

A Comparison of Bayesian Methods and Artificial Neural Networks for Forecasting Chaotic Financial Time Series

Tamer Shahwan¹ and Raed Said²

¹*Al Ain University of Science and Technology, College of Business Administration, Department of Accounting, Finance & Banking, P.O. Box: 64141, Al Ain, United Arab Emirates.*

Email Address: Tamer.shahwan@aau.ac.ae

²*Al Ain University of Science and Technology, College of Business Administration, Department of Management Information Systems, P.O. Box: 64141, Al Ain, United Arab Emirates.*

Email Address: Raed.Tawfek@aau.ac.ae

Received March 16, 2012; Revised April 20, 2012; Accepted April 25, 2012

Abstract: Most recent empirical work implies that the presence of low-dimensional deterministic chaos increases the complexity of the financial time series behavior. In this study we propose the Generalized Multilayer Perceptron (GMLP), and the Bayesian inference via Markov Chain Monte Carlo (MCMC) method for parameter estimation and one-step-ahead prediction. By out-of-sample prediction approach, these proposed methods are compared to autoregressive integrated moving average (ARIMA) models which have been used as a benchmark. The deterministic Mackey-Glass equation with errors that follow an ARCH (p) process (MG-ARCH (p)) is applied to generate the data set used in this study. It turns out that GMLP outperforms the other two forecasting methods using RMSE, MAPE, and MAE criteria of forecasting accuracy.

Keywords: ARIMA; Artificial Neural Networks (ANNs); Bayesian inference; MG-ARCH (p) model; one-step-ahead forecasting

1 Introduction

Obtaining accurate stock prices forecast is one of the main goals of financial and academic research institutions which they seek to achieve for the purpose of supporting key financial decisions such as selling and hedging. However, accurate forecasting of stock prices remains a major challenge under efficient market conditions (Shahwan, 2006). Much effort has been devoted over the past decades to the development of time series forecasting models. Traditionally, Autoregressive integrated moving average (ARIMA) models are considered as some of the most widely used linear models in time series forecasting because of their theoretical elaborateness and accuracy in short-term forecasting (Jhee and Shaw, 1996). However, ARIMA models cannot easily capture non-linear patterns resulting from the existence of a bounded rationality assumption in financial markets (McNelis, 2005). In recent years, however, interest in the use of artificial neural networks (ANNs) for forecasting and time series prediction has grown steadily (Zhang et al. 1998) as has interests in the applications of Bayesian inference in forecasting time series (Mendoza and De Alba, 2006).

In this paper, we will examine the ability of Bayesian inference via Markov Chain Monte Carlo (MCMC) method for forecasting time series using simulation data. MCMC is a sampling based

simulation technique that generates a dependent sample from a certain distribution of interest. Several schemes of implementing MCMC methods are widely used in Bayesian inference such as the Gibbs sampler introduced by (Geman and Geman, 1984) and Metropolis-Hasting method originally developed by (Metropolis et al. 1953) and further generalized by (Hastings, 1970). These two algorithms are simple to implement and are effective in practice when used for Bayesian inference (Surapaitookorn, 2007). The stochastic Mackey-Glass process is generated using Monte Carlo experiment since the chaotic time series has a lot of similarity to economic and financial time series (McNelis, 2005). In this paper we compare Bayesian estimation methods with artificial neural networks. ARIMA models will be used as a benchmark. The comparability or superiority of the proposed models will be investigated using Monte Carlo experiments. The main motivation of this approach is due to the existence of little empirical evidence regarding the performance of ANN, Bayesian estimation methods and ARIMA models under different characteristics of chaotic financial time series. Therefore, using a simulation study which covers a wide range of these characteristics will enable us to identify the conditions under which one of these models is superior to others.

The remaining part of the paper is organized as follows. Section 2 describes different proposed forecasting methods employed for time series forecasting. The numerical simulation in section 3 includes generation of the data using Mackey-Glass stochastic process, specification of the forecasting models and performance measures. The results are presented in section 4. Finally, section 5 gives concluding remarks and some suggestions for future work.

2 Forecasting Methods

2.1 ARIMA Models

Consider the following stochastic process $\{y_t, t \in N_+\}$ that can be expressed in terms of its conditional moments as follows (Campbell et al. 1997, p. 469) and (Shahwan, 2006, p. 8):

$$y_t = g(F_{t-1}) + \sqrt{h(F_{t-1})} \cdot \varepsilon_t \tag{1}$$

where $\varepsilon_t = a_t/\sigma_t$ is a standardized shock. a_t is a white noise series with a mean of zero and a variance σ_t^2 , F_{t-1} is the collection of $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$ and $(a_{t-1}, a_{t-2}, \dots, a_{t-a})$, $g(\cdot)$ represents a linear or nonlinear function, and $h(\cdot) = \sigma_t^2$. Accordingly, the stochastic process of the time series y_t will be a non-linear in mean if the $g(\cdot)$ is a non-linear function, whereas a time series with non-linear $h(\cdot)$ is said to be non-linear in variance.

