Forecasting the Stock Exchange Rate of Thailand Index by Conditional Heteroscedastic Autoregressive Nonlinear Model with Autocorrelated Errors

Autcha Araveeporn [a] Sujit K. Ghosh [b] and Kamon Budsaba[*] [a]

[a] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani 12120, Thailand.

[b] Department of Statistics at NC State University, Raleigh, NC 27695-8203, USA.

* Author for correspondence; e-mail : kamon@mathstat.sci.tu.ac.th

Abstract

        The goal of this work is to develop a nonparametric regression model that not only account for possibly non-linear trend (i.e., conditional mean of the response variable) but also account for possibly non-linear conditional variance of response (i.e., heteroscedasticity) as a function of predictor variables in the presence of auto-correlated errors. The trend and the heteroscedasticity are modeled using a class of penalized spline. The residuals are modeled as a long autoregressive process which can approximate almost any autoregressive moving average (ARMA) process by selecting an appropriate number of lag residuals. Both classical and Bayesian methodologies are developed to obtain the smooth estimates of the conditional mean and variance functions. The resulting estimated residuals are then used to fit a possibly long AR process by suitably choosing the order of AR using the Akaike Information Criteria (AIC). The forecasting performance of the proposed methods is then applied to the series of monthly observations of the Stock Exchange Rate of Thailand (SERT) to illustrate the methodology. The forecasts these methods are compared with those obtained based on future six months of withheld observations.

_____

Keywords: autocorrelation, Bayesian inference, heteroscedasticity, penalized spline.

1. Introduction

The heteroscedasticity or volatility has been modeled in the fields of business and finance in the form of time series that often exhibit nonstationarity. The nonstationarity of time series might be caused by several aspects including changes in trend, volatility and random walk, especially when the data are systematically collected over a long period of time.

The modeling of available explanatory variables has a variety of applications in time series models. The nonparametric method is the choice for estimating regression function between two sets of variables that consist of a vector of predictor and a response variable which may have a nonlinear relationship. Robinson [1] suggested the use of nonparametric estimation in the context of time series model, Marsy and Tjøstheim [2] extended the nonparametric regression to use nonlinear autoregressive conditional heteroscedastic model.

Typically, the nonparametric regression methods are based on a smoothing technique which produces a smoother. A smoother is a tool for summarizing the trend of a response variable as a function of one or more predictor variables. The single predictor case is called scatterplot smoothing that can be used to enhance the visual appearance of the scatterplot of response versus predictor variable, to help our eyes pick out the trend in the plot [3]. There are many smoothing techniques, E.g., a local polynomial regression [4,5], regression splines [6,7], smoothing splines [8,9], and penalized spline [10]. These smoothing techniques are generally based on the assumption of homoscedastic variance model which may not be suitable when the data involves high volatility.

There are several methods to model volatility in time series, such as the autoregressive conditional heteroscedastic model (ARCH) by Engle [11], who was the first to introduce the ARCH model to obtain the predictive variance for U.K. inflation rate. Gouriéroux and Monfort [12] and Masry and Tjøstheim [13] have proposed the conditional heteroscedastic autoregressive nonlinear (CHARN) model in financial time series. For simplicity, the case is one lag of the CHARN model were studied to model the foreign exchange rates [14]. Nonparametric smoothing techniques can be applied for the estimation of CHARN model by considering the response and predictor variables in terms of nonparametric regression by using the nonparametric conditional heteroscedastic autoregressive nonlinear model (NCHARN).

The prediction of nonlinear time series is difficult because of the volatility and autocorrelated errors, so the autoregression have been applied on the error term. Various nonlinear autoregressive model have appeared in literature; Haggan and Ozaki [15] modeled nonlinear vibration by using an amplitude-dependent autoregressive time series model defined as exponential autoregressive (EXPAR) model; Tong [16] introduced the threshold autoregressive(TAR) model in nonlinear time series; and Chan and Tong [17] developed TAR model to smooth-transition autoregressive (STAR) model.

In this paper we focus on NCHARN models with autocorrelated errors using penalized spline and develop Bayesian approach for penalized spline. Section 2 presents the methodology of penalized spline [10] to estimate smoothing trend. Section 3 describes the NCHARN models with autocorrelated errors and applies the methodology from Section 2 to real data in Section 4, we discuss the results in Section 5.

