Developing Nonparametric Conditional Heteroscedastic Autoregressive Nonlinear Model by Using Maximum Likelihood Method

Autcha Araveeporn
Department of Applied Statistics, Faculty of Science, King Monkut’s Institute of Technology Ladkrabang, Bangkok 10520, Thailand.
*Author for correspondence; e-mail: kaautcha@kmitl.ac.th

ABSTRACT
The goal of this work is to develop a nonparametric conditional heteroscedastic autoregressive nonlinear (NCHARN) model by using maximum likelihood method that not only account for possibly non-linear trend but also account for possibly non-linear conditional variance of response 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 and the residuals are modeled as an autoregressive process (AR) by selecting an appropriate number of lag residuals. Both classical penalized spline and AR process of penalized spline under NCHARN model are developed to obtain the smooth estimates of the conditional mean and variance functions. The resulting estimated values are then used the maximum likelihood method to fit a trend, volatility, and a coefficient of 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: autoregressive, heteroscedasticity, penalized spline, volatility.
1. INTRODUCTION

In the fields of business and finance, most data are collected in the form of time series that often exhibit nonstationarity and volatility because of sudden random changes in market, especially when the data are systematically collected over a long period of time. The heteroscedasticity or volatility has been modeled in the literature by various authors, for instance, Anderson and Lund [1] and Gallant and Tauchen [2] evaluated risk in finance, Yao and Tong [3] monitored the reliability of nonlinear prediction.

The modeling of available explanatory variables has a variety of applications in regression model. The parametric and nonparametric method are the choice for estimating regression function between two sets of variables that consist of a vector of predictors and a response variable. A parametric regression model requires an assumption in the form of the underlying regression function. To overcome the difficulty caused by the restrictive assumption of the parametric form of the regression function, one may remove the restriction that the regression function belongs to a parametric family. This approach leads to so-called nonparametric regression.

Penalized spline regression models have found a lot application in recent years, which is used of fitting and flexible choice of knots and smoothing parameter in nonparametric regression model. Penalized spline regression models, often referred to as penalized spline model, is a nonparametric method defined by Eiler and Marx [4]. The class of penalized spline, based on the reduced-knot B-Splines basis with discrete different penalties, motivated as a computationally efficient approximation to the natural cubic spline. Ruppert et al. [5] described penalized spline models, based on reduce-knot truncated power function basis with penalties on the untransformed coefficients, fitted as a mixed model, and motivated as a simple low-rank smoothing spline. The penalized spline model is 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 [6], who was the first to introduce the ARCH model to obtain the predictive variance for U.K. inflation rate. Gouriéroux and Monfort [7] and Marry and Tjostheim [8] 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 (Bossaerts et al. [9]). 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 [10] modeled nonlinear vibration by using an amplitude-dependent autoregressive time series model defined as exponential autoregressive (EXPAR) model; Tong [11] introduced the threshold autoregressive(TAR) model in nonlinear time series; and Chan and Tong [12] developed TAR model to smooth-transition autoregressive (STAR) model.

In this article, we focus on NCHARN models with autocorrelated errors using
penalized spline in which the parameters are estimated maximum likelihood method.

Section 2 presents the methodology of penalized spline (Ruppert et al. [5]) to estimate smoothing trend. Section 3 describes classical penalized spline and AR process of penalized spline under NCHARM model and applies the methodology from Section 2 to maximum likelihood method in Section 4. The main results and discussion are shown in section 5 and we conclude in Section 6.

2. METHODOLOGY

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

$$y_t = \mu(x_t) + \epsilon_t, \quad t = 1, 2, \ldots$$

where $\epsilon_t$ are i.i.d. $N(0, \sigma^2)$, $(y_t, x_t)$ 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 (Ruppert and Carroll [13]), and the penalized spline can be expressed as

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

where $\beta = [\beta_1, \ldots, \beta_K]^T \sim N(0, \sigma^2 \Omega^{-1/2} (\Omega^{1/2})^T)$, and the $(l,k)$th entry of $\Omega$ is $|x_t - \tau_l|^{2m-1}$ and only the coefficient of $|x_t - \tau_k|^{2m-1}$ 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

