Ife Journal of Agriculture, Vol. 19. Nos 1 & 2 1997

COMPOSITE FORECASTING METHODS: AN APPLICATION TO NIGERIA'S PALM PRODUCE PRICES

Tijani A. A. and O. Ajobo
Department of Agricultural Economics
Obafemi Awolowo University, Ile-Ife.

Abstract

This paper uses combined time series and regression analysis rather than either method alone to produce better quality forecast of palm kernel prices.

A regression model was selected for the price data whille a time series model was constructed for the regression residual series.

The result of the regression analysis lends credence to the claim that statistical eleganc of an econometric model is not directly related to its ability to forecast well. Values of the root -mean -square (RMS) simulation error, RMS forecast error and the Theil coefficient (U_2) of the linear model were 43.19, 52.76 and 1.05 respectively whilst those of the double log models (having better fit) were 43.84, 61.08 and 1.21 respectively.

The combined regression- time series model outperformed the regression model. Values of the RMS simulation error, RMS forecast and U_2 of the combined model were 85.27, 44.22 and 0.878 respectively, whilst those of the regression model were 93.73, 52.76 and 1.05 respectively.

INTRODUCTION

Palm produce is of significant importance in the economies of several states in southern Nigeria; however, its contribution to Nigeria's international trade profile has declined in the past three decades.

Since the end of Nigeria's civil war in 1970, the Federal and State Governments of the oil palm belt have been executing projects aimed at sustaining self-sufficiency (and possible surplus for export) in palm produce. Despite the long existence of Nigerian Institute for Oil Palm Research (NIFOR) the problem (decline in production of palm produce) still remains largely unsolved. The reason for this is not unconnected with deficient planning processes (Falcon (1984), Helms (1985), Ojo (1989).

Developments during 1986 - 1996 period have clearly shown that agricultural development in Nigeria would require greater attention to effective planning and policy implementation to sustain the momentum created by the structural adjustment programme of the economy. As a matter of priority, there is need to rationalize the agricultural planning process; strengthen the institutional base for policy implementation and design a sound information base which is required for sound planning and plan evaluation (Ojo, 1989).

Planning involves making decisions which will have their effects in the future and so an estimate of this future is required. This assessment of the future is termed forecasting and is a vital ingredient in any planning process (Firth, 1977).

Most studies on economic forecasting have been handled by analysis based on single econometric models alone (especially in the developing countries such as Nigeria). The end result has invariably been poor quality forecast due mainly to lack of sufficient attention to time series concepts (Steckler (1968), Fair, (1970, 1974), Granger and Newbold (1977).

Granger and Newbold (1974, 1977) criticise econometric models for their inability to give sufficient attention to the specification of lags and error structures even though they can adequately represent the structure of economic relationship in a functional sense. They observed that inadequate attention to time series concepts would lead to uncecessarily poor forecasts. They further contended that economic forecasting methods should incorporate as far as possible the benefits of economic theory, and should equally give due consideration to the problem of time series behaviour, if the hope of producing reasonably accurate forecasts is to materialize. Consequently, they suggest employing time series data to indicate appropriate model for the residuals from the general autoregressive integrated moving average class which could be incorporated into the regression for estimation and forecasts. By so doing, the problem of autocorrelation (in the residuals from the regression equation) that is frequently encountered when dealing with time series data will be minimized.

The objective of this paper therefore is to use autoregressive integrated moving average (ARIMA) method together with the conventional regression method to illustrate the possibility of generating a better quality forecast (of Nigeria's palm produce prices) when account of time series behaviour is taken than would be possible through the use of regression (econometric) method alone. Specifically, this paper (i) evaluates the performance of the regression and the time-series models, using historical simulation of palm kernel price data (1973 - 1987) and (ii) forecast future value of the price series using regression and combined regression-time series models respectively.

MATERIALS AND METHOD

Data Source

Data used in this paper are

(i) Monthly world price data of palm oil and palm kernel (In Naira per tonne) for the period (1973-1987).

(ii) Monthly rate of exchange of dollar/Naira for the period (1973-1987)

These data were obtained from the Central Bank of Nigeria and Nigeria's Federal office of Statistics.

REGRESSION ANALYSIS

The Price Model

The world price (in Naira) for palm kernel (P_{1t}) at time t is hypothesized to be a function of the monthly time trend (T), the exchange rate at time t (EX_t) world price of palm kernel at time t-n (P_{1t-n}), the world price of its joint product (palm oil) at time t-m (P_{2t-m}) and a disturbance term (U_t).

$P_{1t} = f(T, EX_t, P_{1t-n}, P_{2t-m}, U_t)$..(1)

Three structural forms were investigated for the palm kernel price series:

$P_{1t} = f(T, U_t)$..(2)

$P_{1t} = f(T, EX_t, U_t)$..(3)

$P_{1t} = f(T, P_{1t-n}, EX_t, U_t)$..(4)

Under (2), four mathematical forms were examined

$P_{1t} = b_0 + b_1 T + U_t$..(Linear)

$\text{Log}_e P_{1t} = \log_e b_o + b_1 \log_e T + \log_e U_t$..(Double log).

