

Research Article

Nonlinear Autoregressive Neural Network and Extended Kalman Filters for Prediction of Financial Time Series

Ghassane Benrhmach ¹, Khalil Namir,² Abdelwahed Namir,² and Jamal Bouyaghroumni¹

¹Laboratory of Analysis, Modelling and Simulation (LAMS), Faculty of Sciences Ben M'Sik, Hassan II University, P.O. Box 7955, Sidi Othman, Casablanca, Morocco

²Laboratory of Information Technology and Modelling, Faculty of Sciences Ben M'Sik, Hassan II University, P.O. Box 7955, Sidi Othman, Casablanca, Morocco

Correspondence should be addressed to Ghassane Benrhmach; ghassane.benrhmach@gmail.com

Received 27 January 2020; Revised 10 March 2020; Accepted 23 March 2020; Published 27 April 2020

Academic Editor: Man Leung Wong

Copyright © 2020 Ghassane Benrhmach et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Time series analysis and prediction are major scientific challenges that find their applications in fields as diverse as finance, biology, economics, meteorology, and so on. Obtaining the method with the least prediction error is one of the difficult problems of financial market and investment analysts. State space modelling is an efficient and flexible method for statistical inference of a broad class of time series and other data. The neural network is an important tool for analyzing time series especially when it is nonlinear and nonstationary. Essential tools for the study of Box-Jenkins methodology, neural networks, and extended Kalman filter were put together. We examine the use of the nonlinear autoregressive neural network method as a prediction technique for financial time series and the application of the extended Kalman filter algorithm to improve the accuracy of the model. As application on a real example, we are analyzing the time series of the daily price of steel over a 790-day period for establishing the superiority of this method over other existing methods. The simulation results using MATLAB and R software show that the model is capable of producing a reasonable accuracy.

1. Introduction

The prediction plays an important role in the management of several areas, among them the economic domain. Forecasting financial time series that are very noisy and nonstationary is a major problem for financial operators to invest at their best profit. The goal is to model relationships linking data to predict short- or medium-term values. Several techniques available for time series analysis assume the linearity of the relationships between variables, among them the Box and Jenkins [1, 2] methodology, one of the most popular methods of time series modelling through five stages: (1) choosing a class of models to represent the series, (2) the identification of the type of model, (3) the estimation of the coefficients of the identified models, (4) validation of the chosen model, and (5) the forecast of the series over a given horizon.

In reality, the financial series present significant irregularities so we resort to more complex techniques such as the extended Kalman filter (EKF) [3, 4] which consists of a set

of mathematical equations to model nonlinear relations. The Kalman filter has been applied in econometrics for the case where a deterministic system is unknown and must be estimated from the data, see for example Engle and Watson (1987). The Kalman filter algorithm proved to be an additional tool to improve model output [3].

Among the most promising methods are neural networks that have been introduced to solve complex classification problems. They are characterized by their ability to learn (supervised or not) from examples, then generalize to data that have not been presented to them. Neural networks can be seen as a black box that learns to map input models to appropriate output models. Learning is accomplished by changing the neuron connection weights to improve the desired matching between the input and output patterns.

After the introduction of simplified neurons by McCulloch and Pitts in 1943 [5], as models of biological neurons, some years later, exactly in 1957, Rosenblatt's Perceptron was endowed with the apprenticeship called "Perceptron Rule."

But the disadvantage of perceptron is that all the functions that it realizes are linearly separable. It was in 1969 that Minsky and Papert approached the disadvantage of the perceptron in their book "Perceptrons" [6] which will lead to a break of research in the next decade.

In the eighties, we saw the explosion of artificial intelligence techniques, with the two articles of physicist Hopfield [7]. In recent years, neural networks have found wide use in time series modelling (Chakraborty et al., 1992, Weigend Gershenfeld 1993, Gencay 1993, Hoptroff 1993).

Most neural networks used in economic forecasts are organized in layers, so we speak of MLP (multilayered perceptron) [8, 9]. In the field of finance, neural networks give good results in prediction.

This article begins with a section that describes time series analysis and gives an idea about the neural networks.

The second section presents a description of the extended Kalman filter and the proposed combination between the extended Kalman filter and the neural networks. The application of different studied models and the proposed model to the daily price of steel and the comparison between them is presented in the third section. The results and their discussion are discussed in the fourth section.

