PREDICTION OF DRY-LAND CROP YIELDS USING RAINFALL

Hovav Talpaz, Chang Hao Chun and C. Robert Taylor

INTRODUCTION

A significant component of risk in agricultural production systems is yield variability. The ability to predict yield and thus reduce risk would have many potential benefits to both individual decision makers and society. For example, in farm planning, quality crop yield forecasts could be crucial in making decisions on optimal crop combinations. At the aggregate level, an example where a yield forecast would have obvious benefits would be cases in which government manages a grain buffer stock.

A few recent attempts to predict the weather and, in turn, its influence on crop yields have been made. Studies by R. Black [1], C. B. Luttrell and R. A. Gilbert [7] and V. L. Harrison [3] deal with weather forecasts and their impact on selected crops in selected regions of the U.S.[1] Harrison concluded that lower-than-average yields are associated with low sunspot activity (for definition see [3, pp. 1-2]), and high sunspot activity is associated with higher-than-average yields. However, his results were largely inconclusive in many cases, yet the contribution of his work is substantial. Black developed a corn price forecast scheme based on "weather odds" but admitted a lack of climatological foundations in relating weather developments to these odds [1, p. 943].

Luttrell and Gilbert concluded that in the leading producing states there is ". . . little evidence that yields are either cyclical or bunchy as a result of weather . . ." [7, p. 530]. They relate any bunchiness of crop yields to use of inputs such as fertilizer, hybrid seed and changes in relative price. Luttrell and Gilbert's results may provide an explanation of Harrison's inconclusive results, who tried to relate highly cyclical sunspot activity to nonbunchy crop yields.

With these very important studies at hand, a researcher is left with two alternative conclusions: either dry-land crop yields are randomly distributed about a time trend (. . . "Statistical tests show little evidence of nonrandomness in these series . . ." [7, p. 521]), or weather variables which do affect crop yields are not necessarily cyclical. If the first conclusion is valid, then decision makers can be helped very little indeed. But, if the second conclusion holds and a dynamic relationship between weather variable(s) and crop yield can be identified, then it is possible to obtain forecasted yields with a lower variance than that around a time trend.

This study applies Box and Jenkins time series methods [2] to the issue of whether a dynamic relationship between crop yield and rainfall or sunspot activity exists, rendering yields more predictable. Dry-land corn and wheat yields in Nebraska were considered in this study. A state was needed as the unit of analysis with size not too large to cancel out weather variability, yet not too small to be influenced by local extreme random behavior which could overshadow causal effects. Also, the necessary secondary data on sunspot activity and rainfall, the two indicators considered, were available. The wheat forecast model developed is embedded in an optimal wheat stock model to illustrate how it can be used in an aggregate economic problem.

H. Talpaz is Associate Professor, C. H. Chun is Graduate Assistant and C. R. Taylor is Assistant Professor of Agricultural Economics, Texas A & M University.

*Texas Agricultural Experiment Station Technical Article, TA—12936. The authors are indebted to the editor and two anonymous reviewers for their valuable suggestions.

[1] More elaborate reviews of earlier works can be found in Black's and Harrison's articles along with extensive reference lists.

METHODOLOGY

The basic assumption behind this study is that crop yield (Y_t) at time(t) and an input indicator (X_t), like sunspot activity or rainfall level, belong to a class of discrete linear transfer functions (a transfer function relates inputs to outputs in a dynamic fashion) (For an elaborated discussion related to the rest of this section, see [2, chs. 10 and 11]) given by

$$Y_t - \eta_1 Y_{t-1} - \ldots - \eta_r Y_{t-r} = \omega_o X_{t-b} + \omega_1 X_{t-b-1} + \ldots + \omega_s X_{t-b-s} \quad (1)$$

where

$\eta_1, \ldots, \eta_r, \omega_o, \ldots, \omega_s$ = parameters of the system and

b = time delay constant to be estimated among the integers 0, 1, ...,20.