$\text{Log}_e P_{1t} = b_o + b_1 T + U_t$..(Semi - log)

Under (3), three mathematical forms were investigated:

$P_{1t} = b_0 + b_1 T + b_2 EX_t + U_t$..(Linear).......(5)

$\text{Log}_e P_{1t} = \log_e b_o + b_1 \log_e T + b_2 \log_e EX_t + U_t$..........(6)

A total of 12 mathematical forms were investigated under (4).

These include 6 linear functions summarized as:

$P_{1t} = b_o + b_1 P_{1t-n} + b_2 P_{2t-m} + b_3 EX_t + U_t$..........(7)

and 6 double log functions summarized as

$\text{Log}_e P_{1t} = \log_e b_0 + b_1 \log_e P_{1t-n} + b_2 \log_e P_{2t-m} + b_3 \log_e EX_t + \log_e U_t$..................(8)

where

$m = 1, 2; n = 1, 2, 3$

$t = 1, 2, ...n$ (i.e. total number of observations)

The disturbance term in the equations above was assumed to be normally distributed with zero mean and constant variance.

Above regression equations were estimated over the sample period (January 1973 - April 1987) using ordinary least squares method.

Evaluatioin of the Estimated Equations

The essence of the evaluation is to indicate the most appropriate among the estimated models. Some evaluative criteria may produce a consistent ranking of results but no single criterion can adequately portray all the qualities of a model. For this reason, an assortment of certain statistics were used to select the best model for the price series. First, explanatory power (R^2), correctness of coefficient signs, statistical significance of the regression models and low standard error of estimate (SEE) were used to select two leading equations. The mathematical forms of the candidate equations are:

$$P_{1t} = a_0 + a_1T + a_2P_{1t-1} + a_3P_{2t-1} + a_4EX_t + U_t \quad \text{.......(9)}$$

$$Log_e P_{1t} = log_e a_0 + a_1 log_e P_{1t-1} + a_2 log_e EX_t + log_e U_t \quad \text{............(10)}$$

Next, root-mean-square (RMS) of simulation error and forecast error together with estimated Theil coefficient (U_2) were used to select the best regression model for the price series. Then the residual series from this final regression model was subjected to time series analysis. The residual is given by

$$U_t = A_t - P_t \quad \text{.............................(11)}$$

Where

A_t = actual price at time t

P_t = estimated price at time t

<u>Theil Coefficient</u> (U_2)

This is given by

$$U_2 = \sqrt{\sum (P_t - A_t)^2} \Big/ \sqrt{\sum A_t^2} \quad \text{....................(12)}$$

Where P's and A's are defined as changes in predictive and actual values, respectively.

Predicted change $(\Delta P_t) = P_t - P_{t-1}$

Actual change $(\Delta A_t) = A_t - A_{t-1}$

Theil (1966), Leuthold (1975) and Granger and Newbold (1977) argue in support of U_2 over U_1*. For example, Theil

(1966) states two reasons why one might prefer U_2 to U_1. "First, it is related more directly to the concept of the failure of the forecast............Second the alternative coefficient (U_1) depends on the forecasts and it is therefore not true that the coefficient is uniquely determined by the mean square predictive error (given the data on realizations). This is against the idea of a quadratic loss criteria".

$$U_1 = \frac{\sqrt{\frac{1}{N}\sum (P_i - A_i)^2}}{\sqrt{\frac{1}{N}\sum A_i^2} + \sqrt{\frac{1}{N}\sum P_i^2}} \quad \text{.............................(13)}$$

Leuthold (1975) advocates the use of U_2 (rather than U_1) as a means of comparing forecasting accuracy among models, because of its flexibility, relative appropriateness and the ease of understanding as well as interpretation.

In a similar vein, Granger and Newbold (1977) note that the inclusion of $\sqrt{1/N \sum P_i^2}$ in the denominator of U_1 gives it the potential to produce misleading results.

The statistic U_2 exhibits certain characteristics:

i. It takes the value of zero if the forecasts are perfect;
ii. It has no upper bound
iii. It has a value of 1 when the prediction is the naive no-change extrapolation; and
iv. It takes on values greater 1 for models less accurate than naive no-change extrapolations.

RMS (Root-Mean-Square) Simulation Error

RMS simulation error is a quantitative measure designed to test the performance of a model by examining how closely an historical simulation tracks its corresponding historical data series. In other words, rms error is a measure of deviation of the simulated variable from its time path.

$$\text{rms error} = \sqrt{1/N \sum_{t=1}^{N} (P_{st} - P_{at})^2} \quad \ldots\ldots\ldots\ldots 13$$

where P_{st} = simulated value of P_t

P_{at} = actual value of P_t

N = number of available observations

P_t = palm kernel prices.

The model that has smaller rms stimulation error is chosen. Since low rms simulation errors are only one desirable measure of simulation fit, a plot of the simulated series and the actual series is used to determine how well each model simulates turning points in the historical data. (Pindyck and Rubinfeld, 1981).

