Pertanika 9(3), 359 – 367

Crude Palm Oil Price Forecasting: Box-Jenkins Approach

FATIMAH MOHD. ARSHAD and ROSLAN A. GHAFFAR*
Department of Agricultural Economics,
Faculty of Economics and Management,
University Pertanian Malaysia,
43400 Serdang, Selangor, Malaysia.

Key words: Crude palm oil price; univariate; identification; estimation; diagnosis.

ABSTRAK

Model univariate yang dicipta oleh Box-Jenkins telah digunakan untuk meramal harga bulanan minyak kelapa sawit mentah. Model yang telah dikenalpasti sesuai untuk ramalam adalah (0, 2, 1) (0, 1, 1)$_6$. Model ini menunjukkan bahawa siri data harga minyak kelapa sawit mentah iaitu tak pegun dan mengandungi unsur-unsur gandaan, menyarankan kewujudan proses purata bergerak Model ARIMA yang dikenalpasti menjadikan siri data kepada bercorak stokastik, membolehkan model ini meramal harga minyak kelapa mentah dalam jangka masa pendek.

ABSTRACT

A univariate ARIMA model developed by Box-Jenkins was utilised to forecast the short-run monthly price of crude palm oil. The appropriate model for forecasting was found to be (0, 2, 1) (0, 1, 1)$_6$. This model indicates that the original crude palm oil series is non-stationary and contains some elements of multiplicity, hence inheriting moving average process. The identified ARIMA model induced the data series into a stochastic one, making it a suitable model for forecasting crude palm oil prices in the short term.

INTRODUCTION

Forecasting consists basically of using data to predict future values for given variables to facilitate macro and micro level decision-making. In the case of Malaysia's crude palm oil, price forecasts represent valuable and fundamental information to direct and indirect traders in fats and oils market, and to financiers, producers and manufacturers and policy makers. Over 3 mn MT of palm kernel oil and palm oil of one form or another are traded in the world each year, with an f.o.b. value in excess of US$2.5 bn.

There are, of course, many ways to generate prediction, ranging in complexity and data requirements from intuitive judgements through time series analysis to econometric models[1]. The latter two approaches produce what Theil (1966) referred to as "scientific forecasts", in that it is formulated as a verifiable prediction from an explicitly stated method which can be reproduced[2]. A prime goal of forecasting studies is to assess the factors influencing supply and demand by developing estimates of coefficients and providing elasticity and flexibility of estimates. To achieve this goal, econometric models are used almost invariably.

---

*Department of Economics, Faculty of Economics and Management, UPM.

[1] Other techniques of forecasting (for agriculture) include informal models, indicators, balance sheet methods and surveys (see Freebairn, 1975).

[2] The terms prediction and forecast are used interchangeably.

A sophisticated subclass of linear time series models which have been known to have desirable forecast properties are those of the autoregressive integrated moving average (ARIMA) type, developed by Box-Jenkins (1970). To date, the Box-Jenkins technique has only been demonstrably successful in non-agricultural forecasting.

Several attempts have been made to compare the relative effectiveness of various forecasting techniques. For instance, Teigen (1973) compared the performance of econometric models, trend models, price difference models and the futures price to forecast US cattle prices using monthly and quarterly prices for the period of 1967 – 70. He concluded that a simple naive (no change) forecasting method provides better results than sophisticated methods. The analysis of Teigen indicated that projections of current cash prices or the corresponding futures price provided a good estimate of price to prevail in the forecast period. Gellatly (1979) evaluated the performance of several methods used to forecast New South Wales quarterly beef production, one quarter ahead. The forecasting procedures used were a single equation regression model, a Box-Jenkins univariate time series model, committee's judgement and a naive model. Although Gellatly's evaluation gave mixed results, it appears that the forecasting committee performed better than the other forecasting procedures. Nevertheless, the results also indicated that the committee's performance was little better then that of a naive (no change) model, suggesting there is room for improvement in general and as regards both the single-equation and Box-Jenkins models in particular.

