Ann Inst Stat Math (2016) 68:301–327
DOI 10.1007/s10463-014-0489-2

Kernel estimators of mode under ψ-weak dependence

Eunju Hwang · Dong Wan Shin

Received: 16 September 2011 / Revised: 30 May 2014 / Published online: 14 October 2014

Abstract Nonparametric kernel-type estimation is discussed for modes which maximize nonparametric kernel-type density estimators. The discussion is made under a weak dependence condition which unifies weak dependence conditions such as mixing, association, Gaussian sequences and Bernoulli shifts. Consistency and asymptotic normality are established for the mode estimator as well as for kernel estimators of density derivatives. The convergence rate of the mode estimator is given in terms of the bandwidth. An optimal bandwidth selection procedure is proposed for mode estimation. A Monte-Carlo experiment shows that the proposed bandwidth yields a substantially better mode estimator than the common bandwidths optimized for density estimation. Modes of log returns of Dow Jones index and foreign exchange rates of US Dollar relative to Euro are investigated in terms of asymmetry.

Keywords Weak dependence · Kernel estimator · Mode · Consistency · Asymptotic normality · Bandwidth · Asymmetry

1 Introduction

We are interested in estimation of the mode based on nonparametric kernel-type density estimators under a very general dependence condition. Doukhan and Louhichi

E. Hwang
Department of Applied Statistics, Gachon University, SeongNamDaeRo 1342, SuJungGu, Seongnam, Gyeonggi-do, South Korea

D. W. Shin (✉)
Department of Statistics, Ewha Womans University, EwhaYoeDaeGil 52, SeoDaeMunGu, Seoul, South Korea
e-mail: shindw@ewha.ac.kr

(1999) have proposed a weak dependence condition, called ψ-weak dependence, for stationary processes that unifies weak dependence conditions, such as mixing, association, Gaussian sequences and Bernoulli shifts. Essentially all classes of stationary processes of real applications are in the class of ψ-weakly dependent processes as claimed by Ango Nze et al. (2002) and others. Ango Nze and Doukhan (2004) showed that stationary ARMA processes, and important nonlinear processes such as bilinear processes, threshold autoregressive processes, and GARCH processes are examples of ψ-weakly dependent processes, by showing that they are all Markovian processes which in turn are Bernoulli shifts.

Because of their wide range of applicability, there have been recently many studies on probabilistic properties of ψ-weakly dependent processes as given by Coulon-Prieur and Doukhan (2000), Dedecker and Prieur (2004), Kallabis and Neumann (2006), and Doukhan and Neumann (2007). Statistical applications were made by Doukhan and Louhichi (2001) for asymptotics of kernel density estimators, Dedecker et al. (2007) for functional estimation and spectral estimation, Doukhan et al. (2009) for least squares estimation of ARCH(∞) among the ψ-weakly dependent processes and Hwang and Shin (2012) for stationary bootstrapping of kernel density estimators.

As for the mode estimator, Parzen (1962) pioneered a work on the estimation of modes. Convergence rates of the mode estimators were discussed by Eddy (1980, 1982), Romano (1988), Vieu (1996), Herrmann and Ziegler (2004), and Shi et al. (2009a) under i.i.d. assumptions. In particular, Romano (1988) considered data-dependent bandwidth to obtain optimal rates by minimizing the mean-squared error of mode estimator. After Romano (1988), several authors such as Vieu (1996), Mokkadem and Pelletier (2005) and Ferraty et al. (2006) dealt with the density mode estimations. Recently, Shi et al. (2009a) improved the convergence rates of the mode estimators of Vieu (1996), Mokkadem and Pelletier (2005) and Ferraty et al. (2006) by establishing a relationship between the convergence rate of the mode estimator and the bandwidth in light of Shi et al. (2009b), which studied strong convergence rate of change point estimator. Asymptotic normality of mode estimators is related with that of density derivatives for which we know the results of Singh (1977) and Gasser and Müller (1979) under i.i.d. samples, Györfi et al. (1989) and Rio (1997) under strong mixing conditions, Jones (1994) for comparative studies, and Louani (1998) for randomly right censorship models under i.i.d. errors.

Recently, some studies for mode estimation for dependent data have been made. Wieczorek and Ziegler (2010) considered problems of optimal estimation of a non-smooth mode in the fixed-design regression model with α-mixing errors. Meister (2011) studied consistency of deconvolution mode estimation for contaminated data where no smoothness assumption on the underlying curve is needed.

As for the conditional mode estimator, Samanta and Thavaneswaran (1990) dealt with nonparametric estimation of the conditional mode, and Louani and Ould-Saïd (1999) provided asymptotic normality for the kernel estimators of the conditional mode under strong mixing condition.

This paper considers mode estimation based on kernel type density estimators for a sequence of strictly stationary ψ-weakly dependent processes. We establish consistency and asymptotic normality of the mode estimators. The convergence rate of the mode is given in terms of the bandwidth. The consistency result extends the rate result of Shi et al. (2009a) for i.i.d. cases to the much more general ψ-weakly dependent cases. Results for kernel type density derivative estimators are obtained as byproducts.

This asymptotic analysis provides us an asymptotically optimal bandwidth in terms of the true mode. This enables us to develop an asymptotically optimal bandwidth selection procedure. The unknown true mode is first estimated using an existing bandwidth optimized for density estimation by, for example, Sheater and Jones (1991), Rudemo (1982) and Bowman (1984), and then plugged in the optimal bandwidth formula. This gives a data-dependent bandwidth optimized for mode estimator.

A Monte-Carlo experiment compares the mode estimator computed using the proposed bandwidth with that computed using the bandwidth optimized for density estimation. The proposed estimator is less-biased and has substantially smaller mean-squared error (MSE). The experiment also investigates favorably finite sample normality of a t statistic for the significance of the mode.

The proposed method is applied to daily log returns of US Dow Jones index and foreign exchange rates of US Dollar relative to Euro. Our results enable us to conduct statistical tests regarding modes. The test for Dow Jones reveals that the mode is significantly different from zero implying that the distribution of log return is not symmetric about 0 and thus positive log returns have probabilistic features different from those of negative log returns. On the other hand, the test for exchange rates shows no significant mode.

The remaining of the paper is organized as follows. In Sect. 2, ψ-weakly dependent processes are defined. In Sect. 3, main results of consistency and asymptotic normalities are stated, and in Sect. 4 bandwidth selection for the mode estimator is discussed. Section 5 presents small sample analysis for optimal bandwidth and Sect. 6 provides a real data set analysis. Section 7 concludes the paper and the last section gives an appendix containing results for density derivatives ("Large sample results for density derivative estimators") and proofs of the results ("Proofs").

2 ψ-weakly dependent processes

The definition of ψ-weak dependence makes explicit the asymptotic independence between "past" and "future". In terms of the time series, for convenient functions g and h, it is assumed that $\mathrm{Cov}(g_{\mathrm{past}}, h_{\mathrm{future}})$ is small when the distance between the "past" and the "future" is sufficiently large. Asymptotics are expressed in terms of the distance between indices of the initial time series in the "past" and the "future" terms; the convergence is not assumed to hold uniformly on the dimension of the marginal involved as seen in (1) of the definition below.

To define the notion of weak dependence, we first introduce some classes of functions. Let $\mathbb{L}^\infty = \bigcup_{n=1}^{\infty} \mathbb{L}^\infty(\mathbb{R}^n)$, the set of real-valued and bounded functions on the space $\mathbb{R}^n$ for $n = 1, 2, \ldots$. Consider a function $g : \mathbb{R}^n \to \mathbb{C}$ where $\mathbb{R}^n$ is equipped with its l_1-norm (i.e. $||(x_1, \ldots, x_n)||_1 = |x_1| + \cdots + |x_n|$) and define the Lipschitz modulus of g,

$$\mathrm{Lip}(g) = \sup_{x \neq y} \frac{|g(x) - g(y)|}{||x - y||_1}.$$

Let

$$\mathcal{L} = \bigcup_{n=1}^{\infty} \mathcal{L}_n \quad \text{where } \mathcal{L}_n = \{g \in \mathbb{L}^{\infty}(\mathbb{R}^n);\ \text{Lip}(g) < \infty,\ ||g||_{\infty} := \sup_x |g(x)| \leq 1\},$$

the class $\mathcal{L}$ is used together with the following two functions:

$$\psi_1(g, h, n, m) = \min(n, m)\text{Lip}(g)\text{Lip}(h)$$
$$\psi_2(g, h, n, m) = 4(n + m) \min\{\text{Lip}(g), \text{Lip}(h)\}$$

for functions g and h defined on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively.

Let $(X_n)_{n\in\mathbb{Z}}$ denote a strictly stationary sequence of real valued random variables. We restrict to 1-dimensional samples for simplicity in this work but can extend to a multi-dimensional frame.

Definition 1 (Doukhan and Louhichi 1999) The sequence $(X_n)_{n\in\mathbb{Z}}$ is called $(\epsilon, \mathcal{L}, \psi)$-*weakly dependent* (simply, ψ-*weakly dependent*), if there exists a sequence $\epsilon = (\epsilon_r)_{r\in\mathbb{Z}}$ decreasing to zero at infinity and a function ψ with arguments $(g, h, n, m) \in \mathcal{L}_n \times \mathcal{L}_m \times \mathbb{N}^2$ such that for n-tuple $(i_1, \ldots, i_n)$ and m-tuple $(j_1, \ldots, j_m)$ with $i_1 \leq \cdots \leq i_n < i_n + r \leq j_1 \leq \cdots \leq j_m$, one has

$$|\text{Cov}(g(X_{i_1}, \ldots, X_{i_n}), h(X_{j_1}, \ldots, X_{j_m}))| \leq \psi(g, h, n, m)\epsilon_r. \tag{1}$$

According to Doukhan and Louhichi (1999), associated sequences are ψ_1-weakly dependent, and Bernoulli shifts and Markov processes are ψ_2-weakly dependent. Our main attraction is such examples of processes that are weakly dependent, but not mixing.

3 Large sample analysis

We consider mode estimation of densities of strictly stationary ψ-weak dependent sequences $\{X_t\}$ of real-valued random variables. We assume that the marginal density of X_t exists and we denote it by $f(x)$. The standard kernel density estimate based on data $\{X_1, X_2, \ldots, X_n\}$ is defined by

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

where K is some real-valued function called *Kernel function*, integrating to 1, being Lipschizian and being rapidly convergent to 0 at infinity, and $h := h_n > 0$ is called *bandwidth*, such that $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

We estimate the mode via $\hat{f}_n$ as in Parzen (1962) where the mode is defined as the location at which f has maximum. Let $I = [a, b]$ be a compact interval and $J = [a - \varrho, b + \varrho],\ \varrho > 0$ be a slight enlargement of I which is introduced to avoid boundary effects. We assume that f is unimodal in the interval I and denote its mode by θ. Then, the mode satisfies

$$f(\theta) > \sup_{x \in J : |x-\theta| > \varepsilon} f(x),$$

for all $\varepsilon > 0$. A sample mode corresponding to the kernel estimator $\hat{f}_n$ is the estimator $\hat{\theta}_n$ satisfying

$$\hat{f}_n(\hat{\theta}_n) = \sup_{x \in I} \hat{f}_n(x). \tag{2}$$

Note that $\hat{\theta}_n$ exists if K is continuous; however, it may not be unique. It is known that kernel estimators tend to produce some additional and superfluous modality. We develop our consistency results so that the results are valid for any choice of $\hat{\theta}_n$ satisfying (2).

