REVIEW OF MARKETING AND AGRICULTURAL ECONOMICS
VOL. 43, No. 2 (JUNE, 1975)

FORECASTING CROP QUALITY*

G. J. Ryland†

The need for a forecasting system of biological quality arises as a result of the price-quality payment schemes in grower and processor contracts which operate in many agricultural cropping industries. The seasonal nature of the series of vertical quality height gives rise to questions as to the repetitive pattern of the shape and trend translation of the series. These hypotheses can be tested using conventional statistical methods. For non-stationary series, however, a Box-Jenkins type dynamic seasonal model is proposed. These forecasting procedures are applied to a series of sugar cane quality.

1 INTRODUCTION

Quality analysis and economic theory are normally associated with product differentiation and monopolistic competition which are usually analysed within a static framework. In a dynamic economic environment, the importance of product quality variation, or the manner in which consumers adjust ruling market prices over time in response to changes in quality over time of a single product, has recently been demonstrated by Cowling and Rayner [2] in their "demand for brands" analysis of Great Britain's tractor market. Besides this recent study very little empirical work has been done on the dynamic aspects of the quality variation problem. The main reason, of course, is the multi-dimensional character and intangible nature of quality which arises even in the absence of price and time effects. In agricultural production, however, many field crops contain a dominant but quite fundamental and naturally occuring quality attribute which is both measurable and time dependent.

Differences in biological quality of many crops can be measured empirically in terms of a single vertical quality attribute which may be contrasted with the sometimes intangible horizontal and vertical quality differences. For example, in the production of wheat the percentage of gluten or wheat protein can be measured and in oil seeds production the percentage of the oil recovered, while in sugar cane production, the

* Manuscript received October, 1974.

† Lecturer in Economics, University of Queensland, St Lucia, and part-time research consultant, I.A.C.

I wish to thank a colleague, Dr M. N. Bhattacharyya, for helpful discussions throughout the period of this investigation and Dr A. Pagan of the Australian National University, Canberra, for the use of his non-linear least squares programme. Mr G. Gengate assisted with the computations.

commercial cane sugar content can be obtained. The biological quality attribute, which varies with time and location, is translated into differences in factor market prices by purchasers of raw material. In the case of integrated industries, forward contracts between growers and processors usually provide for payment per unit of raw material supplied based on quality.

When changes in growers returns over time are reflected by changes in quality over time in the above manner, we see the need for a forecasting system of crop quality at three levels. First, at the level of the individual producer, forecasts of crop quality can be used to prepare his annual budgets and to make decisions with respect to the optimum allocation of production among different quality grades. Similarly individual processors can make use of the anticipated quality cycle in order to establish the appropriate length of processing season. Guise and Ryland [8] have used the crop quality cycle phenomenon to determine the optimum period of production in the case of a single raw sugar manufacturer. Second, at the industry level the determination of the optimum quality mix of a particular commodity could offer several advantages. Matsumoto and French [11] have recently shown how the optimum quality distribution of brussel sprouts in a timeless and spaceless economic environment can be determined when demand and quality cost of each grade are included. Using a quite different approach, Ryland and Guise [12] have shown the importance of quality differences over both time and space when prices are fixed for determining the optimum competitive schedule of sugar cane production through cane processing plants. Third, there seems a need to investigate regional differences in crop quality so that any potential regional trading advantages which may arise as a result of demands for different qualities may be realized.

Given the need for a forecasting system of crop quality it is necessary to distinguish between two broad groups of available systems which are by no means mutually exclusive. Firstly, there are those techniques which study the influence of independent or exogenous variables on the dependent series and, secondly, those methods which examine the previous behaviour of the series itself.

The first type of forecasting system generally involves the correlation of such variables as rainfall, temperature and soil moisture with the level of crop quality in any given time period. In order to forecast crop quality in any given time period with this approach, we require forecasts of the level of the independent variables. Unfortunately, predictions of naturally occurring phenomena such as amount of rainfall and temperature over time are relatively unreliable. The forecast error using this system is compounded by the error of the forecasts made for each independent variable. To overcome this difficulty it is possible to analyze the behaviour of crop quality by using discrete time periods, for example, weeks as a surrogate for average weekly rainfall, soil moisture and temperature. The level of crop quality in each discrete time period can then be expressed as a function of the time period over which observations were obtained. Given the time period, forecasts can be made using this method provided that the series can be regarded as repetitive with constant amplitude from one season to the next.

Associated with the previous method is a related system in which the dependent series is influenced by a single variable whose level is controlled by the forecaster. Control variables in crop quality prediction could be amount of fertilizer or volume of irrigation water applied from the viewpoint of the single producer. In this case the level of crop quality in each time period could be set as a function of the level of control variables and perhaps lagged values of the series of crop quality.

A second type of forecasting system which does not use a behavioural model, as such, involves an analysis of the time series taken out of context. That is, the only input to the forecasting system is the previous history of the series itself. Forecasts are made by projecting forward on the basis of relationships among previous occurrences. A forecasting model which reproduces previous patterns can then be used to forecast future values within a predetermined degree of reliability. The identification of such a model requires a close examination of the effect of lagged dependent variables on the current observation.

