Future Directions in Risk Analysis for Agricultural Firms. Department of Agricultural and Applied Economics, University of Minnesota, August 1987.

The Evaluation of Probability Distributions with Special Emphasis on Price Distributions Derived from Option Premiums

Paul L. Fackler and Robert P. King

Information on price and yield probabilities is an important resource for farm level decision making in the face of uncertainty. As is true in the allocation of resources used in more tangible production processes, choices about the acquisition and use of probabilistic information should reflect both cost and contribution to the quality of decisions. It is often useful, then, to consider a range of sources of information on probabilities.

Subjective beliefs and empirical data are, of course, the two most familiar sources of probabilistic information. Hogarth, Spetzler and Stael von Holstein, Bessler, and, more recently, Norris and Kramer all provide good overviews of the literature on subjective probability. That literature focuses on alternative techniques for eliciting or encoding subjective probabilities and on the impact of cognitive biases introduced by the heuristics people use when they structure their probabilistic beliefs. Young, Pope and Ziemer, and Black, et. al. examine issues related to the use of probabilities based on empirical data. Those issues include the choice of modeling techniques for conditioning empirical distributions to reflect current information, the performance of alternative families of distributions, and the relationships between farm-level and county-level yield data.

With the initiation of trading in options on agricultural commodity futures contracts, interest in the probabilistic information contained in option premiums has grown. As Gardner suggested, futures price probability distributions can be derived directly from option premiums. These distributions reflect, in some sense, the beliefs of traders about potential price movements. They are attractive for use in farm level risk analyses because they can be constructed from a limited amount of readily available data and can be updated as often as option premiums change.

That probability distributions should reflect the subjective beliefs of the decision maker is a fundamental assumption in applied decision analysis. This suggests that the best source of information on probabilities is direct elicitation. Often, however, this often poses difficult problems in a practical setting. Probabilistic knowledge is frequently ill-structured and incomplete, and the elicitation process, itself, can be time consuming and expensive. There is a need, then, for support tools that help decision makers use other information sources, such as empirical data and option premiums. The Agricultural Risk Management Simulator (ARMS) microcomputer program (King, et.al.) is one example of such a support tool. If such tools are to be widely used, however, there is also a need for a better understanding of the performance of probability assessments based on alternative information sources.

Paul L. Fackler is an Assistant Professor, Department of Economics and Business, North Carolina State University. Robert P. King is an Associate Professor, Department of Agricultural and Applied Economics, University of Minnesota.

In this paper, we first explore the general question of how probability assessments can be evaluated. We then turn our attention to a more detailed examination and evaluation of price probability distributions based on option premiums. We first describe some methods for deriving price probability assessments from commodity option premiums. We then present an empirical examination of the performance of option based price forecasts for live cattle and soybeans. In the concluding section of this paper, we discuss the implications of our findings and needs for further research.

The Evaluation of Probability Assessments

Regardless of how a probability assessment is generated, it is essential that it be evaluated in order to judge its value in the decision making process. For example, if the probability assessments used in a business plan are overly optimistic, then unprofitable business enterprises may be adopted. Similarly, if the dispersion of these assessments is underestimated, then proper steps to protect against risk may not be taken. In either case financial difficulties can result.

There are three levels on which probability assessments can be evaluated. First, one can examine a single assessment for a unique random event. Second, one can evaluate a set of assessments associated with several events that are generated by some person or method. Finally, one can compare two or more sets of assessments to one another. These three levels can be examined using the criteria of coherence, reliability, and accuracy.[1]

Coherence

The first, and most basic level, is the examination of a single probability assessment for a particular random event. Such an assessment is coherent if it conforms to the laws of probability. Thus, if the assessment is expressed in terms of a cumulative distribution function (CDF), this function should be non-decreasing and right-continuous with range [0,1]. This property is easily checked and it will be assumed that such checks are performed. Thus it can be assumed henceforth that probability assessments are in the form of proper CDFs.

In examining a single assessment, there is little that can be said about its quality beyond coherence. In particular, the outcome of the associated random variable provides no information, except in the extreme case that the assessment assigned zero probability to the realized outcome. In general, to confront assessments with outcomes it is necessary to have a set of both. The reliability of the assessments can then be examined.

---

[1] Winkler uses the terms coherence, calibration, and expertise, while Morris and Feinburg discuss calibration and refinement. The discussion presented here has drawn on these articles and a number of the references they contain. The focus here, however, is mainly on the case of continuous random variables, while most of the literature on this topic focuses on the binomial case.

Reliability

The concept of reliability can be most easily introduced using the simplest case. Suppose that a given random variable, X, is believed to be described by a CDF, G(X), and that a large sample of realizations of X can be observed. G is said to be reliable if the proportion of realizations for which X is less than or equal to any given value, x, equals G(x). An equivalent, and more useful, statement of this condition is that the proportion of times G(X) is less than any given value, u, between zero and one, is equal to u.

In this example G is applicable to a replicable random variable. The more general situation, however, is that the probability assessments correspond to possibly unique events. Note that this assumes a subjectivist view of probability, since a strict frequentist view has no meaning in this context.

To illustrate such a situation, suppose that each month a person is asked to assess the probability distribution concerning next month's price of some commodity. In such a case it does not make sense to evaluate the reliability of a single assessment but rather the person or method (the assessor) supplying the assessments is evaluated. An important difference between this case and the more narrow one is that, in general, the CDF assessments will not all be the same, as assessors will condition their assessments on information specifically relevant to next month's price. Because the assessments will generally change each month it is not informative to examine the proportion of realized values below a given level and hence the first definition of reliability suggested above is not useful. There is only one realized price associated with each probability assessment. The second definition, however, is essentially still applicable. A probability assessor is said to be reliable if the proportion of the times the value of assessed CDFs is less than or equal to a given value, u, is equal to u, where each CDF is evaluated at the realized value of the associated random variable.

