On Judgmental Forecasts

by

David A. Bessler and Jon A. Brandt

ON JUDGMENTAL FORECASTS

David A. Bessler and Jon A. Brandt*

In earlier work, empirical results were reported which suggest that expert judgments in forecasting agricultural prices were (in some sense) inferior to more mechanical, less judgmentally based forecasts. Brandt and Bessler (1981) show that expert forecasts of hog prices by the Purdue University Agricultural Economics Department's Outlook Committee were inferior in a mean-squared error sense to simple autoregressive integrated moving average (ARIMA) and reduced form econometric model forecasts. These results were extended to forecasts of cattle and broiler prices in Bessler and Brandt and to economic, as well as statistical, significance in Brandt and Bessler (1983). The initial assumption in each of these studies was that each commodity/forecaster studied represented a different group or population; there was no attempt to generalize results to other commodities or judgmental forecasters.

It is the purpose of this paper to consider the merits of such a generalization. Simply put, the point to be addressed is: are judgmental or expert forecasts generally inferior to statistical forecasts? While the evidence is quite one-sided in supporting an affirmative answer to this question, it does not necessarily imply that expert judgments are valueless. In fact, the evidence also suggests that expert judgments, when used in conjunction with more mechanical methods or in group composites, can yield reliable forecasts.

---

*The authors are Associate Professors of Agricultural Economics at Texas A&M and Purdue Universities, respectively. This paper was motivated by ideas and suggestions by Spyros Makridakas. Carl Shafer provided useful comments on the recommendations made in this paper. All errors and omissions are the authors' responsibility.

The paper is organized in four parts. First, a vast literature from psychology which suggests that our findings in agriculture are not all that novel is reviewed. The experimental work conducted over that last quarter century indicates that individual judgmental forecasts are not very reliable and can usually be improved. Next, a method to model subjective forecasts in a dynamic setting is proposed, and as an example, judgmental forecasts of quarterly U.S. hog and cattle prices by a particular group of experts who regularly develop outlook information are examined. Finally, the implications of the findings on forecasts of economic variables are discussed.

What is Known About Judgmental Forecasts?

While the study of judgmental forecasts falls clearly within the province of psychology, it is somewhat interesting that one of the very early studies on the subject is found in agriculture. In 1923, H. A. Wallace gave a rather spirited summary on an experiment with judges of seed corn carried out early in this century. His words are worth recalling:

> "That the corn judges did not know so very much about the factors which make for yields is indicated by the fact that their scores were correlated with yield to the extent of .2 . The difficulty seems to be that they placed too much emphasis on length of ear and possibly also some fancy points, which caused them to neglect placing as much emphasis on sound, healthy kernel characteristics as they should." (p. 304)

Wallace goes on to suggest that "the things which really are in their (the judges') minds are considerably different from . . . (those which they) professed." (p. 304)

The result from this early study is consistent with results found in psychology. Starting with the work of Meehl, psychologists have (almost always) found clinical (judgmental) predictions of numerical variables inferior to mechanical predictions. That is, in cases where both a clinical judgment and a statistical prediction of a criterion variable are available, such as academic success or prisoner parole recidivism, the statistical prediction is rarely (if ever) inferior to the clinical judgment. The twenty cases studied in Meehl's seminal book generated a plethora of additional studies, all reaching similar conclusions -- "an apparent superiority for mechanical modes . . ." (Sawyer, p. 178).

This result gives further cause for concern as predictive ability of judges does not necessarily increase with substantive training and confidence. Stael von Holstein, for example, found meteorologists' assistants outperformed the meteorologists in simple probability forecasting -- the latter tended to give tight forecasts, while the former gave diffuse distributions. Using experimental methods, Fischhoff and McGregor found that, while confidence in judgment increases with knowledge, it does so too fast -- a doubling in confidence being associated with about a 50 percent increase in knowledge. Earlier, Oskamp found that self-confidence in forecasts by clinicians increased with the amount of information; however, there was no corresponding increase in accuracy. Ryback found similar results in experiments with university students. He concluded: "confidence increased with experience . . . accuracy did not." (p. 331)

These results on forecast accuracy and self-confidence can be explained by the extensive literature on superficial information search and processing biases, a recent review of which can be found in Hogarth and Makridakis. Under the heading of superficial information search, they have identified biases resulting from the following heuristics (for extensive bibliographies see Hogarth and Hogarth and Makridakis):

availability - ease with which specific instances come to mind influences judgments of frequency;

selective perception - anticipation of what one expects to see biases what one thinks he does see;

frequency - probabilities and forecasts are often based on frequency rather than relative frequencies;

ignorance of prior information - some forecasters give predictions which are dominated by sample information relative to prior information;

illusory correlation - people may believe variables vary together when in fact they do not;

data presentation - the order, display, mode or context in which data are presented can influence judgmental forecasts.