We first give consistency of the mode estimates and specifies the rate of convergence under smoothness conditions on f. As in Shi et al. (2009a), we establish the relationship between the convergence rate of the mode estimator and the bandwidth. Convergence in probability is established for the mode under a set of mild conditions (A1)–(A4) below. Under an additional condition (A5) below, we also establish almost complete convergence for modes, which is a stronger convergence than almost sure convergence. A sequence $\{Z_n\}$ of real-valued random variables is said to *converge almost completely* to 0 if for every $\epsilon > 0$ we have $\sum_{n=1}^{\infty} P(|Z_n| > \epsilon) < \infty$. This is denoted by $Z_n = o_{a.co.}(1)$. The complete convergence implies almost sure convergence and convergence in probability. A sequence $\{Z_n\}$ is said to be *almost completely bounded* if there exists an $M > 0$ such that $\sum_{n=1}^{\infty} P(|Z_n| > M) < \infty$. This is denoted by $Z_n = O_{a.co}(1)$. For a sequence $a_n > 0$, $Z_n = o_{a.co.}(a_n)$ and $Z_n = O_{a.co.}(a_n)$ denote $Z_n/a_n = o_{a.co.}(1)$ and $Z_n/a_n = O_{a.co.}(1)$, respectively.

In Assumption 1 below, we describe conditions required for our study.

Assumption 1 (A1) Let ρ denote the regularity of the function f in terms of Hölder spaces, this means that setting $\rho = a + b$ with $a \in \mathbb{N}^+$ and $0 \le b < 1$ there exists a constant $A > 0$ such that f is a-times continuously differentiable with $|f^{(a)}(x) - f^{(a)}(y)| \le A|x-y|^b$ for x, y belonging to an arbitrary compact interval of $\mathbb{R}$, where $f^{(a)}$ is the ath order derivative of f. Assume f has $k+\rho$ times continuously differentiable for $k = 0, 1, \ldots, \rho$.

(A2) ρ is an integer (≥ 2) and the kernel function K is of order ρ, i.e., it satisfies

$$\int x^i K(x)\,\mathrm{d}x = 0 \quad \text{for } i = 1, \ldots, \rho - 1, \quad \text{and} \quad \int x^\rho K(x)\,\mathrm{d}x \ne 0.$$

Also we have $|x|K(x) \to 0$ and $|x|K^{(\rho)}(x) \to 0$ as $x \to \infty$. Here and in the sequel, integrals are all over $(-\infty, \infty)$. Assume that the kernel function K has ρ bounded derivatives, and $K^{(k)}$ is Lipschitz continuous for $k = 0, 1, \ldots, \rho$.

(A3) Let $f_{0,n}$ denote the joint density of X_0 and X_n. There exists some positive constant C such that for any positive integer n,

$$||f||_\infty \le C \quad \text{and} \quad ||f_{0,n}||_\infty \le C. \tag{3}$$

The stationary sequence (X_t) satisfies the following weak dependence condition:

$$\text{for all } g, h \in \mathcal{L}_1, \quad |\mathrm{Cov}(g(X_0), h(X_r))| \leq \mathrm{c}\,\mathrm{Lip}(g)\mathrm{Lip}(h)\epsilon_r, \tag{4}$$

where c is a constant not depending on g, h and r and $\epsilon_r \to 0$ as $n \to \infty$.

(A4) We have

$$f^{(j)}(\theta) = 0 \quad \text{for } j = 1, \ldots, \phi - 1 \text{ and } f^{(\phi)}(\theta) \neq 0 \tag{5}$$

for some positive integer ϕ, and $f^{(\phi)}(x)$ is continuous in a neighborhood of θ.

(A5) The weak dependence coefficients $\{\epsilon_r\}$ satisfy $\epsilon_r = o(r^{-2}), U(r) := \sum_{i=r}^{\infty} \epsilon_i = O(r^{-3})$ and $\lim_{n\to\infty} ||X_t||_\gamma = \lim_{n\to\infty}(E|X_t|^\gamma)^{1/\gamma} < \infty$ where $\gamma \sim \sqrt{n}$.

A common value for ϕ in (A4) is 2 because usually $f^{(1)}(\theta) = 0$ and $f^{(2)}(\theta) \neq 0$. In (A2), if K is an even function as usual, and $x^2 K(x)$ is integrable, then $\rho = 2$ because $\int x K(x)\,\mathrm{d}x = 0$ and $\int x^2 K(x)\,\mathrm{d}x > 0$.

The first three conditions (A1)–(A3) are regularity conditions on the density f, kernel K, and the ψ-weakly dependent process X_t which were adopted by Doukhan and Louhichi (2001) for asymptotic study of kernel density estimators. We also use these three conditions and, additionally, the condition (A4) on derivatives of $f(\theta)$ for establishing convergence in probability and asymptotic normality in Theorems 1(a), 2(a), and 3 below. The last one (A5) is related with weak dependence coefficient ϵ_r and asymptotic norm condition of X_t. Under these conditions, we can establish an exponential inequality, which is given in Lemma 2 in Appendix, and the inequality will be used to show almost complete convergence results in Theorems 1(b) and 2(b) below.

Theorem 1 *Suppose that the stationary sequence $(X_t)_{t\in\mathbb{N}}$ is either $(\epsilon, \mathcal{L}, \psi_1)$- or $(\epsilon, \mathcal{L}, \psi_2)$-weakly dependent. Assume $nh/\log n \to \infty$ and $h = O((\log n/n)^{1/(2\rho+1)})$.*

(a) *If* (A1)–(A4) *hold, then $\hat{\theta}_n - \theta = o_p(1)$ as $n \to \infty$.*
(b) *If* (A1)–(A5) *hold, then $\hat{\theta}_n - \theta = o_{a.co}(1)$ as $n \to \infty$.*

Conditions on the bandwidth in Theorem 1 state that h is of larger order than $\log n/n$ and is of the same order as $(\log n/n)^{1/(2\rho+1)}$. If a rate condition is imposed on h, a rate is obtained for the convergence of $\hat{\theta}_n$ as given in the following theorem which relates convergence rate of $\hat{\theta}_n$ with the bandwidth h.

Theorem 2 *Suppose that the stationary sequence $(X_t)_{t\in\mathbb{N}}$ is either $(\epsilon, \mathcal{L}, \psi_1)$- or $(\epsilon, \mathcal{L}, \psi_2)$-weakly dependent. Assume $nh/\log n \to \infty$ and $h = o((\log n/n)^{\phi/(2\rho(\phi-1)+3\phi)})$.*

(a) *If* (A1)–(A4) *hold, then $\hat{\theta}_n - \theta = O_p\left(\left(\frac{\log n}{nh^3}\right)^{1/(2\phi-2)}\right)$ as $n \to \infty$.*
(b) *If* (A1)–(A5) *hold, then $\hat{\theta}_n - \theta = O_{a.co}\left(\left(\frac{\log n}{nh^3}\right)^{1/(2\phi-2)}\right)$ as $n \to \infty$.*

Remark 1 Note that the rate result $\left(\log n/nh^3\right)^{1/(2\phi-2)}$ in O_p of Theorem 2(a) does not require any condition on the ψ-weak dependent coefficients ϵ_r other than $\epsilon_r \to 0$. The almost complete boundedness in Theorem 2(b), having the same rate as the order-in-probability boundedness in Theorem 2(a), requires more restrictive conditions on the ψ-weak dependent coefficients ϵ_r of (A5). The convergence rate in Theorem 2 is sharp, in that it is the same as that in Theorem 1 of Shi et al. (2009a) and others in i.i.d. cases. In the ψ-weak case, we are able to prove the same convergence rate by suitably bounding a covariance term arising from expanding $\mathrm{Var}\hat{f}_n^{(k)}(x)$, see Lemma 1(b) and Theorem 4(a), (c) and their proofs in Appendix.

We now consider the asymptotic normality of the mode estimator. To establish the asymptotic normality of the mode estimator, we suppose that f is twice differentiable, (i.e., $\phi = 2$). (For general $\phi \geq 2$, the results will be similar.) That is,

$$f^{(1)}(\theta) = 0 \quad \text{and} \quad f^{(2)}(\theta) < 0.$$

A nonparametric kernel estimator of the kth derivative $f^{(k)}$ is given by

$$\hat{f}_n^{(k)}(x) = \frac{1}{nh^{k+1}} \sum_{i=1}^{n} K^{(k)}\left(\frac{x - X_i}{h}\right) \tag{6}$$

where $K^{(k)}$ is the kth derivative of K, $(k = 1, 2, \ldots)$. We choose a twice differentiable kernel function K such that

$$\hat{f}_n^{(1)}(\hat{\theta}_n) = 0 \quad \text{and} \quad \hat{f}_n^{(2)}(\hat{\theta}_n) < 0.$$

By Taylor expansion of $\hat{f}_n^{(1)}(\hat{\theta}_n)$ in a neighborhood of θ, we have $0 = \hat{f}_n^{(1)}(\hat{\theta}_n) = \hat{f}_n^{(1)}(\theta) + (\hat{\theta}_n - \theta)\hat{f}_n^{(2)}(\theta_n^*)$ where θ_n^* is between θ and $\hat{\theta}_n$. Hence, we write

$$\hat{\theta}_n - \theta = \frac{-\hat{f}_n^{(1)}(\theta)}{\hat{f}_n^{(2)}(\theta_n^*)} \tag{7}$$

if the denominator does not vanish. The following theorem states asymptotic normality of the mode estimator.

Theorem 3 *Suppose that the stationary sequence $(X_n)_{n\in\mathbb{N}}$ is either $(\epsilon, \mathcal{L}, \psi_1)$-weakly dependent with $\epsilon_r = O(r^{-12-\nu})$, or $(\epsilon, \mathcal{L}, \psi_2)$-weakly dependent with $\epsilon_r = O(r^{-9-\nu})$, for some $\nu > 0$. Let the conditions* (A1)–(A4) *in Assumption 1 hold. If $h = O(n^{-1/(2\rho+3)})$ and $nh^5 \to \infty$, then*

$$\sqrt{nh^3}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N(b(\theta), \sigma^2(\theta))$$

where

$$b(\theta) = c' \cdot \delta_{\rho,K} \frac{f^{(1+\rho)}(\theta)}{f^{(2)}(\theta)}, \quad \sigma^2(\theta) = \lambda_{1,K} \frac{f(\theta)}{[f^{(2)}(\theta)]^2},$$
$$\delta_{\rho,K} = \frac{(-1)^\rho}{\rho!} \int x^\rho K(x)\mathrm{d}x, \quad \textit{and} \quad \lambda_{1,K} = \int (K^{(1)}(x))^2 \mathrm{d}x,$$

provided $f^{(2)}(\theta) \neq 0$ *for some constant* c'. *If* $h = c \cdot n^{-1/(2\rho+3)}$ *for some constant* c, *then* $c' = -c^{(2\rho+3)/2}$.

Theorem 3 establishes a central limit theorem for nonparametric density estimator of mode under more general class of ψ-weak dependent errors than the class of i.i.d. errors under which Parzen (1962), Eddy (1980, 1982), and Romano (1988) established central limit theorems for mode estimators. The asymptotic mean, bias, and variance in Theorem 3 are determined by marginal quantities and are the same as those in i.i.d. cases.

Theorem 3 can be used for statistical inference on mode. For example, testing for significance of $\hat{\theta}_n$ can be performed by comparing

$$t_{\hat{\theta}} = \left(\sqrt{nh^3}\hat{\theta}_n - b(\hat{\theta}_n)\right) / \sigma(\hat{\theta}_n)$$

with standard normal percentiles as is illustrated in Sects. 5 and 6 below.

Remark 2 Some authors discussed the bandwidth in terms of density mode estimator in i.i.d. cases. Silveran (1978) and Romano (1988) showed that, for consistent estimation of $f^{(2)}(\theta)$, we should have $nh_n^5/\log n \to \infty$. Romano (1988) verified that, by choosing $h \sim n^{-1/7}$, the kernel mode estimator achieves the optimal minimax rate of $n^{2/7}$. After Romano (1988), several authors dealt with the density mode estimations, e.g., Vieu (1996), Mokkadem and Pelletier (2005) and Ferraty et al. (2006). Shi et al. (2009a) established the relationship between the rate of convergence of the kernel mode estimator and the bandwidth. In Remarks 5,6,7 of Shi et al. (2009a), they showed that their almost complete convergence rate for density mode estimator is better than those in Vieu (1996), Mokkadem and Pelletier (2005) and Ferraty et al. (2006).

