A goodness-of-fit test for GARCH innovation density

Hira L. Koul[1] and Nao Mimoto
Michigan State University

Abstract

We prove asymptotic normality of a suitably standardized integrated square difference between a kernel type error density estimator based on residuals and the expected value of the error density estimator based on innovations in GARCH models. This result is similar to that of Bickel-Rosenblatt under i.i.d. set up. Consequently the goodness-of-fit test for the innovation density of GARCH processes based on this statistic is asymptotically distribution free, unlike the tests based on the residual empirical process. A simulation study comparing the finite sample behavior of this test with Kolmogorov-Smirnov test and the test based on integrated square difference between the kernel density estimate and null density shows some superiority of the proposed test.

1 Introduction

The problem of fitting a given distribution function to a random sample, otherwise known as the goodness-of-fit testing problem, is a classical problem in statistics. Often omnibus tests are based on a discrepancy measure between empirical and null distribution functions (d.f.'s). These tests are easy to implement as long as there are no nuisance parameters under the null hypothesis. For example, when fitting a known continuous d.f. to the given data, Kolmogorov-Smirnov test is known to be distribution free for all sample sizes and hence easy to implement. But it looses this property when fitting an error distribution in the one sample location-scale model. In comparison, as noted by Bickel and Rosenblatt (1973), some goodness-of-fit tests based on density estimates do not suffer from this draw back. One such statistic is the integrated square difference between a density estimate and its expected value under the null hypothesis.

More precisely, let $\varepsilon_1, \cdots, \varepsilon_n$ be i.i.d. observations from a density f. Let f_0 be a given density with zero mean and finite variance and consider the problem of testing the hypothesis

$$H_0 : f = f_0, \quad \text{vs.} \quad H_1 : f \neq f_0.$$

[1]Supported in part by the NSF Grant DMS-07-04130
AMS 2000 subject classifications. Primary 62M10; secondary 62G10
Key words and phrases. L_2 difference of densities.

Define the density estimator

$$f_n(x) = \frac{1}{nh}\sum_{k=1}^{n} K\Big(\frac{x-\varepsilon_k}{h}\Big), \quad x \in \mathbb{R} := (-\infty, \infty),$$

where K is a density kernel and $h = h_n$ is a sequence of positive numbers, tending to zero. Bickel and Rosenblatt (1973) proposed to use the statistic

$$T_n = \int \big(f_n(x) - E_0 f_n(x)\big)^2 dx$$

for testing H_0, vs. H_1. Here E_0 is expectation under H_0. They proved, under H_0 and under some conditions that included second order differentiability of f_0, that as $n \to \infty$,

$$n\sqrt{h}\Big(T_n - \frac{1}{nh}\int K^2(t)dt\Big) \to_D \mathcal{N}(0, \tau^2), \quad \tau^2 = 2\int f_0^2(x)dx \int (K * K(x))^2 dx, \quad (1.1)$$

where $g_1 * g_2(x) := \int g_1(x-t)g_2(t)dt$, for any two integrable functions g_1, g_2. Bachmann and Dette (2005) weakened their conditions required for (1.1), and also established the following asymptotic normality result for T_n under the alternatives $\mathcal{H}_1 : f \neq f_0$, $\int (f(x) - f_0(x))^2 dx > 0$. Assuming only f, f_0 to be continuous and square integrable, they proved

$$\sqrt{n}\Big(T_n - \int \big\{K_h * (f - f_0)\big\}^2(x)dx\Big) \to_D N(0, 4\omega^2), \quad n \to \infty, \quad (1.2)$$

where $\omega^2 = \mathrm{Var}[f(\varepsilon_0) - f_0(\varepsilon_0)]$, and $K_h(\cdot) = (1/h)K(\cdot/h)$. Bachmann and Dette (2005) mention they assume f, f_0 to be twice continuously differentiable with bounded second derivatives, but a close inspection of their proofs of (1.1) and (1.2) shows that all they need is f, f_0 to be continuous and square integrable.

Now consider the problem of fitting a zero mean density f_0 to the error density of a stationary linear autoregressive model of a known order. Lee and Na (2002) and Bachmann and Dette (2005) showed that (1.1) and (1.2) continue to hold for an analog of T_n based on autoregressive residuals, under H_0 and $\mathcal{H}_1$, respectively. In other words, not knowing nuisance autoregressive parameters has no effect on asymptotic level of the test based on this analog for fitting f_0 to the error density in this model.

In this paper we consider the problem of fitting density f_0 to the error density of a generalized autoregressive conditionally heteroscedastic (GARCH(p, q)) model, where p and q are known positive integers. We provide some sufficient conditions under which (1.1) continues to hold for $\widehat{T}_n$, an analog of T_n based on GARCH residuals, defined at (2.4) below. In addition, we establish a first order expansion of $\widehat{T}_n$ under $\mathcal{H}_1$. This expansion shows that unlike in linear autoregressive models, the estimation of the model parameters affects the asymptotic distribution of $\widehat{T}_n$ under $\mathcal{H}_1$.

Mimoto (2008) showed that the goodness-of-fit test for the error density in GARCH models based on a suitably standardized sup-norm statistic $\|\widehat{f}_n - f_0\|_\infty$ has the same asymptotic null distribution as in the i.i.d. set up. Cheng (2008) derives a similar result in the case of ARCH models.

This paper is organized as follows. In the next section we describe the model, assumptions and recall some preliminaries from Berkes, Horváth and Kokoszka (2003). Asymptotic normality of $\widehat{T}_n$ under H_0 and a first order approximation of $\widehat{T}_n$ under $\mathcal{H}_1$ are given in section 3 with proofs appearing in section 5. Section 4 contains a simulation study comparing $\widehat{T}_n$ test with the Kolmogorov-Smirnov test and the one based on $\int(\widehat{f}_n - f_0)^2$. The proofs given below use several results from Berkes et al. (2003) and Horváth and Zitikis (2006) about some properties of GARCH models. Many details of the proofs are different from those appearing in these papers and Mimoto (2008).

2 Model, some preliminaries and assumptions

In this section, we describe the model, review some known results about the model, and state the needed assumptions. Let p, q be known positive integers, y_k, $k \in \mathbb{Z} := \{0, \pm 1, \cdots,\}$ be the GARCH(p, q) process satisfying

$$y_k = \sigma_k \varepsilon_k, \qquad \sigma_k^2 = \omega + \sum_{1 \le i \le p} \alpha_i y_{k-i}^2 + \sum_{1 \le j \le q} \beta_j \sigma_{k-j}^2, \tag{2.1}$$

where $\boldsymbol{\theta} = (\omega, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q)^T$ is the parameter vector of the process, with $\omega > 0$; $\alpha_i \ge 0$, $1 \le i \le p$; $\beta_j \ge 0$, $1 \le j \le q$. The innovations ε_k, $k \in \mathbb{Z}$, are assumed to be i.i.d. random variables with a density function f having zero mean and finite variance.

Necessary and sufficient conditions for the existence of a unique stationary solution of (2.1) have been specified by Nelson (1990) for $p = 1,\ q = 1$, and by Bougerol and Picard (1992a, 1992b) for $p \ge 1$, $q \ge 1$. In particular, for GARCH(1,1) model, $E \log(\beta_1 + \alpha_1 \varepsilon_0^2) < 0$ implies stationarity of the process. Because $e^x \ge 1 + x$, for all x, we have $E \log(\beta_1 + \alpha_1 \varepsilon_0^2) \le \beta_1 + \alpha_1 E(\varepsilon^2) - 1$. Thus, if error density is standardized to have zero mean and unit variance, then $\beta_1 + \alpha_1 < 1$ implies the stationarity of the GARCH$(1, 1)$ process.

A major characteristic of the GARCH process is that the past dependency of the observations y_k is only through the unobservable conditional variance σ_k^2. Berkes et al. (2003) show that under suitable conditions, σ_k^2 admits a unique representation as the infinite sum of y_k^2, allowing σ_k^2 to be estimated from the observations. One of them is to assume that the polynomials

$$\alpha_1 x + \cdots + \alpha_p x^p \text{ and } 1 - \beta_1 x - \cdots - \beta_q x^q \text{ are coprimes} \tag{2.2}$$

on the set of polynomials with real coefficients.

This is to ensure that the equations (2.1) are true only with $\boldsymbol{\theta}$ and there is no other parameter that satisfies the equation.