Adding a disturbance variable N_t and condensing (1) into a transfer function format, we have the following special case

$$Y_t = \frac{\omega(B)X_{t-b}}{\eta(B)} + \frac{a_t}{\Phi(B)} \quad (2)$$

with the following quadratic polynomial definitions

$$\omega(B) = \omega_o + \omega_1 B + \omega_2 B^2$$

$$\eta(B) = 1 - \eta_1 B - \eta_2 B^2$$

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2$$

$$a_t = \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q} + N_t - \Phi_1 N_{t-1} - \Phi_2 N_{t-2} \quad (3)$$

and B is the backward time shift operator defined

$$B^k X_t = X_{t-k} \quad (4)$$

and $\theta_1, \ldots, \theta_q$ are the autoregressive parameters of the white noise a_t.

The problem is to estimate $\omega(B)$, $\eta(B)$ and $\Phi(B)$. It can be shown that estimation by the method described below is possible if both Y_t and X_t are stationary time series.[2] The input X_t appears to be stationary, and the crop yield variable Y_t was transformed into a stationary series by removing the time trend in the following manner:

$$\overset{o}{Y}_t = Y_t - (\tilde{\gamma}_o + \tilde{\gamma}_1 t + \tilde{\gamma}_2 t^2 + \tilde{\gamma}_3 t^3) \quad (5)$$

where $\overset{o}{Y}_t$ is the observed series and $\tilde{\gamma}_o, \ldots, \tilde{\gamma}_3$ are the maximum likelihood estimators of the model

$$\overset{o}{Y}_t = \gamma_o + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3 + \epsilon_t \quad (6)$$

Coefficients of a higher degree polynomial were not significantly different from zero at the 5 percent level.[3]

The lag parameter was estimated as follows. The computer program FTTRAN (see IMSL) was first used in estimating two ARMA models[4] to obtain series α_t and β_t by:

$$\alpha_t = \theta^{-1}(B)\,\Psi(B)\,(X_t - \overline{X}_t) \quad (7)$$

$$\beta_t = \theta^{-1}(B)\,\Psi(B)\,(Y_t - \overline{Y}_t) \quad (8)$$

where θ and Ψ are the estimated parameters of the moving average and autoregression, respectively. Then estimates of the impulse response weights V_i (for a definition and discussion of V_i, see [2, pp. 337-338]).

$$\beta_t = V_o \alpha_t + \ldots + V_i \alpha_{t-i}, \quad i = 0, 1, \ldots, 20 \quad (9)$$

and b is taken to be the smallest index of V_i which passes a significance (5%) test for $V_i = 0$. With this b, maximum likelihood estimates for $\omega(B)$ and $\eta(B)$ are made followed by the estimation of the white noise disturbance a_t in equation (2).

If the lag parameter b is greater than or equal to unity, then a one-period (year) forecast is possible by equation (2), with appropriate substitution of equation (3) into (2), using the estimated parameters. If b = 0, it becomes necessary to make the forecast in two steps. First, input X_t is forecast. Second, the forecast of Y_t is obtained, as before, using the forecast X_t and its observed lagged values. In forecasting X_t, it is necessary to estimate the following ARIMA (integrated ARMA) [8, chs. 5, 6]. Initial estimates of the parameters are obtained by least squares regression:

$$Z_t = \sum_{i=1}^{p} \tilde{\lambda}_i Z_{t-i} + \sum_{j=1}^{q} \tilde{\Omega}_j \epsilon_{t-j} + \epsilon_t \quad (10)$$

[2] A time series is classified stationary if the process is nondiversing and possesses a constant mean level (i.e. the mean does not change with time), see [8, pp. 19-22].

[3] Such detrending should be revised as current data become available. This study demonstrates that this procedure allows only short-term forecasts.

[4] ARMA stands for autoregressive moving average, see [8, chs. 4, 5; or 2, chs. 3, 4].