4 Bandwidth selection

This section proposes a bandwidth selection procedure which is optimized for mode estimation. Let the conditions in Theorem 3 hold. The theorem states that the mode has a nontrivial limiting distribution for bandwidth $h = O(n^{-1/(2\rho+3)})$. If $h = c \cdot n^{-1/(2\rho+3)}$, the mean squared error of the mode estimator $\hat{\theta}_n$ is given by

$$E(\hat{\theta}_n - \theta)^2 = \mathrm{Var}\hat{\theta}_n + (E\hat{\theta}_n - \theta)^2 = \frac{1}{nh^3}\left[\sigma^2(\theta) + c^{2\rho+3}\left(\delta_{\rho,K}\frac{f^{(1+\rho)}(\theta)}{f^{(2)}(\theta)}\right)^2\right] + o\left(\frac{1}{nh^3}\right). \tag{8}$$

For $h = c \cdot n^{-1/(2\rho+3)}$, we have

$$E(\hat{\theta}_n - \theta)^2 \cong n^{-2\rho/(2\rho+3)} \left[c^{-3}\sigma^2(\theta) + c^{2\rho}\left(\delta_{\rho,K}\frac{f^{(1+\rho)}(\theta)}{f^{(2)}(\theta)}\right)^2 \right] \equiv n^{-2\rho/(2\rho+3)} g(c).$$

We obtain $c = [3\sigma^2(\theta)(f^{(2)}(\theta))^2/\{2\rho\delta^2_{\rho,K}(f^{(1+\rho)}(\theta))^2\}]^{1/(2\rho+3)}$ that minimizes $g(c)$ from $g'(c) = 0$. Thus, we obtain the optimal bandwidth

$$\hat{h}_\theta = \left[\frac{3\lambda_{1,K} f(\theta)}{2\rho\delta^2_{\rho,K}(f^{(1+\rho)}(\theta))^2}\right]^{1/(2\rho+3)} \cdot n^{-1/(2\rho+3)} \tag{9}$$

which minimizes the MSE $E(\hat{\theta}_n - \theta)^2$ approximately, provided $f^{(1+\rho)}(\theta) \neq 0$.

Even though the optimal bandwidth (9) is valid only if $f^{(1+\rho)}(\theta) \neq 0$ which is not the case for normal distributions, it as well as (8) gives insight into the relation with the MSE and bandwidth in normal cases. For normal distributions, $f^{(1+\rho)}(\theta) = 0$ and MSE in (8) becomes $E(\hat{\theta}_n - \theta)^2 = \sigma^2(\theta)/nh^3 + o\left(1/nh^3\right)$ which decreases as h increases. However, h cannot increase without limit because h should be in the form $c \cdot n^{-1/(2\rho+3)}$.

In the real world, the true distribution of "normally-looking" data would not be exactly normal and should be close to the normal distribution. For such "close-normal" distributions, the optimal bandwidth (9) is valid. For normal or "close-normal" cases, bandwidth will be chosen to be larger than estimated value of (9) because the true values of $f^{(1+\rho)}(\theta)$ are zero or close to zero and hence $\hat{h}_\theta$ with estimates of $f(\theta)$ and $f^{(1+\rho)}(\theta)$ plugged-in underestimates its true value.

To implement the bandwidth in real data analysis, estimates of terms $f(\theta)$ and $f^{(1+\rho)}(\theta)$ should be plugged-in which require an estimate of the density f and an initial estimate of the mode θ. For estimating f, a natural choice is using a good kernel density estimator established in the literature. We use the bandwidth, $\bar{h}$ say, optimized for density estimation ("density-optimal" in the sequel) such as the rule of thumb estimator, plug-in estimator of Sheater and Jones (1991) and others, or the cross-validation estimator of Rudemo (1982) and others. See Sect. 5 for examples of such estimators. Let $\bar{f}_n$ be the kernel density estimator constructed with the density-optimal bandwidth $\bar{h}$. Now, an initial mode estimator $\bar{\theta}_n$, say, can be constructed which maximizes $\bar{f}_n$. Based on (8) and the above discussions on the normal cases, we propose a "feasible" bandwidth

$$\hat{h} = H \cdot \left[\frac{3\lambda_{1,K} \bar{f}_n(\bar{\theta}_n)}{2\rho\delta^2_{\rho,K}(\bar{f}_n^{(1+\rho)}(\bar{\theta}_n))^2}\right]^{1/(2\rho+3)} \cdot n^{-1/(2\rho+3)},$$

where $\bar{f}_n^{(1+\rho)}$ be the estimator of $f^{(1+\rho)}$ constructed using the density-optimal bandwidth $\bar{h}$. This is a bandwidth optimized for mode estimation. In the sequel, we will call $\hat{h}$ "mode-optimal" bandwidth. Here, according to the discussion in the previous paragraph, H is chosen to be greater than 1, for example 2. This choice is a compromise which covers both normal distributions with $f^{(1+\rho)}(\theta) = 0$ and non-normal distributions with $f^{(1+\rho)}(\theta) \neq 0$. A Monte-Carlo experiment in Sect. 5 reveals that MSE performance of $\hat{\theta}_n$ is better than $\bar{\theta}_n$ over a wide range of H.

5 Small sample analysis

A finite sample Monte-Carlo experiment is conducted to compare two mode estimates. One estimate $\bar{\theta}_n$ is constructed using the density-optimal bandwidth and the other estimate $\hat{\theta}_n$ is constructed using the mode-optimal bandwidth. Also, rejection percentage of the level 5 % test $t_{\hat{\theta}} = (\sqrt{n\hat{h}^3}\hat{\theta}_n - b(\hat{\theta}_n))/\sigma(\hat{\theta}_n)$ is compared with its nominal level 5 %.

Four data generating processes are considered:

$$\begin{aligned} &D_1 : X_t = a_t; \quad D_2 : X_t = 0.5X_{t-1} + a_t; \\ &D_3 : X_t = \sigma_{t-1}a_t, \quad \sigma_t^2 = 1 + 0.45\sigma_{t-1}^2 + 0.45X_{t-1}^2; \\ &D_4 : X_t = e^{a_t} - 1.032, \quad t = 1, 2, \ldots, n, \end{aligned}$$

where $X_0 = 0$, $\sigma_0^2 = 1/(1 - 0.45 - 0.45) = 10$ and a_t are i.i.d. N(0,1) errors. Mean of all distributions is 0. The distribution D_1 is considered as a standard distribution. The distributions D_2 and D_3 are considered to investigate effects of AR(1)-type serial correlation and GARCH(1,1)-type conditional heteroscedasticity, respectively. The distribution D_4 is a skewed one. Modes of the first three distributions D_1, D_2 and D_3 are zero and mode of D_4 is -0.093. The first two distributions D_1 and D_2 are normal and the last two distributions D_3 and D_4 are not normal. For sample size n, we consider $n = 100, 1000$.

To compute initial mode estimator $\bar{\theta}_n$ and density function $\bar{f}_n$, three density-optimal bandwidths are considered. The first two bandwidths $\bar{h}_{\mathrm{RT}}$ and $\bar{h}_{\mathrm{SJ}}$ are plug-in estimators which try to minimize a limit

$$\mathrm{AMISE}(h) = \frac{1}{nh}R(K) + h^{2\rho}\left(\int \frac{1}{\rho!}x^{\rho}K(x)\mathrm{d}x\right)^2 R(f^{(2)})$$

of mean integrated squared error $\mathrm{MISE}(h) = E[\mathrm{ISE}(h)]$, where $R(F) = \int F^2(x)\mathrm{d}x$ for a function F and $\mathrm{ISE}(h) = \int\{\hat{f}_h(x) - f(x)\}^2\mathrm{d}x$ is the integrated squared error of $\hat{f}_h$. If we use normal kernel, AMISE(h) is minimized by $h = (1/(2\sqrt{\pi}))^{1/5} \left[nR(f^{(2)})\right]^{-1/5}$. If the distribution is normal, this becomes $h = 1.06\sigma_x n^{-1/5}$, where σ_x^2 is the variance of X_t. Since σ_x^2 is not known in the real world, it is replaced by $\hat{\sigma}_x^2$, the sample variance of X_t, $t = 1, \ldots, n$. This is called the rule of thumb estimator. The resulting one $\bar{h}_{\mathrm{RT}} = 1.06\hat{\sigma}_x n^{-1/5}$ is our first choice. The second one $\bar{h}_{\mathrm{SJ}}$ is the plug-in estimator of Sheater and Jones (1991) who estimate $R(f^{(2)})$ by

$$\hat{R}_{\alpha(h)}(f^{(2)}) = \frac{1}{n(n-1)}\sum\sum_{i\neq j}\frac{1}{\{\alpha(h)\}^5}K^{(4)}\left(\frac{X_i - X_j}{\alpha(h)}\right),$$

where

$$\alpha(h)=(6\sqrt{2})^{1/7}\left(\frac{\hat{R}_a(f^{(2)})}{\hat{R}_b(f^{(3)})}\right)^{1/7}\cdot h^{5/7},\ a=0.920Q_R\cdot n^{-1/7},\ b=0.912Q_R\cdot n^{-1/9},$$

$$\hat{R}_a(f^{(2)})=\frac{1}{n(n-1)}\sum\sum_{i\neq j}\frac{1}{a^5}K^{(4)}\left(\frac{X_i-X_j}{a}\right),$$

$$\hat{R}_b(f^{(3)})=-\frac{1}{n(n-1)}\sum\sum_{i\neq j}\frac{1}{b^7}K^{(6)}\left(\frac{X_i-X_j}{b}\right),$$

and Q_R is the interquartile range. Now, the Sheather and Jones estimator $\bar{h}_{\mathrm{SJ}}$ is the solution to the equation

$$(1/(2\sqrt{\pi}))^{1/5}\hat{R}_{\alpha(h)}(f^{(2)})^{-1/5}n^{-1/5}-h=0.$$

The last one $\bar{h}_{\mathrm{CV}}$ is obtained by the least square cross-validation of Rudemo (1982) and Bowman (1984) which minimizes an estimate of ISE(h) called least-squares cross-validation function given by

$$\frac{1}{2\sqrt{\pi}nh}+\sum\sum_{i\neq j}\frac{1}{\sqrt{2}h}K\left(\frac{X_i-X_j}{\sqrt{2}h}\right)-\frac{2}{h}K\left(\frac{X_i-X_j}{h}\right).$$

Mode estimates $\bar{\theta}_n$ are constructed with density-optimal bandwidths $\bar{h}=\bar{h}_{\mathrm{RT}},\bar{h}_{\mathrm{SJ}},\bar{h}_{\mathrm{CV}}$ and improved mode estimator $\hat{\theta}_n$ is constructed using the corresponding mode-optimal bandwidths. The normal random errors a_t are generated by RNNOA, an IMSL FORTRAN subroutine. As the kernel function, the Gaussian kernel $K(x)=(2\pi)^{1/2}e^{-x^2/2}$ is used. As the tuning parameter H for computing $\hat{\theta}_n$, four values are considered: $H=1.5$, 2, 3 and 4.

Averages and mean squared errors of the estimates are displayed in Table 1, which is based on 1,000 independent replications. Relative efficiency of $\hat{\theta}_n$ over $\bar{\theta}_n$ is also displayed, which is the ratio of the MSE of $\bar{\theta}_n$ over that of $\hat{\theta}_n$. Rejection percentages of the level 5 % test $t_{\hat{\theta}}$ are displayed in Table 2 which reject $H_0:\theta=0$ if $|t_{\hat{\theta}}|>1.96$. For D_1, D_2 and D_3 having zero mode, the percentages are empirical sizes and, for D_4 having mode −0.093, the percentage is empirical power.

General messages of these tables are:

1. The proposed estimator $\hat{\theta}_n$ is substantially better than $\bar{\theta}_n$ in terms of both bias and MSE for all H considered here.
2. The test $t_{\hat{\theta}}$ has reasonable size values for D_1, D_2 and D_3 while having good power values for D_4.
3. If we are interested in statistical testing, $H=2$ seems a good choice giving us a valid empirical size and a reasonable MSE performance.

