Expected Soybean Futures Price Distributions: Option-Based Assessments

Bruce J. Sherrick*
Assistant Professor of Agricultural Economics
University of Illinois at Urbana-Champaign

Scott H. Irwin
Associate Professor of Agricultural Economics
The Ohio State University

D. Lynn Forster
Professor of Agricultural Economics
The Ohio State University

*The authors are grateful to Anne Peck and an anonymous reviewer for helpful comments on an earlier draft and to Paul Fackler for generously providing a source code that greatly simplified our programming. The usual disclaimers apply.
Abstract

No-arbitrage option pricing models are used to derive ex ante expected price distributions. The performance of the method is assessed in the context of the calibration of the derived probability density functions, evaluated at the expiration date prices. It is found that the option-based probability assessments of soybean futures prices display some evidence of miscalibration very near to expiration and far from expiration. Endogenous switchpoint regressions indicate that volatility measures exhibit regime-dependent influences of time-to-maturity and price level.

Introduction

Predicted or expected values of economic variables serve as the primary inputs in business planning and decision-making. Often, these variables are in the form of a mean or average with an interval of possible error. Decision-makers, particularly in a risk-management context, behave as though they consider an entire probability-weighted distribution of future events. Hence, the applicability of mean forecasts should be questioned. More useful would be a characterization of the entire expected distribution of a future event.\(^1\) The range of possible outcomes, along with their associated probabilities, would be highly valuable in a decision context to producers, processors, speculators, and other market participants. Improvements in decision-making methods and risk-management techniques would be facilitated with a means of accurately describing distributions of events rather than of simply providing point forecasts.

In deriving estimates of values of ex ante variables, ex post data are often used. If, in fact, the distributions are nonstationary, use of ex post data may lead to seriously faulted conclusions. More useful would be a set of ex ante distribution parameters or an ex ante description of the stochastic process governing the realizations of the random variables. Unfortunately, direct elicitation of ex ante parameter expectations from market participants is often difficult, if not impossible.

One particularly important uncertain variable is the price of a commodity or security at some future time. The interaction of all participants in futures and options markets results in a collective expression of their beliefs about future or ex ante price distributions in current prices.\(^2\) This seemingly innocuous observation provides an avenue toward the recovery of the expected price distribution parameters without extensive surveys or direct elicitation. Specifically, options written on uncertain assets display valuable properties that may facilitate the examination of ex ante distributional parameters.

Options’ payoffs are contingent upon the possible outcomes of the underlying security’s price. The option price (premium) therefore implicitly contains the assessments and preferences of market participants over the

---

\(^1\) In the case of an expected utility maximization, wherein most financial optimizations may easily be couched, a Taylor series expansion of the utility function introduces all higher relevant moments in some number of terms.

\(^2\) The term future refers to a point in time yet to come, and the term futures refers to the standardized contract for exchange of a commodity in the future. Unless otherwise stated, the price distributions referenced pertain to futures prices. This convention allows us to avoid phrases containing the words future futures price distributions.
distribution of the underlying security's outcomes. In the case of soybean futures markets, current prices may give a good deal of information about expected future prices. For example, the current futures price may correspond to the mean of the expected future price distribution. However, in the absence of relatively sophisticated descriptions of the price diffusion process, and a risk premium if any, the futures price may be silent about other moments of an expected future price distribution. Options' payoffs, however, are contingent upon the entire range of possible outcomes for future price and, as such, contain an assessment of all relevant moments of a future price distribution. Thus, options may potentially be used to reveal information about future price distributions that may not be obtained from current futures prices. And, in developing good estimates of the probabilities of various future prices, it is unlikely that one will find any better estimate than the market's (Gardner [1977]).

Option pricing models and observed market prices often have been used in efforts to recover information about an underlying futures expected price distribution. The most prevalent examples of using options market prices to recover information about an underlying security's price distribution involve the Black-Scholes (B-S) [1973] or Black [1976] option pricing models and recovery of an estimate of the ex ante price volatility (standard deviation), often termed the implied volatility (IV). Then, if an estimate of the mean of the future price distribution is available as well, an entire two-parameter expected price distribution may be constructed for the underlying asset using information from observed option prices. Many studies have then sought to either explain observed biases in pricing related to implied measures from the model or, more directly, to explain changes in the implied distributions through time-series models of the implied volatilities. However, if the assumed pricing model is inaccurate, the derived estimates may be suspect. We use a much less restrictive no-arbitrage option pricing model to recover estimates of future price distributions and investigate the usefulness of the approach to reveal reliable information and to give insights into the changes in the expected price distributions over time.

Only one previous study has used the no-arbitrage option pricing method to derive complete probabilistic descriptions of uncertain future price distributions. Fackler and King [1990] use closing-price option data to examine the calibration of no-arbitrage implied price distributions for four agricultural commodities (including soybeans) and find evidence that the market implied distributions for soybeans overstated resulting variability. However, we expand the time period of analysis, use contemporaneous option and futures price data, and expand the scope of the inquiry in two other dimensions. First, we consider a three-parameter distribution that was first suggested by Fackler [1986] and Fackler and King [1990]. We then

3 For examples, see Latane and Rendleman [1976], Schmalensee and Trippi [1978], Chiras and Manaster [1978], Choi and Longstaff [1985], Park and Sears [1985], Anderson [1985], Milonas [1986], Shastri and Tandon [1987], Jordan et al. [1987], Beckers [1981], Ball and Torous [1986], and many others.
model the changes in the variance of the expected distribution in an endogenous switchpoint regression model.

Two specific issues regarding the "usefulness" of option premia data to assess expected future price distributions are thereby delineated. First, the goodness-of-fit of the implied distributions to the realized outcomes is assessed in terms of a calibration function. The calibration function gives a notion of the appropriateness of using the observed option premia in conjunction with the no-arbitrage valuation procedures to derive complete probabilistic descriptions of expected price distributions. Second, and perhaps more importantly, an examination of the changes of the parameters of expected distributions over time provides insights into the fundamental economic forces that may be reflected in these markets. If for no other reason than the passage of time, expected distributions will collapse to the expiration date price. Therefore, the equilibrium implied expected distributions will also potentially change as the agents' information sets change. The investigation of the equilibrium implied expected distributions would therefore be seriously suspect if, in addition to examining the static properties of the implied distributions, the issue of nonstationarity, or time-varying parameters, was not also considered. We employ an endogenous switchpoint regression model to detect the number and timings of switchpoints in each contract's series of implied distributions allowing for the possibility of varying dependence on time-to-maturity and price level.

The ex ante price distributions derived from option prices provide a unique source of information from which inferences regarding market forces and agents' actions may be assessed. Market agents learn and update their information sets with the passage of time and resolution of uncertainty about future events. Investigations of the forces that impact these distributions and an examination of how they change through time, provide fundamental insights into the workings of speculative markets. Inferences about the speed and types of adjustments in expectations and expectation formation may be gleaned from an examination of the changes in market implied future distributions. Further, the term structure of uncertainty is revealed through a comparison of various maturity options and by examining changes in a particular contract's implied distributions through time. Other studies have indicated that the variance of returns is not constant but varies through time. To the extent that these distributions could be used as an informational component in other decisions, seriously incorrect conclusions could be drawn if the issues of parameter nonstationarity are not first considered.

