An Empirical Likelihood Goodness-of-Fit Test for Time Series

Song Xi Chen
Department of Statistics and Applied Probability
National University of Singapore
Singapore 117543
stacsx@nus.edu.sg

Wolfgang Härdle and Torsten Kleinow
Institut für Statistik und Ökonometrie
Humboldt-Universität zu Berlin
10178 Berlin, Germany

December 23, 2000

ABSTRACT. The testing of a computing model for a stationary time series is a standard task in statistics. When a parametric approach is used to model the time series, the question of goodness-of-fit arises. In this paper, we employ the empirical likelihood for an α-mixing process and formulate a statistic that measures the goodness-of-fit of a parametric model. The technique is based on comparison with kernel smoothing estimators. The goodness-of-fit test proposed is based on the asymptotics of the empirical likelihood, which has two attractive features. One is its automatic consideration of the variation associated with the nonparametric fit due to the empirical likelihood's ability to studentise internally. The other one is that the asymptotic distributions of the test statistic are free of unknown parameters which avoids secondary plug-in estimation. We apply the empirical likelihood based test to a discretised diffusion model which has been recently considered in financial market analysis.

KEY WORDS: Empirical likelihood; Goodness-of-Fit Test; Nadaraya-Watson Estimator; Parametric Models; Power of Test; Square Root Processes; α-mixing; Weakly Dependence.

1 Introduction

The analysis and prediction of time series is standard work in statistics. The techniques employed though rely on the actual model assumed to represent and generate the dynamics of the time series. Mismodelling might result in biased prediction and incorrect parameter specification. The aim of this paper is to show how the empirical likelihood technique (Owen, 1988, 1990) may be used to construct simple test procedures for the goodness-of-fit of standard time series models.

We assume that $\{(X_i, Y_i)\}_{i=1}^n$ is a strictly stationary time series with $Y_i \in \mathbb{R}$ and $X_i \in \mathbb{R}^d$. The explanatory variable X might be the lagged d-dimensional past of a one dimensional time series for example. The conditional mean for an AR modelled process is in this case a linear function of X. In a financial time series setting, the explanatory variables X might control the conditional volatility of an ARCH type process. Standard parametric approaches assume a linear structure in the squares of the past values. This specific form however introduces symmetric shocks as a reaction to recent return values (a so called news impact). Empirical observations though indicate that volatility shocks are more pronounced after negative returns in the period before, see e.g. Engle and Gonzalez-Rivera (1991), Zakoian (1991), Gouriéroux and Monfort (1992) and Gouriéroux (1997). Diffusion models are important in mathematical finance. Interest rates, stock and other financial products are modelled by discretised diffusion processes with specific parametric assumptions on the drift and scale functions, see e.g. Karatzas and Shreve (1998) and Platen (1999). Genon-Catalot, Jeantheau and Laredo (2000) model the stochastic volatility as a diffusion process. They use hidden markov model techniques for the statistical analysis. These three types of time series models, along with many others, may be evaluated and tested for their specific form by the proposed empirical likelihood test.

Figure 1 shows the daily closing value of the S&P 500 share index from the 31st December 1976 to the 31st December 1997, which covers 5479 trading days. In the upper panel, the index series shows a trend of exponential form which is estimated using the method given in Härdle et al. (2000). The lower panel is a residual series after removing the exponential trend. In mathematical finance, one assumes a specific dynamic form of this residual series.

More precisely, Härdle et al. (2000) assume the following model for an index process $S(t)$

$$S(t) = S(0)X(t)\exp\left(\int_0^t \eta(s)ds\right) \tag{1.1}$$

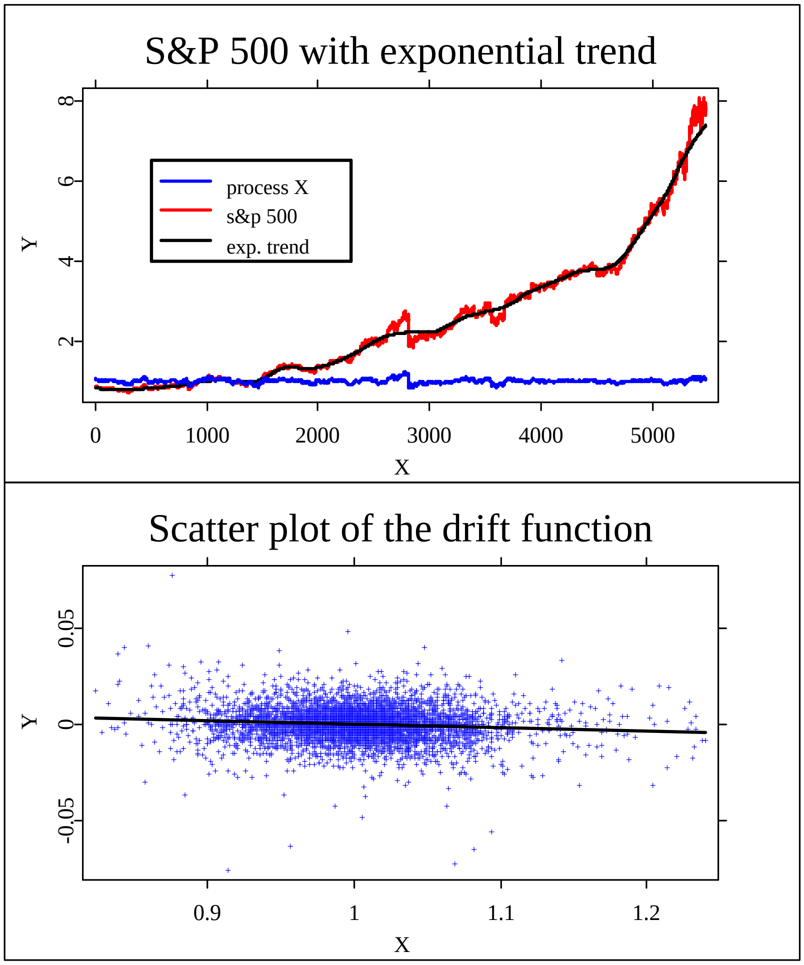

Figure 1. The S&P 500 Data.

with a diffusion component $X(t)$ solving the stochastic differential equation

$$dX(t) = \alpha\{1 - X(t)\}dt + \sigma X^{1/2}(t)dW(t) \tag{1.2}$$

where $W(t)$ is a Brownian motion and α and σ are parameters. Discretising this series with a sampling interval Δ leads to the observations (X_i, Y_i) with $Y_i = X_{(i+1)\Delta} - X_{i\Delta}$ and

$X_i = X_{i\Delta}$, which will be α-mixing and fulfill all the other conditions assumed in the paper based on the results given in Genon-Catalot, Jeantheau and Laredo (2000). Note that this series has a mean function as in Cox, Ingersol and Ross (1985).

Let $m(x) = E(Y|X = x)$ be the conditional mean function, f be the density of the design points X, and $\sigma^2(x) = Var(Y|X = x)$ be the conditional variance function of Y given $X = x \in S$, a set to be specified later. Suppose that $\{m_\theta|\theta \in \Theta\}$ is a parametric model for the mean function m and that $\hat{\theta}$ is an estimator of θ under this parametric model. The interest is to test the null hypothesis:

$$H_0 : m(x) = m_\theta(x) \quad \text{for all}\, x \in S$$

against a series of local smooth nonparametric alternatives:

$$H_1 : m(x) = m_\theta(x) + c_n\Delta_n(x),$$

where c_n is a non-random sequence tending to zero as $n \to \infty$ and $\Delta_n(x)$ is a sequence of bounded functions.

The problem of testing against a nonparametric alternative is not new for an independent and identically distributed setting, see e.g. Härdle and Mammen (1993), Hart (1997) and Horowitz (1997). In a time series context the testing procedure for functional form has been considered in Kreiss, Neumann and Yao (1998) and Herwartz (2000) for example. Development here is somewhat slower since theoretical results on kernel estimators for time series have appeared only very recently, see e.g. Bosq (1998). This is surprising given the interests in time series for financial engineering.

The device we consider here to formulate goodness-of-fit tests for H_0 against H_1 is the empirical likelihood. The empirical likelihood is a computer intensive nonparametric statistical method of inference introduced by Owen (1988,1990) as an alternative to the bootstrap. Instead of resampling with an equal probability weight for all data values like the bootstrap, the empirical likelihood chooses the weights by profiling a multinomial likelihood under a set of constraints which reflect characteristics of the parameters of interest. The empirical likelihood has been shown to share some key properties with the parametric likelihood, for instance Wilks' theorem and the Bartlett correction; see Hall and La Scala (1990) and Qin and Lawless (1994) for details. The empirical likelihood literature which is related to the topic of this paper are Chen (1994) on testing hypothesis; Chen (1996) and Chen and Qin (2000) on confidence intervals in nonparametric curve estimation; Kitamura (1997) on inference for parameters associated with weakly dependent processes; and Baggerly (1998) for using empirical likelihood as a goodness-of-fit measure associated with a parameter.

The proposed goodness-of-fit test is based on a joint "pseudo" empirical likelihood statistic over a grid of points within S, the domain of the design density f. It is shown that the empirical likelihood test statistic is asymptotically equivalent to a studentised version of the L_2 measure of goodness-of-fit considered in Härdle and Mammen (1993). In the case of heteroscedasticity, when the variance function $\sigma^2(x)$ is not constant, it is crucial to studentise the L_2 distance in constructing the goodness-of-fit test in order to put the L_2 distance in the context of its variation. An unique feature of the proposed test is that the studentisation comes automatically due to the empirical likelihood's ability to studentise internally, which means that no estimation of unknown quantities such as $\sigma^2(x)$ and $f(x)$ is required. It is shown that the proposed test statistic converges to an integral of a squared normal process, which leads to several test procedures including those based on the asymptotic normal and chi-square distributions.

