

An Initial Study on the Forecast Model for Unemployment Rate

Mohd Nadzri Mohd Nasir, Kon Mee Hwa and Huzaifah Mohammad¹

Abstract

The purpose of the article is to determine the most suitable technique to generate the forecast of unemployment rate using data from the series of Labour Force Surveys. The models understudied are based on Univariate Modelling Techniques i.e. Naïve with Trend Model, Average Change Model, Double Exponential Smoothing and Holt's Method Model. These models are normally used to determine the short-term forecasts (one quarter ahead) by analyzing the pattern such as quarterly unemployment rates. The performances of the models are validated by retaining a portion of the quarterly observations as holdout samples. In addition, comparisons are made to see how well the historical and forecasted data matched and correlated. The selection of the most suitable model was indicated by the smallest value of mean square error (MSE). Based on the analysis, Holt's Method Model is the most suitable model for forecasting quarterly unemployment rates.

Keywords: *Univariate Modelling Techniques; Forecast Model; Mean Square Error.*

Introduction

Forecasting is defined as the prediction of future events based on known past values of relevant variables (Makridakis, S., Wheelright, S. C. & Hyndman, R. J., 1998). Forecasting unemployment rate accurately is important because it helps economists to have a better idea of what the future economy holds (Lewis, R., & Brown, C., 2001). Besides, it is also important for the government in terms of decision and policy making.

With the support of stable economic growth, Malaysia experienced low unemployment rates in the 1990s with the lowest recorded in 1997 at 2.4 per cent. From 1999 onwards, the unemployment rate has increased as a result of the financial crisis and subsequent economic downturn.

Univariate Modelling Techniques are methods for analyzing data on a single variable at a time. Examples of Univariate Modelling Techniques are the Naive Models, Methods of Average, the Exponential Smoothing Techniques and the

¹ Mohd Nadzri Mohd Nasir is currently the Director of Department of Statistics, Pahang, Kon Mee Hwa is Assistant Director of Services Statistics Division and Huzaifah Mohammad is Assistant Statistical Officer of Price, Income and Expenditure Statistics Division.

Box-Jenkins Methodology. Both Double Exponential Smoothing and Holt's Method illustrated in this study are classified in the Exponential Smoothing Techniques. Other models available in this same category are Single Exponential Smoothing, Adaptive Response Rate Exponential Smoothing (ARRES), Holt's Method and Holt-Winters' Trend & Seasonality.

This paper is divided into several sections. Following the introduction, the second section describes the definitions, objectives and literature review of the study, the third section focuses on the methodology and some of the attempts made to move beyond the models. In this section, a same set of unemployment data were tested using four different univariate forecasting models to obtain MSE value. The fourth section goes beyond the discussion of analysis and results while the fifth section explores selection of models. This is followed by a normality test, paired samples t-test and a correlation test on the chosen model. The final section presents an evaluation of Holt's Method and a brief conclusion.

Definition of unemployment

International Labour Organization (ILO) defines "Unemployed" as all individuals above a specified age who satisfies simultaneously the following criteria:

- a) without work (not in paid or self employment);
- b) currently available for paid employment or self-employment; and
- c) actively seeking work.

The "unemployment rate" is defined as the share of people not working, available and actively seeking for work out of the working age population (ILO Geneva, 1990).

Basic economic theory defines unemployment as a situation where supply of labour exceeds demand. The unemployment rate of 4 per cent or lower indicates that an economy is operating in full employment condition. The Keynesian view states that unemployment is an excess supply of labour resulting from a failure of coordination in the economy market whilst the classical view states that unemployment is a job search for a better match between productive worker and employer (Baharudin, N., 2004).

Objective of the study

The objective of the study is to choose the most suitable model to forecast the unemployment rate in Malaysia. The output of the study will serve as a guide in selecting a model for future forecasting/projection of unemployment rate. Forecasting on unemployment rate is one of the areas that should be developed in fulfilling the requirements at national and international levels.

