AGRICULTURAL TECHNOLOGY IMPROVEMENT PROJECT

(ATIP)

PROGRESS REPORT

PREDICTION OF SORGHUM AND MILLET PROCESSING TIME FOR USE IN THE ECONOMIC ANALYSIS OF ON-FARM TRIALS

NUMBER: ATIP PR M87-3

BY

D. BAKER

DEPARTMENT OF AGRIC. RESEARCH
MINISTRY OF AGRICULTURE
BOTSWANA

MID-AMERICA INTERNATIONAL
AGRICULTURAL CONSORTIUM
(MIAC)

FEBRUARY 1988

PRINTED: 2/2/88

ADDRESSES:

P/BAG 0033
GABORONE

P.O. BOX 10
MAHALAPYE

P.O. BOX 10275
TATITOWN (FRANCISTOWN)

PREFACE

ATIP Progress Reports are prepared and circulated to make ATIP research findings easily available to GOB personnel and researchers interested in Botswana farming systems. A major objective of the Progress Report series is to make research findings available in a timely manner. These reports are not subject to professional review outside the Mahalapye team. Therefore, data and findings presented in the Progress Report series may be subject to further revision and should not be cited without permission of the authors. Therefore they do not necessarily reflect the official views of ATIP, DAR or MIAC.

PREDICTION OF SORGHUM AND MILLET PROCESSING TIME FOR USE IN THE ECONOMIC ANALYSIS OF ON-FARM TRIALS

Doyle Baker

INTRODUCTION

One of the most difficult and expensive tasks in on-farm agronomic experimentation is the collection of data on labour inputs. Few farmers have watches and are able to record and report the time taken for field operations. Even farmers who have watches may not understand when timing should begin and end. As a result, the collection of accurate labour data generally requires that well-trained enumerators be present to time field operations. This severely limits the number of field trials which can be conducted each season.

A top priority for ATIP is to develop low cost procedures for conducting on-farm trials. As one option, we have been exploring the possibility of using predicted labour times for operations which primarily are functions of plot outcomes (or the technical environment). To date, we have concentrated on prediction equations for weeding and processing time. Processing time refers to the total time required for harvesting and threshing.

Use of predicted labour times has three distinct advantages:

(a). Spurious variation (that not due to the treatments being evaluated) is eliminated, allowing a more accurate assessment of the specific effects on farm profitability stemming from the use of different practices.

(b). The need for enumerators to be present during all field operations can be eliminated if reliable prediction can be made on the basis of data which can be collected over a period of days or even weeks.

(c). The marginal cost of including labour time in the analysis of trial outcomes can be greatly reduced if the data required for prediction would have been collected in any event for the agronomic analysis.

This paper reports the results of an attempt to develop a processing time prediction equation for sorghum and millets. The equacion was estimated using an analysis of covariance model. The appendix gives an example of the use of predicted processing labour time in a partial budget analysis.

METHODS

Data collected by the ATIP Francistown team from two trials conducted during the 1985-86 season were used to estimate the prediction equation. The two trials were the Draught Management and Early Ploughing Trial and the Hand Planter Trial[b].

The Early Ploughing Trial involved a comparision of double versus single ploughing. Each participating farmer planted one or two blocks of five plots, of which three were single ploughed and two were double ploughed. Four farmers planted one block and one farmer planted two blocks. Harvest data were collected from four sites, including the farm with two blocks, for a total of 25 plots. Plant stands for the 25 plots ranged from a low of 1,800 plants per hectare to a high of 88,500 plants. Yields ranged from three kgs/ha to 875 kgs/ha. Processing times ranged from 2.5 hrs/ha to 155.8 hrs/ha.

The Hand Planter Trial entailed a comparison of the traditional broadcast-single plough system, the traditional system plus hand gap filling

(where necessary), and row planting with a hand planter. Two replications of each treatment were implemented by three farmers, resulting in a total of 18 plots. The emergence stand counts ranged from a low 16,000 plants per hectare to 98,000 plants per hectare. Yields ranged from 38 kgs/ha to 345 kgs/ha. Processing times ranged from 10 hrs/ha to 50 hrs/ha.