Unfortunately, the well-known Black-Scholes stock option model and the Black futures option pricing model derivations rely on a set of assumptions about the constancy of price volatility that makes term-structure of

---

4 As a simple analogy, suppose the variance of an expected price distribution for three months in the future were "k" and the variance of a price for expiration four months in the future were "5*k", then the time during which the uncertainty is greatest would be between the third and fourth months in the future.
uncertainty investigations inappropriate. The no-arbitrage method exploited herein suffers no such drawback and will in fact be used for a similar type of analysis.

The remainder of the paper is organized as follows. First, option pricing methods are discussed in light of the need to recover a complete probabilistic description of expected prices. Next, candidate parameterizations of expected distributions are given followed by a discussion of the concept of calibration of the estimates. The data are then described and the methods used and results are discussed. The paper then turns to modeling changes of the implied distributions in a switching regression framework and documents the variable influences of time-to-maturity and price level on volatility of the implied distributions. The paper then concludes with a summary and concluding remarks.

### Option Pricing: A No-Arbitrage Approach

A widely accepted basis for asset pricing is based on the set of no-arbitrage restrictions, first proposed by Ross [1978]. Absence of arbitrage is a necessary condition for market equilibrium, so the assumption that assets trade at equilibrium assures that there is no arbitrage. The widespread acceptance of no-arbitrage as a basis for asset pricing suggests that it may serve as a useful basis for option pricing as it does not suffer from many of the same restrictive features of the B-S formula.

Ross [1978] and others (Breeden and Litzenberger [1978], Banz and Miller [1978], Cox and Ross [1976]) show that no-arbitrage implies the existence of a supporting pricing function denoted as $f(s)$. The pricing function may be interpreted as a set of prices of pure contingent claims that pay one dollar if and only if their particular state, $s$, occurs. A one dollar risk-free bond may be constructed by buying one pure contingent claim for each possible state. Hence, the price of a one dollar bond is equal to the sum of the state prices for all possible states. If the states are essentially continuous, $f(s)$ corresponds to the state price density rather than a discrete probability function, but the arguments are otherwise analogous in that the price of a bond that pays one dollar at time $T$ regardless of what state occurs has a current price equal to $b(T) = \int f(s)ds$ where $r_s$ is the return in state $s$ (in this case equal to a constant of one dollar) and $f(s)$ is expected state price density at $T$. Given linearity of the pricing function across assets, any asset with return $r_s$ at time $T$ may be valued like the bond by simply taking the expectation of its returns with respect to $f(s)$. The bond is the simplest because its returns are one dollar in all states. Conversely, if the price of a risk-free bond was known and all possible states were identified, a consistent $f(s)$ could be located. Since the focus is on the use of the model, the discussion is somewhat limited. Excellent treatments of the complete set of no-arbitrage restrictions are given in Fackler [1986] and in Ingersoll [1987].

---

3 In fact, the implications of the Black-Scholes [1973] approach are sometimes in direct conflict with the no-arbitrage approach. For an interesting example, see Grinblatt and Johnson [1988].
Given the system of (1) asset prices, (2) states of the asset economy, and (3) distribution of state prices, any one of the three may be determined if the other two are known. The discussion above used payoffs and a supporting price distribution to determine no-arbitrage consistent prices, but observed asset prices and a payoff function could be used to solve for a set of state probabilities. Or, if the state probabilities are parameterized as a continuous function, the parameters of that function can be estimated. Thus, a payoff function $r_i$, and a distribution of time dependent state prices yield a current value consistent with no arbitrage. Any asset, $V_i$, with returns $r_{i,s}$, in the economy may therefore be priced if its return function is known via the relation:

$$V_i = \int_0^T r_{i,s} f(s) ds.$$  \hspace{1cm} (1)

Call and put options written on futures have clear return functions, $r_{i,a}$. For Call $C_i$, the return at time $T$ in the future is simply $\max(Y_T - K_i, 0)$ where $K_i$ is the exercise price and $Y_T$ is the random futures price at time $T$. For a put, $P_j$, the function is $\max(K_j - Y_T, 0)$. The outcome of $Y$ at $T$ completely determines the relevant state for an option's payoff. By noting that if $Y_T > K_j$, the value of the $j^\text{th}$ put option $(P_j)$ is zero, and that for $Y_T < K_i$, the value of the $i^\text{th}$ call option $(C_i)$ is zero, current values of puts and calls can be expressed as their discounted expected payoff:

$$C_i = b(T) \int_{K_i}^Y f(Y_T) dY_T,$$

$$P_j = b(T) \int_0^{K_j} f(Y_T) dY_T.$$  \hspace{1cm} (2)

The term $b(T)$, as defined earlier, is the price per dollar of a pure discount bond and serves as the discount function to place the option formulas into a current value.

A key distinction between this and the typical option pricing approach is that no assumptions have been made about the underlying price dynamics or changes in the economic environment prior to expiration. The only assumption made is that there are no arbitrage opportunities, thus guaranteeing the existence of $f(Y_T)$.

For obvious reasons, we wish to choose distributions to consider for $f(Y_T)$ that are as unrestrictive as possible. To be practical, a distribution is needed that does not allow negative prices (simple arbitrage), has relatively few parameters, and allows a fairly wide range of shapes to emerge for the CDF. Some studies suggest the lognormal distribution is not very descriptive of reality. Studies find empirical distributions that are more leptokurtic and more or less skewed than that implied by a lognormal distribution of prices.

---

* The relatively small percentage soybean futures options exercised prior to expiration will be taken as evidence that the difference in value between European and American options is small and can be ignored.

* For examples, see Gordon [1985] or Hall, Broersen, and Irwin [1989] and the references therein.
This study uses two distributions as the primary candidates for \( f(Y_T) \): (1) the two-parameter lognormal and (2) the three-parameter Burr-12 or Singh-Maddala (SM) [1976] distribution. Both distributions allow for only positive values of \( Y_T \), and the SM may take on a wide range of skewness and kurtosis (Tadikamalla [1980]). The SM cumulative distribution function (CDF) with parameters \( \alpha, \lambda, \) and \( \tau \) is

\[
F_{SM}(Y|\alpha, \lambda, \tau) = 1 - (\lambda/(\lambda + Y))^\alpha \text{ for } \alpha, \lambda, \tau, Y > 0, \text{ and}
\]

thus the density, or PDF, is

\[
f_{SM}(Y|\alpha, \lambda, \tau) = \alpha \lambda^{\alpha} \tau \gamma - 1 (\gamma + \lambda)^{-(\alpha + 1)}.
\]

The cumulative distribution function for the lognormal distribution with parameters \( \mu \) and \( \sigma \) is

\[
F_{LN}(Y|\mu, \sigma) = N(\ln(Y - \mu)/\sigma).
\]

where \( N(\bullet) \) is the cumulative normal density function. The lognormal density is

\[
f_{LN}(Y|\mu, \sigma) = (2\pi)^{-1/2}(\sigma Y)^{-1}\exp[-(\ln Y - \mu)^2/(2\sigma^2)].
\]

When substituted into the option pricing formulas in (2) and (3), we have the no-arbitrage option pricing formulas with which observed data are used to fit parameters of the two distributions. For example, using the SM distribution the value of a call is

\[
C_t = b(T) \int_{K_t}^{Y_T} \alpha \lambda^{\alpha} \tau \gamma - 1 (\gamma + \lambda)^{-(\alpha + 1)} dY_T.
\]

The parameters may be estimated if observed prices are assumed to behave according to the model by simply applying a penalty function to the difference between observed prices and model prices conditional on the distribution parameters.

