Calibration of Option-Based Probability Assessments in Agricultural Commodity Markets
Author(s): Paul L. Fackler and Robert P. King
Source: *American Journal of Agricultural Economics*, Vol. 72, No. 1 (Feb., 1990), pp. 73-83
Published by: Oxford University Press on behalf of the Agricultural & Applied Economics Association
Stable URL: https://www.jstor.org/stable/1243146

Calibration of Option-Based Probability Assessments in Agricultural Commodity Markets

Paul L. Fackler and Robert P. King

A method for evaluating the reliability of option-based price probability assessments is developed based on the calibration concept. Empirical tests using goodness-of-fit criteria are applied to four agricultural commodities. Results suggest that assessments in the corn and live cattle markets are reliable, but such assessments overstate the volatility of soybean prices and understate the location of hog prices.

Key words: calibration, futures, goodness-of-fit, options, probability assessment.

The initiation of trading on agricultural commodity option futures in 1984 created new risk management opportunities for market participants. This expansion in risk management alternatives directly benefits only those who trade options. As Gardner notes, however, commodity option markets can also provide useful public information about price probability distributions. Probability assessments derived from data on option premiums require only a limited amount of readily available information and can be updated easily.

For decision makers, the performance of option-based price probability assessments is an important issue. This paper focuses on one aspect of their performance: calibration or reliability (DeGroot and Fienberg; Winkler; Lichtenstein, Fischoff, and Phillips). A probability assessment method is well calibrated "if, over the long run, for all propositions assigned a given probability, the proportion that is true equals the probability assigned" (Lichtenstein, Fischoff, and Phillips, p. 307).

Calibration is important for two reasons. First, well-calibrated probability assessment methods generate reliable probability statements. Second, well-calibrated option-based probability assessments are consistent with the absence of risk premiums in options markets. Therefore, tests for calibration provide useful information about the functioning of option markets.

In the sections to follow, a theory of option pricing based on the absence of arbitrage opportunities is outlined and results from that theory are used to show how probability assessments can be derived from option premiums. The concept of calibration is then defined and statistical methods for testing the calibration of option-based probability assessments are described and used to evaluate the performance of price probability assessments for corn, soybeans, live cattle, and hogs. Implications of the findings and needs for further research are found in the concluding section.

Option Pricing Theory

Option contracts guarantee their holder the right to buy or sell a specified asset (the underlying asset) at a given price (the exercise price) on or before a given date (the expiration date). Put and call options provide the right to sell and to buy, respectively. European options allow exercise of the option only on the expiration date, while American options allow exercise any time on or before that date. An option premium is the price paid for the option contract; typically, it is payable immediately and in full.

A general theory of option pricing, based on Cox and Ross, relies on the minimal assumption that no arbitrage opportunities exist. They show that this condition is equivalent to the existence of an artificial probability distribution such that the asset price equals the stream of expected re-

Paul L. Fackler and Robert P. King are, respectively, an assistant professor in the Department of Economics and Business at North Carolina State University, and an associate professor in the Department of Agricultural and Applied Economics at the University of Minnesota.

This is Journal Article No. 16,692 of the Minnesota Agricultural Experiment Station.

turns discounted at the risk-free rate. The idea of an artificial distribution is widely applied in the finance literature. It is called the risk-neutral valuation measure (RNVM) in this paper.

To apply this notion to the pricing of options, note that the return to a European put option with exercise price x expiring at time T (the expiration date) is $\max(0, x - Y_T)$, where Y_T is the (random) price of the underlying asset at time T. If arbitrage opportunities are absent, the current price of the option, $V_p(x)$, can be written as

$$(1) \quad V_p(x) = b(T) \int_0^{\infty} \max(0, x - Y_T)\, dG(Y_t),$$

where $b(T)$ is the current price of a risk-free pure discount bond paying \$1 at time T, and G is the RNVM discussed above. This expression can be rearranged as

$$(2) \qquad V_p(x) = b(T)\{xG(x) - E_G[Y_T; x]\},$$

where

$$E_G[Y_T; x] = \int_0^{x} Y_T\, dG(Y_T)$$

is the incomplete expectation of Y_T with respect to $G(Y_T)$.

Similarly, the value of a European call option can be written as

$$\begin{aligned}(3) \quad V_c(x) &= b(T) \int_0^{\infty} \max(0, Y_T - x)\, dG(Y_T), \\ &= b(T)\{E_G[Y_T] - E_G[Y_T; x] \\ &\quad - x(1 - G(x))\}, \\ &= V_p(x) + b(T)\{E_G[Y_T] - x\}.\end{aligned}$$

This approach can also be used to show that the current value of an asset paying $Y_T/b(T)$ at time T is equal to y_0, the current futures price (Cox, Ingersoll, and Ross). Therefore,

$$\begin{aligned}(4) \qquad y_0 &= b(T) \int_0^{\infty} Y_T/b(T)\, dG(Y_T) \\ &= E_G[Y_T].\end{aligned}$$

Combining (3) and (4) yields the familiar put-call parity relationship for options on futures:

$$(5) \qquad V_c(x) - V_p(x) = b(T)(y_0 - x).$$

Given $G(Y_T)$, the value of an option can be defined. For example, if $G(Y_T)$ is log-normal with mean μ and standard deviation σ, then[1]

$$\begin{aligned}(6) \qquad V_p(x) = b(T)\{x\Phi((z + \sigma/2) \\ - y_0\Phi((z - \sigma/2)\},\end{aligned}$$

where $z = \ln(x/y_0)/\sigma$, and, given the symmetry of the normal distribution,

$$\begin{aligned}(7) \qquad V_c(x) = b(T)\{y_0\Phi((z + \sigma/2) \\ - x\Phi((z - \sigma/2)\}.\end{aligned}$$

These formulas are equivalent to the familiar option pricing formulas derived by Black (p. 177) and Gardner (p. 989).[2]

Option-Based Probability Assessments

Knowledge of the RNVM, $G(Y_T)$, which need not be log-normal, and the price of a risk-free bond, $b(T)$, is sufficient to calculate the premiums for call or put options at any exercise price. Conversely, knowledge of option premiums [and $b(T)$] can be used to infer information about $G(Y_T)$ (Breeden and Litzenberger). Suppose that $G(Y_T)$ can be reasonably well approximated by a parametric family of functions. The particular member of this family can be selected by finding the parameter values that yield predicted option premiums closest to observed market premiums.