Other attempts to evaluate forecasting techniques reveal the supperiority of the Box-Jenkins univariate time-series model for delivering accurate predictions. Helmer and Johansson (1977) compared Box-Jenkins' results with the established econometric models and cross-spectral analysis applied by other authors on the same set of data on advertising-sales relationships. The forecasting ability of Box-Jenkins transfer function models were proven to be far more accurate than econometric models and cross-spectral analysis. Bourke (1979) compared the forecasting accuracy of the Box-Jenkins and other econometric techniques for forecasting manufacturing-grade beef prices in the U.S.A. Criteria used to measure accuracy were Root Mean Squared Error, Theil's Inequality Coefficient and Turning Points. His findings suggested that Box-Jenkins models were in general marginally superior to the econometric method. Using a similar approach as Bourke, Brandt and Bessler (1981) found that the ARIMA model method performed substantially better than either the econometric or the expert opinion methods.

Thus, the application of the Box-Jenkins technique can be regarded as a fairly safe one for forecasting in comparison with other methods. With notable exceptions of Mohamad Napi (1982) and Mohamad Yusof Talib (1985), whose uses of the Box-Jenkins technique produced satisfactory results, little attempt has been made to utilise the technique to forecast the prices of Malaysian local commodity prices. In view of its proven superiority, this article therefore seeks to forecast crude palm oil prices by using the Box-Jenkins model.

THE GENERALISED MODEL

Details of the Box-Jenkins approach may be found in Box and Jenkins (1970) and critical discussions of it by Newbold (1975) and Geurts and Ibrahim (1975). A brief account of the generalised Box-Jenkins model is given below.

Generally speaking time series data can be categorised as stationary and non-stationary data which are mostly generated by a stochastic process. Stationary models are based on the assumption that the process remains in equilibrium about a constant mean level. Suppose we have a stationary series having mean and observations Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}..... are taken at equal intervals. We define a_t, a_{t-1}, a_{t-2}..... as "white noise" or random shocks to the system. Then there are two ways to model the series as an autoregressive (AR) model; and, as a moving average (MA) model.

Autoregressive processes of order p (ARIMA (p, 0, 0)) may be modelled by using p lagged observations of the series to predict the current observation, that is,

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \phi_2 \bar{Z}_{t-2}$$

$$+ \ldots\ldots\ldots \phi_p \bar{Z}_{t-p} + a_t \ldots\ldots\ldots 1 \qquad (1)$$

Which, using the backward shift operator, may be rewritten as,

$$\phi (B) \bar{Z}_t = a_t$$

where $\phi (B) = 1 - \phi_1 B - \phi_2 B^2$

$$\ldots\ldots -\phi_p B^p, \ldots\ldots 2$$

$a_t \sim NI (0, \sigma_a^2)$ and

$$\bar{\bar{Z}}_t = \bar{\bar{Z}}_t - \mu$$

Sometimes a current deviation from the mean period t is made linearly dependent on all prior deviations back to period (t – q). Therefore, the current deviation can be expressed as a linear function of the "white noise" to the system. The above process is called moving average process of order q (ARIMA (0, 0, q)) which can be expressed as:

$$\bar{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

$$- \ldots\ldots \theta_q a_{t-q} \ldots\ldots\ldots\ldots 3$$

or $\bar{Z} = \theta_q (B) a_t$, and

$$a_t \sim NI (0, \sigma_a^2)$$

Often the pattern of data may be described best by a mixed process of AR and MA elements. Almost all the stochastic or deterministic time series encountered in practice exhibit some degrees of non-stationarity which denote the existence of either AR or MA elements separately, or possibility some combination of them both. A preliminary model which incorporates stochastic and deterministic trend characteristics, non-seasonality and seasonality, is necessary before identifying a specific and relevant form of model which will describe the process. The appropriate preliminary model for an autoregressive integrated moving average model (ARIMA (p, d, q)) is:–

$$\phi_p (B) (1 - B)^d (1 - B^s)^D Z_t = \theta_p$$

$$+ \theta_q(B) a_t \ldots\ldots\ldots\ldots 4$$

where $\theta_p = (1 - \phi_1 - \phi_2 - \ldots\ldots\ldots \phi_p) \mu$

$$\phi_p (B) = 1 - \phi_1 B^2 - \ldots\ldots\ldots \phi_p B^p$$

$$\theta_q (B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots\ldots\ldots\ldots \theta_q B^p$$

where

d = amount of regular differencing

s = length of a season.

p = represents a deterministic trend constant. (Since most of the time series data are generated stochastically, the value of p is normally set equal to zero).

D = amount of seasonal differencing

Sometimes Z is in the form of logarithm or power transformation. This is to induce constant amplitude in the series over time so that the residuals from the fitted model will have a constant variance. By appropriately choosing certain levels of p, d, q and D, one can obtain a autoregressive model, a moving average model or some combination of them.