To formalize these ideas, let G_i be an assessment of the probability of a random variable, X_i, with i=1,..,n and let $U_i = G_i(X_i)$. U_i can be thought of as a new random variable, with CDF H_i, i.e. $Prob(U_i \leq u) = H_i(u)$, where u lies on the interval [0,1]. An assessor generating a set of CDFs, G_i, for the random variables X_i, is said to be reliable if $H_i(u) = u$, for all i. Thus the notion of reliability is equivalent to the property that the random variables, U_i, which equal the value of the assessed CDFs at the realized values of the X_i, are identically distributed uniformly on the interval [0,1].[2]

While the reliability of a probability assessor cannot be unambiguously determined in the manner in which the coherence of an assessment can be checked, it is nonetheless amenable to statistical examination. This can be accomplished using goodness of fit tests for the hypothesis that a sample of U's comes from a uniform [0,1] distribution. Stephens discusses and evaluates the power of five non-parametric goodness of fit tests: the Komolgorov, Cramer-von Mises, Kuiper, Watson, and Anderson-Darling. Each of these tests varies in relative power against specific alternative hypotheses and

---

[2] Some assumptions about the nature ofthe random variables associated with the assessments is needed to make the concept of reliability rigorous. See Morris for a discussion of the concept of exchangeability.

examination of all together can provide a more balanced assessment of the possibility of non-uniformity than the examination of any of them alone. However, if a single omnibus test is desired, evidence from Stephens and from further power tests by Quesenberry and Miller, suggest that the Watson test is the most reliable.

An alternate procedure involves the estimation and testing of some parametric form for the function H which is assumed to be stable over the sample. A natural choice of parametric families is the Beta distribution, the parameters (denoted here by p and q) of which can be estimated using maximum likelihood techniques. The Beta is well known, fairly flexible, and contains the uniform distribution as a special case.

This alternate procedure has the advantage of providing a means to adjust the function G, should it prove to be non-reliable. By definition the function $F(X)=H(G(X))$ is a reliable probability distribution. The function H has been called the calibration function (Curtis, et al.), since it provides a way to calibrate probability distributions, i.e., it provides a transformation to ensure the reliability of a probability assessment. Indeed, many authors refer to reliable assessments as well-calibrated (e.g. Morris and Feinburg).

If H is reasonably stable and can be estimated accurately from a sample of U's, then the reliable distribution F can be easily constructed. It should be noted that the issue of the stability of H is not an innocuous one, particularly if an assessor has the capacity to learn from previous assessment performance. From an evaluator's point of view, however, it may not be feasible to evaluate the stability of H, particularly if relatively few assessments exist.

The reliability concept is, however, a limited one. To illustrate its limitations, suppose that an assessor always uses the same probability distribution for next month's price. Even if this distribution is reliable, i.e. it is the long run marginal price distribution, it is unlikely that this will be of great value, since it takes no account of current market conditions. The long run marginal distribution might suggest, for example, that the price is on average three dollars and has a standard deviation of fifty cents. If the current price is one dollar and the market has been quiet for the last month, few people would be interested in this assessor's predictions.

<u>Accuracy</u>

This raises the issue of how different sets of assessments can be compared, which is the third level of evaluation mentioned above. One could address this issue by asking whether one set of assessments are more reliable than another. That set would be judged more reliable whose calibration function, H, is most nearly uniform, using some distance measure. This is not a very satisfactory solution given the limited scope of the reliability criterion. One set of assessments could be reliable by using the long run marginal distribution, while another set might use more information but tend to slightly overestimate the mean and therefore be deemed less reliable. This situation is analogous to the problem of evaluating statistical estimators, with reliability likened to the property of unbiasedness. An estimator can be unbiased but have large variance, while another may be somewhat biased but have low variance. In terms of mean squared error, the latter would therefore be deemed a better estimator.

In the estimation analogy the biased estimator is more accurate than is the unbiased one, using the mean squared error criteria. Some analogous concept of accuracy is needed for the evaluation of probability assessments. A natural candidate uses the likelihood criterion. Given a set of realized outcomes, the higher the assessed likelihood of these outcomes, the more accurate a set of assessments is said to be. The application of the likelihood concept is most easily carried out when the random variable of interest is either discrete or continuous, but not mixed. The continuous case is of primary interest here and it will therefore be assumed that the probability assessment is in a form such that an associated density is obtainable. If the probability assessments and the associated random variables are independent, the joint likelihood for a sample of realized outcomes is the product of the values of the densities evaluated at the realized outcomes. Using the notation above, with g_i being the density associated with G_i, the accuracy of a set of assessments can be measured by the product of the $g_i(x_i)$ or, alternately, as the sum of the $\ln(g_i(x_i))$.

In the literature on the evaluation of probability assessments a concept that has attracted much attention is that of the scoring rule (Savage, Winkler). Such rules associate each probability assessment and realized outcome with a value and the sum of these values becomes an overall score by which an assessor is judged. Much of the literature on scoring rules is concerned with the use of rules that encourage persons to reveal their true beliefs. Such rules are called proper scoring rules. In the interests of avoiding excessive and tangential discussion, scoring rules are not further discussed except to note in passing that the use of likelihood as an accuracy criteria is equivalent to the use of the logarithmic scoring rule, which is a proper scoring rule.

The evaluation methods discussed in this section are implemented in the Section 3 of this paper. First, however, we discuss particular methods for deriving price probability distributions using option premiums. These methods are then applied and the resulting assessments evaluated.

<u>Price Probability Distributions from Option Premiums</u>

In this section methods are outlined by which price probability assessments can be derived using commodity option premiums. It is believed that people working in risk assessment and management will find these methods useful, especially given their relatively minimal input requirements and the ease with which they can be generated.

In order to use options markets as an information source concerning probability distributions it is necessary to have a theory of option pricing. The most general theory, as outlined by Cox and Ross, relies only on the absence of arbitrage, which is a necessary condition for market equilibrium under the weak assumption that agents prefer more to less. A fundamental theorem of modern finance is that any traded asset has current value equal to the discounted value of its expected returns, with the expectation taken with respect to some artificial, or risk adjusted probability distribution.[3] Let $G(Y_T)$ denote such a distribution relevant for an option on a futures contract, where the futures price is denoted Y_t, and T is the option's expiration date. The return on a European put option is $\max(0,x-Y_T)$, where x is the exercise price. Therefore its current value is

$$(1) \quad b(T)\int_{o}^{\infty} \max(0,x-Y_T)dG(Y_T)$$

where b(T) is the current price of a bond paying $1 at time T. This can be written equivalently as