A similar list associated with biases arising from suboptimal information processing is given as :

inconsistency - some forecasters are unable to apply a consistent judgmental strategy over a number of exchangeable cases;

conservatism - people do not revise opinion upon receipt of new information as suggested by Bayes' theorem;

nonlinear extrapolation - inability to extrapolate nonlinear processes;

mental effort heuristics - rules of thumb methods are used to reduce mental effort in the forecasting task. An example is that forecasters may anchor predictions on a particular value and adjust up or down to allow for circumstances.

In addition to these heuristics, other environmental characteristics related to the complexity of the forecasting task or the prassure associated with the task may bear negatively on forecast performance. Goldberg provides a useful summary on the environment and its influence on the judgmental forecaster:

> "(he) is not a machine. While he possesses his full share of human learning and hypothesis - generating skills, he lacks the machine's (statistical model's) reliability. He "has his days": Boredom, fatigue, illness, situational and interpersonal distractions all plague him, with the result that his repeated judgments of the exact same stimulus configuration are not identical ... (p. 423)."

Some analysts, of course, point out that heuristics are used in judgment because they usually work (or at least have worked in the past). However, knowledge that one is using a heuristic and that it may not be applicable in a particular situation is generally not widespread. Furthermore, advances in data acquisition and processing will probably reduce the advisability of using traditionally useful heuristics (see Tversky and Kahneman).

While the evidence supporting the fact that heuristics are used in judgmental forecasting is impressive (at least to us), it should be noted that most forecasts have been studied in environmental settings with little or no feedback. Murphy and Winkler (1977a) find weather forecasters do quite well in daily temperature and precipitation forecasts. Here, the subjects are quite familiar with their subject matter and receive continuous (daily) feedback. Interestingly though, their forecasts of tornados (which are less frequent) are not nearly as good (Murphy and Winkler (1977b)). In an explicit study of feedback in forecasting, Lichtenstein and Fischhoff find that feedback improved the quality of some subjects' assessments. However, their work points to the need for further research on the usefulness of particular forms of feedback which may prove useful in applied settings.

The study of a wide range of individual predictions casts doubt on experts' abilities to consistently provide good judgmental forecasts; however, some have suggested and others have observed that composite or group forecasts will often provide useful forecasts. Goldberg writes: ". . . pooling the responses of a set of judges may generally be a more powerful technique than modeling as a means of improving the accuracy of clinical judgments" (p. 431). Indeed, there is some evidence that the composite forecast may perform quite well. Ebert and Kruse find that the forecasts of common stock returns by individual security analysts could be improved upon if one used a group or composite forecast. Similarly, Zarnowitz recently found that in forecasting numerous aggregate time series, individual experts participating in the quarterly National Bureau of Economic Research and American Statistical Association survey of business conditions perform worse than the group forecast. He finds: "it is difficult for individuals to predict consistently better than the group . . . for most people, most of the time, the predictive record is spotty. . . a series of group averages has the advantage that it is helped by the cancellation of individual errors of the opposite sign." (p. 17) Finally, Brandt and Bessler (1981) show that a composite of expert opinion with forecasts from an ARIMA model and those from an econometric model outperform both the expert and the individual model forecasts of quarterly U.S. hog prices. This latter work is replicated with additional data points in Granger, *et al*.

Modeling Judgmental Forecasts

Whether or not one accepts the generality of the empirical work reviewed above, he may yet find it informative to study the explicit forecasts of a particular expert and their relationship to public information available at the time of the forecast. That is, even if one believes his particular expert performs better than those studied in the psychological or economic experiments described above, he may gain useful insights as to what particular sets of information his expert "seems" to be using. Perhaps the most useful model (to date) of the judgmental or clinical forecast is summarized under the heading of the Lens model, due to Brunswik and others. This model, studies relationships among an expert's judgment, relevant information variables, and the actual variable to be forecasted. Here, analysts are interested in the relationships between information variables and both the subjective forecasts and the actual variable to be forecasted. The Lens model assumes that (see Hogarth, p. 8): judgment results from a series of operations on information which is related to other items of information. Such interrelationships in the human mind have an analogue in the environment (subjective judgment relies on cues because it is believed that a similar process operates in nature). Judgment will be accurate to the extent that an individual's utilization of information cues matches those of the environment (representing the individual's subjective utilization of cue i as r_{is} and the environment's utilization as r_{ie}, judgmental accuracy can be improved by setting $r_{is} = r_{ie}$ for all i cues).

