

Kernel smoothed prediction intervals for ARMA models

Klaus Abberger, University of Konstanz, Germany

Abstract

The procedures of estimating prediction intervals for ARMA processes can be divided into model based methods and empirical methods. Model based methods require knowledge of the model and the underlying innovation distribution. Empirical methods are based on the sample forecast errors. In this paper we apply nonparametric quantile regression to the empirical forecast errors using lead time as regressor. With this method there is no need for a distribution assumption. But for the data pattern in this case a double kernel method which allows smoothing in two directions is required. An estimation algorithm is presented and applied to some simulation examples.

Keywords: Forecasting; Prediction intervals; Non normal distributions; Nonparametric estimation; Quantile regression

1 Introduction

In calculating predictions it is often important to associate an assessment of uncertainty together with a point forecast. In this case an interval forecast as well as a point forecast should be given. The intervals can be used to compare forecasts by different models more thoroughly or the intervals might be used to calculate risks associated with the forecast. E.g. in finan-

cial data one is often interested in the behaviour of quantiles in the tails of the return distributions. The “value at risk“ for a certain asset is measured by low quantiles ($\alpha = 0.01$ or $\alpha = 0.05$) of the conditional distribution of the corresponding series of returns.

In 1993 Chris Chatfield published a very impressive review entitled “Calculating Interval Forecasts“. Calculating prediction intervals the author differentiates between model based methods and empirical methods. The derivation of the former assumes complete knowledge of the model including the distribution. Empirical methods make use either of within sample “forecast“ errors or of post sample forecast errors by splitting the past time series in two parts: fit the model to the first part and make predictions of the second part. In his discussion of Chatfield’s paper Tsay analyses a data example and computes prediction intervals by using the empirical distribution of predictive residuals. In his reply Chatfield states: “The second feature of the proposed method (in my view the more important feature) is that the empirical distribution of forecast errors at different horizons are used to get appropriate percentiles to add and subtract from the point forecasts at different horizons. No normality assumption is required, and there is no need to estimate expected mean squared prediction error. This means in particular that the approach can cope with an asymmetric error distribution,.... The drawback of the method is that only 22 predictive residuals are available for each forecast horizon, and this is dangerously small unless some form of smoothing is used“. And later Chatfield writes: “Therefore, although I sympathize with the idea of using the empirical distribution in some way, I think some sort of smoothing must also be used if the results to be judged acceptable“. This kind of smoothing is discussed in the present

paper. Especially when we work with post sample forecasts so that the data series is splitted we look for methods which allow calculating prediction intervals even with small data sets since the most part of the data is required for model selection and estimation. In this case with a small amount of post sample prediction errors we need “some sort of smoothing“ as Chatfield stated.

2 Nonparametric prediction interval estimation

2.1 Notation and motivation

This article adopts the following notation. Let X_t follow the stationary, invertible ARMA(p,q) process

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \sum_{j=0}^q b_j \epsilon_{t-j}, \quad b_0 = 1. \quad (1)$$

ϵ_t denotes the innovation process. Standing at time n , let the k-step ahead point forecast of X_{n+k} be \hat{X}_{n+k} . The ARMA(p,q) process $a(B)X_t = b(B)\epsilon_t$ can be written as an infinite moving average $X_t = c(B)\epsilon_t$ with $c(B) = c_0 + c_1B + c_2B^2 + \dots$ and B the backshift operator. Then the forecast errors are given by

$$e_n(k) = X_{n+k} - \hat{X}_{n+k} = \sum_{j=0}^{k-1} c_j \epsilon_{n+k-j}. \quad (2)$$

The innovations are usually assumed to be a sequence of independent normally distributed random variables with zero mean and constant variance. Even when we drop the normality assumption we usually keep the zero mean and the constant variance assumptions. Hence the variances of the forecast errors are given by

$$var[e_n(k)] = E[e_n^2(k)] = \sigma_\epsilon^2 \sum_{j=0}^{k-1} c_j^2. \quad (3)$$

Assuming normality, prediction intervals can then be constructed based on this expression. Thus an $(1 - \alpha)$ prediction interval is given by

$$\hat{X}_{n+k} \pm z_{\alpha/2} \sqrt{\text{var}[e_n(k)]}, \quad (4)$$

where $z_{\alpha/2}$ denotes the appropriate percentage point of a standard normal distribution. The derivation of this interval assumes complete knowledge of the model including values of c_j and σ_ϵ^2 and the normality. The coefficients usually have to be estimated. However, this approach to prediction interval estimations is used by several of the widely used forecasting packages.

