BAYESIAN FORECASTING USING NON-LINEAR TIME SERIES MODELS.

Zafar Mahmud

Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan.

Abstract:

It is generally considered that the statistical forecasting methods are superior to the methods, which are based on Non-statistical Principles. Non-statistical methods are less sophisticated and simple to understand for an ordinary person. This paper presents the forecast based on purely statistical methods called Bayesian Forecasting. Asymmetric time series method has been used along with Kalman's Filter results. GARCH process has been used to estimate non-constant variances in the observation equation and in the system equation. Finally exchange rate of Pakistani rupees against UK pounds is used for forecasting prupose.

Keywords

Non-linearity; Bayesian; Forecasting; Asymmetric; Time series; GARCH Variances and Simplex Methods.

1. Introduction

There has been a boom in the development of statistical forecasting methods throughout the twentieth century. It has been noted through some comparative studies on univariate time series forecasting, "that no single procedure or class of procedure is superior in all circumstances"; see Newbold and Granger (1965), Reid (1975), Makridakis and Hibon (1979), Makridakis et al (1982) and Makridakis et al (1993). The statistical model considered in this study is the Nonlinear Dynamic Time Series Model, which may include the non normal and nonlinear forms and in particular it contains complex nonlinear model which is an extension a dynamic linear model discussed by Harrison and Stevens (1976). This extension of dynamic linear model is based on the innovations with one of different rules according to the whether the innovation is positive or negative, which was referred asymmetric time series by Wecker (1981). The asymmetric model is generally required in the situation where the distribution of the residual is squared with non-zero mean. In fitting asymmetric moving average model

(ASMA) with GARCH residuals, it is assumed that the residuals are independent but not identical. This indicate that residuals variance may depend on the past residuals and their past variances (for detail see Mahmud (1991) and Bollersleve (1986). Since the parameters in the Bayesian model are random variables, so these parameters in the mean part of the model may assume different values at time t^{-1} and t. Therefore the difference between the values of the parameters at time t^{-1} and t may or may not have the constant variance. The Kalman's results (1963) have been used to estimate the values of the random variables (parameters) at each t, which provides elegant recursive relationship for updating information concerning the parameter. The variances of these random variables (parameters) were noted to be non-constant and were estimated by the GARCH process. Further in forecasting the future value, when the future input is known and unknown has been considered for comparison. The prior values of the parameters have ben simulated by "The Simplex Method" given by Nelder and Mead (1965).

2. Dynamic Asymmetric Moving Average Non-linear Time Series Model

In this study the asymmetric moving average model is considered as a non-linear model. The dynamic nonlinear model (DNLM)[1] is a system of equations, which specifies how the observations of process or stochastically dependent on the current process parameters and secondly how the process evolves in time (for detail see Harrison (1976) and harvery (1979). This model is stated in terms of equally space intervals of time. Throughout the articles the following notions are the standards.

t = time index (1, 2, ..., n);

y_t = $(n \times 1)$ vector of process observations made at time t;

$\theta_{\sim t}$ = $(m \times 1)$ vector of process parameters at time t;

F_t = $(n \times m)$ matrix of independent variables, known at time t;

G = $(m \times m)$ known system matrix;

The dynamic nonlinear model has the observation equation and the system equation as under respectively.

$$\left.\begin{aligned} Y_t &= F_t\theta + \epsilon_t \\ \Theta_{\sim t} &= G\theta_{\sim t-1} + \omega_{\sim t} \end{aligned}\right\} \qquad (2.1)$$

with $\epsilon_t \sim N(0, \sigma^2_t)$ and $\omega_{\sim t} \sim N(0, \Omega_t)$, if $\Theta_{\sim t} = \Theta_{\sim t-1} = \Theta_{\sim}$ and ϵ_t S are independent. Then the conventional linear normal model of the classical statistics becomes a simple static case of the dynamic linear model (DLM). For constant F_t, the model is a state space representation in which the parameters may be interpreted as process level, process growth and so on. The process $y_t = F_t \Theta_{\sim t} + \epsilon_t$ specifies the stochastic dependence of the observation variable on the (unknown) process "parameters" it being assumed that this completely specifies the distrubution of y_t. The matrix G represents the fixed characteristics of the process, $\Theta_{\sim t} = G\Theta_{\sim t-1} + \omega_{\sim t}$, which defines the deterministic motion of the parameters from movement to the next, superimposed on which is a random component specified by $\omega_{\sim t}$ which we term 'distrubance vector' at time t. According to the study the observation equation is.