Let $\boldsymbol{u} = (r, s_1, \cdots, s_p, t_1, \cdots, t_q)$ denote a generic element of the parameter space

$$\begin{aligned} U &:= \big\{\boldsymbol{u} : t_1 + \cdots + t_q \le \rho_0, \\ &\qquad \underline{u} < \min(r, s_1, \cdots, s_p, t_1, \cdots, t_q) \le \max(r, s_1, \cdots, s_p, t_1, \cdots, t_q) \le \overline{u}\big\}, \end{aligned}$$

where $0 < \underline{u} < \overline{u},\ 0 < \rho_0 < 1,\ q\underline{u} < \rho_0$. With coefficients $c_i(\boldsymbol{u}),\ 0 \le i < \infty$ as in Berkes et al. (2003), let

$$w_k(\boldsymbol{u}) = c_0(\boldsymbol{u}) + \sum_{1 \le i < \infty} c_i(\boldsymbol{u}) y_{k-i}^2.$$

Assuming that $E|\varepsilon_0|^\delta < \infty$ for some $\delta > 0$ and $\boldsymbol{\theta}$ is an interior point of U with none of the coordinates equal to zero, Berkes et al. showed that $\sigma_k^2 = w_k(\boldsymbol{\theta})$, for all $k \in \mathbb{Z}$, a.s. Assuming further that the distribution of ε_0^2 is non-degenerate, this representation is almost surely unique. With given observations $y_k, 1 \le k \le n$, this representation allows one to estimate σ_k^2 by the truncated version

$$\widehat{\sigma}_k^2 = \widehat{w}_k(\boldsymbol{\theta}_n) = c_0(\boldsymbol{\theta}_n) + \sum_{i=1}^{k-1} c_i(\boldsymbol{\theta}_n) y_{k-i}^2, \tag{2.3}$$

where $\boldsymbol{\theta}_n$ is an estimator of $\boldsymbol{\theta}$ based on $y_k, 1 \le k \le n$. This leads to the GARCH residuals $\{\widehat{\varepsilon}_k = y_k/\widehat{\sigma}_k\}$ and to the GARCH error density estimate

$$\widehat{f}_n(x) = \frac{1}{nh} \sum_{k=1}^{n} K\Big(\frac{x - \widehat{\varepsilon}_k}{h}\Big).$$

The proposed test for H_0 is to be based on

$$\widehat{T}_n = \int \Big(\widehat{f}_n(x) - E_0 f_n(x)\Big)^2 dx. \tag{2.4}$$

We shall now state additional needed assumptions for obtaining asymptotic distributions of $\widehat{T}_n$ under H_0 and $\mathcal{H}_1$. About $\boldsymbol{\theta}_n$ assume

$$\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}) = O_p(1). \tag{2.5}$$

Lee and Hansen (1994), Lumsdaine (1996) and Berkes et al. (2003) discuss some sufficient conditions that imply (2.5) for the sequence of quasi-maximum likelihood estimators.

About the kernel K, assume

K is a bounded symmetric density on $[-1, 1]$, vanishing off $(-1, 1)$, and twice differentiable with bounded derivative K' and $\int \big(K''(z)\big)^2 dz < \infty$, (2.6) where g' and g'' denote, respectively, the first and second derivatives of a smooth function g.

In order to use results from Berkes et al. (2003), we need to assume

$$E|\varepsilon_0^2|^\delta < \infty, \qquad \text{for some } \delta > 1. \tag{2.7}$$

We say density f satisfies condition $C(f)$ if the following holds.

$$\begin{aligned}
&f \text{ is absolutely continuous with its a.e. derivative } \dot{f} \text{ satisfying} && (2.8)\\
&I_\ell(f) := \int \Big(\frac{\dot{f}(x)}{f(x)}\Big)^2 f(x)dx < \infty, \quad I_s(f) := \int \Big(1 + x\frac{\dot{f}(x)}{f(x)}\Big)^2 f(x)dx < \infty,\\
&\int\int x^2\big\{\dot{f}(x-zh) - \dot{f}(x)\big\}^2 dxK(z)dz \to 0, \qquad \text{as } h \to 0.
\end{aligned}$$

Note that $I_\ell(f)$ and $I_s(f)$ are, respectively, Fisher information for location and scale parameters in one observation from f. Their finiteness together imply f is bounded, Lipschitz (1/2), cf. Koul (2002, p78), and

$$\int (\dot{f}(x))^2 dx < \infty, \quad \int \big(x\dot{f}(x)\big)^2 dx < \infty, \quad \text{and} \quad \int \big(f(x) + x\dot{f}(x)\big)^2 dx < \infty. \tag{2.9}$$

Also, note that f being bounded implies that $P\{\varepsilon_0^2 \le t\} = o(t^\eta)$, as $t \to 0$, for some $\eta > 0$, which is one of the conditions required in Berkes et al. (2003).

For the bandwidth h, we assume

$$h \to 0, \quad nh^5 \to \infty, \quad \text{as } n \to \infty. \tag{2.10}$$

Condition $C(f)$ avoids assuming higher order differentiability of f. Many smooth densities can be shown to satisfy $C(f)$. An example of non-differentiable density that also satisfies this condition is double exponential density $f(x) = e^{-|x|}/2$. For, clearly a.e. derivative of this $f(x)$ is $\dot{f}(x) =$ -sign$(x)f(x)$. Hence,

$$\begin{aligned}
\mathcal{I} &:= \int\int x^2\big[\dot{f}(x-zh) - \dot{f}(x)\big]^2 dxK(z)dz\\
&= \int\int x^2\big[\text{sign}(x-zh)\big\{f(x-zh) - f(x)\big\} + \big\{\text{sign}(x-zh) - \text{sign}(x)\big\}f(x)\big]^2 dxK(z)dz\\
&\le 2(I_1 + I_2),
\end{aligned}$$

where

$$\begin{aligned}
I_1 &:= \int\int x^2[f(x-zh) - f(x)]^2 dxK(z)dz,\\
I_2 &:= \int\int x^2\big\{\text{sign}(x-zh) - \text{sign}(x)\big\}^2 f^2(x)dxK(z)dz.
\end{aligned}$$

But, for $z > 0$, and because f is a density bounded by $1/2$,

$$\begin{aligned} I_1 &= \int\int x^2\Big(\int_0^{zh} \text{sign}(x-s)f(x-s)ds\Big)^2 dxK(z)dz \\ &\leq 2h^2\int z^2K(z)dz + (2/3)h^4\int |z|^3K(z)dz, \\ I_2 &= 4\int\int_0^{hz} x^2f^2(x)dxK(z)dz \leq (1/3)h^3\int |z|^3K(z)dz. \end{aligned}$$

Similar facts hold for $z < 0$. Hence, $\mathcal{I} \to 0$, as $n \to \infty$, because $h \to 0$. The rest of the conditions in (2.8) are easy to verify in this case.

Throughout the rest of the paper, for any square integrable function g on $\mathbb{R}$, its L_2 norm is denoted by $\|g\|_2 := \big(\int g^2(x)dx\big)^{1/2}$, and all limits are taken as $n \to \infty$, unless stated otherwise.

3 Main Results

In the first theorem below we give some preliminary results about $\widehat{f}_n$. To state this result we need to introduce

$$g_n(x) = -\frac{1}{2}(\boldsymbol{\theta}_n - \boldsymbol{\theta})^T\frac{1}{n}\sum_{k=1}^n \frac{\boldsymbol{w}_k'(\boldsymbol{\theta})}{w_k(\boldsymbol{\theta})}\frac{1}{h^2}E\Big[\varepsilon_0K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big],$$

where $\boldsymbol{w}_k'(\boldsymbol{\theta})$ is the column vector of length $p+q+1$, consisting of the first derivatives of $w_k(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$, and where ^{T} denotes the transpose. We are now ready to state

Theorem 3.1 *Suppose the GARCH(p, q) model (2.1) is stationary with the true error density f satisfying $C(f)$. In addition, assume (2.2), and (2.5)-(2.7) hold. Then,*

$$\|\widehat{f}_n - f_n - g_n\|_2 = O_p(\frac{1}{nh^{5/2}}), \tag{3.1}$$

and

$$\sqrt{n}\|g_n\|_2 = O_p(1). \tag{3.2}$$

Consequently, if also (2.10) holds, then

$$\sqrt{n}\|\widehat{f}_n - f_n\|_2 = O_p(1). \tag{3.3}$$

Theorem 3.1 is useful in establishing an analog of the result (1.1) for $\widehat{T}_n$ as follows. We shall first approximate $\widehat{T}_n$ by T_n. Assume H_0 holds and recall E_0 denotes the expectation under H_0. Direct calculations show that

$$\begin{aligned}
& n\sqrt{h}\big(\widehat{T}_n - T_n\big) \qquad (3.4)\\
= \;& n\sqrt{h}\int \Big(\widehat{f}_n(x) - E_0 f_n(x)\Big)^2 dx - n\sqrt{h}\int \Big(f_n(x) - E_0 f_n(x)\Big)^2 dx \\
= \;& n\sqrt{h}\int \Big(\widehat{f}_n(x) - f_n(x)\Big)^2 dx + 2n\sqrt{h}\int \big(\widehat{f}_n(x) - f_n(x)\big)\big(f_n(x) - E_0 f_n(x)\big) dx.
\end{aligned}$$

By (2.10) and (3.3), the first term is $o_p(1)$. The following proposition shows the same holds for the second.