More detailed discussion of the experiment results follows. In Table 1, we see that, compared with $\bar{\theta}_n$, $\hat{\theta}_n$ is substantially less-biased and more efficient. Exception is only

Table 1 Averages and MSE of mode estimators

	n	Avg. of $\bar{\theta}_n$	Avg. of $\hat{\theta}_n$				MSE of $\bar{\theta}_n$	Rel. eff. of $\hat{\theta}_n$ over $\bar{\theta}_n$			
			H					H			
			1.5	2	3	4		1.5	2	3	4
$\bar{h} = \bar{h}_{\mathrm{RT}}$											
D_1	100	−0.1015	0.0050	0.0045	0.0019	0.0007	0.03955	1.43	2.49	3.69	4.19
D_1	1000	−0.0575	−0.0007	0.0020	0.0020	0.0018	0.01149	1.39	3.02	6.60	9.05
D_2	100	−0.1199	0.0030	0.0052	0.0018	0.0003	0.07441	1.24	1.66	2.04	2.15
D_2	1000	−0.0673	−0.0017	0.0012	0.0024	0.0026	0.01738	1.43	2.33	3.41	3.93
D_3	100	−0.2052	0.0195	0.0178	0.0110	0.0082	0.17889	1.42	2.12	2.71	2.81
D_3	1000	−0.0810	0.0043	0.0050	0.0041	0.0038	0.02595	1.39	2.44	3.65	4.02
D_4	100	−0.1883	−0.1124	−0.0184	0.1150	0.2058	0.02220	1.41	1.04	0.37	0.21
D_4	1000	−0.3012	−0.3268	−0.2460	−0.1215	−0.0288	0.04522	0.80	1.77	14.49	6.88
$\bar{h} = \bar{h}_{\mathrm{SJ}}$											
D_1	100	−0.0911	0.0055	0.0050	0.0021	0.0009	0.03398	1.41	2.28	3.26	3.63
D_1	1000	−0.0535	−0.0006	0.0017	0.0019	0.0017	0.01036	1.38	2.97	6.40	8.56
D_2	100	−0.1106	0.0022	0.0045	0.0015	0.0002	0.06977	1.23	1.60	1.92	2.01
D_2	1000	−0.0617	−0.0012	0.0015	0.0025	0.0027	0.01555	1.36	2.19	3.14	3.57
D_3	100	−0.2049	0.0149	0.0136	0.0075	0.0053	0.18017	1.36	2.04	2.69	2.83
D_3	1000	−0.1047	−0.0004	0.0035	0.0042	0.0042	0.04082	1.44	2.79	4.95	5.94
D_4	100	−0.4709	−0.3924	−0.3174	−0.1995	−0.1093	0.15413	1.46	2.32	5.30	8.03
D_4	1000	−0.5966	−0.5486	−0.4963	−0.4076	−0.3353	0.25712	1.21	1.52	2.42	3.83
$\bar{h} = \bar{h}_{\mathrm{CV}}$											
D_1	100	−0.0176	0.0002	−0.0025	−0.0012	−0.0009	0.05991	1.75	2.75	4.39	5.21
D_1	1000	−0.0112	−0.0018	0.0003	0.0016	0.0018	0.02097	2.08	4.09	9.71	14.27
D_2	100	−0.0363	−0.0045	0.0001	0.0012	0.0011	0.09555	1.41	1.86	2.38	2.55
D_2	1000	−0.0153	0.0022	0.0025	0.0028	0.0030	0.02967	1.82	3.08	4.96	6.03
D_3	100	−0.0441	0.0098	0.0077	0.0033	0.0001	0.33115	1.89	2.85	4.23	4.57
D_3	1000	−0.0075	0.0021	0.0026	0.0050	0.0046	0.09367	2.06	3.93	7.71	10.48
D_4	100	−0.5112	−0.4439	−0.3803	−0.2781	−0.1957	0.19817	1.35	1.87	3.27	5.04
D_4	1000	−0.6108	−0.5724	−0.5334	−0.4590	−0.3947	0.27527	1.16	1.37	1.93	2.71

for D_4 with $\bar{h} = \bar{h}_{\mathrm{RT}}$ and $n = 100$, which is not a disappointing result because the normal-based bandwidth $\bar{h} = \bar{h}_{\mathrm{RT}}$, being based heavily on normality, does not work well for small $n = 100$ under the asymmetric distribution D_4. As H increases from 1.5 up to 3, $\hat{\theta}_n$ gets less-biased having smaller MSE but, as H increases from 3 to 4, there appear several cases of increased bias and larger MSE. The better performance for the larger $H = 2, 3$ than that for the smaller $H = 1.5$ tells us that smoother density estimate with $H = 2, 3$ gives us better mode estimator. However, as H gets too large, the good mode estimation tends to disappear. Among the three initial bandwidths $\bar{h} = \bar{h}_{\mathrm{RT}}$, $\bar{h}_{\mathrm{SJ}}$ and $\bar{h}_{\mathrm{CV}}$, $\bar{h}_{\mathrm{SJ}}$ seems to be the best, giving the smallest MSE for $\hat{\theta}_n$.

Table 2 Rejection frequency (%) of level 5 % test $t_{\hat{\theta}_n}$

	n	$\bar{h} = \bar{h}_{RT}$				$\bar{h} = \bar{h}_{SJ}$				$\bar{h} = \bar{h}_{CV}$			
		1.5	2	3	4	1.5	2	3	4	1.5	2	3	4
Size													
D_1	100	8	3	1	0	8	3	1	0	12	5	2	1
D_1	1000	12	5	3	2	12	5	3	2	13	6	2	1
D_2	100	18	14	9	6	19	14	9	6	21	15	10	6
D_2	1000	16	12	13	12	16	14	13	12	19	14	13	10
D_3	100	5	2	0	0	7	3	0	0	11	5	1	0
D_3	1000	9	4	2	1	10	5	3	2	13	6	3	3
Power													
D_4	100	36	7	1	1	94	86	57	27	93	89	70	52
D_4	1000	100	100	88	20	100	100	100	98	100	100	100	99

In Table 2, the test $t_{\hat{\theta}}$ is slightly over-sized for $H = 1.5$. Size values are reasonably close to the nominal value 5 % for $H = 2$. The test has reasonable power for D_4 except for $h = h_{RT}$ with $n = 100$. This indicates that the rule of thumb bandwidth does not work well for distributions which are much different from normal distributions, especially for small n.

6 Example

The proposed method is applied to investigate asymmetric features of daily log returns of Dow Jones index (DW) and US Dollar exchange rate (EX) relative to Euro for the period from 01-04-2001 to 12-31-2009 whose plots are displayed in Figs. 1, 2. Basic statistics are $n = 2153$, $\bar{x} = -0.00005099$, $\hat{\sigma}_x = 0.01339189$ for DW and $n = 2264$, $\bar{x} = 0.00018637$, $\hat{\sigma}_x = 0.0064877$ for EX.

Finance variables such as stock prices and exchange rates have diverse asymmetric aspects. Specifically, dynamics of the variables at up-times, i.e., the times at which they are increasing, are different from those at down-times rendering the distributions for positive values being different from those for negative values. Diverse models such as TAR models and EGARCH models are employed to analyze the asymmetry.

Important asymmetric features of those variables would be revealed in their density plots. In Figs. 3 and 4, kernel density plots are provided for Dow Jones and exchange rates as well as the density curves based on normal approximations. The curves are constructed using the mode-optimal bandwidths $\hat{h}_{SJ} = 0.0047$ and 0.0031 of Sect. 4, whose computation is described in the later part of this section.

The figures show significant non-normality and asymmetry. Distributions of both DW and EX are much different from normal density plots which correspond to $N(\bar{x}, \hat{\sigma}_x^2)$. Specifically, both DW and EX have heavier tails and sharper peaks than the corresponding normal densities. This is due to volatility clustering as manifested in Figs. 1 and 2: wilder variations for years 2008–2009 and milder variations for other

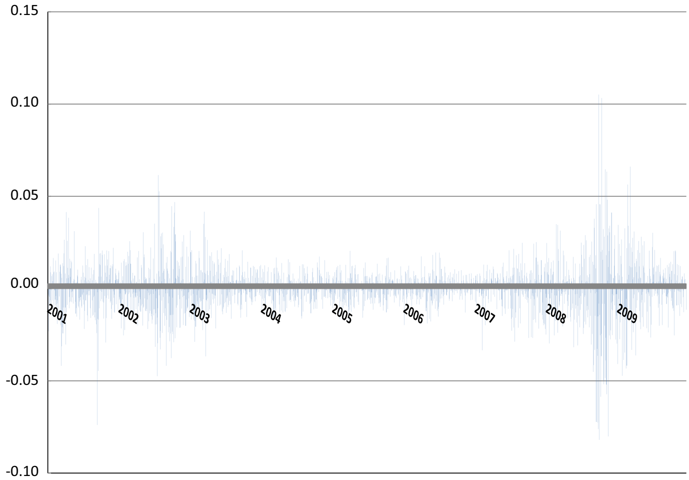

Fig. 1 Log return of daily Dow Jones index, 2001–2009

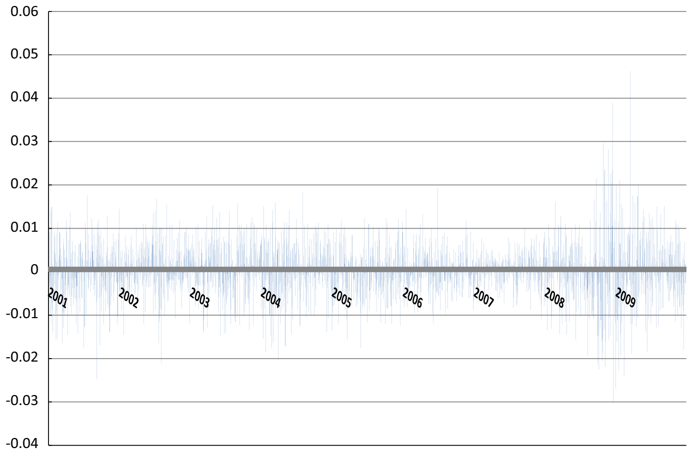

Fig. 2 Log return of daily exchange rates (Dollar/Euro), 2001–2009

times. Stronger conditional heteroscedasticity is observed for DW than for EX. We have fitted GARCH(1,1) models and found that both DW and EX have very significant GARCH coefficients. The data generating processes may well be approximated by GARCH processes, which are special cases of ψ-weakly dependent processes.

The density plots also reveal asymmetric features. Among others, we observe that distribution of DW is skewed to the right, having positive mode while that of EX is not so much. To justify this claim formally, applying the result of Theorem 3, we conduct tests for significance of the modes. If we use the Gaussian kernel and the bandwidth $\bar{h}_{SJ}$ of Sheater and Jones (1991) for initial mode estimation and $H = 2$ for the mode-optimal bandwidth, we have mode $\hat{\theta}_n = 0.0006052$ for DW being significant at 5 % level with t value 2.497 while mode $\hat{\theta}_n = 0.0001929$ for EX is not significant with

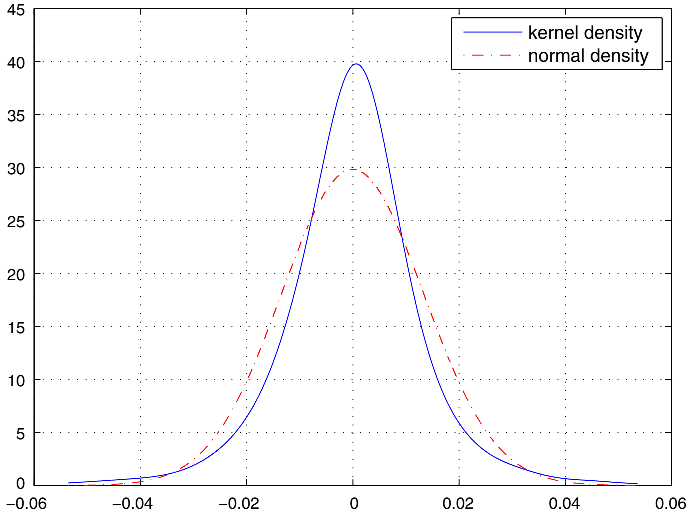

Fig. 3 Kernel density and normal density of daily Dow Jones index

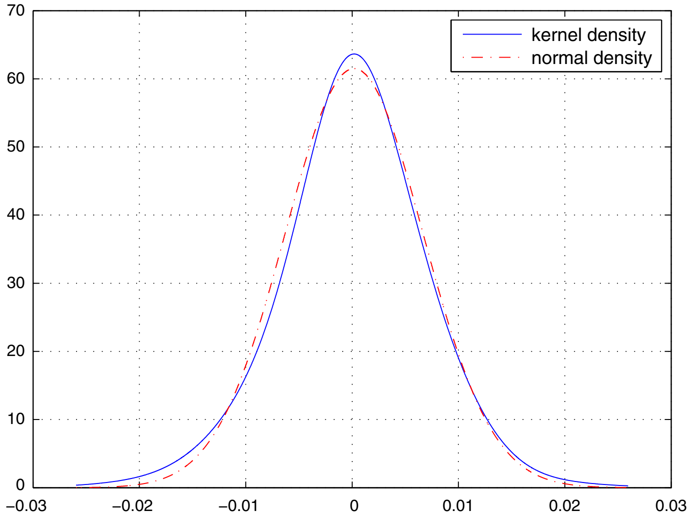

Fig. 4 Kernel density and normal density of daily exchange rates

t value 1.051. Since we are interesting in testing for significance of mode, following the recommendation of Sect. 5, we have used $H = 2$. The other two bandwidths $\bar{h}_{\mathrm{RT}}$ and $\bar{h}_{\mathrm{CV}}$ lead us to the same conclusion as those for the bandwidth $\bar{h}_{\mathrm{SJ}}$.