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

where $\theta = (\alpha_0, \alpha_1, \beta_1, \ldots, \beta_K)^T$ is the vector of regression coefficients, and $\tau_1 < \tau_2 < \ldots < \tau_K$ 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
3. NCHARN Model

Consider the NCHARN model as
\[ y_t = \mu(x_t) + \sigma^2(x_t) \epsilon_t, \quad t = 1, 2, \ldots \]
(5)
where \( \mu(\cdot) \) is a smooth unknown trend (condition mean) function and \( \sigma^2(\cdot) \) is a smooth unknown volatility (condition variance) function. In this structure, \( y_t \) denotes a response variable and \( x_t \) denotes a predictor variable.

However, the error terms are significant when we come to address the time series data problems: the errors terms are assumed i.i.d. N(0,1) defined NCHARN of classical penalized spline (NCHARN(PS)) and penalized spline under autoregressive process defined NCHARN(PS,AR).

3.1 NCHARN(PS) Model

Under the NCHARN model in (5), Araveeporn [14] assumed that trend \( \mu(x_t) \) and volatility \( \sigma^2(x_t) \) are generated by penalized spline form as
\[ \mu(x_t) = a_0 + a_1 x_t + \sum_{k=1}^{K} b_k |x_t - \tau_k|^3 \]
\[ \sigma(x_t) = \sqrt{\log \{ \alpha_0 + \alpha_1 x_t + \sum_{k=1}^{K} \beta_k |x_t - \tau_k|^3 \} } \]
and the error terms are independently and identically distributed in Normal distribution with mean 0 and variance 1.
3.2 NCHARN(PS,AR) Model

Under the NCHARN model in (5), Araveeporn et al. [15] applied the error process \( \{\varepsilon_t, t = 1, 2, \ldots\} \) assumed to follow an autoregressive (AR) process given by

\[
\varepsilon_t = \sum_{j=1}^{k} \rho_j \varepsilon_{t-j} + \varepsilon_t \tag{8}
\]

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

We further assume that \( \{\varepsilon_t; t = 1, 2, \ldots\} \) is a white noise i.e., \( \varepsilon_t \)'s are independently and identically distributed in Normal distribution 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, \( \rho_1, \ldots, \rho_k \).

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

\[
y_t = \mu(x_t) + \sigma(x_t)\varepsilon_t, \quad t = 1, 2, \ldots \tag{9}
\]

where the standardized residuals are rewrite from (8) following

\[
\varepsilon_t = \frac{y_t - \mu(x_t)}{\sigma(x_t)} \tag{10}
\]

and the autoregressive (AR) process given by

\[
\varepsilon_t = \sum_{j=1}^{k} \rho_j \varepsilon_{t-j} + \varepsilon_t \tag{11}
\]

Next, we replace equation (10) and (11) in NCHARN model (9) and we obtain

\[
y_t = \mu(x_t) + \sigma(x_t)\varepsilon_t \\
= \mu(x_t) + \sigma(x_t) \left[ \sum_{j=1}^{k} \rho_j \varepsilon_{t-j} + \varepsilon_t \right] \\
= \mu(x_t) + \sigma(x_t) \left[ \sum_{j=1}^{k} \rho_j \left( \varepsilon_{t-j} - \frac{\mu(x_{t-j})}{\sigma(x_{t-j})} \right) \right] + \sigma(x_t)\varepsilon_t \\
= \mu(x_t) + \sum_{j=1}^{k} \rho_j \sigma(x_{t-j})\mu(x_{t-j}) + \sum_{j=1}^{k} \rho_j \frac{\sigma(x_{t-j})\mu(x_{t-j})}{\sigma(x_{t-j})} y_{t-j} + \sigma(x_t)\varepsilon_t \\
= \mu(x_t) + \mu(x_t) y_{t-1}, \ldots, y_{t-n} + \sigma(x_t)\varepsilon_t \\
= \mu^*(x_t) + \sigma(x_t)\varepsilon_t \tag{12}
\]
where

\[ \mu^*(x_t) = \mu_1(x_t) + \mu_2(y_{t-1}, \ldots, y_{t-p}) \]
\[ \mu_1(x_t) = \mu(x_t) - \sum_{j=1}^{p} \rho_j \frac{\sigma(x_t)}{\sigma(x_{t-j})} \mu(x_{t-j}) \]
\[ \mu_2(y_{t-1}, \ldots, y_{t-p}) = \sum_{j=1}^{p} \rho_j \frac{\sigma(x_t)}{\sigma(x_{t-j})} y_{t-j} \]

where \( t \leq p \), trend \( \mu(x_t) \) and volatility \( \sigma^2(x_t) \) are calculated as the NCHARN(PS) at (6) and (7).

