LIMIT THEOREMS FOR QUADRATIC FORMS OF MARKOV CHAINS

YVES F. ATCHADÉ AND MATIAS D. CATTANEO

(April 2011)

Abstract. We develop a martingale approximation approach to studying the limiting behavior of quadratic forms of Markov chains. We use the framework to examine two statistical problems. Firstly, we study the asymptotic properties of U-statistic of Markov chains with possibly varying U-statistics kernels. We apply these results to derive the "small bandwidth asymptotics" of a class of kernel-based semiparametric estimators of density-weighted averages. In a second application, we employ the same technique to examine the classical and "fixed-b" asymptotic behavior of lag-window estimators for long-run variance estimation, and apply this result to Markov Chain Monte Carlo.

1. Introduction

In this paper we deal with quadratic forms given by

$$U_n(h_n) = \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell,j)h_n(X_\ell,X_j), \qquad n \geq 1,$$

for a stochastic process $\{X_n,\ n \geq 0\}$, weight matrices $w_n : \mathbb{N}\times\mathbb{N} \to \mathbb{R}$ and symmetric kernels $h_n : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$. Quadratic forms of possibly time-dependent random variables naturally arise in a variety of statistical and econometric problems, and their large sample properties are of particular importance to develop asymptotically valid inference procedures.

For an independent sequence $\{X_n,\ n \geq 0\}$, the well known Hoeffding decomposition provides a useful approach to studying the asymptotic properties of $U_n(h_n)$ because it decomposes the statistic into two (uncorrelated) martingale sequences, which are then easily handled by standard martingale theory. See, e.g., Serfling (1980) for a review. This approach, for example, is particularly useful when the kernel h_n varies with the sample size, since in this case both terms in the Hoeffding decomposition contribute in the limit

2000 *Mathematics Subject Classification.* 60J10, 62M10.

Key words and phrases. Averaged derivatives, Central limit theorems, Markov Chains, Markov Chain Monte Carlo, Martingale approximations, Quadratic forms, Small bandwidth asymptotics, U-statistics.

Y. F. Atchadé: University of Michigan, Department of Statistics, Ann Arbor, 48109, MI, United States. *E-mail address:* yvesa@umich.edu.

Matias D. Cattaneo: University of Michigan, Department of Economics, Ann Arbor, 48109, MI, United States. *E-mail address:* cattaneo@umich.edu.

in general (see, e.g., Jammalamadaka and Janson (1986) and de Jong (1987)). When the process $\{X_n,\ n \geq 0\}$ is time-dependent, however, the classical Hoeffding decomposition is not useful because the resulting representation does not have the desirable martingale property in general. As a consequence, the large sample properties of quadratic forms of time-dependent random variables are typically established in a less systematic way. The most well understood case is the case of a standard U-statistics where h_n does not depend on n and $w_n(\ell, j) = 1$ if $\ell \neq j$ and 0 otherwise (Yoshihara (1976); Eagleson (1979); Dehling and Wendler (2010)). Hsing and Wu (2004) considers $U_n(h)$ where neither h_n nor w_n depends on n, whereas Wu and Shao (2007) studies $U_n(h_n)$ when $h_n(x, y) = h(x, y) = xy$ for a martingale-difference sequence (see also Bhansali et al. (2007) for i.i.d. sequences).

This paper develops a general martingale approximation for quadratic forms of Markov chains that allows for a unified asymptotic treatment of these quadratic statistics via martingale limit theory. This martingale approximation has interesting implications for statistics and econometrics because it provides a simple and systematic way of deriving the asymptotic properties of statistics that may be represented as a quadratic form of Markov chains.

To illustrate the broad applicability of this result, we study the large sample properties of two seemingly unrelated statistical problems: kernel-based semiparametric inference for density-weighted averages and lag-window long-run variance estimation. The martingale approximation developed in this paper not only leads to new results for these important statistical problems, but also shows that the two problems share a common structure in general.

The first main application discussed here is related to robust statistical inference in the context of semiparametric models. It is well documented that asymptotic distributional approximations for semiparametric estimators (and similar statistics) tend to be very sensitive to the user-defined choices involved in their construction (see, e.g., Cattaneo et al. (2010)). Non-standard asymptotics usually provide more accurate distributional approximations in finite samples because they capture certain terms that are assumed away by the classical large sample theory routinely employed in the literature (see, e.g., Cattaneo et al. (2011c) and references therein for further discussion). One example of non-standard asymptotic distributional approximations for semiparametric estimators is the "small bandwidth asymptotics" of Cattaneo et al. (2011b). This alternative asymptotic theory, which is developed for a kernel-based semiparametric estimator of density-weighted average derivatives under i.i.d. data, leads to a more general distributional approximation, relaxes side conditions (e.g., restrictions on the kernel and bandwidth) and also removes strong untestable assumptions (e.g., smoothness of the underlying infinite dimensional parameter; c.f., Hristache et al. (2001) and references therein).

In this paper we employ the "small bandwidth asymptotics" framework to analyze kernel-based estimators of semiparametric density-weighted averages for stationary Markov chains. The results obtained here generalize those available for i.i.d. sequences, and also cover new problems of interest in statistics and econometrics, such as those discussed in Robinson (1989), Stoker (1989), Newey et al. (2004, Section 2) and references therein. To develop these results, we first study the asymptotic properties of U-statistics of Markov chains with varying U-statistic kernels using the martingale approximation introduced in this paper. We then apply these general results, together with appropriate regularity conditions, to establish the "small bandwidth asymptotics" of the kernel-based semiparametric estimators. We also report the main findings of a small-scale simulation study to show how this alternative asymptotic approximation leads to confidence intervals that are less sensitive in terms of empirical coverage rates to perturbations in the choice of bandwidth.

The second substantive statistical application considered in this paper involves studying the asymptotic properties of a well known class of lag-window long-run variance estimators (see, e.g., Priestley (1981) for detailed discussion and early references). These estimators also depend on the choice of a weighting function (or kernel) and tuning parameter (or bandwidth), and are typically employed for studentization purposes. It is also well known that these estimators are very sensitive to perturbations of the user-chosen parameters, a fact that has motivated some researchers to consider an alternative, non-standard asymptotic framework for the analysis of its large sample properties. In particular, the "fixed-b asymptotics" (Kiefer and Vogelsang (2005)) is a popular approach in the econometrics literature which makes the tuning parameter sequence entering the long-run variance estimators proportional to the sample size. This alternative asymptotic analysis implies that the long-run variance estimator is inconsistent, and converges weakly to a functional of Brownian motion, and thus leads to an alternative distributional approximation of commonly used studentized statistics.

In this paper we exploit the new decomposition to derive the classical and "fixed-b" asymptotics of lag-window long-run variance estimators of possibly non-stationary Markov chains. Our results have important implications for Markov Chain Monte Carlo (MCMC) simulations, offering in particular new robust procedures for constructing confidence intervals. Our results also substantially improve on the existing literature on the convergence of lag-windows estimators for Markov chains (Flegal and Jones (2010), Atchade (2010)). The findings from a small simulation study show that the robust confidence intervals derived under the "fixed-b asymptotic" framework have better empirical coverage properties than their counterparts derived under classical approximations in samples of moderate size.

The rest of the paper is organized as follows. The rest of the introduction outlines the general setup and introduces the main notation employed throughout, while Section 2 derives the main martingale approximation method. Section 3 studies the large sample properties of U-statistics with possibly varying U-statistic kernels and, in particular, the asymptotic distribution of kernel-based semiparametric density-weighted averages estimators. Section 4 derives the large sample properties of lag-window long-run variance estimators under quite general conditions and, in particular, applies these results to MCMC simulation. Section 5 discusses the connection between the two examples, and concludes. All the proofs are presented in Section 6.

1.1. **Setup and Notation.** Throughout the paper, $\mathcal{X}$ denotes a Polish space equipped with its Borel σ-algebra $\mathcal{B}$ and $\{X_n,\ n \geq 0\}$ is a $\mathcal{X}$-valued Markov chain defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n,\ n \geq 0\})$. We denote P the transition kernel of the Markov chain and μ its invariant distribution, whose existence is assumed. We do not necessarily assume that the Markov chain is stationary and we denote ρ its initial distribution. Let $\mathbb{N}$ be the set of nonnegative integers. For $n \geq 1$, $w_n : \ \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ denotes a weight function. If $h : \ \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a symmetric measurable function and $n \geq 1$ an integer, we define the quadratic form $U_n(h)$ as

$$U_n(h) := \sum_{\ell=1}^{n} \sum_{j=1}^{\ell} w_n(\ell, j) h(X_\ell, X_j), \qquad n \geq 1. \tag{1}$$

We study the limiting behavior of the quadratic form $U_n(h_n)$ for some sequence of symmetric measurable functions $\{h_n : \ \mathcal{X} \times \mathcal{X} \to \mathbb{R},\ n \geq 1\}$.

We will rely on the following set of notation. Let $(\mathbb{T}, \mathcal{A})$ be an arbitrary measure space. If $W : \mathbb{T} \to [1, +\infty)$ is a function, the W-norm of a function $f : \mathbb{T} \to \mathbb{R}$ is defined as $|f|_W := \sup_{x \in \mathbb{T}} |f(x)|/W(x)$. The set of measurable functions $f : \ \mathbb{T} \to \mathbb{R}$ with finite W-norm is denoted by $\mathcal{L}_W(\mathbb{T})$ or simply $\mathcal{L}_W$ when there is no ambiguity on the space $\mathbb{T}$. For a finite real-valued signed measure ν on $\mathbb{T}$, we denote the W-norm of ν as

$$\|\nu\|_W := \int W(x)|\nu|(dx) = \sup_{|f|_W \leq 1} \left| \int f(x)\nu(dx) \right|,$$

where $|\nu|$ is the total variation measure of ν. We denote $\mathcal{M}_W(\mathbb{T})$ the space of all finite real-valued signed measures ν on $\mathbb{T}$ such that $\|\nu\|_W < \infty$. It is well-known that $(\mathcal{M}_W(\mathbb{T}), \|\cdot\|_W)$ is a Banach space. When the measure space $\mathbb{T}$ is understood, we simply write $\mathcal{M}_W$. We will use the notation $\nu(f)$ to denote the integral $\int f(x)\nu(dx)$. If ν_1, ν_2 are two finite signed measures on $(\mathbb{T}, \mathcal{A})$, we define the product signed measure $\nu_1 \bigotimes \nu_2 = \nu_1\nu_2$ on $(\mathbb{T}^2, \mathcal{A}^2)$ as $\nu_1\nu_2(A) = \int \nu_1(dx)\nu_2(A_x)$, where $A_x = \{y \in \mathbb{T} : \ (x, y) \in A\}$. It is obvious that if $|\nu_1|(\mathbb{T}) < \infty$ and $|\nu_2|(\mathbb{T}) < \infty$, then $\nu_1\nu_2$ is a well-defined finite signed measure and $|\nu_1\nu_2| = |\nu_1| \bigotimes |\nu_2|$.

If Q is a transition kernel on $(\mathbb{T}, \mathcal{A})$, and $h : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ a bivariate function then Qh is the bivariate function defined by the rule $Qh(x, y) = \int Q(x, dz)h(z, y)$. If $h : \mathbb{T} \to \mathbb{R}$ is univariate, Qh is defined similarly as $Qh(x) = \int Q(x, dz)h(z)$. Fix Q a Markov kernel, and $V : \mathbb{T} \times \mathbb{T} \to [1, \infty)$. For $p \geq 1$ and a function $h : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$, we define

$$\|h\|_{p,V} := \sup_{x,y \in \mathbb{T}} \frac{\left(\int Q(x, dz)|h(z, y)|^p\right)^{1/p}}{V(x, y)}.$$

For a univariate function $V : \mathbb{T} \to [1, \infty)$ and for $h : \mathcal{X} \to \mathbb{R}$, we define $\|h\|_{p,V}$ similarly as

$$\|h\|_{p,V} := \sup_{x \in \mathbb{T}} V(x)^{-1} \left(\int Q(x, dz)|h(z)|^p\right)^{1/p}.$$

When we use the notation $\|h\|_{p,V}$ below, it will always be with respect to P, the Markov kernel of the reference process $\{X_n,\ n \geq 0\}$, unless stated otherwise.

Definition 1.1. *We say that the transition kernel Q with invariant distribution μ has a short-range dependence if there exist measurable functions $V \leq W : \mathbb{T} \to [1, \infty)$ such that*

$$\sum_{n \geq 0} \|Q^n(x, \cdot) - \mu\|_V \leq cW(x), \quad x \in \mathcal{X}, \tag{2}$$

for some finite constant c. In that case we say that Q satisfies the condition $\mathsf{C}(V, W)$.

Clearly, if Q satisfies $\mathsf{C}(V, W)$ and $\mu(VW) < \infty$ then for any $f \in \mathcal{L}_V$, we have $\sum_{n \geq 0} |\mathsf{Cov}_\mu(f(X_0), f(X_n))| \leq c|f|_V^2 \mu(VW) < \infty$, which explains the terminology.

Throughout the paper, we denote by c a finite constant which depends solely on the kernel P but whose actual value can change from one equation to the next. In particular c does not depend on the family of function $\{h_n,\ n \geq 1\}$ considered. Finally, all limits are taken as $n \to \infty$ unless explicitly noted otherwise.

2. A general decomposition for quadratic forms

For $n, m \in \mathbb{N}$, $x, y \in \mathcal{X}$, we introduce on $\mathcal{X} \times \mathcal{X}$ the finite signed measure

$$\pi_{n,m}(x, y; (du, dv)) := (P^n(x, du) - \mu(du)) (P^m(y, dv) - \mu(dv)).$$

Consider the following assumption.

Assumption A1 There exist measurable functions $V_1 \leq W_1 : \mathcal{X} \to [1, \infty)$, symmetric measurable functions $\overline{V} \leq \overline{W} : \mathcal{X} \times \mathcal{X} \to [1, \infty)$, and a finite constant c such that P satisfies $\mathsf{C}(V_1, W_1)$ and for all $x, y \in \mathcal{X}$, $P\overline{W}(x, y) < \infty$ and

$$\sum_{n \geq 0} \sum_{m \geq 0} \|\pi_{n,m}(x, y; \cdot)\|_{\overline{V}} \leq c\overline{W}(x, y). \tag{3}$$

It is always possible to deduce A1 from a univariate short-range dependence assumption $\mathsf{C}(V_1, W_1)$. Indeed, if P satisfies $\mathsf{C}(V_1, W_1)$ and $\mathsf{C}(V_2, W_2)$, and $PW_1 < \infty$, $PW_2 < \infty$, define $\overline{V}(x,y) = V_1(x)V_2(y)$ and $\overline{W}(x,y) = W_1(x)W_2(y)$. Then $\|\pi_{n,m}(x,y;\cdot)\|_{\overline{V}} = \|P^n(x,\cdot) - \mu\|_{V_1}\|P^m(y,\cdot) - \mu\|_{V_2}$ and therefore A1 holds. But the choice $\overline{V}(x,y) = V_1(x)V_2(y)$, $\overline{W}(x,y) = W_1(x)W_2(y)$ is just one possibility among many others.

Remark 1. The condition $\mathsf{C}(V, W)$ holds for geometrically ergodic Markov kernels (that is, kernels P for which $\|P^n(x,\cdot) - \mu\|_V$ converges to zero exponentially fast for some $V \geq 1$). It also holds for sub-geometrically ergodic Markov kernels ($\|P^n(x,\cdot) - \mu\|_V$ converges to zero sub-geometrically) for which the rate of convergence is summable. It is sometimes possible to check the condition $\mathsf{C}(V, W)$ using Lyapunov drift conditions and their extensions and this has been done for several time series Markov models (Douc et al. (2004); Meitz and Saikkonen (2008); Meyn and Tweedie (2009)). See also Chen and Shen (1998) for examples and application in semiparametric inference.

We show that whenever A1 holds, there exists a martingale approximation to $U_n(h_n)$ that offers a simple route to study the asymptotics of $U_n(h_n)$. This is achieved by introducing a bivariate extension of the well-known resolvent kernel associated to P.

2.1. **The function G_h.** The space $\mathcal{M}_{\overline{V}}(\mathcal{X} \times \mathcal{X})$ of all finite signed measure on $\mathcal{X} \times \mathcal{X}$ with finite $\|\cdot\|_{\overline{V}}$ norm, equipped with the norm $\|\cdot\|_{\overline{V}}$ is a Banach space. Under A1 and for any $x, y \in \mathcal{X}$, the family $\{\pi_{n,m}(x,y;\cdot),\ (n,m) \in \mathbb{N} \times \mathbb{N}\}$ is absolutely summable in $\mathcal{M}_{\overline{V}}(\mathcal{X} \times \mathcal{X})$ and its sum

$$R(x,y;(du,dv)) := \sum_{n\geq 0}\sum_{m\geq 0} \pi_{n,m}(x,y;(du,dv)),$$

is a finite signed measure that belongs to $\mathcal{M}_{\overline{V}}(\mathcal{X} \times \mathcal{X})$ and satisfies

$$\|R(x,y;\cdot)\|_{\overline{V}} \leq c\overline{W}(x,y). \tag{4}$$

Let $h:\ \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a measurable function. We define $Ph(x,y) := \int P(x,dz)h(z,y)$, $\overline{Ph}(x,y) := \int P(y,dz)h(x,z)$, $P^2h(x,y) := P\{\overline{Ph}\}(x,y) = \overline{P\{Ph\}}$,

$$G_h(x,y) := \int R(x,y;(du,dv))h(u,v), \text{ and } \bar{G}_h(x,y) := G_h(x,y) - G_{\overline{Ph}}(x,y),\ x,y \in \mathcal{X}.$$

Notice that if h is symmetric, $\overline{Ph}(x,y) = Ph(y,x)$. Since $\overline{V}$ is symmetric, this implies that if $Ph \in \mathcal{L}_{\overline{V}}$, then so is $\overline{Ph}$. We say that h is degenerate if $\int h(x,y)\mu(dy) = \int h(y,x)\mu(dx) = 0$ for all $x, y \in \mathcal{X}$. The next lemma establishes some basic properties of the functions G_h and $\bar{G}_h$.

