Pertanika J. Sci. & Technol. 25 (3): 787 - 796 (2017)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Combination of Forecasts with an Application to Unemployment Rate

Muniroh, M. F.[1]*, Ismail, N.[2] and Lazim, M. A.[3]

[1]*Actuarial and Statistics Branch, Department of Risk and Research, Social Security Organizational, Kuala Lumpur, Malaysia*

[2]*School of Mathematical Sciences, Faculty of Sciences and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*

[3]*Department of Statistics and Decisions Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 UiTM, Shah Alam, Selangor, Malaysia*

ABSTRACT

Combining forecast values based on simple univariate models may produce more favourable results than complex models. In this study, the results of combining the forecast values of Naïve model, Single Exponential Smoothing Model, The Autoregressive Moving Average (ARIMA) model, and Holt Method are shown to be superior to that of the Error Correction Model (ECM).Malaysia's unemployment rates data are used in this study. The independent variable used in the ECM formulation is the industrial production index. Both data sets were collected for the months of January 2004 to December 2010. The selection criteria used to determine the best model, is the Mean Square Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). Initial findings showed that both time series data sets were not influenced by the seasonality effect.

Keywords: Combination forecast, unemployment rate, error correction model

INTRODUCTION

Forecasting is an important tool for making decisions in a variety of fields. It helps government and top management of firms, in their decision making for strategic planning purposes. There are tremendous diversities in forecasting applications, such as in marketing where it plays a key role in determining the sales targets, pricing and advertising expenditure, in sales and services that allows salespersons to estimate sales to be achieved.

Governments use forecasts to guide monetary and fiscal policy and make plans for

Article history:
Received: 27 May 2016
Accepted: 14 November 2016

E-mail addresses:
Muniroh.fadzil@gmail.com, Muniroh.fadzil@perkeso.gov.my (Muniroh, M. F.),
ni@ukm.edu.my (Ismail, N.),
dralias@tmsk.uitm.edu.my (Lazim, M. A.)
*Corresponding Author

ISSN: 0128-7680

the country's future directions while private firms use them to determine prices and product demand. Currently, there are many forecasting methods but not all are able to provide the best results. This has made the issue pertaining to the accuracy of forecasts to become a major topic in forecasting research. Study done by Dimitrios & John (2004), to compare the forecasting performance of the Naïve model, ARIMA model and Transfer Function (TF) model. In the Transfer Function they were interested in the contemporaneous relationship among variables such as Gross Domestic Product (GDP), Consumer Price Index (CPI), Industrial Production Index (IPI), Real Cost Labour, and OECD Industrial Production Index.

The availability of various forecasts methods allows the forecasters to examine their respective performance and hence to identify strengths and weaknesses. ARIMA models are based on autocorrelations while exponential smoothing is based on a structural view of the data to include level, trend and events. Exponential smoothing attempts to estimate the trend as a part of the modelling process in comparison with the ARIMA methodology which attempts to eliminate the trend before constructing the model. One suggestion regarding forecasting performance is to combine forecast values. This study aims to compare the performance of the combined univariate models with the more complex multivariate models using data of the unemployment rate. The four univariate models analysed are Naïve model, Single Exponential Smoothing, Holt-Winter Smoothing, Autoregressive Moving Average (ARIMA). The multivariate models studied using industrial production index (IPI) as the independent variable is the Error Correction Model (ECM).

Assis, Amran, Remali and Affendy (2010) compared the performance of four models in their prediction of the Cocoa Bean prices. The data used were Tawau Cocoa Bean prices graded SMC 18 for the period of January 1992-December 2006. The four univariate models used to make the comparisons were Single Exponential Smoothing, ARIMA, GARCH and Mixed ARIMA/GARCH models.

From the result, it was found that mixed ARIMA/GARCH model outperformed the Single Exponential Smoothing, ARIMA and GARCH models. The best result is indicated by the smallest error measures, the Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and statistics. This is in agreement with the findings of the study done by Zhou, He & Sun (2006). Zhou, He & Sun (2006) Fatimah and Roslan (1986), Shamsuddin, Rosdi & Ann (1992), and Kahforoushan, Zarif & Mashahir (2010). On the other hand, Kamil, & Noor (2006) found that the GARCH model type has good predictive accuracy.