Where $Z_t = X_t - \overline{X}_t$ and $\epsilon_t = Z_t - \hat{Z}_t$ ($\hat{Z}_t$ is the forecast of Z_t) are the residuals, $\tilde{\lambda}_i$ and $\tilde{\Omega}_j$ are the autoregressive and moving average initial estimates, respectively. Since high correlation exists among Z_t and its lags and among ϵ_t and its lags, equation (10) provides inconsistent least square regression estimates with large variances. Therefore, the initial estimates are used in a maximum likelihood estimation procedure of equation (10). This is obtained by a modified steepest descent nonlinear algorithm [6] aimed at minimizing:

$$S = \sum_{t=1}^{T} \epsilon_t^2 \ . \qquad (11)$$

Results of applying the above methodology to corn and wheat yields in Nebraska are discussed in the next section.

RESULTS

As indicated before by Harrison, sunspot activity proved not to be significant as an explanatory variable for corn and wheat yields. (The time delay parameter b was greater than 20, hence the model was rejected.) Furthermore, no better results were achieved in attempting to estimate and forecast annual rainfall for Nebraska using sunspot activity as an explanatory variable.

However, rainfall proved to be a modestly successful predictor of corn and wheat yields in Nebraska.

Table 1 presents time trend parameters for equations (5) and (6) used to achieve stationarity. Also presented are transfer function estimators and sum of squared residuals.

Since the b parameter was zero in the case of corn, it was necessary to obtain an ARIMA model for the rainfall variable, equation (11). The parameters of this model were estimated as:

$$\hat{\lambda}_1 = 0.0016; \hat{\lambda}_2 = 0.0466; \hat{\lambda}_3 = 0.0927;$$

$$\hat{\Omega}_1 = ... = \hat{\Omega}_q = 0.00; \overline{X} = 19.7237.$$

These estimators are significant at a 5 percent level. X_t was then projected one year ahead and used to forecast corn yield.

Tables 2 and 3, along with Figures 1 and 2, depict comparisons between forecasted and observed yields of the years 1964-1974 and the 1975 real forecast yield. (The 1866-1974 data were used in the estimation process, but limitation of space would not allow presentation of the entire period.) For both crops, the 1975 observed yields were not used in the estimation stage, and their values represent the preliminary USDA estimates.

TABLE 1. THE TIME TREND COEFFICIENTS (γ), TIME DELAY (b), TRANSFER FUNCTION PARAMETERS (ω,η) AND SUM SQUARED ERROR (S) FOR CORN AND WHEAT YIELD (BU/ACRE) USING RAINFALL AS THE INPUT (1866-1974 PERIOD)

Model Parameters	Corn	Wheat
γ_0	28.8037	14.3633
γ_1	0.8559	-0.1636
γ_2	-0.0345	0.0030
γ_3	0.0003	0.0
b	0	1
ω_0	1.1262	0.294786
ω_1	0.2211	-0.226767
ω_2	-0.0263	-0.124904
η_1	0.1684	0.407634
η_2	0.1001	0.163924
Φ_1	0.1927	0.422436
Φ_2	0.1263	0.173070
S	5182.13	1216.05

The corn-yield one-year-ahead forecast yielded an average absolute error of 6.20 bu/acre for the above mentioned periods (or an average of 8.40%). The observed yield of 1974 was extremely low, due to a very unfavorable and unique rain distribution coupled with other bad weather conditions for corn. This led to a large negative deviation which forced the model to forecast small but consistent positive

TABLE 2. WHEAT CROP YIELD ONE-YEAR-AHEAD FORECAST

Year	Forecast	Observation	Difference	%
1964	22.3881	24.5000	2.1119	8.6199
1965	24.9406	20.0000	-4.9406	-24.7030
1966	27.3211	35.0000	7.6789	21.9398
1967	27.4687	26.5000	-0.9687	-3.6554
1968	29.2597	32.0000	2.7403	8.5635
1969	30.9441	31.5000	0.5559	1.7647
1970	31.4305	38.0000	6.5695	17.2881
1971	33.0606	42.0000	8.9394	21.2843
1972	38.0565	37.0000	-1.0565	-2.8553
1973	36.4311	35.0000	-1.4311	-4.0889
1974	36.2488	34.0000	-2.2488	-6.6141
1975	29.7991	32.0000	4.2009	12.3557
Mean absolute difference			3.62	11.14