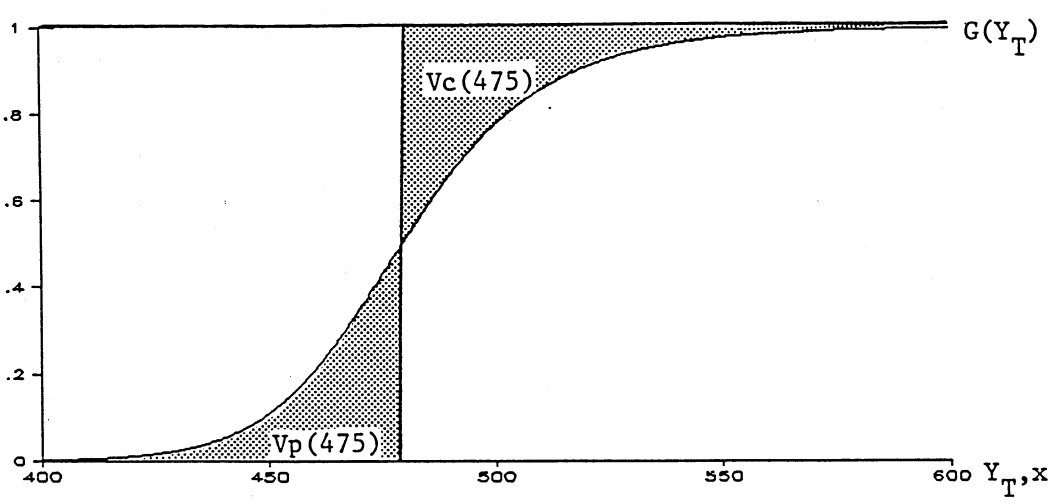
$$(2) \quad b(T)\int_{0}^{x} G(Y_T)dY_T,$$

since G(T) is a probability distribution. A European call can be written

$$(3) \quad b(T)\int_{0}^{\infty} \max(0,Y_T-x)dG(Y_T) \quad = \quad b(T)\int_{x}^{\infty}[1-G(Y_T)]dY_T$$

Let $V_p(x)$ be the interest rate adjusted value of a put with exercise price x, equation (2) divided by b(T). From (2) this is equal to the area under $G(Y_T)$ evaluated from 0 to X. The value of an interest rate adjusted call, $V_c(X)$, can be similarly defined and is equal to the area above G and below 1 evaluated from X to infinity. These option valuation formulas are illustrated in Figure 1.

Figure 1. Option Premiums as Areas Associated with the Risk Adjusted Probability Distribution, G



---

[3] See also Ross, Garman, Breeden & Litzenberger, and Green & Srivastava for further discussion of this theorem. Note that this distribution is not necessarily unique.

These relationships suggest a means by which the function G can be evaluated, given a set of observed market option premiums for several exercise prices. Two methods of approximation are suggested here, one parametric and one nonparametric. The first entails choosing a family of functional forms for G and finding the parameters of that family of functions yielding values of the $V_p(x)$ and $V_c(x)$ that most closely match those observed.

A number of families of distributions exist that could be used to approximate the G function. The lognormal distribution has been used by a number of researchers (Chiras & Manaster, Latane & Rendelman), and is associated with the well known Black commodity option pricing formula. A common procedure has been to use only the at-the-money option (i.e., the option with the exercise price closest to the current asset price) and to solve for the unknown volatility parameter of the Black formula. This completely describes the function G under the assumption of lognormality, since the expected value of the future price is assessed to equal the current futures price. Since the volatility parameter is uniquely determined by the option premium (given its strike price, the current futures price, the price of an appropriate risk free bond, and the time until expiration), this parameter has been called the implied volatility.[4]

Other, more flexible distributions exist that can potentially approximate the G function more closely than does the lognormal. One distribution that has proved useful is the Burr-12 or Singh-Madalla distribution, the CDF of which is

$$(4) \quad G(y) = 1-(1+(y/b)^{a})^{-q}.$$

Details on this distribution can be found in McDonald, Singh and Madalla, Rodriquez, Tadikamalla, and Fackler.

Using such a flexible family of distributions, the fitted values of the option premiums implied by the parameters of the distribution are made as close as possible to the observed market premiums. The least squares criteria is convenient in this context, as the approximation problem then reduces to a non-linear least squares estimation problem, for which computer software is readily available.

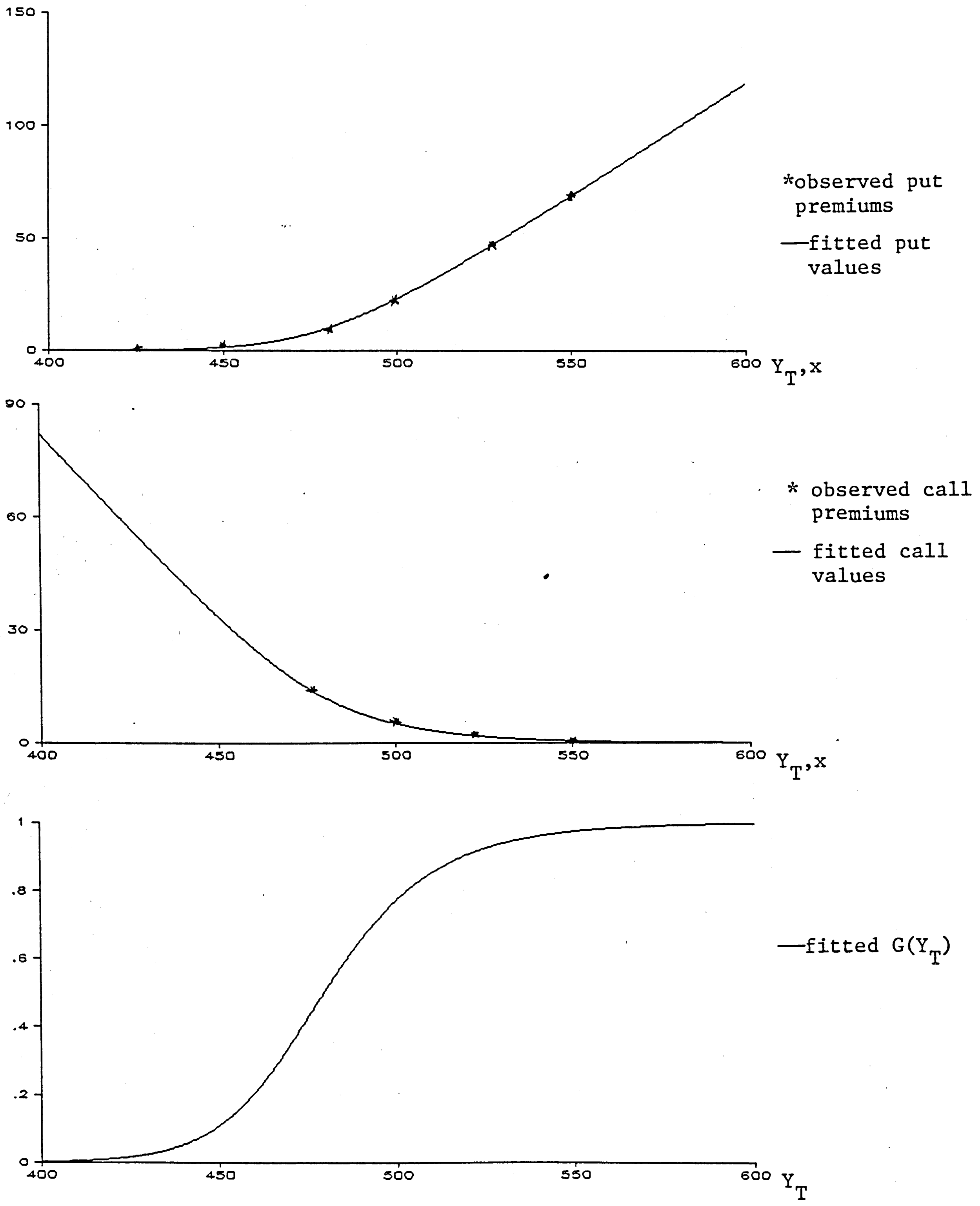
The basic idea is illustrated in Figure 2 with an example from the soybean market using the Burr-12 distribution. The top two figures show the (interest rate adjusted) observed and fitted put and call values, respectively, while the bottom figure shows the fitted G function.