$$Y_t = \Theta^{+}_{t} \epsilon^{+}_{t-1} + \epsilon_t \qquad \ldots\ldots\ldots\ldots\ldots\ldots (2.2)$$

We call (2.2) as asymmetric moving average model of order one abbreviated as "ASMA (1)" where $\epsilon^{+}_{t} = \max(0, \epsilon^{+}_{t})$ and $\epsilon^{+}_{t} = \min(0, \epsilon_t)$ with $E(\epsilon^{+}_{t}) = -E(\epsilon^{-}_{t}) = \sigma_t/\sqrt{2\pi}$ and $Var(\epsilon^{+}_{t}) = Var(\epsilon^{-}_{t}) = \sigma^{2}_{t}/2$ at time t. For $\Theta^{+}_{t} = \Theta^{-}_{t} = \Theta_t$. This model reduced to "MA (1)" see Box and Jenkin (1976). Asymmetric moving average models along with its relationships with asymmetric moving average models have been discussed in Mahmud (1989, 1990). Accordingly if σ^{2}_{t} is non-constant at each t, then it may be estimated by the ARCH or GARCH process defined by Bollerslev (1986) below in (2.3). Bera and Higgins (1993) also gave a comprehensive discussion on the ARCH models

$$\sigma^{2}_{t} = \gamma_0 + \sum_{i=1}^{p} \gamma_i \epsilon^{2}_{t-1} + \sum_{j=1}^{t} \Phi \sigma^{2}_{t-j} \qquad \ldots\ldots\ldots\ldots\ldots (2.3)$$

Now we discuss the observation and system equations of ASMA (1) for the process defined in (2.1) as

$$F_t = [\epsilon^{+}_{t-1}\ \epsilon^{-}_{t-1}],\ \Theta_{\sim t} = \begin{pmatrix} \Theta^{+}_{t} \\ \Theta^{-}_{t} \end{pmatrix}$$

$$G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \omega_{\sim t} = \begin{pmatrix} \omega^{+}_{t} \\ \omega^{-}_{t} \end{pmatrix}$$

For asymmetric moving average of order q.

$\Theta_{\sim t} = \begin{pmatrix} \Theta^{+}_{\sim t} \\ \Theta^{-}_{\sim t} \end{pmatrix}$ and $F_t = [\epsilon^{+}_{t-1}\ \epsilon^{-}_{t-1}]$ where $\epsilon^{+}_{\sim t-1} = [\epsilon^{+}_{t-1} + \epsilon^{+}_{t-2}, \ldots \epsilon^{+}_{t-q}]$

$\epsilon^{-}_{\sim t-1} = [\epsilon^{-}_{t-1} + \epsilon^{-}_{t-2}, \ldots\ldots \epsilon^{-}_{t-q}]$, $\Theta^{+}_{\sim}$ and $\Theta^{-}_{\sim}$

have q component in each. The seasonal components may be introduced in

Θ^+_{-t}, Θ^-_{-t}, $\in^+_{-t-1}$ and $\in^-_{-t-1}$ as the last element. The quantities ω_t^+ and ω_t^- represents the disturbance terms corresponding to the random variables Θ_t^+ and Θ_t^- respectively for ASMA (1) in the system equation defined in (2.1). We know that $\omega_{-t} \sim N(0, \Omega_t)$. So the diagonal components of the square metrix Ω_t are Var $(\omega^+_t) = (\sigma^+_t)^2$ and Var $(\omega^-_t) = (\sigma^-_t)^2$ which may or may not have constant variances with respect to time. If the components of Ω_t have non constant variances with respect to time, then we need a process to estimate the components of Ω_t. The unknown compnents of Ω_t along with asymmetric moving average model at each t may be estimate by using ARCH or GARCH (see Bollersleve 1986) process as given in (2.3).

