Composite Forecasting: some empirical results using BAE short-term forecasts

L. O. Jolly and G. Wong*

The contention advanced in this paper is that forecast performance could be improved if short-term commodity forecasters were to consider formally the use of a variety of forecasting methods, rather than seeking to improve one selected method. Many researchers have demonstrated that a linear combination of forecasts can produce a composite superior to the individual component forecasts. Using a case study of two Bureau of Agricultural Economics' forecast series and alternative, time series model forecasts of the same series, four methods of deriving composite forecasts are applied on an *ex ante* basis and are thus evaluated as a means of improving the Bureau's forecast performance. Despite the fact that the authors could not, by combining the available forecasts, form a superior composite forecast, the application highlights the suitability of this approach for reviewing the performance of forecasting methods on a formal basis, and did prove useful in exposing weaknesses and strengths in BAE market information forecasts which otherwise would not have come to light.

1. Introduction

Forecasting the short-term outlook for Australian agriculture is a major activity of the Bureau of Agricultural Economics (BAE). The mechanism for generating BAE forecasts varies from commodity to commodity. Although the BAE uses econometric and simple time series models for some commodity forecasts, the opinions of expert commodity analysts also heavily influence the final forecast.

Freebairn (1975) was the first to assess the level of accuracy of BAE forecasts. Using a non-parametric sign test, Freebairn found that BAE forecasts of production and prices were significantly more accurate than mechanical forecasts generated from a second-order auto-regressive model. A common approach to assessing the performance of a forecasting system is to compare its forecasts of various measures with those produced by other, mechanical forecasting procedures. Lee and Bui-Lan (1982) argued that this method falls short of pinpointing the possible weaknesses of a forecasting system. They showed that, for this purpose, past predictive errors can be used to reveal systematic biases. They developed a simple technique to correct for bias (thereby improving forecast accuracy) using BAE forecasts of production and prices in the Australian wool market.

Composite forecasting methods have been studied for more than a decade. Bates and Granger (1969) noted that, faced with alternative forecasts, analysts tend to try to discover which is the best and then discard the others. Bates and Granger suggested that, if the objective is to make as good a forecast as possible, alternative forecasts should not be discarded because they nearly always contain some useful independent information. Many researchers have applied a variety of approaches to constructing composite forecasts from two or more component forecasts. Using empirical analyses, researchers have demonstrated that composite forecasting techniques can produce forecasts which are superior to the individual forecasts. Thus, any forecasting effort should seriously consider using such methods.

*Bureau of Agricultural Economics, Canberra. This paper was presented at the 30th Annual Conference of the Australian Agricultural Economics Society. The authors gratefully acknowledge comments given by BAE colleagues, *Review* referees and editors.

The purpose in this paper is to explore the possibility of combining the BAE's market information forecasts with alternative, time series model forecasts as a means of improving forecasting performance. Many commodity analysts already combine forecasts in a subjective manner. Quantitative forecasts derived from econometric models (or simple time trend models) are very often incorporated into final subjective assessments. A formal "marriage" between the different forecasting methods could well lead to an improvement in forecasting performance.

The remainder of this paper is organized in six parts. First the current BAE approach to forecasting is discussed. Second, the various criteria available for assessing forecasts are presented, giving a justification for selecting mean square error measures. Third, the basic concepts underlying composite forecasting are reviewed. Fourth, results of the alternative ARIMA (autoregressive integrated moving-average) and naive "no change" forecasts are presented and evaluated for two commodities—citrus and sugar cane. Fifth, the composite forecasts are derived, presented and evaluated. On the basis of the results, some broad conclusions are reached concerning the usefulness of combining procedures for short-term commodity forecasting purposes.

2. The BAE Forecasting Approach

BAE "first" commodity forecasts (that is, the first forecasts issued by the BAE for the coming financial year) for prices, production and gross value of production can be regarded as "market information forecasts". Some forecasts are made subjectively, based on knowledge of technical and economic factors influencing the production, consumption and trade of the commodity. They are significantly affected by discussions with individuals involved in the trade and with other interested parties. Although some forecasts are based on econometric or simple time series models, the commodity analyst frequently adjusts such forecasts in the light of discussions with industry contacts. This study will be concerned with first forecasts issued by the BAE for the coming financial year. Forecasts are revised on a quarterly basis, and it would be expected that in later quarters, as more market information becomes available, less weight will be given to econometric and time series model forecasts. Since we intend to generate alternative time series model forecasts and to evaluate BAE, time series and composite forecasts, it is necessary to establish some criteria for evaluation.

3. Assessing Forecasting Accuracy

Longmire and Watts (1981) argued that, although accuracy is a key measure of forecast quality, other characteristics such as timeliness, clarity, relevance and credibility must also be taken into account. However, accuracy is by far the most useful criterion for comparative purposes, because it can be measured.

In any forecasting situation, for every forecast error there must be an associated cost of error. In general, a specific cost function must be chosen, though there are many cost functions from which to choose. A common approach has been to select that functional form of the cost function which is most likely to reflect the way in which the forecast error will affect costs. The most popular with researchers has been the quadratic cost-of-error function. The justification for its use often relies on the arguments that it is, *a priori*, not an unreasonable assumption, it is mathematically more tractable than any alternative, and it bears an obvious relationship to the least squares criterion generally used in estimating forecasting models (Granger and Newbold 1977).

Bordley (1982) illustrated theoretically how using a different cost function can lead to a different assessment of forecast accuracy. However, noting Granger's (1969) study concerning the approximation of any cost function by a squared error form, the criterion used to evaluate forecasting accuracy in this study will be based on mean squared error (MSE). MSE is defined as:

$$(1) \qquad MSE\,(F_i) = \frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2$$

where
A_i is the actual value,
F_i is the forecast value, and
n is the number of observations.

A more meaningful indicator of the performance of a forecasting system than the absolute magnitude of MSE is a comparison of the MSE of a forecasting system with the MSE of a simple benchmark, such as the naive no-change forecast. Theil's U_2 statistic, which implicity incorporates an alternative forecast of no change, is defined as:[1]

(2)
$$U_2 = \frac{MSE(F_i)}{MSE(A_{i-1})} = \frac{\sum_{i=2}^{n} (A_i - F_i)^2}{\sum_{i=2}^{n} (A_i - A_{i-1})^2}$$

Leuthold (1975) advocated the use of U_2 as a minimum measure of forecasting accuracy. The U_2 statistic has the following two properties. It is bounded below by zero, this being the case of perfect forecasting (that is, no forecast error). It has no upper limit, but takes the value of unity when the prediction method produces the same mean squared error as the naive no-change forecast. Clearly, to know that BAE forecasts were less accurate than a simple no-change forecast would tell us significantly more than knowing the magnitude of the MSE. For this reason, Theil's U_2 statistic is used in this study.

4. Techniques for Combining Forecasts

Combining forecasts requires choosing a set of weights for the individual forecasts. A number of alternatives are available for deriving these weights. First a simple average of the different forecasts can very easily be derived:

(3)
$$C_n = (F_n^{(1)} + F_n^{(2)} + \dots + F_n^{(p)})/p$$

where C_n is the composite forecast for period n, $F_n^{(i)}$ is the forecast of a particular random variable z_n for the i-th method, and p is the number of forecasts included in the combination (Makridakis and Winkler 1983).

The above approach takes no account of any information concerning the relative accuracy of the individual methods. Bates and Granger (1969), Newbold and Granger (1974), Granger and Newbold (1975, 1977), Dickinson (1973, 1975) and Granger and Ramanathan (1984) developed procedures which take account of such information.

Bates and Granger (1969) considered the following problem: let $f_n^{(1)}$ and $f_n^{(2)}$ be two unbiased forecasts of z_n. A combined forecast can be derived by taking a weighted average of the two individual forecasts: that is,

(4)
$$C_n = WF_n^{(1)} + (1-W)F_n^{(2)}$$

The problem is to find the optimum value of W (which is initially not known) that minimizes the variance of C_n.

Bates and Granger adopted two principles: first, most weight should be given to those forecasts which have performed best in the recent past; and second, the weights should be variable, to allow for the possibility of change in the relationship between forecasting procedures. For an *ex ante* study such as this (that is, one based on observations

1. Both MSE and U_2 statistics are the true statistics therefore there are no distributions associated with MSE and U_2.

preceding the forecast period), such adaptive weighting procedures are difficult to apply because it is only in hindsight that the particular error history which gives the lowest MSE over the evaluation period can be gauged.