2. Methodology

The general methodjology of smoothing technique modeling starting with the simple nonparametric regression model can be written as

$$y_t = \mu(x_t) + \varepsilon_t, \qquad t = 1, 2, \ldots \tag{1}$$

where ε_t are i.i.d. $N(0, \sigma_\varepsilon^2)$, (y_t, x_t) are a set of response and predictor variables, and $\mu(.)$ is a smooth unknown trend function which is also the conditional mean of y_t given to $x = x_t$.

The penalized spline is a method to estimate a unknown smooth function using the truncated power function [18], and the penalized spline can be expressed as

$$\mu(x_t) = \sum_{j=0}^{m-1} \alpha_j x_t^j + \sum_{k=1}^{K} \beta_k \left| x_t - \tau_k \right|^{2m-1} \tag{2}$$

where $\beta = [\beta_1, \ldots, \beta_K]^T \sim N(0, \sigma_\beta^2 \Omega^{-1/2} (\Omega^{1/2})^T)$, and the (l,k)th entry of Ω is $\left| \tau_l - \tau_k \right|^{2m-1}$ and only the coefficient of $\left| x_t - \tau_k \right|^{2m-1}$ are penalized so that a reasonably large order K can be used.

In this case, we focus m=2, or the so-called low-rank thin-plate spline which tend to have very good numerical properties. The low-rank thin-plate spline representation of $\mu(.)$ is

$$\mu(x_t, \theta) = \alpha_0 + \alpha_1 x_t + \sum_{k=1}^{K} \beta_k |x_t - \tau_k|^3 \tag{3}$$

where $\theta = (\alpha_0, \alpha_1, \beta_1, ..., \beta_K)^T$ is the vector of regression coefficients, and

$\tau_1 < \tau_2 < ... < \tau_K$ are fixed knots. The number of knots, K can be selected using a cross validation method or information theoretic methods (e.g., BIC or AIC).

This class of penalized spline smoothers, $\hat{\mu}(.)$, may also be expressed in convenient vector form

$$\hat{\mu} = C(C^T C + \lambda^3 D)^{-1} C^T y \tag{4}$$

where

$$C = \begin{bmatrix} 1 & x_t & |x_t - \tau_k|^3_{1 \le k \le K} \end{bmatrix}_{1 \le t \le n} \quad , \quad D = \begin{bmatrix} 0_{2 \times 2} & 0_{2 \times K} \\ 0_{K \times 2} & (\Omega_K^{1/2})^T \Omega_K^{1/2} \end{bmatrix}$$

and $\lambda = \dfrac{\sigma_\beta^2}{\sigma_\varepsilon^2}$ is a smoothing parameter estimated by the restricted maximum likelihood method.

3. NCHARN Model with Autocorrelated Errors

Consider the NCHARN model as

$$y_t = \mu(x_t) + \sigma(x_t)\, \varepsilon_t \quad , \qquad t=1,2,... \tag{5}$$

where $\mu(.)$ is a smooth unknown trend (condition mean) function and $\sigma^2(.)$ is a smooth unknown volatility (condition variance) function. In this structure, y_t denotes a response variable and x_t denotes a predictor variable.

The error process $\{\varepsilon_t, t = 1,2,...\}$ is assumed to follow an autoregressive (AR) process given by

$$\varepsilon_t = \sum_{j=1}^{k} \rho_j \varepsilon_{t-j} + e_t \tag{6}$$

where $\rho_1,...,\rho_k$ and k will be estimated based on the data $\{(y_t, x_t); t = 1,...,n\}$.

We further assume that $\{e_t; t = 1,2,....\}$ is a white noise i.e., e_t 's are independently

and identically distributed with mean 0 and variance 1. It would be of interest to estimate

the trend, $\mu(.)$, and volatility, $\sigma^2(.)$, the order of AR process k and the AR coefficients,

$\rho_1,...,\rho_k$.

         Next we test these standardized residuals $\varepsilon_1,...,\varepsilon_n$ for possible

autocorrelation. We choose the order k by using Akaike's information criteria (AIC) and

the AR coefficients $\rho_1,...,\rho_k$ by using maximum likelihood method based on the

autoregressive process.