Since we have to estimate the parameters a common procedure is to split the data set. The first part of the data is used to select and estimate the model. The second part of the data is needed to calculate post-sample forecast errors. Because of this data splitting we look for methods which use the data economically. This objective leads us to the idea of smoothing.

For the AR(1) model the variances of the forecast errors are

$$\text{var}[e_n(k)] = \sigma_\epsilon^2 (1 - a_1^{2k}) / (1 - a_1^2) \quad (5)$$

and looking at an ARIMA(0,1,1) model the variances are given by

$$\text{var}[e_n(k)] = \sigma_\epsilon^2 (1 + (k - 1)(1 - b_1)^2). \quad (6)$$

The error variances are functions of k . This motivates the use of nonparametric regression methods with explanatory variable k .

Since calculating interval forecasts is similar to estimate quantiles of the prediction error distribution it is obvious to use nonparametric quantile

regression methods to achieve the kind of smoothing Chatfield demands. An overview about nonparametric quantile regression methods is included in Heiler's (2000) survey on nonparametric time series analysis. A Nadaraya-Watson type method is to estimate the conditional distribution by

$$F_n(y|x) = \frac{\sum_{i=1}^n 1_{(-\infty, y]}(y_i) K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)} \quad (7)$$

with $K()$ an appropriate kernel function, h a bandwidth and 1 the indicator function. From this the empirical quantile function can be derived by

$$q_{n,\alpha}(x) = \inf\{y \in \mathfrak{R} | F_n(y|x) \geq \alpha\}, \quad 0 < \alpha < 1. \quad (8)$$

In the present application the conditioning variable x is similar to the step length k .

Figure 1 shows the usual data pattern we are faced with. Prediction errors are plotted against the step length k . So we have various data points for $k = 1$ which denotes the 1-step-ahead errors. There are further observations at the integers $k = 2, 3, \dots$. One can imagine that in this setting not only smoothing in k is appropriate but also smoothing in the y direction is required. Yu and Jones (1998) use a so called double kernel method which provides smoothing in both directions. An adopted Nadaraya-Watson kernel estimator of the conditional distribution function is

$$\tilde{F}_n(y|x) = \frac{\sum_{i=1}^n K_1\left(\frac{x_i - x}{h_1}\right) F_2\left(\frac{y - y_i}{h_2}\right)}{\sum_{i=1}^n K\left(\frac{x_i - x}{h_1}\right)}, \quad (9)$$

with $F_2(z) = \int_{-\infty}^z K_2(u) du$ the distribution function of the kernel K_2 and a further bandwidth for the y direction denoted h_2 .

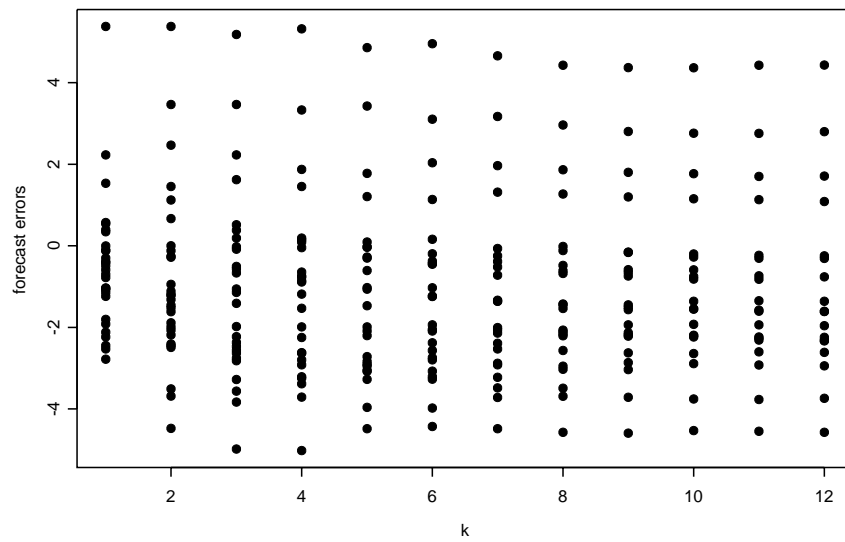


Figure 1: k-step-ahead forecast errors for an AR(1) process with χ_3^2 distributed errors