The choice of a forecasting system (assuming cost considerations are negligible) depends on the purpose for which the forecasts are intended and the amount of information available to the forecaster. For example, a forecasting model for the individual grower might include a behavioural model associated with a control variable such as fertilizer. On the other hand regional crop quality forecasts would include the series itself (and perhaps the time period in which observations were obtained) as it is impossible to place controls in any aggregative fashion. Our aim is to develop an appropriate forecasting system to handle the latter situation.[1]

2 SOME IMPORTANT PROPERTIES OF A REGIONAL FORECASTING SYSTEM FOR CROP QUALITY

Several desirable properties of a forecasting model for crop quality could be summarized as follows:

(i) The model selected from a finite range of alternative formulations is that one which minimizes the error between previous estimated values and actual observations.

(ii) The number of pieces of information required to produce an estimate must be kept to a minimum.

(iii) For ease of subsequent calculations the estimating equation should be continuous and differentiable with respect to time.

However, for forecasting to various lead times, the above properties might conflict with the most desirable property of any forecasting system which is to produce good forecasts. For example, a first-order

[1] Spivey and Wecker [14] have recently reviewed the type of forecasting models applicable to regional economic forecasting. They carefully distinguish between the behavioural or extrinsic models and intrinsic models which extrapolate using the previous values of the series itself.

autoregressive model may be all that is required for making forecasts at a lead time of one period. For longer lead times a more complex model may be more appropriate. Thus to choose among competitive forecasting models at different lead times we require a criterion by which it is possible to compare relative forecasting efficiency. For this purpose we adopt the Thiel coefficient, U, with:

$$U = \sqrt{\frac{\Sigma_t(P_t - A_t)^2}{\Sigma_t(A_t - A_{t-b})^2}}$$

where P_t = predicted value at time t, and A_t = actual observation. The objective is to choose that forecasting model which minimizes U over a range of different lead times for some base period change, $(A_t - A_{t-b})$.[2]

3 A SUGGESTED PROCEDURE FOR FORECASTING REGIONAL CROP QUALITY

Before it is possible to outline the technical procedures involved in forecasting crop quality, we require a fundamental interpretation of the seasonal quality series to be compared with the usual class of time series normally available. An observed time series may be defined as an ordered sequence of observations taken from an infinite population at equidistant points in time. The difficulty of treating the seasonal crop quality series as a typical observed time series arises because observations on crop quality are normally only available throughout the harvesting season each year which varies with the type of crop and the particular region. In the case of wheat, harvesting operations each year are usually confined to a short period of 1-month duration compared with sugar cane which is harvested over a 6-month period. In addition to the problem associated with the truncated nature of the observed series of crop quality each year, is the practical one of relating observations in one season to correspond with observations in successive seasons recorded during the same time interval, as harvesting is not confined exactly to the same period from season to season. Thus we require assumptions relating to the occurrence of missing observations during the non-harvest period and for standardizing the length of the harvesting period each year.

The missing observations on crop quality during the non-harvest period are not random occurrences but are deliberately planned for in the course of farming operations. Consequently these missing observations have zero expectation and may be omitted. On the other hand the variable length of the harvesting season each year may be regarded as a random occurrence as the choice of the harvesting season is usually constrained by seasonal factors. On this basis we can standardize the time period and the length of the harvesting season each year by filling in the missing observations which occur either at the commencement or the completion of the harvesting period.

---

[2] Schmitz and Watts [13] have also used Thiel's coefficient to compare the accuracy of their forecasting models relative to a naive or "rule of thumb" forecasting technique.

To provide estimates of the missing observations on crop quality at the end of the harvesting season each year we require to estimate a relation which faithfully reproduces the intraseasonal behaviour of crop quality. At the regional level, this can be done by setting quality as a function of the parabolic time trend and estimating, using ordinary least squares, an equation of the form:

$$Q_t = f(t) + u_t \quad \quad . \; . \; . \; (1)$$

where Q_t = unit of quality measurement in period t.

$f(t)$ = parabolic function of the time trend variable t, where $t = 1, \ldots, T$

and u_t = random error term.

One of the problems associated with the fitting of a parabolic time trend of the form given by equation (1) is the problem of autocorrelation among the residuals which give inefficient estimates of the partial regression coefficients. However my experience with autocorrelation with this type of estimation problem is that the fitting of a spliced polynomial time trend continuous at the join points, such as those discussed by Fuller [5], not only overcomes the problem of autocorrelation but also considerably improves the significance of the estimates. Estimates obtained using this form of estimating equation can be used to fill in the missing observations at the end of the harvesting season so as to standardize the length of the harvesting period each year.

The slope and trend parameters for each year's fitted function of the form of equation (1) can be compared using an extended dummy variable regression model in the conventional manner. If there is no significant difference in the trend and slope parameters each year compared with a selected base year, there is no need to proceed further. In this case the forecasting function would simply be the function fitted to crop quality data over all seasons.

When there are significant differences in the slope and trend coefficients each year we require a forecasting system which is able to handle non-stationary seasonal series.[3] While there is a divergent stream in the literature on time series analysis to handle the estimation of seasonal series,[4] we have used a mixed autoregressive-integrated moving average (ARIMA) seasonal model (of the type discussed by Box and Jenkins [1]) applied to the output from an appropriate linear filter of the original crop quality series.