Lemma 2.1. *Assume A1. Let $h \in \mathcal{L}_{\overline{V}}$, and h degenerate. Then G_h is well-defined, $G_h \in \mathcal{L}_{\overline{W}}$, and $|G_h|_{\overline{W}} \leq c|h|_{\overline{V}}$. Furthermore if $Ph \in \mathcal{L}_{\overline{V}}$ then*

$$\int P(x,dz)G_h(z,y) = G_{\{Ph\}}(x,y),\ x,y \in \mathcal{X}. \tag{5}$$

If $Ph, \overline{Ph}, P^2h \in \mathcal{L}_{\overline{V}}$, then we have

$$h(x,y) = \bar{G}_h(x,y) - \int P(x,dz)\bar{G}_h(z,y),\ \ x,y \in \mathcal{X}. \tag{6}$$

Proof. See Section 6.1. □

Remark 2. Equation (6) gives a bivariate extension to the well known Poisson's equation (Gordin (1969)) and provides a simple approach for approximating $U_n(h)$ by a martingale, as we shall see shortly. We will also need the univariate Poissson's equation below in (9).

Remark 3. The summability assumption (3) in A1 can be weakened. Maxwell and Woodroofe (2000) constructed a martingale approximation for linear partial sums of Markov chains using an approximate Poisson's equation solution. In principle, their idea can be lifted to the bivariate case. For $\epsilon > 0$, consider the Abel sum

$$R_\epsilon(x,y;(du,dv)) = \sum_{n\geq 0}\sum_{m\geq 0}\left(\frac{1}{1+\epsilon}\right)^{n+m+2}\left(P^n(x,du)-\mu(du)\right)\left(P^m(y,dv)-\mu(dv)\right).$$

Because of the geometric weight $(1+\epsilon)^{-n-m-2}$, for $\epsilon > 0$, $R_\epsilon(x,y;\cdot)$ is well defined under weaker conditions than A1. Define $G_h^{(\epsilon)}(x,y) := \int R_\epsilon(x,y;(du,dv))h(u,v)$, $\bar{G}_h^{(\epsilon)}(x,y) := G_h^{(\epsilon)}(x,y) - G_{\overline{Ph}}^{(\epsilon)}(x,y)$. Then, formally we have

$$(1+\epsilon)^2 h(x,y) = \bar{G}_h^{(\epsilon)}(x,y) - \int P(x,dz)\bar{G}_h^{(\epsilon)}(z,y),\quad x,y \in \mathcal{X}. \tag{7}$$

Equation (7) is an approximate bivariate Poisson's equation for h (and P) and can be used in place of the Poisson's equation (6) to derive a martingale approximation for quadratic form under weaker conditions. To limit the technical details, we do not pursue this idea any further, but we mention this as a promising avenue for future work.

2.2. **The decomposition.** For notational convenience, we shall write $\bar{\mu}$ to denote the product probability measure $\bar{\mu}(du,dv) = \mu(du)\mu(dv)$, where μ is the invariant distribution of the Markov kernel P. Let $\{h_n : \mathcal{X}\times\mathcal{X} \to \mathbb{R}\}$ be a family of symmetric measurable functions such that $\bar{\mu}(|h_n|) < \infty$. Define $\theta_n := \bar{\mu}(h_n)$, $h_{n,1}(x) := \int h_n(x,y)\mu(dy)$, $\bar{h}_{n,1}(x) := h_{n,1}(x) - \theta_n$, $\bar{h}_{n,2}(x,y) := h_n(x,y) - \bar{h}_{n,1}(x) - \bar{h}_{n,1}(y) - \theta_n$. Notice that $\bar{h}_{n,2}$ is degenerate. Assume A1 and suppose that $h_{n,1} \in \mathcal{L}_{V_1}$ and $h_{n,2}, Ph_{n,2}, P^2h_{n,2} \in \mathcal{L}_{\overline{V}}$. Under A1, $\sum_{j\geq 0}|P^j\bar{h}_{n,1}(x)| \leq c|h_{n,1}|_{V_1}W_1(x)$. Therefore $g_n(x) := \sum_{j\geq 0} P^j\bar{h}_{n,1}(x)$ is well-defined, $g_n \in \mathcal{L}_{W_1}$ and we have

$$|g_n|_{W_1} \leq c|h_{n,1}|_{V_1},\quad \text{and}\quad |Pg_n|_{W_1} \leq c|Ph_{n,1}|_{V_1}, \tag{8}$$

where for the second inequality, we assume that $Ph_{n,1} \in \mathcal{L}_{V_1}$. It is also well-known that the function g_n satisfies the Poisson's equation for $h_{n,1}$ and P:

$$\bar{h}_{n,1}(x) = g_n(x) - Pg_n(x), \quad x \in \mathcal{X}. \tag{9}$$

We introduce the functions

$$L_n(x,y) := g_n(x) - Pg_n(y), \quad x, y \in \mathcal{X}, \quad \text{and}$$

$$Q_n(x,y,u,v) := G_{\bar{h}_{n,2}}(x,u) - G_{P\bar{h}_{n,2}}(y,u) - G_{\overline{P\bar{h}_{n,2}}}(x,v) + G_{P^2\bar{h}_{n,2}}(y,v), \quad x,y,u,v \in \mathcal{X}.$$

For $1 \leq \ell \leq n$, $1 \leq j < \ell$, we define the random variables

$$L_{n,\ell} := L_n(X_\ell, X_{\ell-1}), \quad \text{and} \quad Q_{n,\ell,j} := Q_n(X_\ell, X_{\ell-1}, X_j, X_{j-1}).$$

We have $\mathbb{E}\left(L_{n,\ell}|\mathcal{F}_{\ell-1}\right) = \int P(X_{\ell-1}, dz) g_n(z) - Pg_n(X_{\ell-1}) = 0$ almost surely, showing that $\{(L_{n,\ell},\ \mathcal{F}_\ell),\ 1 \leq \ell \leq n\}$ is a martingale-difference array. By combining (9) and (8), we obtain

$$|L_n(x,y)| \leq \left|\bar{h}_{n,1}(x)\right| + c|Ph_{n,1}|_{V_1}\left(W_1(x) + W_1(y)\right), \quad x, y \in \mathcal{X}. \tag{10}$$

For $1 \leq j < \ell$, using (5), we have

$$\begin{aligned}\mathbb{E}\left(Q_{n,\ell,j}|\mathcal{F}_{\ell-1}\right) = \int P(X_{\ell-1}, dz) G_{\bar{h}_{n,2}}(z, X_j) - G_{P\bar{h}_{n,2}}(X_{\ell-1}, X_j)\\ - \int P(X_{\ell-1}, dz) G_{\overline{P\bar{h}_{n,2}}}(z, X_{j-1}) + G_{P^2\bar{h}_{n,2}}(X_{\ell-1}, X_{j-1}) = 0,\end{aligned}$$

which shows that the process $\{(\sum_{j=1}^{\ell-1} Q_{n,\ell,j}, \mathcal{F}_\ell),\ 2 \leq \ell \leq n\}$ is a martingale-difference array. Notice also that from (6),

$$\begin{aligned}Q_n(x_\ell, x_{\ell-1}, x_j, x_{j-1}) = \bar{h}_{n,2}(x_\ell, x_j) + G_{P\bar{h}_{n,2}}(x_\ell, x_j) - G_{P\bar{h}_{n,2}}(x_{\ell-1}, x_j)\\ + G_{\overline{P\bar{h}_{n,2}}}(x_\ell, x_j) - G_{\overline{P\bar{h}_{n,2}}}(x_\ell, x_{j-1}) - G_{P^2\bar{h}_{n,2}}(x_\ell, x_j) + G_{P^2\bar{h}_{n,2}}(x_{\ell-1}, x_{j-1}).\end{aligned}$$

We deduce that

$$\begin{aligned}|Q_n(x_\ell, x_{\ell-1}, x_j, x_{j-1})| \leq \left|\bar{h}_{n,2}(x_\ell, x_j)\right| + c\left(|P\bar{h}_{n,2}|_{\overline{V}} + |P^2\bar{h}_{n,2}|_{\overline{V}}\right)\\ \times \left(\overline{W}(x_\ell, x_j) + \overline{W}(x_{\ell-1}, x_j) + \overline{W}(x_\ell, x_{j-1}) + \overline{W}(x_{\ell-1}, x_{j-1})\right),\end{aligned} \tag{11}$$

for some finite constant c. We will use (10) and (11) to derive bounds on the p-th moment of $|L_{n,\ell}|$ and $|Q_{n,\ell,j}|$, respectively. We also introduce the following sequences

$$\gamma_{n,\ell} := \sum_{j=1}^{\ell} w_n(\ell, j) + \sum_{j=\ell}^{n} w_n(j, \ell), \quad \varpi_n^{(1)}(\ell, j) := w_n(\ell, j) - w_n(\ell - 1, j),$$

$$\varpi_n^{(2)}(\ell, j) := w_n(\ell, j) - w_n(\ell, j-1),$$

$$\text{and} \quad \varpi_n^{(3)}(\ell, j) := w_n(\ell, j) + w_n(\ell - 1, j - 1) - w_n(\ell, j - 1) - w_n(\ell - 1, j).$$

Lemma 2.2. *Assume A1. Let $\{h_n,\ n \geq 1\}$ be a family of symmetric measurable function $h_n : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that for each $n \geq 1$, $\bar{\mu}(|h_n|) < \infty$, $h_{n,1} \in \mathcal{L}_{V_1}$, and $h_{n,2}, Ph_{n,2}, P^2 h_{n,2} \in \mathcal{L}_{\overline{V}}$. Then*

$$U_n(h_n) = \theta_n \sum_{\ell=1}^{n} \sum_{j=1}^{\ell} w_n(\ell, j) + \sum_{\ell=1}^{n} \gamma_{n,\ell} L_{n,\ell} + \sum_{\ell=1}^{n} \sum_{j=1}^{\ell} w_n(\ell, j) Q_{n,\ell,j} + \zeta_n^{(1)} + \zeta_n^{(2)}, \quad (12)$$

where

$$\zeta_n^{(1)} = \gamma_{n,0} P g_n(X_0) - \gamma_{n,n} P g_n(X_n) + \sum_{\ell=1}^{n} (\gamma_{n,\ell} - \gamma_{n,\ell-1}) P g_n(X_{\ell-1}),$$

and

$$\begin{aligned}
\zeta_n^{(2)} = &\sum_{\ell=1}^{n} \sum_{j=1}^{\ell} \varpi_n^{(1)}(\ell, j) \left(G_{P\bar{h}_{n,2}}(X_{\ell-1}, X_j) - G_{P^2\bar{h}_{n,2}}(X_{\ell-1}, X_{j-1}) \right) \\
&+ \sum_{\ell=1}^{n} \sum_{j=1}^{\ell} \varpi_n^{(2)}(\ell, j) \left(G_{\overline{P\bar{h}}_{n,2}}(X_\ell, X_{j-1}) - G_{P^2\bar{h}_{n,2}}(X_{\ell-1}, X_{j-1}) \right) \\
&+ \sum_{\ell=1}^{n} \sum_{j=1}^{\ell} \varpi_n^{(3)}(\ell, j) G_{P^2\bar{h}_{n,2}}(X_{\ell-1}, X_{j-1}) + \sum_{\ell=1}^{n} w_n(\ell, 0) G_{\overline{P\bar{h}}_{n,2}}(X_\ell, X_0) \\
&- w_n(0,0) G_{P^2\bar{h}_{n,2}}(X_0, X_0) + \sum_{\ell=1}^{n} w_n(n, \ell) \left(G_{P^2\bar{h}_{n,2}}(X_n, X_\ell) - G_{P\bar{h}_{n,2}}(X_n, X_\ell) \right) \\
&+ \sum_{\ell=1}^{n} w_n(\ell-1, \ell) G_{P\bar{h}_{n,2}}(X_{\ell-1}, X_\ell) - \sum_{\ell=1}^{n} w_n(\ell, \ell) G_{\overline{P\bar{h}}_{n,2}}(X_\ell, X_\ell).
\end{aligned}$$

Proof. See Section 6.2. □

Remark 4. The usefulness of this decomposition comes from the fact that the remainders $\zeta_n^{(1)}$ and $\zeta_n^{(2)}$ involve either single summations or difference sequences of the weights w_n. As a result (and we will see this on the examples considered below), these remainders are typically negligible compared to the other terms in the decomposition. It does not seem possible to establish the negligibility of these reminders in complete generality. For example, consider the case where the quadratic form kernel h_n is multiplicative: $h_n(x, y) = h(x)h(y)$. Then the first term $\sum_{\ell=1}^{n} \sum_{j=1}^{\ell} \varpi_n^{(1)}(\ell, j) \left(G_{P\bar{h}_{n,2}}(X_{\ell-1}, X_j) - G_{P^2\bar{h}_{n,2}}(X_{\ell-1}, X_{j-1}) \right)$ in the remain $\zeta_n^{(2)}$ becomes the partial sum of a martingale array, a property that can be used to obtain a sharper bound on its p-moments (see Section 4). But it is clear that this martingale property does not hold in general.

3. Asymptotic Properties of U-statistics

We consider the special case where $w_n(\ell,\ell) = 0$ and $w_n(\ell,k) = 1$ for $\ell \neq k$. $U_n(h_n)$ is then a standard bivariate U-statistic with varying kernels

$$U_n(h_n) = \sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1} h_n(X_\ell, X_j).$$

We study the law of large numbers and the central limit theorem for $U_n(h_n)$. For U-statistics with a fix kernel ($h_n = h$), one typically distinguishes the cases h degenerate versus h non-degenerate. When h is non-degenerate, $U_n(h)$ is driven by the linear term $\sum_{\ell=1}^{n} L_\ell$, while when h is degenerate the U-statistic is characterized by the quadratic term $\sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1} Q_{n,\ell,j}$.

This dichotomy is less relevant for varying kernels U-statistics, since in this case either term (or both) may be asymptotically leading. The asymptotic behavior of the varying U-statistic will crucially depend on the rate of convergence of the linear and quadratic terms obtained from the martingale approximation derived in this paper. Specifically, Theorem 3.2-(1) and Theorem 3.3 below are concerned with the case when the quadratic term is asymptotically negligible relative to the linear term. In contrast, Theorem 3.2-(2) and Theorem 3.4 consider the case when the linear part is at most of the same order than the quadratic part in the decomposition of $U_n(h_n)$. This generality in the asymptotic analysis is important (partially) because it allows for non-standard asymptotics in the context of time series semiparametric inference (and related problems), as discussed in Section 3.1.

We work under the following assumption.

Assumption A2 There exist a measurable function $V : \mathcal{X} \to [1,\infty)$, $\kappa \in [0,1)$ such that $\mu(V) < \infty$, $PV(x) < \infty$ for all $x \in \mathcal{X}$, and for any $\beta \in [0, 1-\kappa]$, we have $\sup_{x\in\mathcal{X}} PV^\beta(x)/V^\beta(x) < \infty$ and there exist a sequence of nonnegative numbers $\{\rho_\beta(n),\ n \geq 0\}$ with $\sum_{n\geq 0} \rho_\beta(n) < \infty$ such that

$$\|P^n(x,\cdot) - \mu\|_{V^\beta} \leq \rho_\beta(n)\ V^{\beta+\kappa}(x), \quad n \geq 0. \tag{13}$$

Furthermore, we assume that

$$\sup_{n\geq 0} \mathbb{E}\left(V(X_n)\right) < \infty. \tag{14}$$

Remark 5. It is sometimes possible to verify A2 by establishing that the Markov kernel P satisfies an appropriate drift and minorization conditions (see e.g. Douc et al. (2004); Meyn and Tweedie (2009)).

We introduce $V_{(2)}^\beta(x,y) = V^\beta(x)V^\beta(y)$, $W_{(2)}^\beta(x,y) = V^{\beta+\kappa}(x)V^{\beta+\kappa}(y)$. It is straightforward to check that A2 implies A1 with $V_1 = V^\beta$, $W_1 = V^{\beta+\kappa}$, $\overline{V} = V_{(2)}^\beta$, $\overline{W} = W_{(2)}^\beta$, for any $\beta \in [0, 1-\kappa)$. It is also easy to check that for $h_n \in \mathcal{L}_{V_{(2)}^\beta}$, with $\beta \in [0, 1-\kappa]$, $h_{n,1} \in \mathcal{L}_{V^\beta}$ and $h_{n,2}, Ph_{n,2}, P^2 h_{n2} \in \mathcal{L}_{V_{(2)}^\beta}$ and $|P\bar{h}_{n,2}|_{V_{(2)}^\beta} + |P^2\bar{h}_{n,2}|_{V_{(2)}^\beta} \le c|Ph_n|_{V_{(2)}^\beta}$, for some finite constant c (that depends only on P and V).

Now, by Lemma 2.2 we have

$$U_n(h_n) - \binom{n}{2}\theta_n \ = (n-1)\sum_{\ell=1}^{n} L_{n,\ell} + \sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1} Q_{n,\ell,j} + \zeta_n. \tag{15}$$

It is clear that for $\ell, j \ge 1$ and $|\ell - j| \ge 2$, $\varpi_n^{(1)}(\ell, j) = \varpi_n^{(2)}(\ell, j) = \varpi_n^{(3)}(\ell, j) = 0$. This implies that the double sums in the term $\zeta_n^{(2)}$ of Lemma 2.2 reduces to single sums. Using Lemma 2.1 and (14), the following result on the negligibility of the remainder follows easily.

Proposition 3.1. *Assume A2 and suppose that $h_n \in \mathcal{L}_{V_{(2)}^\beta}$ for each n and for some $\beta \in [0, 1-\kappa]$. For any $p \ge 1$ such that $2p(\beta+\kappa) \le 1$, there exists a finite constant c such that for all $n \ge 1$,*

$$\mathbb{E}^{1/p}\left(|\zeta_n|^p\right) \le cn\left(|Ph_{n,1}|_{V^\beta} + |Ph_n|_{V_{(2)}^\beta}\right). \tag{16}$$

We now turn to the limiting behavior of $U_n(h_n) - \binom{n}{2}\theta_n$, starting with a law of large numbers.

