Linear and nonlinear dynamic systems in financial time series prediction

Salim Lahmiri[*]

[a]*Department of Computer Science, University of Quebec at Montreal, Montreal, Canada. ESCA School of Management, Casablanca, Morocco*

A R T I C L E I N F O	A B S T R A C T
	Autoregressive moving average (ARMA) process and dynamic neural networks namely the nonlinear autoregressive moving average with exogenous inputs (NARX) are compared by evaluating their ability to predict financial time series; for instance the S&P500 returns. Two classes of ARMA are considered. The first one is the standard ARMA model which is a linear static system. The second one uses Kalman filter (KF) to estimate and predict ARMA coefficients. This model is a linear dynamic system. The forecasting ability of each system is evaluated by means of mean absolute error (MAE) and mean absolute deviation (MAD) statistics. Simulation results indicate that the ARMA-KF system performs better than the standard ARMA alone. Thus, introducing dynamics into the ARMA process improves the forecasting accuracy. In addition, the ARMA-KF outperformed the NARX. This result may suggest that the linear component found in the S&P500 return series is more dominant than the nonlinear part. In sum, we conclude that introducing dynamics into the ARMA process provides an effective system for S&P500 time series prediction.

1. Introduction

The modeling and forecasting of time series is an appealing yet difficult task in real world problems. Indeed, the prediction of time series is challenging and has received a large attention over the most recent decades in many applications such as power load (Zhou et al., 2012), traffic flow (Castro-Neto et al., 2009), weather forecast (Awan & Awais, 2011), sunspot prediction (Mirmomeni et al., 2010), electricity price forecasting (Niu et al., 2010) and stock market prediction (Liu et al., 2012); to name a few. A number of techniques to predict time series have been introduced in the literature including the well known autoregressive–integrated-moving-average (ARIMA) processes and artificial neural networks (ANN). Introduced by Box and Jenkins (Box & Jenkins, 1976), ARIMA models originate from the autoregressive (AR) model, moving average (MA) model, and an integrated (I) part to make the data series stationary by differencing. The combination of AR and MA models gives the ARIMA process. However, ARIMA models assume a linear relationship between current and past values of a time series as well as with white noise. Indeed, they fit linear relations better than nonlinear relations.

* Corresponding author. Tel: +1-514-649-7771
E-mail addresses: lahmiri.salim@courrier.uqam.ca (S. Lahmiri)

In sum, the advantage of ARIMA is the fact that is it suitable to capture the stationary linear component, but it fails to model nonlinear structure. Therefore, approximations obtained by ARMA processes may not be adequate for complex nonlinear real-world data. In addition, ARIMA models are not capable to capture the dynamics of a time series. More precisely, the obtained AR and MA coefficients are static which make ARIMA a non dynamic system. On the other hand, ANN are nonlinear methods that can learn from patterns and capture hidden functional relationships in a given data even if the functional relationships are not known or difficult to identify (Zhang et al., 1998). In particular, they are capable of parallel processing of the information with no prior assumption about the model form. In particular, the ANN is determined by the characteristics of the data. In addition, ANN are adaptive; hence the network is capable to model non-stationary and dynamic data and systems. The ANN can be categorized into static and dynamic networks.

Static networks are feed-forward networks and do not have any delay or feedback elements, whilst the output of dynamic networks depends on the current, previous inputs or outputs of the network (Hagan et al., 1996). Although previous studies have reported that ARIMA models outperform ANN in terms of forecasting accuracy (Denton, 1995; Hann & Steurer, 1996; Taskaya & Casey, 2005) several recent papers in the literature have shown the superiority of ANN over ARIMA and that integration of ARIMA and ANN (for instance multi-layer back-propagation neural network) can be an effective way of improving upon their predictive performance (Zhang, 2003; Khashei & Bijari, 2010, 2011, 2012). ARIMA models were also combined with other machine learning techniques for time series prediction. For instance, Tseng et al. (2001) proposed a hybrid system called FARIMA that combines the fuzzy regression and ARIMA models to forecast foreign exchange market time series. Pai and Lin (2005) proposed a hybrid system based on ARIMA and support vector machines (SVM) for stock prices forecasting. Chen and Wang (2007) proposed hybrid model that incorporates seasonal autoregressive integrated moving average (SARIMA) model and support vector machines for seasonal time series forecasting, and Zhou and Hu (2008) proposed a hybrid modeling and forecasting approach based on grey and the Box–Jenkins autoregressive moving average models to predict gyro drift. In all previous papers, the effect of hybridization was significant on the improvement of the forecasting accuracy.