---

[3] An alternative way of viewing the significance of the coefficients in the dummy variable regression model is that the series could be a possible candidate for variance components analysis with random time effects of the type recently discussed by Maddala [10]. Variance components models with random time effects assume a stationary series with random variation over time about a fixed mean. A broader and less restrictive class of models is the linear homogeneous non-stationary models for which stationarity can be induced by a suitable linear filter. We adopt Box and Jenkins' suggestion that it is much more realistic to specify, at least initially, a stochastic non-stationary trend when random changes in the level of the series are observed.

[4] *See*, for example, Hannan [9], Terrel and Tuckwell [15] and Doran and Quilkey [4].

In order to fit a stable model to the series of sugar cane quality the Box-Jenkins iterative model building procedures are employed. These stages are:

(i) Model Identification: The series is differenced using a suitable differenced operator, $\nabla_s^D \nabla^d Q_t$ (where D, d is of low order and s is the seasonal frequency) and the sample autocorrelation and partial autocorrelation functions of the differenced series are examined. A tentative model is selected usually at that level of differencing which reduces variability and induces stationarity in the differenced series.

(ii) Fitting: The parameters of the chosen model are estimated by least squares.[5]

(iii) Diagnostic Checking: Residual Analysis is applied to check the adequacy of the model as a stationary stochastic process.

(iv) Forecasting Performance: Postsample forecasting at various lead times is conducted to evaluate the adequacy of the model as a forecasting system. We have used Theil's Inequality Coefficient for comparing the mean square forecasting error of alternative models for a range of different lead times.

When forecasting performance is the sole objective of the analyst, the choice of an appropriate forecasting system must ultimately depend on its ability to provide good forecasts. The steps above are repeated for alternative models so as to assess the sensitivity of forecasts to changes in model structure for different lead times.

4 AN EMPIRICAL EXAMPLE

The above procedures were used to establish an appropriate forecasting system for Commercial Cane Sugar (C.C.S.) content of sugar cane which follows a distinct parabolic seasonal pattern. The C.C.S. series or maturity cycle of sugar cane for a particular sugar district in Australia for 1971 season is graphed in figure 1. The graph of average weekly C.C.S. plotted over time shows that C.C.S. at first slowly increases, reaches a maximum in mid-October and then falls rather sharply towards the end of the harvesting season. This pattern of rise and fall in sugar content is duplicated in almost all sugar-producing districts where there is a clearly defined Winter and Summer harvesting season. To approximate this behaviour, Guise and Ryland [8] have used a quadratic function of time of the form:

$$\text{C.C.S.}_t = a_o + a_1 t + a_2 t^2 (a_o, a_1 > 0{\cdot}0,\ a_2 < 0{\cdot}0) \quad . \quad . \quad . \quad (2)$$

where:

C.C.S._t = average C.C.S. in time t, at some plant expressed as a percentage.

t = units of one week with some clearly defined origin.

[5] The preliminary identification procedures above may suggest the inclusion of moving average or autoregressive disturbance terms which require non-linear least squares estimation methods. Alternative algorithms for handling this type of estimation problem have recently been discussed by Goldfeld and Quandt [7].

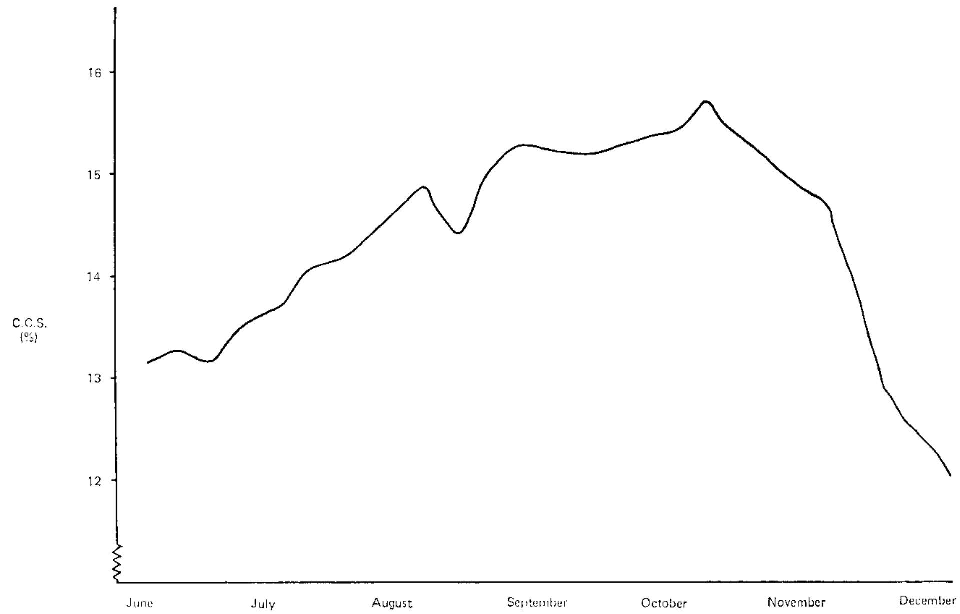

FIGURE 1. *Maturity Cycle of Sugar Cane Central District of Queensland* (1971)

