FORECASTING THE YIELD OF SORGHUM IN NIGERIA USING ALTERNATIVE FORECASTING MODELS

NMADU J.N.* and D.O.A. PHILLIP**

*Dept. of Agricultural Engineering, The Federal Polytechnic, Bida, Nigeria

**Department of Agric. Econs and Rural Sociology, A.B.U. Zaria, Nigeria

ABSTRACT

Grafting technique was used to estimate a forecasting model for yield (kg/ha) of sorghum in Nigeria. The estimated model was compared with the traditional models to show that time series data does not always relate linearly to trend over the entire sample space. The results show that the grafted model was superior to others.

INTRODUCTION

Accurate planning is predicated on a reliable and flexible forecasting model, which also is a function of accurate data. Knowledge situation in agricultural production is still imperfect; hence, it is still dependent on weather, policy and risk (Just 1974, 1975; Meilke 1976 etc). the considerable turbulence in agricultural policy equilibrium also contributes to difficulties in obtaining and maintaining credible data. As a result, planning tools in agriculture

aic ... ~d from estimated data time series, that is, in most developing nations, Nigeria inclusive, the capability and capacity to create and maintain a database and the ability to estimate predictive forms are seriously deficient. Where the data is even available, the accuracy is subject to various errors (e.g Nmadu 1998). In these circumstances, time series data, though may be subject to some error, still become the best data for estimation.

Most of the traditional tools available for time series analysis tend to force the data to a chosen form, perhaps based on theoretical consideration, ease of estimation or interpretation (Fuller 1969). Most of the forms are linear trend or a variant of it. However, it has been shown that the time series might not be linearly related over the entire sample space (Fuller 1969, Phillip, 1990). Therefore, a more flexible form, which co.iforms to the form the data suggest, is required.

The objective of this present study is to determine the appropriate prediction form for yield (kg/ha) of sorghum in Nigeria (1961-1997) and com-

pare it with the linear trend or some of its variants. The form (also called mean equation) would be obtained by grafting technique (Fuller, 1969).

MODEL SPECIFICATION

There are many functional forms of estimating forecasting models. The common ones used include:

$$Y_t = a + btl \qquad \text{linear} \qquad (1)$$

$$Y_t = a + b \ln t \qquad \text{semilog} \qquad (2)$$

$$\ln Y_t = a + bt \qquad \text{growth or exponential} \qquad (3)$$
(Tschirley, 1995)

Where Y_t is yield of sorghum. t is trend variable while a and b are structural coefficients to be estimated. The shortcoming of the above functional forms to forecasting time series, is mainly because time series data might not be linearly related to the series over the entire sample period as the models tend to suggest. Therefore, the above m̲̅̅̅̅̅̅ be improved by dividing the data into different segments and applying different functional forms as suggested by the data rather than force the data to accept a particular form. A

graphical examination of the data (estimated yield of sorghum in Nigeria 1961-1997) shows that it can be divided into different segments; hence the following trend function was suggested:

$$Y_t = \alpha_o + \beta_o t, \quad t \leq 1982 \quad (4)$$

$$Y_t = \alpha_1 + \beta_1 t + \phi_1 t^2, \quad 1982 < t \leq 1988 \quad (5)$$

$$Y_t = \alpha_2 + \beta_2 t, \quad 1988 < t \leq 1994. \quad (6)$$

Where: Y_t = yield of sorghum in kg/ha in year t

t = trend

α's, β's and ϕ = structural parameters to be estimated

Fuller (1969) elucidated the desirable properties of the mean function, which we hope to estimate from equations (4) – (6) above, thus,