Dickinson (1973), Newbold and Granger (1974) and Granger and Newbold (1977) extended the analysis to the combining of more than two forecasts. The special two forecast linear combination of equation (2) can be generalized to:

$$C_n = W'F \tag{5}$$

where F is the $P \times 1$ vector of P unbiased forecasts and W the $P \times 1$ vector of linear weights. W is determined as:

$$W = S^{-1} e(e'S^{-1}e)^{-1} \tag{6}$$

where S is the $P \times P$ covariance matrix of forecast errors and e is a $P \times 1$ unit vector. That is, the composite error variance from combining p forecasts will be minimized subject to the constraint that the weights are convex if the vector of weights W is determined according to equation (4).

Forming a composite forecast using the weights given by equation (4) allows greater weight to be assigned to individual forecast methods which have performed better over the entire historical period.

Dickinson (1973, 1975) had some reservations about the practicalities of the combining process. From an analysis of the sampling distributions of the weights and of the error variance of the combined forecast, Dickinson concluded that, although the combining of forecasts seems very attractive from a theoretical viewpoint, in a practical situation it is difficult to attach an accurate value to the error variance of the combination because errors in the estimation of the covariance matrix can lead to unreliable estimates of the combining weights.

Granger and Ramanathan (1984) and Nelson (1972) noted that the constraint that the weights total to unity is acceptable in the case of unbiased forecasts. Granger and Ramanathan showed (by considering changes in the formula for minimum sum of squared forecast errors) that the alternative of unconstrained weights, as in equation (7), can lead to suboptimal composite forecasts if one of the component forecasts is biased:

$$C_n = W_1F_n^{(1)} + W_2F_n^{(2)}. \tag{7}$$

They showed how the addition of a constant term to the combining equation can produce unbiased composite forecasts, even if the component forecasts are biased. That is:

$$C_n = W_1F_n^{(1)} + W_2F_n^{(2)} + W_3M \tag{8}$$

where M is the mean of the variable that is being forecast.

Not suprisingly, the emphasis in the literature has tended to be on finding robust procedures which can give good composite forecasts over small samples.

The difficulties with the minimum variance criterion applied by such authors as Bates and Granger, Dickinson, and Newbold and Granger centre on obtaining reliable estimates of the full error covariance matrix. The general solution of the above-mentioned authors, where equation (6) is used under the assumption of independent forecast errors, has been criticized as too *ad hoc* by several authors (*e.g.* Bunn 1975 and Bordley 1982), and has led to the development of a Bayesian methodology for the combining of forecasts. This follows the work of Harrison and Stevens (1971) who applied Bayesian theory to the forecasting of time series. To date, however, this method has received little attention in the combining literature compared with the minimum variance criterion, and would be more appropriately examined in another paper.

The different techniques identified above have been applied in a variety of studies including:

(a) combining different time series models (Bates and Granger 1969; Newbold and Granger 1974; Granger and Newbold 1977; Spriggs 1981; and Makridakis and Winkler 1983);

(b) combining econometric and time series models (Nelson 1972; Cooper and Nelson 1975; Granger and Newbold 1975; Longbottom and Holly 1985); and

(c) combining econometric, time series and expert opinion forecasts (Bessler and Brandt 1981; Granger and Ramanathan 1984).

The results of such studies have been generally favourable, particularly in the *ex post* calculation and application of the combining weights, where constant optimal weights are determined after noting all the forecast errors and are then applied over the same sample period to form the composite forecast. In *ex ante* studies the results of composite forecasts prove to be less favourable and much more varied.

The following linear combining methods are tested:

(a) Naive equal weighting:

$$C_n = (F_n^{(1)} + F_n^{(2)} + \ldots + F_n^{(p)})/p \tag{9}$$

(b) Constrained model:

$$C_n = W_1F_n^{(1)} + W_2F_n^{(2)} + \ldots + W_pF_n^{(p)} \tag{10}$$

where $W_1 + W_2 + \ldots + W_p = 1$,

the weights W being estimated by the regression:

$$(A_T - F_T^{(p)}) = W_1(F_T^{(1)} - F_T^{(p)}) + \ldots + W_{p-1}(F_T^{p-1} - F_T^{(p)}) + U_T \tag{11}$$

where
$W_p = 1 - (W_1 + W_2 + \ldots + W_{p-1})$;
A_T is the actual series; and
$T = (1, \ldots n-1)$.

That is, only data available prior to the period to be forecast (n) are used.

(c) Unconstrained combination:

The combining equation is the same as in (b) with weights estimated by the regression:

$$A_T = W_1F_T^{(1)} + \ldots + W_pF_T^{(p)} + U_T \tag{12}$$

$T = (1, \ldots, n-1)$.

(d) Unconstrained combination with constant term:

$$(13) \qquad C_n = W_1 F_n^{(1)} + \ldots + W_p F_n^{(p)} + W_{p+1} M$$

where M is the unconditional mean of the actual series, and weights are estimated by the regression:

$$(14) \qquad A_T = (\hat{\alpha}) + W_1 F_T^{(1)} + \ldots + W_p F_T^{(p)} + u_T$$

$T = (1, \ldots, n-1)$.

5. Methodology and Data

Two time series model forecasting methods have been selected for analysis, comparison and eventual combination with BAE first forecasts. The first method uses Box-Jenkins methodology (Box and Jenkins 1970) for the identification and estimation of an autoregressive integrated moving-average (ARIMA) process; once diagnostic checking has indicated that the ARIMA representation is satisfactory, the estimation process is used for forecasting. The second method is the naive no-change method, where this year's forecast is last year's actual observation.

All forecasts (individual and composite) are *ex ante*—that is, only data available prior to the period to be forecast are used; this applies also to the combining weights, as will be discussed subsequently.

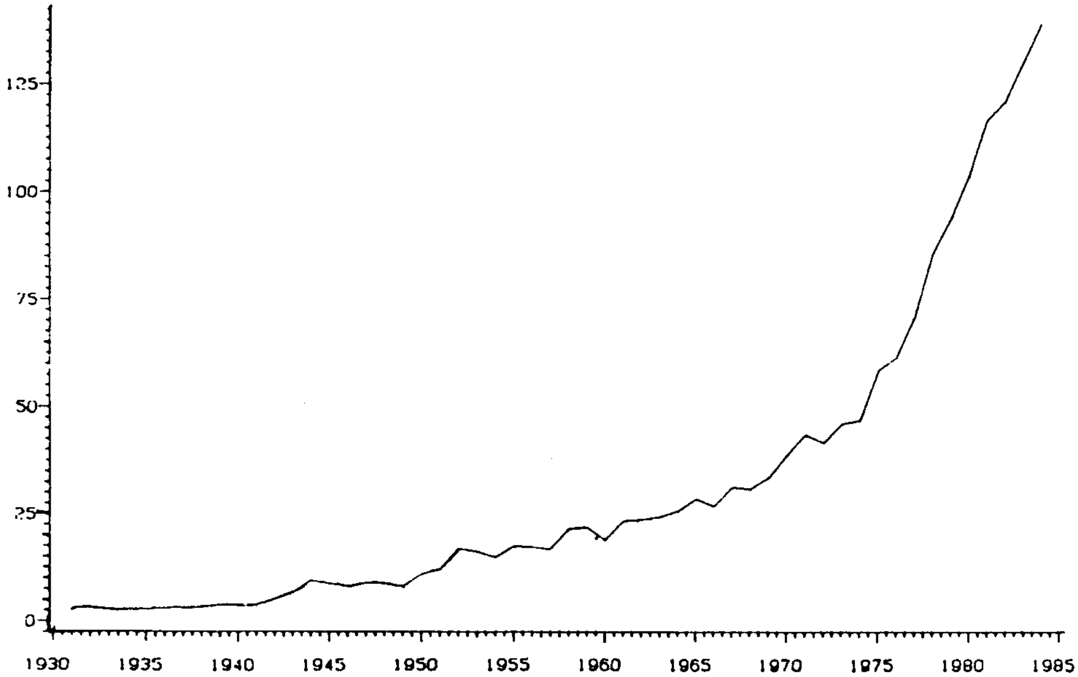
The intention in this paper is to illustrate the application of the above general methodology by using the series for the nominal gross value of production of citrus and sugar cane. Forecasts one year ahead were generated year by year from 1970–71 to 1983–84 inclusive. The starting date for each series was chosen to allow sufficient observations for the identification of an ARIMA process; 1930–31 to 1969–70 and 1935–36 to 1969–70 for citrus and sugar respectively.