A total of 43 observations were available for estimating the prediction equation. It was hypothesized that most of the variation in processing labour time could be explained by the amount of grain harvested. There should be some economies of scale in harvesting, so the relationship was expected to be curvilinear. It was further hypothesized that for any given amount of grain processed, the processing time would be greater if the grain was taken from a larger number of plants (because of the implicitly smaller heads).

A general function for the prediction equation is

(1) $H = f(Y,P)$

where H is the hours spent harvesting and threshing , Y is the amount of grain harvested and threshed, and P is the number of plants from which the grain was harvested.

To take into account the expected scale economies in processing time, the relationship in (1) can be estimated as a double log function. The double log function to be estimated is

(2) $\ln H_j = \ln(a) + b_1(\ln Y_j) + b_2(\ln P_j) + u_j$

where the H_j, Y_j, and P_j are the processing hours, yield and emergence stand counts on a per hectare basis for plot j; and u_j is a stochastic term. The equation is linear in parameters. Estimates of the parameters can be obtained from single equation ordinary least squares (OLS). If (2) is the proper specification of the relationship, then the estimates for b_1 and b_2 will be consistent[c].

There could be a specification error in (2), in which case single equation OLS will provide biased estimates of the coefficients. Specification bias will occur if H_j, Y_j and P_j are functionally related to a non observable input. One non-observable input for processing time might be the distance of the plot from the threshing floor. Another might be unobserved differences in management practices or management ability. For example, some farmers might obtain higher stands and larger yields because they work more quickly (at planting time and when processing). Specification bias could also occur if H, Y and P are all related to the initial tillage-planting treatments.

If there are significant non-observable inputs associated with management or trial treatments, specification bias can be eliminated by using analysis of covariance. In an analysis of covariance, the variation associated with the non-observed variables is removed by including indicator variables for the different farms or planting treatments. Consider

(3) $\ln H_{jf} = \ln(a_c) + \ln(a_f) + \ln(a_p) + b_1(\ln Y_{jf}) + b_2(\ln P_{jf}) + u_{jf}$

where H_{jf}, Y_{jf}, and P_{jf} are the output, yield and plant stand on farm f on plot j, respectively; a_f is a "farm" variable; a_c represents the general mean, and a_p is a "planting treatment" variable. If there is no interaction between processing efficiency and the initial planting treatment, equation (3) reduces to:

(4) $\ln H_{jf} = \ln(a_o') + \ln(a_f) + b1(\ln Y_{jf}) + b2(\ln P_{jf}) + u_{jf}$
where $a_o' = a_o + a_p$.

Whether [2], [3] or [4] is the proper equation can be tested using an F statistic to evaluate whether there is significant reduction in the residual sums of squares due to the inclusion of the indicator variables. If the farm or treatment indicator variables significantly reduce the residual variation, it would support the conclusion that farm or treatment-related non-observable inputs affect both the independent and dependent variables, and therefore an analysis of covariance must be used to eliminate bias.

If an analysis of covariance must be used, there is a problem in developing a prediction equation. A different constant term is needed for each farm or tillage treatment, but this constant will not be known for new farms or tillage treatments. As a result, the entire goal of developing a prediction equation could be negated. To deal with this problem, the farm or treatment indicator variables can be dropped after the equation has been estimated, and then a new constant can be calculated as follows

[5] $a_c = H/[(Yb_1)*(Pb_2)]$

where H, Y and P are entered at their geometric means. The final prediction equation can be derived by taking antilogs.

The usefulness of the prediction equation can then be checked by evaluating the predicted versus actual processing hours. There should be a high and significant correlation between the predicted and actual hours.