That the model should be continuous at some joint points (K's), and

That the model should possess first derivative at the joint points. Specifically. $\alpha_o + \beta_o K_1 = \alpha_1 + \beta_1 K_1 + \phi_1 K$ (7)

$$\alpha_1 + \beta_1 K_2 = \phi_1 K_2^2 = \alpha_2 + \beta_2 K_2 \quad (8)$$

$$\beta_o = \beta_1 + 2\phi_1 K_1 \quad (9)$$

$$\beta_1 + 2\phi_1 K_2 = \beta_2 \quad (10)$$

Where K's = joints of the different segments of the function. In this study, $K_1 = 1982$, while $K_2 = 1988$. There are seven structural parameters and four restrictions on the mean function thus reducing the number of parameters to be estimated to three. Since the primary purpose of estimating the model is to be able to estimate future yields, all the parameters on the later model were retained (Phillip, 1990). Thus in all, parameters α_2, β_2, ϕ_1 were retained for subsequent estimation. Therefore, α_o, β_o, α_1, β_1 were dropped and equation (7) – (10) were rederfined in favour of the dropped parameters as follows:

$$\alpha_o + \alpha_2 + \phi_1 K_2^2 - \phi_1 K_1^2 \quad (11)$$

$$\beta_o = \beta_2 - 2\phi_1 K_2 + \beta_2 K_2 \quad (12)$$

$$\alpha_1 + b\alpha_2 + \phi_1 K_2^2 \quad (13)$$

$\beta_1 = \beta_2 - 2\phi_1 K_2$

$$(14)$$

The grafted or mean equation can now be obtained by substituting α_o, β_o, α_1, β_o in equation $(4)-(6)$. The result is

$Y_t = \mu_o Z_o + \mu_1 Z_1 + \mu_1 Z_2 + U_t$ (15)

(See Appendices A and B for details)

Where $Z_o \quad = 1, \forall t, \; \forall = $ for all

$Z_1 \quad = \quad t, \forall\, t$

$Z_2 \quad = $
$K_2{}^2 - K_1{}^2 - 2t\,(K_2 - K_1),\ t \leq 1982$

$= $
$(t - K_2)^2,\ 1982 < t \leq 1988$

$= $
$0, t > 1988.$

$U_t \quad = $
error term assumed to be well behaved.

Equation (15) is the mean equation; it is continuous with the restrictions given in equations (7) – (10). Ordinary least squares technique (Everitt and Dunn, 1990 and Johnston, 1991) was applied to equations (1), (2) (3) and (15) using the observed data on yield (kg/ha) of sorghum in Nigeria from 1961 to 1994. The observed data for 1995 to 1997 were used to carry out an ex-post evaluation of the estimated equations in order to assess their predictive performance. The data were obtained from FAO (1970-1997). In addition, two separate scenarios were simulated with model (15). The first is when $k_1 = 1971$ and $k_2 = 1981$, corresponding to the years of rise and fall of oil boom in Nigeria respectively. The second is when $K_1 = 1976$ and $K_2 = 1986$, corresponding to the year State Marketing Boards were abolished and the year Structural Adjustment Programme (SAP) was introduced, marking the abolishment of Federal Marketing Boards respectively. The purpose of these simulations was to find out the possible effect these scenarios had on sorghum production. After the simulations, it was observed that the two scenarios were monotonic transformations of the grafted functions, although, their estimating qualities were better. In particular, the scenarios accounted for higher amount of variation than the grafted functions and the F-values were generally higher. Therefore, an attempt was made to bring the two scenarios and the grafted

functions together in order to isolate the actual effect of the scenarios. Consequently, by adding two variables to the grafted functions, a third scenario was simulated. The variables were defined as:

$Z3 = 3$, t<1967, transitional, multinational marketing agencies

2, $1967 \leq t \leq$ 1976, Regional marketing boards

1, $1976 \leq t \leq$ 1986, Federal Marketing Boards

0, t > 1986, end of state control, free-market economy

$Z4 = 1$, $t <$ 1971, agricultural boom era

0, $1971 \leq t \leq$ 1981, oil boom era

1, t > 1981, end of oil boom as oil prices fell Following the estimation of the third scenario, it was discovered that the estimated properties of the function were plague with acute positive autocorrelation. Therefore, the model was re-estimated using the weighted least square technique. However, although the estimating properties of the new model improved tremendously, the model was not very good for forecasting purposes. Hence, the model was left as Scenario 4.

