Forthcoming Omega

July 1986

Comments Welcome

ON BAYESIAN COMPOSITE FORECASTING

1986

by

David A. Bessler

and

Peter J. Chamberlain*

Swine -- Prices

presented @ AAEA 1986

*The authors are Professor and Research Associate at Texas A&M University, College Station, Texas, U.S.A. The authors thank Jon A. Brandt for the expert opinion data and Kusumben Mistry who provided computing assistance. An anonymous referee made useful comments on an earlier draft of this paper.

On Bayesian Composite Forecasting

(Abstract)

Oftentimes decision makers have several forecasts of an uncertain and operationally relevant random variable. A rich literature now exists which argues that in this situation the decision maker should consider forming a forecast as a weighted average of each of the individual forecasts. In this paper, composite forecasting is discussed in a Bayesian context. The ability of the user to control the impact of the data on his composite weights is illustrated by an example using expert opinion forecasts of U.S. hog prices.

On Bayesian Composite Forecasting

INTRODUCTION

Composite Forecasting has been studied recently; papers include: Bates and Granger [1], Bessler and Brandt [3], Bunn [4,5], and Granger and Ramanathan [10]. A prime motivation for considering composite forecasting is that often economic agents have multiple forecasts of random variables, whose previsions are meaningful for rational decision-making (deFinetti, [8]).

Several formulations in composite forecasting employ forecast error variance as the underlying loss function for determining optimal composite weights -- where error variance is estimated from historical forecasts and realizations. Such studies (implicitly) recommend that decision-makers use the derived weights to form a composite forecast, as a linear (convex) combination of the individual forecasts. This procedure has much to recommend it; especially in cases where estimation of a stationary forecast error covariance is possible. However, in many instances, such is not the case. Oftentimes, we are faced with a few forecast realizations of processes which are not (necessarily) stationary. Here analysts will be less well-advised to rely on composite weights based solely on historical error realizations. In fact, some applied researchers recognize this in making the recommendation that composites be formed with equal weights (Clemen and Winkler, [7]). Makridakis, et al. [13] find that a simple average of several alternative forecasts performs well in an out-of-sample forecasting competition with a large number of series.

A problem associated with equal weighting is that it does not allow the analyst to update his composite weights with new information. Equal weighting may be an extreme reaction to the problem of nonstationary error variances. If one method does outperform another, an agent may wish to adjust his weights for the following period based on the most recent observed performance. Bunn [4,5] considers the Bayesian treatment of this problem and provides suggestions for its implementation.

In this paper we consider the Bayesian analysis of composite forecasting, in cases where prior opinion is elicited in the form of a binomial distribution on outperformance probabilities. The paper is very much in the spirit of analysis begun by Bunn [5]. We present composites of actual practicing expert forecasters. We describe an elicitation experiment in which users are asked to assess the the relative likelihoods of one expert outperforming another in a finite number of trials. The beta distribution is used, as a natural conjugate to the Bernoulli distribution, to summarize the elicited prior information. Following Bunn's work, we update the elicited priors using Bayes' Theorem. Forecast performance statistics and posterior weights are presented for alternative priors.

OUTPERFORMANCE PROBABILITIES AND BETA PRIORS

The Bayesian treatment of composite forecasting (as presented in Bunn [5]) requires one to assign an initial weight to each alternative forecasting method. Initial weights are updated, period by period, with Bayes' Theorem. Interpretation of the weights as probabilities that one method will outperform another over a finite number of trials is suggested by Bunn [5]. An alternative interpretation, that assigns a subjective probability to the relative "truthfulness" of each model is not recommended, as it leads to several problems including the impossibility of ever confirming (or refuting) inductive models (Bunn [5]).

Composite forecasting based on outperformance probabilities should be viewed from a decision theory and not a scientific (confirmation and refutation) perspective. As an illustration of this point, one might well expect a univariate time series forecast to outperform a forecast based on a structural econometric model; while still maintaining that the structural model reflects a more appropriate description of the system generating the random variable.