To assess the forecasting performances of each of the models that were developed various error measures were used, and the best model was ascertained on the out-of sample forecast evaluation procedure. Clemen (1989) found that combining forecast can produce better and more accurate results.

The performance of a single variable model such as Naïve, Exponential Smoothing, the ARIMA model and all univariate models show a diversity in their level of accuracy. . The inclusion of exogenous and endogenous variables like the Error Correction Model (ECM) can make them better. The forecasting performances of these models are measured using several statistical criteria such as Akaike Info Criterion (AIC), Schwarz Criterion (BIC) and F-statistics. The objective of this study is to compare the performance of the models based on the out-of-sample evaluation. This paper is organized as follows: Section 2 provides the data and methodology, section 3 presents the empirical results from the various methods of combining forecasts values and section 4 is the conclusion.

METHODOLOGY

An observation on the time series data to ascertain the presence of a trend, seasonal pattern was made. . Following this five models, namely the Naive model, Single Exponential Smoothing, Holt Method, Autoregressive Moving Average (ARIMA) and Error Correction Model were used to fit the data. Analyses was done using EViews software to develop statistical models from the data and forecast the future values of the series.

The Data Series

Data used in this project was obtained from the Department of Statistics, Malaysia for months January 2004 till December 2010.

Modelling Procedures

The development of the five models suitable for forecasting purposes using the data series were based on the three step procedure described below:

1. Firstly the data series was split into two parts. The first part is called model estimation part where all models studied are estimated and fitted. The second part is called the evaluation part where the fitted models are compared out-of-sample to determine which model gives on the average, the best forecast values. In this study, the monthly data from January 2004 until December 2009 were used as the fitted (estimation) part, whereas data from January 2010 until December 2010 is the evaluation part. The sample size of the estimation part comprises of 72 observations and the sample size for the evaluation part 12 observations.

2. Secondly five models , namely Naive model, Single Exponential Smoothing, Holt Method, Autoregressive Integrated Moving Average (ARIMA), acted as univariate models while the Error Correction Model was the multivariate model. These models will be estimated on the estimation part of the data set, comprising 72 observations. In the multivariate model, the study adopted simple model to see the relationship between x_t and y_t where x_t is industrial production index while y_t is unemployment rate. To express this relationship, Error Correction Model (ECM) will be used.

3. Thirdly, these models were evaluated based on out-of sample procedure. The model that produces the best results will be declared the most suitable. The decision to pick the best model will be based on the minimum of the error measures. The error measures used are the Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

The Model Used

Naive Model. This model is founded on the belief what happens today will happen again tomorrow or any other time in the future. Frechtling (1996) Summarized that the forecast value using the naive method is equal to the actual value from the last period available. This model can be used in a stable data series, which include seasonal component or any pattern. Naive model is also used as a benchmark for forecasting model. Mathematically, the model is represented as,

$$F_{t+m} = y_t \tag{1}$$

where *m* refers to the number of periods into future for which the forecast is desired and y_t is actual value at time *t*. For the one-step-ahead forecast model can present as $F_{t+1} = y_t$ where *m=1*.

Single Exponential Smoothing. In single exponential smoothing models, the forecast for the next and all subsequent periods are determined by adjusting the current period forecast by a portion of the difference between the current forecast and the current actual value. The formula for simple exponential smoothing is;

$$F_{t+m}=\alpha\, y_t + (1- \alpha)\, F_t \tag{2}$$

where, $F_{t+m}$is the single exponentially forecast value in period *m* (this is also defined as the forecast value when generated out-of-sample), for $m= 1,2,3\ldots, t$. y_t is the actual value in time period t, α is unknown smoothing constant to be determined for value between 0 and 1, i.e. ($0\leq\alpha\leq1$) and F_t is the forecast value at period t.