The alternate, nonparametric approach relies on application of the mean value theorem for integrals to (2) to derive

$$(5) \quad V_p(x_2) - V_p(x_1) = \int_{x_1}^{x_2} G(Y_T)dY_T = (x_2-x_1)\, G(\hat{x}),$$

[4] The implied volatility of the Black formula should be multiplied by the square root of the time until expiration so it applies to the expiration date futures price.

Figure 2. An Example of an Approximation to the Function G Using the Burr-12 Distribution



where $x_1 \leq \hat{x} \leq x_2$.

Therefore,

$$(6)\quad G(\hat{x}) = \frac{V_p(x_2) - V_p(x_1)}{x_2 - x_1}.$$

Similarly,

$$(7)\quad V_c(x_2) - V_c(x_1) = \int_{x_1}^{x_2} [1-G(Y_T)]dY_T$$
$$= (x_2-x_1)\,[1-G(\hat{x})]$$

So

$$(8)\quad G(\hat{x}) = 1 - \frac{V_c(x_2) - V_c(x_1)}{x_2 - x_1}.$$

An operational procedure can be derived from these relationships by taking $\hat{x}$ to be the midpoint between each consecutive exercise price. If n options are traded this provides n-1 points on G(x) (puts and calls can be evaluated separately and the resulting G(x) averaged). These points are then linearly interpolated. The tail areas can be approximated using the function

$$(9)\quad G(x) = \frac{\left(\frac{x}{b}\right)^a}{1 + \left(\frac{x}{b}\right)^a}$$

where a and b are determined by the first two and last two values of G(x) for the lower and upper tails, respectively.[5] This method is illustrated in Figure 3, using the same data as previously.

Empirical Applications: Cattle and Soybean Futures Prices

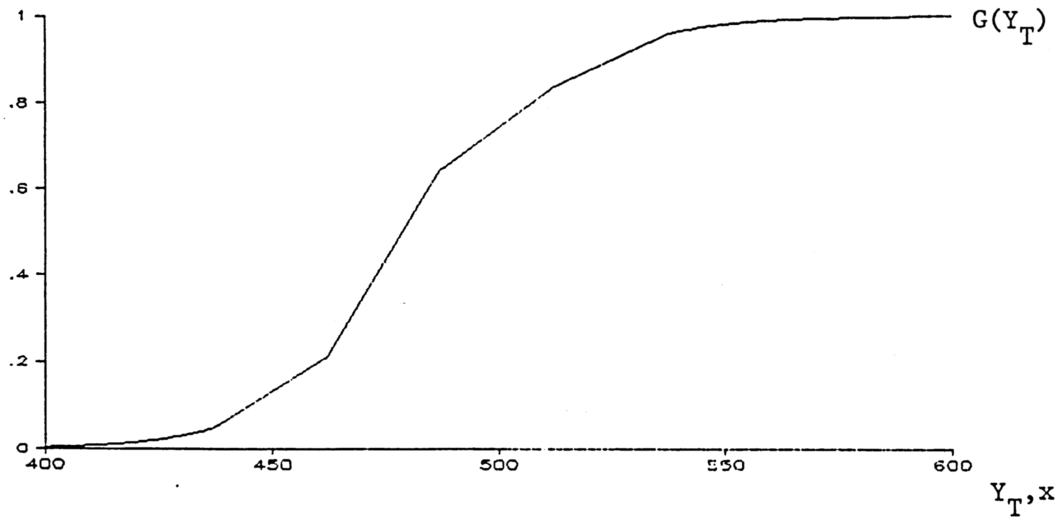
This section applies the methods described in the previous two sections to the cattle and soybean futures markets. Options trading in these markets began at the end of October, 1984. The options are written on futures contracts, which expire every other month beginning in February for cattle and in January for soybeans.[6] In this study 13 cattle contracts were used, beginning with the February, 1985 and ending with the February, 1987 contract, while in soybeans, 12 contracts were used, beginning with the March, 1985 and ending with the January, 1986 contract.

---

[5] This is the CDF for the two parameter Fisk distribution. It is a special case of the Burr-12, with q=1.

[6] The August soybean contract is not used in this study.

Figure 3. An Example of an Approximation to the Function G Using the Non-Parametric Method



For each contract, three sets of probability assessments are made for each of three approximation methods. The three sets of assessments correspond to the periods 8, 6, and 4 weeks prior to the expiration of the options. Three approximation methods are used: the at-the-money implied volatility (IV), the least squares fit using the Burr-12 distribution (BR), and the non-parametric (NP). This yields a total of nine samples for evaluation and allows comparison between the three methods. It should be noted, however, that the samples generated using different periods until expiration should not be thought of as mutually independent, random samples, even though the values within each sample can be taken to be independent.

The data used was taken primarily from the Wall Street Journal, though futures prices and option premiums for soybeans prior to 1986 were obtained from Chicago Board of Trade tapes.

Reliability test results for the cattle market are provided in Tables 1 and 2, and in the top half of Figure 4. Table 1 gives the realized values of the U_i for each of the periods until expiration and approximation methods. Comparison of the values for each of the methods reveals that all three tended to give fairly similar results. Tests evaluating the reliability of these assessment methods are given in Table 2. The first section of this table shows the values of the five modified nonparametric test statistics described by Stephens, applied to each of the nine samples from Table 1. In no case were any of these values significant at the 0.15 level, suggesting that the hypothesis that these methods provide reliable probability assessments cannot be rejected.

