Model Selection for Nonlinear Time Series

Sebastiano Manzan

Center for Nonlinear Dynamics in Economics and Finance (CeNDEF)

Department of Quantitative Economics, University of Amsterdam

Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

Tel.: +31 20 5254248, Fax: +31 20 5254349

e-mail: s.manzan@uva.nl

Abstract

We investigate the finite-sample performance of model selection criteria for local linear regression by simulation. Similarly to linear regression, the penalization term depends on the number of parameters of the model. In the context of nonparametric regression, we use a suitable quantity to account for the Equivalent Number of Parameters as previously suggested in the literature. We consider the following criteria: Rice T, FPE, AIC, Corrected AIC and GCV. To make results comparable with other data-driven selection criteria we consider also Leave-Out CV. We show that the properties of the penalization schemes are very different for some linear and nonlinear models. Finally, we set up a goodness-of-fit test for linearity based on bootstrap methods. The test has correct size and very high power against the alternatives investigated. Application of the methods proposed to macroeconomic and financial time series shows that there is evidence of nonlinearity.

Keywords: Local Linear Regression, Nonlinear time series, Bootstrap, Linearity Test.

I would like to thank Cees Diks, Cars Hommes and an anonymous referee for useful comments that significantly improved the paper.

1 Introduction

Recently, the application of nonlinear methods to economic and financial data has gathered increasing interest. The seminal work of Hamilton (1989) on markov switching models and the successful application to US GNP data emphasized the importance of considering nonlinear effects. The range of nonlinear models used has widened rapidly and Granger (2001) is a recent survey of the different approaches and results.

A flexible way to model nonlinearities in the data are nonparametric regression techniques. Their main advantage is the adaptability in capturing the dependence structure in the data without relying on a specific parametric family. However, their finite samples properties in the high-dimensional case are very misleading, a situation known in the literature as the "curse of dimensionality". Many dimensionality reduction techniques, such as additivity, have been proposed to circumvent this problem. Fan and Yao (2003) is a recent review of these techniques in a time series framework. Applications of nonparametric regression methods to economic time series are Diebold and Nason (1990) to weekly exchange rates, Mizrach (1992) to daily exchange rates and Pagan and Schwert (1990) to estimate the conditional variance of stock prices.

In this paper, we investigate the performance of selection criteria for nonparametric regression. In a time series context, there are 2 parameters to select: the lag order and the bandwidth. Both of them affect the complexity of the model, that is, the number of parameters used to fit the data. The selection criteria try to balance between goodness-of-fit and the number of parameters. Using a quantity that captures the number of parameters implied by both the lag order and the bandwidth, it is possible to extend the selection criteria frequently used in linear analysis. In particular, we investigate the performance of the selection criteria in moderate samples and for large orders where the "curse of dimensionality" is a relevant issue. We find that most of the criteria perform reasonably well. However, Akaike Information Criteria (AIC) and Final Prediction Error (FPE) perform very poorly because they tend to overfit, that is, to use too many parameters compared to optimal.

The paper is organized as follows: in Section (2) we describe the local linear smoother and the selection criteria we use to choose order and bandwidth. Section (3) shows simulation results concerning various linear and nonlinear autoregressive models. In Section (4) a goodness-of-fit test for linearity is proposed and the appropriateness of the selection criteria is emphasized. Section (5) investigates the presence of nonlinearity in some macroeconomic time series. Finally, Section (6) concludes.

2 The Method

Assume $\{x_t\}_{t=-p+1}^{n}$ is a univariate stationary time series generated by the following nonlinear autoregressive model of order p

$$x_{t+1} = m(X_t) + \epsilon_{t+1},$$

where $m(x)$ is a function of unknown form, $X_t = (x_t, ..., x_{t-p+1})'$ denotes the p-dimensional vector of lagged values of the time series and ϵ_t is an $i.i.d.$ disturbance term with mean 0 and variance σ^2. This general form encompasses the AR model as well as many nonlinear time series models like thresold autoregressive (TAR) and exponential autoregressive (EXPAR). We estimate $m(x)$ using nonparametric regression techniques. In particular, we adopt the local linear approach proposed by Cleveland and Devlin (1988) and Fan and Gijbels (1996). The local linear estimator using nearest-neighbors bandwidth is known in the literature as LOESS and it minimizes

$$\sum_{t=0}^{n-1} \{x_{t+1} - \alpha - \beta(X_t - x)\}^2 K\left(\frac{\| X_t - x \|}{d_k(x)}\right), \tag{1}$$

where $\widehat{m}(x) = \hat{\alpha}$, $K(\cdot)$ is the tricube kernel defined as

$$K(u) = \begin{cases} (1-u^3)^3 & \text{for } 0 \leq u < 1 \\ 0 & \text{otherwise,} \end{cases}$$

$\| \cdot \|$ indicates the euclidean distance and

$$d_k(x) = \begin{cases} \| X_{x(k)} - x \| & \text{for } 0 < h \leq 1 \\ \| X_{x(n)} - x \| \, h^{\frac{1}{p}} & \text{for } h > 1. \end{cases}$$

where $X_{x(k)}$ denotes the k-th nearest neighbor of x, h is the bandwidth and k is the integer part of (hn). In the LOESS approach it is often used a tricube kernel but other kernels might deliver similar results. The smoothing scheme can be described as follows: if X_t is among the k nearest neighbors of the design point x, it receives a positive weight given by the tricube kernel $K(\cdot)$. Otherwise, it receives a null weight. A practical advantage of nearest neighbors bandwidths, compared to fixed bandwidths, is that they deliver more reliable and stable variances of the fit in regions where the data are sparse. If we let the bandwidth $h \to \infty$, we include all the data in the regression and the local linear model approaches the linear AR model.