A comparison of the lognormal and SM distributions is made to lift up possible improvements in moving to a three-parameter distribution. There is no guarantee that the parameters of \( f(Y_T) \) will conform to an ex post price distribution with either parameterization. Indeed, they are simply mathematical constructs with probabilitylike properties. The relative performance of the two-candidate distributions is visited in a later section. Before using the derived parameters in an investigation of the economic forces reflected in the markets, an investigation of the properties of the estimates is conducted. This evaluation is the topic of the next section.

**Evaluation of Estimated Distributions: Calibration Tests**

Once estimated, meaningful criteria are needed to evaluate the usefulness or accuracy of the estimates. Calibration, or reliability, refers to the correspondence between a predicted and an actual event. In terms of distributions, calibration describes how close the predicted and resulting functions are. If there was a reason for the market's aggregation of individual expectations to yield estimated parameters that required an adjustment to
correspond to the “true” parameters, then this adjustment is termed the 
*calibration function*. Specifically, if the true parameters of a distribution are 
\( \phi(x) \) and the estimates are \( F(x) \), then \( K(F(x)) = \phi(x) \) implicitly defines a 
transformation \( K(\bullet) \) of \( F \) to generate estimates, \( K(F(x)) \), that are well-
calibrated or reliable. The function \( K(\bullet) \) is called the *calibration function*.
Equivalently, given a subjective or implied PDF, the process generating the 
subjective or implied PDF is said to be *well-calibrated* if the proportion of 
times the realized value lies below the \( r \)th fractile of the implied PDF is 
equal to \( r \) (Curtis et al. [1985]).

A calibration function accounts for more than a simple bias in that it 
corrects all moments of an estimated distribution. The result of calibration is 
to make the long-run probabilities (density) of \( K(F(x)) = \phi(x) \) for any level 
of \( x \). If \( F(x) \) is already well-calibrated, then \( K(\bullet) \) will simply be an identity 
mapping. If, for example, \( F(x) \) places too much weight in the lower tail, 
\( K(\bullet) \) will be lower than a uniform density at low values of \( x \) and higher at 
high values reflecting the reweighting of \( F \) that is necessary to force a 
correspondence to \( \phi(x) \). \( K(\bullet) \) therefore reweights \( F(\bullet) \) and is itself a 
probability measure. The test for calibration then is equivalent to testing the 
uniformity of \( K \). For if \( F(\bullet) \) is calibrated, \( K \) is simply a one-to-one mapping 
whose CDF is a straight line. To implement the test, independent realizations 
(expiration date futures prices) are compared to the estimated ex ante 
distributions for those realizations. If the estimated ex ante distributions are 
accurate, the proportion of realizations falling into the \( r \)th fractile of the 
distributions should equal \( r \).

For the purposes of this study, the calibration function is based on the beta 
distribution with density:

\[
K(x) = x^{p-1}(1-x)^{q-1}/B(p,q). \tag{9}
\]

where \( B(p,q) \) is the beta function with parameters \( p \) and \( q \). As noted in 
Fackler and King [1990], the beta distribution is well-known, flexible, and 
contains the uniform distribution as a special case when \( p = q = 1 \). Fackler 
and King [1990] outline a means of using maximum likelihood estimates of 
the parameters of the beta distribution to explicitly model the calibration 
function. A likelihood ratio statistic is easily constructed for the hypothesis 
that the calibration function is uniform. Shapes other than uniform of the 
fitted calibration curve indicate the “reweighting” of the estimated 
distributions needed to correspond to those subsequently observed.

**Data**

The data consist of a subset of all time-stamped transactions of soybean 
futures and options from the inception of trading in the soybean futures 
options markets on October 31, 1984, and ending September 30, 1988. The 
data were provided on tape from the Chicago Board of Trade. In addition to 
all trades at which a price change occurs, the data set contains bids that 
exceed, and asks that fall below, the previous transactions. Volume data per 
se are not available, but an examination of the mean trading price-change 
time interval was examined as a flag for possible liquidity problems. Lack of
volume does not appear to be a problem per se, but the absence of data on
days that the soybean futures “hit a limit” does filter out several days that
would have been interesting to examine.

Some exclusion criteria were considered to alleviate induced biases. Trades
that occurred more than one year prior to expiration were excluded. Also,
deep in- or out-of-the-money options were scrutinized carefully although
there is no theoretical reason for exclusion. Next, the time interval between
trades of matched prices should be as short as possible. Two earlier studies
(Whaley [1986] and Ogden and Tucker [1987]) require that the futures price
precede the option price. Jordan et al. [1987] simply require that the put,
call, and futures prices each occur within a common 30-second interval.
Since there is little reason (other than volume) to expect that one price
necessarily leads or causes the other, no a priori imposition of order was
made, but the matched futures and option prices were required to be within
90 seconds of each other. Also, based on Bookstaber’s [1981] arguments, it
was felt that synchronous futures and option prices were necessary to avoid
possible distortions found in closing or settlement prices not equally
representative of both markets. And, in order to minimize the day-to-day
effects, the point in time during the day should be during a relatively stable
trading interval and away from the opening and closing periods when price
distortions may be more prevalent.

Synchronous option and futures prices were obtained as follows: For each
type of option, all strike prices that traded during the day were arranged
according to their proximity to 11:00 a.m. (to avoid opening and closing
distortions). Then, for each strike, the option that traded closest to
11:00 a.m. was matched to the nearest futures price. The result each day for
each contract was a complete set of the strikes for both puts and calls that
traded, with each observation as near to 11:00 a.m. as possible, and each
option matched within 90 seconds to the nearest futures price.

The resulting samples are further described in Table 1. For example, the
November 1987 contract distributions were computed on each of the 121
different trading days for which enough different options were traded to
perform the estimation. On average, 11.28 different options were used each
day in the estimation for a total of 1,365 total observed options and futures
prices. The average time difference between 11:00 a.m. and the options
observations was 14.1 minutes and the average length of time between the
option observation and its matched futures price was 13.7 seconds. Sixty-
four percent of the options used for this contract were calls.

In the combined contracts’ samples, 15,020 options were used to compute a
total of 1,715 daily distributions across 26 different soybean futures options
contracts. On average, 8.76 options per day were used to derive an implied
distribution for each contract, with 64 percent of the options being calls.
The mean time difference between 11:00 a.m. and the option’s trade was less
than 20 minutes and the mean difference between collected options prices
and matched futures prices was less than 22 seconds.
Table 1. Descriptive statistics of options samples