Literature Review

A study undertaken by Jaafar, J.² (2006) showed that Holt's Method with two parameters were suitable to forecast the five major labour force indicators i.e. labour force, employed, unemployed, unemployment rate and underemployed. First, the method of Ordinary Least Squares (OLS) was used to estimate the trend line for each labour force indicators to project past pattern into the future for forecasting purposes. Then the seasonal ratios were calculated to identify the seasonal components in the data. By removing the seasonal component from the series, the occurrence of the fluctuations reflected the true movements of the series. Finally, Holt's Method was used to forecast the major labour force indicators while a comparison of the model's forecasting performance was determined by Mean Squared Error (MSE). The estimations were done with the objective of minimising the MSE.

Floros, C. (2005) compared the out-of-sample forecasting accuracy for the United Kingdom unemployment rate using the Root Mean Square, Mean Absolute and Mean Absolute Per cent Errors. He evaluated the performance of the competing models covering the period January 1971 to December 2002. The forecasting sample (January 1996 – December 2002) was divided into four sub-periods. First, for total forecasting sample, it was found that Moving Average, MA (4) – Autoregressive Conditional Heteroscedasticity, ARCH (1) provided superior forecasts of unemployment rate. On the other hand, two forecasting samples showed that the MA(4) model performed well, while both MA(1) and Autoregressive, AR (4) proved to be the best forecasting models for the other two forecasting periods. The empirical evidence derived from his investigation suggested a close relationship between forecasting theory and labour market conditions.

Montgomery, A. L. et al. (1998) compared the forecasting performance for a variety of linear and nonlinear time series models using the United States unemployment rate. Their main emphasis was:

- a. on measuring forecasting performance during economic expansions and contractions by exploiting the asymmetric cyclical behaviour of unemployment numbers;
- b. on building vector models that incorporate initial jobless claims as a leading indicator; and
- c. on utilising additional information provided by the monthly rate for forecasting the quarterly rate.

² The study was undertaken by the Human Resource and Social Statistics Division, Department of Statistics Malaysia with the aim to estimate the five main indicators of labour force. The results were presented in the Workshop in Reviewing the Labour Force/Migration Survey in November 2006, Putrajaya.

Comparisons were also made with the consensus forecasts from the Survey of Professional Forecasters. In addition, the forecasts of nonlinear models were combined with the consensus forecasts. The results showed that significant improvements in forecasting accuracy can be obtained over existing methods.

Methodology

This section described briefly about the statistical techniques applied to analyse the data collected through the Labour Force Survey. Univariate Modelling Techniques were applied to predict future values of unemployment rate based on past observations in a given time series, by fitting a model to the data.

The quarterly Labour Force Survey data from 1998-2007 were used to determine the suitable model. Time series forecasting analysis and forecast models were applied to predict the unemployment rate in Malaysia. Analysis was done using four types of forecast models i.e. *Naïve with Trend Model*, *Average Change Model*, *Double Exponential Smoothing* and *Holt's Method*. Subsequently, predicted unemployment rate with the best model were compared with the actual unemployment rate obtained from the Labour Force Survey to determine the accuracy of prediction. The statistical test and data analysis were done through SPSS and Microsoft Excel.

a. Naïve with Trend Model

This model implies that all future forecast can be set to equal the actual observed value in the most recent time period plus the growth rate. The value $\frac{Y_t}{Y_{t-1}}$ measures the trend. If Y_t is greater than Y_{t-1} then the trend is on the upward and likewise if Y_t is less than Y_{t-1} , then the trend is on the downward side. The one step ahead forecast is represented as, $F_{t+1} = Y_t \frac{Y_t}{Y_{t-1}}$ where Y_t is the actual value in time t, and Y_{t-1} is the actual value in the preceding time period.

This model is highly sensitive to the changes in the actual values. A sudden drop or sharp increase in the values will severely affect the forecast. Furthermore, fitting this model type will result in the loss of the first two observations in the series. On the other hand, this model is only suitable to be used for short time series.

b. Average Change Model

The average change model is based on the premise that the forecast value is equal to the actual value in the current period plus the average of the absolute changes experienced up to that point in time. The one step ahead forecast is given as:

$$F_{t+l} = Y_t + \frac{(Y_t - Y_{t-1}) + (Y_{t-1} - Y_{t-2})}{2}$$

This model is useful when the historical data being analysed are characterised by period-to-period changes that are approximately of the same size. However, this model tends to lag behind turning points and that all periods are weighted equally, irrespective of their importance, when deriving the forecast values.

c. Double Exponential Smoothing

This technique is also known as Brown's method. It is useful for series that exhibit a linear trend characteristic. The following notations are used:

Let, S_t be the exponentially smoothed value of Y_t at time t

S'_t be the double exponentially smoothed value of Y_t at time t

Generally, there are four equations involved.