Gross value series have been used because these are a primary input used by the BAE in deriving its aggregate measures of rural sector performance. The citrus and sugar cane series, besides being data with which the authors are familiar, have the advantage of displaying inherently differing data characteristics. For citrus the data series is "well-behaved", making it amenable to time series modelling (see Figure 1), whereas the series for sugar cane is inherently more variable (see Figure 2). The two series were chosen also because of historically different performance of their BAE first forecasts. For citrus, BAE first forecasts have been less accurate than naive no-change forecasts, whereas those for sugar cane are much superior to the latter (see Tables 3 and 4 below). These differences could be expected to lead to very different results in time series modelling and the combining phase of the exercise.

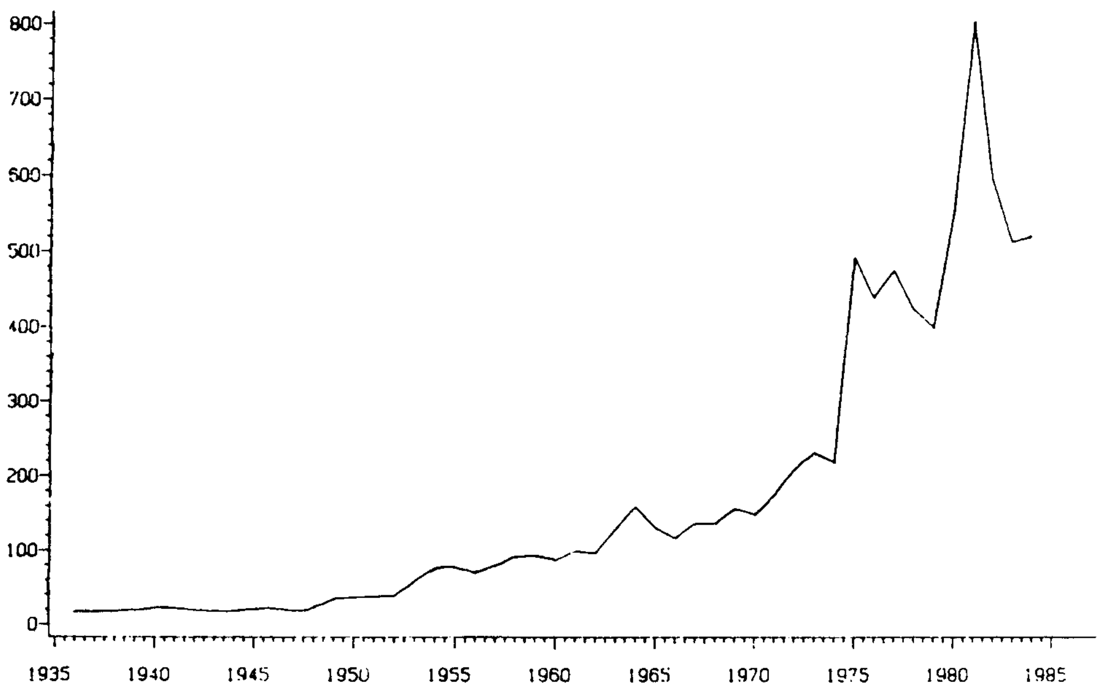
One of the difficult parts of Box-Jenkins methodology is to identify an appropriate model. This is the first step which has to be taken on order to produce the ARIMA forecasts.

The first data series under investigation is the annual gross value of production of citrus. Graphs of the raw data, the first-15 autocorrelation function (ACF,r(ε)) and the partial autocorrelation function (PACF) from 1930–31 to 1969–70 are presented, respectively, in Figures 1, 3 (a) and 3 (b).[2] The raw data show a significant upward trend, and the ACF and PACF also suggest that the data are not stationary. The ACF and PACF after first-differencing are presented in Figures 3 (c)

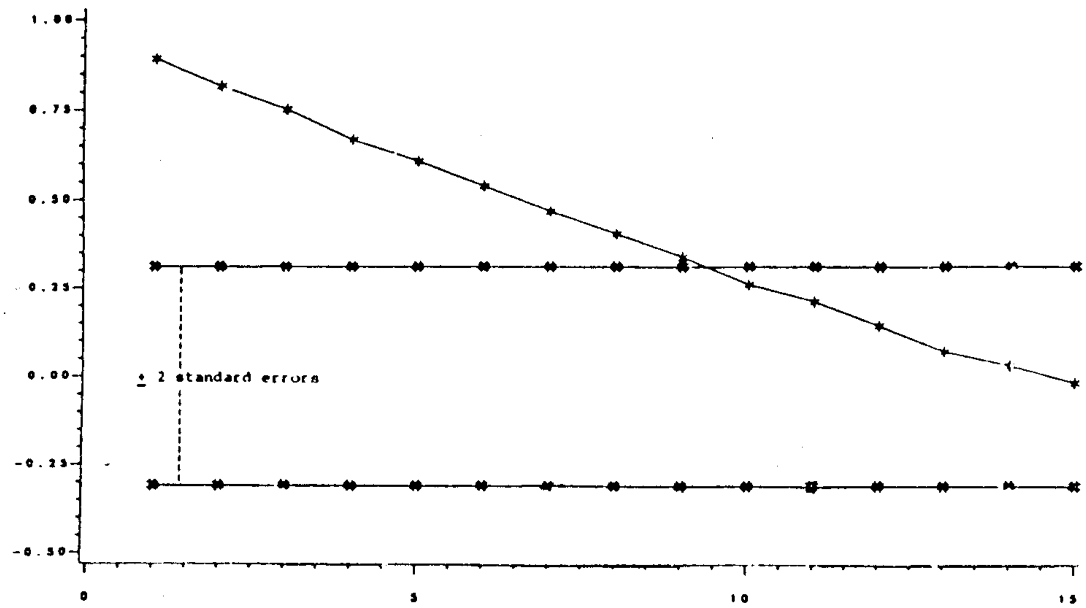
[2] For definition of ACF and PACF see Box and Jenkins (1970).



X–Axis: Financial Year
Y–Axis: $ (million)
Figure 1: Citrus—Nominal Gross Value of Production



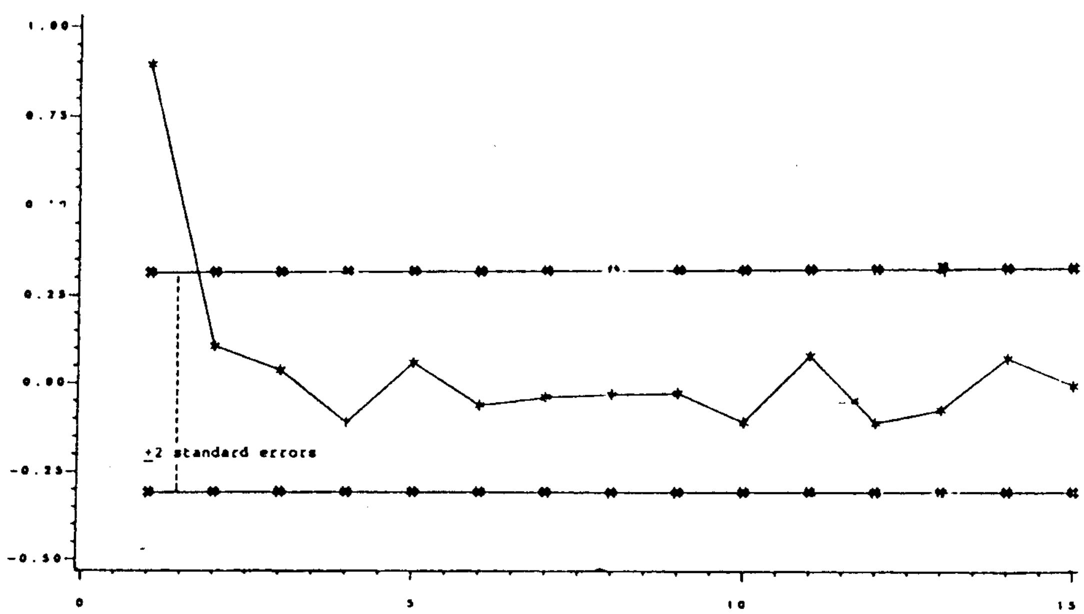
X–Axis: Financial Year
Y–Axis: $ (million)
Figure 2: Sugar—Nominal Gross Value of Production



