ACCURACY OF PAST FORECASTS AND PROPOSALS FOR FUTURE AAEA OUTLOOK SURVEYS

A. Gene Nelson and James Cornelius*

The purpose of this presentataion is two-fold. The first is to review the results of past outlook surveys and evaluate forecasting accuracy and performance. This section of the paper will demonstrate a methodology for conducting such a review and suggest some factors that seem to affect the accuracy of the responses. Is it easier to forecast prices for some commodities than others? In which years were respondents more accurate? Are certain respondents more accurate than others?

The second purpose is to discuss some issues and propose some alternatives for conducting the survey in future years. The discussion here will center on the objectives of the survey, what variables to forecast, probabilistic forecasts, aggregation of survey results, and the selection of survey respondents.

REVIEW OF PAST SURVEYS

The AAEA outlook surveys in 1978 and 1979 have been summarized [Ikerd] and briefly reviewed [Futrell] as a first step in evaluating the survey results. The following section is a somewhat more indepth assessment of the 1978 and 1979 surveys, employing both graphical presentation and various statistical comparisons. This is designed to demonstrate additional methodology for evaluating the outlook survey results, as well as judging the appropriateness of the current form of the survey itself.

---

* The authors are Professor and Assistant Professor of Agricultural and Resource Economics at Oregon State University. This paper was prepared for presentation at the Outlook Symposium, AAEA Meetings, Urbana, Ill., July 27-30, 1980.

Procedures Used

In order to facilitate analysis of survey results, the individual responses were coded on computer cards, identifying the respondents by code and their responses to the survey outlook queries for both 1978 and 1979. This format will allow for adding new respondents in succeeding years and additional years for established respondents. An extensive field is required to encode successive years' responses. For example, to encode the first two years' results for an individual required 10 eighty-field computer cards.

Some additional difficulties became obvious in this coding process. The consistency of response format varied among individuals. Most identified point estimates, but others specified a range of responses. The accuracy of a range estimate is more difficult to assess than a point estimate, but the range estimate probably contains more "probabilistic" information. For this analysis the high and low of the range were averaged to develop a point estimate.

Interpretation of Results

The distributions in Figure 1 (atttached at end of paper) present a graphic review of respondents' forecasting performance. For the purpose of this brief evaluation, only the price forecasts were considered, and then only for major commodities; slaughter cattle, feeder cattle, hogs, wheat, corn, and soybeans.

Each individual histogram can be explained as follows. The commodity forecast is identified by the year of the forecast followed by the year for which the forecast is made. Thus, "78/79" is a forecast made in 1978 for the year 1979. The selection of intervals for the distribution was arbitrary, based on the range of forecast error. The interval lengths were specified

such that a "perfect" forecast falls between two intervals, and due to the discrete nature of the distribution these were accumulated in the first positive interval to the right of zero.

For each of the six commodities shown, the distribution of forecast error is presented for: 1) the 78 forecast of the 78 price; 2) the 78 forecast of 79 price; 3) the 79 forecast of 79 price; and 4) the 79 forecast of 80 price. Regarding the latter, since actual prices for the complete 1980 marketing year are not yet known, representative first quarter, or other abbreviated year estimates are utilized and noted. No suitable 1980 figures were obtained for soybean price, so the 79/80 evaluation is not shown.

In addition to the distribution of forecast error, additional evaluation criteria statistics are listed beneath each graph. These include: N, the number of respondents; ME, or mean error of the forecasts; RMSE, or root mean squared error which is the squareroot of the average of the squared forecast errors; R, or range of forecast error; and the number of outliers falling beyond the interval shown.

In addition, the mean percentage forecast errors are given in Table 1. The above evaluation statistics by no means exhaust the possibilities. Depending upon one's objectives in evaluating economic forecasts, it might be appropriate to include tracking measures such as indices of forecast error over and under actual values, error decompositions, and cyclical or dynamic properties of the responses [Dhrymes, P. et. at., Mincer and Zarnowitz, Granger].

Assessment of the 1978 and 1979 Surveys

Returning to Figure 1, a few tentative conclusions can be drawn regarding the responses. First, for the commodities shown, forecasts made for the same time period tend to be both more accurate, and contain a smaller root

mean square error and percentage error than those estimates made for the next year. Intuitively, there is more information available on the short run than for the longer run. Another way of looking at this is to compare the estimates made in 1978 for the 1979 year with the estimates made in 1979 for the 1979 year. In all cases, the 1979 estimates were more efficient (smaller root mean squared error) than the 1978 estimates for 1979.

A second generalization which can be drawn from these data is that the accuracy of forecasts also varies significantly among commodities. For current year forecasts (that is, forecasts for the year in which the estimate is made) soybean price forecasts have been much more accurate than wheat. This is somewhat surprising given that the current year price forecast for wheat is essentially an ex-post observation, whereas the soybean forecast covers at least a three month projection into the future. In prospective, over the short time interval sampled, the distribution of random error among the commodities selected may be unequal, a longer time interval would be more appropriate for intercommodity forecast error comparisons.

Lastly, there is the intriguing possibility, even with only two years' data, to assess the "track record" of individual forecasters who responded to both surveys. For example. for those respondents who forecast slaughter cattle price in both 1978 and 1979, mean squared errors or average percent errors can be calculated to evaluate "track records" of forecaster accuracy. One such distribution is illustrated in Figure 2. Some revision of the outlook survey procedures may be necessary in order to utilize the results of this type of evaluation, because the present survey format pledges anonymity of the respondents.

Proposals for Future Surveys

After two years of experience with this annual AAEA outlook survey, it is timely to review the activity and consider alternative courses of action.

The following sections discuss some issues and propose some alternatives for conducting the survey in future years.

Objectives of the Survey

The primary purpose when the survey was initiated in 1978 was to "provide a stimulus for discussion of the agricultural outlook at the annual meeting" [Ikerd, 1978]. In 1979, minor revisions were made in the survey form and procedures, but the symposium format and purpose were essentially unchanged (Ikerd, 1979]. It was also indicated that if as a result of this effort the AAEA membership becomes interested in a more scientific survey of the membership, different procedures would have to be developed. The format for this 1980 symposium has been altered to include discussion of outlook methodology and to begin a formalized process for analyzing past results.

The experience to date has been beneficial, and these two years provide a background for considering changes in approach and procedures for future surveys. The first step in this planning process should involve specifying the objectives for this effort. There are a number of possibilities that we will list for consideration.

1. *To share outlook information from the survey with AAEA members attending the symposiums held at the AAEA annual meetings.*

This objective is consistent with the current activities, however, it might be appropriate to consider a wider dissemination of the survey results.

2. *To disseminate the survey results to survey participants and the AAEA membership through publication in the Journal or the AAEA Newsletter.*

A more ambitious objective would be:

3. *To disseminate the survey results to the general public through the agricultural press.*

In addition to the composite forecasts and statistical summaries of the survey results, additional valuable information would be the track record of the forecast.

4. *To provide a data base for assessing the performance of cooperating AAEA members in making outlook forecasts.*

Given the experience from this survey effort over time, there is an opportunity to do additional analysis.

5. *To determine what factors influence forecasting accuracy and evaluate methods for improving forecasting performance based on the analysis of past forecasting experience.*

Given enough time and experience with this annual survey effort, it would be possible to determine what commodities are more amenable to price forecasting and which are more difficult to forecast accurately. It would also be possible to identify which forecasters are more accurate and why. Also, new approaches might be found for developing and disseminating outlook information.

What Variables to Forecast

The 1980 outlook survey form allows for forecasting 62 variables relating to production and prices of agricultural commodities, as well as other cost and return factors. This year for the first time, respondents are asked to specify those areas where they have major forecasting responsibility, and make estimates only in those areas in which they possess "professional competence."

Should variables be added or deleted from this list? The criteria for making these changes depends on the objectives of the effort. The ultimate use of these survey results is for decision making. Both policy makers and agricultural producers are interested in what the future holds for agricultural production and prices. We expect that the primary function of those participating in the survey

is to provide outlook information for agricultural producers. Their interest will likely center on the price forecasts from the survey. While the production, supply, and utilization data may provide the background or justification for the price forecast, the question is what prices will do.

An aspect which should be reviewed is the set of commodities for which forecasts are made. The tendency is to deal with those commodities that have the largest value of production, however, it may be that this outlook information would be equally or more valuable for some of the less significant commodities whose prices are subject to greater uncertainty and for which futures markets do not exist [Gardner].

Another dimension to these forecast variables is the time period for which the forecasts are made. There is some ambiguity in the questionnaire regarding calendar years and market years. In the first two surveys the wheat price is made after the marketing year has ended. While the "official" average price has not been released when the forecasts are made, making a forecast for this time period is a moot point.

Another time related issue concerns the use of this information in a decision-making context. Annual forecasts are of limited use to decision makers. While annual prices might provide some basis for determining the mix of commodities to produce, buying, storage, and other marketing decisions require forecasts indicating trends within years. Therefore, we would recommend that quarterly forecasts be requested for at least the major crop and livestock commodities. This would multiply the number of variables to be forecast. But there would appear to be opportunities for replacing some of the lower priority variables that are now included, such as cash receipts from all crops and products, cash receipts from all livestock and products, land values, livestock and poultry production, and crop supply and utilization.