A useful simplification in applying outperformance probabilities to a particular problem is the use of a conjugate prior distribution. Here we are interested in selecting a particular class of priors to summarize (up to a reasonable approximation) the true prior belief of a forecast user on the outperformance of each forecasting method. In our particular problem (the combination of two forecasting methods) each observation can be treated as a Bernoulli trial (method 1 will or will not outperform method 2 on any particular case). Combining prior information, which is summarized in a beta distribution, with data, generated as a sequence of Bernoulli trials, results, via Bayes' Theorem, in a beta posterior (DeGroot [9]). If prior and marginal distributions are correct, the posterior mean of the weights minimizes posterior expected squared-error loss.

For example, consider the beta prior, B, characterized by the parameters a_1 and a_2, and the sequence of indicators δ_i (δ_i=1 if method 1 outperforms method 2 on trial i, otherwise $\delta_i = 0$) over j realizations. The posterior mean on the weight associated with method 1(k) is given as:

$$\bar{k} = (a_1+s_j) \div (a_1+a_2+j)$$

where $s_j = \Sigma\delta_i$, i=1,...,j.

The parameters a_1 and a_2 reflect the initial shape of the prior distribution; that is, its center and dispersion. Prior weighting favoring model 1 is reflected by beta parameters a_1 greater than a_2; favoring model 2, a_1 less than a_2, and equal weighting by a_1 equal to a_2. Higher values of a_1 and a_2 reflect tighter prior density (see figure 1). For example, a uniform prior would obtain for a_1 and a_2 equal to one. This might be used in cases where little (nothing) is known on the forecast performance of two alternative models (people). Here posterior weights change quickly with each new observation. Tighter prior density (higher values of a_1 and a_2) will reflect more confidence in prior weights as they change more slowly with new observations. In cases where a_1 does not equal a_2, the prior distribution will not be centered over .5.

Several methods for assessing prior distributions have been described in the literature. Bessler [2] characterizes these as motivating and nonmotivating methods and suggests the former be used in applied settings. Below we elicit prior probabilities on outperformance in terms of discrete outcomes on a binominal distribution. That is to say, we set up an outperformance competition between two forecasting methods (people) and ask our assessor (a person) the probability of method 1 outperforming method 2 j times, where j runs from zero to N, the number of forecasts. The competition is set up such that, after the fact, we can observe the number of actual outperformances and reward the assessor according to a proper scoring rule. The beta function is fit to the discrete responses for each assessor using nonlinear least squares (Chamberlain [6]). The beta function is then standardized such that it integrates to unity.

COMPOSITES ON U.S. HOG PRICES

Table 1, column (2) gives actual, average seven market, quarterly hog prices from 1976 through 1984, as reported by the USDA. Columns (2) and (3) give the one-quarter lead-time forecasts made by the faculty in the Agricultural Economics Departments at the University of Missouri (column (3)) and Purdue University (column (4)). Both forecasts are expert opinion-type.

In terms of individual forecast performance, the Missouri forecasts have a mean squared error of 14.20; while the Purdue forecasts have a mean squared error of 19.14. Missouri outperforms Purdue in 19 cases; Purdue outperforms Missouri in 16 cases and there is one tie. In terms of runs, the longest run of outperformance is five (Missouri outperforms Purdue the last three quarters of 1979 and the first two quarters of 1980). Purdue had a run of four consecutive quarters of outperformance, beginning in the second quarter of 1976 and ending in the first quarter of 1977. There are seventeen runs; which does not seem to be too many or too few to believe that the forecasts are exchangeable with respect to order (non-Bayesians would say that for our problem the number seventeen falls within acceptable lower and upper limits at the five percent level of significance; see Hoel, appendix table 9).

We consider two types of analyses with these forecasts. First, we explore various tightness levels on the beta prior. Our prior will be centered, in all cases on equal weights (.5). Here, interest is in forecast performance associated with alternative levels of tightness on the beta prior. One might expect that as the forecast performance of an individual model deteriorates, the poorer will be the performance associated with tighter priors. Our second point of focus with these forecasts is on substantive experts' priors relative to outperformance.

Equal Prior Weights

Table 2 gives mean squared errors associated with six levels of tightness on the beta prior. The table presents composites of the Purdue-Missouri forecasts, the Purdue - a random walk forecast, Purdue - a constant forecast, Missouri - a random walk forecast, and Missouri - a constant forecast. The random walk forecast gives the most recent observed quarterly price (price in period t) as the forecast for the next period (forecast price in period t+1). The constant forecast gives the observed quarterly price in the fourth quarter of 1975 as the forecast over all 36 periods (obviously, a poor forecast). The random walk forecast has a mean squared error of 28.93; while the constant forecast has a mean squared error of 80.50.