3.1 Trend and Volatility Estimation using Classical Penalized Spline

         The trend $\mu(x_t)$ and volatility $\sigma^2(x_t)$ can also be considered in NCHARN

model. As an initial step, we start by estimating the trend $\mu(x_t)$ using a homoscedastic

nonparametric regression model written as

$$y_t = \mu(x_t) + \delta_t , \qquad t = 1,2,\ldots \tag{7}$$

where $\delta_t = \sigma(x_t) \varepsilon_t$. Next, we obtain $\hat{\mu}(x_t)$ from the method of penalized spline

smoother in Section 2 with package *SemiPar* in R Program which downloaded from

 http://cran.r-project.org. The residuals can be estimated as

$$\hat{\delta}_t = y_t - \hat{\mu}(x_t)$$

$$(\hat{\delta}_t)^2 = (\sigma(x_t) \varepsilon_t)^2 \tag{8}$$

We transform $\sigma(x_t) = \exp\left\{\dfrac{h(x_t)}{2}\right\}$, and take log with residuals in (8)

$$\log \hat{\delta}_t^2 = h(x_t) + \log \varepsilon_t^2 \tag{9}$$

$$\log \hat{\delta}_t^2 - E[\log \varepsilon_t^2] = h(x_t) + \log \varepsilon_t^2 - E[\log \varepsilon_t^2] \tag{10}$$

If we require ε_t to be normally distributed with mean 0 and variance 1, then

$E[\log \varepsilon_t^2] = $ -1.2704 and hence we can apply in penalized spline to obtain

$$\log \hat{\delta}_t^2 + 1.2704 = h(x_t) + \log \varepsilon_t^2 + 1.2704 \qquad (11)$$

$$\tilde{y}_t = h(x_t) + \tilde{\varepsilon}_t \qquad (12)$$

where $\tilde{y}_t = \log \hat{\delta}_t^2 + 1.2704$ and $\tilde{\varepsilon}_t = \log \varepsilon_t^2 + 1.2704$. Next, we obtain a

smooth estimate $\hat{h}(x_t)$ using penalized spline by using (12) and update the volatility

estimate to be

$$\hat{\sigma}(x_t) = \exp\left\{\frac{\hat{h}(x_t)}{2}\right\} \qquad (13)$$

At the second stage of estimation we update the trend estimate by using the following

model

$$y_t = \mu(x_t) + \exp\left\{\frac{\hat{h}(x_t)}{2}\right\}\varepsilon_t \qquad (14)$$

$$\exp\left\{-\frac{\hat{h}(x_t)}{2}\right\}y_t = \exp\left\{-\frac{\hat{h}(x_t)}{2}\right\}\mu(x_t) + \varepsilon_t \qquad (15)$$

$$\breve{y}_t = g(x_t) + \varepsilon_t \qquad (16)$$

where $\breve{y}_t = \exp\left\{-\frac{\hat{h}(x_t)}{2}\right\}y_t$ and $g(x_t) = \exp\left\{-\frac{\hat{h}(x_t)}{2}\right\}\mu(x_t)$. Finally, if

$\hat{g}(x_t)$ is the estimate obtained by using penalized spline, the second stage estimate

of $\mu(x_t)$ is given by

$$\hat{\mu}(x_t) = \exp\left\{\frac{\hat{h}(x_t)}{2}\right\}\hat{g}(x_t) \qquad (17)$$

Finally, when the estimates of $\mu(x_t)$ and $\sigma(x_t)$ converge we obtain

$$\hat{\varepsilon}_t = \frac{y_t - \hat{\mu}(x_t)}{\hat{\sigma}(x_t)}, \ t = 1,2,\ldots \qquad (18)$$

as the standardized residuals based on the converged values of $\hat{\mu}(x_t)$ and

$\hat{\sigma}(x_t)$ as describe above by using the AR process.