A linear autoregressive moving average (ARMA) model of order (p, q) implies that the current value of y_t of the process can be expressed as a linear combination of its past values $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$ and a random shock series $(a_{t-1}, a_{t-2}, \dots, a_{t-a})$. Thus, the ARMA model can be expressed as follows (Tsay, 2005):

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i a_{t-i} \tag{2}$$

where $(\phi, \theta) = (\phi_0, \dots, \phi_p, \theta_1, \dots, \theta_q)$ are the parameters of the model. p and q are non negative integers. Since many time series are non-stationary, differencing one or more times is required. This leads to the well-known autoregressive integrated moving average (ARIMA) model. By using the back shift operator B with $B(y_t) = y_{t-1}$, ARIMA model can then be written as follows (Pindyck and

Rubinfeld, 1991):

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t, \quad (3)$$

with $|\phi_1 + \phi_2 + \dots + \phi_p| < 1$

where d is the level of differences used to change non-stationary time series into stationary time series.

Basically, the fitting of ARIMA (p, d, q) model to a given time series consists of the following three phases: (i) Model identification, by analyzing the behavior of the autocorrelation function (ACF) and the partial autocorrelation function (PACF), aims to determine the proper orders of (p, q) , (ii) Parameters estimation by using the maximum likelihood technique, and (iii) Forecasting of new values based on the estimated parameters. It is worthwhile to mention that ARIMA models cannot capture non-linear patterns in time series data. In the following section, we examine artificial neural networks, as a non-parametric data-driven approach with the capability of capturing non-linear patterns that expected to be existed in the data of volatile financial markets.

2.2 Artificial Neural Networks (ANNs)

The artificial neural networks (ANNs), as representative of a more general class of non-linear models, are probably one of the most frequently used tools in finance and economics. Well-known applications of this model include credit approval, bankruptcy prediction, and time series prediction [See Jensen (1992); Raghupathi et al. (1996) and Tam and Kiang (1996)]. With regard to one-step-ahead prediction \hat{y}_{t+1} , artificial neural networks determine the function (f) using the historical observations $(y_t, y_{t-1}, \dots, y_{t-n})$ with $(n = 1, \dots, N)$ denoting the number of input units as follows (Shahwan, 2006):

$$\hat{y}_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-n}) \quad (4)$$

An ANN model is defined in terms of subsidiary decisions such as selecting the architecture of the artificial neural networks, determining the number of hidden layers, the number of hidden neurons, the number of input nodes in the input layer, the type of activation function, the values of the learning rate and the way the data is to be divided into training, cross validation, and test sets (Shahwan, 2006). Based on Occam's razor principle, the simpler network will be selected to strike a balance between training performance and network complexity (Azoff, 1994). A more detailed review of guidelines, rules, and their implications to the ANN output when specifying the optimal ANN is provided by Azoff (1994); Zhang et al. (1998); and Shahwan (2006). In the current study, the Multilayer feedforward with jump connections will be used to conduct one-step-ahead forecasting. This type of neural networks can be considered as a generalization of the MLP (GMLP). Fig. 1 depicts the idea of GMLP where each input node in the input layer is directly connected to the output layer. The GMLP with a "tanh sigmoidal activation function" in the hidden layer and an "identity transfer function" in the output layer has the following structure (Tsay, 2005):

$$n_{k,t} = w_{k,0} + \sum_{i=1}^I w_{k,i} x_{i,t} \quad (5)$$

$$M_{k,t} = f(n_{k,t}) = \frac{e^{n_{k,t}} - e^{-n_{k,t}}}{e^{n_{k,t}} + e^{-n_{k,t}}} \quad (6)$$

$$y_t = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k M_{k,t} + \sum_{i=1}^I v_i x_{i,t} \tag{7}$$

where $x_{i,t}, i = 1, 2, \dots, I$, are input variables. $w_{k,0}$ and γ_0 are constants terms, $w_{k,i}$ are the synaptic weights of input variables and $n_{k,t}$ is a linear combination of these input variables observed at time $t, t = 1, 2, \dots, T$. Hence, the $n_{k,t}$ is squashed by the tanh sigmoid activation function and becomes $M_{k,t}$ at time t . Note that a set of k^* neurons $\{k\}$ can be found in the hidden and output layers, γ_k denotes the coefficient vector between the hidden and output layer, and v_i is the coefficient vector between the input and output layer. Thus, a feedforward network jump connection with a linear function in the output layer can be considered as a generalization of the linear regression model with non-linear terms (Shahwan, 2006). Consequently, if the underlying function between the input and the output is a pure linear, the coefficient sets γ and w will be zero, yielding a linear model [see Gonzalez (2000), Tsay (2005) and McNelis (2005)].

