# A Model for Forecasting Wheat Production in the Punjab

*by*

B.A. AZHAR, M. GHAFFAR CHAUDHRY,

AND

M. SHAFIQUE*

## INTRODUCTION

The purpose of the study is to explore the possibility of developing an econometric model which would help in explaining and predicting changes in wheat production in the Punjab. Wheat is a major crop in the province occupying nearly 40 per cent of the total cultivated area and about 2/3 of the area under Rabi crops. It is by far the most important food crop providing nearly 2/3 of the total calorie intake. Accurate forecasting of wheat production is critical for planning process. Wheat production in the province increased from 27.2 lakh tons in 1949-50 to 56 lakh tons in 1972-73 or 106 per cent. The increase in production was achieved largely by the increase in wheat area which rose from 59.1 lakh acres in 1949-50 to 107.9 lakh acres in 1972-73 or nearly 83 per cent. About 80 per cent of the wheat acreage is irrigated and only 20 per cent is grown in the *Barani* land. The yield per acre which was 12.38 maunds in 1949-50 declined to only 9.17 maunds in 1959-60 but increased sharply to 14.14 maund in 1972-73.

In developing the model we tried first to identify the various factors which affect wheat production. A priori, area under wheat, particularly the area under the high yielding dwarf varieties, fertilizer use, canal water availability, rainfall, temperature and humidity would all affect wheat output. What variables were really significant could be decided only after preliminary graphic analysis of the data. It was found that the most significant variables were: area under high yielding varieties, area under local varieties, fertilizer use and rainfall. Further more it was discovered that the significance of the variables differed between the *Barani* and the irrigated districts. It was decided, therefore, to conduct the analysis separately for: (i) each of the 19 districts in the

*B.A. Azhar is Joint Economic Advisor, Government of Pakistan, Ministry of Finance, Planning and Development, and is Honorary Research Advisor to PIDE; M. Ghaffar Chaudhry and M. Shafique are Research Economist and Research Assistant respectively of PIDE. The authors gratefully acknowledge the help and cooperation of the Planning Unit, (Agriculture Wing), Ministry of Food and Agriculture, Islamabad, in general, and of Mr. S.M. Aslam Jafry, Statistical Officer of the Unit, in particular, for the provision of statistical data required for this study.

province separately, (ii) Four *Barani* districts as a group, (iii)15 irrigated districts as a group; and (iv) 19 districts as a whole. Multiple linear regression was used to analyse the influence of each of the variables listed earlier. Time-series data for the years 1962-63 to 1971-72 and cross section data for various districts were pooled for the analysis. As wheat production showed a spectacular increase from 1967-68 onwards largely due to the large scale introduction of mexican dwarf varieties, it was decided to carry out the analysis separately for the period 1967-68 to 1971-72 in addition to the one for the over-all period.

## II. THE REGRESSION EQUATION AND THE VARIABLES

The methodology consists in specifying the model and determining its parameters. The model used to perform the analysis is linear in form and may be formulated as follows:

$$Y = a_o + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \ldots\ldots a_nx_n$$ where

y is the total production of wheat in thousand tons

$a_o$ is the constant term

$a_1, a_2, a_3, a_4 \ldots\ldots\ldots a_n$ are the parameter estimates corresponding to the explanatory variables $x_1, x_2, x_3, x_4, \ldots\ldots\ldots x_n$

The explanatory variables used in the present model are identified as follows:—

$x_1$ is the area under Mexi-Pak wheat, in thousand acres

$x_2$ is the area under local varieties, in thousand acres

$x_3$ represents the total nutrient tons of fertilizer applied to total wheat acreage

$x_4$ is the total rainfall in inches during the months of November, December and January of each wheat season

The other explanatory variables, that is, temperature, humidity and canal withdrawals were discarded after several trials as their explanatory power turned out to be negligible. While the insignificance of temperature and humidity can be attributed to the constancy or very slow changes of the two variables, the insignificance of canal withdrawals may not make sense to the layman. During the period under study the private tubewells had emerged as an important source of additional irrigation water. Any defficiency of irrigation water resulting from canal withdrawal is more than made up by a greater pumpage from private tubewells. With flexible and assured supply of irrigation water from tubewells, the crop production is unlikely to suffer from the slackness in canal withdrawals.