We consider data-driven (or automatic) methods to select h such as minimizing the RSS (Residuals Sum of Squares)

$$\frac{1}{n}\sum_{t=0}^{n-1} \{x_{t+1} - \widehat{m}_h(X_t)\}^2, \tag{2}$$

The RSS trivially achieves a minimum for $h \to 0$ because it implies $\widehat{m}(X_t) \to x_{t+1}$. An approach to solve this problem is the Leave-One-Out Cross-Validation method, which minimizes the following function

$$CV(h) = \frac{1}{n}\sum_{t=0}^{n-1}\{x_{t+1} - \widehat{m}_{h,-t}(X_t)\}^2, \qquad (3)$$

where $\widehat{m}_{h,-t}(x)$ indicates the fitted value obtained by excluding the t-th observation. In a time series context, Härdle and Vieu (1992) proved the asymptotic optimality of the selection method. An alternative approach consists of multiplying the RSS by a penalization factor that corrects the tendency of h to go to 0. These methods are inspired by selection criteria used for linear models that choose the order p that minimizes

$$SC(p) = \log RSS + \phi(p), \qquad (4)$$

where the first term indicates the goodness-of-fit of the model and the second term penalizes the inclusion of more parameters, measured by p. This approach can be extended to nonparametric regressions because they are linear smoothers. The fitted regression function can be expressed as

$$\widehat{y} = Hy,$$

where $y = (x_1, ..., x_n)'$ and H is the $n \times n$ hat matrix that depends only on lagged values. Similarly to the linear case, we define the number of parameters involved in the regression by

$$\pi = \frac{tr(H)}{n},$$

If the bandwidth tends to ∞ then $tr(H)$ will approximate p. However, for $h \to 0$ it will approach n, the case in which we fit as many parameters as data points. The extension to the nonparametric case of the criterion in (4) is

$$SC(\pi) = \log RSS + \phi(\pi), \qquad (5)$$

where π quantifies the complexity of the model implied by both the choice of the bandwidth, h, and the number of lags, p. Considering more lags and smaller bandwidths, increases π and $\phi(\pi)$ attributes a larger penalization to the goodness of fit measure. In the literature π is called the Equivalent Number of Parameters (ENP) by Cleveland and Devlin (1988), to stress the analogy with the linear regression case.

There are many proposals concerning the form of the penalization function $\phi(\cdot)$. We consider here the most often used:

1. Akaike Information Criteria (AIC): $\phi(\pi) = 2\pi$
2. Corrected AIC (AICC): $\phi(\pi) = \frac{1+\pi}{1-\pi-\frac{2}{n}}$
3. Final Prediction Error (FPE): $\phi(\pi) = \log\left(\frac{1+\pi}{1-\pi}\right)$
4. Generalized Cross-Validation (GCV) : $\phi(\pi) = -2\log(1-\pi)$

5. Rice T (T): $\phi(\pi) = -\log(1 - 2\pi)$

A discussion of these criteria can be found in Härdle (1990). A bias corrected version of AIC, indicated as AICC, has been recently proposed by Hurvich *et al.* (1998). Selection criteria are used in nonparametric regression to select order or bandwidth and in theoretical work these two problems are kept separate. Some references on lag selection using nonparametric regression are Tschernig and Yang (2000), Tjøstheim and Auestad (1994) and Cheng and Tong (1991); for bandwidth selection see Hurvich *et al.* (1998), Yao and Tong (1998) and Härdle and Vieu (1992).

In this paper, we adopt the point of view of an applied analyst that faces the problem of selecting both bandwidth and lag order. Instead of considering them separately, we jointly select these 2 parameters by minimization of Equation (5). In the following section we compare via simulation the finite-sample performances of the selection criteria.

3 Simulation

We simulate 1000 samples of length 100 for linear and nonlinear models. We consider a maximum number of lags of 4 and the bandwidth h is varied from 0.1 to 1 at steps of 0.02. Increasing the order p we consider more lags in the regression function instead of the procedure adopted by Tjøstheim and Auestad (1994) that performs a specification search for the lags to be included in the estimation.

For all simulated models, the noise term ϵ_t is distributed standard normally. We compare the ENP selected by the selection criteria, π^{crit}, to the optimal ENP, π^{opt}, that is selected by minimizing the Average Squared Error (ASE) defined as

$$\frac{1}{n}\sum_{t=0}^{n-1}[\widehat{m}_h(X_t) - m(X_t)]^2 \tag{6}$$

with respect to h, whereas the order p is assumed to be known. The bandwidth selected by minimizing 6, h^{opt}, should be interpreted as the optimal degree of smoothing for the simulated time series. We report also the average (over the simulations) of the ratio of the ASE calculated for the different selection criteria and for the optimal bandwidth.

AR(1) The AR(1) process is defined as

$$x_{t+1} = 0.6x_t + \epsilon_{t+1}.$$

The top plot in Figure (1) shows the smoothed density of the log-ratio of the ENP selected by the criterion, π^{crit}, to the optimal one, π^{opt}. The plot gives information on the performance of the criteria in selecting h and p with respect to the optimal choices. Positive values of the log-ratio, $\log(\pi^{crit}/\pi^{opt})$, imply that the criterion selects more parameters than optimal, a situation that we indicate as overfitting. However, if the log-ratio is negative, the criterion is affected by underfitting, that is, it selects a parsimonious model compared to optimal. The chosen ENP, π^{crit}, may be higher than the optimal value (overfitting) because the selected bandwidth, h^{crit}, is too small and/or because the criterion selected a large number of lags, p^{crit}. The phenomenon of small bandwidth compared to optimal is called undersmoothing. The interpretation is similar to the linear case, with the only difference of the additional role played by the bandwidth in increasing the (equivalent) number of parameters in the regression. The middle plot in Figure (1) shows the density of the selected bandwidth for the different selection criteria. Finally, the Table reports the frequency of selection of the order and the average ASE ratio.

Figure (1) here

Figure (1) clearly suggests that AIC and FPE have a tendency to overfit. The distribution of $\log(\pi^{crit}/\pi^{opt})$ for AIC and FPE is bimodal with one mode around zero and the other in the positive region. They select too frequently an ENP that is larger than optimal. The distributions for the other criteria are centered around zero but are skewed to the right.

The reason for the bad performance of AIC and FPE are clear from the distribution of h^{crit} and p^{crit}. These criteria select often bandwidth that are small compared to the smoothing required by the model. Instead, the other criteria correctly select most often the largest bandwidth of 1 that corresponds to give an equal weight to all the observations. The Table shows the frequency of selection of the order. AIC is more likely to select order 4 (49% of the times) than the true order 1 (36%). A similar problem occurs also for FPE, that selects order 4 in 20% of the simulations. T and AICC select the true order in approximately 77% of the cases while GCV and CV are correct in 72 and 68%, respectively. In term of ASE it is also clear the poor performance of AIC and FPE that score the highest error. Instead, the lowest error is achieved by AICC and by Rice T criteria.

Summarizing, for the AR(1) model we found that severe overfitting occurs for AIC and FPE. It arises both because of undersmoothing (the bandwidth selected is too small) and too frequent selection of large orders. The best criteria are AICC and T that perform reasonably well both in the selection of the bandwidth and order. The skewness in the distribution of $\log(\pi_{crit}/\pi_{opt})$ is probably due to a small sample effect that disappears for larger samples.