Table 1. Cattle Market - Values of U using the IV, BR, and NP Approximation Methods

Contract	8 Week IV	8 Week BR	8 Week NP	6 Week IV	6 Week BR	6 Week NP	4 Week IV	4 Week BR	4 Week NP
2/85	0.315	0.323	0.299	0.411	0.403	0.388	0.231	0.230	0.236
4/85	0.091	0.105	0.103	0.032	0.043	0.048	0.107	0.096	0.107
6/85	0.179	0.168	0.158	0.407	0.391	0.395	0.589	0.570	0.553
8/85	0.001	0.006	0.001	0.001	0.003	0.002	0.024	0.026	0.032
10/85	0.722	0.738	0.700	0.433	0.420	0.395	0.668	0.657	0.675
12/85	0.892	0.905	0.911	0.739	0.748	0.762	0.684	0.666	0.679
2/86	0.150	0.148	0.154	0.288	0.271	0.288	0.319	0.303	0.324
4/86	0.274	0.258	0.269	0.296	0.282	0.285	0.345	0.325	0.313
6/86	0.202	0.192	0.203	0.415	0.385	0.408	0.641	0.640	0.629
8/86	0.897	0.903	0.924	0.896	0.903	0.918	0.625	0.609	0.602
10/86	0.694	0.707	0.690	0.680	0.697	0.680	0.800	0.813	0.785
12/86	0.852	0.856	0.843	0.866	0.865	0.864	0.884	0.893	0.891
2/87	0.499	0.494	0.489	0.729	0.744	0.733	0.811	0.819	0.802

Table 2. Cattle Market - Reliablity Tests

Modified Nonparametric Goodness of Fit Test Statistics*

	8 Week			6 Week			4 Week		
Test Statistic	IV	BR	NP	IV	BR	NP	IV	BR	NP
Kolmolgorov	0.839	0.809	0.898	0.687	0.733	0.778	0.766	0.698	0.634
Cramer-von Mises	0.089	0.097	0.100	0.042	0.041	0.044	0.048	0.036	0.031
Kuiper	1.172	1.174	1.202	1.213	1.194	1.295	1.224	1.120	1.107
Watson	0.065	0.078	0.073	0.055	0.052	0.056	0.064	0.055	0.051
Anderson-Darling	0.783	0.726	0.889	0.715	0.540	0.574	0.412	0.357	0.341

Beta Distribution Parameters

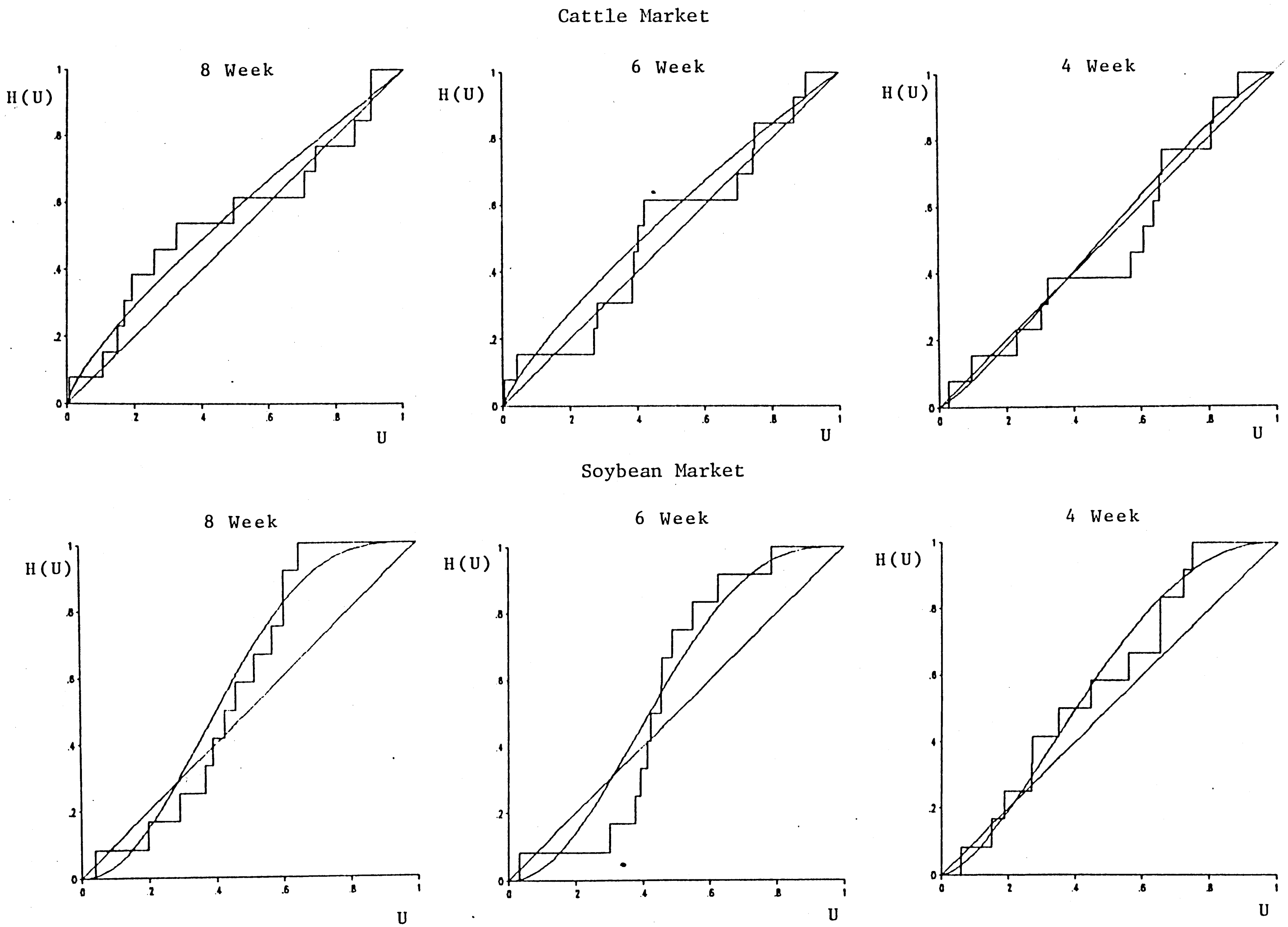
	8 Week			6 Week			4 Week		
Beta Parameter	IV	BR	NP	IV	BR	NP	IV	BR	NP
p	0.672	0.747	0.610	0.692	0.805	0.759	1.233	1.192	1.289
q	0.914	0.955	0.840	0.931	1.023	0.973	1.272	1.249	1.343

Beta Distribution LR Test Statistics and p-values

	8 Week			6 Week			4 Week		
	IV	BR	NP	IV	BR	NP	IV	BR	NP
LR Statistic	1.836	1.018	2.641	1.614	0.745	0.963	0.429	0.353	0.625
p-value	0.399	0.601	0.267	0.446	0.689	0.618	0.807	0.838	0.732

*None of these test statistics are significant at the 0.15 level.

