left[(Y_t - g(X_t))^2 | X_t = x\right] < \infty$, for every $x \in \mathrm{Supp}(X)$.

**Assumption B2** (Probability density functions).

(a) The density function $f(x)$ of $X_t$ is continuous, positive, and bounded everywhere.

(b) The joint density function $f_\tau(u, v)$ of $(X_t, X_{t+\tau})$ is uniformly bounded for $\tau \geq 1$.

**Assumption B3** (Regression functions).

(a) $g^+(x)$, $g^-(x)$, $\overline{\rho}_\tau(u, v)$, and $\sigma(x)$ are continuous functions.

(b) $g(x)$ is Lipschitz, and $\lim_{\epsilon \to 0}[g(x + \epsilon u) - g(x)]/\epsilon = r'(x, u)$ such that $\int_{-1}^{1} r'(x, u)du < \infty$ for every $x \in \mathrm{Supp}(X)$.

**Assumption B4** (Smoothing parameters).

(a) $k_T$ verifies $k_T T^{-2/3} = o(1)$.

(b) $h_T$ verifies $h_T/T = o(1)$ and $\lim_{T \to \infty} h_T T^{-2/3} = 2\beta f(x)$ for some $0 \leq \beta < \infty$.

(c) $m_T$ verifies $m_T^{1+2q}/T = O(1)$, for some $q \in (1/2, \infty)$, and $m_T/\min(k_T, h_T) = o(1)$.

**Assumption B5** (Strong mixing).

(a) The strong-mixing coefficient of $X_t$ satisfies $\lim_{n \to \infty} n \sum_{j=n}^{\infty} \alpha(\mathcal{F}_{X,0}, \mathcal{F}_X^j) = 0$.

(b) The strong-mixing coefficient of $(Y_t, X_t)$ satisfies

(i) $\lim_{n \to \infty} n \sum_{\tau=n}^{\infty} \{\alpha(\mathcal{F}_0, \mathcal{F}^\tau)\}^{1-1/w} = 0$ with $w = s/(s-2)$, where $s$ satisfies Assumption B1a;

(ii) $\sum_{\tau=1}^{\infty} \{\alpha(\mathcal{F}_0, \mathcal{F}^\tau)\}^{2(1/r-1/8)} < \infty$ for some $r \in (2, 4]$ such that $r > 2 + 1/q$, where $q$ satisfies Assumption B4c.

(c) The conditional strong-mixing coefficient of $Y_t$ (on $X_t$) verifies

(i) $\sum_{\ell=2}^{\infty} \{\alpha^*(\ell)\}^{(1-1/q-1/r)/2} < \infty$;

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

(ii) $\lim_{T\to\infty} k_T^{-1/2} \sum_{\tau=1}^{m_T} w_{\tau m_T} \frac{1}{\sqrt{T-\tau}} \sum_{\ell=2}^{T-\tau} \{\alpha^*(\min(\tau, \ell-1, |\ell-\tau-1|))\}^{(1-\sum_{i=1}^4 1/p_i)/2} = 0$, where $q$, $r$, $p_1$, $p_2$, $p_3$, and $p_4$ satisfy Assumption B1b; and $\{k_T, m_T\}$ satisfy Assumption B4.

Most conditions above are similar to those in Boente and Fraiman (1991, 1995) and Hansen (1992). These are required to show the weak convergence of the nearest-neighbor (NN) estimators $\widehat{f}_T(.)$ and $\widehat{g}_T(.)$, and the weak consistency of the standard heteroskedasticity and autocovariance matrix (HAC) estimators, respectively. Assumptions B1, the second part of B4c, and B5c are needed as well. Assumption B1, loosely speaking, just imposes finite bounds on the higher-order moments of $\frac{Y_t - g(X_t)}{f(X_t)}$. These moments arise in the proof of the $L_p$-convergence-to-zero of the biases of the NN estimators employed.[9] The second part of Assumption B4c, $m_T/\min(k_T, h_T) = o(1)$, controls the growth of the lag truncation parameter $m_T$ with respect to the growth of the smoothing parameters, $k_T$ and $h_T$, as $T$ diverges to infinity. That is, $m_T$ must diverge slower than $\min(k_T, h_T)$. Condition B5c is closer in spirit to Assumption A3a above. It essentially imposes a summability requirement of the conditional strong-mixing coefficient $\alpha^*(.)$ as $k_T$ and $m_T$ diverge with $T$. This condition can be effectively strengthened to a condition involving the summability of the standard strong-mixing coefficients by using the inequality employed to derive Assumption A3′ above.

THEOREM 2. *Let Assumptions* A1 *and* B1–B5 *hold. Then as* $T \longrightarrow \infty$, *we have that*

$$s^2_{\widehat{W}^*,2} \xrightarrow{P} \sigma^2_{W^*}.$$

Finally, notice that for i.i.d. data, the estimators $s^2_{\widehat{W}^*,1}$ and $s^2_{\widehat{W}^*,2}$ simplify to

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

The *strong* consistency of $s^2_{\widehat{W}^*,3}$ is established under substantially weaker conditions in the following theorem.[10]

THEOREM 3. *Suppose that* $\{Y_t, X_t\}_{t=1}^T$ *is a sequence of i.i.d. data and that* $\widehat{g}_T(\cdot)$ *is a uniformly strong consistent estimator for* $g(\cdot)$, *i.e.,* $\sup_{x\in\mathbb{R}} |\widehat{g}_T(x) - g(x)| = o_{a.s.}(1)$. *Under Assumption A2d, 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*

$$s^2_{\widehat{W}^*,3} \xrightarrow{a.s.} \sigma^2_{W^*} \equiv \mathrm{E}[W_0^{*2}] \tag{2.2}$$

*where* $W_0^* = \frac{Y_0 - g(X_0)}{f(X_0)}$.

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

## 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 $Y_t = 2X_t(1+e_t)\mathbb{I}(0 < X_t < 1)$, and $\{X_t\}_{t=1}^T$ follows the following data generating processes (DGPs), 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),

$$X_t = \begin{cases} .6X_{t-1} + Z_t & ; \quad X_{t-1} < 1, \\ -.5X_{t-1} + Z_t & ; \quad X_{t-1} \geq 1. \end{cases}$$

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

**DGP 4.** Tem Map model (TEM MAP),

$$X_t = \begin{cases} X_{t-1}/\alpha & ; \quad 0 < X_{t-1} < \alpha, \\ (1 - X_{t-1})/(1-\alpha) & ; \quad \alpha \leq X_{t-1} \leq 1, \end{cases}$$

where $\alpha = .49999$ and $Y_0$ is generated from $U[0,1]$.

**DGP 5.** Markov Regime-Switching model (MARKOV RS),

$$X_t = \begin{cases} .6X_{t-1} + Z_t & ; \quad S_t = 0, \\ -.5X_{t-1} + Z_t & ; \quad S_t = 1, \end{cases}$$

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)), $X_t = .5X_{t-1} + 10X_{t-1} \exp\left(-X_{t-1}^2\right) + Z_t$.

DGP 1 corresponds to Lewbel and Schennach's (2007) original design, while DGPs 2–6 are taken from Escanciano and Jacho-Chávez (2010). In each of 1,000 replications we generate sets of $T = 100$, 200, and 400 pseudo-random numbers following these DGPs. Estimators (1.4) and (1.5) are calculated for different values of $k = 1, 2, \ldots, 30$. Figures 1 and 2 display the finite sample behavior of the simulated $\{\widehat{\theta}^{(s)} - \theta_0\}_{s=1}^{1000}$ and $\{\widetilde{\theta}^{(s)} - \theta_0\}_{s=1}^{1000}$ as functions of $k$ respectively in the

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

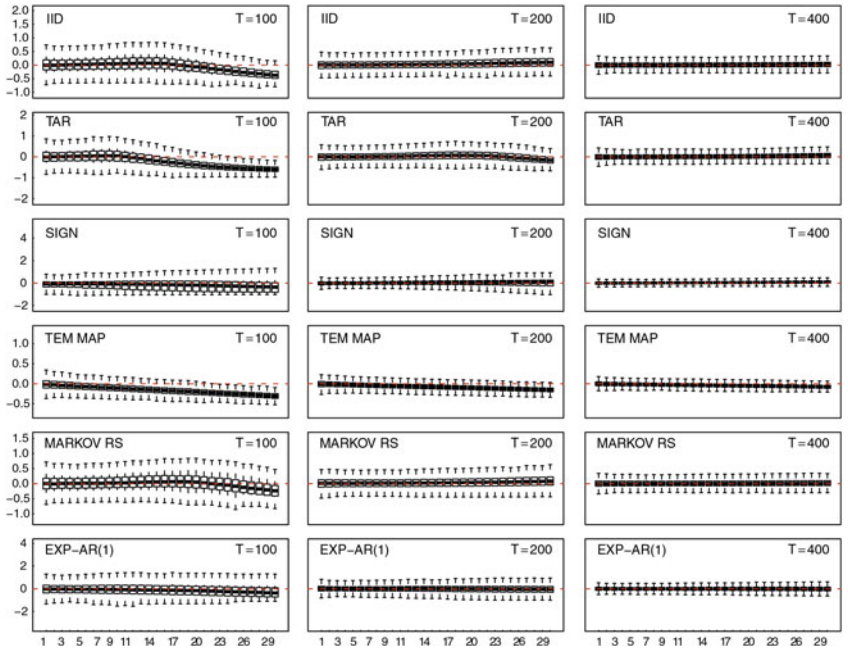



**FIGURE 1.** Semiparametric ordered-data estimator. Note: Box plots of 1,000 simulated differences between the ordered-data estimator and $\theta_0$, where $\widetilde{\theta} = k^{-1}\sum_1^{T-k} Y_{[t]}\left(X_{(t+k)} - X_{(t)}\right)$ for $k = 1, \ldots, 30$.

form of box plots for each design. These figures contrast the estimators' different bias and variance behaviors as $k$ changes. Overall, the results are qualitatively 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 the 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 A4, combined with the two-sided nature of the proposed estimator. On the other hand, the ordered-data estimator's simulated bias and variance display the standard trade-off usually found in nonparametric smoothers; see $T = 100$ in Figure 1. The simulated bias and variability of both estimators decrease 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 qualitatively similar for both estimators, but the proposed estimator displays less variation for DGPs 2 and 3. Furthermore, as predicted by our theory, both estimators have numerically similar variances in DGP 1.

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

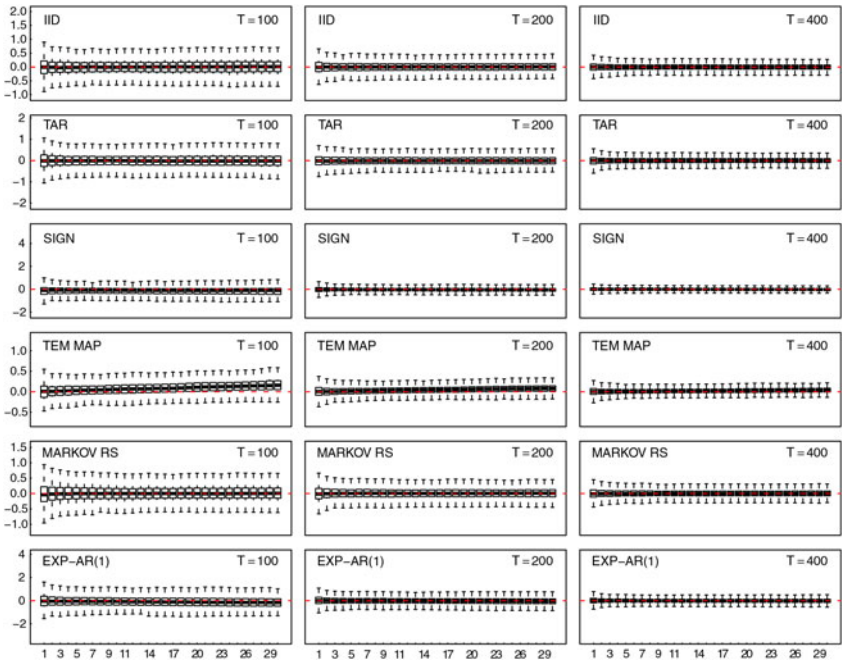



**FIGURE 2.** Semiparametric $k$-nearest neighbor based estimator. Note: Box plots of 1,000 simulated differences between the $k$-nearest neighbor based estimator and $\theta_0$, where $\widehat{\theta} = (2/k)\sum_1^T Y_t R_T(X_t, k)$ for $k = 1, \ldots, 30$.