AR(1)-GARCH(1,1) If we allow for heteroscedasticity in the disturbance term of the GARCH(1,1) type, the AR(1) process becomes

$$x_{t+1} = 0.6x_t + \epsilon_{t+1}, \qquad \epsilon_{t+1} \sim N(0, h_{t+1})$$

$$h_{t+1} = 1 + 0.1\epsilon_t^2 + 0.8h_t.$$

The heteroscedasticity in the innovations does not change the previous analysis for the AR(1) case. Figure (2) shows that AIC and FPE are affected by overfitting that can be partly explained by undersmoothing, that is, the selection of bandwidths that are smaller than optimal. The other criteria are slightly skewed to the right as in the case with homoscedastic innovations. This is probably due to the effect of the selection of p^{crit}.

Figure (2) here

The table confirms that AIC and FPE select often the largest order. AICC and T perform better and select the correct order in approximately 75%. This suggests that AIC and FPE do not penalize enough the selection of a large number of parameters and choose too often a small bandwidth or a large order.

TAR(1) The model is

$$x_{t+1} = -0.5x_t I(x_t \leq 1) + 0.6x_t I(x_t > 1) + \epsilon_{t+1}.$$

Figure (3) confirms that also for this nonlinear model AIC and FPE tend to overfit. The distributions of $\log(\pi^{crit}/\pi^{opt})$ for these criteria are bimodal and a large part lays in the positive region.

Figure (3) here

The other criteria have a distribution centered around zero but with a slight skewness in the left tail that indicates a tendency to underfit. The distribution of h^{crit} shows that AICC, GCV and T are likely to choose a bandwidth in the range 0.3 to 0.8 while AIC and FPE most often select the smallest value in the range. Also in this case it is clear that undersmoothing seriously affects the performances of the criteria.

The table shows that AIC selects often larger orders than optimal: lag 4 is selected 40% of the times while the first order (the true order) in 49% of the cases. FPE behaves better selecting in 74% of the simulations the correct order. The other criteria identify correctly the true lag more than 90% of the times. We conclude that, for AIC, the order selection contributes to the bimodality of the distribution of $\log(\pi^{crit}/\pi^{opt})$. This is less the case for FPE and the other criteria that have lower mean ASE. T and AICC are again the best in term of this measure of error.

EXPAR(1) The model is

$$x_{t+1} = \{0.5 + 1.3\exp(-0.5x_t^2)\}x_t + \epsilon_{t+1}.$$

The model has nonlinear dependence in the first lag.

Figure (4) here

Figure (4) shows that also this model is characterized by the bimodality of the distribution of $\log(\pi^{crit}/\pi^{opt})$ when using AIC and FPE selection criteria. The distributions for the remaining criteria are centered around zero and are not significantly skewed. The density of the bandwidth for most criteria are centered around 0.55 while for AIC and FPE there is significant undersmoothing. AICC and T perform best in term of mean ASE. They select the correct order in approximately 90% of the simulations. Instead, AIC correctly select the right order only 46% of the times. The weak penalization of AIC and FPE is also clear from this Table because they have a tendency to oversmooth (that is, to select orders too large).

EXPAR(2) The model is

$$x_{t+1} = \{0.5 + 1.3\exp(-0.5x_{t-1}^2)\}x_t + \epsilon_{t+1},$$

that has the same structure of the previous model but the dependence occurring on the first 2 lags. This allows to test the performance of the selection criteria in identifying the correct order in the nonlinear case.

Figure (5) here

The top plot of Figure (5) shows that also for this model AIC and FPE are affected by overfitting. In this case, also GCV and CV are slightly biased toward using too many parameters. T and AICC seem to behave correctly. In term of the selected bandwidth, it is clear that AIC and FPE severely undersmooth because of the weak penalization used. GCV confirms that it has a slight propensity to undersmooth while T and AICC do not show any problem related to the choice of h. Most of the criteria correctly identify the second order dependence in the simulated time series. T and AICC select order two in 95% of the cases, GCV and CV in around 86%, FPE in 68% and AIC in only 45%. Again, AIC selects the true order and order four with approximately the same frequency. In term of mean ASE, the best performing are T and AICC while AIC and FPE are the worst.

4 A Test for Linearity

We set up a test for linearity based on the comparison of the goodness-of-fit of the parametric and nonparametric regression. The specification test can be interpreted as a Generalized Likelihood Ratio Test in the sense of Fan *et al.* (2001). Recently, Cai *et al.* (2000) and Lee and Ullah (2002) adopted a similar testing strategy. The null hypothesis of the test is

$$\mathrm{H}_0 : E(x_{t+1}|X_t) = X_t'\theta$$

$$\mathrm{H}_1 : E(x_{t+1}|X_t) = m(X_t),$$

where θ is the coefficients vector of the AR(p) model and $m(\cdot)$ is a nonlinear function. Let RSS^P and RSS^{NP} denote the parametric and nonparametric RSS, respectively, defined as

$$RSS^P = \frac{1}{n}\sum_{t=0}^{n-1}\{x_{t+1} - X_t\hat{\theta}\}^2 = \frac{1}{n}\sum_{t=0}^{n-1}\{\hat{u}_{t+1}^P\}^2,$$
$$RSS^{NP} = \frac{1}{n}\sum_{t=0}^{n-1}\{x_{t+1} - \widehat{m}_h(X_t)\}^2 = \frac{1}{n}\sum_{t=0}^{n-1}\{\hat{u}_{t+1}^{NP}\}^2.$$

We use the same selection criteria in order to choose the lag order (and the bandwidth) of the parametric and nonparametric regressions.

The test statistic is defined as

$$B = \frac{RSS^P - RSS^{NP}}{RSS^{NP}}. \tag{7}$$

To evaluate the significance of the test statistic we use bootstrap methods. We account for the heteroscedasticity that is observed in many economic time series by resampling the residuals of the nonparametric regression using the wild bootstrap approach proposed by Liu (1988). The test procedure is as follows:

1. calculate the test statistic, B, for the original time series.

2. generate bootstrap innovations, u_{t+1}^*, from the centered fitted residuals of the nonparametric regression, $\tilde{u}_{t+1} = \hat{u}_{t+1}^{NP} - \frac{1}{n}\sum_{t=0}^{n-1}\hat{u}_{t+1}^{NP}$, as

$$u_{t+1}^* = \begin{cases} a\tilde{u}_{t+1} & \text{with probability } p = (\sqrt{5}+1)/(2\sqrt{5}) \\ b\tilde{u}_{t+1} & \text{with probability } 1-p, \end{cases}$$

   where $a = -(\sqrt{5}-1)/2$ and $b = (\sqrt{5}+1)/2$.