Proposition 3.1 *Suppose the conditions of Theorem 3.1 hold with $f = f_0$. Then, under (2.10) and H_0,*

$$n\sqrt{h}\int \big(\widehat{f}_n(x) - f_n(x)\big)\big(f_n(x) - E_0 f_n(x)\big) dx \to_p 0.$$

This proposition together with (3.4) yields the following corollary.

Corollary 3.1 *Suppose the GARCH(p, q) model (2.1) is stationary and (2.5) - (2.7), $C(f_0)$, and (2.10) hold. Then, under H_0,*

$$n\sqrt{h}\Big(\widehat{T}_n - T_n\Big) = o_p(1),$$

and hence,

$$n\sqrt{h}\Big(\widehat{T}_n - \frac{1}{nh}\int K^2(x)dx\Big) \to_D \mathcal{N}(0, \tau^2), \quad \tau^2 = 2\int f_0^2(x)dx \int (K * K)^2(x)dx. \quad (3.5)$$

Remark 3.1 An alternative test of H_0 using density estimates could be based on $\widetilde{T}_n := \int(\widehat{f}_n(x) - f_0(x))^2 dx$. Upon taking $v = 2$ in (3.5) of Horváth and Zitikis (2006) one obtains that under (2.10) and $E|\varepsilon_0|^{3+\delta} < \infty$ with some $\delta > 0$, and under some conditions on K and f that are stronger than those given above, $n\sqrt{h}\Big(\widetilde{T}_n - \int K^2(x)dx/nh\Big) \to_D \mathcal{N}(0, \tau^2)$.

It is important to point out that under our assumptions, (3.5) does not follow directly from asymptotic normality of $\widetilde{T}_n$, because $n\sqrt{h}|\widehat{T}_n - \widetilde{T}_n| \to_p \infty$. To see this, consider

$$\begin{aligned}
\widehat{T}_n - \widetilde{T}_n &= -\int (E_0 f_n - f_0)^2(x)dx + 2\int (f_0 - E_0 f_n)(\widehat{f}_n - E_0 f_n)(x)dx \\
&= -\int (E_0 f_n - f_0)^2(x)dx + 2\int (f_0 - E_0 f_n)(\widehat{f}_n - f_n)(x)dx \\
&\qquad + 2\int (f_0 - E_0 f_n)(f_n - E_0 f_n)(x)dx.
\end{aligned}$$

But,

$$\begin{aligned}
\int (E_0 f_n - f_0)^2(x)dx &= \int \Big(\int K(z)(f_0(x - zh) - f_0(x))dz\Big)^2 dx \\
&= \int \Big(\int K(z)\int_0^{zh} \dot{f}_0(x - s)ds dz\Big)^2 dx \;=\; O_p(h^2).
\end{aligned}$$

Hence, under (2.10),

$$\sqrt{nh}\int (E_0 f_n - f_0)^2(x)dx \;\; = \;\; O_p(\sqrt{nh^5}) \to_p \infty.$$

Next, by the Cauchy-Schwarz inequality, (1.1) and (3.3),

$$\begin{aligned}
&\sqrt{nh}\Big|\int (f_0 - E_0 f_n)(\widehat{f}_n - f_n)(x)dx\Big| \\
&\leq \;\Big(\int (E_0 f_n - f_0)^2(x)dx\Big)^{1/2}\Big(nh\int (\widehat{f}_n - f_n)^2(x)dx\Big)^{1/2} = o_p(1) \\
&\sqrt{nh}\Big|\int (f_0 - E_0 f_n)(f_n - E_0 f_n)(x)dx\Big| \\
&\leq \;\Big(\int (E_0 f_n - f_0)^2(x)dx\Big)^{1/2}\Big(nh\int (f_n - E_0 f_n)^2(x)dx\Big)^{1/2} = o_p(1)
\end{aligned}$$

Therefore, $\sqrt{nh}|\widehat{T}_n - \widetilde{T}_n| \to_p \infty$, and also $n\sqrt{h}|\widehat{T}_n - \widetilde{T}_n| \to_p \infty$.

The next theorem gives the first order limiting behavior of $\widehat{T}_n$ under the alternative $\mathcal{H}_1$.

Theorem 3.2 *Suppose the GARCH(p, q) model (2.1) is stationary, and (2.5) - (2.7), $C(f)$, $C(f_0)$, and (2.10) hold. Then, under $\mathcal{H}_1$,*

$$\begin{aligned}
&\sqrt{n}\Big(\widehat{T}_n - \int \Big\{K_h * (f - f_0)(x)\Big\}^2 dx\Big) \\
&= \;\sqrt{n}\Big(T_n - \int \Big\{K_h * (f - f_0)(x)\Big\}^2 dx\Big) \\
&\qquad - \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T E\Big[\frac{\boldsymbol{w}_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}\Big] \int \big(f(x) + x\dot{f}(x)\big)(f(x) - f_0(x))dx + o_p(1).
\end{aligned}$$

This result is unlike the result (1.2) in AR models because of the presence of the second term in the right hand side above. A primary reason that an analog of this term is absent in the AR model is that the analog of $\boldsymbol{w}_0'(\boldsymbol{\theta})/w_0(\boldsymbol{\theta})$ in the AR(p) model is $(y_{1-p}, y_{2-p}, \cdots, y_0)^T$ whose expected value is zero when fitting an error density with zero mean. But here these entities come from a scale factor and hence their expectation can not be zero.

4 Simulation Study

This section contains results of a simulation study illustrating a finite sample performance of the goodness-of-fit tests based on $\widehat{T}_n$ of Corollary 3.1 and $\widetilde{T}_n$ of Remark 3.1 against the Kolmogolov-Smirnov (KS) test based on $n^{1/2}\sup_{x\in\mathbb{R}}|\widehat{F}_n(x) - F_0(x)|$, where

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(\widehat{\epsilon}_i \leq x), \quad x \in \mathbb{R}.$$

In the study, GARCH(1,1) process with $\omega = 0.5, \alpha_1 = .4, \beta_1 = .2$ (so that $\boldsymbol{\theta} = (.5, .4, .2)^T$) of varying length n were simulated, each iterated 10,000 times. Note that $\alpha_1 + \beta_1 = .6 < 1$, and to ensure stationarity further, first 500 observations were not included in each simulation. Estimator $\boldsymbol{\theta}_n$ was obtained by the quasi-maximum likelihood method.

Seven densities were used in the study: standard normal density (N), Student-t densities with degrees of freedom 40, 20, 10, and 5 ($T40, T20, T10, T5$), double exponential density (D), and logistic density (L). All densities were standardized to have mean zero and variance 1. For f_0, we chose normal, double exponential, and logistic densities, all standardized. For the kernel function, we used $K(u) = (3/4)(1-u^2)I(|u| \le 1)$ and bandwidth

$$h = \Big(\int K^2(x)dx \Big/ \int (f_0''(x))^2 dx (\int x^2 K(x) dx)^2\Big)^{1/5} n^{-1/t}$$

with $t = 5.1$. Note that the above h with $t = 5$ is an optimum bandwidth which minimizes the asymptotic mean integrated square error of kernel density estimators. For this simulation, $t = 5.1$ is chosen to satisfy assumption (2.10).

Test based on empirical (asymptotic) critical values of $\widehat{T}_n$ is denoted by $\widehat{T}_{n,e}$ ($\widehat{T}_{n,a}$). Define $\widetilde{T}_{n,e}$ and $\widetilde{T}_{n,a}$ similarly. Tables 1-3 contain empirical sizes and powers of these tests and of the KS test using empirical critical values only. The first row entries in all tables are empirical sizes and should be close to the nominal level .05.

In Table 1, f_0 is the standard normal. Test $\widehat{T}_{n,e}$ clearly outperforms all the other tests for all chosen sample sizes.

In Table 2, $f_0 = D$, the double exponential density. Empirical powers of all the tests quickly becomes large for $n = 500$, 1000. For all sample sizes chosen, KS test has worse empirical power compared to that of $\widehat{T}_{n,e}$ and $\widetilde{T}_{n,e}$, with $\widetilde{T}_{n,e}$ dominating $\widehat{T}_{n,e}$ for $n = 100$. For all the values of n considered, $\widetilde{T}_{n,a}$ test suffers from large bias of the kernel density estimation around its peak.

In Table 3, $f_0 = L$, the logistic density. Against N, T40, T20 and T10 alternatives, $\widetilde{T}_{n,e}$ performs the best. On the other hand, against the alternatives T5 and D, $\widehat{T}_{n,e}$ performs the best for all n considered, and KS test has higher empirical power than $\widetilde{T}_{n,e}$ for $n = 100, 500$.

