To Appear: Journal of Modern Applied Statistical Methods, 2008

A Weighted Moving Average Process for Forecasting

Shou Hsing Shih
Chris P. Tsokos

Department of Mathematics and Statistics
University of South Florida, USA

Abstract

The object of the present study is to propose a forecasting model for a nonstationary stochastic realization. The subject model is based on modifying a given time series into a new *k*-time moving average time series to begin the development of the model. The study is based on the autoregressive integrated moving average process along with its analytical constrains. The analytical procedure of the proposed model is given. A stock XYZ selected from the Fortune 500 list of companies and its daily closing price constitute the time series. Both the classical and proposed forecasting models were developed and a comparison of the accuracy of their responses is given.

Keywords: ARIMA; Moving Average; Stock; Time Series Analysis

Introduction

Time series analysis and modeling plays a very important role in forecasting, especially when our initial stochastic realization is nonstationary in nature. Some of the interesting and useful publications related to the subject area are Akaike (1974), Banerjee et al. (1993), Box et al. (1994), Brockwell and Davis (1996), Dickey and Fuller (1979), Dickey et al. (1984), Durbin and Koopman (2001), Gardner et al. (1980), Harvey (1993), Jones (1980), Kwiatkowski et al. (1992), Rogers (1986), Said and Dickey (1984), Sakamoto et al. (1986), Shumway and Stoffer (2006), Tsokos (1973), Wei (2006).

The subject of the present study is to begin with a given time series that characterizes an economic or any other natural phenomenon and as usual, is nonstationary. Box and Jenkins (1994) have introduced a popular and useful classical procedure to develop forecasting models that have been shown to be quite effective. In the present study we introduce a procedure for developing a forecasting model that is more effective than the classical approach. For a given stationary or nonstationary time series, $\{x_t\}$, we generate a k-day moving average time series, $\{y_t\}$, and our developmental process begins.

We review certain basic concepts and analytical methods that are essential in structuring the proposed forecasting model. The review is based on the autoregressive integrated moving average processes. The accuracy of the proposed forecasting model is illustrated by selecting from the list of Fortune 500 companies, company XYZ, and considering its daily closing prices for 500 days. We develop the classical time series model for the subject information along with the proposed process. A statistical comparison based on the actual and forecasting residuals is given, both in tabular and graphical form.

The Proposed Forecasting Model: *k*-th Moving Average

Before introducing our proposed forecasting model, we shall first define several important mathematical concepts that are essential in developing the analytical process. It is known that one cannot proceed in building a time series model without conforming to certain mathematical constrains such as stationarity of a given stochastic realization. Almost always the time series that we are given are nonstationary in nature and then, we must proceed to reduce it into being stationary. Let $\{x_t\}$ be the original time series. The difference filter is given by

$$(1-B)^d, \qquad (1)$$

where $B^j x_t = x_{t-j}$, and *d* is the degree of differencing of the series.

In time series analysis, the primary use for the *k*-th moving average process is for smoothing a realized time series. It is very useful in discovering a short-term, long-term trends and seasonal components of a given time series. The *k*-th moving average process of a time series $\{x_t\}$ is defined as follows:

$$y_t = \frac{1}{k}\sum_{j=0}^{k-1} x_{t-k+1+j} \ , \tag{2}$$

where $t = k, k+1, ..., n$.

It can be seen that as *k* increases, the number of observations *k* of $\{y_t\}$ decreases, and $\{y_t\}$ gets closer and closer to the mean of $\{x_t\}$ as *k* increases. In addition, when $k = n$, $\{y_t\}$ reduces to only a single observation, and equals to μ , that is

$$y_t = \frac{1}{n}\sum_{j=1}^{n} x_j = \mu \, . \tag{3}$$

We proceed to develop our proposed model by transforming the original time series $\{x_t\}$ into $\{y_t\}$ by applying (2). After establishing the new time series, usually nonstationary, we begin the process of reducing it into a stationary time series. Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. introduced the useful KPSS Test (1992) to check the level of stationarity of a time series. We apply the differencing order *d* to our new time series $\{y_t\}$ for $d = 0,1,2,...$, then verify the stationarity of the series with the KPSS test until the series become stationary. Therefore, we can reduce the nonstationary time series into a stationary one after a proper number of differencing. We then proceed the model building procedure of developing the proposed forecasting model.