In this study $G(Y_T)$ is assumed log-normal to simplify the analysis and facilitate comparison with other studies. A preliminary analysis using a more flexible distribution, the Burr-12 or Singh-Madalla (Singh and Madalla, McDonald), did not greatly improve the fit between predicted and actual option premiums. Under log-normality, the option valuation formulas in equations (6) and (7) have only one unobserved parameter, σ. Given an observed option premium, a value of σ can be found that yields a predicted premium exactly equal to the observed value.[3] The re-

[1] These formulas make use of the following facts (Hogg and Klugman, p. 229):

$$G(x) = \Phi((\ln(x) - \mu)/\sigma),$$
$$E_G[Y_T] = \exp(\mu + \sigma/2), \text{ and}$$
$$E_G[Y_T; x] = E_G[Y_T]\Phi((\ln(x) - \mu)/\sigma - \sigma),$$

where Φ is the standard normal CDF.

[2] A superficial difference between the formulas involves the $sqrt(T)$ term in Black's formulation. This difference is explained by the fact that the variance term, σ^2, used in this paper refers to the variance parameter associated with the expiration date price, which is T times the daily variance used by Black.

[3] Computationally, this involves a numerical search algorithm because no analytical solution exists. The bisection method was used in this study because it is easy to implement and reasonably fast. The evaluation of the option values also requires the numerical evaluation of the standard normal CDF. See Kennedy and Gentle (pp. 90–93) for a discussion of numerical techniques for this problem.

sulting value of σ is commonly termed the implied volatility. It is uniquely determined by the option premium, given its exercise price, the current futures price, the price of an appropriate risk-free bond, and the time until expiration, and has been used in a number of studies (e.g., Latané and Rendleman, Chiras and Manaster, Schmalensee and Trippi, Beckers).

Options typically are traded at a number of different exercise prices. Several implied volatility estimates, therefore, can be calculated for any given trading period. Options with exercise prices that are not close to the asset's current price tend to be thinly traded; thus, they may nor accurately reflect equilibrium relationships. The just-out-of-the-money option is usually the most heavily traded and is used in this study. Both the just-out put and just-out call were used to calculate implied volatilities, and the two volatility estimates were then averaged.

Given an estimate of σ the implied value of μ is given by

$$\mu = \ln(y_0) - \sigma^2/2. \tag{8}$$

Parameters μ and σ provide a complete description of the RNVM, $G(Y_T)$.

The example in table 1 illustrates the process for generating probability assessments. The data come from the *Wall Street Journal* and include the current and expiration date futures prices, y_0 and Y_T, the just-out put and call premiums and their strike prices and the bid and ask discount rates on the appropriate T-bill. In this example, there are eight weeks (56 days) remaining until expiration. The bid and ask bond rates are combined to obtain an average rate, r, from which the bond price is calculated using $b = 1 - r*56/360$. The implied volatilities using the put and call contracts are calculated numerically as the solution to (6) and (7). The average of these volatilities yields the implied volatility measure, σ. The "implied mean," μ, is calculated from (8).

Table 1. An Example of the Data and Calculation of the Log-Normal *RNVM*

Contract	November 1985 Soybeans
Current date	16 August 1985
Expiration date	11 October 1985
Current futures (y_0)	513.5
500 put	8.625
525 call	9.25
T-bill bid rate	7.09
T-bill ask rate	7.05
Average T-bill rate	7.07
Bond price (b)	0.989
Implied volatility from put	0.07165
Implied volatility from call	0.06942
Average implied volatility (σ)	0.07053
Implied mean (μ)	6.23876

The relationship between the artificial probability distributions generated by this process and the actual stochastic behavior of associated expiration date futures prices is an important empirical stochastic issue. If the G's reliably represent the stochasitc behavior of the associated Y_T's (i.e., if the G's are well calibrated), option-based probability assessments are a reliable, readily available source of probabilistic information. Furthermore, if the G's are well calibrated, equations (2) and (3) imply that option premiums equal the present value of the expected returns and are, therefore, an actuarially fair form of price insurance.

Generally, little can be said theoretically about the relationship between G and the stochastic behavior of Y_T. However, poor calibration of G can be identified for several reasons. First, if option writers require risk premiums because they cannot diversify away the risk of an options position, then the value of the option would be greater than under risk-neutral conditions. If G is log-normal, this condition would result in upward bias in the implied volatility. Second, risk premiums may be required by speculators in the underlying futures (see Kamara for a review). For example, if normal backwardation exists ($y_0 < E[Y_T]$), the location of G would be biased downward. Third, G may be unreliable because probability assessments of market participants are unreliable. None of these factors affect the validity of the RNVM approach to pricing options; instead, they affect the relationship between the RNVM and an objective stochastic measure of price behavior.