Figure 4. Empirical CDF and Fitted Beta Distributions for BR Case



The second section of the table gives the maximum likelihood values of the Beta distribution fitted to each of the samples and provides a parametric forms for the calibration function, H. Of interest is the fact that, for the 8 and 6 week cases, the values of these parameters are typically less than 1, while in the 4 week case they are greater than 1. This lends further support to the hypothesis of reliability, since the nature of potential bias suggested by the calibration function changes for the different time periods. The Beta likelihood ratio test is given in the third section of the table. In all cases, the p-values for these tests are high, again supporting the hypothesis that the methods provide reliable assessments.

Figure 4 provides a graphical documentation of these results for the BR case. Each graph shows the empirical distribution of a sample of the U's, together with the fitted Beta calibration function. The uniform CDF (the 45 degree line) is shown as a reference; large divergences from this line indicate non-reliable assessments. In these cases the divergences are not large.

Measures of accuracy are provided in Table 3, which gives the log likelihood values for each of the assessments, together with the sums for each method and time until expiration. The sum of the individual log likelihoods serve as a basis for comparing the three different methods. For each of the 3 periods until expiration the ranking of the methods is the same, with the BR method favored over the IV, which in turn performed better than the NP method. It is difficult to know, however, whether the size of the differences exhibited are significant and while this is an important issue, it deserves more attention than is possible here. It is also of considerable interest to evaluate other methods of probability assessment, such as the use of time series procedures. These topics, however, are left for further study and the accuracy results presented should be viewed as illustrative.

The results for the soybean market tell a somewhat different story than those for the cattle market. Reliability results analogous to those presented for cattle are given in Tables 4 and 5, and in the bottom half of Figure 4. Again, each of the methods yielded values of the U's that seem fairly close. The test statistics given in Table 5, however, indicate that the reliability hypothesis is not supported in the soybean market as it was for cattle. In particular, the hypothesis is rejected at at least the 0.05 level of significance with both the Kuiper and the Watson test statistics, for the 8 and 6 week cases and for all of the methods. These two test statistics tend to be most sensitive to assessments that misrepresent the dispersion of the distribution.

This interpretation is reinforced by the values of the beta parameters, which are, in all cases, greater than one. Such values are associated with a bell shaped beta density. In general, areas of the density associated with the calibration function, H, which have more mass than does the uniform density are associated with too little mass in the probability assessment. A bell shaped calibration density is therefore associated with too little mass in the central area of the distribution and too much in the tails. These conclusions can also be reached by examination of the empirical and fitted beta representations of the H functions depicted in Figure 4 for the BR cases. The likelihood ratio tests provide further confirmation that the hypothesis of reliability should be rejected. The p-values associated with these tests are all, with the exception

Table 3. Cattle Market - Log likelihood Values
Using the IV, BR, and NP Approximation Methods

	8 Week			6 Week			4 Week		
Contract	IV	BR	NP	IV	BR	NP	IV	BR	NP
2/85	-2.14	-2.10	-2.30	-1.98	-1.98	-2.03	-1.95	-1.83	-2.35
4/85	-2.93	-3.10	-3.22	-3.58	-3.60	-3.62	-2.56	-2.69	-2.09
6/85	-2.66	-2.78	-2.69	-2.07	-2.02	-2.00	-1.88	-1.83	-1.83
8/85	-6.60	-5.84	-7.77	-7.24	-6.29	-6.77	-4.01	-4.15	-4.02
10/85	-2.80	-2.79	-2.97	-2.37	-2.37	-2.07	-2.19	-2.14	-2.60
12/85	-3.38	-3.40	-3.41	-2.63	-2.56	-2.19	-2.31	-2.19	-2.11
2/86	-2.93	-3.10	-3.07	-2.48	-2.49	-2.55	-2.30	-2.28	-2.34
4/86	-2.82	-2.82	-3.01	-2.57	-2.58	-2.62	-2.44	-2.47	-2.77
6/86	-2.79	-2.85	-2.89	-2.55	-2.51	-2.87	-2.57	-2.48	-2.67
8/86	-3.53	-3.67	-3.63	-3.48	-3.57	-3.53	-2.48	-2.38	-2.59
10/86	-2.92	-2.88	-2.89	-2.54	-2.50	-2.64	-2.64	-2.69	-2.73
12/86	-3.03	-3.09	-2.87	-2.86	-2.88	-3.50	-2.93	-2.98	-2.90
2/87	-2.32	-2.23	-2.36	-2.46	-2.45	-2.71	-2.36	-2.44	-2.67
SUM	-40.86	-40.64	-43.08	-38.83	-37.80	-39.09	-32.62	-32.54	-33.67

Table 4. Soybean Market - Values of U Using the IV, BR, and NP Approximation Methods

	8 Week			6 Week			4 Week		
Contract	IV	BR	NP	IV	BR	NP	IV	BR	NP
3/85	0.442	0.426	0.438	0.623	0.625	0.603	0.623	0.652	0.626
5/85	0.567	0.569	0.563	0.772	0.786	0.751	0.627	0.652	0.618
7/85	0.281	0.292	0.286	0.539	0.549	0.557	0.697	0.721	0.722
9/85	0.097	0.039	0.050	0.108	0.031	0.038	0.080	0.058	0.083
11/85	0.379	0.390	0.376	0.375	0.378	0.379	0.431	0.447	0.453
1/86	0.581	0.603	0.576	0.415	0.413	0.422	0.736	0.747	0.704
3/86	0.481	0.515	0.515	0.419	0.423	0.426	0.339	0.350	0.371
5/86	0.372	0.368	0.393	0.415	0.454	0.445	0.287	0.267	0.321
7/86	0.601	0.650	0.630	0.424	0.488	0.470	0.440	0.558	0.503
9/86	0.188	0.197	0.219	0.338	0.392	0.364	0.183	0.187	0.204
11/86	0.576	0.458	0.578	0.443	0.456	0.445	0.273	0.272	0.309
1/87	0.571	0.605	0.625	0.291	0.300	0.309	0.147	0.149	0.175