Detailed procedures for computing the mode values and their t values are described. Description is made for the data set DW with $\bar{h}_{\mathrm{SJ}}$. For DW, with $n = 2153$, $\hat{\sigma}_x = 0.01339189$, the density-optimal bandwidth of Sheater and Jones (1991) is $\bar{h}_{\mathrm{SJ}} = 0.0028$. From the kernel density estimator constructed with this initial bandwidth, we get an initial mode estimator $\bar{\theta}_n = -0.0000644$. A mode-optimal bandwidth

Table 3 Mode estimation

Initial BW	$\bar{\theta}_n$	$\hat{\theta}_n$	t	$\bar{h}$	$\hat{h}$	$b(\hat{\theta}_n)$	$\hat{\sigma}(\hat{\theta}_n)$
	Dow Jones Index, $n = 2153$, $\bar{x} = -0.0000510$, $\hat{\sigma}_x = 0.0133919$						
$\bar{h}_{\mathrm{RT}}$	−0.0000644	0.0006052	2.438	0.0031	0.0050	−0.000000313	0.000004249
$\bar{h}_{\mathrm{SJ}}$	−0.0000644	0.0006454	2.497	0.0028	0.0047	−0.000000244	0.000003862
$\bar{h}_{\mathrm{CV}}$	−0.0000644	0.0006856	2.640	0.0012	0.0042	−0.000000172	0.000003396
	Exchange rate, $n = 2264$, $\bar{x} = .0001864$, $\hat{\sigma}_x$ =0.0064877						
$\bar{h}_{\mathrm{RT}}$	−0.0000537	0.0002058	1.137	0.0015	0.0032	0.000000130	0.000001441
$\bar{h}_{\mathrm{SJ}}$	−0.0000796	0.0001929	1.051	0.0013	0.0031	0.000000125	0.000001387
$\bar{h}_{\mathrm{CV}}$	−0.0000537	0.0002058	1.137	0.0015	0.0032	0.000000130	0.000001441

$\hat{h}_{\mathrm{SJ}} = 0.0047$ is obtained from the procedure in Section 4 with $H = 2$. Using this mode-optimal bandwidth, we obtain mode estimator $\hat{\theta}_n = 0.0006454$. According to Theorem 3, when $\theta = 0$, $\hat{\theta}_n$ is asymptotically normally distributed with mean $b(\hat{\theta}_n)/\sqrt{n\hat{h}_{\mathrm{SJ}}^3} = -0.0000163$ and variance $\sigma^2(\hat{\theta}_n)/(n\hat{h}_{\mathrm{SJ}}^3) = 0.0002583^2$. Therefore, the t statistic for significance of $\hat{\theta}_n$ is $(0.0006454 + .000000244)/0.0002583 = 2.497$ being significant at 5 % level (Table 3).

7 Conclusion

Consistency and asymptotic normality have been established for mode estimators based on kernel-type densities under a class of ψ-weakly dependent processes. The convergence rate of the kernel density estimator of the mode is given in terms of the bandwidth. A bandwidth selection procedure is proposed which is based on an asymptotic optimal bandwidth. A finite-sample experiment shows that the mode estimator based on the mode-optimal bandwidth is substantially better than that based on density-optimal bandwidth. The proposed methods are applied to log returns of US Dow Jones index and foreign exchange rates of US Dollar relative to EUR. The analysis shows that the former has asymmetry of having mode significantly different from zero while the latter has not.

Appendix

Large sample results for density derivative estimators

This subsection provides some results for density (derivative) estimators. Expressions for large sample expectations and variances are given. Almost complete uniform bounds are given for deviations of density (derivative) estimators from their expectations. Asymptotic normalities are established. These results are used in establishing consistency and asymptotic normality of mode estimators. Throughout this subsection, we assume that $k \in \{0, 1, 2, \ldots, \rho\}$ is given. In the following lemmas and theorem, we suppose that the stationary sequence $(X_t)_{t\in\mathbb{N}}$ is either $(\epsilon, \mathcal{L}, \psi_1)$- or $(\epsilon, \mathcal{L}, \psi_2)$-weakly dependent and assume $nh/\log n \to \infty$. The asymptotic results are for $n \to \infty$.

Lemma 1 *If* (A1)–(A3) *are fulfilled, then, for all* $x \in \mathbb{R}$,

$$(a)\quad E\hat{f}_n^{(k)}(x) = f^{(k)}(x) + h^\rho \delta_{\rho,K} f^{(k+\rho)}(x) + o(h^\rho),$$
$$(b)\quad Var\hat{f}_n^{(k)}(x) = \frac{1}{nh^{2k+1}} f(x) \int (K^{(k)}(x))^2 \mathrm{d}x + \frac{2}{nh^{2k+1}}\Delta_h + o\left(\frac{1}{nh^{2k+1}}\right)$$

for some Δ_h, *covariance term, satisfying*

$$\Delta_h \leq c_1 mh \left(\int |K^{(k)}(x)|\mathrm{d}x\right)^2 + \frac{c_2}{h^3} Lip^2(K^{(k)}) \sum_{i=m}^{\infty} \epsilon_i$$

for some constants c_1, c_2, *and for any sequence* $m = m_n$ *with* $m < n$ *and* $m \to \infty$.

Lemma 2 *Let* $\{X_i\}$ *be a stationary sequence of* ψ*-weak dependent process with* ψ*-weak dependence coefficient sequence* $\{\epsilon_i\}$ *and with mean zero. Let* $S_n = \sum_{i=1}^n X_i$ *and* $\sigma_n^2 = Var(S_n)$. *If* $\epsilon_r = o(r^{-2})$, *then for any* $t > 0$ *and for sufficiently large n, we have*

$$P(|S_n| \geq t) \leq C_0 \log n \exp\left(-\frac{t^2}{A_n + B_n t}\right)$$

provided $||X_i||_\gamma < \infty$ *where* $\gamma \sim \sqrt{n}$ *for sufficiently large n, where* A_n *can be chosen as any number greater than or equal to* σ_n^2 *and* $B_n = n^{3/4}\log n/A_n$ *for some constant* $C_0 > 0$.

The proof of Lemma 2 is given in Hwang and Shin (2014) [also see Hwang and Shin (2013)] which establishes new exponential inequalities under some mild conditions.

Theorem 4 *If* (A1)–(A4) *are fulfilled, then*

$$\text{(a)}\quad \sup_{x\in I} |\hat{f}_n^{(k)}(x) - E\hat{f}_n^{(k)}(x)| = O_p\left(\sqrt{\frac{\log n}{nh^{2k+1}}}\right),$$
$$\text{(b)}\quad \sup_{x\in I} |E\hat{f}_n^{(k)}(x) - f^{(k)}(x)| = O(h^\rho).$$

If (A1)–(A5) *are fulfilled, then*

$$\text{(c)}\quad \sup_{x\in I} |\hat{f}_n^{(k)}(x) - E\hat{f}_n^{(k)}(x)| = O_{a.co}\left(\sqrt{\frac{\log n}{nh^{2k+1}}}\right).$$

Remark 3 (a) By Theorem 4, we have

$$\sup_{x\in I}|\hat{f}_n^{(k)}(x)-f^{(k)}(x)| = O_p\left(\sqrt{\frac{\log n}{nh^{2k+1}}}\right)+O(h^{\rho}).$$

Therefore, if $h = O((\log n/n)^{1/(2\rho+1)})$ as in Theorem 1, then

$$\sup_{x\in I}|\hat{f}_n^{(k)}(x)-f^{(k)}(x)| = O_p\left(\left(\frac{\log n}{n}\right)^{(\rho-k)/(2\rho+1)}\right). \tag{10}$$

(b) If the condition (A5) holds additionally, then (10) holds with O_p replaced by $O_{a.co}$.

Theorem 4 gives a key result of our work, and will be used in the proofs of Theorems 1 and 2. In the proof of Theorem 4(c), Lemma 2 will be applied.

Remark 4 For stationary-associated random processes, Douge (2007) derived an exponential inequality assuming boundedness and a convergence rate under conditions $nh/\log^2 n \to \infty$ and $h = O((\log^2 n/n)^{1/(2\rho+1)})$ as follows:

$$\sup_{x\in I}|\hat{f}_n(x)-f(x)| = O_{a.s.}\left(\left(\frac{\log^2 n}{n}\right)^{\rho/(2\rho+1)}\right).$$

Remark 3(a), (b) can be regarded as an extension of the result of Douge (2007) to ψ-weak dependent processes.

Theorem 5 *We suppose the same assumptions as in Lemma 1, and suppose moreover that the sequence $(X_t)_{t\in\mathbb{N}}$ is either $(\epsilon, \mathcal{L}, \psi_1)$-weakly dependent with $\epsilon_r = O(r^{-12-\nu})$, or $(\epsilon, \mathcal{L}, \psi_2)$-weakly dependent with $\epsilon_r = O(r^{-9-\nu})$, for some $\nu > 0$. Then*

(a) *if $nh^{2k+1} \to \infty$, then for all $x \in \mathbb{R}$,*

$$\sqrt{nh^{2k+1}}[\hat{f}_n^{(k)}(x) - E\hat{f}_n^{(k)}(x)] \xrightarrow{\mathcal{D}} N(0,\ \lambda_{k,K} f(x)),$$

where $\lambda_{k,K} = \int (K^{(k)}(x))^2 \mathrm{d}x$,

(b) *if $h = O(n^{-1/(2k+2\rho+1)})$, then for all $x \in \mathbb{R}$,*

$$\sqrt{nh^{2k+1}}[\hat{f}_n^{(k)}(x) - f^{(k)}(x)] \xrightarrow{\mathcal{D}} N(b_{k,\rho}(x),\ \lambda_{k,K} f(x))$$

where $b_{k,\rho}(x) = c'\cdot\delta_{\rho,K} f^{(k+\rho)}(x)$ for some constant c'. If $h = c\cdot n^{-1/(2k+2\rho+1)}$ then $c' = c^{k+\rho+1/2}$.

Proofs

Proofs are arranged in the order that makes proofs proceed naturally without referring later results. We first give proofs of Lemma 1 and Theorem 4 because they are used in proving Theorems 1 and 2.