## 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 a strong mixing condition, which is derived from Cox and Kim's (1995) inequality.

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 $(Y_t, X_t)$, for example. Similarly, as in the current literature of semiparametric inference using the $k$-NN method, the question of how one chooses the number of nearest neighbors, $k$, in a given application remains open. An automatic lag selection mechanism of $m$ can potentially be developed by non-trivial extensions of asymptotic arguments in Newey and West (1994). This is left for

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

**TABLE 1.** Monte Carlo performance at simulated optimal $k$

| DGP | $T$ | $k$ | Bias | Std. Dev. | RMSE | MAE | IQR |
|---|---|---|---|---|---|---|---|
| IID | 100 | 4 | 0.010 | 0.255 | 0.255 | 0.201 | 0.349 |
| | | 20 | 0.002 | 0.250 | 0.250 | 0.200 | 0.356 |
| | 200 | 6 | 0.002 | 0.169 | 0.169 | 0.135 | 0.230 |
| | | 16 | 0.000 | 0.170 | 0.169 | 0.135 | 0.222 |
| | 400 | 9 | 0.005 | 0.109 | 0.109 | 0.087 | 0.148 |
| | | 30 | 0.003 | 0.109 | 0.109 | 0.087 | 0.150 |
| TAR | 100 | 3 | 0.013 | 0.305 | 0.306 | 0.241 | 0.394 |
| | | 11 | −0.010 | 0.301 | 0.301 | 0.238 | 0.399 |
| | 200 | 3 | −0.004 | 0.210 | 0.209 | 0.166 | 0.280 |
| | | 15 | −0.010 | 0.210 | 0.210 | 0.167 | 0.286 |
| | 400 | 5 | −0.003 | 0.141 | 0.141 | 0.111 | 0.182 |
| | | 26 | −0.004 | 0.140 | 0.140 | 0.110 | 0.180 |
| SIGN | 100 | 3 | −0.064 | 0.452 | 0.490 | 0.341 | 0.449 |
| | | 10 | −0.081 | 0.441 | 0.490 | 0.336 | 0.436 |
| | 200 | 3 | 0.000 | 0.317 | 0.322 | 0.202 | 0.245 |
| | | 11 | −0.035 | 0.313 | 0.325 | 0.201 | 0.225 |
| | 400 | 4 | 0.011 | 0.173 | 0.173 | 0.114 | 0.175 |
| | | 10 | 0.000 | 0.174 | 0.175 | 0.113 | 0.165 |
| TEM MAP | 100 | 3 | −0.039 | 0.120 | 0.125 | 0.101 | 0.162 |
| | | 7 | 0.041 | 0.142 | 0.148 | 0.119 | 0.196 |
| | 200 | 5 | −0.024 | 0.082 | 0.087 | 0.068 | 0.103 |
| | | 9 | 0.031 | 0.094 | 0.098 | 0.078 | 0.124 |
| | 400 | 5 | −0.011 | 0.059 | 0.061 | 0.048 | 0.078 |
| | | 9 | 0.015 | 0.065 | 0.067 | 0.053 | 0.086 |
| MARKOV RS | 100 | 5 | 0.011 | 0.254 | 0.254 | 0.199 | 0.322 |
| | | 15 | 0.004 | 0.253 | 0.253 | 0.198 | 0.327 |
| | 200 | 8 | 0.014 | 0.169 | 0.170 | 0.135 | 0.220 |
| | | 24 | 0.010 | 0.171 | 0.172 | 0.136 | 0.223 |
| | 400 | 9 | −0.003 | 0.116 | 0.116 | 0.090 | 0.151 |
| | | 13 | −0.002 | 0.119 | 0.119 | 0.093 | 0.157 |
| EXP-AR(1) | 100 | 3 | −0.022 | 0.503 | 0.503 | 0.388 | 0.642 |
| | | 13 | −0.086 | 0.469 | 0.476 | 0.379 | 0.610 |
| | 200 | 6 | 0.007 | 0.308 | 0.308 | 0.239 | 0.379 |
| | | 13 | −0.009 | 0.309 | 0.309 | 0.242 | 0.382 |
| | 400 | 6 | 0.004 | 0.196 | 0.195 | 0.152 | 0.252 |
| | | 13 | −0.002 | 0.199 | 0.199 | 0.155 | 0.256 |

*Note:* Simulated bias (Bias), standard deviation (Std. Dev.), root mean squared error (RMSE), mean absolute error (MAE), and interquartile range (IQR) of 1,000 replications for $k$ minimizing the Monte Carlo RMSE.

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

future research. 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 $\mathrm{E}[W_t^2] = \infty$, $W_t = Z_t - \mathrm{E}[\mathrm{E}[Y_t|X_t]/f(X_t)]$, is also possible as in Khan and Tamer (2011) but beyond the scope of this paper. Finally, relaxation of Assumption A1 to allow for a larger class of dependent processes is also worth exploring, especially for possibly nonstationary processes.

## NOTES

1. Although other weighting schemes can be chosen as suggested in Mack and Rosenblatt (1979), the usage of (1.3) in (1.2) leads to various simplifications in the statistical theory.
2. However, notice that the contributions to the sums in (1.4) and (1.5) are the same for $X_{(1)}$.
3. This allows for probability density functions such as the standard normal and Student's $t$ commonly found in practice.
4. 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.
5. We are indebted to an anonymous referee for pointing this out and providing this counter example.
6. We thank an anonymous referee for pointing this out.
7. If $\ell = 2$, then $k = O(T^{3/4})$.
8. Other possibilities include Andrews (1991) and Smith (2005).
9. Note that the orders of these moments could potentially be lowered if the almost sure convergence-to-zero is used instead.
10. The proof is available from the authors upon request.

## REFERENCES

Aaronson, J., R. Burton, H. Dehling, D. Gilat, T. Hill, & B. Weiss (1996) Strong laws for *L*- and *U*- statistics. *Transactions of the American Mathematical Society* 348, 2845–2865.

Abadie, A. & G. Imbens (2006) Large sample properties of matching estimators for average treatment effects. *Econometrica* 74(1), 235–267.

Abadie, A. & G. Imbens (2011) Bias corrected matching estimators for average treatment effects. *Journal of Business Economics and Statistics* 29(1), 1–11.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59(3), 817–858.

Bhattacharya, P.K. & Y.P. Mack (1987) Weak convergence of *k*-nn density and regression estimators with varying *k* and applications. *Annals of Statistics* 15(3), 976–994.

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

Boente, G. & R. Fraiman (1990) Asymptotic distribution of robust estimators for nonparametric models from mixing processes. *Annals of Statistics* 18, 891–906.

Boente, G. & R. Fraiman (1991) Strong order of convergence and asymptotic distribution of nearest neighbor density estimates from dependent observations. *Sankhyā: Series A* 53(2), 194–205.

Boente, G. & R. Fraiman (1995) Asymptotic distribution of smoothers based on local means and local medians under dependence. *Journal of Multivariate Analysis* 54(1), 77–90.

Bradley, R.C. (2007) *Introduction to Strong Mixing Conditions*, vol. 1. Kendrick Press.

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

Clauser, J.F. & M.A. Horne (1974) Experimental consequences of objective local theories. *Physical Review D* 10, 526–535.
Cox, D. & T.Y. Kim (1995) Moment bounds for mixing random variables useful in nonparametric function estimation. *Stochastic Processes and Their Applications* 56, 151–159.
Delgado, M.A. (1992) Semiparametric generalized least squares in the multivariate nonlinear regression model. *Econometric Theory* 8, 203–222.
Delgado, M.A. & T. Stengos (1994) Semiparametric testing in non-nested econometric models. *Review of Economic Studies* 61(2), 291–303.
Escanciano, J.C. & D.T. Jacho-Chávez (2010) Approximating the critical values of Cramér-von Mises tests in general parametric conditional specifications. *Computational Statistics & Data Analysis* 54(3), 625–636.
Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. Academic Press.
Hall, P. & A. Yatchew (2005) Unified approach to testing functional hypotheses in semiparametric contexts. *Journal of Econometrics* 127(2), 225–252.
Hansen, B.E. (1992) Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* 60(4), 967–972.
Härdle, W. & T.M. Stoker (1989) Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* 84, 986–995.
Hausman, J.A. & W.K. Newey (1995) Nonparametric estimation of exact consumers surplus and deadweight loss. *Econometrica* 63(6), 1445–1476.
Heckman, J.J., H. Ichimura, & P. Todd (1998) Matching as an econometric evaluation estimator. *Review of Economic Studies* 65(2), 261–94.
Hong, Y. & H. White (2005) Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* 73(3), 837–901.
Honoré, B.E. & A. Lewbel (2002) Semiparametric binary choice panel data models without strictly exogenous regressors. *Econometrica* 70(5), 2053–2063.
Ibragimov, I.A. & Y.V. Linnik (1971) *Independent and stationary sequences of random variables*. Wolters-Noordhoff.
Jacho-Chávez, D.T. (2008) $k$ nearest-neighbor estimation of inverse density weighted expectations. *Economics Bulletin* 3(48), 1–6.
Jacho-Chávez, D.T. (2010) Optimal bandwidth choice for estimation of inverse conditional density weighted expectations. *Econometric Theory* 26(01), 94–118.
Jun, S.J. & J. Pinkse (2009a) Efficient semiparametric seemingly unrelated quantile regression estimation. *Econometric Theory* 25(05), 1392–1414.
Jun, S.J. & J. Pinkse (2009b) Semiparametric tests of conditional moment restrictions under weak or partial identification. *Journal of Econometrics* 152(1), 3–18.
Jun, S.J. & J. Pinkse (2009c) Testing Under Weak Identification with Conditional Moment Conditions. Unpublished manuscript.
Khan, S. & A. Lewbel (2007) Weighted and two-stage least squares estimation of semiparametric truncated regression models. *Econometric Theory* 23(2), 309–347.
Khan, S. & E. Tamer (2011) Irregular identification, support conditions, and inverse weight estimation. *Econometrica* 78(6), 2021–2042.
Lewbel, A. (1997) Semiparametric estimation of location and other discrete choice moments. *Econometric Theory* 13(01), 32–51.
Lewbel, A. (1998) Semiparametric latent variable model estimation with endogenous or mismeasured regressors. *Econometrica* 66(1), 105–121.
Lewbel, A. (2000) Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics* 97(1), 145–177.
Lewbel, A. (2007) Endogenous selection or treatment model estimation. *Journal of Econometrics* 141(2), 777–806.
Lewbel, A., O.B. Linton, & D. McFadden (2011) Estimating features of a distribution from binomial data. *Journal of Econometrics* 162(2), 170–188.

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

Lewbel, A. & S.M. Schennach (2007) A simple ordered data estimator for inverse density weighted expectations. *Journal of Econometrics* 127(1), 189–211.

Li, Q. (2006) Testing Parametric Regression Functional Forms Using *k*-Nearest Neighbor Method. Unpublished manuscript.

Liu, Z. & X. Lu (1997) Root-*n*-consistent semiparametric estimation of partially linear models based on *k*-NN method. *Econometric Reviews* 16(4), 411–420.

Loftsgaarden, D.O. & C.P. Quesenberry (1965) A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics* 36(3), 1049–1051.

Lu, Z. & P. Cheng (1998) Strong consistency of nearest neighbor kernel regression estimation for stationary dependent samples. *Science in China (Series A)* 41(9), 918–926.

Mack, Y.P. & M. Rosenblatt (1979) Multivariate *k*-nearest neighbor density estimates. *Journal of Multivariate Analysis* 9(1), 1–15.

McFadden, D. (1999) Computing willingness to pay in random utility models. In J.C. Moore, R. Reizman, & J.R. Melvin (eds.), *Trade, Theory and Econometrics*, chap. 15, pp. 253–274. Routledge.