Care should also be exercised in defining the weights and grades for the commodities to be forecast. For example: the weight range for the choice slaughter steer price at Omaha would be helpful. The same is true for barrows and gilts, seven markets. Also, because of the large number of classes for wheat, additional categories and market locations might be added.

Probabilistic or Point Forecasts

The present survey calls for point estimates regarding the future variables. A few respondents, however, have indicated ranges or intervals rather than point estimates. But no information was provided to help in interpreting these intervals. Are they 100 percent confidence intervals or 50 percent intervals? With the growing acceptance of probabilistic information, there is an alternative to this traditional approach of making point estimates.

While there are a few examples of the estimation and use of probabilities in business management such as banking [Kabus], the potential has yet to be realized. The most extensive experience with the use of probabilities is in weather forecasting. Since 1965, probability of precipitation forecasts have been routinely formulated and disseminated to the general public by the National Weather Service. Although there was some initial resistance, it is now generally agreed that these probabilities are an important and integral part of the Services' public weather information [Murphy and Winkler].

Because publicly-employed outlook specialists are not usually directly involved in the production or marketing of the commodity for which they are making forecasts, they can be more objective in quantifying their probabilities. They are not as likely to be influenced by what they hope will happen as those whose financial situation is dependent upon the outcomes. Outlook specialists can place their probabilities in a broader perspective, providing a clearer indication of the worst and best that can happen.

There is a lack of research testing alternative methods for requesting probabilistic forecasts. Two approaches that appear to be the most feasible for the outlook survey would be (1) use of the triangular probability function and (2) assigning "weights" measuring strengths of conviction.

Triangular probability distributions [Cassidy, et. al.] can be quantified by specifying three parameters: the lowest possible event, highest possible, and modal, i.e., most likely, event. This approach is an easy method for eliciting respondents' beliefs. However, accuracy may be lost due to the rigidity of the form of the distribution.

A procedure of assigning weights which measures the strength of the respondent's conviction that particular events will occur has been informally tested with a variety of audiences. Briefly, the approach can be summarized as follows:

1. Discrete events are defined, using a total of 8 to 12 outcomes to describe the phenomenon.
2. The respondents are asked to identify the lowest and highest events thought possible. All events outside these extremes are assigned weights of zero.
3. They are then asked to consider the general shape of their probability distributions, e.g., uniform, normal, skewed right, etc.
4. They assign a weight of 100 to the most likely event, and weights of 1 to 100 to the remaining events, consistent with their information and beliefs.
5. The weights for each individual event are divided by the total of all the weights to calculate the probability for each event.

This approach is easily explained. However, additional work is needed to compare this approach with other alternatives to determine how well this approach reflects the respondent's true beliefs. For a more thorough treatment of the various methods for eliciting probabilities, see Spetzler and

Stael von Holstein. They recommend a structured interview process and various techniques to reduce biases in quantifying the subjective probabilites.

Probabilities, by no means, preclude the use of data analysis, including econometric models and time series techniques, in outlook forecasting. Weather forecasters often use "guidance" from statistical models in forming their probability of precipitation forecasts [Murphy and Winkler]. Hogarth argues that probability assessment can add something to predictive accuracy over and above that which can be achieved by the best available statistical model. This has been shown to be the case he states, for substantive experts in medicine and meteorology.

How to Aggregate the Results

Probabilities provide the mechanism for aggregating the information from several experts and efficiently communicating it. A weighted average of the individual assessments can be used to pool the information and form the aggregate probability distribution for the group of respondents. Initially, a simple weighting scheme which allocates equal weight to the probabilities of each specialist on the panel could be used. After more experience accumulates, differential weights might be based on the past performance of the individual cooperating specialists. There are various "scoring rules" which could be used for determining these differential wieghts.

Bessler [1977] discusses some alternative scoring rules that might be used. The "proper" scoring rule that appears to be most applicable in this situation is the quadratic scoring rule. To illustrate how this rule might be used to derive weights for aggregating the probability forecasts, suppose there are n mutually exclusive and collectively exhaustive events E_1, E_2, ..., E_n. The specialists probability forecasts are denoted as p_1, ...,

p_n, and d_1, ..., d_n represent the actual outcomes, where $d_j = 1$ if E_j occurs and zero, otherwise.

The quadratic scoring rule is defined as:

$$Q\ (p,d) = [1- \sum_{i=1}^{n} (P_i - d_i)^2]$$

If event E_j occurs, $d_j = 1$ and $d_i = 0$ for all $i \neq j$. Thus,

$$Q_j\ (p,d) = (2p_j - \sum_{i} p_i^2)$$

Based on this score, the weight to assign to the probabilities of these events next year could be set as follows:

$$w = 2 + Q_j\ (p,d)$$

With this weighing scheme, a weight of four would be used for a "perfect" forecaster.[1/] The forecaster who misses it completely would have weight of 1.

The weights for this set of events would be summed for all m forecasters, divided by the total, and multiplied times their probability forecasts to derive the aggregate probabilities.[2/] The aggregate probability of event E_i would be:

$$P_i = \left(\sum_{k=1}^{m} w_k p_{ik} \right) / \sum_{k=1}^{m} w_k$$

As additional years of experience accumulate for each forecaster, a weighted average scheme could be used to derive the weights. This scheme should weight the more recent scores more heavily. An exponential moving average would be a good possibility because of its computational ease [Brown]. The following formula would base half the weight on the score for last year's forecast and the other half on previous years performance for forecaster k:

[1/] To be perfect, the forecaster must have assigned a probability of one to the event that actually occurs.

[2/] Forecasters with no history could be given weights of 1.

$$w'_{kt} = .5\, w_{k\ t-1} + .5\, w'_{k\ t-1}$$

where,

w'_{kt} is the average weight for the forecasts in year t

$w_{k\ t-1}$ is the actual weight based on the forecasts made in year t-1

$w'_{k\ t-1}$ is the average weight for the forecasts in year t-1

Care and Feeding of Survey Participants

In the past three surveys respondents have been "selected" on an informal, volunteer basis. This approach presents some problems in developing continuity. It is difficult to accumulate a data base for evaluating track records if the pool of respondents changes in composition from year-to-year. Alternatives should be explored for follow-up and feedback to maintain a more consistent set of respondents each year. The goal might be to develop a panel for each commodity consisting of those "experts" with the best forecasting performance.

Winkler and Murphy describe the expertise required for making probability forecasts. There is "substantive" expertise that relates to knowledge about the specific phenomena involved. Then there is "normative" expertise which concerns the ability to express this knowledge as probability forecasts. They observed that it is easier to train substantive experts to have normative skills, than it is to take individuals with normative abilities and also make them substantive experts.

To help individuals on the panel become "better" forecasters, feedback is needed. The results from comparing past forecasts to actual outcomes can help the forecasters to evaluate and improve performance. This feedback information (possibly in the form of scoring rule scores) should help the outlook specialist understand the correspondence between beliefs and

probabilities. Also, by providing comparisons among forecasters, those with "poorer" performance records could be lead to reevaluate the substantive and normative processes they used to make their probability assessments. For further discussion of how to identify and improve forecasting performance see Winkler and Murphy.

Further research is needed in this area of agricultural price forecasting performance. There would appear to be significant opportunities for improvement. A good, recent example of such research is provided by Brandt and Bessler. This symposium represents a positive step toward focusing increased attention on this important function of agricultural economists.

Summary

The review of past surveys has demonstrated both graphical and statistical techniques for evaluating the performance of outlook forecasts. Preliminary comparisons of both intercommodity and intertime period survey results point to some intriguing issues in forecasting accuracy. However, the limitations inherent in a two year data series must be recognized in drawing conclusions concerning forecasting performance.

Relative to the planning of future surveys there are several questions to be discussed:

1. What objective should be set for this annual survey?
2. What information should be disseminated from the survey and to whom?
3. What variables should be forecast and for what time periods?
4. Should respondents provide probabilistic forecsts rather than point estimates?
5. What type of weighting scheme should be used for aggregating or summarizing the forecasts?

6. Should procedures be changed to maintain the same panel of respondents each year?

7. What methods should be used for reviewing the performance of the outlook panelists and providing feedback?

8. What research should be initiated to improve overall forecasting performance?

References

1. Bessler, David. "Foresight and Inductive Reasoning: Analysis of Expectations on Economic Variables with California Field Crop Farmers." Ph.D. dissertation, University of California-Davis, 1977.

2. Brandt, J. A., and D. A. Bessler. "Composite Forecasting: An Application with U.S. Hog Prices." American Journal of Agricultural Economics, 63 (1981): forthcoming.

3. Brown, R. G. Smoothing, Forecasting and Prediction of Discrete Time Series. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

4. Cassidy, P. A., J. L. Rogers, and W. O. McCarthy. "A Simulation Approach to Risk Assessment in Investment Analysis." Review of Marketing and Agricultural Economics, 38 (1970): 3-24.

5. Dhrymes, P., E. P. Howrey, et. al. "Criteria for Evaluation for Econometric Model," Annals of Economic and Social Measurement, 1 (1972): 291.

