STATISTICS IN TRANSITION-new series, August 2011
Vol. 12, No. 1, pp. 179—192

IDENTIFYING AN APPROPRIATE FORECASTING MODEL FOR FORECASTING TOTAL IMPORT OF BANGLADESH

TANVIR KHAN[1]

ABSTRACT

Forecasting future values of economic variables are some of the most critical tasks of a country. Especially the values related to foreign trade are to be forecasted efficiently as the need for planning is great in this sector. The main objective of this research paper is to select an appropriate model for time series forecasting of total import (in taka crore) of Bangladesh. The decision throughout this study is mainly concerned with seasonal autoregressive integrated moving average (SARIMA) model, Holt-Winters' trend and seasonal model with seasonality modeled additively and vector autoregressive model with some other relevant variables. An attempt was made to derive a unique and suitable forecasting model of total import of Bangladesh that will help us to find forecasts with minimum forecasting error.

Key words: ARIMA model, Holt Winters' trend and seasonality method, VAR model, Forecasting accuracy, Out-of-sample accuracy measurement.

1. Introduction

An important economic concept involves international trade and finance. International trade in goods and services allows nations to raise their standards of living by exporting and importing goods and services. In a modern economy the economic condition is highly affected by the amount of its foreign trade and its balance of trade. Imports, along with exports, form the basis of trade. Bangladesh has had a negative trade balance since its independence, and the gap between

[1] MS Student. Institute of Statistical Research and Training, University of Dhaka, Dhaka-1000, Bangladesh. E-mail: tkhan1@isrt.ac.bd.

export and import are still widening. The country is importing a lot of goods from foreign countries and the aim of this paper is to find a forecasting model that will help us to get ideas about the future values of total import of Bangladesh.

1.1. Data and variables

This study is conducted on total import of Bangladesh. The data set has 133 observations, during the time period from July 1998 to July 2009, in the initialization set and 14 observations, during the time period from August 2009 to September 2010, in the test set. For vector autoregressive model two more variables - total export and net foreign assets, are also used. The data were obtained from the following sources for validation:

1. Statistical Department of Bangladesh Bank
2. 'Economic Trends', a monthly report published by Bangladesh Bank.

2. Methodology

In this study three methods are used, i.e. seasonal ARIMA model, Holt Winters' trend and seasonality method and vector autoregressive model. The goal is to find an appropriate model that has both in-sample and out-of-sample forecasting errors as small as possible.

2.1. ARIMA model

A model containing p autoregressive terms and q moving average terms is classified as ARMA(p,q) model. If the series is differenced d times to achieve stationary, the model is classified as ARIMA(p,d,q), where the symbol 'I' signifies 'integrated'. The equation for the ARIMA (p,d,q) model is as follows:

$$Y_t = c + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots . + a_p Y_{t-p} + e_t - b_1 e_{t-1} - b_2 e_{t-2} - \ldots . - b_q e_{t-q} \qquad (1)$$

Or, in backshift notation:

$$\left(1 - a_1 L - a_2 L^2 - \ldots . a_p L^p\right)(1 - L) Y_t = c + (1 - b_1 L - b_2 L^2 - \ldots . - b_q L^q) e_t \qquad (2)$$

The ARIMA notation can be extended readily to handle seasonal aspects, and a Seasonal ARIMA (p,d,q)(P,D,Q) or SARIMA model can be represented as

$$\left(1 - L - L^2 - L^3 - \ \dots\ L^d\right)\left(1 - L^S - L^{2(S)} - \ \dots\dots\ L^{D(S)}\right)\left(1 - a_1L - a_2L^2 - \dots\ a_pL^p\right)(1 - A_1L - A_2L^2 \dots A_PL^P)Y_t = \left(1 - b_1L - b_2L^2 - \ \dots. -b_qL^q\right)\left(1 - B_1L - B_2L^2 - \ \dots - B_QL^Q\right) \tag{3}$$

2. 2. Holt Winters trend and seasonality method

The Holt-Winters' method is based on three smoothing equations-one for the level, one for trend, and one for seasonality. In fact, there are two different Holt-Winters' methods, depending on whether seasonality is modeled in an additive or multiplicative way. The basic equations for Holt-Winters' multiplicative method is as follows:

Level: $L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1})$ (4)

Trend: $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$ (5)

Seasonal: $S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s}$ (6)

Forecast: $F_{t+m} = (L_t + b_tm)S_{t-s+m}$ (7)

Where s is the length of seasonality (e.g. number of month or quarter in a year), L_t represents the level of the series, b_t represents the trend, S_t is the seasonal component, and F_{t+m} is the forecast for m period ahead. The first two equations for additive model are identical to the first two equations of the multiplicative method. The only difference is in the third equation, that is the seasonal indices are now added and subtracted, i.e.:

Seasonal: $S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s}$ (8)

2.3. VAR model

In a VAR model the variables which have bilateral causality with each other have been included. A VAR model consists of a set of variables $Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{Kt})$ which can be represented as

$$Y_t = \alpha + A_1Y_{t-1} + A_2Y_{t-2} + \ \dots + A_pY_{t-p} + u_t \tag{9}$$

With A_i are (K×K) coefficient matrix for i=1,2,….p and u_t is a p dimensional process with $\mathrm{E}(u_t) = 0$ and covariance matrix $E(u_tu_t^T) = \sum_u$.

3. COMPARISON AMONG THE FORECASTING METHODS

To make comparison among the methods some well known measures of forecast error are used. The model that gives the minimum measures of these errors will be the expected model for further forecasting. The measures used are cited below.

Mean Error (ME): The mean error gives the average forecast error, i.e.:

$$ME = \frac{1}{n}\sum_{i=1}^{n} e_t$$

Where

$$e_t = Y_t - F_t$$

Y_t= The observation at time t, F_t=Forecasted value at time t, n= the number of observation.

Mean Absolute Error (MAE): The MAE is first defined by making each error positive by taking its absolute value, and then averaging the result, i.e.:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_t|$$

Mean Squared Error (MSE): The MSE is defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n} {e_t}^2$$

Mean Percentage Error (MPE): The MPE is the mean of the relative or percentage error and is given by

$MPE = \frac{1}{n}\sum_{i=1}^{n} PE_t$; Where PE_t=$\frac{Y_t - F_t}{Y_t}$×100 is the relative or percentage error at time t.

Mean Absolute Percentage Error (MAPE): The MAPE is defined as

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} |PE_t|$$

3.1. Out-of-sample accuracy measurement

The summary statistics described so far measures the goodness of fit of the model to historical data. Such fitting does not necessarily imply good forecasting. An MSE or MAPE of zero can always be obtained in the fitting phase by using a polynomial of sufficiently high order. These problems can be overcome by measuring true out-of-sample forecast accuracy. That is, the total data are divided into an 'initialization' set and a 'test' set or 'holdout' set. Then the initialization set is used to estimate any parameters and to initialize the method. Forecasts are made for the test set. The accuracy measures are computed for the errors in the test set only.

4. Analysis of data

The initialization set of the data has 133 data points, starting from July 1998 to July 2009. The first step of the analysis is to plot the whole dataset to visualize the nature of it. The time plot for total import (in taka crore) of Bangladesh is shown in the figure given below.

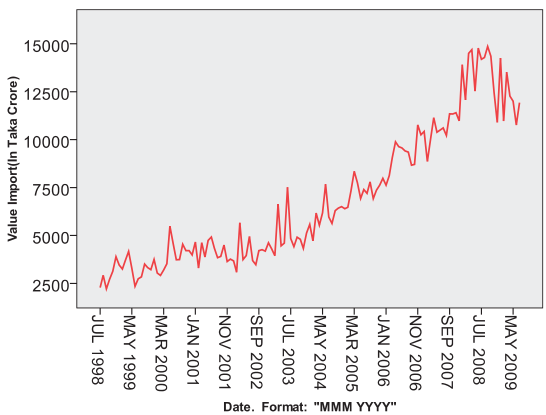

Figure 1. Time series plot for actual data

4.1. ARIMA model

It can be seen from Figure 1 that the data had an upward trend but from October 2008 it started to show a downward trend. From simple view of the plot we can have an idea about non-stationarity in mean but stationarity in variance. To become more certain about stationarity of the data, Augmented Dickey Fuller test has been conducted. The calculated value of Dickey Fuller is -2.0622 with P value .5506 at suggested lag 5. So, we fail to reject the null hypothesis of non-stationarity.