The wheat price, although possibly an important variable for explaining production variations, has been ignored because response to price of wheat is largely reflected either in wheat acreage or higher use of fertilizer. Since, our analysis included both variables, variations in wheat production due to this variable have been measured implicitly under the cover of wheat acreage and fertilizer use.

## III. EMPIRICAL RESULTS

Estimates of the regression equation for 19 districts of the province as a whole for the periods 1962-63 to 1971-72 and 1967-68 to 1971-72 are shown below:

### TABLE I

*Results of the Regression Analysis for all Districts for (i) 1962-63 to 1971-72 and (ii) 1967-68 to 1971-72*

| Regression No. | Constant | Coefficients | | | | Adjusted $R_2$ | D.W. Statistic | Period |
|---|---|---|---|---|---|---|---|---|
| | | $X_1$ | $X_2$ | $X_3$ | $X_4$ | | | |
| 1.1 | −89.01 | 0.44<br>(10.89) | 0.49<br>(21.09) | 0.03<br>(10.89) | −2.90<br>(−1.34) | 0.92 | 1.36 | 1962-63/ 1971-72 |
| 1.2 | −105.20 | 0.49<br>(8.99) | 0.53<br>(12.65) | 0.02<br>(7.74) | −2.69<br>(−0.77) | 0.92 | 1.45 | 1967-68/ 1971-72 |

Figures in parenthesis indicate t-statistics.

The first regression (1.1) is based on pooled data of 19 districts for 10 years. It has thus 190 observations and 185 degrees of freedom. The second regression (1.2) is based on 95 observations and has 90 degrees of freedom. It can be observed from the table that t-statistic which is used to test whether the parameter estimates are significantly different from zero, show that at a significance level of 5 per cent, coefficients of all explanatory variables except $X_4$ (rainfall) are highly significant. While the first three variables are positively correlated to output Y, the coefficient of rainfall has a negative sign but its influence is statistically insignificant. As will be seen later, rainfall has a significant effect only in the *Barani* areas. In the irrigated areas, whatever effect rainfall has on production is picked up by the variables $X_1$ and $X_2$ i.e., the acreage under wheat. Rainfall seems to have little effect on wheat production in the irrigated districts. The coefficients of multiple determination ($R^2$), adjusted for degrees of freedom, which states the proportion of total variation explained by the independent variables, indicate that both the regressions have very good explanatory power. The Durbin-Watson statistic used to test the presence of serial correlation among the residuals does show some serial correlation but not one that will lead us to reject our findings.

The corresponding regression equations can be formulated as follows:

$$(1.1) = -\ 89.01 + 0.44\,X_1 + 0.49\,X_2 + 0.03\,X_3 - 2.90\,X_4$$

$$(1.2) = -\ 105.20 + 0.49\,X_1 + 0.53\,X_2 + 0.02\,X_3 - 2.69\,X_4$$

The results of regression analysis for the irrigated districts above are given below in table II.

TABLE II

*Results of the Regression Analysis for Irrigated Districts for (i) 1962-60 to 1971-72 and (ii) 1967-68 to 1971-72*

| Regres-sion No. | Cons-tant | Coefficients | | | | Adjusted $R^2$ | D.W. Statistic | Period |
|---|---|---|---|---|---|---|---|---|
| | | $X_1$ | $X_2$ | $X_3$ | $X_4$ | | | |
| 2.1 | − 89.34 | 0.51 (14.93) | 0.54 (26.60) | 0.02 (9.78) | −2.77 (−1.14) | 0.95 | 1.68 | 1962-63/ 1971-72 |
| 2.2 | −113.59 | 0.55 (12.31) | 0.64 (16.91) | 0.02 (6.85) | −6.37 (−1.61) | 0.95 | 1.97 | 1967-68/ 1971-72 |

Figures in parenthesis indicate t-statistic.

