Statistics and Its Interface Volume 0 (2008) 1–14

Bootstrap tests for simple structures in nonparametric time series regression*

Jens-Peter Kreiss, Michael H. Neumann and Qiwei Yao

This paper concerns statistical tests for simple structures such as parametric models, lower order models and additivity in a general nonparametric autoregression setting. We propose to use a modified L_2-distance between the nonparametric estimator of regression function and its counterpart under null hypothesis as our test statistic which delimits the contribution from areas where data are sparse. The asymptotic properties of the test statistic are established, which indicates the test statistic is asymptotically equivalent to a quadratic form of innovations. A regression type resampling scheme (i.e. wild bootstrap) is adapted to estimate the distribution of this quadratic form. Further, we have shown that asymptotically this bootstrap distribution is indeed the distribution of the test statistics under null hypothesis. The proposed methodology has been illustrated by both simulation and application to German stock index data.

AMS 2000 subject classifications: Primary 62G10; secondary 62M10.
Keywords and phrases: Absolute regularity, Additive models, Autoregression, Kernel estimation, Local polynomial estimation, Lower order models, Nonparametric regression, Parametric models, Wild bootstrap.

1. INTRODUCTION

Testing on parametric structures is an important issue in nonparametric statistics. In the context of time series modeling, this problem has also been addressed by many authors. For example, Hjellvik and Tjøstheim (1995, 1996) proposed linearity tests based on nonparametric estimates of conditional means and conditional variances. Their tests are based on average L_2-distances between parametric and nonparametric estimators of mean (or conditional variance) functions. Hjellvik, Yao and Tjøstheim (1998) established the asymptotic theory of the tests. Further, simulation conducted in that paper clearly demonstrates that the approximation from the first order asymptotic theory is far too crude to be useful in practice unless the sample size is tremendously large. Following the lead of Hjellvik and Tjøstheim (1995, 1996), Hjellvik, Yao and Tjøstheim (1996, 1998) adopted a parametric bootstrap scheme to estimate the critical values of tests, which amounted to resampling estimated residuals from the *best* fitted linear autoregressive model. This bootstrap procedure was proposed in Kreiss (1988); see also Bühlmann (1997) and Kreiss (1997). Again by simulations, Hjellvik, Yao and Tjøstheim (1998) demonstrate that the bootstrap approximation for the distribution of the test statistic is much more accurate than a first-order asymptotic approximation. However, there has been no theoretical justification of using bootstrap method in this context. One goal of this paper is to fill in this gap.

In this paper, we propose statistical tests for simple structures such as parametric models, lower order models and additivity in a general setting of stochastic regression model which includes autoregression as a special case. Our test statistic can be viewed as a generalized form of L_2-distance between nonparametric regression and its counterpart under null hypothesis. The idea to use the L_2-distances as test statistics goes back to Härdle and Mammen (1993), where the regression is considered with independent observations. In fact, Härdle and Mammen considered various kinds of bootstrap methods and concluded that the wild bootstrap method was most relevant to regression type of problems. Our test statistic is an improved version of that used by Härdle and Mammen. The improvement is effectively due to the introduction of a weight function in the statistic, which is proportional to the squared marginal density of the regressor. This not only stabilizes the statistic against the so-called boundary effect in nonparametric regression, but also delimits the influence from the areas where data are sparse. Furthermore, it simplifies theoretical derivation considerably. Following Härdle and Mammen's suggestion, we also use wild bootstrap method. However different from Härdle and Mammen, we only use it to estimate the distribution of a quadratic form of innovations which has an uniform form for all the three types of null hypotheses considered in the paper. Indeed this quadratic form is asymptotically equivalent to the test statistics under the null hypotheses. This means that practically we bootstrap from a population which always reflects the null hypothesis concerned (Hall and Wilson (1991)). This resampling scheme is nonparametric, which retains conditional heteroscedasticity in the model. For further discussion on using regression types of resampling techniques in autoregression, we refer to Neumann and Kreiss (1998) and Franke, Kreiss and Mammen (2002).

*Partially supported by an EPSRC Research Grant.

The rest of paper is organized as follows. We present the bootstrap test and the three types of null hypotheses in Section 2. In Section 3, the finite sample properties of the proposed methodology will be demonstrated by simulation and later by the application to German stock index data. The asymptotic properties in Section 4 guarantee that the bootstrap distributions are asymptotically the same as the null hypothesis distributions of the test statistics. All technical proofs are relegated in the Appendix.

2. BOOTSTRAP TESTS

2.1 Model and hypotheses

Suppose that $\{\mathbf{X}_t, Y_t\}$ is a strictly stationary discrete-time stochastic process with $\mathbf{X}_t \in \mathbb{R}^d$ and $Y_t \in \mathbb{R}$. Given observations $\{(\mathbf{X}_t, Y_t) : 1 \le t \le T\}$, we are interested in testing whether the conditional expectation $m(\mathbf{x}) = \mathrm{E}\{Y_t | \mathbf{X}_t = \mathbf{x}\}$ is of some simple structure. We write

$$Y_t = m(\mathbf{X}_t) + \varepsilon_t, \quad t \ge 1, \tag{2.1}$$

where $\mathrm{E}\{\varepsilon_t | \mathcal{F}_t\} = 0$ for all t, and $\mathcal{F}_t$ is the σ-algebra generated by $\{(\mathbf{X}_s, Y_{s-1}), s = t, t-1, \ldots\}$. This setting includes the autoregressive model as a special case in which $\mathbf{X}_t$ consists of some lagged values of Y_t. We do not assume that the random variables $\varepsilon_t,\ t \ge 1$, are independent. This, in particular, allows us to include the conditional heteroscedasticity in the model (see an application in Subsection 3.2 below). In fact, our original motivation is to test whether an autoregressive function has some simple forms such as, for example, a given parametric representation.

In this paper, we consider three types of null hypotheses on $m(\cdot)$:

$$\begin{aligned} H_p &: \quad m(\cdot) \in \left\{ m_{\boldsymbol{\theta}}(\cdot) \mid \boldsymbol{\theta} \in \Theta \right\}, \\ H_0 &: \quad m(x_1, \ldots, x_d) = m_0(x_1), \\ H_a &: \quad m(x_1, \ldots, x_d) = m_1(x_1) + \ldots + m_d(x_d). \end{aligned}$$

As a simple example of H_p, we may think of testing for a linear regression, namely $m_{\boldsymbol{\theta}}(x_1, \ldots, x_d) = \sum_{i=1}^d \theta_i x_i$. Another example would be to test for a parametric threshold model (Tong (1990)). For applications in econometrics, it is interesting to test for a so-called ARCH-structure (Engle (1982)), i.e. to test the validity of the model

$$X_t = \sigma_{\boldsymbol{\theta}}(X_{t-1}, \ldots, X_{t-d}) e_t, \qquad \text{where } \sigma_{\boldsymbol{\theta}}(x_1, \ldots, x_d) = \sqrt{\theta_0 + \sum_{i=1}^d \theta_i x_i^2}.$$

In the above expression, it is assumed that $Ee_t = 0$ and $Ee_t^2 = 1$. This problem can be formulated as a special case of testing for H_p by writing $Y_t = X_t^2$. (See Subsection 3.2 below.) Although hypothesis H_0 specifies a one-dimensional model only, our approach can be applied to test for the hypothesis of a d_0-dimensional model for some $d_0 < d$. In view of the curse of dimensionality which makes nonparametric methods in high dimensions problematic, it is often appealing to assume, for example, the additivity in nonparametric modeling. To date, most work on additive modeling has focused on the estimation aspect, whereas little attention has been paid on testing the validity of the additivity. The method proposed in this paper provides a bootstrap test for this purpose.

All the tests concerned in this paper are omnibus in the sense that they are designed for the testing against the alternative (2.1) which is a very general d-dimensional regression model. Therefore they are typically less powerful than the tests constructed for the alternatives with explicitly specified structures. We refer to Fan and Jiang (2007) for an overview on function-based nonparametric tests with the structured alternative hypotheses.

2.2 The test statistic

Let $\widetilde{m}(\cdot)$ be a corresponding estimator of $m(\cdot)$ under the relevant null hypothesis, namely,

$$\widetilde{m}(x_1, \ldots, x_d) = \begin{cases} m_{\widehat{\boldsymbol{\theta}}}(x_1, \ldots, x_d) & \text{if } H_p \text{ holds,} \\ \widehat{m}_0(x_1) & \text{if } H_0 \text{ holds,} \\ \widehat{m}_1(x_1) + \ldots + \widehat{m}_d(x_d) & \text{if } H_a \text{ holds.} \end{cases} \tag{2.2}$$

We propose to use the test statistic

$$S_T = \int_{\mathbb{R}^d} \left(\frac{1}{T} \sum_{t=1}^T K_h(\mathbf{x} - \mathbf{X}_t) \{ Y_t - \widetilde{m}(\mathbf{X}_t) \} \right)^2 w(\mathbf{x})\, d\mathbf{x}, \tag{2.3}$$

where $K_h(\cdot) = h^{-d} K(\cdot/h)$, $K(\cdot)$ is a kernel function on $\mathbb{R}^d$, $h > 0$ is a bandwidth, and $w(\cdot)$ denotes a weight function.