RESULTS AND DISCUSSION

Estimates of the hypothesized structural parameters of equations (1),(2),(3),(15 is presented in Table 1.

The main objective of estimating equations (1),(2) and (3) was the need to evaluate and compare the predictive and forecasting ability of the mean or grafted equations with them. The result shows that all the structural parameters were significantly different from zero except for Z3 and Z4 in Scenario III. It therefore means that the quadratic falls in the yield (kg/ha) of sorghum between 1982 and 1988 was not incidental. The drought of 1983 and the considerable turbulence in the amount and distribution of rainfall that followed might have caused this. In terms of rank, Scenario I seems to be the best accounting for about 46% of the observed variation in the yield of sorghum in Nigeria during the estimation period; while the Growth model

Table 1: Estimates of Alternative Forecasting Schemes for Yield of Sorghum in Nigeria

Variable	Grafted	Linear	Growth	Semilog	Scenario I	Scenario II	Scenario III	Scenario IV
Constant	130306-3	-24701.95						
	-18.989	-192186.9	-82198.2	-112732	-125586	1.772		
Z1	65.998a	12.89a						
	.013b	25426.39	41.843a	57.188a	63.597a	0.459		
Z2	(3.4)	(2.8)						
	(2.5)	(2.8)	(5.27)	(4.773)	(3.076)	(15.253		
Z3	5.49a	-	-	-	3.321a	3.407a	5.428b	-2.65
Z3	(2.8)	-	-	-	(4.122)	(3.895)	(1.674)	(-2.991)
-	-	-	-	-	-	-	25.255	201.547
-	-	-	-	-	-	-	(0.22)	(5.408)
R^2	0.343	0.202	0.164	0.2001	0.4683	0.4474	-59.664	-181.436a
F	8.829	8.091	6.259	8.041	14.561	13.424	(0.331)	(-2.954)
Df	31	32	32	32	31	31	0.30515	0.993
							4.227	2125
							29	29

57

was the poorest accounting for only 17% of the variation. However, more evidence is needed to reaffirm this fact, as we shall do presently. A diagnostic check on the various models estimated is presented in Table 2.

However, Pindyck and Rubinfeld (1976) have asserted that poor estimating properties of single-equation models do not necessarily affect their forecasting ability.

After the estimation, the models were used to forecast the variables during the estimation period and the forecast compared with the observed data. The result is presented in Table 1. An examination of the Figure shows that the linear models gave a monotonic increase (or decrease) of the variables during the estimation period. However, the grafted models gave a quadratic fall (or rise) at some regions, which were very consistent

Table 2: Diagnostic checks on the estimated models

	MARE	U	PTPE	RSME
	Mare = $^1/_{n \, abs}(Y_t$-$y_t/$ Y_t)	U = $\sqrt{(Y_t$-$y_t)^2}/$ $\sqrt{\sum Y_t^2}+\sqrt{\sum y_t^2}$	PTPE = $^1/_n$ $\sum TPE_t$	RMSE = $^1/_n \sum (Y_t$- $y_t)^2$
Grafted	0.0614	0.1367	0.5000	50487.52
Linear	0.0877	0.1537	0.5000	63252.04
Growth	0.0420	0.1583	0.5294	63968.80
S e - milog	0.0879	0.1538	0.5294	63331.86
S c e - nario 1	0.0503	0.1225	0.4412	40858.30
S c e - nario 2	0.0504	0.1250	0.4412	42466.50
S c e - nario 3	0.0001	0.0128	0.5000	429.89
S c e - nario 4	0.1960	0.1639	0.5588	80291.55