The PACF of the data and ACF of the first differenced data are:

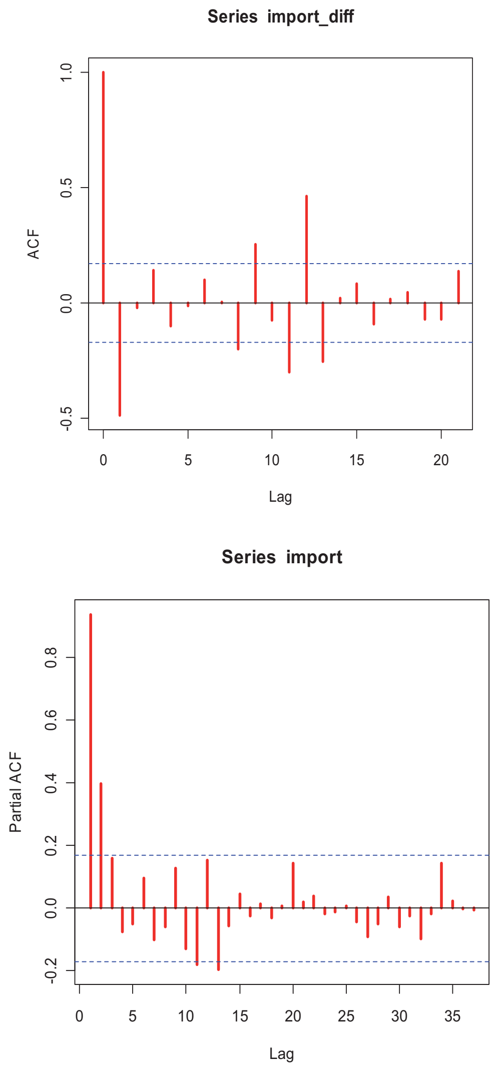

Figure 2. PACF of original data and ACF of first differenced data

From the PACF plot it can be seen that the partial autocorrelation at lag 12 is insignificant and so are at further seasonal lags (i.e. 24, 36 etc.). Thus, it can be assumed that the data do not have strong seasonality.

To obtain stationarity the original data are differenced at order one. As the data set do not have strong seasonality, a first order non-seasonal difference is taken rather than a seasonal difference and ADF test is conducted again. The value of Dickey Fuller for the first differenced data is -5.3819 with P value less than .01 at suggested lag 5. This clearly indicates that our data are now stationary. The ACF plot of the first differenced data shows significant spikes at lag 12, which gives a clear indication that a seasonal term must be included in the model. To be certain about the form of the appropriate model, the AIC and BIC values are checked for all the probable models and it is found that ARIMA (0,1,1,)(1,0,0)[12] model has the smallest AIC and BIC values, so, this model should be the desired ARIMA model.

The estimated ARIMA (0,1,1) (1,0,0)12 model is given below:

$$\widehat{M}_t = 0.4733M_{t-12} - M_{t-1} + 0.4733M_{t-13} + .5088e_{t-1} \qquad (10)$$

The value of Box-Pierece Q statistic 30.6526 is with degrees of freedom 42 with P value .9026 at lag 44. So, the null hypothesis that residuals of ARIMA (0,1,1)(1,0,0)12 model are white noise fails to be rejected. Again, the value of Ljung-Box Q statistic 36.4945 is with degrees of freedom 42 with P value 0.7107 at lag 44. So, it can be said that the residuals are white noise for ARIMA(0,1,1)(1,0,0)12 model.