Theorem 3.2. *Assume A2 and suppose that $h_n \in \mathcal{L}_{V_{(2)}^\beta}$ for each n and for some $\beta \in [0, 1-\kappa]$. For $\lambda > 1$, $n^{-\lambda}\left(U_n(h_n) - \binom{n}{2}\theta_n\right)$ converges to zero in probability under any one of the following two conditions.*

(1) There exists $p > 1$ such that $2p(\beta+\kappa) \le 1$ and

$$n^{-(\lambda-1)+\frac{1}{2}\vee\frac{1}{p}}\left(\|h_{n,1}\|_{p,V^\beta} + \|h_n\|_{p,V_{(2)}^\beta}\right) \to 0. \tag{17}$$

(2) $4(\beta+\kappa) + 2\kappa \le 1$ and

$$n^{-\lambda+1}\left(\sqrt{n}\,\|h_{n,1}\|_{2,V^\beta} + \|h_n\|_{2,V_{(2)}^\beta}\right) \to 0. \tag{18}$$

Proof. See Section 6.3. □

We now consider the rate of convergence of $\binom{n}{2}^{-1}\left(U_n(h_n) - \binom{n}{2}\theta_n\right)$. We focus on normal central limits. We introduce the product measure $\{\mu P\}(dx, dy) = \mu(dx)P(x, dy)$ and the variance term

$$\sigma_{n,1}^2 \ = \ \int \{\mu P\}(dx, dy) L_n^2(x, y) \ = \ \mathsf{Var}_\mu\left(h_{n,1}(X_0)\right) + 2\sum_{\ell\ge 1}\mathsf{Cov}_\mu\left(h_{n,1}(X_0), h_{n,1}(X_\ell)\right).$$

Theorem 3.3. *Assume A2 and suppose that $h_n \in \mathcal{L}_{V^\beta_{(2)}}$ for each n and for $\beta > 0$ such that $4(\beta+\kappa)+2\kappa \le 1$. Suppose also that $\sigma_{n,1} > 0$ and there exists $p > 1$ such that*

$$\|h_{n,1}\|_{2,V^\beta} = O(\sigma_{n,1}),$$
$$n^{-1+\frac{1}{p}}\sigma_{n,1}^{-2}\|h_{n,1}\|^2_{2p,V^\beta} \to 0 \text{ and } n^{-1}\sigma_{n,1}^{-2}\|h_n\|^2_{2,V^\beta_{(2)}} \to 0. \quad (19)$$

Then $\sigma_{n,1}^{-1}n^{1/2}\binom{n}{2}^{-1}\left(U_n(h_n)-\binom{n}{2}\theta_n\right)$ converges weakly to $\mathcal{N}(0,4)$.

Proof. See Section 6.4. □

The key assumption $\|h_n\|_{2,V^\beta_{(2)}} = o(\sqrt{n}\sigma_{n,1})$, which makes the quadratic term asymptotically negligible, is restrictive in general. To remove that assumption, we need to account for the term $\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} Q_{n,\ell,j}$ in the limit. We do this at the expense of higher moment assumptions. We define

$$\sigma_{n,2}^2 := \int \{\mu P\}(dx,dy)\int\{\mu P\}(du,dv)\{Q_n(y,x;v,u)\}^2,$$
$$\text{and } \sigma_n^2 := 2(n-1)\sigma_{n,1}^2 + \sigma_{n,2}^2.$$

Theorem 3.4. *Assume A2 and suppose that $h_n \in \mathcal{L}_{V^\beta_{(2)}}$ for each n and for $\beta > 0$ such that $8(\beta+\kappa)+4\kappa \le 1$. Suppose also that $\sigma_{n,1} > 0$, $\sigma_{n,2} > 0$,*

$$\sigma_{n,1}^{-1}\|h_{n,1}\|_{2,V^\beta} + \sigma_{n,2}^{-1}\|h_n\|_{2,V^\beta_{(2)}} + \sigma_{n,2}^{-2}|\delta_n|_{V^{2\beta}} = O(1), \quad (20)$$

and

$$\sigma_n^{-2}|Ph_n|^2_{V^\beta_{(2)}} + n^{-1/2}\sigma_{n,1}^{-2}\|h_{n,1}\|^2_{4,V^\beta} + n^{-1}\sigma_{n,2}^{-2}\|h_n\|^2_{4,V^\beta_{(2)}} \to 0, \quad (21)$$

where $\delta_n(x) := \int \mu(dz)h_n^2(x,z)$. Then $\sigma_n^{-1}\binom{n}{2}^{-1/2}\left(U_n(h_n)-\binom{n}{2}\theta_n\right)$ converges weakly to $\mathcal{N}(0,1)$.

Proof. See Section 6.5. □

3.1. **Application to Semiparametric Density-weighted Averages.** We illustrate the applicability of the results obtained in the previous section by deriving the asymptotic distribution of a class of kernel-based semiparametric estimators of density-weighted averages employing the "small bandwidth asymptotics" regime. The new results obtained herein generalize those available in the literature by covering new estimands and allowing for time series data.

Consider the class of estimands introduced in Newey et al. (2004, Section 2) but with stationary possibly time-dependent data. To describe these estimands let $\{X_n = (Y_n, W_n, Z_n')',\ n \ge 0\}$ be a stationary Markov chain with $Y_n \in \mathcal{Y} \subset \mathbb{R}$, $W_n \in \mathcal{W} \subset \mathbb{R}$, $Z_n \in \mathcal{Z} \subset \mathbb{R}^d$, and with stationary distribution μ on $\mathcal{X} = \mathcal{Y}\times\mathcal{W}\times\mathcal{Z}$. Further assume that Z_n is continuously distributed with (marginal) distribution $\mu_z(\cdot) = \int \mu(dy, dw, \cdot)$, and (Lebesgue) stationary density $f(\cdot)$. To conserve notation, set $e_w(z) = g_w(z)f(z)$ with $g_w(Z_n) = \mathbb{E}[W_n|Z_n]$, and let $s = (s_1, \cdots, s_d)' \in \mathbb{Z}_+^d$ be a multi-index with the usual notations (e.g., $|s| = s_1 + \cdots + s_d$, $z^s = z_1^{s_1} \cdots z_d^{s_d}$, $\partial^s f(z) = \partial^{|s|} f(z)/(\partial^{s_1} z_1 \cdots \partial^{s_d} z_d)$, etc.). For fixed multi-index $s \in \mathbb{Z}_+^d$, the estimand of interest is given by

$$\theta = \mathbb{E}[\gamma(Z_n)Y_n], \qquad \gamma(z) = \partial^s e_w(z),$$

where here and in the sequel dependence on s in omitted whenever it causes no confusion.

This class of estimands covers many interesting problems in statistics and econometrics. For instance, when $Y_n = 1$ and $W_n = 1$ for all $n \geq 0$, the estimand reduces to $\theta = \mathbb{E}[f_s(Z_n)]$ with $f_s(z) = \partial^s f(z)$, which in particular includes the classical problem of estimating the integrated squared density when $|s| = 0$. A second very popular example is obtained when $|s| = 1$ and $W_n = -2$ for all $n \geq 0$, leading to $\theta = -2\mathbb{E}[f_s(Z_n)Y_n]$, which corresponds to the well known problem of estimating the density-weighted average derivative (Stoker (1986)). Robinson (1989) and Stoker (1989) offer several other applications for the case $|s| \geq 1$. Many other estimands (as defined by the choice of s, Y_n and W_n) are discussed in, for example, Powell and Stoker (1996) and Aradillas-Lopéz et al. (2007). For substantive empirical applications see, e.g., Härdle et al. (1991) and Deaton and Ng (1998).

To verify the conditions of the central limit theorems developed in the previous section we need to impose additional assumptions. To save notation, set $e_y(z) = g_y(z)f(z)$ with $g_y(Z_n) = \mathbb{E}[Y_n|Z_n]$, $e_{yw}(z) = g_{yw}(z)f(z)$ with $g_{yw}(Z_n) = \mathbb{E}[Y_nW_n|Z_n]$, $v_y(Z_n) = \mathbb{E}[Y_n^2|Z_n]$ and $v_w(Z_n) = \mathbb{E}[W_n^2|Z_n]$. In addition, define (recall $x = (y, w, z')'$)

$$4\sigma_1^2 = \mathbb{E}[\psi(X_0)^2] + 2\sum_{\ell\geq 1} \mathbb{E}[\psi(X_0)\psi(X_\ell)], \quad \psi(x) = y\partial^s e_w(z) + w(-1)^{|s|}\partial^s e_y(z) - 2\theta,$$

and

$$\sigma_2^2 = \frac{1}{2}\mathbb{E}[(v_w(Z_n)v_y(Z_n) + (-1)^{|s|}g_{yw}(Z_n)^2)f(Z_n)].$$

The following assumption gives a set of simple sufficient conditions on the stationary distribution μ. Let $\mathcal{I}_s = \{\ell \in \mathbb{Z}_+^d : \ell \leq s\}$ and $\mathcal{I}(k) = \{\ell \in \mathbb{Z}_+^d : |\ell| \leq k\}$.

Assumption S: (a) $\mathbb{E}[|Y_n|^8 + |W_n|^8] < \infty$, $0 < \sigma_1^2$ and $0 < \sigma_2^2$.

(b) $\partial^\ell e_w(z)$ and $\partial^\ell e_y(z)$ exist and are bounded for all $\ell \in \mathcal{I}_s$.

(c) For some $Q \geq 2$: $\partial^{s+\ell} e_w(z)$ exists and is bounded for all $\ell \in \mathcal{I}(Q)$.

(d) $\partial^{s+\ell} e_y(z)$, $\partial^\ell e_{yw}(z)$ and $\partial^\ell v_w(z)f(z)$ exist and are bounded for all $\ell \in \mathcal{I}(1)$.

(e) $\lim_{\|z\|\to\infty} \left[|\partial^\ell e_w(z)| + |\partial^\ell e_y(z)|\right] = 0$ for all $\ell \in \mathcal{I}_s$.

To describe the class of kernel-based semiparametric estimators considered here, we introduce the following assumption characterizing a class of kernel functions that will be used to nonparametrically estimate the unknown function $\gamma(\cdot)$.

Assumption K: (a) $K : \mathbb{R}^d \to \mathbb{R}$ is even and bounded.

(b) $\partial^\ell K(u)$ exists and is bounded for all $\ell \in \mathcal{I}_s$, and $0 < \int_{\mathbb{R}^d} (\partial^s K(u))^2 du$.

(c) For some $\varrho \geq 2$: $\int (1 + \|u\|^\varrho)|K(u)|du < \infty$, $\int (1 + \|u\|^2)|\partial^s K(u)|du < \infty$, and $\int u^\ell K(u)\mathrm{d}u = \mathbf{1}_{\{|\ell|=0\}}$ for all $\ell \in \mathcal{I}(\varrho - 1)$.

A plug-in leave-one-out kernel-based estimator of θ is given by

$$\hat{\theta}_n = \binom{n}{2}^{-1} \sum_{i=1}^n \hat{\gamma}_i(Z_i)Y_i, \qquad \hat{\gamma}_i(z) = \frac{1}{n-1} \sum_{j=1, j\neq i}^n K_{s,b_n}(z - Z_j)W_j,$$

with $K_{s,b}(u) = \partial^s K_b(u)$, $K_b(u) = K(u/b)/b$, and where b_n is a positive bandwidth sequence. It follows immediately that this class of kernel-based estimators can be recast as n-varying U-statistics with $\hat{\theta}_n = \binom{n}{2}^{-1} U_n(h_n)$ where, for any $x_1, x_2 \in \mathcal{X}$,

$$h_n(x_1, x_2) = \frac{K_{s,b_n}(z_1 - z_2)w_2 y_1 + K_{s,b_n}(z_2 - z_1)w_1 y_2}{2}.$$

Powell et al. (1989) and Robinson (1989), under the random sampling (for $|s| = 1$ and $W_n = -2$) and under absolute regularity (for $|s| \geq 1$ and W_n degenerated), respectively, stablish a "classical" central limit theorem for the kernel-based estimator $\hat{\theta}_n$. Their approach relies on the Hoeffding decomposition for i.i.d. data, which under appropriate regularity conditions leads to a simple asymptotic representation of the estimator that is then used to verify its asymptotic normality. Specifically, under Assumptions S and K (and appropriate restrictions on the data time-dependence), the classical Hoeffding decomposition gives

$$\hat{\theta}_n - \theta = \underbrace{\theta_n - \theta}_{O\left(b_n^{\varrho \wedge Q}\right)} + \underbrace{\frac{2}{n}\sum_{i=1}^n \bar{h}_{n,1}(X_i)}_{O_p\left(n^{-1/2}\right)} + \underbrace{\binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=1}^{i-1} \bar{h}_{n,2}(X_i, X_j)}_{O_p\left(n^{-1} b_n^{-d/2-|s|}\right)},$$

where the convergence rates are obtained by simple, problem-specific bounding arguments. This representation, coupled with an appropriate central limit theorem for possibly weakly dependent random variables, implies that if $nb_n^{\varrho \wedge Q} \to 0$ and $nb_n^{d+2|s|} \to \infty$ then

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{2}{\sqrt{n}} \sum_{i=1}^n \bar{h}_{n,1}(X_i) + o_p(1) \to_d \mathcal{N}(0, 4\sigma_1^2).$$

This result corresponds to the classical approach to establish a central limit theorem for a statistic, which is based on the Hajék projection of the U-statistic. Indeed, in the i.i.d. case it is easy to show that $\mathbb{E}[(\bar{h}_{n,1}(X_i) - \psi(X_i))^2] = o(1)$ thereby verifying that $\psi(x)$ is the (semiparametric efficient) influence function for θ. A crucial implication of this approach is that the resulting limiting distribution of the statistic of interest is invariant to the nonparametric procedure employed (for discussion see, e.g., Newey (1994) and references therein), which in turn results in an important lack of robustness of the distributional approximation in general.

In an attempt to increase the robustness of the asymptotic distributional approximation for $\hat{\theta}_n$, Cattaneo et al. (2011b) develop a generalized central limit theorem based also on the Hoeffding decomposition under random sampling for the special case of the density-weighted average derivatives under i.i.d. data. Specifically, employing a standard martingale CLT it is possible to show that both the (linear) Hajék projection and the (quadratic) remainder of the Hajék projection converge in distribution jointly to a Gaussian random variable under conditions that are substantially weaker than those entertrained by the classical large sample result. In particular, under the i.i.d. assumption, $|s| = 1$ and $W_n = -2$ for all $n \geq 0$, the authors show that if $\min(1, nb_n^{d+2})nb_n^{2(\varrho\wedge Q)} \to 0$ and $n^2 b_n^d \to \infty$ then

$$\frac{\hat{\theta}_n - \theta}{\sqrt{n^{-1}4\sigma_1^2 + \binom{n}{2}^{-1} b_n^{-d-2}\sigma_2^2 \int_{\mathbb{R}^d} (\partial^s K(u))^2 du}} \to_d \mathcal{N}(0,1).$$

In this section we generalize this result to cover the class of density-weighted averages introduced above under possibly time-series data. The potential (weak) dependence of the data introduces additional complications that render the classical approach based on the i.i.d. Hoeffding decomposition difficult to apply (i.e., $\hat{\theta}_n - \theta_n$ is no longer a martingale difference sequence). The martingale approximation obtained in this paper and Theorem 3.4 give an alternative approach to verify the joint asymptotic normality of the linear and quadratic terms in $\hat{\theta}_n - \theta$, under essentially the same conditions imposed in the i.i.d. case.

Assumption A3 (a) $\{X_n = (Y_n, W_n, Z_n),\ n \geq 0\}$ is a stationary Markov chain with transition kernel $P(x_1, dx_2) = Q_z(z_1, dz_2) Q_{yw|z}(z_2; dy_2, dw_2)$ where $Q_z(z_1, dz_2) = q_z(z_1, z_2)dz_2$ is a transition kernel with invariant distribution μ_z.

(b) Q_z satisfies A2 for some measurable function $V_z : \mathcal{Z} \to [1, \infty)$ and $\kappa_z = 0$.

(c) For $1 \leq p \leq 4$,

$$\sup_{z_1} V_z^{-p/8}(z_1) \left(m_{y,p}(z_1) + m_{w,p}(z_1)\right) < \infty,$$

and

$$\sup_{z_1,z_2} V_z^{-p/8}(z_1) q_z(z_1, z_2) \left(m_{y,p}(z_2) + m_{w,p}(z_2)\right) < \infty,$$

where $m_{y,p}(z) = \int |y|^p Q_{yw|z}(z; dy, dw)$ and $m_{w,p}(z) = \int |w|^p Q_{yw|z}(z; dy, dw)$.

Assumption A3-(a) is restrictive because it rules out any potential "feedback" of the outcome variables Y_n and W_n. We also impose $\kappa_0 = 0$. These restrictions are imposed for simplicity and may be relaxed. Since $\mathbb{E}\left(h(Y_n, W_n, Z_n)\right) = \mathbb{E}\left(\int Q_{yw|z}(Z_n; dy, dw) h(y, w, Z_n)\right)$, it is easy to verify that Assumption A3 implies A2 with $V(x) = V_z(z) + |w|^8 + |y|^8$ and

$\kappa = 0$. The short-range dependence assumed in A1 is crucial to obtain asymptotic normality in general, as shown by example in Cheng and Robinson (1994) for the case of $|s| \geq 1$ and W_n degenerated.

The main result of this section is given in the following theorem.

Theorem 3.5. *Assume that (S), (K) and A3 hold. If*

$$\min(1, nb_n^{d+2|s|})nb_n^{2(\varrho\wedge Q)} \to 0 \qquad \textit{and} \qquad n^2 b_n^d \to \infty$$

then

$$\Sigma_n^{-1}(\hat{\theta}_n - \theta) \to_d \mathcal{N}(0,1), \qquad \Sigma_n^2 = n^{-1}4\sigma_{n,1}^2 + \binom{n}{2}^{-1}\sigma_{n,2}^2,$$

with

$$\Sigma_n^2 = n^{-1}[4\sigma_1^2 + o(1)] + \binom{n}{2}^{-1} b_n^{-d-2|s|}\left[\sigma_2^2 \int_{\mathbb{R}^d} (\partial^s K(u))^2 du + o(1)\right].$$

Proof. See Section 6.6. □

Remark 6. (a) Although beyond the scope of this paper, feasible inference procedures can be constructed by developing appropriate consistent standard errors under the "small bandwidth asymptotic" regime. Assuming i.i.d. data, $|s| = 1$ and $W_n = -2$ for all $n \geq 0$, Cattaneo et al. (2011b) discuss in detail three alternative approaches to construct such standard-error estimators, which may be extended to the context of time-series data. This extension is important partially because it is shown in Cattaneo et al. (2011a) that the standard nonparametric bootstrap does not provide a valid asymptotic approximation under the general asymptotics considered in Theorem 4.5.
(b) The classical and alternative bandwidth selection procedures discussed in Cattaneo et al. (2010) may be easily extended to cover the more general setup considered in this section, although to conserve space we do not discuss the details here.

3.1.1. *Simulation Evidence.* This section summarizes the results of a small-scale simulation experiment conducted to explore the implications of the (small bandwidth) asymptotic distributional approximation obtained in Theorem 3.5 in samples of moderate size. To facilate the comparison with the results known in the context of i.i.d. data, we consider the same simulation setup as in Cattaneo et al. (2011b, 2010), but with time-dependent covariates. Specifically, we consider the density-weigthed average derivative estimand (i.e., $|s| = 1$ and $W_n = -2$ for all $n \geq 0$) in the context of single-index models. Let $\{\varepsilon_n,\ n \geq 0\}$ be a sequence of i.i.d. standard Gaussian random variables independent of $\{(Y_n, Z_n')',\ n \geq 0\}$, and assume that $Y_n = \tau(Z_n'\beta + \varepsilon_n)$ for $\beta \in \mathbb{R}^d$ and for some unknown weakly increasing real-valued link function $\tau(\cdot)$. Three natural choices are $\tau(u) = u$ (linear models), $\tau(u) = \mathbf{1}_{\{u>0\}}$ (binary choice models) and $\tau(u) = u\mathbf{1}_{\{u>0\}}$ (censored models).