The purpose of this paper is twofold. First, it presents a hybrid model that combines ARIMA and Kalman filter (Kalman, 1960) to model and forecast time series. Since the Kalman filter is a well-known as an efficient recursive algorithm that estimates the state of dynamic systems, it used to update the ARIMA process coefficients. In other words, ARIMA is represented in a state space form. As a result, it is no longer static since its coefficients are dynamic and change over time. Unlike the standard ARIMA which is a static model, the main advantage of the ARIMA-Kalman model is to account for the dynamics of the time series and the possibility to improve the forecasting accuracy. Second, the nonlinear autoregressive moving average with exogenous inputs (NARX) which is a dynamic neural network suitable for modeling dynamic systems and time series prediction is used for comparison purpose. Indeed, the NARX has been successfully applied in modeling the dynamics of time series and complex systems (Menezes & Barreto, 2008; Arash & Atry, 2009; Pisoni et al., 2009; Pham et al., 2010). In particular, it is suitable for long-term prediction of univariate time series (Hagan et al., 1996). The performances of ARIMA-Kalman and NARX are compared in the context of financial time series prediction. For instance, the two models are applied to the problem of Standard&Poors (S&P500) returns forecasting.

The rest of the paper is organized as follows. In Section 2, ARIMA, Kalman filter and NARX are described. Experiments of the ARIMA, ARIMA-Kalman and NARX models are compared in Section 3. Finally, the conclusions are presented in Section 4.

2. Methodology

2.1 The ARMA model

The ARMA(p,q) model represents the future value of a variable as a linear function of past observations and random errors. For instance, the ARMA(p,q) process of a stationary time series Z is given by:

$$Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j}, \tag{1}$$

where t is time script, p is the order of the auto-regressive component, q is the order of the moving average component, ϕ and θ are coefficients to be estimated, and a are random errors which are assumed to be

independently and identically distributed with a mean of zero and a constant variance σ^2. The Box and Jenkins (1976) methodology includes three steps to identify the ARIMA parameters p and q. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the sample data are used to identify the orders p and q. Finally, the equation (1) is estimated by maximum likelihood estimation.

2.2 The ARMA-KF model

The Kalman filter (KF) is an adaptive method of state variables and data modeling and estimation. The KF is the minimum variance estimation in linear systems within the realm of stochastic dynamic mode (Nasseri et al., 2011). The method recursively estimates the state variables of a system after measuring them. Thus, the KF is capable of making estimations of past, present and future states based on linearity and Gaussian assumptions of the underlying process. Consider the ARMA(p,q) process in equation (1) and let $m = \max\{p, q\}$ and let $\phi_i = 0$ for $i > 0$ and $\phi_j = 0$ for $j > q$. Then, the ARMA(p,q) process can be rewritten as:

$$Z_t = \sum_{i=1}^{m} \phi_i Z_{t-i} + a_t - \sum_{i=1}^{m} \theta_i a_{t-i} \tag{2}$$

Thus, if the state space S_t is defined by:

$$S_t = \left(Z_{t|t-1}, Z_{t+1|t-1}, \ldots, Z_{t+m-1|t-1} \right)'. \tag{3}$$

Then, the observational equation is given by:

$$Z_t = [1, 0, \ldots, 0] S_t + a_t \tag{4}$$

and the state-transition equation that relates current and previous state variables is given by

$$S_{t+1} = FS_t + Ga_t \tag{5}$$

where,

$$F = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & & \\ \phi_m & \phi_{m-1} & \ldots & \phi_2 & \phi_1 \end{bmatrix}, \qquad G = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix}, \tag{6}$$

$$\psi_{m|} = \sum_{i=1}^{m} \phi_i \psi_{m-i|} - \theta_m$$

More details about state space representation of ARMA models by Kalman filter can be found in Harvey (1993).

2.3 The NARX neural network

The Nonlinear Auto-Regressive with Exogenous inputs (NARX) neural network (Leontaritis & Billings, 1985) is a class of discrete-time nonlinear systems that can represent a variety of nonlinear dynamic systems. In particular, the NARX network is a dynamic neural network that contains recurrent feedbacks from several layers of the network to the input layer. It can be mathematically represented as follows:

$$y(t+1) = f\left(y(t), y(t-1), \ldots, y(t-d_y), u(t), u(t-1), \ldots, u(t-d_u); W \right) \tag{7}$$

where $u(t)$ and $y(t)$ are respectively the input and the output of the system at time t, $d_u \geq 1$, $d_y \geq 1$, $d_y \geq d_u$, W is a weights matrix, and f is an unknown nonlinear function to be approximated by a multi-layer perceptron. As a result, the general NARX network equation can be written as:

$$y(t+1) = f_o\left(b_o + \sum_{h=1}^{Nh} w_{ho}.f_h\left(b_h + \sum_{i=0}^{d_u} w_{ih}u(t-i) + \sum_{j=0}^{d_y} w_{jh}y(t-j)\right)\right) \quad i=1,2,...,d_u; j=1,2,...,d_y; h=1,2,...,Nh \tag{8}$$

where f_h and f_o are the hidden and output functions, w_{ih}, w_{jh} and w_{ho} are the weights, and b_h and b_o are biases. All data fed to NARX network are normalized in [-1,1] interval.

3. Simulation results

The data analyzed in this study include the daily return of the S&P500 price index from January 27th, 2000 to January 31st 2008; a total of 2014 observations. Figure 1 exhibits the daily returns of the S&P500 calculated as first logarithmic differences. The data set is divided into two parts. The first group of the data (90% of observations) is used for training the model. The second group of the data (10% of observations) is used for testing. For evaluation purpose, two typical indicators the mean absolute errors (MAE) and the mean absolute deviation (MAD) are used to compare predicted and actual values to validate each forecasting system. The error can be estimated by MAE and MAD defined as:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|F_t - R_t|, \tag{9}$$

$$MAD = \frac{1}{N}\sum_{t=1}^{N}|F_t - \overline{F}|, \tag{10}$$

where R_t, F_t, and $\overline{F}$ are respectively the observed return, forecasted return and the average of forecasted return over the testing (out-of-sample) period; for example $t = 1$ to N. The simulation results are shown in Figure 2. A clear comparison of the three methods (ARMA, ARMA-KF, NARX neural network) for forecasting of the S&P500 is reported via MAE and MAD. The obtained results indicate that the prediction performance of the ARMA-KF as a dynamic linear prediction system is better than the performance of the single ARMA time series forecasting model. This shows that introducing dynamics into the standard ARMA process is an effective and promising approach to improve its performance in S&P500 time series forecasting.

In addition, it is quite clear that the NARX neural network which is a nonlinear dynamic system performs better than the standard ARMA process. However, the NARX neural network underperforms the ARMA-KF model. At first sight, it may be surprising that the dynamic nonlinear model obtains worst results in terms of MAE and MAD than the linear model ARMA-KF. This finding could be explained by the fact that the nonlinear structure of the considered time series could be sufficiently weak so that the data generating process could be better approximated by a dynamic linear process such as the ARMA-KF.
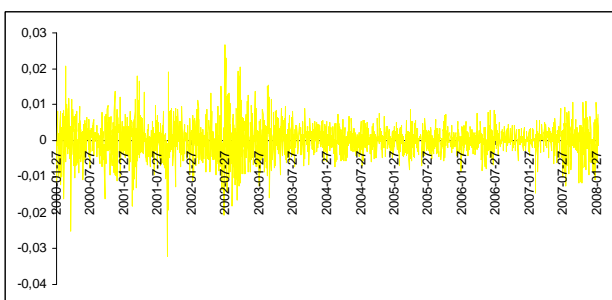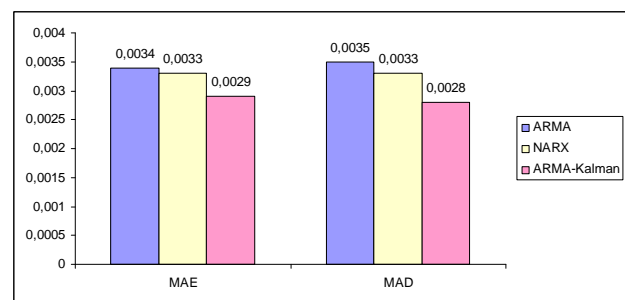

Fig. 1. S&P return series

Fig. 2. Performance measures

4. Conclusion

In this study, ARMA process, ARMA coupled with Kalman filter (ARMA-KF), and the Nonlinear Auto-Regressive with Exogenous inputs (NARX) neural network have been compared for financial time series prediction. In terms of empirical results, we have found that across these predictive systems for the test case of the S&P500 variations and on the basis of two evaluation criteria namely the mean absolute errors (MAE) and mean absolute deviation (MAD), the ARMA-KF which is a linear dynamic predictive system performs the

best. The findings demonstrate the predicting potential of the ARMA-KF system in financial applications. In particular, the results indicate that ARMA-KF can be a useful tool for S&P500 returns prediction. For future work, a combination of the ARMA-KF and NARX systems could be considered to check whether this hybrid system could improve the forecasting accuracy. Other market indices will be also considered.