Out - Of - Sample Forecast Evaluation

In the *ex. post* forecast, the forecast results are compared with available data. The rms forecast error, i.e., the rms simulation error computed over the forecast range provides a measure of the ability of each model to forecast well. The smaller the value, the better the forecast quality is considered to be.

TIME SERIES ANALYSIS

Autoregressive Integrated Moving Average (ARIMA) Model: Theoretical Considerations

ARIMA model can be purely autoregressive or purely moving average or a mixture of both. In a moving average model, the process U_t is described completely by a weighted sum of current and lagged disturbances.

A moving average process of order q may be expressed as:

$$U_t = E_t + b_1E_{t-1} + b_2E_{t-2} + \ldots\ldots + b_qE_{t-q} \ldots\ldots (14)$$

or briefly

$$U_t = b_q(B)\,E_t \ldots\ldots (15)$$

In the autoregressive model, U_t depends on a weighted sum of its past value and random disturbance term. An autoregressive process of order P may be expressed as:

$$U_t = a_1U_{t-1} + a_1U_{t-2} + \ldots\ldots + a_pU_{t-p} + E_t \ldots\ldots (16)$$

or briefly

$$a_p(B)\,U_t = E_t \ldots\ldots (17)$$

In a mixed autoregressive-moving average model the process U_t is a function of both lagged random disturbances and its past values, as well as current disturbance term.

A general ARIMA process is given as

$$a(B)\,(1-B)^d\,(U_{t-m}) = b(B)E_t \ldots\ldots (18)$$

Where

$$a(B) = (1 - a_1B\ a_2B^2 \ldots\ldots a_pB^p)$$

$$b(B) = (1 + b_1B + b_2B^2 + \ldots\ldots b_qB^q)$$

M = Means of the series Ut

B = Lag operator (BUt = Ut-1)

$$(1-B)\,U_t = U_t - U_{t-1}$$

and E_t = random error

The interger d is the degree of differencing required to produce a stationary process U_t. It is usually of low order 0, 1 or 2.

After the removal of any deterministic component including a non-zero mean, and/or the application of some suitable transformation to the data, the model can be rewritten thus

$$A(B)\,(1-B)^d\,U_t = b(B)\,E_t \ldots\ldots (19)$$

Furthermore, the seasonality model can be written as:

$a(B)a_s(B^s)(1\text{-}B)^d(1\text{-}B^s)^D U_t = b(B)b_s(B^s)E_t$............ (20)

where

$a_s(B) = 1\text{-}a_1B\text{-}a_2B^2 .... a_pB^p$

$a_s(B^s) = 1\text{-}a_{1,s}B^s - a_{2,s}B^{2s} .... ap_{,s}B^{PS}$

$b(B) = (1 + b_{1,s}B + b_{2,s}B^2 + ... b_{q,s}B^q$

$b_s(B^s) = (1 + b_{1,s}B^s + b_{2,s}B^{2s} + ... b_{q,s}B^{Qs}$

S is the seasonal order; and D is the degree of seasonal differencing.

More briefly, the model can be described as being of the form (p, d, q) x (P, D, Q)s where

P = number of regular autoregressive parameters

q = number of regular moving averages

P = number of seasonal autoregressive parameters

Q = number of sesonal moving average parameters.

The distinguishing characteristics of ARIMA model include the following.

i. Specification of the model can only be achieved by examining the acutal data. This is unlike econometric modeling that is based on the assumption that the structure of the model can be identified a priori; and

ii. The building process is a systematic process involving idenitification, estimation and diagnostic checking. This system process is discussed as follows:

Identification

This stage involves the selection of value for p, d and q for a non-seasonal model together with P, D and Q for a seasonal model. (Box and Jenkins, 1970; Granger and Newbold 1977).

The two most useful tools in any attempt at model indentification are sample autocorrelation function and sample partial autocorrelation function.*

This is accomplished by estimating the sample autocorrelation function of the series using expression (21) and estimating sample partial autocorrelations using expression (22)

$$r_k = \frac{C_k}{C_o} \text{..........................(21)}$$

where $$C_k = \frac{1}{n}\sum_{t=1}^{n-k}(U_t - \bar{U})(U_{t+k} - \bar{U})$$

* See Box and Jenkins (1970) for the relation between the theoretical autocorrelation and partial autocorrelation functions.

$$C_o = \frac{1}{n}\sum_{t=1}^{n}\left(U_t - \overline{U}\right)^2$$

$$\overline{U} = \frac{1}{n}\sum_{t=1}^{n} U_t$$

n = number of available observations

K = 0, 1, 2, ... K - (K is 20 in this study).

$$\hat{a}_{kk} = Y_1 \qquad K = 1$$

$$\hat{a}_{kk} = \frac{Y_k - \sum_{j=1}^{k-1} a_{k-1,j} Y_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{a}_{k-1,j} Y_j} \quad ....22$$

K = 2, 3, ... 20

where

$$a_{kj} = a_{k-1,j} - a_{kk} a_{k-1,k-j}$$

J = 1, 2, ... 19

(Box and Jenkins, 1970).