Semiparametric density-weighted derivatives are of interest partially because, under stationarity and assumption S, integration by parts implies that $\theta = \mathbb{E}[\partial^s g_y(Z_n) f(Z_n)] \propto \beta$.

The time series dependence is modelled by assuming that each component in Z_n follows an autoregressive process of order one. Specifically, we impose $Z_n = AZ_{n-1} + v_n$ where for simplicity A is diagonal with all elements equal to $1/2$, and for all $n \geq 0$ we impose $v_n \sim \mathcal{N}(0, \sigma_v^2)$ independently of all other random variables and across time. Finally, Z_0 and σ_v^2 are chosen so that $Z_n \sim \mathcal{N}(0,1)$ for all $n \geq 0$.

To conserve space, and because qualitatively similar results were obtained in all cases, in this section we only summarize the main findings for the first component of $\theta = (\theta_1, \theta_2)'$ in the context of a binary choice model with $d = 2$ and $\varrho \in \{2, 4\}$. Under this configuration, the true data generating process is characterized by $\theta_1 = \frac{1}{8\pi^{3/2}}$, $4\sigma_1^2 \simeq 0.003667$, and $\sigma_2^2 \simeq 0.026552$. The estimator $\hat{\theta}_n$ is constructed using $K(\cdot)$ chosen to be a Gaussian product kernel of order ϱ, which implies that $\int_{\mathbb{R}^d} (\partial^s K(u))^2 du$ equals to $\frac{1}{8\pi}$ and $\frac{1485}{2048\pi}$ when $\varrho = 2$ and $\varrho = 4$, respectively.

To assess the relative improvements in the distributional approximation under classical asymptotics and under "small bandwidths asymptotics", we study the empirical coverage rate of the two competing (unfeasible) confidence intervals for θ_1 obtained under each regime. Specifically, the classical asymptotic approximation obtained in, e.g., Robinson (1989) (and also justified by Theorem 4.3 above) imply the following confidence intervals

$$\hat{\theta}_n \pm z_{1-\alpha/2} \sqrt{n^{-1} 4\sigma_1^2},$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. In contrast, the "small bandwidth asymptotics" (Theorem 4.5) justify the following alternative confidence intervals

$$\hat{\theta}_n \pm z_{1-\alpha/2} \sqrt{n^{-1} 4\sigma_1^2 + \binom{n}{2}^{-1} b_n^{-d-2} \sigma_2^2 \int_{\mathbb{R}^d} (\partial^s K(u))^2 du}.$$

Figure 1 reports the empirical converage as a function of the choice of bandwidth b_n for each of the competing confidence intervals in a simulation experiment with $n = 500$ and $S = 10,000$. The first panel corresponds to the case $\varrho = 2$ (theoretically not valid for the classical asymptotics), while the second panel reports the results for $\varrho = 4$ (where both confidence intervals are theoretically justified). The improvements of the small bandwidth asymptotics over the classical asymptotics are substantial, leading to confidence intervals with good empirical size properties across a non-trivial range of bandwidths.

4. Asymptotic variance estimation

Quadratic forms also appear in the problem of asymptotic variance estimation. Suppose that $\{X_n,\ n \geq 0\}$ is a nonstationary Markov chain with transition kernel P, invariant

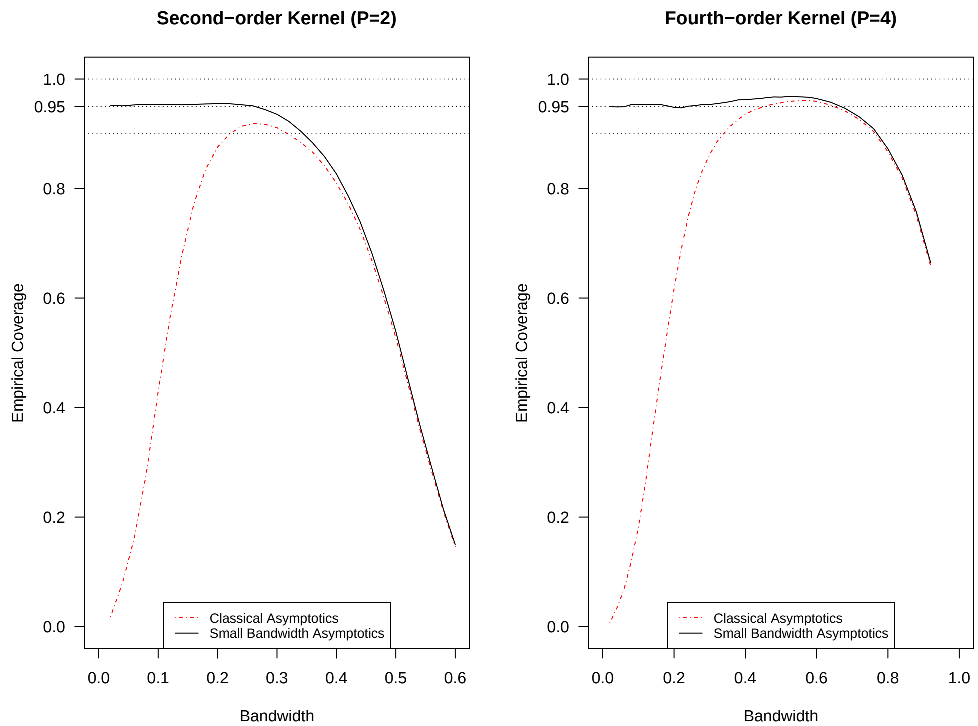

FIGURE 1. Empirical Coverage Rates for Infeasible 95% Confidence Intervals ($d = 2$, $n = 500$, $S = 10,000$)

distribution μ and some initial distribution ρ. Let $h: \mathcal{X} \to \mathbb{R}$ be a measurable function such that $\mu(|h|^2) < \infty$. We assume without any loss of generality that $\mu(h) = 0$. We are interested in the estimation of the long-run variance (or the asymptotic variance) of h defined as:

$$\sigma^2(h) = \mathsf{Var}_\mu(h(X_0)) + 2\sum_{\ell \geq 1} \mathsf{Cov}_\mu\left(h(X_0), h(X_\ell)\right), \tag{22}$$

which plays a role in time series analysis and in Markov Chain Monte Carlo.

We consider here the class of lag-window estimators (see, e.g., Priestley (1981) for detailed discussion and early references). For $0 \leq k \leq n-1$, we define the k-th order sample autocovariance $\Upsilon_{n,k} := n^{-1}\sum_{j=1}^{n-k}\left(h(X_j) - \mu_n(h)\right)\left(h(X_{j+k}) - \mu_n(h)\right)$, where $\mu_n(h) = n^{-1}\sum_{j=1}^{n} h(X_j)$. To describe the estimator, we consider a weight function with the following properties.

Assumption W: For $b > 0$, $w_b: \mathbb{R} \to [0,1]$ is a continuous function with support $[0,b]$, of class $\mathcal{C}^2$ on $(0,b)$, such that $w_b(b) = 0$ and $w_b(0) = 1$.

This assumption allows for the use of all commonly employed weighting functions, including the Bertlett and Parzen kernels. Let $\{c_n,\ n \geq 1\}$ be an increasing sequence of positive numbers such that $c_n \leq n$ and $c_n \to \infty$. The lag-window estimator of $\sigma^2(h)$ is given by

$$\Gamma^2_{n,b}(h) := \Upsilon_{n,0} + 2\sum_{k=1}^{n-1} w_b(kc_n^{-1})\Upsilon_{n,k}. \tag{23}$$

When $w_b(x) = w(x/b)$, an equivalent parametrization is $w_b(x) = w(x)$ and $c_n = bc_n$. It is well known that for $\Gamma^2_{n,b}(h)$ to consistently estimate $\sigma^2(h)$, we need the condition $c_n = o(n)$. But it is also well documented that for $c_n = o(n)$, lag-window estimators often have poor finite-sample properties, particularly for highly correlated time-series. Alternatively, it has been suggested to use $\Gamma^2_{n,b}(h)$ with $c_n = n$, the so-called "fixed-b asymptotics" (Kiefer and Vogelsang (2005)). In the regime $c_n = n$, $\Gamma^2_{n,b}(h)$ is no longer a consistent estimator of $\sigma^2(h)$, but can still be used to derive asymptotically valid confidence intervals and statistical tests for $\mu(h)$. These statistical procedures often have better finite sample properties than their counterparts based on $c_n = o(n)$ (see, e.g., Jansson (2004) and Sun et al. (2008) for some theoretical justifications).

Using Lemma 2.2, we derive a decomposition of the lag-window estimator $\Gamma^2_{n,b}(h)$ that sheds some light on the asymptotic and finite-sample behavior of the estimator, in both cases $c_n = n$ and $c_n = o(n)$. Then we study the asymptotics of $\Gamma^2_{n,b}(h)$ for nonstationary Markov chain and under conditions that are more easily verifiable. The conditions given below substantially improve on Flegal and Jones (2010) and Atchade (2010). We impose the following assumption.

Assumption A4 There exist measurable functions $V_k : \mathcal{X} \to [1,\infty)$ $(k = 1,2,3)$ such that $V_1 \le V_2$, $V_2^2 \le V_3$ such that $PV_3(x) < \infty$, and P satisfies the assumptions $\mathsf{C}(V_1, V_2)$ and $\mathsf{C}(V_2^2, V_3)$. Furthermore there exists $q > 1$ such that

$$\sup_{n\ge 0} \mathbb{E}\left(V_3^q(X_n)\right) < \infty. \tag{24}$$

A4 implies A1 with $W_1 = V_2$, $\bar V(x,y) = V_1(x)V_1(y)$ and $\bar W(x,y) = V_2(x)V_2(y)$. Define the partial sums $S_{n,k} := \sum_{j=k+1}^{k+n} h(X_j)$, the weight $w_{n,b}(0) = n^{-1}$ and $w_{n,b}(k) = 2n^{-1}w_b(kc_n^{-1})$ for $k > 0$. We write $\Upsilon_{n,k}$ as $\Upsilon_{n,k} = n^{-1}\sum_{j=1}^{n-k} h(X_j)h(X_{j+k}) + n^{-3}(n-k)S^2_{n,0} - n^{-2}S_{n,0}\left(S_{n-k,0} + S_{n-k,k}\right)$, so that

$$\Gamma^2_{n,b}(h) = \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_{n,b}(\ell - j)h(X_j)h(X_\ell) + R_n, \tag{25}$$

where

$$R_n = 2n^{-2}S^2_{n,0}\sum_{k=1}^{n-1} w_b(kc_n^{-1})\left(1 - \frac{k}{n}\right) - 2n^{-2}S_{n,0}\left(\sum_{j=2}^{n} h(X_j)\sum_{k=1}^{j-1} w_b(kc_n^{-1}) + \sum_{j=1}^{n-1} h(X_j)\sum_{k=1}^{n-j} w_b(kc_n^{-1})\right) - n^{-2}S^2_{n,0}. \tag{26}$$

If we set aside the term R_n, the expression (25) is of the form (1) with $h_n(x,y) = h(x)h(y)$ and $w_n(\ell,j) = w_{n,b}(\ell-j)$. Here we have $h_{n,1}(x) = \int h(x)h(y)\mu(dy) = 0$, $\theta_n = 0$, and $h_{n,2}(x,y) = h(x)h(y)$. Therefore $L_{n,\ell} = 0$. Define

$$G(x) := \sum_{j\geq 0} P^j h(x), \quad \text{and} \quad PG(x) = \int P(x,dz)G(z), \;\; x \in \mathcal{X}.$$

Then $G_h(x,y) = G(x)G(y)$, $G_{Ph}(x,y) = PG(x)G(y)$, $G_{\overline{Ph}}(x,y) = G(x)PG(y)$ and $G_{P^2h}(x,y) = PG(x)PG(y)$. Therefore

$$Q_{n,\ell,j} = Q_\ell Q_j, \quad \text{where} \quad Q_\ell = G(X_\ell) - PG(X_{\ell-1}).$$

As above, $\{(G_\ell,\mathcal{F}_\ell),\ \ell \geq 1\}$ is a martingale: $\mathbb{E}\left(Q_\ell|\mathcal{F}_{\ell-1}\right) = 0$. Lemma 2.2 yields for all $n \geq 1$,

$$\Gamma^2_{n,b}(h) = n^{-1}\sum_{\ell=1}^n Q_\ell^2 + \sum_{\ell=1}^n\sum_{j=1}^{\ell-1} w_{n,b}(\ell-j)Q_\ell Q_j + \zeta_n + R_n, \tag{27}$$

where by rearranging appropriately the terms we get for $n \geq 3$ and writing $\Delta_{n,b}(k) = w_{n,b}(k) - w_{n,b}(k-1)$,

$$\begin{aligned}
\zeta_n = \sum_{\ell=3}^n PG(X_{\ell-1})\sum_{j=1}^{\ell-2}\Delta_{n,b}(\ell-j)Q_j - \sum_{\ell=2}^n Q_\ell \sum_{j=1}^{\ell-1}\Delta_{n,b}(\ell-j+1)PG(X_{j-1}) \\
+ \sum_{\ell=3}^n PG(X_{\ell-1})\sum_{j=1}^{\ell-2}\left(\Delta_{n,b}(\ell-j) - \Delta_{n,b}(\ell-j+1)\right)PG(X_{j-1}) \\
+ \left(2w_{n,b}(0) - w_{n,b}(1)\right)\sum_{\ell=1}^n \left(PG(X_{\ell-1})Q_\ell + (PG(X_{\ell-1}))^2 - G(X_\ell)PG(X_\ell)\right) \\
- \Delta_{n,b}(2)\sum_{\ell=2}^n PG(X_{\ell-1})PG(X_{\ell-2}) + PG(X_0)\sum_{\ell=1}^n w_{n,b}(\ell)Q_\ell - PG(X_n)\sum_{\ell=1}^n w_{n,b}(n-\ell)Q_\ell \\
+ \sum_{\ell=2}^n \left(w_{n,b}(\ell) - w_{n,b}(\ell-1)\right)PG(X_0)PG(X_{\ell-1}) - 2w_{n,b}(n)PG(X_0)PG(X_n) \\
+ PG(X_n)\sum_{\ell=1}^{n-1}\left(w_{n,b}(n-\ell) - w_{n,b}(n-\ell+1)\right)PG(X_{\ell-1}) - w_{n,b}(0)G(X_n)PG(X_n) \\
+ \Delta_{n,b}(1)(PG(X_0))^2 + w_{n,b}(0)(PG(X_n))^2 + \Delta_{n,b}(1)PG(X_n)PG(X_{n-1}).
\end{aligned}$$

In fact the decomposition (27) is similar to the decomposition derived in Atchade (2010) Theorem 4.1, with the exception that the quadratic term is not distinctly identified as in (27). Under A4 and using the martingale property of $\{Q_\ell,\ \ell \geq 1\}$, the smoothness of w_b, and the same technique as in Atchade (2010), it can be shown that the exists $p > 1$ such that

$$\mathbb{E}^{1/p}\left(|\zeta_n|^p\right) \leq c c_n^{-1+\frac{1}{2}\vee\frac{1}{p}}, \quad n \geq 3, \tag{28}$$

for some finite constant c. We have the following theorem.

Theorem 4.1. *Assume (A4) and (W) and $h \in \mathcal{L}_{V_1}$. Let $\Gamma^2_{n,b}(h)$ be as in (23), where $c_n \uparrow \infty$, and let $\{B(t),\ 0 \le t \le 1\}$ be the standard Brownian motion. There exist $p > 1$ and a finite constant c such that for all $n \ge 3$,*

$$\mathbb{E}^{1/p}\left(|\zeta_n|^p\right) \le c c_n^{-1+\frac{1}{2}\vee\frac{1}{p}}, \quad \mathbb{E}^{1/p}\left(|R_n|^p\right) \le c n^{-1} c_n,$$
$$\text{and} \quad \mathbb{E}^{1/p}\left(\left|\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} w_{n,b}(\ell-j) Q_\ell Q_j\right|^p\right) \le c\left(\frac{c_n}{n}\right)^{\frac{1}{2}} n^{-\frac{1}{2}+\frac{1}{p}\vee\frac{1}{2}}.$$

(1) If $c_n = o(n)$, then $\Gamma^2_{n,b}(h)$ converges in probability to $\sigma^2(h)$. Furthermore, assuming $\Gamma^2_{n,b}(h) > 0$ almost surely, $\{n\Gamma^2_{n,b}(h)\}^{-1/2}\sum_{j=1}^n (h(X_j) - \mu(h)) \to_d \mathcal{N}(0,1)$.

(2) If $c_n = n$, then $\Gamma^2_{n,b}(h) \to_d \sigma^2(h)\mathsf{f}_b$, where

$$\mathsf{f}_b = 1 + 2\int_0^1\int_0^t w_b(t-s)dB(s)dB(t) - 2B(1)\int_0^1 g_b(t)dB(t) + 2B^2(1)\int_0^1 (1-t)w_b(t)dt,$$

where $g_b(t) = \int_0^t w_b(u)du + \int_0^{1-t} w_b(u)du$. Furthermore, assuming $\Gamma^2_{n,b}(h) \ne 0$ almost surely,

$$\frac{n^{-1/2}\sum_{j=1}^n (h(X_j) - \mu(h))}{\sqrt{|\Gamma^2_{n,b}(h)|}} \to_d \frac{B(1)}{\sqrt{|\mathsf{f}_b|}}.$$

Proof. See Section 6.7. □

Although the limiting distribution $B(1)/\sqrt{|\mathsf{f}_b|}$ is non-standard, it can be simulated, for example by Euler discretization of the stochastic integrals in f_b. We report in Table 1 the 95% quantiles of the distribution of $B(1)/\sqrt{|\mathsf{f}_b|}$ using $w_b(x) = \mathbf{1}_{(0,b)}(x)$, $w_b(x) = (1 - x/b)\mathbf{1}_{(0,b)}(x)$ and $w_b(x) = (1 - (x/b)^2)\mathbf{1}_{(0,b)}(x)$, and for different values of b, based on $10,000$ replications of $B(1)/\sqrt{|\mathsf{f}_b|}$. The distribution departs further from the standard normal distribution as b increases.

