Dedicated to
Professor Jakub Gutenbaum
on his 70th birthday

Control and Cybernetics

vol. **29** (2000) No. 1

Testing for a difference between conditional variance functions of nonlinear time series

by

J. Ćwik[1], J. Koronacki[1,2] and J. Mielniczuk[1,2]

[1] Institute of Computer Science, Polish Academy of Sciences
Ordona 21, 01-237 Warsaw, Poland

[2] Polish-Japanese Institute of Information Technologies,
Koszykowa 86, 02-008 Warsaw, Poland

Abstract: In this report, the problem of testing for a difference between conditional variance functions (or volatilites) of two independent nonlinear time series is investigated by means of an extensive simulation study. Empirical results on the properties of the test proposed confirm the test's validity, at least for some types of heteroscedasticity as contrasted with homoscedastic errors as well as for some types of differences in heteroscedasticity. Moreover, interesting properties of several estimators of conditional mean, variance and fourth moment functions are empirically found too.

Keywords: nonparametric autoregression, nonlinear time series, conditional variance, volatility, homogeneity test for conditional variances.

1. Introduction

Let

$$X_t = \lambda(X_{t-1}) + \delta(X_{t-1})\epsilon_t \tag{1}$$

be an autoregressive model, where $X_0 = x_0$ and $(\epsilon_t)_{t=0}^{\infty}$ is a sequence of i.i.d. random variables such that the ϵ_t's are independent of the sigma-field $\sigma(X_0, X_1, \ldots, X_{t-1})$. Clearly,

$$E(X_t|X_{t-1} = x) = \lambda(x)$$

and, with $\sigma_0^2 = \mathrm{Var}(\epsilon_t)$,

$$\mathrm{Var}(X_t|X_{t-1} = x) \equiv \gamma_2(x) = \delta^2(x)\sigma_0^2.$$

The conditional variance or volatility becomes homoscedastic if $\delta(\cdot)$ is a constant; in general, it is heteroscedastic.

The problem of identification and, to some extent, testing of nonlinear and/or heteroscedastic time series has received much attention in the last decade. For excellent surveys see Tjøstheim (1994) and Tong (1995); see also Tong (1990) for a relatively early but fundamental treatment of nonlinear time series. An analysis of ARCH (Autoregressive Conditionally Heteroscedastic) models can be found in Gouriéroux (1997). For recent developments in nonparametric estimation of $\lambda(\cdot)$ and $\delta^2(\cdot)\sigma_0^2$, see in particular Hoffman (1999), Neumann and Kreiss (1998), Härdle and Tsybakov (1997), and the literature there for related earlier work. In Hoffman (1999), wavelet threshold estimators for both $\lambda(\cdot)$ and $\delta^2(\cdot)\sigma_0^2$ have been investigated. In Neumann and Kreiss (1998), estimators for $\lambda(\cdot)$, based on local polynomial estimators (LPE's) have been dealt with using their strong approximations to LPE's for corresponding regression models. In this way, bootstrap methodology for nonparametric autoregressions has been simplified and, hence, bootstrap confidence bands and (composite) goodness-of-fit supremum-type tests have been provided for $\lambda(\cdot)$ as well. Härdle and Tsybakov (1997) used local polynomial fits to estimate the conditional variance function.

Regressogram estimators for cumulative versions of $\lambda(\cdot)$ and $\delta^2(\cdot)\sigma_0^2$ (Subsection 2.3 below), along with corresponding confidence bands and some goodness-of-fit tests, have been given by McKeague and Zhang (1994). Recently, Hafner and Herwatz (1999) have examined empirical properties of several testing procedures for autoregressive dynamics of order one (against pure noise null hypothesis) and Liero (1999) has proposed a nonparametric test of homoscedasticity against heteroscedasticity for the nonlinear regression problem.

The main goal of this report is to investigate small-sample properties of methods for distinguishing between models which are given by the same conditional mean function $\lambda(\cdot)$ but differ in conditional variance functions $\delta^2(\cdot)\sigma_0^2$. Thus, in contrast to *goodness-of-fit type* tests, where the model is known exactly under a simple null hypothesis or it is known up to a parameter in case of a composite null hypothesis, we aim at tests of *homogeneity* of two autoregressive models (actually, we are interested in testing homogeneity, i.e., equality, of two conditional variance functions). Our study relies on simulations for samples of size 500, whose size we consider small to medium if one samples from a time series starting from a fixed initial value. More precisely, when deciding on the sample size, we aimed at obtaining clear distinguishability of hypotheses under scrutiny. Our preliminary simulation studies (not reported here) showed that rather similar results can be hoped for for, say, samples of size 400 and still meaningful but much weaker results for samples of size 300.

One way to obtain a test for a difference between such models is to construct a confidence band for the difference between two conditional variance functions in the two-sample problem with two independent time series (which have or have not identical conditional mean functions). Under the hypothesis of equality of conditional variance functions, the confidence band should include the zero function. If either the band's upper bound goes negative or the band's lower bound goes positive, the two conditional variance functions are likely to be different. In Section 3 we study empirical properties of tests based on this idea.

Section 2 begins with presentation of the models studied. Except for models Cminus and E, they are either borrowed from or are heteroscedastic variations of the models discussed by Auestad and Tjøstheim (1990) and McKeague and Zhang (1994). Following the latter of those papers, estimators of cumulative versions of the conditional mean and variance functions are used to construct the confidence bands needed.