The paper is structured as follows. An empirical likelihood test statistic is established for the testing purpose and its properties are studied in Section 2. The goodness-of-fit test procedures are proposed in Section 3. The proposed tests are applied to evaluate a parametric diffusion model for S&P 500 index data in Section 4, which also include results from a simulation study. All proofs and conditions assumed are given in the appendix.

2 Kernel Estimator and Empirical Likelihood

We first introduce a nonparametric kernel estimator for m. Let $S = \{x \in \mathbb{R}^d | f(x) \geq C_1\}$ for some $C_1 > 0$ be a compact set. Without loss of generality we assume that S is the d-dimensional unit cube, $S = [0,1]^d$.

Let K be a d-dimensional bounded probability density function with a compact support on the d-dimensional cube $[-1,1]^d$ that satisfies moment conditions:

$$\int uK(u)du = 0, \int uu^T K(u)du = \sigma_K^2 \mathcal{I}_d$$

where $\mathcal{I}_d$ is the d-dimension unit matrix and σ_K^2 is a positive constant. Let h be a positive smoothing bandwidth which will be used to smooth in every component of X implying that the scale in each component is roughly the same. When the scale of the variables are different, they can be standardised by their standard deviation. We may also use a general bandwidth matrix without altering the main results of the paper.

Let $K_h(u) = h^{-d}K(h^{-1}u)$. The nonparametric estimator considered is the Nadaraya-Watson (NW) estimator

$$\hat{m}(x) = \frac{\sum_{i=1}^n Y_i K_h(x - X_i)}{\sum_{i=1}^n K_h(x - X_i)}. \tag{2.1}$$

Let

$$\tilde{m}_{\hat{\theta}}(x) = \frac{\sum K_h(x - X_i) m_{\hat{\theta}}(X_i)}{\sum_{i=1}^n K_h(x - X_i)}$$

be the smoothed parametric model. The test statistics we are going to consider are based on the difference between $\tilde{m}_{\hat{\theta}}$ and $\hat{m}$, rather than directly between $\hat{m}$ and $m_{\hat{\theta}}$, in order to avoid the issue of bias associated with the nonparametric fit.

The local linear estimator can be used to replace the NW estimator in estimating m. However, as we compare $\hat{m}$ with $\tilde{m}_{\hat{\theta}}$ in formulating the goodness-of-fit test, the possible bias associated with the NW estimator is not an issue here. In addition, the NW estimator provides a better handling of sparse design for multivariate data and it has a simpler analytic form. Extension of the results to the local linear estimator based test can be derived in a similar fashion, although the proof will be more involved in the multivariate case.

Let us now as in Owen (1988, 1990) introduce the empirical likelihood concept for the testing problem that we consider here. At an arbitrary $x \in S$, let $p_i(x)$ be nonnegative numbers representing weights allocated to each (X_i, Y_i). The empirical likelihood (EL) for $\tilde{m}_{\hat{\theta}}(x)$ is

$$L\{\tilde{m}_{\hat{\theta}}(x)\} = \max \prod_{i=1}^n p_i(x) \tag{2.2}$$

subject to $\sum_{i=1}^n p_i(x) = 1$ and $\sum_{i=1}^n p_i(x) K\left(\frac{x - X_i}{h}\right) \{Y_i - \tilde{m}_{\hat{\theta}}(x)\} = 0$.

By introducing the Lagrange multipliers, we obtain as a solution to (2.2) the optimal weights

$$p_i(x) = n^{-1}\left[1+\lambda(x)K\left(\frac{x-X_i}{h}\right)\{Y_i-\tilde{m}_{\hat{\theta}}(x)\}\right]^{-1} \tag{2.3}$$

where $\lambda(x)$ is the root of

$$\sum_{i=1}^{n}\frac{K\left(\frac{x-X_i}{h}\right)\{Y_i-\tilde{m}_{\hat{\theta}}(x)\}}{1+\lambda(x)K\left(\frac{x-X_i}{h}\right)\{Y_i-\tilde{m}_{\hat{\theta}}(x)\}}=0. \tag{2.4}$$

The maximum EL is achieved at $p_i(x) = n^{-1}$ corresponding to the nonparametric curve estimate $\hat{m}(x)$. The log-EL ratio is

$$\ell\{\tilde{m}_{\hat{\theta}}(x)\} = -2log[L\{\tilde{m}_{\hat{\theta}}(x)\}n^n].$$

To study properties of the empirical likelihood based test statistic we need to evaluate $\ell\{\tilde{m}_{\hat{\theta}}(x)\}$ at an arbitrary x first, which requires the following lemma on $\lambda(x)$ whose proof is given in the appendix.

Lemma 1. Under the assumptions (i)-(vi),

$$\sup_{x\in S}|\lambda(x)| = o_p\{(nh^d)^{-1/2}\log(n)\}.$$

Let $\gamma(x)$ be a random process with $x \in S$. Throughout this paper we use the notation $\gamma(x) = \tilde{O}_p(\delta_n)$ ($\tilde{o}_p(\delta_n)$) to denote the facts that $\sup_{x\in S}|\gamma(x)| = O_p(\delta_n)$ ($o_p(\delta_n)$) for a sequence δ_n.

Let $\bar{U}_j(x) = (nh^d)^{-1}\sum_{i=1}^{n}\left[K\left(\frac{x-X_i}{h}\right)\{Y_i-\tilde{m}_{\hat{\theta}}(x)\}\right]^j$ for $j = 1, 2, \ldots$. From Lemma 1, an expansion from (2.4) yields

$$\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\{Y_i-\tilde{m}_{\hat{\theta}}(x)\}\Big[\sum_{j=0}^{\infty}\lambda^j(x)K^j\left(\frac{x-X_i}{h}\right)\{Y_i-\tilde{m}_{\hat{\theta}}(x)\}^j\Big]=0.$$

Inverting the above expansion, we have

$$\lambda(x) = \bar{U}_2^{-1}(x)\bar{U}_1(x) + \tilde{o}_p\{(nh^d)^{-1}\log^2(n)\}. \tag{2.5}$$

This together with (2.3) and Lemma 1 implies that

$$\ell\{\tilde{m}_{\hat{\theta}}(x)\} \;=\; -2log[L\{\tilde{m}_{\hat{\theta}}(x)\}n^n] = 2\sum_{i=1}^{n}\log[1+\lambda(x)K\left(\frac{x-X_i}{h}\right)\{Y_j-\tilde{m}_{\hat{\theta}}(x)\}]$$

$$
\begin{aligned}
&= 2(nh^d)\lambda(x)\bar{U}_1 - (nh^d)\lambda^2(x)\bar{U}_2 + \tilde{o}_p\{(nh^d)^{-1/2}\log^3(n)\} \\
&= (nh)^d \bar{U}_2^{-1}(x)\bar{U}_1^2(x) + \tilde{o}_p\{(nh^d)^{-1/2}\log^3(n)\}. \qquad (2.6)
\end{aligned}
$$

For any $x \in S$, let $v(x;h) = h^d \int_{y\in S} K_h^2(x-y)dy$ and $b(x;h) = h^d \int_{y\in S} K_h(x-y)dy$ be the variance and the bias coefficient functions associated with the NW estimator, respectively. Let $S_I = \{x = (x_1,\dots,x_d) \in \mathbb{R}^d | \min_i(|x_i - 1|, |x_i|) > h\}$ be the set of interior points in S, and $S_B = S - S_I$ be the boundary set. When $x \in S_I$, $v(x;h) = R(K) =: \int K^2(x)dx$ and $b(x;h) = 1$. Define $V(x;h) = v(x;h)\sigma^2(x)/\{f(x)b^2(x;h)\}$. Clearly, $V(x;h)/(nh^d)$ is the asymptotic variance of $\hat{m}(x)$ when $nh^d \to \infty$ which is one of the conditions we assumed.

It can be shown from Condition (iii) and the proof of Lemma 1 given in the appendix that

$$
\begin{aligned}
\bar{U}_1(x) &= n^{-1}\sum_{i=1}^n K_h(x-X_i)\{Y_i - \tilde{m}_{\hat{\theta}}(x)\} \\
&= n^{-1}\sum_{i=1}^n K_h(x-X_i)\{Y_i - m_\theta(X_i)\} + \tilde{O}_p(n^{-1/2}) \\
&= \hat{f}(x)\{\hat{m}(x) - \tilde{m}_\theta(x)\} + \tilde{O}_p(n^{-1/2}) \\
&= f(x)b(x;h)\{\hat{m}(x) - \tilde{m}_\theta(x)\} + \tilde{O}_p\{n^{-1/2} + (nh^d)^{-1}\log^2(n)\}. \qquad (2.7)
\end{aligned}
$$

From the proof of Lemma 1, we have $\sup_{x\in S}|\bar{U}_2(x) - f(x)v(x;h)\sigma^2(x)| = O_p(h)$. These and (2.6) mean that

$$
\begin{aligned}
\ell\{\tilde{m}_{\hat{\theta}}(x)\} &= \bar{U}_2^{-1}\bar{U}_1^2 + \tilde{o}_p\{(nh^d)^{-1/2}\log^3(n)\} \\
&= V^{-1}(x;h)\{\hat{m}(x) - \tilde{m}_\theta(x)\}^2 + \tilde{O}\{(nh^d)^{-1}h\log^2(n)\} \qquad (2.8)
\end{aligned}
$$

Therefore, $\ell\{\tilde{m}_{\hat{\theta}}(x)\}$ is asymptotically equivalent to a studentised L_2 distance between $\tilde{m}_{\hat{\theta}}(x)$ and $\hat{m}(x)$. It is this property that leads us to use $\ell\{\tilde{m}_{\hat{\theta}}(x)\}$ as the basic building block in the construction of a global test statistic for distinction between $\tilde{m}_{\hat{\theta}}$ and $\hat{m}$ in the next section. The use of the empirical likelihood as a distance measure and its comparison with other distance measures have been discussed in Owen (1991) and Baggerly (1998).