An equation typical of the form of (1) fitted to twenty-four observations of weekly average C.C.S. data for a particular plant in the Central District of Queensland in the 1971 season is:

$$C.C.S._{t} = 11{\cdot}86587 + 0{\cdot}51107t - 0{\cdot}01838t^2 \qquad \ldots \quad (3a)$$

$$(9.80506) \quad (0{\cdot}76645)$$

$$R^2 = 0{\cdot}8221$$

$$d = 0{\cdot}5522$$

$$C.C.S._{t} = 9{\cdot}63952 + 0{\cdot}80130t - 0{\cdot}02721t^2 \qquad \ldots \quad (3b)$$

$$(4{\cdot}39695) \quad (5{\cdot}09854)$$

$$R^2 = 0{\cdot}6610$$

$$d = 1{\cdot}874$$

t = units of one week with origin at 12 noon, Wednesday 16th June, 1971.

In equations (3a) and (3b), the values in parenthesis under each coefficient are "t" values obtained when each coefficient is tested against the zero-valued null hypothesis. Clearly, a non-zero significant partial regression coefficient is hypothesized at least 95 per cent of the time in all cases. The Durbin-Watson statistic, d, detects a significant positive autocorrelation effect in equation (3a) and one available remedy is to transform the original series into a related series by a first order probability difference operation. Equation (3b) was derived using a simple Cochrane-Orcutt transformation scheme.

The behavioural pattern of C.C.S. using equation (3b) provides quite reliable estimates if the degree of explanation R^2 is regarded as the criteria for best fit. However equation (3b) is somewhat of a disappointment when compared with the actual observations. Firstly, equation (3b) consistently underestimates the actual observations recorded towards the end of the season. Secondly, equation (3b) fails by over 2 weeks to predict the turning point of the cycle (maximum value of C.C.S. occurs at $t = 14{\cdot}72$ weeks using equation (3b), while in 1971 week 17 using our arbitrary origin recorded the maximum value of C.C.S.). Both these points serve to cast doubt on the reliability of equation (3b) and the form of equation (2) in adequately reproducing the C.C.S. cycle for the single season.

From the graph of average weekly C.C.S. (figure 1) observations on C.C.S. seem to follow a linear trend between commencement of harvesting up to a certain time period, say K_1. After K_1 is reached the points form a semi-ellipse around the maximum value until a time period, say K_2, where a persistent linear trend downwards can be observed. To obtain a *single* equation continuous at K_1 and K_2 we can, following Fuller [5], estimate an equation which is linear up to K_1, quadratic between K_1 and K_2 and linear after K_2 of the form:

$$\text{C.C.S.}_t = a_0 + a_1 t + b_1 Z_1 + b_2 Z_2 \qquad \text{. . . (4)}$$

where:

$$Z_1 = (t - K_1)^2 \text{ for } K_1 < t < K_2$$

$$Z_2 = (K_1^2 - K_2^2) + 2(K_2 - D_1)t \text{ for } t > K_2.$$

The overall improvement from splicing the original quadratic equation (2) can be gauged by the increase in its explanatory power given by R^2. Also since we are interested in minimizing the error between the actual and estimated observations reflected by the improvement in R^2 a search for the optimum integer value for K_1 and K_2 where R^2 is maximized may also be made.[6] Using the same raw data on C.C.S. as that which was used to estimate equations (3a) and (3b), three variations of the model outlined in equation (4) were analysed. These were a linear followed by a quadratic segment, quadratic then linear, in addition to the completely specified model. For each equation optimal integer join points were determined. The results of this analysis for the 1971 data are given in table 1; of the three variations of equation (4) tested, the completely specified model with $K_1 = 14$ and $K_2 = 22$ was globally optimum having an R^2 of $0{\cdot}9575$. The equation estimated for $K_1 = 14$ and $K_2 = 22$ is as follows:

$$\text{C.C.S.}_t = 13{\cdot}11821 + 0{\cdot}16046t - 0{\cdot}03047Z_1 - 0{\cdot}04086Z_2 \qquad \text{. . . (5)}$$

$$\qquad\qquad (15{\cdot}13349) \quad (8{\cdot}52752) \quad (20{\cdot}15239)$$

$$R^2 = 0{\cdot}9575$$

$$d = 1{\cdot}6886$$

[6] Recently, Gallant and Fuller [6] have developed a non-linear least squares algorithm for computing the optimum (not necessarily integer) join point. The importance of the integer requirement in our case may be gauged by the sensitivity of the sum of squares surface to changes in the level of join points about the computed optimum. From table 1 we may observe that this variation is small indicating a relatively "flat" surface about the optimum.

where:

$Z_1 = (t - 14)^2$ for $14 < t < 22$, and

$Z_2 = -288 + 16t$ for $t > 22$.

The marked improvement in the explanatory power of equation (5) compared with the simple quadratic relationship, equation (3), gives an overall improvement of approximately 30 per cent in explanatory power and this improvement is certainly worth the additional computational effort involved. The improved estimating power of equation (5) may be gauged from the margin of error in predicting the turning point of the C.C.S. cycle. The maximum value of C.C.S. using equation (5) is reached at time period $t = 16{\cdot}63$ weeks which may be compared with $t = 17$ weeks from the raw data. The use of equation (3) rather than equation (5) would tend to overstate revenue per tonne of sugar cane particularly in the latter periods of the harvest if C.C.S. forecasts were to be put to this purpose.