After choosing a proper degree of differencing *d*, we can proceed the model building process by assuming different orders for the autoregressive integrated moving average model, ARIMA(p,d,q), also known as Box and Jenkins method, where (p,d,q) represent the order of the autoregressive process, the order of differencing and the order of the moving average process, respectively. The ARIMA(p,d,q) is defined as follows

$$\phi_p(B)(1-B)^d y_t = \theta_q(B)\varepsilon_t\,, \tag{4}$$

where $\{y_t\}$ is the realized time series, ϕ_p and θ_q are the weights or coefficients of the AR and MA that drive the model, respectively, and ε_t is the random error. We can write ϕ_p and θ_q as

$$\phi_p(B) = (1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)\,, \tag{5}$$

and

$$\theta_q(B) = (1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)\,. \tag{6}$$

In time series analysis, sometimes it is very difficult to make a decision in selecting the best order of the ARIMA(p,d,q) model when we have several models that all adequately represent a given set of time series. Hence, Akaile's information criterion (AIC) (1974), plays a major role when it comes to model selection. AIC was first introduced by Akaike in 1973, and it is defined as follows:

$$\text{AIC}(M) = -2 \ln [\text{maximum likelihood}] + 2M, \tag{7}$$

where M is the number of parameters in the model and the unconditional log-likelihood function suggested by Box, Jenkins, and Reinsel (1994), is given by

$$\ln L(\phi,\mu,\theta,\sigma_\varepsilon^2) = -\frac{n}{2}\ln 2\pi\sigma_\varepsilon^2 - \frac{S(\phi,\mu,\theta)}{2\sigma_\varepsilon^2}\,, \tag{8}$$

where $S(\phi,\mu,\theta)$ is the unconditional sum of squares function given by

$$S(\phi,\mu,\theta) = \sum_{t=-\infty}^{n}[E(\varepsilon_t|\phi,\mu,\theta,y)]^2 \tag{9}$$

where $E(\varepsilon_t|\phi,\mu,\theta,y)$ is the conditional expectation of ε_t given ϕ,μ,θ,y.

The quantities $\hat{\phi}$, $\hat{\mu}$, and $\hat{\theta}$ that maximize (8) are called unconditional maximum likelihood estimators. Since $\ln L(\phi,\mu,\theta,\sigma_\varepsilon^2)$ involves the data only through $S(\phi,\mu,\theta)$, these unconditional maximum likelihood estimators are equivalent to the unconditional least squares estimators obtained by minimizing $S(\phi,\mu,\theta)$. In practice, the summation in (9) is approximated

by a finite form

$$S(\phi,\mu,\theta)=\sum_{t=M}^{n}[E(\varepsilon_t|\phi,\mu,\theta,y)]^2 \tag{10}$$

where M is a sufficiently large integer such that the backcast increment $|E(\varepsilon_t|\phi,\mu,\theta,y)-E(\varepsilon_{t-1}|\phi,\mu,\theta,y)|$ is less than any arbitrary predetermined small ε value for $t\le-(M+1)$. This expression implies that $E(\varepsilon_t|\phi,\mu,\theta,y)\cong\mu$; hence, $E(\varepsilon_t|\phi,\mu,\theta,y)$ is negligible for $t\le-(M+1)$.

After obtaining the parameter estimates $\hat{\phi}$, $\hat{\mu}$, and $\hat{\theta}$, the estimate $\hat{\sigma}_\varepsilon^2$ of σ_ε^2 can then be calculated from

$$\hat{\sigma}_\varepsilon^2=\frac{S(\hat{\phi},\hat{\mu},\hat{\theta})}{n}. \tag{11}$$

For an ARMA(p,q) model based on n observations, the log-likelihood function is

$$\ln L=-\frac{n}{2}\ln 2\pi\sigma_\varepsilon^2-\frac{1}{2\sigma_\varepsilon^2}S(\phi,\mu,\theta). \tag{12}$$

We proceed to maximize (12) with respect to the parameters $\phi,\mu,\theta,$ and σ_ε^2, from (11), we have

$$\ln\hat{L}=-\frac{n}{2}\ln\hat{\sigma}_\varepsilon^2-\frac{n}{2}(1+\ln 2\pi). \tag{13}$$

Since the second term in expression (13) is a constant, we can reduce the AIC to the following expression

$$\text{AIC}(M) = n \ln \hat{\sigma}_{\varepsilon}^{2} + 2M \ . \tag{14}$$

Thus, we generate an appropriate time series model and select the statistical process with the smallest AIC. The model that we have identified will possess the smallest average mean square error. In addition, we summarize the development of the model as follows:

- Transform the original time series $\{x_t\}$ into a new series $\{y_t\}$
- Check for stationarity of the new time series $\{y_t\}$ by determining the order of differencing *d*, where $d = 0,1,2,...$ according to KPSS test, until we achieve stationarity
- Deciding the order m of the process, for our case, we let $m = 5$ where $p + q = m$
- After (*d*, m) being selected, listing all possible set of (*p*, *q*) for $p + q \leq m$
- For each set of (p, q), estimating the parameters of each model, that is, $\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q$
- Compute the AIC for each model, and choose the one with smallest AIC

According to the criterion that we mentioned above, we can obtain the ARIMA(*p*,*d*,*q*) model that best fit a given time series, where the coefficients are $\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q$.

Using the model that we developed for $\{y_t\}$ and subject to the AIC criteria, we forecast values of $\{y_t\}$ and proceed to apply the back-shift operator to obtain estimates of the original phenomenon $\{x_t\}$, that is,

$$\hat{x}_t = k\,\hat{y}_t - x_{t-1} - x_{t-2} - ... - x_{t-k+1} \,. \tag{15}$$

The proposed model and the corresponding procedure discussed in this section shall be illustrated with real economic application and the results will be compared with the classical time series model.

Application: Forecasting Stock XYZ

We selected a stock from fortune 500 companies that we identify as (XYZ). We shall use the daily closing price for 500 days that will constitute the time series $\{x_t\}$. A plot of the actual information is given by Figure 1.

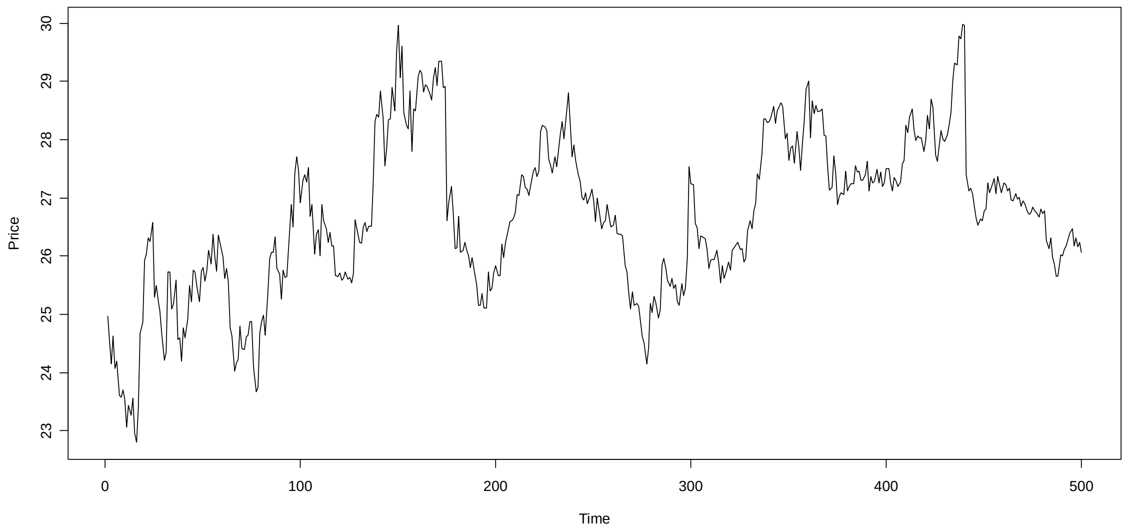

Figure 1. Daily Closing Price for Stock XYZ

First, we shall develop a time series forecasting model of the given nonstationary data using the ordinary Box and Jenkins methodology. Secondly, we shall modify the given data, Figure 1, to develop our proposed time series forecasting model. A comparison of the two models will be given.