3 Empirical Likelihood Goodness-of-fit Statistic

To extend the EL ratio statistic to a global measure of goodness-of-fit, we choose k_n-equally spaced lattice points $t_1, t_2, \cdots, t_{k_n}$ in $[0,1]^d$ where $t_1 = (0, \cdots, 0)$, $t_{k_n} = (1, \cdots, 1)$ and $||t_i|| \le ||t_j||$ for $1 \le i < j \le k_n$. Here $|| \cdot ||$ is the Euclidian distance in $\mathbb{R}^d$. We let $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. This essentially divides S into k_n small bins (hypercubes) of size $(k_n)^{-1}$. A simple choice is to let $k_n = [(2h)^{-d}]$ where $[a]$ is the largest integer less than a. This choice as justified later ensures asymptotic independence among $\ell\{\tilde{m}_{\hat{\theta}}(t_j)\}$ at different t_js. Bins of different size can be adopted to suit situations where there are areas of low design density. This corresponds to the use of different bandwidth values in adaptive kernel smoothing. The main results of the paper is not affected by un-equal bins. For the purpose of easy presentation, we will treat bins of equal size.

As $\ell\{\tilde{m}_{\hat{\theta}}(t_j)\}$ measures the goodness-of-fit at a fixed t_j, an empirical likelihood based statistic that measures the global goodness-of-fit is defined as

$$\ell_n(\tilde{m}_{\hat{\theta}}) = \sum_{j=1}^{k_n} \ell\{\tilde{m}_{\hat{\theta}}(t_j)\}.$$

by combining $\ell\{\tilde{m}_{\hat{\theta}}(t_j)\}$ at each t_j. It will be the statistic we use to derive the goodness-of-fit test.

Theorem 1. Under the assumptions (i) - (vi),

$$k_n^{-1}\ell_n(\tilde{m}_{\hat{\theta}}) = (nh^d)\int \frac{\{\hat{m}(x) - \tilde{m}_\theta(x)\}^2}{V(x)}dx + O_p\{k_n^{-1}\log^2(n) + h\log^2(n)\}. \tag{3.1}$$

Härdle and Mammen (1993) proposed the L_2 distance

$$T_n = nh^{d/2}\int \{\hat{m}(x) - \tilde{m}_{\hat{\theta}}(x)\}^2 \pi(x)dx$$

as a measure of goodness-of-fit where $\pi(x)$ is a given weight function. Theorem 1 indicates that the leading term of $k_n^{-1}\ell_n(\tilde{m}_{\hat{\theta}})$ is $h^{d/2}T_n$ with $\pi(x) = V^{-1}(x)$. The differences between the two test statistics are (a) the empirical likelihood test statistic automatically studentises via its internal algorithm conducted at the background, so that there is no need to explicitly estimate $V(x)$; (b) the empirical likelihood statistic is able to capture other features such as skewness and kurtosis exhibited in the data without using the bootstrap resampling which involves more technical details when data are dependent. If we choose $k_n = (2h)^{-d}$ as prescribed, then the remainder term in (3.1) becomes $O_p\{h \log^2(n)\}$.

Theorem 2. Suppose assumptions (i) - (vi), then $k_n^{-1}\ell_n(\tilde{m}_{\hat{\theta}}) \xrightarrow{\mathcal{L}} \int_S \mathcal{N}^2(s)ds$ where $\mathcal{N}$ is a normal process on $S = [0,1]^d$ with mean $E\{\mathcal{N}(s)\} = h^{d/4}\Delta_n(s)/\sqrt{V(s)}$ and covariance

$$\Omega(s,t) = Cov\{\mathcal{N}(s), \mathcal{N}(t)\} = \sqrt{\frac{f(s)\sigma^2(s)}{f(t)\sigma^2(t)}} \frac{W_0^{(2)}(s,t)}{\sqrt{W_0^{(2)}(s,s)W_0^{(2)}(t,t)}}$$

where

$$W_0^{(2)}(s,t) = \int_{y\in S} h^{-d}K\{(s-y)/h\}K\{(t-y)/h\}dy. \tag{3.2}$$

As K is a compact kernel on $[-1,1]^d$, when both s and t are in S_I (the interior part of S),

$$W_0^{(2)}(s,t) = \int K(u)K\{u-(s-t)/h\}du = K^{(2)}\left(\frac{s-t}{h}\right) \tag{3.3}$$

where $K^{(2)}$ is the convolution of K. The compactness of K also means that $W_0^{(2)}(s,t) = 0$ if $|s-t| > 2h$ which implies $\Omega(s,t) = 0$ if $|s-t| > 2h$. Hence $\mathcal{N}(s)$ and $\mathcal{N}(t)$ are independent if $|s-t| > 2h$. As $f(s)\sigma^2(s) = f(s)\sigma^2(t) + O(h)$ when $|s-t| \le 2h$,

$$\Omega(s,t) = \frac{W_0^{(2)}(s,t)}{\sqrt{W_0^{(2)}(s,s)W_0^{(2)}(t,t)}} + O(h), \tag{3.4}$$

So, the leading order of the covariance function is free of σ^2 and f and is completely known.

Let $\mathcal{N}_0(s) = \mathcal{N}(s) - h^{d/4}\Delta_n(s)f(s)/\sqrt{V(s)}$. Then $\mathcal{N}_0(s)$ is a Normal process with zero mean and covariance Ω. The boundedness of K implies $W_0^{(2)}$ being bounded, and hence $\int_S \Omega(t,t)dt < \infty$. Let $T = \int_S \mathcal{N}^2(s)ds =: T_1 + T_2 + T_3$ where

$$\begin{aligned} T_1 &= \int_S \mathcal{N}_0^2(s)ds, \quad T_2 = 2h^{d/4}\int_S V^{-1/2}(s)\Delta_n(s)\mathcal{N}_0(s)ds \quad \text{and} \\ T_3 &= h^{d/2}\int_S V^{-1}(s)\Delta_n^2(s)ds. \end{aligned}$$

From some basic results on stochastic integrals and (3.4),

$$E(T_1) = \int_S \Omega(s,s)ds = 1 \quad \text{and}$$

$$
\begin{aligned}
Var(T_1) &= 2\int_S\int_S \Omega^2(s,t)dsdt \\
&= 2\int_S\int_S \{W_0^{(2)}(s,t)\}^2\{W_0^{(2)}(s,s)W_0^{(2)}(t,t)\}^{-1}dsdt\,\{1+O(h^2)\}
\end{aligned}
$$

From (3.3) and the fact that the size of the boundary region S_B is $O(h)$, we have

$$
\begin{aligned}
&\int_S\int_S \{W_0^{(2)}(s,t)\}^2\{W_0^{(2)}(s,s)W_0^{(2)}(t,t)\}^{-1}dsdt \\
&= \{K^{(2)}(0)\}^{-2}\int_S\int_S [K^{(2)}\{(s-t)/h\}]^2 dsdt\,\{1+o(1)\} \\
&= h^d K^{(4)}(0)\{K^{(2)}(0)\}^{-2} + o(h^d).
\end{aligned}
$$

Therefore,

$$
Var(T_1) = 2h^d K^{(4)}(0)\{K^{(2)}(0)\}^{-2} + o(h^{2d}).
$$

It is obvious that $E(T_2) = 0$ and

$$
Var(T_2) = 4h^{d/2}\int\int V^{-1/2}(s)\Delta_n(s)\Omega(s,t)V^{-1/2}(t)\Delta_n(t)dsdt.
$$

As Δ_n and V^{-1} are bounded in S, there exist constants C_1 and C_2 such that

$$
Var(T_2) \le C_1 h^{d/2}\int\int \Omega(s,t)dsdt \le C_2 h^{3d/2}.
$$

As T_3 is non-random, we have

$$
\begin{aligned}
E(T) &= 1 + h^{d/2}\int V^{-1}(s)\Delta_n^2(s)ds + o(h^{d/2}) \quad \text{and} &(3.5)\\
Var\{T\} &= 2h^d K^{(4)}(0)\{K^{(2)}(0)\}^{-2} + o(h^d) &(3.6)
\end{aligned}
$$

4 Goodness-of-Fit Test

We now turn our interest to the derivation of the asymptotic distribution of $k_n^{-1}\ell_n(\tilde{m}_{\hat{\theta}})$. We do this by discretising $\int_S \mathcal{N}^2(s)ds$ as $(k_n)^{-1}\sum_{j=1}^{k_n}\mathcal{N}^2(t_j)$ where $\{t_j\}_{j=1}^{k_n}$ are the mid-points of the original bins in formulating $\ell_n(\tilde{m}_{\hat{\theta}})$. If we choose $k_n = (2h)^{-d}$ such that $||t_j - t_k|| \ge 2h$ for any $j \ne k$, then $\{\mathcal{N}(t_j)\}$ are independent and each $\mathcal{N}(t_j) \sim N(h^{1/4}\Delta_n(t_j)/\sqrt{V(t_j)}, 1)$. This means that under the alternative H_1

$$
\sum_{j=1}^{k_n}\mathcal{N}^2(t_j) \sim \chi^2_{k_n}(\gamma_{k_n}),
$$

a non-central chi-square random variable with k_n degree of freedom and the non-central component $\gamma_{k_n} = h^{d/4}\{\sum_{j=1}^{k_n} \Delta_n^2(t_j)/V(t_j)\}^{1/2}$. Under H_0, $\sum_{j=1}^{k_n} \mathcal{N}^2(t_j) \sim \chi^2_{k_n}$. This leads to a chi-square test which rejects H_0 if $\ell_n(\tilde{m}_{\hat{\theta}}) > \chi^2_{k_n,\alpha}$ where $\chi^2_{k_n,\alpha}$ is the upper α-quantile of $\chi^2_{k_n}$. The asymptotic power of the chi-square test is $Pr\{\chi^2_{k_n}(\gamma_{k_n}) > \chi^2_{k_n,\alpha}\}$, which is sensitive to alternative hypotheses differing from H_0 in all directions.