6. Futrell, Gene. "Review of 1978 AAEA Outlook Survey Results." Iowa State University, n.d.

7. Gardner, B. L. "Futures Prices in Supply Analysis," American Journal of Agricultral Economics, 58 (1976): 81-84.

8. Granger, C. Forecasting in Business and Economics. New York: Academic Press, 1980.

9. Hogarth, R. M. "Cognitive Processes and the Assessment of Subjective Probability Distributions." Journal of American Statistical Association, 70 (1975): 271-294

10. Ikerd, J. E. "AAEA Survey of Annual Outlook Information: Summary of 1978 Survey." Paper presented at AAEA Meetings, San Diego, Calif., July 1978.

11. Ikerd, J. S. "AAEA Survey of Annual Outlook Information: Summary of 1979 Survey." Paper presented at AAEA Meetings, Blacksburg, Virginia, July 1979.

12. Kabus, Irwin. "You can Bank on Uncertainty." Harvard Business Review, 54(3) (1976): 95-105.

13. Mincer, J. and V. Zarnowitz. "The Evaluation of Economic Forecast." Economic Forecast and Expectations, J. Mincer (Editor), New York: National Bureau of Economic Research, 1969.

14. Murphy, A. H. and R. L. Winkler. "Probabilistic Temperature Forecasts: The Case for an Operational Program." Bulletin of American Meteorological Society, 60 (1979): 12-19.

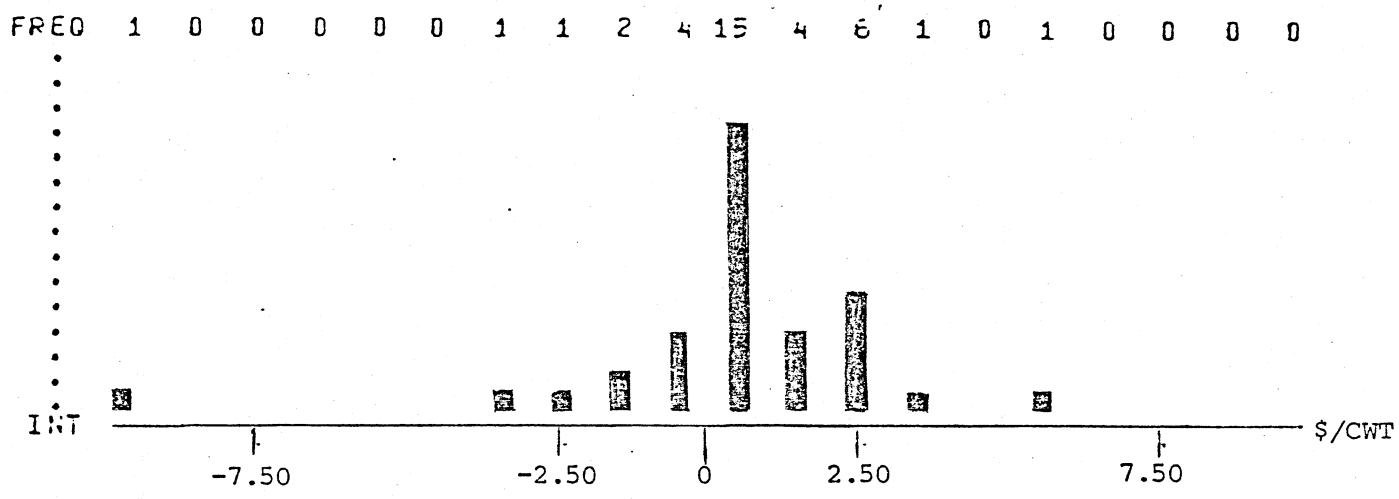
15. Spetzler, C. S. and C. A. Stael von Holstein. "Probability Encoding in Decision Analysis." Management Science, 22 (1975): 340-358.

16. Winkler, R. L., and A. H. Murphy. "Good Probability Assessors." Journal of Applied Meteorology, 7 (1968): 751-758.

Figure 1.