TABLE 1

Results of Search for Best Fitting Spliced Polynomial with Integer Join Points†

Model	K_1	$K_2 = 21$	$K_2 = 22$	$K_2 = 23$	$K_2 > 27$
Linear-Quadratic. . . .	15	....	....	....	0·9478
	16	....	....	....	0·9607
	17	....	....	....	0·9441
Linear-Quadratic-Linear	13	0·9194	0·9574	0·9550	....
	14	0·9250	0·9575*	0·9566	....
	15	0·9310	0·9570	0·9571	....
	16	0·9368	....	0·9552	....
	17	0·9405	....	....	....
	18	0·9385	....	....	....
Quadratic-Linear. . . .	0	0·8766	0·9422	0·9302	....

* Global optimum.

† Improvement in R^2 (adjusted for degrees of freedom) was taken as the criterion for best fit. All equations satisfied the Durbin-Watson test for positive autocorrelation and in each equation all partial regression coefficients were significant.

Before it is possible to use equation (5) as a predictive tool, it is necessary to determine whether the C.C.S. cycle is in fact stable from season to season. We can do this quite simply by extending the spliced Polynomial model, equation (4), to include proxy or shift variables for intercept and slope coefficients for each season compared with the average over all seasons with 1971 set as the base year. These proxy variables gauge the differences up or down which have taken place in the level of these coefficients between seasons. Those years in which the regression coefficients for the intercept and slope parameters were not significant were omitted after the initial equation was formed and the significant estimates only were obtained. These procedures result in the regression coefficients given in table 2.

Two major points may be inferred from table 2. Firstly, there is an apparent significant divergence between seasons in the linear trend effect or translation of the series. *A priori*, one would expect a continual upward trend in vertical quality height reflected in the negative sign of each trend coefficient for each season prior to 1971. This is not the case indicating that seasonal effects are of overriding importance in determining vertical quality height.

Secondly, the linear slope coefficients are only significant for two years out of the previous ten (1963 and 1968). The significance of those coefficients reflect the unusually high incidence of rainfall experienced during the harvesting period in these years. This implies that the slope of the C.C.S. curve in normal seasons is virtually stable while the only non-stationary aspects of the C.C.S. series is the translation or vertical quality height. These conclusions suggest the degree and type of differencing which would be required to induce stationarity. As only the trend is non-stationary, a first difference may be a stationary model.

Before it is possible to operate on the entire series of sugar cane quality it is necessary to standardize the series in terms of the length of the harvesting period. The missing observations were estimated using the best fitting form of the spliced polynomial model, equation (4), applied

TABLE 2

Comparative Statistical Analysis of the C.C.S. Cycle for a Particular Plant in the Central District 1961–71*

Term	Regression coefficient	$\lvert t \rvert$ value (246 observations)
Average Intercept	12·16903	....
Significant Differences—		
1961	1·19903	9·8850
1963	1·22382	4·730
1966	0·51925	4·2566
1968	0·64916	2·7788
1969	−0·51637	4·0259
1970	0·94624	7·00689
Average Linear Slope (t)	0·22233	....
Significant Differences—		
1963	0·08979	4·45424
1968	0·02885	....
Average Quadratic Slope (Z_1)	−0·03768	....
Average Linear Slope (Z_2)	−0·03980	....
		$R^2 = 0·8130$

* No test for significant differences over time in Z_1 and Z_2 were conducted because of the sparse data towards the end of season.

to each seasons data with the origin for each season established at 12 noon on the Wednesday of the second week in June.[7]

A total of 27 time periods was considered each year finishing in each case on the Wednesday of the week prior to Christmas. The series from 1961–71, a total of 297 observations on C.C.S. was regarded as the fit period while the data for the season 1971–3 became the test period for assessing the adequacy of the fitted model as a forecasting system.

Prior to linear filtering, the fit period data were first transformed using natural logarithms as it is the percentage fluctuations in the data which are expected to be comparable.[8]

$$Z_t = \ln . \text{C.C.S.}_t \qquad \ldots \quad (6)$$

where:

Z_t = logged series of crop quality in each period t.

To identify a stationary model for sugar cane quality the procedure is to difference the series and then from the lagged autocorrelation coefficients of the differenced series choose an appropriate model. Figures 2 and 3 supply, respectively the sample autocorrelation and partial autocorrelation functions over 81 lags for the series Z_t, ∇Z_t and $\nabla_{27} Z_t$, as the length of the seasonal frequency in this case is of order 27. The differenced series which extracts the minimum variance component is ∇Z_t. At this level of differencing, the autocorrelation function gives significant coefficients at lags 12, 27 and 54, at the 5 per cent level (standard error of the autocorrelation coefficients is approximately 0·06). On the other hand, the partial autocorrelation function for ∇Z_t has a non-zero coefficient at lag 54 which suggests an autoregressive operator at this period lag. These preliminary identification procedures suggest the following ARIMA model.

$$(1 - \varphi B^{54})\omega_t = (1 - \theta_1 B^{12} - \theta_2 B^{27})a_t \qquad \ldots \quad (7)$$

where:

$\omega_t = (1 - B)Z_t = \nabla Z_t = Z_t - Z_{t-1}$

a_t = uncorrelated stationary stochastic process, and

B^k = lag or shift operator of order k.