3. generate iteratively a new bootstrap time series as

$$x_{t+1}^* = X_t^*\hat{\theta} + u_{t+1}^*,$$

   where X_t^* is a p-dimensional vector of lagged values.

5. calculate the test statistic B^* on the bootstrap time series using the same orders and bandwidth selected for the original time series[1].

6. Repeat steps (1) and (2) M times.

7. Calculate the one sided p-value as

$$\hat{p} = \frac{1 + \#\{B^* > B\}}{1 + M},$$

   and reject if $\hat{p} < \alpha$, where α denotes the significance level.

[1]We avoid the selection of the bandwidth and the order for each of the bootstrap replications for computational reasons. We think that this simplyfying assumption will not have dramatic effects on our results.

We perform a one-sided test because deviations from the null hypothesis are expected to occur for positive values of the test statistic. The consistency of the bootstrap procedure derive from the fact that the residuals of the nonparametric regression are always consistent both under the null and the alternative. See Cai *et al.* (2000) for details.

Table (1) here

We simulate 1000 samples of length 100 and the number of bootstrap replications set to 199 for the models examined in the previous Section. Given the results in the previous section, we investigate the size and power properties of the test only for the T, AICC, GCV and CV. Table (1) shows that the test has good size properties also in the presence of heteroscedasticity that is known to cause size distortions. Under the alternative examined, the test has high power: for the AICC selection criteria it is 96% against TAR(1), 85% against EXPAR with dependence in the first lag and 91% when the dependence is in the second lag.

5 Empirical Applications

The properties of the test for linearity suggest that AICC, T and GCV have reasonable power against a wide range of dependence structures, such as linear and nonlinear models. We apply the previous methods to investigate the presence of nonlinearity in U.S. macroeconomic and financial time series. In the order selection we extend the search up to 6 lags. For all series, we test the growth rate of the variable.

We plot the selection criterion, $SC(\pi)$ in Equation (5), against h and for p given. It allows us to qualitatively evaluate the nonlinearity in the data by comparing the error curve implied by the parametric and nonparametric regression. We normalize the criterion by the log of the standard deviation of the time series: if it has a value close to 1, there is no evidence for (linear or nonlinear) dependence in the time series. On the other hand, there is strong evidence of dependence when the normalized $SC(\pi)$ is smaller than 1. We plot only the selection criterion function for AICC. The graphical evidence of nonlinearity contained in the error curves is evaluated statistically using the linearity test.

We analyzed the following time series:

GNP GNP (Gross National Product) in real terms and seasonally adjusted from the first quarter of 1947 to the third quarter of 2001. Some nonlinear models that have been proposed for this time series are TAR models by Potter (1995) and Tiao and Tsay (1994) and markov-switching models by Hamilton (1989). The common feature of these models is that they explain the time series in term of transition (deterministic or stochastic) between different regime (expansion and recession).

Figure (6) here

The plot in Figure (6) compares the $SC(\pi)$ of the nonparametric regression to that of a linear AR model. The normalized error curve of the nonparametric regression using AICC or T criteria achieves its minimum at $h^{crit} = 0.86$ and the goodness-of-fit test rejects at 10% the null hypothesis of linearity. T and AICC give a similar answer selecting a moderate bandwidth and pointing to dependence in the first 5 lags. However, GCV selects a smaller bandwidth and a higher lag order. It has also a smaller p-value.

Industrial Production from the first quarter of 1947 to the end of 2001. The data are seasonally adjusted.

Figure (7) here

In Figure (7) it is clear that the nonparametric regression lowers significantly the error compared to the linear regression. T and AICC select lags up to 5 and $h^{crit} = 0.74$. GCV selects a smaller bandwidth whereas CV a higher one. The test for linearity strongly rejects the null hypothesis for all the selection criteria. Hence, we can conclude that for US industrial production there is robust evidence of nonlinear structure.

Unemployment Rate from the first quarter of 1947 to the last quarter of 2001. The data are seasonally adjusted. For previous nonlinear analysis of this time series see Montgomery *et al.* (1998). Figure (8) shows that the selection criterion curve achieves a much lower normalized error compared to the AR model.

Figure (8) here

The minimum for AICC and T occurs at $h^{crit} = 0.66$ and $p^{crit} = 6$ while for GCV the optimal bandwidth is 0.56. The linearity test rejects at 1% significance level the null hypothesis. Hence, the smaller error achieved by the nonparametric regression in the plot of $SC(\pi)$ is statistically significant. For this time series CV gives results different from the selection criteria: it selects a large value for the bandwidth such that the nonparametric fit approximate very closely the linear one. Consequently, the linearity test does not reject the null.

T-bills Monthly 3-months Treasury Bills interest rates from 1950 to the end of 2001. The plot in Figure (9) shows that the normalized criterion of the linear model is very close to 1.

Figure (9) here

The nonparametric regression improves consistently the fit by achieving a much lower error. The selected bandwidth (for $p^{crit} = 6$) is 0.70 for T, 0.66 for AICC and 0.32 for GCV. There is significant evidence to reject linearity for the returns on the Treasury Bills: the test for linearity rejects for all three criteria used. For this series, a linear model performs poorly but a nonparametric approach is able to capture the significant nonlinear structure in the data.

S&P 500 S&P500 Index from the first quarter of 1947 to the third quarter of 2001.

Figure (10) here

The selection criterion curve for both the parametric and nonparametric regressions are very close to each other and to 1. A qualitative interpretation of the plot suggests no evidence to reject linearity. All the criteria select the same order and bandwidth. The p-values of the linearity test are equal to 0.16. Hence, we can conclude that there is no evidence to reject linearity. The test shows that there is no evidence of nonlinear dependence in the conditional mean of the returns on the S&P500 index.

**DM/$** DM/$ exchange rate from the first quarter of 1974 to the end of 2001. The plot in Figure (11) shows that the $SC(\pi)$ achieves a minimum at the largest bandwidth of 1.

Figure (11) here

The value of the normalized criterion is close to the linear criterion and close to 1. Thus, also for this series we should expect weak dependence. The linearity test suggests that for T and AICC there is no evidence to reject the null hypothesis. Instead, GCV achieves a minimum for $h^{crit} = 0.82$ and a p-value of 0.09. The simulation study indicates that T and AICC have better finite sample properties compared to GCV. Hence, we attribute more relevance to the results of T and AICC and conclude that there is no evidence of nonlinear dynamics for this series.