In the ordinary regression problem smoothing in y is less regarded than smoothing in x direction. The double kernel approach suffers from the disadvantage of the necessity to specify a second bandwidth h_2 as well as the bandwidth h_1 . But even Yu and Jones state that “...it turns out that the estimates are not very sensitive to the value of h_2 ,.... But the choice $h_2 = 0$ is not attractive to us, because it results in a discontinuous conditional quantile estimate (which is inelegant, at least for small samples)“. Unsurprisingly, for estimating interval forecasts smoothing in y is not such trifle, because of the concentration of observations at the integers $k = 1, 2, \dots$

2.2 Algorithm

Since a double kernel method is appropriate for the current problem of prediction interval estimation two bandwidths have to be chosen. A practical procedure for bandwidth choice and α -quantile estimation consists in the following steps.

(i)

The “single kernel“ method (7) in conjunction with the cross-validation technique is applied to chose h_1 , the bandwidth in x direction. This cross-validation method is further discussed in Abberger (1998). The bandwidth which minimizes

$$h_1^{CV} = \min_h \left\{ \sum_{i=1}^n \varphi(y_i - q_{n,\alpha}^{(-i)}(x_i)) \right\} \quad (10)$$

is chosen as cross-validation bandwidth. $q_{n,\alpha}^{(-i)}$ denotes the so called “leave-one-out“ estimator which estimates the conditional quantile at (y_i, x_i) without using the observation (y_i, x_i) . φ is the quantile function $\varphi(u) = \alpha 1_{\{u \geq 0\}}(u)$.

$u + (\alpha - 1)1_{\{u < 0\}}(u) \cdot u$, introduced by Koenker and Basset (1978) in the manner of linear quantile regression. Since the minimum problem is solved for each k separately local bandwidths $h_1(k)$ are calculated.

(ii)

The resulting $h_1(k)$ and the “single kernel“ method is used for a first estimate of the conditional quantile $q_\alpha(k)$.

(iii)

For the choice of the second bandwidth h_2 methods from kernel density estimation are adopted. Monographs about these methods are Wand and Jones (1995) and Silverman (1986). The usual kernel density estimator

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{y - y_i}{h}\right), \quad (11)$$

has the approximate mean squared error

$$MSE[\hat{f}(y)] \approx \frac{1}{4}h^4 f''^2(y) \left[\int t^2 K(t) dt \right]^2 + \frac{1}{nh} f(y) \int K^2(t) dt. \quad (12)$$

In density estimation the mean integrated squared error is calculated. Then procedures which estimate the bandwidth minimizing this criterion are developed. Since in the present application the aim is to have a density estimation at one point namely at the α quantile it is appropriate to ground on the mean squared error instead of the mean integrated squared error. The bandwidth which minimizes (12) is

$$h_*^5(y) = \frac{f(y) \int K^2(t) dt}{n f''^2(y) [\int t^2 K(t) dt]^2} \quad (13)$$

which depends besides on n and the kernel on the unknown quantities $f(y)$ and $f''^2(y)$. These expressions have to be estimated, which also can be

achieved by kernel methods. But therefore another primary bandwidth must be used. This primary bandwidth is also calculated by (13) but assuming normality. So f and f''^2 can be calculated at the theoretical α quantiles of the standard normal distribution. Now in the current application of prediction interval estimation the precise procedure is:

Fix x and take the y observations at x . Then compute the primary bandwidth. Calculate f and f''^2 at the first estimate of the conditional α -quantile resulting from step (ii). With these preparations the final step can be tackled.

(iv)

The final double kernel estimate (9) is calculated with h_1 from step (i) and h_2 calculated by expression (13) replacing n by the number of observations within the x window. Now the conditional quantile can be computed similar to (8).

3 Simulation examples

The developed algorithm is applied to some simulated datasets. The settings in the calculated examples are:

- For all kernel functions the Epanechnikow kernel of the form $3/4(1 - u^2)1_{\{[-1,1]\}}(u)$ is chosen.
- In each repetition of the simulations a time series of length 10000 is generated for the model under consideration. But only the first 50 observations are used to calculate the α prediction interval. The

estimated intervals are applied to the rest of the time series and the observations falling below and above the interval are counted. Thus we can calculate a value $\hat{\alpha}$ for the prediction interval.

- The $\hat{\alpha}$ values computed for the various repetitions are presented graphically for the three methods: normal prediction interval, empirical distribution, kernel smoothing. In each simulation example the amount of repetitions is 500. The quantiles for the empirical distribution are calculated with the S-Plus function `quantile(.)` which lineary interpolates between the order statistics.