Poor performance of the tests $\widehat{T}_{n,a}$ and $\widetilde{T}_{n,a}$ based on asymptotic critical values results from the fact that Monte Carlo distributions of the statistics $\widehat{T}_n$ and $\widetilde{T}_n$ do not appear to approximate their asymptotic distributions well even for $n = 1000$. This is also a reason for no substantial improvement in the empirical power of these tests when n is increased. Empirical size of the $\widetilde{T}_{n,a}$ is especially worse when $f_0 = D$.

Table 1: Empirical sizes and powers of the tests when $f_0 = N(0,1)$.

n	$f\backslash$ Tests	$\widehat{T}_{n,e}$	$\widetilde{T}_{n,e}$	KS	$\widehat{T}_{n,a}$	$\widetilde{T}_{n,a}$
100	f_0=N	0.050	0.050	0.050	0.009	0.019
	T40	0.046	0.039	0.045	0.008	0.016
	T20	0.049	0.034	0.047	0.010	0.015
	T10	0.072	0.031	0.052	0.013	0.016
	T5	0.219	0.051	0.091	0.054	0.030
	D	0.539	0.109	0.230	0.178	0.042
	L	0.090	0.030	0.056	0.015	0.015
500	f_0=N	0.050	0.050	0.050	0.009	0.029
	T40	0.056	0.029	0.051	0.010	0.015
	T20	0.083	0.022	0.056	0.016	0.012
	T10	0.264	0.038	0.098	0.087	0.023
	T5	0.912	0.502	0.478	0.752	0.421
	D	1.000	0.998	0.976	0.999	0.996
	L	0.537	0.106	0.179	0.252	0.071
1000	f_0=N	0.050	0.050	0.050	0.011	0.032
	T40	0.061	0.024	0.051	0.013	0.016
	T20	0.131	0.024	0.069	0.038	0.016
	T10	0.521	0.098	0.170	0.280	0.072
	T5	0.997	0.930	0.865	0.988	0.906
	D	1.000	1.000	0.999	1.000	1.000
	L	0.860	0.370	0.349	0.678	0.313

Table 2: Empirical sizes and powers of the tests when $f_0 = D$.

n	$f\backslash$ Tests	$\widehat{T}_{n,e}$	$\widetilde{T}_{n,e}$	KS	$\widehat{T}_{n,a}$	$\widetilde{T}_{n,a}$
100	f_0=D	0.050	0.050	0.050	0.026	0.173
	N	0.732	0.831	0.278	0.525	0.971
	T40	0.671	0.782	0.252	0.470	0.956
	T20	0.592	0.722	0.226	0.396	0.939
	T10	0.445	0.582	0.188	0.271	0.869
	T5	0.210	0.295	0.121	0.110	0.617
	L	0.329	0.458	0.152	0.184	0.783
500	f_0=D	0.050	0.050	0.050	0.016	0.386
	N	1.000	1.000	0.993	1.000	1.000
	T40	1.000	1.000	0.980	1.000	1.000
	T20	1.000	1.000	0.957	1.000	1.000
	T10	0.998	1.000	0.857	0.992	1.000
	T5	0.839	0.928	0.424	0.707	0.996
	L	0.983	0.995	0.718	0.954	1.000
1000	f_0=D	0.050	0.050	0.050	0.020	0.564
	N	1.000	1.000	1.000	1.000	1.000
	T40	1.000	1.000	1.000	1.000	1.000
	T20	1.000	1.000	1.000	1.000	1.000
	T10	1.000	1.000	0.996	1.000	1.000
	T5	0.986	0.998	0.717	0.966	1.000
	L	1.000	1.000	0.976	1.000	1.000

Table 3: Empirical sizes and powers of the tests when $f_0 = L$.

n	$f\backslash$ Tests	$\widehat{T}_{n,e}$	$\widetilde{T}_{n,e}$	KS	$\widehat{T}_{n,a}$	$\widetilde{T}_{n,a}$
	f_0=L	0.050	0.050	0.050	0.015	0.042
	N	0.118	0.210	0.077	0.029	0.178
	T40	0.095	0.164	0.064	0.023	0.138
	T20	0.082	0.132	0.063	0.019	0.109
100	T10	0.060	0.080	0.058	0.018	0.067
	T5	0.077	0.045	0.060	0.030	0.040
	D	0.199	0.027	0.112	0.056	0.024
	f_0=L	0.050	0.050	0.050	0.016	0.059
	N	0.558	0.777	0.184	0.320	0.806
	T40	0.393	0.613	0.140	0.192	0.649
	T20	0.255	0.441	0.106	0.104	0.478
500	T10	0.087	0.153	0.058	0.028	0.174
	T5	0.202	0.026	0.090	0.102	0.030
	D	0.961	0.592	0.621	0.897	0.628
	f_0=L	0.050	0.050	0.050	0.015	0.068
	N	0.886	0.972	0.379	0.740	0.982
	T40	0.721	0.898	0.260	0.514	0.924
	T20	0.492	0.728	0.180	0.297	0.780
1000	T10	0.130	0.241	0.071	0.053	0.287
	T5	0.379	0.054	0.169	0.245	0.067
	D	1.000	0.981	0.953	0.999	0.987

5 Proofs

This section contains the proofs of some of the claims of section 3. The following Cauchy-Schwarz inequality is used repeatedly in the proofs. For any real sequences a_k, b_k, $1 \le k \le n$,

$$\Big(\sum_{k=1}^{n} a_k b_k\Big)^2 \le \Big(\sum_{k=1}^{n} |a_k|\Big)\Big(\sum_{k=1}^{n} |a_k||b_k|^2\Big). \tag{5.1}$$

We also need to recall the following facts from Berkes et al. (2003). Facts 5.1 - 5.5 below are Lemmas 2.2, 2.3, 5.1, 5.6, and (5.35) in Berkes et al. (2003), respectively.

Fact 5.1 *Let* $\log^+ x = \log x$ *if* $x > 1$*, and* 0 *otherwise. If* $\{\zeta_k, 0 \le k < \infty\}$ *is a sequence of identically distributed random variables satisfying* $E \log^+ |\zeta_0| < \infty$*, then* $\sum_{0 \le k < \infty} \zeta_k z^k$ *converges a.s., for all* $|z| < 1$.

Under the assumptions of Theorem 3.1 the following facts hold.

Fact 5.2 *There exists a* $\delta^* > 0$*, depending on* δ *of (2.7), such that* $E|y_0^2|^{\delta^*} + E|\sigma_0^2|^{\delta^*} < \infty$.

Fact 5.3 $E\big[\sup_{\boldsymbol{u}\in U} \sigma_k^2/w_k(\boldsymbol{u})\big]^{\nu} < \infty$*, for any* $0 < \nu < \delta$*, where* δ *is as in (2.7).*

Fact 5.4

$$E \sup_{\boldsymbol{u}\in U} \Big\| \frac{\boldsymbol{w}_0'(\boldsymbol{u})}{w_0(\boldsymbol{u})} \Big\|^{\nu} < \infty \quad \textit{and} \quad E \sup_{\boldsymbol{u}\in U} \Big\| \frac{\boldsymbol{w}_0''(\boldsymbol{u})}{w_0(\boldsymbol{u})} \Big\|^{\nu} < \infty, \quad \forall\, \nu > 0,$$

where $\|\cdot\|$ *denotes the maximum norm of vectors and matrices. This implies that*

$$E \sup_{\boldsymbol{u}\in U} \big\|\boldsymbol{w}_0''(\boldsymbol{u})/w_0(\boldsymbol{u})\big\|_E^{\nu} < \infty, \quad \forall\, \nu > 0,$$

where, for any $m \times r$ *matrix* $A = ((a_{ij}))$, $\|A\|_E := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{r} a_{ij}^2}$.

Fact 5.5 *For all* $\boldsymbol{u} \in U$,

$$\frac{|w_k(\boldsymbol{u}) - \widehat{w}_k(\boldsymbol{u})|}{\widehat{w}_k(\boldsymbol{u})} \le \frac{C_2}{C_1} \rho_0^{k/q} \sum_{0 \le j < \infty} \rho_0^{j/q} y_{-j}^2$$

where ρ_0 *is from the definition of* U*, and* $0 < C_1, C_2 < \infty$.

We would like to point out that Fact 5.4 is the corrected version of Lemma 3.4 of Mimoto (2008) where $w_0^2(\boldsymbol{u})$ appears in the denominators.