	$w(x) = 1 - x/b$	$w(x) = 1 - (x/b)^2$	$w(x) = \mathbf{1}_{(0,1)}(x)$
$b = 0.3$	2.828	4.134	5.496
$b = 0.5$	3.557	6.580	6.299
$b = 0.9$	4.735	12.575	13.045

TABLE 1. 0.975-quantile of the distribution of $B(1)/\sqrt{|\mathsf{f}_b|}$.

4.1. **Application to Markov Chain Monte Carlo.** Markov Chain Monte Carlo (MCMC) is a popular computational tools to obtain random samples from intractable and high-dimensional distributions (see e.g. Roberts and Rosenthal (2004) for a survey and for additional references). Suppose that we interested in sampling from the probability measure μ and compute the integral $\mu(h) = \int h(x)\mu(dx)$. Let $\{X_n,\ n \geq 0\}$ be a Markov chains with transition kernel P, invariant distribution μ and initial distribution ρ. By simulating the Markov chain, we approximate $\mu(h)$ by the Monte Carlo average $\mu_n(h) = n^{-1}\sum_{k=1}^n h(X_k)$. Furthermore, under A4, $\lim_{n\to\infty} n^{1/2}\mathsf{Var}\left(\mu_n(h)\right) = \sigma^2(h)$, as given by (22), and a central limit theorem holds: $n^{-1/2}\sum_{k=1}^n \left(h(X_k) - \pi(h)\right) \to_d N(0, \sigma^2(h))$. By Theorem 4.1 (1) a consistent $(1-\alpha)$-confidence interval for $\mu(h)$ is

$$\mu_n(h) \pm z_{1-\alpha/2}\frac{\hat{\sigma}_n(h)}{\sqrt{n}},$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution and where $\hat{\sigma}_n(h) = \sqrt{\Gamma^2_{n,b}(h)}$, with $b = 1$, $c_n = o(n)$. Typical choice of c_n includes $c_n = n^{-\delta}$, $\delta \in \{1/3, 1/2, 2/3\}$.

Theorem 4.1 (2) provides another confidence interval for $\mu(h)$:

$$\mu_n(h) \pm t_{1-\alpha/2}\frac{\tilde{\sigma}_n(h)}{\sqrt{n}},$$

where $t_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the distribution of $B(1)/\sqrt{|\mathsf{f}_b|}$ and where $\tilde{\sigma}_n(h) = \sqrt{\Gamma^2_{n,b}(h)}$, with $c_n = bn$. We compare the finite sample properties of these two confidence intervals in terms of coverage probability and interval length.

4.1.1. *Illustration.* We undertake the comparison using a log-linear model taken from Gelman et al. (2004). For $e = 1, \ldots, N_e$ and $p = 1, \ldots, N_p$, the variables y_{ep} are conditionally independent given $(\{\beta_p\}, \{\varepsilon_{ep}\}) \in \mathbb{R}^{N_p} \times \mathbb{R}^{N_e N_p}$, with conditional distribution

$$y_{ep} \sim \mathcal{P}\left(n_{ep}e^{\mu+\alpha_e+\beta_p+\varepsilon_{ep}}\right),\ \ e = 1, \ldots, N_e,\ \ p = 1, \ldots, N_p, \tag{29}$$

where $\mathcal{P}(\lambda)$ is the Poisson distribution with parameter λ. In the above display, $\{n_{ep}\}$ is a deterministic baseline covariate, and $\mu \in \mathbb{R}$, $\{\alpha_e\} \in \mathbb{R}^{N_e}$ are parameters. We assume that $\{\beta_p\}$ and $\{\varepsilon_{ep}\}$ are independent with distributions

$$\beta_p \overset{iid}{\sim} N(0, \sigma^2_\beta), \quad \varepsilon_{ep} \overset{iid}{\sim} N(0, \sigma^2_\varepsilon), \tag{30}$$

For some parameters $\sigma^2_\beta > 0$, $\sigma^2_\varepsilon > 0$. We assume a diffuse prior for $(\mu, \alpha, \sigma^2_\beta, \sigma^2_\varepsilon)$ $(\sigma^2_\epsilon > 0, \sigma^2_\beta > 0)$ with the additional constraint that $\alpha_{N_e} = -\sum_{k=1}^{N_e-1}\alpha_k$. Let $\theta = (\mu, \alpha, \beta, \epsilon, \sigma^2_\epsilon, \sigma^2_\beta) \in \mathbb{R}^{3+N_e-1+(N_p+1)N_e}$. The posterior distribution of θ given $\mathcal{D} = (y_{ep}, n_{ep})$

takes the form

$$\pi(\theta|\mathcal{D}) \propto \exp\Bigg\{\sum_{e,p} y_{e,p}(\mu+\alpha_e+\beta_p+\epsilon_{e,p}) - n_{ep}e^{\mu+\alpha_e+\beta_p+\epsilon_{ep}} \\ -\frac{N_eN_p}{2}\log\sigma_\epsilon^2 - \frac{N_p}{2}\log\sigma_\beta^2 - \frac{1}{2\sigma_\epsilon^2}\sum_{e,p}\epsilon_{e,p}^2 - \frac{1}{2\sigma_\beta^2}\sum_{p=1}^{N_p}\beta_p^2\Bigg\}. \quad (31)$$

This posterior distribution is typical of the probability distributions for which MCMC is useful. We set $N_e = 3$ and $N_p = 20$. Consider the posterior mean of the parameter α_1, i.e. $\int \alpha_1\pi(\theta|\mathcal{D})d\theta$. To compare the different confidence interval for α_1, we generate an artificial dataset with $(\alpha_1, \alpha_2, \mu, \sigma_\varepsilon^2, \sigma_\beta^2) = (0.35, 0.15, -1.0, 0.1, 0.3)$. We run a preliminary MCMC sampler for 6 millions (6×10^6) iterations and compute its sample mean. We obtain $\bar{\alpha}_1 = 0.3309$. We take this value to be $\int \alpha_1\pi(\theta)d\theta$.

To compare the two confidence interval methods, we use a Random Walk Metropolis (RWM) algorithm with proposal kernel $\mathcal{N}(0, \kappa\Sigma)$. We want to compare the confidence interval methods in various scenarios of mixing of the MCMC chain. We work with two versions of the RWM: a naive algorithm in which $\kappa = 0.015$ and $\Sigma = I_{85}$, the identity matrix. The other version is an optimized RWM where κ and Σ are selected (after a preliminary simulation) to improve the mixing of the chain. We run both chains for $50,000$ iterations. We repeat the simulations 400 times in order to estimate the coverage probabilities and interval lengths. The results are given in Table 2 and 3. As expected, the regime $c_n = n$ have a much better finite sample behavior. The improvement is particularly drastic when the underlying Markov chain is slowly mixing.

	$n^{1/3}$	$n^{1/2}$	$n^{2/3}$
Coverage	9.25	16.5	39.25
Length	0.006	0.015	0.034

	$b = 0.3$	$b = 0.5$	0.9
Coverage	77.50	82.25	84.00
Length	0.097	0.115	0.123

TABLE 2. Coverage probability and confidence interval length for α_1. $w_b(x) = (1 - x/b)\mathbf{1}_{(0,b)}(x)$. From $50,000$ iterations of the naive RWM algorithm. Right table is the regime $c_n = n$.

	$n^{1/3}$	$n^{1/2}$	$n^{2/3}$
Coverage	47.0	79.5	92.00
Length	0.007	0.015	0.021

	$b = 0.3$	$b = 0.5$	0.9
Coverage	95.0	95.0	96.0
Length	0.027	0.031	0.031

TABLE 3. Coverage probability and confidence interval length for α_1. $w_b(x) = (1 - x/b)\mathbf{1}_{(0,b)}(x)$. From $50,000$ iterations of the optimized RWM algorithm. Right table is the regime $c_n = n$.

5. Discussion and Final Remarks

The martingale approximation introduced in Lemma 2.2 for quadratic forms of Markov Chains permits a simple and systematic analysis of many interesting problems in statistics and econometrics via martingale limit theory, as illustrated for the important problems considered in Sections 4 and 5. This decomposition also leads to an interesting connection between the two type of non-standard asymptotic experiments disscused in the previous sections, which partially sheds light on the important improvements achieved by these alternative large sample approximations.

To see this connection, note first that Lemma 2.2 implies that each of the quadratic statistics in Sections 4 and 5 can be decomposed into a single average and a double average, each forming a martingale sequence. Specifically, for the "small bandwidth asymptotics" considered in Section 4, and assuming $nb_n^{d+2|s|} \to \kappa \in (0, \infty]$ for simplicity (this assumption corresponds to the $\sqrt{n}$-case), Theorem 3.5 gives

$$\sqrt{n}(\hat{\theta}_n - \theta) = \underbrace{\sqrt{n}(\theta_n - \theta)}_{O(\sqrt{n}b_n^{P\wedge Q})} + \underbrace{\frac{2}{\sqrt{n}}\sum_{i=1}^{n} L_{n,i}}_{O_p(1)} + \underbrace{\sqrt{n}\binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{i-1} Q_{n,i,j}}_{O_p\left(\frac{1}{\sqrt{nb_n^{d+2|s|}}}\right)} + \underbrace{\sqrt{n}\zeta_n}_{o_p(1)},$$

while for the "fixed-b asymptotics" studied in Section 5 (with $p = 2$), Theorem 5.1 shows that

$$\Gamma_{n,b}^2(h) = \underbrace{n^{-1}\sum_{\ell=1}^{n} Q_\ell^2}_{O_p(1)} + \underbrace{\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1} w_{n,b}(\ell - j)Q_\ell Q_j + R_n}_{O_p\left(\sqrt{\frac{c_n}{n}} + \frac{c_n}{n}\right)} + \underbrace{\zeta_n}_{o_p(1)},$$

where R_n is also a double average. Thus, both statistics are represented by a sum of a single average, a double average and remainder in general.

This representation shows that the classical large sample distributional approximations correspond to tuning parameter sequences that assume away the asymptotic contribution of the double average obtained in each example. In contrast, both the "small bandwidth asymptotic" framework and the "fixed-b asymptotic" framework correspond to sequences of tuning parameters that render this double average of (at least) the same stochastic order as the single average, and hence leading to the joint weak convergence of both terms.

In the context of i.i.d. data, this connection also arises when comparing the "small bandwidth asymptotics" with other non-standard asymptotic approximations such as the "many instruments asymptotics" and the "many regressors asymptotics" (see Cattaneo et al. (2011c) for a discussion). In all these cases, the alternative asymptotic experiments explicitly capture a quadratic term (i.e., a double average), which is otherwise ignored by the classical distributional approximations typically employed in the literature, thereby providing a better distributional approximation for the underlying statistics of interest.

6. Proofs

6.1. **Proof Lemma 2.1.** Since $R(x, y; \cdot) \in \mathcal{M}_{\bar{V}}(\mathcal{X} \times \mathcal{X})$, and $h \in \mathcal{L}_{\overline{V}}$, $G_h(x, y) = \int\int h(u, v)R(x, y; (du, dv))$ is well-defined and $|G_h(x, y)| \leq c|h|_{\bar{V}}\overline{W}(x, y)$, for all $x, y \in \mathcal{X}$.

Suppose that $h, Ph \in \mathcal{L}_{\overline{V}}$. Let $\phi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ be a bijection (that is, an indexing of $\mathbb{N} \times \mathbb{N}$). We write $\phi(n) = (\phi_1(n), \phi_2(n))$. For $N \in \mathbb{N}$, define $g_N(x, y) = \sum_{n=0}^{\phi_1(N)} \sum_{m=0}^{\phi_2(N)} \int\int h(u, v)\pi_{n,m}(x, y; (du, dv))$. Then under A1, $g_N(x, y)$ converges to $G_h(x, y)$ for all $x, y \in \mathcal{X}$ as $N \to \infty$. Also, $|g_N(x, y)| \leq c\overline{W}(x, y)$, for all $x, y \in \mathcal{X}$ and uniformly in N, and $\int P(x, dz_1) \int \delta_y(dz_2)\overline{W}(z_1, z_2) = \int P(x, dz)W(z, y) < \infty$, by assumption. By dominated convergence,

$$
\begin{aligned}
\int P(x, dz)G_h(z, y) &= \int\int P(x, dz_1)\delta_y(dz_2)G_h(z_1, z_2) \\
&= \lim_{N\to\infty} \sum_{n=0}^{\phi_1(N)} \sum_{m=0}^{\phi_2(N)} \int P(x, dz_1) \int \delta_y(dz_2) \int\int \pi_{n,m}(z_1, z_2; (du, dv))h(u, v) \\
&= \lim_{N\to\infty} \sum_{n=0}^{\phi_1(N)} \sum_{m=0}^{\phi_2(N)} \int\int \pi_{n,m}(x, z; (du, dv))\{Ph\}(z, y) = G_{\{Ph\}}(x, y),
\end{aligned}
$$

which proves (5).

From the assumptions, it follows that the functions G_h, $G_{\{Ph\}}$, $G_{\{\overline{Ph}\}}$, and $G_{\{P^2h\}}$ are well defined. Since h is degenerate, it is easy to check that for all $n, m \geq 0$, $\pi_{n,m}(Ph) = \pi_{n+1,m}(h)$, $\pi_{n,m}(\overline{Ph}) = \pi_{n,m+1}(h)$ and $\pi_{n,m}(P^2h) = \pi_{n+1,m+1}(h)$. Therefore for any $N, M \geq 1$:

$$
\begin{aligned}
&\sum_{n=0}^{N}\sum_{m=0}^{M} \int\int h(u, v)\pi_{n,m}(x, y; (du, dv)) - \sum_{n=0}^{N}\sum_{m=0}^{M} \int\int \{Ph\}(u, v)\pi_{n,m}(x, y; (du, dv)) \\
&-\sum_{n=0}^{N}\sum_{m=0}^{M} \int\int \{\overline{Ph}\}(u, v)\pi_{n,m}(x, y; (du, dv)) + \sum_{n=0}^{N}\sum_{m=0}^{M} \int\int \{P^2h\}(u, v)\pi_{n,m}(x, y; (du, dv)) \\
&= h(x, y) - \int \left(P^{N+1}(x, du) - \mu(du)\right) h(u, y) - \int \left(P^{M+1}(y, dv) - \mu(dv)\right) h(x, v) \\
&\qquad + \int\int \left(P^{N+1}(x, du) - \mu(du)\right)\left(P^{M+1}(y, dv) - \mu(dv)\right) h(u, v).
\end{aligned}
$$

Letting $N, M \to \infty$ yields $h = G_h - G_{\{\overline{Ph}\}} - G_{\{Ph\}} + G_{\{P^2h\}}$. Combined with (5) and the definition $\bar{G}_h := G_h - G_{\{\overline{Ph}\}}$ yields the identity (6).

□

6.2. **Proof Lemma 2.2.** From the definition, we have $h_n(x,y) = \theta_n + \bar{h}_{n,1}(x) + \bar{h}_{n,1}(y) + \bar{h}_{n,2}(x,y)$, and we deduce after some rearrangements that

$$U_n(h_n) = \theta_n \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell,j) + \sum_{\ell=1}^{n}\gamma_{n,\ell}\bar{h}_{n,1}(X_\ell) + \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell,j)\bar{h}_{n,2}(X_\ell,X_j).$$

The handling of the term $\sum_{\ell=1}^{n}\gamma_{n,\ell}\bar{h}_{n,1}(X_\ell)$ is standard. Using (9), we obtain

$$\sum_{\ell=1}^{n}\gamma_{n,\ell}\bar{h}_{n,1}(X_\ell) = \sum_{\ell=1}^{n}\gamma_{n,\ell}L_{n,\ell} + \zeta_n^{(1)}, \quad n \geq 1$$

where $L_{n,\ell} = g_n(X_\ell) - Pg_n(X_{\ell-1})$, and

$$\zeta_n^{(1)} = \gamma_{n,0}Pg_n(X_0) - \gamma_{n,n}Pg_n(X_n) + \sum_{\ell=1}^{n}(\gamma_{n,\ell} - \gamma_{n,\ell-1})Pg_n(X_{\ell-1}).$$

Using (6), we write $\bar{h}_{n,2}(X_\ell,X_j) = G_{\bar{h}_{n,2}}(X_\ell,X_j) - G_{P\bar{h}_{n,2}}(X_\ell,X_j) - G_{\overline{P\bar{h}}_{n,2}}(X_\ell,X_j) + G_{P^2\bar{h}_{n,2}}(X_\ell,X_j)$. Thus

$$\begin{aligned}
\sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell,j)\bar{h}_{n,2}(X_\ell,X_j) &= \sum_{\ell=1}^{n}\sum_{j=1}^{\ell} w_n(\ell,j)Q_{n,\ell,j} \\
&+ \sum_{\ell=1}^{n}\sum_{j=1}^{\ell}\varpi_n^{(1)}(\ell,j)G_{P\bar{h}_{n,2}}(X_{\ell-1},X_j) + \sum_{\ell=1}^{n}\sum_{j=1}^{\ell}\varpi_n^{(2)}(\ell,j)G_{\overline{P\bar{h}}_{n,2}}(X_\ell,X_{j-1}) \\
&- \sum_{\ell=1}^{n}\sum_{j=1}^{\ell}\left(w_n(\ell,j) - w_n(\ell-1,j-1)\right)G_{P^2\bar{h}_{n,2}}(X_{\ell-1},X_{j-1}) + \epsilon_n, \quad (32)
\end{aligned}$$

where ϵ_n is comprised of the remainder telescoping sums. We obtain

$$\begin{aligned}
\epsilon_n = \sum_{\ell=1}^{n}&\Big(w_n(\ell-1,\ell)G_{P\bar{h}_{n,2}}(X_{\ell-1},X_\ell) - w_n(\ell,\ell)G_{\overline{P\bar{h}}_{n,2}}(X_\ell,X_\ell)\Big) \\
&+ \sum_{\ell=1}^{n} w_n(\ell,0)G_{\overline{P\bar{h}}_{n,2}}(X_\ell,X_0) - w_n(0,0)G_{P^2\bar{h}_{n,2}}(X_0,X_0) \\
&+ \sum_{\ell=1}^{n} w_n(n,\ell)\left(G_{P^2\bar{h}_{n,2}}(X_n,X_\ell) - G_{P\bar{h}_{n,2}}(X_n,X_\ell)\right).
\end{aligned}$$

The decomposition follows by rearranging the terms in (32).