Figure 2 shows the results for simulations of the AR(1) model

$$X_t = 0.7X_{t-1} + \epsilon_t \quad (14)$$

and $\epsilon_t \sim \chi_3^2$. At this stage of analysis the problem of model choice and parameter estimation is left out. So the forecast errors are calculated with knowledge of the true model. The 0.05 prediction quantiles are estimated and visualized by the boxplots for step length 1 to 12. In Figure 3 the results for step length $k = 6$ are compared. The boxes of the normal intervals are all below 0.05 indicating that at least 75% of the normal estimates are too low.

Compared with the empirical distribution method kernel smoothing leads to better estimates in the middle and reduces the variations of the estimates. Similar results are obtained for the 0.95 prediction intervals for the same model presented in Figure 4.

The simulation results visualized in Figure 5 ground on the ARMA(1,1) model

$$X_t = 0.7X_{t-1} + 0.5\epsilon_{t-1} + \epsilon_t \quad (15)$$

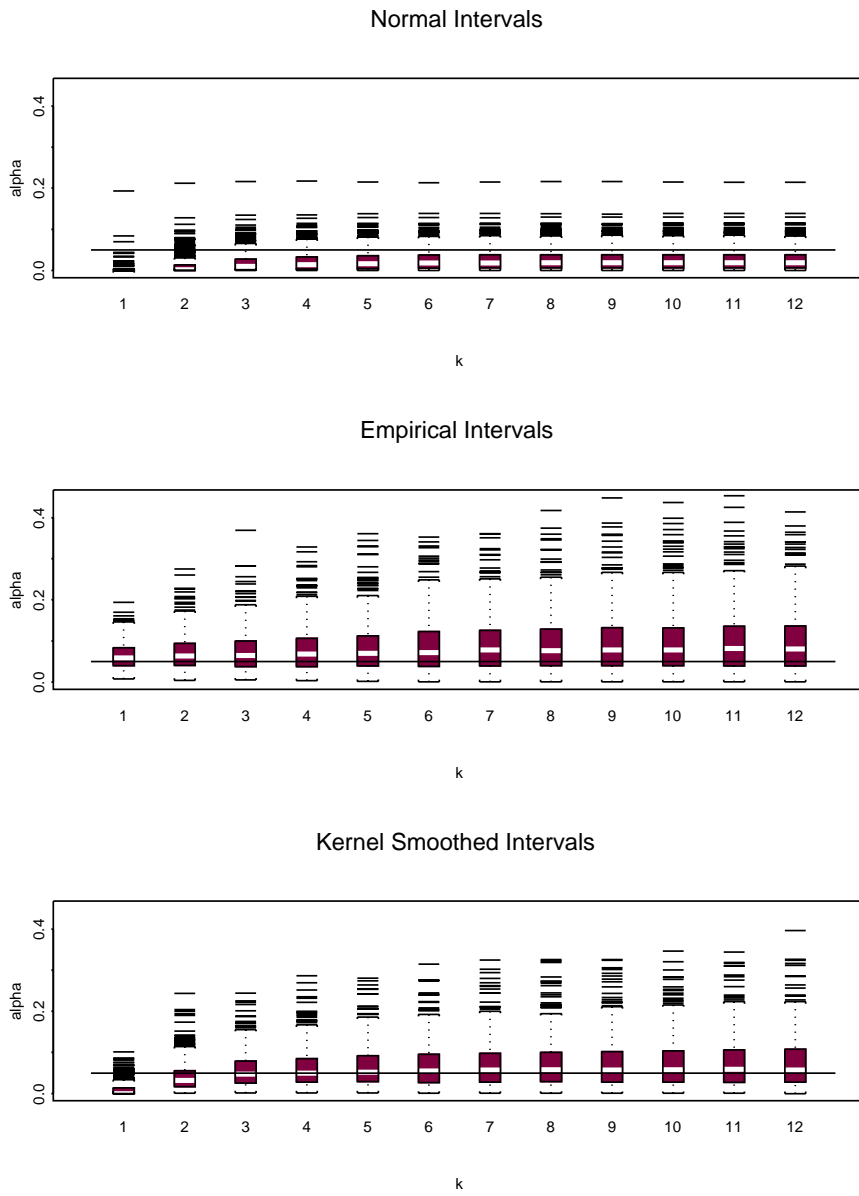


Figure 2: Simulated $\hat{\alpha}$ for the AR(1) model (14) with χ_3^2 distributed errors ($\alpha = 0.05$, 500 repetitions)