We now proceed with the proof of Theorem 3.1. Throughout the proof below, let

$$\boldsymbol{r}_k(\boldsymbol{u}) := \frac{\boldsymbol{w}_k'(\boldsymbol{u})}{w_k(\boldsymbol{u})}, \quad \boldsymbol{R}_k(\boldsymbol{u}) := \frac{\boldsymbol{w}_k''(\boldsymbol{u})}{w_k(\boldsymbol{u})}, \quad \boldsymbol{r}_k := \frac{\boldsymbol{w}_k'(\boldsymbol{\theta})}{w_k(\boldsymbol{\theta})}, \quad \boldsymbol{S}_n := \sum_{k=1}^{n} \boldsymbol{r}_k, \quad \overline{\boldsymbol{S}}_n := \frac{1}{n}\boldsymbol{S}_n. \tag{5.2}$$

Because the underlying process is stationary, by the Ergodic Theorem, in view of Fact 5.4,

$$\overline{\boldsymbol{S}}_n \to E\boldsymbol{r}_0 = E\frac{\boldsymbol{w}_0'(\boldsymbol{\theta})}{w_0(\boldsymbol{\theta})}, \quad \text{a.s.} \tag{5.3}$$

This notation and fact is often used in the sequel.

Proof of Theorem 3.1. Define

$$\sigma_{kn}^2 = w_k(\boldsymbol{\theta}_n) = c_0(\boldsymbol{\theta}_n) + \sum_{1\le i<\infty} c_i(\boldsymbol{\theta}_n) y_{k-i}^2.$$

Let $\{\widetilde{\varepsilon}_k = y_k/\sigma_{kn}\}$ be the non-truncated version of the residuals and let

$$\widetilde{f}_n(x) := \frac{1}{nh}\sum_{k=1}^n K\Big(\frac{x-\widetilde{\varepsilon}_k}{h}\Big).$$

Now write $\widehat{f}_n - f_n - g_n = \widehat{f}_n - \widetilde{f}_n + \widetilde{f}_n - f_n - g_n$, so that

$$\|\widehat{f}_n - f_n - g_n\|_2 \le \|\widehat{f}_n - \widetilde{f}_n\|_2 + \|\widetilde{f}_n - f_n - g_n\|_2. \tag{5.4}$$

We claim

$$\big\|\widehat{f}_n - \widetilde{f}_n\big\|_2 = O\big(\frac{1}{nh^{3/2}}\big), \quad \text{a.s.} \tag{5.5}$$

Use the Mean-Value Theorem and the triangle inequality to obtain

$$\begin{aligned} |\widehat{f}_n(x) - \widetilde{f}_n(x)| &= \Big|\frac{1}{nh}\sum_{k=1}^n \Big[K\Big(\frac{x-\widehat{\varepsilon}_k}{h}\Big) - K\Big(\frac{x-\widetilde{\varepsilon}_k}{h}\Big)\Big]\Big| \\ &\le \frac{1}{nh^2}\sum_{k=1}^n |\widetilde{\varepsilon}_k - \widehat{\varepsilon}_k|\big|K'\big(\frac{x-\eta_k}{h}\big)\big|, \end{aligned}$$

where $\eta_k = \varepsilon_k + c^*(\widehat{\varepsilon}_k - \varepsilon_k)$, for some $0 < c^* < 1$. Hence, (5.1) and a routine argument yields

$$\begin{aligned} \int \big|\widehat{f}_n(x) - \widetilde{f}_n(x)\big|^2 dx &\le \frac{1}{n^2h^3}\Big(\sum_{k=1}^n |\widetilde{\varepsilon}_k - \widehat{\varepsilon}_k|\Big)\Big(\sum_{k=1}^n |\widetilde{\varepsilon}_k - \widehat{\varepsilon}_k| \int \frac{1}{h}\Big|K'\Big(\frac{x-\eta_k}{h}\Big)\Big|^2 dx\Big) \\ &= \frac{1}{n^2h^3}\Big(\sum_{k=1}^n |\widetilde{\varepsilon}_k - \widehat{\varepsilon}_k|\Big)\Big(\sum_{k=1}^n |\widetilde{\varepsilon}_k - \widehat{\varepsilon}_k| \int \big|K'(z)\big|^2 dx\Big) \\ &= \frac{1}{n^2h^3}\Big(\sum_{k=1}^n |\widetilde{\varepsilon}_k - \widehat{\varepsilon}_k|\Big)^2 \int \big|K'(z)\big|^2 dz. \end{aligned}$$

We shall show

$$\sum_{k=1}^\infty |\widetilde{\varepsilon}_k - \widehat{\varepsilon}_k| = O(1), \quad \text{a.s.} \tag{5.6}$$

This fact and (2.6) then completes the proof of (5.5).

To prove (5.6), observe that

$$\begin{aligned}
|\widetilde{\varepsilon}_k - \widehat{\varepsilon}_k| &= |y_k|\Big|\frac{1}{\sqrt{w_k(\boldsymbol{\theta}_n)}} - \frac{1}{\sqrt{\widehat{w}_k(\boldsymbol{\theta}_n)}}\Big| \\
&= \frac{|y_k|}{\sqrt{w_k(\boldsymbol{\theta}_n)}}\ \Big|\frac{w_k(\boldsymbol{\theta}_n) - \widehat{w}_k(\boldsymbol{\theta}_n)}{\sqrt{\widehat{w}_k(\boldsymbol{\theta}_n)}\Big(\sqrt{w_k(\boldsymbol{\theta}_n)} + \sqrt{\widehat{w}_k(\boldsymbol{\theta}_n)}\Big)}\Big| \\
&\le \sup_{\boldsymbol{u}\in U}\frac{|y_k|}{\sqrt{w_k(\boldsymbol{u})}}\ \Big|\frac{w_k(\boldsymbol{u}) - \widehat{w}_k(\boldsymbol{u})}{2\widehat{w}_k(\boldsymbol{u})}\Big| \\
&\le \sup_{\boldsymbol{u}\in U}\frac{|y_k|}{\sqrt{w_k(\boldsymbol{u})}}\ \Big(\frac{C_2}{2C_1}\rho_0^{k/q}\sum_{j=0}^{\infty}\rho_0^{j/q}y_{-j}^2\Big), \quad \forall\, k \ge 1,
\end{aligned}$$

where $0 < C_1, C_2 < \infty$, and ρ_0 is from the definition of U. We obtain the last but one upper bound above by the fact that $\widehat{w}_k(\boldsymbol{u}) \le w_k(\boldsymbol{u})$, for all $k \ge 1$ and $\boldsymbol{u} \in U$, and the last upper bound by Fact 5.5. Therefore,

$$\sum_{k=1}^{\infty}|\widetilde{\varepsilon}_k - \widehat{\varepsilon}_k| \ \le\ \frac{C_2}{2C_1}\Big(\sum_{j=0}^{\infty}\rho_0^{j/q}y_{-j}^2\Big)\Big(\sum_{k=1}^{\infty}\sup_{\boldsymbol{u}\in U}\frac{|y_k|}{\sqrt{w_k(\boldsymbol{u})}}\rho_0^{k/q}\Big).$$

Note that $\sup_{\boldsymbol{u}\in U}|y_k|/\sqrt{w_k(\boldsymbol{u})}$ is a stationary sequence, and so is y_{-j}^2. By Fact 5.2, (2.7) implies $E|y_0^2|^{\delta^*} < \infty$, for some $\delta^* > 0$. By the independence of ε_k and $w_k(\boldsymbol{u})$ and Fact 5.3,

$$E\Big[\sup_{\boldsymbol{u}\in U}\frac{|y_k|}{\sqrt{w_k(\boldsymbol{u})}}\Big] = E\Big[|\varepsilon_0|\Big]\Big(E\Big[\sup_{\boldsymbol{u}\in U}\frac{\sigma_0^2}{w_0(\boldsymbol{u})}\Big]^{1/2}\Big) < \infty.$$

Since $\log^+$ moments of both $\sup_{\boldsymbol{u}\in U} y_k/\sqrt{w_k(\boldsymbol{u})}$ and y_{-j} are finite, and $|\rho_0| < 1$, Fact 5.1 implies

$$\sum_{0\le j<\infty} y_{-j}^2\rho_0^{j/q} < \infty \quad \text{and} \quad \sum_{1\le k<\infty}\sup_{\boldsymbol{u}\in U}\frac{|y_k|}{\sqrt{w_k(\boldsymbol{u})}}\rho_0^{k/q} < \infty, \quad \text{a.s.,}$$

thereby completing the proof of (5.6).

Next, we shall analyze $\|\widetilde{f}_n - f_n - g_n\|_2$, the second term of the upper bound in (5.4). By the Taylor expansion up to the second order,

$$\begin{aligned}
\widetilde{f}_n(x) - f_n(x) - g_n(x) &= \frac{1}{nh}\sum_{k=1}^{n}\Big\{K\Big(\frac{x-\widetilde{\varepsilon}_k}{h}\Big) - K\Big(\frac{x-\varepsilon_k}{h}\Big)\Big\} - g_n(x) \qquad (5.7)\\
&= \frac{1}{nh^2}\sum_{k=1}^{n}(\widetilde{\varepsilon}_k - \varepsilon_k)K'\Big(\frac{x-\varepsilon_k}{h}\Big) - g_n(x) \\
&\qquad + \frac{1}{2nh^3}\sum_{k=1}^{n}(\widetilde{\varepsilon}_k - \varepsilon_k)^2K''\Big(\frac{x-\xi_k}{h}\Big),
\end{aligned}$$

where $\xi_k = \varepsilon_k + c^*(\widetilde{\varepsilon}_k - \varepsilon_k)$, for some $0 < c^* < 1$.

