Fuzzy Time Series Method for Forecasting Taiwan Export Data

P. Arumugam [*], V. Anithakumari [#]

[*] Reader Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli-627012, India
[#] Research Scholar Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli-627012, India

Abstract: Forecasting accuracy is one of the most favorable critical issues in Autoregressive Integrated Moving Average (ARIMA) models. The study compares the application of two forecasting methods on the amount of Taiwan export, the Fuzzy time series method and ARIMA method. Model discussed for the ARIMA method and Fuzzy time series method include the Sturges rules. When the sample period is extend in our models, the ARIMA models shows smaller than predicted error and closer predicted path to the realistic trend than those of the Fuzzy models, resulted in more accurate forecast of the export amount the Autoregressive Integrated Moving Average models. In the economic viewpoints, the amount of Taiwan export is mainly attributable to external factors. However, this impact reduces with time and export amount in the time series analysis. The ARIMA models can be utilized to predicted export value accurately, when all of value or data is available.

Keywords: Time series; Stationary Stochastic Process; Fuzzy Time series model; ARIMA models; Sturges rule and Taiwan export.

I. INTRODUCTION

Owing to the relatively small domestic market and the insufficient nature resources in Taiwan, the economic development of Taiwan relies heavily on the international trades. In the past, Taiwan used its low labor cost advantages to manufacture labor-intensive products as a competitive edge to expand export worldwide and gained great economic growth. However, the low labor cost advantage has no longer existed in recent year that the standard wage is comparatively much higher in Taiwan than in the mainland China and Southeast Asia, Taiwan is facing with sever challenges from these low labor cost countries in the international markets and the amount of export has shrunken rapidly. Consequently, over the 1990-2002 Taiwan economic is stagnating and the unemployment rate researches its highest record.

In January 2002 Taiwan joined the World Trade Organization (WTO), becoming an official partner in the world trading system. Today the government is vigorously promoting knowledge based economy and industrial modernization to transform Taiwan into a "green silicon island" of high value added production. Taiwan has had one of the fastest growing economics for the past five decade and its development has been praised as an "economic miracle" Taiwan gradually high teached its industries over the past two decades and currently has the fourth largest information hardware and semiconductor industries in the world. Taiwan is the 17 th largest economy in the world, 14 th largest exporter and 16 th largest importer and third largest holder of foreign exchange reserves with over US$180 billion. Give the highly economic dependent nature on the export business of Taiwan; better understanding toward future trends in the international business to enhance market competition is an important task. Therefore, it is vital to have more accurate forecast and prediction models to assist and enact economic policies and business strategies particularly in the international export trade.

II. REVIEW OF LITERATURE

The problem of traditional forecasting methods is that they cannot contract with forecasting problems in which the historical data are represented by linguistic values. However, using fuzzy time series to deal with forecasting problem can overcome this problem. For instance, based on fuzzy theory, Song and Chissom (1993a, b, 1994), construct the one factor fuzzy time series. Sullivan and Woodall (1994), Chen (1996) and Huarng (2001a), the application of the fuzzy logic for dynamic analysis are increasing quickly in recent years. This research attempts to use information concerning Taiwan's export values as an example to test whether the fuzzy time series is indeed practical in its forecast of macro-economic variables. It also compares the Autoregressive Integrated Moving Average (ARIMA).

III. METHODOLOGY

A. Time Series Analysis

A time series is a sequence of observation, usually ordered in time although in some cases the ordering may be according to another dimension. That is in general, measurement on some particular characteristic over a period of time constitute a time series. It may be the hourly record of temperature at a metrological station, the daily record of Price of a commodity, the monthly sales value of a company, etc. The main feature of time series analysis, which distinguishes it from other statistical analyses, is the exploit recognition of the importance of the order in which the observation are made. While in many problem the observation are statistically independent, in time series successive observation are almost always dependent.

The main purpose of time series analysis is to study and understand how the characteristic under study evolves with respect to time so that the underlying probabilistic mechanism generating the time series is determined with reliable accuracy, once this is done, one can try to predict the future behavior of the characteristic. The main objective in analyzing time series is to understand, interpret and evaluate changes in the phenomena in the hope of more correctly anticipating the course of future events. A good understanding of the mechanism generating the series may also help us to control the phenomena involved in the generating mechanism and thereby control the future behavior of the process.

The various univariate time series models that are currently in use may be broadly classified under (A) linear time series models and (B) non linear time series models. The models under the format category can be again grouped under the heads (1) purely random (2) Autoregressive (3) Moving average (4) Autoregressive moving average and (5) Autoregressive integrated moving average models. Even though lot of work has been reported under the last category B, the models that are currently used in practice in time series analysis belong only to the category A. There does not seem to be any book discussing the various models in the second category B. A detailed study of most of the models belonging to the category A can be found in Brock well, The main objective of analyzing time series is to Understand, interpret and evaluate changes in economic phenomena in the hope of more. Correctly anticipating the course of future events.