**Yen/$** Yen/$ exchange rate for the same period of DM/$.

Figure (12) here

The normalized $SC(\pi)$ curve in Figure (12) suggests that also for this exchange rate there is no evidence of linearity. All the criteria agree in selecting a bandwidth of 1 and order 1. The p-values of the test statistic are large enough to conclude that we cannot reject the null hypothesis of linearity.

6 Conclusion

In this paper we show that nonparametric regression is a useful tool to detect nonlinearities in time series. As a first step in the analysis of time series, it guarantees flexibility in the type of dependence it captures. In addition, the simulation results indicate that they are reasonably accurate in moderate samples. We show that selecting bandwidth and lag order can be reliably carried out by using AICC or T criteria, which are not affected by significant problems of overfitting, in contrast to AIC and FPE. In addition, a goodness-of-fit test based on the comparison of parametric and nonparametric regression is a powerful test to detect deviations from the linearity assumption. If the linearity test rejects the null hypothesis, further steps could be to apply a battery of parametric linearity tests to identify which nonlinear dependence structure is more suitable to explain the time series dynamics. The application of nonparametric autoregression to macroeconomic time series shows that there is statistical evidence of nonlinearity for some of them. For growth rates of US real GNP, Industrial Production and Unemployment Rate the null hypothesis of linearity is rejected when using T and AICC as selection criteria. However, returns of financial time series do not show evidence to reject linearity with the exception of the US T-Bills for which there is strong evidence of nonlinearity.

References

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, **95,** 941–956.

Cheng, B. and Tong, H. (1991). On consistent nonparametric order determination and chaos. *Journal of the Royal Statistical Society* B, **54,** number 2, 427–449.

Cleveland, W.S. and Devlin, S.J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, **83,** number 403, 596–610.

Diebold, F.X. and Nason, J.A. (1990). Nonparametric exchange rate prediction? *Journal of International Economics*, **28,** 315–332.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series.* New York: Springer-Verlag.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and wilks phenomenon. *Annals of Statistics*, **29,** 153–193.

Granger, C. W. (2001). Overview of nonlinear macroeconometric empirical models. *Macroeconomic Dynamics*, **5,** number 4, 466–481.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57,** 357–384.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press.

Härdle, W. and Vieu, P. (1992). Kernel regression smoothing of time series. *Journal of Time Series Analysis*, **13,** number 3, 209–224.

Hurvich, C.M., Simonoff, J.S. and Tsai, C.L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike Information Criterion. *Journal of the Royal Statistical Society* B, **60,** 271–293.

Lee, T.-H. and Ullah, A. (2002). Nonparametric bootstrap tests for neglected nonlinearity in time series regression models. *Journal of Nonparametric Statistics* , forthcoming.

Liu, R.Y. (1988). Bootstrap procedure under some non-i.i.d. models. *Annals of Statistics*, **16,** 1696–1708.

Mizrach, B. (1992). Multivariate nearest neighbors forecast of ems exchange rates. *Journal of Applied Econometrics*, **7,** S151–S163.

Montgomery, A.L., Zarnowitz, V., Tsay, R.S. and Tiao, G.C. (1998). Forecasting the U.S. unemployment rate. *Journal of the American Statistical Association*, **93,** number 442, 478–493.

Pagan, A.R. and Schwert, W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics*, **45,** 267–290.

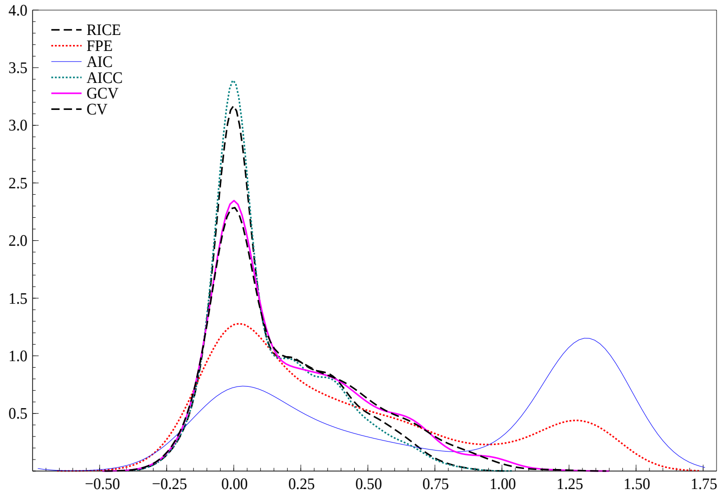
Potter, S. (1995). A nonlinear approach to US GNP. *Journal of Applied Econometrics*, **10,** 109–125.

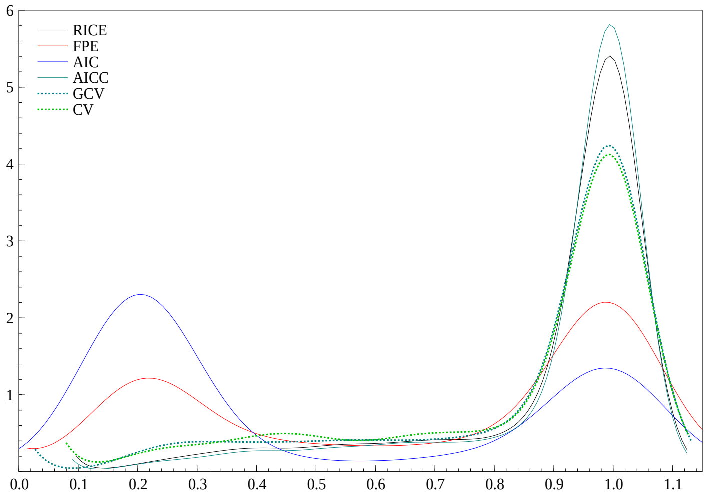
Tiao, G.C. and Tsay, R.S. (1994). Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting*, **13,** 109.

Tjøstheim, D. and Auestad, B.H. (1994). Nonparametric identification of nonlinear time series: Selecting significant lags. *Journal of the American Statistical Association*, **89,** number 428, 1410–1419.

Tschernig, R. and Yang, L. (2000). Nonparametric lag selection for time series. *Journal of Time Series Analysis*, **21,** number 4, 457–487.