Table 5. Soybean Market - Reliability Tests

Modified Nonparametric Goodness of Fit Test Statistics

Test Statistic	8 Week IV	8 Week BR	8 Week NP	6 Week IV	6 Week BR	6 Week NP	4 Week IV	4 Week BR	4 Week NP
Kolmolgorov	1.441@	1.266*	1.337*	1.136	1.055	1.136	1.005	0.914	1.005
Cramer-von Mises	0.259	0.259	0.259	0.343	0.310	0.343	0.147	0.115	0.147
Kuiper	1.865#	1.754@	1.865@	1.990#	1.877#	1.990#	1.363	1.175	1.363
Watson	0.228#	0.208@	0.228@	0.305#	0.285#	0.305#	0.094	0.058	0.094
Anderson-Darling	1.427	1.418	1.427	1.737	1.594	1.737	0.963	0.810	0.963

Beta Distribution Parameters

Beta Parameter	8 Week IV	8 Week BR	8 Week NP	6 Week IV	6 Week BR	6 Week NP	4 Week IV	4 Week BR	4 Week NP
p	3.283	2.282	2.527	3.510	2.196	2.475	1.874	1.640	2.119
q	4.514	3.257	3.425	4.660	2.940	3.391	2.773	2.303	2.926

Beta Distribution LR Test Statistics and p-values

	8 Week IV	8 Week BR	8 Week NP	6 Week IV	6 Week BR	6 Week NP	4 Week IV	4 Week BR	4 Week NP
LR Statistic	9.398	6.314	6.790	9.775	5.456	6.689	4.998	3.572	5.403
p-value	0.009	0.043	0.034	0.008	0.065	0.035	0.082	0.168	0.067

*: significant at the 0.1 level
@: significant at the 0.05 level
#: significant at the 0.025 level

of the 4 week BR case, lower that 0.1 and in the 8 and 4 week case mostly below the 0.05 level.

Accuracy results for the soybean market are provided in Table 6. The relative ranking of the three methods is again unambiguous, with the BR method performing best, followed by the IV method and then the NP method. As already mentioned, there is no presumption that the differences exhibited by the methods are significant. Indeed, the methods seem to yield fairly similar results in as much as the realized values of the assessed distributions (Tables 1 and 4) are typically within a few percentage points of one another. In a business context, it is an open question whether such small differences would affect decisions significantly.

Summary and Conclusions

The evaluation of probability assessments has been discussed in this paper. Where such assessments are used in business decision making settings, such evaluation is critical if the use of probabilistic information is to contribute to the quality of the decisions made. The properties of coherence, reliability, and accuracy are suggested as criteria by which probability assessments can be judged. The latter two criteria, in particular, provide a way by which probability assessments can be evaluated relative to the realized outcomes of their associated random variables.

We have presented only a brief introduction to the subject of the evaluation of probability assessments. One important issue that has not been discussed is that of how assessors can learn from previous outcomes to adjust their future assessments. This involves the use of feedback rules that would enable an assessor to dynamically calibrate non-reliable assessment methods or to seek new information sources if assessments are not sufficiently accurate. A related issue is how assessments from disparate sources, such as subjective assessments and those based on probabilistic models and empirical data, should best be combined. To date, the literature on these topics, which sometimes appears under the heading of the expert problem, is far from satisfactory.

The primary focus of this paper is on methods for assessing price probability distributions using option market data. The newness of agricultural option markets makes this a timely topic, but also makes evaluation of these methods somewhat preliminary. Nonetheless, the applications in the cattle and soybean market demonstrate the potential of the methods. Option premiums are a relatively inexpensive source of frequently updated information on price probability distributions. Options data are easily obtained at low cost and can be regularly updated to provide assessments based on current market conditions. Furthermore, software to implement these methods is relatively simple. This is particularly true for the nonparametric and implied volatility methods, and the former method has been incorporated into the ARMS software previously mentioned. While the parametric method, which relies on the solution to a non-linear least squares problem, requires more programming, library routines that can be incorporated into software are readily available.

Our results indicate that option based assessments of cattle price distributions are reliable, while those for soybean price distributions tend to

Table 6. Soybean Market - Log likelihood values using the IV, BR, and NP Approximation Methods

	8 Week			6 Week			4 Week		
Contract	IV	BR	NP	IV	BR	NP	IV	BR	NP
3/85	-4.68	-4.60	-4.82	-4.51	-4.41	-4.43	-4.33	-4.26	-4.38
5/85	-4.65	-4.57	-4.53	-4.67	-4.70	-4.84	-4.22	-4.13	-4.39
7/85	-4.60	-4.47	-4.85	-4.22	-4.17	-4.14	-4.23	-4.31	-4.02
9/85	-5.38	-5.26	-5.86	-5.41	-5.63	-5.92	-5.13	-5.28	-5.24
11/85	-4.47	-4.39	-4.59	-4.30	-4.19	-4.32	-4.01	-3.94	-3.97
1/86	-4.29	-4.21	-4.41	-4.16	-4.08	-4.18	-4.34	-4.31	-4.38
3/86	-4.52	-4.42	-4.49	-4.56	-4.43	-4.58	-4.42	-4.19	-4.33
5/86	-4.26	-4.14	-4.28	-4.04	-3.91	-3.98	-4.17	-3.98	-4.11
7/86	-4.26	-4.31	-4.10	-4.29	-4.17	-4.17	-3.86	-3.87	-3.84
9/86	-4.75	-4.28	-4.64	-4.33	-4.02	-4.33	-4.61	-4.40	-4.97
11/86	-4.53	-4.41	-4.35	-4.14	-4.13	-3.95	-4.09	-4.03	-4.05
1/87	-4.22	-4.22	-4.06	-4.23	-4.07	-4.58	-4.22	-4.14	-4.97
SUM	-54.59	-53.29	-54.97	-52.86	-51.91	-53.43	-51.63	-50.85	-52.64

overstate price variability. These findings demonstrate the ability of statistical tests presented here to both confirm and reject the null hypothesis of reliability. More important, they point to the need for further examination of the causes for apparently systematic biases in the pricing of soybean options.