The statistic defined above can be viewed as a modified version of the following statistic used by Härdle and Mammen (1993) for testing the hypothesis H_p based on independent observations

$$\int_{\mathbb{R}^d} \left(\frac{\sum_{t=1}^T K_h(\mathbf{x} - \mathbf{X}_t) \{ Y_t - m_{\widehat{\boldsymbol{\theta}}}(\mathbf{X}_t) \}}{\sum_{t=1}^T K_h(\mathbf{x} - \mathbf{X}_t)} \right)^2 w(\mathbf{x}) d\mathbf{x}. \tag{2.4}$$

Their basic idea is to use the average L_2-distance between a parametric estimator $m_{\widehat{\boldsymbol{\theta}}}(\cdot)$ and a nonparametric estimator

$$\widehat{m}_h(\cdot) = \sum_{t=1}^T Y_t K_h(\cdot - \mathbf{X}_t) \Big/ \sum_{t=1}^T K_h(\cdot - \mathbf{X}_t) \tag{2.5}$$

as a test statistic. To compensate the bias in nonparametric estimation under H_0, they smooth $m_{\widehat{\boldsymbol{\theta}}}(\cdot)$ as well. We omit the estimator of the stationary density $\pi(\cdot)$ of $\mathbf{X}_t$ in the denominator of the integrand in (2.3), which could be interpreted as that we add a factor $\pi^2(\cdot)$ into the weight function in (2.4). This means that we consider the difference of the two estimators only at the values of $\mathbf{x}$ within the support of $\pi(\cdot)$ and pay substantially less attention in areas where the data are sparse (see Subsection 3.2, especially Figure 3.5). Further, this modification not only simplifies the theoretical derivations, but also makes the statistic stable in practice – regardless of the choice of weight function $w(\cdot)$. In fact, we can choose $w(\cdot) \equiv 1$ for testing H_p. Note that Fan and Zhang (2004) Subsection 3.1 presents an interesting interpretation of the above bias correction via smoothing in terms of reparametrization. It also points out the link to the prewhitening technique of Press and Tukey (1956).

Our test statistics for the three different null hypotheses H_p, H_0 and H_a have a common representation S_T as given in (2.3). The respective estimators for the regression function $m(\cdot)$ under different hypotheses are building blocks in defining S_T (see (2.2)). We specify those estimators as follows.

For testing a parametric hypothesis H_p, we assume that $\widehat{\boldsymbol{\theta}}$ is a $\sqrt{T}$-consistent estimator of $\boldsymbol{\theta}_0$ (the true parameter) for which

$$(2.6)\quad m_{\widehat{\boldsymbol{\theta}}}(\cdot) - m_{\boldsymbol{\theta}_0}(\cdot) = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\tau \dot{m}_{\boldsymbol{\theta}_0}(\cdot) + O_P\left(\frac{||\cdot||^2}{\sqrt{T}\log T}\right),$$

where $\dot{m}_{\boldsymbol{\theta}}(\cdot)$ denotes the derivative of $m_{\boldsymbol{\theta}}(\cdot)$ with respect to $\boldsymbol{\theta}$, and $||\cdot||$ denotes the Euclidean norm.

For testing the one-dimensional nonparametric regression model H_0, we use a local polynomial estimator of order p, where $[p/2] > 5d/16$, ($[p/2]$ denotes the integer part of $p/2$), i.e. we estimate $m_0(x_1)$ by $\widehat{m}_g(x_1) = \widehat{a}$, where

$$(2.7)\quad (\widehat{a}, \widehat{b}_1, \ldots, \widehat{b}_p) = \arg\min_{a, b_1, \ldots, b_p} \sum_{t=1}^{T} \Big\{ Y_t - a - b_1(x_1 - X_{t,1}) - \cdots - b_p(x_1 - X_{t,1})^p \Big\}^2 W\left(\frac{x_1 - X_{t,1}}{g}\right),$$

W is a kernel function on $\mathbb{R}$, $g > 0$ is a bandwidth, and $X_{t,1}$ is the first component of $\mathbf{X}_t$. We use a local polynomial estimator with a sufficiently high order (i.e. $[p/2] > 5d/16$) rather than a conventional kernel (i.e. local constant) estimator in order to keep the bias in estimation of $m_0(\cdot)$ small enough in the first place. Note that the way of defining the statistic S_T involves the further smoothing on the estimator of $m_0(\cdot)$, which inevitably increases its bias further under H_0.

We use the so-called nonparametric integration estimators for the additive conditional mean function $m(\mathbf{x}) = m_1(x_1) + \cdots + m_d(x_d)$, which, as proved by Fan, Härdle and Mammen (1998) achieve the usual one-dimensional rate of nonparametric curve estimators. This indicates that obtained results for testing on a one-dimensional nonparametric hypothesis immediately carry over to the additive nonparametric case.

2.3 Bootstrapping

It is easy to see that

$$(2.8)\quad Th^{d/2}S_T = Th^{d/2}S'_T - \frac{2h^{d/2}}{T}\int \sum_{t=1}^{T} K_h(\mathbf{x} - \mathbf{X}_t)\varepsilon_t \times \sum_{s=1}^{T} K_h(\mathbf{x} - \mathbf{X}_s)\{\widetilde{m}(\mathbf{X}_s) - m(\mathbf{X}_s)\} w(\mathbf{x})d\mathbf{x} + \frac{h^{d/2}}{T}\int\left(\sum_{t=1}^{T} K_h(\mathbf{x} - \mathbf{X}_t)\{\widetilde{m}(\mathbf{X}_t) - m(\mathbf{X}_t)\}\right)^2 w(\mathbf{x})d\mathbf{x},$$

where

$$(2.9)\quad S'_T = \frac{1}{T^2}\int\left(\sum_{t=1}^{T} K_h(\mathbf{x} - \mathbf{X}_t)\varepsilon_t\right)^2 w(\mathbf{x})d\mathbf{x},$$

which is a quadratic form of the innovations $\{\varepsilon_t\}$ and is invariant under the three null hypotheses. Theorem 1 in Section 4 below shows that under the null hypotheses, $Th^{d/2}S_T$ is asymptotically normal, and more importantly its asymptotic distribution is the same as that of $Th^{d/2}S'_T$. (The dominating role played by the quadratic term was also observed by Härdle and Mammen (1993) for regression with independent observations.) This indicates that we may mimic the distribution of S_T by bootstrapping the quadratic form S'_T only. Note that the distribution of S'_T does not depend on whether the null hypothesis holds or not, although S_T does. Therefore, the derived bootstrap test automatically follows the first guideline set by Hall and Wilson (1991). Namely our bootstrap approximation to the null hypothesis distribution of S_T is always valid even the data $\{(Y_t, \mathbf{X}_t)\}$ were drawn from a population under which the null hypothesis does not holds. (See Figure 3.3 below for an illustration.) This ensures the reasonable power of the bootstrap test against the departure from the null hypothesis.

Härdle and Mammen (1993) studied three different bootstrap procedures and concluded that the wild bootstrap is the most pertinent method for testing the regression structure. Following their lead, we adopt a wild bootstrap scheme to estimate the distribution of (2.9). To this end, we define the bootstrap statistic

$$(2.10)\quad S^*_T = \frac{1}{T^2}\int\left(\sum_{t=1}^{T} K_h(\mathbf{x} - \mathbf{X}_t)\varepsilon^*_t\right)^2 w(\mathbf{x})d\mathbf{x},$$

where the bootstrap innovations $\varepsilon^*_1, \ldots, \varepsilon^*_T$ are conditionally independent given the observed data $\{(\mathbf{X}_t, Y_t) : 1 \le t \le T\}$, and

$$E^*\varepsilon^*_t = 0 \quad \text{and} \quad E^*(\varepsilon^*_t)^2 = \widehat{\varepsilon}_t^{\,2} = (Y_t - \widehat{m}_h(\mathbf{X}_t))^2,$$

where E^* denotes the expectation under bootstrap distribution (i.e. the conditional distribution given $\{(\mathbf{X}_t, Y_t) : 1 \le t \le T\}$), $\widehat{m}_h(\cdot)$ is defined as in (2.5). In practice, we can define $\varepsilon_t^* = \widehat{\varepsilon}_t \cdot \eta_t$, where $\{\eta_t\}$ is a sequence of i.i.d. random variables with zero mean and unit variance. We reject the null hypothesis if $S_T > t_\alpha^*$, where t_α^* is the upper α-point of the conditional distribution of S_T^*. The latter can be evaluated via repeated bootstrap samplings. In fact, the p-value of the test is the relative frequency of the event $\{S_T^* \ge S_T\}$ in the bootstrap replications. We have proved that this bootstrap test is asymptotically correct in the sense that its significance level converges to α as $T \to \infty$ (Corollary 1 in Section 4 below).

3. NUMERICAL PROPERTIES

In this section, we investigate the finite sample properties of the proposed method by both simulation and application with a real data set. As an illustration, we deal only with the parametric hypothesis H_p. We always use the kernel $K(u) = 3/4\ (1-u^2)\ I_{[-1,1]}(u)$ in our calculations, whereas the standard Gaussian kernel is also possible, and weight function $w(\cdot) \equiv 1$. We use the cross-validation to choose bandwidths for nonparametric regression estimation.

3.1 Simulations

We conduct simulations with five different models. It turns out that the bootstrap scheme provides fairly accurate approximations to the significance levels of the tests. The simulated power of tests are also reported. Finally, we demonstrate by example that the bootstrap approximation stays closely to the distribution of S_T', which is equal to the null hypothesis distribution of S_T asymptotically, even when S_T is calculated from the data generated from a nonlinear model.

We consider three linear autoregression models

(M1) $X_t = -0.9 \cdot X_{t-1} + \varepsilon_t,\ t = 1, \dots, T,$
(M2) $X_t = 0.9 \cdot X_{t-1} - 0.5 \cdot X_{t-2} + \varepsilon_t,\ t = 1, \dots, T,$
(M3) $X_t = 0.9 \cdot X_{t-1} - 0.5 \cdot X_{t-2} + 0.3 \cdot X_{t-3} + \varepsilon_t,\ t = 1, \dots, T,$

and two nonlinear autoregression models

(M4) $X_t = 0.9 \cdot \sin(X_{t-1}) + \varepsilon_t,\ t = 1, \dots, T,$
(M5) $X_t = -0.9 \cdot X_{t-1} + \sin(X_{t-2}) + \varepsilon_t,\ t = 1, \dots, T.$

We always assume that innovations in the above models are i.i.d.. Their distribution may be normal, double exponential (heavier tails), logistic or shifted exponential (in order to have zero mean). All the five models are stationary. We replicate simulation 500 times with sample size $T = 100, 200$ and 500 respectively. We replicate bootstrap sampling 500 times.