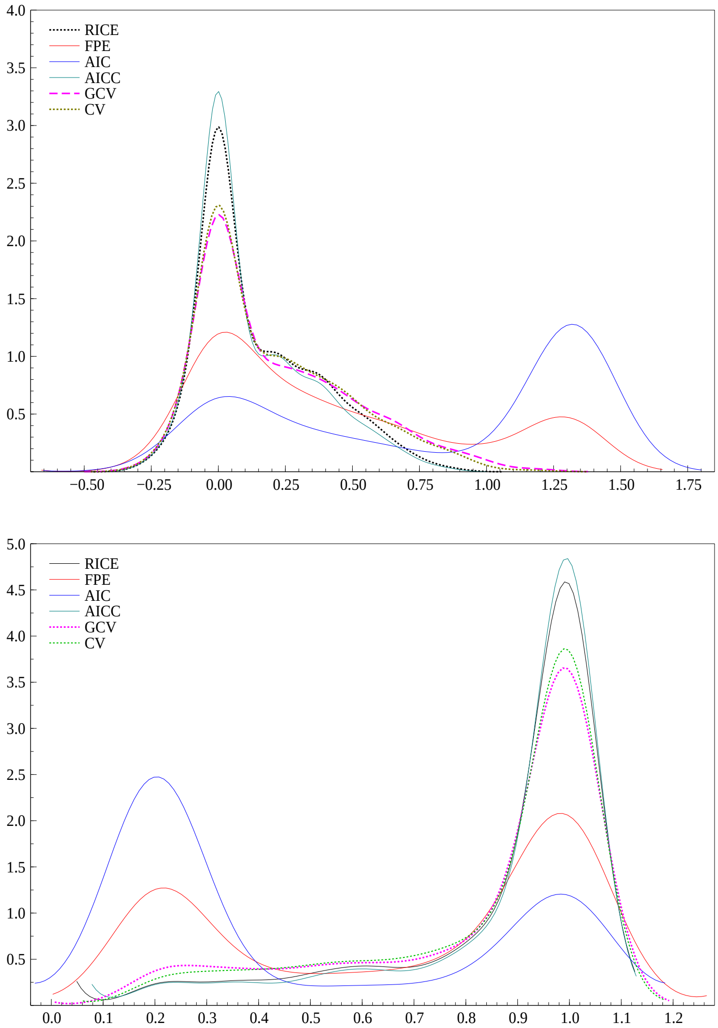
Yao, Q. and Tong, H. (1998). Cross-validatory bandwidth selections for regression estimation based on dependent data. *Journal of Statistical Planning and Inference*, **68,** 387–415.





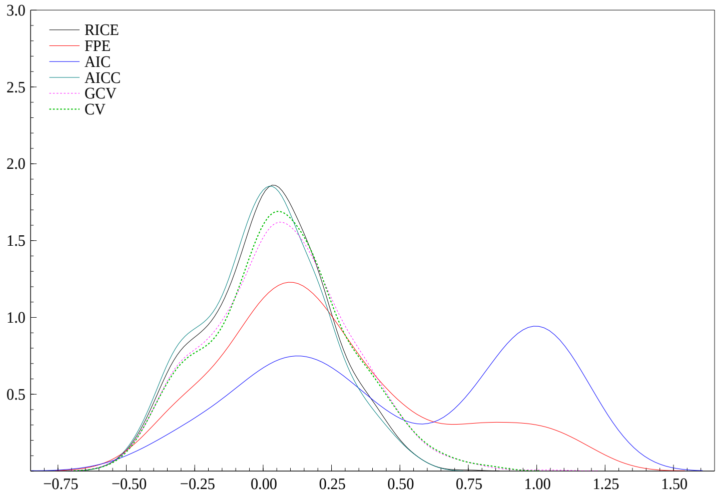
	T	FPE	AIC	$AICC$	GCV	CV
$p = 1$	0.77	0.55	0.36	0.78	0.72	0.68
$p = 2$	0.15	0.15	0.09	0.15	0.17	0.17
$p = 3$	0.05	0.10	0.06	0.05	0.07	0.09
$p = 4$	0.03	0.20	0.49	0.02	0.04	0.06
$\overline{ASE}$	2.49	5.62	8.43	2.43	3.10	2.86

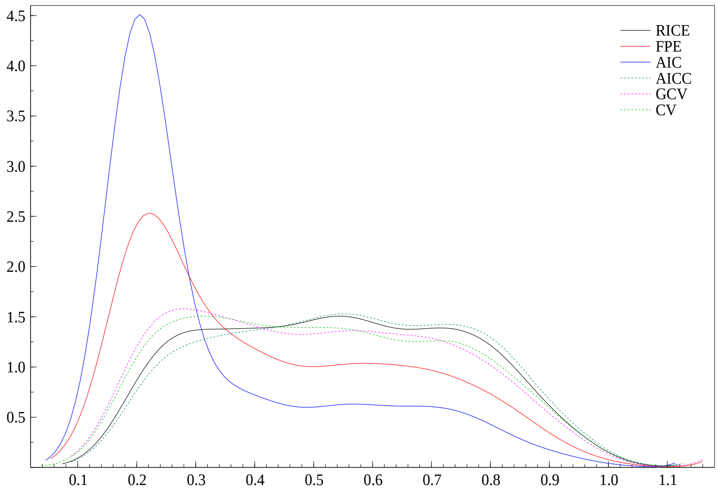
Figure 1: *AR(1) model: smoothed densities of the log of* π^{crit}/π^{opt} *(top), smoothed densities of* h^{crit} *(middle) and mean ASE ratio and frequency of selection of the order (bottom).*



	T	FPE	AIC	$AICC$	GCV	CV
$p=1$	0.74	0.55	0.34	0.75	0.70	0.68
$p=2$	0.16	0.18	0.11	0.16	0.17	0.18
$p=3$	0.08	0.11	0.08	0.07	0.09	0.09
$p=4$	0.02	0.16	0.47	0.02	0.04	0.05
$\overline{ASE}$	3.00	6.06	8.31	2.89	3.31	3.27

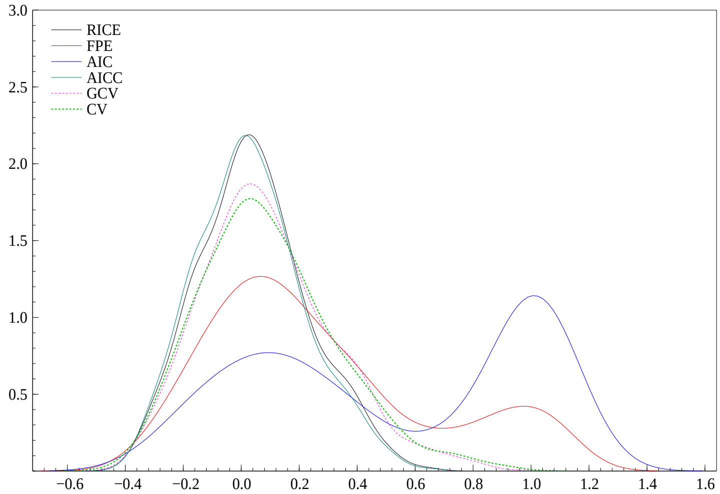
Figure 2: *AR(1)-GARCH(1,1) model: smoothed densities of the log of* π^{crit}/π^{opt} *(top), smoothed densities of* h^{crit} *(middle) and mean ASE ratio and frequency of selection of the order (bottom).*