Proof of Lemma 1 Result (a) can be easily given by the stationarity of (X_n), and using Taylor's formula. To find the asymptotic variance of (b), let $K_i^{(k)} = K^{(k)}((x - X_i)/h)$ for simplicity. We have $\operatorname{Var} \hat{f}_n^{(k)}(x) =$

$$\frac{1}{n^2 h^{2k+2}} \operatorname{Var}\left[\sum_{i=1}^{n} K_i^{(k)}\right] = \frac{1}{nh^{2k+2}} \operatorname{Var} K_1^{(k)} + \frac{2}{nh^{2k+2}} \sum_{i=1}^{n} \left(1 - \frac{i}{n}\right) \operatorname{Cov}\left(K_1^{(k)}, K_{1+i}^{(k)}\right).$$

We observe that

$$\frac{1}{h} \operatorname{Var} K_1^{(k)} = f(x) \int (K^{(k)}(x))^2 \mathrm{d}x + O(h),$$

since $EK_1^{(k)} = O(h^{k+1})$ and $E(K_1^{(k)})^2 = \int \left(K^{(k)}\left(\frac{x-z}{h}\right)\right)^2 f(z)\,\mathrm{d}z = \int \left(K^{(k)}(s)\right)^2 \times f(x - hs)\, h\, \mathrm{d}s = h \int \left(K^{(k)}(s)\right)^2 f(x) \mathrm{d}s + O(h^2)$.

We now observe that

$$|\Delta_h| := \left| \frac{1}{h} \sum_{i=1}^{n} \left(1 - \frac{i}{n}\right) \operatorname{Cov}(K_1^{(k)}, K_{1+i}^{(k)}) \right|$$

which is less than

$$\frac{1}{h} \sum_{i=1}^{n} |\operatorname{Cov}(K_1^{(k)}, K_{1+i}^{(k)})| \leq \frac{1}{h} \sum_{i<m} |\operatorname{Cov}(K_1^{(k)}, K_{1+i}^{(k)})| + \frac{1}{h} \sum_{i \geq m} |\operatorname{Cov}(K_1^{(k)}, K_{1+i}^{(k)})| \tag{11}$$

for any sequence $m := m_n$. In the first term on the righthand side of (11), we use condition (3) to get $|\operatorname{Cov}(K_1^{(k)}, K_{1+i}^{(k)})| \leq c_1 h^2 (\int |K^{(k)}(x)| \mathrm{d}x)^2$ for some c_1. Thus, the first term is less than or equal to $c_1 m h (\int |K^{(k)}(x)| \mathrm{d}x)^2$. In the second term of the right side of (11), under the assumption (4) of the weak dependence, we have $|\operatorname{Cov}(K_1^{(k)}, K_{1+i}^{(k)})| \leq c_2 \frac{1}{h^2} \operatorname{Lip}^2(K^{(k)}) \epsilon_i$ for some c_2 and hence $\frac{1}{h} \sum_{i \geq m} |\operatorname{Cov}(K_1^{(k)}, K_{1+i}^{(k)})| \leq \frac{c_2}{h^3} \operatorname{Lip}^2(K^{(k)}) \sum_{i=m}^{\infty} \epsilon_i$. Thus the desired result in Lemma 1(b) holds with inequality $\Delta_h \leq c_1 m h (\int |K^{(k)}(x)| \mathrm{d}x)^2 + \frac{c_2}{h^3} \operatorname{Lip}^2(K^{(k)}) \sum_{i=m}^{\infty} \epsilon_i$. □

Proof of Theorem 4 For the compact interval I [working as in Roussas (1988, 1990)], we divide I into b_n subintervals of each length δ_n with $b_n \delta_n \leq C$ for some constant C, $\delta_n \to 0$, and $\delta_n \log n = O(1/h^3)$ as $n \to \infty$. Let x_{nl} be an arbitrary point in the lth subinterval, $l = 1, 2, \ldots, b_n$.

To show (a), we use Chebyshev inequality, and Lemma 1(b). For $0 < M < \infty$,

$$
\begin{aligned}
&P\left(\sup_{x\in I}|\hat{f}_n^{(k)}(x) - E\hat{f}_n^{(k)}(x)| \geq M\sqrt{\frac{\log n}{nh^{2k+1}}}\right)\\
&\leq \sum_{l=1}^{b_n} P\left(|\hat{f}_n^{(k)}(x_{nl}) - E\hat{f}_n^{(k)}(x_{nl})| \geq M\sqrt{\frac{\log n}{nh^{2k+1}}}\right)\\
&= \sum_{l=1}^{b_n} P\left(\frac{|\hat{f}_n^{(k)}(x_{nl}) - E\hat{f}_n^{(k)}(x_{nl})|}{\sqrt{Var(\hat{f}_n(x_{nl}))}}\right.\\
&\left.\geq \frac{M\sqrt{\log n}}{\sqrt{f(x_{nl})\int (K^{(k)}(x))^2 \mathrm{d}x + c_1 mh(\int |K^{(k)}(x)|\mathrm{d}x)^2 + c_2 \mathrm{Lip}^2(K^{(k)})\sum_{i=m}^{\infty}\epsilon_i/h^3}}\right)\\
&\leq \frac{1}{M^2\log n}\sum_{l=1}^{b_n}\left[f(x_{nl})\int (K^{(k)}(x))^2\mathrm{d}x\right.\\
&\left.+c_1 mh\left(\int |K^{(k)}(x)|\mathrm{d}x\right)^2 + \frac{c_2}{h^3}\mathrm{Lip}^2(K^{(k)})\sum_{i=m}^{\infty}\epsilon_i\right]\\
&\leq \frac{C_1}{\delta_n \log n} + \frac{C_2 mh}{\delta_n\log n} + \frac{C_3\sum_{i=m}^{\infty}\epsilon_i}{h^3\delta_n\log n}
\end{aligned}
$$

for generic constants C_1, C_2, C_3. Since $\delta_n \log n = O(1/h^3) \to \infty$, the first term tends to zero. For the second and third terms, we choose the sequence $m = (\delta_n \log n/h)^{1-\nu}$ for some $0 < \nu < 1$, then $m \to \infty$, $mh/(\delta_n \log n) \to 0$ and $\sum_{i=m}^{\infty}\epsilon_i \to 0$. Thus, result (a) holds. Note that this result is shown without any condition on ϵ_r. Result (b) follows from Lemma 1(a).

To show (c), we apply Lemma 2 above. Let $Y_i(x) = K^{(k)}((x-X_i)/h) - EK^{(k)}((x-X_i)/h)$, and $S_n(x) = \sum_{i=1}^{n} Y_i(x)$, (script k is omitted on $Y_i(\cdot)$ and $S_n(\cdot)$ for notational simplicity). Then, $\hat{f}_n^{(k)}(x) - E\hat{f}_n^{(k)}(x) = S_n(x)/(nh^{k+1})$. For any $0 < M < \infty$, and for $t = M\sqrt{nh\log n}$ in applying Lemma 2,

$$
\begin{aligned}
&P\left(\sup_{x\in I}|\hat{f}_n^{(k)}(x) - E\hat{f}_n^{(k)}(x)| \geq M\sqrt{\frac{\log n}{nh^{2k+1}}}\right)\\
&\leq \sum_{l=1}^{b_n} P\left(|\hat{f}_n^{(k)}(x_{nl}) - E\hat{f}_n^{(k)}(x_{nl})| \geq M\sqrt{\frac{\log n}{nh^{2k+1}}}\right)\\
&= \sum_{l=1}^{b_n} P\left(|S_n(x_{nl})| \geq M\sqrt{nh\log n}\right) \leq C_0 b_n \log n \exp\left(-\frac{t^2}{A_n + B_n t}\right)\\
&= C_0 b_n \log n \exp\left(-\frac{M^2 nh\log n}{A_n + B_n M\sqrt{nh\log n}}\right) \qquad (12)
\end{aligned}
$$

where A_n can be chosen as any number greater than or equal to $Var(S_n)$ and $B_n = n^{3/4}\log n/A_n$. By Lemma 1(b), $Var(S_n(x_{nl})) = nhf(x_{nl})\int (K^{(k)}(x))^2 \mathrm{d}x + 2nh\Delta_h +$

$o(nh)$, and thus we choose $A_n = nh||f||_\infty \int (K^{(k)}(x))^2 \mathrm{d}x + 2c_1 nmh^2 (\int |K^{(k)}(x)|\mathrm{d}x)^2 + 2c_2 nLip^2(K^{(k)}) \sum_{i=m}^{\infty} \epsilon_i / h^2 \geq Var(S_n(x_{nl}))$, then the fraction in the exponent of the right term of (12) is;

$$\frac{M^2 nh \log n}{C_1 nh + C_2 nmh^2 + C_3 n \sum_{i=m}^{\infty} \epsilon_i / h^2 + M n^{3/4} \log n \sqrt{nh \log n} / A_n}$$

for some constants C_1, C_2, C_3 with $C_1 = ||f||_\infty \int (K^{(k)}(x))^2 \mathrm{d}x$. The fraction is greater than or equal to

$$\frac{M^2 nh \log n}{C_1 nh + C_2 nmh^2 + C_3 n \sum_{i=m}^{\infty} \epsilon_i / h^2 + M n^{5/4} (\log n)^{3/2} \sqrt{h} / (C_1 nh)}$$
$$= \frac{M^2 \log n}{C_1 + C_2 mh + C_3 \sum_{i=m}^{\infty} \epsilon_i / h^3 + M (\log n)^{3/2} / (C_1 n^{3/4} h^{3/2})} =: \lambda_n.$$

Note that $\lambda_n / n \to 0$ and $(\log n)^{3/2} / (C_1 n^{3/4} h^{3/2}) \to 0$. Thus

$$\lambda_n \sim \frac{M^2}{C_2 mh + C_3 \sum_{i=m}^{\infty} \epsilon_i / h^3} \log n.$$

Now we choose a sequence $m \to \infty$ such that $p_n := M^2 / (C_2 mh + C_3 \sum_{i=m}^{\infty} \epsilon_i / h^3) \to p$ for some $p > 1$. Under condition $U(m) := \sum_{i=m}^{\infty} \epsilon_i = O(m^{-3})$ in assumption (A5), we may choose $m = O(h^{-1})$, and then $C_2 mh + C_3 \sum_{i=m}^{\infty} \epsilon_i / h^3 \to c$ for some $c > 0$, and thus $p_n \to M^2 / c =: p$. If we choose sufficiently large M such that $M^2 / c > 1$, then $\lambda_n \sim p \log n$ with $p > 1$, and (12) is less than or equal to $C_0 b_n \log n \exp(-\lambda_n)$. Using $b_n = O(1/\delta_n) = O(h^3 \log n)$,

$$\frac{b_n \log n}{e^{\lambda_n}} \sim \frac{h^3 (\log n)^2}{e^{p \log n}} = \frac{h^3 (\log n)^2}{n^p} \quad \text{with } p > 1.$$

Therefore,

$$\sum_{n=1}^{\infty} P\left(\sup_{x \in I} |\hat{f}_n^{(k)}(x) - E \hat{f}_n^{(k)}(x)| \geq M \sqrt{\frac{\log n}{nh^{2k+1}}} \right) \leq C_0 \sum_{n=1}^{\infty} \frac{h^3 (\log n)^2}{n^p} < \infty,$$

and result (c) holds. □

Proof of Theorem 1 Since f is uniform continuous in the interval I and has a unique mode θ in I, proceeding as in Parzen (1962), we have, for $|\theta - \hat{\theta}_n| \geq \varepsilon$, there exists $\delta > 0$ such that $|f(\theta) - f(\hat{\theta}_n)| \geq \delta > 0$. Hence $P(|\theta - \hat{\theta}_n| \geq \varepsilon) \leq P(|f(\theta) - f(\hat{\theta}_n)| \geq \delta)$. By the inequality $|f(\theta) - f(\hat{\theta}_n)| \leq |f(\theta) - \hat{f}_n(\hat{\theta}_n)| + |f(\hat{\theta}_n) - \hat{f}_n(\hat{\theta}_n)| \leq \sup_{x \in I} |f(x) - \hat{f}_n(\hat{\theta}_n)| + \sup_{x \in I} |f(x) - \hat{f}_n(x)| \leq 2 \sup_{x \in I} |f(x) - \hat{f}_n(x)|$, we have $P(|\theta - \hat{\theta}_n| \geq \varepsilon) \leq P\left(\sup_{x \in I} |f(x) - \hat{f}_n(x)| \geq \delta/2 \right)$. By Theorem 4 with $k = 0$, if $h = O((\log n / n)^{1/(2\rho+1)})$, then (10) in Remark 3(a) holds to give us

$\hat{\theta}_n - \theta = o_p(1)$, yielding (a). Also, under addition of condition (A5), according to Remark 3(b), $\hat{\theta}_n - \theta = o_{a.co}(1)$, yielding (b). □

Proof of Theorem 2 To prove Theorem 2, we apply the approach of Shi et al. (2009a). As in the proof of Theorem 1, we apply Theorem 4(a) for proving (a) and Theorem 4(c) for proving (b).