X–Axis: Lags

Y–Axis: Values of ACF of raw data

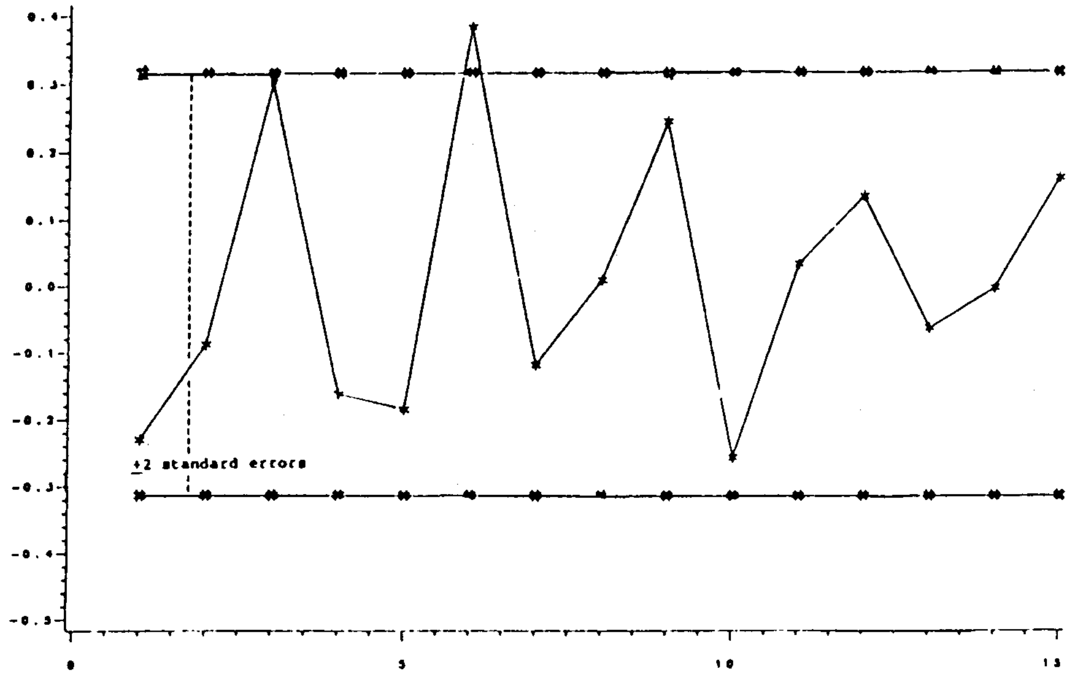
Figure 3 (a): Autocorrelation Function of Raw Data for Citrus



X–Axis: Lags

Y–Axis: Values of PACF of raw data

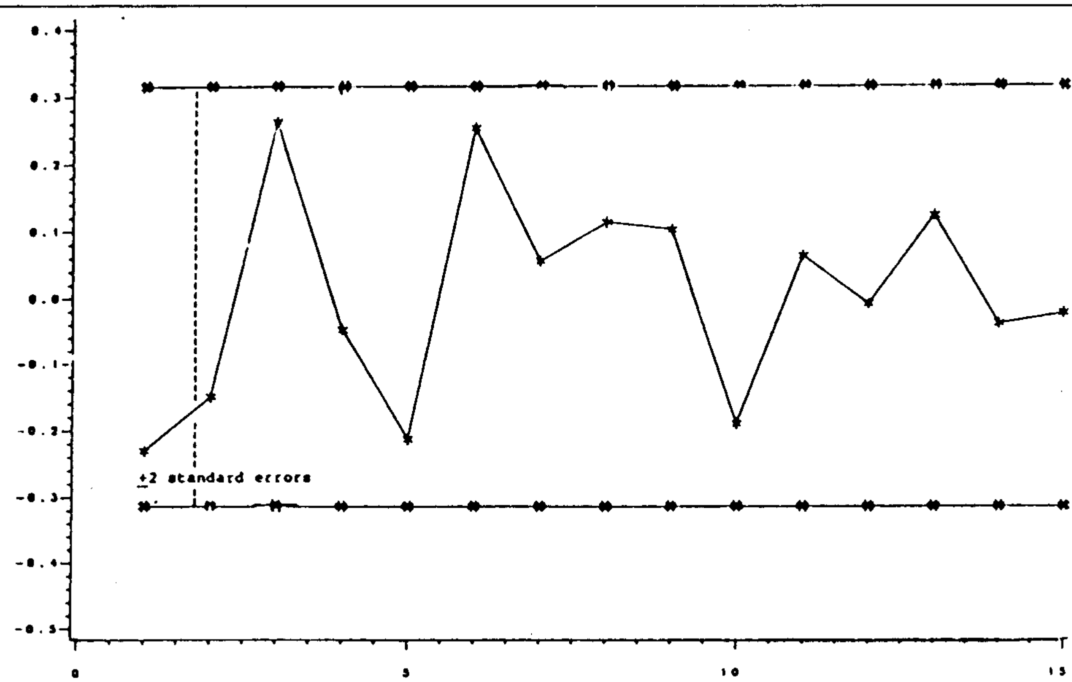
Figure 3 (b): Partial Autocorrelation Function of Raw Data for Citrus



X–Axis: Lags

Y–Axis: Values of ACF of First Differenced Data for Citrus

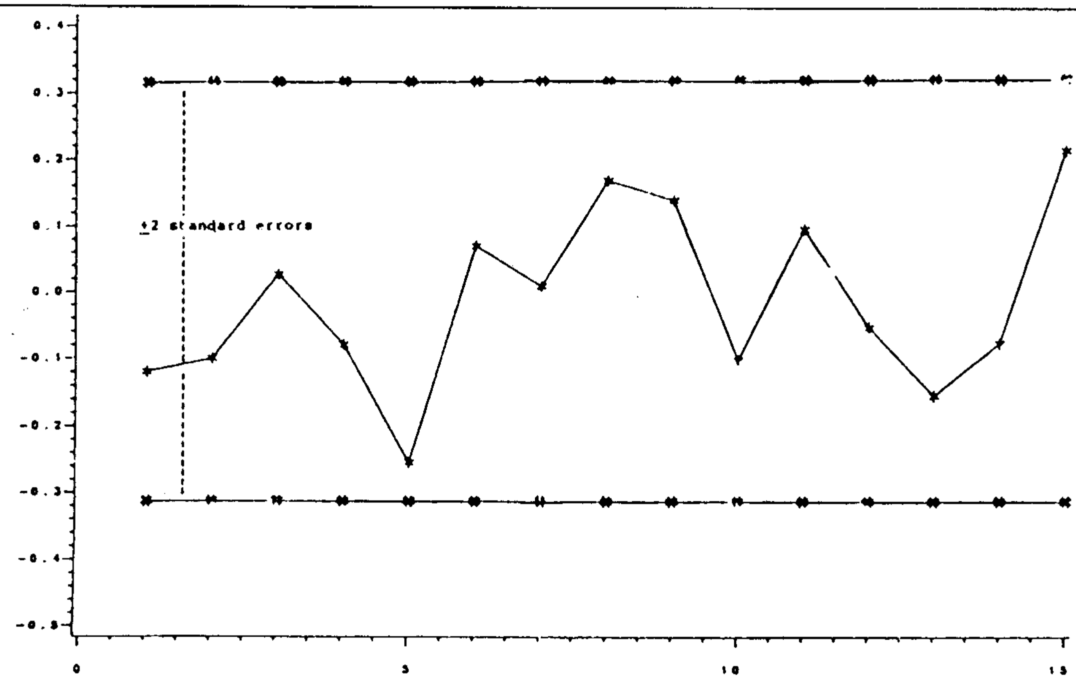
Figure 3 (c): Autocorrelation Function of First Differenced Data for Citrus