□

6.3. **Proof Theorem 3.2.** The starting point is (15):

$$n^{-\lambda}\left(U_n(h_n) - \binom{n}{2}\theta_n\right) = n^{-\lambda}\sum_{\ell=2}^{n}\left((n-1)L_{n,\ell} + \sum_{j=1}^{\ell-1}Q_{n,\ell,j}\right) + n^{-\lambda}\zeta_n.$$

Note that for $|Ph_{n,1}|_{V^\beta} \le \|h_{n,1}\|_{2,V^\beta}$ and $|Ph_n|_{V^\beta_{(2)}} \le \|h_n\|_{2,V^\beta_{(2)}}$. In view of (16),

$$n^{-\lambda}\mathbb{E}\left(|\zeta_n|\right) \le cn^{-(\lambda-1)}\left(\|h_{n,1}\|_{2,V^\beta} + \|h_n\|_{2,V^\beta_{(2)}}\right),$$

which together with (17) or (18) shows that $n^{-\lambda}\zeta_n$ converges in probability to zero.

By conditioning on $\mathcal{F}_{\ell-1}$ and using (10), (11) and (14), we get for any $p \ge 1$ such that $2p(\beta+\kappa) \le 1$,

$$\sup_{1\le \ell\le n} \mathbb{E}^{1/p}\left(|L_{n,\ell}|^p\right) \le c\,\|h_{n,1}\|_{p,V^\beta}\,, \quad \text{and} \quad \sup_{1\le j<\ell\le n} \mathbb{E}^{1/p}\left(|Q_{n,\ell,j}|^p\right) \le c\,\|h_n\|_{p,V^\beta_{(2)}}\,. \tag{33}$$

For any $p > 1$ such that $2p(\beta+\kappa) \le 1$, by Burkeholder's inequality and (33),

$$n^{-\lambda}\mathbb{E}^{1/p}\left[\left|\sum_{\ell=2}^{n}\left((n-1)L_{n,\ell} + \sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)\right|^p\right] \le cn^{-(\lambda-1)+\frac{1}{2}\vee\frac{1}{p}}\left(\|h_{n,1}\|_{p,V^\beta} + \|h_n\|_{p,V^\beta_{(2)}}\right),$$

for some finite constant c, which proves that (1) is sufficient. On the other hand,

$$\begin{aligned}\mathbb{E}\left[\left((n-1)L_{n,\ell} + \sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^2\right] &= (n-1)^2\mathbb{E}\left(L_{n,\ell}^2\right) + \sum_{j=1}^{\ell-1}\mathbb{E}\left(Q_{n,\ell,j}^2\right) \\ &\quad + 2(n-1)\sum_{j=1}^{\ell-1}\mathbb{E}\left(L_{n,\ell}Q_{n,\ell,j}\right) + 2\sum_{k=1}^{\ell-2}\sum_{j=k+1}^{\ell-1}\mathbb{E}\left(Q_{n,\ell,j}Q_{n,\ell,k}\right).\end{aligned}$$

We shall show that there exists a finite constant c such that for all $n \ge \ell \ge 1$,

$$\begin{aligned}&\sum_{j=1}^{\ell-1}\left|\mathbb{E}\left(L_{n,\ell}Q_{n,\ell,j}\right)\right| \le c\,\|h_{n,1}\|_{2,V^\beta}\,\|h_n\|_{2,V^\beta_{(2)}}\,, \\ &\qquad\qquad \text{and} \quad \sum_{k=1}^{\ell-2}\sum_{j=k+1}^{\ell-1}\left|\mathbb{E}\left(Q_{n,\ell,j}Q_{n,\ell,k}\right)\right| \le c\,\|h_n\|_{2,V^\beta_{(2)}}^2\,.\end{aligned} \tag{34}$$

Then, using the martingale property, (34) and (33), we obtain

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{\ell=1}^{n}\left((n-1)L_{n,\ell} + \sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)\right)^2\right] &= \sum_{\ell=1}^{n}\mathbb{E}\left[\left((n-1)L_{n,\ell} + \sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^2\right] \\ &\le c\left(n^3\,\|h_{n,1}\|_{2,V^\beta}^2 + n^2\,\|h_n\|_{2,V^\beta_{(2)}}^2 + n^2\,\|h_{n,1}\|_{2,V^\beta}\,\|h_n\|_{2,V^\beta_{(2)}}\right).\end{aligned}$$

and the result follows from (18).

It remains to prove (34). We prove the second inequality. The first is obtained similarly. By the Markov property and A1, we have almost surely

$$\begin{aligned}
&\mathbb{E}\left(Q_{n,\ell,j}Q_{n,\ell,k}|\mathcal{F}_{k-1}\right) \\
&\quad = \int P(X_{k-1},dx_k)\int\left\{P^{j-1-k}(x_k,dx_{j-1})-\mu(dx_{j-1})\right\}\int P(x_{j-1},dx_j)\times \\
&\qquad\qquad \int\left\{P^{\ell-1-j}(x_j,dx_{\ell-1})-\mu(dx_{\ell-1})\right\}\Delta(x_{\ell-1},x_j,x_{j-1},x_k,X_{k-1}),
\end{aligned}$$

where

$$\Delta(x_{\ell-1},x_j,x_{j-1},x_k,x_{k-1})=\int P(x_{\ell-1},dx_\ell)Q_n(x_j,x_{j-1};x_\ell,x_{\ell-1})Q_n(x_k,x_{k-1};x_\ell,x_{\ell-1}).$$

By the Cauchy-Schwartz inequality and taking into account (11), we see that

$$\begin{aligned}
&|\Delta(x_{\ell-1},x_j,x_{j-1},x_k,x_{k-1})| \\
&\quad\le c\,|\!|\!|h_n|\!|\!|^2_{2,V^\beta_{(2)}}\,V^{2(\beta+\kappa)}(x_{\ell-1})\left(V^{\beta+\kappa}(x_j)+V^{\beta+\kappa}(x_{j-1})\right)\left(V^{\beta+\kappa}(x_k)+V^{\beta+\kappa}(x_{k-1})\right).
\end{aligned}$$

Using (13), we deduce that there exist a finite constant c and a convergent series $\sum\rho_1(n)<\infty$ such that

$$\begin{aligned}
&\left|\int P(x_{j-1},dx_j)\int\left\{P^{\ell-1-j}(x_j,dx_{\ell-1})-\mu(dx_{\ell-1})\right\}\Delta(x_{\ell-1},x_j,x_{j-1},x_k,x_{k-1})\right| \\
&\qquad\le c\,|\!|\!|h_n|\!|\!|^2_{2,V^\beta_{(2)}}\,\rho_1(\ell-j-1)\left(V^{\beta+\kappa}(x_k)+V^{\beta+\kappa}(x_{k-1})\right) \\
&\qquad\times\int P(x_{j-1},dx_j)V^{2(\beta+\kappa)+\kappa}(x_j)\left(V^{\beta+\kappa}(x_j)+V^{\beta+\kappa}(x_{j-1})\right) \\
&\qquad\le c\,|\!|\!|\bar{h}_n|\!|\!|^2_{2,V^\beta_{(2)}}\,\rho_1(\ell-j-1)V^{3(\beta+\kappa)+\kappa}(x_{j-1})\left(V^{\beta+\kappa}(x_k)+V^{\beta+\kappa}(x_{k-1})\right).
\end{aligned}$$

Using (13) again, there exists a positive series $\sum\rho_2(n)<\infty$ such that

$$\begin{aligned}
&|\mathbb{E}\left(Q_{n,\ell,j}Q_{n,\ell,k}|\mathcal{F}_{k-1}\right)| \\
&\le c\,|\!|\!|h_n|\!|\!|^2_{2,V^\beta_{(2)}}\,\rho_1(\ell-j-1)\rho_2(j-k-1)\int P(X_{k-1},dx_k)V^{3(\beta+\kappa)+2\kappa}(x_k)\left(V^{\beta+\kappa)}(x_k)+V^{(\beta+\kappa)}(x_{k-1})\right) \\
&\qquad\le c\,|\!|\!|h_n|\!|\!|^2_{2,V^\beta_{(2)}}\,\rho_1(\ell-j-1)\rho_2(j-k-1)V^{4(\beta+\kappa)+2\kappa}(X_{k-1}).
\end{aligned}$$

Since $4(\beta+\kappa)+2\kappa\le 1$ and $\sup_{n\ge 0}\mathbb{E}\left(V(X_n)\right)<\infty$, we obtain (34).

□

6.4. **Proof Theorem 3.3.** By (15), we have:

$$\begin{aligned}
\sigma_{n,1}^{-1}n^{1/2}\left(\binom{n}{2}^{-1}U_n(h_n)-\theta_n\right) &= \frac{2}{\sigma_{n,1}\sqrt{n}}\sum_{\ell=1}^{n}L_{n,\ell} \\
&\quad+\sigma_{n,1}^{-1}n^{1/2}\binom{n}{2}^{-1}\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1}Q_{n,\ell,j}+\sigma_{n,1}^{-1}n^{1/2}\binom{n}{2}^{-1}\zeta_n.
\end{aligned}$$

By Proposition 3.1 it is easy to see that under (19), $\sigma_{n,1}^{-1}n^{1/2}\binom{n}{2}^{-1}\zeta_n$ converges in probability to zero. By the martingale property and (34),

$$\mathbb{E}\left[\left(\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1}Q_{n,\ell,j}\right)^2\right]=\sum_{\ell=1}^{n}\mathbb{E}\left[\left(\sum_{j=1}^{\ell-1}Q_{n,\ell,j}\right)^2\right]\leq cn^2\left\|h_n\right\|_{2,V_{(2)}^{\beta}}^2 .$$

Therefore, given (19), we see that $\sigma_{n,1}^{-1}n^{1/2}\binom{n}{2}^{-1}\sum_{\ell=1}^{n}\sum_{j=1}^{\ell-1}Q_{n,\ell,j}$ converges to zero in probability.

We apply the martingale central limit theorem (Hall and Heyde (1980) Theorem 3.2) to conclude that the martingale array $\frac{2}{\sigma_{n,1}\sqrt{n}}\sum_{\ell=1}^{n}L_{n,\ell}$ converges weakly to $N(0,4)$. Note that $L_{n,\ell}=L_n(X_\ell,X_{\ell-1})$ and also that $\sigma_{n,1}^{-2}\int\{\mu P\}(dx,dy)L_n^2(x,y)=1$. To check the Lindeberg condition, we take $p>1$ as in the theorem and do

$$\sigma_{n,1}^{-2p}n^{-p}\sum_{\ell=1}^{n}\mathbb{E}\left(|L_{n,\ell}|^{2p}\right)\leq cn^{-p+1}\sigma_{n,1}^{-2p}\left\|h_{n,1}\right\|_{2p,V^\beta}^{2p}\to 0,$$

by (19). To check the law of large numbers, we notice that

$$\mathbb{E}\left(L_{n,\ell}^2|\mathcal{F}_{\ell-1}\right)=\int P(X_{\ell-1},dz)\left\{L_n(x,X_{\ell-1})\right\}^2\leq c\left\|h_{n,1}\right\|_{2,V^\beta}^2V^{2(\beta+\kappa)}(X_{\ell-1}).$$

Under A1, P satisfies $\mathsf{C}(V^{2(\beta+\kappa)},V^{2(\beta+\kappa)+\kappa})$ and by the remark following Proposition 7.1, $\sigma_{n,1}^{-2}n^{-1}\sum_{\ell=1}^{n}\mathbb{E}\left(L_{n,\ell}^2|\mathcal{F}_{\ell-1}\right)$ converges in probability to one. This ends the proof of the theorem.

□

6.5. **Proof Theorem 3.4.** Again by (15), we have:

$$\sigma_n^{-1}\binom{n}{2}^{-1/2}\left(U_n(h_n-\theta_n)\right)=\sum_{\ell=1}^{n}D_{n,\ell}+\sigma_n^{-1}\binom{n}{2}^{-1/2}\zeta_n, \tag{35}$$

where $D_{n,\ell}=\sigma_n^{-1}\binom{n}{2}^{-1/2}\left((n-1)L_{n,\ell}+\sum_{j=1}^{\ell-1}Q_{n,\ell,j}\right)$. Proposition 3.1 and (20-21) shows that $\sigma_n^{-1}\binom{n}{2}^{-1/2}\zeta_n$ converges in probability to zero. The bulk of the proof consists in showing that the martingale array $\{\sum_{\ell=2}^{k}D_{n,k},\ \mathcal{F}_k\}$ satisfies the conditions of the martingale CLT (Theorem 3.2 of Hall and Heyde (1980)). By (34),

$$\mathbb{E}\left(\sum_{\ell=1}^{n}D_{n,\ell}^2\right)=O\left(\frac{\left\|h_{n,1}\right\|_{2,V^\beta}^2}{\sigma_{n,1}^2}+\frac{\left\|h_n\right\|_{2,V_{(2)}^\beta}^2}{\sigma_{n,2}^2}+n^{-1/2}\frac{\left\|h_{n,1}\right\|_{2,V^\beta}}{\sigma_{n,1}}\frac{\left\|h_n\right\|_{2,V_{(2)}^\beta}}{\sigma_{n,2}}\right)=O(1),$$

by (20). Next, we show that $\max_{1\leq\ell\leq n}|D_{n,\ell}|$ converges in probability to zero. It is enough to show that $\sum_{\ell=1}^{n}\mathbb{E}\left(D_{n,\ell}^4\right)\to 0$. By convexity,

$$\mathbb{E}\left(D_{n,\ell}^4\right)\leq 2^3\sigma_n^{-4}\binom{n}{2}^{-2}\left(n^4\mathbb{E}\left(L_{n,\ell}^4\right)+\mathbb{E}\left[\left(\sum_{j=1}^{\ell-1}Q_{n,\ell,j}\right)^4\right]\right).$$

And $\sigma_n^{-4}\binom{n}{2}^{-2}n^4\sum_{\ell=1}^n \mathbb{E}\left(L_{n,\ell}^4\right) \leq cn^{-1}\sigma_{n,1}^{-4}\left\|h_{n,1}\right\|_{4,V^\beta}^4 \to 0$, by (21). For the second term we have

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^4\right] &= \sum_{j=1}^{\ell-1}\mathbb{E}\left(Q_{n,\ell,j}^4\right) + 6\sum_{1\leq j_1<j_2\leq \ell-1}\mathbb{E}\left(Q_{n,\ell,j_1}^2 Q_{n,\ell,j_2}^2\right)\\ &\quad + 4\sum_{1\leq j_1<j_2\leq \ell-1}\mathbb{E}\left(Q_{n,\ell,j_1}^3 Q_{n,\ell,j_2} + Q_{n,\ell,j_1}Q_{n,\ell,j_2}^3\right)\\ &+ 12\sum_{1\leq j_1<j_2<j_3\leq \ell-1}\mathbb{E}\left(Q_{n,\ell,j_1}^2 Q_{n,\ell,j_2}Q_{n,\ell,j_3} + Q_{n,\ell,j_1}Q_{n,\ell,j_2}^2 Q_{n,\ell,j_3} + Q_{n,\ell,j_1}Q_{n,\ell,j_2}Q_{n,\ell,j_3}^2\right)\\ &\quad + 24\sum_{1\leq j_1<j_2<j_3<j_4\leq \ell-1}\mathbb{E}\left(Q_{n,\ell,j_1}Q_{n,\ell,j_2}Q_{n,\ell,j_3}Q_{n,\ell,j_4}\right).\end{aligned}$$

And $\sum_{j=1}^{\ell-1}\mathbb{E}\left(Q_{n,\ell,j}^4\right) \leq cn\left\|h_n\right\|_{4,V_{(2)}^\beta}^4$. Using the same argument as in the proof of (34), it is easy (but a bit tedious) to show that

$$\sum_{1\leq j_1<j_2<j_3<j_4\leq \ell-1}\left|\mathbb{E}\left(Q_{n,\ell,j_1}Q_{n,\ell,j_2}Q_{n,\ell,j_3}Q_{n,\ell,j_4}\right)\right| \leq c\left\|h_n\right\|_{4,V_{(2)}^\beta}^4,$$

and that

$$\begin{aligned}&\sum_{1\leq j_1<j_2\leq \ell-1}\left|\mathbb{E}\left(Q_{n,\ell,j_1}^3 Q_{n,\ell,j_2} + Q_{n,\ell,j_1}Q_{n,\ell,j_2}^3\right)\right|\\ &+ \sum_{1\leq j_1<j_2<j_3\leq \ell-1}\left|\mathbb{E}\left(Q_{n,\ell,j_1}^2 Q_{n,\ell,j_2}Q_{n,\ell,j_3} + Q_{n,\ell,j_1}Q_{n,\ell,j_2}^2 Q_{n,\ell,j_3} + Q_{n,\ell,j_1}Q_{n,\ell,j_2}Q_{n,\ell,j_3}^2\right)\right|\\ &\leq cn\left\|h_n\right\|_{4,V_{(2)}^\beta}^4.\end{aligned}$$

Thus

$$\mathbb{E}\left[\left(\sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^4\right] = 6\sum_{1\leq j_1<j_2\leq \ell-1}\mathbb{E}\left(Q_{n,\ell,j_1}^2 Q_{n,\ell,j_2}^2\right) + O\left(n\left\|h_n\right\|_{4,V_{(2)}^\beta}^4\right). \qquad (36)$$

Then with (11),

$$\mathbb{E}\left(Q_{n,\ell,j_1}^2 Q_{n,\ell,j_2}^2\right) = \mathbb{E}\left(\bar{h}_{n,2}^2(X_\ell, X_{j_1})\bar{h}_{n,2}^2(X_\ell, X_{j_2})\right) + O\left(\left\|h_n\right\|_{4,V_{(2)}^\beta}^2\left\|h_n\right\|_{2,V_{(2)}^\beta}^2\right),$$

and

$$\mathbb{E}\left(\bar{h}_{n,2}^2(X_\ell, X_{j_1})\bar{h}_{n,2}^2(X_\ell, X_{j_2})\right) = \mathbb{E}\left(\int P^{j_2-j_1}(X_{j_1}, dx_{j_2}) \int \left(P^{\ell-1-j_2}(x_{j_2}, dx_{\ell-1}) - \mu(dx_{\ell-1})\right)\right.$$
$$\left. \times \int P(x_{\ell-1}, dx_\ell)\bar{h}_{n,2}^2(x_\ell, X_{j_1})\bar{h}_{n,2}^2(x_\ell, x_{j_2})\right)$$
$$+\mathbb{E}\left(\int \left(P^{j_2-j_1}(X_{j_1}, dx_{j_2}) - \mu(dx_{j_2})\right) \int \mu(dx_{\ell-1}) \int P(x_{\ell-1}, dx_\ell)\bar{h}_{n,2}^2(x_\ell, X_{j_1})\bar{h}_{n,2}^2(x_\ell, x_{j_2})\right)$$
$$+\mathbb{E}\left(\int \mu(dx)\bar{h}_{n,2}^2(x, X_{j_1}) \int \mu(dx_{j_2})\bar{h}_{n,2}^2(x, x_{j_2})\right)$$
$$\leq c(\rho(\ell-1-j_2)+\rho(j_2-j_1)) \left\|\bar{h}_{n,2}\right\|_{4,V_{(2)}^\beta}^4 + \mathbb{E}\left(\int \mu(dx)\bar{h}_{n,2}^2(x, X_{j_1}) \int \mu(dx_{j_2})\bar{h}_{n,2}^2(x, x_{j_2})\right),$$

for a summable series of positive numbers $\{\rho(n)\}$. The last inequality uses A2. Now, noting that

$$\mathbb{E}\left(\int \mu(dx)\bar{h}_{n,2}^2(x, X_{j_1}) \int \mu(dx_{j_2})\bar{h}_{n,2}^2(x, x_{j_2})\right) = \int \mu(dx)\left(\int \bar{h}_{n,2}^2(x,u)\mu(du)\right)^2$$
$$+ \int \mu(dx)\mathbb{E}\left(\bar{h}_{n,2}^2(x, X_{j_1}) - \int \bar{h}_{n,2}^2(x,u)\mu(du)\right) \int \mu(dx_{j_2})\bar{h}_{n,2}^2(x, x_{j_2}),$$

we conclude easily that there exists a finite constant c such that

$$\mathbb{E}\left[\left(\sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^4\right] \leq cn^2\left(\int \mu(dx)\left(\int \mu(dz)h_n^2(x,z)\right)^2 + \left\|h_n\right\|_{4,V_{(2)}^\beta}^2 \left\|h_n\right\|_{2,V_{(2)}^\beta}^2\right)$$
$$+ cn\left\|h_n\right\|_{4,V_{(2)}^\beta}^4$$
$$\leq cn^2\left(|\delta_n|_{V^{2\beta}}^2 + \left\|h_n\right\|_{4,V_{(2)}^\beta}^2 \left\|h_n\right\|_{2,V_{(2)}^\beta}^2\right) + cn\left\|h_n\right\|_{4,V_{(2)}^\beta}^4.$$

Thus under (21), we see that $\sum_{\ell=1}^n \mathbb{E}\left(D_{n,\ell}^4\right) \to 0$.