3.2 Trend and Volatility Estimation using Bayesian Penalized Spline

      To perform Bayesian data analysis for the penalized spline, it helps to set up the model as a three-level hierarchical model. At the first level, the conditional distribution of the observation y_t's is specified, given the unobserved random coefficients $\underline{a}, \underline{b}, \underline{\alpha}$ and $\underline{\beta}$; at the second level, the fixed effect is parameter $\underline{a}$ and $\underline{\alpha}$ from multivariate normal distribution, and the random effect of $\underline{b}$, and $\underline{\beta}$ is specified given the parameter Σ_b and Σ_β from multivariate normal distribution; and finally at the last level, the prior distribution of Σ_b and Σ_β is specified from Inverse Wishart distribution. In this section, we will assume that ε_t's are iid $N(0, \sigma_\varepsilon^2)$.

      We are able to express the penalized spline model in the following hierarchical structure,

$$y_t \big| \underline{a}, \underline{b}, \underline{\alpha}, \underline{\beta} \quad \sim \quad N(\mu(x_t), \sigma^2(x_t))$$

$$\mu(x_t) = a_0 + a_1 x_t + \sum_{k=1}^{K} b_k \big| x_t - \tau_k \big|^3$$

$$\sigma^2(x_t) = \Big| \log\{\alpha_0 + \alpha_1 x_t + \sum_{k=1}^{K} \beta_k \big| x_t - \tau_k \big|^3 \} \Big|$$

$$\underline{a} \sim MVN(0, \Sigma_a) \quad , \quad \underline{\alpha} \sim MVN(0, \Sigma_\alpha)$$

$$\underline{b} \big| \Sigma_b \sim MVN(0, \Sigma_b) \quad , \quad \underline{\beta} \big| \Sigma_\beta \sim MVN(0, \Sigma_\beta)$$

$$\Sigma_b \sim IW(R_b, K_b) \quad , \quad \Sigma_\beta \sim IW(R_\beta, K_\beta) \tag{19}$$

where $\underline{a} = (a_0, a_1)^T$, $\underline{b} = (b_1, .., b_K)^T$, $\underline{\alpha} = (\alpha_0, \alpha_1)^T$, $\underline{\beta} = (\beta_1, .., \beta_K)^T$.

      We use the so-called Markov Chain Monte Carlo (MCMC) methods to generate samples from the posterior distribution of $\theta = (\underline{a}, \underline{b}, \underline{\alpha}, \underline{\beta}, \Sigma_b, \Sigma_\beta)^T$. MCMC methods consist of algorithms to construct a Markov chain of the parameters such that its stationary distribution is the posterior distribution of the parameters. Hence, under certain regularity conditions, the realization of the Markov chain can be thought of as approximate values sampled from the posterior distribution of θ given the y_t's. We

carry out the Metropolis-Hastings sampler a widely used MCMC method, to obtain dependent samples from the posterior distribution using the *WinBUGS* that installed from *http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml.* The model (19) is specified in *WinBUGS* as follows

for (t in 1:n)

{y[t]~dnorm(mu[t], tau[t])

 mu[t] <- inprod(X[t,],beta.mu[1:2]) + inprod(Z[t,],gamma.mu[1:n.knots])

 tau[t] <- pow(log(sd[t]),-2)

 sd[t] <- inprod(X[t,],beta.tau[1:2]) + inprod(Z[t,],gamma.tau[1:n.knots]) }

The number of sample sizes, n, is a constant in the program. The first statement specifies that the *t*-th y has a normal distribution with mean, μ_t , and precision $\tau_t = \sigma_t^{-2}$. Both the matrix **X** and **Z** are obtained outside *WinBUGS* and the code of matrix is referred in Crainiceanu, Ruppert, and Wand [19].

       We consider Metropolis-Hastings algorithm to obtain dependent samples from the posterior distribution $p(\theta|y) \propto h(\theta) \equiv f(y|\theta)\pi(\theta)$, a powerful Markov Chain method to simulate multivariate distribution. To evolution of a Markov Chain depends on the transition kernel density (TKD), $K(\theta,\theta')$ and also known as the proposal density. Starting with $\theta^{(0)}$ iteration for k = 1,2,…, m+M, and the TKD is $K(\theta,\theta')$.