Poor calibration may also reflect problems with the construction of the RNVM rather than the nature of the market. For example, equations (6) and (7) strictly apply only to European options, while the options traded in the United States are American options. Failure to consider that American options are more valuable than their European counterparts (Ramaswamy and Sundaresan) could cause an upward bias in the implied volatility. However, after calculating values for both American- and European-type options, Plato concluded (p. 9) that the difference between the two for near-the-money option values is negligible. Methodological problems may also be caused by nonsynchronous option and futures prices and prices that represent nontrades, though this problem is minimized by the use of heavily traded options.

This study does not attempt to explain lack of calibration in option-based probability assessments. Rather, the goals are to examine whether such assessments are well calibrated, to describe any calibration problems, and to suggest how the assessments can be improved. The derivation of these probability forecasts is purposely simplified by not considering all possible sources of poor calibration.

The Concept of Calibration

Approximations of the RNVF derived from option premiums will not necessarily generate reliable probability statements for the expiration date futures price. An empirical examination of this issue is based on the concepts of calibration and the calibration function (Bunn, chap. 8).

To make the calibration concept more explicit, let G_i for $i = 1, \ldots, n$ be a set of probability assessments, expressed as CDFs, for a set of realizations of a random variable, Z. Given a large sample of independent assessments, G_i, and realized outcomes of Z, z_i, the probability assessment process generating the G_i is calibrated if the proportion of times $G_i(z_i)$ is less than or equal to any given value, u, on the interval [0, 1] is u. More formally, let U be a random variable on the interval [0, 1] with CDF $C(U)$, defined by $U_i = G_i(Z_i)$. The process used to generate the G_i is calibrated if U is uniformly distributed on the interval [0, 1], i.e., if $C(u) = u$ for all u on the interval [0, 1].

In this study, option-based probability assessments are the G_i. For any option-based assessment, the realized outcome, z_i, is the expiration date price of the underlying futures contract, Y_T. Evaluating the G_i at Y_T generates a CDF value, a u_i. This is illustrated by the example in table 1. The realized price associated with the assessment (the expiration date futures price) is 502. The associated CDF value, the u_i, is calculated as

$$(9) \quad \begin{aligned} u_i &= \Phi((\ln(Y_T) - \mu/\sigma) \\ &= \Phi((\ln(502) - 6.23876)/0.07053) \\ &= 0.3875. \end{aligned}$$

If the option-based assessment process is well calibrated, the u_i calculated in this manner from a sample of independent assessments will be uniformly distributed.

The function C, the CDF of U, is called a calibration function (Bunn; for alternate terminology see Curtis, Ferrell, and Solomon; Morris). The calibration function provides a means of transforming a noncalibrated assessment into a calibrated one. The distribution defined by the transformation $F_i(Z) = C(G_i(Z))$ will always be a calibrated distribution because, for any Z,

$$(10) \quad \begin{aligned} C(G_i(z_i)) &= \text{Prob}(G_i(Z_i) \leq G_i(z_i)), \\ &= \text{Prob}(Z_i \leq z_i). \end{aligned}$$

When the G_i are calibrated, $F_i = G_i$. When the G_i are not calibrated, the calibration function can provide insights into the nature of the problems in assessment process and their resolution.

In the example in table 1 G_i is log-normal with parameters $(\mu, \sigma) = (6.239, 0.0705)$. Suppose, however, that the options-based volatility assessment is inflated because options writers require a substantial risk premium. More specifically, suppose that dividing the volatility parameter, σ, by 2 results in calibrated assessments, and therefore F_i is log-normal with parameters $(\mu, \sigma/2)$. The calibration function associated with this case is illustrated in the upper left-hand panel of figure 1. Because the assessed distributions tend to overstate variability, U_i is more likely to fall in the center of the distribution than in its tails, resulting in an S-shaped calibration function. The original G_i and the calibrated F_i are illustrated in the upper right-hand panel of figure 1. The function C serves to calibrate G_i by shifting probability weight from the tails to the center of the distribution, thus yielding a calibrated assessment with lower variance than the original assessment.

In a second example, shown in the lower panel of figure 1, the G_i's tend to understate the mean parameter, μ, by 0.01. Such a situation might reflect normal backwardation in the underlying futures. U_i is more likely to fall above 0.5 than below it, resulting in the "J-shaped" calibration function shown in the lower left-hand panel. In this case the calibration function calibrates G_i by shifting probability weight into the right-hand tail.

These examples illustrate how the calibration function is associated with specific kinds of probability assessment problems. A straight (45°) line indicates that no such problems exist. In general, flat spots in the calibration function (parts with slope less than one) are associated with areas of the assessed distribution that receive too much weight, while steep parts are associated with areas that receive too little weight. Overassessment of the volatility, for example, places too much weight in the tail areas of the assessed distribution. The calibration function must, there-

