Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1989 - 1st Annual Conference Proceedings

FORECASTING CORN EAR WEIGHT USING SURFACE AREA AND VOLUME MEASUREMENTS: A PRELIMINARY REPORT

Fatu Bigsby

Follow this and additional works at: <https://newprairiepress.org/agstatconference>

Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)

This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Bigsby, Fatu (1989). "FORECASTING CORN EAR WEIGHT USING SURFACE AREA AND VOLUME MEASUREMENTS: A PRELIMINARY REPORT," *Conference on Applied Statistics in Agriculture*.
<https://doi.org/10.4148/2475-7772.1463>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

FORECASTING CORN EAR WEIGHT USING SURFACE AREA AND VOLUME MEASUREMENTS: A PRELIMINARY REPORT

Fatu Bigsby
National Agricultural Statistics Service
U.S. Department of Agriculture
Washington, D.C.

ABSTRACT

Data from the Corn Ear Weight Study were used to analyze the forecast performance of models estimated using surface area and volume measurements to predict corn ear weight. Two models based on research measurements were compared to models estimated using the operational procedures from the Corn Objective Yield Survey. Research and operational models were estimated both within and across years using data from the 1986 and 1987 Michigan Corn Ear Weight Study. Results show that research models based on surface area and volume measurements have mean square errors that are 32 to 52 percent lower than models estimated using the operational procedure. The differences between forecast errors of the research and operational models are statistically significant.

KEY WORDS: corn; objective yield; ear weight; volume; surface area; forecast error.

1. INTRODUCTION

The National Agricultural Statistics Service (NASS) uses the Corn Objective Yield Survey to forecast corn yield, acreage and production [4]. In the Objective Yield Survey, gross yield is defined as the number of corn ears times the average final ear weight. Net yield, estimated at the state level, is gross yield minus estimated harvest loss. The Corn Ear Weight Study examines one component used to determine yield, specifically, the ear weight.

The purpose of the Corn Ear Weight Study is to develop a more accurate ear weight estimator using surface area and/or volume measurements of the corn ear in addition to the length measurements made in the operational program. Surface area and volume measurements not only have a potential of producing superior forecasts during normal conditions, but especially during drought years. Using ear length alone may not accurately forecast ear weight because of a reduction in the total ear size.

The first part of the paper gives the background of the study. Subsequent sections give the methodology used to derive and validate the forecast models. Results obtained from these methods using 1986 and 1987 data for Michigan are also given.

BACKGROUND

In the Corn Objective Yield Survey a composite of two forecasts (EW_{1ij} and EW_{2ij}) produces the monthly ear weight forecast (EW_{ij}). Current measurements used in regression models provide values for EW_{1ij} and EW_{2ij} . The models give forecasts for maturity categories 3(blister), 4(milk), 5(dough) and 6(dent), which occur before the crop is harvested.

The regression models are constructed from sample level data from the previous five years. A sample contains two units. A unit is defined as two 15 foot rows of corn. Growing season measurements in the units (cob length) and on the first five ears of corn adjacent to one of the units (kernel row length) provide independent variables for the models. The dependent variable is final ear weight. When the corn is harvested, both field and laboratory procedures determine the final ear weight. The enumerator husks and weighs the corn in row 1 of each unit to obtain a field weight. The laboratory technician determines the moisture content and shelling fraction of the third and fourth ears of each unit. An average final ear weight for the sample unit is produced by using the moisture content and shelling fraction to adjust the field weight.

The operational models estimated are:

$$\text{Model 1: } EW_{1ij} = \alpha_0 + \alpha_1(\text{Kernel Row Length}_{ij}) \quad (1)$$

$$\text{Model 2: } EW_{2ij} = \beta_0 + \beta_1(\text{Cob Length}_{ij}) \quad (2)$$

The method used to obtain a weighted estimate follows:

$$\text{Let } R1 = \sum(R^2_{1j} * n_j) / \sum n_j \quad (3)$$

$$R2 = \sum(R^2_{2j} * n_j) / \sum n_j \quad (4)$$

where

R^2_{1j} is the coefficient of determination for the kernel row length model and R^2_{2j} is the coefficient of determination for the cob length model estimated for maturity category j . The number of observations in the j th maturity category is n_j . $R1$ and $R2$ are calculated by summing across maturity categories 3 through 6.

$$W = R1 + R2 \quad (5)$$

$$W1 = R1/W, W2 = 1-W1, \text{ and} \quad (6)$$

$$\text{then the ear weight forecast} = (W1 * EW_{1ij}) + (W2 * EW_{2ij}) \quad (7)$$

Kernel row length (1) and cob length (2) were chosen by prior NASS researchers because of their obvious relationship to ear weight.