The plot of forecasted value along with the original time series is

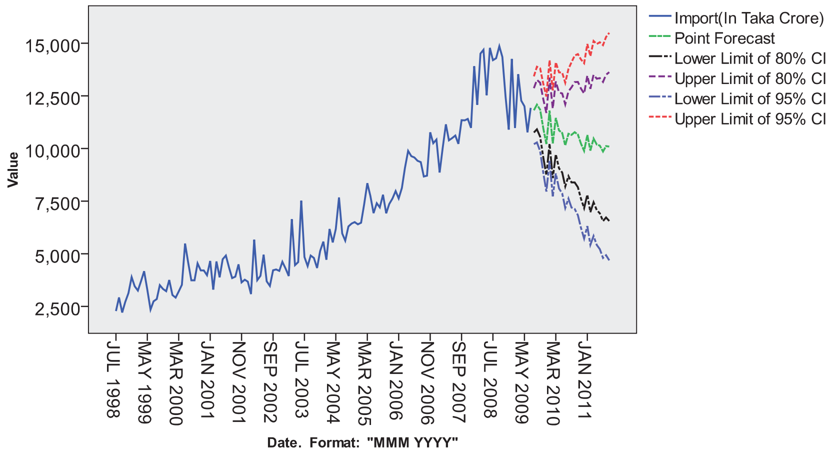

Figure 3. Point and interval forecasting with ARIMA (0,1,1)(1,0,0)12 model

4.2. Holt Winters trend and seasonality method

To initialize the Holt-Winters' forecasting method, initial values of the level L_t, the trend b_t, and the seasonal indices S_t are needed. To determine initial estimates of the seasonal indices 12 data points were used as the data set is monthly. The estimated model is

$$L_t = .3925046(M_t - S_{t-12}) + (1 - .3925046)(L_{t-1} + b_{t-1}) \quad (11)$$

$$b_t = b_{t-1} \quad (12)$$

$$S_t = 0.8301905(M_t - L_t) + (1 - 0.8301905)S_{t-12} \quad (13)$$

The plot of forecasted value obtained from Holt Winters model is given below.

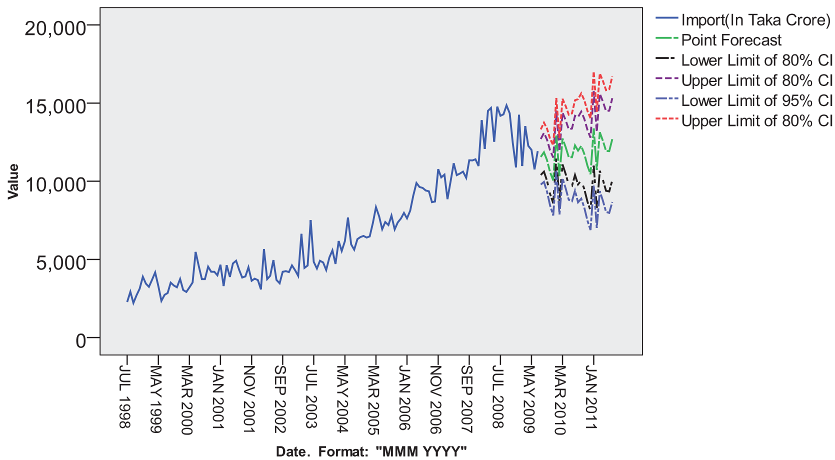

Figure 4. Point and Interval Forecasts obtained from Holt-Winters Model

4.3. VAR model

To fit a VAR model, at first an attempt was made to find relevant variables that have bilateral causality with total import of Bangladesh as well as with each other. Primarily 5 variables were chosen. They are: total export, net foreign assets, domestic credit, exchange rate and inflation rate. Among these variables, it was found that only total export and net foreign assets have bilateral relationship with each other as well as with total import. The results or Granger causality tests are given below.

Table 1. Result of Granger causality test

relationship	Lag	F-statistic	P-value
Export→Import	3	3.1070343	2.901318e-02
Import→Export	3	7.5056727	1.180776e-04
Net Foreign Assets→Import	5	2.582494	2.966103e-02
Import→Net Foreign Assets	5	2.460508	3.696343e-02
Export→Import	9	2.758777	0.0061726486
Import→Export	9	2.954010	0.0036209259
Net Foreign Assets→Export	9	3.881317	0.0002814135
Export→Net Foreign Assets	9	2.346345	0.0187481360