McLeish, D.L. (1975) Invariance principles for dependent variables. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 32, 165–178.

Newey, W.K. (1990) Efficient instrumental variables estimation of nonlinear models. *Econometrica* 58(4), 809–37.

Newey, W.K. (1997) Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79(1), 147–168.

Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In D. McFadden & R.F. Engle (eds.), *Handbook of Econometrics*, vol. IV, pp. 2111–2245. Elsevier, North-Holland.

Newey, W.K. & P.A. Ruud (2005) Density weighted linear least squares. In D.W.K. Andrews & J.H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, chap. 23, p. 554–573. Cambridge University Press.

Newey, W.K. & K.D. West (1987) A simple positive definite heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.

Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies* 61(4), 631–653.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55(2), 277–301.

Pinkse, J. (2006) Heteroskedasticity Correction and Dimension Reduction. Unpublished manuscript.

Rao, B.L.S.P. (1999) *Semimartingales and their Statistical Inference*. Chapman & Hall/CRC.

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

Robinson, P.M. (1987) Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 55(4), 875–891.

Robinson, P.M. (1995) Nearest-neighbour estimation of semiparametric regression models. *Journal of Nonparametric Statistics* 5(1), 33–41.

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

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

Roussas, G.G. & D. Ioannides (1987) Moment inequalities for mixing sequences of random variables. *Stochastic Analysis and Applications* 5(1), 61–120.

Severini, T.A. & G. Tripathi (2001) A simplified approach to computing efficiency bounds in semiparametric models. *Journal of Econometrics* 102, 23–66.

Smith, R.J. (2005) Automatic positive semidefinate HAC covariance matrix and GMM estimation. *Econometric Theory* 21(1), 158–170.

Tran, L.T. (1993) Nonparametric function estimation for time-series by local average estimators. *Annals of Statistics* 21(2), 1040–1057.

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

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

Truong, Y.K. & C.J. Stone (1992) Nonparametric function estimation involving time series. *Annals of Statistics* 20(1), 77–97.

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

White, H. (2001) *Asymptotic Theory for Econometricians*, revised ed. Academic Press.

Yakowitz, S. (1987) Nearest-neighbor methods for time series analysis. *Journal of Time Series Analysis* 8(2), 235–247.

# APPENDIX A: Main Proofs

In this section we use the notation $k_T \doteq k$, $h_T \doteq h$, $m_T \doteq m$, and $\sum_1^T \doteq \sum_{t=1}^T$ interchangeably throughout.

**Proof of Theorem 1.** The proof of this theorem follows four steps.

*Step 1*. First, notice that

$$\begin{aligned}\widehat{\theta}_{k_T} - \theta_0 &= \frac{2}{k_T}\sum_{t=1}^{T} Y_t R(X_t, k_T) - \int_{\mathbb{R}} g(x)dx \\ &= \frac{1}{T}\sum_{t=1}^{T}\frac{Y_t - g(X_t)}{f(X_t)} + \sum_{t=1}^{T}\left\{\frac{2(Y_t - g(X_t))R(X_t, k_T)}{k_T} - \frac{1}{T}\frac{Y_t - g(X_t)}{f(X_t)}\right\} \\ &\quad + \sum_{t=1}^{T}\frac{2g(X_t)R(X_t, k_T)}{k_T} - \int_{\mathbb{R}} g(x)dx \\ &= \mathcal{L}_{T;1} + \mathcal{L}_{T;2} + \mathcal{L}_{T;3},\end{aligned}$$

where $g(x) \doteq \mathrm{E}[Y|x]$. Hence, the asymptotic behavior of $\widehat{\theta}_{k_T} - \theta_0$ 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 $h(X, Y) = (Y - g(X))/f(X)$, thus $m_h(X_t) = \mathrm{E}[h(X_t, Y_t)|X_t] = 0$. Since $h(x, y)$ is a measurable function of $(x, y)$, it then follows that $h(X_t, Y_t)$ is also strongly mixing with the same mixing coefficient as $(X_t, Y_t)$. In view of Assumptions A2e and A3a, an application of the central limit theorem for strongly mixing stationary sequences (see, e.g., Ibragimov and Linnik, 1971, Thm. 18.5.3, p. 346) yields

$$\sqrt{T}\mathcal{L}_{T;1} \xrightarrow{W} N\left(0, \sigma_{W^*}^2\right). \tag{A.1}$$

*Step 3*. We now verify that $\sqrt{T}\mathcal{L}_{T;2} = o_P(1)$. First notice that

$$\mathcal{L}_{T;2} = \frac{1}{T}\sum_{1}^{T}[Y_t - g(X_t)]\left\{\frac{2T R(X_t, k)}{k} - \frac{1}{f(X_t)}\right\}.$$

Assumption A1 implicitly implies that the law of iterated expectation yields $\mathrm{E}[\sqrt{T}\mathcal{L}_{T;2}] = T^{-1/2}\sum_{t=1}^{T}\mathrm{E}[\{2T R(X_t, k)/k - 1/f(X_t)\}\mathrm{E}[Y_t - g(X_t)|X_t]] = 0$, and we only need to show that $\lim_{T\longrightarrow\infty}\mathrm{E}[(\sqrt{T}\mathcal{L}_{T;2})^2] = 0$. We proceed to write

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

$$\begin{aligned}\mathrm{E}\left[\left(\sqrt{T}\mathcal{L}_{T;2}\right)^2\right] &= \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left[[Y_t - g(X_t)]^2\left\{\frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)}\right\}^2\right] \\ &\quad + \frac{1}{T}\sum_{\substack{s=1\\t=1\\s\neq t}}^{T}\mathrm{E}\left[[Y_t - g(X_t)][Y_s - g(X_s)]\left\{\frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)}\right\}\right. \\ &\quad \left.\times\left\{\frac{2TR(X_s,k)}{k} - \frac{1}{f(X_s)}\right\}\right] \\ &= \mathcal{L}_{T;2;I} + \mathcal{L}_{T;2;II}.\end{aligned}$$

Then by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathcal{L}_{T;2;I} &= \mathrm{E}\left[[Y_t - g(X_t)]^2\left\{\frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)}\right\}^2\right] \\ &\leq \mathrm{E}^{1/2}\left[\sup_{1\leq t\leq T}\left|\frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)}\right|^4\right]\mathrm{E}^{1/2}\left[[Y_t - g(X_t)]^4\right];\end{aligned}$$

and by Assumption A1, the law of the iterated expectation, the Jensen inequality, and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}&\left|\mathcal{L}_{T;2;II}\right| \\ &= \frac{1}{T}\left|\sum_{\substack{s=1\\t=1\\s\neq t}}^{T}\mathrm{E}\left[\left(\frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)}\right)\left(\frac{2TR(X_s,k)}{k} - \frac{1}{f(X_s)}\right)\right.\right. \\ &\qquad \left.\left.\times\mathrm{E}\left[[Y_t - g(X_t)][Y_s - g(X_s)]|X_s, X_t\right]\right]\right| \\ &\leq \frac{1}{T}\sum_{\substack{s=1\\t=1\\s\neq t}}^{T}\mathrm{E}\left[\sup_{1\leq t\leq T}\left|\frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)}\right|\sup_{1\leq s\leq T}\left|\frac{2TR(X_s,k)}{k} - \frac{1}{f(X_s)}\right|\right. \\ &\qquad \left.\times\left|\mathrm{E}[[Y_t - g(X_t)][Y_s - g(X_s)]|X_s, X_t]\right|\right] \\ &\leq \left\|\sup_{1\leq t\leq T}\left|\frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)}\right|\right\|_{2p}^2\sum_{\tau=1}^{\infty}\left\|\mathrm{E}\left[Y_\tau Y_0|X_\tau, X_0\right] - g(X_\tau)g(X_0)\right\|_{p/(p-1)}.\end{aligned}$$

Assumptions A2c, A2d, A3, and Lemma 1 imply that $\mathrm{E}^{1/2}\left[[Y_t - g(X_t)]^4\right] < \infty$ and $\sum_{\tau=1}^{\infty}\left\|\mathrm{E}\left[Y_\tau Y_0|X_\tau, X_0\right] - g(X_\tau)g(X_0)\right\|_{p/(p-1)} < \infty$. Thus, in order to prove that $\mathcal{L}_{T;2;I} = o(1)$ and $\mathcal{L}_{T;2;II} = o(1)$, we need to show that

$$\mathrm{E}\left[\sup_{1\leq t\leq T}\left|\frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)}\right|^4\right] = o(1), \quad \text{and} \tag{A.2}$$

$$\mathrm{E}\left[\sup_{1\leq t\leq T}\left|\frac{2TR(X_t,k)}{k} - \frac{1}{f(X_t)}\right|^{2p}\right] = o(1). \tag{A.3}$$

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

In fact, (A.2) and (A.3) immediately follow from the dominated convergence theorem if one can show that

$$\lim_{T\to\infty}\sup_{1\le t\le T}\left|\frac{2TR(X_t,k)}{k}-\frac{1}{f(X_t)}\right|=0 \qquad \text{P-a.s.} \tag{A.4}$$

Hence, it is necessary to demonstrate that

$$\lim_{T\to\infty}\left|\frac{2TR(X_t,k)}{k}-\frac{1}{f(X_t)}\right|=0 \tag{A.5}$$

almost everywhere (*pointwise*) for every $X_t$ in the sample. First, we shall define a random event, $\mathcal{A}_t=\{\omega\in\Omega:\ |2TR(X_t,k)/k-1/f(X_t)|<\epsilon\}$, where $\epsilon$ is some arbitrarily small positive generic constant. Thus, in light of the Borel-Cantelli lemma, (A.5) follows if

$$\lim_{T\to\infty}\sum_{t=1}^{T}P(\mathcal{A}_t)=\infty, \quad \text{which is analogous to} \lim_{T\to\infty}\sum_{t=1}^{T}P(\mathcal{A}_t^c)<\infty. \tag{A.6}$$

In order to prove (A.6) we notice that $\mathcal{A}_t^c$ can be rewritten as $\mathcal{A}_t^c=\mathcal{A}_{t;1}^c\cup\mathcal{A}_{t;2}^c$, where $\mathcal{A}_{t;1}^c \doteq \{\omega\in\Omega:\ 2TR(X_t,k)/k-1/f(X_t)\ge\epsilon\}$ and $\mathcal{A}_{t;2}^c \doteq \{\omega\in\Omega: 2TR(X_t,k)/k-1/f(X_t)\le-\epsilon\}$. This implies that

$$P\left(\mathcal{A}_t^c\right)\le P\left(\mathcal{A}_{t;1}^c\right)+P\left(\mathcal{A}_{t;2}^c\right). \tag{A.7}$$

Now in order to bound the probabilities $P(\mathcal{A}_{t;1}^c)$ and $P(\mathcal{A}_{t;2}^c)$ in (A.7), we notice that $\{2TR(X_t,k)/k-1/f(X_t)\ge\epsilon\}\equiv\{R(X_t,k)\ge k(2T)^{-1}((1+\epsilon f(X_t))/f(X_t))\}$; and for some $s\neq t$ we define $q_{tT}=P\{\|X_s-X_t\|\le\delta_{1T}\}$, where $\delta_{1T}=k(2T)^{-1}[(\epsilon f(X_t)+1)/f(X_t)]$. Lemma 2 implies

$$\lim_{T\to\infty}\frac{q_{tT}}{\frac{k}{T}\left(\frac{\epsilon f(X_t)+1}{f(X_t)}\right)}=f(X_t)\Longleftrightarrow\lim_{T\to\infty}\frac{Tq_{tT}}{k}=\epsilon f(X_t)+1. \tag{A.8}$$