The results for the irrigated districts, it will be observed, show an improvement over the results of all the districts. A large proportion of the variation in output is explained by the regressions 2.1 and 2.2 than by regressions 1.1 and 1.2. All explanatory variables have a positive relationship with wheat output and their coefficients are highly significant with the only exception of $X_4$ (rainfall) which is negatively related but is statistically insignificant. The effect of rainfall in the months of September, October and November on wheat production is covered under acreage which is a separate explanatory variable. It is only the effect of rainfall on yield per acre that is intended to be captured by $X_4$ which represents rainfall in the months of November, December and January. The results show that the effect of rainfall on yields per acre is insignificant in the irrigated districts. This is not unexpected. Irrigated districts have assured water supply from the canals round the year and as was pointed out earlier, any defficiency in canal withdrawals is more than made up by a greater pumpage from private tubewells. Moreover, natural precipitation in the months of November, December and January is so meagre that it hardly makes any contribution to yield in the irrigated areas.

TABLE III

*Results of Regression Analysis for Barani Districts*

| Regres-sion No. | Cons-tant | Coefficients | | | | Adjusted $R^2$ | D.W. Statistic | Period |
|---|---|---|---|---|---|---|---|---|
| | | $X_1$ | $X_2$ | $X_3$ | $X_4$ | | | |
| 3.1 | — 28.28 | 0.41 (5.55) | 0.20 (7.63) | 0.02 (2.77) | 4.02 (2.81) | 0.73 | 1.47 | 1962-63/ 1966-67 |
| 3.2 | — 43.79 | 0.46 (5.71) | 0.21 (6.41) | 0.03 (3.58) | 3.79 (2.08) | 0.83 | 1.88 | 1667-68 1971-72 |

Figures in parenthesis show t-statistic.

It will be observed that the parameter estimates of all the independent variables, including $X_4$ (rainfall), are very significant and all variables are positively related to wheat production. Unlike the irrigated areas, rainfall in the months of November, December and January does have a significant influence on yields of wheat. The rainfall in these months would appear to be vital for wheat yield in the *barani* districts. It may be mentioned here that the impact of rainfall from the months of November onward cumulatively till February was investigated in order to discover the most significant period. Further, the average of the rainfall records from different stations in each district was used rather than the readings of one station alone which were found to be undependable.

TABLE IV

*Matrix of Inter-Correlation Coefficients among Explanatory Variables (separately for all, irrigated and Barani Districts and for the two periods*

| Regions and Variables | Matrices of inter-correlation coefficients based on | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1962-63 to 1971-72 data | | | | 1967-68 to 1971-72 data | | | |
| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
| A. All Punjab | | | | | | | | |
| $X_1$ | +1.00 | —0.47 | +0.82 | —0.23 | +1.00 | —0.50 | +0.81 | —0.43 |
| $X_2$ | | +1.00 | —0.12 | +0.14 | | +1.00 | —0.12 | +0.41 |
| $X_3$ | | | +1.00 | —0.19 | | | +1.00 | —0.25 |
| $X_4$ | | | | +1.00 | | | | +1.00 |
| B. Irrigated Areas | | | | | | | | |
| $X_1$ | +1.00 | —0.49 | +0.82 | —0.22 | +1.00 | —0.50 | +0.79 | —0.43 |
| $X_2$ | | +1.00 | —0.12 | +0.22 | | +1.00 | —0.06 | +0.53 |
| $X_3$ | | | +1.00 | —0.14 | | | +1.00 | —0.20 |
| $X_4$ | | | | +1.00 | | | | +1.00 |
| C. Barani Area | | | | | | | | |
| $X_1$ | +1.00 | —0.20 | +0.56 | —0.17 | +1.00 | —0.23 | +0.39 | —0.38 |
| $X_2$ | | +1.00 | —0.09 | —0.15 | | +1.00 | —0.21 | +0.23 |
| $X_3$ | | | +1.00 | —0.11 | | | +1.00 | —0.18 |
| $X_4$ | | | | +1.00 | | | | +1.00 |