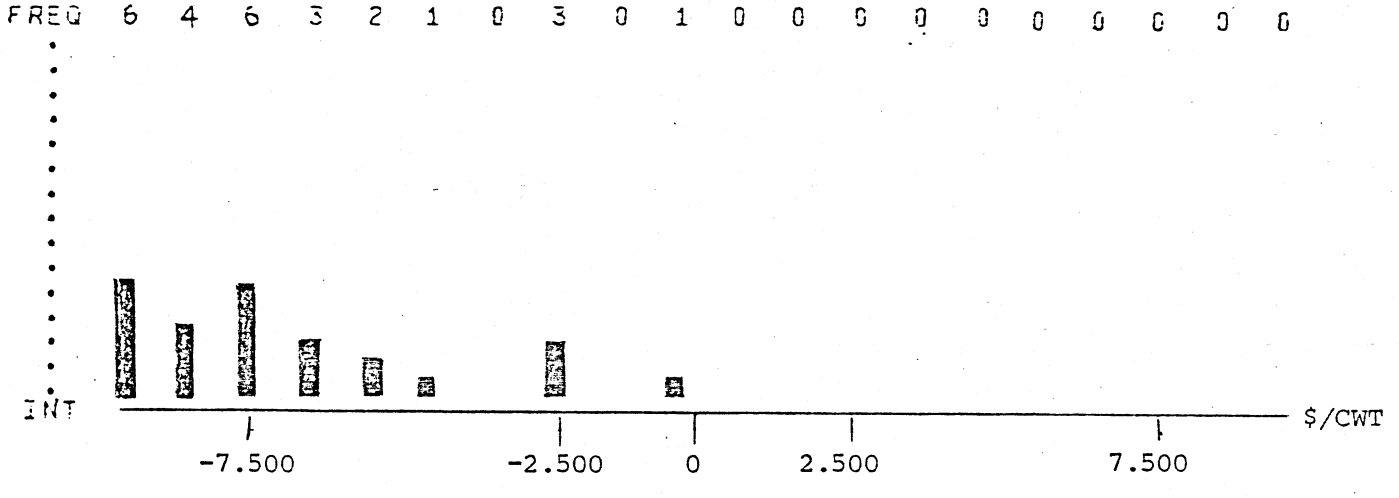
DISTRIBUTION OF PRICE PREDICTION ERROR; PREDICTED MINUS ACTUAL

1. CHOICE SLAUGHTER STEERS - 1978/78



N = 36; ME = .58; RMSE = 2.34; R: -9.05 to 5.66; Outliers = Ø

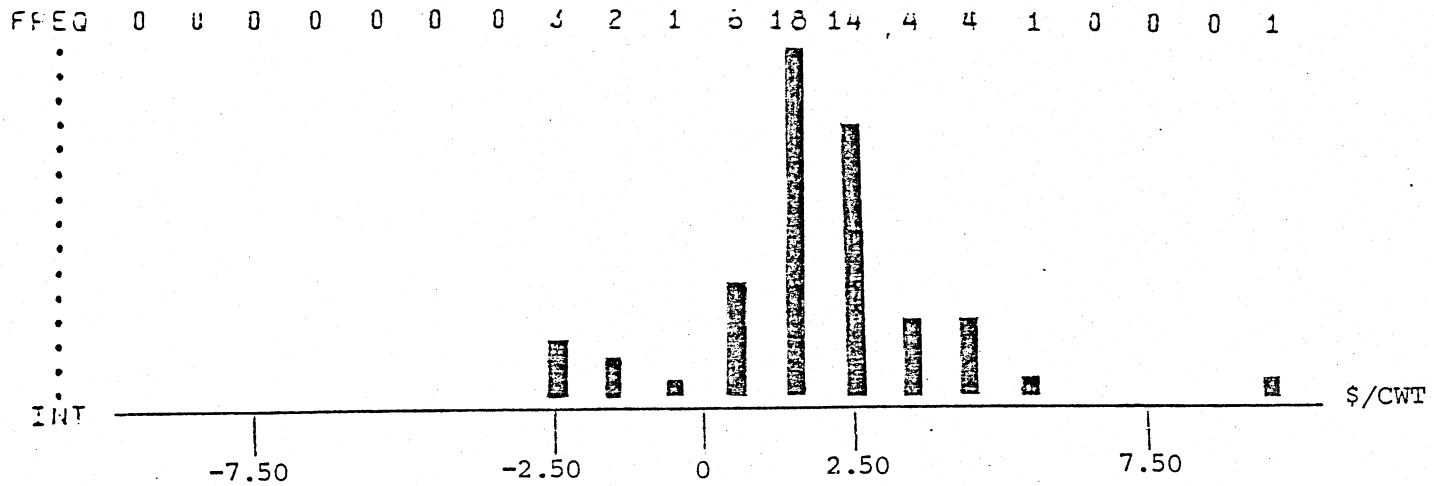
2. CHOICE SLAUGHTER STEERS - 1978/79



N = 35; ME = -8.77; RMSE = 4.14; R: -23.45 to -.79; Outliers = 9

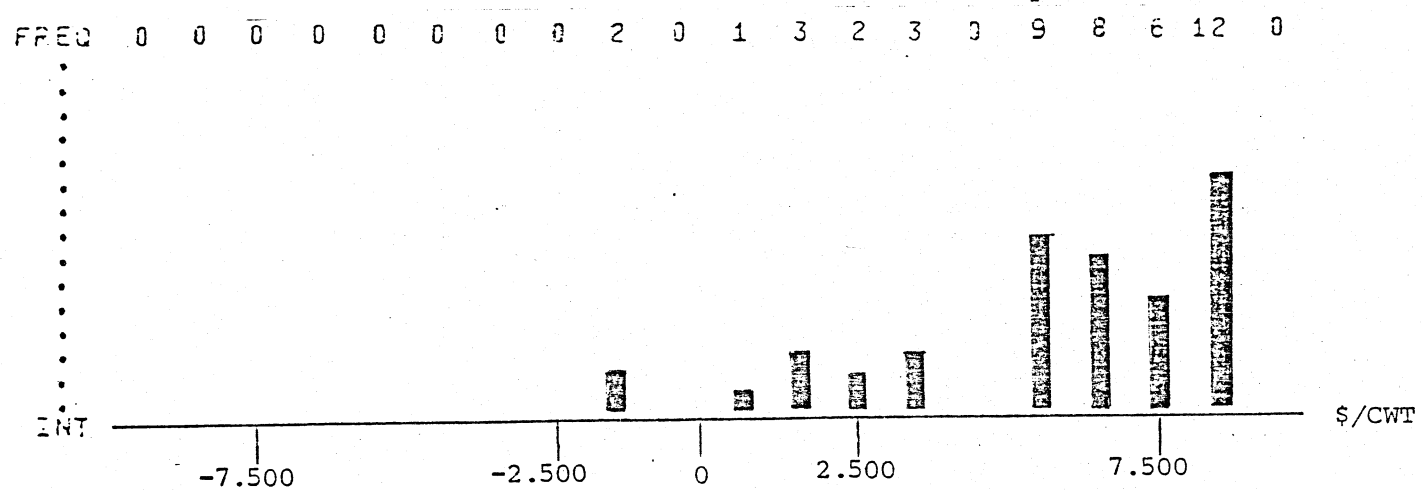
Figure 1 (continued)

3. CHOICE SLAUGHTER STEERS - 1979/79



N = 55; ME = 1.34; RMSE = ~~4.14~~ 3.49; R: -19.67 to 9.33; Outliers = 1

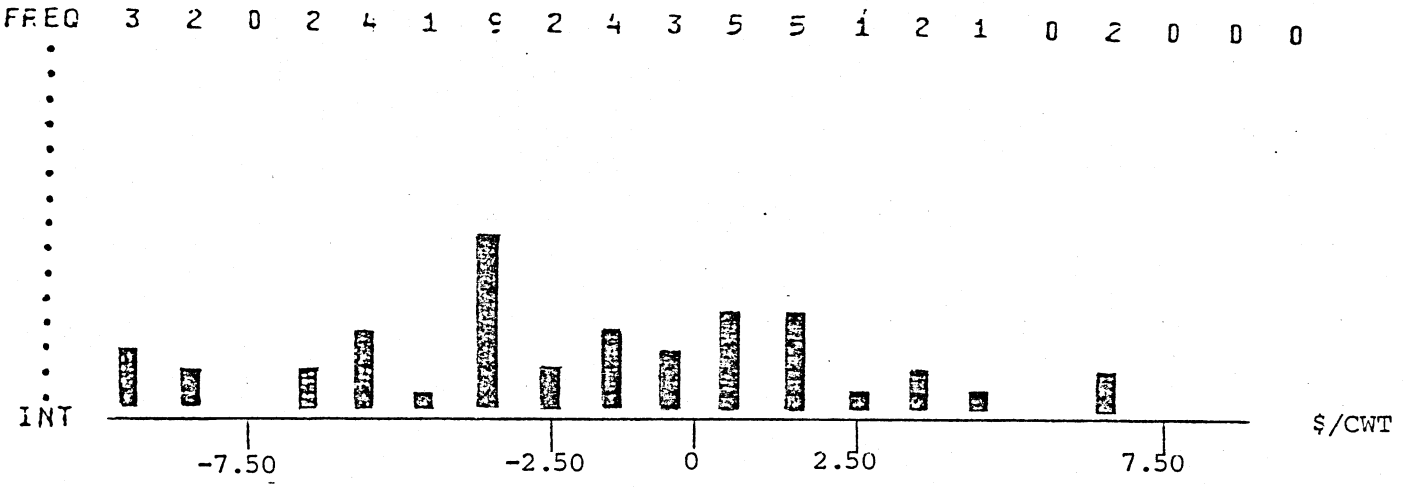
4. CHOICE SLAUGHTER STEERS - 1979/80 (first quarter 80)



N = 55; ME 6.02; RMSE = 5.13; R: -20.85 to 19.15; Outliers = 9

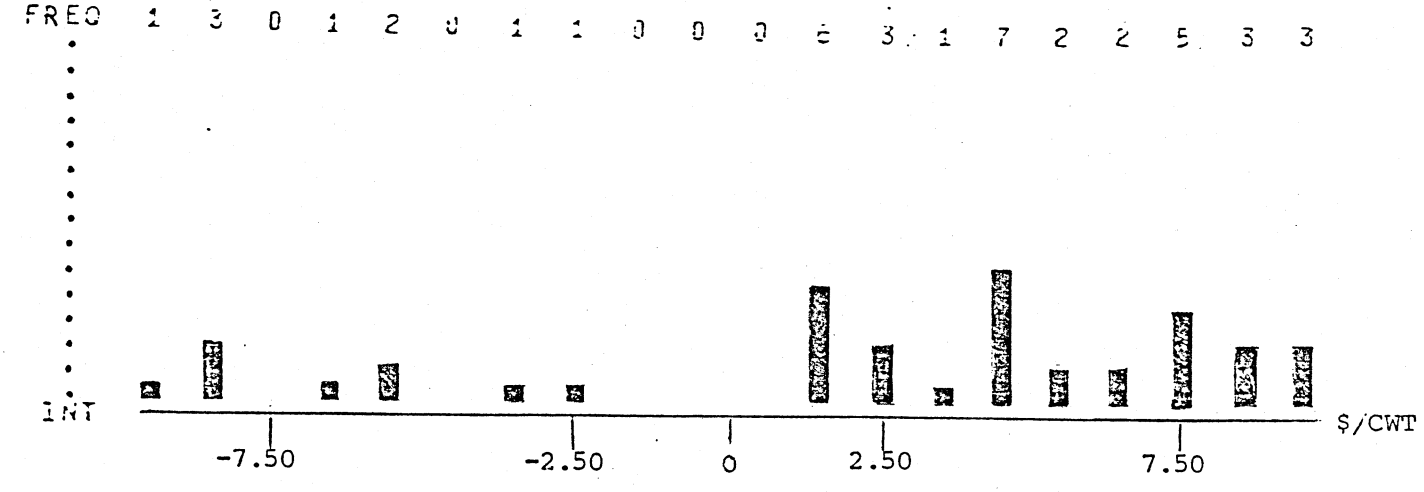
Figure 1 (continued)

5. CHOICE FEEDER STEERS - 1979/79



N = 48; ME = -2.56; RMSE = 5.67; R: -28.11 to 6.80; Outliers = 2

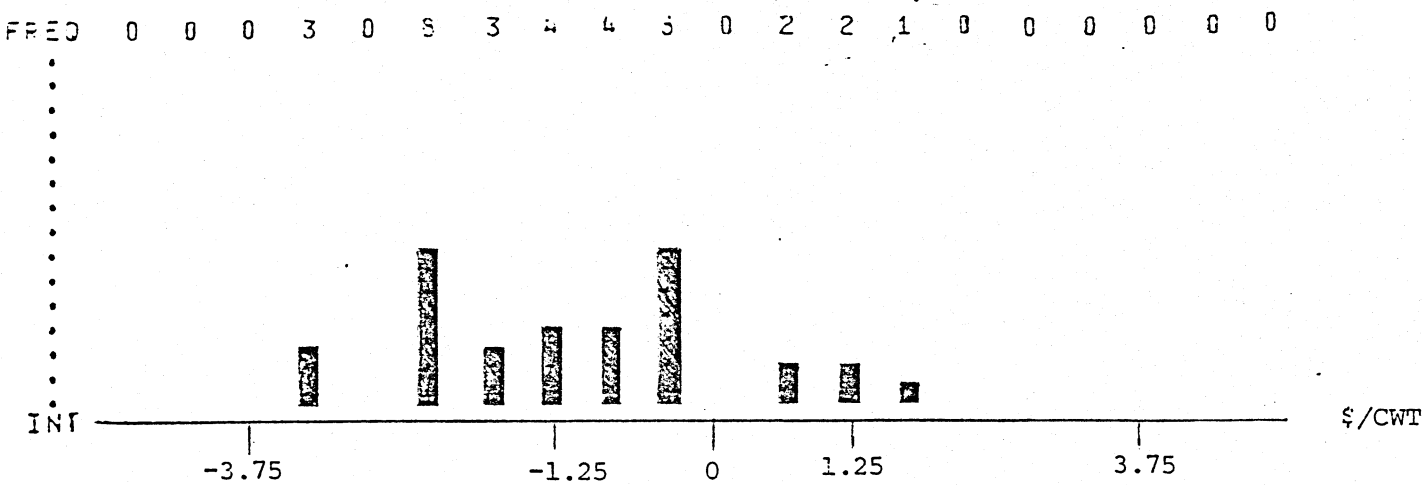
6. CHOICE FEEDER STEERS - 1979/80 (first quarter 80)