Tables 3.1 and 3.2 report the actual levels of the proposed bootstrap tests for all five models with different innovation distributions. For the first three models we test for linearity, while for model four and five we test for the parametric hypothesis $m(x) \in \{\theta\ \sin(x)\}$ and $m(x_1, x_2) \in \{\theta_1\ x_1\ +\ \theta_2\ \sin(x_2)\}$ respectively. It can be seen from Tables 3.1 and 3.2 that the actual levels of the proposed bootstrap tests are very stable around or below the nominal level α. Even when the distribution of innovations in model (M2) is exponential, which is strongly asymmetric, the proposed test tends to make the right decision. Note that it is not always trivial to separate nonlinearity from non-normality, and some classical test procedures would reject a linearity hypothesis for a linear model with strongly skewed innovations.

Table 3.1. Nominal level $\alpha = 0.05$

model	T=100	$\mathcal{L}(\varepsilon_1)$	T=200	$\mathcal{L}(\varepsilon_1)$	T=500	$\mathcal{L}(\varepsilon_1)$
M1	0.048	logistic	0.036	logistic	0.050	logistic
M2	0.066	logistic	0.048	logistic	0.036	logistic
M2	0.040	exponential	0.018	exponential	0.022	exponential
M3	0.066	normal	0.045	normal	0.030	normal
M4	0.052	double exp.	0.048	double exp.	0.046	double exp.
M4	0.026	exponential	0.028	exponential	0.024	exponential
M5	0.048	normal	0.034	normal	0.028	normal

Table 3.2. Nominal level $\alpha = 0.10$

model	T=100	$\mathcal{L}(\varepsilon_1)$	T=200	$\mathcal{L}(\varepsilon_1)$	T=500	$\mathcal{L}(\varepsilon_1)$
M1	0.074	logistic	0.055	logistic	0.086	logistic
M2	0.106	logistic	0.100	logistic	0.068	logistic
M2	0.078	exponential	0.044	exponential	0.062	exponential
M3	0.106	normal	0.080	normal	0.082	normal
M4	0.124	double exp.	0.084	double exp.	0.058	double exp.
M4	0.076	exponential	0.060	exponential	0.064	exponential
M5	0.096	normal	0.092	normal	0.066	normal

Table 3.3. Underlying model (M4), test on first order linear autoregression

level α	T=100	$\mathcal{L}(\varepsilon_1)$	T=200	$\mathcal{L}(\varepsilon_1)$	T=500	$\mathcal{L}(\varepsilon_1)$
0.05	0.540	double exp.	0.878	double exp.	1.000	double exp.
0.05	0.432	exponential	0.806	exponential	1.000	exponential
0.10	0.676	double exp.	0.950	double exp.	1.000	double exp.
0.10	0.614	exponential	0.914	exponential	1.000	exponential

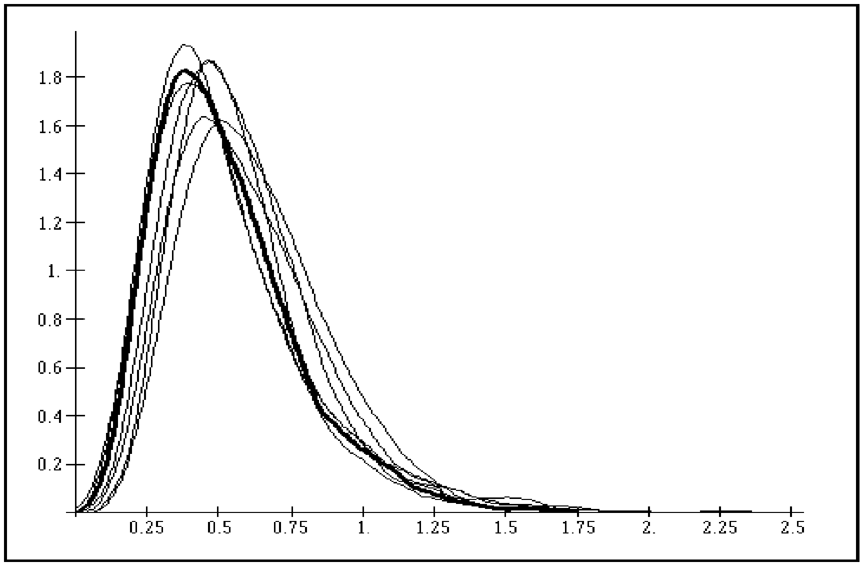


Figure 3.1. $T = 200$. Simulated density of $\mathcal{L}(Th^{d/2}S_T)$ (thick) and six bootstrap approximations (thin).

Table 3.4. Underlying model (M5), test on second order linear autoregression

level α	T=100	$\mathcal{L}(\varepsilon_1)$	T=200	$\mathcal{L}(\varepsilon_1)$	T=500	$\mathcal{L}(\varepsilon_1)$
0.05	0.992	normal	1.000	normal	1.000	normal
0.10	0.998	normal	1.000	normal	1.000	normal

Now we apply the bootstrap test for the linearity hypothesis for models (M4) and (M5). Tables 3.3 and 3.4 report the simulated values of the power function of the proposed bootstrap test. Comparatively, the proposed test is more powerful to detect the nonlinearity in model (M5) than that in (M4). The wider dynamic range of X_t in model (M5) than that in (M4) is certainly more helpful to identify the nonlinearity.

Finally, we look more closely at models (M2) and (M4). We plot the density functions of the test statistic $Th^{d/2}S_T$ (obtained from a simulation with 1000 Monte Carlo replications) and a couple of its bootstrap approximations in Figure 3.1 for model (M2) with $T = 200$ and in Figure 3.2 for model (M4) with $T = 100$. The null hypothesis concerned here is the correct parametric form specified in model (M2) and (M4), respectively. For testing the linearity for model (M4), we plot the distributions of $Th^{d/2}S_T$ and $Th^{d/2}S'_T$ together in Figure 3.3. Since now the null hypothesis no longer holds, the distributions of $Th^{d/2}S_T$ and $Th^{d/2}S'_T$ are quite different. The bootstrap approximations are always close to the null hypothesis distribution of $Th^{d/2}S_T$ whenever the underlying models reflect null hypothesis (Figures 3.1 and 3.2) or not (Figure 3.3).

One pillar of the consistency of our bootstrap proposal is that the more complicated test statistic S_T (cf. (2.3)) can be approximated by a simpler quadratic form S'_T (cf. (2.9)). For a proof of this approximation it is necessary that the bandwidth h fulfills the restrictions of assumption **A5**, i.e. especially converges to zero as sample size increases (cf. proof of Theorem 1 (i)). On the other hand it seems reasonable that the approximation of the bootstrap statistic S^*_T (cf. (2.10)) to the quadratic form S'_T (cf. (2.9)) is less dependent on the choice of the bandwidth h. For a larger h we even are more close to a parametric situation which indicates that the accuracy of the bootstrap approximation to the quadratic form S'_T tends to be better the larger the bandwidth h is chosen.

3.2 Application

We apply our test to the daily German stock index DAX (S_t) for the period January 2, 1990 — December 30,

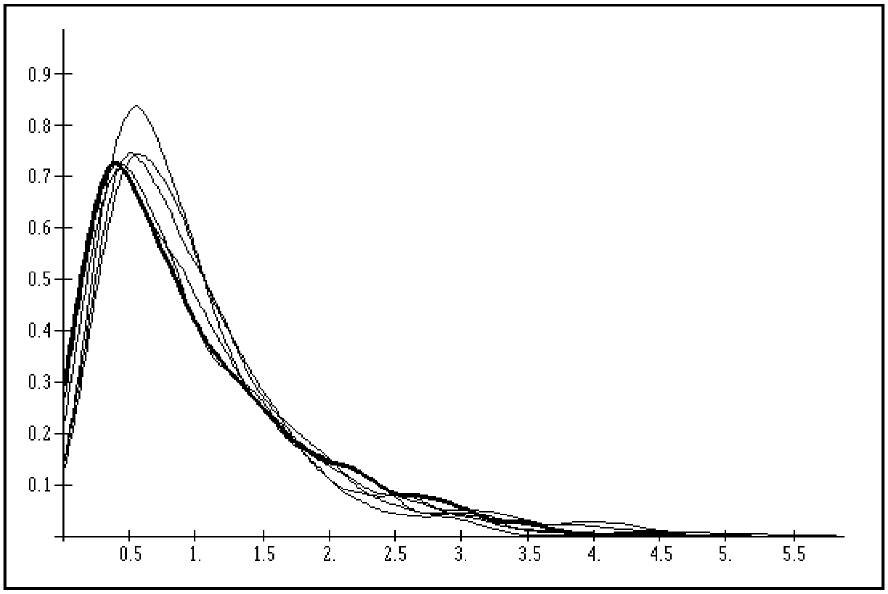


Figure 3.2. $T = 100$. Simulated density of $\mathcal{L}(Th^{d/2}S_T)$ (thick) and five bootstrap approximations (thin).