In Subsection 2.2, we get results on estimation of $\lambda(\cdot)$, in particular, on how different estimators behave for homoscedastic as opposed to heteroscedastic errors. Also, we get results concerning different estimators of $\gamma_2(\cdot)$ and $\gamma_4(\cdot)$, where

$$\gamma_4(x) = \mathrm{Var}((X_t - \lambda(x))^2 | X_{t-1} = x).$$

All these estimators are needed to construct the confidence bands, which are described in Subsection 2.3.

All in all, our simulation results confirm the validity of the approach proposed. Indeed, the suggested method of testing for a difference between the conditional variance functions has proved truly promising already for small to medium samples and it deserves further study. To the best of our knowledge the method is new. For its possible applications one can confer, e.g., Tong (1990) and McKeague and Zhang (1994).

2. Models and estimators considered

2.1. Autoregressive models

In this study, the following homoscedastic models are taken into consideration. In all the models except in model E, we assume that $\epsilon_t \sim N(0, 0.01)$.

Model A. (Linear autoregressive, AR(1))

$$X_t = 0.8X_{t-1} + \epsilon_t.$$

Model B. (Exponential autoregressive)

$$X_t = (0.8 - 1.1\exp\{-30X_{t-1}^2\})X_{t-1} + \epsilon_t.$$

Model C. (Threshold autoregressive)

$$X_t = 0.8X_{t-1}I\{X_{t-1} > 0\} - 0.3X_{t-1}I\{X_{t-1} \leq 0\} + \epsilon_t.$$

Model Cminus.

$$X_t = -0.8X_{t-1}I\{X_{t-1} > 0\} + 0.3X_{t-1}I\{X_{t-1} \leq 0\} + \epsilon_t.$$

Model D. (Random coefficient autoregressive)

$$X_t = (1 + \tilde{\epsilon}_t)(0.8X_{t-1}I\{X_{t-1} > 0\} - 0.3X_{t-1}I\{X_{t-1} \leq 0\}) + \epsilon_t$$

with $\tilde{\epsilon}_t$ independent of ϵ_t and $\sigma(X_0, X_1, \ldots, X_{t-1})$, and assuming value 0.25 with probability 1/2 and value -0.25 with probability 1/2.

Model E. (Discontinuous)

$$X_t = \frac{2\text{sgn}X_{t-1}}{3 + |X_{t-1}|} + \epsilon_t$$

where $\epsilon_t \sim N(0, 0.5^2)$.

In the original version of model B, as introduced by Auestad and Tjøstheim (1990), the factor in the exponent is 50, not 30. We have used the latter factor to get a function whose nonlinearity is more apparent. It should also be noted that model D does not belong to the class of autoregressive models but is a special case of the hidden Markov chain models. It is, however, worth consideration, since its conditional mean function is the same as that of model C. The last of the models is interesting due to its discontinuity at zero.

Models A, B,...,E will be referred to as the primal ones. In addition, two modifications of each of these models will be considered, to be referred to as primed models and double-primed models, respectively.

Primed models are obtained by including errors of the form $\delta(X_{t-1})\epsilon_t$, where

$$\delta(x) = \begin{cases} 0, & \text{if } x \leq -1, \\ 2x + 2, & \text{if } -1 < x \leq 0, \\ -2x + 2, & \text{if } 0 < x < 1, \\ 0, & \text{if } x \geq 1. \end{cases}$$

Double-primed models are obtained by including ARCH-like errors with

$$\delta(X_{t-1}) = \sqrt{1 + (1 - X_{t-1})^2}.$$

It is worth observing that, within each triple of models (consisting of a primal model, primed model and double-primed model), the conditional mean function is the same but the conditional variance functions are different. In order to

obtain the latter, it suffices to use formula (1) for families A, B, C and E, and to note that model D can be written as

$$X_t = (1 + \tilde{\epsilon}_t)\lambda(X_{t-1}) + \delta(X_{t-1})\epsilon_t, \tag{2}$$

where all the $\lambda(\cdot)$ and $\delta(\cdot)$ are known. The resulting conditional variance functions are as follows. For all the models except model D, the variance of the primal model is σ_0, of the primed model is $\delta^2(x)\sigma_0^2$ and of the double-primed model is $(1+(1-x)^2)\sigma_0^2$. For model D we have, respectively, $\sigma_0^2 + (1/16)\lambda^2(x)$, $\delta^2(x)\sigma_0^2 + (1/16)\lambda^2(x)$ and $(1 + (1-x)^2)\sigma_0^2 + (1/16)\lambda^2(x)$.

2.2. Estimators

The following two well-known types of nonparametric estimators have been used to estimate conditional mean function: kernel estimator and LOWESS estimator. The former is of the form

$$\hat{\lambda}(x) = \frac{\frac{1}{n-1}\sum_{j=1}^{n-1} X_{j+1} K\left(\frac{x-X_j}{h}\right)}{\frac{1}{n}\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)}, \tag{3}$$

where $X_1, X_2, \ldots, X_n$ are the observations, h is the bandwidth or smoothing factor and $K(\cdot)$ is the kernel. In our simulations, the Gaussian kernel was taken, i.e., $K(x) = (2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$. For a discussion of asymptotic properties of kernel estimators see Tjøstheim (1994) and references there (throughout this paper, we skip asymptotic considerations, since our interest has been focused on small- to medium-size samples). In the simulations, either a bandwidth of fixed width was used (chosen a priori by trial and error) or its width was chosen adaptively, namely, the k-nearest neighbour ($k - nn$) approach was used to determine h locally for each x.