The correlation matrix of independent variables shows that $X_1$ (area under Mexi-Pak) and $X_3$ (fertilizer use) are highly correlated particularly in the irrigated areas. In the *barani* areas, the correlation between $X_1$ and $X_3$, is much weaker than in the irrigated areas. Interestingly, the correlation between $X_2$ (area under local varieties) and $X_3$ is very low and negative. There are interesting relationships and explain the factual and observed situation in Pakistan's agriculture. The high yielding varieties (HYVS) of wheat can stand heavy doses of fertilizer, and the response of the farmers to this quality of HYVS has been fairly good. However the response of HYVS to heavy applications of fertilizer is only meagre if these are not accompanied by a corresponding increase in water application because of toxic effect of fertilizer. Since in the irrigated areas, there is a regular system of water supply from canals which is supplemented by liberal supplies from tubewells, the indicated relationship between $X_1$ and $X_3$ is scientifically true. The *barani* areas, on the other hand, have no such mechanism of controlling water supply and hence use little fertilizer for fear of involving toxic effect. The insignificant relation between fertilizer application and local varieties is also understandable. Since the local varieties are liable to lodging even at lighter doses of fertilizer, the farmers hesitate to use fertilizer for fear of suffering a heavier loss due to lodging. The relationship between $X_1$ and $X_2$ is negative, the larger the area under Mexi-Pak, the smaller would be the area under local varieties. The correlation coefficient between $X_1$ and $X_2$ is larger in the irrigated areas than in the *barani* areas.

The high correlation between $X_1$ and $X_3$ introduces multicollinearity in the analysis as a result of which the accuracy of the individual regression coefficients is adversely affected but the regression equation remains valid for prediction purposes.

**Influence of Individual Explanatory Variables**

In order to judge the degree of influence exercised by individual explanatory variables on wheat production it is necessary to have a look at the Beta Coefficients which are given below:—

TABLE V

*Simple Correlation and Beta Coefficients of the Dependent Variable with respect to Independent Variables*

| Region and Dependent Variables | Beta Coefficients based on | |
|---|---|---|
| | 1962-63 to1971-72 data | 1967-68 to 1971-72 data |
| A. All Punjab | | |
| $X_1$ | 0.54 | 0.62 |
| $X_2$ | 0.59 | 0.51 |
| $X_3$ | 0.48 | 0.46 |
| $X_4$ | —0.29 | —0.26 |
| B. Irrigated Areas | | |
| $X_1$ | 0.67 | 0.73 |
| $X_2$ | 0.68 | 0.65 |
| $X_3$ | 0.38 | 0.35 |
| $X_4$ | —0.22 | —0.50 |
| C. Barani Areas | | |
| $X_1$ | 0.59 | 0.71 |
| $X_2$ | 0.66 | 0.65 |
| $X_3$ | 0.28 | 0.38 |
| $X_4$ | 0.24 | 0.22 |

It would be seen that the most important explanatory variable both in the irrigated and the *barani* areas is the acreage under wheat ($X_1$ and $X_2$). Area under Mexi-Pak varieties is the single most important variable in the period 1967-68 to 1971-72. The next in importance is the variable $X_2$ (area under local varieties). Fertilizer use ($X_3$) comes third in the order of importance. Rainfall ($X_4$) is important only in the *barani* areas and occupies the fourth position. In the canal-irrigated areas, rainfall in the months of November to January has little influence on wheat yield.

**Forecasting**

The equations of the regression model can be used to serve as a forecasting device. As with the introduction of Mexi-Pak varieties, a technological break-through was achieved in the production of wheat, the parameters of the regression equation under-went a change in the second period under analysis. It was, therefore, appropriate to use the equations based on this period for forecasting wheat production.

(*i*) *For Irrigated Districts*

(2.2) $= -113.59+0.55 X_1+0.64 X_2+0.02 X_3-6.37 X_4 \ (R^2=0.95)$

(*ii*) *For Barani Districts*

(3.2) $= -43.79+0.46 X_1+0.21 X_2+0.03 X_3+3.79 X_4 \ (R^2=0.83)$

(*iii*) *For all Districts*

(1.2) $= -105.20+0.49 X_1+0.53 X_2+0.02 X_3-2.69 X_4 \ (R^2=0.92)$

In order to guage the usefulness of these equations in forecasting future output of wheat, it would be instructive to have a look at the observed and the estimated values of output in the past years according to the foregoing equations which are given in table VI below.