It remains to show that the law of large numbers holds, that is, $\sum_{\ell=2}^n D_{n,\ell}^2$ converges in probability to 1.

$$\sum_{\ell=2}^n D_{n,\ell}^2 = \sigma_n^{-2}\binom{n}{2}^{-1}\left(n-1)^2\sum_{\ell=1}^n L_{n,\ell}^2 + \sum_{\ell=2}^n\sum_{j=1}^{\ell-1} Q_{n,\ell,j}^2\right)$$
$$+ 2\sigma_n^{-2}\binom{n}{2}^{-1}\left((n-1)\sum_{\ell=2}^n L_{n,\ell}\sum_{j=1}^{\ell-1} Q_{n,\ell,j} + \sum_{\ell=2}^n\sum_{j_1=1}^{\ell-1}\sum_{j_2=1}^{j_1-1} Q_{n,\ell,j_1}Q_{n,\ell,j_2}\right).$$

We have:

$$\left(\sum_{\ell=2}^{n} L_{n,\ell}\sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^2 = \sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1} L_{n,\ell}^2 Q_{n,\ell,j}^2 + 2\sum_{\ell=2}^{n}\sum_{1\le j_1<j_2<\ell} L_{n,\ell}^2 Q_{n,\ell,j_1}Q_{n,\ell,j_2}$$
$$+ 2\sum_{1<\ell_1<\ell_2\le n}\sum_{j_1=1}^{\ell_1-1}\sum_{j_2=1}^{\ell_2-1} L_{n,\ell_1}L_{n,\ell_2}Q_{n,\ell_1,j_1}Q_{n,\ell_2,j_2}.$$

By considering separately the different cases $j_1 < j_2 < \ell_1 < \ell_2$, $j < \ell_1 < \ell_2$, $j_1 < \ell_1 = j_2 < \ell_2$, $j_1 < \ell_1 < j_2 < \ell_2$, and proceeding as in the proof of (34), we obtain

$$\mathbb{E}\left[\left(\sum_{\ell=2}^{n} L_{n,\ell}\sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^2\right] = \sum_{\ell=2}^{n}\sum_{j=1}^{\ell-1}\mathbb{E}\left(L_{n,\ell}^2 Q_{n,\ell,j}^2\right) + O\left(n^3 \|h_n\|_{4,V_{(2)}^\beta}^2 \|h_n\|_{2,V_{(2)}^\beta}^2\right).$$

Using (10) and (11), and the same calculations as done above for $\mathbb{E}\left(Q_{n,\ell,j_1}^2 Q_{n,\ell,j_2}^2\right)$, we get

$$\begin{aligned}
\mathbb{E}\left(L_{n,\ell}^2 Q_{n,\ell,j}^2\right) &= \mathbb{E}\left(\bar{h}_{n,1}^2(X_\ell)\bar{h}_{n,2}^2(X_\ell,X_j)\right)\\
&\quad + O\left(\|h_{n,1}\|_{4,V^\beta}^2 \|h_n\|_{2,V_{(2)}^\beta}^2 + \|h_{n,1}\|_{2,V^\beta}^2 \|h_n\|_{4,V_{(2)}^\beta}^2 + \|h_{n,1}\|_{2,V^\beta}^2 \|h_n\|_{2,V_{(2)}^\beta}^2\right)\\
&\le c\left(\|h_{n,1}\|_{4,V^\beta}^2 |\delta_n|_{V^{2\beta}} + \rho(j)\|h_{n,1}\|_{4,V^\beta}^2 \|h_n\|_{2,V_{(2)}^\beta}^2 + \rho(\ell-j-1)\|h_{n,1}\|_{4,V^\beta}^2 \|h_n\|_{4,V_{(2)}^\beta}^2\right)\\
&\quad + c\left(\|h_{n,1}\|_{4,V^\beta}^2 \|h_n\|_{2,V_{(2)}^\beta}^2 + \|h_{n,1}\|_{2,V^\beta}^2 \|h_n\|_{4,V_{(2)}^\beta}^2 + \|h_{n,1}\|_{2,V^\beta}^2 \|h_n\|_{2,V_{(2)}^\beta}^2\right),
\end{aligned}$$

for a summable series of positive numbers $\{\rho(n)\}$. It follows that

$$\sigma_n^4\binom{n}{2}^{-2}(n-1)^2\mathbb{E}\left[\left(\sum_{\ell=2}^{n} L_{n,\ell}\sum_{j=1}^{\ell-1} Q_{n,\ell,j}\right)^2\right] \le \sigma_n^{-4}\|h_{n,1}\|_{4,V^\beta}^2 |\delta_n|_{V^{2\beta}}$$
$$+ cn^{-1}\sigma_n^{-4}\|h_{n,1}\|_{4,V^\beta}^2 \|h_n\|_{4,V_{(2)}^\beta}^2 + c\sigma_n^{-4}\left(\|h_{n,1}\|_{4,V^\beta}^2 \|h_n\|_{2,V_{(2)}^\beta}^2 + \|h_{n,1}\|_{2,V^\beta}^2 \|h_n\|_{4,V_{(2)}^\beta}^2\right).$$

Under (21), we conclude that $2\sigma_n^{-2}\binom{n}{2}^{-1}(n-1)\sum_{\ell=2}^{n} L_{n,\ell}\sum_{j=1}^{\ell-1} Q_{n,\ell,j}$ converges in probability to zero. With the same arguments we obtain

$$\sigma_n^{-4}\binom{n}{2}^{-2}\mathbb{E}\left[\left(\sum_{\ell=2}^{n}\sum_{j_1=1}^{\ell-1}\sum_{j_2=1}^{j_1-1} Q_{n,\ell,j_1}Q_{n,\ell,j_2}\right)^2\right] \le cn^{-1}\sigma_n^{-4}\left(|\delta_n|_{V^{2\beta}}^2 + \|h_n\|_{4,V_{(2)}^\beta}^2 \|h_n\|_{2,V_{(2)}^\beta}^2\right)$$
$$+ cn^{-2}\sigma_n^{-4}\|h_n\|_{4,V_{(2)}^\beta}^4.$$

It follows that $\sigma_n^{-2}\binom{n}{2}^{-1}\sum_{\ell=2}^{n}\sum_{j_1=1}^{\ell-1}\sum_{j_2=1}^{j_1-1} Q_{n,\ell,j_1}Q_{n,\ell,j_2}$ converges in probability to zero.

For the remaining term in $\sum_{\ell=2}^n D_{n,\ell}^2$, we write:

$$\sigma_n^{-2}\binom{n}{2}^{-1}\left((n-1)^2\sum_{\ell=1}^n L_{n,\ell}^2+\sum_{\ell=2}^n\sum_{j=1}^{\ell-1}Q_{n,\ell,j}^2\right)=1+\frac{2}{n}\sum_{\ell=1}^n(n-1)\left(\sigma_n^{-2}L_{n,\ell}^2-\sigma_n^{-2}\sigma_{n,1}^2\right)$$
$$+\binom{n}{2}^{-1}\sum_{\ell=2}^n\sum_{j=1}^{\ell-1}\left(\sigma_n^{-2}Q_{n,\ell,j}^2-\sigma_n^{-2}\sigma_{n,2}^2\right).$$

We deal with these terms by working with the Markov chain $\{Y_n,\ n\geq 1\}$ where $Y_n=(X_{n-1},X_n)$. Clearly $\{Y_n,\ n\geq 1\}$ is a Markov chain with transition kernel $\bar{P}((u,v);(dx,dy))=\delta_v(dx)P(x,dy)$ and invariant distribution $\{\mu P\}$. $\bar{P}$ inherits the ergodicity of P and it is easy to check that if P satisfies A2, $\bar{P}$ also satisfies A2 with respect to the function $V_1(x,y)=V(x)+V(y)$. Set $\bar{\beta}=2(\beta+\kappa)$. By (10), $L_n^2(x,y)=(g_n(x)-Pg_n(y))^2$ belongs to $\mathcal{L}_{V_1^{\bar{\beta}}}$ and for $p\geq 1$, $\left\|\left|L_n^2\right|\right\|_{p,V_1^{\bar{\beta}}}\leq c\left\|h_{n,1}\right\|_{2p,V^\beta}^2$, for some finite constant c, where the norm $\left\|\left|L_n^2\right|\right\|_{p,V_1^{\bar{\beta}}}$ is taken wrt the kernel $\bar{P}$. With $p=2$, we have $n^{-1+\frac{1}{p}}\left\|\left|(n-1)\sigma_n^{-2}L_n^2\right|\right\|_{p,V_1^{\bar{\beta}}}\leq cn^{-1/2}\sigma_{n,1}^{-2}\left\|h_{n,1}\right\|_{4,V^\beta}^2\to 0$, by (21). From (the remark following) Proposition 7.1, we conclude that the term $\frac{1}{n}\sum_{\ell=1}^n(n-1)\left(\sigma_n^{-2}L_{n,\ell}^2-\sigma_n^{-2}\sigma_{n,1}^2\right)$ converges in probability to zero.

Recall that $Q_{n,\ell,j}=Q_n(X_\ell,X_{\ell-1},X_j,X_{j-1})=Q_n(Y_\ell,Y_j)=Q_n(Y_j,Y_\ell)$. Recall also that $\sigma_{n,2}^2=\int\{\mu P\}(dy_1)\int\{\mu P\}(dy_2)Q_n^2(y_1,y_2)$. Therefore we can handle the term $\binom{n}{2}^{-1}\sum_{\ell=2}^n\sum_{j=1}^{\ell-1}\sigma_n^{-2}\left(Q_{n,\ell,j}^2-\sigma_{n,2}^2\right)$ as a centered U-statistics in $\{Y_n,\ n\geq 1\}$ and apply Theorem 3.2-(2). Set $\bar{V}_{1,(2)}(x_\ell,x_{\ell-1},x_j,x_{j-1})=\bar{V}_1(x_\ell,x_{\ell-1})\bar{V}_1(x_j,x_{j-1})$. We check that $Q_n^2\in\mathcal{L}_{\bar{V}_{1,(2)}^{\bar{\beta}}}$ and that $\left\|\left|Q_n^2\right|\right\|_{p,\bar{V}_{1,(2)}^{\bar{\beta}}}\leq c\left\|h_n\right\|_{2p,V_{(2)}^\beta}^2$, provided $2p(\beta+\kappa)\leq 1$. Also, for $Q_{n,1}^2(u,v):=\int\mu(dz_1)\int P(z_1,dz_2)Q_n^2(z_2,z_1,u,v)$, we check that

$$\left\|\left|Q_{n,1}^2\right|\right\|_{2,\bar{V}_1^{\bar{\beta}}}\leq c\left(|\delta_n|_{V^{2\beta}}^{1/2}+|Ph_n|_{V_{(2)}^\beta}\right)^2,$$

provided $4(\beta+\kappa)\leq 1$, where $\delta_n(x)=\int\mu(dz)h_n^2(z,x)$. It follows that

$$n^{-1}\left(\sqrt{n}\left\|\left|\sigma_n^{-2}Q_{n,1}^2\right|\right\|_{2,\bar{V}_1^{\bar{\beta}}}+\left\|\left|\sigma_n^{-2}Q_n^2\right|\right\|_{2,\bar{V}_{1,(2)}^{\bar{\beta}}}\right)$$
$$\leq n^{-1/2}\sigma_n^{-2}|\delta_n|_{V^{2\beta}}+n^{-1/2}\sigma_n^{-2}|Ph_n|_{V_{(2)}^\beta}^2+n^{-1}\sigma_n^{-2}\left\|h_n\right\|_{4,V_{(2)}^\beta}^2\to 0,$$

by (20) and (21). By Theorem 3.2-(2), $\binom{n}{2}^{-1}\sum_{\ell=2}^n\sum_{j=1}^{\ell-1}\left(\sigma_n^{-2}Q_{n,\ell,j}^2-\sigma_n^{-2}\sigma_{n,2}^2\right)$ converges in probability to zero.

□

6.6. **Proof of Theorem 3.5.** The proof employs Theorem 3.4 and the following two lemmas. Recall that $V(x)=V_z(z)+|w|^8+|y|^8$, and we set $\beta=1/8$.

Lemma 6.1. (i) $\theta_n = \theta + O(b_n^{P\wedge Q})$*;* (ii) $\|h_{n,1}\|^p_{p,V^\beta} = O(1)$ *for* $p \in [1,4]$*;* (iii) $\|h_n\|^p_{p,V^\beta_{(2)}} = O(b_n^{-(p-1)d-p|s|})$ *for* $p \in [1,4]$*; and* (iv) $\sup_x V^{-2\beta}(x)\int \mu(dx_2)h_n^2(x,x_2) = O(b_n^{-d-2|s|})$*.*

Proof. For (i), change of variables, integration by parts, and a Taylor's series expansion give

$$\begin{aligned}\theta_n &= \int\int h_n(x_1,x_2)\mu(dx_1)\mu(dx_2) = \int\int K_{s,b_n}(z_1-z_2)g_w(z_2)g_y(z_1)\mu(dx_1)\mu(dx_2)\\ &= \int\left[\int K(u)\partial^s e_w(z_2-ub_n)du\right]e_y(z_2)dz_2 = \theta + O(b_n^{P\wedge Q}).\end{aligned}$$

For (ii), elementary bounds, change of variables and integration by parts give

$$\begin{aligned}\|h_{n,1}\|^p_{p,V^\beta} \le c\sup_{x_1}\frac{\int P(x_1,dx_2)(|y_2|^p+|w_2|^p)}{V^{\beta p}(x_1)}&\\ \le c\sup_z V^{-p\beta}(z)\int Q_z(z,dz_1)\left(m_{y,p}(z_1)+m_{w,p}(z_1)\right) < \infty,&\end{aligned}$$

under the assumptions imposed.

For (iii), elementary bounds and change of variables give

$$\begin{aligned}\int P(x_1,dx_2)\,|h_n(x_2,x_3)|^p \le \frac{c|w_3|^p}{b_n^{(p-1)d+p|s|}}\int|\partial^s K(u)|^p m_{y,p}(z_2-ub_n)q_z(z_1,z_2-ub_n)du&\\ +\frac{c|y_3|^p}{b_n^{(p-1)d+p|s|}}\int|\partial^s K(u)|^p m_{w,p}(z_2-ub_n)q_z(z_1,z_2-ub_n)du,&\\ \le cb_n^{-(p-1)d-p|s|}V^{p/8}(z_1)\left(|w_3|^p+|y_3|^p\right).&\end{aligned}$$

and therefore

$$\|h_n\|^p_{p,V^\beta_{(2)}} \le \frac{c}{b_n^{(p-1)d+p|s|}}\sup_{x_1,x_3}\frac{V^{p/8}(z_1)}{V^{\beta p}(x_1)}\left(\frac{|w_3|^p}{V^{\beta p}(x_3)}+\frac{|y_3|}{V^{\beta p}(x_3)}\right) < \infty.$$

For (iv), by elementary bounds and change of variables

$$\begin{aligned}\int\mu(dx_2)h_n^2(x_1,x_2) &\le \frac{cw_1^2}{b_n^{d+2|s|}}\int(\partial^s K(u))^2 e_y(z_1-ub_n)du\\ &\quad+\frac{cy_1^2}{b_n^{d+2|s|}}\int(\partial^s K(u))^2 e_w(z_1-ub_n)du,\end{aligned}$$

and the result follows.

□

Note also that $|P\bar{h}_{n,1}|_{V^\beta} \le \left\|\bar{h}_{n,1}\right\|_{p,V^\beta}$ and $|P\bar{h}_{n,2}|_{V^\beta_{(2)}} \le \left\|\bar{h}_{n,2}\right\|_{p,V^\beta_{(2)}}$ for all $p \ge 1$. Thus, Lemma 6.1(ii) implies $|P\bar{h}_{n,1}|^p_{V^\beta} = O(1)$ and Lemma 6.1(iii) implies $|P\bar{h}_{n,2}|^p_{V^\beta_{(2)}} = O(b_n^{-p|s|})$.