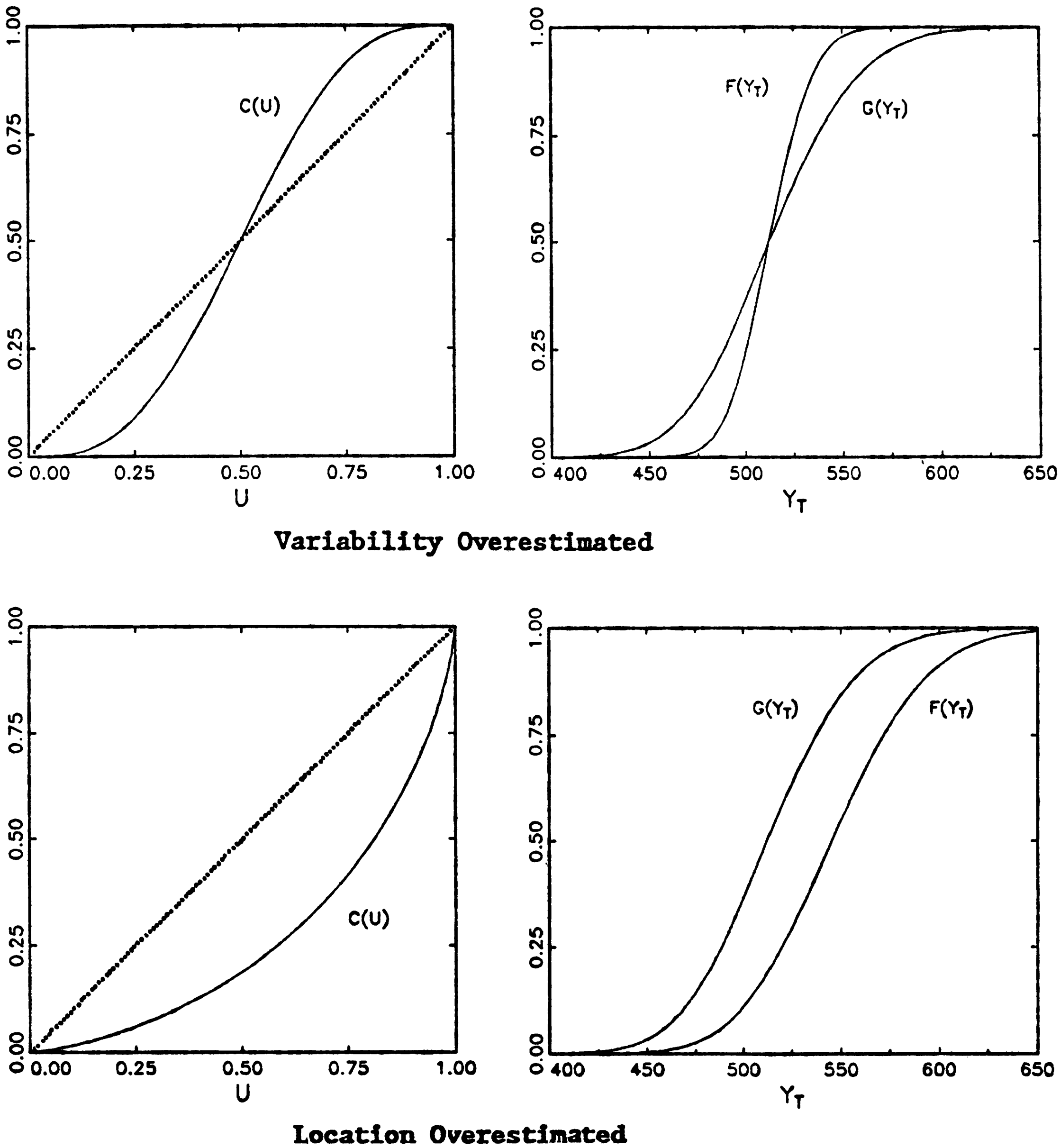

Figure 1. Examples of calibration functions: $F(Y_T) = C(G(Y_T)) = C(U)$

fore, be flat in the tails and steep in the center (S-shaped). If the assessed location is too far to the left, the calibration function must put more weight in the right tail and therefore will exhibit a flat-steep (J-shaped) pattern. Steep-flat-steep and steep-flat patterns are indicative of under-assessment of volatility and overassessment of location, respectively.

The value of the calibration function at 0.5 is a useful indictor of bias in location (median). If $C(0.5) < 0.5$ (> 0.5), more weight is in the right (left) half of the calibration function; hence, the assessed distribution is located too far to the left (right). A comparable indicator of problems in assessing dispersion is the range of the calibration function over the middle half of its domain (the interquartile range): $IQR = C(0.75) - C(0.25)$. An $IQR > 0.5$ (< 0.5) indicates that too few (many) observations fall in the tails of the assessed distribution.

Estimation of Calibration Functions

The calibration function associated with an assessment method can be estimated statistically given a random sample of values of U. A single observation on U provides no information on the reliability of an assessment process; a process can be evaluated only when patterns emerge in a sample of assessments. Furthermore, the U_i used to estimate a calibration function should be independent random variables. New option-based probability assessments can be constructed almost continuously; however, they will not necessarily yield independent U_i values because all assessments for a particular option contract share the same realized price. A random sample can be constructed by considering only one option-based assessment per futures contract and by ensuring that the time periods between trading and expiration dates do not overlap for any sample assessments.

The simplest estimator of C is the empirical CDF of a set of U_i values. After sorting the n-values of U_i in ascending order, the empirical CDF is obtained by setting $C(U_i) = i/n$. This procedure makes no assumption about the form of C. Alternatively, a parametric representation of C can be obtained by choosing a suitable functional form for C. The parametric approach offers both parsimony and ease of interpretation. Parametric forms can be fully described by a few parameters, while the empirical CDF requires the whole sample. Furthermore, if the parametric form is well chosen, specific values of the parameters will be associated with likely calibration problems in the probability assessment process. The principal disadvantage of a parametric approach is that the form chosen may not represent the calibration function well in particular cases.

The beta distribution, defined on the interval [0, 1], is a natural candidate for representing the calibration function. Its probability density function (PDF), $c(u)$, is

$$(16) \qquad c(u) = u^{p-1}(1 - u)^{q-1}/B(p, q),$$

where B is the beta function (with the associated CDF denoted $B(u; p, q)$). The beta distribution is well known, flexible, and contains the uniform distribution as a special case ($p = q = 1$). The two parameters of this distribution can be easily estimated by applying maximum likelihood methods to the sample values of U_i.