Note::Y_t = observed, Y_t = estimated, n = sample size, TPE_t = 0 if $Y_t - Y_t -1 > 0$ else, TPE_t = 1, $\forall t$, $y_t - y_t - 1$, MARE = mean absolute relative error, U = Theil inequality coefficient, PTPE = percentage turning point error, RMSE = root mean-square simulation error
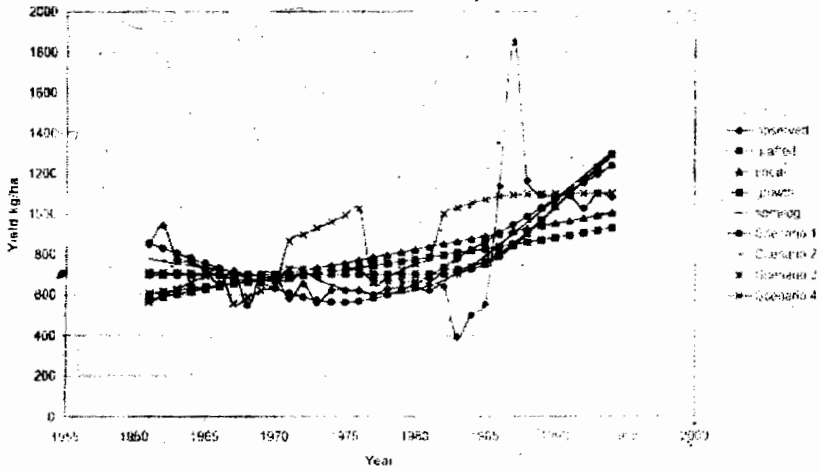
Figure 1. [illegible caption]
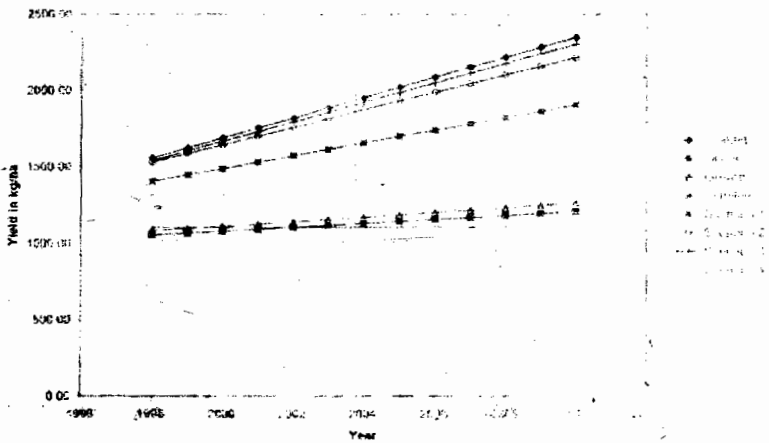
Figure 2. [illegible caption]

with the observed data. This shows that the time series did not linearly relate to trend during the whole time space and that the grafting technique better fit a time series data rather than linear models. Scenario 3 exhibited another important feature. It was observed that there was a considerable back-pull of the variables at the time of changeover from one policy regime to the other, especially when the policy change causes a major turbulence to the economy.

At the time of the data collection for this research, observed data were available for the years 1961 to 1997. The data for the subperiod 1961-1994 were used for the estimation while the data for the subperiod 1995-1997 were reserved for performing an Ex-post forecast of the variables. The Ex-post evaluation of the models was carried out and compared with observed values. The result of the forecasts is presented in Table 3. The evaluation shows that all the linear trends underestimated while all the grafted functions overestimated except for Scenario 4, which overestimated at the beginning and underestimated at the end showing a decreasing trend. According to Fotopoulos (1995), trend is a bundle of unmeasured

'technical progress'. With regards to yield, we expect improvement in seed technology; hence, as a result, better seed with higher yielding potentials should be expected with growth in trend. Therefore, if that is the case, then the grafted models (except Scenario 4) are consistent with expectation and therefore better forecasting models. Hence, this goes to show that time series data are better modeled using grafting techniques. The other important factor to consider is that since we are dealing with a future of which knowledge certainty cannot be guaranteed, it is better to overestimate rather than otherwise. A comparison of this result with Phillip (1990) shows a similar but opposite trends.