       The metropolis-Hasting algorithm is

   1.   Draw $\theta^{new} \sim K_0(\theta^{(k-1)},\theta')$

   2.   Draw $u \sim U(0,1)$ and set

$$\theta^{(k)} = \begin{cases} \theta^{new}, if \ u \le \rho(\theta^{(k-1)},\theta^{new}) \\ \theta^{(k-1)}, \qquad otherwise \end{cases}$$

where the acceptance probability $\rho(\theta,\theta')$ is defined as

$$\rho(\theta,\theta') = \min\left\{\frac{h(\theta')K_0(\theta',\theta)}{h(\theta)K_0(\theta,\theta')},1\right\}$$

where m is burn-in and M is the number of samples generated after burn-in.

       Repeating the above sampling steps, we obtain a discrete-time Markov chain

$$\left\{ \left(\underline{a}^{(k)}, \underline{b}^{(k)}, \underline{\alpha}^{(k)}, \underline{\beta}^{(k)}, \Sigma_b^{(k)}, \Sigma_\beta^{(k)} \right); \ k = 1,2,... \right\}$$ whose stationary distribution is the

joint posterior density of the parameters.

The MCMC samples (after sufficient number of turn-in samples) of θ are obtained via *WinBUGS* to compute approximate posterior summaries of the parameters as the posterior estimation of θ. In particular, we use the posterior median as point estimates to estimate of $\mu(x_t)$ and $\sigma^2(x_t)$. Let

$$\hat{\mu}(x_t) = \hat{a}_0 + \hat{a}_1 x_t + \sum_{k=1}^{K} \hat{b}_k \left| x_t - \tau_k \right|^3$$

and

$$\hat{\sigma}^2(x_t) = \left| \log\{ \hat{\alpha}_0 + \hat{\alpha}_1 x_t + \sum_{k=1}^{K} \hat{\beta}_k \left| x_t - \tau_k \right|^3 \} \right|$$

denote posterior estimate of $\mu(x_t)$ and $\sigma^2(x_t)$, where $\hat{\underline{a}}, \hat{\underline{b}}, \hat{\underline{\alpha}}$ and $\hat{\underline{\beta}}$ denote the component-wise posterior median of $\underline{a}, \underline{b}, \underline{\alpha}$ and $\underline{\beta}$.

Finally the standardized residuals are calculated by

$$\hat{\varepsilon}_t = \frac{y_t - \hat{\mu}(x_t)}{\hat{\sigma}(x_t)}, \ t = 1,2,...$$

to obtain the estimate the order of the AR process of the errors.

3.3 Autoregressive Process

The maximum likelihood estimator of error process is evaluated from the probability density function of each ε_t is

$$(2\pi)^{-1/2} \exp\left\{ -\frac{e_t^2}{2} \right\} \quad , \quad -\infty < e_t < \infty \tag{20}$$

The likelihood function in form of ε_t is written as

$$L(\rho_k) = (2\pi)^{-(n-k)/2} \exp\left\{ -\frac{1}{2} \sum_{t=2}^{n} \left(\varepsilon_t - \sum_{j=1}^{k} \rho_j \varepsilon_{n-j} \right)^2 \right\} \tag{21}$$

The log-likelihood function denoted $l(\rho_k)$ is given by

$$l(\rho_k) = -\frac{n-k}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{n}\left(\varepsilon_t - \sum_{j=1}^{k}\rho_j\varepsilon_{n-j}\right)^2 \tag{22}$$

The maximum likelihood estimator of ρ_k causes the equation $\dfrac{\partial l(\rho_k)}{\partial \rho_k}$ =0 and we get

the maximum likelihood estimator of ρ_k is

$$\hat{\rho}_k = \frac{\displaystyle\sum_{t=k+1}^{n}\varepsilon_t\varepsilon_{t-k}}{\displaystyle\sum_{t=k+1}^{n}\varepsilon_{t-k}^2 + \sum_{j=1}^{k-1}\left\{\sum_{t=k+1}^{n}\varepsilon_{t-j}\varepsilon_{t-k}\right\}} \quad , \quad k=1,2,\ldots \tag{23}$$

We choose the order k by using Akaike's information criteria, AIC, as

$$\text{AIC} = -2\ln L(\theta) + 2p \tag{24}$$

where $L(\theta)$ is the likelihood function evaluated at the maximum likelihood

estimates and p is the total number of parameter estimated.