The beta distribution can assume any of the four basic patterns of calibration functions discussed above (in addition to $p = q = 1$): flat-steep-flat ($p, q > 1$), steep-flat-steep ($p, q < 1$), flat-steep ($p > 1, q < 1$) and steep-flat ($p < 1, q > 1$). The condition associated with no location bias, i.e., $C(0.5) = 0.5$, occurs when $p = q$. No simple comparable condition exists for bias in dispersion. However, if p and q are nearly equal, indicating little location bias, values of p and q greater than one indicate overassessment of dispersion, while values of p and q less than one indicate underassessment.

It is possible for the calibration of an assessment process to change over time or to be dependent on conditioning variables. For example, if option premiums incorporate a risk premium which increases with the volatility of the price of the underlying asset, then the divergence between the RNVM and an objective probability measure would also increase with volatility. Furthermore, option-based probability assessments implicitly depend on the assessments made by options traders, who can learn if their assessments are unreliable. This might result in option-based assessments becoming more calibrated as an option market matures. Unfortunately, the recent opening of agricultural option markets limits available data; therefore, a simple unconditional calibration function is estimated for each sample examined in this study.

Tests of Calibration

An examination of whether observed departures from calibration are statistically significant can be based on goodness-of-fit tests for uniformity on [0, 1]. Three approaches to testing for uniformity are used in this study, with the main focus on detecting potential problems in the assessment of location and dispersion.

The first test is based on the percentage of the u_i values lying within an interval of length 0.5 on the interval [0, 1]. Under the null hypothesis, the chance that a realization of U will lie in any such range is 50%. The interval [0, 0.5] provides information on bias in location (median), while the interval [0.25, 0.75] provides information on bias in the dispersion (interquartile range). This test is a variant of the sign test, which rejects the null hypothesis if the percentage falling into such an interval is either too large or too small relative to a binomial distribution in which the probability of a "success" is 0.5.

Nonparametric goodness-of-fit tests based on the empirical CDF comprise the second approach. Stephens discusses and evaluates a number of such tests, including the familiar Komolgorov test. Each of the tests varies in relative power against alternative hypotheses. The Watson U^2 and the Cramer-von Mises tests appear to be relatively powerful against alternatives associated with nonuniform dispersion and location, respectively. Further power tests by Quesenberry and Miller support this finding and indicate that the Watson test is a good candidate for a single omnibus test of uniformity.

A third approach uses the maximum likelihood parameter estimates of the fitted beta calibration function to construct a likelihood ratio statistic for the hypothesis that $p = q = 1$, the parameter values for the uniform special case. Because the loglikelihood of the uniform distribution is always equal to zero, the likelihood ratio statistic is twice the maximum likelihood. This test statistic has an asymptotic chi-square distribution with two degrees of freedom. Monte Carlo simulation results performed by the authors suggest that the test statistic should be multiplied by a small sample correction factor of $(1 + 1/n)$, where n is the sample size. Further simulation results indicate that this test statistic compares favorably in power to the nonparametric tests examined by Stephens.

Testing the Calibration of Option-Based Probability Assessments in Agricultural Markets

Option trading for soybeans and cattle began in October 1984, while trading for corn and hogs began in February 1985. The options are written on futures contracts. For cattle and hogs, those futures contracts expire in February, April, June, August, October, and December. For soybeans, they expire in January, March, May, July, September, and November; for corn they expire in March, May, July, September, and December.[4] For all of these commodity options, trading ends on a Friday one to two business weeks prior to the beginning of the futures contract delivery month. The expiration date futures price is taken to be the close price on the last trading date of the options. Samples of size 15, 20, 20, and 18 (July 1985–May 1988, March 1985–May 1988, Feb. 1985–Apr. 1988, and June 1985–Apr. 1988) were used for corn, soybeans, cattle, and hogs, respectively. Futures prices and option premiums for corn and soybeans prior to 1986 are from Chicago Board of Trade tapes. All other data are from the *Wall Street Journal*.

Probability assessments for each commodity were generated on dates eight and four weeks prior to expiration. These assessments were used to construct two samples (per commodity) of realized CDF values, i.e., of U_i. The CDF values within each sample can be viewed as independent random variables because the periods between assessment and expiration dates do not overlap. Thus, each sample can be used to test for calibration. The two samples for each commodity are not independent because the period between assessment and expiration dates of the four-week assessment is always encompassed by that of the eight-week assessment. Thus, it is inappropriate to pool the samples. However, presenting the results for both samples does indicate possible changes in the calibration of option-based assessments as the forecast period changes.

CDF values associated with the expiration date futures price for each of the eight samples were used to construct the empirical and fitted beta calibration functions shown in figure 2.[5] The calibration functions estimated by the two methods have similar shapes, supporting the use of the beta distribution as a parametric form for representing calibration functions. The main exceptions are in the eight-week samples for corn and cattle, both of which have empirical, but not beta, calibration functions that exhibit flat areas in the center portion of the [0, 1] interval. One should suspect therefore that, in these cases, test results based on the beta distribution may be in conflict with those based on nonparametric methods.

The shapes of the calibration functions vary considerably across these markets. For corn and cattle the calibration functions are difficult to characterize and are not consistent between the eight- and four-week samples. The eight-week sample for corn has a steep-flat-steep pattern, while the four-week sample has a flat-steep-flat pattern. These patterns are roughly associated with problems in the under- and over-assessment of dispersion. The eight-week cattle sample has a steep-flat-steep pattern, while the four-week sample has a flat-steep pattern. For soybeans, however, both samples clearly exhibit a flat-steep-

[4] August soybean and July hog contracts are also traded. However, these contracts were not used so that two-month periods between expiration dates could be maintained, thereby eliminating the possibility of overlapping forecast periods. The expiration dates are spaced two or three months apart in the corn market.