The locally weighted scatter plot smoothing (or LOWESS for short) estimator belongs to the family of LPE's. More precisely, in our implementations, it is a locally linear estimator based on weighted least squares fits over local neighbourhoods of observations. The task is to estimate the regression function from the sample $((X_i, X_{i-1}))_{i=2}^n$ on a fixed interval $[a, b]$. For any given sample, we choose $a = \min\{X_i\}_{i=1}^n$ and $b = \max\{X_i\}_{i=1}^n$. For each fixed observation (or design point) X_i, its neighbourhood $N(X_i)$ is constituted as including $k = 0.3n$ nearest observations to X_i. Then, for each X_i and neighbourhood $N(X_i)$, observations in $N(X_i)$ are assigned weights using the tri-cube weight function:

$$W\left(\frac{|X_i - X_j|}{\Delta(X_i)}\right),$$

where $W(u) = (1-u^3)^3$ for $u \in [0, 1)$ and $W(u) = 0$ otherwise, and $\Delta(X_i)$ is the largest distance between X_i and another observation in $N(X_i)$. Now, for each X_i, $\hat{\lambda}(X_i)$ is obtained using the weighted least squares fit over $N(X_i)$. For x

different from the observed points (or design points), $\hat{\lambda}(x)$ can be obtained, for example, via interpolation. In our simulations, the S-Plus implementation of the LOWESS estimator was used (for a general description and asymptotic analysis of the estimator, as well as for those of its multidimensional counterpart, known as LOESS, see, e.g., Fan and Gijbels (1996)).

Both kernel and LOWESS estimators can in turn be applied to build estimators of the conditional variance function $\gamma_2(\cdot)$. In general, estimation of this function can be performed in at least two ways. The first is of the following general form

$$\widehat{\gamma}_2(x) = \widehat{E}(X_t^2 \mid X_{t-1} = x) - (\hat{\lambda}(x))^2, \tag{4}$$

where $\widehat{\ }$ refers to an estimator of conditional expectation needed. The second approach to estimating $\gamma_2(\cdot)$ consists in computing an estimate of the regression function from the sample

$$((X_t - \hat{\lambda}(x))^2, X_{t-1})_{t=2}^n. \tag{5}$$

Estimators based on (4) are referred to as K1 (or K1-fixed_h) estimators if kernel estimators are used to estimate the two conditional expectations required, and as L1 if instead LOWESS estimators are used (the forms of suitable estimators are not given here, since they can be obtained by trivially modifying corresponding estimators of $\lambda(\cdot)$; e.g., as regards kernel estimators, it suffices to note that the estimator given by (3) is a kernel estimator for regression $\lambda(x) = E(X_t|X_{t-1} = x)$.

Analogously, estimators based on (5) are referred to as K2 and L2, respectively; in this case, estimators of the same type have been used for estimating both $\lambda(\cdot)$ and then $E((X_t - \hat{\lambda}(x))^2|X_{t-1} = x)$. In fact, a simple algebra shows that estimator K2 can be disregarded: if in (3) the denominator is replaced by

$$\frac{1}{n-1}\sum_{j=1}^{n-1} K\left(\frac{x - X_j}{h}\right),$$

the "new" kernel estimator obtained, which is asymptotically strictly equivalent and practically always equivalent to the "old" one, provides the same estimate as K1. Accordingly, in the sequel, only estimator K1 is taken into account.

Confidence band for the conditional variance function requires an estimate of the fourth conditional moment,

$$\gamma_4(x) = \mathrm{Var}((X_t - \lambda(x))^2 \mid X_{t-1} = x).$$

Perhaps the most natural or the most immediate candidates for an estimator of $\gamma_4(\cdot)$ are the following two estimators:

$$\widehat{\gamma}_4(x) = \widehat{E}((X_t - \widehat{\lambda}(x))^4 \mid X_{t-1} = x) - \{\widehat{\gamma}_2(x)\}^2, \tag{6}$$

where $\widehat{\gamma}_2(x)$ is an estimator of the conditional variance function, and a regression function estimator based on the sample

$$(((X_t - \hat{\lambda}(x))^2 - \widehat{\gamma}_2(x))^2, X_{t-1})_{t=1}^n. \tag{7}$$

The rationale behind estimator (6) is clear, while that for estimator (7) follows from the fact that, for families A, B, C and E, $\gamma_4(x) = E(((X_t - \lambda(x))^2 - \gamma_2(x))^2) \mid X_{t-1} = x)$. One more reason to consider the latter estimator will be given in the next section.

In order to introduce one more candidate for an estimator of $\gamma_4(\cdot)$, let us observe first that for families A, B, C, and E (recall that the ϵ_t are normal and hence the last equality follows)

$$\begin{aligned}\gamma_4(x) &= E(\{\delta(X_{t-1})\epsilon_t\}^4 \mid X_{t-1} = x) - \{\gamma_2(x)\}^2 \\ &= \delta^4(x)E\epsilon_t^4 - \{\delta^2(x)\sigma_0^2\}^2 = 2\sigma_0^4\delta^4(x).\end{aligned} \tag{8}$$

Analogously, using (2) and performing some elementary calculations we get for family D

$$\begin{aligned}\gamma_4(x) &= E((\tilde{\epsilon}_t\lambda(x) + \delta(X_{t-1})\epsilon_t)^4 \mid X_{t-1} = x) - \{\gamma_2(x)\}^2 \\ &= 2\sigma_0^4\delta^4(x) + 0.25\sigma_0^2\lambda^2(x)\delta^2(x).\end{aligned} \tag{9}$$

Finally, let us observe that for models from families A, B, C and E

$$Var(X_t^2 \mid X_{t-1} = x) = 2\sigma_0^4\delta^4(x) + 4\sigma_0^2\lambda^2(x)\delta^2(x).$$

Thus, in view of (8) and (9), the following conservative estimator can also be proposed

$$\widehat{\gamma}_4(x) = \widehat{E}(X_t^4 \mid X_{t-1} = x) - \{\widehat{E}(X_t^2 \mid X_{t-1} = x)\}^2. \tag{10}$$

Before concluding this subsection we will forward one more remark. Confidence bands of our interest require also that a stationary marginal density of the observations X_t be estimated. A natural estimator is, of course, the usual kernel estimator with smooth kernel function. Since the estimator appears in the formulas below in the denominator, we modify it slightly to make it bounded away from zero. Namely, we use the estimator of the form

$$\widehat{g}(x) = \max\{\widehat{\widehat{g}}(x), 0.05\},$$

where

$$\widehat{\widehat{g}}(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right),$$

K is the Gaussian kernel and h is the same bandwidth as that in (3).