N = 48; ME = 3.03; RMSE = 7.91; R: -26.44 to 19.55; Outliers = 7

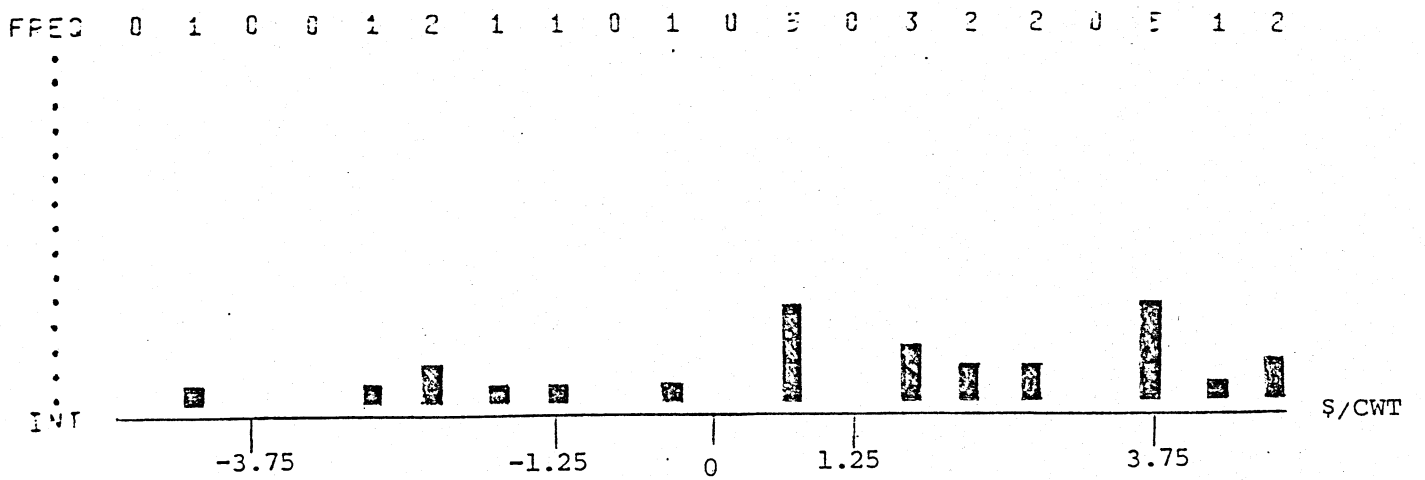
Figure 1 (continued)

7. BARROWS AND GILTS - 1978/78



N = 36; ME = -1.57; RMSE = 2.34; R: -13.13 to 1.51; Outliers = 1

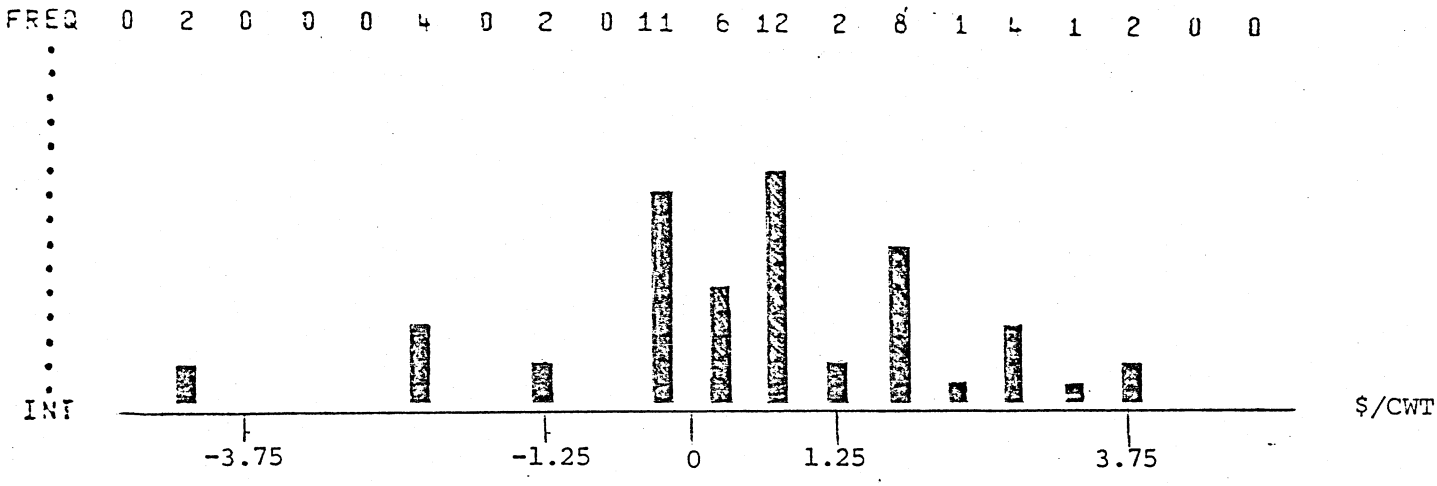
8. BARROWS AND GILTS - 1978/79



N = 35; ME = 1.98; RMSE = 3.73; R: -9.01 to 9.52; Outliers = 8

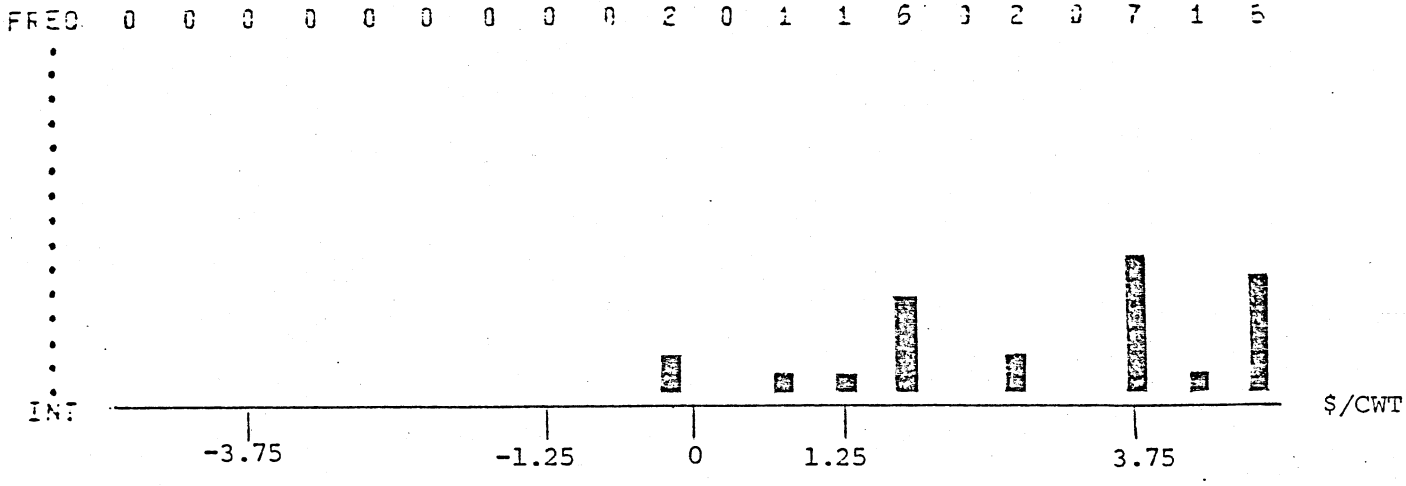
Figure 1 (continued)