[5] For ease of interpretation, the empirical calibration function is presented as a piecewise linear function that linearly interpolates the points $\{u_i, (i - 0.5)/n\}$, with u_i sorted and using (0, 0) and (1, 1) as endpoints.

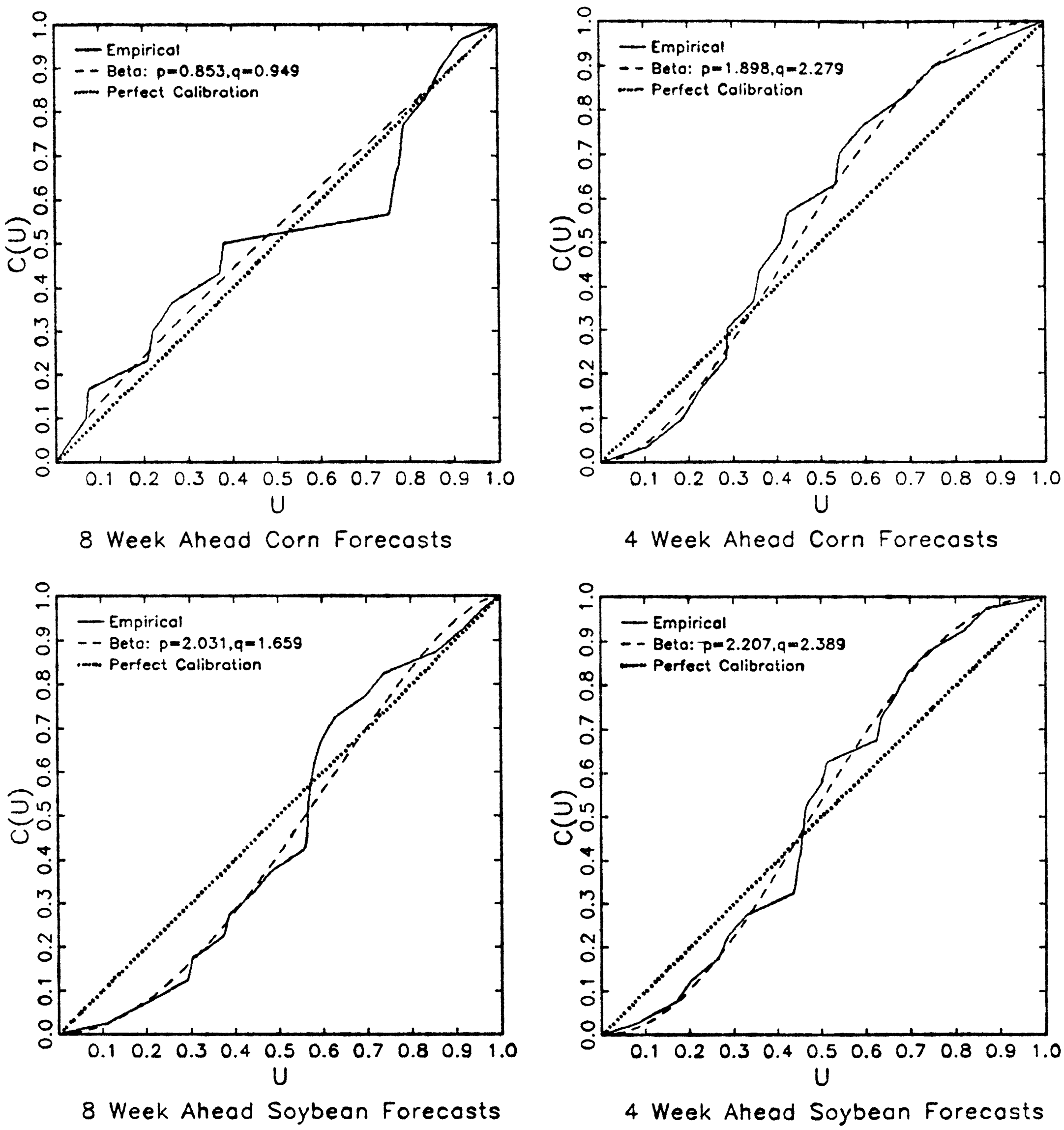

Figure 2. Empirical and beta calibration functions

flat pattern, while for hogs both exhibit a flat-steep pattern. These patterns suggest that probability forecasts in the soybean and hog markets exhibit overassessment of dispersion and underassessment of location, respectively.

Statistical results in table 2 largely confirm the graphical results. The sign test results show the percentage of observations in each sample falling into the [0, 0.5] and [0.25, 0.75] intervals, as well as the associated *p*-values for the two-sided test. Tests on the first of these intervals suggest problems in assessing location in the four-week hog sample, and, weakly, in the four-week cattle sample. The second of these intervals suggests problems in assessing dispersion in both soybean samples and both corn samples, although only weakly in the four-week corn case. For corn, however, the eight-week sample has too few observations in the [0.25, 0.75] interval, while the four-week sample has too many.