4. Application to SERT Index

In this section, we apply the methods described in Section 2 to the price index of Thailand. The Stock Exchange Rate of Thailand (SERT) index is an important index in Thailand that started trading on April 30, 1975. The data consisted of 396 records of the monthly volume of SERT index from January 1976 to December 2008 that can be found at *http://www.set.or.th/th/market/market_statistics.html*.

Let y_t denote the SRET Index of month t where t=1 represents January of 1976 and t=390 represents June of 2008. The method of the classical penalized spline and Bayesian penalized spline are used to forecast future values of SERT index for July, 2008 to December, 2008 given in Table 1. The estimate the order and coefficients of the AR process of the standardized residuals by the classical penalized spline and Bayesian penalized spline are shown in Tables 2 and 3.

Table 1. The actual SERT Index, the forecast values (FV), 95% prediction interval of future values (95%PI), mean absolute deviation (MAD) for classical penalized spline and Bayesian penalized spline

m	SERT Index	Classical Penalized Spline		Bayesian penalized spline	
		FV	95% PI	FV	95% PI
1	676.32	697.48	(611.96,783.01)	711.65	(601.23,822.08)
2	684.44	632.96	(503.33,762.59)	643.10	(496.08,790.13)
3	596.54	574.04	(422.69,725.40)	577.56	(405.26,749.86)
4	416.53	507.95	(348.74,667.68)	527.06	(343.98,710.14)
5	401.84	464.88	(305.09,624.68)	466.74	(278.53,654.94)
6	449.96	412.81	(259.98,565.63)	448.86	(259.03,638.70)
MAD		47.79	-	45.36	-

From Table 1, it is apparent that the Mean Absolute Deviation(MAD) by Bayesian penalized spline method is slightly smaller than that of the classical penalized spline method. However it can not be concluded that Bayesian penalized spline method performs significantly better than classical penalized spline, because the predictive intervals obtained by both methods over lap each other in all of the six test cases. The classical penalized spline method being non iterative is computationally easier to implemented than that of the Bayesian penalized spline.
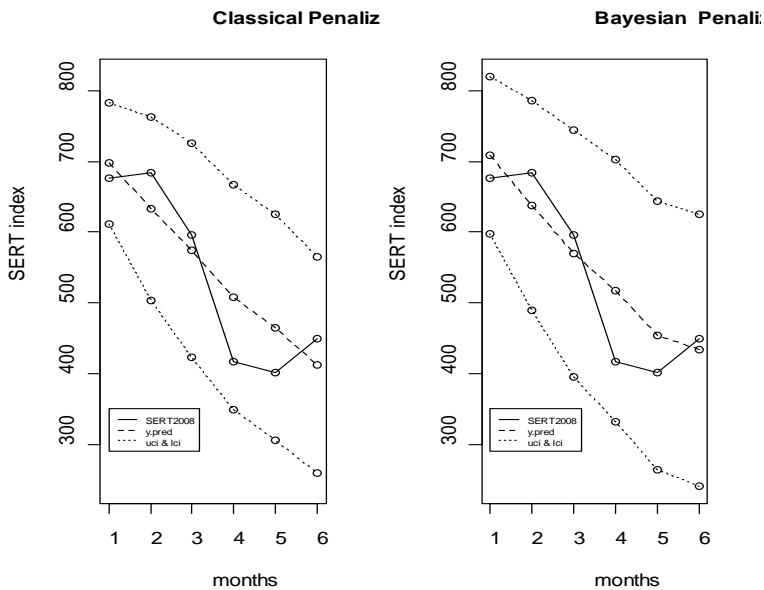
Figure 1. The actual SERT index(SERT 2008), forecast values(y.pred) and 95% prediction interval of future values (uci and lci) for classical penalized spline and Bayesian penalized spline.

Figure 1 is shown the forecast values and 95% prediction interval of withheld SERT values obtained by the classical penalized spline method and Bayesian penalized spline method. It follows from the figure that the both methods provides predictive intervals that capture all future values.

Comparing MAD values from Table 1, it can be seen that the Bayesian approach for penalized spline method provides the smaller MAD. However, it should be noted that both the Bayesian estimate (based on penalized spline) are iterative and may not be computationally fast. On the other hand, the classical penalized spline are non-iterative and provide reasonably good estimates in terms of minimizing the MAD.