Early work on the Lens model (see Hursch, *et al.)* assumed known environmental relationships (known to the investigator). Very little dynamic structure was assumed in these early works. That is, early work considered experimental situations where the subject was presented with a particular set of information whose relationship to the variable to be forecasted was known and static. However, many judgmental forecasts are made in dynamic settings in which neither the variables used in the subjective forecasts or the particular lag structure are known *a priori* (see Camerer). Moreover, most psychological research does not allow past values of the subject's judgments to influence either his current judgment or the current environment. Apparently, no one has attempted to model feedback in judgmental forecasts.

In this analysis, a dynamic model is proposed in which all variables are studied in a vector autoregression. Equation [1] gives a rather general representation of the model:

$$[1] \qquad \begin{bmatrix} \phi_{11}(B) & \phi_{12}(B) & \cdots & \phi_{1m}(B) \\ \phi_{21}(B) & \phi_{22}(B) & \cdots & \phi_{2m}(B) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \phi_{m1}(B) & \phi_{m2}(B) & \cdots & \phi_{mm}(B) \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \\ \cdot \\ \cdot \\ \cdot \\ X_m(t) \end{bmatrix} = \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \\ \cdot \\ \cdot \\ \cdot \\ \xi_m(t) \end{bmatrix}$$

where $\phi_{ijk}(B) = \phi_{ij0}(B^0) + \phi_{ij1}(B^1) + \ldots + \phi_{ijk}(B^k)$ are polynominals in the lag operator B $(B^k X_t = X_{t-k})$ for which $\phi_{ij0}(B^0) = 1$ for $i = j$ and 0, $i \neq j$; $X_1(t)$, $X_2(t)$ and $X_i(t)$, $i = 3, \ldots, m$ are the time-ordered equivalents of the environmental, judgmental, and information variables as discussed above; and $\xi_1(t)$, $\xi_2(t)$, and $\xi_i(t)$ are dynamic random noise or innovations associated with each equation.

The stochastic representation given in equation [1] differs from usual Lens model representations by not explicitly defining lag length. Furthermore, lagged (past) values of all variables are permitted to influence the current value of each. In many forecasting applications, subjective judgment may be well represented (at least in very early stages of investigation) by equation [1]. As investigators acquire more information about a particular judgmental structure, a more detailed dynamic model may prove useful.

In practice, the set of information variables must be selected *a priori.*

That is, one must use theory or other *a priori* knowledge to determine what information series are relevant to the judgmental task (one could, perhaps, ask the expert what information he deems relevant). Theory can also be used to select the lag length (k in equation 1) associated with the autoregressive process. Where theory or other prior information gives little information on lag length, analysts must rely on statistical measures (see Sims (1980, p. 17) or Geweke and Meese).

Given a finite vector autoregression which adequately summarizes the empirical relationships between judgmental forecasts, the variable to be forecasted, and the relevant information, a number of additional operations can be performed to study the dynamics of the judgmental system. In particular, given some rather general properties of the system, one can invert the parameter matrix of equation [1]. This operation yields the (so-called) moving average equivalent in which the studied series are expressed as an infinite series of historical shocks (surprises). The moving average expression gives a parameter matrix at successive lags (t-1, t-2, t-3, . . .). The matrix at lag t-j, for example, summarizes the impact of shocks j periods ago on levels of each variable in period t. Analysts will prefer to study the moving average representation (rather than the autoregressive representation) because it gives a rather concise summary of the dynamics implied by the interaction of the entire set of autoregressive coefficients at successive lags.

Following Sims (1980) one can study the reaction of the expert in future periods to a shock in a particular information variable in the current period. One can compare the expert forecast path with that of the variable being forecasted. Similarly, one can study the response of the variable to be forecasted to a shock in the judgmental forecast. Under normal circumstances, if the forecaster possesses substantive information beyond what is modeled in the vector autoregression, one might expect that a shock in judgmental forecasts would lead to a subsequent positive response in the variable forecasted.

In actual study of the dynamic behavior of an estimated vector autoregression, one needs to be explicit about contemporaneous error correlations. Simulations of the system to a shock in a particular series will be misleading if same period error correlations are ignored. That is, where contemporaneous error correlations are high, one must decide on a particular causal chain among the series studied. In this study, the system is ordered: $Y_s(t)$, $X_1(t)$, . . . , $X_n(t)$, $Y_e(t)$ in which case, the subjective forecast for period t ($Y_s(t)$) and all information variables for period t ($X_i(t)$, i = 1, . . . , n) are prior in a Granger-type causality sense to the variable to be forecasted ($Y_e(t)$) (see Sims (1981, pp. 288-90) for details). The methods outlined here are used to study the judgmental forecasts of prices of two agricultural commodities.