TABLE 3. CORN CROP YIELD ONE-YEAR-AHEAD FORECAST

Year	Forecast	Observation	Difference	%
1964	52.5790	54.0000	1.4210	2.6315
1965	57.7399	70.0000	12.2610	17.5145
1966	80.1683	80.0000	-0.1683	-0.2103
1967	70.1825	74.0000	3.8175	5.1588
1968	70.2765	74.0000	3.7235	5.0317
1969	87.7740	93.0000	5.2660	5.6194
1970	75.3228	76.0000	0.6772	0.8911
1971	79.7966	85.0000	5.2034	6.1216
1972	98.3774	104.0000	5.6226	5.4063
1973	80.9222	94.0000	13.0778	13.9125
1974	94.6325	68.0000	-26.6325	-39.1654
1975	88.1079	86.0000	-2.1079	-0.0245
Mean absolute difference			6.20	8.40

deviations for the previous seven years. However, forecasts closely follow the observed values for all other years (except 1974) including 1975, which was largely recognized as a minor drought year. The wheat-yield one-year-ahead forecast yielded an average absolute difference of 3.62 bu/acre (or 11.14%) for the same period mentioned above. The larger absolute percent deviation can be explained by the fact that the crop yield is dependent largely on last year's rain (b=1), but other weather conditions which are not represented here do affect current-year yield. Since some weather conditions such as temperature, humidity, etc., are correlated with current

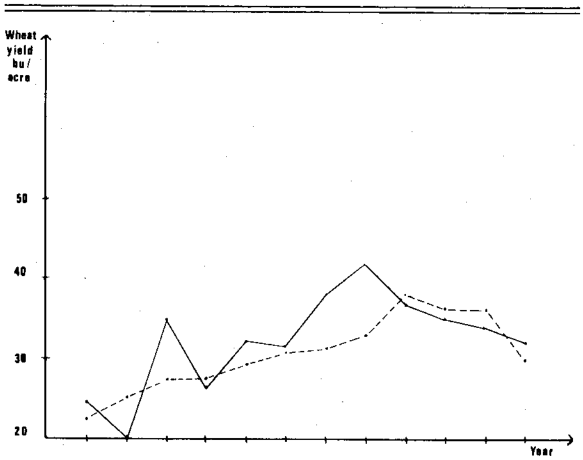

FIGURE 1. WHEAT CROP YIELDS (BU/ACRE), ONE-YEAR-AHEAD FORECASTS (BROKEN LINE) AND OBSERVED VALUES (SOLID LINE)

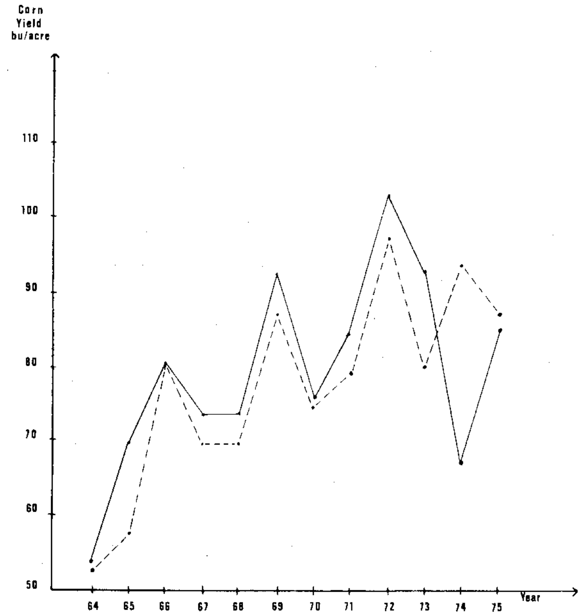

FIGURE 2. CORN CROP YIELDS (BU/ACRE), ONE-YEAR-AHEAD FORECASTS (BROKEN LINE) AND OBSERVED VALUES (SOLID LINE)

rainfall; this was a minor problem in the corn forecast model which depends on current rainfall (b=0).

Previous studies did not attempt a one-year-ahead forecast, so it is impossible to compare these results with others. Objectively, a diagnostic checking [2, p. 397] applied to the two cases provided no evidence that the model is inadequate.