These results are generally consistent with those of the nonparametric Watson and Cramer-von Mises tests. No assessment problems are indicated for corn and cattle. On the other hand, the Watson test rejects the null hypothesis in both

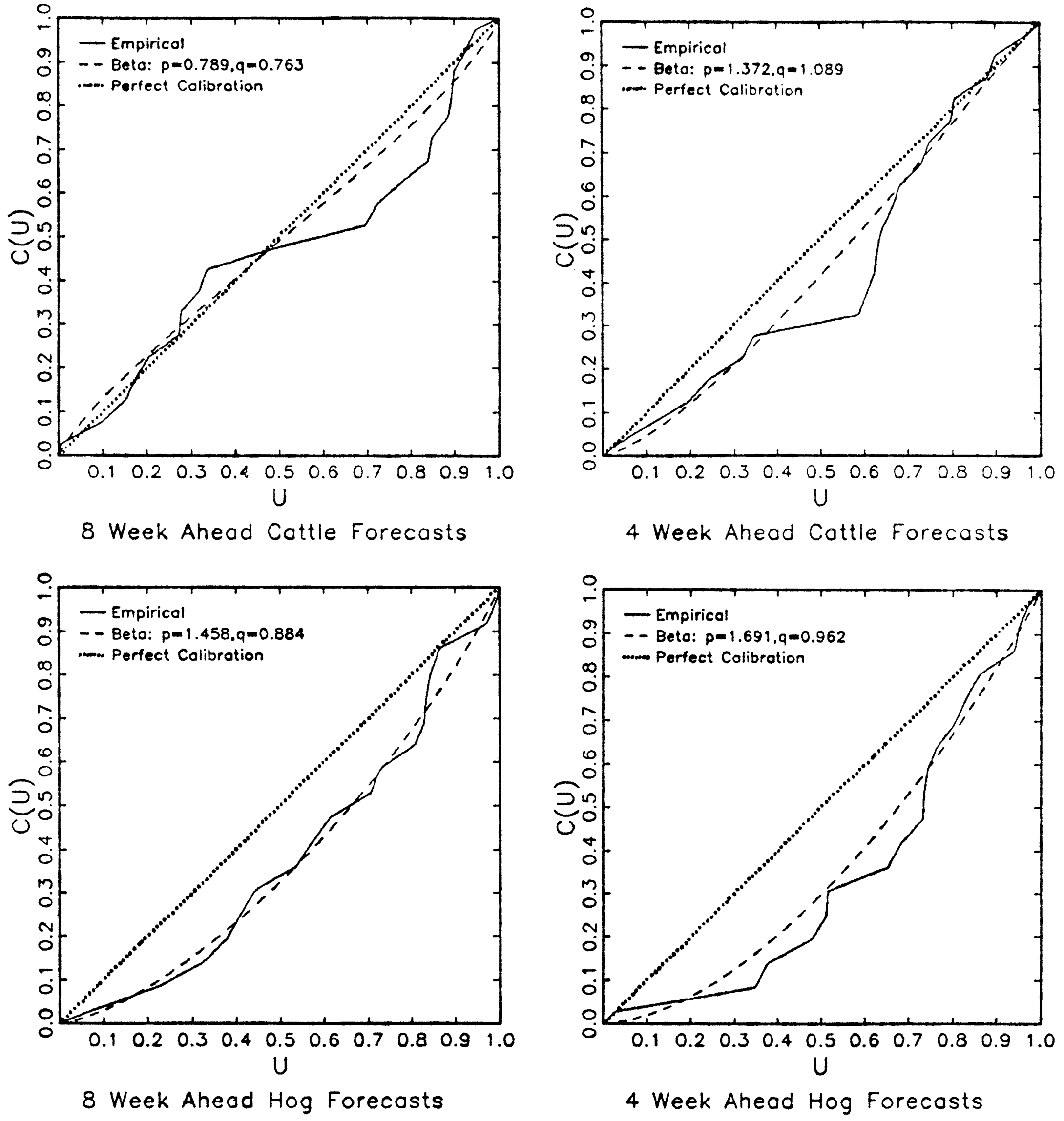

Figure 2. Continued

soybean samples. For hogs both tests reject the null hypothesis at the 0.05 significance level in the four-week case, while the null hypothesis is rejected at the 0.15 level by the Cramer-von Mises test in the eight-week sample.

The parametric (beta distribution) test results provide only weak evidence of noncalibration in the four-week sample in the corn market and none in the cattle market. Calibration problems are indicated in both soybean and hog markets, with the problems in both cases most evident in the four-week samples.

Taken together, these results suggest calibration problems in option-based probability forecasts for both the soybean and hog markets. For soybeans these problems involve overassessment of dispersion; for hogs they involve underassessment of location, although the evidence for this is not quite as strong.[6] The corn market results are somewhat ambiguous, but they do not

[6]An examination of hog futures prices in the period 1976–84 (prior to options trading) reveals no evidence that the probability of an up or down movement was other than 50% at either the 4- or 8-week horizon. The results for the 1985–88 period therefore reflect recent developments in this market or are attributable to sampling error.

Table 2. Calibration Test Results

Samples	Sign Tests (%) [0, 5]	Sign Tests (%) [.25, .75]	Cramer-von Mises W^2	Watson U^2	Beta LR
Corn					
8-Week	53.3 (.607)[a]	20.0 (.035)	0.97	.113	0.277 (.871)
4-Week	60.0 (.302)	66.7 (.118)	.148	.117	4.398 (.111)
Soybeans					
8-Week	40.0 (.503)	75.0 (.012)	.182	.157@	4.457 (.108)
4-Week	60.0 (.263)	75.0 (.012)	.147	.154@	6.868 (.032)
Live cattle					
8-Week	50.0 (.824)	35.0 (.263)	.144	.127	1.057 (.590)
4-Week	30.0 (.115)	55.0 (.503)	.237	.130	1.290 (.525)
Hogs					
8-Week	33.3 (.238)	50.0 (.815)	.320*	.065	3.585 (.167)
4-Week	22.2 (.031)	55.6 (.481)	.675#	.187#	4.664 (.097)

[a] p-values are shown in parentheses; asterisk indicates significant at the 0.15 level; @ indicates significant at the 0.10 level; # indicates significant at the 0.05 level.

suggest that option-based probability forecasts exhibit a consistent noncalibration pattern. Instead, a change from under- to overassessment of dispersion is indicated when the forecast horizon is changed from eight to four weeks. Without further study, this apparent calibration problem is perhaps best attributed to sampling fluctuation.