We claim

$$\int \Big[\frac{1}{nh^3}\sum_{k=1}^n (\widetilde{\varepsilon}_k - \varepsilon_k)^2 K''\Big(\frac{x-\xi_k}{h}\Big)\Big]^2 dx = O_p\big(\frac{1}{n^2h^5}\big). \tag{5.8}$$

Let $\boldsymbol{\Delta}_n = \boldsymbol{\theta}_n - \boldsymbol{\theta}$, and write

$$(\widetilde{\varepsilon}_k - \varepsilon_k) \;=\; y_k\Big(\frac{1}{\sqrt{w_k(\boldsymbol{\theta}_n)}} - \frac{1}{\sqrt{w_k(\boldsymbol{\theta})}}\Big) = y_k\Big(\frac{1}{\sqrt{w_k(\boldsymbol{\Delta}_n + \boldsymbol{\theta})}} - \frac{1}{\sqrt{w_k(\boldsymbol{\theta})}}\Big)$$

Recall (5.2). The first and second order Taylor expansions of $1/\sqrt{w_k(\boldsymbol{\Delta}_n + \boldsymbol{\theta})}$ around $\boldsymbol{\theta}$ yield the following two equations, respectively.

$$(\widetilde{\varepsilon}_k - \varepsilon_k) \;=\; -\frac{\varepsilon_k}{2}\boldsymbol{\Delta}_n^T \boldsymbol{r}_k(\boldsymbol{\theta}_1^*) \tag{5.9}$$

$$\;=\; -\frac{\varepsilon_k}{2}\boldsymbol{\Delta}_n^T \boldsymbol{r}_k \;+\; \frac{3\varepsilon_k}{8}\Big(\boldsymbol{\Delta}_n^T \boldsymbol{r}_k(\boldsymbol{\theta}_2^*)\Big)^2 \;-\; \frac{\varepsilon_k}{4}\boldsymbol{\Delta}_n^T \boldsymbol{R}_k(\boldsymbol{\theta}_2^*)\boldsymbol{\Delta}_n \tag{5.10}$$

where $\boldsymbol{\theta}_1^* = \boldsymbol{\theta}_n + c_1^*\boldsymbol{\Delta}_n$, and $\boldsymbol{\theta}_2^* = \boldsymbol{\theta}_n + c_2^*\boldsymbol{\Delta}_n$, for some $0 < c_1^*,\, c_2^* < 1$.

By (5.9), the left hand side of (5.8) is equal to

$$\frac{1}{4n^4h^6}\int \Big[\sum_{k=1}^n \varepsilon_k^2\big(\sqrt{n}\boldsymbol{\Delta}_n^T \boldsymbol{r}_k(\boldsymbol{\theta}_1^*)\big)^2 K''\Big(\frac{x-\xi_k}{h}\Big)\Big]^2 dx \tag{5.11}$$

$$\leq \; \frac{1}{4n^2h^5}\Big[\frac{1}{n}\sum_{k=1}^n \varepsilon_k^2\big(\sqrt{n}\boldsymbol{\Delta}_n^T \boldsymbol{r}_k(\boldsymbol{\theta}_1^*)\big)^2\Big]^2 \int \Big(K''(z)\Big)^2 dz,$$

where the last inequality is obtained by applying (5.1) with $a_k = \varepsilon_k^2\big(\sqrt{n}\boldsymbol{\Delta}_n^T \boldsymbol{r}_k(\boldsymbol{\theta}_1^*)\big)^2$, $b_k = K''\big((x-\xi_k)/h\big)$, and by a change of variable in the integration. The above bound is $O_p(1/n^2h^5)$, by (2.5), (2.6), (2.7) and Fact 5.4, thereby proving (5.8).

Upon combining (5.7) with (5.8), we obtain

$$\|\widetilde{f}_n - f_n - g_n\|_2^2 \tag{5.12}$$

$$= \int \Big[\frac{1}{nh^2}\sum_{k=1}^n (\widetilde{\varepsilon}_k - \varepsilon_k) K'\Big(\frac{x-\varepsilon_k}{h}\Big) - g_n(x)\Big]^2 dx + O_p\Big(\frac{1}{n^2h^5}\Big).$$

By (5.10),

$$\frac{1}{nh^2}\sum_{k=1}^n (\widetilde{\varepsilon}_k - \varepsilon_k) K'\Big(\frac{x-\varepsilon_k}{h}\Big) \tag{5.13}$$

$$\;=\; -\frac{1}{2n^{3/2}h^2}\sum_{k=1}^n \Big(\sqrt{n}\boldsymbol{\Delta}_n^T \boldsymbol{r}_k\Big)\varepsilon_k K'\Big(\frac{x-\varepsilon_k}{h}\Big)$$

$$+\frac{1}{nh^2}\sum_{k=1}^n \Big(\boldsymbol{\Delta}_n^T \boldsymbol{r}_k(\boldsymbol{\theta}_2^*)\Big)^2 \;\frac{3\varepsilon_k}{8} K'\Big(\frac{x-\varepsilon_k}{h}\Big)$$

$$-\frac{1}{nh^2}\sum_{k=1}^n \boldsymbol{\Delta}_n^T \boldsymbol{R}_k(\boldsymbol{\theta}_2^*)\boldsymbol{\Delta}_n \;\frac{\varepsilon_k}{4} K'\Big(\frac{x-\varepsilon_k}{h}\Big).$$

We shall now show

$$\int \Big[\frac{1}{nh^2}\sum_{k=1}^{n}\varepsilon_k\Big(\boldsymbol{\Delta}_n^T\boldsymbol{r}_k(\boldsymbol{\theta}_2^*)\Big)^2K'\Big(\frac{x-\varepsilon_k}{h}\Big)\Big]^2dx = O_p\big(\frac{1}{n^2h^3}\big). \tag{5.14}$$

$$\int \Big[\frac{1}{nh^2}\sum_{k=1}^{n}\varepsilon_k\boldsymbol{\Delta}_n^T\boldsymbol{R}_k(\boldsymbol{\theta}_2^*)\boldsymbol{\Delta}_nK'\Big(\frac{x-\varepsilon_k}{h}\Big)\Big]^2dx = O_p\big(\frac{1}{n^2h^3}\big). \tag{5.15}$$

To prove (5.14), use (5.1) in a similar fashion as for (5.11) and a change of variable formula to obtain that the left hand side of (5.14) is bounded above by

$$\frac{1}{n^2h^3}\Big[\frac{1}{n}\sum_{k=1}^{n}|\varepsilon_k|\big(\sqrt{n}\boldsymbol{\Delta}_n^T\boldsymbol{r}_k(\boldsymbol{\theta}_2^*)\big)^2\Big]^2\quad\int\Big(K'(z)\Big)^2dz.$$

This bound in turn is $O_p(1/n^2h^3)$, by (2.5), (2.6), (2.7), and Fact 5.4. This proves (5.14). The proof of (5.15) is exactly similar.

Thus, upon combining (5.12) to (5.15), we obtain

$$\begin{aligned}&\|\widetilde{f}_n - f_n - g_n\|_2^2 \qquad\qquad (5.16)\\ &= \int\Big[-\frac{1}{2n^{3/2}h^2}\sum_{k=1}^{n}\Big(\sqrt{n}\boldsymbol{\Delta}_n^T\boldsymbol{r}_k\Big)\varepsilon_kK'\Big(\frac{x-\varepsilon_k}{h}\Big)-g_n(x)\Big]^2dx + O_p\Big(\frac{1}{n^2h^5}\Big).\end{aligned}$$

Let Z_n denote the first term in the right hand side above. To obtain its rate of convergenc, introduce

$$G_k(x) \;=\; \varepsilon_kK'\Big(\frac{x-\varepsilon_k}{h}\Big) - E\Big[\varepsilon_0K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big].$$