2. Time Series and Neural Network

2.1. Time Series. A time series is a parameter that changes over time (price, cost, turnover, stocks, etc); it is the set of observations of a quantity ordered according to their indices. In the following, we will have a noted series:

$$Y_t \text{ with } t = 1, \dots, n, \text{ with } n \in \mathbb{N}^*. \quad (1)$$

The study of a time series makes it possible to analyze, describe, and explain a phenomenon over time and to draw consequences for decision-making. One of the main objectives of the time series study is prediction which consists of predicting the future values of the series from its observed values.

The notion of stationarity is indispensable for the analysis of time series. A stationary series Y_t is a series whose properties are unchanged by the change of time. We are led to the following definition.

Definition 1. A stochastic process $(Y_t, t \in \mathbb{Z})$ is (weakly) stationary, if for any finite sequence of instants $t_1, \dots, t_k, k \in \mathbb{N}^*$, and for any integer t , the joint law of $Y_{t_1+t}, \dots, Y_{t_k+t}$ does not depend on t .

Definition 2. A process $(Y_t, t \in \mathbb{Z})$ is stationary, if:

- (1) $\forall t \in \mathbb{Z}, E[Y_t] = \mu$ (independent of t),
- (2) $\forall t \in \mathbb{Z}, E(Y_t^2) < \infty$ (independent of t),
- (3) $\forall t \in \mathbb{Z}, \forall k \in \mathbb{Z}, \text{cov}(Y_t, Y_{t+k}) = \gamma(k)$ (independent of t).

In sum, if the statistical characteristics of the stochastic process studied vary during the measurement period, it is said that the latter is nonstationary. Stationarity can be summarized as temporal homogeneity.

2.2. Neural Networks. Network learning takes place as the weights are adjusted along the layers, according to the relationship between the inputs and the desired outputs. One of the most basic models is multilayer perceptron (MLP) network, which is widely used in the approximation of nonlinear functions that describe complex relationships between independent and dependent variables in many applications.

Multilayer perceptron (MLP) was first introduced to solve complex classification problems. But because of their universal approximation property [9], they were quickly used as nonlinear regression models and then for time series modelling and forecasting.

However, the estimation and identification of these models use sophisticated techniques and it is not easy to determine the correct architecture. Indeed, these models are by definition overparametrized, the error functions to be minimized have many local minima, and the implementation is often difficult.

The nonlinear autoregressive neural network (NAR) as shown in Figure 1 can be trained to predict a time series from that series past values $Y(t-1), Y(t-2), \dots, Y(t-d)$ called feedback delays, with d is the time delay parameter.

The network is created and trained in an open loop, using the real target values as a response and making sure of greater quality being very close to the true number in training. After training, the network is converted into a closed loop and the predicted values are used to supply new response inputs to the network. A nonlinear autoregressive neural network applied to time series forecasting, describe a discrete, nonlinear autoregressive model that can be written in this form:

$$Y_t = h(Y_{t-1}, Y_{t-2}, \dots, Y_{t-d}) + \varepsilon_t. \quad (2)$$

The function $h(\cdot)$ is unknown in advance, and the training of the neural network is aimed at approximating the function by means of the optimization of the network weights and neuron bias.

So a model (NAR) is defined precisely by an equation of the type

$$Y_t = \alpha_0 + \sum_{j=1}^k \alpha_j \phi \left(\sum_{i=1}^a \beta_{ij} Y_{t-i} + \beta_{0j} \right) + \varepsilon_t, \quad (3)$$

where a is the number of entries, k is the number of hidden layers with activation function Φ , and β_{ij} is the parameter corresponding to the weight of the connection between the input unit i and the hidden unit j , α_j is the weight of the connection between the hidden unit j and the output unit, and β_{0j} and α_0 are the constants that correspond, respectively, to the hidden unit j and the output unit.

The optimization of the architecture is aimed at reducing as much as possible the number of synapses (weights) and

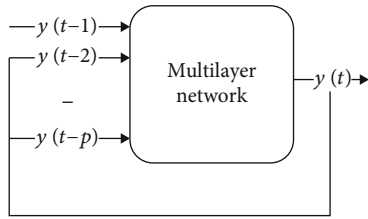


FIGURE 1: Architecture of the nonlinear autoregressive neural network.

neurons in order to reduce the complexity of the network, improve computing times, and maintain the generalization capabilities. Concerning the optimization of the network architecture, two main approaches have been proposed in the literature:

- Selection approach: consists of starting with the construction of a complex network that contains a large number of neurons, then this approach is to try to reduce the number of unnecessary neurons and remove redundant connections during or at the end of learning
- Incremental approach: we start with the simplest possible network, then we add neurons or layers, until we have an optimal architecture

An effective approach is to estimate the prediction error using a set of data that was not used to construct the predictor, i.e., not used for learning. This dataset is called a test set.