We may also establish the asymptotic normality of $(k_n)^{-1}\sum_{i=1}^{k_n} \mathcal{N}^2(t_j)$ by applying the central limit theorem for a triangular array, which together with (3.5) and (3.6) means that

$$k_n^{-1}\ell_n(\tilde{m}_{\hat{\theta}}) \xrightarrow{\mathcal{L}} N\Big(1 + h^{1/2}\int \Delta_n^2(s)V^{-1}(s)ds, 2hK^{(4)}(0)\{K^{(2)}(0)\}^{-2}\Big).$$

A test for H_0 with an asymptotic significance level α is to reject H_0 if

$$k_n^{-1}\ell_n(\tilde{m}_{\hat{\theta}}) > 1 + z_\alpha\{K^{(2)}(0)\}^{-1}\sqrt{2hK^{(4)}(0)} \tag{4.1}$$

where $P(Z > z_\alpha) = \alpha$ and $Z \sim N(0,1)$. The asymptotic power of this test is

$$1 - \Phi\Big\{z_\alpha - \frac{K^{(2)}(0)\int \Delta_n^2(s)V^{-1}(s)ds}{\sqrt{2K^{(4)}(0)}}\Big\}. \tag{4.2}$$

We see from the above that the binning based on the bandwidth value h provides a key role in the derivation of the asymptotic distributions. However, the binning discretises the null hypothesis and unavoidably leads to some loss of power as shown in the simulation reported in the next section. From the point of view of retaining power, we would like to have the size of the bins smaller than that prescribed by the smoothing bandwidth in order to increase the resolution of the discretised null hypothesis to the original H_0. However, this will create dependence between the empirical likelihood evaluated at neighbouring bins and make the above asymptotic distributions invalid. One possibility is to evaluate the distribution of $\int_S \mathcal{N}_0^2(s)ds$ by using the approach of Wood and Chan (1994) by simulating the normal process $\mathcal{N}^2(s)$ under H_0. However, this is not our focus here and hence is not considered in this paper.

5 Simulation and Application

We probe our testing procedure in a simulation and in a test on a financial market model. In the simulation we consider the time series model

$$Y_i = 2Y_{i-1}/(1+Y_{i-1}^2) + c_n Sin(Y_{i-1}) + \sigma(Y_{i-1})\eta_i$$

where $\{\eta_i\}$ are independent and identically distributed uniform random variables in $[-1,1]$, η_i is independent of $X_i = Y_{i-1}$ for each i, and $\sigma(x) = exp(-x^2/4)$. Note that the mean and the variance functions are both bounded which ensures the series is asymptotically stationary. To realise the stationarity, we pre-run the series 100 times with an initial value $Y_{-100} = 0$. The sample size considered are $n = 500$ and 1000 and c_n takes value of 0, 0.03 and 0.06. As the simulation shows that the two empirical likelihood tests have very similar power performance, we will report the results for the test based on the chi-square distribution only. To gauge on the effect of the smoothing bandwidth h on the power, ten levels of h are used for each simulated sample to formulate the test statistic.

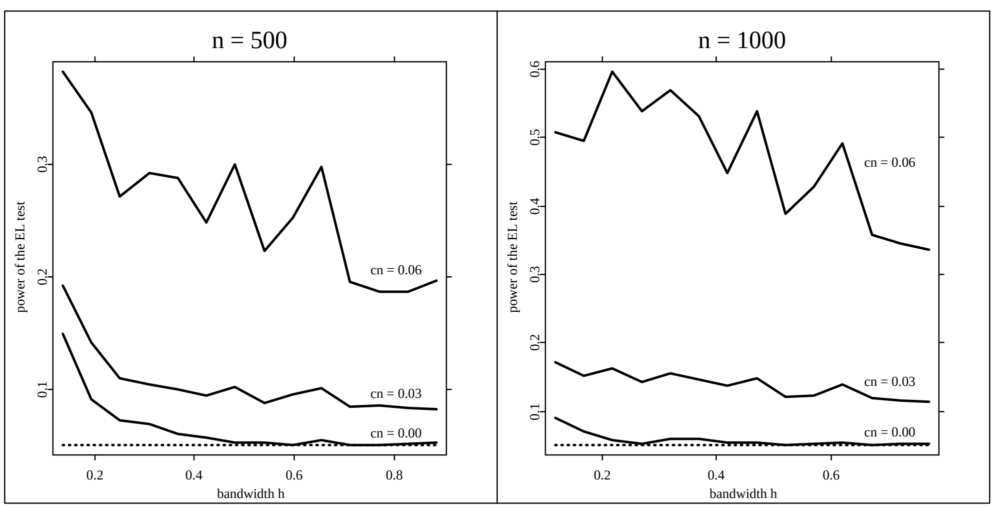

Figure 2. Power of the empirical likelihood test. The dotted lines indicate the 5% level.

Figure 2 presents the power of the empirical likelihood test based on 5000 simulation with a nominal 5% level of significance. We notice that when $c_n = 0$ the simulated significance level of the test is very close to the nominal level for large range of h values which is especially the case for the larger sample size $n = 1000$. When c_n increases, for each fixed h the power increases as the distance between the null and the alternative hypotheses becomes larger. For each fixed c_n, there is a general trend of decreasing power when h increases. This is due to the discretisation of H_0 by binning as discussed at the end of previous section. We also notice that the power curves for $c_n = 0.06$ are a little erratic although they maintain the same trend as in the case of $c_n = 0.03$. This may be due to the fact that when the difference between H_0 and H_1 is large, the difference between the nonparametric and the parametric fits becomes larger and the test procedure becomes more sensitive to the bandwidths.

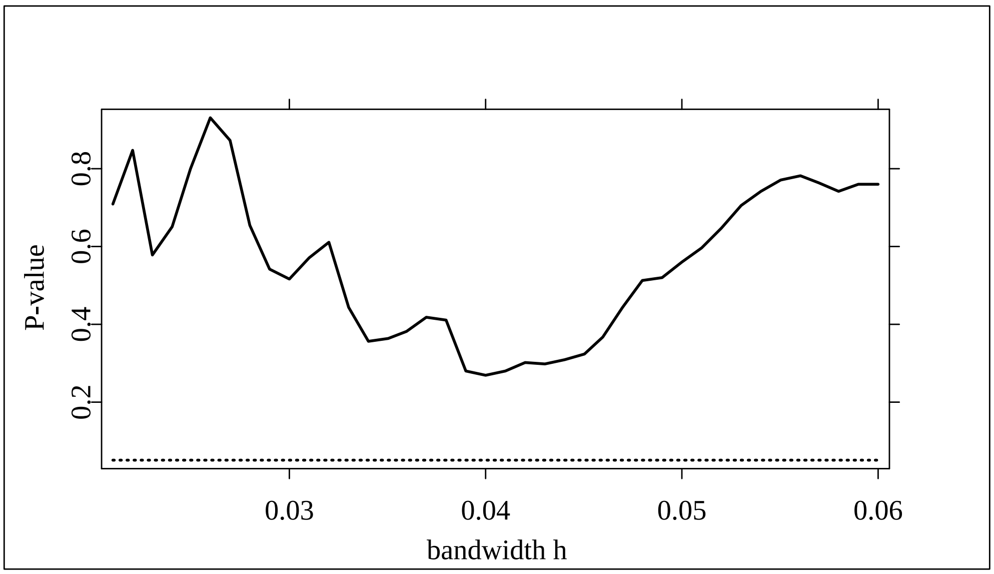

Figure 3. P-values of the empirical likelihood test for the S&P 500 data. The dotted line indicates the 5% level.

We now apply the empirical likelihood test procedure on the S&P 500 data presented in Figure 1 to test the parametric mean function $m(x) = a(1 - x)$ given in the Cox-Ingersoll-Ross diffusion model (1.2). The process X is restored from the observed residuals by the approach introduced in (Härdle et al. (2000)). The parametric estimate for a is $\hat{a} = 0.00968$ by using methods based on the marginal distribution and the autocorrelation structure of X. For details about the procedure see (Härdle et al. (2000)). The cross validation is used to find the bandwidth h. However, the score function is monotonic decreasing for $h < 0.15$ and then become a flat line for $h \in [0.15, 0.8]$. This may be caused by the different intensity level of the design points as revealed in Figure 1. Further investigation shows that a h-value larger (smaller) than 0.06 (0.02) produces an oversmoothed (undersmoothed) curve estimate. Therefore, the test is carried out for a set of h values ranging from 0.02 to 0.06. The P-values of the test as a function of h is plotted in Figure 3. The P-values indicate that there is insufficient evidence to reject the diffusion model.

Acknowledgement: The authors acknowledge support by the Deutsche Forschungsgemeinschaft via Sonderforschungsbereich 373 "Quantifikation and Simulation ökonomischer Prozesse" at Humboldt-Universität zu Berlin.

REFERENCES

Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika* **85**, 535-547.

Billingsley, P. (1999). *Convergence of Probability Measures*. New York: Wiley.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*, Lecture Note in Statistics, **110**. Heidelberg: Springer-Verlag.

Bibby, B. M. and Sørensen, M. (1996). On estimation of discretely observed diffusion, a review. *Theory of Stochastic Processes*, **2**, 49-56.

Chen, S.X. (1994). Comparing empirical likelihood and bootstrap hypothesis tests. *J. Mult. Anal.* **51**, 277-293.

Chen, S.X. (1996). Empirical likelihood for nonparametric density estimation. *Biometrika*. **83** 329-341.