In addition to the model suggested in equation (7) (model A) three alternative models were identified (models B, C and D). Model B is simply the unlogged version of model A. While simple linear filtering extracts the minimum variance component it is possible that this level of differencing is too harsh given the inherent seasonal component in the original series. Thus the seasonal differenced series of order 27, ∇_{27}, was used to select models C and D, the logged and unlogged versions of the seasonal differenced models.

---

[7] In all cases the best fitting model was a linear-quadratic-linear model with join points $K_1 = 10, \ldots, 15$ and $K_2 = 22$.

[8] In a recent empirical study by Chatfield and Prothero [3] using the Box-Jenkins' seasonal models, good forecasts were not realized as a result of overtransformation of the original series using logarithms. We compare forecasts using both logged and unlogged data. However no attempt is made to determine the "optimal" transformation which remains somewhat of a controversial matter.

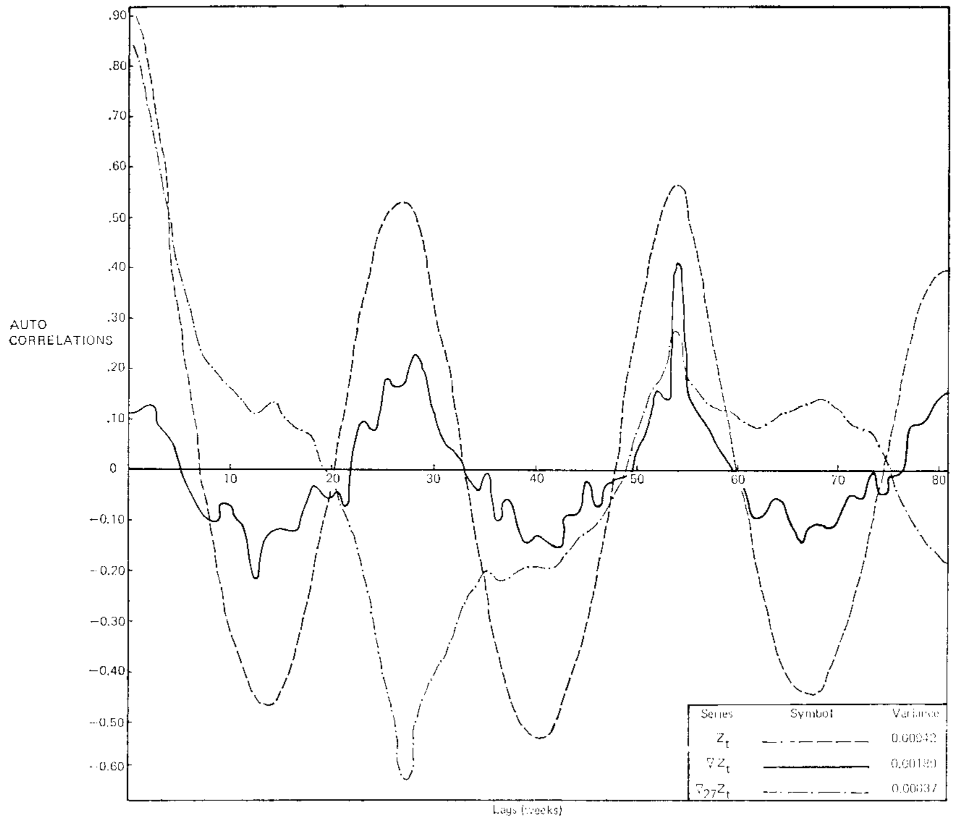

FIGURE 2. *Sample Autocorrelation Functions of Logged C.C.S. Series* (Z_t)