Equation 1: Computes the single exponentially smoothed value

$$S_t = \alpha Y_t + (1 - \alpha) S_{t-1}$$

Equation 2 : Computes the double exponentially smoothed value

$$S'_t = \alpha S_t + (1 - \alpha) S'_{t-1}$$

Equation 3 : Computes the difference between the exponentially smoothed values

$$a_t = 2S_t - S'_t$$

Equation 4 : Computes the adjustment factor

$$b_t = \frac{\alpha}{1 - \alpha} (S_t - S'_t)$$

Therefore, the l -step-ahead forecast is computed by using the equation,

$$F_{t+l} = a_t + b_t l$$

where F_{t+l} is the l -step-ahead forecast at period l made in period t for $l=1,2,3,\dots$

There are several advantages that can be obtained when using double exponential smoothing technique:

- a) exponential smoothing models mesh very easily with computer system and hence, simple spreadsheet program such as Microsoft Excel can be used to generate new forecasts;
- b) data storage requirements are minimal when compared to other forecasting models;
- c) it embodies the advantages of a weighted moving average since current observations are assigned larger weights;
- d) exponential smoothing models react more quickly to changes in data patterns than the moving average; and
- e) it does not require as much data as the Box-Jenkins methodology or the econometric modelling technique.

The main difficulty encountered when using this method is the determination of the size of α . The criterion is to choose α such that the MSE is minimum. However, with the assistance of solver facility available in Microsoft Excel, the enormous amount of work is lessened when searching for the best parameter α value.

d. Holt's Method

Holt's Method is a technique that takes into account to smooth the trend and the slope directly by using different smoothing constants. It also provides more flexibility in selecting the parameter value which the trend and slopes are tracked. Holt's Method consists of three basic equations that define the exponential smoothed series and the trend estimate. The Holt's Method equations are represented as follows:

Exponentially smoothed series:

$$S_t = \alpha Y_t + (1 - \alpha)(S_{t-1} + T_{t-1})$$

Trend estimate:

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1}$$

Therefore, the one step ahead forecast is:

$$F_{t+1} = S_t + T_t \quad (1)$$

where

S_t = exponentially smoothed series

Y_t = actual values

T_t = trend estimate

α = smoothing constant ($0 < \alpha < 1$)

β = smoothing constant for the trend estimate ($0 < \beta < 1$)

Mean Squared Error (MSE)

MSE is the standard error measure for assessing the model's fitness to a particular data and comparing the model's forecasting performance. For the one step a head forecast, the MSE is written as:

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

for which $e_t = Y_t - \hat{Y}_t$

where,

Y_t is the actual observation at the time t .

$\hat{Y}_t$ is the fitted value in time t generated from the origin ($t = 1, 2, 3, \dots, n$)

n is the number of out-of-sample error terms generated by the model.

Estimation and Evaluation Procedures

Basically, there are three stages involved:

- i. In the first stage, the series is divided into two parts. The first part is called model estimation part (or fitted part) and the second part is the evaluation part (or holdout part), which will be used to evaluate the model's forecasting performance;
- ii. In the second stage, the models are tested using various forms of functional relationship and variable selections; and
- iii. In the third stage, the minimum value of α and β are determined by 'Solver' facility available in Microsoft Excel which derived parameter values from data series for the related model. Then, all the models with the smallest MSE value are evaluated by comparing the MSE value of each model. Usually, a benchmark model is used as the basis of comparison against the actual data.

The model that meets all the criteria is thus selected as the most suitable model. The selection criterion is based on the results of comparing their respective error measures.

Analysis and Results

Univariate forecasting model was used to predict unemployment rate for the first quarter of year 2008 in Malaysia. The estimations were done using data from first quarter of 1998 to the fourth quarter of 2003 while the remaining observations from first quarter of 2004 to the fourth quarter of 2007 were used to evaluate the model's forecasting performance. Unemployment rate (%) for the period of study from the first quarter 1998 to fourth quarter 2007 is as shown in Table 1.