Thus, it can be assumed that total export and net foreign assets can be used as endogenous variables with total import in the desired VAR model. The natural log forms of all the three variables are taken and for dealing with seasonality all the three variables are seasonally adjusted by the classical multiplicative decomposing method. The modified data set are then analyzed. As at lag 2 the model gives the smallest AIC, FP and HQ values, the model is fitted with lag 2. The estimated model has the form

$$\widehat{M}_t = 0.52533 + 0.37965M_{\mathrm{t}-1} - 0.02456M_{\mathrm{t}-2} - 0.15031X_{\mathrm{t}-1} + 0.5257X_{\mathrm{t}-2} + 0.02215F_{t-1} + 0.18856F_{\mathrm{t}-2} \tag{14}$$

$$\hat{X}_t = 0.0342 - 0.1211M_{\mathrm{t}-1} + 0.0234M_{\mathrm{t}-2} + 0.14806X_{\mathrm{t}-1} - 0.0225X_{t-2} + 1.07017F_{\mathrm{t}-1} - 0.09139F_{\mathrm{t}-2} \tag{15}$$

$$\hat{F}_t = 0.0342 - 0.12107M_{t-1} + 0.02343M_{t-2} + 0.14806X_{t-1} - 0.0225X_{t-2} + 1.07017F_{t-1} - 0.09139F_{t-2} \tag{16}$$

For checking whether the required type of causality exists in our model or not, Granger causality test was done again.

Table 2. Results of Multivariate Granger Causality test

Null Hypothesis	F value	P value	Decision
Total export does not Granger cause total import and net foreign assets	6.6997	3.244e-05	Null hypothesis is rejected with more than 99% confidence
total import does not Granger cause total export and net foreign assets	3.7346	0.0054	Null hypothesis is rejected with more than 99% confidence
net foreign assets does not Granger cause total import and total export	2.9141	0.02141	Null hypothesis is rejected with more than 95% confidence

At lag 31, asymptotic Portmanteau test has the Chi-Squared value 250.7905 with P value 0.664 at lag 31 and the adjusted Portmanteau test has Chi-Squared value 286.9564 with P value 0.1294. Thus, the model has white noise residuals. The forecasted values of the model are then back-transformed. The plot of original data and forecasted values is

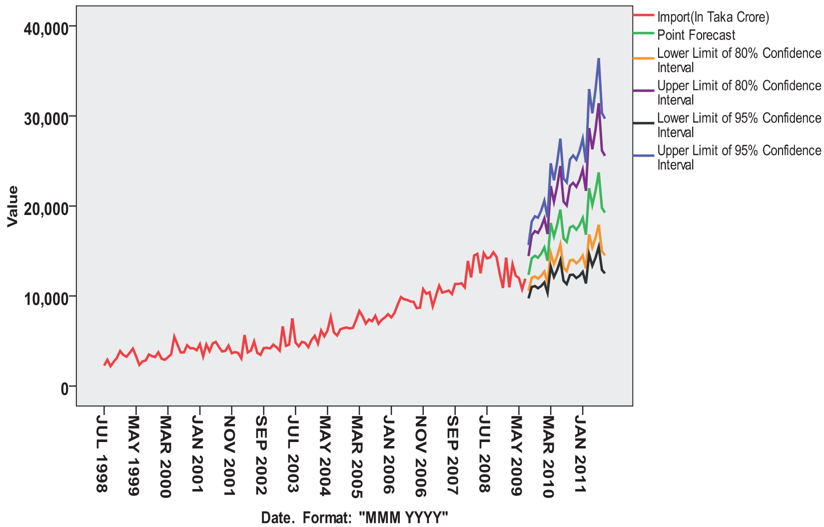

Figure 5. Point and Interval Forecasts obtained from VAR model

5. Comparison among arima, holt winters trend and seasonality method and VAR model

The forecasting performance of these three models have been compared with each other with respect to different measures of error and the summary measures are listed in the table below.