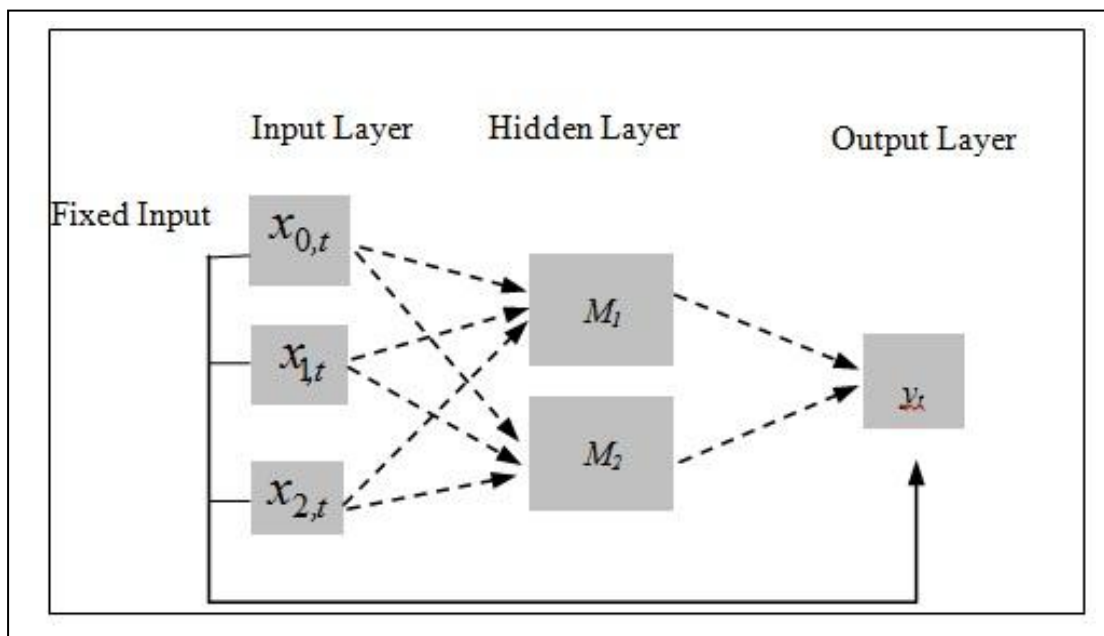


Fig. 1. Feedforward network with jump connection (Shahwan, 2006, p.23).

2.3 Markov Chain Monte Carlo (MCMC) Method

As an alternative to the classical ARIMA models, it is possible to apply Bayesian inference. De Alba and Mendoza (2007) proved that Bayesian procedures can be more effective, if only a small amount of past data is available, than the classical ARIMA models. It is also preferred to use it in the presence of cycles and trends in the time series data. In this section, we describe how to carry out Bayesian inference for the simulated data using a Gibbs sampling method. Following Koop (2003), Gibbs sampling for independent Normal-Gamma prior is implemented to obtain the Bayesian estimates. As mentioned in section (2.1), an autoregressive model of order p , $AR(p)$ as a linear combination of its p past values is defined as follows:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + a_t \tag{8}$$

where $(\beta_0, \beta_1, \dots, \beta_p)$ are the parameters of $AR(p)$ model. The Bayesian approach to the parameters estimation of a stochastic process starts by determining the likelihood function $p(y / \theta)$ where $y = (y_1, y_2, \dots, y_N)$ is the observed time series and $\theta = (\beta, h) = (\beta_0, \beta_1, \dots, \beta_p, h)$ is the vector of unknown parameters.

$$p(y / \beta, h) = \frac{1}{(2\pi)^{\frac{N}{2}}} \{h^{\frac{1}{2}} \exp[-\frac{h}{2}(\beta - \hat{\beta})' X' X(\beta - \hat{\beta})]\} \{h^{\frac{v}{2}} \exp[-\frac{hv}{2s^{-2}}]\} \tag{9}$$

Where $X = \begin{bmatrix} y_p & \dots & y_1 \\ y_{p+1} & \dots & y_2 \\ \vdots & & \vdots \\ y_{N-1} & \dots & y_{N-p} \end{bmatrix}$, $\hat{\beta} = (X'X)^{-1} X'y$, $s^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/v$, $v = N - p$, and $y = \begin{bmatrix} y_{p+1} \\ y_{p+2} \\ \vdots \\ y_N \end{bmatrix}$.