9. BARROWS AND GILTS - 1979/79



N = 55; ME = .32; RMSE = 1.68; R: -4.48 to 3.52; Outliers = ø

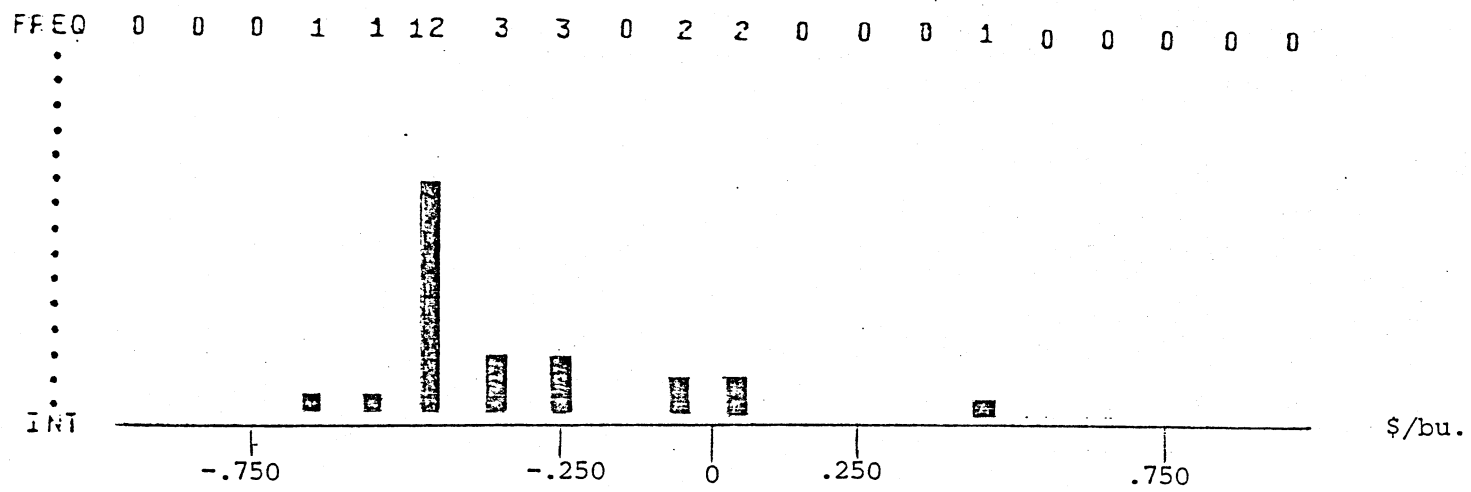
10. BARROWS AND GILTS - 1979/80 (first quarter)



N = 54; ME = 5.96; RMSE = 2.62; R: -.31 to 23.69; Outliers = 29

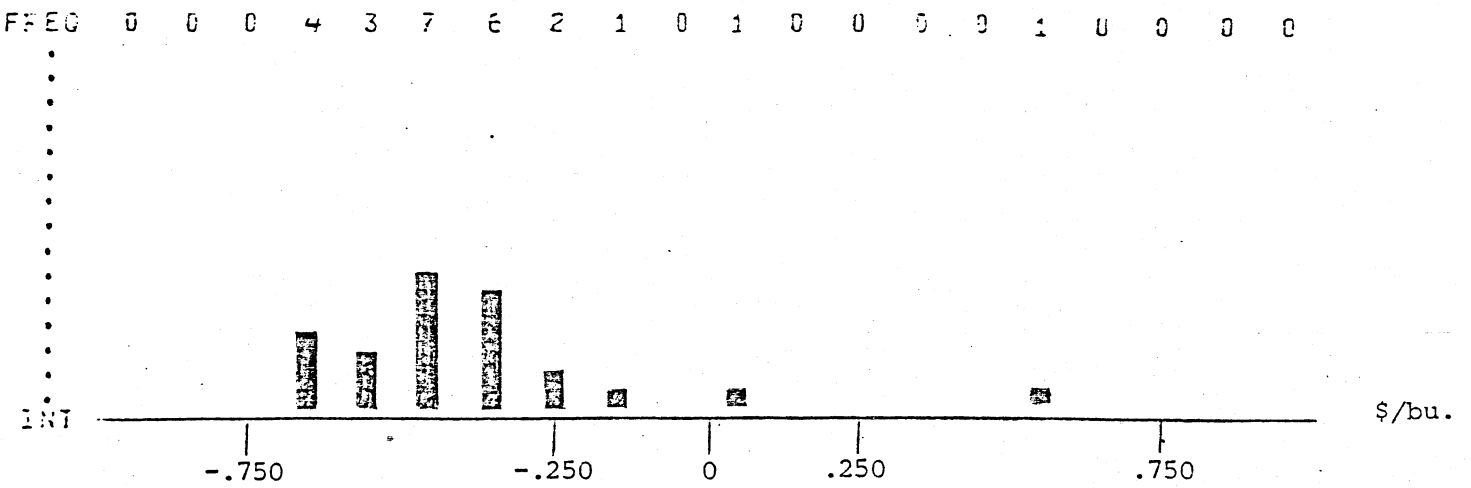
Figure 1 (continued)

11. WHEAT - 1978/78



N = 25; ME = -.30; RMSE = .24; R: -.69 to .48; Outliers = 0

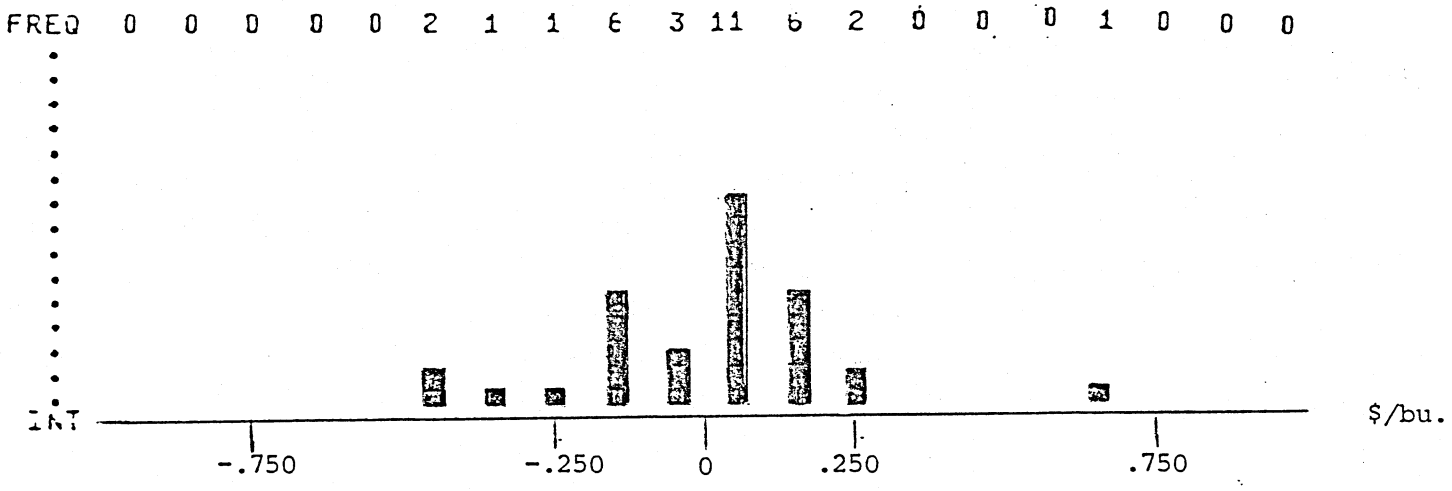
12. WHEAT - 1978/79



N = 26; ME = -.43; RMSE = .28; R: -1.04 to .52; Outliers = 1

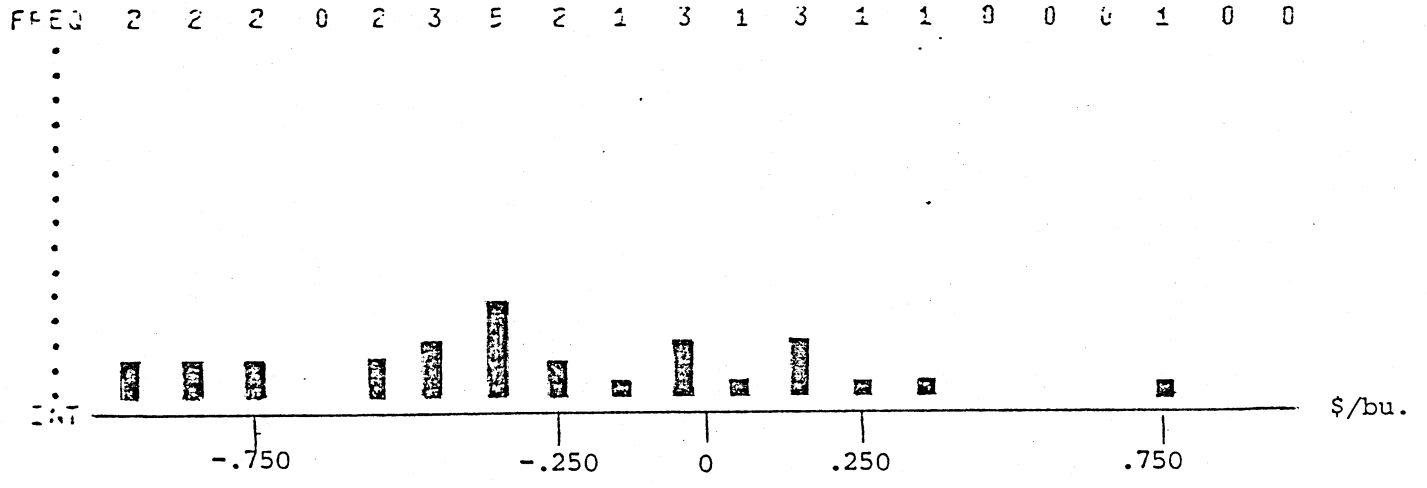
Figure 1 (continued)

