A comparison of Bayesian methods and Artificial Neural Networks for forecasting chaotic financial time series

Tamer Shahwan\textsuperscript{1,}\textsuperscript{*} and Raed Said\textsuperscript{2}

\textsuperscript{1}Department of Accounting, Finance & Banking, Al Ain University of Science and Technology, College of Business Administration, P.O. Box: 64141, Al Ain, United Arab Emirates, Tamer.shahwan@aau.ac.ae, Funded by Al Ain University of Science and Technology.

\textsuperscript{2}Department of Management Information Systems, Al Ain University of Science and Technology, College of Business Administration, P.O. Box: 64141, Al Ain, United Arab Emirates, Raeed.Tawfeq@aau.ac.ae, Funded by Al Ain University of Science and Technology.

ICM 2012, 11-14 March, Al Ain

\section*{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. \textbf{Keywords:} ARIMA, Artificial Neural Networks (ANNs), Bayesian Inference, MG-ARCH (p) Model, One-step-ahead forecasting

\section{INTRODUCTION}

Obtaining accurate prices forecasting derives the attention of many financial institutions and academic research to support financial decisions such as hold and sell decisions and hedging decisions. However, an accurate forecast of prices remains a major problem under the existence of efficient market hypothesis (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 Adjacent forecasting problems, artificial neural networks (ANNs) have also received increasing interest in forecasting and time series prediction (ZHANG et al. 1998). Additionally, there have been recent interests in Bayesian inference for forecasting time series (MENDOZA and DE ALBA, 2006).

\textsuperscript{*}To whom correspondence should be addressed.

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 (SURAPAITOOLKORN, 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). One proposed Bayesian estimation method is compared to artificial neural networks and the ARIMA model as an attempt to investigate the comparability or superiority of these models.

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, performance measures, and the results. Finally, section 4 gives concluding remarks and some suggestions for future work.

\section{FORECASTING METHODS}

\subsection{ARIMA Models}

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

\[ y_t = g(F_{t-1}) + \sqrt{h(F_{t-1})} \cdot e_t \] \hspace{1cm} (1)

where \( e_t = \frac{a_t}{\sigma} \) is a standardized shock. \( a_t \) is a white noise series with a mean of zero and a variance \( \sigma_t^2 \). 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 \( \{y_{t-1}, y_{t-2}, \ldots, y_{t-p}\} \) and a
random shock series \( \{e_{t}, e_{t-1}, \ldots, e_{t-q}\} \). Thus, the ARMA model
can be expressed as follows (Tsay, 2005):

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

where \( \phi_0, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q \) 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):

\[
(1 - \theta_1 B - \cdots - \theta_p B^p)(1 - \phi_1 B - \cdots - \phi_q B^q)y_t = \epsilon_t,
\]

with \( \epsilon_t \) is a white noise process with zero mean and
constant variance \( \sigma^2 \).

(3)

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 \((p,d,q)\). (ii) Parameters estimation by
using the maximum likelihood technique, and (iii) Forecasting of
new values based on the estimated parameters.

### 2.2 Artificial Neural Networks (ANNs)

The artificial neural networks (ANNs), as representative of a most
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
detection, and forecasting. 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 \((p,d,q)\). (ii) Parameters estimation by
using the maximum likelihood technique, and (iii) Forecasting of
new values based on the estimated parameters.

The artificial neural networks (ANNs), as representative of a most
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
detection, and forecasting. 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 \((p,d,q)\). (ii) Parameters estimation by
using the maximum likelihood technique, and (iii) Forecasting of
new values based on the estimated parameters.

The following structure

\[
M_{h,f} = f(h_{k,i}) = e^{n_{k,i} - \alpha_{k,i}} / e^{n_{k,i} + \alpha_{k,i}}
\]

\[
y_t = y_0 + \sum_{k=1}^{k} w_{k,i} M_{k,0} + \sum_{i=1}^{l} v_{i} x_{i,0}
\]