X–Axis: Lags

Y–Axis: Values of PACF of First Differenced Data for Citrus

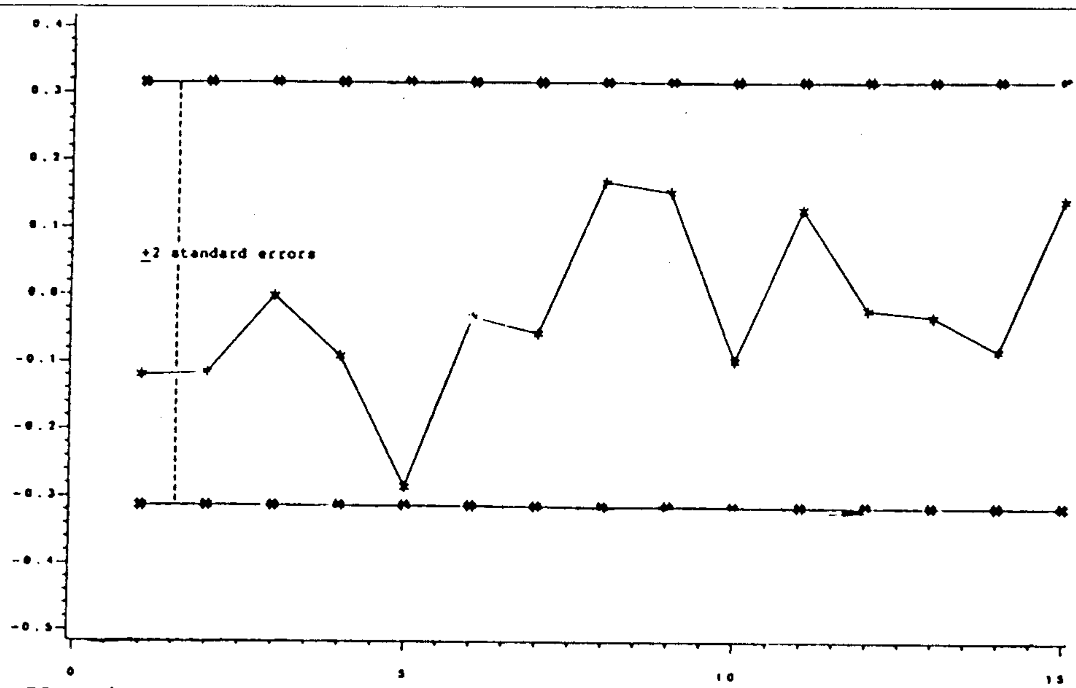
Figure 3 (d): Partial Autocorrelation Function of First Differenced Data for Citrus



X–Axis: Lags

Y–Axis: Value of ACF of First Differenced Data with MA Filter

Figure 3 (e): Autocorrelation Function of First Differenced Data with Moving Average Filter



X–Axis: Lags

Y–Axis: Values of PACF of First Differenced Data with MA Filter

Figure 3 (f): Partial Autocorrelation Function of First Differenced Data with Moving Average Filter.

and 3 (d). From the ACF, notice that $r_3(\hat{\varepsilon})$ and $r_6(\hat{\varepsilon})$ are quite close to the boundary of the confidence interval. In addition, from the Ljung-Box Q statistic,[3]

$$(15) \quad Q = n(n+2) \sum_{k=1}^{j} (n-k)^{-1} r_k^2 (\hat{\varepsilon})$$
$$= 27.02$$
$$\chi^2(0.95,15) = 25,$$

where

d = degree of differencing
n = $N - d$
N = number of observations
$r^2(\hat{\varepsilon})$ = the square of the ACF of residuals
j = number of relevant lags in the test (in this case $j = 15$)

the null hypnothesis of white-noise residuals is rejected at the 5 per cent level. In this situation a moving average model of the following form is tentatively fitted:

$$(16) \quad (1-B)Y_t = (1-\theta_3 B^3 - \theta_6 B^6)\varepsilon_t$$

with $\theta_3 = -0.21$ $\theta_6 = -0.81$
(−2.0) (−10.8)

where B is the backward operator (*i.e.* $(1\text{-}B)Y_t = Y_t - Y_{t-1}$ or $(1 - \theta_3 B^3)\,\varepsilon_t = \varepsilon_t - \theta_3\,\varepsilon_{t-3}$ etc.) and θ_i are the moving average parameters.

The resulting ACF and PACF are shown in Figures 3 (e) and 3 (f). The Q statistic,

$$(17) \quad Q = 13.12$$

$$\chi^2\,(0.95,\ 13) = 22.36,$$

implies that the null hypothesis of white noise residuals is not rejected at the 5 per cent level.

In order to calculate the combining weights, 14 one-step-ahead forecasts are needed. Each time, the ARIMA, model is re-identified and re-estimated, and the Q-statistic is used to test the adequacy of the model. The results are presented in Table 1.

Turning to the series for gross value of sugar production, examination of the raw data, ACF and PACF and the Q-statistic, indicates that the series is not stationary. After first-differencing the data, a moving-average model of the form:

$$(18) \quad (1-B)Y_t = (1-\theta_{10}B^{10})\varepsilon_t$$

with $\theta_{10} = -0.649$
(−5.29)

is fitted to the data. The resulting Q statistic is:

$$(19) \quad Q = 18.25$$

$$\chi^2\,(0.95,\ 14) = 23.68$$

implies that the null hypnothesis of white noise residuals is not rejected at the 5 per cent level.[4]

Following the same procedure as for the citrus series, 14 one-step-ahead forecasts are calculated; the results are presented in Table 2.

[3] See Ljung-Box.

[4] Statistical evidence will be provided upon request.

TABLE 1
Estimated ARIMA Models and One-Year-Ahead Forecasts: Citrus

Time period	The model	Q-sum 15 lags	χ^2 (0.95)	Forecast
1930-31 to 1969-70	$\Delta Y_t=(1-0.21B^3-0.81B^6)\epsilon_t$ (-2) (-10.8)	13.12	22.36	39.69
1930-31 to 1970-71	$\Delta Y_t=(1-0.795B^6)\epsilon_t$ (-9.4)	16.36	23.68	42.77
1930-31 to 1971-72	$\Delta Y_t=(1-0.803B^6)\epsilon_t$ (-8.8)	20.06	23.68	43.86
1930-31 to 1972-73	$\Delta Y_t=(1-0.807B^6)\epsilon_t$ (-10.3)	20.23	23.68	44.92
1930-31 to 1973-74	$\Delta Y_t=(1-0.804B^6)\epsilon_t$ (-9.6)	21.29	23.68	47.66
1930-31 to 1974-75	$\Delta Y_t=(1-0.755B^6)\epsilon_t$ (-4.7)	6.06	23.68	61.68
1930-31 to 1975-76	$\Delta Y_t=(1-0.747B^6)\epsilon_t$ (-5.0)	6.84	23.68	63.56
1930-31 to 1976-77	$\Delta Y_t=(1-0.777B^6)\epsilon_t$ (-7.9)	8.94	23.68	69.33
1930-31 to 1977-78	$(1+0.508B^2+0.582B^3)\Delta Y_t=\epsilon_t$ (4.4) (4.9)	8.11	22.36	91.47
1930-31 to 1978-79	$(1+0.523B^2+0.569B^3)\Delta Y_t=\epsilon_t$ (4.8) (5.0)	8.07	22.36	105.88
1930-31 to 1979-80	$(1+0.499B^2+0.579B^3)\Delta Y_t=\epsilon_t$ (4.6) (4.9)	12.78	22.36	115.59
1930-31 to 1980-81	$(1+0.47B^2+0.587B^3)\Delta Y_t$ (4.4) (5.5) $=(1+0.254B^5)\epsilon_t$ (1.5)	8.61	21.03	124.83
1930-31 to 1981-82	$(1+0.433B^2+0.618B^3)\Delta Y_t$ (3.9) (5.6) $=(1+0.255B^5)\epsilon_t$ (1.7)	10.06	21.03	130.15
1930-31 to 1982-83	$(1+0.431B^2+0.616B^3)\Delta Y_t$ (4.0) (5.6) $=(1+0.3B^5)\epsilon_t$ (2.0)	9.15	21.03	139.04

Note: Figures in parentheses are t ratios.