4. Trend and Volatility Estimation Using Maximum Likelihood Method

In a general framework, the method of maximum likelihood has been widely used in estimation for any set of observations. The estimation of volatility model, Engle [6] introduced ARCH model and described the maximum likelihood estimations. Next, we propose the NCHARN(PS) and NCHARN(PS,AR) using maximum likelihood method to evaluate the estimated \( \hat{\mu}(x_t) \) and \( \hat{\sigma}^2(x_t) \) described in section 3.

4.1 Ncharn(PS) Model

To perform maximum likelihood method for the penalized spline, it helps to set up the model in normal distribution as

\[ y_t \sim N(\mu(x_t), \sigma^2(x_t)) \]
\[ \mu(x_t) = a_0 + a_1 x_t + \sum_{k=1}^{K} b_k \left| x_t - \tau_k \right|^3 \] (13)
\[ \sigma(x_t) = \sqrt{\log \left\{ \log \left( a_0 + \alpha_1 x_t + \sum_{k=1}^{K} \beta_k \left| x_t - \tau_k \right|^3 \right) \right\}} \] (14)

where \( \mu(.) \) is a smooth unknown trend function and \( \sigma^2(.) \) is a smooth unknown volatility function.

We use the concept of maximum likelihood estimation to estimate parameter

\[ \theta = (a_0, a_1, b_1, \ldots, b_K, \alpha_0, \alpha_1, \beta_1, \ldots, \beta_K) \].

The probability density function of each \( y_t \) is

\[ f(y_t \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2(x_t)}} \exp \left\{ -\frac{(y_t - \mu(x_t))^2}{2\sigma^2(x_t)} \right\}, -\infty < y_t < \infty \] (15)
The likelihood function is given by

\[
L(\theta \mid y_i) = \prod_{i=1}^{n} f(y_i \mid \theta) = \frac{1}{\left(2\pi \sigma^2(x_i)\right)^{n/2}} \exp\left\{ -\sum_{i=1}^{n} \frac{(y_i - \mu(x_i))^2}{2\sigma^2(x_i)} \right\}
\]

(16)

The log-likelihood function denoted \( l(\theta \mid y_i) \) written as

\[
l(\theta \mid y_i) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2(x_i) - \frac{1}{2} \sum_{i=1}^{n} \frac{(y_i - \mu(x_i))^2}{2\sigma^2(x_i)}
\]

(17)

Now we have

\[
\frac{\partial}{\partial \hat{\theta}(a_0, a_1, b_1, \ldots, b_1)} l(\theta \mid y_i) = 0
\]

The maximum likelihood estimator of \((a_0, a_1, b_1, \ldots, b_1)\) of NCHARN(PS) only is

\[
\hat{a}_o = \frac{\sum_{i=1}^{n} \frac{y_i}{\hat{\sigma}^2(x_i)} - \hat{a}_1 \sum_{i=1}^{n} \frac{x_i}{\hat{\sigma}^2(x_i)} - \sum_{k=1}^{K} \frac{1}{\hat{\sigma}^2(x_i)} \sum_{i=1}^{n} \frac{|x_i - \tau_k|^3}{\hat{\sigma}^2(x_i)}}{\sum_{i=1}^{n} \frac{1}{\hat{\sigma}^2(x_i)}}
\]