The estimates of the parameters h and β are performed by MCMC method via Gibbs sampling technique. The form of the above likelihood suggests that the natural conjugate prior is an independent Normal-Gamma. In particular, we assume $p(\beta, h) = p(\beta).p(h)$ with $p(\beta)$ being Normal and $p(h)$ being Gamma:

$$p(\beta) = \frac{1}{(2\pi)^{\frac{p}{2}}} |V|^{-\frac{1}{2}} \exp[-\frac{1}{2}(\beta - \underline{\beta})' V^{-1}(\beta - \underline{\beta})] \tag{10}$$

$$p(h) = c_G^{-1} h^{\frac{v-2}{2}} \exp(-\frac{hv}{2s^{-2}}) \tag{11}$$

where s^{-2} and \underline{v} are the prior mean and degrees of freedom of h , $\underline{\beta} = E(\beta / y)$ is the prior mean of β , and c_G is the integrating constant for the Gamma probability density function.

Bayes' theorem allows us to combine the likelihood function with the prior in order to form the conditional distribution of h and β given the observed data y , that is,

$$p(\beta, h / y) \propto \{ \exp[-\frac{1}{2}\{h(y - X\hat{\beta})'(y - X\hat{\beta}) + (\beta - \underline{\beta})' V^{-1}(\beta - \underline{\beta})\}] \} h^{\frac{N-v-2}{2}} \exp[-\frac{hv}{2s^{-2}}] \tag{12}$$

The aim of MCMC simulation is to generate a sample $\{\theta_{(j)} = (\beta_{(j)}, h_{(j)}), j = 1, 2, \dots, m\}$ from the conditional densities $p(h / \beta, y)$ and $p(\beta / h, y)$ obtained from the joint posterior density $p(\beta, h / y)$. This sample is then used to infer the point estimates of the parameters β and h . For instance, the Bayesian point estimates $\hat{\beta}$ and \hat{h} are given by:

$$\hat{\beta} = \frac{1}{m} \sum_{j=1}^m \beta_{(j)} \text{ and } \hat{h} = \frac{1}{m} \sum_{j=1}^m h_{(j)} \tag{13}$$

As mentioned above, we will use Gibbs sampler as a simulation strategy in our study. One of the main advantages of this sampler is that it is often easier to implement than any of the other MCMC methods. The Gibbs sampler is also flexible in the sense that its output may be used in order to make a variety of posterior and predictive inferences. Following (Geman and Geman, 1984), The Gibbs-Sampler algorithm is briefly described in the following steps:

- Step (1) Assign initial values to the parameters: $\beta_{(0)}$ and $h_{(0)}$;
- Step (2) Obtain a new observation $\beta_{(j+1)}$ from the conditional density $p(\beta/h_{(j)}, y)$;
- Step (3) Obtain a new observation $h_{(j+1)}$ from the conditional density $p(h/\beta_{(j+1)}, y)$;
- Step (4) Stop if the convergence of the Markov chains has been detected. Otherwise, do $j = j + 1$ and return to step 2.

After a sufficiently large number of iterations, the set of observations $(\beta_{(j)}, h_{(j)})$ converges and it can be treated as a sample from the joint posterior density $p(\beta, h/ y)$.

3 Numerical Simulation and Model Specification Methods

In the current simulation experiment, we aim to simulate the price behavior of financial time series using the stochastic Mackey-Glass process. The Mackey-Glass equation was originally developed for modeling white blood cells production (Calvo and Jabri, 2000). The prime motive in selecting this stochastic process is that real economical dynamics is a mixture of deterministic and stochastic chaos (Holyst et al. 2001). Following Kyrtsou and Terraza (2003), the discrete version of the deterministic Mackey-Glass equation is:

$$r_t = d \frac{r_{t-\tau}}{1 + r_{t-\tau}^c} - \lambda r_{t-1} \tag{14}$$

where r_t is the return of the time series. We must note that the choice of lags c and τ are vital in determining the dimensionality of the system.

In finance, asset return volatility exhibits volatility clustering in the sense that periods of high volatility tend to be followed by high volatility and periods of low volatility tend to be followed by low volatility (Poon, 2005, p. 7). Hence, the basic assumption of the current simulation is that the conditional variance of the stochastic Mackey-Glass process follows an autoregressive conditional heteroscedastic process of order one, ARCH (1). A time discretized realization of that process is:

$$y_t = y_{t-1} \cdot e^{r_t} \tag{15}$$

$$r_t = d \frac{r_{t-\tau}}{1 + r_{t-\tau}^c} - \lambda r_{t-1} + a_t \tag{16}$$

$$a_t = \sigma_t \cdot \varepsilon_t \tag{17}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \cdot a_{t-1}^2 \tag{18}$$

where y_t denotes the price at time t . α_0 is constant, σ is the volatility, α_1 is the weight assigned to a_{t-1}^2 , and ε_t is a random sample drawn from a standardized normal distribution with mean zero and a standard deviation of one.