ECONOMIC IMPLICATIONS: AN APPLICATION

The potential contribution of a good quality crop yield forecast could be highly valuable. At the farm level, a producer may improve his planning by utilizing more precise data. For policy analysis at the government level, a better forecast could become crucially important. Many applications could be considered; however, space does not allow an elaborated discussion. Therefore, the following is a demonstrated application, not necessarily the most important among those mentioned above.

In an unrelated study, Taylor and Talpaz [9] developed a model for optimizing wheat grain stocks for the U.S. Briefly, that model gives the level of stocks that maximize consumers' plus producers' surplus less storage costs subject to an econometric model of the wheat sector.

Let us assume, for the sake of this demonstration, that the average wheat yield for the U.S. is identical to Nebraska's for the years 1971-75.

Obviously, this is a strong assumption. Consider three simulators:

I. With the crop yield as observed.
II. With the crop yield as projected by time trend only, eq. (6).
III. With crop yield as forecast by the transfer function method.

Suppose there is an interest in predicting wheat price at the farm level with the assumptions underlying Taylor and Talpaz's model. Then the predicted wheat price can be arrived at by supplying crop yields under I, II and III, holding everything else constant. The wheat price is resolved through optimization of wheat stock levels subject to the stock identity relationship (demand equals supply plus change in stocks). Table 4 depicts results of these deterministic simulation runs. The last two columns show the absolute deviations in price of wheat ($/bu.) between the five years' prices of the observed crop yields and the two forecasted crop yields—transfer function and time trend, respectively. Deviations under the transfer function forecasts are smaller than those under the time trend (total sum of absolute deviations are 2.73 and 3.71, respectively). This was primarily because smaller errors in yield forecasts led to better planning of wheat storage levels.

TABLE 4. WHEAT CROP YIELD AND ABSOLUTE WHEAT PRICE DIFFERENCES FROM OBSERVED YIELDS

Year	Wheat Crop Yield			Absolute Differences of Wheat Prices	
	Observed	Transfer Func. Forecast	Time Trend Forecast	Transfer Func. Forecast	Time Trend Forecast
1971	42.00	33.06	30.26	$ 0.16	$ 0.76
1972	37.00	38.06	30.73	1.40	0.65
1973	35.00	36.43	31.21	0.01	0.43
1974	34.00	36.25	31.69	1.10	1.37
1975	32.00	29.80	32.17	0.06	0.50

CONCLUDING REMARKS

This study reports an attempt to forecast dry-land crop yields using advanced and powerful time series analysis methods. Sunspot activity, as an explanatory variable, fails to serve as an input in the transfer function. This conclusion was observed by others also. Rainfall, however, is a relatively good explanatory variable and allows estimation of the transfer function parameters, which in turn allows computation of one-year-ahead forecasts.

Presently, it is difficult to assess the extent of the contribution of such a study in providing one-year outlook information for farmers. A further effort towards this assessment should be carried out. Also, future research should extend this study to include additional states, and possibly the entire Great Plains region.

This study shows some significant promise in adapting transfer function methodology for the analysis and forecast of dry-land crop yields which are so crucial in economic planning at all levels.

REFERENCES

[1] Black, R. "Weather Variation as a Cost-Price Uncertainty Factor as It Affects Corn and Soybean Production," *American Journal of Agricultural Economics*, 57 (1975): 940-944.

[2] Box, P. O. and G. Jenkins. *Time Series Analysis, Forecasting and Control*, Revised Edition, San Francisco: Holden-Day, Inc., 1976.

[3] Harrison, V. L. "Crop Yields and Sunspot Cycles," USDA, National Economic Analysis Division, February 1975 (Mimeo.)

[4] IMSL. *The IMSL Library*, 5th Edition, Houston, Texas: IMSL, Inc., 1976.

[5] Johnson, D. G. "World Agriculture, Commodity Policy, and Price Variability," *American Journal of Agricultural Economics*, 57 (1975): 823-828.

[6] Luenberger, D. G. *Introduction to Linear and Nonlinear Programming*, Reading, Mass.: Addison-Wesley, 1973.