13. WHEAT - 1979/79



N = 33; ME = -.008; RMSE = .21; R: -.48 to .62; Outliers = Ø

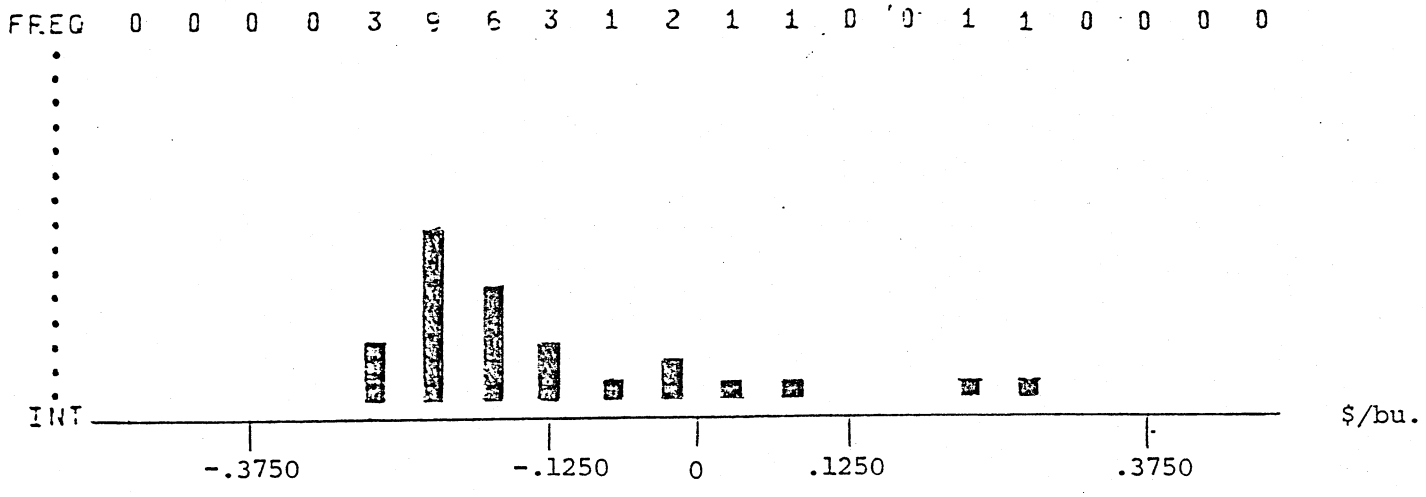
14. WHEAT - 1979/80 (ten months)



N = 33; ME -.43; RMSE = .49; R: -1.52 to .70; Outliers = 4

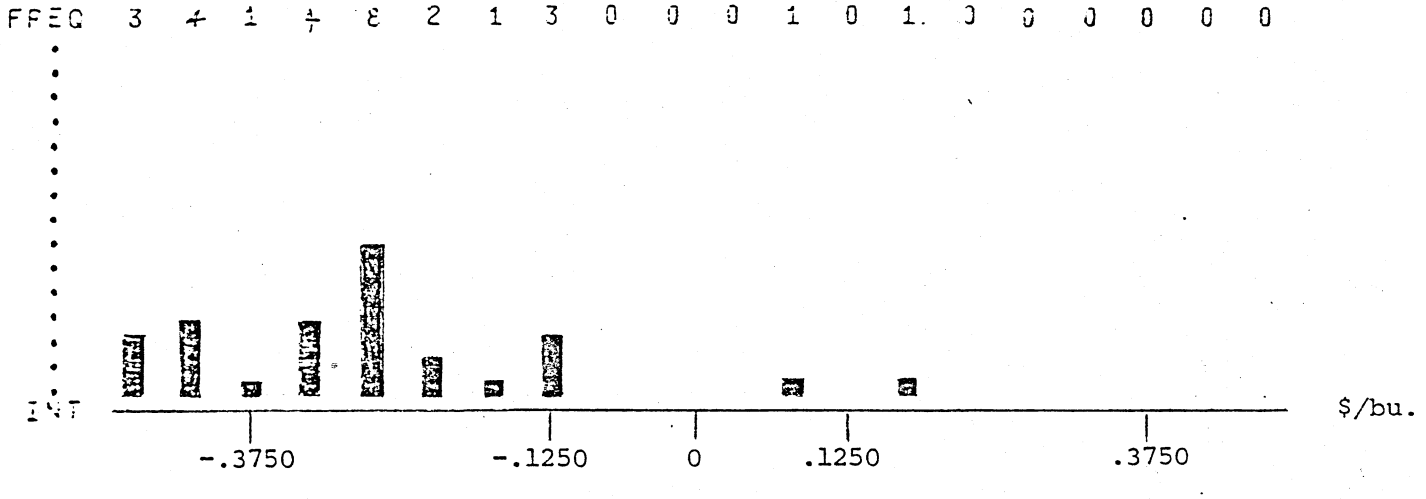
Figure 1 (continued)

15. CORN - 1978/78



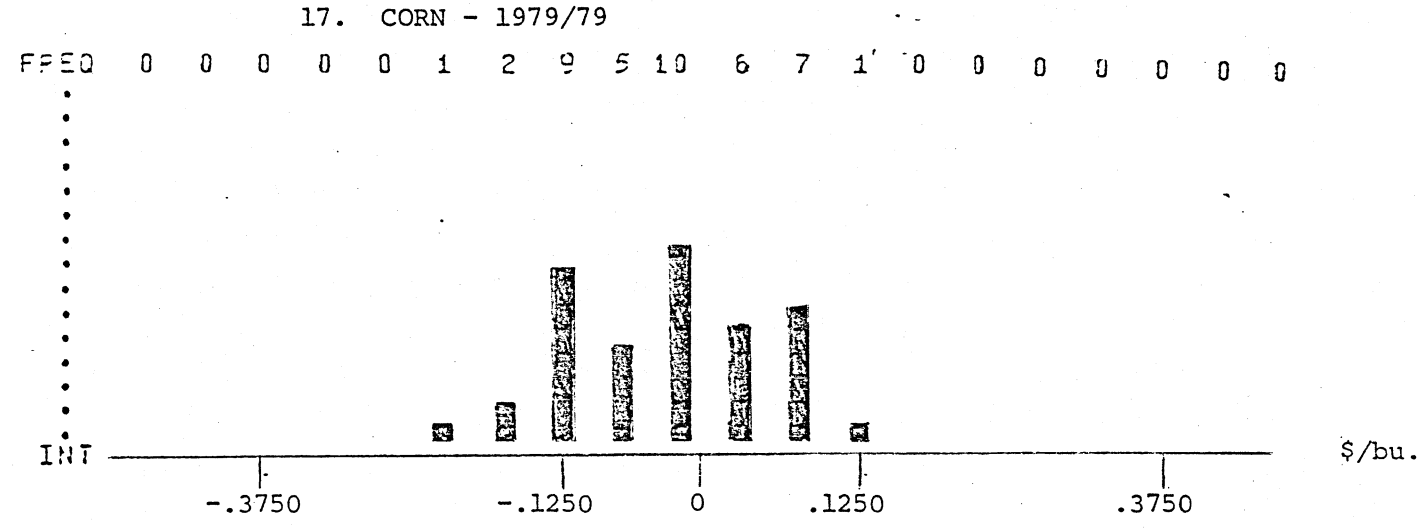
N = 28; ME = -.13; RMSE = .14; R: -.26 to .29; Outliers = ø

16. CORN - 1978/79

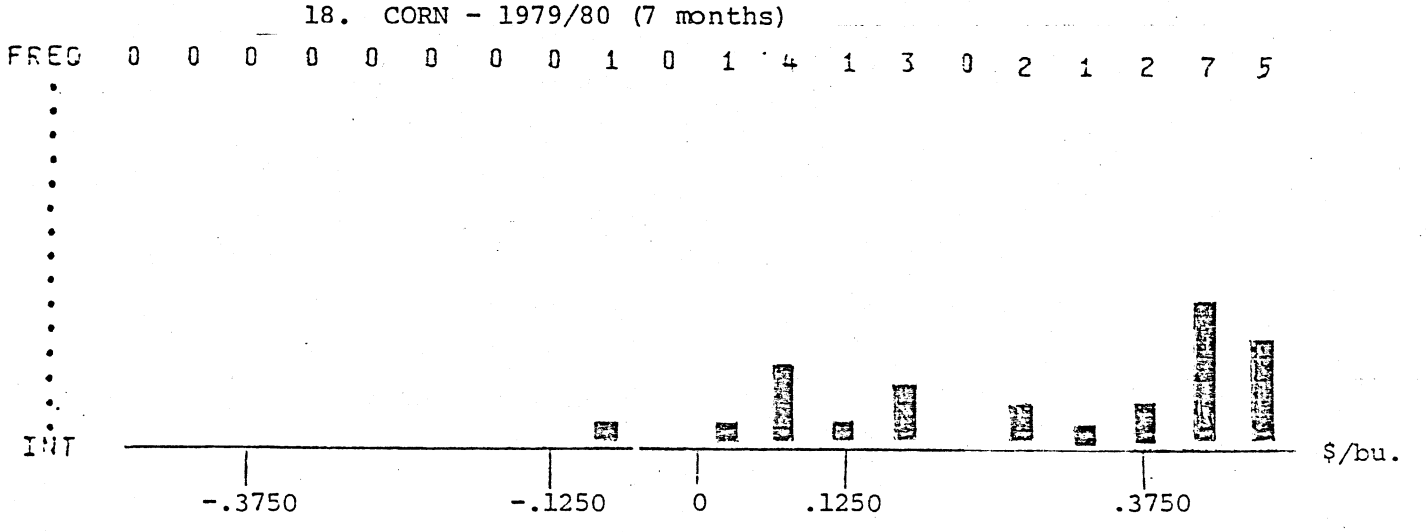


N = 29; ME = -.30; RMSE = .15; R: -.51 to 16; Outliers = 1

Figure 1 (continued)