Empirical Study of the Price Forecasts of Two Agricultural Commodities

For several years, a committee within the Department of Agricultural Economics at Purdue University has given short-term forecasts on prices of agricultural commodities produced by midwestern farmers. The committee is composed of professional agricultural economists who have many years of experience working with producers and processors.

Bessler and Brandt studied the committee's quarterly forecasts of hog, cattle, and broiler prices over the years 1976-1979. Expert opinion was found to be inferior to ARIMA and/or reduced form econometric model forecasts in mean-squared-error loss evaluations. This result is consistent with prior work (some of which is reviewed above). The intent here is to study the dynamic relationship between the cattle and hog expert forecasts (the broiler forecasts have not been issued regularly enough to continue their use here) and a set of seemingly relevant information variables.

The 33 quarterly, one-period-ahead forecasts begin with the last quarter of 1974 and extend through the last quarter of 1982. Participation by individuals on the Purdue Outlook Committee did change over this period -- so the analysis cannot be construed to reflect on any particular forecaster. Although a particular person was largely responsible for preparing an initial forecast for each commodity in each time period, the final forecast was based on subsequent roundtable discussion by the entire committee.

The number of different information variables under study is severely limited due to having only 33 data points on each forecast. A system of m variables requires the estimation of m^2k autoregressive coefficients (here k is the number of lags studied) plus $(m \times (m + 1))/2$ unique elements of the variance-covariance matrix. For $m = 4$ and $k = 4$, this results in a rather meager total-observations-to-parameter ratio of $132 \div 74$. Most time-series analysts would prefer a higher value for this ratio, although there are no clear guidelines available and evidently systems have been successfully estimated with fewer data points relative to parameters (Wallis). Certainly, as longer sets of data and expert forecasts become available, richer sets of conditioning information ought to be studied.

In this study, the hog analysis includes quarterly observations on the hog price forecasts, actual hog prices, nominal U.S. disposable income and sows farrowing. The cattle model studies relationships between quarterly data on cattle forecasts, actual cattle prices, nominal U.S. disposable income and cattle placed on feed. As each equation in each VAR contains the same set of regressors, ordinary least squares regression, applied equation by equation, will provide efficient estimates of the VAR coefficients.

Tables 1 and 2 list calculated chi-squared statistics associated with likelihood ratio tests of successive lag lengths of the two vector autoregressions. This test considers sequential tests of the restriction that all parameters of the matrix at lag k of equation number [1] are zero. Only lags 1 through 4 are considered in these tests because of data requirements. Higher order lags should be considered as more data become available. Statistics in both tables indicate that the data are generated by at least a fourth-order autoregression. All subsequent discussion will be with respect to a fourth-order model for both commodities.

In tables 2 and 3, the approximate probabilities associated with exogeneity tests on each variable in each equation of the VAR and the approximate probabilities that the residuals in each regression are uncorrelated (based on summary Q-statistics) are listed. A high probability in these tables suggests that the current variable is not responsive to (affected by) the lagged values of the corresponding variable. With respect to hog prices (table 2), coefficients on past expert opinion and disposable income in the current expert opinion equation are not significantly different from zero (based on standard F-tests on grouped coefficients). The hog price expert seems to be relying on just past hog prices and sows farrowing. However, both past expert opinion and past disposable income have grouped coefficients different from zero in the actual hog price equation (column 4). Evidently, past expert opinion and disposable income include (or are correlated with) information to which the other variables in the VAR are not responding. The expert, however, ignores both (further evidence on this offered below).

Table 1. Likelihood Ratio Tests on Vector Autoregressions on and U.S. Cattle Market (lags 1 through 4).

k	m(k)	
	hogs	cattle
1	356.91*	378.23*
2	31.98*	16.16
3	50.77*	32.35*
4	40.53*	26.29

Note: Under the null hypothesis, that the autoregressive matrix at lag k is equal to zero, m(k) is distributed chi-squared with 16 degrees of freedom. An asterisk indicates rejection of the null hypothesis at the 5% level of significance.

With respect to cattle prices (table 3), differences between coefficients in the expert opinion and actual equation, show up in F-tests on past disposable income (again) and cattle on feed. Both are significantly different from zero (at least at the 12% level) in the actual equation and not so in the expert opinion equation. Interestingly, past values of expert opinion are not significantly different from zero in the actual price equation, a result which differs from the hog model.