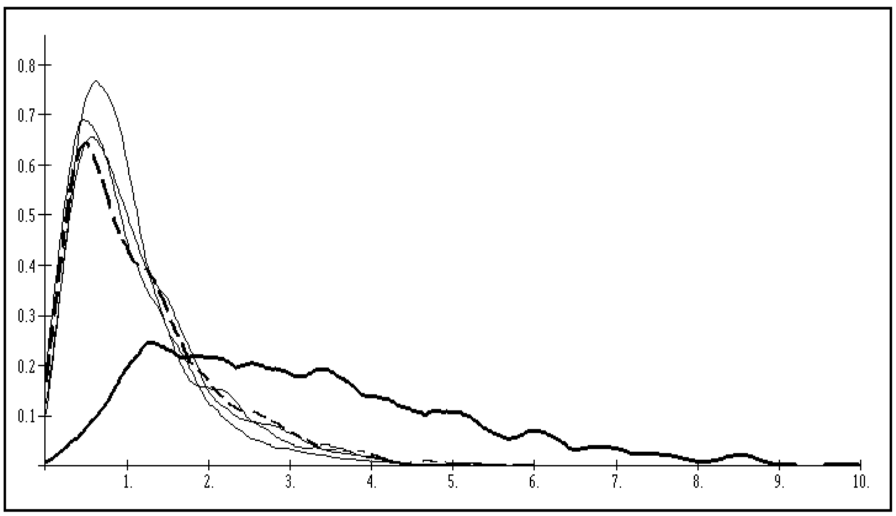


Figure 3.3. $T = 100$. Simulated density of test statistic (thick), simulated density of quadratic form (broken) and three bootstrap approximations (thin).

1992 plotted in Figure 3.4. It is of practical interest to test whether the first order ARCH-model is an appropriate parametric form for the so-called returns $R_t \equiv \log S_t - \log S_{t-1},\ t = 1, \ldots, T = 746$. The implied ARCH model is

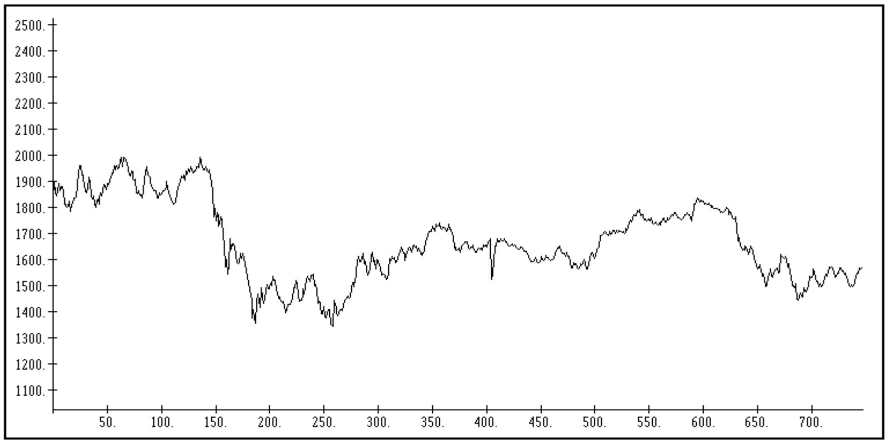
$$R_t = \sqrt{\alpha_0 + \alpha_1 R_{t-1}^2} \cdot e_t,$$

which can equivalently be expressed as,

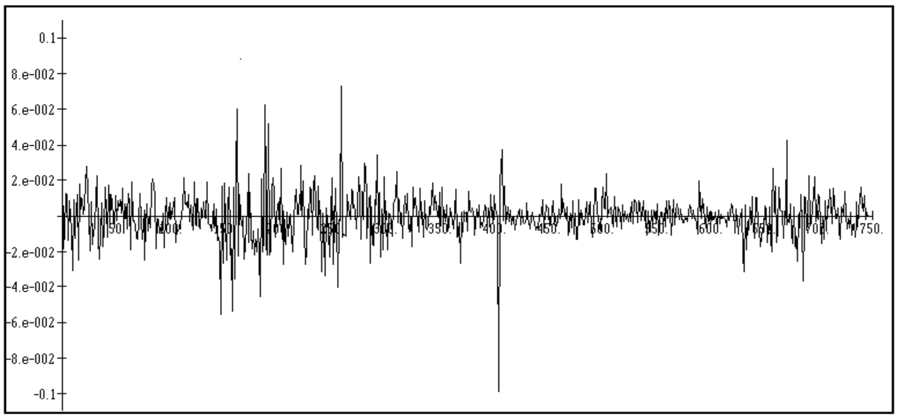
$$R_t^2 = \alpha_0 + \alpha_1 R_{t-1}^2 + \left(\alpha_0 + \alpha_1 R_{t-1}^2\right) \cdot (e_t^2 - 1),$$

where the innovations $\{e_t\}$ are assumed to be i.i.d. random variables with zero mean and unit variance. Our test statistic is based on the average L_2-distance of the parametric estimator $\widehat{\alpha}_0 + \widehat{\alpha}_1 x^2$ and the nonparametric estimator of $E\left[R_t^2 | R_{t-1} = x\right]$, the volatility function, over the interval $[-0.02, 0.02]$ (in which we find 91.4% of our data). Note that we are now dealing with a model with (conditional) heteroscedasticity.

In a first step we use the statistic (2.3) with weight function w equal to one. As is explained in Subsection 2.2 this means that we implicitly make use of an intrinsic weight



German Stock Index DAX (1990-1992)



Returns of German Stock Index DAX (1990-1992)

Figure 3.4.

function proportional to the square of the stationary density, i.e. we weight down regions where the observations are sparse. The value of the test statistic is $T\sqrt{h}S_T = 1.2 \cdot 10^{-6}$, where the bandwidth $h = 9.0 \cdot 10^{-3}$ has been selected by cross-validation. From 5000 bootstrap replications we obtain the bootstrap critical value $t^*_{0.10} = 3.1 \cdot 10^{-6}$ (at a level of 10 per cent), which implies that the first order parametric ARCH-model for the returns of the German stock index (1990–1992) cannot be rejected. The related p-value, obtained from the bootstrap simulations, reads 0.367. Figure 3.5 depicts both parametric estimator and nonparametric estimator of the regression function, together with the estimated marginal density function. It is clear that the ARCH structure is predominant when the density function is reasonably large and it fades away when we look at more extreme values of returns (which could be positive or negative). Note that the estimated density function takes very small values in the areas where the returns take extreme values. The intrinsic weight function in our test statistic weighs down the discrepancy of the two estimators in those areas automatically.

In a second step we don't use the simplified statistic (2.3) but instead the statistic (2.4) with two different weight functions w given below. This means that we don't want to weight down regions where the data are sparse as we did above. In order to be able to detect asymmetry of the conditional expectation of the squared returns we use the following two weight functions $w_1 = 1_{[-0.020,-0.005]}$ and $w_2 = 1_{[0.005,0.020]}$, i.e. we test, separately, for the same parametric ARCH-structure on a part of the negative and positive axes. Recall that we could not reject the ARCH-model at the level of 10% above. Now, at the stricter level of 5%, the bootstrap test applied to the test statistic (2.4) with weight function w_2 yields a clear cutoff rejection (bootstrap p-value obtained reads 0.023) while no rejection for the same test with w_1 of the parametric ARCH-structure is obtained (bootstrap p-value 0.205). The above analysis suggests that while ARCH(1) may provide a reasonable fitting in a middle area, it certainly fails to quantify the volatilities due to extreme negative returns. Our analysis also reinforces the common knowledge that volatility functions are not symmetric.

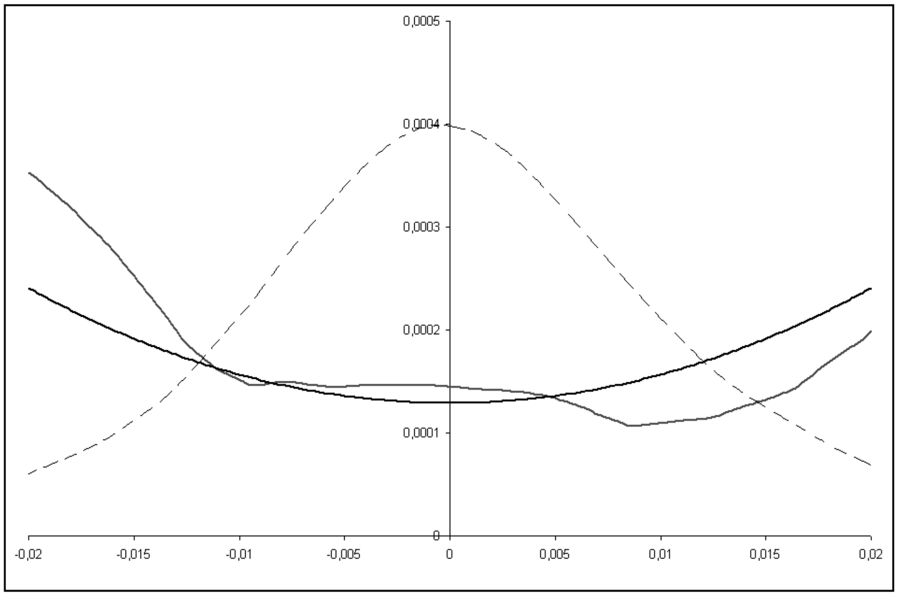


Figure 3.5: Nonparametric estimator $\widehat{m}_h(x)$, parametric estimator $\widehat{\alpha}_0 + \widehat{\alpha}_1 x^2$ of $E[R_t^2|R_{t-1} = x]$ and density of the underlying data (scaled by factor 10^{-5}) (broken).

4. ASYMPTOTIC PROPERTIES

To study the asymptotic properties of the proposed method, we need to introduce some regularity conditions as follows.