METHODOLOGY

Box-Jenkins models are built through an iterative identification/estimation/diagnosis strategy, a procedure which is repeated until a satisfactory model is obtained. The identification of a tentative model or set of models from the general class requires prior knowledge of the data pattern. Particular attention should be paid to the sources of nonstationarity which may be visible in the series plot. Nonstationarity due only to systematic trend (or drift) is easily detected by an inspection of the series of autocorrelation functions. However, nonstationarity associated with other causes (for example, variance nonstationarity) can be detected through an inspection of the time series plot. Generally, the empirical basis for identifying a model are the patterns of autocorrelation found in the autocorrelation and partial autocorrelation functions estimated from the series.

If data shows nonstationarity, differencing is needed to transform it to a stationary series. A stationary series is required for identification because the theoretical autocorrelation and partial autocorrelation of stationary series have distinct patterns for various models. In general, data should be differenced either regularly and/or seasonally to a degree that is just sufficient to induce stationarity.

For each differencing pattern specified by d, D and s, Box-Jenkins calculate sample autocorrelation function, r of lag k as:

$$r_k = \frac{\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^{n} (Z_t - \bar{Z})^2} \quad \ldots\ldots\ldots\ldots\ldots 5$$

and the theoretical autocorrelation function, ρ_k,

$$\rho_k = \frac{E(Z_t - \bar{\mu})(Z_{t-k} - \bar{\mu})}{\sigma Z^2} \quad \ldots\ldots\ldots\ldots\ldots 6$$

For the autotegressive model of order k for example, there exists, an autocorrelation function such that:

$$\rho_j = \phi_{k1}\, \rho_{j-1} + \ldots\, \phi_{k(k-1)}\, \rho_{j-1}$$

$$+ \phi_{kk}\, \rho_{j-k} \quad \ldots\ldots\ldots\ldots\ldots 7$$

$j = 1, 2, 3 \ldots\ldots\ldots k$

Where ϕ_{kk} is the last coefficient. This will lead to a Yule-Walker Equation which may be written in matrix form as:

$$P_k\, \phi_k = \rho_k$$

The quantity ϕ_{kk} regarded as a function of lag k is the partial autocorrelation function. The estimates of partial autocorrelations as shown by Quenouille (1949), of order p + 1 or higher are approximately independently distributed with variance: $\mathrm{var}(\phi_{kk}) = \frac{1}{n}\,; k \leqslant p + 1$

when n = number of observations.

$$S.E(\phi_{kk}) = \frac{1}{\sqrt{n}}$$

As a general rule when the autocorrelations drop off exponentially to zero as k increases, this implies an AR model whose order is determined by the number of partial autocorrelation which are significantly different from zero. If the partial autocorrelations drop off exponentially to zero as k increases, the model is MA, and its order is determined by the number of statistically significant autocorrelations. When both autocorrelations and partial autocorrelations drop off exponentially to zero, the model is ARIMA.

Having tentatively identified an ARIMA (p, d, q) model for the time series, ϕ_p and θ_q can be estimated by minimising the sum of squares residual from $(E(Z_t - \bar{Z}_t)^2 = \Sigma a_t^2)$; where $\bar{Z}$ is the estimated value of period t. This is done by using non-linear least square estimation.

The final stage for model determination is the diagnostical checking. A statistically adequate model is defined as one whose residuals are distributed as white noise, i.e., normally and independently distributed with mean zero and constant variance, σ_a^2(a NI $(0, \sigma_a^2)$). If the tentative model is not statistically adequate by this criterion, the situation has to be remedied by reexamining the autocorrelations, identifying a better model and repeating the whole process.

RESULTS

Box-Jenkins technique was applied to a set of 132 of monthly cash price of crude palm oil for the years 1974 – 1984[1] The pattern of original data is shown in *Figure 1*.

The plot exhibits a slow upward tendency and a periodic component consisting of irregular seasonal pattern. Besides, the series values also show an irregular peak in the early part of 1984 which was due to excessive speculation in palm oil futures trading in the Kuala Lumpur Commodity Exchange (Fatimah M.A., 1985). Aggressive speculative buying and desperate long-positions taken by hedgers to cover unfulfilled forward contracts resulted in a "market-squeeze" situation — which pushed the near futures and cash prices to an all-time high[2]. Hence, the series is nonstationary and heteroscedastic in nature which is reduced through log transformation.