The F-tests relate to the significance of past coefficients in each VAR equation. The tests do not shed light on the current prediction and current actual price (actual price and forecasted price for period t are indexed by the same value of t). Figures 1 and 2 give the dynamic response of expert opinion and actual price to one-time-only shocks in expert opinion, actual prices, disposable income, and either sows farrowing (hog model) or cattle on feed (cattle model). The contemporaneous covariance in each vector autoregression was ordered as follows: expert opinion, income, sows farrowing (cattle on feed for the cattle VAR), and actual price. Because expert opinion is formed prior to period t, this ordering should not distort the actual flow of information. Putting expert forecasts last would, of course, result in such a distortion.

The responses of actual hog and cattle prices to shocks in expert opinion are different. Recall from table 2 and 3 that past hog expert opinion was significant in the actual hog price regression while the expert opinion on cattle price was not significant in the actual cattle price regression. In figures 1 and 2, similar patterns emerge. A shock in expert opinion on hog price in period zero causes (in Granger's sense) actual hog prices in period zero to be higher by .35 . This result is consistent with the hypothesis that the expert is providing a degree of substantive knowledge;

Table 2. Approximate Probabilities on Significance of Lagged Values of Variables in Four Lag Vector Autoregression on U.S. Hog Market.[a]

	Current Variables			
Lagged Variables	Expert	Income	Sows Farrowing	Actual
Expert	.78	.64	.80	.01
Income	.59	.00	.01	.00
Sows Farrowing	.02	.55	.00	.01
Actual	.02	.66	.08	.01
Q	.42	.20	.20	.12

[a] The probabilities are derived from an F-test on lagged values of each variable in an ordinary least-squares regression of current values of each variable on four lags of each variable. High probabilities suggest that we do not reject the null hypothesis that the coefficients on the lagged variables equal zero (i.e., that they are not used in the prediction of the current variable). The coefficients associated with the bottom "Q" row are the approximate probabilities that the residuals in each regression are uncorrelated. That is, a high probability is associated with the failure to reject the null hypothesis of white noise residuals.

Table 3. Approximate Probabilities on Significance of Lagged Values of Variables in Four Lag Vector Autoregression on U.S. Cattle Market.[a]

	Current Variables			
Lagged Variables	Expert	Income	Cattle on Feed	Actual
Expert	.99	.89	.31	.48
Income	.49	.00	.59	.03
Cattle on Feed	.86	.84	.05	.12
Actual	.03	.84	.80	.01
Q	.17	.78	.01	.20

[a] See table 2.

however, a positive shock in the cattle expert forecast is not followed by a significant positive response by actual cattle prices. That is, an innovation in the expert's forecast of cattle price for period t is not subsequently verified in the actual price series.

The figures also suggest (at least tentatively) that the cattle expert forecast is not based on the past disposable income or cattle on feed, as seen from the evidence discussed above, and the response to shocks in disposable income and actual cattle price. Notice that actual cattle price responds significantly (positively as theory would suggest) to a shock in disposable income in periods 0, 2, 3, 4, 5, 6; while, expert opinion responds (apparently unknowingly) to a shock in disposable income in periods 3, 4, 5, 6, 7 with the channel of influence being through actual prices. That is, expert opinion responds to disposable income through actual price. A shock in disposable income influences cattle price in periods t, t+1, etc., which in turn influences expert forecasts in periods t+1, t+2, etc. The expert forecasts of cattle prices seem to be merely tracking past cattle prices.

Expert opinion and actual prices are reacting to the sows farrowing information similarly with respect to expert opinion on hog price and the information summarized in the VAR (argument given above).

Discussion

In this paper, the use of prior knowledge on judgmental forecasting is investigated. Early work from psychology indicates that judgmental forecasts are generally inferior to mechanical forecasts or group composites. This work was found to be consistent with that recently reported in economics. Simple ARIMA and reduced form econometric models have been shown to outperform expert forecasts. In addition, group composites of either multiple experts or experts and mechanical models have outperformed the expert.

Early work in psychology gives one reason to view these empirical findings as special cases of general behavior. Heuristics in information acquisition and processing have been demonstrated to exist in experimental and real-world settings. These heuristics generally result in suboptimal forecasts.