In addition, after defining the Bernoulli random variable, $K_s=\mathbb{I}(\|X_s-X_t\|\ge\delta_{1T})$. It immediately follows that $P(\mathcal{A}_{1t}^c)=P\{R(X_t,k)\ge\delta_{1T}\}\le P\{\sum_{s=1}^{T}K_s>T-k\}=P\{\sum_{s=1}^{T}(K_s-\mathrm{E}[K_s])>T-k-T\mathrm{E}[K_s]\}=P\{\sum_{s=1}^{T}(K_s-\mathrm{E}[K_s])>Tq_{tT}-k\}$. Equation (A.8) implies that $k/Tq_{tT}\le\epsilon\sup_{x\in C_{X_t}}f(x)+1$, 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 $a_{\epsilon t}\doteq 1+\epsilon\sup_{x\in C_{X_t}}f(x)$, we then obtain

$$\begin{aligned}
P\left(\mathcal{A}_{1t}^c\right)&\le P\left\{\sum_{s=1}^{T}(K_s-\mathrm{E}[K_s])>Tq_{tT}\left(1-\frac{k}{Tq_{tT}}\right)\right\}\\
&\le P\left\{\sum_{s=1}^{T}(K_s-\mathrm{E}[K_s])>Tq_T\frac{a_{\epsilon t}-1}{a_{\epsilon t}}\right\}\\
&\le\frac{\mathrm{E}\left[\left\{\sum_{t=1}^{T}(K_s-\mathrm{E}[K_s])\right\}^{2\ell}\right]}{(Tq_{tT}a_{\epsilon t})^{2\ell}},
\end{aligned} \tag{A.9}$$

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

where the last inequality follows from Tchebyshev's inequality with $\ell$ representing a positive integer. Now since $K_s$ is the indicator function of $X_s$, one can show that $K_s - \mathrm{E}[K_s]$ is also mixing with the same coefficient, $\alpha(s)$, as $X_s$ by using the inequality in Truong and Stone (1992, Lem. 1, p. 82), i.e., $|\mathrm{E}[UV] - \mathrm{E}[U]\mathrm{E}[V]| \leq 4B_1B_2\alpha(t-s)$, 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 3 yields

$$P\left(\mathcal{A}_{1t}^c\right) \leq C_\ell \left\{ \frac{T^\ell}{k^{2\ell}} \frac{M_{t,p\ell}^{2\ell} \Sigma_P^\infty i^{\ell-1}\alpha^{1-2/p}(i)}{\left(\frac{Tq_{tT}}{k}\right)^{2\ell} a_{\epsilon t}^{2\ell}} + \sum_1^\ell \frac{T^s}{k^{2\ell}} \frac{P^{2\ell-s} v_t^s}{\left(\frac{Tq_{tT}}{k}\right)^{2\ell} a_{\epsilon t}^{2\ell}} \right\},$$

where $M_{t,p\ell} = \left[q_{tT}^{p\ell}(1-q_{tT}) + (1-q_{tT})^{p\ell}\right]^{1/p\ell} \leq 1$ and $v_t = q_{tT}^k(1-q_{tT}) + (1-q_{tT})^k q_{tT}$. Hence, we have

$$\begin{aligned}\sum_{t=1}^T P\left(\mathcal{A}_{1t}^c\right) &\leq C_\ell \left\{ \frac{T^\ell}{k^{2\ell}} \left(\sum_P^\infty i^{\ell-1}\alpha^{1-2/p}(i)\right) \sum_{t=1}^T \frac{M_{t,p\ell}^{2\ell}}{\left(\frac{Tq_{tT}}{k}\right)^{2\ell} a_{\epsilon t}^{2\ell}} + \sum_{s=1}^\ell \frac{T^s}{k^{2\ell}} P^{2\ell-s} \sum_1^T \frac{v_t^s}{\left(\frac{Tq_{tT}}{k}\right)^{2\ell} a_{\epsilon t}^{2\ell}} \right\} \\ &\leq C_\ell \left\{ \frac{T^{\ell+1}}{k^{2\ell}} \left(\sum_P^\infty i^{\ell-1}\alpha^{1-2/p}(i)\right) \sup_t \left| \frac{M_{t,p\ell}^{2\ell}}{\left(\frac{Tq_{tT}}{k}\right)^{2\ell} a_{\epsilon t}^{2\ell}} \right| \sum_{s=1}^\ell \frac{T^s}{k^{2\ell}} P^{2\ell-s} \sup_t \left| \frac{v_t^s}{\left(\frac{Tq_{tT}}{k}\right)^{2\ell} a_{\epsilon t}^{2\ell}} \right| \right\}. \end{aligned} \tag{A.10}$$

Assumptions A3b, A4, and the fact that $M_{t,p\ell}^{2\ell}/(Tq_{tT}/k)^{2\ell} a_{\epsilon t}^{2\ell} \longrightarrow 1/(\epsilon f(X_t)+1)^{2\ell} a_{\epsilon t}^{2\ell} < \infty$ and $v_t^s/(\frac{Tq_{tT}}{k})^{2\ell} a_{\epsilon t}^{2\ell} \longrightarrow 1/(\epsilon f(X_t)+1)^{2\ell} a_{\epsilon t}^{2\ell} < \infty$ as $T \longrightarrow \infty$ then imply

$$\sum_{t=1}^\infty P\left(\mathcal{A}_{1t}^c\right) < \infty. \tag{A.11}$$

Using the exact same arguments, we can also prove that

$$\sum_{t=1}^\infty P\left(\mathcal{A}_{2t}^c\right) < \infty. \tag{A.12}$$

Hence, results (A.11) and (A.12) prove (A.6) in view of (A.7). It then follows that

$$\sqrt{T}|\mathcal{L}_{T;2}| = o_P(1). \tag{A.13}$$

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

$$\mathcal{L}_{T;3} = \frac{1}{T}\sum_1^T \left\{ \frac{2g(X_t)TR(X_t,k)}{k} - \mathrm{E}\left[\frac{g(X_t)}{f(X_t)}\right] \right\}.$$

Now using the variable transformation for integrals we obtain for finite $T$, $\mathrm{E}[g(X_t)/f(X_t)] \approx T^{-1}\sum_{t=1}^{T-1} \int_{F^{-1}(\frac{t}{T})}^{F^{-1}((t+1)/T)} g(u)/f(u) = T^{-1}\sum_{t=1}^{T-1} g(\xi_{(t)})/f(\xi_{(t)})$ we obtain an exact equality as $T$ diverges to infinity, where $F^{-1}(s) \doteq \inf\{x : F(x) \geq s\}$ and $\xi_{(t)}$ is some

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

real number in the interval $\left[F^{-1}(\frac{t}{T}), F^{-1}(\frac{t+1}{T})\right)$, we further obtain

$$\begin{aligned}\mathcal{L}_{T;3} &\approx \frac{1}{T}\sum_{1}^{T}\left\{\frac{2g\left(X_{(t)}\right)TR\left(X_{(t)},k\right)}{k} - \frac{g\left(\xi_{(t)}\right)}{f\left(\xi_{(t)}\right)}\right\} \\ &= \frac{1}{T}\sum_{1}^{T}\left\{\frac{2g(X_t)TR(X_t,k)}{k} - \frac{g(X_t)}{f(X_t)}\right\} + \frac{1}{T}\sum_{1}^{T}\left\{\frac{g\left(X_{(t)}\right)}{f\left(X_{(t)}\right)} - \frac{g\left(\xi_{(t)}\right)}{f\left(\xi_{(t)}\right)}\right\} \\ &= \mathcal{L}_{T;3a} + \mathcal{L}_{T;3b}.\end{aligned}$$

Note that it is straightforward to verify that the boundary term in this equation disappears at the order $o_p(1)$. In view of Assumption A2a and equation (A.5), we can immediately prove that $\sqrt{T}|\mathcal{L}_{T;3a}| = o_{a.s.}(1)$. However, the second term, $\mathcal{L}_{T;3b}$, requires some extra steps as follows.

An application of the stochastic mean value theorem yields

$$\begin{aligned}\sqrt{T}\left|\mathcal{L}_{T;3b}\right| &= \frac{1}{\sqrt{T}}\left|\sum_{1}^{T}\frac{\partial}{\partial x}\left(\frac{g(x)}{f(x)}\right)\Big|_{x=\eta_{(t)}}\left(X_{(t)} - \xi_{(t)}\right)\right| \\ &\leq \left(\frac{1}{T}\sum_{1}^{T}\left\{\frac{\partial}{\partial x}\left(\frac{g(x)}{f(x)}\right)\Big|_{x=\eta_{(t)}}\right\}^2\right)^{1/2}\left(\sum_{1}^{T}\left(X_{(t)} - \xi_{(t)}\right)^2\right)^{1/2} \\ &= \bar{\mathcal{L}}_{T;3b;I}\bar{\mathcal{L}}_{T;3b;II},\end{aligned}$$

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 A2b gives

$$\bar{\mathcal{L}}_{T;3b;I} \longrightarrow \mathrm{E}^{1/2}\left[\left\{\frac{\partial}{\partial x}\left(\frac{g(x)}{f(x)}\right)\Big|_{x=X_t}\right\}^2\right] < \infty.$$

Moreover, note that

$$\begin{aligned}\sum_{1}^{T}\left[X_{(t)} - F^{-1}(t/T)\right]^2 &= \sum_{1}^{T}\left[F_T^{-1}(t/T) - F^{-1}(t/T)\right]^2 \\ &\leq \sup_{u\in[0,1]}\left|F_T^{-1}(u) - F^{-1}(u)\right|\sum_{1}^{T}\left|F_T^{-1}(t/T) - F^{-1}(t/T)\right|. \quad \textbf{(A.14)}\end{aligned}$$

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

$$\lim_{T\to\infty}\sup_{u\in[0,1]}\left|F_T^{-1}(u) - F^{-1}(u)\right| = 0 \qquad \text{P-a.s., and}$$

$$\sum_{1}^{T}\left|F_T^{-1}(t/T) - F^{-1}(t/T)\right| < \infty \qquad \text{P-a.s.}$$

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

Then in view of equation (A.14), we have

$$\lim_{T\to\infty}\sum_{1}^{T}\left[X_{(t)}-F^{-1}(t/T)\right]^2=0 \qquad \text{P-a.s.} \tag{A.15}$$

Considering the fact that $|F^{-1}(u)-F^{-1}(v)|\longrightarrow 0$ as $|u-v|\longrightarrow 0$ 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.15) yields

$$\lim_{T\to\infty}\bar{\mathcal{L}}_{T;3b;II}=0 \qquad \text{P-a.s.,} \tag{A.16}$$

and it immediately follows that $\sqrt{T}\left|\mathcal{L}_{T;3b}\right|=o_{a.s.}(1)$. Thus we obtain

$$\sqrt{T}\left|\mathcal{L}_{T;3}\right|=o_{a.s.}(1). \tag{A.17}$$

Therefore, equations (A.1), (A.13), and (A.17) imply the main result of this theorem. ■

**Proof of Theorem 2.** The proof of this result follows 3 steps.

*Step 1*. First, define

$$s^2_{W^*,2}=\frac{1}{T}\sum_{t=1}^{T}W_t^{*2}+2\sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{T-\tau}\sum_{t=1}^{T-\tau}W_t^*W_{t+\tau}^*.$$

The triangle inequality yields

$$\begin{aligned}\left|s^2_{\widehat{W^*},2}-\sigma^2_{W^*}\right| &\le \left|s^2_{W^*,2}-\sigma^2_{W^*}\right|+\left|s^2_{\widehat{W^*},2}-s^2_{W^*,2}\right|\\ &=\mathcal{H}_{T;1}+\mathcal{H}_{T;2}.\end{aligned}$$

By virtue of Lemma 7, Assumptions B1a, B4c, and B5b imply that

$$\mathcal{H}_{T;1}=o_p(1). \tag{A.18}$$

Thus, it remains to show that $\mathcal{H}_{T;2}=o_p(1)$ in the next step.