3. Estimation and Forecasting

For the estimation of the parameter included in the model, we shall use the powerful results given by Kalman (1963). At time $t = 0$, it is assumed that the prior information concerning the parmeters θ_{-0} is described in the form normal distrubition as $\theta_{-0} \sim N(m1_{-0}, \Sigma_0)$. Now the Kalman results provides elegant recurrence relationship for updating and revising the information's concerning the parameter vectors. Since the objective of the article is to provide Bayesian forecasting, therefore we can write the results in the form, most suitable for the application. The distribution of θ_{-0} prior to the 1st information is $N(m_{-0}, \Sigma_0)$ and the posterior distribution of θ_{-t} is also normally distributed i.e., $(\theta_{-t}/Y_t) \sim N(m_{-t}, \Sigma_t)$ where the values of m_{-t} and Σ_t can be obtained recursively from the Kalman results.

$$\text{If } m_{-0} = \begin{pmatrix} \theta^+_0 \\ \theta^-_0 \end{pmatrix} \text{ and } \Sigma_0 = \begin{pmatrix} \text{Var}(\theta^+_0) & 0 \\ 0 & \text{Var}(\theta^-_0) \end{pmatrix}$$

Let $\hat{Y}_t = F_t m_t$ then $\in_t = \hat{Y}_t - Y_t$, $R_t = G\Sigma_{t-1}G' + \Omega_t$, $\hat{\sigma}_{y,t} = F_t R_t F'_t + \sigma^2_t$

And $A_t = R_t F_t (\sigma^2_{y,t})^{-1}$ then $m_t = Gm_{t-1} + A_t \in_t$ and $\Sigma_t = R_t - A_t \sigma_{y,t}$

A_t, then the posterior of θ_{-t} at time t prior to the last observation is $N(m_t\ \Sigma_t) \sim t$. The quantities in A_t are explained. Since $\in_t$ is one step ahead forecast error (conditional), so $\hat{Y}_t$ and $\hat{\sigma}_{y,t}$ are the expectation and variance of Y_t (conditional on F_t) the quantity $A_t \in_t$ is refereed as Kalman's gain. The quantities $\in_t$ and A_t clearly works like the one step ahead forecast error within sample period and "smoothing constant" respectively in many conventional system (A_t is not in general constant and is in fact a matrix of order $m \times n$). Posterior density of θ_{-t} obtained $t = 1, 2, \ldots$ may help to find the $\in_t$ one step ahead forecast error within sample period and also may be used to find the maximum of the Joint posteriors at t = 1, 2, ..., n, which may help to estimate the parameters of vector $\theta_{\sim t}$ at time t = 1, 2, ... and the information may be used to infer the distrubition of the future observation Y_{t+k} (k = 1, 2, ...) and these predictions are purely extrapolative in nature. From the equations of the Nonlinear Dynamic Models. Dynamic Models (NLDM) we may write at time t, a future observation value as

$$Y_{t+k} = F_{t+k}\,\theta_{\sim t+k} + \epsilon_{t+k} \quad \text{.................}(3.1)$$

$$\theta_{\sim t+k} = G\theta_{\sim t+k-1} + \omega_{\sim t+k} \quad \text{.................}(3.2)$$

The prediction of future values of $Y_{\sim t+k}$ requires inferences about ~t and independent variable matrix F_{t+k}. Assuming

$$\hat{M}_{k,t} = E(\theta_{\sim t+k}) \,\big|\, (Y_t, F_t) \quad \text{.................}(3.3)$$

$$\Sigma_{k,t} = Var(\theta_{\sim t+k}) \,\big|\, (Yt, Ft) \quad \text{.................}(3.4)$$

At k = 0(3.3) and (3.4) are known from the Kalman filter algorithm. so from (3.3) and (3.4). We have

$$\hat{M}_{k,t} = G\hat{M}_{k-1,t} \text{ and } \hat{\Sigma}_{k,t} = G\hat{\Sigma}_{k,t-1}G' + \Omega_{t+k} \text{ for } k = 1, 2, \ldots\ldots$$