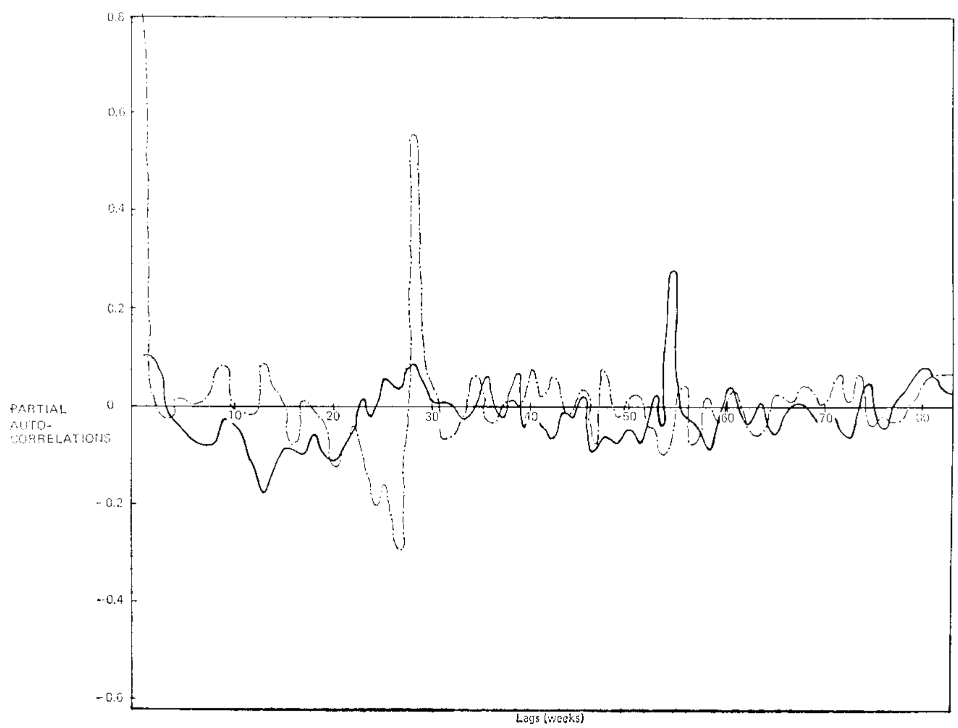

FIGURE 3. *Sample Partial Autocorrelation Function of Logged C.C.S. Series* (Z_t)

The autocorrelation function in figure 2 for the seasonal differenced series shows the autocorrelation functions declining by a factor of approximately 0·8 over a seasonal cycle indicating an autoregressive operator of order 1. In addition a second autoregressive operator of approximately −0·5 at the seasonal frequencies is also evident. The partial autocorrelation function tends to confirm these preliminary identifications with non-zero coefficients at lags 1 and 27. Thus model C selected from the class of seasonal models is an autoregressive (AR) model of the form

$$(1 - \varphi_1 B - \varphi_2 B^{27})\nu_t = a_t \qquad . \; . \; . \; (8)$$

where:

$$\nu_t = (1 - B^{27})Z_t = Z_t - Z_{t-27}$$

Model D was fitted as the unlogged analogue of model C.

As moving average terms were included in models A and B a non-linear least squares method using Hartley's algorithm[9] was used to provide least squares estimates of the parameters of each model. Prior to fitting the models the effect of imposing the linear filter on the original series was tested against the unconstrained autoregressive model. The autoregressive coefficients of $\varphi_1 B$ and $\varphi_2 B^{27}$ were 0·999 and 0·993 for models A and C respectively which confirms the linear homogeneous hypothesis in each case.

The models specified in equations (7) and (8) and fitted to the data are reported in table 3. For each model only the significant estimates are supplied. A seasonal moving average term was included in models C and D after the seasonal model, equation (8), was initially fitted in order to accommodate a significant autocorrelation coefficient occurring at lag 27 in the residual series. The effect of this adjustment was to reduce to insignificance the seasonal autoregressive term.

The models summarized in table 3 were diagnostically checked for significant period patterns (Kolgomorov-Smirnov) and for non-randomness in the residuals (Box-Pierce). Both hypotheses were rejected at the levels of significance commonly assumed. On both these criteria as well as the residual standard error ($\hat{\sigma}_a$) it is not possible to choose among the fitted models a model on which to base a suitable forecasting system. To evaluate the forecasting ability of the models at different lead times we chose as the naive "no change" extrapolation the values of C.C.S. for the corresponding week in the previous season, for example, the denominator of Thiel's coefficient was $A_t - A_{t-27}$.

---

[9] Hartley's algorithm is a modified Newton-type Gradient Method using a quadratic approximation to determine the optimal step length at each iteration. With respect to the latter, Hartley's technique bears a family resemblance to those iterative non-linear least squares methods discussed by Quandt and Goldfeld [7]. A practical computational problem which we have found with this technique is singularity of the matrix which, of course, inhibits further iteration. One way to overcome this problem is by the use of a more robust matrix inversion technique.

TABLE 3

*Summary of Box-Jenkins Seasonal Models Fitted to Sugar Cane Quality Series**

Model	No. of observ- ations	Para- meters	Coefficients	$\mid t \mid$ values	Q	$\hat{\sigma}_a$
A $\omega_t + \varphi\omega_{t-54} + a_t + \theta a_{t-12}$	242	φ θ	0·46985 −0·11721	8·143 1·981	37·20	0·040
B $\omega_t^1 = \varphi\omega^1_{t-54} + a_t + \theta a_{t-12}$	242	φ θ	0·44960 −0·14221	7·755 2·201	40·76	0·530
C $\nu_t = \varphi\nu_{t-1} + a_t + \theta a_{t-27}$	269	φ θ	0·81703 −0·84159	22·903 23·537	47·73	0·039
D $\nu_t^1 = \varphi\nu^1_{t-1} + a_t + \theta a_{t-27}$	269	φ θ	0·83449 −0·82520	24·497 22·113	47·15	0·525

* $\omega_t = (1 - B)Z_t$, $\omega_t^1 = (1 - B)$ C.C.S.$_t$,
$\nu_t = (1 - B^{27})Z_t$, $\nu_t^1 = (1 - B^{27})$ C.C.S.$_t$.