The failure of the sample autocorrelation to dampen out at high lags is an indication of the need for differencing.

The identification procedure can result in two or more distinctly different models which give reasonable representation of the empirical series (Box and Jenkins, 1970). Each of the models is considered a candidate for further analysis of parameter estimation and diagnostic checking.

ESTIMATION

Having selected a particular model the next step is to obtain the estimates.

a_i......................a_p; b_i..................b_q.

For a purely autogressive model the estimation process is essentially a linear regression i.e.,

$a(B)U_t = E_t$ or $U_t = a_i U_{t-1} + ... + a_p U_{t-p} + E_t$ .....: (23)

Estimation process becomes more difficult if the model contains a moving average component as well. In this case, the model is given as

$b^{-1}(B)a(B)U_t = E_t$ .... (24)

This can be accomplished by using iterative or non-linear least squares (Marquant algorithm) method.

Diagonostic Checking

This is carried out to determine whether or not the chosen model adequately represents the given set of data. Any inadequacy revealed may suggest an alternative model specification.

When a model is correctly specified and a's and b's are close to their "true values", the residuals Et (which are estimates of the unobservable error terms), should resemble a white noise process. Should residuals show significant autocorrelations, the model is considered inadequate. The sample autocorrelation function of the residuals ($r_k(E_t)$) is calculated by

$$\hat{Y}_k(E_t) = \frac{\sum_t \hat{E}_t \hat{E}_{t-1}}{\sum \hat{E}_t^2} \quad \text{..............................(25)}$$

k = 1 20

Examination of the residual autocorrelations (from time series model) to diagnose whether the autocorrelations arise as a result of white noise only are often executed by

(i) Comparing the sample autocorrelations of the residual with bounds $\pm 2n^{-\frac{1}{2}}$ which will provide a general indication of possible departure from 'white noise' behaviour in E_t.

(ii) Comparing the statistics

$$Q = n\sum_{k-1}^{m} \hat{Y}_k^2 \quad \text{................................ (26)}$$

with tabulated values of chi-squared statistics for m-p-q degrees of freedom. The hypothesis of white noise behaviour in the residuals is rejected at high values of Q(m being moderately large generally at least 20).

Table 1: Sample Autocorrelations and Partial Autocorrelations for data on World Prices of Palm Kernel

	K	1	2	3	4	5	6	7	8	9	10
	A	-0.7	-0.4	0.37	-0.07	0.00	0.11	0.01	-0.05	0.04	-0.02
	P	-0.07	-0.05	0.37	-0.02	0.02	-0.03	-0.05	-0.06	-0.01	-0.04
Ut											
	K	11	12	13	14	15	16	17	18	19	20
	A	0.00	-0.04	-0.03	-0.02	-0.11	-0.11	0.05	-0.11	-0.01	0.08
	P	0.04	-0.07	-0.02	-0.05	-0.09	-0.13	0.07	-0.05	0.11	0.04
$(1-B)U_t$	K	1	2	3	4	5	6	7	8	9	10
	A	-0.51	-0.17	0.38	-0.22	-0.02	0.08	0.00	-0.06	0.06	-0.03
	P	-0.51	-0.59	-0.13	-0.15	-0.09	-0.19	-0.07	0.12	-0.02	-0.10
	K	11	12	13	14	16	17	18	19	20	
		0.02	-0.02	0.00	0.05	-0.05	-0.07	0.15	-0.13	0.02	0.17
		0.02	-0.04	0.00	0.04	0.07	-0.13	0.01	-0.16	-0.07	0.09
	K	1	2	3	4	5	6	7	8	9	10
	A	-0.02	0.01	-0.31	-0.02	0.08	-0.17	0.03	-0.08	0.05	-0.03
	P	-0.02	0.01	-0.31	-0.04	0.09	-0.30	0.01	-0.03	-0.14	-0.01
$(1-B^3)U_t$	K	11	12	13	14	15	16	17	18	19	20
	A	0.04	-0.01	0.05	-0.03	-0.05	-0.16	0.03	0.07	0.08	0.07
	P	0.05	-0.12	0.07	-0.09	-0.13	-0.16	-0.02	-0.03	-0.01	0.05