Chen, S.X. and Qin, Y. S. (2000). Empirical likelihood confidence intervals based on a local linear smoother. *Biometrika*. **87**, 946-953.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). A theory of term structure of interest rates. *Econometrica*, **53**, 385-407.

Doukhan, P. (1994). *Mixing*. Lecture Note in Statistics, **85**. Heielberg: Springer-Verlag.

Engle, R. F. and Gonzalez-Rivera, G. (1991). Semiparametric ARCH models, *Journal of Business and Econometric Statistics*, **9**, 345-360.

Genon-Caralot, V., Jeantheau, T. and Laredo, C. (2000). Stochastic Volatility models as hidden markov models and statistical applications. Preprint, Universite De Marne-La-Vallee, Forthcoming in Bernoulli.

Gouriéroux, Ch. and Monfort, A. (1992). Qualitative threshold ARCH models, *Journal of Econometrics*, **52**, 159-199.

Gouriéroux, Ch. (1997). *ARCH Models and Financial Applications.* Heidelberg: Springer-Verlag.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58**, 109–127.

Härdle, W., Kleinow, T., Korostelev, A., Logeay, C. and Platen, E. (2000). Diffusion Estimation and Modeling of a Stock Market Index. Discussion Paper, SFB 373, Humboldt-Universität zu Berlin

Härdle, W. and Mammen, E. (1993). Comparing Nonparametric versus parametric regression fits. *Annals of Statistics*, **21**, 1926-1947.

Herwartz, H. (2000). Weekday Dependence of German Stock Market Returns. *Applied Stochastic Models in Business and Industry*, **16**, 47–71.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications.* London: Chapman and Hall.

Hart, J. (1997). *Nonparametric Smoothing and Lack-of-fit Tests.* Heidelberg: Springer-Verlag.

Karatzas, I. and Shreve, S. E. (1998). *Methods of Mathematical Finance.* Springer Verlag New York.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.*, **25**, 2084-2102.

Kreiss, J., Neumann, M. H. and Yao, Q. (1998). Bootstrap tests for simple structures in nonparametric times series regression.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90–120.

Owen, A (1991). Empirical likelihood for linear model. *Ann. Statist.* **19** 1725-1747.

Platen, E. (1999). Risk premia and financial modelling without measure transformation. Pre-print, School of Finance & and Economics, University of Technology Sydney.

Pham, T. D. and Tran, L. T. (1985). Some mixing properties of time series models. *Stoch. Process. Appl.*, **19**, 297-303.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating functions. *Ann. Statist.* **22**, 300–325.

Rosenblatt, M. (1956). A central limit theorem and a the α-mixing condition. *Proc. Nat. Acad. Sc. U.S.A.*, **42**, 43-47.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Heidelberg: Springer-Verlag.

Wood, A. T. A. and Chan, G. (1994). Simulation of stationary Gaussian process in $[0,1]^d$. *J. Comp. Graph. Stat.*, **3**, 409-432.

Zakoian, J. M. (1991). Threshold ARCH models, CREST Discussion Papers, Ecole Nationale de la Statistique et de l'Administration Economique.

Appendix: Technical Details

Let $\mathcal{F}_k^l$ be the σ-algebra of events generated by $\{(X_i, Y_i), k \le i \le l\}$ for $l \ge k$. The measure for dependence between the time series is the α-mixing coefficient

$$\alpha(k) = \sup_{\mathcal{A} \in \mathcal{F}_1^i, \mathcal{B} \in \mathcal{F}_{i+k}^\infty} |P(AB) - P(A)P(B)|$$

introduced by Rosenblatt (1956).

Assumptions

The assumptions required to establish the results given in the paper are the following:

(i) The kernel K is Lipschitz continuous in $[-1,1]^d$, that is $|K(t_1) - K(t_2)| \leq C||t_1 - t_2||$ where $||\cdot||$ is the Euclidean norm, and $h = O\{n^{-1/(4d+1)}\}$;

(ii) f, m and σ^2 have continuous derivatives up to the second order in S.

(iii) $\hat{\theta}$ is a parametric estimator of θ within the family of the parametric model, and

$$\sup_{x\in S} |m_{\hat{\theta}}(x) - m_\theta(x)| = O_p(n^{-1/2}).$$

(iv) $\Delta_n(x)$, the local shift in the H_1, is uniformly bounded with respect to x and n, and $c_n = n^{-1/2}h^{-d/4}$ which is the order of the difference between H_0 and H_1.

(v) The process $\{(X_i, Y_i)\}$ is strictly stationary and α-mixing, and $\alpha(k) \leq a\rho^k$ for some $a > 0$ and $\rho \in [0, 1)$.

(vi) $E\{\exp(a_0|Y_1 - m(X_1)|)\} < \infty$ for some $a_0 > 0$; The conditional density of X given Y $f_{X|Y} \leq A_1 < \infty$, and the joint conditional density of (X_1, X_l) given (Y_1, Y_l) is bounded for all $l > 1$.

Assumptions (i) and (ii) are standard in nonparametric curve estimation and are satisfied for example for bandwidths selected by cross validation, whereas (iii) and (iv) are common in nonparametric goodness-of-fit tests. Assumption (v) means the data are weakly dependent. It can be seen from the proof that the geometric the α-mixing condition can be weakened to $\alpha(k) \leq Ck^{-s(d)}$ where $s(d) > 2$ and is a monotone function of d. It is convenient technically to assume geometric the α-mixing. For a univariate linear causal process (which includes ARMA models)

$$Y_t = \sum_{s=0}^{\infty} g_{t-s}\xi_s$$

with independent and identically distributed innovation $\{\xi_s\}_{s=0}^{\infty}$, Gorodeskii (1977) showed that the linear process is the α-mixing under certain conditions and established the rate for the the α-mixing coefficient. Pham and Tran (1985) show that if the each coefficient g_t of the process is $O(\gamma^t)$, $0 < \gamma < 1$, then the process is geometric the α-mixing. For the p-Markovian processes $Y_i = m(X_i) + \epsilon_i$ where ϵ is an i.i.d. process whose marginal distribution is equivalent to the Lebesgue measure and $X_i = (Y_{i-1}, \ldots, Y_{i-p})$. The geometric the α-mixing condition is satisfied if m is bounded. It is true also if there is a compact set where m is bounded and outside of which $m(y_1, \ldots, y_p) \leq a_1|y_1| + \ldots + a_p|y_p|$, and if the roots of the polynomial $t^p - a_1 t^{p-1} - \ldots - a_p$ lie in the open unit disk. These and other cases of processes satisfying the condition are available in Doukhan (1994).

Throughout the proof we will use C to denote positive constants which may take different values.

Proof of Lemma 1:

Recall that $\bar{U}_j(x) = (nh^d)^{-1} \sum_{i=1}^n \left[K\left(\frac{x-X_i}{h}\right) \{Y_i - \tilde{m}_{\hat{\theta}}(x)\} \right]^j$. Following Owen (1990), we need to show that:

$$\sup_{x \in S} |\bar{U}_1(x)| = o_p\{(nh^d)^{-1/2} \log(n)\}, \tag{A.1}$$

$$Pr\{\inf_{x \in S} \bar{U}_2(x) \geq d_0\} = 1 \quad \text{for a positive } d_0 > 0, \text{ and} \tag{A.2}$$

$$\max_{1 \leq j \leq n} \sup_{x \in S} |g_i(x)| = o_p\{(nh^d)^{1/2} \log^{-1}(n)\}. \tag{A.3}$$

To prove (A.1), we define $\epsilon_i = Y_i - m(X_i)$ and write $\bar{U}_1(x) = I_1(x) + I_2(x) + I_3(x)$ where

$$I_1 = n^{-1} \sum_{i=1}^n K_h(x - X_i)\{m_\theta(X_i) - \tilde{m}_{\hat{\theta}}(x)\} = \hat{f}(x)\{\tilde{m}_\theta(x) - \tilde{m}_{\hat{\theta}}(x)\},$$

$I_2 = n^{-1} \sum_{i=1}^n K_h(x - X_i)\epsilon_i$ and $I_3 = n^{-1} c_n \sum_{i=1}^n K_h(x - X_i)\Delta_n(X_i)$. As

$$\sup_{x \in S} |n^{-1} \sum_{i=1}^n K_h(x - X_i) - f(x)| \xrightarrow{a.s.} 0$$

as shown in Bosq (1998, p.49), condition (iv) implies,

$$\sup_{x \in S} |I_1(x)| = O_p(n^{-1/2}) \quad \text{and} \quad \sup_{x \in S} |I_3(x)| = O_p(c_n). \tag{A.4}$$

Let $M_n = b_0 \log(n)$ for some positive constant b_0. Split $I_2(x)$ into two parts:

$$I_2^+(x) = n^{-1} \sum_{i=1}^n K_h(x - X_i)\epsilon_i I(|\epsilon_i| \geq M_n) \quad \text{and} \quad I_2^-(x) = n^{-1} \sum_{i=1}^n K_h(x - X_i)\epsilon_i I(|\epsilon_i| < M_n).$$

As $\sup_{x \in S} |I_2^+(x)| \leq C(nh^d)^{-1} \sum_{i=1}^n |\epsilon_i| I(|\epsilon_i| \geq M_n)$ for some $C > 0$, the Cauchy-Schwartz inequality implies that

$$E\Big[\sup_{x \in S} |I_2^+(x) - E\{I_2^+(x)\}|\Big] \leq 2C(nh^d)^{-1} \sum_{i=1}^n \{E(|\epsilon_i|^2) P(|\epsilon_i| \geq M_n)\}^{1/2}.$$