Table 2. The estimate the order(k) and coefficients (ρ_k) of the AR process of the standardized residuals by the classical penalized spline

k	1	2	3	4	5	6
ρ_k	1.3492	-0.4421	0.0375	0.0360	-0.2005	0.2010
k	7	8	9	10	11	12
ρ_k	-0.0503	-0.1620	0.1946	-0.1419	0.0987	-0.1501
k	13	14	15	16	17	18
ρ_k	0.1092	-0.0427	-0.0822	0.1219	-0.1241	-0.0081
k	19	20	21	22	23	
ρ_k	0.1624	-0.0862	0.0990	-0.2722	0.1361	

Table 3. The estimate the order(k) and coefficients (ρ_k) of the AR process of the standardized residuals by the Bayesian penalized spline

k	1	2	3	4	5	6
ρ_k	0.8779	0.0407	-0.1910	0.0358	-0.0247	-0.0772
k	7	8	9	10	11	
ρ_k	0.1895	-0.0696	-0.0706	0.2080	-0.2069	

From Tables 2 and 3, it is apparent that the order of k of AR process from classical penalized spline is higher than the Bayesian penalized spline. However, the variance of classical penalized spline is 0.006334 that is smaller than the variance of the Bayesian penalized spline at 23.45.

5. Discussion

In this article, we have investigated and compared the classical penalized spline method and the Bayesian penalized spline method to estimate smooth unknown trend and smooth unknown volatility based on a conditional heteroscedastic autoregressive nonlinear model. The Bayesian method performs slightly better than the classical penalized spline based method in terms of minimizing the MAD. The autoregressive process is useful for prediction of future values.

References

[1] Robinson, P.M., Non-parametric estimation for time series models. *Journal of Time Series Analysis*, 1983;4:185-208.

[2] Marsy, E., and Tjøstheim, D., *Non-parametric estimation and identification of ARCH and ARX nonlinear time series: convergence properties and rates*. Preprint, Dept. of Mathematics, Univ. of Bergen., 1992.

[3] Hastie, T.J., and Tibshirani, R., *Generalized Additive Models*. Chapman and Hall, London, 1990.

[4] Wand, M.P., and Jones, M.C., *Kernel Smoothing*. Chapman and Hall, London, 1995.

[5] Fan, J., and Gijbels, I., *Local polynomial modeling and its applications*. Chapman and Hall, London, 1996.

[6] Eubank, R.L., *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York, 1988.

[7] Eubank, R.L., *Nonparametric Regression and Spline Smoothing*. Marcel Dekker, New York, 1999.

[8] Wahba, G., *Spline Models for Observational Data*. SIAM, Philadelphia, PA, 1990.

[9] Green, P.J., and Silverman, B. W., *Nonparametric Regression and Generalized Linear Models*: *A Roughness Penalty Approach.* Chapman and Hall, London, 1994.

[10] Ruppert, D., Wand, M.P., and Carroll, R.J., *Semiparametric Regression*. Cambridge University Press, 2003.

[11] Engle, R F., Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 1982;50:987-1007.

[12] Gouriéroux, C., and Monfort, A., Qualitative threshold ARCH models. *Journal of Econometrics*, 1992;52:159-199.

[13] Masry, E., and Tjøstheim, D., Non-parametric estimation and identification of ARCH nonlinear time series: strong convergence and asymptotic normality. *Econometric Theory*, 1995a;11:258-289.

[14] Bossaerts, P., Härdle, W., and Hafner, C., Foreign exchange-rates have surprising volatility. In Athens Conference on Applied Probability and Time Series, Vol. 2 (ed. P.M. Robinson). *Lecture Notes in Statistics 115,* New York: Springer, pp. 55-72, 1996.

[15] Haggan, V., and Ozaki, T., Modeling nonlinear vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, 1981;68: 189-196.

[16] Tong, H., Threshold models in nonlinear time series analysis. *Lecture Notes in Statistics 21*, Heidelberg: Springer, 1983.

[17] Chan, K.S., and Tong, H., On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 1986; 7: 179-190.

[18] Ruppert, D., and Carroll, R.J. Spatial-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics*, 2000; 42: 205-224.

[19] Crainiceanu, C.M., Ruppert, D., and Wand, M.P. Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software*, 2005; 14: 1-24.