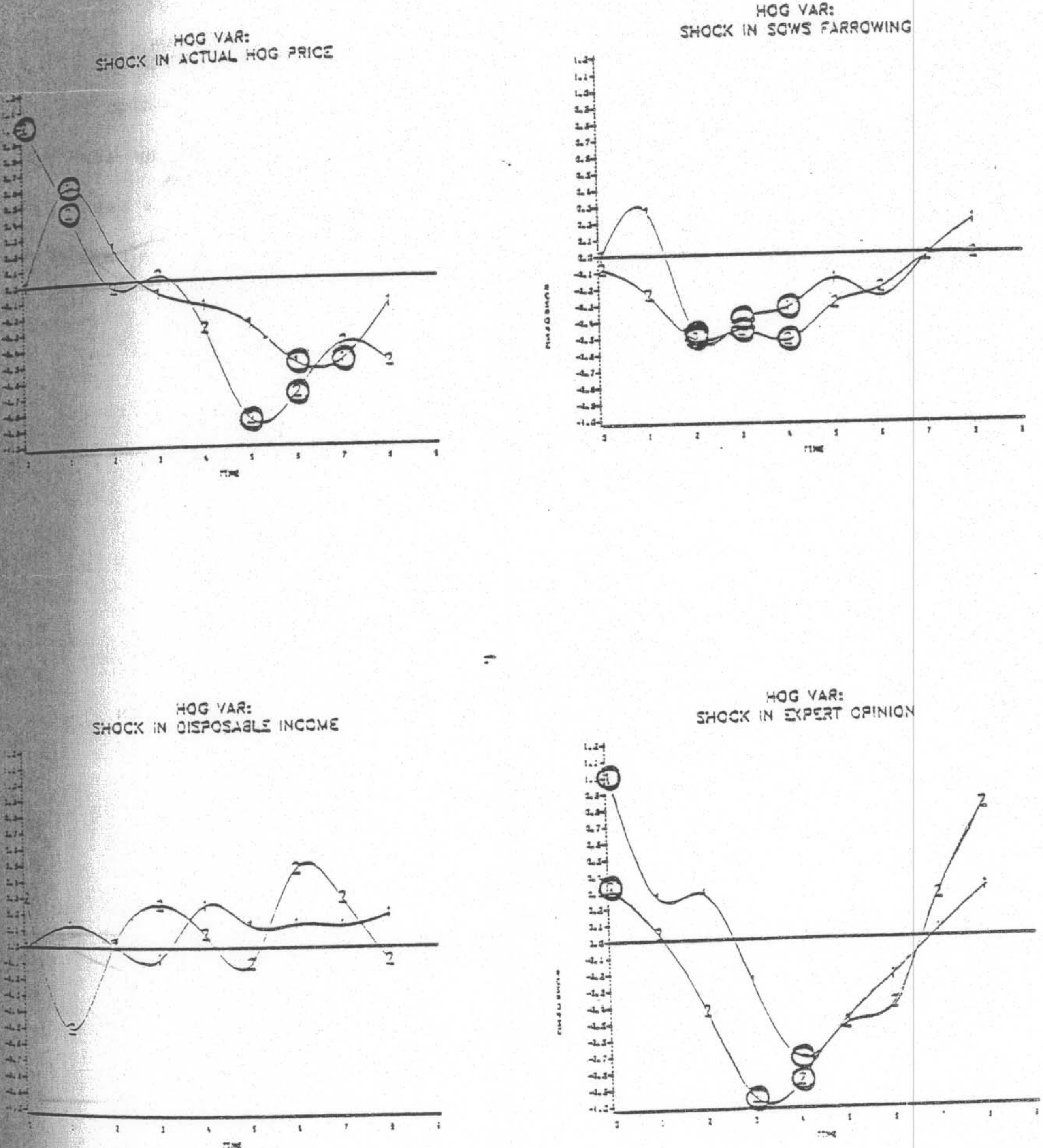
It is argued here (as others have elsewhere) that in nonexperimental settings where control over the set and levels of conditioning information is not possible, a strict application of the basic Brunswikian Lens model is not possible. Rather, an analysis of expert forecasts in general vector autoregression is proposed. In undertaking such an analysis, the researcher must make rather strong assumptions on the particular information set which the expert brings to his assessment task. In addition, the analyst must be willing to allow the data actually chosen to do "quite a lot of the talking" -- as lag length for time-ordered data are usually unknown.

In the particular applications presented here, the results indicate the series of expert forecasts of cattle price were (at least) consistent with the hypothesis that only past cattle prices entered the utilized information set. Other (seemingly) relevant information which did appear significant in the actual cattle price equation was apparently not used in the cattle expert forecast. On the other hand, evidence suggests that the hog price expert opinion did pay attention to sows farrowing and past hog prices in forming the forecast of hog prices. Apparently, neither cattle nor hog experts paid attention to disposable income.

These findings are of course subject to criticism and should be considered tentative. They are not based on the more elaborate versions of the Lens model which psychologists have used with experimental data. Nonlinear relationships and richer information sets may in fact show the expert forecasters in a different light. However, given the overwhelming evidence which psychologists have uncovered on judgmental forecasts and the evidence (albeit meager) uncovered so far in economics, it is our recommendation that judgmental forecasters (for at least the time being) strike a humble pose.