The general theoretical form of the ARIMA(*p*,*d*,*q*) is given by

$$\phi_p(B)(1-B)^d x_t = \theta_q(B)\varepsilon_t . \qquad (16)$$

Following the Box and Jenkins' methodology (1994), the classical forecasting model with the best AIC score is the ARIMA(1,1,2). That is, a combination of first order autoregressive (AR) and a second order moving average (MA) with a first difference filter. Thus, we can write it as

$$(1-.9631B)(1-B)x_t = (1-1.0531B+.0581B^2)\varepsilon_t . \qquad (17)$$

After expanding the autoregressive operator and the difference filter, we have

$$(1-1.9631B+.9631B^2)x_t = (1-1.0531B+.0581B^2)\varepsilon_t \quad (18)$$

and we may rewrite the model as

$$x_t = 1.9631x_{t-1} - .9631x_{t-2} + \varepsilon_t - 1.0531\varepsilon_{t-1} + .0581\varepsilon_{t-2} \quad (19)$$

by letting $\varepsilon_t = 0$, we have the one day ahead forecasting time series of the closing price of stock XYZ as

$$\hat{x}_t = 1.9631x_{t-1} - .9631x_{t-2} - 1.0531\varepsilon_{t-1} + .0581\varepsilon_{t-2}. \quad (20)$$

Using the above equation, we graph the forecasting values obtained by using the classical approach on top of the original time series, as shown by Figure 2.

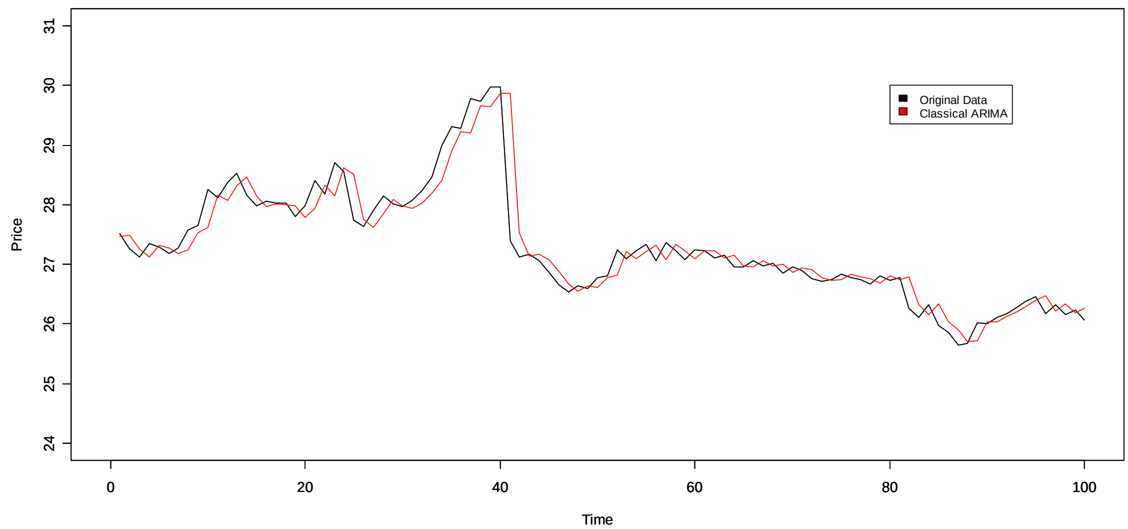

Figure 2. Comparisons on Classical ARIMA Model VS. Original Time Series

The basic statistics that reflect the accuracy of model (20) are the mean $\bar{r}$, variance S_r^2, standard deviation S_r and standard error $S_r/\sqrt{n}$ of the residuals. Figure 3 gives a plot of the residual and Table 1 gives the basic statistics.