Source: Data analysis

ARIMA Method

Table 1 shows the sample autocorrelations (A) and partial autocorrelations (P) for palm kernel regression residual series and its first and 3-month difference respectively. The sample autocorrelations of the undifferenced series dampened out at high lags, suggesting stationary behaviour. Some evidence of seasonality is indicated by relatively high values at lags 3 and 6, (though less than overwhelming), thus the operator $1\text{-}B^3$ was employed. The sample autocorrelations of the 3-month difference exhibit stationary behaviour, alternate in sign and show some significant values (compared with $\pm 2/\sqrt{162} = \pm .16$) at lag 3 and 6 respectively thus suggesting an autoregressive model. For this reason, the model'

$$(1 - a_{1,3}B - a_{2,3}B^2)(1\text{-}B^3)U_t = E_t \text{ .... } (27)$$

was fitted to the data. Since the model is essentially a linear regression, it was estimated using ordinary least squares method. Next, the residual series was computed and its autocorrelations estimated. Test of the sample autocorrelations for significance revealed that the model is adequate, i.e., none of the residual sample autocorrelations lies outside the range $+2/\sqrt{162}$ ($\pm.16$), and so these autocorrelations provide no basis to question the fitted model. To further establish the adequacy of the model, an augmented model:

$$(1 - a_{1,3}B - a_{2,3}B^2 - a_{3,3}B^3)(1\text{-}B^3)U_t = E_t \text{ ...... } (28)$$

was fitted to the data and also estimated using ordinary least squares method. The estimated standard error of the extra coefficient indicates that its true value does not differ significantly from zero at 5% level, and hence this check gives no grounds to question the adequacy of the originally chosen model.

Having specified the model for the palm kernel residual series, historical simulation of data over a 24 month period (May 1985 to April 1987) was then performed. This was done to assess the forecasting ability of the model. Next, the model was combined with its lead corresponding regression model to obtain a combined regression-time series model.

Historical simulation of the price data over the 24-month period as well as 6-month ex-post forecast (a means of out-of-sample performance) were undertaken using the combined model (with associated statistics - rms errors and Theil coefficient). The rms errors and the Theil coefficient of the combined model were then compared with those of the best regression model to test the possible superior efficiency of the former (taking some account of time series concept) over the latter. In addition, a plot of simulated series and actual series was also made for the combined model to determine how well the simulated series fits the actual series.

The general form of the combined model is given by

$$Y_t = X^1_t B + b^{-1}(B)\, a(B)E_t \text{ .... } (29)$$

where

$X_t^1 B$ is a structural (economic) component in which X^1_t constitutes a vector of observations on $X_1, X_2 \text{..................} X_K$ made at time t, so that $X_t^1 = (X_{1t} \text{................} X_{kt})$ and B is a vector of estimated constant coefficient $B = (B_1, B_2 \text{................} Bk)$ while $b^{-1}(B)a(B)E_t$ is the time series component i.e., estimated ARIMA model for the residual series of the best regression model in which E_t is normally distributed error term which may have a different variance from the original error term (U_t).

Results and Discussion

Regression run results

The two lead equations for palm kernel prices are presented below (estimated standard errors are in parentheses).

Linear equation

$$P_{1t} = 130.396 + 0.983P_{1t-1} - 61.32EX_t \quad 0.074P_{2t-1} - 0.66T \; .... (30)$$
$$(33.072) \quad (0.067) \quad (16.158) \quad (0.044) \quad (0.091)$$

$R^2 =$ 0.8518 $\bar{R}^2 = 0.8482$ $F = 238.502$ $DW = 2.143$

Double Logarithmic Equation

$$Log_e P_{1t} = 0.990 + 1.053 log_e P_{1t-1} - 0.167 log_e EX_t - 0.205 log_e P_{2t-1} ....(31)$$
$$(0.225) \quad (0.060) \quad (0.040) \quad (0.074)$$

$R^2 = 0.8886$ $\bar{R}^2 = 0.8866$ $F = 444.007$ $DW = 2.04$

where

P_{1t} = World price (in Naira) of palm kernel at time t

P_{1t-1} = World price (in Naira) of palm kernel at time t-1

EX_t = Exchange rate of Dollar/Naira.

P_{2t-1} = World price (in Naira) of palm oil at time t-1

T = Time trend variable (Jan, 1973=1, ...etc)

In equation 31, the coefficients of P_{1t-1} and Et are significantly different from zero at 5% level. P_{1t-1} is positively related to P1t whilst Et is negatively related to P_{1t}. These are in conformity with a priori expectations. P_{1t-1} and T are however wrongly signed and statistically insignificant at 5% level (but P_{2t-1} is statistically signifiant at 10% level)

The double log function is not better than its linear counterpart in terms of the signing of its parameters but has all its parameters significant at 5% level as well as higher R^2 $\bar{R}^2$ ($\bar{R}^2 = .89$) and F value than the linear function ($\bar{R}^2 = .84$). In other words, the double logarithmic function has better statistical fit than the linear function. Since statistical elegance is not directly correlated with forecasting ability, both models were considered for predictive purposes (Steckler (1968).

Table 2 shows the value of Theil coefficient, RMS forecast error and RMS simulation error of the respective equation.