RESULTS

To assess the explanatory power of Y and P alone, equation [2] was first estimated. The results were as follows:

	b	std error	t-statistic		
Constant	.2863			S.E. est.	= .3963
lnY	.6735	.0674	9.9875	F[2,40]	= 57.7
lnP	-.0588	.0796	-.7391	Adj R^2	= .70

Just using the two independent variables, approximately three quarters of the variation in processing time was explained. This degree of explanatory power would ordinarily be acceptable for predicting processing times for the economic analysis of trials. However, the potential problem of specification bias remained.

In a second iteration, equation [4] was estimated using 0,1 indicator variables for six of the seven farms from which data were collected. The hypothesis of no farm effect (af = 0 for all f) was tested using an F test. The F value was 35.6. With 3 and 36 degrees of freedom, this is highly significant and therefore the hypothesis of equal farm (or management) efficiency was rejected. Exclusion of the farm variables should bias the estimated coefficients, b_1 and b_2.

The third step was to estimate the full model (equation 3), including indicator variables for the various tillage-planting treatments. The hypothesis of no tillage-planting treatment effects could not be rejected. As orginally posited, it would not appear that processing time is related to the tillage-planting treatment, other than through plot outcomes and unobserved inputs associated with differences in farm circumstances. Therefore, the model implicit in equation [4] would appear to be the best.

The next step in developing a prediction equation was to eliminate the farm variables not needed from the equation estimated using [4], to improve the efficiency of the estimation. The coefficients for three of the six included farm variables were not significantly different from zero (as determined by the t-statistics of the farm coefficients), indicating there

were not significant inter-farm differences between those farms and the excluded farm. The equation was then re-estimated without the insignificant farm indicator variables. The results of this estimation were as follows:

	b	std error	t-statistic
Constant	-.5041		
lnY	.6543	.0408	16.0424
lnp	.1762	.0476	3.7023
F1	-.4069	.0993	-4.0987
F2	.9973	.1141	8.7404
F3	.4897	.1329	3.5331

S.E. est. = .2069
F(5,37) = 121.2
Adj R^2 = .93

The above equation should enable reliable prediction for the seven farms involved in the trials. This was confirmed by comparing the predicted values for H against the actual values. The correlation between predicted and actual values exceeded .97. A standard deviation was calculated for the differences between the predicted and actual values and only 2 of 43 predicted values were more than two standard deviations from the actual value. Since 2 of 43 observations falling outside of two standard deviations might be expected (if the residuals are normally distributed), no outliers were removed.

To obtain a more general prediction equation, the farm indicator variables were dropped and a new constant was derived using equation (5). After taking antilogs, the final prediction equation was

(6) $H = .8768 \cdot Y^{0.1762} \cdot p^{0.1762}$

where H is hours per hectare, Y is kgs per hectare of threshed grain, and P is plants per hectare divided by 1,000.

To evaluate the usefulness of (6), the predicted values for H were calculated and compared to the actual values for H. The correlation coefficient for H and predicted H dropped to 0.71, still acceptable. However, the distribution for H was skewed, affecting the validity of the correlation coefficient (which assumes a normally distributed population). Consequently, a Spearman rank correlation coefficient was calculated. It was found to be a highly significant 0.83 (with 41 degrees of freedom).

The differences between the actual and predicted values were also calculated and then transformed to eliminate skewness9. The mean and standard deviation of the difference between the actual and predicted values were calculated. Of 43 predicted values, only three fell more than two standard deviations from the mean difference.

DISCUSSION

A problem in the estimation labour prediction equations is the potential for specification bias due to farm specific differences in efficiency. Therefore, the prediction equation was estimated using analysis of covariance. It was not necessary to include indicator variables representing the initial tillage-planting treatments.

It was found that much of the variation in processing time could be explained on the basis of yield and the number of plants in a plot. A high degree of correlation was obtained between the predicted and actual values for processing time even after the farm indicator variables were eliminated and an adjusted constant was calculated. Since measurements of stand counts and yields generally are made for agronomic trials, the need for additional data collection to take into acccount processing time in economic analysis could eventually be eliminated.