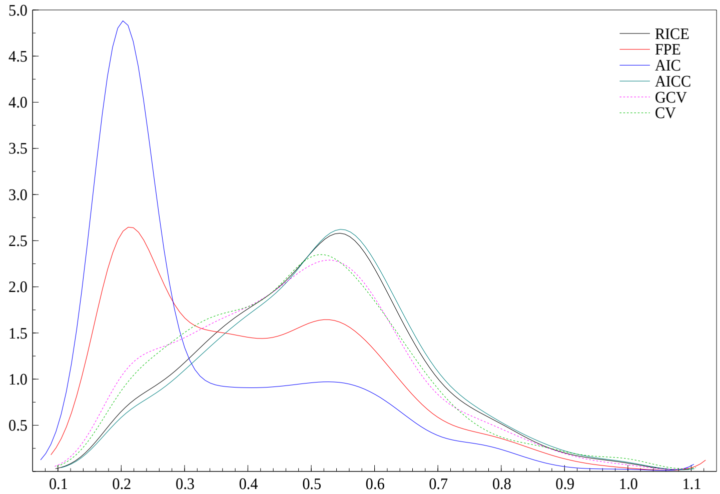




	T	FPE	AIC	$AICC$	GCV	CV
$p = 1$	0.94	0.74	0.49	0.94	0.90	0.88
$p = 2$	0.06	0.10	0.06	0.06	0.08	0.09
$p = 3$	0.00	0.05	0.05	0.00	0.02	0.01
$p = 4$	0.00	0.11	0.40	0.00	0.00	0.01
$\overline{ASE}$	1.31	1.42	1.52	1.31	1.35	1.34

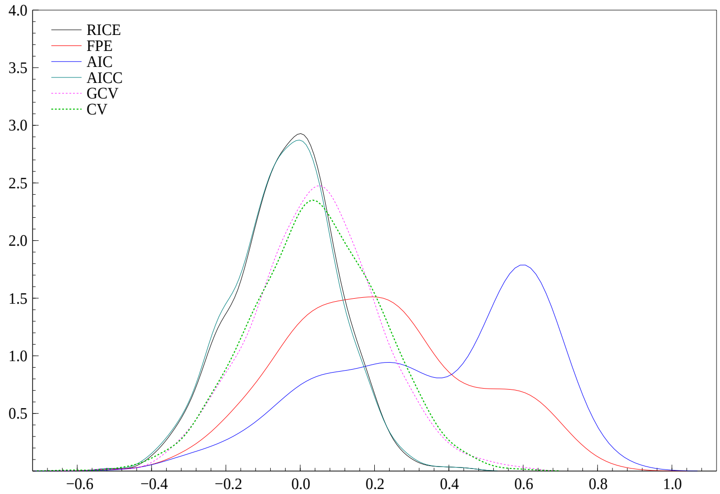
Figure 3: *TAR(1) model: smoothed densities of the log of π^{crit}/π^{opt} (top), smoothed densities of h^{crit} (middle) and mean ASE ratio and frequency of selection of the order (bottom).*

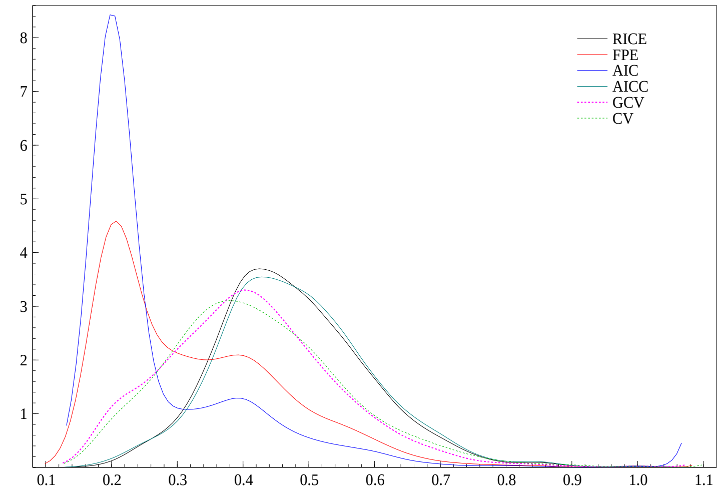




	T	FPE	AIC	$AICC$	GCV	CV
$p = 1$	0.89	0.68	0.46	0.91	0.84	0.82
$p = 2$	0.09	0.14	0.09	0.08	0.12	0.14
$p = 3$	0.01	0.06	0.05	0.01	0.03	0.03
$p = 4$	0.01	0.12	0.40	0.00	0.01	0.01
$\overline{ASE}$	1.43	1.69	1.91	1.42	1.47	1.47

Figure 4: *EXPAR1 model: smoothed densities of the log of* π^{crit}/π^{opt} *(top), smoothed densities of* h^{crit} *(middle) and mean ASE ratio and frequency of selection of the order (bottom).*



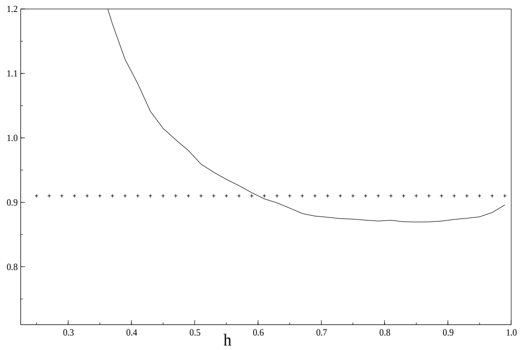


	T	FPE	AIC	$AICC$	GCV	CV
$p = 1$	0.00	0.00	0.00	0.00	0.00	0.00
$p = 2$	0.95	0.68	0.45	0.96	0.89	0.86
$p = 3$	0.05	0.16	0.12	0.04	0.10	0.12
$p = 4$	0.00	0.16	0.43	0.00	0.01	0.02
$\overline{ASE}$	1.13	1.30	1.39	1.14	1.17	1.16