From the Chebyshev inequality and condition (vi), for a positive constant η_0,

$$Pr\Big[M_n^{-1}(nh^d)^{1/2}\sup_{x\in S}|I_2^+(x)-E\{I_2^+(x)\}|\geq \eta_0)\Big]\leq 2C\eta_0^{-1}n^{1/2}h^{-d/2}M_n^{-1}\exp\{-\tfrac{1}{2}a_0b_0\log(n)\}.$$

By properly choosing b_0, the RHS converges to zero as $n\to\infty$. This means that

$$\sup_{x\in S}|I_2^+(x)-E\{I_2^+(x)\}|=o_p\{(nh^d)^{-1/2}\log(n)\}. \tag{A.5}$$

Let $\phi_i(x)=K(\frac{x-X_i}{h})\epsilon_i I(|\epsilon_i|<M_n)$, $Z_i(x)=\phi_i(x)-E\{\phi_i(x)\}$. Clearly, at each fixed x, $\{Z_i(x)\}$ has zero mean, is bounded by $b=C_1M_n$ and geometrical the α-mixing. Put $\eta=(h^d/n)^{1/2}M_n\eta_0$. From Theorem 1.3 of Bosq (1998),

$$\begin{aligned} Pr[|I_2^-(x)-E\{I_2^-(x)\}|>(nh^d)^{-1/2}M_n\eta_0] &= P\left(|\sum Z_i(x)|>n\eta\right)\\ \leq\ & 4\exp[-\eta^2q/\{8v^2(q)\}]+22(1+4C_1M_n/\eta)^{1/2}q\alpha\{[n/(2q)]\} \end{aligned} \tag{A.6}$$

where $q=\eta_0M_n^2\sqrt{n}h^{-d/2}$, $p=n/q$, $v^2(q)=\frac{2}{p^2}\sigma^2(q)+\frac{b\eta}{2}$ and

$$\sigma^2(q)=\max_{0\leq j\leq 2q-1}E\{\beta_1(p)Z_{[jp]+1}(x)+\sum_{i=2}^{p}Z_{[jp]+i}(x)+\beta_2(p)Z_{[jp]+1}(x)\}^2.$$

In the last equation $\beta_1(p)=[jp]+1-jp$ and $\beta_2(p)=(j+1)p-[(j+1)p]$. By the stationarity of $\{(X_i,Y_i)\}$,

$$\sigma^2(q)\leq(p+2)E\{Z_1^2(x)\}+J \tag{A.7}$$

where

$$J=2p\sum_{l=1}^{p}(1-\frac{l-1}{p})|Cov\{Z_1(x),Z_{l+1}(x)\}|+2|Cov\{Z_1(x),Z_{p+1}(x)\}|.$$

Condition (vi) implies that $E(|\epsilon|^\delta)<\infty$ for some $\delta>2$. Using the Davydov's lemma,

$$\begin{aligned}|Cov\{Z_1(x),Z_{l+1}(x)\}| &\leq 2\delta(\delta-2)^{-1}\{E|K(\frac{x-X_i}{h})\epsilon_i|^\delta\}^{2/\delta}\alpha^{1-2/\delta}(l)\\ &\leq Ch^d\alpha^{1-2/\delta}(l).\end{aligned} \tag{A.8}$$

Following the approach used in Fan and Gijbles (1996), we let $d_n\to\infty$ be a sequence of integers such that $d_nh^d\to 0$ and split J as

$$\begin{aligned} J_1 &= 2p\sum_{l=1}^{d_n-1}(1-\frac{l-1}{p})|Cov\{Z_1(x),Z_{l+1}(x)\}| \quad \text{and}\\ J_2 &= 2p\sum_{l=d_n}^{p}(1-\frac{l-1}{p})|Cov\{Z_1(x),Z_{l+1}(x)\}|+2|Cov\{Z_1(x),Z_{p+1}(x)\}|. \end{aligned} \tag{A.9}$$

As $Cov\{Z_1(x), Z_{l+1}(x)\}| \le Var\{Z_1(x)\} \le C$, $J_1 \le Cpd_n = o(ph^d)$. From (A.8) and condition (v), we have $J_2 = o(ph^d)$ as well. These imply that

$$J = J_1 + J_2 = o(ph^d) \tag{A.10}$$

and hence $\sigma^2(q) \le Cph^d$. The particular forms of q, b and η mean that $v^2(q) \le Cqh^d/n$ and

$$\exp[-\eta^2 q/\{8v^2(q)\}] \le \exp\left(-C_1 M_n^2 \eta_0^2\right). \tag{A.11}$$

The geometric the α-mixing condition implies:

$$(1 + 4Ch^d/\eta)^{1/2} q\alpha\{n/(2q)\} \le C_2 (nh^{-d})^{-3/4} M_n^2 \rho^{\frac{1}{2}\eta_0 M_n^{-2}(nh^d)^{1/2}}. \tag{A.12}$$

Combining (A.11), (A.12) with (A.6) and noticing that both (A.11) and (A.12) are free of x, we have

$$\begin{aligned} & \sup_{x\in S} Pr[|I_2^-(x) - E\{I_2^-(x)\}| \ge (nh^d)^{-1/2}\log(n)\eta_0] \\ \le\ & \exp\left(-C_1 b_0^2 \log^2(n)\eta_0^2\right) + C_2 h^{-3d/4} n^{3/4} M_n^2 \rho^{\frac{1}{2}\eta_0 b_0^{-2}\log^{-2}(n)(nh^d)^{1/2}}. \end{aligned} \tag{A.13}$$

Let $\{B_k\}_{k=1}^{v_n}$ be a set of equal volume disjoint hypercubes with centers $\{s_k\}_{k=1}^{v_n}$ such that $S = \bigcup_{k=1}^{v_n} B_k$, $v_n = [n^{t_0}]$ for some $t_0 > 0$ and $\sup_{x\in B_k} ||x - s_k|| \le cv_n^{-1}$. Based on this partition of S, and let $I_2^{-\star}(x) = I_2^(x) - E\{I_2^-(x)\}$

$$\sup_{x\in S} |I_2^-(x) - E\{I_2^-(x)\}| \le \max_{k=1,\dots,v_n} |I_2^{-\star}(s_k)| + \sup_{x\in S} |I_2^{-\star}(x) - I_2^{-\star}(s_{k(x)})|$$

where $k(x)$ being the index of the hypercube containing x. Note that

$$P\{\max_{k=1,\dots,v_n} |I_2^{-\star}(s_k)| \ge (nh^d)^{-1/2}\eta_0 M_n\} \le n^{t_0} \sup_{x\in S} P\{|I_2^-(x) - E\{I_2^{-1}(x)\}| \ge (nh^d)^{-1/2} M_n \eta_0\},$$

By properly choosing b_0, (A.13) implies that

$$\max_{k=1,\dots,v_n} |I_2^{-\star}(s_k)| = o_p\{(nh^d)^{-1/2}\log(n)\}. \tag{A.14}$$

As K is Lipschitz continuous,

$$\sup_{x\in S} |I_2^{-\star}(x) - I_2^{-\star}(s_{k(x)})| \le Ch^{-1}n^{-t_0}\left(n^{-1}\sum_{i=1}^{n} |\epsilon_i| + E|\epsilon_i|\right).$$

Note that $n^{-1}\sum|\epsilon_i| \stackrel{w.s.}{\rightarrow} E|\epsilon_i|$, and $E|\epsilon_i| \le C$. We get with probability one

$$\sup_{x\in S}|I_2^{-\star}(x) - I_2^{-\star}(s_{k(x)})| \le Ch^{-1}n^{-t_0}.$$

By choosing $t_0 > 3/\{2(d+1)\}$, we have

$$Pr\{\sup_{x\in S}|I_2^{-\star}(x) - I_2^{-\star}(s_{k(x)})| \ge (nh^d)^{-1/2}\log(n)\eta_0\} \to 0,$$

which means that

$$\sup_{x\in S}|I_2^{-\star}(x) - I_2^{-\star}(s_{k(x)})| = o_p\{(nh^d)^{-1/2}\log(n)\}. \tag{A.15}$$

Clearly, (A.5), (A.14) and (A.15) imply (A.1).

We need to do a few things before proving (A.2). Similar to the derivation of (A.1) and the proof of Theorem 2.2 of Bosg (1998), it can be shown that for any smooth function g in $\mathbb{R}^d$

$$\begin{aligned} & \sup_{x\in S}|n^{-1}h^d\sum_{i=1}^n K_h^2(x-X_i)g(X_i) - f(x)v(x;h)g(x)| \\ = \ & O_p\{(nh^d)^{-1/2}\log(n) + h\}, \qquad \text{(A.16)} \\ & \sup_{x\in S}|n^{-1}h^d\sum_{i=1}^n K_h^2(x-X_i)\epsilon_i^2 - f(x)v(x;h)\sigma^2(x)| \\ = \ & O_p\{(nh^d)^{-1/2}\log(n) + h\} \qquad \text{(A.17)} \end{aligned}$$

and

$$\sup_{x\in S}|n^{-1}h^d\sum_{i=1}^n K_h^2(x-X_i)\epsilon_i| = O_p\{(nh^d)^{-1/2}\log(n)\} \tag{A.18}$$

where the h-order terms in the remainders are due to the bias associated with the kernel estimator. Note that

$$\bar{U}_2(x) = n^{-1}h^d\sum_{i=1}^n K_h^2(x-X_i)\{m_\theta(X_i) - \tilde{m}_{\hat{\theta}}(x) + \epsilon_i + c_n\Delta_n(X_i)\}^2 = \sum_{l=1}^6 J_l(x)$$

where, from (A.16) to (A.18),

$$J_1(x) = n^{-1}h^d\sum_{i=1}^n K_h^2(x-X_i)\{m_\theta(X_i) - \tilde{m}_{\hat{\theta}}(x)\}^2 = \tilde{O}_p\{n^{-1/2} + h\}$$

$$
\begin{aligned}
J_2(x) &= n^{-1}h^d \sum_{i=1}^n K_h^2(x-X_i)\epsilon_i^2 \to f(x)v(x;h)\sigma^2(x) + \tilde{O}_p\{(nh^d)^{-1/2}\log(n)+h\} \\
J_3(x) &= n^{-1}h^d c_n^2 \sum_{i=1}^n K_h^2(x-X_i)\Delta_n^2(X_i) = \tilde{O}_p(c_n^2) \\
J_4(x) &= 2n^{-1}h^d c_n \sum_{i=1}^n K_h^2(x-X_i)\{m_\theta(X_i)-\tilde{m}_{\hat{\theta}}(x)\}\Delta(X_i) = \tilde{O}_p\{c_n(n^{-1}+h)\} \\
J_5(x) &= 2n^{-1}h^d \sum_{i=1}^n K_h^2(x-X_i)\{m_\theta(X_i)-\tilde{m}_{\hat{\theta}}(x)\}\epsilon_i = \tilde{o}_p(n^{-1/2}) \\
J_6(x) &= 2n^{-1}h^d c_n \sum_{i=1}^n K_h^2(x-X_i)\epsilon_i\Delta_n(X_i) = \tilde{O}_p\{c_n(nh^d)^{-1/2}\log(n)\}.
\end{aligned}
$$