Table 3. In-sample error measures of the methods

Measures of Error	ARIMA (0,1,1)(1,0,0)12	Holt-Winters' Method	VAR Model
Mean Error	58.61606930	80.57279	32.63461
Mean Absolute Error	588.55960438	676.039	616.8842
Mean Squared Error	671123.9	808836.5	780904.8
Mean Percentage Error	0.01089789	0.08561744	-0.7211356
Mean Absolute Percentage Error	9.48140336	10.56153	9.086229

In-sample error measures do not necessarily imply good forecasting model. To find which forecasting method is better, true out-of-sample forecast accuracy was measured.

The plot of the test set observations with the forecasted values using all three methods is given below.

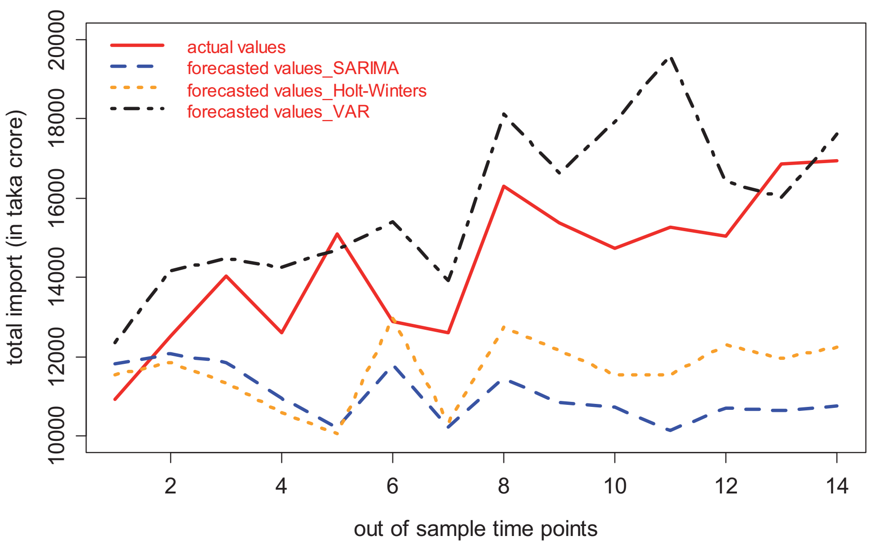

Figure 6. Plot of test set of the data and forecasted values obtained from the three methods

The out-of-sample accuracy measures are given below.

Table 4. Out-of-sample accuracy measurements

Measures of Error	ARIMA(0,1,1)(1,0,0)12	Holt-Winters' Method	VAR Model
Mean Error	3350.570	2712.007	-1460.056
Mean Absolute Error	3477.229	2815.95	1636.428
Mean Squared Error	15747374	10227508	3775589
Mean Percentage Error	21.81878	17.79637	-10.5399
Mean Absolute Percentage Error	22.97802	18.72796	11.62293

It is clear from the above table that VAR model is giving minimum values for all the measures of forecast error. Therefore, one may take VAR model of total import, total export and net foreign assets as an appropriate model for forecasting total import of Bangladesh.

6. Conclusion

The basic aim of this paper is to select an appropriate model that will be helpful for understanding future behavior of total import of Bangladesh. Three methods are used - seasonal ARIMA, Holt-Winters' trend and seasonality method as well as VAR model. Various measures of forecasting accuracy were also measured for all the three models. The comparison shows that VAR model of total import with total export and net foreign assets as other endogenous variables is better than the other two methods as it is producing both in-sample and out-of-sample forecasting errors. But this is not the end. More researches have to be done to handle seasonality in a better way and to find more relevant variables that are useful for forecasting total import of Bangladesh.

REFFERENCES:

AMISANO, G. AND GIANNINI, C. (1997). Topics in Structural VAR Econometrics, Springer-Verlag, Berlin, 2nd edition.

GUJRATI, D. N. AND SANGEETHA (2007). Basic Econometrics, McGraw-Hill Book Co, New York.

HAMILTON,J.D. (1994). Time Series Analysis, Princeton University Press, Princeton.

MAKRIDAKIS, S., WHEELWRITGHT, S. C. AND HYNDMAN, R. J. (1998). Forecasting Methods and Applications, John Wiley and Sons, Ink., New York.

PFAFF, B. (2008). VAR, SVAR and SVEC Models: Implementation within R Package vars, New York. URL: http://CRAN.R-project.org/package=vars.