[7] Luttrell, C. B., and R. A. Gilbert. "Crop Yields: Random, Cyclical or Bunchy," *American Journal of Agricultural Economics*, 58 (1976): 521-531.

[8] Nelson, C. R. *Applied Time Series Analysis for Managerial Forecasting*, San Francisco: Holden-Day, Inc., 1973.

[9] Taylor, C. R. and Hovav Talpaz. *First Order Optimal Carryover Levels for Wheat in the United States*, Texas Agricultural Experiment Station Technical Article #TA-13243.

DATA REFERENCE

I. Wheat:

a. 1866-1943
U.S.D.A. "Wheat: Acreage, Yield, Production by States, 1866-1943," SB 158, February 1955.

b. 1944-1949
U.S.D.A. "Crop production, Revised Estimates, 1944-49, Acreage, Yield and Production of Principal Field Crops," SB 108, March 1952.

c. 1950-1954
U.S.D.A. "Field Crops by States, 1949-54, Acreage, Yield, Production, Revised Estimates," SB 185, June 1956.

d. 1955-1959
U.S.D.A. "Field Crops by States, 1954-59, Acreage, Yield, Production, Revised Estimates," SB 290, June 1961.

e. 1960-1963
U.S.D.A. Statistical Reporting Service, "Crop Production Annual Summary," CrPr 2, various years.

f. 1964-1969
U.S.D.A. "Field Crops, Revised Estimates by States, 1964-1969, Acreage, Yield, Production," SB 498, November 1972.

g. 1970-1973
U.S.D.A. Stastical Reporting Service, "Crop Production Annual Summary," CrPr 2, various years.

II. Corn

a. 1866-1928
U.S.D.A. Bureau of Agricultural Economics. "Corn Statistics," Statistical Bulletin 28, Washington, D.C., January 1930.

b. 1929-1933
U.S.D.A. Yearbook of Agriculture, U.S. Government Printing Office, Washington, D.C., various years, 1931-35.

c. 1934-1943
U.S.D.A. Agricultural Statistics, U.S. Government Printing Office, Washington, D.C., 1936-45.

d. 1944-1949
U.S.D.A. Crop Reporting Board, Bureau of Agricultural Economics, "Crop Production, Revised Estimates, 1944-49, Acreage, Yield and Production of Principal Field Crops," SB 108, March 1952.

e. 1950-1954
U.S.D.A. Agricultural Marketing Service, Crop Reporting Board, "Field Crops by States, 1949-54, Acreage, Yield, Production, Revised Estimates," SB 185, June 1956.

f. 1955-1959
U.S.D.A., SRS. "Field Crops by States, 1954-1959, Acreage, Yield, Production, Revised Estimates," SB 290, June 1961.

g. 1960-1963, 1970-1974
U.S.D.A., SRS. "Crop Production Annual Summary," CrPr 2, various years.

h. 1964-1969
U.S.D.A., SRS. "Field Crops, Revised Estimates by States, 1964-1969, Acreage, Yield, Production," SB 498, November 1972.

III. Rainfall

a. 1866-1876
U.S.D.A., Weather Bureau. Bulletin Q, Climatology of the U.S., 1906, U.S. Government Printing Office, Washington, D.C.

b. 1876-1946
U.S. Department of Commerce, Weather Bureau. Climatological Data for the United States by Sections, Washington, D.C.

c. 1946-1960
U.S. Department of Commerce, Weather Bureau. Decennial Census of the U.S. Climate—Monthly Average for State Climatic Divisions, 1931-1960; Nebraska Climatography of the United States, No. 85-21, Washington, D.C., 1963.

d. 1960-1974

U.S. Department of Commerce, Weather Bureau. Climatological Data, Nebraska, Asheville, N.C., July 65, Changes to Environmental Science Services Administration, Environmental Data Service.

IV. Sunspots

a. Mrs. Viola Miller

World Data Center A for Solar-Terrestrial Physics

National Oceanic and Atmospheric Administration,

Boulder, Colorado 80302

Tel. 303-499-1000. Ext. 6467