TABLE 2

Estimated ARIMA Models and One Year Ahead Forecasts: Sugar Cane

Time period	The model	Q-sum 15 lags	χ^2 (0.95)	Forecast
1935-36 to 1969-70	$\Delta Y_t = (1-0.6494B^{10})\varepsilon_t$ (-5.29)	18.25	23.68	155.275
1935-36 to 1970-71	$(1-0.297B^2)\Delta Y_t = (1-0.721B^6)\varepsilon_t$ (-1.26) (-2.48)	12.85	23.36	178.95
1935-36 to 1971-72	$(1-0.463B^2)\Delta Y_t=(1.0758B^6)\varepsilon_t$ (-2.08) (-5.76)	14.67	22.36	195.15
1935-36 to 1972-73	$\Delta Y_t=(1+0.37B^2-0.676B^6)\varepsilon_t$ (3.13) (-8.37)	18.20	72.30	224.02
1935-36 to 1973-74	$\Delta Y_t=(1+0.516B^2-0.713B^6)\varepsilon_t$ (4.77) (-8.03)	21.09	22.36	292.743
1935-36 to 1974-75	$\Delta Y_t=(1-0.267B^2-0.812B^6)\varepsilon_t$ (-2.96) (-7.95)	3.61	22.36	555.61
1935-36 to 1975-76	$\Delta Y_t=(1+0.1665B)\varepsilon_t$ (1.35)	3.1	22.36	437.269
1935-36 to 1976-76	$(1-0.18789B+0.155B^2)\Delta Y_t=\varepsilon_t$ (-1.18) (0.96)	3.58	22.36	456.743
1935-36 to 1977-78	$\Delta Y_t=(1-0.172B)\varepsilon_t$ (-1.57)	4.21	23.68	428.35
1935-36 to 1978-79	$(1-0.286B)\Delta Y_t=(1+0.942B)\varepsilon_t$ (1.87) (44.02)	4.30	23.68	419.982
1935-36 to 1979-80	$(1-0.6525B)\Delta Y_t=\varepsilon_t$ (-4.99)	14.30	23.68	585.249
1935-36 to 1980-81	$(1-0.2999B^5-0.7927B^6$ (-2.09) (-5.88) $+0.2288B^7)\Delta Y_t=\varepsilon_t$ (1.37)	4.22	22.36	711.517
1935-36 to 1981-82	$(1-0.2034B^5-0.587B^6$ (-1.62) (-2.86) $+0.32B^7)\Delta Y_t=\varepsilon_t$ (1.80)	11.82	22.36	594.086
1935-36 to 1982-83	$\Delta Y_t=(1-0.497)B^5\varepsilon_t$ (-2.27)	17.06	23.68	500.26

Note: Figures in parentheses are t ratios.

6. Comparing Accuracy of ARIMA and BAE Forecasts

The forecasts and forecast performance measures are given in Tables 3 and 4. Over the evaluation period, it is clear that no one method consistently outperforms the others. For citrus, the ARIMA model forecasts are best (in terms of MSE), but for sugar cane the ARIMA forecasts are the poorest—even slightly worse than the naive no-change forecast. This result can be explained by considering the contrasting data patterns to which the ARIMA models are fitted. The large changes in sugar cane gross value in 1974–75, 1980–81 and 1981–82 are difficult to forecast on an *ex ante* basis (that is, on the basis of historical data) and correspondingly large errors result. In contrast the lower variability of the citrus series, with no major turning points, is well suited to Box-Jenkins methodology.

Of considerable importance are that naive no-change forecasts perform slightly better than the ARIMA model for sugar cane, and much better than BAE forecasts for citrus. Matridakis and Hibson (1979) argued that the naive no-change extrapolation can perform well, in terms of mean absolute percentage error, because the forecasts are hedged toward the middle. Thus, the chance of large errors is smaller when the pattern of the data changes. The more sophisticated ARIMA procedures, on the other hand, are designed to follow the pattern of the data as closely as possible. When there is no change from the previous pattern, they are very accurate, as can be clearly seen by examining a well-behaved series such as citrus. When ARIMA forecasts fail to predict a change in the pattern of the data, however, errors are large—as is the case with sugar cane. (If forecasting directional changes were the single objective, of course, the naive forecast would perform just as badly as the ARIMA forecasts.)

To a large extent the difference in performance of BAE first forecasts as between the two crops can be explained by differences in industry structure. For sugar cane, the tight regulation of this industry, together with climatic factors, makes production relatively easy to forecast. Although world prices for sugar cane vary considerably over time, such movements are not fully reflected in the unit value return to Australian growers. This is because a large proportion of the total crush earns a set price on the domestic market and fixed prices on long-term negotiated contracts. (In total, about 40 per cent of 1985–86 sugar production was sold at a fixed price.) More importantly, the levels of these set prices are known with reasonable certainty at the time the commodity specialist is making the first forecast for the coming financial year. For citrus, errors in forecasting the production series arise from yield variations. Unit value forecasts are subject to the considerable uncertainty of forecasting price movements in international markets for fresh and processed citrus, and to greater uncertainty in forecasting structural changes than in the sugar industry.

7. Correcting for Bias in the BAE Citrus Forecasts

In the theory of combining, it is recognized that the combining of biased forecasts will result in biased composite forecasts. The point is illustrated by Granger and Ramanathan (1984); most previous studies had been based on the assumption that component forecasts had been corrected for any inherent biases.

For the ARIMA forecasts the assumption of lack of bias would seem reasonable, since the usual requirements for estimation are satisfied. In BAE forecasts, on the other hand, systematic bias could easily arise, especially if BAE workers are conservative in making their forecasts (perhaps because an implicit cost function quite different from the quadratic cost function can be imputed to the BAE). Lee and Bui-Lan (1982) suggest that BAE forecasters are likely to moderate their forecasts when they indicate a major departure from the latest observed value.

Upon examining Table 3, it appears that first forecasts for citrus by the BAE are consistently underestimated from 1977

TABLE 3

One-Year-Ahead Forecasts of Gross Value of Production: Citrus

Year	Actual	BAE		BAE corrected[a]		ARIMA		No change	
		Forecast	Squared error	Forecast	Squared error	Forecast	Squared error	Forecast	Squared error
1970-71	43.2	35.0	67.24	35.0	67.24	39.7	12.33	38.5	22.09
1971-72	41.4	38.9	6.25	38.9	6.25	42.8	1.89	43.2	3.24
1972-73	45.7	41.9	14.44	41.9	14.44	43.9	3.37	41.4	18.49
1973-74	46.5	41.0	30.25	41.0	30.25	44.9	2.51	45.7	0.64
1974-75	58.2	54.5	13.69	54.5	13.69	47.7	111.17	46.5	136.89
1975-76	61.2	69.0	60.84	69.0	60.84	61.7	0.23	58.2	9.00
1976-77	70.5	66.1	19.36	66.1	19.36	63.6	48.16	61.2	86.49
1977-78	85.0	70.4	213.16	74.8	104.04	69.3	245.60	70.5	210.25
1978-79	93.0	72.0	256.00	86.5	42.45	91.5	2.35	85.0	64.00
1979-80	103.2	85.0	331.24	96.7	42.77	105.9	7.15	93.0	104.00
1980-81	115.8	93.0	519.84	108.2	57.76	115.6	0.04	103.2	158.76
1981-82	120.0	112.0	64.00	126.0	36.00	124.8	23.32	115.8	17.64
1982-83	129.0	103.0	686.44	118.7	109.41	130.1	0.90	120.0	84.64
1983-84	138.0	117.0	441.00	133.4	21.16	139.0	1.07	129.0	77.44
Mean squared error			194.5		44.41		32.86		70.97
Root mean squared error			13.9		6.7		5.7		8.4
Mean absolute percentage error			13.2		8.5		5.4		9.1
Theil's U_2 statistic			1.6		0.76		0.68		1.00

[a] Corrected for inherent bias

TABLE 4

One-Year-Ahead Forecasts of Gross Value of Production: Sugar Cane

Year	Actual	BAE		ARIMA		No change	
		Forecast	Squared error	Forecast	Squared error	Forecast	Squared error
1970-71	173.3	163.0	106.09	155.3	324.00	148.1	635.04
1971-72	207.4	159.5	2294.41	178.9	809.46	173.3	1162.81
1972-73	230.2	212.0	331.24	195.1	1228.43	207.4	519.84
1973-74	218.9	217.6	1.68	224.0	26.29	230.2	127.69
1974-75	490.7	333.1	24837.70	292.7	39187.00	218.9	73875.10
1975-76	435.6	435.0	0.36	555.6	14401.60	490.7	3035.99
1976-77	472.2	504.0	1011.24	437.3	1220.17	435.6	1339.55
1977-78	420.5	440.0	380.25	456.7	1313.55	472.2	2672.88
1978-79	396.5	452.0	3080.25	428.4	1014.16	420.5	576.00
1979-80	548.2	435.0	12814.20	419.9	16439.90	396.5	23012.90
1980-81	799.7	790.0	94.09	585.2	45989.20	548.2	63252.00
1981-82	590.2	700.0	12056.00	711.5	14717.80	799.7	43890.20
1982-83	508.9	480.0	835.20	594.1	7256.66	590.2	6609.70
1983-84	517.0	435.0	6724.00	500.3	280.39	508.9	65.62
Mean squared error			4611.91		10300.70		15769.70
Root mean squared error			67.9		101.5		125.6
Mean absolute percentage error			11.2		16.0		18.0
Theil's U_2 statistic			0.54		0.81		1.0

onwards. To check whether there is any systematic downward bias, application of cumulative sum techniques (cusum)—as illustrated by Lee and Bui-Lan (1982)—reveals systematic bias in the citrus forecasts. Using the simple correction technique also illustrated by Lee and Bui-Lan, a substantial improvement in forecasting accuracy is achieved over the 1977–78 to 1983–84 period. The root mean square error (RMSE) was improved by 59 per cent (from 17.79 to 7.35).