<table>
<thead>
<tr>
<th>Contract</th>
<th>Days</th>
<th>Total</th>
<th>Obs.</th>
<th>Dif 1</th>
<th>DifFO</th>
<th>% Calls</th>
<th>Calls</th>
<th>Cents/bu</th>
<th>STD</th>
<th>Cents/bu</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 85</td>
<td>28</td>
<td>192</td>
<td>11.6</td>
<td>14.2</td>
<td>50.5</td>
<td>0.166</td>
<td>0.422</td>
<td>0.137</td>
<td>1.300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 85</td>
<td>57</td>
<td>400</td>
<td>16.4</td>
<td>18.1</td>
<td>61.3</td>
<td>0.071</td>
<td>0.962</td>
<td>0.120</td>
<td>0.401</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 85</td>
<td>72</td>
<td>482</td>
<td>16.6</td>
<td>21.2</td>
<td>64.7</td>
<td>-0.069</td>
<td>1.667</td>
<td>-0.076</td>
<td>1.233</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jul 85</td>
<td>93</td>
<td>684</td>
<td>17.2</td>
<td>21.8</td>
<td>65.8</td>
<td>-0.301</td>
<td>0.996</td>
<td>-0.145</td>
<td>0.962</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug 85</td>
<td>32</td>
<td>213</td>
<td>21.0</td>
<td>28.5</td>
<td>65.8</td>
<td>-1.259</td>
<td>1.768</td>
<td>-0.738</td>
<td>1.277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep 85</td>
<td>22</td>
<td>169</td>
<td>22.4</td>
<td>33.2</td>
<td>71.0</td>
<td>1.440</td>
<td>14.706</td>
<td>-5.032</td>
<td>6.348</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov 85</td>
<td>138</td>
<td>1310</td>
<td>16.9</td>
<td>18.2</td>
<td>63.3</td>
<td>-0.694</td>
<td>2.316</td>
<td>-0.364</td>
<td>2.155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 86</td>
<td>58</td>
<td>432</td>
<td>16.1</td>
<td>21.8</td>
<td>58.3</td>
<td>0.021</td>
<td>0.669</td>
<td>-0.018</td>
<td>1.272</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 86</td>
<td>79</td>
<td>619</td>
<td>14.8</td>
<td>19.3</td>
<td>64.8</td>
<td>0.356</td>
<td>1.806</td>
<td>0.250</td>
<td>0.684</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 86</td>
<td>86</td>
<td>569</td>
<td>18.8</td>
<td>24.5</td>
<td>65.9</td>
<td>-0.077</td>
<td>1.553</td>
<td>0.027</td>
<td>1.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jul 86</td>
<td>120</td>
<td>870</td>
<td>18.8</td>
<td>25.3</td>
<td>71.3</td>
<td>0.223</td>
<td>4.267</td>
<td>-0.532</td>
<td>2.589</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug 86</td>
<td>11</td>
<td>157</td>
<td>23.9</td>
<td>35.1</td>
<td>70.7</td>
<td>6.798</td>
<td>22.345</td>
<td>1.439</td>
<td>3.249</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep 86</td>
<td>20</td>
<td>184</td>
<td>23.8</td>
<td>39.5</td>
<td>66.1</td>
<td>-1.120</td>
<td>3.902</td>
<td>0.202</td>
<td>3.544</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct 86</td>
<td>74</td>
<td>1338</td>
<td>17.2</td>
<td>19.7</td>
<td>60.4</td>
<td>6.116</td>
<td>1.065</td>
<td>2.042</td>
<td>0.746</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 87</td>
<td>44</td>
<td>326</td>
<td>19.0</td>
<td>26.5</td>
<td>61.3</td>
<td>-0.100</td>
<td>0.788</td>
<td>-0.073</td>
<td>4.890</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 87</td>
<td>32</td>
<td>310</td>
<td>19.9</td>
<td>28.9</td>
<td>62.0</td>
<td>-0.007</td>
<td>1.257</td>
<td>-2.585</td>
<td>0.910</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 87</td>
<td>20</td>
<td>242</td>
<td>21.9</td>
<td>33.4</td>
<td>67.4</td>
<td>-0.024</td>
<td>0.508</td>
<td>-0.093</td>
<td>0.589</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jul 87</td>
<td>87</td>
<td>687</td>
<td>18.1</td>
<td>25.2</td>
<td>68.1</td>
<td>1.170</td>
<td>5.934</td>
<td>0.008</td>
<td>2.526</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug 87</td>
<td>49</td>
<td>420</td>
<td>18.6</td>
<td>33.0</td>
<td>67.2</td>
<td>2.833</td>
<td>11.146</td>
<td>1.662</td>
<td>9.329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep 87</td>
<td>42</td>
<td>335</td>
<td>19.4</td>
<td>35.5</td>
<td>58.5</td>
<td>0.095</td>
<td>3.804</td>
<td>-0.364</td>
<td>2.571</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov 87</td>
<td>121</td>
<td>1365</td>
<td>14.1</td>
<td>13.7</td>
<td>64.0</td>
<td>0.338</td>
<td>2.314</td>
<td>0.268</td>
<td>1.960</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 88</td>
<td>81</td>
<td>653</td>
<td>14.2</td>
<td>20.5</td>
<td>64.0</td>
<td>-0.070</td>
<td>1.024</td>
<td>-0.089</td>
<td>5.520</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 88</td>
<td>106</td>
<td>953</td>
<td>14.5</td>
<td>16.8</td>
<td>67.2</td>
<td>0.545</td>
<td>6.517</td>
<td>0.362</td>
<td>0.798</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 88</td>
<td>92</td>
<td>787</td>
<td>15.2</td>
<td>17.3</td>
<td>62.9</td>
<td>0.526</td>
<td>4.964</td>
<td>1.318</td>
<td>6.925</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jul 88</td>
<td>128</td>
<td>1058</td>
<td>17.6</td>
<td>19.1</td>
<td>68.5</td>
<td>2.083</td>
<td>8.790</td>
<td>0.792</td>
<td>9.276</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug 88</td>
<td>23</td>
<td>265</td>
<td>23.3</td>
<td>28.3</td>
<td>66.8</td>
<td>-0.912</td>
<td>2.359</td>
<td>0.088</td>
<td>1.549</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Days .......... Total number of trade days for which option premia were sufficient to recover parameters of implied distributions

Total Obs ..... Total number of observations that remained in the contract after the deletion/estimation criteria

Dif 1 ...... Mean absolute difference in minutes of all strike prices used from 11:00 a.m.

DifFO ...... Mean absolute difference in seconds between the option and futures prices

% Calls ...... Percent of the sample represented by calls

Cents/bu ...... Mean of futures price minus first moment of implied distribution

STD .......... Standard deviation of within contract daily cents/bu

To solve for the parameters of the expected price distributions, a risk-free rate was also needed that corresponds to \( b(T) \). The rate used is based on the daily discount-basis yield of three-month T-bills as provided by the Federal Reserve Bank of Cleveland.

Methods and Results

If the no-arbitrage option pricing model is assumed to be correct, observed option prices may be used to recover estimates of the parameters of the candidate price distributions by searching for the values of the parameters that would most nearly result in the observed option prices. The procedure is
similar to IV studies that search for a volatility measure that would equate observed prices to those predicted by the market, but the dimension of the choice variable vector is increased to the dimension of the parameters of the candidate distribution.

The expression:

\[
\min_{\beta} \left\{ \sum_{i=1}^{n} \left( \frac{C_i}{K_i} \int_{K_i}^{\infty} f(Y_T|\beta) (Y_T - K_i) dY_T \right)^2 \right\} + \sum_{j=1}^{m} \left( \frac{P_j}{\sigma_j} \int_{0}^{\sigma_j} f(Y_T|\beta) (K_j - Y_T) dY_T \right)^2 \right\},
\]

(10)

was used to solve for a daily set of implied distribution parameters. \(\beta\), that most nearly results in the observed option prices for both the SM and LN distributions. Daily samples of \(m\) puts and \(n\) calls were used subject to the requirement that \((m + n)\) be greater than the dimension of \(\beta\) (greater than three for the SM and greater than two for the LN). Some studies have weighted the individual options differently based upon the elasticity of price with respect to variance, or by the degree to which the option was in- or out-of-the-money. Simple nonlinear least squares were used for simplicity. Implied ex ante price distributions were computed for a total of 1,715 contract days (or 3,430 distributions) over the 26 contracts with an average of 8.76 different options per day. Sherrick [1989] provides daily parameter estimates for each distribution.