2.3. Confidence bands

For any two independent time series, let the estimators of their conditional mean and variance functions be $\hat{\lambda}$ and $\hat{\gamma}_2$ for one series and $\tilde{\lambda}$ and $\tilde{\gamma}_2$ for the other.

As has been already mentioned in the Introduction, the confidence bands to be dealt with are based on cumulative versions of the conditional mean and variance functions,

$$\Lambda(\cdot) = \int_a^{\cdot} \lambda(x)dx$$

and

$$\Gamma(\cdot) = \int_a^{\cdot} \gamma_2(x)dx,$$

where, as usual, $\gamma_2(x) = Var(X_t|X_{t-1} = x)$ and a is an appropriately chosen point in the state space (we restrict ourselves to estimation and testing on some interval [a,b]).

Let for any $x \in [a, b]$

$$\hat{\Lambda}(x) = \int_a^x \hat{\lambda}(s)ds - \int_a^x \tilde{\lambda}(s)ds.$$

It follows from McKeague and Zhang (1994) that under appropriate conditions the asymptotic $100(1-\alpha)\%$ confidence band to test for a difference between two conditional mean functions should have the form

$$\hat{\Lambda}(x) \pm c_\alpha n^{-1/2}(\hat{H}(b))^{1/2}(1 + \frac{\hat{H}(x)}{\hat{H}(b)})$$

where

$$\hat{H}(x) = \int_a^x \frac{\hat{\gamma}_2(s)}{\hat{g}(s)}ds + \int_a^x \frac{\tilde{\gamma}_2(s)}{\tilde{g}(s)}ds,$$

$\hat{g}$, $\tilde{g}$ are estimators of stationary marginal densities of the two series and c_α is a constant depending on the confidence level $1-\alpha$, e.g.,

$1-\alpha$	c_α
0.90	1.133
0.95	1.273
0.99	1.552

In turn, let for $x \in [a, b]$

$$\hat{\Gamma}(x) = \int_a^x \hat{\gamma}(s)ds - \int_a^x \tilde{\gamma}(s)ds.$$

Then, it again follows from McKeague and Zhang (1994) that (under the same conditions as before if $X_0 = x_0$ is fixed) the asymptotic $100(1-\alpha)\%$ confidence band to test for a difference between two conditional variance functions should have the form

$$\hat{\Gamma}(x) \pm c_\alpha n^{-1/2} (\hat{I}(b))^{1/2} (1 + \frac{\hat{I}(x)}{\hat{I}(b)})$$

where

$$\hat{I}(x) = \int_a^x \frac{\hat{\gamma}_4(s)}{\hat{g}(s)} ds + \int_a^x \frac{\tilde{\gamma}_4(s)}{\tilde{g}(s)} ds,$$

and $\hat{\gamma}_4$, $\tilde{\gamma}_4$ are corresponding estimators of $\gamma_4(x)$ for the two series considered.

Again, we omit stating assumptions under which the asymptotic results of McKeague and Zhang are valid. However, a few comments are in place here. First, McKeague and Zhang considered only histogram estimators for marginal densities and histogram-like estimators (or regressograms) for $\lambda(\cdot)$ and $\gamma_2(\cdot)$, all these with nonadaptive (i.e., deterministic or fixed in advance) bandwidth h (satisfying conditions $nh^2 \to \infty$ and $nh^4 \to 0$ as the sample size $n \to \infty$). Still, e.g., Theorem 23.2.1 in Shorack and Wellner (1986) indicates that the given confidence bands remain valid if, in particular, kernel estimators considered in Subsection 2.2 are used with the same h as that for the regressograms and histograms.

Second, the results of McKeague and Zhang do not carry over to the case with kernel estimators which have variable bandwidths, let alone to the case with the LOWESS estimator. At the same time, some way of locally adapting bandwidths and neighbourhoods to data is strongly recommended in practice, in particular when sample sizes are small to moderate. Moreover, well-known analytical results for $k-nn$ kernel estimators and LPEs also provide a reliable justification for consideration of these estimators within the context of current investigation.

3. Simulation results

In this report, we confine ourselves to the case when two independent time series have the same conditional mean function. The reason is that the procedure to distinguish between two models consists in fact of two stages. First, equality of two conditional mean functions is tested and, if the hypothesis of their equality is not rejected, the two models are tested for a difference between their conditional variance functions.

In order to construct reliable confidence bands, all the suggested estimators of $\lambda(\cdot)$, $\gamma_2(\cdot)$ and $\gamma_4(\cdot)$ had to be evaluated beforehand. In all simulations, samples of size 500 were taken and the starting value X_0 was always set at 0. For any given sample $X_1, \ldots, X_{500}$, the estimation interval was taken to be

$[a, b]$ with $a = \min\{X_i\}_{i=1}^{500}$ and $b = \max\{X_i\}_{i=1}^{500}$. (All the simulations, including generation of random samples, were performed using S-Plus 5.)