*Step 2*. We rewrite the term $\mathcal{H}_{T;2}$ as

$$\begin{aligned}\mathcal{H}_{T;2}&=\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{W}_t^{*2}-W_t^{*2}\right)+2\sum_{\tau=1}^{m}\left(1-\frac{\tau}{m+1}\right)\frac{1}{T-\tau}\sum_{t=1}^{T-\tau}\left(\widehat{W}_t^*\widehat{W}_{t+\tau}^*-W_t^*W_{t+\tau}^*\right)\\ &=\mathcal{H}_{T;2;1}+2\mathcal{H}_{T;2;2}.\end{aligned}$$

To show that $\mathcal{H}_{T;2;1}=o_p(1)$, using the inequality $\left||x|^p-|y|^q\right|\le p2^{p-1}\left\{|y|^{p-1}|x-y|+|x-y|^p\right\}$ one can write this term as

$$\begin{aligned}\mathcal{H}_{T;2;1}&=\frac{1}{T}\sum_{t=1}^{T}\left(\left\{\frac{Y_t-\widehat{g}(X_t)}{\widehat{f}(X_t)}\right\}^2-\left\{\frac{Y_t-g(X_t)}{\widehat{f}(X_t)}\right\}^2\right)+\frac{1}{T}\sum_{t=1}^{T}\left(\left\{\frac{Y_t-g(X_t)}{\widehat{f}(X_t)}\right\}^2-\left\{\frac{Y_t-g(X_t)}{f(X_t)}\right\}^2\right)\\ &\le\frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\}^2+2\frac{1}{T}\sum_{t=1}^{T}f(X_t)\left(\frac{g(X_t)-\widehat{g}(X_t)}{\widehat{f}^2(X_t)}\right)\left(\frac{Y_t-g(X_t)}{f(X_t)}\right)\\ &\quad+4\frac{1}{T}\sum_{t=1}^{T}\left|\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)}\right|\left(\frac{Y_t-g(X_t)}{f(X_t)}\right)^2+4\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)}\right)^2\left(\frac{Y_t-g(X_t)}{f(X_t)}\right)^2\\ &=\mathcal{H}_{T;2;1;a}+\mathcal{H}_{T;2;1;b}+\mathcal{H}_{T;2;1;c}+\mathcal{H}_{T;2;1;d}.\end{aligned}$$

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

$\mathcal{H}_{T;2;1;a} = o_p(1)$: An application of Minkowski's inequality yields $\left\|\mathcal{H}_{T;2;1;a}\right\|_2 \leq \sup_{1\leq t\leq T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_4^2$. In light of Assumptions B1a, B2, B3, B4a, B4b, B5a, and B5b, the conditioning device together with an application of Lemmas 5 and 6 yield $\widehat{f}(X_t) \xrightarrow{p} f(X_t)$ and $h_T^{1/2}\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)} \xRightarrow{W} \Xi$ where $\Xi \sim N\left(\frac{b}{f(X_t)}, \frac{\sigma^2(X_t)}{f(X_t)}\right)$ (conditional on $X_t$), $\max\left(\frac{b}{f(X_t)}, \frac{\sigma^2(X_t)}{f(X_t)}\right) < \infty,\ \forall\ x \in \text{Supp}(X)$, by Assumptions B1c and B3b. In view of Bradley (2007, Thm. 1.19), the asymptotic normality of a strictly stationary sequence of random variables, $h_T^{1/2}\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}$, implies that it is uniformly integrable (conditional on $X_t$). That is, it is also stochastically uniformly integrable in the sense of Rao (1999). Hence, an application of the stochastic dominated convergence theorem (see, e.g., Rao, 1999, p. 304) yields

$$\mathrm{E}\left[\left|h_T^{1/2}\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right|^p \Big| X_t\right] \xRightarrow{P} \mathrm{E}[\Xi^p|X_t] < \infty.$$

It thus follows that

$$\mathrm{E}\left[\left|h_T^{1/2}\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right|^p\right] \Longrightarrow \mathrm{E}\left[\Xi^p\right] < \infty, \qquad \forall\, t \in [1,T].$$

We then have that

$$\sup_{1\leq t\leq T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_p = O\left(h_T^{-1/2}\right) \quad \text{for some } p \geq 1. \tag{A.19}$$

Therefore, $\mathcal{H}_{T;2;1;a} = o_p(1)$.

$\mathcal{H}_{T;2;1;b} = o_p(1)$: Using Minkowski's and Hölder's inequalities, we obtain

$$\left\|\mathcal{H}_{T;2;1;b}\right\|_2 \leq 2\left(\max_x f(x)\right)\left\|\frac{Y_t-g(X_t)}{f(X_t)}\right\|_4 \sup_{1\leq t\leq T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}^2(X_t)}\right\|_4.$$

Similar arguments used in the proof of the first term yield $\sup_{1\leq t\leq T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}^2(X_t)}\right\|_4 = O(h_T^{-1/2})$. This along with Assumptions B1a and B2a implies that $\mathcal{H}_{T;2;1;b} = o_p(1)$.

$\mathcal{H}_{T;2;1;c} = o_p(1)$: Since

$$\left\|\mathcal{H}_{T;2;1;c}\right\|_2 \leq 4\left\|\frac{Y_t-g(X_t)}{f(X_t)}\right\|_8^2 \sup_{1\leq t\leq T}\left\|\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)}\right\|_4,$$

in view Assumption B1a, one merely needs to show that $\sup_{1\leq t\leq T}\left\|\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)}\right\|_4 = o(1)$. Reminiscent of the argument used to prove the first term, an application of Lemma 5 yields

$$\sup_{1\leq t\leq T}\left\|\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)}\right\|_p = O\left(k_T^{-1/2}\right) \quad \text{for some } p > 1. \tag{A.20}$$

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

Thus, equation (A.20) proves this result.

$\mathcal{H}_{T;2;1;d} = o_p(1)$: Since

$$\left\|\mathcal{H}_{T;2;1;d}\right\|_2 \le 4 \left\| \frac{Y_t - g(X_t)}{f(X_t)} \right\|_8^2 \sup_{1\le t\le T} \left\| \frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)} \right\|_8^2,$$

Assumption B1a and equation (A.20) yield the result. Hence, we obtain

$$\mathcal{H}_{T;2;1} = o_p(1). \tag{A.21}$$

*Step 3*. Next, we shall show that $\mathcal{H}_{T;2;2} = o_p(1)$. Note that, since $\widehat{W}_t^* = W_t^* + \left(\frac{Y_t - g(X_t)}{f(X_t)}\right)\left(\frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)}\right) + \frac{g(X_t) - \widehat{g}(X_t)}{\widehat{f}(X_t)}$, this term can be written as

$$\begin{aligned}
\mathcal{H}_{T;2;2} &= \sum_{\tau=1}^{m_T} \left(1 - \frac{\tau}{m_T+1}\right) \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left(\frac{Y_t - g(X_t)}{f(X_t)}\right)\left(\frac{Y_{t+\tau} - g(X_{t+\tau})}{f(X_{t+\tau})}\right)\left(\frac{f(X_t) - \widehat{f}(X_t)}{\widehat{f}(X_t) f(X_t)}\right) \\
&\quad + \sum_{\tau=1}^{m_T} \left(1 - \frac{\tau}{m_T+1}\right) \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left(\frac{Y_{t+\tau} - g(X_{t+\tau})}{f(X_{t+\tau})}\right)\left(\frac{g(X_t) - \widehat{g}(X_t)}{\widehat{f}(X_t)}\right) \\
&\quad + \sum_{\tau=1}^{m_T} \left(1 - \frac{\tau}{m_T+1}\right) \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left(\frac{Y_t - g(X_t)}{f(X_t)}\right)\left(\frac{Y_{t+\tau} - g(X_{t+\tau})}{f(X_{t+\tau})}\right)\left(\frac{f(X_{t+\tau}) - \widehat{f}(X_{t+\tau})}{\widehat{f}(X_{t+\tau})}\right) \\
&\quad + \sum_{\tau=1}^{m_T} \left(1 - \frac{\tau}{m_T+1}\right) \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left\{\left(\frac{Y_t - g(X_t)}{f(X_t)}\right)\left(\frac{Y_{t+\tau} - g(X_{t+\tau})}{f(X_{t+\tau})}\right)\right. \\
&\qquad \left. \times \left(\frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)}\right)\left(\frac{\widehat{f}(X_{t+\tau}) - f(X_{t+\tau})}{\widehat{f}(X_{t+\tau})}\right)\right\} \\
&\quad + \sum_{\tau=1}^{m_T} \left(1 - \frac{\tau}{m_T+1}\right) \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left(\frac{Y_{t+\tau} - g(X_{t+\tau})}{f(X_{t+\tau})}\right)\left(\frac{\widehat{g}(X_t) - g(X_t)}{\widehat{f}(X_t)}\right)\left(\frac{\widehat{f}(X_{t+\tau}) - f(X_{t+\tau})}{\widehat{f}(X_{t+\tau})}\right) \\
&\quad + \sum_{\tau=1}^{m_T} \left(1 - \frac{\tau}{m_T+1}\right) \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left(\frac{Y_t - g(X_t)}{f(X_t)}\right)\left(\frac{g(X_{t+\tau}) - \widehat{g}(X_{t+\tau})}{\widehat{f}(X_{t+\tau})}\right) \\
&\quad + \sum_{\tau=1}^{m_T} \left(1 - \frac{\tau}{m_T+1}\right) \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left(\frac{Y_t - g(X_t)}{f(X_t)}\right)\left(\frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)}\right)\left(\frac{\widehat{g}(X_{t+\tau}) - g(X_{t+\tau})}{\widehat{f}(X_{t+\tau})}\right) \\
&\quad + \sum_{\tau=1}^{m_T} \left(1 - \frac{\tau}{m_T+1}\right) \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left(\frac{\widehat{g}(X_t) - g(X_t)}{\widehat{f}(X_t)}\right)\left(\frac{\widehat{g}(X_{t+\tau}) - g(X_{t+\tau})}{\widehat{f}(X_{t+\tau})}\right) \\
&= \mathcal{K}_{T;1} + \mathcal{K}_{T;2} + \mathcal{K}_{T;3} + \mathcal{K}_{T;4} + \mathcal{K}_{T;5} + \mathcal{K}_{T;6} + \mathcal{K}_{T;7} + \mathcal{K}_{T;8}.
\end{aligned}$$

We shall now show that $\mathcal{K}_{T;1}$ ($\mathcal{K}_{T;3}$ follows similar arguments), $\mathcal{K}_{T;2}$ ($\mathcal{K}_{T;6}$ follows similar arguments), $\mathcal{K}_{T;4}$, $\mathcal{K}_{T;5}$ ($\mathcal{K}_{T;7}$ follows similar arguments), and $\mathcal{K}_{T;8}$ converge in probability to zero.

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

$\mathcal{K}_{T;1} = o_p(1)$: We first need to bound the moment

$$\begin{aligned}\mathcal{K}^*_{T;1} &= \mathrm{E}\left[\left|\frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{Y_t - g(X_t)}{f(X_t)}\right)\left(\frac{Y_{t+\tau} - g(X_{t+\tau})}{f(X_{t+\tau})}\right)\left(\frac{f(X_t) - \widehat{f}(X_t)}{\widehat{f}(X_t)f(X_t)}\right)\right|^2\right] \\ &= \frac{1}{T-\tau}\sum_{t=1}^{T-\tau}\mathrm{E}\left[\left|Z_t Z_{t+\tau}\left(\frac{f(X_t) - \widehat{f}(X_t)}{\widehat{f}(X_t)f(X_t)}\right)\right|^2\right] \\ &\quad + \frac{1}{T-\tau}\sum_{\substack{t=1\\ s=1\\ t\neq s}}^{T-\tau}\mathrm{E}\left[Z_t Z_s Z_{t+\tau} Z_{s+\tau}\left(\frac{f(X_t) - \widehat{f}(X_t)}{\widehat{f}(X_t)f(X_t)}\right)\left(\frac{f(X_s) - \widehat{f}(X_s)}{\widehat{f}(X_s)f(X_s)}\right)\right] \\ &= \mathcal{K}^*_{T;1;1} + \mathcal{K}^*_{T;1;2},\end{aligned}$$

where $Z_t = \frac{Y_t - g(X_t)}{f(X_t)}$. An application of Hölder's inequality yields

$$\mathcal{K}^*_{T;1;1} \leq \sup_{1\leq t\leq T}\left\|\frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)f(X_t)}\right\|_6^2 \left\|\frac{Y - g(X_t)}{f(X_t)}\right\|_6^4. \tag{A.22}$$