Our reason for studying the random walk and constant forecast is that we are interested in composite forecasting performance under less than ideal conditions. That is, both the constant and random walk forecasts are poorer in mean squared error sense than either of the University forecasts. While the random walk forecast does have some merit (Litterman [12]); the constant has little to recommend it. One might expect that increasing tightness on the beta prior, which is centered on .5, will result in poor forecast performance in composites formed with the constant forecast.

Notice from Table 2 that composites formed between the Purdue and Missouri forecasts improve with increased prior tightness. That is, as we move from the uniform prior (1:1), where the data determine the weights entirely, to the very tight prior (200:200), where the data have very little influence on outperformance weights, the mean squared error falls about 6 percent. Results are mixed with respect to the random walk composites. Tightness improves the Purdue-random walk composite; while it does not improve the Missouri-random walk composite. The results are as we expected with respect to the constant forecast. Increased tightness - which gives less attention to the data - results in poor forecast performance for both Purdue and Missouri composites.

The results on tightness are perhaps generalizable and are certainly worth further study. In cases where the forecast user is fairly sure that he has two competent forecasts, tight *a priori* weights are perhaps advisable. In our case, one would not expect the Missouri forecast to always outperform the Purdue forecast; as there is incentive for individual forecasters to adapt after each realization (Clemen and Winkler [7] suggest that this behavior is plausible). On the other hand, when one method is suspect (constant forecast) it is perhaps wise to allow the data to quickly move the weights away from their prior mean.

Elicited Prior Weights

Figure 2 gives the prior beta distributions fit from elicited distributions of three agricultural economists on Purdue outperforming Missouri. The experts are all Ph.D. economists. Each is a faculty member in the Department of Agricultural Economics at Texas A&M University. Each also has knowledge of hog prices and is aware of both outlook groups. The subjective beliefs were assessed according to a discrete logarithmic scoring rule with small monetary rewards (the elicitation questionnaire is available from the senior author).

Note that expert two gives a probability edge to the Purdue forecasts; expert one gives Purdue a slight edge; while expert three gives favor to the Missouri forecasts. Expert one is slightly looser in his prior than both experts two and three (his beta parameters sum to a smaller number than do the parameters of assessors two and three.)

In Table 3 we present the results associated with composites, based on the expert opinion probabilities. Notice that expert three's prior probabilities result in the best composite forecast mean squared error. Recall, his beliefs were not symmetric - he favored the Missouri forecast.

DISCUSSION

In this paper we have considered the use of Bayesian outperformance probabilities in constructing composite forecasts. The Bayesian treatment of composite forecasting allows weights to change according to forecasting performance. However, the user is able to control the degree of change through his prior or initial beliefs on outperformance. Tighter initial distributions allow the posterior (final) weights to change much slower than do less tight prior distributions. The beta distribution is discussed as a convenient form in which to summarize prior beliefs when two forecasts are to be combined.

We illustrated Bayesian combinations with forecasts of U.S. hog prices. Using a prior centered on equal weights, we showed that increased tightness did improve forecast performance, when the individual forecasts had reasonable credibility. In addition, we presented a case where one forecast was not credible (constant). Here, increased tightness did not serve to improve performance. In addition to the illustrations on tightness, we elicited distributions on outperformance, which were not centered on equal initial weighting. Elicited priors of three agricultural economists were used to fit parameters of a beta distribution. In the particular example studied, this nonsymmetric prior performed well.

In all cases, peformance of the composite was better than the worst individual forecast. While, performance of the composite rarely improved on the best individual method; it was usually quite close to the best individual method. If one has two individual methods, both of which he considers credible, combining does seem to be a reasonable approach -- rather than selecting one method or the other.

If a user of forecasts has three or more individual methods, the outperformance idea can be extended through the Dirichlet or Matrix Beta distributions. Computer code for actual calculations of these as well the beta prior is available in Bunn [5] or from the senior author of this paper.

Table 1. Actual Quarterly Hog Prices* (2); Missouri (3) and Purdue (4) Forecasts by Time Period (1).