(18)

\[
\hat{a}_1 = \frac{\sum_{i=1}^{n} \frac{y_i x_i}{\hat{\sigma}^2(x_i)} - \hat{a}_0 \sum_{i=1}^{n} \frac{x_i}{\hat{\sigma}^2(x_i)} - \sum_{k=1}^{K} \frac{1}{\hat{\sigma}^2(x_i)} \sum_{i=1}^{n} \frac{|x_i - \tau_k|^3 x_i}{\hat{\sigma}^2(x_i)}}{\sum_{i=1}^{n} \frac{x_i^2}{\hat{\sigma}^2(x_i)}}
\]

(19)

\[
\hat{b}_h = \frac{\sum_{i=1}^{n} \frac{y_i}{\hat{\sigma}^2(x_i)} - \hat{a}_0 \sum_{i=1}^{n} \frac{|x_i - \tau_k|^3}{\hat{\sigma}^2(x_i)} - \hat{a}_1 \sum_{i=1}^{n} \frac{x_i}{\hat{\sigma}^2(x_i)} - \sum_{k=1}^{K} \frac{1}{\hat{\sigma}^2(x_i)} \sum_{i=1}^{n} \frac{|x_i - \tau_k|^3}{\hat{\sigma}^2(x_i)}}{\sum_{i=1}^{n} \frac{|x_i - \tau_k|^3}{\hat{\sigma}^2(x_i)}}
\]

\[
, h = 1, 2, \ldots, K
\]

(20)
The maximum likelihood estimators of $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1, \ldots, \hat{\beta}_k$ can be obtained by calculating $\hat{\sigma}(x_i)$ that consisted of $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1, \ldots, \hat{\beta}_k$, where $\hat{\sigma}(x_i)$ are the minimizer of the log-likelihood function.

Unfortunately, the log-likelihood function is nonlinear with respect to $\mu(x_i)$ and $\sigma(x_i)$, and the minimization can be readily done by any numerical routines. We carry out a limited memory algorithm for bound constrained optimization or so-called L-BFGS-B algorithm in traduced by Byrd et al. [16] which allows box constraints, that each variable can be given a lower and/or upper bound and initial value must satisfy the constraints. The L-BFGS-B algorithm has ability for solving large nonlinear optimization problems with simple bounds by using package stat4 in R Program.

The maximum likelihood estimates $\hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ are obtained by using the L-BFGS-B algorithm following

$$l(\hat{\theta} \mid y_i) = -\inf_{\theta} l(\theta \mid y_i)$$

From the results of the L-BFGS-B algorithm, we can be written the maximum likelihood estimator of $\mu(x_i)$ and $\sigma(x_i)$,

$$\hat{\mu}(x_i) = \hat{\alpha}_0 + \hat{\alpha}_1 x_i + \sum_{k=1}^{K} \hat{\beta}_k |x_i - \tau_k|^3$$  \hspace{1cm} (21)

$$\hat{\sigma}(x_i) = \sqrt{\log \left\{ \frac{\hat{\alpha}_0 + \hat{\alpha}_1 x_i + \sum_{k=1}^{K} \hat{\beta}_k |x_i - \tau_k|^3}{1} \right\}}$$  \hspace{1cm} (22)

### 4.2 Ncharn(Ps,Ar) Model

The parameter of maximum likelihood estimation is denoted $\theta = (a_0, a_1, b_1, \ldots, b_k, a_0, a_1, \beta_1, \ldots, \beta_k, \rho_1, \ldots, \rho_p, p)$ as the NCHARN(PS) model. The probability density function of each $y_i$ is

$$f(y_i \mid \theta) = \frac{1}{\sqrt{2\pi \sigma^2(x_i)}} \exp \left\{ -\frac{(y_i - \mu^*(x_i))^2}{2\sigma^2(x_i)} \right\}, -\infty < y_i < \infty$$  \hspace{1cm} (23)

The likelihood function is given by

$$L(\theta \mid y_i) = \frac{1}{(2\pi \sigma^2(x_i))^{n/2}} \exp \left\{ -\frac{\sum_{i=1}^{n} (y_i - \mu^*(x_i))^2}{2\sigma^2(x_i)} \right\}$$  \hspace{1cm} (24)
The log-likelihood function denoted \( l(\theta \mid y_i) \), it can be written as

\[
l(\theta \mid y_i) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2(x_i) - \frac{1}{2} \left\{ \sum_{i=1}^{n} \frac{(y_i - \mu^*(x_i))^2}{\sigma^2(x_i)} \right\}
\]  