By virtue of Assumptions B1a, B2, B4a, and B5a, Lemma 5 asserts that (after conditioning on $X_t$), $\widehat{f}(X_t) \xrightarrow{P} f(X_t)$, and thus, $\sqrt{k_T}\left(\frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)f(X_t)}\right) \xrightarrow{W} N\left(0, 1/f^2(X_t)\right)$. Similar arguments used to prove the first term in Step 2 yield that, for every $t = 1, \ldots, T$,

$$\left\|\frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)f(X_t)}\right\|_p = O\left(k_T^{-1/2}\right) \quad \text{for some } p \geq 1. \tag{A.23}$$

Assumption B1a and equation (A.23) imply

$$\mathcal{K}^*_{T;1;1} = O\left(k_T^{-1/2}\right). \tag{A.24}$$

Moreover, using the law of iterated expectations together with Assumption A1 yields

$$\begin{aligned}\mathcal{K}^*_{T;1;2} &= \frac{1}{T-\tau}\sum_{\substack{t=1\\ s=1\\ t\neq s}}^{T-\tau}\mathrm{E}\left[\left(\frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)f(X_t)}\right)\left(\frac{\widehat{f}(X_s) - f(X_s)}{\widehat{f}(X_s)f(X_s)}\right)\right. \\ &\qquad \left.\times\, \mathrm{E}\left[Z_t Z_s Z_{t+\tau} Z_{s+\tau}\,|\mathcal{F}_{X,\max(s+\tau,t+\tau)}\right]\right].\end{aligned}$$

The Hölder's inequality and the conventional property of strictly stationary processes leads to

$$\mathcal{K}^*_{T;1;2} \leq 2\left\{\sup_{1\leq t\leq T}\left\|\frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)f(X_t)}\right\|_3\right\}^2 \sum_{\ell=2}^{T-\tau}\left\|\mathrm{E}\left[Z_1 Z_{\tau+1} Z_\ell Z_{\ell+\tau}\,|\mathcal{F}_{X,\max(\tau+1,\ell+\tau)}\right]\right\|_3.$$

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

Since $\mathrm{E}[Z_i|\mathcal{F}_{X,\max(\tau+1,\ell+\tau)}]=0$ for $i=1,\tau+1,\ell,\ell+\tau$, Lemma 4 asserts that

$$\mathcal{K}^*_{T;1;2} \le 60\left\{\sup_{1\le t\le T}\left\|\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)f(X_t)}\right\|_3\right\}^2\prod_{i=1}^4\left\|\|Z_1\|_{p_i,\mathcal{F}_{X_1}}\right\|_{12}\sum_{\ell=2}^{T-\tau}\left\{\alpha^*(\tau^*)\right\}^{1-\sum_{i=1}^4 1/p_i},$$

where $\sum_{i=1}^4\frac{1}{p_i}<1$ and $\tau^*=\min(\tau,\ell-1,|\ell-\tau-1|)$. In view of equations (A.22) and (A.25), an application of the Minkowski's inequality and the elementary inequality, $|x+y|^{1/2}\le|x|^{1/2}+|y|^{1/2}$, yields

$$\begin{aligned}\|\mathcal{K}_{T;1}\|_2 &\le \sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{\sqrt{T-\tau}}\mathcal{K}^{*1/2}_{T;1}\\ &\le \left\|\frac{Y_t-g(X_t)}{f(X_t)}\right\|_6^2\left\{\sup_{1\le t\le T}\left\|\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)f(X_t)}\right\|_6^2\right\}^{1/2}\sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{\sqrt{T-\tau}}\\ &\quad+\sqrt{60}\sup_{1\le t\le T}\left\|\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)f(X_t)}\right\|_3\prod_{i=1}^4\left\|\|Z_1\|_{p_i,\mathcal{F}_{X_1}}\right\|_{12}^{1/2}\\ &\quad\times\sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{\sqrt{T-\tau}}\sum_{\ell=2}^{T-\tau}\left\{\alpha^*(\tau^*)\right\}^{\frac{1}{2}\left(1-\sum_{i=1}^4 1/p_i\right)}.\end{aligned}$$

Moreover, since $\frac{m_T}{T-m_T}\longrightarrow 0<1$ as $T\longrightarrow\infty$, it follows that, as $T\longrightarrow\infty$, $\frac{1}{\sqrt{T-\tau}}\ll\frac{1}{\sqrt{T-m_T}}\ll\frac{1}{\sqrt{m_T}}$. Consequently, $\sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{\sqrt{T-\tau}}\ll\sqrt{m_T}$ as $T\longrightarrow\infty$. Therefore, equation (A.23) and Assumptions B1 and B5c imply that $\|\mathcal{K}_{T;1}\|_2=O\left(\sqrt{\frac{m_T}{k_T}}\right)=o(1)$, which proves that $\mathcal{K}_{T;1}=o_p(1)$.

$\mathcal{K}_{T;2}=o_p(1)$: As before, we start by bounding the moment:

$$\begin{aligned}\mathcal{K}^*_{T;2} &\le \mathrm{E}\left[\left|\frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{Y_{t+\tau}-g(X_{t+\tau})}{f(X_{t+\tau})}\right)\left(\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right)\right|^2\right]\\ &=\frac{1}{T-\tau}\sum_{t=1}^{T-\tau}\mathrm{E}\left[Z^2_{t+\tau}\left(\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right)^2\right]\\ &\quad+\frac{1}{T-\tau}\sum_{\substack{t=1\\s=1\\t\ne s}}^{T-\tau}\mathrm{E}\left[Z_{t+\tau}Z_{s+\tau}\left(\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right)\left(\frac{\widehat{g}(X_s)-g(X_s)}{\widehat{f}(X_s)}\right)\right]\\ &=\mathcal{K}^*_{T;2;1}+\mathcal{K}^*_{T;2;2}.\end{aligned}$$

An application of Hölder's inequality yields

$$\mathcal{K}^*_{T;2;1}\le\sup_{1\le t\le T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_4^2\|Z_t\|_4^2. \tag{A.25}$$

Moreover, using the law of iterated expectations, Hölder's inequality, and the conventional property of strict stationarity, one can derive

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

$$\mathcal{K}^*_{T;2;2} \le 2\left\{\sup_{1\le t\le T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_3\right\}^2\sum_{\ell=2}^{T-\tau}\left\|\mathrm{E}[Z_1Z_\ell|\mathcal{F}_{X,\ell}]\right\|_3.$$

Lemma 4 asserts that

$$\begin{aligned}\mathcal{K}^*_{T;2;2} \le{}& 20\left\{\sup_{1\le t\le T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_3\right\}^2\left\|\|Z_1\|_{q,\mathcal{F}_{X_1}}\right\|_6\left\|\|Z_1\|_{r,\mathcal{F}_{X_1}}\right\|_6\\ &\times\sum_{\ell=2}^{T-\tau}\{\alpha^*(\ell)\}^{1-1/q-1/r}, \end{aligned}\tag{A.26}$$

where $1/q+1/r<1$. In view of equations (A.25) and (A.26), an application of Minkowski's inequality yields

$$\begin{aligned}\left\|\mathcal{K}_{T;2}\right\|_2 \le{}& \left\{\sup_{1\le t\le T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_4^2\right\}^{1/2}\|Z_1\|_4\sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{\sqrt{T-\tau}}\\ &+\sqrt{20}\left\{\sup_{1\le t\le T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_3\right\}\sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{\sqrt{T-\tau}}\\ &\times\left\|\|Z_1\|_{q,\mathcal{F}_{X_1}}\right\|_6^{1/2}\left\|\|Z_1\|_{r,\mathcal{F}_{X_1}}\right\|_6^{1/2}\sum_{\ell=2}^{T-\tau}\{\alpha^*(\ell)\}^{(1-1/q-1/r)/2}\\ \le{}& \sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{\sqrt{T-\tau}}\left\{\left\{\sup_{1\le t\le T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_4^2\right\}^{1/2}\|Z_1\|_4\right.\\ &+\sqrt{20}\left\{\sup_{1\le t\le T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_3\right\}\left\|\|Z_1\|_{q,\mathcal{F}_{X_1}}\right\|_6^{1/2}\left\|\|Z_1\|_{r,\mathcal{F}_{X_1}}\right\|_6^{1/2}\\ &\left.\times\sum_{\ell=2}^{\infty}\{\alpha^*(\ell)\}^{(1-1/q-1/r)/2}\right\}.\end{aligned}$$

In light of (A.19) and Assumptions B1b and B5c, it is implied that $\left\|\mathcal{K}_{T;2}\right\|_2 = O\left(\sqrt{\frac{m_T}{h_T}}\right) = o(1)$, which proves the desired result.

$\mathcal{K}_{T;4} = o_p(1)$: First, we need to bound the moment

$$\begin{aligned}\mathcal{K}^*_{T;4} ={}& \mathrm{E}\left[\left|\frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}Z_tZ_{t+\tau}\left(\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)}\right)\left(\frac{\widehat{f}(X_{t+\tau})-f(X_{t+\tau})}{\widehat{f}(X_{t+\tau})}\right)\right|^2\right]\\ ={}& \frac{1}{T-\tau}\sum_{t=1}^{T-\tau}\mathrm{E}\left[\left|Z_tZ_{t+\tau}\left(\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)}\right)\left(\frac{\widehat{f}(X_{t+\tau})-f(X_{t+\tau})}{\widehat{f}(X_{t+\tau})}\right)\right|^2\right]\\ &+\frac{1}{T-\tau}\sum_{\substack{t=1\\ s=1\\ t\ne s}}^{T-\tau}\mathrm{E}\left[\left(\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)}\right)\left(\frac{\widehat{f}(X_{t+\tau})-f(X_{t+\tau})}{\widehat{f}(X_{t+\tau})}\right)\left(\frac{\widehat{f}(X_s)-f(X_s)}{\widehat{f}(X_s)}\right)\right.\end{aligned}$$

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

$$\times \left( \frac{\widehat{f}(X_{s+\tau}) - f(X_{s+\tau})}{\widehat{f}(X_{s+\tau})} \right) Z_t Z_{t+\tau} Z_s Z_{s+\tau} \Bigg]$$

$$\leq \|Z_1\|_8^4 \left\{ \sup_{1\leq t\leq T} \left\| \frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)} \right\|_8^2 \right\}^2$$

$$+ 60 \left\{ \sup_{1\leq t\leq T} \left\| \frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)} \right\|_5 \right\}^4 \prod_{i=1}^4 \left\| \|Z_1\|_{p_i,\mathcal{F}_{X_1}} \right\|_{20} \sum_{\ell=2}^{T-\tau} \{\alpha^*(\tau^*)\}^{1-\sum_{i=1}^4 1/p_i},$$

where the last inequality is, as before, derived by applying Hölder's inequality and Lemma 4. It then follows that, by Minkowski's inequality,

$$\left\|\mathcal{K}_{T;4}\right\|_2 \leq \|Z_1\|_8^2 \sup_{1\leq t\leq T} \left\| \frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)} \right\|_8^2 \sum_{\tau=1}^{m_T} \left(1 - \frac{\tau}{m_T+1}\right) \frac{1}{\sqrt{T-\tau}}$$

$$+ \sqrt{60} \left\{ \sup_{1\leq t\leq T} \left\| \frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)} \right\|_5 \right\}^2 \prod_{i=1}^4 \left\| \|Z_1\|_{p_i,\mathcal{F}_{X_1}} \right\|_{20}^{1/2}$$

$$\times \sum_{\tau=1}^{m_T} \left(1 - \frac{\tau}{m_T+1}\right) \frac{1}{\sqrt{T-\tau}} \sum_{\ell=2}^{T-\tau} \{\alpha^*(\tau^*)\}^{\frac{1}{2}(1-\sum_{i=1}^4 1/p_i)}.$$

Reminiscent of (A.20), Assumptions B1b, and B5c, the same arguments of the proof $\|\mathcal{K}_{T;1}\|_2 = o(1)$ yield that $\left\|\mathcal{K}_{T;4}\right\|_2 = O\left(\frac{\sqrt{m_T}}{k_T}\right) = o(1)$. This proves $\mathcal{K}_{T;4} = o_p(1)$ as required.