Table 1: Quarterly Unemployment Rates, Malaysia, 1998 to 2007

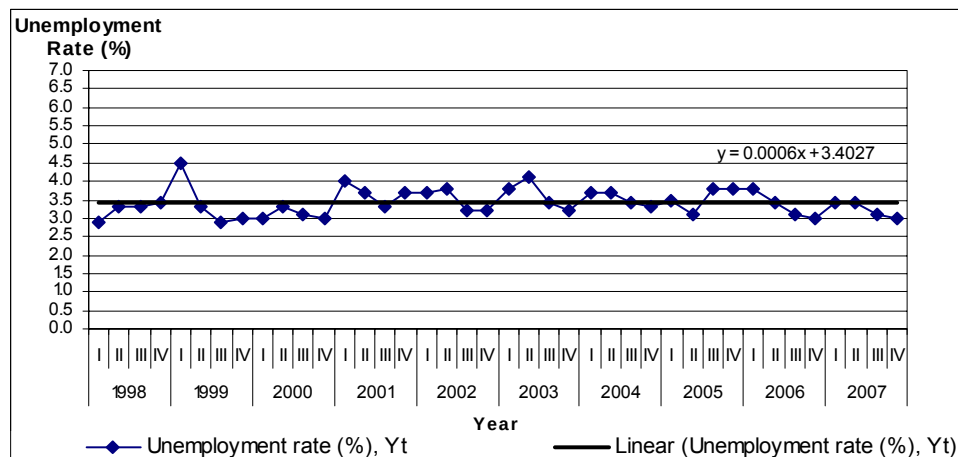
Year	Quarter	Unemployment rate, (%), Y_t
1998	I	2.9
	II	3.3
	III	3.3
	IV	3.4
1999	I	4.5
	II	3.3
	III	2.9
	IV	3.0
2000	I	3.0
	II	3.3
	III	3.1
	IV	3.0
2001	I	4.0
	II	3.7
	III	3.3
	IV	3.7
2002	I	3.7
	II	3.8
	III	3.2
	IV	3.2
2003	I	3.8
	II	4.1
	III	3.4
	IV	3.2

Year	Quarter	Unemployment rate, (%), Y_t
2004	I	3.7
	II	3.7
	III	3.4
	IV	3.3
2005	I	3.5
	II	3.1
	III	3.8
	IV	3.8
2006	I	3.8
	II	3.4
	III	3.1
	IV	3.0
2007	I	3.4
	II	3.4
	III	3.1
	IV	3.0

Source: Labour Force Survey Report, Department of Statistics Malaysia

Figure 1 shows the graph and the trend line of the unemployment rates from first quarter of 1998 to fourth quarter of 2007. Values indicate that maximum rates are reached during the first and second quarter, while minimum rates are reached during the third and fourth quarter. The overall trend line equation for quarterly data is given by $y = 0.0006x + 3.4027$. The trend line indicates that the underlying pattern of the data follows a relatively stable upward trend.

Figure 1: Quarterly Unemployment Rate, Malaysia, 1998 - 2007

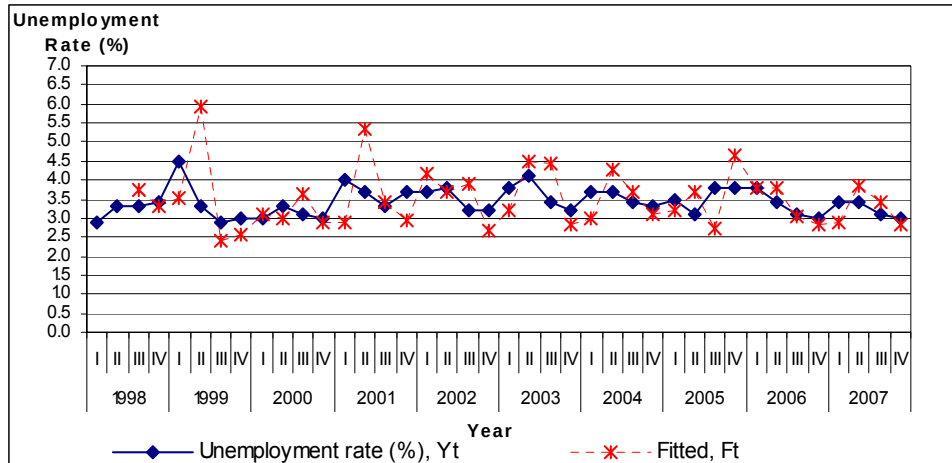


Univariate Modelling Techniques

The estimations were done with the objective of minimising Mean Squared Error (MSE). Results of the corresponding MSE value for each model are shown below.