An example of the daily estimated distributions is shown in Figure 1, panel A, which illustrates the implied ex ante price densities for the July 1988 soybean futures contract that was implicit in the options prices on the 64th day prior to expiration of the option. Panel B shows the corresponding CDFs for both distributions.

Because the current futures prices do not enter directly into the estimation of the distributions [equation (10)], comparisons of the first moment of the implied distributions with the current futures price gives an indication of the option market's expected direction of futures price movements. For example, if the futures price is currently lower (higher) than the mean of the implied distribution, it indicates that the futures price is expected to increase (decrease) according to the option-based estimates. This type of information is generally not available if the futures price is used in the estimation of the ex ante distribution. Further, this approach may provide additional evidence as to the unbiasedness of current futures prices as predictors of future prices. Table 1 also summarizes the mean difference between the average futures price and \(E(Y_T)\) as reflected in the implied distributions. Note the increased differences during the near-harvest contracts. For the positive (negative) differences, it may indicate that the options market expectation is for the price to rise (fall) as contract expiration approaches. Also, there is considerable increased variability in this measure around the harvest months possibly indicating a greater likelihood of reflecting expected futures price movements through the options premia. The SM distribution yields smaller differences between futures prices and
Figure 1.
Panel A. Ex ante PDFs for July 1988

Panel B. Ex ante CDFs for July 1988
Figure 2. Panel A. Calibration function

Beta (1.41, 0.93)

Panel B. Ex ante CDFs for May 1988

Original SM

Calibrated SM

Expected price - 40 days out
the first moments of the implied distributions than for the lognormal
distribution. The true distribution of expected prices may be better reflected
by the three-parameter SM distribution than the more restrictive two-
parameter lognormal distribution.

Calibration tests are used to assess whether the estimated distributions
generate reliable descriptions of future distributions. For each of the 26
soybean contracts, ex ante distributions were examined at seven fixed
independent (nonoverlapping) intervals prior to expiration\(^8\) (7, 10, 20, 40,
60, 80, and 100 days), and comparisons were made with the contract prices
at expiration. The fitted beta distribution [equation (9)] was examined as an
indication of the shape of the calibration function. Figure 2, panel A, shows
a representative beta calibration function. The ex ante distribution \([F(Y_p)]\)
and the calibrated distribution \([K(F(Y_p))]\) are given in panel B. The slope
of the calibration function in A corresponds to the reweighting necessary to
arrive at the calibrated function in B. The ex ante distribution shown is for
the May 1988 soybean contract at 40 days prior to expiration parameterized
as an SM distribution. The beta calibration function \((p = 1.41 \text{ and } q = .93)\)
is derived by assessing the accuracy of 21 different contracts that traded 40
days out and then is used to recalibrate the expected price distribution. The
calibrated distribution in panel B now corresponds to an adjusted probability
assessment that, in the long run, most nearly corresponds to the pattern of
realizations experienced.

Table 2 gives a summary of the estimated values of \(p\) and \(q\) for the seven
fixed intervals used in the calibration exercise. The interpretation of the
 calibration function is that it serves to reweight the estimated CDF to arrive
at one that would have allowed the realizations to occur with highest
probability. The likelihood ratio statistic is also calculated and the
 corresponding \(p\)-value for the null of uniformity is given.

A "menu" for interpretation is included near the bottom of the table. For
both the SM and lognormal distributions, there is some evidence that ex
ante distributions are not well calibrated one week prior to contract
expiration. The general shape of the beta function indicates a tendency for
the futures prices to rise at expiration. More precisely, the option-based
estimates were drawn from a distribution that was overdispersed and located
to the left of the unknown distribution. The results for location are
consistent with Paul's [1986] observations of a liquidation bias in futures
price spreads and with theories of backwardation.

Over the range of 10-40 days prior to expiration, there is little evidence of
miscalibration. The August contract was dropped in tests for 40 days and
beyond in order to avoid overlapping of intervals from which the calibration
functions were estimated. Other differences in sample sizes arise due to

\(^8\) For 80-100 days, the samples are not strictly independent so the results must be interpreted
with caution. The portion of the intervals that overlap is roughly 10 percent so distortions on
this account are not likely to be large.
Table 2. Calibration statistics

<table>
<thead>
<tr>
<th></th>
<th>Days to maturity</th>
<th>Number of observations</th>
<th>Parameters of beta calibration function</th>
<th>LR statistic</th>
<th>prob &gt; LR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>22</td>
<td>5.274, 3.660</td>
<td>19.978</td>
<td>0.000</td>
</tr>
<tr>
<td>Lognormal</td>
<td>10</td>
<td>24</td>
<td>1.455, 1.117</td>
<td>2.067</td>
<td>0.356</td>
</tr>
<tr>
<td>distribution</td>
<td>20</td>
<td>25</td>
<td>1.331, 1.228</td>
<td>1.070</td>
<td>0.586</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>21</td>
<td>1.184, 0.724</td>
<td>4.480</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>20</td>
<td>1.288, 0.791</td>
<td>3.854</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>19</td>
<td>0.898, 0.538</td>
<td>7.544</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>19</td>
<td>0.894, 0.506</td>
<td>9.444</td>
<td>0.009</td>
</tr>
<tr>
<td>SM</td>
<td>5</td>
<td>22</td>
<td>4.678, 2.901</td>
<td>17.713</td>
<td>0.000</td>
</tr>
<tr>
<td>distribution</td>
<td>10</td>
<td>24</td>
<td>1.428, 1.091</td>
<td>1.982</td>
<td>0.371</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>25</td>
<td>1.359, 1.261</td>
<td>1.234</td>
<td>0.540</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>21</td>
<td>1.410, 0.930</td>
<td>2.975</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>21</td>
<td>1.463, 1.031</td>
<td>2.410</td>
<td>0.300</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>19</td>
<td>1.259, 0.845</td>
<td>2.438</td>
<td>0.295</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>19</td>
<td>1.442, 0.895</td>
<td>3.451</td>
<td>0.178</td>
</tr>
</tbody>
</table>

"Menu" to interpret shape of beta calibration function

- \( p = q = 1 \) .......... uniform distribution
- \( p = q > 1 \) .......... "S"-shaped; (similar to normal CDF)
- \( p = q < 1 \) .......... reverse "S"-shaped; (mirror image of "S" across 45° line)
- \( p > 1, q = 1 \) .......... "C"-shaped; (humped over uniform CDF)
- \( p > 1, q < 1 \) .......... "U"-shaped; (dropped under uniform CDF)

limit days or other days on which lack of data precluded the estimation of parameters at a particular interval prior to expiration. At 80-100 days prior to expiration, the lognormal distribution appears to be miscalibrated, but the SM distribution appears reasonably well calibrated. In neither case is the evidence strong enough to suggest definitively that recalibration is necessary for reasonably accurate estimates to be obtained with the present methods.

In short, the calibration results indicate that the estimated ex ante price distributions generally are appropriate for describing distributions of realized outcomes. The exception is that the ex ante price distributions estimated one week prior to expiration may not accurately reflect realized outcomes. These results are different from Fackler and King [1990], which finds some evidence of miscalibration using closing data at four weeks prior to expiration using a lognormal distribution.