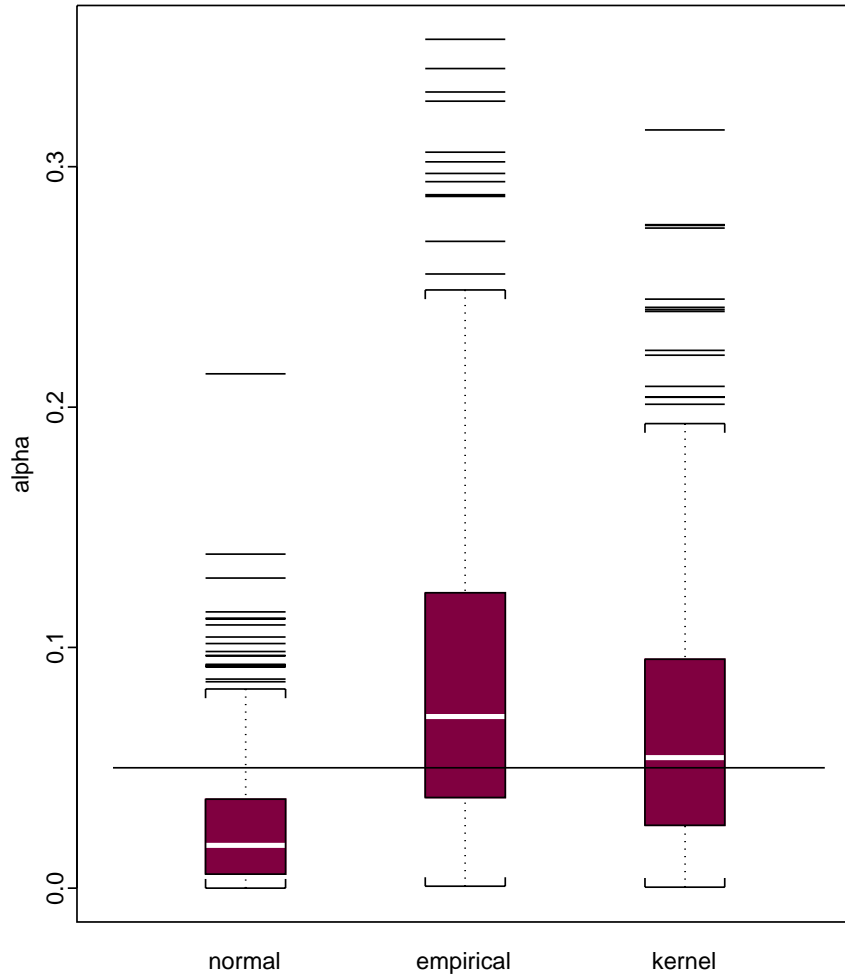


Figure 3: Simulated $\hat{\alpha}$ for the AR(1) model (14) with χ_3^2 distributed errors and step length $k=6$ ($\alpha = 0.05$, 500 repetitions)