(A1) The process $\{(\mathbf{X}_i, Y_i)\}$ is absolutely regular, i.e.

$$\beta(j) \equiv \sup_{i\geq 1} E\left\{ \sup_{A\in\mathcal{F}_{i+j}^{\infty}} |P(A|\mathcal{F}_1^i) - P(A)| \right\} \to 0,$$

as $j \to \infty$,

where $\mathcal{F}_i^j$ is the σ-field generated by $\{(\mathbf{X}_k, Y_k) : k = i, \ldots, j\}$ $(j \geq i)$. Further it is assumed that $\{\beta(j)\}$ decay at a geometric rate.

(A2) $\mathbf{X}_t$ has a bounded density function $\pi(\cdot)$. Further, the joint density of distinct elements of $(\mathbf{X}_1, Y_1, \mathbf{X}_s, Y_s, \mathbf{X}_t, Y_t)$ $(t > s > 1)$ is continuous and bounded by a constant independent of s and t.

(A3) $E\{\varepsilon_t|\mathbf{X}_t, \mathcal{F}_1^{t-1}\} = 0$ for all t, $\sigma^2(\mathbf{x}) = \mathrm{Var}\{Y_t|\mathbf{X}_t = \mathbf{x}\} = E[\varepsilon_t^2|\mathbf{X}_t = \mathbf{x}]$ is continuous, and $\mathrm{E}\{[m(\mathbf{X}_t)]^{16} + Y_t^{16}\} < \infty$.

(A4) K is a product kernel, i.e. $K(x) = \prod_{i=1}^d W(x_i)$, and W is a symmetric density function with a bounded support in $\mathbb{R}$, and $|W(x_1) - W(x_2)| \leq c|x_1 - x_2|$ for all x_1 and x_2 in its support.

(A5) $h \in [aT^{-\frac{1}{d+4}}/\log T, bT^{-\frac{1}{d+4}}\log T]$, where $0 < a < b < \infty$ are some constants.

(A6) This assumption differs for the three different null hypotheses.

- For testing the hypothesis H_p, it is assumed that $E_{\boldsymbol{\theta}_0}||\dot{m}_{\boldsymbol{\theta}_0}(\mathbf{X}_1)||^2 < \infty$, and $w(\cdot) \equiv 1$.
- For testing H_0, it is assumed that $m_0(\cdot)$ is $(p+1)$-times differentiable with a bounded $(p+1)$-th order derivative, and $g \in [aT^{-\frac{1}{4[p/2]+5}}/\log T, bT^{-\frac{1}{4[p/2]+5}}\log T]$, where $[p/2]$ denotes the integer part of $p/2$ and $0 < a < b < \infty$ are some constants. Further, we assume that $[p/2] > 5d/16$. The weight function $w(\cdot)$ has a compact support contained in the support of $\pi(.)$.
Further, for any $M < \infty$ and arbitrary compact subset B contained in the support of $X_{t,1}$, there exists a constant $C_{M,B} < \infty$ such that
$$\sup_{x\in B}\left\{ E\left(|\varepsilon_t|^M \middle| X_{t,1} = x\right)\right\} \leq C_{M,B} \quad \text{for all } t.$$
- For testing H_a, smoothness conditions on $m_k(\cdot)$ $(1 \leq k \leq d)$ and suitable assumptions on the bandwidth are assumed in order to ensure that all the estimators for $\{m_k(\cdot)\}$ achieve the one-dimensional convergence rate and the uniform convergence over compact sets contained in the support of $\pi(\cdot)$.

Some remarks are now in order.

Remark 1. We impose the boundedness on the support of $W(\cdot)$ for brevity of the proofs; it may be removed at the cost of lengthier proofs. In particular, the Gaussian kernel is allowed. The assumption of the convergence rate of $\beta(j)$ is also imposed for technical convenience.

Remark 2. We assume all the bandwidths taking values around their optimal orders (with *symmetric* kernels) in the sense which minimize the risks in estimation of regression functions. (For practical implementation we recommend to use data-driven bandwidths such as cross-validation which achieve these orders.) Lepski and Spokoiny (1999) showed that the bandwidths which provide the most powerful tests against *local* alternatives are of slightly different orders. To achieve the best power, they proposed to use the supremum of a family of statistics instead of just one single statistic. They adopted a slightly conservative rule to determine the critical value based on Bonferroni's inequality. See also the multi-frequency test of Fan (1996) and the multi-scale test of Fan, Zhang and Zhang (2001). We do not opt for those approaches simply to keep our method simple and easy to implement.

Remark 3. The theoretical results presented in this paper are proved for nonrandom bandwidths. It is conceivable that they should also hold for some data-driven bandwidths, for which it remains to be proved that the difference between the test statistics based on two types of bandwidths are negligible. Neumann (1995) proved such a result in the context of confidence intervals of regression function.

Remark 4. In testing H_0, we need to use p-th order local polynomial estimator for $m_0(\cdot)$ with $[p/2] > 5d/16$, which always favors an even value of p if we wish to keep p as small as possible. For example, we have to use at least local quadratic estimation in order to test whether the model is one-dimensional against a two- or three-dimensional alternative.

Remark 5. Concerning suitable assumptions in order to ensure (A6) for testing H_a we refer to Yang, Härdle and Nielsen (1999). Also see Fan, Härdle and Mammen (1998).

Theorem 1. *Suppose that one of the null hypotheses H_p, H_0 or H_a holds, and that the statistic S_T given in (2.3) is defined in terms of one of the estimators specified in (2.6)–(2.8) according to the null hypothesis concerned. We also suppose that assumptions (A1)–(A6) hold. Then, as $T \to \infty$,*

(i) $S_T = S_T' + o_p(T^{-1}h^{-d/2})$, where S_T' is defined as in (2.9).

(ii) $(Th^{d/2})\{S_T' - E(S_T')\} \xrightarrow{d} N(0, V)$ as $T \to \infty$, where

$$E(S_T') = \frac{1}{Th^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K^2(\mathbf{u})w(\mathbf{x}+h\mathbf{u})\pi(\mathbf{x})\sigma^2(\mathbf{x})\,d\mathbf{x}\,d\mathbf{u},$$

$$V = 2\int_{\mathbb{R}^d}\sigma^4(\mathbf{x})\pi^2(\mathbf{x})w(\mathbf{x})\,d\mathbf{x} \times \int_{\mathbb{R}^{3d}} K(\mathbf{u})K(\mathbf{v})K(\mathbf{u}-\mathbf{z})K(\mathbf{v}-\mathbf{z})\,d\mathbf{u}\,d\mathbf{v}\,d\mathbf{z}.$$

Theorem 2. *Assume that the conditions of Theorem 1 hold. For the bootstrap statistic S_T^* defined as in (2.10), we have that as $T \to \infty$,*

$$Th^{d/2}[S_T^* - E^*(S_T^*)] \xrightarrow{d} N(0, V)$$

conditionally on $\{(\mathbf{X}_t, Y_t),\ 1 \le t \le T\}$, where V is the same as given in Theorem 1, and $Th^{d/2}|E(S_T') - E^*(S_T^*)| \to 0$ in probability.

Corollary 1. *Assume that the conditions of Theorem 1 hold. Let t_α^* be the upper α-point of the conditional distribution of S_T^* given $\{(\mathbf{X}_t, Y_t),\ 1 \le t \le T\}$ and $\alpha \in (0,1)$. Then as $T \to \infty$, $P\{S_T > t_\alpha^*\} \to \alpha$ under the corresponding null hypothesis.*

The above corollary follows immediately from Theorems 1 and 2. The proofs of Theorems 1 and 2 are given in the Appendix.

APPENDIX: PROOFS

A.1 Proof of Theorem 1

We first prove Theorem 1 (ii). Then we present the proof of Theorem 1 (i) for the case of testing H_0 only, since it is technically more involved than the case of testing H_p. We always use δ to denote an arbitrarily small positive constant.

Proof of Theorem 1 (ii). It is easy to see that

$$\begin{aligned} Th^{d/2}S_T' &= \frac{1}{Th^{3d/2}}\sum_{t=1}^{T}\int K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h}\right)w(\mathbf{x})\,d\mathbf{x}\ \varepsilon_t^2 \\ &\quad + \frac{2}{Th^{3d/2}}\sum_{1\le t<s\le T}\int K\left(\frac{\mathbf{x}-\mathbf{X}_s}{h}\right) \\ &\quad \times K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h}\right)w(\mathbf{x})\,d\mathbf{x}\ \varepsilon_t\varepsilon_s. \end{aligned}$$

By the Ergodic Theorem, the first term on the right-hand side of the above expression is equal to

$$E\left\{h^{-3d/2}\int K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h}\right)w(\mathbf{x})\sigma^2(\mathbf{X}_t)\,d\mathbf{x}\right\} + O_P(1/\sqrt{Th^d}),$$

where

$$\begin{aligned} & E\left\{h^{-3d/2}\int K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h}\right)w(\mathbf{x})\sigma^2(\mathbf{X}_t)\,d\mathbf{x}\right\} \\ &= h^{-d/2}\int\int K^2(\mathbf{u})w(\mathbf{x}+h\mathbf{u})\pi(\mathbf{x})\sigma^2(\mathbf{x})\,d\mathbf{x}\,d\mathbf{u}. \end{aligned}$$

Assumption (A3) ensures that the second term has mean 0. By Theorem A of Hjellvik, Yao and Tjøstheim (1996), this term is asymptotically normal with mean 0 and asymptotic variance

$$\begin{aligned} & 2h^{-3d}\int \varepsilon_t^2\varepsilon_s^2\left\{\int K\left(\frac{\mathbf{z}-\mathbf{X}_t}{h}\right)K\left(\frac{\mathbf{z}-\mathbf{X}_s}{h}\right)d\mathbf{z}\right\}^2 \\ & dP(\mathbf{X}_t, Y_t)dP(\mathbf{X}_s, Y_s) \\ &= 2h^{-3d}\int\sigma^2(\mathbf{u_1})\sigma^2(\mathbf{u_2})\pi(\mathbf{u_1})\pi(\mathbf{u_2})K\left(\frac{\mathbf{u_1}-\mathbf{z_1}}{h}\right)\cdot \\ & K\left(\frac{\mathbf{u_2}-\mathbf{z_1}}{h}\right)K\left(\frac{\mathbf{u_1}-\mathbf{z_2}}{h}\right)\cdot \\ & K\left(\frac{\mathbf{u_2}-\mathbf{z_2}}{h}\right)d\mathbf{u_1}d\mathbf{u_2}d\mathbf{z_1}d\mathbf{z_2} \to V. \end{aligned}$$