Varying Parameter Models of Volatility

The ex ante assessments of price distribution provide a unique set of data from which inferences regarding market forces and agents' actions may be assessed. If the market correctly aggregates uncertain price information and the options premiums accurately reflect that information, then changes in the distributions reflect both decreases in time to expiration and exogenous events that manifest themselves in parameter changes. Before using the derived parameters in an investigation of the economic forces reflected in the markets, the time-series properties of the parameters must be considered. The evidence points most consistently to time-to-expiration and
price level as two of the more important potentially relevant explanatory variables.

The fact that each successive day’s estimate of \( f(Y_r) \) is for a subperiod of the previous observations suggests that a time series model might be appropriate to explain changes in volatility and changes in other aspects of the distribution. However, weekends, missing days, holidays, thin markets, and the like would render this approach quite difficult as no consistent interval exists between “daily” observations and, hence, the interpretation of the autocorrelation measures and issues of nonstationarity would be unclear. Instead, a regression framework is imposed that takes direct recognition of the potential serial dependence and inconsistent observation interval. The posited model is quite simple in keeping some comparability to earlier approaches, while allowing very general parameter variation.

It is quite reasonable to expect the volatility of the implied distributions to be a function of time-to-maturity (TTM) by the very nature of the variable. Resolution of uncertainty and the collapse of ex ante distributions to the realized price evolve through time. Also, the price of the underlying variable may potentially influence volatility if there is a fairly constant coefficient of variation describing the uncertainty [consider the Constant Elasticity of Variation (Cox [1975]-CEV) model as a parallel]. However, with the problems of time series modeling noted above, alternate models are pursued.

It is assumed instead that the following model holds:

\[
\Theta_t = \beta_{1r} + \beta_{2r}(P_t) + \beta_{3r}(\text{TTM}_t) + u_{ir},
\]

where \( \Theta \) is the characteristic of the distribution being investigated (in this case, standard deviation), the \( j \) subscript is the regime index running from 1 to \( r \) implying \( r-1 \) switchpoints, and \( P_t \) is the futures price on day \( t \). If \( r = 1 \) (no switchpoints), the model reduces to the common ordinary least squares. If the index \( j \) depends upon some other possible stochastic variable \( z \), so that the regime of influence depends upon a threshold value of some \( z_j \geq z^* \) for \( j \) to be in a new regime, then the model is of the class of switching regimes regressions. For this purpose, the most likely regime index is obviously time, but the cutoff values, \( z^* \), at which one regime supersedes the next must be estimated as well.

Brown, Durbin, and Evans (BDE [1975]) were among the first to investigate the detection of switchpoints in such a context. Because of potential nonnormality problems, their test, while powerful at detecting the presence of a switchpoint, is not well-suited to defining the precise location of that point. The BDE model makes use of recursive residuals and one-period-ahead forecast errors. Intuitively, each of the one-period-ahead forecast errors should contribute to the sum of squared residuals in about the same proportion. Put another way, the cumulative sum of squared errors, given

\footnotesize
9 See Choi and Longstaff [1985] for an application of the Cox (1975)-CEV model.

401
that the model is correct, should increase in proportion to the number of forecasts.\textsuperscript{10}

In order to test for the existence of multiple regime shifts, the BDE test was applied over sequentially longer intervals until a switchpoint was located. Then the most likely point in the interval was found with the log-likelihood ratio test (LLR) by computing the test statistic for each possible switchpoint and picking the one that maximized the ratio (Johnston [1984]). Next, the interval up to and including the switchpoint was excluded from the sample and the process repeated. If no new switchpoints were located, then only two regimes are said to exist. If a new switchpoint was located, each possible subset of the original data, as partitioned by the switchpoints, was tested for the most likely occurrence. The process then searches for a third switch and so on until no new switchpoints are found. Table 3 reports the location and significance of the switchpoints by contract.

The most notable result is the prevalence of switches located within approximately two months prior to expiration. It is suspected that the expiration of the contract immediately preceding each contract would have a spill-over effect as activities are concentrated in what has become the new, nearby contract. However, these effects would tend to be two months apart, except for the September and August contracts, which have contracts that expire one month earlier. There are calendar effects and exchange rules that could cause the range of time between adjacent option contract expirations to vary significantly. For all contracts, there may also be a gradual change in volume over the life of the futures as different participants acquire new reasons to trade various maturity instruments. A changing volume may be associated with new information that would manifest itself as a parameter change. Also, note the high levels of significance for many of the switchpoints. In all but four of the soybean contracts examined, a switchpoint was located with significance greater than .10 (smaller p-value). The caution is simply that there are frequent structural changes that require consideration before using long time series in other economic analyses.

At this point, the switchpoints are still defined only in general terms. That is, if $r^*$ is the switchpoint, the tests will detect either $\beta_{t+r^*} \neq \beta_{t+r}$ or nonconstant error variance across the regimes. Following Goldfeld and

\begin{equation}
\sigma_r = \left( \frac{\sum_{j=k}^{T-r} w_j^2}{T-k} \right) / \left( \frac{\sum_{j=k}^{T-r} w_j^2}{T-k} \right), \text{ for each } r = k, k+1, \ldots, T
\end{equation}

where $w_j^2$ are the recursive residuals (see Brown et al. [1975]) and by definition $s_k = 0$ and $s_T = 1$, and if all forecast errors are identically distributed, the $s_r$ are shown to have a beta distribution with mean $(r-k)/(T-k)$. Departures from the uniform line of the plot of $s_r$ indicate a significant switchpoint is in the interval $k+1, T$. A confidence interval can be set at $(r-k)/(T-k) = C_\alpha$, with $C_\alpha$ a function of the level of significance as tabulated in BDE and elsewhere. Note that if the null of no switch holds, the $w_j$ are independent and the use of such residuals avoids the serial correlation and nonnormality problems associated with ordinary OLS residuals, and thus allows for much more powerful tests (Hays and Upton [1986]).
Table 3. Location and significance of switchpoints