(25)

The maximum likelihood estimator of \((a_0, a_1, b_1, \ldots, b_K)\) of NCHARN (PS, AR) based on the \( \mu^*(x_i) \), hence we consider in term of \( \mu^*(x_i) \) following

\[
l(\theta \mid y_i) \approx -\frac{1}{2} \left\{ \sum_{i=1}^{n} \left[ y_i - \frac{\mu(x_i) - \mu(y_{i-j})}{\sigma(x_i)} \right]^2 \right\}
\]

\[
\approx -\frac{1}{2} \left\{ \sum_{i=1}^{n} \left[ \frac{y_i}{\sigma(x_i)} - \frac{\mu(x_i)}{\sigma(x_i)} - \frac{\sum_{j=1}^{\rho} \rho_j}{\sigma(x_{i-j})} \mu(x_{i-j}) - \frac{\sum_{j=1}^{\rho} \rho_j}{\sigma(x_{i-j})} y_{i-j} \right]^2 \right\}
\]

\[
\approx -\frac{1}{2} \left\{ \sum_{i=1}^{n} \left[ \frac{y_i}{\sigma(x_i)} - \frac{\mu(x_i)}{\sigma(x_i)} - \frac{\sum_{j=1}^{\rho} \rho_j}{\sigma(x_{i-j})} \mu(x_{i-j}) - \frac{\sum_{j=1}^{\rho} \rho_j}{\sigma(x_{i-j})} y_{i-j} \right]^2 \right\}
\]

\[
\approx -\frac{1}{2} \left\{ \sum_{i=1}^{n} \left[ \frac{y_i}{\sigma(x_i)} - \frac{a_0}{\sigma(x_i)} - \frac{a_1 x_i}{\sigma(x_i)} - \sum_{k=1}^{K} b_k |x_i - t_k|^p \right]^2 \right\}
\]

\[
\approx -\frac{1}{2} \left\{ \sum_{i=1}^{n} \left[ \frac{y_i}{\sigma(x_i)} - \frac{1}{\sigma(x_i)} \left( \sum_{i=1}^{n} \frac{p_j}{\sigma(x_{i-j})} x_i - \sum_{j=1}^{n} \frac{p_j}{\sigma(x_{i-j})} y_{i-j} \right) \right]^2 \right\}
\]

\[
\approx -\frac{1}{2} \left\{ \sum_{i=1}^{n} \left[ \frac{y_i}{\sigma(x_i)} - \frac{1}{\sigma(x_i)} \left( \sum_{i=1}^{n} \frac{p_j}{\sigma(x_{i-j})} x_i - \sum_{j=1}^{n} \frac{p_j}{\sigma(x_{i-j})} y_{i-j} \right) \right]^2 \right\}
\]

\[
\approx -\frac{1}{2} \left\{ \sum_{i=1}^{n} \left[ \frac{y_i}{\sigma(x_i)} - \frac{1}{\sigma(x_i)} \left( \sum_{i=1}^{n} \frac{p_j}{\sigma(x_{i-j})} x_i - \sum_{j=1}^{n} \frac{p_j}{\sigma(x_{i-j})} y_{i-j} \right) \right]^2 \right\}
\]

The maximum likelihood estimator of \( \theta \) cause the equation \( \frac{\partial l(\theta \mid y_i)}{\partial \theta} = 0 \) and we obtain

\[
\hat{\theta}_0 = \frac{\sum_{i=1}^{n} \delta \psi_i - \hat{\alpha}_n \sum_{i=1}^{n} \gamma_i \psi_i - \hat{\beta}_K \sum_{i=1}^{n} \phi_i \psi_i - \sum_{i=1}^{n} \phi_i \psi_i}{\sum_{i=1}^{n} \psi_i^2}
\]