□

To prove Theorem 1 (i) (for the case of H_0), we introduce two lemmas first. Recall that $\widehat{m}_g(\cdot)$ is the local polynomial estimator of $m_0(\cdot)$ with bandwidth g (see (2.7)). We write

$$\mathbf{X}_t = (X_{t,1}, \ldots, X_{t,d})^\tau, \qquad \mathbf{x} = (x_1, \ldots, x_d)^\tau.$$

Let B denote any compact subset contained in the support of $X_{t,1}$. It follows from Propositions 2.1 and 2.2 of Neumann and Kreiss (1998) that uniformly in $x_1 \in B$,

$$\begin{aligned}(4.11)\qquad & \widehat{m}_g(x_1) - m_0(x_1) \\ &= \sum_{t=1}^T \bar{w}_g(x_1, X_{t,1})\varepsilon_t + b_\infty(x_1) \\ &\quad + O_P\left(\frac{(\log T)^{3/2}}{Tg} + \frac{g^{p+1}\log T}{\sqrt{Tg}}\right).\end{aligned}$$

The last term does not depend on x_1 and b_∞ denotes a non-random function with

$$(4.12)\qquad \sup_{x_1 \in B} |b_\infty(x_1)| = O(g^{p+1}),$$

and the weights $\bar{w}_g(x_1, X_{t,1})$ are given as

(4.13)

$$\bar{w}_g(x_1, X_{t,1}) = \sum_{k=0}^p d_k^{(\infty)}(x_1) W\left(\frac{x_1 - X_{t,1}}{g}\right)\left(\frac{x_1 - X_{t,1}}{g}\right)^k.$$

$d_k^{(\infty)}(x_1)$ denotes the $(1, k+1)$-th element of the inverse of a $(p+1) \times (p+1)$ matrix with $TE\Big\{K\left(\frac{x_1 - X_{t,1}}{g}\right) \times \left(\frac{x_1 - X_{t,1}}{g}\right)^{i+j-2}\Big\}$ as its (i,j)-th element. The minimal eigenvalue of this matrix is of order Tg, which immediately implies $d_k^{(\infty)}(x_1) = O(1/(Tg))$.

Lemma 1. *Suppose that assumptions (A1)–(A6) hold. Under hypothesis H_0,*

$$(4.14)\qquad \sup_{x_1 \in B}\left|\sum_{t=1}^T \bar{w}_g(x_1, X_{t,1})\varepsilon_t\right| = O_P\left(\log T/\sqrt{Tg}\right),$$

$$\sup_{x_1 \in B}\left|\sum_{t=1}^T \dot{\bar{w}}_g(x_1, X_{t,1})\varepsilon_t\right| = O_P\left(\log T/\sqrt{Tg^3}\right),$$

$$\sup_{x_1 \in B}\left|\sum_{t=1}^T \ddot{\bar{w}}_g(x_1, X_{t,1})\varepsilon_t\right| = O_P\left(\log T/\sqrt{Tg^5}\right),$$

where $\dot{\bar{w}}_g(x_1, .)$ and $\ddot{\bar{w}}_g(x_1, .)$ denote the first and second order derivative with respect to x_1.

Proof of Lemma 1. We prove (4.14) only, the two other equations can be proved in a similar way.

Without loss of the generality we assume $B = [a, b]$. First we divide $[a, b]$ into $I \equiv I(T) = O(T^2)$ small intervals with the same length. Let $b_0 = a < b_1 \ldots < b_I = b$ be the endpoints of the intervals and $B_i = [b_{i-1}, b_i]$. It is obvious that

$$\begin{aligned}(4.15)\quad & \sup_{x_1 \in B}\left|\sum_{t=1}^T \bar{w}_g(x_1, X_{t,1})\varepsilon_t\right| \\ &\le \max_{1\le i\le I}\sup_{x_1 \in B_i}\left|\sum_{t=1}^T \{\bar{w}_g(x_1, X_{t,1}) - \bar{w}_g(b_i, X_{t,1})\}\varepsilon_t\right| \\ &\quad + \max_{1\le i\le I}\left|\sum_{t=1}^T \bar{w}_g(b_i, X_{t,1})\varepsilon_t\right|.\end{aligned}$$

Since $W(\cdot)$ is bounded and has a compact support, it follows from (4.13) that $|\dot{\bar{w}}_g(x_1, X_{t,1})| = O_P(T^{-1}g^{-2})$ holds uniformly in $x_1 \in B$ and $t = 1, \ldots, T$. Therefore,

$$\begin{aligned}(4.16)\qquad & \max_{1\le i\le I}\sup_{x_1 \in B_i}\left|\sum_{t=1}^T \{\bar{w}_g(x_1, X_{t,1}) - \bar{w}_g(b_i, X_{t,1})\}\varepsilon_t\right| \\ &\le \max_{1\le i\le I}\sup_{x_1 \in B_i}\frac{1}{Tg^2}\sum_{t=1}^T |\varepsilon_t||x_1 - b_i| \\ &= O_P\left(\frac{1}{g^2 I}\right) = O_P\left(\frac{\log T}{\sqrt{Tg}}\right).\end{aligned}$$

Now we apply Lemma 2.1 (ii) of Neumann and Kreiss (1998) to the second summand on the right hand side of (4.15). Because of (A3) we have that $Tg\ \bar{w}_g(b_i, X_{t,1})\varepsilon_t$ satisfies the assumptions of that lemma. Since

$$\operatorname{Var}(Tg\bar{w}_g(b_i, X_{t,1})\varepsilon_t) = O(g)$$

we obtain for some $C > 0$

$$P\left\{\left|\sum_{t=1}^T \bar{w}_g(b_i, X_{t,1})\varepsilon_t\right| > C\frac{\log T}{\sqrt{Tg}}\right\} = O(T^{-\lambda}),$$

where λ denotes an arbitrarily large constant. Consequently,

$$\begin{aligned} & P\left\{\max_{1\le i\le I}\left|\sum_{t=1}^T \bar{w}_g(b_i, X_{t,1})\varepsilon_t\right| \ge \frac{C\log T}{\sqrt{Tg}}\right\} \\ &\le \sum_{i=1}^I P\left\{\left|\sum_{t=1}^T \bar{w}_g(b_i, X_{t,1})\varepsilon_t\right| \ge \frac{C\log T}{\sqrt{Tg}}\right\} = o(1).\end{aligned}$$

Combining this with (4.16) and (4.15), we have completed the proof of (4.14). □

Lemma 2. *Suppose that assumptions (A1)–(A6) hold. Under hypothesis H_0,*

$$\sup_{x_1 \in B}|\widehat{m}_g(x_1) - m_0(x_1)| = O_P\left(\log T/\sqrt{Tg} + g^{p+1}\right).$$

Lemma 2 follows immediately from Lemma 1 and (4.11).

Proof of Theorem 1 (i). We decompose $Th^{d/2}S_T$ as in (2.8). The first term on the right-hand side of (2.8) is $Th^{d/2}S_T'$.

We denote the last two terms by $-2R_{T,1}$ and $R_{T,2}$. Theorem 1 (i) follows from (a) $R_{T,1} \to 0$ and (b) $R_{T,2} \to 0$ in probability. We establish (a) and (b) in the sequel.

Substituting $\{\widehat{m}_g(\cdot) - m_0(\cdot)\}$ by the right-hand side of (4.11), we have that

$$|R_{T,1} - R'_{T,2} - R'_{T,3}| \leq R'_{T,1}, \tag{4.17}$$

where

$$\begin{aligned}
R'_{T,1} &= \frac{h^{d/2}}{T}\int \left|\sum_{t=1}^{T} K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t\right| \sum_{s=1}^{T} K_h(\mathbf{x}-\mathbf{X}_s)w(\mathbf{x})d\mathbf{x} \\
&\quad \times O_P\left(\frac{(\log T)^{3/2}}{Tg} + \frac{\log T g^{p+1}}{\sqrt{Tg}}\right), \\
R'_{T,2} &= \frac{h^{d/2}}{T}\int \sum_{t=1}^{T} K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t \\
&\quad \times \sum_{s=1}^{T} K_h(\mathbf{x}-\mathbf{X}_s)b_\infty(X_{s,1})w(\mathbf{x})d\mathbf{x}, \\
R'_{T,3} &= \frac{h^{d/2}}{T}\int \sum_{t=1}^{T} K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t \sum_{s=1}^{T} K_h(\mathbf{x}-\mathbf{X}_s) \\
&\quad \times \sum_{k=1}^{T} \bar{w}_g(X_{s,1}, X_{k,1})\varepsilon_k w(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

We split $R'_{T,2}$ into the following two terms

$$\begin{aligned}
&\frac{h^{d/2}}{T}\int \sum_{t=1}^{T} K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t \sum_{s=1}^{T}\{K_h(\mathbf{x}-\mathbf{X}_s)b_\infty(X_{s,1}) \\
&\quad - E[K_h(\mathbf{x}-\mathbf{X}_s)b_\infty(X_{s,1})]\}w(\mathbf{x})d\mathbf{x},
\end{aligned} \tag{4.18}$$

$$h^{d/2}\int \sum_{t=1}^{T} K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t E\{K_h(\mathbf{x}-\mathbf{X}_1)b_\infty(X_{1,1})\}w(\mathbf{x})d\mathbf{x}. \tag{4.19}$$