$\mathcal{K}_{T;5} = o_p(1)$: As before, we shall bound the moment

$$\mathcal{K}_{T;5}^* = \mathrm{E}\left[ \left| \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} Z_{t+\tau} \left( \frac{\widehat{g}(X_t) - g(X_t)}{\widehat{f}(X_t)} \right) \left( \frac{\widehat{f}(X_{t+\tau}) - f(X_{t+\tau})}{\widehat{f}(X_{t+\tau})} \right) \right|^2 \right].$$

The same arguments used before lead to the inequality

$$\mathcal{K}_{T;5}^* \leq \|Z_1\|_6^2 \sup_{1\leq t\leq T} \left\| \frac{\widehat{g}(X_t) - g(X_t)}{\widehat{f}(X_t)} \right\|_6^2 \sup_{1\leq t\leq T} \left\| \frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)} \right\|_6^2$$

$$+ 20 \left( \sup_{1\leq t\leq T} \left\| \frac{\widehat{g}(X_t) - g(X_t)}{\widehat{f}(X_t)} \right\|_5 \right)^2 \left( \sup_{1\leq t\leq T} \left\| \frac{\widehat{f}(X_t) - f(X_t)}{\widehat{f}(X_t)} \right\|_5 \right)^2$$

$$\times \left\| \|Z_1\|_{q,\mathcal{F}_{X_1}} \right\|_{10} \left\| \|Z_1\|_{r,\mathcal{F}_{X_1}} \right\|_{10} \sum_{\ell=2}^{T-\tau} \{\alpha^*(\ell)\}^{1-1/q-1/r}.$$

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

An application of Minkowski's inequality yields

$$\begin{aligned}\left\|\mathcal{K}_{T;5}\right\|_2 \leq \|Z_1\|_6 &\left\{\sup_{1\leq t\leq T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_6^2 \sup_{1\leq t\leq T}\left\|\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)}\right\|_6^2\right\}^{1/2} \\ &\times\left\{\sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{\sqrt{T-\tau}}\right\}+\sqrt{20}\left(\sup_{1\leq t\leq T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_5\right) \\ &\times\left(\sup_{1\leq t\leq T}\left\|\frac{\widehat{f}(X_t)-f(X_t)}{\widehat{f}(X_t)}\right\|_5\right)\left\|\|Z_1\|_{q,\mathcal{F}_{X_1}}\right\|_{10}^{1/2}\left\|\|Z_1\|_{r,\mathcal{F}_{X_1}}\right\|_{10}^{1/2} \\ &\times\sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{\sqrt{T-\tau}}\sum_{\ell=2}^{\infty}\{\alpha^*(\ell)\}^{\frac{1}{2}(1-1/q-1/r)}.\end{aligned}$$

In view of equations (A.19) and (A.20), Assumptions B1 and B5c lead to $\left\|\mathcal{K}_{T;5}\right\|_2 = O\left(\sqrt{\frac{m_T}{k_T h_T}}\right) = o(1)$. This proves that $\mathcal{K}_{T;5} = o_p(1)$.

$\mathcal{K}_{T;8} = o_p(1)$: An application of Minkowski's inequality yields

$$\left\|\mathcal{K}_{T;8}\right\|_2 \leq m_T\left\{\sup_{1\leq t\leq T}\left\|\frac{\widehat{g}(X_t)-g(X_t)}{\widehat{f}(X_t)}\right\|_2\right\}^2.$$

In view of (A.19), we have that $\left\|\mathcal{K}_{T;8}\right\|_2 = O\left(\frac{m_T}{h_T}\right) = o(1)$, which proves that $\mathcal{K}_{T;8} = o_p(1)$.

As a result, the above itemized substeps imply that

$$\mathcal{H}_{T;2;2} = o_p(1). \tag{A.27}$$

In light of equations (A.18), (A.21), and (A.27), the result follows. ■

# APPENDIX B: Auxiliary Results

The following lemmas are used to prove Theorems 1 and 2.

LEMMA 1. *Given integers, $p > 1$ and $q \geq p$, such that $1/p + 1/q < 1$, let $\left\|\|Y_0\|_{q,\mathcal{F}_{X,0}}\right\|_{2p/(p-1)} < \infty$ and $\sum_{t=1}^{\infty}\{\alpha^*(t)\}^{(p-1)/p-1/q} < \infty$ in Assumption A3. If Assumption A1 holds, then*

$$\sum_{\tau=1}^{\infty}\left\|\mathrm{E}\left[Y_0 Y_\tau | X_0, X_\tau\right] - g(X_0)g(X_\tau)\right\|_{p/(p-1)} < \infty, \quad \textit{and} \tag{B.1}$$

$$\sum_{\tau=1}^{\infty}\left\|\left\|\mathrm{E}[Y_\tau | X_\tau, \mathcal{F}_0] - g(X_\tau)\right\|_{2,\mathcal{F}_0}\right\|_{p/(p-1)}^{1/2} < \infty. \tag{B.2}$$

**Proof.** We only prove (B.1) as the proof of (B.2) follows similar steps (see, e.g., McLeish, 1975). In order to establish (B.1), we follow the same strategy in Hall and

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

Heyde (1980, p. 276) by first demonstrating that if $|Y_0| \le C_1$ and $|Y_t| \le C_2$ for some finite constants, $C_1$ and $C_2$, then

$$|\mathrm{E}[Y_0 Y_t | X_0, X_t] - g(X_0)g(X_t)| \le 4C_1C_2\alpha^*(t). \tag{B.3}$$

Assumption A1 implies

$$\begin{aligned}|\mathrm{E}[Y_0 Y_t | X_0, X_t] - g(X_0)g(X_t)| &= |\mathrm{E}[Y_0 \mathrm{E}[Y_t | X_0, Y_0, X_t] | X_0, X_t] - \mathrm{E}[Y_0|X_0]\mathrm{E}[Y_t|X_t]| \\ &\le C_1 \mathrm{E}\left[|\mathrm{E}[Y_t|X_0, Y_0, X_t] - \mathrm{E}[Y_t|X_t]| \,\middle|\, X_0, X_t\right] \\ &\le C_1 \left|\mathrm{E}\left[\xi\{\mathrm{E}[Y_t|X_0, Y_t, X_t] - \mathrm{E}[Y_t|X_t]\}\middle| X_0, X_t\right]\right|,\end{aligned} \tag{B.4}$$

where $\xi \doteq \mathrm{sgn}\,(\mathrm{E}[Y_t|X_0, Y_0, X_t] - \mathrm{E}[Y_t|X_t])$ is a Rademacher random variable measurable with respect to $\mathcal{F}_{X,0} \bigcap \mathcal{F}_{Y,0} \bigcap \mathcal{F}_X^t$; and

$$\begin{aligned}&|\mathrm{E}[\xi \mathrm{E}[Y_t|X_0, Y_0, X_t] - \xi \mathrm{E}[Y_t|X_t] | X_0, X_t]| \\ &\quad = |\mathrm{E}[\xi \mathrm{E}[Y_t|X_0, Y_0, X_t] | X_0, X_t] - \mathrm{E}[\xi|X_0, X_t]\mathrm{E}[Y_t|X_t]| \\ &\quad = |\mathrm{E}[\xi Y_t|X_0, Y_0, X_t] - \mathrm{E}[\xi|X_0, X_t]\mathrm{E}[Y_t|X_t]| \\ &\quad = |\mathrm{E}[Y_t\{\mathrm{E}[\xi|X_0, X_t, Y_t] - \mathrm{E}[\xi|X_0, X_t]\}|X_0, X_t]| \\ &\quad \le C_2|\mathrm{E}[\eta\{\mathrm{E}[\xi|X_0, X_t, Y_t] - \mathrm{E}[\xi|X_0, X_t]\}|X_0, X_t]| \\ &\quad = C_2\,|\mathrm{E}[\eta\xi|X_0, X_t] - \mathrm{E}[\eta|X_0, X_t]\mathrm{E}[\xi|X_0, X_t]|,\end{aligned} \tag{B.5}$$

where $\eta \doteq \mathrm{sgn}\,(\mathrm{E}[\xi|X_0, X_t, Y_t] - \mathrm{E}[\xi|X_0, X_t])$ is another Rademacher random variable measurable with respect to $\mathcal{F}_{X,0} \bigcap \mathcal{F}_X^t \bigcap \mathcal{F}_Y^t$.

Let us define the following random sets: $A = \{\omega\colon\ \xi = 1 | X_0 = x_0, X_t = x_t\} \in \mathcal{F}_{Y,0}$ and $B = \{\omega\colon\ \eta = 1 | X_0 = x_0, X_t = x_t\} \in \mathcal{F}_Y^t$ for some given observations, $x_0$ and $x_t$. Then we have

$$\begin{aligned}&|\mathrm{E}[\eta\xi|X_0 = x_0, X_t = x_t] - \mathrm{E}[\eta|X_0 = x_0, X_t = x_t]\mathrm{E}[\xi|X_0 = x_0, X_t = x_t]| \\ &\quad = \left|P(AB) + P\left(A^cB^c\right) - P\left(A^cB\right) - P\left(AB^c\right)\right. \\ &\qquad \left. - \left(P(A) - P\left(A^c\right)\right)\left(P(B) - P\left(B^c\right)\right)\right| \\ &\quad \le |P(AB) - P(A)P(B)| + |P\left(A^cB^c\right) - P\left(A^c\right)P\left(B^c\right)| \\ &\qquad + |P\left(A^cB\right) - P\left(A^c\right)P(B)| + |P\left(AB^c\right) - P(A)P\left(B^c\right)| \\ &\quad \le 4\alpha^*(t),\end{aligned} \tag{B.6}$$

where $A^c$ is the complement of $A$.

Hence, (B.3) follows from (B.4), (B.5), and (B.6).

Next, we shall establish that, if $Y_t \le C$ for a generic constant, $C$, and $\left\| \|Y_0\|_{p,\mathcal{F}_{X,0}} \right\|_{p/(p-1)} < \infty$, which is obvious in view of the first assumption and the inequality $\|Y\|_{p,\mathcal{I}} \le \|Y\|_{q,\mathcal{I}}$ if $p \le q$, for some $p > 1$, then

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

$$\|\mathrm{E}[Y_0Y_t|X_0,X_t]-\mathrm{E}[Y_0|X_0]\mathrm{E}[Y_t|X_t]\|_{p/(p-1)} \le 6C \left\| \|Y_0\|_{p,\mathcal{F}_{X,0}} \right\|_{p/(p-1)} \{\alpha^*(t)\}^{(p-1)/p}. \tag{B.7}$$

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

$$\begin{aligned}
&|\mathrm{E}[Y_0Y_t|X_0,X_t]-\mathrm{E}[Y_0|X_0]\mathrm{E}[Y_t|X_t]| \\
&\quad \le |\mathrm{E}[Y_{0,N}Y_t|X_0,X_t]-\mathrm{E}[Y_{0,N}|X_0]\mathrm{E}[Y_t|X_t]| + |\mathrm{E}[Y_{0,N}^{'}Y_t|X_0,X_t] \\
&\qquad -\mathrm{E}[Y_{0,N}^{'}|X_0]\mathrm{E}[Y_t|X_t]| \\
&\quad \le \mathcal{D}_1 + \mathcal{D}_2,
\end{aligned}$$

where $\mathcal{D}_1 \le 4CN\alpha^*(t)$ from (B.3). Moreover, Hölder's inequality together with Tchebyshev's inequality yield $\mathcal{D}_2 \le 2C\mathrm{E}[|Y_{0,N}^{'}|\,|X_0] \le 2CN^{1-p}\mathrm{E}[|Y_0|^p\,|X_0]$ for a generic constant $C$. Then by setting $N = \|Y_0\|_{p,\mathcal{F}_{X,0}}\{\alpha^*(t)\}^{-1/p}$, we have

$$|\mathrm{E}[Y_0Y_t|X_0,X_t]-\mathrm{E}[Y_0|X_0]\mathrm{E}[Y_t|X_t]| \le 6C\|Y_0\|_{p,\mathcal{F}_{X,0}}\{\alpha^*(t)\}^{(p-1)/p}.$$

Then (B.7) follows after taking $L_{p/(p-1)}$-norm on both sides of the above equation.