(27)
\[
\hat{\alpha}_i = \frac{\sum_{j=1}^{n_i} \delta_i \gamma_{ij} - \hat{\alpha}_0 \sum_{j=1}^{n_i} \gamma_{ij} - \frac{\sum_{j=1}^{n_i} \hat{\beta}_i \sum_{j=1}^{n_i} v_{ij} - \sum_{j=1}^{n_i} \phi_i \gamma_{ij}}{\sum_{j=1}^{n_i} \gamma_{ij}^2}}{\sum_{j=1}^{n_i} \gamma_{ij}} 
\]

\[
\hat{\beta}_h = \frac{\sum_{i=1}^{n} \delta_i \nu_{ih} - \hat{\alpha}_0 \sum_{j=1}^{n_i} \delta_i \nu_{ij} - \hat{\alpha}_1 \sum_{i=1}^{n} \sum_{j=1}^{h \times K} \nu_{ij} \nu_{ij} - \sum_{j=1}^{n} \phi_i \nu_{ij}}{\sum_{j=1}^{n} \nu_{ij}^2}, \quad h = 1, 2, \ldots, K
\]

where

\[
\delta_i = \frac{y_i}{\hat{\sigma}(x_i)}, \quad \psi_i = \frac{1}{\hat{\sigma}(x_i)} - \sum_{j=1}^{p} \hat{\rho}_j \hat{\sigma}(x_{ij})
\]

\[
\gamma_i = \frac{x_i}{\hat{\sigma}(x_i)} - \sum_{j=1}^{p} \hat{\rho}_j \hat{\sigma}(x_{ij}) x_{ij}, \quad \phi_i = \sum_{j=1}^{p} \hat{\rho}_j \hat{\sigma}(x_{ij}) \nu_{ij}
\]

\[
v_i = \frac{|x_{ij} - \tau_i|^3}{\hat{\sigma}(x_i)} - \sum_{j=1}^{p} \hat{\rho}_j \hat{\sigma}(x_{ij}) |x_{ij} - \tau_i|^3
\]

\[
v_h = \frac{|x_{ij} - \tau_h|^3}{\hat{\sigma}(x_i)} - \sum_{j=1}^{p} \hat{\rho}_j \hat{\sigma}(x_{ij}) |x_{ij} - \tau_h|^3
\]

The log likelihood from (24) in term of \( \rho(.) \) is

\[
= -\frac{1}{2} \sum_{i=1}^{n} \left[ \left[ \frac{y_i - a_0 x_i}{\sigma(x_i)} - \frac{a_1 x_i}{\sigma(x_i)} \sum_{j=1}^{p} \hat{\beta}_j \frac{|x_{ij} - \tau_i|^3}{\sigma(x_i)} \right]^2 + \sum_{j=1}^{p} \hat{\rho}_j \left[ \frac{a_0 x_{ij}}{\sigma(x_{ij})} - \frac{a_1 x_{ij}}{\sigma(x_{ij})} \sum_{k=1}^{h \times K} \nu_{ijk} \frac{|x_{ijk} - \tau_h|^3}{\sigma(x_{ijk})} \right] \right]
\]
Next, the maximum likelihood of \( \rho_h, h = 1, \ldots, K \), cause the equation
\[
\frac{\partial }{\partial \theta} \log L(\theta | y_t) = 0
\]
and we obtain
\[
\hat{\rho}_h = \frac{\sum_{t=1}^{n} y_t \eta_h - \hat{a}_0 \sum_{t=1}^{n} \frac{\eta_h}{\hat{\sigma}^2(x_t)} - \hat{d}_i \sum_{t=1}^{n} x_t \eta_h - \sum_{k=1}^{K} \hat{b}_k \sum_{t=1}^{n} \frac{\eta_h | x_t - \tau_h |^3}{\hat{\sigma}^2(x_t)} - \frac{\sum_{t=1}^{n} \eta_h^2}{\hat{\sigma}^2(x_t)}}{\sum_{t=1}^{n} \eta_h^2}, h = 1, 2, \ldots, p
\]
(30)
where
\[
\eta_t = \frac{\hat{a}_0}{\hat{\sigma}(x_{t-j})} - \frac{\hat{a}_i x_{t-j}}{\hat{\sigma}(x_{t-j})} - \sum_{k=1}^{K} \hat{b}_k | x_{t-j} - \tau_h |^3 \frac{y_{r-j}}{\hat{\sigma}(x_{t-j})}
\]
\[
\eta_h = \frac{\hat{a}_0}{\hat{\sigma}(x_{t-j})} - \frac{\hat{a}_i x_{t-j}}{\hat{\sigma}(x_{t-j})} - \sum_{k=1}^{K} \hat{b}_k | x_{t-j} - \tau_h |^3 \frac{y_{r-j}}{\hat{\sigma}(x_{t-j})}
\]