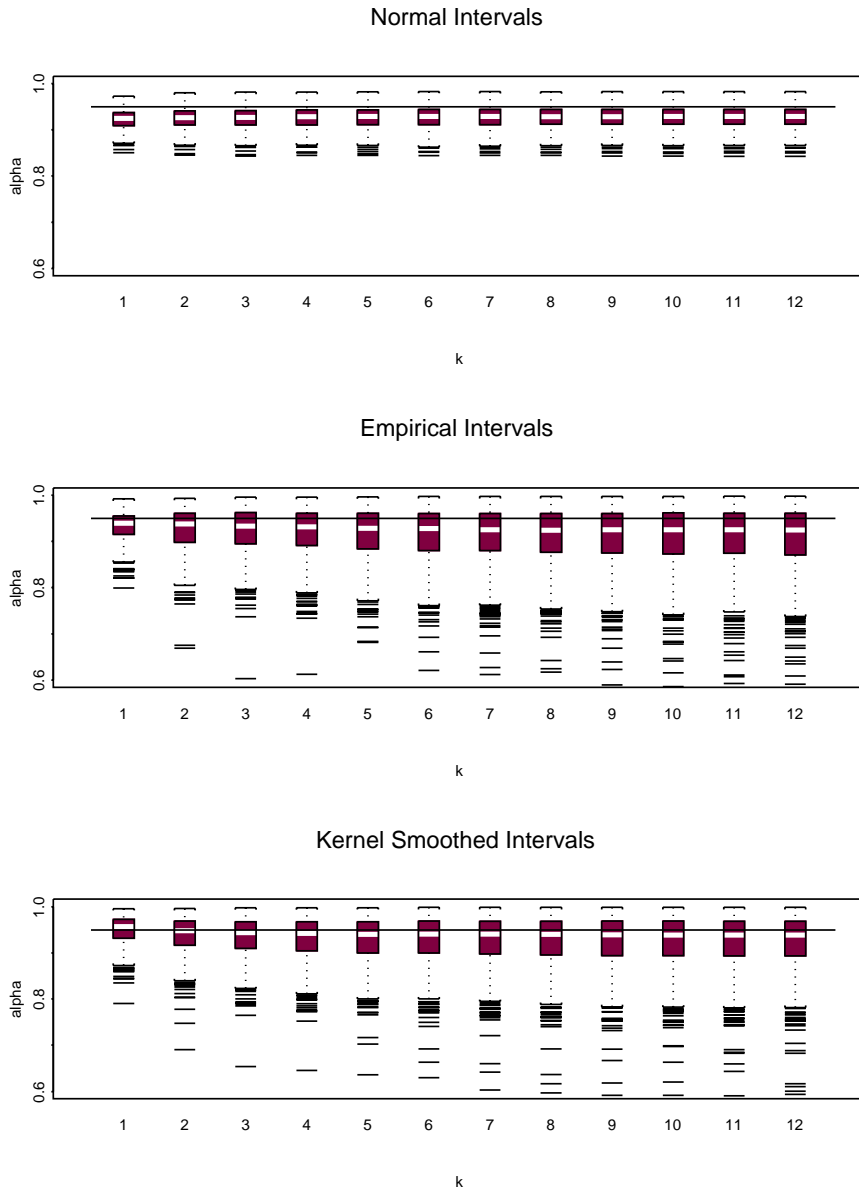


Figure 4: Simulated $\hat{\alpha}$ for the AR(1) model (14) with χ_3^2 distributed errors ($\alpha = 0.95$, 500 repetitions)