Kernel estimators with fixed h were always chosen to have $h = 0.1$ for families A, B, C and D, and $h = 0.3$ for family E. For kernel estimators with $k - nn$ bandwidth, k was always equal to $0.3n$, with n being the sample size. Near the boundaries of the $[a, b]$ interval, however, bandwidth h of the $k - nn$ kernel estimator was modified in order to avoid undesirable boundary effects. The $k - nn$ rule was applied for $x \in [a + 0.2\Delta, b - 0.2\Delta]$, where $\Delta = b - a$, while for $x \in [a, a + 0.2\Delta)$ and $x \in (b - 0.2\Delta]$, neighbourhoods of length equal to that of the neighbourhood of the closest point in $[a + 0.2\Delta, b - 0.2\Delta]$ were taken.

The estimators were evaluated by comparing their Empirical Integrated Squared Errors,

$$\text{EISE} = \frac{1}{\#\{x_i\}} \sum_{x_i} (\hat{f}(x_i) - f(x_i))^2,$$

where the x_i's are equidistant points in $[a, b]$, $\#\{x_i\}$ is the number of points x_i in $[a, b]$, $f(\cdot)$ is a function to be estimated, and $\hat{f}(\cdot)$ is its estimator. Actually, for each model considered, 200 repetitions of the experiment were conducted, i.e., 200 sets of samples were generated, and estimation of each of the functions of interest was performed for each sample. In this way, for each model and each function of interest, densities of EISE (based on 200 repetitions) for each estimator were obtained. In all cases, $\#\{x_i\}$ was taken to be equal 100.

Let us note that kernel estimators with fixed h were used only for comparative purposes. Indeed, in practice, when the true underlying density is unknown, the bandwidth should be chosen adaptively, without human intervention. However, in a simulation study, one can take the fact that one knows the true density to his or her advantage. In particular, an optimal (or near to optimal) bandwidth can be found by trial and error. Results obtained using estimators with bandwidths thus determined can then be used as a reference for those obtained using adaptive methods. Clearly, adaptive estimators should be required to provide results similar to those obtainable by using estimators with (nearly) optimal bandwidths.

Let us begin by briefly discussing the results obtained for the estimators of $\lambda(\cdot)$. In this case, a brief discussion will suffice, since estimation of the conditional mean function is by far the simplest problem to deal with in this study. In general, all the estimators proved reliable in all cases with the LOWESS estimator being superior to the two kernel estimators. The $k - nn$ kernel estimator performed worst, except for example E when it performed equally well as the LOWESS estimator. Except for example E, heteroscedasticity had seemingly no effect on the performance of the LOWESS estimator. Rather surprisingly, reliability of the kernel estimator with fixed h was improved by switching from homoscedastic to heteroscedastic errors.

Results for the estimators of $\gamma_2(\cdot)$ and $\gamma_4(\cdot)$ are summarized in Figs. 1-3 for the former function and in Figs. 4-6 for the latter. In the figures, K1 stands

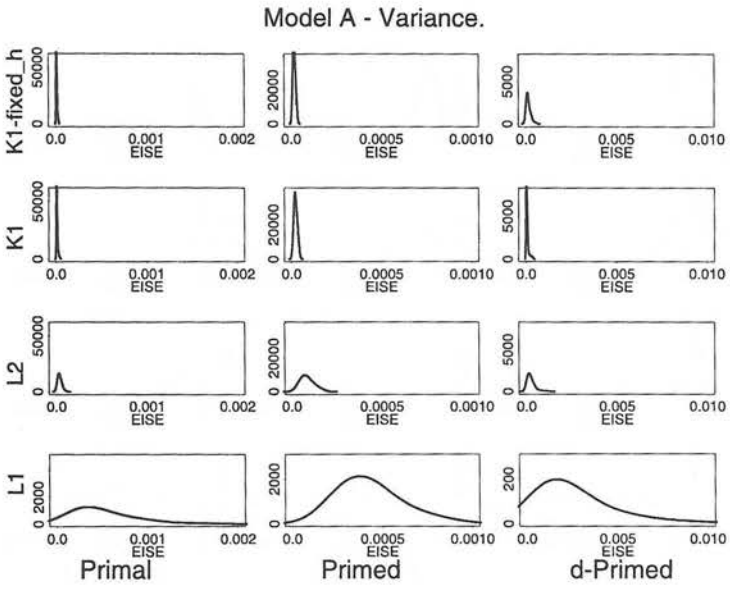

Figure 1. Estimating conditional variance functions in family A: Distributions of EISE for K1-fixed_h, K1, L1 and L2

for the $k - nn$ kernel estimator and K1-fixed_h for the kernel estimator with fixed bandwidth. K1w6 (L2w6) stands for estimator (6) of $\gamma_4(\cdot)$ with K1 (L2) as the estimator of $\lambda(\cdot)$. Analogous convention applies to estimator L2w7 (and to other estimators referred to in the sequel). In the figures, only the results for families A, B and C are given, as the other follow essentially the same pattern.

Given that the $k - nn$ kernel estimator proved reliable as an estimator of $\lambda(\cdot)$ (albeit inferior to the other two estimators) and taking into account results presented in Figs. 1-3 and 4-6, one concludes that it is estimator K1 which should be recommended for use in constructing confidence bands. It also follows from Figs. 1-3 that estimator L1 is unacceptable – the bulk of the density of EISE for estimator L1 lies in fact outside of the supports of the densities for other estimators (note that the densities for L1 estimator required sometimes a different scale than that for other estimators). Estimator L2 is better than L1 as an estimator of $\gamma_2(\cdot)$, but it leads to unacceptable results if it is used to build an estimator of $\gamma_4(\cdot)$, regardless of whether one relies on estimator (6) or (7).