Period	Dollars Per Hundredweight		
(1)	(2)	(3)	(4)
197601	47.99	47.00	47.00
197602	49.19	47.00	48.50
197603	43.88	47.00	45.00
197604	34.25	36.00	35.00
197701	39.08	34.50	35.00
197702	40.87	35.50	32.50
197703	43.85	42.50	44.00
197704	41.38	37.50	37.50
197801	47.44	39.50	36.00
197802	47.84	48.50	42.00
197803	48.52	46.50	51.00
197804	50.05	48.50	45.00
197901	51.98	48.50	51.00
197902	43.04	45.00	48.00
197903	38.52	40.00	43.00
197904	36.39	35.00	32.50
198001	36.74	38.50	32.50
198002	31.18	35.50	37.00
198003	46.23	37.50	42.00
198004	46.44	43.50	45.00
198101	41.13	45.00	50.00
198102	43.62	42.00	44.00
198103	50.42	52.00	52.50
198104	43.63	49.50	48.00
198201	48.17	44.50	48.50
198202	56.46	53.00	51.00
198203	61.99	58.00	60.00
198204	55.12	59.50	61.00
198301	55.00	57.50	58.50
198302	46.74	53.50	52.00
198303	46.90	46.50	45.00
198304	42.18	41.50	40.50
198401	47.68	48.50	50.50
198402	48.91	49.50	52.00
198403	51.21	56.50	55.00
198404	47.65	45.50	46.50

*Forecasts are for one quarter lead times.

Table 2. Mean Squared Errors Associated with Composites of Individual Forecasts and Varying Levels of Prior Tightness.*

Beta Parameters	Composite				
	Purdue Missouri	Purdue Random Walk	Missouri Random Walk	Purdue Constant	Missouri Constant
1:1	15.56	18.71	17.05	18.31	12.83
2:2	15.39	18.53	17.06	18.09	12.94
10:10	14.93	18.18	17.16	20.10	15.97
20:20	14.79	18.11	17.24	21.64	18.07
80:80	14.64	18.08	17.38	23.93	21.25
200:200	14.61	18.08	17.43	24.63	22.25

*Composites are formed as convex combinations of each individual forecast; where weights are calculated from Bayes' Theorem - based on prior weights and sequentially updated with sample information. The update is over the period 1976-1984, quarterly data.

Table 3. Results on Composite Forecasting of U.S. Hog Prices Based on Elicited Prior Probabilities on Outperformance.*

Expert	Beta Parameters	MSE	Weight to Purdue	
			Minimum	Maximum
1	2.78:2.55	16.12	.603	.767
2	3.52:2.07	15.11	.429	.633
3	2.37:3.08	13.77	.231	.275

*Beta parameters are based on fit of discrete outperformance probabilities elicited with logarithmic scoring rule.

Figure 1. Beta Distributions for Alternative Parameters, a(1) equals a(2).