Naïve with Trend

Figure 2: Fitted Naïve With Trend Model, Malaysia, 1998 - 2007



Fitted Period (1998 – 2003):

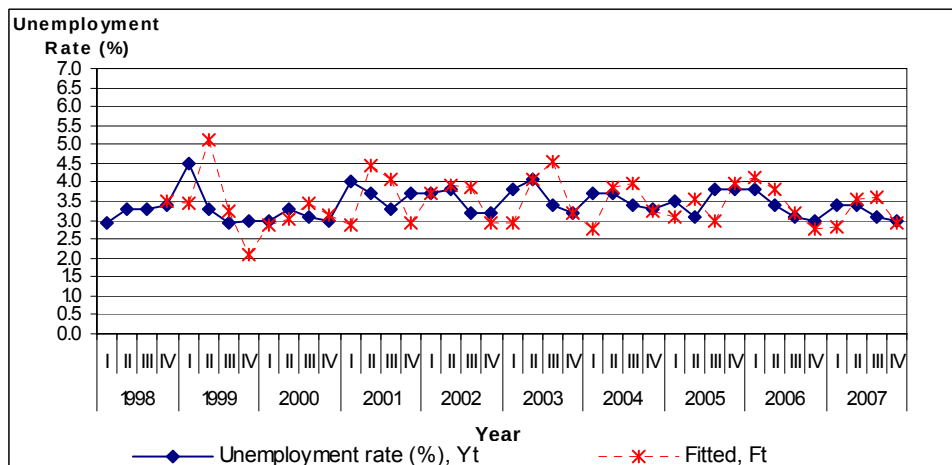
MSE = 0.6762

Evaluation Period (2004 – 2007):

MSE = 0.2503

Average Change Model

Figure 3: Fitted Average Change Model, Malaysia, 1998 - 2007



Fitted Period (1998 – 2003):

MSE = 0.4647

Evaluation Period (2004 – 2007):

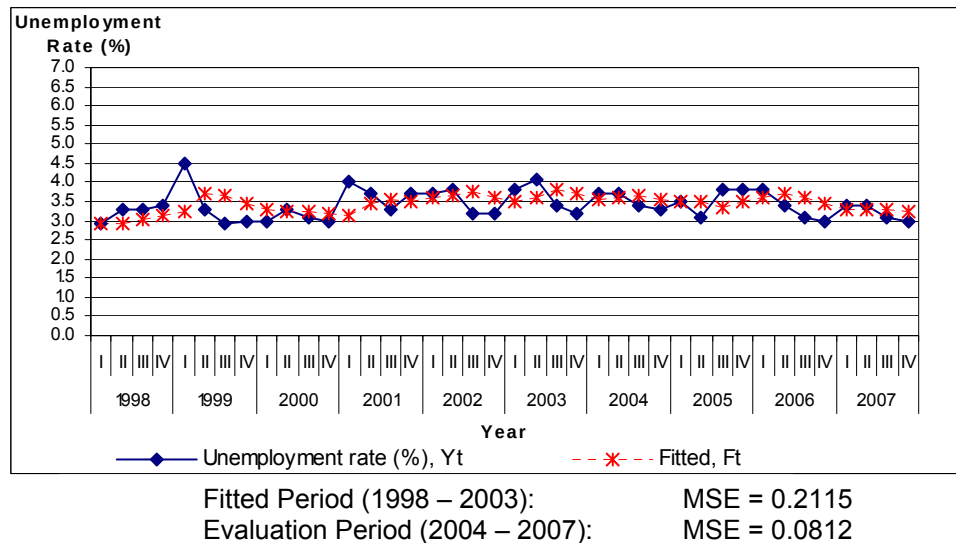
MSE = 0.2028

The determinant for the best forecasting model is proven by a smaller MSE value. For example in the above cases, Naïve with Trend Model (MSE = 0.2503) generated a better forecast data if compared with Average Change Model (MSE = 0.2028) which has higher MSE values. Therefore, MSE is the best standard error measure for assessing the fitness and the forecasting performance by comparing the MSE value of each model.