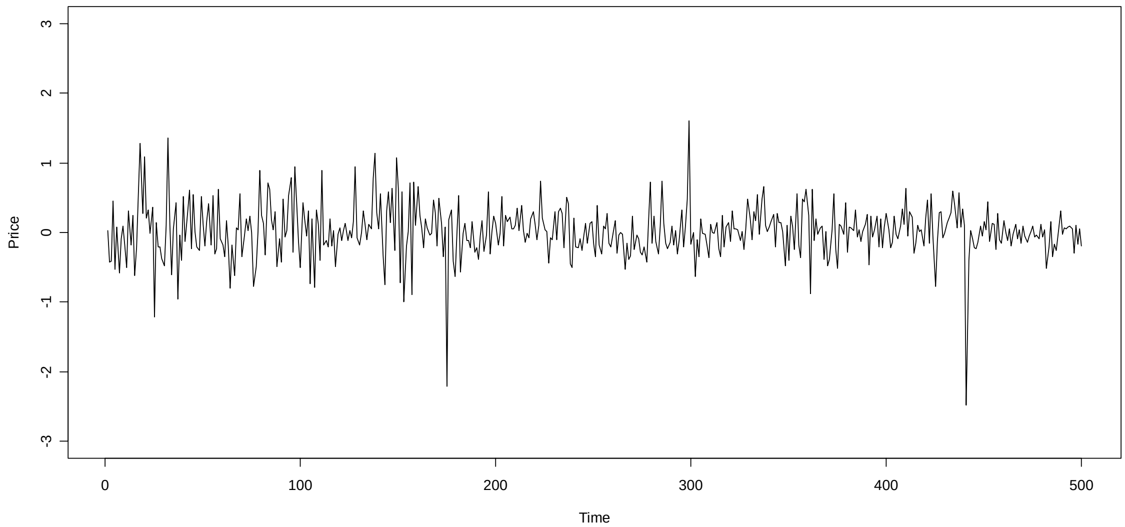

Figure 3. Time Series Plot of the Residuals for Classical Model

Table 1. Basic Evaluation Statistics

$\bar{r}$	S_r^2	S_r	$S_r/\sqrt{n}$
0.02209169	0.1445187	0.3801562	0.0170011

Furthermore, we restructure the model (20) with $n = 475$ data points to forecast the last 25 observations only using the previous information. The purpose is to see how accurate our forecast prices are with respect to the actual 25 values that have not been used. Table 2 gives the actual price, predicted price, and residuals between the forecasts and the 25 hidden values.

Table 2. Actual and Predicted Price

N	Actual Price	Predicted Price	Residuals
476	26.78	26.8473	-0.0673
477	26.75	26.7976	-0.0476
478	26.67	26.7673	-0.0972
479	26.8	26.6922	0.1078
480	26.73	26.8064	-0.0764
481	26.78	26.7490	0.0310
482	26.27	26.7911	-0.5211
483	26.12	26.3277	-0.2077
484	26.32	26.1631	0.1569
485	25.98	26.3364	-0.3564
486	25.86	26.0349	-0.1749
487	25.65	25.9068	-0.2568
488	25.67	25.6670	0.0031
489	26.02	25.7119	0.3081
490	26.01	26.0335	-0.0235
491	26.11	26.0427	0.0674
492	26.18	26.1343	0.0457
493	26.28	26.2032	0.0768
494	26.39	26.2986	0.0914
495	26.46	26.4043	0.0557
496	26.18	26.4743	-0.2943
497	26.32	26.2219	0.0981
498	26.16	26.3354	-0.1754
499	26.24	26.1953	0.0447
500	26.07	26.2602	-0.1902

The average of these residuals is $\bar{r} = -0.05608$.

We proceed to develop the proposed forecasting model. The original time series of stock XYZ daily closing prices is given by Figure 1. The new time series is being created by $k = 3$ days moving average and the analytical form of $\{y_t\}$ is given by

$$y_t = \frac{x_{t-2} + x_{t-1} + x_t}{3}. \tag{21}$$

Figure 4 shows the new time series $\{y_t\}$ along with the original time series $\{x_t\}$, that we shall use to develop the proposed forecasting model.

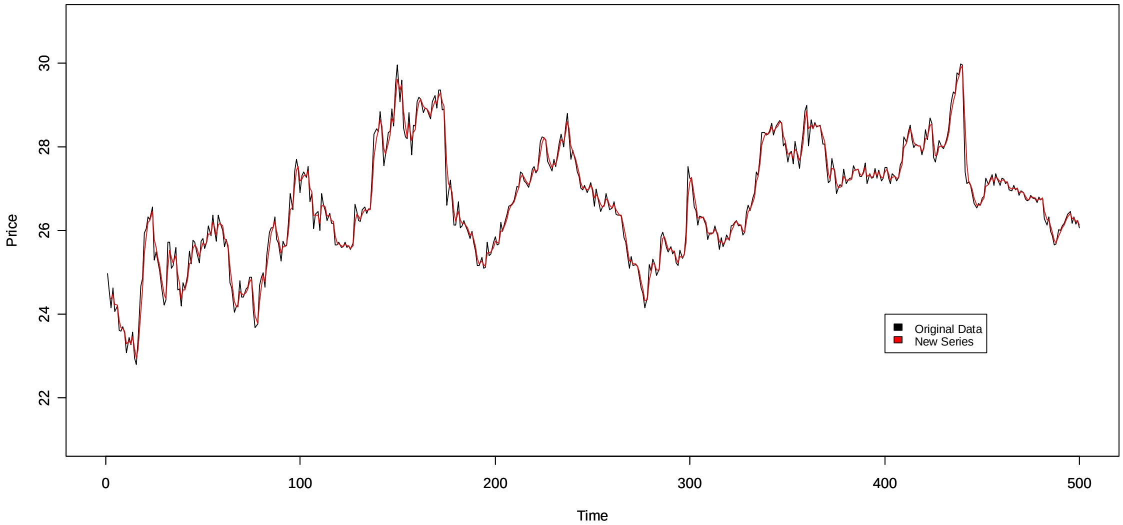

Figure 4. Three Days Moving Average on Daily Closing Price of Stock XYZ Vs. the original time series