Divide the dataset into three kinds of target timesteps as follows:

- Training: these datasets are presented to the network during training and the network is adjusted according to its error
- Validation: these datasets are used to measure network generalization and to halt training when generalization stops improving
- Testing: these datasets have no effect on training and so provide an independent measure of network performance during and after training.

3. Extended Kalman Filter

Kalman filtering (KF) is a technique that gives estimates of unknown variables using a series of measurements containing statistical noise. It is a recursive way of doing things that process new data as they arrive being suited for inline real-time processing. The KF can only work with linear equations, When the system under consideration is nonlinear, the extended Kalman filter (EKF) is applied.

A short description of the EKF is given below. Consider a nonlinear system described by the following 2 equations:

$$\begin{aligned} y_{t+1} &= f(y_t) + u_t, \\ z_t &= h(y_t) + r_t, \end{aligned} \quad (4)$$

where y_t is a vector that describes the system state, z_t is the observation vector (values obtained through a direct measurement of the system), u_t is the process noise vector, r_t is the measurement noise vector, $f(\dots)$ is a nonlinear function that gives the state transition of the system, and $h(\dots)$ is the observation (nonlinear) function.

Kalman filter gives a method for the recursive estimation of the unknown state y based on all observation values z up to time t [10].

Kalman filter can be used for the improvement of neural network forecasts. The evolution of the filter from time $t-1$ to t is described by the following equations:

- (1) Initialization:

$$y_t^a = y_0 \text{ with } P_0 \text{ is the error covariance matrix}$$

- (2) The t -th predictor step:

$$y_t^p = f(y_{t-1}^a) \quad (5)$$

$P_t^p = J_f(y_{t-1}^a) \cdot P_{t-1} \cdot J_f^k(y_{t-1}^a) + Q_{t-1}$ where $Q_t = E[u_t \cdot u_{t-1}^k]$ is the process noise covariance matrix.

- (3) The t -th corrector step:

$$\begin{aligned} y_t^a &= y_t^p + K_t \cdot (z_t - h(y_t^p)), \\ K_t &= P_t^p \cdot J_h^k \cdot (J_h(y_t^p) \cdot P_t^p \cdot J_h^k(y_t^p) + R_t)^{-1} \text{ where } R_t = E[r_t \cdot r_t^k], \\ P_t &= (I - K_t \cdot J_h(y_t^p)) \cdot P_t^p, \end{aligned} \quad (6)$$

where a is the actual value of the variable, p for the predicted value, J_f is the Jacobian matrix of the $f(\dots)$ function, and J_h is the Jacobian matrix of $h(\dots)$ function and K_t is a matrix called the Kalman gain that arranges how easily the filter adjusts to possible new conditions or alternations of the type of data.

3.1. Proposed Prediction Model. Financial time series are very noisy and nonstationary, the presence of these two constraints pushed us to propose a model of prediction (NAR-EKF) which is a combination between the extended Kalman filter (EKF) and the multilayer perceptron (MLP) nonlinear autoregressive neural network (NAR).

The main idea of forecasting time series using the extended Kalman filter and neural networks (NAR-EKF) is to use the data processed by the extended Kalman filter of the Y_t series as the input for the nonlinear autoregressive neural network (described in Section 2.2), according to the following steps:

- Step 1: a set of historical data is collected

- Step 2: these data are evaluated using the extended Kalman filter to create a version with good approximation properties
- Step 3: nonlinear autoregressive neural network models are trained and validated
- Step 4: the created neural network model is used to predict the data

3.2. *Evaluation Criteria.* Various performance indices can be used to evaluate the model performances. In this study, performance indicators such as mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) used to assess model results.