where \( x_{i,j}, i = 1,2, \ldots, I \), are input variables, \( w_{k,i} \) and \( y_0 \) are constants terms, \( w_{k,i} \) are the synaptic weights of input variables and \( n_{k,i} \) is a linear combination of these input variables observed at
time \( t, t = 1,2, \ldots, T \). Hence, the \( n_{k,i} \) is squashed by the tanh
sigmoid activation function and becomes \( M_{k,0} \) at time \( t \). Note
that a set of \( k \) neurons \( \{k\} \) can be found in the hidden and output
layers, \( \gamma_{k} \) denotes the coefficient vector between the hidden and
output layer, and \( \nu_{i} \) 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 generali-
ization of the linear regression model with non-linear terms
(Shahwan, 2006). Consequently, if the underlying function be-
 tween the input and the output is a pure linear, the coefficient sets
\( \gamma \) and \( \nu \) will be zero, yielding a linear model [see Gonzalez
(2000), Tsay (2005) and McNeils (2005)].

![Feedforward network with jump connection](image)

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

### 2.3 Markov Chain Monte Carlo (MCMC) Method

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 men-
tioned 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:

\[
y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + a_t
\]

\[
(8)
\]

where \( \beta_0, \beta_1, \ldots, \beta_p \) are the parameters of \( AR(p) \) model. The
Bayesian approach to the parameters estimation of a stochastic
process starts by determining the likelihood function \( p(y_0/\theta) \)
where \( y_0 = (y_1, y_2, \ldots, y_N) \) is the observed time series and
\( \theta = (\beta, h) = (\beta_0, \beta_1, \ldots, \beta_p, h) \) is the vector of unknown parameters.
\[ p(y/\beta, h) = \frac{1}{(2\pi)^\frac{1}{2}} \exp\left[ -\frac{1}{2} \left( \frac{y - \beta}{\sigma} \right)^2 \right] \] (9)

Where \( X = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{j-1} \end{bmatrix}, \beta = (XX')^{-1}XYs^2 = (y - X\hat{\beta})(y - X\hat{\beta})' \), \( v = N - p \), and \( y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_y \end{bmatrix} \). The estimates of the parameters \( h \) and \( \beta \) 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 \( p(\beta, h) = p(\beta) \cdot p(h) \) with \( p(\beta) \) being Normal and \( p(h) \) being Gamma:

\[ p(\beta) = \frac{1}{\left(p \sqrt{2\pi\nu}\right)} \exp\left[ -\frac{1}{2} \left( \beta - \nu \right)^2 \right] \] (10)

\[ p(h) = \frac{c_G^{-1}h^{\frac{\nu-2}{2}} \exp\left[ \frac{hv}{2} \right]}{\Gamma\left(\frac{\nu}{2}\right)} \] (11)

Where \( s^2 \) and \( \nu \) are the prior mean and degrees of freedom of \( h \), \( \beta = E(\beta / y) \) is the prior mean of \( \beta \), 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 \( \beta \) given the observed data \( y \), that is,

\[ p(\beta, h / y) \propto \exp\left[ -\frac{1}{2} \left( \beta - \mu \right)^2 \right] \cdot \exp\left[ -\frac{1}{2} \left( h - \nu \right)^2 \right] \] (12)

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

\[ \hat{\beta} = \frac{1}{m} \sum_{j=1}^{m} \beta_{(j)} \quad \text{and} \quad \hat{h} = \frac{1}{m} \sum_{j=1}^{m} h_{(j)} \] (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

3.1 Data Generation and Performance Measures

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 Kyrtou and Terraza (2003), the discrete version of the deterministic Mackey-Glass equation is:

\[ r_t = d \cdot \frac{r_t - r_0}{1 + r_t^\tau} - \lambda r_{t-1} \] (14)

where \( r_t \) is the return of the time series. We must note that the choice of lags \( c \) and \( \tau \) 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 (Poorn, 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:

\[ y_t = y_{t-1} \cdot e^{\sigma_t} \] (15)

\[ r_t = d \cdot \frac{r_t - r_0}{1 + r_t^\tau} - \lambda r_{t-1} + \alpha_t \] (16)

\[ \alpha_t = \sigma_t \cdot \varepsilon_t \] (17)

\[ \sigma_t^2 = \alpha_0 + \alpha_1 \cdot r_t^2 \] (18)

where \( y_t \) denotes the price at time \( t \). \( \alpha_0 \) is constant, \( \sigma \) is the volatility, \( \alpha_1 \) is the weight assigned to \( \alpha_{t-1} \) and \( \varepsilon_t \) 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 \( y_0 = 1.25 \) as an initial price, \( \sigma_0 = 0.0225 \) used as an initial variance, \( \tau = 1 \), \( c = 2 \), \( d = 2.1 \), \( \lambda = 0.05 \), \( \alpha_0 = 0.2 \), and \( \alpha_1 = 0.5 \).