The set of parameters used to generate the aforementioned stochastic process are $y_0 = 12.5$ as an initial price. Another important parameter that must be determined is the volatility σ . Following Shahwan (2006), the estimated volatility of hog prices in the German spot market is about 14.60%. Accordingly, σ_0^2 is set to be 0.0225 as an initial variance. Each value of $\tau = 1$, $c = 2$, $d = 2.1$, $\lambda = 0.05$, $\alpha_0 = 0.2$, and $\alpha_1 = 0.5$ are derived from Kyrtsou and Terraza (2003, p. 261).

The sample size for the generated data consists of 500 observations. The forecasting performance of the proposed model is assessed by an out-of-sample technique. Each time series is divided into 450 observations as a training set and 50 observations for testing. The training set is used for model specification and then the testing set is used to evaluate the established model. Three criteria are used to evaluate the accuracy of each model. The root mean square error (RMSE), Mean absolute percentage error (MAPE) and Mean absolute error (MAE) are employed to measure the forecasting error. These statistical measures of out-of-sample predictions are:

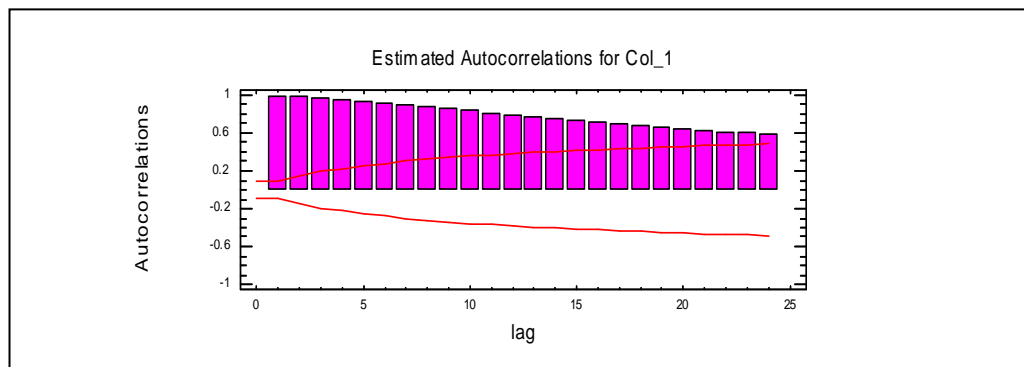
$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2} \quad (19)$$

$$MAPE = \frac{100}{T} \sum_{t=1}^T \left| \frac{\hat{y}_t - y_t}{y_t} \right| \quad (20)$$

$$MAE = \frac{1}{T} \sum_{t=1}^T |\hat{y}_t - y_t| \quad (21)$$

where y_t and \hat{y}_t are the actual and the predicted price, respectively, at time t . T is the number of observations in the test set.

The Structure of ARIMA model is determined through the following steps: (i) a natural logarithm is applied to each value of y_t as an attempt to stabilize the data set. (ii) Investigation of the time series stationarity by applying the augmented-Dickey Fuller test (ADF). ADF statistics is (-0.543) and lies inside the acceptance region at 5% level of significance. Therefore, we can not reject the presence of unit root which indicates the non-stationary of the time series. Therefore, the first order difference is applied. (iii) Analyzing the Autocorrelation (ACF) and the partial autocorrelation (PACF) for the time series as shown in fig. (2), each of p and q can be inferred. The best estimated ARIMA model for the data set has the structure (1,1,2). Without prejudging the nature of nonlinearity existed in our data set, the residuals of ARIMA model have been tested for the presence of nonlinearity using the BDS test¹. Following Kanzler (1999, p. 33), the dimensional distance of 1.5 has been selected to yield a better approximation as shown in Table (1). These results indicate that there is a non-linear structure in our data set. The significant evidence of nonlinearity implies that the use of nonlinear model such as ANN might be accurate in fitting the time series. Additionally, we test for the presence of GARCH effects using Engle's ARCH test and Ljung-Box Q-statistic. The results in Table (2) indicate that such effect exist in the data. Hence, the time series is nonlinear in terms of variance.



¹ The BDS test examines for (iid) versus general nonlinearity in the time series.