- (i) $MAE = 1/N \sum_{i=1}^N |(y_i - \hat{y}_i)|$,
- (ii) $RMSE = \sqrt{1/N \sum_{i=1}^N (y_i - \hat{y}_i)^2}$,
- (iii) $R^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 / \sum_{i=1}^N (y_i - \bar{y}_i)^2$,
- (iv) With $\bar{y}_i = 1/N \sum_{i=1}^N y_i$.

4. Analysis of a Real World Time Series

4.1. *Description of Data.* The data used in this study represent the daily price of steel between 2013 and 2016. This data was collected from the <https://www.investing.com/> website

Figure 2 shows the price of steel per day between January 2013 and March 2016. A nonstationary series is one which generally increases (decreases) with time; the observation of this figure shows that the curve decreases steadily over time, thus revealing the presence of a long-term trend. The nonstationarity of this series (Figure 1) is confirmed by the Dickey-Fuller test ($p.value = 0.645 > 5\%$) and Phillips-Perron ($p.value = 0.649 > 5\%$) [11, 12].

To overcome this problem, the series of daily price of steel is transformed into a daily series of logarithmic difference of the price of steel. The stationarity of the series of daily returns of steel price (Figure 3) is confirmed by the Dickey-Fuller test ($p.value = 0, 01 < 5\%$).

4.2. *Application of Box-Jenkins Analysis.* In this section, we will model the daily steel price. The data was centred and reduced and then used to develop a descriptive model. In order to find the most suitable model to best explain the data, we used the Box-Jenkins methodology to estimate the parameters that characterize the series. The most suitable models for the daily price of steel are an ARMA (1,0) and an ARMA-GARCH(1,1). The estimation of the parameters of ARMA (1,0) can be carried out by several methods. In this study, they were estimated using the maximum likelihood method for the dataset [13, 14].

4.3. *Application of Neural Networks and EKF.* We have chosen to apply a nonlinear autoregressive neural network model to our time series. The concept is forecasting by using the data which is preprocessed through the extended Kalman filter as the input for the neural network. The financial time series is high fluctuation and time varying, and the extended

Kalman filter has a good dynamic real-time tracking characteristics. The advantage that the EKF provides is smoothing and denoising the time series. After this, smoothed time series might be predicted using the nonlinear autoregressive neural networks.

- Training: 570 observations are used for training the network
- Validation: 114 observations are used to measure network generalization
- Testing: the last 76 observations are used for testing the network

Some preliminary tests made it possible to define the structure of the network. A hidden layer of 25 neurons was used, and the activation function for the hidden layer and the output neuron is the Sigmoid function ($f(x) = 1 / [1 + \exp(-x)]$). We evaluated the effect of the time delay parameter d on the performance of the training process, evaluated using the mean-squared error (MSE) and the coefficient of determination R , which is a goodness-of-fit measure for linear regression between the target and the predictions. We set d from 1 to 6. The NAR-EKF model presents a very accurate fit and a small MSE independent of the value of d over all 6 trials.

Training of the network use Levenberg-Marquardt back-propagation. Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples. Other specifications of the NAR-EKF model are mentioned in Table 1 below.

In order to enable a comparison between the classical methods, extended Kalman filter, nonlinear autoregressive neural network, and the combination between the extended Kalman filter and the nonlinear autoregressive neural network, we adopted the same evaluation criteria RMSE and MAE.

The results from a real example (daily price of steel) shown in Table 2, proof that the combination of nonlinear autoregressive neural network model with the extended Kalman filter allows to produce high precision of prediction than that obtained by tested methods. This result is coherent with the literature that state nonlinear autoregressive neural network models are less sensitive to long-term time dependencies and presenting better learning capabilities [12].

5. Conclusions

This paper proposes a combination of a nonlinear autoregressive neural network model with the extended Kalman filter for predicting the financial time series. The results were compared according to the calculation of the root mean-squared error (RMSE) and the mean absolute error (MAE) for the six proposed models.

The combination of the two methods (NAR and EKF) for the prediction of the financial series (which are in step of daily time) seems able to improve the forecasts of the series studied by bringing the possibility of a continuous correction.

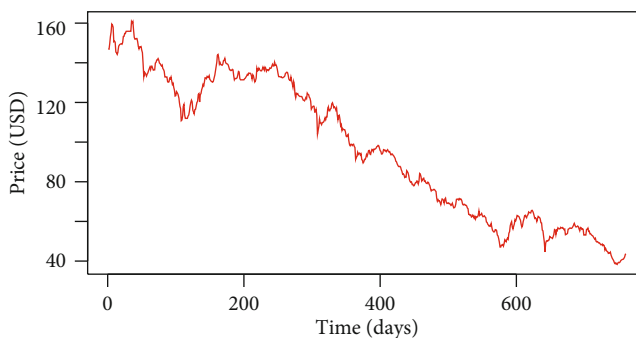


FIGURE 2: Daily price of steel from 2013 to 2016.