Table 2: Evaluative Statistics of Palm Kernel regression Equation

Statistics	Linear Function	Duble log Function
Theil coefficient	1.05	1.21
RMS forecast error	25.76	61.08
RMS simulation error	43.187	43.84

Source: Data analysis

As indicated in table 2 above, the linear function has relatively lower values for the three statistics than the double log equation, i.e., linear equation outperforms the double log model as a forecasting tool (and therefore chosen as the better regression model for forecasting world price series of palm kernel). Given the U_2 values (1.05 and 1.21), the two models are less accurate than the naive no-change extrapolations. An historical simulation of the linear model is shown in figure 1. And an *ex-post* forecast of the equation over the period (May 1987 to October 1987) is shown in fig. 2. The simulation tracks fairly the general movement of the series though it falls to simulate the upturn in September 1986 reasonably. The forecast is rather poor for it is unable to capture the downward and upward trends in prices effectively. For example, the downward trend in price of July 1987 is not captured.

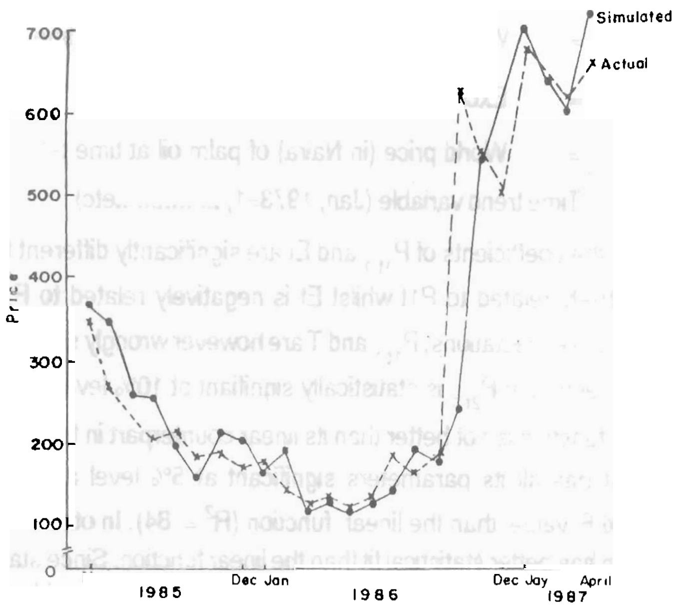
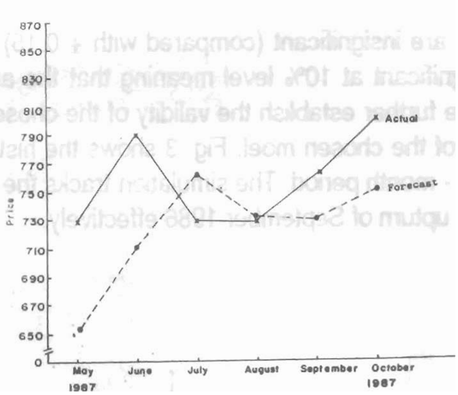


Fig. 1: Historical Simulation of Palm Kernel World Price Series Using the Lead Regression Model. Time bound: (May 1985 - April 1987)



Time

Fig. 2: Ex - Post Forecast of Palm Kernel World Prices Using the Lead Regression Equation.
Time bounds: (May 1987 - October 1987).

Time series analysis

For the regression residual series, the lead equation is

$$\Delta^3 U_t = 0.9496 - 0.4717\Delta^3 U_{t-1} - 0.4176\Delta^3 U_{t-2}$$
$$(0.08159) \quad (0.08523)$$

$$R^2 = 0.227 \; X^2(2,20) = 10.61.$$

The fitted augmented model gives the equation:

$$(1 + 0.4919B + 0.4452B^2 + 0.1062B^3 \; (1-B^3)^{U}t = 1.0162 + E_t \quad \text{............ (33)}$$
$$(0.08411) \quad (0.09027) \quad (0.1075)$$
$$R^2 = 0.235, \quad X^2(3,20) = 10.66$$

The extra coefficient (0.10620) in the augmented model is not significantly different from zero even at 25 percent level thus lending credence to the adequacy of the chosen model.

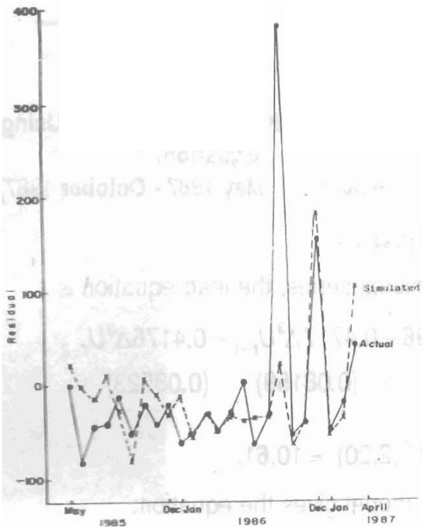
The first 20 autocorrelations of the errors from the fitted ARIMA $(2,0,0)_3$ model are shown in Table 3.

Table 3: Autocorrelations of errors from ARIMA (2,0,0)3 model fitted to series on residual of world prices of palm kernel.