Model Identification

To identify a tentative model requires the examination of the patterns of autocorrelation found in the autocorrelations and partial autocorrelation functions estimated from the series. Patterns of autocorrelation observed in the data are then compared with the pattern expected of various ARIMA models. For instance an ARIMA (1, 0, 0) process, autocorrelation function is expected to show a 'perfect' exponential decay and to have a spike at partial autocorrelation functions.

The values of the partial autocorrelation parameters of the series are initially large and their magnitude decreases as the time lag increases. They do not seem to have a cut off after p time lags, instead they continue and slowly trail off to zero. In other words, there is an exponential decrease in partial autocorrelation from large to smaller values as the time lags of autocorrelation strengthen (*Figures 2* and *3*).

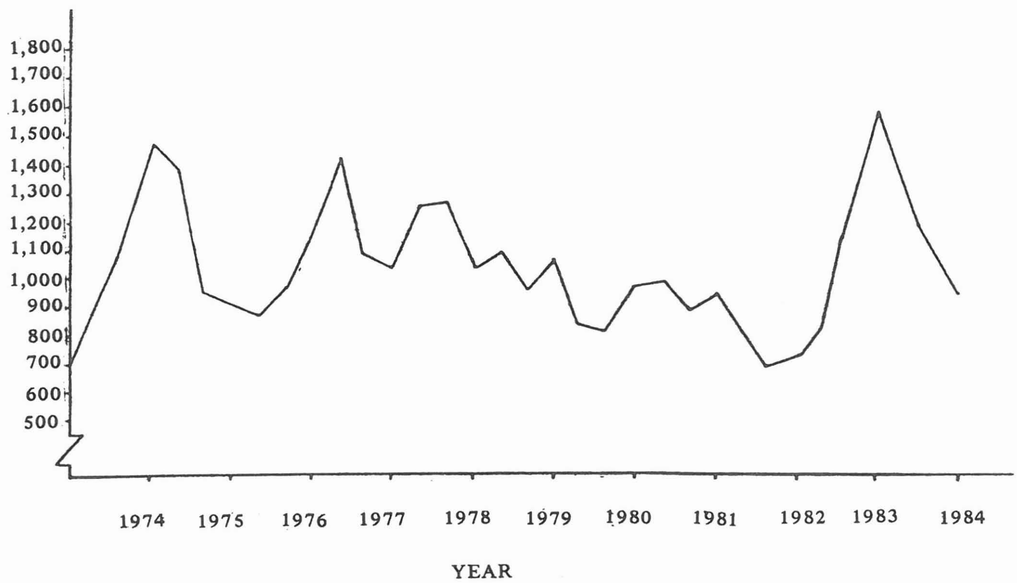

Fig. 1: Crude palm oil prices, 1974 – 1984

[1] These prices are local delivered net prices. Data was collected from the Department of Statistics, Ministry of Agriculture.
[2] The market was later "cornered" singlehandedly by one speculator through a heavy short-selling tactic which brought the palm oil futures market to a temporary halt.

Autocorrelation function for variable price
autocorrelations ★
Two standard error limits •

LAG	AUTO. CORR.	STAND. ERR.
1	-.566	.086
2	.066	.086
3	.021	.086
4	.063	.085
5	.115	.085
6	-.266	.085
7	.157	.084
8	-.035	.084
9	.031	.084
10	-.027	.083
11	-.024	.083
12	.056	.082
13	-.056	.082
14	.085	.082
15	-.105	.081
16	.044	.081
17	.041	.081
18	-.053	.080
19	.053	.080
20	-.106	.080
21	.147	.079
22	-.130	.079
23	.036	.078
24	.071	.078
25	-.084	.078
26	.036	.077
27	-.021	.077
28	.052	.077
29	-.029	.079
30	-.087	.076
31	.130	.075
32	-.041	.075
33	-.007	.075
34	-.035	.074
35	.043	.074
36	.047	.073