Double Exponential Smoothing

Computation of the minimum value of α was determined by solver facility available in Microsoft Excel. Based on the solver result, the best α to use is 0.1767 since it minimises the error measure (Figure 4).

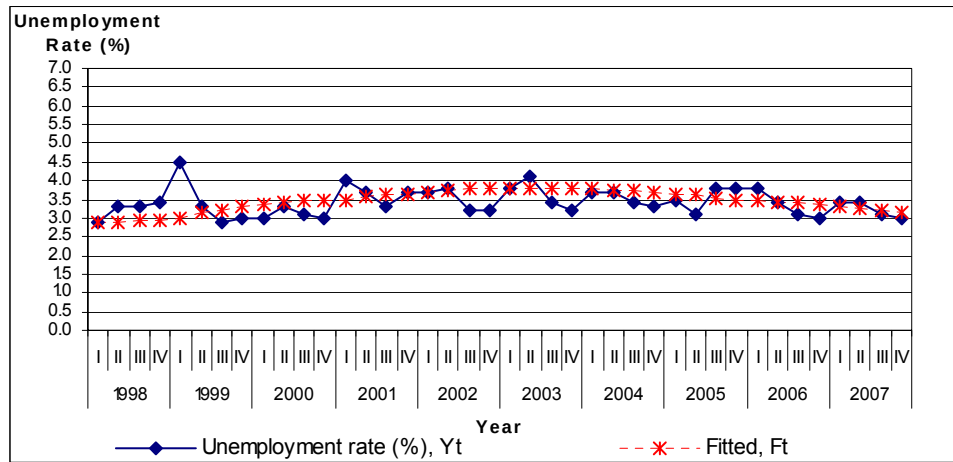
Figure 4: Fitted Double Exponential Smoothing Model, Malaysia, 1998 - 2007



Holt's Method

Computation of the minimum value of α and β were done using solver facility available in Microsoft Excel. Based on the solver result, the best parameter to use is $\alpha = 0.03$ and $\beta = 1.0$, since it minimises the error measure (Figure 5).

Figure 5: Fitted Holt's Method, Malaysia, 1998 - 2007



Fitted Period (1998 – 2003):

MSE = 0.2130

Evaluation Period (2004 – 2007):

MSE = 0.0722

Selection of Model

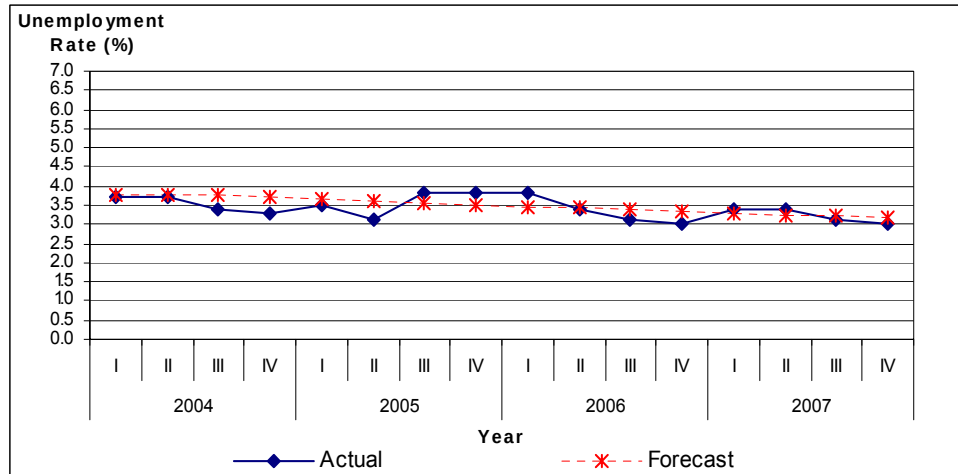
Table 2 presents the summaries and comparison on MSE figures for Naïve with Trend Model, Average Change Model, Double Exponential Smoothing and Holt's Method. On the basis of the size of MSE calculated over the evaluation period, it can be concluded that the most suitable model to forecast the unemployment rate is Holt's Method with $\alpha = 0.03$ and $\beta = 1.0$ since it has the smallest value of MSE compared to other forecasting techniques.