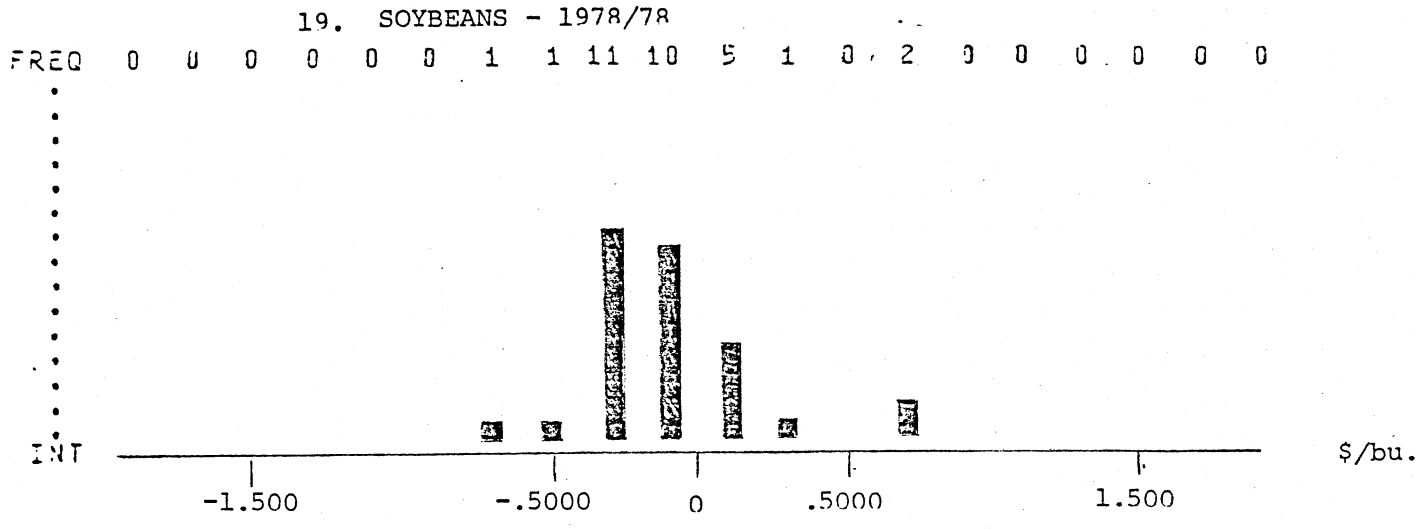


N = 41; ME = -.05; RMSE = .09; R: -.24 to .11; Outliers = Ø

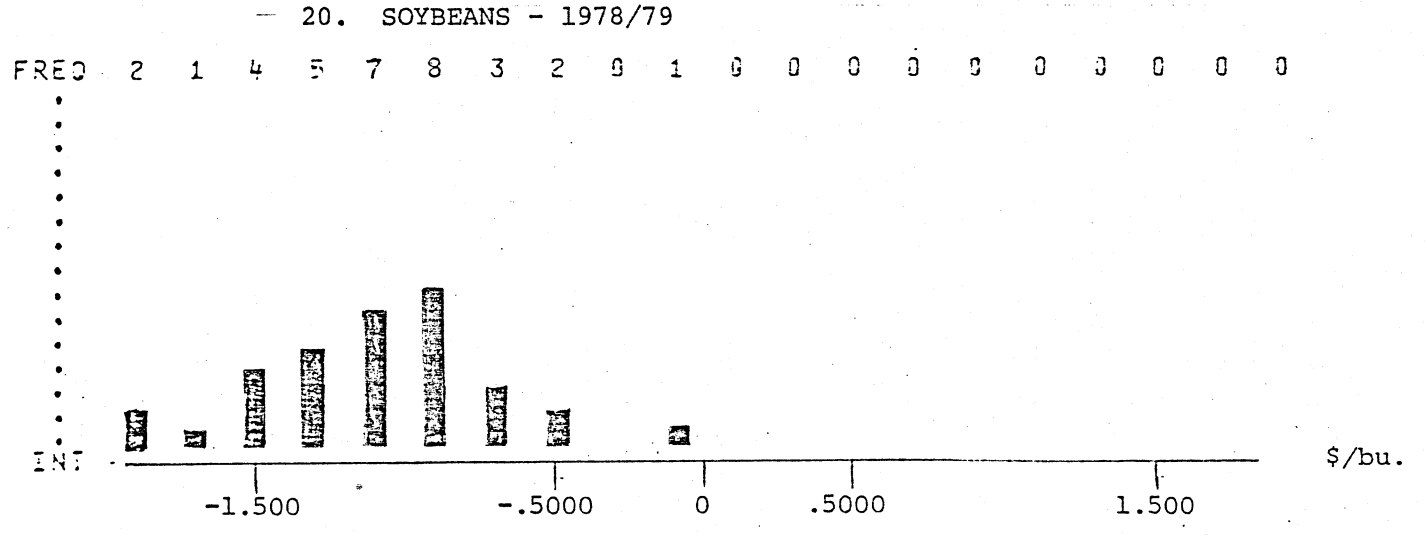


N = 41; ME = .45; RMSE = .25; R: -.06 to .89; Outliers = 14

Figure 1 (continued)

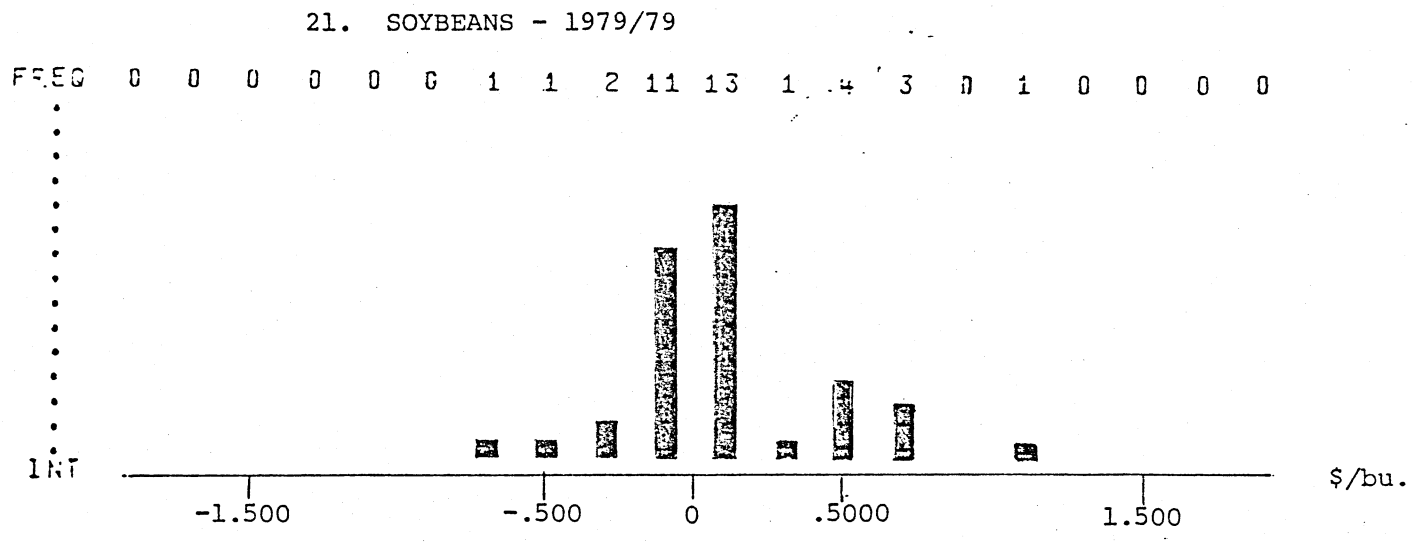


N = 31; ME = -.10; RMSE = .27; R: -.61 to .64; Outliers = ø



N = 33; ME = -1.12; RMSE = .40; R: -1.99 to -.09; Outliers = ø

Figure 1 (continued)



N = 37; ME = .08; RMSE = .35; R: -.69 to 1.06; Outliers = Ø

Table 1. Mean Percentage Price Forecast Error

Commodity \ Period	1978/78	1978/79	1979/79	1979/80[a/]
	----------	----------	% ----------	----------
Slaughter Cattle	1.1	13.0	2.0	9.0
Feeder Cattle	--	--	3.0	3.7
Barrows & Gilts	3.2	4.7	0.8	16.4
Wheat	11.0	12.4	0.2	10.0
Corn	6.0	12.0	2.0	19.0
Soybeans	0.01	16.0	0.01	--

[a/] 1980 actual prices based on partial year reports; figures should be regarded as illustrative only.

Chart 1. "Track Record" of Live Cattle Forecasters, 1978 and 1979