Ronald Steele initiated the Corn Ear Weight Study in 1985 [3]. He modeled the relationship of individual surface area and volume measurements on 78 ears of corn to their final ear weight. The ears were obtained from two purposefully selected fields in Nebraska. Analyses of the Nebraska data showed that models derived from surface area and volume measurements were promising. The Corn Ear Weight Study began formally in 1986 in Michigan. The study differs from the Nebraska analysis in that a probability sample is used to forecast ear weight at the sample level. The Michigan State Statistical Office staff collected the data. The author conducted the analyses using SAS[2].

2.

METHODOLOGY

Data Collection

Data collection for the study continued for two consecutive years during the Corn Objective Yield Survey in Michigan [5]. September 1 and October 1 were the survey periods for data collection. The maturity categories for the modeling effort were 4 (milk), 5 (dough) and 6 (dent). Maturity category 3 (blister) was excluded from the study because of the assumption that size of the corn ear, at this stage of maturity, was not large enough to show a significant relationship to the final ear weight.

The enumerators made three measurements on ears in row 1 of the first and second sample units with a vernier caliper. The measurements were: Length of the kernel row, diameter of the cob two inches from the tip, and diameter of the cob one inch from the butt.

Research Models

The ear weight model defined using Least Squares is

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e \quad (8)$$

where,

Y_i = final ear weight for the i th sample field,
 X_1 or X_2 = a surface area or volume or length variable
 e_i = the difference between final ear weight
and the estimate produced by the model.

Although ear length variables were available for inclusion in the research models, the models were forced to contain a surface area and/or volume variable. The maximum number of independent variables in a model was restricted to two because the use of more than two independent variables did not improve the forecasting ability of the models. The models were estimated across maturity categories.

The research models estimated were:

$$\text{Model R1: Ear Weight}_i = \beta_0 + \beta_1 S1_i + \beta_2 BT_i + e_i \quad (9)$$

$$\text{Model R2: Ear Weight}_i = \beta_0 + \beta_1 V2_i + \beta_2 BD_i + e_i \quad (10)$$

Where,

$$\begin{aligned} S1 &= (\text{Husk Length}) * \text{Butt Diameter} \\ BT &= \text{Butt Diameter} * \text{Tip Diameter} \\ V2 &= (\text{Husk Length} * \text{Tip Diameter}^2) / 4 * \pi \\ BD &= \text{Butt Diameter} \end{aligned}$$

The research models were selected because they had lower regression mean square errors and higher R^2 's than other models that were examined.

The operational model provided a benchmark for assessing performance of a research model. The author estimated the operational model using two methods. The background section of this paper gave the first method (1) - (7). The exception occurs in calculating R1 and R2 in (3) and (4). R^2 's for the kernel row length(KL) and husk length(HL) models result from fitting the models across all maturity categories. The operational model estimated using this method is referred to as model O_S.

The second method estimates the relationship of the two operational independent variables(kernel row length and cob length) to the dependent variable in the same model. Estimation of the operational model using this method produces a more meaningful comparison of the operational and research models because the research model defined in (8) is estimated in this manner. O_J, the operational model estimated using the second method is defined as

$$O_J_i = \beta_0 + \beta_1 (\text{Kernel Row Length})_i + \beta_2 (\text{Cob Length}) BT_i + e_i \quad (11)$$

Table 1 contains independent variables, estimated slope coefficients, P values, estimation mean square errors(MSEs) and R^2 's for the research and operational models. It shows that parameters estimated for variables in all models, research and operational, have observed significance levels equal to .0001. Estimation MSEs of the research models are smaller than those of

the operational models and R^2 's of the research models are almost twice as high as those of the operational models. Both these measures of model efficiency confirm the improved properties of the research models over the operational models.

3. VALIDATION OF RESEARCH MODELS

Validation gives an indication of the forecasting ability of a model by determining how the model performs using data other than that used for its estimation. The 1987 data was used for validating the 1986 models and vice versa. Data for validating the models based on the combined data for 1986 and 1987 included a random sample of one-half of the observations from September. These earlier observations provided a better test of the ability of the models to forecast than do the October observations.

The validation procedure used mean square errors to compare the performance of two or more models. The mean square error was computed from the forecast error. The forecast error is the difference between the final ear weight for the i th sample field and the ear weight forecast for the same field. Table 2 gives the MSEs of the research and operational models and shows that both research models have smaller mean square errors than the operational models for all validation data. Furthermore, Table 2 shows that the reduction in MSE when forecasting with the research model is at least 32 percent.