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. 20% of the training set (90 observations) had to be used for cross-validation in the back-propagation approach. 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:

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

\[
MAPE = \frac{100}{T} \frac{1}{T} \sum_{t=1}^{T} \left| \frac{\hat{y}_t - y_t}{y_t} \right| \tag{20}
\]

\[
MAE = \frac{1}{T} \sum_{t=1}^{T} |\hat{y}_t - y_t| \tag{21}
\]

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.

### 3.2 The Forecasting Model’s Specifications

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 test. 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. 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. Batch updating is chosen as the sequence in which the patterns are presented to the network.

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.

<table>
<thead>
<tr>
<th>Dimensional distance of 1.5</th>
<th>Embedding dimension (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

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

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 is normalized as follows:

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

where \( \mu \) and \( \sigma \) 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 \( \beta \), 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 \( \sigma^2 \), i.e., \( h_0 = 1 / S_1^2 = 1 / (0.07692)^2 \) \( \geq \) 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:

\[
\text{CD} = \frac{\sum_{i=1}^{S} \hat{y}_{t+i} - \frac{1}{S} \sum_{i=1}^{S} \hat{y}_{t+i}}{\sigma^2/N} \tag{23}
\]
\[ CD = \frac{\hat{S}_A - \hat{S}_C}{\sigma_A + \sigma_C} \]  
(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 \( S_A = 0.15 \) and \( S_C = 0.45 \) works well in many forecasting application. Let \( \hat{S}_A \) and \( \hat{S}_C \) be the estimate of \( E(\theta)/\gamma \) using the first \( \hat{S}_A \) replications after the burn-in and the last \( \hat{S}_C \) replications, respectively. \( \sigma_A / \sqrt{\hat{S}_A} \) and \( \sigma_C / \sqrt{\hat{S}_C} \) 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 \( \beta_0, \beta_1, \beta_3 \).

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

<table>
<thead>
<tr>
<th>Test</th>
<th>Lags</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q(5)</td>
</tr>
<tr>
<td>The Ljung-Box Q statistic on the residual of ARIMA</td>
<td>2.35</td>
</tr>
<tr>
<td>The Ljung-Box Q statistic on the squared residual of ARIMA</td>
<td>92.34*</td>
</tr>
<tr>
<td>Engle’s ARCH test</td>
<td>70.13*</td>
</tr>
</tbody>
</table>

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

### 5 CONCLUSION

In this study we explore the usefulness of the Bayesian method and artificial neural networks for forecasting chaotic financial time series. Traditional ARIMA models are applied as a benchmark. Our primary findings imply that ANNs are beneficial to fit a high-dimensional chaotic process rather than other two methods. However, ANNs demand a lot of specification procedure in determining its optimal structure. So, it is a time consuming model comparing with ARIMA. This study also confirms that there is no improvement in the forecast accuracy gained by using the MCMC methods. As a suggestion for further improvement in the Bayesian methods, it is worthwhile to extend the presented model of MCMC by using other priors, and different MCMC algorithms. Moreover, there is a need to investigate the accuracy of all proposed models using different simulated data sets as well as real data sets.

### REFERENCES


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

<table>
<thead>
<tr>
<th>Methods</th>
<th>RMSE</th>
<th>MAPE</th>
<th>MAE</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>0.1146</td>
<td>0.5522</td>
<td>0.0839</td>
<td>2</td>
</tr>
<tr>
<td>GMLP</td>
<td>0.0981*</td>
<td>0.4994*</td>
<td>0.0773*</td>
<td>1</td>
</tr>
<tr>
<td>Bayesian</td>
<td>0.6193</td>
<td>3.1812</td>
<td>0.4854</td>
<td>3</td>
</tr>
</tbody>
</table>

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

Bold letter indicates minimal error.