K	1	2	3	4	5	6	7	8	9	10
A	0.01	0.02	-.02	-.03	0.09	-.03	0.03	-.08	-.08	-.04
K	11	12	13	14	15	16	17	18	19	20
A	-.03	0.01	0.00	-.08	-.08	-.14	0.05	0.01	0.03	0.05

Source: Data analysis

The autocorrelations are insignificant (compared with ± 0.15). Furthermore, the chi-squared value is not significant at 10% level meaning that the autocorrelations are not generally too high. These further establish the validity of the chosen model. Fig. 3 shows the historical simulation of the chosen moel. Fig. 3 shows the historical simulation of the ARIMA model over a 24 - month period. The simulation tracks the actual series fairly well but does not capture the upturn of September 1986 effectively.



Time

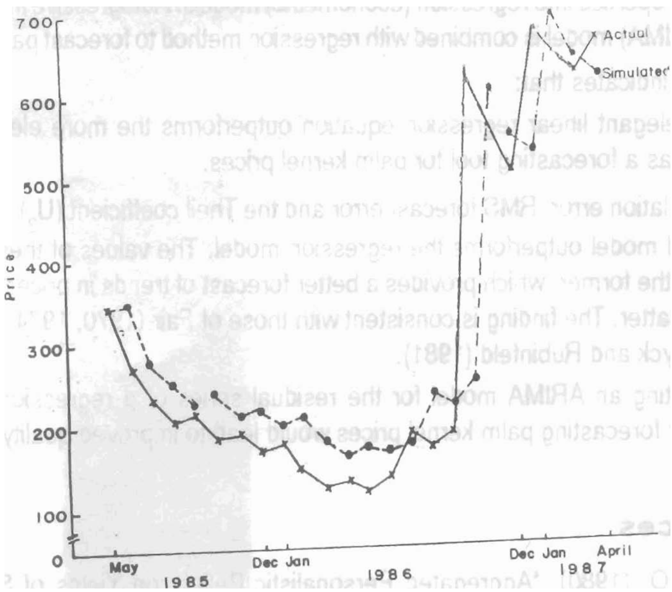
Fig 3: Historical Simulation of (Palm Kernel) Residual Series using ARIMA (2, 1, 0)3
Time bounds: (May - April 1987).

The combined model derived from (30) and (32) is expressed as

$$P_H = 130.96 + 0.983_{11!} - 61.32EX_t - 0.074P_{2t1} - 0.66T + U_t$$

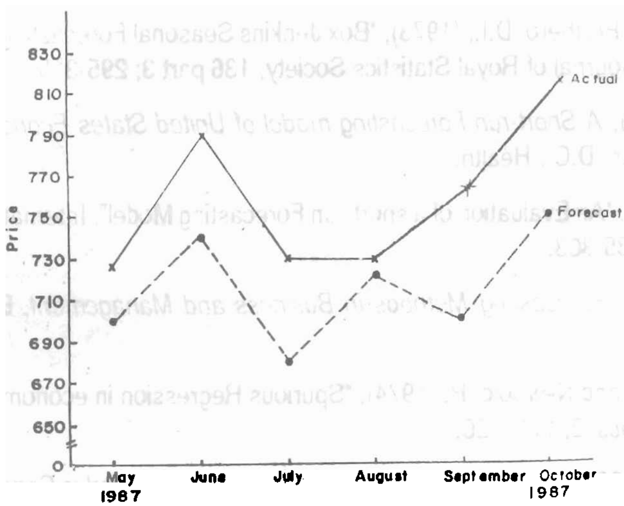
where $U_t = 0.09496 - 0.4717\Lambda^3 U_{t\ 1} - 0.4176\Lambda^3 U_{t\ 1}$............(34)

The historical sumulation of the combined model shows that the RMS simulation error has dropped to 85.27. An historical simulation of the combined model is shown in fig. 4. The simulation tracks the actual series much more closely than when the regression model is used alone. The rms forecast drops from 52.76 to 44.22 whilst the Theil coefficient drops from 1.05 to 0.878. A 6-month *ex post* forecast of palm kernel world price is shown in fig. 5. The forecast captures the trend in price of palm kernel better than when the regression model alone is used. All the evaluative statistics employed thus indicate higher efficiency in forecasting if and when time series properties are incorporated into a regression (econometric) model than otherwise.



Time

Fig. 4: Historical Simulaton of Palm Kernel World Price Series Using Combined Regression - Time Series Model.
Time bounds: (May 1985 - April 1987).



Time

Fig. 5: Ex-Post Forecast of Palm Kernel Prices using Combined Regression - Time Series Model.
Time bounds: (May 1987 - October 1987).

Summary

This work illustrates the improvement in the quality of forecast resulting from incorporating time series properties into regression (econometric) models. Autogressive integrated moving average (ARIMA) model is combined with regression method to forecast palm kernel prices.

Analysis indicates that:

i. The less elegant linear regression equation outperforms the more elegant double log equation as a forecasting tool for palm kernel prices.

ii. RMS simulation error, RMS forecast error and the Theil coefficient (U_2) all show that the combined model outperforms the regression model. The values of these statistics are lower for the former (which provides a better forecast of trends in prices of palm kernel) than the latter. The finding is consistent with those of Fair (1970, 1974), Ludwig (1974) and Pindyck and Rubinfeld (1981).

iii. Constructing an ARIMA model for the residual series of a regression (econometric) model for forecasting palm kernel prices would lead to improved quality of forecast.