Certainly, the abandonment of judgmental forecasting is not proposed here. Most forecasts which are required in everyday settings are done regularly enough so that sufficient outcome feedback justifies high levels of confidence. In addition, the cost of constructing statistical (mechanical) forecasts may not justify their use. The costs of incorrect forecasts are oftentimes not severe (say, one walks over to the student union for a cup of coffee and finds that it's closed; what has he lost?). However, where historical data are available and the consequences of alternative actions are diverse, judgmental forecasting should be one component of an integrated system.

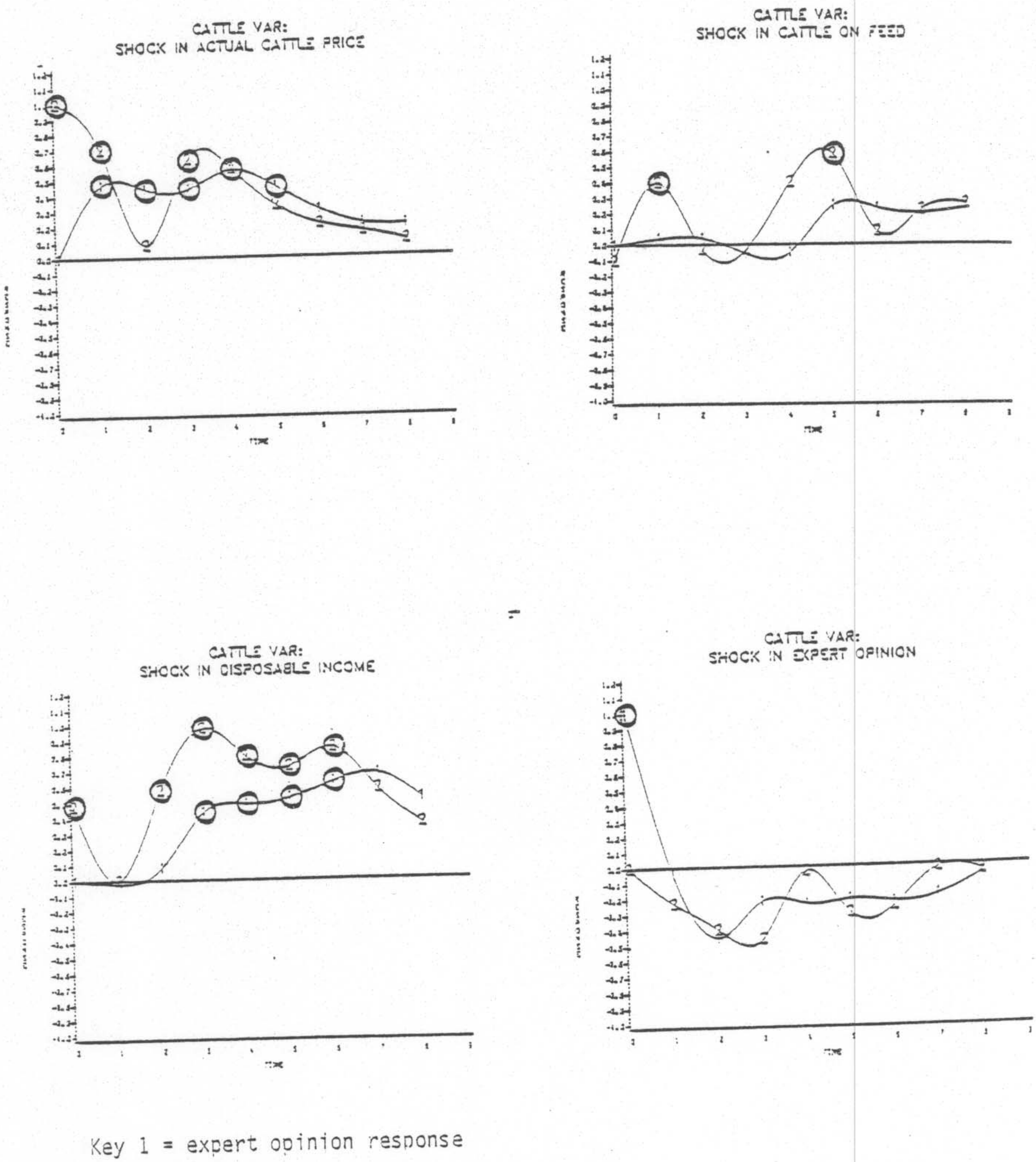
Figure 1. Impulse Responses on Shocks to Variables in Hog Vector Autoregression



Key 1 = expert opinion response

2 = actual price response

Figure 2. Impulse Responses on Shocks to Variables in Cattle Vector Autoregression



Key 1 = expert opinion response

2 = actual price response

Notes on figures 1 and 2. The figures represent the responses of expert opinion (1) on hog or cattle prices and actual (2) hog or cattle prices to one-time-only, one-standard-error-shock in each series of the four variable vector autoregression. Reponses are standardized by standard errors of either the expert opinion or actual price variable. A circle "○" is associated with a response which exceeds a two standard error interval. Standard errors of responses are calculated with Monte Carlo methods as programmed by Doan and Litterman.