Comparison of the properties of L1 and L2 estimators (see (4) and (5)) suggests that using the LOWESS estimator in (10) can hardly lead to acceptable results (it seems that "centering" an estimator, as done in (5) by subtracting $\lambda(\cdot)$ from X_t, may improve LOWESS estimation). This conjecture was confirmed by simulations (not reported here). By the same token, it could be believed that

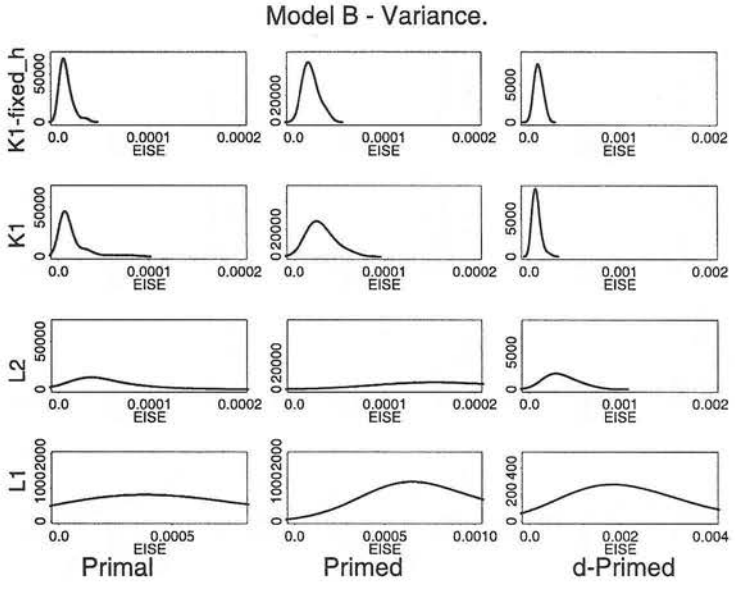

Figure 2. Estimating conditional variance functions in family B: Distributions of EISE for K1-fixed_h, K1, L1 and L2

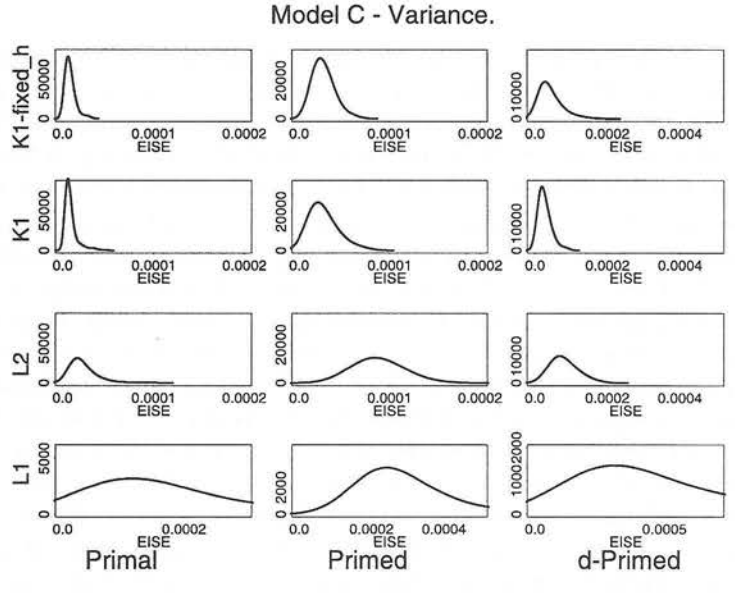

Figure 3. Estimating conditional variance functions in family C: Distributions of EISE for K1-fixed_h, K1, L1 and L2

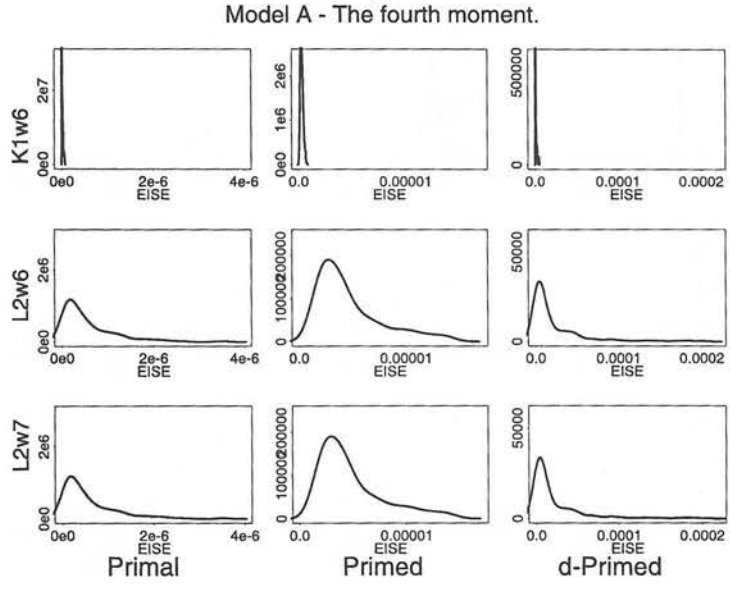

Figure 4. Estimating $\gamma_4(\cdot)$ in family A: Distributions of EISE for K1w6, L2w6 and L2w7