Holt Methods. Holt method provides more flexibility in selecting the parameter value, which include the trend and slopes. This method also can be used to forecast when linear trend is present in the data. The application of the Holt's Method requires three equations;

The exponentially smoothed series;

$$S_t = \alpha y_t - (1-\alpha)(S_{t-1} + T_{t-1}) \tag{3}$$

The trend estimate;

$$T = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1} \tag{4}$$

The forecast for m period;

$$F_{t+m} = S_t + T_t * m \tag{5}$$

Autoregressive Moving Average (ARIMA). The Box-Jenkins methodology of forecasting differs from the methods described earlier because it does not assume any particular pattern in the historical data of the series to be forecast. It uses an iterative approach to identify a possible model from a general class of models. The chosen models are then checked against the historical data to see whether they accurately describe the series. This model also requires expert criteria, stationary data, models' diagnostic and etc. Funke (1992) and Floros (2005) also used a univariate ARIMA model to forecast the German and United Kingdom unemployment rates respectively.

As with any good modelling practice the first step when developing a Box-Jenkins model is to understand the characteristics of the series involved, that is how it behaves over time. Basically, it is assumed that the data series is stationary. A series is said to be 'stationary' if it does not show growth or decline over time. In other words, the data series does not indicate presence of trend component. A series that does not depict this characteristic is called 'non-stationary series' and such series can be made stationary (by removing the trend) by taking successive differences of the data. The basic model of Box-Jenkins methodology comprises of the autoregressive (AR) part and the moving average (MA) part. However, when the variable is non-stationary, then the differencing or integrated part is also included.

The general Autoregressive (AR) model of p^{th} term is given as,

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \tag{6}$$

The general Moving Average (MA) of q^{th} term is given as,

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \tag{7}$$

Mixed Autoregressive and Moving Average (ARMA) model of p^{th} and q^{th} term, respectively (Where y_t is assumed stationary) is given as,

$$\begin{aligned} y_t &= \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} \\ &+ \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \end{aligned} \tag{8}$$

Mixed Autoregressive Integrated Moving Average (ARIMA)

$$\begin{aligned} w_t &= \mu + \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots \\ &- \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots + \varepsilon_t \end{aligned} \tag{9}$$

where $w_t = y_t - y_{t-1}$ is stationary and y_t is the original series p and q can take any value which define the lag length of y_t and ε_t, respectively, y_{t-p} is the p^{th} order of the legged dependent and μ, ϕ_p and θ_q are unknown parameters to be estimated.

Error Correction Model (ECM). According to Engle & Granger (1987) if the set of series x_t and y_t are co-integrated, then there exists a generating mechanism called 'Error Correction Model', which forces the variables to move closer over time whilst allowing a range of short run dynamics. For example, if two variables y_t and x_t are co-integrated of order one, that is~ CI (1, 1), then the general error correction model (ECM) can be written as:

$$\Delta y_t = \alpha_0 + \varphi v_{t-1} + \sum_{j=1}^{J} \gamma_j \Delta y_{t-j} + \sum_{k=1}^{K} \sum_{j=1}^{J_k} \beta_{kj} \Delta x_{k(t-j)} + \varepsilon_t \tag{10}$$

where $\Delta y_t = y_t - y_{t-1}$, $\Delta x_t = x_t - x_{t-1}$, $\varepsilon_t \sim iid(0, \sigma^2{}_\varepsilon)$ φ necessarily negative and $v_t = \hat{y}_t - (\beta_o + \hat{\beta}_1 t)$

Combining Forecasts. Following the discussions made earlier in this study, this section continues with the discussion on combining the forecasts in order to obtain much better forecast values. Combination of forecasts will involve four models of univariate type and the result obtained will be compared against the Error Correction model. In this study, a simple procedure to combine the forecasts used is to take the average of the four models. To generate the average forecast values, assume all forecasts are independent with the same σ^2. A arithmetic mean of these four forecast values are calculated as,

$$SAM = \frac{f_1 + f_2 + f_3 + f_4}{4} \tag{11}$$

Another method of combining forecast is to use the regression analysis**.** The most common procedure used to estimate the combining weight is to perform the Ordinary Least Squares (OLS) regression. The Regression Method (RM) can be expressed by the following equation:

$$y_{t+1} = a_o + \sum_{j=1}^{k} a_j f_{t,j} + \varepsilon_{t+1} \tag{12}$$

where $f_{t,j}$ is the one step ahead forecasts made at time t of y_{t+1} with model i: a_0 is a constant term and a_j is the regression coefficient. As an improvement to the combining methods, Granger and Ramanathan (1984) showed that the optimal method is equivalent to a least squared regression in which the constant is suppressed and the weight are constrained to sum to one. They point out that values from discarded forecasting models still obtain useful information about the behaviour of y_t. If the biased forecasts are included in the least squares equation, the intercept adjusts for the bias. Hence, it is important to use least squares method with an intercept. Therefore, this study used both combination methods to evaluate the performance of forecast models. The smallest error measure is the criterion used to show that the combined of univariate model is better than the single complex models.

ANALYSIS AND RESULTS

Figure 1 shows the monthly observations of Malaysia's Unemployment Rate from January 2004 until November 2010. From the plot, it was found that the Malaysia's Unemployment rate had been on the decline with the exception of the periods Dec 2005 until Mac 2006 where the rate recorded its highest levels. The series however, does not indicate presence of seasonality effect, though significantly large values in June and February in year 2007 and 2008 respectively were observed. This phenomenon could be explained by the fact that labour market conditions had weakened at end of year 2008 as reflected in higher retrenchments whilst demands for labour were moderate.

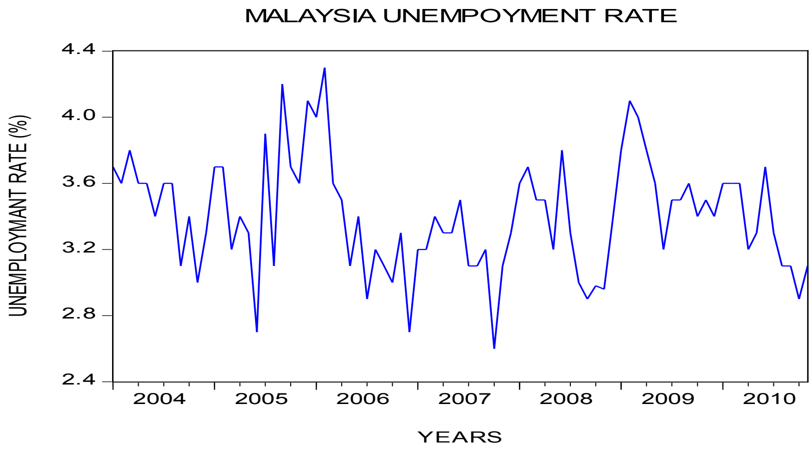

Figure 1. Malaysia unemployment rate

Table 1
Results of Unit Root ADF Test for Unemployment Rate and Industrial Production Indices

Variables	ADF Test Statistics	t-statistics	Probability
Unemployment Rate			
Before differencing	-3.4753	-3.1803	0.0980
After 1st Differencing	-3.4753	-14.2636	0.0001
Industrial Production Indices (IPI)			
Before differencing	-3.4805	-0.2539	0.9903
After 1st Differencing	-3.4805	-4.5019	0.0033

Table 2
Results of Unit Root ADF Results Test for Co-Integrated

Variables	ADF Test Statistics Critical Value at 5% Level	t-statistics	Probability
Before differencing	-3.4753	-3.4501	0.0531
After 1st Differencing	-3.4753	-14.5626	0.0001

ECM versus Univariate Models

Table 3
The Estimation and Evaluation Models

Model Types	Error Measure		
	MSE	**MAE**	**MAPE**
Estimation Period: January 2004-December 2009			
Naïve Model	0.167800	0.289634	8.475200
Single Exponential Smoothing Alpha = 0.5100	0.033300	0.138410	4.112000
Holt Method (Alpha=0.52)	0.025100	0.121633	3.610700
ARIMA (1,1,3)	0.071560	0.201220	6.006736
ECM	0.085671	0.219243	6.523044
Combination Forecast (SAM)	0.004878	0.051120	1.553687
Combination Forecast (RM)	0.000531	0.016369	0.484067
Evaluation Period: January 2010-December 2010			
Naïve Model	0.049400	0.162993	4.984000
Single Exponential Smoothing Alpha = 0.5100	0.012400	0.081667	2.506600
Holt Method (Alpha=0.52)	0.012240	0.080419	2.935200
ARIMA (1,1,3)	0.035928	0.145557	4.388857
ECM	0.029192	0.114284	3.464466
Combination Forecast (SAM)	0.000996	0.022435	0.675566
Combination Forecast (RM)	0.000077	0.007968	0.238952