If F_{t+k} is known then according to Harrison and Stevens (1976)

$$\hat{Y}_{+tk} = F_{t+k}\hat{M}_{k,t} \text{ and } \sigma^2_{y,t} = F_{t+k}\hat{\Sigma}_{k,t}\, F_{t+k} + \sigma^2_{t+k}$$

where $\hat{M}_t$ and $\hat{\Sigma}_t$ are obtained recursively from (3.3) and (3.4). The future forecasting when the elements of F_{t+k}, σ^2_{t+k} and Ω_{t+k} are not known, may be replaced by their expected values available in the past in estimation process of this model. The order identification of asymmetric moving average model and GARCH order have been discussed in Mahmud (1990, 1991).

3. Application

The aim is to study the results the Pakistani data. This is to done by considering the exchange rate of Pakistani rupee versus UK pound, that covers the period from June 1991 to March 1998 (monthly basis). We used 1st 72 observations for estimation of the model's parameters and next 10 observations for testing the forecasting efficiency of the fitted model. The time series is non-stationary and also its variance changes over the time period of observations. These changes in the variance seem not be directly related to changes in the level and so it turns out that taking logarithms does not stabilize the variance. So the series used for the model fitting and forecasting is reduced to stationary by differencing the logarithmic series. The model fitted to this series is asymmetric series. The model fitted to this series is asymmetric moving average of order one ASMA (1) is chosen by using BIC criterion given by Akaike (1978, 1979)

With σ^2_t calculated at each t. In Bayesian analysis ~t is the random variable which may have nonconstant variance and is estimated by using the Kalman's results. We fitted the asymmetric moving average model with different specification, along with GARCH variance in observation and system equation given in (2.1). The prior parametric values were simulated by using simplex method explained by Nelder and Mead (1965). The simulated prior values of mean parameters and variance parameters are given as:-

$$m_{\sim 0} = \begin{pmatrix} \theta^+_0 \\ \theta^-_0 \end{pmatrix} = \begin{pmatrix} 0.0416 \\ 0.0410 \end{pmatrix} \text{ and } \Sigma_0 = \begin{pmatrix} 0.1552 & 0 \\ 0 & 0.1557 \end{pmatrix}$$

So the model fitted is described as under

$$Y_t = \theta^+_t \in^+_{t-1} + \theta \in^-_{t-1} + \in^-_t$$

Since Σ_t has nonconstant variance, estimated by using GARCH process is as.

$$\sigma^2_t = 0.07 + 0.154 \in^2_{t-1}$$

Since $\theta_{\sim t}$ is random variable so its components θ^+_t and θ^-_t are random variable in Bayesian models, the variances of θ^+_t and θ^-_t are found by the recursive process defined by the Kalman's alter. The variances of the random disturbances ω^+_t and ω^-_t corresponding to the parameters θ^+_t and θ^-_t are represented at Var (ω^+_t) = $(\sigma^+_t)^2$ and Var (ω^-_t) = $(\sigma^-_t)^2$ respectively, which have been estimated by using GARCH process, given as under.

$$(\sigma^+_t)^2 = 0.1125 + 0.1517\,(\omega^+_{t-1})^2 + 0.1539\,(\sigma^-_{t-1})^2$$
$$\text{and } (\sigma^-_t)^2 = 0.1517 + 0.1535\,(\omega_{t-1})^2$$

In diagnosytic checking it has been noted that ACF of the standrdized residuals and squared standaridized residuals remain with in $\pm$ 2σ limites. This indicates that standardized residuals are independent and identical. The mean of the standardized residuals is not significant from zero at 5% level of significance. The Bartlet test statistics also indicated that the variance of the standardized residuals is constant over the sample period. Also the measures of the Skewness and kurtosis indicated no Skewness and slightly leptokuritic (which is a feature of financial time series) respectively. The forecasting behavior of this model is when the future inputs are know and unknown are presented below.