If \( t \leq p \), the log-likelihood function denoted \( L(\theta | y_t) \) and the maximum likelihood estimator of \( (a_0, a_i, b_1, \ldots, b_K) \) are similar the NCHARN(PS) at (18), (19), and (20).

In particular, we use the concept of L-BFGS-B algorithm to estimate the maximum likelihood estimator for NCHARN(PS,AR) ,
\[
\hat{\theta} = (\hat{a}_0, \hat{a}_i, \hat{b}_1, \ldots, \hat{b}_K, \hat{a}_1, \hat{a}_2, \ldots, \hat{b}_K, \hat{\beta}_1, \ldots, \hat{\beta}_p, \hat{\rho}_1, \ldots, \hat{\rho}_p, \hat{p})
\]
to estimate \( \hat{\mu}(x_t) \) following:
\[
\hat{\mu}(x_t) = \hat{\mu}_1(x_t) + \hat{\mu}_2(y_{r-1}, \ldots, y_{r-p})
\]
\[
\hat{\mu}_1(x_t) = \hat{\mu}(x_t) - \sum_{j=1}^{p} \hat{\rho}_j \frac{\hat{\sigma}(x_t)}{\hat{\sigma}(x_{t-j})} \hat{\mu}(x_{t-j})
\]
\[
\hat{\mu}_2(y_{r-1}, \ldots, y_{r-p}) = \sum_{j=1}^{p} \hat{\rho}_j \frac{\hat{\sigma}(x_t)}{\hat{\sigma}(x_{t-j})} y_{r-j}
\]
where \( t \leq p \) the \( \hat{\mu}(x_t) \) and \( \hat{\sigma}(x_t) \) are calculated as the NCHARN(PS) at (21) and (22).
4.3 Forecasting Technique

We obtain forecast values of \( y_{n+1}, \ldots, y_{n+m} \) using the forecast trend and volatility based on the NCHARN(PS,AR) model:

\[
\hat{y}_{n+m} = \hat{\mu}^*(x_{n+m}) + \hat{\sigma}(x_{n+m})\hat{e}_{n+m}
\]
\[
\hat{e}_{n+m} = \sum_{j=1}^{\hat{p}} \hat{\rho}_j \hat{e}_{n+m-j}
\]
\[
\hat{\mu}^*(x_{n+m}) = \hat{\mu}_1(x_{n+m}) + \hat{\mu}_2(y_{n+m-1}, \ldots, y_{n+m-\hat{p}})
\]
\[
\hat{\mu}_1(x_{n+m}) = \hat{\mu}(x_{n+m}) - \sum_{j=1}^{\hat{p}} \hat{\rho}_j \frac{\hat{\sigma}(x_{n+m})}{\hat{\sigma}(x_{n+m-j})} \hat{\mu}(x_{n+m-j})
\]
\[
\hat{\mu}_2(y_{n+m-1}, \ldots, y_{n+m-\hat{p}}) = \sum_{j=1}^{\hat{p}} \hat{\rho}_j \frac{\hat{\sigma}(x_{n+m})}{\hat{\sigma}(x_{n+m-j})} y_{n+m-j}
\]

where \( n+m \leq \hat{p} \), the \( \hat{\mu}(x_{n+m}) \) and \( \hat{\sigma}(x_{n+m}) \) are estimated as the NCHARN(PS) model:

\[
\hat{\mu}(x_{n+m}) = \hat{\alpha}_0 + \hat{\alpha}_1 x_{n+m} + \sum_{k=1}^{K} \hat{\beta}_k \big| x_{n+m} - \tau_k \big|^3
\]

\[
\hat{\sigma}(x_{n+m}) = \sqrt{\log \{ \hat{\alpha}_0 + \hat{\alpha}_1 x_{n+m} + \sum_{k=1}^{K} \hat{\beta}_k \big| x_{n+m} - \tau_k \big|^3 \} }
\]

We choose the model by using Akaike’s information criteria, AIC, as

\[
AIC = -2h \ L(\theta) + 2p
\]

where \( L(\theta) \) is the likelihood function evaluated at the maximum likelihood estimates and \( p \) is the total number of the parameter estimated.