The results presented are preliminary, in part because they are based on

relatively small plots. Also, prediction equations should be verified using a new data set. If verified, the prediction equation (8) should be useful for the economic analysis of trials involving a comparison of broadcasted plots. Use of predicted processing times could in fact be preferable to use of actual times -- aside from savings in data collection -- because variation which is not actually associated with cropping outcomes will not incorrectly influence the assessment of trial treatments.

It eventually may be possible to use a comparable approach to predict ploughing and weeding times. Whether this can be done will depend on the quality of the prediction equations which can be developed using easily, and commonly, measured agronomic and environmental variables.

ENDNOTES

a. Attempts also have been made to establish functional relationships between weeding times and weed burdens (measured as the percent weed coverage and the number of types of weeds) on the basis of weeding times collected in FI and AI trials in Mahalapye during the 1985-86 season. Reasonable relationships were established. Selected regression results are presented in D. Baker "Weed Control During Drought," ATIP MP 87-14.

b. Details on trial implementation and outcomes are presented in ATIP PR F86-3 and ATIP PR F86-4.

c. A curvilinear relationship can also be represented using a quadratic function. In a preliminary analysis, both functional forms were estimated. The double log function was chosen on the basis of a smaller standard error of the estimated equation, after standardizing to a common unit of the dependent variable. The double log form also had a higher R^2 and F statistic for the ANOVA of the regression.

d. The F statistic for this test is based on the ratio of: (a) the difference in the residual sums of squares for the full and the reduced models divided by the difference in the number of independent variables in the two models, to (b) the residual sums of squares of the reduced model divided by the number observations minus one more than the parameters to be estimated. The denominators represent the appropriate degrees of freedom for the F test.

e. The appropriate means would be the geometric means since the prediction equation has been estimated in logs.

f. An indicator variable cannot be included for all farms because that would make the matrix of independent variables singular, preventing inversion. The coefficients of the indicator variables for the included farms express differences from the excluded farm.

g. The monotonic transformation entailed adding a constant and taking the natural log.

APPENDIX:

A PARTIAL BUDGET ANALYSIS USING PREDICTED PROCESSING TIME

The value of labour time prediction equations can be illustrated through a simple example based on the results of a farmer implemented double ploughing trial conducted in the Central Region. A partial budget analysis of double ploughing requires a comparison of the net costs and benefits associated with a shift to double ploughing. We will assume that a farmer is considering whether to double plow or to single plough a given hectare.

Based on trials data, it was found that the average ploughing time, when done by donkeys, was 18.1 hrs/ha using two workers, with no significant difference between the first and second ploughing. The average weeding times were 19.8 hrs/ha for single ploughed plots and 16.1 hours for double ploughed plots. Average plant stands were 44.6 (x1,000) plants per hectare for single ploughing and 45.1 (x1,000) plants for double ploughing. Average yields were 89.5 kgs/ha for single ploughing and 173.7 kgs/ha for double ploughing. The predicted values for processing times, using (8) are 25.0 hrs/ha for the single ploughed plots and 38.7 hrs/ha for the double ploughd plots. Valuing grain at 43 thebe/kg and labour at 38 thebe/hr, the net gain associated with a shift from single ploughing to double ploughing (for a user of owned donkey traction) would be as follows:

	Changes	Value	
Reduced Cost:			
SP Weeding Time	19.8 hrs	7.52	
Sp Process Time	25.0 hrs	9.50	
Added Benefit:			
DP Yield	173.7 kgs	74.69	
GAIN SUB-TOTAL			91.71
Added Cost:			
DP Second Plough	2 X 18.1 hrs	13.78	
DP Weeding Time	16.1 hrs	6.12	
DP Process Time	37.8 hrs	14.36	
Reduced Benefit:			
SP Yield	89.5 kgs	38.49	
LOSS SUB-TOTAL			-71.73
NET GAIN			18.98