Forecasting Behavior of the Model (Logarithmic Series)

Years	Month	Logarithmic Values	Forecast F_t known	Forecast F_t unknown
1997	June	4.19360	4.19430	4.19430
1997	July	4.21896	4.16400	4.17737
1997	August	4.17881	4.16400	4.13413
1997	September	4.17582	4.18939	4.13068
1997	October	4.24187	4.23588	4.19666
1997	November	4.31161	4.27499	4.26639
1997	December	4.29576	4.27193	4.25053
1998	January	4.27944	4.28754	4.23421
1998	February	4.28160	4.28347	4.23687
1998	March	4.29613	4.29351	4.25090

The post-Sample Predictive test (given by Harvey 1981 p-180 when F is known suggests that the observations with in the sample and post sample period are generated by the same model, indicates the suitability of the fitted model Similarly the forecasting behavior of the model when F_t is unknown is slightly under estimating the original values as given in the table above.

REFERENCES

1. Akaike, H. (1978); A Bayesian Analysis of minimum AIC procedure Ann. Inst. Statist. ***Math 30A, 9 - 14.***

2. Akaike, H. (1979); A Bayesian extension of the minimum AIC procedure of the Authoregressive model fitting. ***Biometrica, 66, 237 - 242.***

3. Box, G.E.P. & Jenkin, G.M (1976); Time Series Analysis Forecasting & Control. ***Holden-day, San Francisco.***

4. Bera. A.K. & Higgins. M.L (1993); ARCH Models, Properties, Estimatikon and Testing, Journal of Economic survey, ***Vol. 7, No. 4, 305 - 362.***

5. Bollerslev, Tim, 91986); Generalized Autoregressive Conditional Heteroscedasticity, ***Journal of Econometrics 31, 301 - 327.***

6. Harvey, A.C. and Philips, D.G.A (1979); Maximum Likelihood Extimajktion of regression models with Auto regressive Moving Average disturbance, ***Biometrika, 66,1, pp. 49 - 58.***

7. Harvey, A.C. (1981); The Econometric Analysis of Time Series, Philip A llan Ltd.

8. Harrison, P.J and Stevens, C.F. (1976); Bayesian Forecasting Journal of Royal Statistical Society, ***B, 38, 205 - 228.***

9. Kalman, R.E (1963); New Method in wiener filtering theory. In proceeding of the symposium on Engineering application of random function theory and probability. ***(J.L. Boldonoff and F. Kozin eds.) New York: Wiley.***

10. Ma Kridakis, S. and Hibon, M. (1979); Accuracy of Forecasting: An empirical investigation (with disc ussion), Journal of the Royal Statistical Society, ***Series A, 142, 97 - 142.***

11. Makridakis, S., Anderson, R. Carbone, r. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen and R.W Inkkler, (1982); The Accuracy of extrapolation (time series methods: Results of a forecasting competition, ***Journal of forecasting, 1, 111 - 153.***

12. Makridakis, S., C.Chatfield, M. Hibon, M. Lawrence, T.Mills, K. Ord and L.F. Simmons (1993); The M2 - Competition: A real time judge mentally based forecasting study, ***International journal of forecasting 9, 5 - 22.***

13. Mahmud, Z. (1989): Heteroscedasticity and Nonlinearity in Financial Time Series. unpublished Ph.D. thesis Univ. ***College of Wales, Aberystwyth, U.K.***

14. Mahmud, Z. (1990); Asymmetric Moving Average Time Series Models. Journal of Research (Science), ***Vol. 2, No. 3, pp. 5 - 11.***

15. Mahmud, Z. (1991); Time Series Models with GARCH errors. Karachi Univ. ***J.Sc. 19 (1 & 2), 111 - 110.***

16. Nelder, J.A and Mead, R. (1965); A Simplex Method for function Minimization Computer Journal, ***Vol. 7, pp. 308 - 313.***

17. Newbold, p and Granger, C.W.J (1974); Experience with Forecasting Univarate Time Series and combination of forecasts (with conclusion), Journal of the Royal statistical Society, ***Series A, 137, 131 - 165.***

18. Reid, D.J. (1975); A Review of Short Term Projection Technique, in H.A. Gordon ed., practical aspects of forecasting ***(Operational Research Society, London).***

19. Wecker, W.E (1981); Asymmetric time Series, Journal of American Statistical Association, ***Vol. 76, No. 373.***