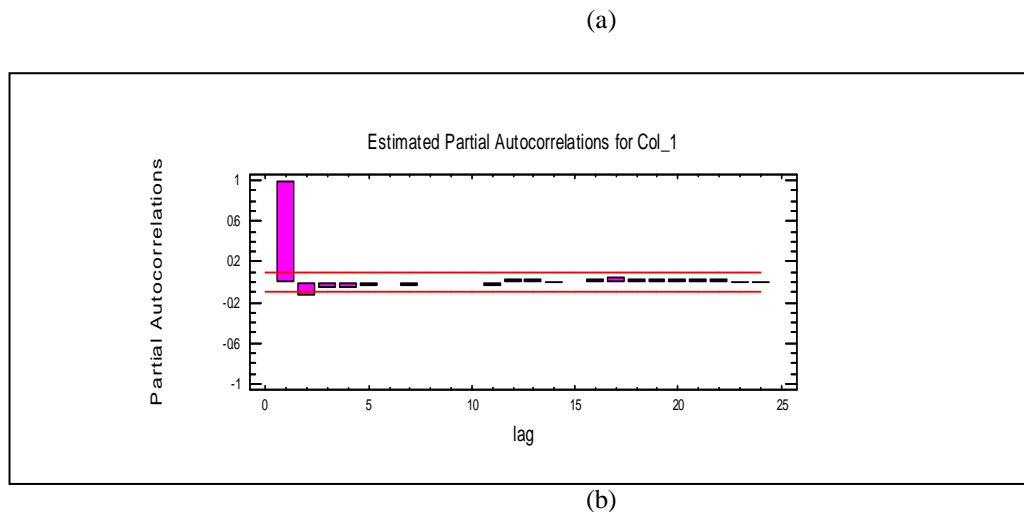


Fig. 2. ACF (a) and PACF (b) for the time series

Table 1. The BDS test on the residuals of ARIMA model for the Mackey-Glass series.

| Dimensional distance of 1.5 | Embedding dimension (m) | | | |
|-----------------------------|-------------------------|---------|---------|---------|
| | 2 | 3 | 4 | 5 |
| | 8.4377* | 9.3745* | 9.5924* | 9.3857* |

* indicates the rejection of null hypothesis of (i.i.d) at the 5 % significance level

Table 2. The Ljung-Box Q statistic and Engle's ARCH test for the Mackey-Glass series.

| Test | Lags | | | |
|---|--------|--------|--------|--------|
| | Q(5) | Q(10) | Q(15) | Q(20) |
| The Ljung-Box Q statistics on the residual of ARIMA | 2.35 | 13.38 | 18.92 | 25.03 |
| The Ljung-Box Q statistics on the squared residual of ARIMA | 92.34* | 129.3* | 131.7* | 137.3* |
| Engle's ARCH test | 70.13* | 80.15* | 84.47* | 86.77* |

* indicates statistical significance for the presence of GARCH effects at the 5% level

The specification of ANN model is now in turn, the generalized MLP network used has six inputs, one hidden layer and one output unit. A genetic algorithm is used to optimize numbers of the hidden nodes, the value of the learning rate, the momentum term and the weight decay constant. The hyperbolic tangent function is chosen as a transfer function between the input and the hidden layer. The identity transfer function connects the hidden with the output layer. The GMLP is trained by back propagation algorithm. To avoid the memorization problem "overfitting" during the training process of ANN model, a cross-validation approach will be applied. 20% of the training set (90 observations) had to be used for cross-validation in the back-propagation approach. Haykin (1999), Principe et al. (2000) and Shahwan (2006) illustrate that the use of cross-validation approach enhances the generalization capability of the neural network model. Batch updating is chosen as the sequence in which the patterns are presented to the network.

To set up MCMC methods, the actual observations of the time series y_t are transformed by applying a logarithmic transformation. Then, the transformed time series Z_t is normalized as

follows:

$$Z_t = \frac{\log y_t - \mu}{\sigma} \tag{22}$$

where μ and σ are the sample mean and standard deviation of the generated time series. The sample mean and variance of our data are (-4.9433) and (0.071692), respectively. We allowed the MCMC simulation to run $S_0 = 1000$ iterations in order to burn-in the Gibbs sampler and remove the effect of the starting values of β , i.e., $\beta_0 = (1, -4, -4.5)$, and then allowed it to run for an additional $S_1 = 9000$ iterations in order to generate a random sample from the posterior distribution. We set the initial draw for the error prediction to be equal to the inverse of OLS prediction estimate of σ^2 , i.e., $h_0 = 1/S^2 = 1/(0.07692)^2 \cong 194$.

To see whether the estimated results using MCMC methods are reliable or not we obtain what we call as MCMC diagnostics. Geweke (1992) suggested a convergence diagnostic (CD) given by;

$$CD = \frac{\hat{g}_{S_A} - \hat{g}_{S_C}}{\frac{\hat{\sigma}_A}{\sqrt{S_A}} + \frac{\hat{\sigma}_C}{\sqrt{S_C}}} \tag{23}$$

where S_A and S_C are the set of first and the last draws, respectively. Following Koop (2003), it has been found that setting $S_A = 0.1S_1$, and $S_C = 0.4S_1$ works well in many forecasting application. Let \hat{g}_{S_A} and \hat{g}_{S_C} be the estimate of $E(g(\theta)/y)$ using the first S_A replications after the burn-in and the last S_C replications, respectively. $\hat{\sigma}_A/\sqrt{S_A}$ and $\hat{\sigma}_C/\sqrt{S_C}$ are the numerical standard errors of CD estimates. The convergence of MCMC algorithm has occurred as CD is less than 1.96 in absolute value for all $\beta_0, \beta_1, \beta_3$. Fig. (3) shows the histogram of the samples generated via MCMC simulation using the Normal Gamma prior density.