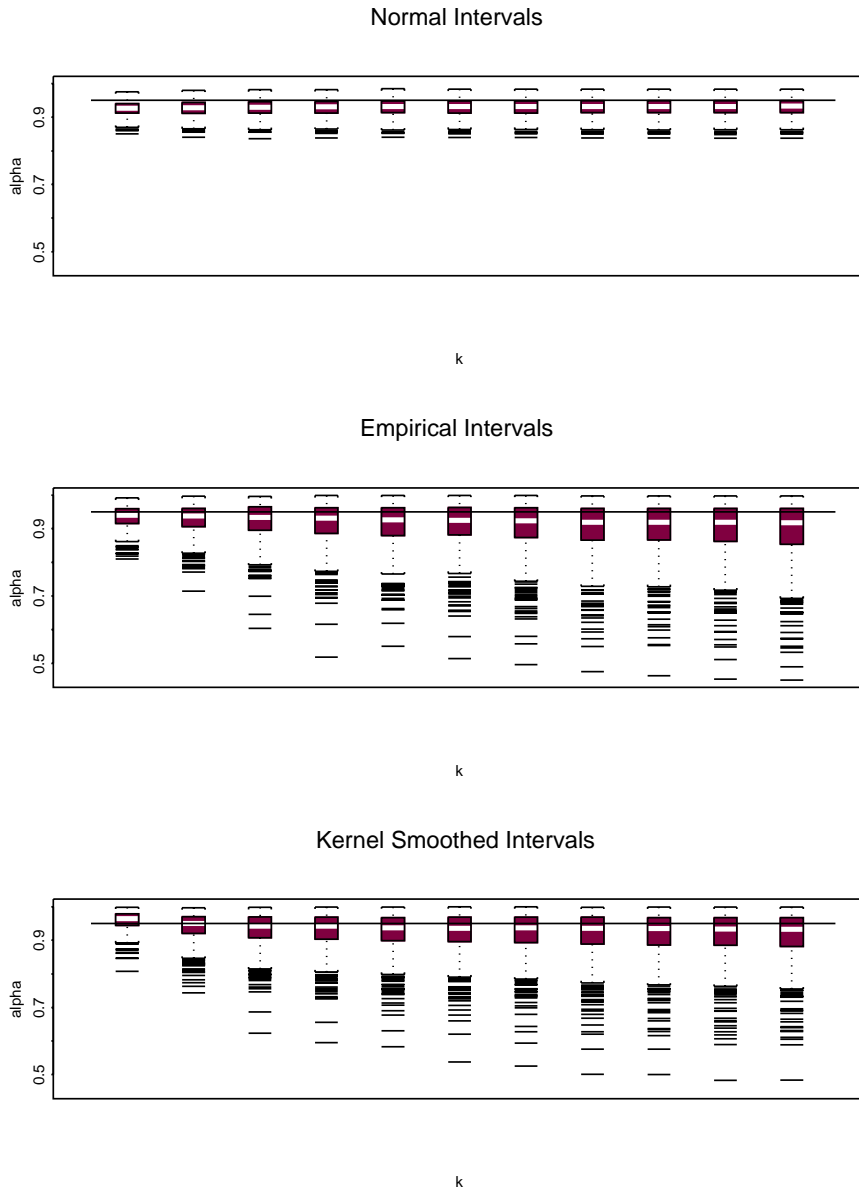


Figure 5: Simulated $\hat{\alpha}$ for the ARMA(1,1) model (15) with χ_3^2 distributed errors ($\alpha = 0.95$, 500 repetitions)

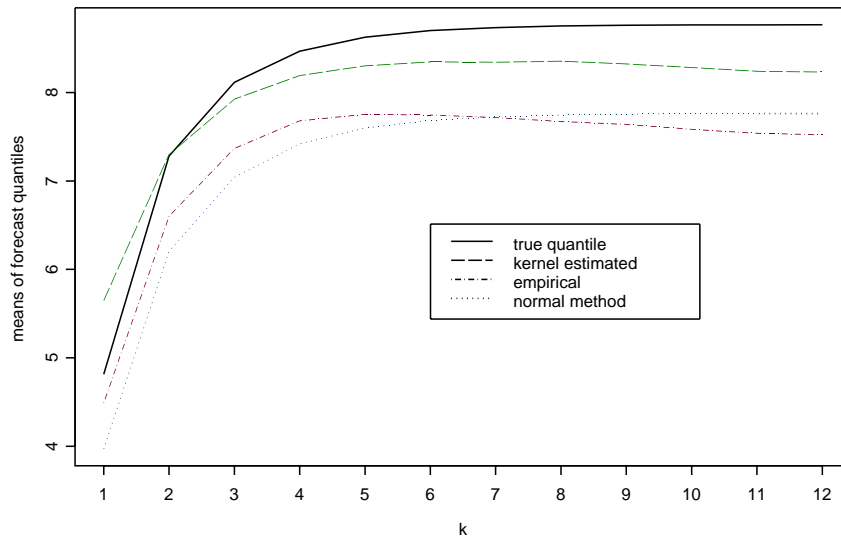


Figure 6: True and means of estimated 0.95 forecast quantiles for the ARMA(1,1) model (15) with χ_3^2 distributed errors ($\alpha = 0.95$, 500 repetitions)

with $\epsilon \sim \chi_3^2$. The calculations are carried out with the knowledge of the underlying model. The plotted results for the 0.95 quantiles show a similar behaviour as the results for the former AR(1) model. Figure 6 shows the true 0.95 quantile and the means of the 500 estimates for each of the three methods. Compared with the normal method, the kernel estimates are in the mean closer to the true quantiles.

Now the problem of parameter estimation is added. Figure 7 presents the results for the 0.95 forecast quantiles based on the ARMA(1,1) model (15). But now the parameters are estimated with the first 50 observations of each series. Comparing the results with the case where the true model is known the statements remain unchanged. The boxes for the normal method are completely below the 0.95 line.

Summarizing the results of these simulations it can be stated that calculating the normal procedure in cases with non normal data gives biased results. Using the empirical forecast errors in connection with nonparametric quantile regression reduces the bias. The described algorithm can be applied in the same way to various distributions and ARMA models of different orders. The price we have to pay is a higher variability of the results. But compared with the proposal of using only the empirical distribution smoothing leads to a reduction of this variability.