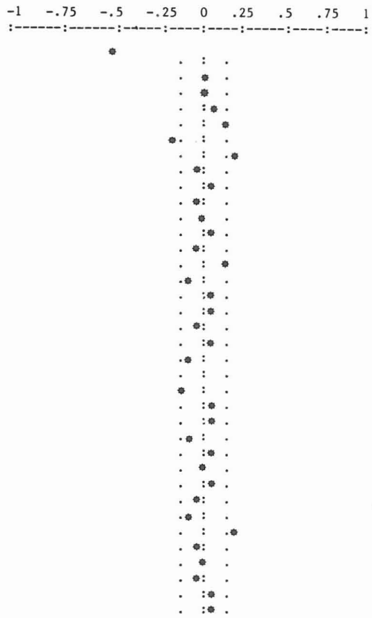

Fig. 2: Autocorrelation function for crude palm oil prices

Partial autocorrelation function for variable price
partial autocorrelations ★
Two standard error limits •

LAG	PR-AUT CORR.	STAND ERR.
1	-.566	.088
2	-.462	.088
3	-.439	.088
4	-.367	.088
5	.035	.088
6	-.063	.088
7	-.002	.088
8	-.072	.088
9	-.094	.088
10	-.080	.088
11	-.032	.088
12	.003	.088
13	-.027	.088
14	.086	.088
15	.019	.088
16	-.025	.088
17	.026	.088
18	.009	.088
19	.095	.088
20	.010	.088
21	.096	.088
22	-.040	.088
23	-.117	.088
24	-.011	.088
25	.021	.088
26	.016	.088
27	.056	.088
28	.041	.088
29	.051	.088
30	-.124	.088
31	-.105	.088
32	-.048	.088
33	.026	.088
34	.061	.088
35	.052	.088
36	.100	.088

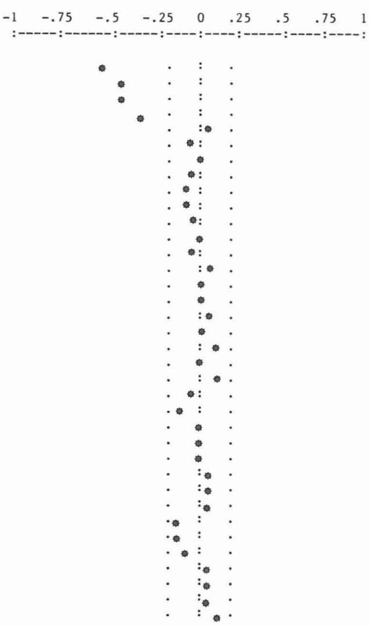

Fig. 3: Partial autocorrelation function for crude palm oil prices

These behaviours of autocorrelation and partial autocorrelation suggest an applicability of MA (1) model (Makridakis, 1978). A periodicity of 6 is suggested by the significant autocorrelation values at lag 6. In addition, spikes are observed at the first and second Q seasonal lags of the autocorrelation and decaying at the successive lags towards zero which suggests MA $(2)_6$ process (Mc Cleary, 1980). Hence, the mathematical model for (0, 1, 1) $(0, 2, 2)_6$ model is: –

$$(1 - B)^2 (1 - B^6) Z_t = (1 - \phi B) (1 - \phi_6 B^6 - \phi_{12} B^{12} \quad \ldots($$

After estimating its parameter, the model is subjected to diagnostic checking to ensure its appropriateness. The key to this diagnosis is the residual autocorrelation function; that is, the residual series should be random (white noise) and thus uncorrelated for all lags. For the crude palm oil price series, the residual autocorrelation plot indicates that this criteria has been achieved (*Figure 4*).

Residual autocorrelation function for variable price
Autocorrelation ★
Two standard error limits •

LAG	AUTO. CORR.	STAND. ERR.
1	.116	.084
2	-.002	.084
3	-.017	.083
4	.087	.083
5	.079	.083
6	.038	.082
7	-.056	.082
8	-.118	.082
9	-.084	.081
10	-.057	.081
11	-.046	.081
12	-.055	.080
13	-.055	.080
14	-.132	.080
15	-.144	.079
16	-.057	.079
17	-.047	.079
18	-.055	.078
19	-.049	.078
20	-.080	.078
21	.005	.077
22	-.012	.077
23	.029	.077
24	-.021	.076
25	.022	.076
26	.020	.076
27	.033	.075
28	.038	.075
29	-.002	.075
30	-.029	.074
31	.159	.074
32	.127	.074
33	.055	.073
34	.002	.073
35	.023	.072
36	-.044	.072