B. Stationary Stochastic Process:

A type of stochastic process that has received a great deal of attention and scrutiny by time series analysts is so called stationary stochastic process. Broadly speaking, a stationary process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the distance or gap or lag between the two prime periods and not the actual time at which the covariance is completed. In the time series literature, such a stochastic process is known as a weakly stationary or covariance stationary or second order stationary or wide sense stochastic process.

Weak stationary, let Y_t be a stochastic time series with these properties:

$$\text{Mean } E(Y_t) = \mu$$

$$\text{Variance } \operatorname{var}(Y_t) = E(Y_t - \mu)^2 = \sigma^2$$

$$\text{cov}ariance\gamma_k = E[(Y_t - \mu)(Y_{t-1} - \mu)]$$

Where γ_k the covariance at lag k, is the covariance between the values of Y_t and Y_{t-1} that is, between two Y values k periods apart. If k = 0, we obtain γ_0, which is simply the variance of $Y(=\sigma^2)$; if k = 1, γ_1 is the variance between two adjacent values of Y. Stationary Time series is a mean and variances are constant term. If a time series is not stationary we say it as non stationary time series. In other words, a non stationary time series will have a time varying mean or a time-varying variance or both.

C. Time Series: Forecasting

Forecasting is an important part of econometric analysis for some probably the most important. In this chapter we discuss two methods of forecasting that have become quite popular.

- Auto Regressive Integrated Moving Average (ARIMA) popularly known as the Box-JENKINS methodology.
- Vector Auto regression (VAR).

The topic of forecasting is infinite and specialized books have been written on this subject. Fortunately most modern econometric and statistical packages have user friendly introductions to several techniques.

D. ARIMA Model

Time Series ARIMA model was proposed by Box-Jenkins in 1970, the model examines each variable by using auto regression, AR(p) and Moving Average (q) to investigate the historical data and economic fluctuations. The algorithm is to be had as follows.

1) Data Interpretation: The first step in developing a Box-Jenkins model is to decide if the series is stationary and if there is any significant seasonality that needs to be modeled. The autocorrelation functions (ACF) are used to define the distribution of sample data.

2) Model Identification: Identifying the phase of the series by using autocorrelation function (ACF and partial autocorrelation function (PACF).

3) Inference: The conditional likelihood and exact likelihood are used to estimate the parameters.

4) Diagnostic Checking: The process of diagnostic check involves testing the assumptions of the model to identify any areas where the model is inadequate. The statistical identification process includes whether the parameter achieves statistical significance or multicollinearity and whether the residual term is white noise or not. If the model is found to be insufficient, it is necessary to remedy and repeat step (4) until a better model is identified.

E. Fuzzy Set and Forecasting

The fuzzy sets theory can be defined as a mathematical formalization that enables us to eliminate indefiniteness and deal with incomplete, inaccurate information of both qualitative and quantitative by nature. The fuzzy sets theory, advanced one of the well-known representatives of modern applied mathematics, by excluding any definite description of the task offers such a solution scheme of the problem that a subjective reasoning and evaluation plays a principal role in evaluating indefinite, unclear fact. Thus anyone, encountering indefinite, incomplete information data, can form some conclusion, if even in a rough way, by passing through his/her reasoning all these realities. The use of fuzzy verbal notions in every-day speech (much, more, little, small, many, a Number of etc.) enables us to give a qualitative description of the problem which must be tackled and take account of its indefinite nature as well as attain the description of the factors that can't be described qualitatively.

The advent of fuzzy logic made it possible to tackle a great many problems with fuzzy input data. One of them was a forecasting problem. Many of the structural elements of the latter (input data and interdependence between its components, interval evaluation of indicators and their interdependence, expert evaluations and judgments etc.) are either of a fuzzy nature or, by being in fuzzy relationships, condition the fuzzy description of the problem.

The application of fuzzy logic to the handling of forecasting problems was undertaken by the researches in which the mathematical models of fuzzy time series were described in a fuzzy form for handling the problem with fuzzy input data This approach was developed later by other scientists dealing with the solution of analogous problems To tackle the task, the authors proposed a model of fuzzy time series and tried to reduce the average forecasting error by making adequate alterations in the model.