Lemma 6.2. (i) $\sigma^2_{n,1}\to\sigma^2_1$*; and* (ii) $b_n^{d+2|s|}\sigma^2_{n,2}\to\sigma^2_2$*.*

Proof. For (i), first let $\varsigma(x) = \bar{h}_{n,1}(x) - \psi(x)/2$, and note that by previous calculations $\varsigma(x) \le c b_n^{\varrho\wedge Q}(|y|+|w|)$. Next define $L(x_1,x_2) = L\psi(x_1,x_2) = \int \Lambda(x_1,x_2;ds)\psi(s)$ and note that, using (10), we obtain $\int\{\mu P\}(dx_1,dx_2)\,(L_n(x_1,x_2) - L(x_1,x_2)/2)^2 = O(b_n^{2(\varrho\wedge Q)})$. Therefore, since $\sigma_{n,1}^2 = \int\{\mu P\}(dx_1,dx_2)L_n^2(x_1,x_2)$, the result follows by Cauchy-Schwarz Inequality.

For (ii), the result follows by analogous arguments to those given in part (i). □

To complete the proof first note that $\Sigma_n^{-1} = O(\min(n^{1/2}, nb_n^{d/2+|s|}))$, which gives

$$\Sigma_n^{-1}(\theta_n - \theta) = O(\min(1, n^{1/2}b_n^{d/2|s|})n^{1/2}b_n^{\varrho\wedge Q}) \to 0.$$

Therefore, it is suffices to analyze

$$\Sigma_n^{-1}(\hat{\theta}_n - \theta_n) = \sigma_n^{-1}\binom{n}{2}^{-1/2}\left(U_n(h_n) - \binom{n}{2}\theta_n\right),$$

where $\sigma_n^{-2} = O(\min(n^{-1} + b_n^{d+2|s|}))$, which may be easily handled by Theorem 3.4. Specifically, the assumptions imposed together with Lemmas 6.1 and 6.2 imply:

$$\left|\!\left|\!\left|\bar{h}_{n,1}\right|\!\right|\!\right|_{2,V^\beta}^2 = O(\sigma_{n,1}^2), \qquad \left|\!\left|\!\left|\bar{h}_{n,2}\right|\!\right|\!\right|_{2,V_{(2)}^\beta}^2 = O(\sigma_{n,2}^2),$$

$$\sigma_n^{-1}|P\bar{h}_{n,2}|_{V_{(2)}^\beta} = O(\min(n^{-1/2} + b_n^{d/2+|s|}))O(b_n^{-|s|}) \to 0,$$

$$n^{-1}\sigma_{n,1}^{-4}\left|\!\left|\!\left|\bar{h}_{n,1}\right|\!\right|\!\right|_{4,V^\beta}^4 = O(n^{-1}) \to 0,$$

$$n^{-2}\sigma_{n,2}^{-4}\left|\!\left|\!\left|\bar{h}_{n,2}\right|\!\right|\!\right|_{4,V_{(2)}^\beta}^4 = n^{-2}O(b_n^{2d+4|s|})O(b_n^{-3d-4|s|)}) = O(n^{-2}b_n^{-d}) \to 0,$$

$$\sigma_{n,2}^2|\delta_n|_{V^{2\beta}} = O(b_n^{d+2|s|})O(b_n^{-d-2|s|}) = O(1).$$

where $\delta_n(x) = \int \mu(dx_2)h_n^2(x,x_2)$. Therefore, the continuous mapping theorem and Theorem 3.4 give the results.

□

6.7. **Proof theorem 4.1.** A2 implies that we can find $p > 1$ such that

$$\sup_{n\ge 0}\mathbb{E}\left(V_2^{2p}(X_n)\right) < \infty. \tag{37}$$

The p-moment of ζ_n is (28). By martingale approximation for linear partial sums (see e.g. the proof of Proposition 7.1 below), for any sequence of real numbers $\{a_{n,\ell},\ 1 \le \ell \le n\}$,

$$\sum_{\ell=1}^n a_{n,\ell}h(X_\ell) = \sum_{\ell=1}^n a_{n,\ell}Q_\ell + \epsilon_{n,1}, \quad \text{where} \quad \mathbb{E}\left(\left|\sum_{\ell=1}^n a_{n,\ell}Q_\ell\right|^\alpha\right) \le c\left(\sum_{\ell=1}^n |a_{n,\ell}|^{\alpha\wedge 2}\right)^{1\vee\frac{\alpha}{2}},$$

$$\text{and } \mathbb{E}\left(|\epsilon_{n,1}|^\alpha\right) \le c\left(|a_{n,1}| + |a_{n,n}| + \sum_{\ell=2}^n |a_{n,\ell} - a_{n,\ell-1}|\right)^\alpha, \tag{38}$$

provided $\sup_{n\geq 0}\mathbb{E}\left(V_2^{\alpha}(X_n)\right)<\infty$. We use (38) to bound (26) and obtain for all $n\geq 1$:

$$\mathbb{E}\left(|R_n|^p\right)\leq cn^{-p}c_n^p,$$

for some finite constant c. Hence, under the assumption $c_n=o(n)$, R_n converges in probability to zero.

By standard martingale inequalities, we obtain the bound

$$\mathbb{E}\left(\left|\sum_{\ell=1}^n\sum_{j=1}^{\ell-1}w_{n,b}(\ell-j)Q_\ell Q_j\right|^p\right)\leq c\left(\frac{c_n}{n}\right)^{\frac{p}{2}}n^{-\frac{p}{2}+1\vee\frac{p}{2}}.$$

Thus if $c_n=o(n)$, and $p\geq 2$, that term vanishes as well. Given the ergodicity assumption $\mathsf{C}(V_2^2,V_3)$ and $\sup_{k\geq 0}\mathbb{E}(V_3^q(X_k))<\infty$, it is easy to show that the term $n^{-1}\sum_{\ell=1}^n Q_\ell^2$ converges in probability to the limit

$$\int\mu(dx)\int P(x,dy)\left(G(y)-PG(x)\right)^2,$$

which is easily seen to be equal to $\sigma^2(h)$. This proves the first part of the theorem.

From now on, we assume that $c_n=n$. Define $W_{n,\ell}=\frac{Q_\ell}{\sqrt{n}\sigma(h)}$. Then by (38) with $a_{n,\ell}\equiv\frac{1}{\sqrt{n}\sigma(h)}$,

$$\sum_{\ell=1}^n\frac{h(X_\ell)}{\sqrt{n}\sigma(h)}=\sum_{\ell=1}^n W_{n,\ell}+\epsilon_{n,2},\quad\text{where}\quad \epsilon_{n,2}\text{ converges in probability to zero}$$

Define $\lfloor x\rfloor$ as the largest integer smaller or equal to x and for $0\leq t\leq 1$, we introduce

$$B_n(t)=\sum_{\ell=1}^{\lfloor nt\rfloor}W_{n,\ell},\ \text{ and }\ Z_n(t)=\int_0^t w_b(t-u)dB_n(u).$$

Since B_n has jumps only at times $\ell/n=\ell/c_n$, we see that $Z_n(\ell c_n^{-1})=\sum_{j=0}^{\ell-1}w_b((\ell-j)c_n^{-1})W_{n,j+1}$. It is also easy to see that the term R_n in (26) can be written as

$$R_n=2B_n^2(1)\int_0^1(1-u)w_b(u)du-2B_n(1)\int_0^1\left(\int_0^t w_b(u)du+\int_0^{1-t}w_b(u)du\right)dB_n(t)+\epsilon_{n,3},$$

where $\epsilon_{n,3}$ converges in probability to zero. Thus

$$\begin{aligned}\Gamma^2_{n,b}(h)&=n^{-1}\sum_{\ell=1}^n Q_\ell^2+2\sigma^2(h)\sum_{\ell=1}^n\frac{Q_\ell}{\sigma(h)\sqrt{n}}\sum_{j=1}^{\ell-1}w_b\left(\frac{\ell-j}{c_n}\right)\frac{Q_j}{\sigma(h)\sqrt{n}}+R_n+\zeta_n\\
&=\sigma^2(h)\sum_{\ell=1}^n W_{n,\ell}^2+2\sigma^2(h)\sum_{\ell=1}^n W_{n,\ell}Z_n\left((\ell-1)c_n^{-1}\right)+R_n+\zeta_n\\
&=\sigma^2(h)\sum_{\ell=1}^n W_{n,\ell}^2+2\sigma^2(h)\int_0^1 Z_n(t)dB_n(t)+2B_n^2(1)\sigma^2(h)\int_0^1 w_b(u)(1-u)du\\
&\qquad-2B_n(1)\sigma^2(h)\int_0^1 g_b(u)dB_n(t)+\epsilon_{n,4},\end{aligned}$$

where $g_b(u)=\int_0^t w_b(u)du+\int_0^{1-t}w_b(u)du$, and $\epsilon_{n,4}$ converges in probability to zero.

From the assumptions, $\sup_{\ell\geq 0}\mathbb{E}\left(|Q_\ell|^{2+\epsilon}\right) < \infty$, for some $\epsilon > 0$. Therefore, by the functional central limit theorem for martingales, $B_n \to_d B$, where $B = \{B(t),\ 0 \leq t \leq 1\}$ is the standard Brownian motion. By the continuous mapping theorem, $(B_n, Z_n) \to_d (B, Z)$, where $Z(t) = \int_0^t w(t-u)dB(u)$. And by the weak convergence of stochastic integrals (see, e.g., Theorem 2.2 in Kurtz and Protter (1991)),

$$\left\{\left(B_n(t), \int_0^t Z_n(t)dB_n(t), \int_0^1 g(u)dB_n(u), \sum_{\ell=1}^n W_{n,\ell}^2\right), 0 \leq t \leq 1\right\}$$

converges weakly to the stochastic process

$$\left\{\left(B(t), \int_0^t Z(u)dB(u), \int_0^1 g(u)dB(u), 1\right),\ 0 \leq t \leq 1\right\}.$$

As the remainders $(\epsilon_{n,2}, \epsilon_{n,4})$ converges in probability to 0, this entails that $\left(\sum_{\ell=1}^n \frac{h(X_\ell)}{\sqrt{n}\sigma(h)}, \Gamma_n^2(h)\right)$ converges weakly to the limit

$$\left(B(1), \sigma^2(h)\left(1 + 2\int_0^1 Z(u)dB(u) + 2B^2(1)\int_0^1 (1-u)w(u)du - 2B(1)\int_0^1 g(u)dB(t)\right)\right).$$

The conclusion of the theorem follows by the continuous mapping theorem.

□

7. Appendix: some technical results

Proposition 7.1. *Let $\{X_n,\ n \geq 0\}$ be a Markov chain with invariant distribution μ and transition kernel P. Suppose that there exist measurable functions $V_1 \leq V_2 :\ \mathcal{X} \to [1, \infty)$ such that*

$$\sum_{k\geq 0} \|P^k(x,\cdot) - \mu\|_{V_1} \leq cV_2(x),\ x \in \mathcal{X}, \tag{39}$$

for some finite constant c. Suppose also that $\mathsf{v}_n := \mathbb{E}(V_2^p(X_n)) < \infty$ for each $n \geq 0$ and for some $p \in (1, 2]$. Let $\{f_n,\ n \geq 1\}$ be such that $f_n, Pf_n \in \mathcal{L}_{V_1}$ and let $\{a_{n,k}, 0 \leq k \leq n\}$ be a sequence of real numbers such that

$$|\!|\!| f_n |\!|\!|_{p,V_1}^p \left(\sum_{k=1}^n |a_{n,k}|\right)^{-p} \sum_{k=1}^n |a_{n,k}|^p \mathsf{v}_k^p \to 0,$$

$$\text{and}\quad |Pf_n|_{V_1}\left(\sum_{k=1}^n |a_{n,k}|\right)^{-1} \sum_{k=1}^n |a_{n,k} - a_{n,k-1}|\mathsf{v}_{k-1} \to 0.$$

Then, as $n \to \infty$, $\left(\sum_{k=1}^n |a_{n,k}|\right)^{-1} \sum_{k=1}^n a_{n,k}\left(f_n(X_k) - \mu(f_n)\right)$ converges in probability to zero.

Proof. Define $S_n = \sum_{k=1}^n a_{n,k}\left(f_n(X_k) - \mu(f_n)\right)$ and $g_n(x) = \sum_{j\geq 0}(P^j f_n(x) - \mu(f_n))$. Under (39), $|g_n(x)| \leq c|f_n|_{V_1}V_2(x)$ and $|Pg_n(x)| \leq c|Pf_n|_{V_1}V_2(x)$. By the Poisson equation,

$f_n(x) - \mu(f_n) = g_n(x) - Pg_n(x)$ which implies that

$$S_n = \sum_{k=1}^n a_{n,k}\left(g_n(X_k) - Pg_n(X_{k-1})\right) + \sum_{k=1}^n (a_{n,k} - a_{n,k-1}) Pg_n(X_{k-1}) \\ + \left(a_{n,0} Pg_n(X_0) - a_{n,n} Pg_n(X_n)\right).$$

where the martingale array $\sum_{k=1}^n a_{n,k}\left(g_n(X_k) - Pg_n(X_{k-1})\right)$ satisfies

$$\mathbb{E}\left(|\sum_{k=1}^n a_{n,k}\left(g_n(X_k) - Pg_n(X_{k-1})\right)|^p \right) \leq C \, \|\!| f_n |\!\|_{p,V_1}^p \sum_{k=1}^n |a_{n,k}|^p \mathsf{v}_k^p.$$

The last inequality follows by noting that $g_n(x) - Pg_n(y) = f_n(x) - \mu(f_n) - Pg_n(x) - Pg_n(y)$ and by conditioning on $\mathcal{F}_{k-1}$. Thus, under the stated assumptions, $\left(\sum_{k=1}^n |a_{n,k}|\right)^{-1} S_n$ converges in probability to zero. □

Remark 7. An important special case is the case where $a_{n,\ell} = 1$ and $\sup_{n\geq 0} \mathbb{E}\left(V_2^p(X_n)\right) < \infty$. In this case it is enough to have $n^{-1+1/p} \, \|\!| f_n |\!\|_{p,V_1} \to 0$. If in addition it is true that $\sup_{x\in\mathcal{X}} PV_1^p(x)/V_1^p(x) < \infty$, then clearly $\|\!| f_n |\!\|_{p,V_1} \leq c|f_n|_{V_1}$ and the law of large number holds if $n^{-1+1/p}|f_n|_{V_1} \to 0$.

Acknowledgments: This work is partly supported by NSF grant DMS 0906631 (Atchade) and NSF grant SES 0921505 (Cattaneo). We are grateful to Michael Jansson, Lutz Kilian and Yizao Wang for helpful discussions, comments and suggestions.

References

Aradillas-Lopéz, A., Honoré, B. E. and Powell, J. L. (2007). Pairwise difference estimation with nonparametric control variables. *International Economic Review*, **48** 1119–1158.

Atchade, Y. (2010). Kernel estimators of asymptotic variance for adaptive Markov Chain Monte Carlo. *Annals of Statistics (to appear).*

Bhansali, R. J., Giraitis, L. and Kokoszka, P. S. (2007). Approximations and limit theory for quadratic forms of linear processes. *Stochastic Process. Appl.*, **117** 71–95.

Cattaneo, M. D., Crump, R. K. and Jansson, M. (2010). Robust data-driven inference for density-weighted average derivatives. *Journal of the American Statistical Association*, **105** 1070–1083.

Cattaneo, M. D., Crump, R. K. and Jansson, M. (2011a). Bootstrapping density-weighted average derivatives. Working paper, University of Michigan.

Cattaneo, M. D., Crump, R. K. and Jansson, M. (2011b). Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory (to appear).*

Cattaneo, M. D., Jansson, M. and Newey, W. K. (2011c). Alternative asymptotics and the partially linear model with many regressors. Working paper, University of Michigan.

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

Cheng, B. and Robinson, P. M. (1994). Semiparametric estimation from time series with long-range dependence. *Journal of Econometrics*, **64** 335–353.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, **75** 261–277.

Deaton, A. and Ng, S. (1998). Parametric and nonparametric approaches to price and tax reform. *Journal of the American Statistical Association*, **93** 900–909.

Dehling, H. and Wendler, M. (2010). Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *J. Multivariate Anal.*, **101** 126–137.

Douc, R., Fort, G., Moulines, E. and Soulier, P. (2004). Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, **14** 1353–1377.

Eagleson, G. K. (1979). Orthogonal expansions and U-statistics. *Austral. J. Statist.*, **21** 221–237.

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov Chain Monte Carlo. *Ann. Statist.*, **38** 1034–1070.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian data analysis.* 2nd ed. Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, FL.

Gordin, M. I. (1969). The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR*, **188** 739–741.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit theory and its application.* Academic Press, New York.

Härdle, W., Hildenbrand, W. and Jerison, M. (1991). Empirical evidence on the law of demand. *Econometrica*, **59** 1525–1549.

Hristache, M., Juditsky, A. and Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-index model. *Annals of Statistics*, **29** 595–623.

Hsing, T. and Wu, W. B. (2004). On weighted U-statistics for stationary processes. *Ann. Probab.*, **32** 1600–1631.

Jammalamadaka, S. R. and Janson, S. (1986). Limit theorems for a triangular scheme of U-statistics with applications to inter-point distances. *Ann. Probab.*, **14** 1347–1358.

Jansson, M. (2004). The error in rejection probability of simple autocorrelation robust tests. *Econometrica*, **72** 937–946.

KIEFER, N. M. and VOGELSANG, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, **21** 1130–1164.

KURTZ, T. and PROTTER, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability*, **19** 1035–1070.

MAXWELL, M. and WOODROOFE, M. (2000). Central limit theorems for additive functional of markov chains. *Annals of Probability*, **28** 713–724.

MEITZ, M. and SAIKKONEN, P. (2008). Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory*, **24** 1291–1320.

MEYN, S. P. and TWEEDIE, R. L. (2009). *Markov chains and stochastic stability.* Cambridge University Press; 2 edition, London.

NEWEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, **62** 1349–1382.

NEWEY, W. K., HSIEH, F. and ROBINS, J. M. (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica*, **72** 947–962.

POWELL, J. L., STOCK, J. H. and STOKER, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, **57** 1403–1430.

POWELL, J. L. and STOKER, T. M. (1996). Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics*, **75** 291–316.

PRIESTLEY, M. B. (1981). *Spectral analysis and time series. Vol. 1.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London. Univariate series, Probability and Mathematical Statistics.

ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.*, **1** 20–71 (electronic). URL `http://dx.doi.org.proxy.lib.umich.edu/10.1214/154957804100000024`.

ROBINSON, P. M. (1989). Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies*, **56** 511–534.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

STOKER, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica*, **54** 1461–1481.

STOKER, T. M. (1989). Tests of additive derivative constraints. *Review of Economic Studies*, **56** 535–552.

SUN, Y., PHILLIPS, P. C. B. and JIN, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, **76** 175–194.

WU, W. B. and SHAO, X. (2007). A limit theorem for quadratic forms and its applications. *Econometric Theory*, **23** 930–951.

YOSHIHARA, K.-I. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **35** 237–252.