Further work on evaluating the usefulness of option based assessment methods needs to be expanded to compare the accuracy of such methods with that of assessments based on other methodologies, such as time series modeling. Another fruitful area for future inquiry is the development methods for combining probabilistic information from disparate sources, such as time series models and options based assessments, in a way that fully accounts for the information about the entire probability distribution contained in each.

References

Beckers, S. "Standard Deviations Implied in Option Prices as Predictors of Future Stock Price Variability." Journal of Banking and Finance. 5(1981):363-81.

Bessler, D.A. "Subjective Probability." Risk Management in Agriculture, ed. Peter J. Barry, pp. 43-52. Ames, Iowa: Iowa State University Press, 1984.

Black, F. "The Pricing of Commodity Contracts." Journal of Financial Economics. 3(1976):167-79.

Black, F. & M. Scholes. "The Valuation of Option Contracts and a Test for Market Efficiency." Journal of Finance. 27(1972):399-417.

Black, F. & M. Scholes. "The Pricing of Options and Corporate Liabilities." Journal of Political Economy. 81(1973):637-59.

Black, J.R., R.P. King, J. Skees, B. Scherle, & J. Driscol. "Crop Yield Variability: Concepts and Empirical Applications." Invited Paper presented at the Joint Meeting of the Western and Canadian Agricultural Economics Associations. Saskatoon, Saskatchewan, Canada. July, 1985.

Breeden, D.T. & R.H. Litzenberger. "Prices of State-contingent Claims Implicit in Option Prices." Journal of Business. 51(1978):621-51.

Chiras, D.P. & S. Manaster. "The Informational Content of Option Prices and a Test of Market Efficiency." Journal of Financial Economics. 6(1978):213-34.

Cox, J.C., J.E. Ingersoll, & S.A. Ross. "The Relationship Between Forward Prices and Futures Prices." Journal of Financial Economics. 9(1981):321-46.

Cox, J.C. & S.A. Ross. "A Survey of Some New Results in Financial Option Pricing." Journal of Finance. 31(1976B):383-402.

Curtis, P.E., W.R. Ferrell, & I. Solomon. "A Model of Calibration for Subjective Probability Distributions and its Application to Aggregation of Judgements." Journal of Forecasting. 4(1985):349-61.

De Finneti, B. "Does it Make Sense to Speak of 'Good Probability Appraisers'?" in Probability, Induction and Statistics: The Art of Guessing. Wiley, 1972.

DeGroot, M.H. & S.E. Fienberg. "The Comparison and Evaluation of Forecasters." The Statistician. 32(1983):12-22.

Fackler, P.F. "The Informational Content of Option Premiums." Unpublished Ph.D. Dissertation. University of Minnesota, 1986.

Fackler, P.F. & R.P. King. "Deriving Probability Assessments from Option Premiums." In draft, 1987.

Gardner, B.L. "Commodity Options for Agriculture." American Journal of Agricultural Economics. 59(1977):986-92.

Garman, M.B. "An Algebra for Evaluating Hedge Portfolios." Journal of Financial Economics. 3(1976):403-27.

Green, R.C. & S. Srivastava. "Risk Aversion and Arbitrage." The Journal of Finance. 40(1985):257-268.

Hogarth, R.M. "Cognitive Processes and the Assessment of Subjective Probability Distributions." Journal of the American Statistical Association. 70(1975):271-89.

Hogg, R.V. & S.A. Klugman. Loss Distributions. John Wiley & Sons, New York. 1984.

King, R.P., J.R. Black, F.J. Benson, and P.A. Pavkov. "The Agricultural Risk Management Simulator Microcomputer Program." Paper presented to the Southern Agricultural Management Research Information Exchange Group, Fayetteville, Arkansas, June 1986.

Latane, H. & R. Rendleman. "Standard Deviations of Stock Price Ratios Implied by Option Premia." Journal of Finance. 31(1976):369-81.

Matheson, J.E. & R.L. Winkler. "Scoring Deviations of Stock Price Ratios Implied by Option Premia." Journal of Finance. 31(1976):369-81.

McDonald, J.B. "Some Generalized Functions for the Size Distribution of Income." Econometrica. 52(1984):647-63.

Morris, P.E. "Combining Expert Judgements: A Bayesian Approach." Management Science. 23(1977): 679-93.

Norris, P.E. & R.A. Kramer. "The Elicitation of Subjective Probabilities with Applications in Agriculatural Economics: A Survey of the Literature." Agricultural Economics Report #56. Department of Agricultural Economics, VPI and State University. 1986.

Pope, R.D. & R.F. Ziemer. "Stochastic Efficiency, Normality, and Sampling Errors in Agricultural Risk Analysis." American Journal of Agricultural Economics. 66(1984):31-40.

Quesenberry, C.P. & F.L. Miller, Jr. "Power Studies of Some Tests for Uniformity." Journal of Statistical Computing and Simulation. 5(1977):169-91.

Ramaswamy, K. & S.M. Sundaresan. "The Valuation of Options on Futures Contracts." Journal of Finance. 40(1984):1319-40.

Rodriguez, R.N. "A Guide to the Burr Type XII Distributions." Biometrika. 64(1977):129-34.

Ross, S.A. "A Simple Approach to the Valuation of Risky Streams." Journal of Business. 51(1978):453-75.

Savage, L.J. "Elicitation of Personal Probabilities and Expectations." Journal of the American Statistical Association. 66(1971):783-801.

Schmalensee, R. & R.R. Trippi. "Common Stock Volatility Expectations Implied by Option Premia." Journal of Finance. 33(1978):129-47.

Singh, S.K. & G.S. Maddala. "A Function for the Size Distribution of Incomes." Econometrica. 44(1976):963-70

Spetzler, C.A. and C.A.S. Stael von Holstein. "Probability Encoding in Decision Analysis," Management Science. 22(1975):340-58.

Stephens, M.A. "EDF Statistics for Goodness of Fit and Some Comparisons." Journal of the American Statistical Association. 69(1974):730-37.

Tadikamalla, P.R. "A Look at the Burr and Related Distributions." International Statistical Review. 48(1980):337-44.

Winkler, R.L. "On 'Good Probability Appraisers'", chap. 17 of Goel, P. & A. Zellner, Bayesian Inference and Decision Techniques. Elsevier Science Publishers, 1985.

Young, D.L. "Evaluating Procedures for Computing Objective Risk from Historical Time Series." in Risk Analysis in Agriculture: Bhavioral, Managerial, and Policy Issues. Illinois Agricultural Experiment Station Bulletin AE-4478.