References

Bessler, D. A. and J. A. Brandt. "Forecasting Livestock Prices with Individual and Composite Methods." *Applied Economics* 13 (1981): 513-22.

Brandt, J. A. and D. A. Bessler. "Composite Forecasting: An Application with U.S. Hog Prices." *American Journal of Agricultural Economics* 63 (1981): 135-40.

Brandt, J. A. and D. A. Bessler. "Price Forecasts and Evaluation: An Application in Agriculture." *Journal of Forecasting* 2 (1983): 237-48.

Brunswick, E. "Systematic and Representative Design of Psychological Experiments." *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability* (ed., J. Neyman), Berkeley: University of California Press, 1949.

Camerer, C. "General Conditions for the Success of Bootstrapping Models." *Organizational Behavior and Human Performance* 27 (1981): 411-22.

Doan, T. and R. Litterman. "RATS: Users' Manual." Minneapolis: VAR Econometrics, 1981.

Ebert, R. J. and T. E. Kruse. "General Conditions for the Success of Bootstrapping the Security Analyst." *Journal of Applied Psychology* 3 (1978): 110-19.

Fischhoff, B. and McGregor. "Subjective Confidence in Forecasts." *Journal of Forecasting* 1 (1982): 155-72.

Geweke, J. and R. Meese. "Estimating the Order of an Autoregression of Finite but Unknown Order." *International Economic Review, (1981):* .

Goldberg, L. "Man Versus Model of Man: A Rationale, Plus Some Evidence for a Method of Improving on Clinical Forecasts." *Psychological Bulletin* 73 (1970): 422-32.

Granger, C. W. J., R. Ramanathan and I. Domowitz. "Improved Methods of Combining Forecasts." Unpublished Manuscript, University of California, San Diego, 1982.

Hogarth, R. M. *Judgment and Choice: The Psychology of Decision.* Chicester, England: John Wiley & Sons, Pub. Co., 1980.

Hogarth, R. M. and S. Makridakis. "Forecasting and Planning: An Evaluation." *Management Science* 27 (1981): 115-38.

Hursch, C. J., K. Hammond and J. Hursch. "Some Methodological Considerations in Multiple-cue Probability Studies." *Psychological Review* 71 (1964): 42-60.

Lichtenstein, S. and B. Fischhoff. "Training for Calibration." *Organizational Behavior and Human Performance* 26 (1980): 149.

Meehl, P. E. *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.* Minneapolis: University of Minnesota Press, 1954.

Murphy, A. H. and R. L. Winkler. "Probabilistic Tornado Forecasts: Some Experimental Results." *Tenth Conference on Severe Local Storms,* American Meteorological Society, 1977, Boston, Massachusetts.

Murphy, A. H. and R. L. Winkler. "Subjective Probability Forecasts of Precipitation and Temperature." *Journal of the Royal Statistical Society* Series C, 26 (1977): 41-47.

Oskamp, S. "Overconfidence in Case-Study Judgments." *Journal of Consulting Psychology* 29 (1965): 261-65.

Ryback, D. "Confidence and Accuracy as a Function of Experience in Judgment-Making in the Absence of Systematic Feedback." *Perceptual and Motor Skills* 24 (1967): 331-34.

Sawyer, J. "Measurement and Prediction, Clinical and Statistical." *Psychological Bulletin* 66 (1966): 178-200.

Sims, C. "An Autoregressive Index Model for the U.S., 1948-1975." *Large Scale Macro-Econometric Models* (eds., J. Kmenta and J. Ramsey), Amsterdam: North-Holland Pub. Co., (1981): 283-327.

Sims, C. "Macroeconomics and Reality." *Econometrica* 48 (1980): 1-48.

Stael von Holstein, S. Carel-Axel. *Assessment and Evaluation of Subjective Probability Distributions.* Economic Research Institute, Stockholm, Sweden, 1970.

Tversky, A. and D. Kahneman. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185 (1974): 1124-31.

Wallace, H. A. "What is in the Corn Judge's Mind?" *Journal of the American Society of Agronomy* 15 (1923): 300-04.

Wallis, K. "Multiple Time Series Analysis and the Final Form of Econometric Models." *Econometrica* 45 (1977): 1481-98.

Zarnowitz, V. "The Accuracy of Individual and Group Forecasts from Business Outlook Surveys." National Bureau of Economic Research Working Paper, No. 1053, 1983.