Thus, the major purpose of the proposed approach is methodological: 1) putting forth an evaluation method based on fuzzy time series for estimating model parameters; 2) testing the extent to which the model is adequate to reflect the real process, that is to say, computing the method error; 3) conducting the comparative analysis of computation results; 4) revealing the practical and theoretical importance of the model.

F. Fuzzy Time Series

Fuzzy Time series is assumed to be a fuzzy variable along with associated member ship function Song and Chissom [1993] have proposed a procedure for solving fuzzy time series model described in the following steps.Time series represents a consecutive series of observation that is conducted by equal time intervals and lies at the root of exploring real processes in economics, meteorology and natural sciences etc.

The analysis of time series of observation consists of the followings: 1) constructing the mathematical model of time series of observation of real processes; 2) model identification or selection of quantitative evaluation/ estimation method for assessing model parameters in order to test the extent to which the model is adequate to reflect the real process; 3) the conversion of identification model into time series through the statistical evaluation of model parameters. Formally, time series can be defined as a discrete function x(t) whose argument and function values are dependent on discrete time moments as well as argument values, function values at different time intervals.

It is assumed, the time interval $0 \le t \le T$ of process $X(t)$ is observed, that is to say, the parameter t various along the time interval [0,T] assumes any integer belonging to this interval. For every fixed time moment t=s, the value of function, beginning from this moment, is generally determined by the values of function arguments at all the time moments ranging from t=0 to t=s-1, and value of function at all the time moments ranging from t=0 to t=s-2. The membership ship function is the key idea introduced in fuzzy set theory to measure the degree to which the fuzzy set element meet the specific proprieties, i.e. to measure the degree of belongingness of an element in a specific fuzzy set .Consequently, the propositions used need not be true or false, but can be to any degree partially true.

Using a membership function μ, we can define a fuzzy set F on a universe of discourse U as $\mu_F(x): U \to [0,1]$ which is nothing but a mapping from the universe of discourse U into the unit interval [0,1] and $\mu_F(x)$ represent the extent to which x belongs to fuzzy set F. The concept of membership functions allows any element within the universe of discourse to have partial membership to a specific fuzzy set and also to have partial membership to other fuzzy sets.

Song and Chissom applied time invariant and time variant models to forecast the enrollment at the University of Alabama. The time Invariant and Invariant model include the following steps:

- Define universe of discourse and the intervals
- Divider the intervals
- Classify the fuzzy sets
- Fuzzify the data
- Create the fuzzy relationships
- Defuzzify the forecasting results
- Forecasting

G. Fuzzy time series using Sturgis rule:

A best forecasting accuracy using time variant and time invariant fuzzy time series and it is emphasized that the forecast uses only historical data. The significance of thesis is to reduce the mean square error when compared with the existing forecast approaches.

The appropriate number of intervals is computed as

$$I = 1 + 3.322 \log_{10} N$$

Where, I is the number of classes in the data. N is the number of observation in the data.

To compute the length of intervals as

$$L = \frac{\max - \min}{I}$$

The process of diagnostic check involves testing the assumptions of the model to recognize any areas where the model is inadequate. The statistical identification process includes whether the parameter achieves statistical significance or multicollinearity and whether the residual term is white noise or not. If the model is found to be insufficient, it is necessary to remedy and repeat step (4) until a better model is identified.

IV. RESULT AND DISCUSSION

A. Method of Forecast

The implementation of the above algorithm for the export forecasting of the Taiwan export is based on the 22 year (1990 to 2011) time series export data of the Taiwan.

Step 1: A best forecasting accuracy using time variant fuzzy time series and it is emphasized that the forecast uses only historical data. The significance of thesis is to reduce the mean square error when compared with the existing forecast approaches.

The appropriate number of intervals is computed as

$$I = 1 + 3.322 \log_{10} N$$

Where, I is the number of classes in the data. n is the number of observation in the data. Proposed in Yule's Pravidlo Rule is given by $I = 2.5 * \sqrt[4]{n}$

To compute the length of intervals as

$$L = \frac{\max - \min}{I}$$

We are discussing our proposed method. The historical data and proposed method are Table1. The universe of discourse U = [150000 497500]; the length of intervals is 69500

Step 2: Partition the universe of discourse in to five equal length intervels $u_1, u_2, \ldots, u_5$, where u_1 = [150000 2195000] ,u_2 = [295000 289000] , u_3 = [289000 3585000] ,u_4 = [3585000 428000],u_5 = [428000 497500].

If we take into account the fact that forecasting with fuzzy time series exhibits the least average error, it's necessary to find the middle point of the intervals u_m1=184750,u_m2=254250,u_m3=323750,u_m4=393250 U_m5=462750.