Figure 5: *EXPAR2 model: smoothed densities of the log of* π^{crit}/π^{opt} *(top), smoothed densities of* h^{crit} *(middle) and mean ASE ratio and frequency of selection of the order (bottom).*

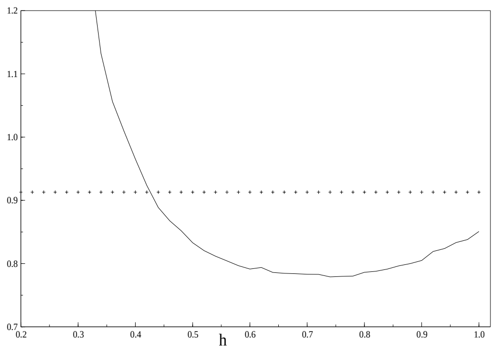
	T	$AICC$	GCV	CV
$AR(1)$	0.06	0.06	0.07	0.06
$AR(1) - GARCH(1,1)$	0.07	0.07	0.08	0.08
$TAR(1)$	0.96	0.96	0.96	0.96
$EXPAR(1)$	0.84	0.85	0.85	0.83
$EXPAR(2)$	0.91	0.91	0.95	0.95

Table 1: *Frequency of rejection of the null hypothesis of linearity at 5% significance level.*



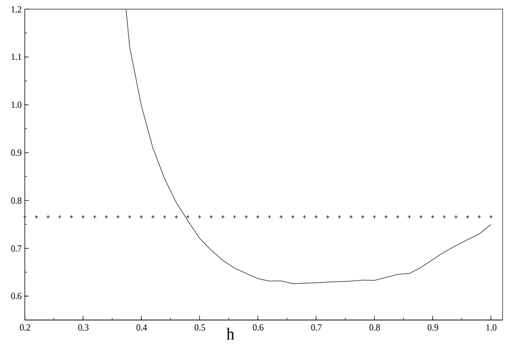
	h	p	p_{lin}	$p - value$
T	0.86	5	4	0.06
AICC	0.86	5	4	0.06
GCV	0.66	6	4	**0.01**
CV	0.87	5	1	**0.03**

Figure 6: *Plot of $SC(h,p)$, selected parameters and p-values of the linearity test for US GNP.*



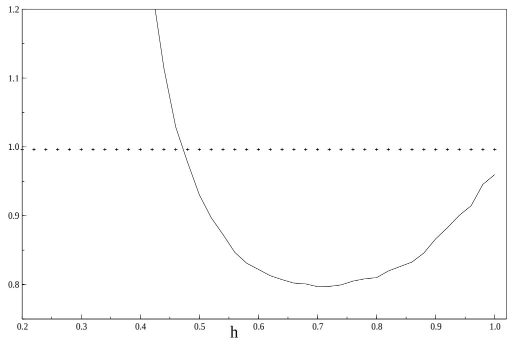
	h	p	p_{lin}	$p - value$
T	0.74	5	5	**0.01**
AICC	0.74	5	5	**0.01**
GCV	0.52	5	5	**0**
CV	0.90	4	5	**0.01**

Figure 7: *Plot of $SC(h,p)$, selected parameters and p-values of the linearity test for US Industrial production.*



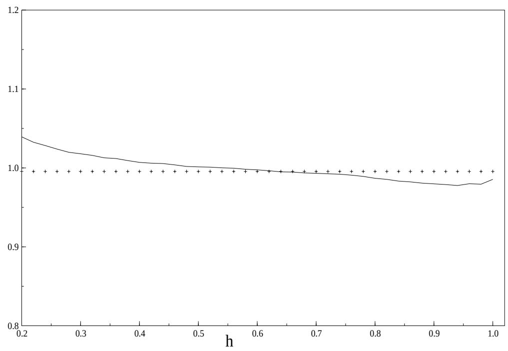
	h	p	p_{lin}	$p-value$
T	0.66	6	4	**0.01**
AICC	0.66	6	4	**0.01**
GCV	0.56	6	4	**0.01**
CV	0.96	4	4	0.18

Figure 8: *Plot of $SC(h,p)$, selected parameters and p-values of the linearity test for the unemployment rate.*



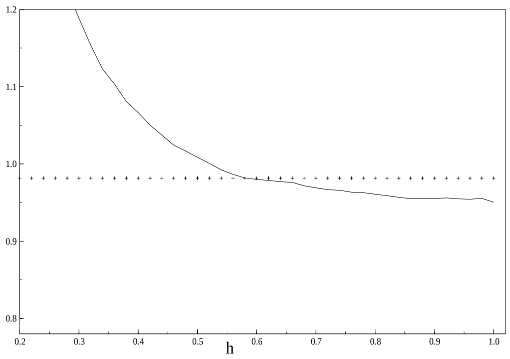
	h	p	p_{lin}	$p-value$
T	0.70	6	6	**0.01**
AICC	0.66	6	6	**0.01**
GCV	0.32	6	6	**0.01**
CV	1	2	1	0.10

Figure 9: *Plot of $SC(h,p)$, selected parameters and p-values of the linearity test for T-bills.*



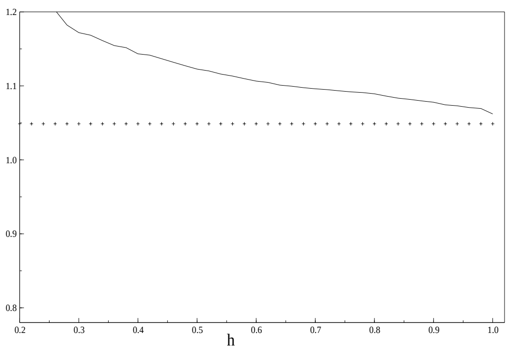
	h	p	p_{lin}	$p-value$
T	0.94	2	1	0.16
AICC	0.94	2	1	0.16
GCV	0.94	2	1	0.16
CV	0.94	2	1	0.16

Figure 10: *Plot of $SC(h,p)$, selected parameters and p-values of the linearity test for S&P500.*



	h	p	p_{lin}	$p-value$
T	1	3	3	0.18
AICC	1	3	3	0.18
GCV	0.82	5	3	0.09
CV	0.88	3	3	0.28

Figure 11: *Plot of $SC(h, p)$, selected parameters and p-values of the linearity test for DM/US exchange rate.*



	h	p	p_{lin}	$p-value$
T	1	1	3	0.54
AICC	1	1	3	0.54
GCV	1	1	3	0.54
CV	1	1	3	0.54

Figure 12: *Plot of $SC(h, p)$, selected parameters and p-values of the linearity test for Yen/US exchange rate.*