Since the MSEs indicated differences in the abilities of the models to forecast ear weight, conducting a one-way analysis of variance (ANOVA) determined if the differences were statistically significant.

A one-way ANOVA defined the following model:

$$Y_{ij} = \mu. + \tau_j + e_{ij} \quad (12)$$

Where,

Y_{ij} = the absolute value of the forecast error
for the i th sample field and j th model,
 $\mu.$ = the overall mean,
 τ_j = the effect of of the forecast error for
the j th model, and
 e_{ij} = the deviation of $(\mu. + \tau_j)$ from Y_{ij} .

The absolute value of the forecast error was considered a response because of interest in the absolute deviation of the forecast from the final ear weight. The analysis of variance used the model in (12) to test the following hypotheses:

$H_0: \tau_1 = \tau_2 = \tau_3 = \dots \tau_k$, k = the number of models
 H_a : not all τ_j 's are zero, $j = 1, 2, \dots, k$

The test chosen was nonparametric since the statistical distribution of the absolute values of the forecast errors was unknown. The test statistic [1] used was:

$$F_R = \frac{[\sum \{ (R_j^2/n_j) - (n(n+1)^2/4) \}] / (k-1)}{\sum \sum (R(X_{ij}) - R_j)^2 / (n-k)} \quad (13)$$

where,

$R(X_{ij})$ = rank of absolute value of the
 ith forecast error for the jth
 model when forecast errors for
 all models are ranked jointly
 from 1 to n,
 R_j = sum of the ranks for model j's
 forecasts,
 $R_{.j}$ = average rank for model j's
 forecasts,
 n_j = the number of observations for
 model j, and
 n = total number of observations.

The numerator in F_R (13) represents differences between ranks of the k models and the denominator represents differences within ranks of the k th model. If the null hypothesis is true and the forecast effect is essentially zero, then F_R will be small. The test rejects H_0 for values of F_R larger than the $(1-\alpha; k-1, n-k)$ percentile of the F distribution. The null hypothesis is rejected for observed significance levels (P values) smaller than or equal to 0.10.

Table 3 gives the nonparametric ANOVA results. For each type of validation data, two types of comparisons were made. The first comparison examined differences among forecast error means of research model R1 (9) and operational models O_S (7) and O_J (11). Likewise, the second comparison examined differences among forecast error means of research model R2 (10) and operational models O_S (7) and O_J (11). The last column of the table shows observed significance levels for each type of validation data. The observed levels are all less than .02 which leads to rejection of the null hypothesis that the forecast effect is essentially zero.

Although Table 3 shows that at least two of the three models in each type of comparison have significantly different forecast errors, the table does not show which two are different. The Tukey multiple comparison procedure determined which forecasts were different from each other and controlled the experimentwise

error rate at .10. The procedure was performed on ranks of absolute values of the forecast errors. Confidence intervals were constructed simultaneously for all pairs of forecasts.

The interval

$$D - Ts(D) \leq \mu_j - \mu_j' \leq D + Ts(D) \quad (14)$$

determined the difference between errors of the j th and j th' forecasts. D is the difference between average ranks of the errors, s is the standard deviation of D , and T is the $(1-\alpha, k, n-k)$ percentile of the studentized range distribution. Only if the interval constructed in (14) contains zero will the forecast errors be statistically the same. Otherwise the forecast errors are different. Table 4 gives simultaneous upper and lower confidence levels constructed for the validation data sets. The table shows that differences between forecast errors of the research and operational models are significant.

4.

SUMMARY

Data from the 1986 and the 1987 Michigan Corn Ear Weight Study were used to estimate ear weight models employing surface area or volume measurements as independent variables. These models had mean square errors that are 32 to 52 percent lower than those for operational models for the 1986 and 1987 validation data. Mean square errors of the research models estimated using data for both years were 44 to 48 percent lower than those of the operational models for the same period. Tests of statistical hypotheses and multiple comparison procedures conducted suggested that differences between forecast errors of the research and operational models were significant.

The Corn Ear Weight Study continued in Michigan in 1988 and was also expanded to Missouri to evaluate the performance of the surface area and volumes models in an environment that is more drought prone than Michigan. Results of analyses on the 1988 data will be available at a later date.

5.

TABLES

Table 1. Estimation statistics for Objective Yield
research and operational models
constructed within and across years.