Over the test period (1971–3) the one step ahead forecasts for each model gave similar "*U*" coefficients (table 4) and hence are the same in terms of forecasting accuracy. Thus for one step ahead forecasts and for adaptively revising intraseasonal forecasts either form of model is suitable. For interseasonal forecasts, however, the first differenced models are clearly inferior relative to the seasonal differenced models as models for adequately reproducing the seasonal series. Thiel's coefficient for the seasonal differenced models are approximately half those of the first differenced models at the extended lead times but are nearly three times that of the coefficient calculated for the lead time of one period. The sensitivity of forecasting performance to changes in model structure over different lead times, which we have shown, suggests that a model which may prove adequate for forecasting one period ahead may be clearly inferior at extended lead times. The culprit in this test case was first differencing which, although extracting the minimum variance component of the original series, dominated the inherent seasonal frequency in the model.

TABLE 4

Thiel's Coefficient Calculated at Different Lead Times over the Test Period 1971–1973

Model	Lead times (weeks)		
	1	27	54
A	0·28	1·61	1·62
B	0·25	1·62	1·63
C	0·30	0·84	0·85
D	0·29	0·84	0·85

5 CONCLUDING DISCUSSION

The forecasting techniques which we have used to forecast biological crop quality have also been applied to forecasting crop yields [13] but with limited success. This is probably due to the less variability associated with the quality attribute series compared with yields and the aggregative nature of the study with which the latter was primarily concerned. A seasonal model which we have fitted to the crop quality series would also probably produce adequate yield forecasts. Unfortunately in many rural industries, yields are subject to regulatory controls and this factor would not permit a proper evaluation of the seasonal component of the yields only series.

Our specific conclusions as they relate to sugar cane quality forecasting are that the fitting of an ARIMA model with or without seasonal differencing (logged or unlogged) performs quite adequately for making routine forecasts confined within a single harvest as indicated by Thiel's coefficient. For interseasonal forecasts a model differenced at seasonal frequencies is required. An autoregressive-moving average model selected from the class of seasonal models reduces the mean square forecasting error by approximately 15 per cent relative to a "no change" seasonal extrapolation.

The procedures for making reliable projections of crop quality have been discussed using aggregate crop quality data generated over time at the regional level. Consequently there is a need to refine the technique to include control variables relevant to the individual forecaster. When control variables are explicitly included in the forecasting system it is then possible to predetermine the responsiveness of quality improvement to changes in the level of the control parameters. Economic optimal control : quality improvement relationships can then be determined.

REFERENCES

[1] Box, G. E. P. and G. M. Jenkins, *Time Series Analysis—forecasting and control*, (Englewood Cliffs, New Jersey, Holden-Day, 1970).

[2] Cowling, K. and A. J. Rayner, "Price, Quality and Market Share", *Journal of Political Economy*, Volume 78, No. 6, (Nov.-Dec., 1970), pp. 1292–1309.

[3] Chatfield, C. and D. L. Prothero, "Box-Jenkins Seasonal Forecasting: Problems in a Case Study", *Journal of the Royal Statistical Society*, Series A, Volume 136, part 3 (1973), pp. 295–352.

[4] Doran, H. E. and J. J. Quilkey, "Harmonic Analysis of Seasonal Data: Some Important Properties", *American Journal of Agricultural Economics*, Volume 54, No. 4, (Nov., 1972), pp. 646–651.

[5] Fuller, W. A., "Grafted Polynomials as Approximating Functions", *Australian Journal of Agricultural Economics*, Volume 13, No. 1 (June, 1969), pp. 35–46.

[6] Gallant, A. L. R. and W. A. Fuller, "Fitting Segmented Polynomial Regression Models Whose Join Points have to be Estimated", *Journal of the American Statistical Association*, Volume 68, No. 341 (1974), pp. 144–147.

[7] Goldfeld, S. M. and R. E. Quandt, *Nonlinear Methods in Econometrics*, (Amsterdam: North Holland, 1972).

[8] Guise, J. W. B. and G. J. Ryland, "Production Scheduling and Allocation: A Normative Decision Model for Sugar Milling", *Australian Journal of Agricultural Economics*, Volume 13, No. 1 (June, 1969), pp. 8–24.

[9] Hannan, E. J., "The Estimation of Seasonal Variation", *Australian Journal of Statistics*, Volume 2, (1960), pp. 1–15.

[10] Maddala, G. S., "The Use of Variance Components in Pooling Cross-Section and Time Series Data", *Econometrica*, Volume 39, (March, 1971), pp. 341–358.

[11] Matsumoto, M. and Ben C. French, "Empirical Determination of Optimum Quality-Mix", *Agricultural Economics Research*, Volume 23, (Jan., 1971), pp. 1–10.

[12] Ryland, G. J. and J. W. B. Guise, "A Spatio-temporal Quality Equilibrium Model of the Australian Sugar-Cane Processing Industry", *American Journal of Agricultural Economics* (article scheduled for August, 1975).

[13] Schmitz, A. and D. G. Watts, "Forecasting Wheat Yields: An Application of Parametric Time Series Modelling", *American Journal of Agricultural Economics*, Volume 51 (1969), pp. 247–254.

[14] Spivey, W. and W. E. Wecker, "Regional Economic Forecasting", *Papers of the Regional Science Association*, Volume 28, (1970), pp. 257–276.

[15] Terrel, R. D. and N. E. Tuckwell, "The Efficiency of Least Squares in Estimating a Stable Seasonal Pattern", *Journal of the American Statistical Association*, Volume 66, (June, 1971), pp. 354–362.