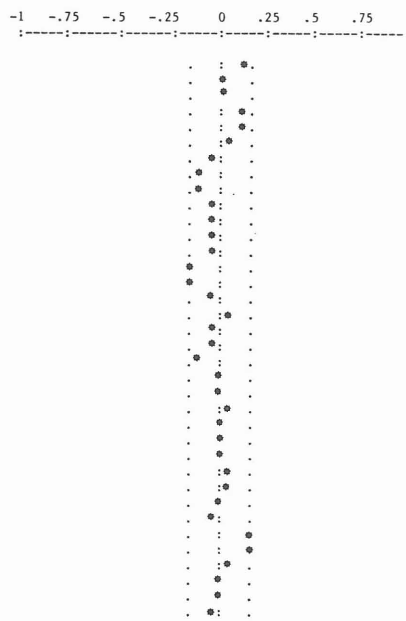

Fig. 4: Residual autocorrelation functions for crude palm oil prices

Estimation of Parameters

Having identified an MA (0, 1, 1) $(0, 2, 2)_6$ model for the crude palm oil series, the parameters of the model were estimated. The estimated parameters were found to be statistically significant. Thus, the resulting model is as follows: —

$$(1-B)^2(1-B^6)Z_t = (1-0.7757^B)(1-0.9084B^6 + 0.2906B^{12})\,a_t \quad \ldots\ldots\ldots 10$$

In the light of encouraging diagnosis, especially the residual autocorrelation function and the correlation matrix of estimated parameters, the model is deemed adequate for forecasting. The estimated monthly price for the next seven months in 1985 are shown in Table 1, together with the confidence intervals, actual values and percentage errors. The graphic displays for forecasting at one time lead are presented in *Figure 5*. As shown in the table, the average percentage error forecasting for crude palm oil for the seven months of 1985 are 9.7%. Besides, graphically, the forecast points follow the original data very closely within 95% confidence limits *(Figure 5)*.

Diagnosis

Having identified a tentative MA (0, 1, 1) (0, 2, 2) model, and having satisfactorily estimated its parameters, the model further needs diagnosis checking to see whether improvements can be attainable. A statistically adequate model is defined as one whose residuals are distributed as white noise, i.e., they are independent, and normally distributed with mean zero and constant variance and hence are uncorrelated for all lags. The residual series of autocorrelation for crude palm oil prices is shown in *Figure 4*. As indicated in the figure, the residuals stay within the confidence interval; and there are no significant spikes at low lags. Besides, the chi-quare diagnosis indicated that the residual series of autocorrelation is independent and normally distributed (Table 2).

TABLE 1
Forecasts for crude palm oil prices, origin at December 1984 and 95% confidence limits

Month	Lower confidence limit	Forecast	Upper confidence limit	Actual values	Percentage of errors
January	1008.1	1278.6	1621.6	1212	5.4
February	936.5	1363.6	1985.5	1194	14.2
March	764.1	1272.1	2117.7	1288	1.2
April	644.9	1231.0	2349.8	1513	23.3
May	591.2	1300.6	2860.7	1415	8.1
June	484.4	1235.0	3148.4	1239	0.3
July	391.6	1177.6	3540.6	1020	15.4
August	338.9	1209.6	4316.7	847	
September	261.8	1116.7	4761.9	735	
October	215.6	1105.7	5670.1	735	
November	190.8	1184.6	7351.6		
December	116.3	1113.1	10645.0		
Average					9.7

Graphic display of forecasts for variable price

Definitions of symbols
Data – ★
Forecasts at lead 1 – +
Estimated 95% confidence limits – •
Forecast function – o
Overlap – *

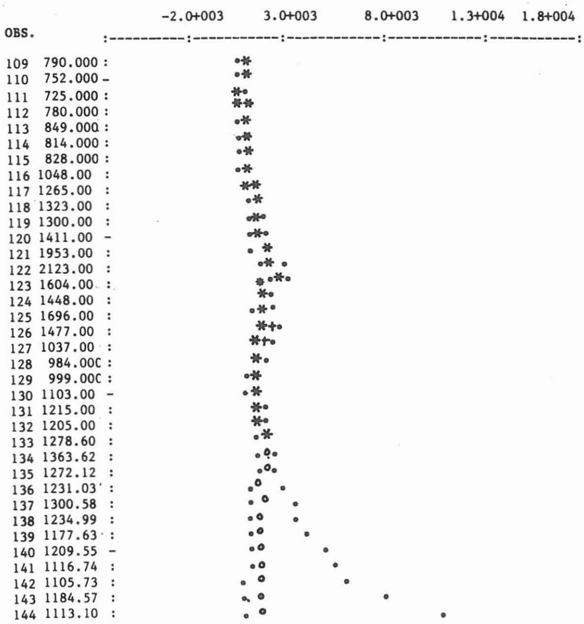

Fig. 5: Forecasted values of crude palm oil prices.