Table 2: MSE values by type of model

Period	Type of model			
	Naïve with Trend	Average Change Model	Double Exponential Smoothing, $\alpha=0.1767$	Holt's Method $\alpha=0.03, \beta=1.0$
Fitted Period: (1998 – 2003)	0.6762	0.4647	0.2115	0.2130
Evaluation Period: (2004 – 2007)	0.2503	0.2028	0.0812	0.0722

As mentioned earlier, 16 data points (first quarter of 2004 to fourth quarter 2007) are used as holdouts or evaluation period for the purpose of model validation. Figure 6 shows how the Holt's Method with $\alpha = 0.03$ and $\beta = 1.0$ forecasts compared with those actual values for the holdout sample. The results indicated that out-of-sample forecasts track closely the actual data.

Figure 6: Results from the Holt's Method Model



The forecasts and the actual unemployment rate of Holt's Method Model showed the same trend. Several tests were performed to ascertain the normality, level of significance and correlation between the actual and forecast values obtained from the recommended model.

Before any statistical technique is used to compare means, it is important that the normality assumptions are met. Therefore, normality test on the actual and forecast data was done.

Based on the Kolmogorov-Smirnov Test (significance value greater than 0.05), the distribution of the actual and forecast values are normal (Table 3).

Table 3: One Sample Kolmogorov-Smirnov Test for actual and forecast values of unemployment rate

		Actual	Forecast
N		16	16
Normal Parameters ^{a,b}	Mean	3.406	3.494
	Standard Deviation	0.291	0.208
Most Extreme Differences	Absolute	0.166	0.152
	Positive	0.166	0.113
	Negative	-0.156	-0.152
Kolmogorov-Smirnov Z		0.665	0.607
Asymptotic Significance (2-tailed)		0.768	0.855

a. Test distribution is Normal.

b. Calculated from data.

Paired samples t-test

The paired samples t-test is used to compare the means of actual and forecast values of unemployment rates.

Table 4: Results of paired samples t-test for actual and forecast

		Mean	N	Standard Deviation	Standard Error Mean
Pair 1	ACTUAL	3.4063	16	0.29090	0.07273
	FORECAST	3.4938	16	0.20807	0.05202

		Paired Differences					t	df	Significance (2-tailed)
		Mean	Standard Deviation	Standard Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Actual - Forecast	-0.0875	0.26552	0.06638	-0.2290	0.0540	-1.318	15	0.207

Table 4 presents the results of paired samples t-test using SPSS. The results showed there is no significant difference in the mean of actual and forecast unemployment rates. The mean absolute percentage error for the two values is 0.0875 per cent. Therefore, it can be concluded that the Holt's Method is more appropriate in forecasting unemployment rate.

Coefficient of Correlation

The coefficient of correlation measures the degree of linear and direction of the relationship of variables.

Table 5 : Results of Pearson Correlation

		Actual	Forecast
Actual	Pearson Correlation	1.000	0.474*
	Significance (1-tailed)	.	0.032
	N	16	16
Forecast	Pearson Correlation	0.474*	1.000
	Significance (1-tailed)	0.032	.
	N	16	16

*. Correlation is significant at the 0.05 level (1-tailed).

Table 5 exhibits the results of Pearson correlation using SPSS which concluded that there is significant correlation of p-value equals 0.032 and positive moderate correlation of 0.474 ($r \approx 0.5$) between the actual and forecast values. This suggests that the Holt's Method with $\alpha=0.03$ and $\beta=1.0$ is suitable in forecasting the unemployment rate.

An Evaluation of Holt's Method

The forecasting of unemployment rate has become one of the major fields of research in recent years. It serves as an important indicator in human resource development planning and policy formulation.

MSE and correlation were used to determine the suitable forecast model. From the results of the analysis, Holt's Method with $\alpha=0.03$ and $\beta=1.0$ and the resulting correlation of 0.474 and MSE of 0.0722 seems to be the most reliable model in generating the forecast value of unemployment rate. The model forecasted 3.1 per cent for the first quarter of 2008 as compared to the actual rate of 3.6 per cent.