An attempt was made to make an Ex-ante forecast of the variables to year 2010. The forecast serves two purposes:

it reinforces the validity of the models, most especially beyond the Ex-post period, and it gives an insight into what we expect the variables to grow to in future, based on past and present trends thereby aiding planning.

The result is presented in figure 2. The year 2010 was chosen because all targets being set now for

Nigeria are geared towards that year. The results show consistency with the estimated results and the expectation about the future. The mean function gave a decreasing function up till 1983 and then a point of inflexion, and finally an increasing function. However, a careful examination shows that the grafted functions are generally, in a family of good forecasting models except for Scenario 4, which is inconsistent with all the results and expectations. This goes to prove that significant t, F and R^2 in single-equation models does not necessarily give good forecast period after period (Pindyck and Rubinfeld, 1976). These models can therefore be used to set targets for sorghum production in Nigeria. For example, the government can set targets based on population growth of the country and the food needs of the populace year by year and thereafter task Scientists and Researchers to develop sorghum varieties that can be used to achieve the levels of output required to achieve the food needs. This is the only way to fight food insecurity in Nigeria.

CONCLUSION

The forecasting ability of commonly used regression models was compared to a model which try to incorporate all the observed functional segments as suggested by the observed data on yield kg/ha of sorghum in Nigeria from 1981 to 1997. Results demonstrated that economic time series, especially agricultural, do not always relate linearly on trend. Therefore, the resulting mean equation, incorporating local trends tends to provide a more reliable forecasting tool, thus aiding policy direction.

REFERENCES

Everitt, B.S. and G. Dunn (1991). Applied Multivariate Data analysis. London, Edward Arnold.

FAO. 1980-1995. Production Yearbook, Various Issues Rome, Italy

FAO. 1975-1992. Monthly Statistically Bulletin, Various Issues Rome, Italy

FAO. 1992-1997. Quarterly Statistically Bulletin, Various Issues Rome, Italy

Fotopoulos, C.V. 1995. "A Partial Adjustment Model with Rational Expectations Applied to the goat Sector in Greece", Agricultural Economics 12: 163-70.

Fuller W.A. 1969. Grafted Polynomials as Approximating Functions. Australian Journal of Agricultural Economics. 35-46.

Johnston J. 91991). Econometric methods, International Edition, New York, McGraw-Hill.

Just E.R. 1974. an Investigation of the Importance of Risk in Farmers' Decisions, American Journal of Agricultural Economics. Decmember, 836-43.

Meilke K.D. 1976. Acreage response to Policy in the Prairie Provinces, American Journal of Agricultural Economics, August, 572-77.

Nmadu J.N. 1998. Impact of Bejin-Doko Community Bank on Income of Small Scale Farmers in Doko and Jima Districts of Niger State. M.Sc thesis, Dept. of Agric. Econ. And Rural Sociology, ahmadu Bello University, Zaria.

Phillip D.O.A. 1990. Grafted Polynomials as Forecasting Functions: Application to Cotton Production in Nigeria. Savanna Journal 11 (1) :-108.

Pindyck, R.S. and D.L. I infeld. 1976. Econometric Models and Economics Forecasts. McGraw-Hill Book Company, New York, U.S.A.