For $0 < M < \infty$, we have

$$P\left(|\hat{\theta}_n - \theta| > M\left(\frac{\log n}{nh^3}\right)^{1/(2\phi-2)}\right)$$
$$\leq P(|\hat{\theta}_n - \theta| \geq \varepsilon) + P\left(|\hat{\theta}_n - \theta| > M\left(\frac{\log n}{nh^3}\right)^{1/(2\phi-2)}, |\hat{\theta}_n - \theta| < \varepsilon\right).$$

Theorem 1 yields $P(|\hat{\theta}_n - \theta| \geq \varepsilon) \to 0$ under (A1)–(A4), and $\sum_{n=1}^{\infty} P(|\hat{\theta}_n - \theta| \geq \varepsilon) < \infty$ under (A1)–(A5). We now consider the other term. Let $Q_n = \left\{x \in I \,\middle|\, M\left(\frac{\log n}{nh^3}\right)^{1/(2\phi-2)} < |x-\theta| < \varepsilon\right\}$. By definition of mode estimator, we have

$$P\left(|\hat{\theta}_n - \theta| > M\left(\frac{\log n}{nh^3}\right)^{1/(2\phi-2)}, |\hat{\theta}_n - \theta| < \varepsilon\right)$$
$$= P[\hat{\theta}_n \in Q_n] \leq P\left(\sup_{x\in Q_n} \hat{f}_n(x) \geq \hat{f}_n(\theta)\right)$$
$$\leq P\left(\sup_{x\in Q_n} \frac{\hat{f}_n(x) - E\hat{f}_n(x) - (\hat{f}_n(\theta) - E\hat{f}_n(\theta))}{|x-\theta|} \geq \inf_{x\in Q_n} \frac{E\hat{f}_n(\theta) - E\hat{f}_n(x)}{|x-\theta|}\right) =: p_n.$$

We will show

$$p_n \leq P\left(\sup_{x\in I} |\hat{f}_n^{(1)}(x) - E\hat{f}_n^{(1)}(x)| \geq c\sqrt{\frac{\log n}{nh^3}}\right). \tag{13}$$

Then under (A1)–(A4), by Theorem 4(a) with $k = 1$, we have $p_n \to 0$ and thus $\hat{\theta}_n - \theta = O_p\left(\left(\log n/(nh^3)\right)^{1/(2\phi-2)}\right)$, arriving at (a). Under (A1)–(A5), by Theorem 4(c) with $k = 1$, we have $\sum_{n=1}^{\infty} p_n < \infty$ and hence $\sum_{n=1}^{\infty} P\left(|\hat{\theta}_n - \theta| > M\left(\frac{\log n}{nh^3}\right)^{1/(2\phi-2)}\right) < \infty$ and thus $\hat{\theta}_n - \theta = O_{a.co}\left(\left(\log n/(nh^3)\right)^{1/(2\phi-2)}\right)$, arriving at (b). It remains to show (13).

By Theorem 4 with $k = 0$, we can write

$$\sup_{x\in Q_n} \frac{\hat{f}_n(x) - E\hat{f}_n(x) - (\hat{f}_n(\theta) - E\hat{f}_n(\theta))}{|x-\theta|} = \sup_{x\in Q_n} \frac{\hat{f}_n(x) - f(x) - (\hat{f}_n(\theta) - f(\theta))}{|x-\theta|} + o(1),$$

and, again by Theorem 4 with $k = 1$, we have $\sup_{x\in Q_n} |\hat{f}_n^{(1)}(x) - f^{(1)}(x)| = \sup_{x\in Q_n} |\hat{f}_n^{(1)}(x) - E\hat{f}_n^{(1)}(x)| + o(1)$. Thus,

$$\sup_{x\in Q_n} \frac{\hat{f}_n(x) - E\hat{f}_n(x) - (\hat{f}_n(\theta) - E\hat{f}_n(\theta))}{|x-\theta|} \leq \sup_{x\in I} |\hat{f}_n^{(1)}(x) - E\hat{f}_n^{(1)}(x)| + o(1). \tag{14}$$

On the other hand, we consider $[E\hat{f}_n(\theta) - E\hat{f}_n(x)]/|x-\theta|$, which can be splitted into following three parts, using Taylor's expansion, $E\hat{f}_n(x) = f(x) + h^\rho \delta_{\rho,K} f^{(\rho)}(x) + o(h^\rho)$:

$$\begin{aligned}\frac{E\hat{f}_n(\theta) - E\hat{f}_n(x)}{|x-\theta|} &= \frac{f(\theta)-f(x)}{|x-\theta|} + O(h^\rho)\frac{f^{(\rho)}(x) - f^{(\rho)}(x)}{|x-\theta|} + o(h^\rho)\frac{1}{|x-\theta|} \\ &:= T_1 + T_2 + T_3.\end{aligned}$$

Without loss of generality, we only consider the case that $x > \theta$ for $x \in Q_n$. By (5) and by application of the Taylor theorem, we have $T_1 \geq c|x-\theta|^{\phi-1} \geq cM\sqrt{\frac{\log n}{nh^3}}$. By Hölder continuity of f, we have $T_2 = O(h^\rho)/|x-\theta|$. Since $x \in Q_n$,

$$\frac{h^\rho}{|x-\theta|} < \frac{h^\rho}{M}\left(\frac{nh^3}{\log n}\right)^{1/(2\phi-2)} = o\left(\sqrt{\frac{\log n}{nh^3}}\right),$$

the last equality holding under condition $h = o((\log n/n)^{\phi/(2\rho(\phi-1)+3\phi)})$. Thus, for generic c,

$$\frac{E\hat{f}_n(\theta) - E\hat{f}_n(x)}{|x-\theta|} \geq cM\sqrt{\frac{\log n}{nh^3}} + o\left(\sqrt{\frac{\log n}{nh^3}}\right), \quad \inf_{x\in Q_n} \frac{E\hat{f}_n(\theta) - E\hat{f}_n(x)}{|x-\theta|} \geq c\sqrt{\frac{\log n}{nh^3}}. \tag{15}$$

By (14) and (15), we obtain the result in (13). □

To prove Theorem 3, we need the results of kernel estimator of density derivatives in Lemma 1, Theorem 4 and Theorem 5. Therefore, proof of Theorem 5 is given prior to that of Theorem 3.

Proof of Theorem 5 Let $S_n = \sqrt{nh^{2k+1}}[\hat{f}_n^{(k)}(x) - E\hat{f}_n^{(k)}(x)] \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n L_{ni}^{(k)}$ where $L_{ni}^{(k)} = \frac{1}{\sqrt{h}}[K^{(k)}(\frac{x-X_i}{h}) - EK^{(k)}(\frac{x-X_i}{h})]$. Let $\alpha = \alpha_n$ and $\beta = \beta_n$ be some integer-valued sequences with $\alpha \to \infty$, $\beta \to \infty$ and $\frac{\beta}{\alpha} + \frac{\alpha}{n} \to 0$ as $n \to \infty$. Let $\tau = \lfloor n/(\alpha+\beta)\rfloor$ be the integer part of $n/(\alpha+\beta)$. For $j = 1, \ldots, \tau$, we define $U_j = \frac{1}{\sqrt{\alpha}}\sum_{i\in I_j} L_{ni}^{(k)}$ with $I_j = [\,(j-1)(\alpha+\beta)+1,\ (j-1)(\alpha+\beta)+\alpha\,]$ and define $T_\tau = \frac{1}{\sqrt{\tau}}\sum_{j=1}^\tau U_j$.

To prove (a), we use three steps below. Then according to Step 1, $\prod_{j=1}^\tau E\exp(\iota t U_j/\sqrt{\tau}) \to \exp(-\frac{t^2}{2}\lambda_{k,K} f(x))$ as $\tau \to \infty$, (where $\iota = \sqrt{-1}$). From Step 2, $T_\tau \xrightarrow{\mathcal{D}}$

$N(0, \lambda_{k,K} f(x))$ as $\tau \to \infty$ and hence $S_n \xrightarrow{\mathcal{D}} N(0, \lambda_{k,K} f(x))$ as $n \to \infty$ by Step 3.

Step 1 Shows that $(\text{Var}U_1)^{-3/2} \sum_{j=1}^{\tau} E|U_j/\sqrt{\tau}|^3 \to 0$ as $n \to \infty$. By the similar argument as in the proof of Lemma 1, we have $\text{Var}U_j = \text{Var}U_1 = \frac{1}{h}\text{Var}K_1^{(k)} + o(1) \to f(x)\int (K^{(k)}(x))^2 \text{d}x = f(x)\lambda_{k,K}$. Also, by the stationarity and by Hölder inequality, we have

$$\sum_{j=1}^{\tau} E\left|\frac{U_j}{\sqrt{\tau}}\right|^3 = \frac{E|U_1|^3}{\sqrt{\tau}} \le \frac{[E|U_1|^4]^{3/4}}{\sqrt{\tau}}.$$

By Marcinkiewicz–Zygmund inequality under the weak dependence (see Doukhan and Louhichi 1999), $E|U_1|^4 = \frac{1}{\alpha^2} E[\sum_{i=1}^{\alpha} L_{ni}^{(k)}]^4 \le \frac{1}{\alpha^2} \cdot B\alpha^2 = B$ for some positive constant B not depending on n and α. Thus, the desired result is completed.

Step 2 Shows that $|E\exp(\iota t T_\tau) - \prod_{j=1}^{\tau} E\exp(\iota t[U_j/\sqrt{\tau}])| \to 0$ as $n \to \infty$, $(\iota = \sqrt{-1})$. Note that the left term is less than or equal to

$$\left|\text{Cov}\left(\exp(\iota t T_{\tau-1}), \exp\left(\iota t \frac{U_\tau}{\sqrt{\tau}}\right)\right)\right| + \left|E\exp(\iota t T_{\tau-1}) - \prod_{j=1}^{\tau-1} E\exp\left(\iota t \frac{U_j}{\sqrt{\tau}}\right)\right|.$$

Also note that $\exp(\iota t T_{\tau-1})$ and $\exp(\iota t U_\tau/\sqrt{\tau})$ can be written, respectively, as $h(X_0, \ldots, X_{(\tau-2)(\alpha+\beta)+\alpha})$ and $g(X_{(\tau-1)(\alpha+\beta)+1}, \ldots, X_{(\tau-1)(\alpha+\beta)+\alpha})$ for some measurable functions h and g, (with $||h||_\infty = ||g||_\infty \le 1$), composed of the exponential function and the kth derivative of the kernel function.

Under $(\epsilon, \mathcal{L}, \psi_1)$-weak dependence,

$$\left|\text{Cov}\left(\exp(\iota t T_{\tau-1}), \exp\left(\iota t \frac{U_\tau}{\sqrt{\tau}}\right)\right)\right| \le \psi_1(h, g, \alpha(\tau-1), \alpha)\epsilon_\beta \le \frac{t^2}{\tau h^3}\epsilon_\beta$$

since $\min(\alpha(\tau-1), \alpha) = \alpha$ and $\text{Lip}(h) = \text{Lip}(g) \le t/\sqrt{\alpha\tau h^3}$. Thus, we have

$$\left|E\exp(\iota t T_\tau) - \prod_{j=1}^{\tau} E\exp\left(\iota t \frac{U_j}{\sqrt{\tau}}\right)\right| \le \frac{t^2}{h^3}\epsilon_\beta$$

which tends to 0 for $\epsilon_\beta = O(\beta^{-12-\nu})$ if we choose $\beta \sim h^{-1}$ and $\alpha \sim h^{-1-2k}$.