CONCLUSION

Despite its proclaimed superiority in forecasting, the Box-Jenkins model is limited to short-term predictions. As shown in Table 1, the 95% confidence limits become wider as the time lead increases; that is the forecast error variance increases simultaneously with the time lead. Thus to retain accuracy, there is a need to update the parameter estimates through incorporting data as available. Alternatively, accuracy could be maintained by updating the forecast to a new origin (t + 1) whenever new information is available.

For a long term forecast, however, a non-stochastic model which is more adequate and more comprehensive than Box-Jenkins is needed. For instance, an econometric model incorporating those variables such as global palm oil stocks, supply forecasts and prices and corresponding data for soyabean oil and competing fats and oils which have direct bearing on palm oil prices both for supply and demand aspects, would be an appropriate one for forecasting.

ACKNOWLEDGEMENTS

The authors record their thanks to Mohd. Napi Daud of the Rubber Research Institute Malaysia for his valuable advice and Zawazi Jawahir for data processing assistance.

REFERENCES

BOURKE, I.J. (1979): Comparing the Box-Jenkins and Econometric Techniques for Forecasting Beef Prices, *Rev. of Mktg. and Agric. Economics,* 47(2).

BOX, G.E.O. and G.M. JENKINS, (1976): Time Series Analysis, Revised Edition, Holden-Day San Francisco.

TABLE 2
Diagnostic chi-square statistics for residual series

Lag	Chi-square	Degrees of freedom	Probability
6	4.12	3	0.1333
12	8.68	9	0.2254
18	16.43	15	0.4957
24	18.12	21	0.7804
30	18.84	27	0.9053
36	27.29	33	0.8258

BRANDT, J.A. and D.A. BESSLER, (1981): Composite Forecasting: An Application with U.S. Hog Prices, *Am. J. of Agric. Econ.*, **63**.

FATIMAH, M.A. (1985): The Impact of Futures Trading on the Crude Palm Oil Cash Price Variability, a working paper presented in PETA Conference, Current Issues in Agriculture, 14 – 15 May.

FREEBAIRN, J.W.(1975): Forecasting for Australian Agriculture, *Aust. J. of Agric. Econ.*

GELLATLY, C. Forecasting N.S.W. Beef Production: An Evaluation of Alternative Techniques, *Rev. of Mktg. and Agric. Econ.*, **47**(2).

GRANGER, C.W.J. and P. NEWBOLD, (1977): Forecasting Economic Time Series, Academic Press, New York.

GUERTS, M.D. and I.B. IBRAHIM, (1975): Comparing The Box-Jenkins Approach with the Exponentially Smoothed Forecasting Model with an Application to Hawaii Tourists, *J. of Mktg. Res.*, **12**(May).

HELMER, R.M. and J.K. JOHANSSON, (1977): An Exposition of the Box-Jenkins Transfer Function Analysis with an Application to the Advertising – Sales Relationship, *J. of Mktg. Res.*, **14**(May).

MAKRIDAKIS, S. and S. WHEELRIGHT, (1978): Forecasting methods and Applications, New York, John Wiley and Sons, 1978.

MC CLEARY, R. and R.A. HAY, (1980): Applied Time Series for the Social Sciences, Sage Publications, Beverly Hills.

MOHAMAD NAPI DAUD. A Forecasting Methodology as Applied to Rubber *J. of Rubb. Res. Inst. Malaysia*, **31**(3).

MOHAMMAD YUSOF TALIB. (1985): Stohastic Model for Malaysia's Palm Oil Production, PORIM Bulletin, May.

QUENOUILLE, M.H. (1949): Approximate Test of Correlation in Time Series, *J. of R. Statist. Soc.* **B 11** **(68)**.

TEIGEN, L. (1979): Costs, Loss and Forecasting Error: An Evaluation of Models for Beef Prices, Unpublished Ph.D. thesis, Michigan State University, 1973, quoted in Bourke.

THEIL, H. (1965): Applied Economic Forecasting, Amsterdam: North-Holland Publishing Co.

(Received 13 February, 1986)