8. Composite Forecasts: The Results

The composite methods are applied *ex ante* to forecasts over a five-year period 1979–80 to 1983–84 inclusive. This means that weights for the first composite forecast are derived from observed forecast error series for the years 1970–71 to 1978–79. The combining weights are then re-estimated year by year over the evaluation period as each successive forecast error observation is added to the historical series. For example, referring to equations (9)–(14) inclusive,

for n = 1979–80, then T = 1970–71, . . . , 1978–79
n = 1980–81, then T = 1970–71, . . . , 1979–80
etc.

The following combined forecasts are derived: (a) BAE and ARIMA, (b) BAE and naive no-change, (c) all three, (d) ARIMA and naive no-change, and are used for the following methods: (A) simple average, (B) constrained weights, (C) unconstrained weights, (D) unconstrained weights with constant term.

The weights are estimated using ordinary least squares (OLS) regressions, using the TROLL time series data analysis system.[5] In total, 90 regressions are estimated.

The root mean square error (RMSE), mean absolute percentage error (MAPE), and Theil's U_2 statistic are given in Table 5 for each composite method for the two commodities. The MAPE, which has been included to discover the implications of using relative errors instead of absolute errors as a measure of forecast accuracy, is defined as:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|A_i - F_i|}{A_i} \qquad (20)$$

The following results apply only to BAE's first forecasts, and not to the revised forecasts produced in later quarters of the year. The immediate result is that the RMSEs of the composite forecasts generally lie within the bounds of the RMSEs of the component forecasts. This is true also using the alternative criteria of accuracy (MAPE and Theil's U_2). The different combining methods (A,B,C,D) produce RMSE figures of varying magnitudes within these same bounds of the component forecasts.

In the case of citrus, none of the composite forecasts proved superior to the ARIMA model forecasts (the best of the component forecasts). With the sugar cane forecasts, the BAE forecasts out-performed both the ARIMA and composite forecasts.

In contrast to the results of such authors as Granger and Ramanathan (1984), the "unconstrained weights with constant term" method is here not always superior to the other combining procedures. Indeed, no one combining method is consistently best throughout the two series examined, although method D is most often the better method for sugar cane.

In Tables 6–9 the estimated combining weights are reported. In method B the weights are constrained to add to unity; in methods C and D they are not so constrained; and the constant term in method D is included to take account of any bias that may be present in the component forecasts.

[5] Time Reactive On Line Laboratory computer system, from the Massachusetts Institute of Technology.

TABLE 5

Composite Forecasts: 1979-80 to 1983-84

Forecast	Citrus RMSE	Citrus MAPE	Citrus Theil's U_2	Sugar cane RMSE	Sugar cane MAPE	Sugar cane Thiel's U_2
Original						
BAE (corrected)	7.3	5.8	0.81	80.7	12.4	0.42
ARIMA	2.6	1.6	0.27	130.1	18.1	0.77
No change	9.4	7.6	1.00	165.4	22.4	1.00
Combining method A (naive equal weighting)						
BAE and ARIMA	3.85	2.9	0.46	93.5	14.1	0.51
BAE and no change	7.93	6.0	0.85	112.1	16.3	0.63
All three	6.02	4.6	0.66	116.8	16.9	0.66
Combining method B (constrained)						
BAE and ARIMA	3.93	3.0	0.44	82.7	13.5	0.44
BAE and no change	7.62	6.2	0.84	108.82	17.0	0.63
All three	4.44	2.6	0.53	84.2	11.2	0.51
Combining method C (unconstrained)						
BAE and ARIMA	6.46	4.1	0.74	93.8	15.4	0.54
BAE and no change	5.31	2.7	0.64	115.9	18.6	0.72
All three	6.06	4.0	0.65	89.7	13.8	0.52
Combining method D (unconstrained with constant)						
BAE and ARIMA	5.31	3.2	0.62	84.0	13.9	0.44
BAE and no change	5.39	2.7	0.65	94.18	14.6	0.54
All three	6.74	4.0	0.76	93.7	15.9	0.54

From comparison of Tables 6 and 7, it is obvious that negative weights are prominent for the time series forecasts of the sugar cane series. This reflects the general inferiority of these methods to the BAE forecasts. This does not mean, however, that the time series forecasts should not have been considered in forming the composite forecast. Nelson (1971) noted that the relative accuracy is not an appropriate basis for excluding any one forecast in favour of another. He suggested that even a very inaccurate forecast would generally be included in a minimum variance composite. Negative weights appear more frequently when all three forecasts are combined.

Another possibility is to combine the two time series forecasts (ARIMA and naive no-change). The composite forecast using this procedure, was not able to improve upon the BAE forecasts for sugar cane, and the results are not reported here. However, in all cases except the naive equal weighting, the composite forecast out-performed the better of the component forecasts (the ARIMA). For the citrus series, the composite forecast did in fact out-perform the better component forecast in three instances, and was in all cases superior to the BAE/ARIMA composite forecast.

TABLE 6

Combining Weights: Citrus

Forecast	1979-80	1980-81	1981-82	1982-83	1983-84
Combining method A (naive equal weighting)					
(i) BAE	0.5	0.5	0.5	0.5	0.5
ARIMA	0.5	0.5	0.5	0.5	0.5
(ii) BAE	0.5	0.5	0.5	0.5	0.5
No change	0.5	0.5	0.5	0.5	0.5
Combining method B (constrained)					
(i) BAE	0.66	0.56	0.47	0.45	0.35
ARIMA	0.34	0.55	0.53	0.55	0.65
(ii) BAE	0.85	0.94	1.06	0.90	0.87
No change	0.15	0.06	0.06	0.10	0.13
Combining method C (unconstrained)					
(i) BAE	0.58	0.71	0.73	0.58	0.56
ARIMA	0.50	0.36	0.33	0.45	0.48
(ii) BAE	0.42	0.42	0.42	0.23	0.25
No change	0.68	0.68	0.68	0.85	0.83
Combining method D (unconstrained with constant)					
(i) Constant	2.15	3.58	3.44	6.85	6.12
BAE	0.59	0.65	0.65	0.49	0.42
ARIMA	0.45	0.36	0.36	0.46	0.54
(ii) Constant	-1.29	-0.63	-0.70	2.90	2.49
BAE	0.40	0.41	0.41	0.31	0.27
No change	0.72	0.69	0.69	0.74	0.78

9. The Estimated Weights: Some Comments

The Ordinary Least Squares (OLS) regression estimates of the weights give an insight into the relative merits of the component forecasts, and foreshadow the likely success of the combining process before the composite forecasts are actually calculated. In many instances, estimated weights are not statistically significant and Durbin-Watson statistics are often unacceptably high or low, indicating the presence of autocorrelation and model mis-specification. Such results suggest that combining the component forecasts would be of little use, because the regression diagnostics show that they contain little independent information. In interpreting the results, it should be noted that the authors are not trying to identify a "true model". The regression parameters could not be improved upon without the addition of other component forecasts which do bring in some useful independent information. At this stage, no attempt was made to respecify the model by adding forecasts generated by other methods as explanantory variables; but this is certainly an avenue for further research.

TABLE 7
Combining Weights: Sugar Cane

Forecast	1979-80	1980-81	1981-82	1982-83	1983-84
Combining method A (naive equal weighting)					
(i) BAE	0.5	0.5	0.5	0.5	0.5
ARIMA	0.5	0.5	0.5	0.5	0.5
(ii) BAE	0.5	0.5	0.5	0.5	0.5
No change	0.5	0.5	0.5	0.5	0.5
Combining method B (constrained)					
(i) BAE	1.12	1.19	1.09	1.12	1.06
ARIMA	-0.12	-0.19	-0.09	-0.12	-0.06
(ii) BAE	1.60	1.74	1.25	1.34	1.27
No change	-0.60	-0.74	-0.25	-0.34	-0.27
Combining method C (unconstrained)					
(i) BAE	1.13	1.23	1.04	1.13	1.05
ARIMA	-0.13	-0.18	-0.01	-0.14	-0.05
(ii) BAE	1.62	1.72	1.23	1.37	1.27
No change	-0.63	-0.70	-0.22	-0.41	-0.28
Combining method D (unconstrained with constant)					
(i) Constant	68.8	56.88	58.81	84.46	77.23
BAE	0.96	1.10	1.07	1.14	1.03
ARIMA	-0.13	-0.20	-0.17	-0.33	-0.18
(ii) Constant	71.19	60.57	82.02	85.52	76.53
BAE	1.45	1.58	1.28	1.28	1.08
No change	-0.65	-0.73	-0.48	-0.48	-0.24

10. Conclusions

The basic motivation of this study has been to investigate the combining of subjectively formed BAE forecasts with alternative time-series model forecasts as a means of improving BAE forecasting performance. In the analysis, the performance of alternative ARIMA and naive no-change forecasting models, as well as the BAE forecasts, were evaluated. For the two selected series—gross value of production of citrus and sugar-cane—substantial differences in the performance of all three forecasting methods were established. These differences highlighted the fundamental differences in industry structure, and in the nature of the underlying generating processes, between the two crops.