By Cauchy-Schwarz inequality, the expectation of the absolute value of (4.18) is bounded by

$$\begin{aligned}
&\frac{h^{d/2}}{T}\int \left\{E\left(\sum_{t=1}^{T} K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t\right)^2 \right. \\
&\quad \times E\left(\sum_{s=1}^{T}\{K_h(\mathbf{x}-\mathbf{X}_s)b_\infty(X_{s,1})\right. \\
&\quad \left.\left. - E[K_h(\mathbf{x}-\mathbf{X}_s)b_\infty(X_{s,1})]\}\right)^2\right\}^{1/2} w(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

Assumption (A3) implies that

$$E\left(\sum_{t=1}^{T} K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t\right)^2 = O(Th^{-d}).$$

Recall that the absolute regularity with geometrically decaying mixing coefficients implies strong mixing with mixing coefficients decaying at the same rate. Applying the covariance inequality for strong mixing processes (Corollary 1.1, Bosq (1996)), we have that

$$\begin{aligned}
&E\left(\sum_{s=1}^{T}\{K_h(\mathbf{x}-\mathbf{X}_s)b_\infty(X_{s,1}) - E[K_h(\mathbf{x}-\mathbf{X}_s)b_\infty(X_{s,1})]\}\right)^2 \\
&\leq O(T)\cdot\left(E\left|K_h(\mathbf{x}-\mathbf{X}_s)b_\infty(X_{s,1})\right|^{2+\delta}\right)^{2/2+\delta} \\
&= O(Th^{-d(1+\delta)}g^{2(p+1)}).
\end{aligned}$$

The last equality uses the fact (4.12). Hence, (4.18) is of the order

$$\begin{aligned}
&O_P\left(T^{-1}h^{d/2}\left\{Th^{-d}Th^{-d(1+\delta)}g^{2(p+1)}\right\}^{1/2}\right) \\
&= O_P\left(g^{p+1}h^{-d(1+\delta)/2}\right) = o_P(1).
\end{aligned}$$

The expectation of the square of (4.19) is equal to

$$\begin{aligned}
&h^d\sum_{t=1}^{T} E\left\{\int K_h(\mathbf{x}-\mathbf{X}_t)E[K_h(\mathbf{x}-\mathbf{X}_1)b_\infty(X_{1,1})]\sigma^2(\mathbf{x})w(\mathbf{x})d\mathbf{x}\right\}^2 \\
&= Th^dE\left\{\int K(\mathbf{u}-\mathbf{X}_t/h)\right. \\
&\quad \left.\times E[K_h(h\mathbf{u}-\mathbf{X}_1)b_\infty(X_{1,1})]\sigma^2(h\mathbf{u})w(h\mathbf{u})\,d\mathbf{u}\right\}^2 \\
&= O(Th^dg^{2(p+1)}) \to 0.
\end{aligned}$$

To obtain the last equality we make use of (4.12). We have proved that both (4.18) and (4.19) converge to 0 in first or second moment. Consequently, $R'_{T,2} \to 0$ in probability. In a similar way, it can be proved that $R'_{T,1} \to 0$ in probability.

To deal with $R'_{T,3}$, we first make a Taylor expansion

$$\begin{aligned}
\bar{w}_g(X_{s,1}, X_{k,1}) &= \bar{w}_g(x_1, X_{k,1}) + \dot{\bar{w}}_g(x_1, X_{k,1})(X_{s,1}-x_1) \\
&\quad + \frac{1}{2}\ddot{\bar{w}}_g(\widetilde{X}_{s,1}, X_{k,1})(X_{s,1}-x_1)^2,
\end{aligned}$$

where $\widetilde{X}_{s,1}$ is between $X_{s,1}$ and x_1 (and also possibly depends on $X_{k,1}$). Accordingly, we split $R'_{T,3}$ into the following three terms:

$$\begin{aligned}
&\frac{h^{d/2}}{T}\int \sum_{t=1}^{T} K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t \sum_{s=1}^{T} K_h(\mathbf{x}-\mathbf{X}_s) \\
&\quad \times \sum_{k=1}^{T} \bar{w}_g(x_1, X_{k,1})\varepsilon_k\; w(\mathbf{x})d\mathbf{x},
\end{aligned} \tag{4.20}$$

$$\begin{aligned}
&\frac{h^{d/2}}{T}\int \sum_{t=1}^{T} K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t \sum_{s=1}^{T} K_h(\mathbf{x}-\mathbf{X}_s)(X_{s,1}-x_1) \\
&\quad \times \sum_{k=1}^{T} \dot{\bar{w}}_g(x_1, X_{k,1})\varepsilon_k w(\mathbf{x})d\mathbf{x},
\end{aligned} \tag{4.21}$$

(4.22)
$$\frac{h^{d/2}}{2T}\int\sum_{t=1}^{T}K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t\sum_{s=1}^{T}K_h(\mathbf{x}-\mathbf{X}_s)(X_{s,1}-x_1)^2 \times\sum_{k=1}^{T}\ddot{w}_g(\widetilde{X}_{s,1},X_{k,1})\varepsilon_k w(\mathbf{x})d\mathbf{x}.$$

We further split (4.20) into the following two terms:

(4.23)
$$\frac{h^{d/2}}{T}\int\sum_{t=1}^{T}K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t\sum_{s=1}^{T}\{K_h(\mathbf{x}-\mathbf{X}_s)-EK_h(\mathbf{x}-\mathbf{X}_s)\} \times\sum_{k=1}^{T}\bar{w}_g(x_1,X_{k,1})\varepsilon_k w(\mathbf{x})d\mathbf{x}$$

(4.24)
$$h^{d/2}\int\sum_{t=1}^{T}K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t EK_h(\mathbf{x}-\mathbf{X}_1) \times\sum_{k=1}^{T}\bar{w}_g(x_1,X_{k,1})\varepsilon_k w(\mathbf{x})d\mathbf{x}.$$

Using Lemma 1, we can prove that (4.23) is of order $O_P\left(\log T/\sqrt{Tgh^{d(1+\delta)}}\right)=o_P(1)$ in the same way as the proof for that of (4.18). Using Lemma 1 again, (4.24) can be bounded by

$$\sqrt{\frac{h^d}{Tg}}\log T\int_{\mathbb{R}}\left|\left\{\int_{\mathbb{R}^{d-1}}K_h(\mathbf{x}-\mathbf{X}_t) \times EK_h(\mathbf{x}-\mathbf{X}_1)w(\mathbf{x})dx_2\dots dx_d\right\}\varepsilon_t\right|dx_1\cdot O_P(1).$$

The expectation of the whole integral in the above expression is less than

$$\begin{aligned}&\int_{\mathbb{R}}\left[E\left(\sum_{t=1}^{T}\varepsilon_t\int_{\mathbb{R}^{d-1}}K_h(\mathbf{x}-\mathbf{X}_t) \times EK_h(\mathbf{x}-\mathbf{X}_1)w(\mathbf{x})dx_2\dots dx_d\right)^2\right]^{1/2}dx_1\\
&=\sqrt{T}\int_{\mathbb{R}}\left[E\left(W_h(x_1-X_{1,1})\int_{\mathbb{R}^{d-1}}\left\{\prod_{i=2}^{d}W_h(x_i-X_{1,i})\right\} \times\pi(\mathbf{x})\sigma^2(\mathbf{x})w(\mathbf{x})dx_2\dots dx_d\right)^2\right]^{1/2}dx_1\\
&=\sqrt{T}\int_{\mathbb{R}}\left[E\left(W_h(x_1-X_{1,1})\int_{\mathbb{R}^{d-1}}\left\{\prod_{i=2}^{d}W(\mathbf{u}_i-X_{1,i}/h)\right\} \times\pi(x_1,hu_2,\dots,hu_d)\cdot\sigma^2(x_1,hu_2,\dots,hu_d) \times w(x_1,hu_2,\dots,hu_d)du_2\dots du_d\right)^2\right]^{1/2}dx_1\\
&\le O(\sqrt{T})\cdot\int_{\mathbb{R}}\left[EW_h^2(x_1-X_{1,1})\right]^{1/2}dx_1=O(\sqrt{T/h}).\end{aligned}$$

Therefore, (4.24) is of the order $O_P(h^{(d-1)/2}g^{-1/2}\log T)=o_P(1)$. Combining what we have shown for (4.23) and (4.24), we conclude that (4.20) converges to 0 in probability. In a similar way, we can also show that (4.21) is of the order

$$O_P\left(g^{-3/2}\log T\{T^{-1/2}h^{-d(1+\delta)/2+1}+Th^{(d+1)/2}\}\right)=o_P(1).$$

Using Lemma 1, (4.22) may be bounded by

$$\frac{h^{d/2}\log T}{T^{3/2}g^{5/2}}\int\left|\sum_{t=1}^{T}K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t\right| \left|\sum_{s=1}^{T}K_h(\mathbf{x}-\mathbf{X}_s)(x_1-X_{s,1})^2\right|w(\mathbf{x})d\mathbf{x}\cdot O_P(1).$$

The integral in the above expression is smaller than the sum of the following two terms:

(4.25)
$$\int\left|\sum_{t=1}^{T}K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t\right|\left|\sum_{s=1}^{T}\{K_h(\mathbf{x}-\mathbf{X}_s)(x_1-X_{s,1})^2 - E[K_h(\mathbf{x}-\mathbf{X}_s)(x_1-X_{s,1})^2]\}\right|w(\mathbf{x})d\mathbf{x},$$

(4.26)
$$T\int\left|\sum_{t=1}^{T}K_h(\mathbf{x}-\mathbf{X}_t)\varepsilon_t\right|E[K_h(\mathbf{x}-\mathbf{X}_s)(x_1-X_{s,1})^2]w(\mathbf{x})d\mathbf{x}.$$

Along the same lines as the proof of that for (4.18), we can show that (4.25) is of order $O_P(Th^{2-d(1+\delta)})$. Note that $E[K_h(\mathbf{x}-\mathbf{X}_s)(x_1-X_{s,1})^2]=O(h^2)$, which entails that the expectation of (4.26) is of order $O(T^{3/2}h^{(4-d)/2})$. Consequently, we have that (4.22) is of the order

$$\begin{aligned}&O_P\left(\log T\sqrt{\frac{h^d}{T^3g^5}}\{Th^{2-d(1+\delta)}+T^{3/2}h^{(4-d)/2}\}\right)\\
&=O_P\left(\log T\left\{\sqrt{\frac{h^{4-d(1+2\delta)}}{Tg^5}}+\sqrt{h^4/g^5}\right\}\right)=o_P(1).\end{aligned}$$

Since we have proved that all three terms in (4.20)–(4.22) converge to 0 in probability, we obtain that $R'_{T,3}\to 0$ in probability. Now it follows from (4.17) that (a) has been established.