Following the procedure we have stated above, the best model that characterizes the behavior of $\{y_t\}$ to be ARIMA(2,1,3). That is,

$$(1-.8961B-.0605B^2)(1-B)y_t = (1+.0056B-.0056B^2-B^3)\varepsilon_t. \quad (22)$$

Expanding the autoregressive operator and the first difference filter, we have

$$(1-1.8961B+.8356B^2+.0605B^3)y_t = (1+.0056B-.0056B^2-B^3)\varepsilon_t. \quad (23)$$

Thus, we can write (23) as

$$y_t = 1.8961y_{t-1} - .8356y_{t-2} - .0605y_{t-3} + \varepsilon_t + .0056\varepsilon_{t-1} - .0056\varepsilon_{t-2} - \varepsilon_{t-3}. \quad (24)$$

The final analytical form of the proposed forecasting model can be written as

$$\hat{y}_t = 1.8961y_{t-1} - .8356y_{t-2} - .0605y_{t-3} + .0056\varepsilon_{t-1} - .0056\varepsilon_{t-2} - \varepsilon_{t-3}. \quad (25)$$

Using the above equation, a plot of the developed model (25), showing a one day ahead forecasting along with the new time series, $\{y_t\}$, is displayed by Figure 5.

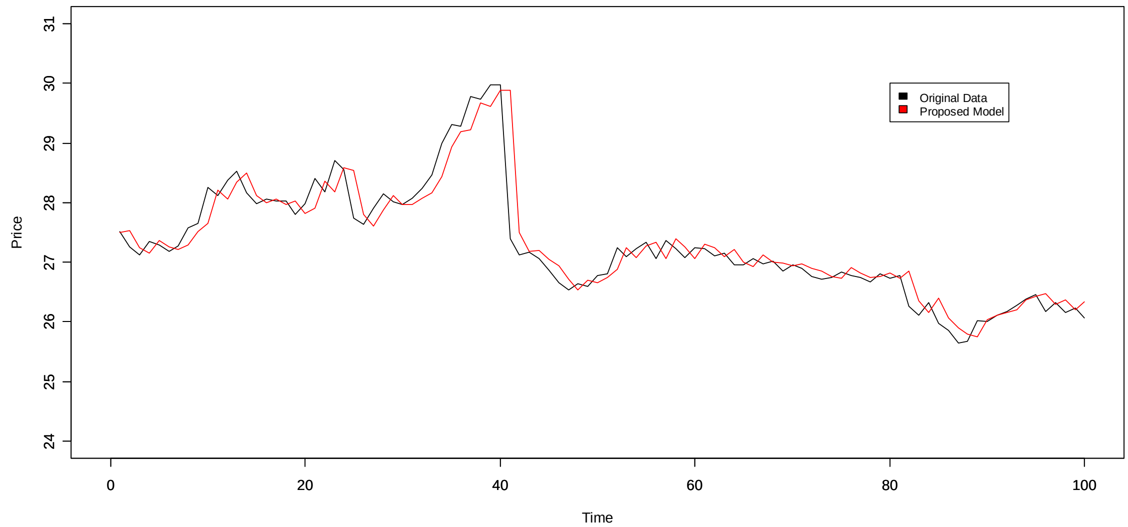

Figure 5. Comparisons on Our Proposed Model VS. Original Time Series

Note the very closeness of the two plots that reflect the quality of the proposed model.