Application of the combining methods showed that with one major exception, composite forecasts derived on an *ex ante* basis could not improve upon the performance of the component forecasts. The exception was the combining of ARIMA and naive no-change forecasts. The general result suggests that there is little independent information embodied in the individual forecasts—as is corroborated by the statistical

TABLE 8

Combining Weights: Combining All Three Forecasts: Citrus

Forecast	1979-80	1980-81	1981-82	1982-83	1983-84
Combining method A (naive equal weighting)					
BAE	0.33	0.33	0.33	0.33	0.33
ARIMA	0.33	0.33	0.33	0.33	0.33
No change	0.33	0.33	0.33	0.33	0.33
Combining method B (constrained)					
BAE	0.66	0.68	0.69	0.52	0.45
ARIMA	0.75	0.64	0.68	0.60	0.70
No change	-0.41	-0.32	-0.37	-0.12	-0.15
Combining method C (unconstrained)					
BAE	0.49	0.45	0.44	0.26	0.27
ARIMA	-0.68	-0.25	-0.17	-0.19	-0.19
No change	1.31	0.92	0.85	1.03	1.02
Combining method D (unconstrained with constant)					
Constant	-3.25	-3.85	-4.16	2.87	2.34
BAE	0.46	0.41	0.41	3.07	0.28
ARIMA	-0.81	-0.46	-0.44	-0.01	-0.03
No change	1.53	1.23	1.22	0.74	0.81

insignificance of the majority of the estimating weights.

A more general result is that forecasters can gauge the extent of independent information contained in the forecasts derived from the different approaches by examining the significance of the estimated weights, even before assessing the accuracy of the composite forecasts relative to the component forecasts.

Even though combining forecasts on an *ex ante* basis has not been successful for these two series over the particular evaluation period chosen, the study has been useful in identifying problems and weaknesses in the citrus series that otherwise would not have come to light. The fact that some BAE forecasts are less accurate than the ARIMA and no-change forecasts does not mean that the ARIMA model should be chosen as the "correct" forecasting method to the exclusion of the BAE commodity specialist forecasts. Rather, a more appropriate response would be to review and revise the methodology of the commodity specialists. When a revised BAE methodology has been devised, however, there may still be gains to be exploited by continuing to apply the combining methodology.

Although this study has examined only BAE forecasts and time series model forecasts, it is desirable to consider all available information, including forecasts from other sources. For instance, it may be beneficial to combine forecasts from time series models, econometric models, and subjectively formed forecasts. There is no *a priori* reason that econometric model forecasts should perform better than time

TABLE 9

Combining Weights: Combining All Three Forecasts: Sugar Cane

Forecast	1979-80	1980-81	1981-82	1982-83	1983-84
Combining method A (naive equal weighting)					
BAE	0.33	0.33	0.33	0.33	0.33
ARIMA	0.33	0.33	0.33	0.33	0.33
No change	0.33	0.33	0.33	0.33	0.33
Combining method B (constrained)					
BAE	1.38	1.47	1.03	1.03	0.99
ARIMA	0.92	1.03	1.16	1.18	1.23
No change	-1.30	-1.50	-1.19	-1.21	-1.22
Combining method C (unconstrained)					
BAE	1.38	1.48	1.08	1.06	0.99
ARIMA	1.15	1.16	1.32	1.14	1.23
No change	-1.58	-1.66	-1.45	-1.23	-1.24
Combining method D (unconstrained with constant)					
Constant	72.15	61.62	75.01	55.77	49.20
BAE	1.20	1.34	1.13	1.07	0.98
ARIMA	1.16	1.17	1.24	0.86	0.01
No change	-1.61	-1.69	-1.60	-1.08	-1.11

series or subjectively formed forecasts, although a properly specified model will always perform better than a mis-specified model. What the combining methodology can show is the extent to which the forecasts from the different sources contain useful independent information. If they do, then forming a composite forecast will improve forecasting accuracy.

References

BATES, J. M. and GRANGER, C. W. J. (1969), "The combination of forecasts", *Operational Research Quarterly* 20 (4), 451–68.

BESSLER, D. A. and BRANDT, J. A. (1981), "Forecasting livestock prices with individual and composite methods", *Applied Economics* 13 (4), 513–22.

BORDLEY, R. F. (1982), "The combination of forecasts: a Bayesian approach", *Journal of the Operational Research Society* 33 (2), 171–4.

BOX, G. E. P. and JENKINS, G. M. (1970), *Time Series Analysis*, Holden-Day, San Francisco.

BUNN, D. W. (1975), "A Bayesian approach to the linear combination of forecasts", *Operational Research Quarterly* 26 (2, i), 325–9.

COOPER, J. P. and NELSON, C. R. (1975), "The *ex ante* prediction performance of the St Louis and FRB-MIT-PENN econometric models and some results on composite predictions", *Journal of Money, Credit and Banking* 7, 1–32.

DICKINSON, J. P. (1973), "Some statistical results in the combination of forecasts", *Operational Research Quarterly* 24 (2), 253–60.

—— (1975), "Some comments on the combination of forecasts", *Operational Research Quarterly* 26 (1, ii), 205–10.

FREEBAIRN, J. W. (1975), "Forecasting for Australian agriculture", *Australian Journal of Agricultural Economics* 19 (3), 154–74.

GRANGER, C. W. J. (1969), "Prediction with a generalised cost of error function", *Operational Research Quarterly* 20 (2), 199–207.

—— and NEWBOLD, P. (1975), "Economic forecasting; the atheist's viewpoint", in G. Renton (ed.), *Modelling the Economy*, Heinemann Educational Books, London.

—— and —— (1977), *Forecasting Economic Time Series*, Academic Press, New York.

GRANGER, C. W. S. and RAMANATHAN, R. (1984), "Improved methods of combining forecasts", *Journal of Forecasting* 3 (2), 197–204.

HARRISON, P. J. and STEVENS, C. F. (1971), "A Bayesian approach to short-term forecasting", *Operational Research Quarterly* 22 (4), 341–52.

LEE, B. M. S. and BUI-LAN, A. (1982), "Use of errors of prediction in improving forecast accuracy: an application to wool in Australia", *Australian Journal of Agricultural Economics* 26 (1), 49–62.

LEUTHOLD, R. M. (1975), "On the use of Theil's inequality coefficents", *American Journal of Agricultural Economics* 57 (2), 344–6.

LJUNG, G. M. and BOX, G. E. P. (1978), "On a measure of lack of fit in time series models", *Biometrika*, 66, 67–72.

LONGBOTTOM, J. A. and HOLLY, S. (1985), "The role of time series analysis in the evaluation of econometric models", *Journal of Forecasting* 4 (1), 75–82.

LONGMIRE, J. and WATTS, G. (1981), "On Evaluating Forecasts and Forecasting Methods", Paper presented at the 25th Annual Conference of the Australian Agricultural Economics Society, Christchurch, New Zealand, 10–12 February, 1981.

MAKRIDAKIS, S. and HIBSON, M. (1979), "Accuracy of forecasting: An empirical investigation", *Journal of the Royal Statistical Society A., Part 2* 142, 97–145.

MAKRIDAKIS, S. and WINKLER, R. L. (1983), "Averages of forecasts: some empirical results", *Management Science* 29 (9), 987–96.

NELSON, C. R. (1972), "The prediction performance of the FRB-MIT-PENN model of the U.S. economy", *American Economic Review* 62 (5), 902–17.

NEWBOLD, P. and GRANGER, C. W. J. (1974), "Experience with forecasting univariate time series and the combination of forecasts", *Journal of the Royal Statistical Society (A)* 137 (2), 131–46.

SPRIGGS, J. (1981), "Forecasts of Indiana monthly farm prices using univariate Box-Jenkins analysis and corn futures prices", *North Central Journal of Agricultural Economics* 3 (1), 81–2.