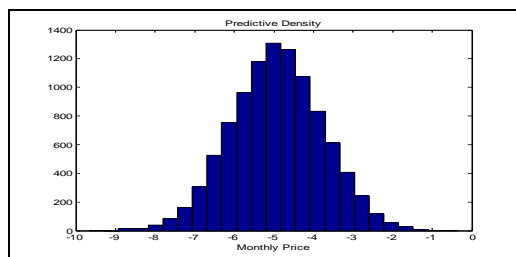


Fig. 3. The predictive density for MG- ARCH (1) series

4 Results

ARIMA model is estimated with the help of Statgraphics Plus software. The software package NeuroSolutions V5.05 was employed for the estimation of the Generalized Multilayer Perceptron (GMLP). The Bayesian estimates using the Normal-Gamma prior via the Gibbs-Sampling algorithm were carried by Koop's Matlab code which was slightly improved to be popular with our data set. We compare the accuracy of the proposed Bayesian method for forecasting time series with GMLP and ARIMA models using the out-of-sample approach. Table (3) shows the comparison of forecasting errors using different criteria for the proposed methods of our MG-GARCH (1) series.

The accuracy criteria of RMSE, MAPE, and MAE shown in table (3) indicate that the GMLP method outperforms the other two methods in one-step-ahead forecasting. Based on Morgan-

Granger-Newbold test, the difference in prediction errors of the three different forecasting methods are statistically significant at the 5% level. These results are with our expectation that GMLP will dominate the linear models when the stochastic process of our data set follows a more complex and nonlinear patterns. It is also remarkable to note that the estimated ARIMA $(1,1,2)$ model dominates the MCMC according to the above mentioned accuracy criteria. Our findings are compatible with the results of De Alba and Mendoza (2007) that standard forecasting procedures, like ARIMA, models will yield better forecasting than our proposed Bayesian method for lengthy time series.

Table 3. Forecasting errors of ARIMA, GMLP, and Bayesian methods for MG-ARCH (1) series

| Methods | RMSE | MAPE | MAE | Rank |
|----------|----------------|----------------|----------------|------|
| ARIMA | 0.1146 | 0.5522 | 0.0839 | 2 |
| GMLP | 0.0981* | 0.4994* | 0.0773* | 1 |
| Bayesian | 0.6193 | 3.1812 | 0.4854 | 3 |

*indicates statistical significance in the forecasting accuracy at the 5 % level

Bold letter indicates minimal error

5 Conclusions

This study has compared the Bayesian inference with artificial neural networks (ANNs) methods in forecasting chaotic financial time series. It has also applied traditional ARIMA models as a benchmark. The findings imply that ANNs are more relevant to fit a high-dimensional chaotic process than the Bayesian and ARIMA methods. However, ANNs demand a lot of specification procedure in determining its optimal structure. It is a time consuming model compared to ARIMA. Unfortunately, there is no fixed method for specifying of ANNs' parameters. The selection almost depends on heuristics. The study also confirms that there is no improvement in the forecast accuracy gained by using the MCMC method. Therefore, it is recommended to extend the MCMC model adopted in this study by using other prior distributions, and different MCMC algorithms. Moreover, there is a need to investigate the accuracy of these forecasting methods; MCMC; ANNs; ARIMA; using different simulated data as well as real data sets.

Acknowledgment

We gratefully acknowledge financial support from Al Ain University of Science and Technology. We also thank conference participants at the Fourth International conference on Mathematical Sciences ICM 2012- hold in United Arab Emirates University for helpful comments and advice.

References

- [1] E. M. Azoff, Neural Network Time Series Forecasting of Financial Markets. John Wiley & Sons Ltd, Chichester, England, (1994).
- [2] J. M. Bernardo, A. F. M. Smith, Bayesian Theory. John Wiley and Sons, Ltd., Chichester, (1994).
- [3] R. Calvo, M. Jabri, Benchmarking Bayesian Neural Networks for Time Series Forecasting, (2000). URL: <http://www.cs.cmu.edu>, 17.10.2011.
- [4] J. Y. Campbell, A. W. Lo, A. C. Mackinlay, The Econometrics of Financial Markets. Princeton University Press, Princeton, New Jersey, (1997).
- [5] G. Casella, E.I. George, Explaining the Gibbs Sampler. *The American Statistician*, 46 (1992) 167-174.