Under $(\epsilon, \mathcal{L}, \psi_2)$-weak dependence,

$$\left|\text{Cov}\left(\exp(\iota t T_{\tau-1}), \exp\left(\iota t \frac{U_\tau}{\sqrt{\tau}}\right)\right)\right| \le \psi_2(h, g, \alpha(\tau-1), \alpha)\epsilon_\beta \le 4\alpha\tau \frac{t}{\sqrt{\alpha\tau h^3}}\epsilon_\beta.$$

Thus we have

$$\left| E\exp(\iota t T_\tau) - \prod_{j=1}^{\tau} E\exp\left(\iota t \frac{U_j}{\sqrt{\tau}}\right) \right| \le 4t\sqrt{\frac{\alpha\tau^3}{h^3}}\epsilon_\beta.$$

If we choose $\beta \sim n^b$ and $\alpha \sim n^{b+1/2k}$ for $2/(9+\nu) < b < 3/(9+\nu)$, then

$$\sqrt{\frac{\alpha\tau^3}{h^3}}\epsilon_\beta \sim \left(\frac{1}{\alpha+\beta}\right)\sqrt{\frac{\alpha}{\alpha+\beta}}\left(\frac{n}{h}\right)^{3/2}\frac{1}{\beta^{9+\nu}}$$

which tends to 0 since $1/\sqrt{(nh)^3} < n^{3/2}/(h^{3/2}\beta^{9+\nu}) < 1/\sqrt{nh^3}$.

Step 3 Shows that $S_n - T_\tau = o_p(1)$ as $n \to \infty$, $(\tau \to \infty)$. We write $S_n - T_\tau =$

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} L_{ni}^{(k)} - T_\tau &= \frac{1}{\sqrt{n}}\left(\sum_{j=1}^{\tau}\sum_{i\in I_j} L_{ni}^{(k)} + \sum_{j\in J} L_{nj}^{(k)}\right)\\ -T_\tau &= \left(\frac{\sqrt{\alpha\tau}}{\sqrt{n}} - 1\right)T_\tau + \frac{1}{\sqrt{n}}\sum_{j\in J} L_{nj}^{(k)} \equiv V_1 + V_2\end{aligned}$$

where J is a set of cardinal $n - \alpha\tau$ $(\sim \beta\tau)$. Thus, $\mathrm{Var}(S_n - T_\tau) \le 2(\mathrm{Var}V_1 + \mathrm{Var}V_2)$. Clearly, $E(S_n - T_\tau) \to 0$. Observe that

$$\mathrm{Var}V_1 = \left(\frac{\sqrt{\alpha\tau}}{\sqrt{n}} - 1\right)^2 \mathrm{Var}T_\tau \quad \text{and} \quad \mathrm{Var}V_2 = \frac{|J|}{n}\mathrm{Var}\left(\frac{1}{\sqrt{|J|}}\sum_{j\in J} L_{nj}^{(k)}\right).$$

Note that $\alpha\tau/n \to 1$ and $|J|/n \to 0$. Also, by the similar argument as in the proof of Lemma 1, we have

$$\mathrm{Var}T_\tau = \frac{1}{\alpha\tau}\mathrm{Var}\left[\sum_{j=1}^{\tau}\sum_{i\in I_j} L_{ni}^{(k)}\right] = \frac{1}{\alpha\tau h}\mathrm{Var}\left[\sum_{j=1}^{\tau}\sum_{i\in I_j} K_i^{(k)}\right] = \frac{1}{h}\mathrm{Var}K_1^{(k)} + o(1) < \infty,$$

and

$$\mathrm{Var}\left(\frac{1}{\sqrt{|J|}}\sum_{j\in J} L_{nj}^{(k)}\right) = \frac{1}{h}\mathrm{Var}K_1^{(k)} + o(1) < \infty.$$

Hence, $\mathrm{Var}(S_n - T_\tau) = o(1)$, and we complete Step 3, and thus part (a). Result (b) follows from Lemma 1 and (a). □

Proof of Theorem 3 By (7), we write

$$\sqrt{nh^3}(\hat{\theta}_n - \theta) = -\frac{1}{\hat{f}_n^{(2)}(\theta_n^*)}\left[\sqrt{nh^3}\{\hat{f}_n^{(1)}(\theta) - E\hat{f}_n^{(1)}(\theta)\}\right] - \sqrt{nh^3}\left(\frac{E\hat{f}_n^{(1)}(\theta)}{\hat{f}_n^{(2)}(\theta_n^*)}\right).$$

We first show that $\hat{f}_n^{(2)}(\theta_n^*) \xrightarrow{p} f^{(2)}(\theta)$. Since θ_n^* is between $\hat{\theta}_n$ and θ, we have $\theta_n^* \xrightarrow{p} \theta$ as $n \to \infty$: for $\varrho > 0$, we have $P(|\theta_n^* - \theta| > \varrho) \to 0$. Let $J' := [\theta - 2\varrho, \theta + 2\varrho]$. According to Theorem 4(a) and Remark 3(a) with $k = 2$, $\sup_{x\in J} |\hat{f}_n^{(2)}(x) - f^{(2)}(x)| \xrightarrow{p} 0$. Therefore, $|\hat{f}_n^{(2)}(\theta_n^*) - f^{(2)}(\theta)| \le |\hat{f}_n^{(2)}(\theta_n^*) - f^{(2)}(\theta_n^*)| + |f^{(2)}(\theta_n^*) - f^{(2)}(\theta)| \xrightarrow{p} 0$. Hence

$$\sqrt{nh^3}(\hat{\theta}_n - \theta) \cong -\frac{1}{f^{(2)}(\theta)}\left[\sqrt{nh^3}\{\hat{f}_n^{(1)}(\theta) - E\hat{f}_n^{(1)}(\theta)\}\right] - \sqrt{nh^3}\left(\frac{E\hat{f}_n^{(1)}(\theta)}{f^{(2)}(\theta)}\right).$$

The first term in the right-hand side converges to $N(0, \lambda_{1,K} f(\theta)/[f^{(2)}(\theta)]^2)$ in distribution by Theorem 5(a) with $k = 1$, while the second term tends to $-c^{(2\rho+3)/2} \cdot \delta_{\rho,K} f^{(1+\rho)}(\theta)/f^{(2)}(\theta)$ as $n \to \infty$ if $h = c \cdot n^{-1/(2\rho+3)}$ by Lemma 1. The desired result is obtained. □

Acknowledgments The authors are very grateful to two anonymous referees for many helpful comments which lead us to a substantial improvement. This work was supported by Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology(2012-000 6691).

References

Ango Nze, P. A., Doukhan, P. (2004). Weak dependence: models and applications to econometrics. *Econometric Theory*, *20*, 995–1045.

Ango Nze, P. A., Bühlmann, P., Doukhan, P. (2002). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Annals of Statistics*, *30*, 397–430.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, *71*, 353–360.

Coulon-Prieur, C., Doukhan, P. (2000). A triangular central limit theorem under a new weak dependent condition. *Statistics and Probability Letters*, *47*, 61–68.

Dedecker, J., Prieur, C. (2004). Coupling for τ-dependent sequences and applications. *Journal of Theoretical Probabability*, *17*, 861–885.

Dedecker, J., Doukhan, P., Lang, G., Leon, R., Jose Rafael, R., Louhichi, S., Prieur, C. (2007). *Weak dependence: with examples and applications* (Vol. 190). Lecture Notes in Statistics. New York: Springer.

Douge, L. (2007). Convergence rates in the law of large numbers and in density estimation for associated random variables. *Comptes Rendus de l'Academie des Sciences-Series I-Mathematics*, *344*(8), 515–518.

Doukhan, P., Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, *84*, 313–342.

Doukhan, P., Louhichi, S. (2001). Functional estimation of a density under a new weak dependence condition. *Scandinavian Journal of Statistics*, *28*, 325–341.

Doukhan, P., Neumann, M. H. (2007). Probability and moment inequalities for sums of weakly dependent random variables with applications. *Stochatis Processes and their Applications*, *117*, 878–903.

Doukhan, P., Mayo, N., Truquet, L. (2009). Weak dependence, models and some applications. *Metrika*, *69*, 199–225.

Eddy, W. F. (1980). Optimum kernel estimators of the mode. *Annals of Statistics*, *8*, 870–882.
Eddy, W. F. (1982). The asymptotic distributions of kernel estimators of the mode. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *59*(3), 279–290.
Ferraty, F., Laksaci, A., Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*, *9*, 47–76.
Gasser, T., Müller, H. G. (1979). Kernel estimation of regression functions. In T. Gasser, M. Rosenblatt (Eds.), *Smoothing techniques for curve estimation*. Heidelberg: Springer.
Györfi, L., Härdle, W., Sarda, P., Vieu, P. (1989). *Nonparametric curve estimation from time series* (Vol. 60). Lecture Notes in Statistics. New York: Springer.
Herrmann, E., Ziegler, K. (2004). Rates of consistency for nonparametric estimation of the mode in absence of smoothness assumptions. *Statistics and Probability Letters*, *68*, 359–368.
Hwang, E., Shin, D. W. (2012). Stationary bootstrap for kernel density estimators under ψ-weak dependence. *Compurarional Statistics and Data Analysis*, *56*, 1581–1593.
Hwang, E., Shin, D. W. (2013). A study on moment inequalities under a weak dependence. *Journal of the Korean Statistical Society*, *42*, 133–141.
Hwang, E., Shin, D. W. (2014). A note on exponential inequalities of ψ-weakly dependent sequences. *Communications for Statistical Applications and Methods*, *21*(3), 245–251.
Jones, M. C. (1994). On kernel density derivative estimation. *Communications in Statistics-Theory and Methods*, *2*(8), 2133–2139.
Kallabis, R. S., Neumann, M. H. (2006). An exponential inequality under weak dependence. *Bernoulli*, *12*, 333–350.
Louani, D. (1998). On the asymptotic normality of the kernel estimators of the density function and its derivatives under censoring. *Communications in Statistics-Theory and Methods*, *27*(12), 2909–2924.
Louani, D., Ould-Saïd, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *Journal of Nonparametric Statististics*, *11*(4), 413–442.
Meister, A. (2011). On general consistency in deconvolution mode estimation. *Journal of Statistical Planning and Inference*, *141*, 771–781.
Mokkadem, A., Pelletier, M. (2005). Moderate deviations for the kernel mode estimator and some applications. *Journal of Statistical Planning and Inference*, *135*, 276–299.
Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, *33*, 1065–1076.
Rio, A. Q. (1997). Nonparametric estimation of density derivatives of dependent data. *Journal of Statistical Planning and Inference*, *61*, 155–174.
Romano, J. (1988). On weak convergence and optimality of kernel density estimators of the mode. *Annals of Statististics*, *16*, 629–647.
Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random variables. *Journal of Statistical Planning and Inference*, *18*(2), 135–149.
Roussas, G. G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Processes and their Applications*, *36*(1), 107–116.
Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, *9*, 65–78.
Samanta, M., Thavaneswaran, A. (1990). Nonparametric estimation of the conditional mode. *Communications in Statistics-Theory and Methods*, *14*, 1123–1136.
Sheater, S. J., Jones, M. C. (1991). A relaible data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society Series B. Statistical Methodology*, *53*, 683–690.
Shi, X., Wu, Y., Miao, B. (2009a). A note of the convergence rate of the kernel density estimator of the mode. *Statistics and Probability Letters*, *79*, 1866–1871.
Shi, X., Wu, Y., Miao, B. (2009b). Strong convergence rate of estimators of change point and its application. *Computational Statistics and Data Analysis*, *53*(4), 990–998.
Silveran, B. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Annals of Statistics*, *6*, 177–184.
Singh, R. S. (1977). Improvement on some known nonparametric uniformly consistent estimator of derivatives of a density. *Annals of Statistics*, *5*(1), 61–120.
Vieu, P. (1996). A note on density mode estimation. *Statistics and Probability Letters*, *26*, 297–307.
Wieczorek, B., Ziegler, K. (2010). On optimal estimation of a non-smooth mode in a nonparametric regression model with α-mixing errors. *Journal of Statistical Planning and Inference*, *140*(2), 406–418.