<table>
<thead>
<tr>
<th>Futures contract</th>
<th>First switch*</th>
<th>p-value</th>
<th>Second switch*</th>
<th>p-value</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 1985</td>
<td>23</td>
<td>.0600</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 1985</td>
<td>64</td>
<td>.0123</td>
<td>50</td>
<td>.0300</td>
<td></td>
</tr>
<tr>
<td>May 1985</td>
<td>46</td>
<td>b</td>
<td>28</td>
<td>.0002</td>
<td>c</td>
</tr>
<tr>
<td>Jul 1985</td>
<td>33</td>
<td>b</td>
<td>32</td>
<td>.0200</td>
<td></td>
</tr>
<tr>
<td>Aug 1985</td>
<td>24</td>
<td>.0012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep 1985</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov 1985</td>
<td>56</td>
<td>.0130</td>
<td>(day 28)c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 1986</td>
<td>30</td>
<td>.0300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar 1986</td>
<td>39</td>
<td>b</td>
<td>56</td>
<td>.1600</td>
<td></td>
</tr>
<tr>
<td>May 1986</td>
<td>59</td>
<td>.0004</td>
<td>51</td>
<td>.1800</td>
<td>(day 130)c</td>
</tr>
<tr>
<td>Jul 1986</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aug 1986</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep 1986</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov 1986</td>
<td>44</td>
<td>.0010</td>
<td>29</td>
<td>.0050</td>
<td></td>
</tr>
<tr>
<td>Jan 1987</td>
<td>57</td>
<td>.1100</td>
<td>50</td>
<td>.1200</td>
<td></td>
</tr>
<tr>
<td>Mar 1987</td>
<td>69</td>
<td>.0010</td>
<td>51</td>
<td>.0010</td>
<td></td>
</tr>
<tr>
<td>May 1987</td>
<td>26</td>
<td>.0540</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jul 1987</td>
<td>23</td>
<td>b</td>
<td>32</td>
<td>.0010</td>
<td></td>
</tr>
<tr>
<td>Aug 1987</td>
<td>49</td>
<td>.0100</td>
<td>38</td>
<td>.0250</td>
<td></td>
</tr>
<tr>
<td>Sep 1987</td>
<td>30</td>
<td>.0900</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov 1987</td>
<td>72</td>
<td>.0010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 1988</td>
<td>57</td>
<td>b</td>
<td>21</td>
<td>.0050</td>
<td></td>
</tr>
<tr>
<td>Mar 1988</td>
<td>39</td>
<td>b</td>
<td>45</td>
<td>.0300</td>
<td>(day 113)c</td>
</tr>
<tr>
<td>May 1988</td>
<td>93</td>
<td>.0250</td>
<td>10</td>
<td>.7700</td>
<td></td>
</tr>
<tr>
<td>Jul 1988</td>
<td>67</td>
<td>.1400</td>
<td></td>
<td></td>
<td>(day 116)c</td>
</tr>
<tr>
<td>Aug 1988</td>
<td>31</td>
<td>.0600</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* . . . . . . days from expiration
b . . . . . . p-value less than .0001
c . . . . . . other days significant at p = .2, or spurious switchpoints found
NS . . . . . . no switchpoint found

Quandt [1973, 1976], Quandt [1958, 1960], and Kane and Unal [1987, 1990], an endogenous switchpoint regression was next used to simultaneously estimate the $\beta_n$, equivalent $t^*$, and an additional parameter, $\sigma^*$, indicating the gradualness of change as one regime supersedes the previous. Note the immediacy of the use of a switching regression technique in event studies. Switchpoints are located first and then compared with plausible events. Running the test in this order conserves degrees of freedom and helps avoid the tendency to form illusory correlations.

There are likely to be only a few relevant switchpoints if the model posited is in fact correct and is relatively stable within given regimes. However, since the number and location of the switchpoints is unknown, a set of dummies to detect the switches cannot be used in any parsimonious fashion. That is, to detect the unknown location of one switchpoint, there would need to be $(n-2) \times 3$ sets of dummies, which obviously exceeds the total degrees of freedom. A feasible problem emerges, however, if a structure is imposed on the set of dummies such that a relatively small number of parameters
describe an entire set of regime dummies. Then the parameters of the regression and the parameters of the dummy equations may be simultaneously estimated from the data. The model used is a variant of the “D-method” of Goldfeld and Quandt [1973, 1976].

To allow the switchpoints to be endogenized, introduce transitional dummy variables, modeled as normal CDFs, $D_{ij}$,

$$D_{ij} = \int_{-\infty}^{z_i} [(2\pi)(\sigma^*)^2]^{-1/2} \exp[-1/2((\Phi - z_i^*)/(\sigma^*)^2)] d\Phi,$$  

(13)

where $j = 1, \ldots, k, \ldots, r$ and by definition $D_{i0} = 1$ and $D_{ir} = 0$. Then the $k$th regime is multiplied by

$$\tau_{ik} = \prod_{j=0}^{k-1} D_{ij} \prod_{j=k}^{r} (1 - D_{ij}) \text{ for } t = 1, \ldots, T,$$  

(14)

thereby approximating a step into and out of the regime.

The $r$ regime equations are then summed to get

$$\sum_{k=1}^{r} \theta_{i}\tau_{ik} = \sum_{k=1}^{r} \{(\beta_{1k} + \beta_{2k}P_i + \beta_{3k}TTM_i + e_{ik})(\tau_{ik})\}.  

(15)$$

Assuming a normal distribution for $\theta$ with mean and variance of

$$\mu_{i\theta} = \sum_{k=1}^{r} \{(\beta_{1k} + \beta_{2k}P_i + \beta_{3k}TTM_i)(\tau_{ik})\}, \quad \text{and}$$

$$\sigma^2_{i\theta} = \sum_{k=1}^{r} \sigma^2_{\theta k}(\tau^2_{ik}).$$  

(16)

(17)

where $\sigma^2_{\theta k}$ is the $k$th regime error variance implies a log-likelihood function.

$$L = -(T/2)\log 2\pi - (1/2) \sum_{t=1}^{T} \log \sigma_{i\theta}^2$$

$$- (1/2) \sum_{t=1}^{T} [(\sum_{k=1}^{r} \theta_{i}\tau_{ik} - \mu_{i\theta})^2]/\sum_{t=1}^{T} \sigma_{i\theta}^2.$$  

(18)

Maximizing $L$ with respect to its unknown parameters gives the switchpoints, $z^*$; the gradualness of change, $\sigma^*$; and the parameters of the original regression relation. The endogenous nature of the D-method lends itself to event study as the switchpoints are taken as events and estimated rather than specified a priori.

The above equation was estimated via maximum likelihood for various contracts that displayed regime dependence in the earlier tests of BDE and the LLR test. Starting values for $\beta$ were taken as the OLS coefficients of the regression over the subperiods as delineated by the likelihood ratio test, and the error variance in separate regimes was initially assumed identical. The value of $z^*$, or the most likely switchpoint, was started at the earlier located switchpoint and not permitted to vary beyond the ends of the sample. After some experimentation, the variance of the dummy equation was set to initial values near 0 and permitted to range to 20 percent of the maximum time index. Greater values tended to obliterate the discrimination between regimes.
The contracts' switching regression results are given in Table 4 for 15 of the contracts that displayed regime dependence earlier. As expected, the switchpoints, $z^*$, correspond well to those located by the earlier methods. Also, the coefficients of the TTM and price level differ by regime indicating that there is a need to consider regime-dependent influences before using these types of implied parameters in other economic contexts. Finally, the switching regression methods serve to confirm dates that may serve as events. For the present purposes, the most obvious and prevalent switches appear to be within bimonthly intervals prior to expiration. Further classifying the causes and modeling them directly present topics for further study.