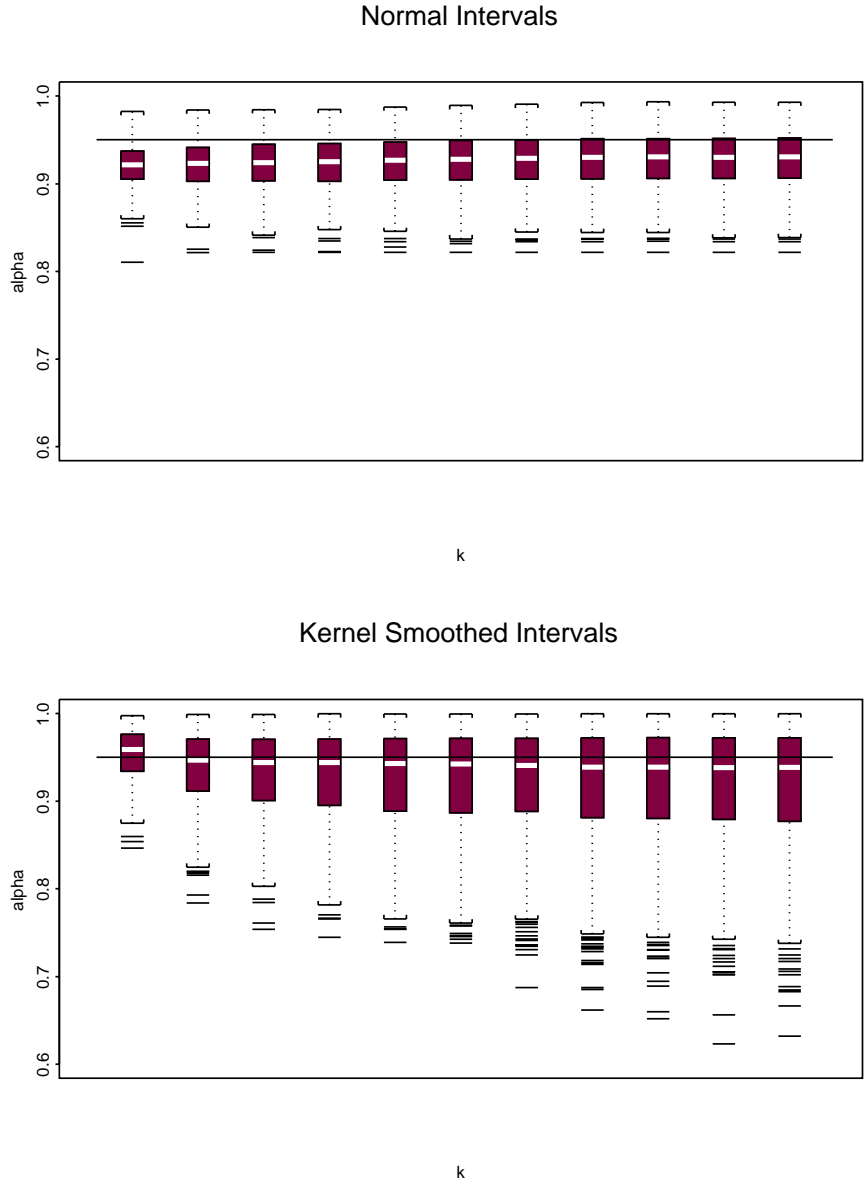


Figure 7: Simulated $\hat{\alpha}$ for the ARMA(1,1) model (15) with χ_3^2 distributed errors and estimated model parameters ($\alpha = 0.95$, 500 repetitions)

4 Literature

Abberger K. (1998): Cross-Validation in Nonparametric Quantile Regression. *Allgemeines Statistisches Archiv*, 82, 149-161.

Chatfield C. (1993): Calculating Interval Forecasts (with discussion). *Journal of Business & Economic Statistics*, 11, 121-144.

Heiler S. (2000): Nonparametric Time Series Analysis. In: *A Course in Time Series Analysis*, edited by D. Pena and G.C. Tiao. John Wiley, London.

Koenker R., Basset G. (1978): Regression Quantiles. *Econometrica*, 46, 33-50.

Silverman B.W. (1986): *Density Estimation*. Chapman & Hall, London.

Tsay R. S. (1993): Adaptive Forecasting (comment to C. Chatfield: Calculating Interval Forecasts). *Journal of Business & Economic Statistics*, 11, 140-142.

Wand M. P., Jones M. C. (1995): *Kernel Smoothing*. Chapman & Hall, London.

Yu K., Jones M. C. (1998): Local Linear Quantile Regression. *Journal of the American Statistical Association*, 93, 228-237.