TABLE VI

*Observed and Estimated Output during* 1967-68 *to* 1971-72

(Figures in million tons)

| Years | Observed values of total output | | | Estimated values of total output | | |
|---|---|---|---|---|---|---|
| | Irrigated | Barani | Total | Irrigated | Barani | Total |
| A. From equations 2.2 and 3.2 | | | | | | |
| 1967-68 | 4.5 | 0.5 | 5.0 | 4.2 | 0.4 | 4.6 |
| 1968-69 | 4.7 | 0.4 | 5.1 | 4.9 | 0.4 | 5.3 |
| 1969-70 | 5.1 | 0.4 | 5.5 | 5.1 | 0.4 | 5.5 |
| 1970-71 | 4.6 | 0.3 | 4.9 | 4.8 | 0.3 | 5.1 |
| 1971-72 | 4.8 | 0.4 | 5.2 | 4.7 | 0.4 | 5.1 |

B. From equation 1.2

| | | | | | | |
|---|---|---|---|---|---|---|
| 1967-68 | 4.5 | 0.5 | 5.0 | 3.8 | 0.6 | 4.4 |
| 1968-69 | 4.7 | 0.4 | 5.1 | 4.6 | 0.7 | 5.3 |
| 1969-70 | 5.1 | 0.4 | 5.5 | 5.0 | 0.7 | 5.7 |
| 1970-71 | 4.6 | 0.3 | 4.9 | 4.6 | 0.6 | 5.2 |
| 1971-72 | 4.8 | 0.4 | 5.2 | 4.7 | 0.5 | 5.2 |

It is evident from the data in the above table that the observed and estimated values of output are very close to each other. The difference between the two values further reduces to insignificance if the estimated values are calculated on the basis of equations separately for irrigated districts, 2.2, and *barani* districts, 3.2. Thus equations 2.2 and 3.2 can be used more successfully in the forecasting of wheat output for the period beyond 1972-73.

In order to forecast the value of the output for 1973-74 and beyond, expected values of the explanatory variables have to be inserted into the predictive equations. These values would be known, at least, three or four months in advance of the actual harvest. District-wise estimates of the area under Mexi-Pak and the other varieties of wheat are generally known by the end of January. Information about the fertilizer use and the rainfall in various districts could also be compiled by the middle of February. All these data can be fed into the regression equations (2.2 and 3.2) and forecasts for each year can be obtained at least three months ahead of the harvest.

It may be mentioned here that a good forecasting model is one which has a high coefficient of multiple determination ($R^2$) even if it may not really explain the underlying structural relationship. The predictive equations developed in this study have high coefficients of multiple determination and should be valid for the next few years.

## IV. CONCLUSIONS

The objective of this paper was to develop a model which could explain and predict changes in wheat production in the Punjab. After trying a number of explanatory variables, the following were selected: area under Mexi-Pak varieties, area under local varieties, fertilizer use, and rainfall. The analysis was done for two periods separately, i.e., for 1962-63 to 1971-72 and 1967-68 to 1971-72. Also, the regressions were run separately for irrigated and *barani* districts in addition to the ones for all the districts together. The analysis was based on pooled data for 10 years and 19 districts.

The results show that the most important variable explaining changes in wheat production during the period 1967-68 to 1971-72 was the area under Mexi-Pak varieties of wheat. The other variables in decending order of importance were areas under local varieties, fertilizer use and rainfall (in *barani* areas). Equations (2.2) and (3.2) were found to be suitable for forecasting purposes. The actual production in recent years was found to be very close to the production that could be forecast by the use of the equations (2.2) and (3.2). Thus a basis has been found for forecasting wheat production in the Punjab for 1973-74 and future years about three months ahead of the harvest.

## REFERENCES

1. Heady, Earl. O. and John L. Dillon, *Agricultural Production Functions*, Iowa State University Press, Ames, Iowa, 1961.

2. Hussain, S.M. "Price Incentives for the Production of High Yielding Varieties of Wheat." *Pakistan Development Review*, Winter, 1970.

3. Oury, Bernard, "Allowing for Weather in Crop Production Model Building" *Journal of Farm Economics*, Vol. 47, No. 2, May, 1965.

4. Oury, Bernard, *Time Series Analysis of West Punjab District Wheat Production Data* (Lahore District), IBRD, Economics Department Working Paper No. 19 (Mimeographed), June, 1968.

5. Qureshi, S.K., "Rainfall, Acreage and Wheat Production in West Pakistan: A Statistical Analysis," *Pakistan Development Review*, Winter, 1969.