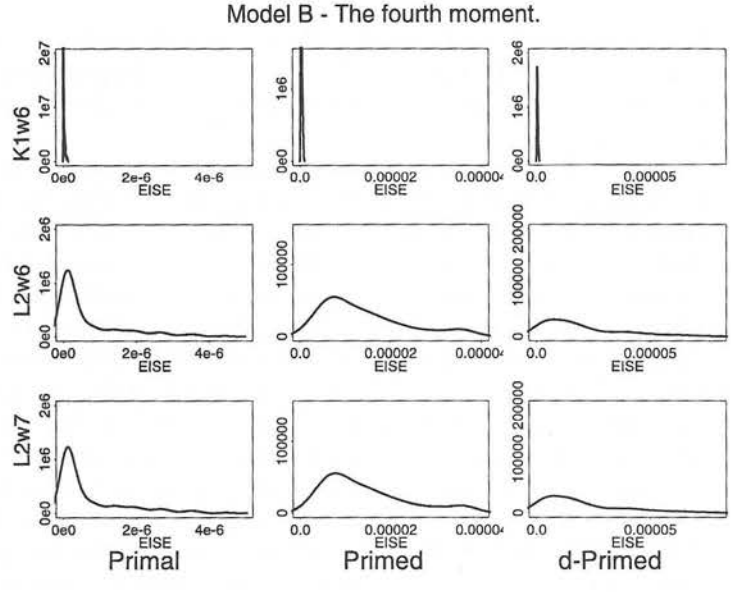

Figure 5. Estimating $\gamma_4(\cdot)$ in family B: Distributions of EISE for K1w6, L2w6 and L2w7

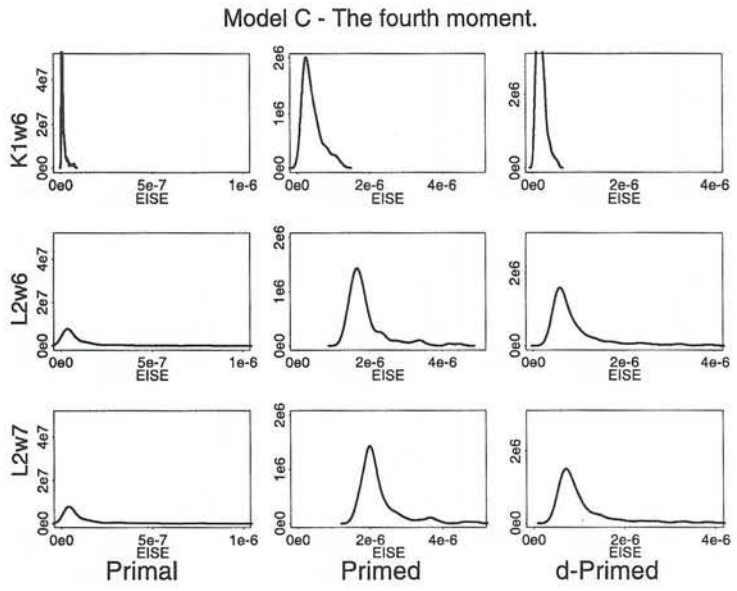

Figure 6. Estimating $\gamma_4(\cdot)$ in family C: Distributions of EISE for K1w6, L2w6 and L2w7

estimator L2w7 would perform better than L2w6. Unfortunately, as Figs. 4-6 show, it has proved not to be the case. In any case, the disappointing results for the LOWESS-type estimators are rather surprising and their explanation requires further study.

As should have been expected, K1w7 performs just as K1w6 does (we still use the same notational convention). All in all, therefore, we are left with estimator K1 of $\gamma_2(\cdot)$ and we have to choose between K1w6 and K1w10 to estimate $\gamma_4(\cdot)$. This final choice is made by a suitable comparison of the procedures to test for a difference between two conditional variance functions, these procedures obtained using either K1w6 or K1w10 to construct the confidence bands required. To be concise, we present our results in the form of suitable tables (see Tables 1-4).

Samples of size 500 (and with $x_0 = 0$) from the given models were generated and 95% confidence bands were constructed over interiors of the supports of the estimated stationary marginal densities. Proper choice of such interiors requires some care. First, one has to note that the confidence band is based on comparing two series and, thus, one has to deal with two supports, one for each series. Accordingly, for each series, the interval between estimated quantiles of prespecified orders, ϕ and ψ, was constructed and then the intersection of the intervals for both series was obtained (this intersection to be referred to as [qϕ,qψ]). Finally, the confidence band was built over the interval [qϕ,qψ].

Appropriate values for ϕ and ψ were found empirically, as the tables illustrate

Model		interval (q.10,q.90)	interval (q.15,q.85)	interval (q.20,q.80)	interval (Q1,Q3)
A	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	100.0	99.0	100.0	95.0
	primed, d-primed	20.0	48.6	66.0	77.0
B	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	100.0	100.0	100.0	99.0
	primed, d-primed	92.0	96.8	93.0	93.0
C	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	92.0	90.8	87.0	71.0
	primed, d-primed	89.0	87.6	78.0	57.0
C-minus	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	100.0	100.0	100.0	100.0
	primed, d-primed	80.0	83.8	80.0	82.0
D	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	87.5	86.6	79.0	58.0
	primed, d-primed	81.5	82.2	77.0	61.0
E	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	98.0	96.0	73.0	32.0
	primed, d-primed	97.0	99.8	100.0	100.0

Table 1. Percentages of rejections of the hypothesis of equality of two variances: K1 estimation with $\widehat{\gamma}_4$ given by (10).