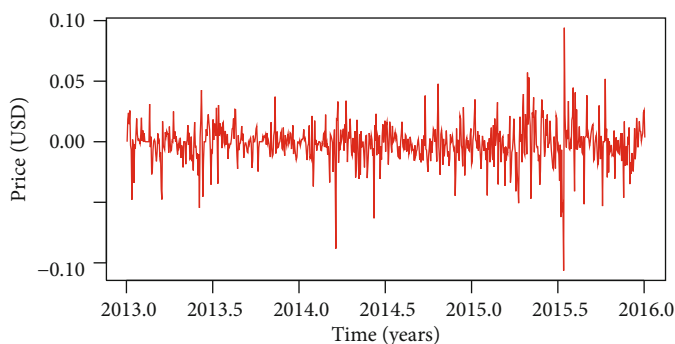


FIGURE 3: Series of daily returns of steel price.

TABLE 1: NAR-EKF parameter settings.

No.	Parameter	Data/technique used
1	Number of input neuron(s)	One
2	Number of output neuron(s)	One
3	Number of hidden layer(s)	One
4	Number of neurons in hidden layer(s)	25
5	Number of training, validation and test dataset	570, 114, and 76 observations
6	Transfer function	Sigmoid function
7	Error function(s)	Mean-squared error (MSE) function

TABLE 2: Error comparison.

	ARMA	ARMA-GARCH	EKF	NAR	NAR-EKF
RMSE	0.0492365	0.0231953	0.0190617	0.0150147	0.0096924
MAE	0.0579361	0.0186651	0.0154820	0.0134820	0.0084524

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] G. FBOX and J. M. JENKINS, *Time Series Analysis: Forecasting and Control*, Holdenday, San Fransisco, 1970.
- [2] R. Y. Aragon, *Séries Temporelles*, Springer, Paris, 1st edition, 2011.
- [3] G. Galanis, P. Louka, P. Katsafados, I. Pytharoulis, and G. Kallos, “Applications of Kalman filters based on non-

- linear functions to numerical weather predictions,” *Annales Geophysicae*, vol. 24, no. 10, pp. 2451–2460, 2006.
- [4] P. J. Bolland and J. T. Connor, “A robust non-linear multivariate Kalman filter for arbitrage identification in high frequency data,” *Neural Networks in Financial Engineering (Proceedings of the NNCFM-95)*, A.-P. N. Refenes, Y. Abu-Mostafa, J. Moody, and A. S. Weigend, Eds., pp. 122–135, World Scientific, 1996.
- [5] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [6] M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry*, M.I.T. Press, Cambridge, Mass, 1969.
- [7] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proceedings of the National Academy Of Sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.
- [8] B. Krose and P. V. D. Smagt, *An Introduction to Neural Networks*, University of Amsterdam, 1996.
- [9] A. Fiordaliso, “Une application des réseaux de neurones artificiels MLP à la prévision du prix d’une option négociable,” *Économie & Prévision*, vol. 127, no. 1, pp. 47–62, 1997.
- [10] M. Martin, “Kalman filtering of a seasonal time series. Application and forecasting of electricity consumption,” *Review of Applied Statistics*, vol. 47, pp. 69–86, 1999.
- [11] E. Cesar and B. Richard, *Les Séries Temporelles*, Université de Versailles, Saint-Quentin-en-Yvelines, 2006.
- [12] B. Dola, *Problèmes économétriques d’analyse des séries temporelles à mémoire longue. Économétrie de la finance*, Université Pantheon-Sorbonne - Paris I, Thèse de doctorat, 2012.
- [13] I. Alon, M. Qi, and R. J. Sadowski, “Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods,” *Journal of Retailing and Consumer Services*, vol. 8, no. 3, pp. 147–156, 2001.
- [14] H. Liu and J. Shi, “Applying ARMA–GARCH approaches to forecasting short-term electricity prices,” *Energy Economics*, vol. 37, pp. 152–166, 2013.