In summary of the above results, we have

$$\sup_{x\in S}|\bar{U}_2(x) - f(x)v(x;h)\sigma^2(x)| = O_p(h). \tag{A.19}$$

As $f(x)v(x;h)\sigma^2(x)$ is uniformly bounded below,

$$\inf_{x\in S} f(x)v(x;h)\sigma^2(x) \geq d_0 \quad \text{for some } d_0 > 0. \tag{A.20}$$

Since

$$\inf_{x\in S}|U_2(x)| \geq -\sup_{x\in S}|U_2(x) - f(x)v(x;h)\sigma^2(x)| + \inf_{x\in S}|f(x)v(x;h)\sigma^2(x)|,$$

(A.2) is implied by (A.19) and (A.20).

Let $w_i = \sup_{x\in S}|K\left(\frac{x-X_i}{h}\right)\{Y_i - \tilde{m}_{\hat{\theta}}(x)\}|$. As K, m and Δ_n are bounded in S, $w_i \leq C_1|\epsilon_i| + C_2$. From the Chebyshev inequality and Condition (vi)

$$
\begin{aligned}
Pr\left(w_i > (nh^d)^{1/2}\{\log(n)\}^{-1}\right) &\leq Pr\left(|\epsilon_i| \geq C_3(nh^d)^{1/2}\{\log(n)\}^{-1}\right) \\
&\leq C_4\exp\{-C_5(nh^d)^{1/2}\log^{-1}(n)\}
\end{aligned}
$$

Thus, $\sum_{n=1}^{\infty} Pr\left(w_i > (nh^d)^{1/2}\{\log(n)\}^{-1}\right) < \infty$. According to the Boreal-Cantelli lemma, $w_i > (nh^d)^{1/2}\{\log(n)\}^{-1}$ finitely often with probability 1. This means that $Z_n = \max_{1\leq i\leq n} w_i > (nh^d)^{1/2}\{\log(n)\}^{-1}$ finitely often, which implies (A.3).

Proof of Theorem 1: From (2.7) and (A.19)

$$\bar{U}_2^{-1}\bar{U}_1^2 = \left[n^{-1}\sum_{i=1}^n W_h(x-X_i)\{\epsilon_i + c_n\Delta_n(X_i)\}\right]^2 + \tilde{O}_p\{(nh^d)^{-1}h\log^2(n)\} \tag{A.21}$$

where $W_h(x - X_i) = K_h(x - X_i)/\{f(x)v(x;h)\sigma^2(x)\}^{1/2}$. Note that $(nh^d)^{1/2}c_n = O(h^{d/4})$. Let

$$\begin{aligned} A &= k_n^{-1}(nh^d)\sum_{j=1}^{k_n}\int_{B_j}\Big[n^{-1}\sum_{i=1}^{n}W_h(t_j - X_i)\{\epsilon_i + c_n\Delta_n(X_i)\}]^2 \\ &\quad -[n^{-1}\sum_{i=1}^{n}W_h(t - X_i)\{\epsilon_i + c_n\Delta_n(X_i)\}]^2\Big]dt \\ &= k_n^{-1}\sum_{j=1}^{k_n}\int_{B_j}T_{1j}(t)T_{2j}(t)dt \qquad (A.22) \end{aligned}$$

where for $t \in B_j$

$$\begin{aligned} T_{1j}(t) &= n^{-1/2}\sum_{i=1}^{n}\{W_h(t_j - X_i) - W_h(t - X_i)\}\{\epsilon_i + c_n\Delta_n(X_i)\}, \\ T_{2j}(t) &= n^{-1/2}\sum_{i=1}^{n}\{W_h(t_j - X_i) + W_h(t - X_i)\}\{\epsilon_i + c_n\Delta_n(X_i)\}. \end{aligned}$$

Let $M_n = b_0\log(n)$ for a positive constant b_0 and $\omega_i = \epsilon_i + c_n\Delta_n(X_i)$. Define

$$\begin{aligned} T_{1j}^{+}(t) &= (nh^{-d})^{-1/2}\sum_{i=1}^{n}\{W_h(t_j - X_i) - W_h(t - X_i)\}\omega_i I\left(|\omega_i| > M_n\right), \\ T_{1j}^{-}(t) &= (nh^{-d})^{-1/2}\sum_{i=1}^{n}\{W_h(t_j - X_i) - W_h(t - X_i)\}\omega_i I\left(|\omega_i| \le M_n\right). \end{aligned}$$

Similar definitions apply for $T_{2j}^{+}(t)$ and $T_{2j}^{-}(t)$. It may be shown similar to the derivation of (A.5) that for $l = 1$ and 2

$$\max_{j=1,\ldots,k_n}\sup_{t\in B_j}|T_{1j}^{+}(t) - E\{T_{1j}^{+}(t)\}| = o_p\{(nh^d)^{-1/2}\log(n)\}. \qquad (A.23)$$

Let $\phi_i(t) = h^d\{W_h(t_j - X_i) - W_h(t - X_i)\}\omega_i I\left(|\omega_i| < M_n\right)$ and $Z_i(t) = \phi_i(t) - E\{\phi_i(t)\}$. Then, for $u_n \to \infty$ (the exact order of u_n will be decided later)

$$Pr\{|T_{1j}^{-}(t) - E\{T_{1j}^{-}(t)\} > u_n^{-1}\eta_0\} = Pr\{|\sum_{i=1}^{n}Z_i(t)| > n\eta\}$$

where $\eta = (h^d/n)^{1/2}u_n^{-1}\eta_0$. Note that $|Z_i(t)| \le CM_nk_n^{-1}h^{-1}$. Let $b = CM_nk_n^{-1}h^{-1}$ and $q = n^{1/2}M_n\eta_0u_n^{-1}h^{-1}$. Similar to the derivation of (A.13) and employing again Theorem 1.3

of Bosq (1998), we have

$$Pr\{|\sum_{i=1}^{n} Z_i(t)| > n\eta\} \quad \leq \quad 4\exp\{-\frac{\eta^2 q}{8v^2(q)}\} + 22(1+b/\eta)^{1/2} q\alpha\left([n/(2q)]\right)$$

where $v^2(q) \leq Cqh^d(nk_n)^{-1}$. The upper bound for $v^2(q)$ can be obtained using the same approach in deriving a similar bound for the same name quantity as given between (A.6) and (A.10). By choosing $u_n = b_1 k_n^{1/2} \log^{-1}(n)$ for some positive b_1,

$$\exp\{-\frac{\eta^2 q}{8v^2(q)}\} \leq \exp\left(-C\frac{k_n \eta_0}{u_n^2}\right) = \exp\{-Cb_1\eta_0 \log(n)\}$$

and

$$(1+b/\eta)^{1/2} q\alpha\left([n/(2q)]\right) \leq Cn^{3/4} M_n^{3/2} u_n^{-1/2} h^{-3/2-d/4} k_n^{-1/2} \rho^{M_n^{-1}\eta_0^{-1} n^{1/2} u_n h}.$$

As the right hand sides of the above two inequalities are free of t, we have

$$\begin{aligned} & \sup_{t \in B_j} Pr[|T_{1j}^-(t) - E\{T_{1j}^-(t)\}| \geq b_1 k_n^{-1/2} \log(n)\eta_0] \\ \leq \quad & \exp\{-Cb_1\eta_0 \log(n)\} + Cn^{3/4} M_n^{3/2} u_n^{-1/2} h^{-3/2-d/4} k_n^{-1/2} \rho^{M_n^{-1}\eta_0^{-1} n^{1/2} u_n h}. \end{aligned} \quad \text{(A.24)}$$

Let $\{B_{jl}\}_{l=1}^{v_j}$ be a partition of B_j of equal size hypercubes B_{jl} where v_j be an integer tending to ∞ as $n \to \infty$. Employing similar derivations to those in deriving (A.14) and (A.15) and utilizing (A.24), it can be shown that

$$\sup_{t \in [t_j, t_{j+1}]} |T_{1j}^-(t) - E\{T_{1j}^-(t)\}| = O_p\{k_n^{-1/2} \log(n)\}. \quad \text{(A.25)}$$

A similar derivation will show that

$$\sup_{t \in [t_j, t_{j+1}]} |T_{2j}^-(t) - E\{T_{2j}^-(t)| = O_p\{k_n^{-1/2} \log(n)\}. \quad \text{(A.26)}$$

From (A.23), (A.25) and (A.26) we have for $l = 1$ and 2

$$\sup_{t \in [t_j, t_{j+1}]} |T_{lj}(t)| = O_p\{(nh^d)^{-1/2} \log(n) + k_n^{-1/2} \log(n)\}. \quad \text{(A.27)}$$

These together with (A.22) complete the proof.