- [6] E. De Alba, M. Mendoza, Bayesian Forecasting Methods for Short Time Series. *The International Journal of Applied Forecasting*, 8 (2007) 41-44.
- [7] A. K. Dixit, R. S. Pindyck, Investment under Uncertainty. Princeton University Press, Princeton, New Jersey, (1994).
- [8] A. E. Gelfand, A. F. M. Smith, Sampling Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85 (1990) 398-409.
- [9] S. Geman, D. Geman, Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6 (1984) 721-741.
- [10] J. Geweke, Evaluating the Accuracy of Sampling- Based Approaches to the Calculation of Posterior Moments. In J. Bernardo, A. Dawid, A. Smith (Eds.), *Bayesian Statistics*. Oxford, Clarendon Press, 4th edition, (1992), 641- 649.
- [11] S. Gonzalez, Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models. Working paper, Department of Finance Canada, Canada, (2000). URL: <http://www.fin.gc.ca/access/wpliste.html>, 06.2.2012.
- [12] W. K. Hastings, Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57 (1970) 97-109.
- [13] S. Haykin, Neural Network: A Comprehensive Foundation. Second Edition, Prentice-Hall Inc., New Jersey, (1999).
- [14] J. A. Holyst, M. Zebrowska, K. Urbanowicz, Observations of Deterministic Chaos in Financial Time Series by Recurrence Plots, Can One Control Chaotic Economy? *The European Physical Journal B*, 20 (2001) 531-535.
- [15] H. L. Jensen, Using Neural Networks for Credit Scoring. *Managerial finance*, 18 (1992) 15-26.
- [16] W. C. Jhee, M. J. Shaw, Time Series Prediction Using Minimally Structured Neural Networks: An Empirical Test. In R. R. Trippi, E. Turban, (Eds.), *Neural Networks in Finance and Investing*. Revised Edition, Irwin, Chicago, (1996).
- [17] L. Kanzler, Very Fast and Correctly Sized Estimation of the BDS Statistics. Working paper, Department of Economics, University of Oxford, Christ church, UK, (1999).
- [18] G. Koop, Bayesian Econometrics. John Wiley and Sons Ltd, England, (2003).
- [19] C. Kyrtsov, M. Terraza, Is It Possible to Study Chaotic and ARCH Behaviour Jointly? Application of a Noisy Mackey-Glass Equation with Heteroskedastic Errors to the Paris Stock Exchange Returns Series. *Computational Economics*, 21 (2003) 257-276.
- [20] P. D. McNelis, Neural Networks in Finance: Gaining Predictive Edge in the Market. Elsevier Academic Press, (2005).
- [21] M. Mendoza, E. De Alba, Forecasting an Accumulated Series Based on Partial Accumulation II: A New Bayesian Method for Short Time Series with Stable Seasonal Patterns. *International Journal of Forecasting*, 22 (2006) 781-798.
- [22] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21 (1953) 1087-1092.
- [23] R. S. Pindyck, D. L. Rubinfeld, *Econometric Models and Economic Forecasts*. McGraw-Hill, Inc., New York, USA, (1991).
- [24] S.-H. Poon, *A Practical Guide to Forecasting Financial Market Volatility*. John Wiley & Sons, Ltd., New York, (2005).
- [25] J. C. Principe, N. R. Euliano, W. C. Lefebvre, *Neural and Adaptive Systems: Fundamentals through Simulations*. John Wiley & Sons, Inc., (2000).
- [26] W. Raghupathi, L. L. Schkade, B. S. Raju, A Neural Network Approach to Bankruptcy Prediction. In R. R. Trippi, E. Turban (Eds.), *Neural Networks in Finance and Investing*. Revised Edition, Irwin, Chicago, (1996).
- [27] T. Shahwan, Forecasting Commodity Prices Using Artificial Neural Networks. PhD Thesis, Humboldt-Universität Zu Berlin, Berlin, Germany, (2006).

- [28] W. Surapaitoorkorn, Bayesian Markov Chain Monte Carlo (MCMC) for Stochastic Volatility Model Using FX data. *7th Annual Hawaii International Conference on Business*, Honolulu, USA, (2007).
 - [29] K. Tam, M. Kiang, Predicting Bank Failures: A Neural Network Approach. In R. R. Trippi, E. Turban (Eds.), *Neural Networks in Finance and Investing*. Revised Edition, Irwin, Chicago, (1996).
 - [30] R. S. Tsay, *Analysis of Financial Time Series*. John Wiley and Sons, Inc., New Jersey, (2005).
 - [31] G. Zhang, B. E. Patuwo, M. Y. Hu, Forecasting with Artificial Neural Networks: The State of the Art. *International Journal of Forecasting*, 14 (1998) 35-62.
-