Similar to the classical model approach that we discussed earlier, we shall use the first 475 observations $\{y_1, y_2, ..., y_{475}\}$ to forecast $\hat{y}_{476}$. Then we use the observations $\{y_1, y_2, ..., y_{476}\}$ to forecast $\hat{y}_{477}$, and continue this process until we obtain forecasts all the observations, that is, $\{\hat{y}_{476}, \hat{y}_{477}, ..., \hat{y}_{500}\}$. From equation (21), we can see the relationship between the forecasting values of the original series $\{x_t\}$ and the forecasting values of 3 days moving average series $\{y_t\}$,that is,

$$\hat{x}_t = 3\hat{y}_t - x_{t-1} - x_{t-2}. \tag{26}$$

Hence, after we estimated $\{\hat{y}_{476}, \hat{y}_{477}, ..., \hat{y}_{500}\}$, we can use the above equation, (26), to solve the forecasting values for $\{x_t\}$. Figure 6 is the residual plot generated by our proposed model, and followed by Table 3, that includes the basic evaluation statistics.

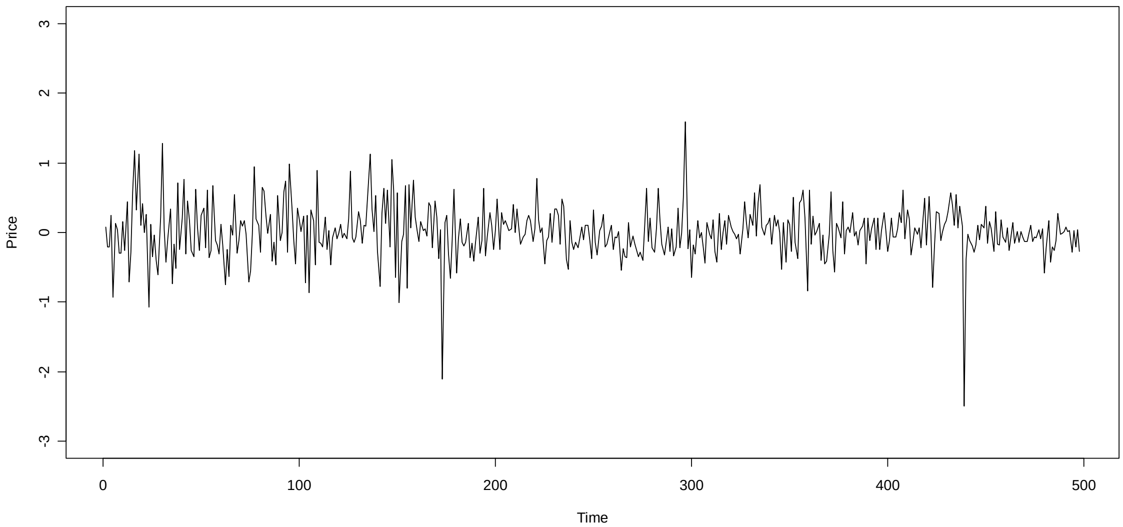

Figure 6. Time Series Plot for Residuals for Our Proposed Model

Table 3. Basic Evaluation Statistics

$\bar{r}$	S_r^2	S_r	$S_r / \sqrt{n}$
0.01016814	0.1437259	0.3791119	0.01698841

Both of the above displayed evaluations reflect on accuracy of the proposed model. The actual daily closing prices of stock XYZ from the 476th day along with the forecasted prices and residuals are given in Table 4.

Table 4. Actual and Predicted Price

N	Actual Price	Predicted Price	Residuals
476	26.78	26.8931	-0.1131
477	26.75	26.7715	-0.0215
478	26.67	26.7121	-0.0421
479	26.8	26.7239	0.0761
480	26.73	26.7854	-0.0554
481	26.78	26.6892	0.0908
482	26.27	26.8292	-0.5592
483	26.12	26.3027	-0.1827
484	26.32	26.0808	0.2392
485	25.98	26.3603	-0.3803
486	25.86	25.9868	-0.1268
487	25.65	25.8443	-0.1943
488	25.67	25.7115	-0.0414
489	26.02	25.6499	0.3701
490	26.01	25.9650	0.0450
491	26.11	26.0526	0.0574
492	26.18	26.0912	0.0888
493	26.28	26.1449	0.1351
494	26.39	26.3090	0.0810
495	26.46	26.3752	0.0848
496	26.18	26.4223	-0.2423
497	26.32	26.2461	0.0739
498	26.16	26.2964	-0.1364
499	26.24	26.1437	0.0963
500	26.07	26.2678	-0.1978

The Results given above attest to the good forecasting estimates for the hidden data.