Year	Model	Independent Variables	Estimated Coefficients	P Value	Estimation MSE	R ²
1986	R1	S1	.000519	.0001		
		BT	.000104	.0001	.00199	.6092
	R2	V2	.000103	.0001		
		BD	.010303	.0001	.00205	.5975
	O_J	KL	.020516	.0001		
		HL	.030677	.0001	.00322	.3670
	O_S	KL	.024564	.0001	.00419	.1698
		HL	.034228	.0001	.00378	.2513
1987	R1	S1	.000386	.0001		
		BT	.000141	.0001	.00238	.7256
	R2	V2	.000115	.0001		
		BD	.009666	.0001	.00237	.7261
	O_J	KL	.021675	.0001		
		HL	.031384	.0001	.00476	.4512
	O_S	KL	.032526	.0001	.00586	.3138
		HL	.044579	.0001	.00568	.3396
1986 & 1987	R1	S1	.000432	.0001		
		BT	.000126	.0001	.00242	.6738
	R2	V2	.000109	.0001		
		BD	.009709	.0001	.00242	.6740
	O_J	KL	.021245	.0001		
		HL	.030124	.0001	.00428	.4230
	O_S	KL	.030663	.0001	.00525	.2884
		HL	.041547	.0001	.00510	.3079

Table 2. Mean square errors for objective yield research and operational models.

Estimation Data	Validation Data	Model	Variables	MSE	Percent Reduction in Forecast Error	
					R1	R2
1986	1987	R1	(S1,BT)	.00258		
		R2	(V2,BD)	.00257		
		O_J	(HL,KL)	.00476	45.8	46.0
		O_S	(HL,KL)	.00540	52.2	52.4
1987	1986	R1	(S1,BT)	.00219		
		R2	(V2,BD)	.00221		
		O_J	(HL,KL)	.00327	33.0	32.4
		O_S	(HL,KL)	.00328	33.2	32.6
1986 & 1987	1968 & 1987	R1	(S1,BT)	.00177		
		R2	(V2,BD)	.00183		
		O_J	(HL,KL)	.00327	45.9	44.0
		O_S	(HL,KL)	.00345	48.7	47.0

Table 3. Analysis of variance results for determining significance of differences between means of forecast errors of operational models and each research model.

Estimation Data	Validation Data	Models	Number of Obs.	MSTR	MSE	F _R	Pr > F
1986	1987	R1,O_S,O_J	393	93868	12488	7.52	.0006
		R2,O_S,O_J	393	91566	12500	7.33	.0008
1987	1986	R1,O_S,O_J	390	57244	12477	4.59	.0107
		R2,O_S,O_J	390	54348	12492	4.35	.0135
1986 & 1987	1986 & 1987	R1,O_S,O_J	228	36012	4070	8.85	.0002
		R2,O_S,O_J	228	30962	4114	7.53	.0007

Table 4. Tukey's studentized range test for differences
between ranks of forecast errors
(*** Comparison significant at the .10 level).

Estimation Data	Validation Data	Treatment Comparison	Simultaneous	
			Estimation Confidence Limits	Validation
1986	1987	R1-O_J	(-60.74 , -3.90)	***
		R1-O_S	(-81.54 , -24.70)	***
		O_J-O_S	(-49.22 , 7.62)	
		R2-O_J	(-60.06 , -3.19)	***
		R2-O_S	(-80.95 , -24.08)	***
		O_J-O_S	(-49.32 , 7.55)	
1987	1986	O1-O_J	(-65.38 , -8.34)	***
		R1-O_S	(-64.33 , -7.29)	***
		O_J-O_S	(-27.46 , 29.57)	
		R2-O_J	(-64.20 , -7.13)	***
		R2-O_S	(-63.69 , -6.62)	***
		O_J-O_S	(-28.02 , 29.05)	
1986 & 1987	1986 & 1987	R1-O_J	(-55.20 , -12.51)	***
		R1-O_S	(-61.98 , -19.29)	***
		O_J-O_S	(-28.12 , 14.57)	
		R2-O_J	(-52.75 , -9.82)	***
		R2-O_S	(-59.20 , -16.27)	***
		O_J-O_S	(-29.92 , 15.01)	

REFERENCES

1. Conover, W.J., and Ronald L. Iman. Rank Transformations as a Bridge Between Parametric and Non-parametric Statistics, The American Statistician, August 1981, Vol. 35, No. 3.
2. SAS Institute Inc. SAS/STAT Guide for Personal Computers, Version 6 Edition. Cary, NC: SAS Institute inc., 1985
3. Steele, Ronald J., 1985. Preliminary Investigation For An Improved Corn Grain Weight Forecast Model, Unpublished.
4. U.S. Department of Agriculture, 1987 Corn Objective Yield Survey: Supervising & Editing Manual, National Agricultural Statistics Service, Washington, D.C., 1987
5. U.S. Department of Agriculture, 1987 Michigan Corn Research Study, National Agricultural Statistic Service, Washington, D.C., 1987