For the Error Correction Model, both unemployment rate and industrial production indices show that they are co-integrated. That is proven using the ADF test which shows that at least two variables are integrated of the first order (Table 1 and Table 2). Looking at the results of the model evaluation from the table, we can see that the ECM performed worst when compared against the deterministic models of Single Exponential Smoothing and Holt Method. However, both ARIMA (1,1,3) and the ECM were beaten when pitted against the combined (in this case the average forecast) forecast of all the deterministic models. The table also shows that the combined Simple Average Method and Regression Method forecasts are superior than ARIMA (1,1,3) and ECM. But, the combination forecasts from Regression Methods perform slightly better when compared to Simple Average Method.

CONCLUSION

In this study five univariate models were estimated and evaluated. The univariate models are Naive, Single Exponential Smoothing, Holt Method, Autoregressive Integrated Moving Average (ARIMA) and the multivariate model is the Error Correction Model (ECM).

From the results it can be concluded that the Holt Method shows the best result. However, after combining the forecasts using Simple Average Method and Regression Method, the combined forecast values explain better than Holt Method, proving combining several competing forecasts can help reduce errors and improve accuracy. In this regard the regression method of combination proved better than the simple average method.

The study also shows that variations in performance between univariate and multivariate analysis show that combining forecast provides better results.

REFERENCES

Assis, K., Amran, A., Remali, Y., & Affendy, H. (2010). A Comparison of Univariate Time Series Methods for Forecasting Cocoa Bean Prices. *Trends in Agricultural Economics*, *3*(4), 207-215.

Clemen, R. T. (1989). Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting*, *5*(4), 559-583.

Engle, R. F., & Granger, C. W. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, *55*(2), 1057-1072

Fatimah, M. A., & Roslan, A. G. (1986). Univariate Approach towards Cocoa Price Forecasting. *The Malaysian Journal of Agricultural Economics*, *3*, 1-11.

Floros, C. (2005). Forecasting the UK Unemployment Rate: Model Comparisons. *International Journal of Applied Econometrics and Quantitative Studies*, *2*(4), 57-72.

Frechtling D. C. (1996). *Practical Tourism Forecasting*. Great Britain: Butterworth-Heinemann.

Funke, M. (1992). Time-Series Forecasting of the German Unemployment Rate. *Journal of Forecasting, 11*(2), 111-125.

Granger, C. W., & Ramanathan, R. (1984). Improved Methods of Combining Forecasts. *Journal of Forecasting, 3*(2), 197-204.

Kahforoushan, E., Zarif, M., & Mashahir, E. B. (2010). Prediction of Added Value of Agricultural Subsections using Artificial Neural Networks: Box-Jenkins and Holt-Winters Methods. *Journal of Development and Agricultural Economics, 2*(4), 115-121.

Kamil, A. A., & Noor, A. M. (2006). Time Series Modeling of Malaysian Raw Palm Oil Price: Autoregressive Conditional Heteroskedasticity (ARCH) Model Approach. *Discovering Mathemathics*, *28*(2), 19-32.

Shamsudin, M. N., Rosdi, M. L., & Ann, T. C. (1992). An Econometric Analysis of Cocoa Prices: A Structural Approach. *Malaysia Journal of Economic, 25*(June 1992), 3-17.

Thomakos, D. D., & Guerard, J. B. (2004). NaïVe, ARIMA, Nonparametric, Transfer Function and VAR Models: A Comparison of Forecasting Performance. *International Journal of Forecasting, 20*(1), 53-67.

Zhou, B., He, D., & Sun, Z. (2006). *Modelling and Simulation Tools for Emerging Telecommunication Networks*. United States of America, USA: Springer.