Step 3: Fuzzy set are defined on the universal set U.In this case "the variation in total export" is a linguistic variable that assumes the following linguistic values: A_1=(very low level export), A_2=(low level export), A_3=(normal level export), A_4= (high level exort), A_5= (very high level export). To every linguistic value here corresponds a fuzzy variable which, according to a certain rule is assigned against a corresponding fuzzy set determining the meaning of this variable.

If the value of variable U in formula is accepted as the middle point of the corresponding interval of uzzy set A_i(i=1...5) will be defined as follows:

$$A_i = (\mu_{Ai}(u_i)/u_i)u_i \ \varepsilon U, \ \mu_{Ai}(u_i)\varepsilon[0,1]$$

is a fuzzy set.

Step 4: Forecasting result for the year are calculated F(1995)=245798,that is to say, anticipated export value for year 1995 equal to 18082 persons. In orders to estimate the forecasted total export for year 1995, we must add the calculated export growth to the total export for the year 1994. In other words

$$N(1995) = 204668 + 18082 = 202508$$

Step 5: The error of the proposed method is computed by the following formula

$$Error = ((AV - FV)/Observedvalue)*100$$

Where AV is the variation in total Export for the t-th year, FV is the variation in total Export for the t-th year, Observed value for the t-th year.

The existing study applies Box-Jenkins (1970) forecasting model popularly known as ARIMA model. The ARIMA is an extrapolation method, which requires historical time series data of underlying variable. The model in specific and general forms may be expressed as follows. Let Y_t is a discrete time series variable which takes different values over a period of time. The corresponding $AR(p)$ model of Y_t series, which is the generalizations of autoregressive model, can be expressed as:

$$Y_t = \mu + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + . . . + \Phi_p Y_{t-p} + \varepsilon_t$$

Where, Y_t is the response variable at time t, $Y_{t-1}, Y_{t-2}, . . . , Y_{t-p}$ is the respective variables at different time with lags; $\phi_0, \phi_1, . . . , \phi_p$ are the coefficients; and ε_t is the error factor. Similarly, the MA (q) model which is again the generalizations of moving average model may be specified as: MA (q)

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - . . . - \theta_q \varepsilon_{t-q} - \varepsilon_t$$

Where, μ is the constant mean of the series $\theta_1, \theta_2, . . . , \theta_q$ is the coefficients of the estimated error term; ε_t is the error term. Combining both the model is called as ARIMA models, which has general form as:

$$Y_t = \mu + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + . . . + \Phi_p Y_{t-p} + \varepsilon_t$$

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - . . . - \theta_q \varepsilon_{t-q} - \varepsilon_t$$

If Y_t Y_t is stationary at level or I(0) or at first difference I(1) determines the order of integration, which is called as ARIMA model. To identify the order of p and q the ACF and PACF is applied. The details of the estimation and forecasting process are discussed below. The different model can be obtained for various combinations of autoregressive and moving average. The best model is obtained with the following diagnostics low Akaike Information Criteria (AIC) which is defined by

$$AIC = -2\log L + 2m$$

Where $m = p + q$ and L is the likelihood function. Since $-2\log L$ -2 log L is approximately equal to $n(1 + \log 2\pi) + n\log\sigma^2$ Where σ^2 is the mean square error. Also AIC can be written as:

$$AIC = \left(n(1 + \log 2\pi) + n\log\sigma^2 + 2m\right)$$

And Schwartz Bayesian Criteria (SBC) is defined by:

$$SBC = \frac{\log\sigma^2 - (m\log n)}{n}$$

To check the adequacy for the residuals using Q statistic. A modified Q statistic is the Box-Ljung Q statistic is defined by

$$Q = \frac{n(n+2)\Sigma rk2}{(n-k)}$$

Where $rk -$, the residual autocorrelation at lag k, N = the number of residuals. The Q statistic is compared to critical value from Chi square distribution. If the p-value associated with Q statistic is small $(p < \alpha)$, the model is considered in adequate. Forecasting the future periods using the parameters for the tentative model has been selected.