The proof of (b) is much simpler. It follows from Lemma 2 that

$$\begin{aligned}|R_{T,2}|&\le\left\{\sup_{u\in B}|\widehat{m}_g(u)-m_0(u)|\right\}^2\frac{h^{d/2}}{T} \times\int\left\{\sum_{t=1}^{T}K_h(\mathbf{x}-\mathbf{X}_t)\right\}^2w(\mathbf{x})d\mathbf{x}\\
&=O_P\left((\log T)^2\{h^{d/2}g^{-1}+\frac{1}{Tgh^{d(1+2\delta)/2}}\}\right)=o_P(1).\end{aligned}$$

The first equality in the above expression makes use of the fact that the expectation of the integral is of the order $(T^2 + T/h^{d(1+\delta)})$, which has been proved before. This completes our proof. □

A.2 Proof of Theorem 2

Proof of Theorem 2. It is easy to see that $Th^{d/2}[S_T^* - E^*(S_T^*)] \xrightarrow{d} N(0, V^*)$, where $V^* = V + o_P(1)$. In contrast, the proof of $Th^{d/2}|E(S_T') - E^*(S_T^*)| \to 0$ requires more work.

We have

$$\mathrm{E}^* Th^{d/2} S_T^* = \frac{h^{d/2}}{T} \sum_t \int K_h^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h}\right) w(\mathbf{x}) d\mathbf{x}\, \widehat{\varepsilon}_t^{\,2}.$$

First, we split up

$$\widehat{\varepsilon}_t^{\,2} = \varepsilon_t^2 + [\widehat{m}_h(\mathbf{X}_t) - m(\mathbf{X}_t)]^2 + 2\varepsilon_t\, [\widehat{m}_h(\mathbf{X}_t) - m(\mathbf{X}_t)].$$

We get, by the Ergodic theorem,

$$\begin{aligned} &\frac{h^{d/2}}{T} \sum_t \int K_h^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h}\right) w(\mathbf{x}) d\mathbf{x}\, \varepsilon_t^2 \\ &\quad = h^{-d/2} \int\int K^2(\mathbf{u}) w(\mathbf{x}+h\mathbf{u}) \pi(\mathbf{x}) \sigma^2(\mathbf{x})\, d\mathbf{x}\, d\mathbf{u} \\ &\qquad + O_P\left(1/\sqrt{Th^d}\right). \end{aligned}$$

Further,

$$\begin{aligned} &\frac{h^{d/2}}{T} \sum_t \int K_h^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h}\right) w(\mathbf{x})\, d\mathbf{x}\, [\widehat{m}_h(\mathbf{X}_t) - m(\mathbf{X}_t)]^2 \\ &\qquad = \quad O_P\left(h^{-d/2}\left[\frac{1}{Th^d} + h^{2p+2}\right]\right) = o_P(1), \end{aligned}$$

since $2p + 2 > d/2$ and therefore $h \gg T^{-2/(3d)}$.

Analogously to (4.11), we can show that, uniformly in $\mathbf{x} \in B$,

$$\begin{aligned} \widehat{m}_h(\mathbf{x}) - m(\mathbf{x}) &= \sum \overline{w}_h(\mathbf{x}, \mathbf{X}_s)\varepsilon_s + b_\infty \\ &\quad + O_P\left(\frac{(\log T)^{3/2}}{Th^d} + \frac{h^{p+1}\log T}{\sqrt{Th^d}}\right), \end{aligned}$$

where $b_\infty = O(h^{p+1})$. This implies

$$\begin{aligned} &\frac{h^{d/2}}{T} \sum_t \int K_h^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h}\right) w(\mathbf{x}) d\mathbf{x}\, [\widehat{m}_h(\mathbf{X}_t) - m(\mathbf{X}_t)]\varepsilon_t \\ &= \frac{h^{d/2}}{T} \sum_{s,t} \int K_h^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h}\right) w(\mathbf{x})\, d\mathbf{x}\, \overline{w}_h(\mathbf{X}_t, \mathbf{X}_s)\, \varepsilon_s\varepsilon_t \\ &\quad + O_P\left(\frac{h^{d/2}}{T} h^{p+1} h^{-d}\sqrt{T}\right) \\ &\quad + O_P\left(h^{-d/2}\left[\frac{(\log T)^{3/2}}{Th^d} + \frac{h^{p+1}\log T}{\sqrt{Th^d}}\right]\right) \\ &= O\left(h^{-3d/2}T^{-1} + h^{-d}T^{-1}\right) + O_P\left(h^{p+1}/\sqrt{Th^d}\right) \\ &\quad + O_P\left(h^{-d/2}\left[\frac{(\log T)^{3/2}}{Th^d} + \frac{h^{p+1}\log T}{\sqrt{Th^d}}\right]\right) \\ &= o_P(1), \end{aligned}$$

which completes the proof of the desired result. □

Received 10 October 2007

REFERENCES

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes. Lecture Notes in Statistics* **110**, Springer, New York. MR1441072

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3** 123–148. MR1466304

Doukhan, P. (1995). *Mixing: Properties and Examples. Lecture Notes in Statistics* **85**, Springer, New York. MR1312160

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50** 987–1008. MR0666121

Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Ameri. Statist. Assoc.* **91** 674–688. MR1395735

Fan, J., Härdle, W. and Mammen, E. (1998). Direct estimation of additive linear components for high dimensional data. *Ann. Statist.* **26** 943–971. MR1635422

Fan, J. and Jiang, J. (2007). Nonparametric inference with generalized likelihood ratio tests (with discussion). *Test* **16** 409–478. MR2365172

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood test statistic and Wilks phenomenon. *Ann. Statist.* **29** 153–193. MR1833962

Fan, J. and Zhang, W. (2004). Generalized likelihood ratio tests for spectral density. *Biometrika* **91** 195–209. MR2050469

Franke, J., Kreiss, J.-P. and Mammen, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* **8** 1–37. MR1884156

Hall, P. and Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics* **47** 757–762. MR1132543

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. MR1245774

Hjellvik, V. and Tjøstheim, D. (1995). Nonparametric tests of linearity for time series. *Biometrika* **82** 351–368. MR1354234

Hjellvik, V. and Tjøstheim, D. (1996). Nonparametric statistics for testing of linearity and serial independence. *J. Nonparametric Statist.* **6** 223–251. MR1383053

Hjellvik, V., Yao, Q. and Tjøstheim, D. (1996). Linearity testing using local polynomial approximation. Discussion Paper **60**, Sonderforschungsbereich 373, Humboldt Universität zu Berlin.

Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *J. Statist. Plann. Inference* **68** 295–321. MR1629587

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Rel. Fields* **75** 261–277. MR0885466

Kreiss, J.-P. (1988). *Asymptotic Statistical Inference for a Class of Stochastic Processes, Habilitationsschrift.* University of Hamburg.

Kreiss, J.-P. (1997). Asymptotical properties of residual bootstrap for autoregressions. Technical Report, TU Braunschweig.

Lepski, O. V. and Spokoiny, V. G. (1999). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative, *Bernoulli* **5** 333–358. MR1681702

Neumann, M. H. (1995). Automatic bandwidth choice and confidence intervals in nonparametric regression, *Ann. Statist.* **23** 1937–1959. MR1389859

Neumann, M. H. and Kreiss, J.-P. (1998). Regression-type inference in nonparametric autoregression. *Ann. Statist.* **26** 1570–1613. MR1647701

Press, H. and Tukey, J. W. (1956). *Power Spectral Methods of Analysis and their Application to Problems in Airplane Dynamics.* Bell Telephone System Monograph 2606. MR0080430

Tong, H. (1990). *Non-Linear Time Series: a Dynamical System Approach*. Oxford: Oxford University Press. MR1079320

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1343. MR0868303

Yang, L., Härdle, W. and Nielsen, J. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *J. Time Series Analysis* **20** 579–604. MR1720162

Jens-Peter Kreiss
Institut für Mathematische Stochastik
Technische Universität Braunschweig
Pockelsstrasse 14
D-38106 Braunschweig, Germany
E-mail address: j.kreiss@tu-bs.de

Michael H. Neumann
Institut für Stochastik
Friedrich-Schiller-Universität Jena
Ernst-Abbe-Platz 2
D-07743 Jena, Germany
E-mail address: mneumann@mathematik.uni-jena.de

Qiwei Yao
Department of Statistics
London School of Economics
Houghton Street, London
WC2A 2AE, United Kingdom
E-mail address: q.yao@lse.ac.uk