For those markets with calibration problems, the option-based assessments can be transformed using the estimated calibration function (either the nonparametric or the beta). To illustrate, consider the probability distribution associated with the July 1988 soybean futures at the expiration of the associated option (17 June 1988). On 22 April 1988, the log-normal option-based assessment was $G(Y_T) = \Phi((\ln(Y_T) - 6.508)/0.08174)$. However, given the relevant calibration function (8-week soybean) shown in figure 1, an improved assessment could be obtained by using $F(Y_T) = B(G(Y_T), 2.031, 1.659)$.

Table 3 contains a comparison of selected percentiles of these two distributions. The calibrated distribution has shifted the median up by nearly 8.5¢ and reduced the interquartile range by nearly 30% from 73.0¢ to 52.5¢. Because of unanticipated drought conditions, the realized expiration date price of 971.5 was in the extreme right-hand tail of both distributions.

Table 3. A Comparison of Selected Fractiles for a Noncalibrated and a Calibrated Probability Assessment

U	Noncalibrated ($G^{-1}(U)$)	Calibrated ($F^{-1}(U)$)
0.05	586.16	618.21
0.10	603.82	631.11
0.25	634.54	653.26
0.50	670.51	678.89
0.75	708.51	705.74
0.90	744.55	731.12
0.95	767.00	746.93

Concluding Comments

This paper focuses on methods for assessing and evaluating price probability distributions based on options market data. The recent development of agricultural option markets makes this a timely topic, but it also allows for only a preliminary evaluation of these methods. Nonetheless, the applications in the four commodity markets demonstrate the potential of the methods. Option premiums can be used to provide information on price probability distributions, yielding a complete assessment of the stochastic nature of prices that is easily obtainable at low cost and that can be regularly updated.

The results of this paper suggest that there are differences among markets in the reliability of option-based probability assessments. Little evidence was found of any systematic assessment problems in the corn and live cattle markets. However, the results suggest that option-based assessments of price distributions overstate the variability of soybean prices and underpredict the location of hog prices.

The results presented here are preliminary, pending maturity of the options markets under study. However, they present both a framework for the study of the reliability of probability assessments and demonstrate the ability of the statistical tests used to both support and reject the null hypothesis of calibration. Further research could usefully explore the causes of apparent noncalibration, particularly in the soybean market. It would also be useful to compare the reliability of option-based assessments with other assessment methods. Clearly, options markets provide a rich data base for further study.

[Received January 1986; final revision received April 1989.]

References

Beckers, S. "Standard Deviations Implied in Option Prices as Predictors of Future Stock Price Variability." *J. Bank. and Finan.* 5(1981):363–81.

Black, F. "The Pricing of Commodity Contracts." *J. Finan. Econ.* 3(1976):167–79.

Breeden, D. T., and R. H. Litzenberger. "Prices of State-contingent Claims Implicit in Option Prices." *J. Bus.* 51(1978):621–51.

Bunn, D. W. *Applied Decision Analysis.* New York: McGraw-Hill Book Co., 1984.

Chiras, D. P., and S. Manaster. "The Informational Content of Option Prices and a Test of Market Efficiency." *J. Finan. Econ.* 6(1978):213–34.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. "The Relationship between Forward Prices and Futures Prices." *J. Finan. Econ.* 9(1981):321–46.

Cox, J. C., and S. A. Ross. "A Survey of Some New Results in Financial Option Pricing." *J. Finan.* 31(1976B):383–402.

Curtis, P. E., W. R. Ferrell, and I. Solomon. "A Model of Calibration for Subjective Probability Distributions and Its Application to Aggregation of Judgments." *J. Forecast.* 4(1985):349–61.

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

Gardner, B. L. "Commodity Options for Agriculture." *Amer. J. Agr. Econ.* 59(1977):986–92.

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

Kamara, A. "The Behavior of Futures Prices: A Review of Theory and Evidence." *Finan. Analysts J.*, July–Aug. 1984, pp. 68–75.

Kennedy, W. J., Jr., and J. E. Gentle. *Statistical Computing*, vol. 33. New York and Basel: Marcel Dekker, 1980.

Latané, H., and R. Rendleman. "Standard Deviations of Stock Price Ratios Implied by Option Premia." *J. Finan.* 31(1976):369–81.

Lichtenstein, S., B. Fischoff, and L. D. Phillips. "Calibration of Probabilities: The State of the Art in 1980." *Judgement Under Uncertainty: Heuristics and Biases*, ed. D. Kahneman, P. Slovic, and A. Tversky, pp. 306–34. Cambridge: Cambridge University Press, 1982.

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." *Manage. Sci.* 23(1977):679–93.

Plato, Gerald. "Valuing American Options on Commodity Futures Contracts." *Agr. Econ. Res.* 37(1985):1–14.

Quesenberry, C. P., and F. L. Miller, Jr. "Power Studies of Some Tests for Uniformity." *J. Statist. Computing and Simulation* 5(1977):169–91.

Ramaswamy, K., and S. M. Sundaresan. "The Valuation of Options on Futures Contracts." *J. Finan.* 40(1984):1319–40.

Schmalensee, R., and R. R. Trippi. "Common Stock Volatility Expectations Implied by Option Premia." *J. Finan.* 33(1978):129–47.

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

Stephens, M. A. "EDF Statistics for Goodness of Fit and Some Comparisons." *J. Amer. Statist. Assoc.* 69(1974):730–37.

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