Proof of Theorem 2:

We first derive the mean and the covariance of $\hat{m}(x) - \tilde{m}_\theta(x)$. We use $\tilde{O}()$ and $\tilde{o}()$ to denote quantities which are $O()$ and $o()$ uniformly with respect to $x \in S$. It is noted that

$$\begin{aligned} & E\{\hat{m}(x) - \tilde{m}_\theta(x)\} \\ = \; & E\Big[\frac{n^{-1}\sum_{i=1}^n W_h(x - X_i)\{\epsilon_i + c_n\Delta_n(X_i)\}}{b(x;h)f(x)}\{1 + \frac{\hat{f}(x) - b(x;h)f(x)}{b(x;h)f(x)} + \ldots\}\Big] \\ = \; & c_n\Delta_n(x)\{1 + \tilde{O}(h)\} \end{aligned}$$

When x is in the interior of S, the above $\tilde{O}(h)$ term will be $\tilde{O}(h^2)$. This means that

$$E\Big[(nh^d)^{1/2}V^{-1/2}(x)\{\hat{m}(x) - \tilde{m}_\theta(x)\}\Big] = (nh^d)^{1/2}c_n\Delta_n(x)V^{-1/2}(x)\{1 + \tilde{o}(1)\}. \qquad \text{(A.28)}$$

Let $\omega_i = \epsilon_i + c_n\Delta_n(X_i)$. Then,

$$\begin{aligned} & V^{1/2}(s;h)V^{1/2}(t;h)Cov\{\hat{m}(s) - \tilde{m}_\theta(s), \hat{m}(t) - \tilde{m}_\theta(t)\} \\ = \; & Cov\{n^{-1}\sum_{i=1}^n W_h(s - X_i)\omega_i, n^{-1}\sum_{i=1}^n W_h(t - X_i)\omega_i\}\{1 + \tilde{o}(1)\} \\ = \; & \Big[n^{-1}Cov\{W_h(s - X_1)\omega_1, W_h(t - X_1)\omega_1\} \\ & + n^{-1}\sum_{l=2}^n (1 - l/n)Cov\{W_h(s - X_1)\omega_1, W_h(s - X_l)\omega_l\}\Big]\{1 + \tilde{o}(1)\} \end{aligned}$$

Standard derivations show

$$Cov\{W_h(s - X_1)\omega_1, W_h(t - X_1)\omega_1\} = h^{-d}\sqrt{\frac{f(s)\sigma^2(s)}{f(t)\sigma^2(t)}}\frac{W_0^{(2)}(s,t)}{\sqrt{W_0^{(2)}(s,s)W_0^{(2)}(t,t)}} + \tilde{o}(h^{-d}),$$

where $W_0^{(2)}$ is defined in (3.2) and $W_0^{(2)}(t,t) = v(t;h)$. Using the same arguments which establish (A.10) in the proof of Lemma 1, we can show that

$$\sum_{l=2}^n (1 - l/n)Cov\{W_h(s - X_1)\omega_1, W_h(s - X_l)\omega_l\} = \tilde{o}(h^{-d}).$$

Thus,

$$\begin{aligned} & (nh^d)Cov\Big[\frac{\{\hat{m}(s) - \tilde{m}_{\hat{\theta}}(s)\}}{\sqrt{V(s)}}, \frac{\{\hat{m}(t) - \tilde{m}_{\hat{\theta}}(t)\}}{\sqrt{V(t)}}\Big] \\ = \; & \sqrt{\frac{f(s)\sigma^2(s)}{f(t)\sigma^2(t)}}\frac{W_0^{(2)}(s,t)}{\sqrt{W_0^{(2)}(s,s)W_0^{(2)}(t,t)}}\{1 + \tilde{o}(1)\}. \end{aligned} \qquad \text{(A.29)}$$

Next we want to show that for k distinct $t_1, t_2, \cdots, t_k \in [0,1]^d$,

$$(nh^d)^{1/2}\Big(\frac{\{\hat{m}(t_1) - \tilde{m}_\theta(t_1)\}}{V(t_1)}, \cdots, \frac{\{\hat{m}(t_k) - \tilde{k}_\theta(t_k)\}}{V(t_k)}\Big) \xrightarrow{\mathcal{L}} N_k(\mu_k, \Omega_k). \tag{A.30}$$

Here $N_k(\mu_k, \Omega_k)$ is a k-dimensional normal distribution with mean vector

$$\mu_k = (nh^d)^{1/2} c_n \left(\Delta_n(t_1) f^{1/2}(t_1) V^{-1/2}(t_1), \cdots, \Delta_n(t_k) f^{1/2}(t_k) V^{-1/2}(t_k)\right)^T$$

and covariance matrix $\Omega_k = (\omega_{ij})_{k \times k}$, where

$$\omega_{ij} = \sqrt{\frac{f(t_i)\sigma^2(t_i)}{f(t_j)\sigma^2(t_j)}} \frac{W_0^{(2)}(t_i, t_j)}{\sqrt{W_0^{(2)}(t_i, t_i) W_0^{(2)}(t_j, t_j)}}.$$

From Theorem 3.4 of Bosq (1998), $V^{-1/2}(t_i)\{\hat{m}(t_i) - \tilde{m}_\theta(t_i)\}$ is asymptotically normally distributed at each t_i. Then (A.30) is obtained by applying the Cramér-Wold device.

From Theorem 1.5.4 of van der Vaart and Wellner (1996), we only need to show that $(nh^d)^{1/2}\hat{m}()/V^{-1/2}()$ is asymptotically tight in $C([0,1]^d)$. To simplify the presentation, we only prove the case for $d = 1$. From Theorem 8.1 and Theorem 12.3 of Billingsley (1968), we need only to show that

$$(nh^d)^{1/2} V^{-1/2}(0)\{\hat{m}(0) - \tilde{m}_{\hat{\theta}}(0)\} \text{ is tight and} \tag{A.31}$$

$$\begin{aligned} & P\{(nh^d)^{1/2} |V^{-1/2}(t_1)\{\hat{m}(t_1) - \tilde{m}_\theta(t_1)\} - V^{-1/2}(t_2)\{\hat{m}(t_2) - \tilde{m}_\theta(t_2)\}| > \eta\} \\ & \le C(t_1 - t_2)^\alpha / \eta_0^\gamma, \end{aligned} \tag{A.32}$$

for any $\eta_0 > 0$, some $\gamma > 0$ and $\alpha > 1$.

As $V^{-1/2}(0)\{\hat{m}(0) - \tilde{m}_{\hat{\theta}}(0)\}$ has finite mean and variance, (A.31) is readily established from the Markov inequality. Note that

$$\begin{aligned} & (nh^d)^{1/2}\Big[V^{-1/2}(t_1)\{\hat{m}(t_1) - \tilde{m}_\theta(t_1)\} - V^{-1/2}(t_2)\{\hat{m}(t_2) - \tilde{m}_\theta(t_2)\}\Big] \\ = \; & (nh^d)^{1/2} n^{-1} \sum_{i=1}^n \{W_h(t_1 - X_i) - W_h(t_2 - X_i)\}\{\epsilon_i + c_n \Delta(x_i)\} + o_p(1) \\ = \; & (nh^d)^{1/2} n^{-1} \sum_{i=1}^n \{W_h(t_1 - X_i) - W_h(t_2 - X_i)\}\epsilon_i + o_p(1) \end{aligned}$$

So, it is sufficient to prove for any $\eta > 0$,

$$P\{h|\sum_{i=1}^{n} Z_i| > \sqrt{nh}\eta_0\} \leq C(t_1 - t_2)^\alpha/\eta_0^\gamma. \qquad \text{(A.33)}$$

where $Z_i = h\{W_h(t_1 - X_i) - W_{0h}(t_2 - X_i)\}\epsilon_i$. Split Z_i into $Z_{i1} = Z_i I(|\epsilon_i| < M_n)$ and $Z_{i2} = Z_i I(|\epsilon_i| > M_n)$ where M_n is a larger number slowly tending to ∞. Clearly, $|Z_{i1}| \leq b =: C|t_1 - t_2|M_n/h$. Using again Theorem 1.3 of Bosq (1996),

$$\begin{aligned} Pr\{|\sum_{i=1}^{n} Z_{i1}| > \tfrac{1}{2}(nh^d)^{1/2}\eta_0\} &= Pr\left(|\textstyle\sum_{i=1}^{n} Z_{i1}| > n\eta\right) \\ \leq \quad 4\exp\{-\frac{\eta^2 q}{8v^2(q)}\} &+ C(b/\eta)^{1/2} q\alpha\{[n/(2q)]\} \end{aligned}$$

where $q = n^{1/2}h^{-3/2}M_n\eta$ and $v^2(q) = Cqh|t_1 - t_2|/n$. Thus,

$$\exp\{-\frac{\eta^2 q}{8v^2(q)}\} \leq \exp\left(-C\eta_0^2|t_1 - t_2|^{-1}\right) \leq C|t_1 - t_2|^2\eta_0^{-2}$$

and condition (v) implies that $(b/\eta)^{1/2}q\alpha\{[n/(2q)]\} \to 0$. Therefore,

$$Pr\{|\sum_{i=1}^{n} Z_{i1}| > \tfrac{1}{2}(nh^d)^{1/2}\eta_0\} \leq C|t_1 - t_2|^2\eta_0^{-2}. \qquad \text{(A.34)}$$

Standard techniques, similar to those used in studying the properties of I_2^+ in the proof of Lemma 2, show that as $n \to \infty$

$$Pr\{|\sum_{i=1}^{n} Z_{i2}| > \tfrac{1}{2}(nh^d)^{1/2}\eta_0\} \to 0.$$

This and (A.34) prove (A.33), and complete the proof for the tightness.