Also, write $\boldsymbol{\Delta}_n = (\Delta_{n,1},\ldots,\Delta_{n,p+q+1})$, and $\boldsymbol{r}_k = (r_{k,1}, r_{k,2},\cdots,r_{k,p+q+1})'$. Then,

$$-\sum_{k=1}^{n}\sqrt{n}\boldsymbol{\Delta}_n^T\boldsymbol{r}_k\varepsilon_kK'\Big(\frac{x-\varepsilon_k}{h}\Big) - g_n(x) = -\sqrt{n}\boldsymbol{\Delta}_n^T\sum_{k=1}^{n}\boldsymbol{r}_kG_k(x),$$

and by the Cauchy-Schwarz inequality,

$$\begin{aligned} Z_n &= \int\Big[\frac{1}{n^{3/2}h^2}\sqrt{n}\boldsymbol{\Delta}_n^T\sum_{k=1}^{n}\boldsymbol{r}_kG_k(x)\Big]^2dx\\ &= \frac{1}{n^3h^4}\int\Big[\sum_{j=1}^{p+q+1}\sqrt{n}\Delta_{n,j}\sum_{k=1}^{n}r_{k,j}G_k(x)\Big]^2dx\\ &\le \frac{1}{n^3h^4}\sum_{j=1}^{p+q+1}\big(\sqrt{n}\Delta_{n,j}\big)^2\sum_{j=1}^{p+q+1}\int\Big[\sum_{k=1}^{n}r_{k,j}G_k(x)\Big]^2dx.\end{aligned}$$

Fix a $1\le j\le p+q+1$. Let $\mathcal{F}_\ell$ denote the $\sigma-$algebra generated by the r.v.'s $\{\varepsilon_k, k\le\ell\}$. Then,

$$E\big(r_{k,j}G_k(x)\big|\mathcal{F}_{k-1}\big) = E\big(r_{k,j}\big)EG_0(x) = 0,\quad\forall\, k,\, x.$$

Hence,

$$\int E\Big[\sum_{k=1}^{n} r_{k,j}G_k(x)\Big]^2 dx = \int \sum_{k=1}^{n} E\Big[r_{k,j}G_k(x)\Big]^2 dx = nEr_{0,j}^2 \int E\big[G_0^2(x)\big]dx.$$

But, since G_0 is centered,

$$\begin{aligned}\int EG_0^2(x)dx &\le \int E\Big[\varepsilon_0 K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big]^2 dx = h\int\int_{-1}^{1}(x-zh)^2\big(K'(z)\big)^2 f(x-zh)dzdx \\ &\le 2h\,\|K'\|_\infty \int y^2 f(y)dy = O(h),\end{aligned}$$

because K' is bounded and supported in $[-1,1]$. We have $\int x^2 f(x)dx < \infty$ from the moment assumption (2.7). This together with Fact 5.4 proves that $Z_n = O_p(1/n^2h^3)$, which in turns, together with (5.16), (5.5) and (5.4) completes the proof of (3.1).

Next, we prove (3.2). With $\overline{\boldsymbol{S}}_n$ defined at (5.2), observe that

$$n\|g_n\|_2^2 = \Big[\frac{1}{2}\sqrt{n}\boldsymbol{\Delta}_n^T\overline{\boldsymbol{S}}_n\Big]^2 \int \Big[\frac{1}{h^2}E\Big\{\varepsilon_0 K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big\}\Big]^2 dx.$$

Let $\psi(x) := f(x) + x\dot{f}(x)$. We shall shortly prove

$$\int \Big[\frac{1}{h^2}E\Big\{\varepsilon_0 K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big\} - \psi(x)\Big]^2 dx \to 0. \tag{5.17}$$

Consequently, in view of (5.3),

$$n\|g_n\|_2^2 = \Big[\frac{1}{2}\sqrt{n}\boldsymbol{\Delta}_n^T E\boldsymbol{r}_0\Big]^2 \int \psi^2(x)dx + o(1), \quad \text{a.s.} \tag{5.18}$$

This result together with (2.5) completes the proof of (3.2).

To prove (5.17), recall K is a density on $[-1,1]$, vanishing at the end points. Use the change of variable formula, integration by parts and f being absolutely continuous to write

$$\begin{aligned}E\Big\{\frac{1}{h^2}\varepsilon_0 K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big\} &= \frac{1}{h}\int (x-zh)K'(z)f(x-zh)dz \\ &= \frac{1}{h}\int (x-zh)f(x-zh)dK(z) \\ &= \int K(z)\big\{f(x-zh) + (x-zh)\dot{f}(x-zh)\big\}dz.\end{aligned}$$

Hence,

$$\begin{aligned}&\int \Big[E\Big\{\frac{1}{h^2}\varepsilon_0 K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big\} - \psi(x)\Big]^2 dx \\ &= \int\Big[\int \big\{f(x-zh)-f(x)\big\}K(z)dz + x\big\{\dot{f}(x-zh)-\dot{f}(x)\big\}K(z)dz \\ &\qquad\qquad -h\int z\dot{f}(x-zh)K(z)dz\Big]^2 dx \\ &\le 4\big(B_1 + B_2 + h^2 B_3\big),\end{aligned}$$

where

$$B_1 := \int \Big[\int \{f(x-zh) - f(x)\} K(z) dz \Big]^2 dx,$$

$$B_2 := \int \Big[\int x\{\dot f(x-zh) - \dot f(x)\} K(z) dz \Big]^2 dx, \quad B_3 = \int \Big[\int z \dot f(x-zh) K(z) dz \Big]^2 dx.$$

By the Cauchy-Schwarz inequality and Fubini Theorem, and because K is a density,

$$\begin{aligned} B_1 &\le \int\int \{f(x-zh) - f(x)\}^2 dx K(z) dz = \int\int \Big\{ \int_{x-zh}^{x} \dot f(s) ds \Big\}^2 dx K(z) dz \\ &\le \int\int |z| h \int_{x-h|z|}^{x} \big(\dot f(s)\big)^2 ds dx K(z) dz \le h^2 \int z^2 K(z) dz \int \big(\dot f(s)\big)^2 ds = O(h^2), \end{aligned}$$

by assumption (2.9). Similarly, the same assumption implies

$$B_2 \le \int\int x^2 \{\dot f(x-zh) - \dot f(x)\}^2 dx K(z) dz \to 0.$$

Finally, in view of (2.9), $\int \big(\dot f(s)\big)^2 ds < \infty$, and

$$B_3 \le \int z^2 \int \big(\dot f(x-zh)\big)^2 dx K(z) dz = \int z^2 K(z) dz \int \big(\dot f(s)\big)^2 ds = O(1).$$

This completes the proof of (5.17). Claim (3.3) follows from (3.1) and (3.2), thereby completing the proof of Theorem 3.1.

Proof of Proposition 3.1. To begin with note that by the Cauchy-Schwarz inequality, (1.1), (3.1) and (2.10),

$$\begin{aligned} &\Big| n\sqrt{h} \int (\widehat f_n(x) - f_n(x) - g_n(x))(f_n(x) - E_0 f_n(x)\Big) dx \Big| \\ &\le \sqrt{n} \big\| \widehat f_n - f_n - g_n \big\|_2 \sqrt{nh} \big\| f_n - E_0 f_n \big\|_2 = O_p(n^{-1/2} h^{-5/2})\, O_p(1) = o_p(1). \end{aligned}$$

Hence,

$$2n\sqrt{h} \int \big(\widehat f_n - f_n\big)(x)\big(f_n - E_0 f_n\big)(x) dx = 2n\sqrt{h} \int g_n(x)\big(f_n - E_0 f_n\big)(x) dx + o_p(1). \quad (5.19)$$

Moreover,

$$\begin{aligned} &2n\sqrt{h} \int g_n(x)\big(f_n(x) - E_0 f_n(x)\big) dx \quad (5.20) \\ &= -\sqrt{n} \boldsymbol{\Delta}_n^T \overline{\boldsymbol{S}}_n \sqrt{nh} \int E\Big[\frac{\varepsilon_0}{h^2} K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big] \Big(f_n(x) - E_0 f_n(x)\Big) dx. \end{aligned}$$

Let $\psi_0(x) = f_0(x) + x \dot f_0(x)$. We claim

$$\sqrt{nh} \int \Big[\frac{1}{h^2} E_0\Big\{\varepsilon_0 K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big\} - \psi_0(x)\Big] \Big(f_n(x) - E_0 f_n(x)\Big) dx \to_p 0. \quad (5.21)$$

This follows from (5.17) and the fact $E_0\big(nh\big\|f_n - E_0 f_n\big\|_2^2\big) = O(1)$, because the square of the left hand side of (5.21) is bounded above by

$$\int \Big[\frac{1}{h^2} E_0\Big\{\varepsilon_0 K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big\} - \psi_0(x)\Big]^2 dx\, nh\big\|f_n - E_0 f_n\big\|_2^2.$$

Therefore, in view of (5.19) and (5.20),

$$2n\sqrt{h}\int \big(\widehat{f}_n - f_n\big)\big(f_n - E_0 f_n\big)dx = -\sqrt{n}\boldsymbol{\Delta}_n^T \overline{\boldsymbol{S}}_n \sqrt{nh}\int \psi_0(x)[f_n(x) - E_0 f_n(x)]dx + o_p(1).$$