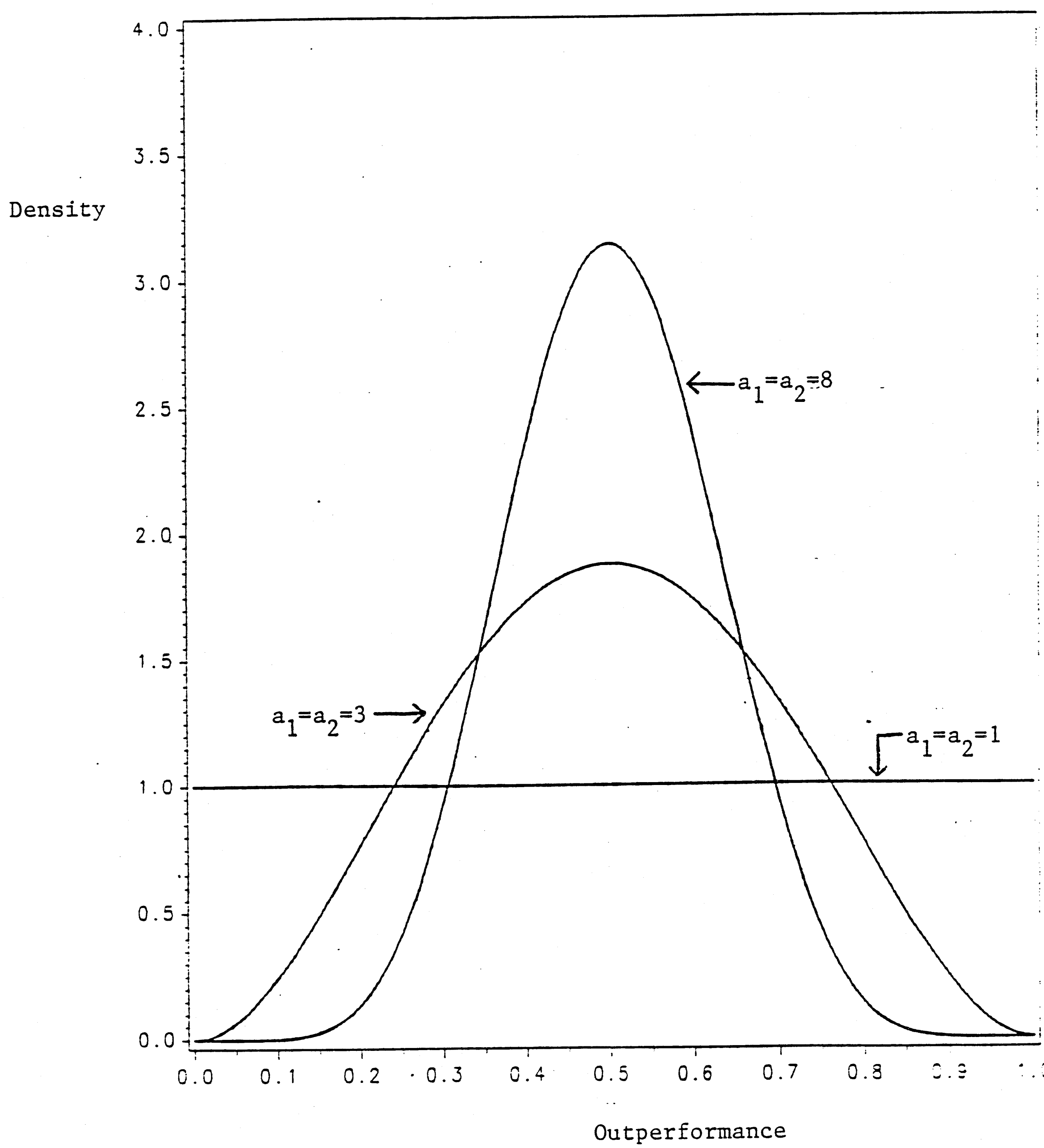
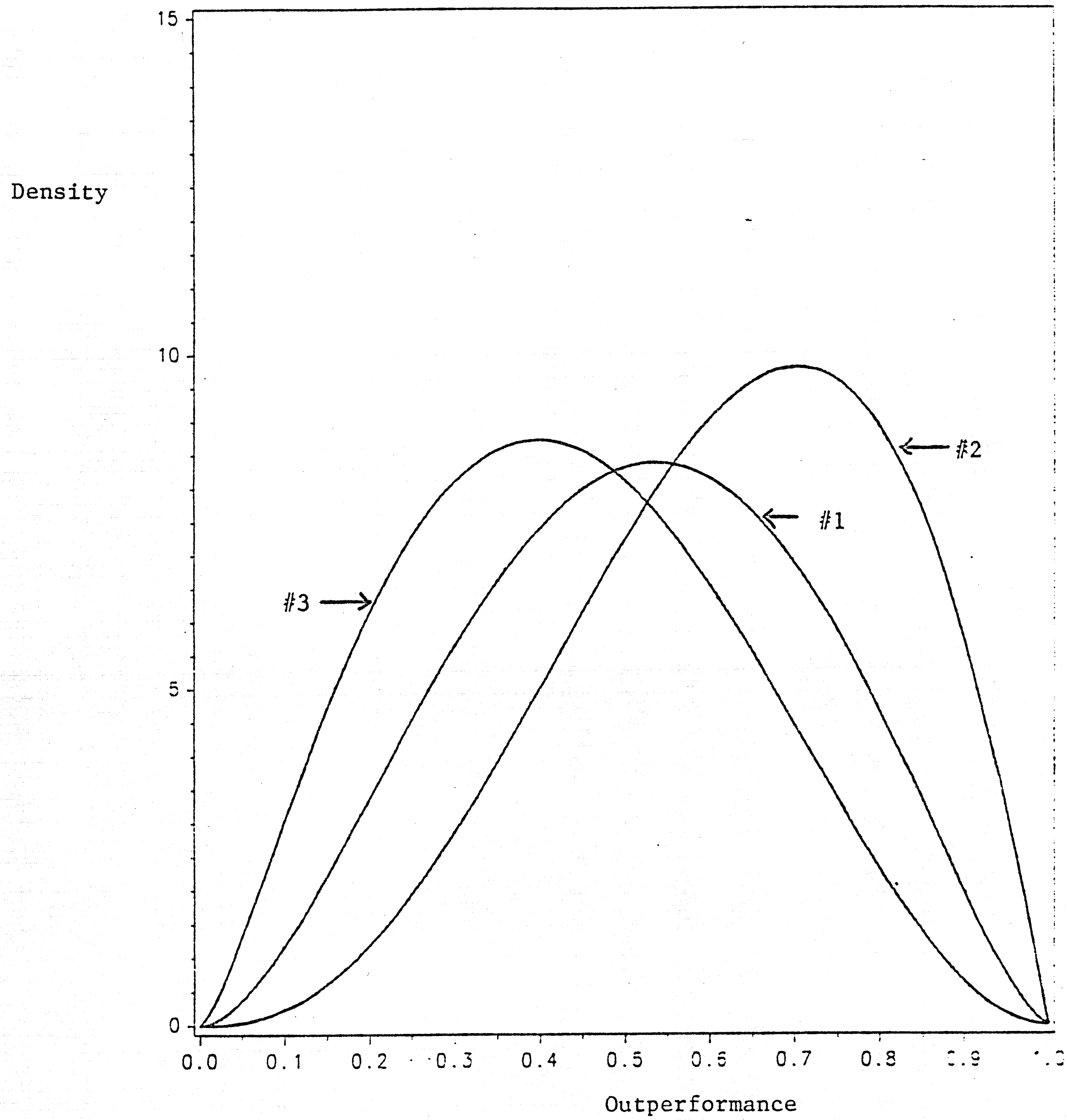