The advantage of using the Holt's Method Model is that recent observations are given relatively more weight in forecasting than the older observations. The one-step-ahead forecast is made in the fit region (in sample) and the multi-step-ahead forecasts are made in the holdout sample region (out of sample). However, in forecasting for longer term, it is less statistically fit than the one-step-ahead forecast because the more recent data may depart from the previous underlying process.

Conclusion

Based on the one step ahead forecast analysis, Holt's Method Model is the most suitable model for forecasting quarterly unemployment rate. Each model type has unique characteristic which fits to a particular data series. More forecasting techniques should be explored to ensure fitness to longer series of unemployment rate. Univariate Modelling Techniques are basically single variable models that use their past information as the basis to generate the forecast values. This is made on the assumption that the forecast values are dependent solely on the past pattern of the data series.

REFERENCES

- Anderson, D. R., Sweeney, D. J. & Williams, T. A. (2005). *Statistics For Business and Economics*. Thomson South-Western.
- Aripin, R. (2004). *Course Notes : A Guide To Data Management and Analysis*. Universiti Teknologi Mara.
- Baharudin, N. (2004). *Unemployed Graduates: Pre and Post 1997 Crisis*. Statistics Malaysia, Journal of the Department of Statistics, Malaysia vol. 1, 2004. Putrajaya.
- Bowerman, B. L., O'Connell, R. T. & Koehler, A. B. (2005). *Forecasting, Time Series, And Regression*. Thomson Brooks/Cole.
- Bozdogan, H. (1994). *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modelling: An Informational Approach Volume 1 Theory and Methodology of Time Series Analysis*. Kluwer Academic Publishers. Canada.
- Cryer, J. D. (1986). *Time Series Analysis*. Duxbury Press, Boston.
- Department of Statistics, Malaysia. (2007). Available from: www.statistics.gov.my
- Department of Statistics, Malaysia. (2005). *Yearbook of Statistics*. Putrajaya.
- Department of Statistics (2005). *Labour Force Survey Report 2004*. Putrajaya.
- Floros, C. (2005). *Forecasting the United Kingdom Unemployment Rate: Model Comparisons*. The International Journal of Applied Econometrics and Quantitative Studies. Vol. 2 - 4(2005): 57-72. [Online]
Available from: www.usc.es/economet/reviews/ijaeqs244.pdf
- Funke, M. (1992). *Time Series Forecasting of the German Unemployment Rate*. *Journal of Forecasting* Vol. 11 No. 2 January 1991;11(2): 111-125.
- Hussmans, R., Merhan, F., Verma, V. (1990). *Survey of economically active population, employment, unemployment and underemployment: an ILO manual on concepts and methods*, ILO Geneva.
- Jaafar, J. (2006). *Peramalan Indikator Utama Tenaga Buruh*, Research and Development Division, Department of Statistics Malaysia.
- Janacek, G. & Swift, L. (1993). *Time Series Forecasting, Simulation, Applications*. Ellis Horwood Limited.

- Keller, G. (2000). *Statistics For Management and Economics*. Duxbury.
- Lau, T. K. & Zainuddin, A. (2000). *Statistik Asas ITM*. Fajar Bakti Sdn. Bhd. Kuala Lumpur.
- Lazim, M. A. (2001). *Introductory Business Forecasting A Practical Approach*. Univision Press Sdn. Bhd.
- Lewis, R. & Brown, C. (2001). *A Forecast of North Carolina Unemployment Rates For 2001 – 2002 Using Past Inflation* [online]
Available from: paws.wcu.edu/mulligan/www/NCES2001lewis.html
- Makridakis, S., Wheelright, S. C. & Hyndman, R.J. (1998). *Forecasting Methods and Applications* (3rd ed). John Wiley & Sons Inc.
- Montgomery, A. L., Zarnowitz, V., Ruey S. T., Tiao, G. C. (1998). Forecasting the U.S. Unemployment Rate. *Journal of the American Statistical Association* Vol. 93 No. 6 Jun 1998; 442(6): 478-493.
- Sanders, D. H. & Smidt, R. K. (2000). *Statistics : A First Course, Sixth Edition*. McGraw-Hill Companies, Inc.
- The Daily. (2007). *Study : Chronic Unemployment* [online]
Available from: www.statcan.ca/Daily/English/050906/d050906a.htm