Tschirley D. (1995). "Using Microcomputer Spreadsheets for Spatial and Temporal Price Analysis: an Application to Rice and Maize in Ecuador". In: Scott G.J. (ed). Prices, Products and People: Analysing Agricultural Markets in Developing countries. Boulder, Colorado, Lynne Reinner Publishers Inc. Pp. 277-300

Table 3: Ex-post Forecast of Yield Using the Estimated Models (1995-1997) [23]

Year	Observed	Grafted	Linear	Growth	Semilog	Scenario 1	Scenario	Scenario	Scenario4
1995	1060	1359.71	1013.60	1038.99	1012.96	1278.59	1358.06	1349.68	1098.91
		(290.71)	(-55.40)	(-30.01)	(-56.04)	(209.59)	(289.06)	(280.68)	(29.91)
1996	1144	1425.71	1026.49	1052.58	1025.71	1329.43	1415.25	1413.28	1099.37
		(281.71)	(-117.51)	(-91.42)	(-118.29)	(176.43)	(271.25)	(269.28)	(-44.63)
1997	1167	1491.71	1039.38	1066.35	1038.44	1362.27	1472.44	1476.87	1099.83
		(324.71)	(-127.62)	(-100.65)	(-128.56)	(195.27)	(305.44)	(309.87)	(-67.17)

Notes: Values in parenthesis are computed errors in estimation.

Appendix A

The full derivation of equations (11), (12) (13) and (14) are given below:-

By inspecting equations (7), (8), (9) and (10), it is obvious that it is better to start from (10) because it has only one term, which we intend to drop i.e.

(A_1) $\beta_1 = \beta_2 - 2\phi_1 K_2$

From (9) substituting (A_1), we obtain

(A_2) $\beta_0 = \beta_2 - 2\phi_1 K_2 + \phi_1 K_1$

We now estimate a_1 from (8) substituting (A_1) and (A_2) i.e.

$a_1 = a_2 + \beta_2 K_2 - K_2(\beta_2 - 2\phi_1 K_2) - \phi_1 K_2) - \phi_1 K_2^2 - \phi_1 K_2^2$

(A_3) $a_1 = a_2 + \phi_1 K_2^2$

Finally we estimate a_0 by making use of (A_1), (A_2) and (A_2)

$a_0 = a_2 + \phi_1 K_2^2 \beta_2 K_2 - K_1(\beta_2 - 2\phi_1 K_2) + \phi_1 K_1^2 - K_1(\beta_2 - 2\phi_1 K_2 + 2\phi_1 K_1$

$= a_2 + \phi_1 K_2^2 + K_1 \beta_2 - 2\phi_1 K_2 K_1 + \phi_1 K_1^2 - K_1 \beta_2 + 2\phi_1 K_2 K_1 - 2\phi_1 K_1^2$

(A_4) $a_0 = a_2 + \phi_1 K_2^2 - \phi_1 K_1^2$

Appendix B

We now derive the grafted trend equation (15). This is done by simply substituting

63

$(A_1), (A_2), (A_2)$ and (A_4) as obtained above in equations (4) – (6).

From equation (4), substituting for a_o and β_{o1}

$Y_t = (a_2 + \phi_1 K_2^2 - 2\phi_1 K_1^2) + \beta_2 - 2\phi_1 K_2 + 2\phi_1 K_1)t$

$= a_2 + \phi_1 K_2^2 - \phi_1 K_1^2 + \beta_2 t - 2\phi_1 K_2 t + 2\phi_1 K_1 t$

$(B_1)\ Y_t = a_2 + \beta_2 t + \phi1(K_2^2 - K_1^2 + - 2K_2 t + 2K_1 t),\ t <= 1982$

From equation (5), substituting for a_1 and β_1,

$Y_t = (a_2 + \phi_1 K_2^2) + (\beta_2 - 2\phi_1 K_2) t + \phi_1 t^2$

$= a_2 + \phi_1 K_2^2 + \beta_2 t - 2\phi_1 K_2 t + \phi_1 t^2$

$(B_2)\ Y_t = a_2 + \beta_2 t + \phi1(K_2^2 - 2K_2 t + t^2),\ 1982 < t \leq 1988$

From equation (6). all coefficients were retained for forecasting purposes.

$(B_3)\ Y_t = a_2 + \beta_2 t$

The grafted equation, (15), was then formed by inspection of (B_1), (B_2) and (B_3) above. The mean equation is continuous on the data set.