Figure 2. Fitted Distributions on the Probability that Purdue will Outperform Missouri in Hog Price Forecasts; Three Experts (1,2,3).*



*Parameters a_1 and a_2 are the estimated beta coefficients determined using nonlinear least squares. Experts are labeled #1, #2 and #3.

References

1. Bates, J. M. and C. W. J. Granger (1969) The Combination of Forecasts. *Operational Research Quarterly* 20, 451-468.

2. Bessler, D. A. (1984) Subjective Probability. Chapter 4 in P. Barry, editor, *Risk Management in Agriculture.* Ames: Iowa State University Press.

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

4. Bunn, D. W. (1985) Statistical Efficiency in the Linear Combination of Forecasts. *International Journal of Forecasting* 1, 151-163.

5. ______ (1978) *The Synthesis of Forecasting Models in Decision Analysis,* Birkhauser, Basel.

6. Chamberlain, P. J. (1986) Movement between Alternative Forage Levels: A Weiner Process. Dept. of Agricultural Economics, Texas A&M University.

7. Clemen, R. T. and R. L. Winkler (1986) "Combining Economic Forecasts," *Journal of Business and Economic Statistics* 4, 39-46.

8. deFinetti, B. (1974) *The Theory of Probability* New York: Wiley.

9. DeGroot, M. (1970) *Optimal Statistical Decisions* New York: McGraw-Hill.

10. Granger, C. W. J. and R. Ramanathan (1984) Improved Methods of Combining Forecasts *Journal of Forecasting* 3, 197-204.

11. Hoel, P. G. (1960) *Elementary Statistics* New York: Wiley.

12. Litterman, R. (forthcoming) Forecasting Vector Autoregressions, in P. Goel and A. Zellner, editors; *Bayesian Inference: Theory and Applications, Essays in Honor of Bruno deFinetti* North Holland.

13. Makridakis, S., A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen and R. Winkler (1982) The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition *Journal of Forecasting* (1), 111-153.