Table 4. Switching regression results for model of expected soybean futures price distributions

<table>
<thead>
<tr>
<th>Contract</th>
<th>$\beta_{11}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{31}$</th>
<th>$\beta_{12}$</th>
<th>$\beta_{22}$</th>
<th>$\beta_{32}$</th>
<th>$\sigma^*$</th>
<th>$z^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 86</td>
<td>333.454</td>
<td>-0.493</td>
<td>1.396</td>
<td>-94.032</td>
<td>0.343</td>
<td>0.764</td>
<td>4.006</td>
<td>26.779</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Mar 86</td>
<td>-245.656</td>
<td>0.589</td>
<td>2.327</td>
<td>243.311</td>
<td>-0.186</td>
<td>0.174</td>
<td>7.238</td>
<td>30.443</td>
</tr>
<tr>
<td></td>
<td>(0.196)</td>
<td>(0.100)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>May 86</td>
<td>-308.930</td>
<td>0.725</td>
<td>0.937</td>
<td>-220.936</td>
<td>0.604</td>
<td>0.706</td>
<td>5.765</td>
<td>58.352</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Jul 86</td>
<td>-639.824</td>
<td>1.300</td>
<td>2.208</td>
<td>-329.536</td>
<td>0.814</td>
<td>0.526</td>
<td>9.976</td>
<td>37.691</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.042)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Nov 86</td>
<td>-267.819</td>
<td>0.696</td>
<td>1.338</td>
<td>-262.005</td>
<td>0.865</td>
<td>0.203</td>
<td>5.096</td>
<td>72.787</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Jan 87</td>
<td>-175.458</td>
<td>0.459</td>
<td>1.465</td>
<td>183.187</td>
<td>-0.202</td>
<td>0.694</td>
<td>3.472</td>
<td>51.457</td>
</tr>
<tr>
<td></td>
<td>(0.157)</td>
<td>(0.063)</td>
<td>(0.000)</td>
<td>(0.002)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Mar 87</td>
<td>206.990</td>
<td>-0.283</td>
<td>0.736</td>
<td>-141.021</td>
<td>0.427</td>
<td>0.769</td>
<td>7.182</td>
<td>64.884</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.481)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>May 87</td>
<td>314.278</td>
<td>-0.499</td>
<td>0.420</td>
<td>188.478</td>
<td>-0.239</td>
<td>0.508</td>
<td>1.923</td>
<td>25.083</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.135)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Jul 87</td>
<td>-132.042</td>
<td>0.372</td>
<td>3.108</td>
<td>-125.925</td>
<td>0.486</td>
<td>0.080</td>
<td>7.675</td>
<td>22.059</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(0.019)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Sep 87</td>
<td>-378.198</td>
<td>0.914</td>
<td>1.585</td>
<td>-359.497</td>
<td>0.988</td>
<td>0.434</td>
<td>9.965</td>
<td>40.807</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.003)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Nov 87</td>
<td>-172.275</td>
<td>0.475</td>
<td>1.035</td>
<td>-366.892</td>
<td>1.029</td>
<td>0.360</td>
<td>9.531</td>
<td>72.116</td>
</tr>
<tr>
<td></td>
<td>(0.166)</td>
<td>(0.037)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Jan 88</td>
<td>191.226</td>
<td>-0.196</td>
<td>4.491</td>
<td>-45.968</td>
<td>0.317</td>
<td>0.348</td>
<td>7.257</td>
<td>22.761</td>
</tr>
<tr>
<td></td>
<td>(0.277)</td>
<td>(0.514)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Mar 88</td>
<td>-152.404</td>
<td>0.389</td>
<td>2.137</td>
<td>-121.303</td>
<td>0.552</td>
<td>0.056</td>
<td>14.607</td>
<td>63.771</td>
</tr>
<tr>
<td></td>
<td>(0.263)</td>
<td>(0.078)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>May 88</td>
<td>-38.000</td>
<td>0.235</td>
<td>1.237</td>
<td>323.480</td>
<td>-0.151</td>
<td>0.248</td>
<td>12.975</td>
<td>100.262</td>
</tr>
<tr>
<td></td>
<td>(0.772)</td>
<td>(0.232)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Jul 88</td>
<td>554.510</td>
<td>-0.372</td>
<td>-2.048</td>
<td>-532.197</td>
<td>1.008</td>
<td>1.148</td>
<td>12.371</td>
<td>52.041</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.006)</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

Notes: Entries are estimated coefficients from equation (11) that correspond to the coefficients from equation (4) and parameters from the dummy equation (6). For example, $\beta_{ij}$ is for the constant in regime $j$, $\beta_{zj}$ is for price level in regime $j$, and $\beta_{zj}$ is for the time-to-maturity in regime $j$.

$P$-values are given in parentheses.
Several other interesting results are indicated. First, note that in four fifths
of the soybean contracts, the coefficient on TTM<sub>1</sub> is greater than on TTM<sub>2</sub>,
indicating that the rate of decline in volatility increases as the time to
maturity decreases. The fact that the formulation actually permits the
regimes to be continuously mixed, with the strength of each regime’s
variables to depend on the dummy parameters, allows a great deal of
smoothing of the coefficients over time. Nonetheless, there appear to be high
levels of discrimination among regimes as evidenced by the low p-values on
the coefficients for z<sup>*</sup> and TTM<sub>1</sub> and TTM<sub>2</sub>. Two thirds of the first regime
price coefficients are positive and five are negative. In the second regime,
four fifths of the price coefficients are positive. The preponderance of
evidence is weakly in support of a positive price-level effect and rather
strongly in support of a multiple regime influence of TTM on volatility.

It is also interesting to consider the results from an event-study standpoint.
Each of the switches admits itself as a candidate event that corresponds to a
temporal change in the coefficients of the model. The mere existence of
significant switchpoints renders estimates unreliable if estimated from the
entire sample. The lack of switches causes little concern, other than the loss
of degrees of freedom in the current model, as similar and significant
coefficients on TTM and price could exist across regimes and the
discrimination by the dummy could be very poor. In fact, an alternate
interpretation suggested by Kane and Unal [1987] along these lines uses the
variance of the dummy, or σ<sup>*</sup>, to construct regions of time through which a
specified proportion of the change between regimes takes place. Although
the use of the D-method by Goldfeld and Quandt [1973, 1976] was
suggested strictly to make the problem estimable, if the resulting
formulation of the model is descriptive of the structure of the process
generating the observations, the interpretation of the transition parameter as
σ<sup>*</sup> is plausible. In either case, it is interesting to allow a more general
evolution of parameter change through time.

Summary and Concluding Comments

No-arbitrage pricing methods were used to derive estimates of expected
future price distributions. The improved data and parameter estimation
techniques of this study provide an interesting backdrop for empirical study.
Risk-management techniques rely on accurate descriptions of uncertainty.
This study demonstrates one such technique for describing an uncertain
price distribution. The tests of calibration fail to reject the hypothesis that
the method is well-calibrated over intermediate time ranges in the soybean
futures market. Thus, these techniques provide promising alternatives to
derive useful information about future price distribution. And, although we
found low significance of the tests for calibration, the technique of
“recalibrating” an implied distribution was demonstrated.

A switching regression model was used to describe the parameter variation
in endogenously selected regimes. The notable findings include the fact that
the switches tend to occur within bimonthly intervals prior to expiration and
that the volatility declines at a greater rate as the time to maturity decreases.
No consistent price-level effects were detected, but they may in fact differ
by time regime. A direct application is the generation of an ex ante
by time regime. A direct application is the generation of an ex ante benchmark for event studies. Further, the detection of numerous switchpoints in the implied volatility series points out that inappropriate sample intervals may at times be candidate causes for anomalous results in previous studies that examined time-to-maturity effects without considering possibly varying parameters. The greatest caution pointed out is that inappropriate conclusions may be drawn if the issues of nonstationarity are not first considered.

A possible extension of this research involves pricing and trading strategies. If accurate descriptions of future price distributions are available, a simple option pricing model emerges as was demonstrated for the SM distribution. If low-cost estimates of future price distributions are available and reliable, they may be easily incorporated in forecasted outcomes of various trading strategies. Or, if “better” descriptions of future uncertainties are available, pricing models may be developed that accurately reflect the inner workings of the particular market.
References


Cox, J. C., 1975, “Notes on Option Pricing #1: Constant Elasticity of Variance Diffusions,” unpublished manuscript, Stanford University, Stanford, California.