References

Arash, A., & Atry, F. (2009). Multi-step ahead forecasts for electricity prices using NARX: a new approach, a critical analysis of one-step ahead forecasts. *Energy Conversion and Management*, 50 (3), 739–747.

Awan, M.S.K., & Awais, M.M. (2011). Predicting weather events using fuzzy rule based system. *Applied Soft Computing*, 11, 56–63.

Box, G. E. P., & Jenkins, G. (1976). Time series analysis, forecasting and control. San Francisco: Holden-Day.

Castro-Neto, M., Jeong, Y.-S., Jeong, M.-K., & Han, L.D. (2009). Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Systems with Applications*, 36, 6164–6173.

Chen, K.Y., & Wang, C.H. (2007). A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Systems with Applications*, 32, 254–264.

Denton, J.W. (1995). How good are neural networks for causal forecasting? *The Journal of Business Forecasting*, 14 (2), 17–20.

Hagan, M.T., Demuth, H.B., & Beale, M.H. (1996). *Neural Network Design*. Boston: PWS Publishing.

Hann, T.H., & Steurer, E. (1996). Much ado about nothing? Exchange rate forecasting: neural networks vs. linear models using monthly and weekly data. *Neurocomputing*, 10, 323–339.

Harvey, A.C. (1993). *Time Series Models.* 2nd ed., Harvester Wheatsheaf, Hemel Hempstead.

Jin, M., Zhou, X., Zhang, Z.M., & Tentzeris, M.M. (2012). Short-term power load forecasting using grey correlation contest modeling. *Expert Systems with Applications*, 39, 773–779.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. Trans ASME, Series D, *Journal of Basic Engineering*, 82, 35–45.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, 11, 2664–2675.

Khashei, M., & Bijari, M. (2012). A new class of hybrid models for time series forecasting. *Expert Systems with Applications*, 39, 4344–4357.

Khashei, M., Bijari, M. (2010). An artificial neural network (p,d,q) model for time series forecasting. *Expert Systems with Applications*, 37, 479–489.

Leontaritis, I.J., & Billings, S.A. (1985). Input–output parametric models for non-linear systems. Part I: Deterministic non-linear systems. *International Journal of Control*, 41 (2), 303–328.

Liu, C.-F., Yeh, C.-Y., & Lee, S.-J. (2012). Application of type-2 neuro-fuzzy modeling in stock price prediction. *Applied Soft Computing*, 12, 1348–1358.

Menezes, J.M.P., & Barreto, G.A, (2008). Long-term time series prediction with the NARX network: An empirical evaluation. *Neurocomputing*, 71, 3335–3343.

Mirmomeni, M., Lucas, C., Moshiri, B., & Araabi, B.N. (2010). Introducing adaptive neurofuzzy modeling with online learning method for prediction of time-varying solar and geomagnetic activity indices. *Expert Systems with Applications*, 37, 8267–8277.

Nasseri, M., Moeini, A., & Tabesh, M. (2011). Forecasting monthly urban water demand using Extended Kalman Filter and Genetic Programming. *Expert Systems with Applications*, 38, 7387–7395.

Niu, D., Liu, D., & Wu, D.D. (2010). A soft computing system for day-ahead electricity price forecasting. *Applied Soft Computing*, 10, 868–875.

Pai, P.F., & Lin, C.S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, 33, 497–505.

Pham, H.T., Tran. V.T., & Yang, B.-S. (2010). A hybrid of nonlinear autoregressive model with exogenous input and autoregressive moving average model for long- term machine state forecasting. *Expert Systems with Applications*, 37 (4), 3310-3317.

Pisoni, E., Farina, M., Carnevale, C., & Piroddi, L. (2009). Forcasting peak air pollution levels using NARX models. *Engineering Applications of Artificial Intelligence*, 22, 593–602.

Taskaya, T., & Casey, M.C. (2005). A comparative study of autoregressive neural network hybrids, *Neural Networks*, 18, 781–789.

Tseng, F.M., Tzeng, G., Yu, H.C., & Yuana, B.J.C. (2001). Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy Sets and Systems*, 118, 9–19.

Zhang, G., Patuwo, B.E., & Hu, M.Y. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, 14 (1), 35–62.

Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175.

Zhou, Z.J., & Hu, C.H. (2008). An effective hybrid approach based on grey and ARMA for forecasting gyro drift. Chaos. *Solitons and Fractals*, 35, 525–529.