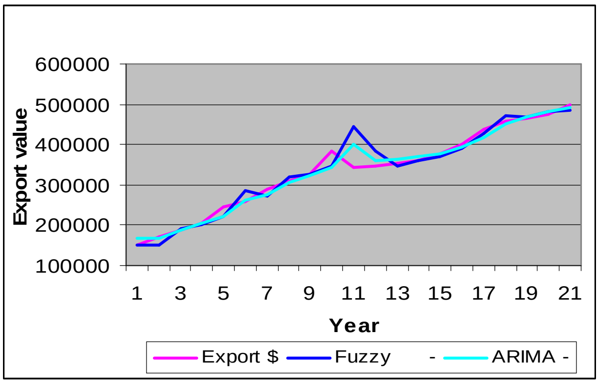

Fig 1: Actual and Fuzzy, ARIMA predictive value of Taiwan Export

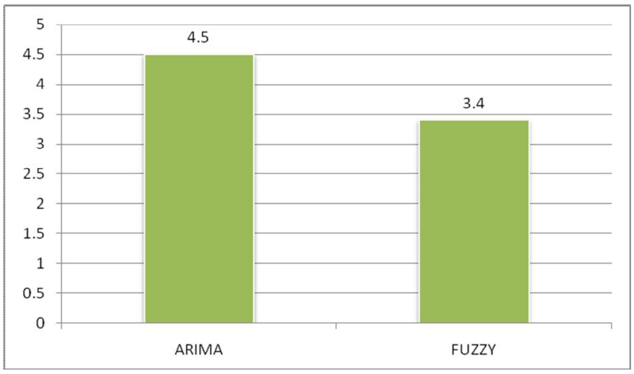

Fig 2: The Average Error result of the Fuzzy & ARIMA Model

TABLE 1
A COMPARISION OF THE AVERAGE ERROR OF PROPOSED MODEL

TABLE 2
FORECASTED TOTAL EXPORT VALUE OVER THE 2012-2016 YEAR

Year	Fuzzy	ARIMA
2012	518835	513647
2013	540522	530146
2014	562209	546645
2015	583896	563144
2016	605583	579643

V. CONCLUSION

The methodology proposed in this paper enable us to forecast Taiwan export value on the basis of fuzzy time series. A peculiar quality of the methodology consists of its capability to forecast the required indicator by utilizing incomplete, fuzzy input data. The described approach, by entering the total export value until some previous year into an experimental base, helps to make forecast calculations for any distant perspective. This, in its turn, allows us to take into account the trend of previous export value rate and as a result achieve more accurate forecast.

In time series and fuzzy time series we can predict the future values. Already we know that using time series analysis one can predict the future values. In this paper, the result of ARIMA (1 1 1) model had the lowest BIC value and its model is $X_t = 4252657 - 0.23905 - 0.40574 + \varepsilon_t$ It means that the amount of export is affected by the residual error of lag 1 and lag 13. We have proved that fuzzy time series also used to predict the future values. In time series using Box Jenkins Methodology one can forecast the average error .In fuzzy time series also we can able to find the average error using the formula ((AV-FV)/Observed value)*100, It was observed average error of fuzzy time series is less than the time series average error value. So we conclude that if we follow fuzzy time series we can get better result than time series.

ACKNOWLEDGMENT

The authors would like to thank the reference for given that very helpful comments and suggestions, their insight and comments led to a better presentation of the idea expressed in this paper. This research work was supported by the university Grant Commission, Delhi.

REFERENCES

Models	ARIMA	FUZZY
Average Error (%)	4.5	3.4

[1] Chi-chen wang," A comparison study between fuzzy time series model and ARIMA model for Forecasting Taiwan export, *Expert system with Application*, Vol.*38,Pp. 9296-9304.2011.*

[2] Yun-sheng Hsu et.al "A Comparison of ARIMA forecasting and heuristic modeling", *Applied Financial Economics*,Vol. *21,Pp. 1095-1102,2011.*

[3] Arumugam .P and Anithakumari. V, "A comparison of Refined Arithmetic and Fuzzy Time series Model for Taiwan Export Data, *Proceeding of the International Conference on Mathematical Modeling and Applied Soft Computing,Volume-2, Pp.210-220,2012*

[4] AAbbasov A.M., Mamedova M.H., Gasimov V.A. Fuzzy Relational Model for Knowledge Processing and Decision Making –*Advance in Mathematics, New – York, 2002, vol.1, pp, 191-223.*

[5] Narendarakumar et.al," Fuzzy time series forecasting of Wheat production", *International on Computer Science and Engineering, Issu l02, Vol No.03, Pp*, 65-64, 2010.

[6] Shyi-ming Chen,"Forecasting enrollment Based on high order fuzzy time series," *Cybernetics and system, An International Journal, copyright, Taylor and Francis* Vol, *33, Pp, 1-16, 2002.*

[7] Chung-ming own pao-Ta-Yu ,"Forecasting fuzzy time series on a heuristic high order mode"l, *Cybernetics and System, An International Journal, Taylor& Francis Inc.* Vol, *36,Pp:705-717, copyright 2005,*

[8] Kunhuang Huarng.,"Heuristic model of fuzzy time series for forecasting", *Fuzzy set and system, Elsever Vol: 123, Pp.369-386, 2001.*