References

Bassler, A. O. (1980), 'Aggregated Personalistic Beliefs on Yields of Selec· ·l Crops Estimated Using ARIMA Processes". *American Journal of Agricultural Economics*, Vol. 6 (4): 666-674.

Box, G.E.P. and Jenkins, G.M. (1970), *Time Series Analysis: Forecasting and Control.* San. Francisco: Holden Day.

Chaffield, C. and Prothero, D.I., (1973), 'Box Jenkins Seasonal Forecasting: Problems in a case study". Journal of Royal Statistics Society, 136 part 3; 295-315.

Fair, R.C. (1970), *A Short-run Forecasting model of United States Economy*, Lexington, Massachusetts, D.C., Health.

Fair, R.C. (1974), 'An Evaluation of a short run Forecasting Model", International Economic Review 15: 285-303.

Firth, M. (1977), *Forecasting Methods in Business and Management*. Edward Arnold, London.

Granger, C.W.J. and Newbold, P. (1974), 'Spurious Regression in econometrics". *Journal of Econometrics*, 2; 111- 120.

Lesser, C.V.U. (1969), *Econometric Techniques and Problems*, Charles Griffin and Company Ltd., London.

Leuthold, R.M. (1975), 'On the Use of Theil's Inequality Coefficients" *American Journal of Agricultural Economics*, 57; 344 - 346.

Ludwig, R. S. (1974), *Forecasting Short-Term Savings Deposit Flows: An Application of*

Time series and a Regional Analysis, Unpublished Master's Thesis, Sloan School of Management, M.I.T.

Naylor, T.H., Seeks, T. H. and Wichern. D.W. (1972), "Box-Jenkins Methods: *An Alternative Econometrica Models" International Statistical Review,* 40: 121 - 37.

Nelson, C.K. (1970), *Applied Time Series Analysis* San Francisco: Holden Day.

Newbold, P. and Granger, C.W.J., (1974), "Experience with forecasting Univariate time series and the combination of forecasts". *Journal of Royal Statistical Society,* 137: 131 - 146.

Ojo, M.O. (1989), *The Structural Adjustment Programme and Nigeria's Export Crop Subsector.* A paper presented at the National Conference on the Impact of S.A.P. on Nigerian Agriculture and Rural Life, NISER, Ibadan.

Olayide, S. O. and Olayemi, J. K. (1977) *Forecasting the Nigerian Cocoa Economy,* TECHNICAL REPORT AETR NO. 7711.

Olufokunbi, B. (1982), "An Econometric Analysis of Demand for Beef in Ile-Ife, Nigeria (1975 - 78)" *Nigerian Journal of Agricultural Economics,* 1(1): 29-39

(1983), The Box-Jenkins Models or the Exponentially Weighted Average: An Enquiry into the Performance of Forecasting Models for Agricultural Commodity Prices in Nigeria, An Unpublished Paper.

Pindyck, S. R. and D.L. Rubinfeld (1981), *Econometric Models and Economic Forecasts* McGraw, Inc.

Reid, D.J. (1969). *A Comparative Study of time series prediction techniques on economic data.* Ph.D Thesis. Dept. of Mathematics, University of Nottingham.

Schmitz A., and Walta D.G. (1970), "Forecasting Wheat Yields: An Application of Parametric Time Series Modelling". *American Journal of Agricultural Economics,* 52: 247 - 254.

Steckler, H. O. (1968), "Forecasting Econometric Models: An Evaluation" *Econometrica* 36: 437-463

Theil, H (1966) *Applied Economic Forecasting* Amsterdam: North Holland Publishing Company

APPENDIX I

Actual and Predicted Prices of Palm Kernel (May 1985 to April 1987).

Palm Kernel

Actual	Predicted Price (Regression)	Predicted Price (Combined)
347.7	346.6	367.46
268.5	351.3	349.3
232.7	278.0	262.1
205.0	247.0	256.1
214.3	227.5	199.4
184.5	237.0	154.7
189.7	209.9	216.7
172.0	217.4	205.4
178.2	199.9	166.7
144.1	207.6	196.2
123.7	177.4	119.0
132.4	162.7	130.0
120.2	170.4	119.7
132.8	164.6	129.8
184.1	181.0	142.7
167.5	233.5	196.3
187.6	221.8	183.9
629.1	249.4	240.8
551.5	606.9	543.9
501.5	543.5	502.1
675.3	521.4	703.8
641.3	691.5	638.7
616.9	637.7	601.6
659.8	617.8	721.9

APPENDIX II

Actual and Forecast Price Data for the Period (May 1987 to Oct. 1987).

Palm Kernel

Actual	Forecast (Regression)	Forecast (Combined)
728	653.1	702.3
789.3	712.5	741.6
728 7	64.8	679.4
728 7	32.3	710.1
762.8	729.3	701.5
805.6	753.8	754.8