Finally, we now prove that (B.1) is implied by (B.3) and (B.7) under the assumptions of the lemma: First, define the truncated random variables, $Y_{t,M} \doteq Y_t\mathbb{I}(|Y_t| \le M)$ and $Y_{t,M}^{'} \doteq Y_t - Y_{t,M}$, where $M = \{\alpha^*(t)\}^{-1/q}\left\|\mathrm{E}[|Y_0|^q|X_0]\right\|_2^{1/(q(p-1)+p)}$ with $1/p + 1/q < 1$. Then we have

$$\begin{aligned}
&\|\mathrm{E}[Y_0Y_t|X_0,X_t]-\mathrm{E}[Y_0|X_0]\mathrm{E}[Y_t|X_t]\|_{p/(p-1)} \\
&\quad \le \|\mathrm{E}[Y_0Y_{t,M}|X_0,X_t]-\mathrm{E}[Y_0|X_0]\mathrm{E}[Y_{t,M}|X_t]\|_{p/(p-1)} + \|\mathrm{E}[Y_0Y_{t,M}^{'}|X_0,X_t] \\
&\qquad -\mathrm{E}[Y_0|X_0]\mathrm{E}[Y_{t,M}^{'}|X_t]\|_{p/(p-1)} \\
&\quad \le 6M \left\| \|Y_0\|_{p,\mathcal{F}_{X,0}} \right\|_{p/(p-1)} \{\alpha^*(t)\}^{(p-1)/p} \\
&\qquad + 2\left\| \|Y_0\|_{p,\mathcal{F}_{X,0}} \|Y_{t,M}^{'}\|_{p/(p-1),\mathcal{F}_{X_t}} \right\|_{p/(p-1)},
\end{aligned}$$

where the last inequality follows from (B.7) and Hölder's inequality. Moreover, an application of Tchebyshev's inequality yields

$$\begin{aligned}
&\left\| \|Y_0\|_{p,\mathcal{F}_{X,0}} \|Y_{t,M}^{'}\|_{p/(p-1),\mathcal{F}_{X_t}} \right\|_{p/(p-1)} \\
&\quad \le 2\frac{\left\| \|Y_0\|_{p,\mathcal{F}_{X,0}} \right\|_{2p/(p-1)} \left\| \mathrm{E}^{(p-1)/p}[|Y_0|^q\,|X_0] \right\|_{2p/(p-1)}}{M^{1-q\frac{p-1}{p}}} \\
&\quad = 2\left\| \|Y_0\|_{p,\mathcal{F}_{X,0}} \right\|_{2p/(p-1)} \left\| \mathrm{E}[|Y_0|^q\,|X_0] \right\|_2 \{\alpha^*(t)\}^{(p-1)/p-1/q} xt.
\end{aligned}$$

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

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

$$\begin{aligned}\|\mathrm{E}[Y_0 Y_t|X_0, X_t] - \mathrm{E}[Y_0|X_0]\mathrm{E}[Y_t|X_t]\|_{p/(p-1)} \\ \leq \left\| \|Y_0\|_{p,\mathcal{F}_{X,0}} \right\|_{2p/(p-1)} \left\{ 6 \left\| \mathrm{E}[|Y_0|^q|X_0] \right\|_2^{1/(q(p-1)+p)} \right. \\ \left. + 2 \left\| \mathrm{E}[|Y_0|^q \big| X_0] \right\|_2 \right\} \{\alpha^*(t)\}^{(p-1)/p-1/q},\end{aligned}$$

and from the assumptions of the lemma the right-hand side of the above inequality is finite and (B.1) follows immediately. ■

LEMMA 2 (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 $\lim_{h \longrightarrow 0} \int_{S_{x,h}} [f(u)/L(S_{x,h})]du = f(x)$, where $L(\cdot)$ is the Lebesgue measure on $\mathbb{R}$.*

**Proof.** See Wheeden and Zygmund (1977, p. 191). ■

LEMMA 3 (Cox and Kim, 1995). *Let $\xi_t$ be strong–mixing processes. Let $\ell$ be a positive integer and assume that $\mathrm{E}[\xi_t] = 0$, 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 $\mathrm{E}[|\xi_t^k|] \leq \nu$, 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*

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

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

**Proof.** See Cox and Kim (1995, Lem. 1, p. 155). ■

LEMMA 4. *Suppose that $\mathcal{F}_{s_i}^{t_i} = \sigma\left(\xi_{s_i}, \xi_{s_i+1}, \ldots, \xi_{t_i}\right)$, for $i = 1, \ldots, m$, are the Borel $\sigma$-algebras generated by a strong-mixing process, say $\xi_\tau$. Let $\xi_{\tau_i} \in \mathcal{F}_{s_i}^{t_i}$, where $s_1 \leq t_1 < s_2 \leq t_2 < \cdots < s_m \leq t_m$. Under the following condition,*

$$\left\| \left\| \xi_{\tau_i} \right\|_{p_i, \mathcal{B}_0} \right\|_{q_i} < \infty,$$

*where $\mathcal{B}_0$ represents a conditioning Borel sub $\sigma$-algebra, for some $p_i > 1$ and $q_i > 1$ such that $P = \sum_{i=1}^m \frac{1}{p_i} < 1$ and $\sum_{i=1}^m \frac{1}{q_i} = \frac{1}{r}$ with $r \geq 1$, respectively. Then*

$$\left\| \mathrm{E}\left[\prod_{i=1}^{m} \xi_{\tau_i} | \mathcal{B}_0\right] - \prod_{i=1}^{m} \mathrm{E}\left[\xi_{\tau_i} | \mathcal{B}_0\right] \right\|_r \leq 10(m-1)\left\{\alpha^*(\tau)\right\}^{1-P} \prod_{i=1}^{m} \left\| \left\| \xi_{\tau_i} \right\|_{p_i, \mathcal{B}_0} \right\|_{q_i},$$

*where $\alpha^*(\tau)$, with $\tau = \min\left\{s_{i+1} - t_i;\ \forall\, i = 1, \ldots, m-1\right\}$, is the conditional strong-mixing coefficient that was proposed by Rao (2009).*

**Proof.** This inequality can be immediately derived by using the induction hypothesis (as in the proof of Theorem 7.4 in Roussas and Ioannides, 1987) and the same conditioning method as in the proof of Lemma 1 above, together with Hölder's inequality for mixed norms; i.e., $\left\| \prod_{j=1}^k U_j \right\|_r = \prod_{j=1}^k \left\| U_j \right\|_{q_j}$ for $\frac{1}{r} = \sum_{j=1}^k \frac{1}{q_j}$. ■

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

LEMMA 5 (Boente and Fraiman, 1991). *Given a strictly stationary strong-mixing vector-valued process, $X_t$, the h-NN estimate of the probability density function $f(x)$ is defined as $\widehat{f}_T(x) = \frac{k_T}{T R_T^p(x,k_T)\|V_1\|}$, where $k_T$ is a fixed sequence of smoothing parameters and $\|V_1\|$ represents the volume of a unit ball. Suppose that the following conditions hold:*

(a) *The strong-mixing coefficients satisfy $n\sum_{j=n}^{\infty}\alpha(j) \longrightarrow 0$ as $n \longrightarrow \infty$;*

(b) *$k_T$ verifies $k_T = o\left(T^{2/(p+2)}\right)$, where $p$ is the dimension of $X_t$;*

(c) *$f(x)$ is locally Lipschitzian in the sense that $|f(u) - f(x)| \leq c\,|u - x|$ for some $c > 0$ and all $u$ in a $\delta$-neighbourhood of $x$;*

(d) *the joint density $f_\tau(u, v)$ of $(X_t, X_{t+\tau})$ is uniformly bounded in $\tau$.*

*Then we have that $k_T^{1/2}\left(\widehat{f}_T(x) - f(x)\right) \xRightarrow{W} N(0, f^2(x))$.*

**Proof.** See Boente and Fraiman (1991, Thm. 4, p. 201). ■

LEMMA 6 (Boente and Fraiman, 1995). *Given a strictly stationary strong-mixing vector-valued process, $(Y_t, X_t)$, the k-NN estimates of the conditional expectation $g(x) = \mathrm{E}[Y|X = x]$ are defined as*

$$\widehat{g}(x, R_T(x, h_T)) = \frac{\sum_{t=1}^{T} K\left(\frac{X_t - x}{R_T(x,h_T)}\right) Y_t}{\sum_{t=1}^{T} K\left(\frac{X_t - x}{R_T(x,h_T)}\right)},$$

*where $K(u) = \frac{1}{\|V_1\|}\mathbb{I}(u \in V_1)$ and $h_T$ is a fixed sequence of smoothing parameters. Suppose that the following conditions hold:*

(a) *$\|Y\|_s < \infty$ for some $s > 2$;*

(b) *The strong-mixing coefficients satisfy $\lim_{n\to\infty} n\sum_{\tau=n}^{\infty}\{\alpha(\tau)\}^{1-1/w} = 0$ with $w = s/(s-2)$;*

(c) *The probability density function of $X_t$ is continuous and positive everywhere;*

(d) *The joint probability density function of $(X_t, X_{t+\tau})$ is uniformly bounded for $\tau \geq 1$;*

(e) *Let $g^+(u) = \mathrm{E}[\max(Y,0)|X = u]$, $g^- = \mathrm{E}[\min(Y,0)|X = u]$, $\rho_\tau(u,v) = \mathrm{E}[Y_1 Y_{\tau+1}|X_1 = u, X_{\tau+1} = v]$, $\overline{\rho}_\tau(u,v) = \mathrm{E}[|Y_1 Y_{\tau+1}||X_1 = u, X_{\tau+1} = v]$, and $\sigma^2(u) = \mathrm{E}\left[(Y - g(X))^2|X = u\right]$. Then*

   (i) *$g^+$, $g^-$, $\overline{\rho}_\tau$, and $\sigma^2$ are continuous functions;*

   (ii) *$g(x)$ is Lipschitzian and $\lim_{\epsilon\to 0}[g(x+\epsilon u) - g(x)]/\epsilon = r'(x,u)$;*

(f) *There exists $0 \leq \beta < \infty$ such that $\lim_{T\to\infty} h_T^{1/p} T^{-2/[p(p+2)]} = \beta(\|V_1\| f(x))^{1/p}$, where $p$ is the dimension of $X_t$.*

*Then we have that $h_T^{1/2}\left(\widehat{g}(x, R_T(x,h_T)) - g(x)\right) \xRightarrow{W} N\left(b, \sigma^2(x)\right)$, where $b = b_1[\|V_1\| f(x)]^{1/2}$ with $b_1 = \beta^{1+p/2}\|V_1\|^{-1}\int_{V_1} r'(x,u)du$.*

**Proof.** See Boente and Fraiman (1995, Thm. 2.1, p. 81). ■

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press

LEMMA 7 (Hansen, 1992). *The Bartlett estimate of the long-run variance,* $\Omega$*, of an underlying strictly stationary strong-mixing zero-mean process,* $V_t$*, is defined as*

$$\widehat{\Omega} = \frac{1}{T}\sum_{t=1}^{T} V_t^2 + 2\sum_{\tau=1}^{m_T}\left(1-\frac{\tau}{m_T+1}\right)\frac{1}{T-\tau}\sum_{t=1}^{T-\tau} V_t V_{t+\tau},$$

*where* $m_T$ *denotes the lag-truncation parameter. Suppose that the following conditions hold:*

*(a)* $m_T^{1+2q}/T = O(1)$ *for some* $q \in (1/2, \infty)$*;*

*(b)* $\sum_{\tau=1}^{\infty}\{\alpha(\tau)\}^{2(1/r-1/p)} < \infty$ *for some* $r \in (2,4]$ *such that* $r > 2 + 1/q$*, and some* $p > r$*;*

*(c)* $\|V_t\|_p < \infty$.

*Then we have that, as* $T \longrightarrow \infty$, $\widehat{\Omega} \xrightarrow{p} \Omega$.

**Proof.** See Hansen (1992, Thm. 1, p. 968) ■

https://doi.org/10.1017/S026646661100079X Published online by Cambridge University Press