(in the tables, Q1 and Q3 denote the first and third quartile, respectively). Results in the tables are based on repeated simulations of each comparison between the models at hand: the results in Table 1 are based on 500 repetitions of each experiment, while those in Table 2-4 are based on 1000 repetitions. The results in Tables 1-3 concern the test for a difference between the conditional variance functions (results on the power of the test are given in Tables 1 and 2, and those on the size of the test are given in Table 3). For the sake of completeness, in Table 4, results on the size of the test for a difference between the conditional mean functions are presented (recall that, within each family of models, the conditional means are the same).

In all the simulations, confidence bands were calculated at 100 equidistant points. The tested hypothesis of equality of two functions was rejected if the rule that the band's upper bound should stay positive and the band's lower bound should stay negative was violated at least at three points (in all simulations, c_α corresponding to $1 - \alpha = 0.95$ was used).

Simulations have shown that already about 200 repetitions suffice to provide rather stable results and there is virtually no difference between results for 500 and 1000 repetitions. Using (6) to obtain $\widehat{\gamma}_4(\cdot)$ has proved decidedly better than relying on (10). Upon the results obtained, the interval [q0.15,q0.85] has been found most suitable to build confidence bands.

Model		interval (q.10,q.90)	interval (q.15,q.85)	interval (q.20,q.80)	interval (Q1,Q3)
A	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	100.0	100.0	100.0	99.3
	primed, d-primed	72.3	84.6	89.1	87.8
B	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	100.0	100.0	99.9	99.2
	primed, d-primed	96.4	97.3	96.2	93.8
C	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	100.0	99.8	99.0	95.5
	primed, d-primed	98.8	98.0	96.4	89.7
C-minus	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	100.0	100.0	99.9	99.7
	primed, d-primed	89.7	91.1	90.9	85.5
D	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	99.9	99.7	98.9	93.9
	primed, d-primed	98.7	98.4	96.7	90.6
E	primal, primed	100.0	100.0	100.0	100.0
	primal, d-primed	100.0	100.0	95.9	76.8
	primed, d-primed	99.6	100.0	100.0	100.0

Table 2. Percentages of rejections of the hypothesis of equality of two variances: K1 estimation with $\widehat{\gamma}_4$ given by (6).

Model		interval (q.15, q.85)
A	primal, primal	0.6
	primed, primed	1.6
	d-primed, d-primed	0.6
B	primal, primal	1.6
	primed, primed	1.2
	d-primed, d-primed	1.8
C	primal, primal	2.4
	primed, primed	1.6
	d-primed, d-primed	1.8
C-minus	primal, primal	1.6
	primed, primed	2.0
	d-primed, d-primed	1.0
D	primal, primal	1.6
	primed, primed	0.4
	d-primed, d-primed	2.2
E	primal, primal	0.8
	primed, primed	2.2
	d-primed, d-primed	0.6

Table 3. Percentages of rejections of the hypothesis of equality of two means: variances: K1 estimation with $\widehat{\gamma}_4$ given by (6).

Model		interval (q.15, q.85)
A	primal, primed	0.9
	primal, d-primed	1.8
	primed, d-primed	1.7
B	primal, primed	1.6
	primal, d-primed	1.3
	primed, d-primed	2.0
C	primal, primed	1.6
	primal, d-primed	1.9
	primed, d-primed	1.7
C-minus	primal, primed	0.8
	primal, d-primed	1.4
	primed, d-primed	1.4
D	primal, primed	1.0
	primal, d-primed	1.4
	primed, d-primed	1.8
E	primal, primed	2.4
	primal, d-primed	1.9
	primed, d-primed	3.3

Table 4. Percentages of rejections of the hypothesis of equality of two means: K1 estimation.

In summary, the simulation results strongly suggest that the proposed method of testing for a difference between the conditional variance functions is surprisingly reliable. Interestingly, the test for a difference between the conditional mean functions correctly recognizes equality of the functions also when the conditional variance functions differ.

Acknowledgement

The authors thank the referees for valuable suggestions for improving the first version of the paper.

References

AUESTAD, B. and TJØSTHEIM, D. (1990) Identification of nonlinear time series: first order characterization and order determination. *Biometrika*, **77**, 4, 669-687.

FAN, J. AND GIJBELS, I. (1996) *Local polynomial modelling and its applications.* Chapman & Hall.

GOURIÉROUX, C. (1997) *ARCH models and financial applications.* Springer.

HAFNER, C.M. AND HERWATZ, H. (1999) Testing for autoregressive dynamics under heteroscedasticity. *Discussion Paper No. 7/1999, Sonderforschungsbereich*, **373**, Humboldt University.

HÄRDLE, W. AND TSYBAKOV, A.B. (1997) Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econometrics*, **81**, 223-242.

HOFFMAN, M. (1999) On nonparametric estimation in nonlinear AR(1)-models. *Statist. & Probab. Letters*, **44**, 29-45.

LIERO, H. (1999) Testing heteroscedasticity in nonparametric regression. Preprint of the Inst. of Math., University of Potsdam.

MCKEAGUE, I.W. AND ZHANG M-J. (1994) Identification of nonlinear time series from first order cumulative characteristics. *Ann. Statist.*, **22**, 1, 495-514.

NEUMANN, M.H. AND KREISS, J.P. (1998) Regression-type inference in nonparametric autoregression. *Ann. Statist.*, **26**, 1570-1613.

SHORACK, G.R. AND WELLNER, J.A. (1986) *Empirical Processes with Applications to Statistics.* Wiley.

TJØSTHEIM, D. (1994) Nonlinear time series: A selective review. *Scandinavian J. Statist.*, **21**, 97-130.

TONG, H. (1990) *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press.

TONG, H. (1995) A personal overview of non-linear time series analysis from a chaos perspective, with discussion. *Scandinavian J. Statist.*, **22**, 399-445.