5. RESULTS AND DISCUSSION

In this section, we apply the methods described in Section 2, 3, and 4 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 414 records of the monthly volume of SERT index from January 1976 to June 2010 that can be found at http://www.set.or.th/th/market/market\_statistics.html.
Let $y_t$ denote the SERT Index of month $t$ where $t=1$ represents January of 1976 and $t=408$ represents December of 2009. The method of the NCHARN(PS) and NCHARN(PS,AR) are used to forecast future values of SERT index for January, 2010 to June, 2010 given in Table 1.

**Table 1.** The actual SERT Index, the forecast values, mean absolute deviation (MAD) for NCHARN(PS) and NCHARN(PS,AR).

<table>
<thead>
<tr>
<th>$m$</th>
<th>SERT</th>
<th>NCHARN(PS)</th>
<th>NCHARN(PS,AR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>696.55</td>
<td>549.17</td>
<td>697.62</td>
</tr>
<tr>
<td>2</td>
<td>721.37</td>
<td>555.67</td>
<td>746.08</td>
</tr>
<tr>
<td>3</td>
<td>787.98</td>
<td>562.22</td>
<td>785.45</td>
</tr>
<tr>
<td>4</td>
<td>763.51</td>
<td>568.81</td>
<td>813.55</td>
</tr>
<tr>
<td>5</td>
<td>750.43</td>
<td>575.45</td>
<td>836.42</td>
</tr>
<tr>
<td>6</td>
<td>797.31</td>
<td>582.15</td>
<td>855.50</td>
</tr>
<tr>
<td>MAD</td>
<td>-</td>
<td>187.28</td>
<td>37.09</td>
</tr>
<tr>
<td>AIC</td>
<td>-</td>
<td>1,804.76</td>
<td>1,754.76</td>
</tr>
</tbody>
</table>

From Table 1, it is apparent that the Mean Absolute Deviation (MAD) and AIC values by NCHARN(PS,AR) model smaller than that of the NCHARN(PS) model. Therefore, it should be noted that NCHARN(PS,AR) model performs significantly better than NCHARN(PS) since the NCHARN(PS,AR) model contains a past data, trend, and volatility in complicated model whereas the NCHARN(PS) is the simple model. In this respect the NCHARN(PS) is much easier to implement than the NCHARN(PS,AR) model but the forecast values are not accurate.

**Figure 1.** The actual SERT index (SERT 2010), forecast values (y.pred) and 95% prediction interval of future values (uci and lci) for NCHARN(PS,AR).
Figure 1 is shown the forecast values and 95% prediction interval of withheld SERT values obtained by NCHARN(PS,AR) in terms of minimizing the MAD. It follows from the figure that the NCHARN(PS,AR) provides predictive intervals that capture in some cases because there began Bangkok political crisis during April and May 2010, so the SERT index has been quickly changed in these periods.

6. CONCLUSIONS

In this article, we have investigated and compared the classical penalized spline, NCHARN(PS), and the penalized spline under autoregressive process, NCHARN(PS,AR), to estimate smooth unknown trend and smooth unknown volatility based on a nonparametric conditional heteroscedastic autoregressive nonlinear model. The NCHARN(PS,AR) model performs better than the NCHARN(PS) based method in terms of minimizing the MAD. The autoregressive process of error term, past of data, past of volatility is useful for prediction of future values shown that the past of time series data are correlated in each time points and effected on the prediction.

As a part of future studies, we can develop the NCHARN(PS,AR) to estimate trend, volatility, and autoregressive using the Bayesian approach.

ACKNOWLEDGEMENTS

The author would like to acknowledge Professor Dr. Sujit K. Ghosh for valuable suggestion.

REFERENCES