Next, we shall show that

$$V_0 := \mathrm{Var}_0\Big(\sqrt{nh}\int \psi_0(x)[f_n(x) - E_0 f_n(x)]dx\Big) = o(1). \tag{5.22}$$

We have

$$V_0 \;=\; nh\int\int \psi_0(x)\psi_0(y)E_0\Big\{\big(f_n(x) - E_0 f_n(x)\big)\big(f_n(y) - E_0 f_n(y)\big)\Big\}dxdy.$$

But

$$\begin{aligned}
&E_0\Big\{\big(f_n(x) - E_0 f_n(x)\big)\big(f_n(y) - E_0 f_n(y)\big)\Big\}\\
&= \frac{1}{nh^2}\Big\{E_0\Big[K\Big(\frac{x-\varepsilon_0}{h}\Big)K\Big(\frac{y-\varepsilon_0}{h}\Big)\Big] - E_0 K\Big(\frac{y-\varepsilon_0}{h}\Big)E_0 K\Big(\frac{x-\varepsilon_0}{h}\Big)\Big\}\\
&= \frac{1}{nh}\int K\Big(\frac{x-y}{h} + w\Big)K(w)f_0(y - wh)dw\\
&\qquad\qquad -\frac{1}{n}\int K(t)f_0(x - th)dt\int K(s)f_0(y - sh)ds\\
&=: A_{n,h}(x,y) - B_{n,h}(x,y), \quad \text{say.}
\end{aligned}$$

Hence, one can write $V_0 = V_{01} - V_{02}$, where

$$\begin{aligned}
V_{01} &:= nh\int\int \psi_0(x)\psi_0(y)A_{n,h}(x,y)dxdy\\
&= \int\int\int \psi_0(x)\psi_0(y)K\Big(\frac{x-y}{h} + w\Big)K(w)f_0(y - wh)dwdxdy\\
&= h\int\int\int \psi_0(t - sh)\psi_0(t)K(s + w)K(w)f_0(t - wh)dwdsdt\\
&\le \|f_0\|_\infty\, h\int (K*K)(s)\int \psi_0(t - sh)\psi_0(t)dtds \to 0,
\end{aligned}$$

by $C(f_0)$. Similarly,

$$\begin{aligned}
V_{02} &:= nh\int\int \psi_0(x)\psi_0(y)B_{n,h}(x,y)dxdy\\
&= h\int\int\int\int |\psi_0(x)\psi_0(y)|K(t)f_0(x - th)K(s)f_0(y - sh)dtdsdxdy\\
&\le \|f_0\|_\infty^2\, h\int\int\int\int |\psi_0(x)\psi_0(y)|K(t)K(s)dtdsdxdy = \|f_0\|_\infty^2\, h\Big(\int |\psi_0(x)|dx\Big)^2 \to 0.
\end{aligned}$$

Therefore, in view of (5.3), (5.22) and (2.5),

$$2n\sqrt{h}\int\big(\widehat{f}_n - f_n\big)(x)\big(f_n - E_0 f_n\big)(x)dx = o_p(1).$$

This concludes the proof of Proposition 3.1.

Proof of Theorem 3.2. Let $\Delta := f - f_0$. We have

$$\begin{aligned}
&\sqrt{n}\Big(\widehat{T}_n - \int\Big(K_h * \Delta\Big)^2(x)dx\Big)\\
&= \sqrt{n}\int\Big(\widehat{f}_n(x) - f_n(x) + f_n(x) - E_0 f_n(x)\Big)^2 dx - \sqrt{n}\int\Big(K_h * \Delta\Big)^2(x)dx\\
&= \sqrt{n}\int\Big(\widehat{f}_n(x) - f_n(x)\Big)^2 dx\\
&\quad+\sqrt{n}\Big(\int\Big(f_n(x) - E_0 f_n(x)\Big)^2 dx - \int\Big(K_h * \Delta\Big)^2(x)dx\Big)\\
&\quad+2\sqrt{n}\int\big(\widehat{f}_n(x) - f_n(x)\big)\big(f_n(x) - E_0 f_n(x)\big)dx\\
&\quad+2\sqrt{n}\int\big(\widehat{f}_n(x) - f_n(x)\big)\big(Ef_n(x) - E_0 f_n(x)\big)dx
\end{aligned}$$

The first term is $o_p(1)$ by Theorem 3.1. The second term is exactly the left hand side of (1.2). The third term is $o_p(1)$ by Proposition 3.1.

Next, observe that

$$\begin{aligned}
&\Big|\sqrt{n}\int\big(\widehat{f}_n - f_n\big)(x)\big(Ef_n - E_0 f_n\big)(x)dx - \sqrt{n}\int g_n(x)(Ef_n - E_0 f_n)(x)dx\Big|\\
&\leq \sqrt{n}\Big\|\widehat{f}_n - f_n - g_n\Big\|_2\Big\|E_0 f_n - E_0 f_n\Big\|_2 = o_p(1),
\end{aligned}$$

by Theorem 3.1, and $C(f)$ and $C(f_0)$. Therefore,

$$\begin{aligned}
&2\sqrt{n}\int\big(\widehat{f}_n(x) - f_n(x)\big)\big(Ef_n(x) - E_0 f_n(x)\big)dx \qquad (5.23)\\
&= \sqrt{n}\int g_n(x)(Ef_n - E_0 f_n)(x)dx + o_p(1).
\end{aligned}$$

Note that

$$\begin{aligned}
&2n\sqrt{h}\int g_n\Big(Ef_n(x) - E_0 f_n(x)\Big)dx\\
&= -\sqrt{n}\boldsymbol{\Delta}_n^T\overline{\boldsymbol{S}}_n\sqrt{nh}\int E\Big[\frac{\varepsilon_0}{h^2}K'\Big(\frac{x-\varepsilon_0}{h}\Big)\Big]\Big(Ef_n(x) - E_0 f_n(x)\Big)dx.
\end{aligned}$$

Recall $\psi(x) = f(x) + x\dot{f}(x)$. Because of $C(f)$ and $C(f_0)$, and (5.17), one readily sees

$$\begin{aligned}
&\int\frac{1}{h^2}E\Big[\epsilon_0 K'\Big(\frac{x-\epsilon_0}{h}\Big)\Big]\Big(Ef_n(x) - E_0 f_n(x)\Big)dx\\
&\qquad-\int\psi(x)\big(Ef_n(x) - E_0 f_n(x)\big)dx = o_p(1).
\end{aligned}$$

But, since under $C(f), C(f_0)$, f and f_0 are bounded and $f(x)+x\dot{f}(x)$ is square integrable, cf. (2.9), the dominated convergence theorem implies,

$$\int \psi(x)[Ef_n(x) - E_0 f_n(x)]dx$$
$$= \int \psi(x) \int K(z)[f(x-zh) - f_0(x-zh)]dzdx \to \int \psi(x)[f(x) - f_0(x)]dx.$$

This, in view of (2.5) and (5.3), completes the proof of Theorem 3.2.

Acknowledgement. Authors are grateful to the two diligent and meticulous referees of the paper whose numerous comments helped to improve the presentation.

References

Bachmann D. and Dette H. (2005) A note on the Bickel-Rosenblatt test in autoregressive time series. *Statist. Probab. Lett.*, **74**, 221-234.

Berkes I., Horváth L., Kokoszka P. (2003) GARCH process: structure and estimation. *Bernoulli*, **9**, 201-227.

Bickel, P. and M. Rosenblatt (1973) On some global measures of the deviations of density function estimates. *Annals of Statistics*, **1**, 1071-1095.

Bougerol, P. and Picard, N. (1992a) Strict stationarity of generalized autoregressive process. *Ann. Probaab*, **20**, 1714-1730.

Bougerol, P. and Picard, N. (1992b) Stationarity of GARCH processes and of some non-negative time series. *J. Econometrics*, **52**, 115-117.

Cheng, F. (2008) Asymptotic properties in ARCH(p)-time series. *Journal of Nonparametric Statistics* **20**, 47-60.

Horváth L. and Zitikis, R. (2006) Testing goodness of fit based on densities of GARCH innovations. *Econometric Theory*, **22**, 457-482.

Koul, H.L. (2002). *Weighted empirical processes in dynamic nonlinear models.* Second edition of *Weighted empiricals and linear models* [Inst. Math. Statist., Hayward, CA, 1992. **Lecture Notes in Statistics, 166**. Springer-Verlag, New York, 2002.

Lee S. and Na S. (2002) On the BickelRosenblatt test for first-order autoregressive models. *Statist. Probab. Lett.* **56**, 23-35.

Lee, S.W. and B.E. Hansen. (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, **10** 29-52.

Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica*, **6**, 575-596.

Mimoto, N. (2008) Convergence in distribution for the sup-norm of a kernel density estimator for GARCH innovations *Statist. Probab. Lett.,* **78**, 915-923.

Nelson, D.B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, **6**, 318-334.