Comparison of the Forecasting Models

In this section, we shall compare the two developed models. The classical process is given by

$$\hat{x}_t = 1.9631 x_{t-1} - .9631 x_{t-2} - 1.0531 \varepsilon_{t-1} + .0581 \varepsilon_{t-2} . \qquad (27)$$

In our proposed model, we shall use the following inversion to obtain the estimated daily closing prices of stock XYZ, that is,

$$\hat{y}_t = 1.8961 y_{t-1} - .8356 y_{t-2} - .0605 y_{t-3} + .0056 \varepsilon_{t-1} - .0056 \varepsilon_{t-2} - \varepsilon_{t-3} \tag{28}$$

in conjunction with

$$\hat{x}_t = 3\hat{y}_t - x_{t-1} - x_{t-2}\,. \tag{29}$$

The table given below is a comparison of the basic statistics used to evaluate the two models under investigation.

Table 5. Basic Comparison on Classical Approach Vs. Our Proposed model

	$\bar{r}$	S_r^2	S_r	$S_r / \sqrt{n}$
Classical	0.02209169	0.1445187	0.3801562	0.0170011
Proposed	0.01016814	0.1437259	0.3791119	0.01698841

The average mean residuals between the two models shown that the proposed model is overall approximately 54% more effective in estimating one day ahead the closing price of Fortune 500 stock XYZ.

Conclusion

In the present study we introduced a new time series model that is based on the actual stochastic realization of a given phenomenon. The proposed model is based on modifying the given economic time series, $\{x_t\}$, and smoothing it with *k*-time moving average to create a new time series, $\{y_t\}$. We developed the basic analytical procedures through the developing process of a forecasting model. A step-by-step procedure is memorized for the final computational procedure for a nonstationary time series. To evaluate the effectiveness of our proposed model we selected a company from the Fortune 500 list, company XYZ, the daily closing prices of the stock for 500 days was used as our time series data, $\{x_t\}$, which was as usually nonstationary. We develop the classical time series forecasting model using the Box and Jenkins, methodology and also our proposed model, $\{y_t\}$, based on a 3-day moving average smoothing procedure. The analytical form of the two forecasting models is presented and a comparison of them also given. Based on the average mean residuals the proposed model was significantly more effective in such term of predicting of the closing daily prices of stock XYZ.

References

Akaike, H. (1974). A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, *AC-19*, 716-723.

Banerjee, A., Dolado, J. J., Galbraith, J. W., & Hendry, D. F. (1993). *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, Oxford.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd ed., Prentice Hall, Englewood Cliffs, NJ., 89-99.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994) *Time Series Analysis: Forecasting and Control*, 3rd ed., Prentice Hall, Englewood Cliffs, NJ., 224-247.

Brockwell, P. J., & Davis, R. A. (1996). *Introduction to Time Series and Forecasting*., Springer, New York., Sections 3.3 and 8.3.

Dickey, D. A., & Fuller, W. A. (1979) Distribution and the Estimators for Autoregressive Time Series With a Unit Root., *Journal of the American Statistical Association*, Vol. 74, No. 366, 427-431.

Dickey, D. A., Hasza, D. P., & Fuller, W. A. (1984). Testing for Unit Roots in Seasonal Time Series., *Journal of the American Statistical Association*, Vol. 79, No. 386, 355-367.

Durbin, J., & Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*., Oxford University Press.

Gardner, G., Harvey, A. C., & Phillips, G. D. A. (1980). Algorithm AS154. An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering., *Applied Statistics*, *29*, 311-322.

Harvey, A. C. (1993). *Time Series Models*, 2nd Edition, Harvester Wheatsheaf., sections 3.3 and 4.4.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations., *Technometrics*, *20*, 389-395.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root., *Journal of Econometrics*, *54*, 159-178.

Rogers, A. J. (1986). Modified Lagrange Multiplier Tests for Problems with One-Sided Alternatives, *Journal of Econometrics*, North-Holland., *31*, 341-361.

Said, S. E., & Dickey, D. A. (1984) Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order., *Biometrika*, *71*, 599-607.

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike Information Criterion Statistics*., D. Reidel Publishing Company.

Shumway, R. H., & Stoffer, D. S. (2006). *Time Series Analysis and Its Applications: with R Examples*, 2nd ed., Springer, New York.

Tsokos, C. P. (1973). Forecasting Models from Non-Stationary Time Series-Short Term Predictability of Stocks., *Mathematical Methods in Investment and Finance*., North Holland Publishing Co., 520-63.

Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed., Pearson Education, Inc.