DERIVING IMPLIED DISTRIBUTIONS FROM COMMODITY OPTION PRICES: AN APPLICATION TO SOYBEAN, CORN AND WHEAT USING MIXTURE OF LOGNORMALS

by

Rui Fan

A PLAN B PAPER

Submitted to
Agricultural Economics Department
In partial fulfillment of the requirement for the degree of

MASTER OF SICENCE

Department of Agricultural Economics

2001

ACKNOLEDGEMENTS

I would like to acknowledge all the help from Dr. Steven Hanson and Dr. Robert Myers, from the initial ideas for the foundation for this paper, to the proof reading of the final draft. Without them, this would not have been possible.

Also I would like to thank Dr. James Hilker, Dr. Roy Black for providing guidance to my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Section 1. Introduction

Commodity prices are highly volatile, particularly compared to prices of manufactured consumer goods (Newbery and Stiglitz 1981). This high volatility is a result of both seasonal changes and daily supply and demand shocks. When important information about supply and demand conditions arrives, significant price changes may occur. These price movements generate price uncertainty, which can pose major problems to policy makers and industry participants, particularly in countries where export earnings and GDP depend heavily on sales of primary commodities.

In order to reduce the likelihood of unfavorable outcomes, agricultural producers use various strategies to manage price risks. One strategy is to use forward pricing instruments such as forward contracts, futures and options to hedge the risk in selling or buying the physical commodity. Futures and options markets provide important hedging instruments for many of the major commodities in the United States including corn, soybean, wheat, hogs, cattle, etc. These contracts offer speculative instruments to investors and additional pricing strategies to agricultural producers and agribusinesses.

A good estimate of the distribution of commodity prices is important for several reasons. First, efficient pricing of derivatives assets such as commodity options and the effective use of these derivatives requires a good estimate of the underlying commodity price distributions. Second, policy makers need to be able to observe market reactions to their policies or to exogenous shocks. Third, knowledge of the perceived commodity price distribution could also be of value to the private sector. In managing risk in physical commodity trading, investors could make decisions based on the market perception of commodity price risk. Corporations could make hedging decisions based on the uncertainty perceived by the market. Therefore, a lot of effort has been put into forecasting the distribution of commodity prices. However, forecasting commodity price distributions is extremely difficult due to two important features displayed by most commodity prices. First, they often display non-stationarity. Second, there are periods when price jumps abruptly to very high levels (or falls to very low levels) relative to their long run average. These two features of commodity price series make forecasting the distribution of commodity price extremely difficult.

Traditional price forecasting approaches use historical data. One approach is to develop market structure models of commodity prices according to economic theory and supply demand conditions. Much of this research relies on a standard set of economic methods. Forecasts can be generated from either a single equation or from the unrestricted or restricted reduced-form models. Streeter and Tomek (1992) found that the variance of futures price changes depends on a variety of factors, including time-to-maturity, seasonality, economic conditions and market structure. Structural models have the potential to provide useful information in explaining commodity price behavior. However, one difficulty in generating forecasts from structural models is the timing of when information is known. Variables whose value will not be known at the time that actual forecasts are made cannot be incorporated into the model. If there is such a variable in the model, then its value must be forecasted independently and, hence, treated as exogenous in the model.

Traditional approaches also include time series models. These models have been successfully applied to capture the time-varying volatility of commodity prices, such as GARCH (Myers and Hanson, 1993), time-independent mixture-of-normals distribution

(Kim and Kon, 1994) and exponentially weighted moving average (EWMA) models (Venkateswanan and Neenakshi, 1993). However, all these methods only involve historical price data and don't include all available market information. In commodity price time series even when stochastic volatility is explicitly modeled, as in a GARCH framework, parameters are often updated quite slowly even when major regime changes have occurred.

A well-known procedure that is different from the traditional historical data approach incorporates available market information into forecasts of commodity price distributions. This method derives the distribution imbedded in the price of options written on commodity futures prices traded on an exchange. An option is a contract which permits, but does not require, the holder to buy (sell) the underlying asset at a predetermined strike price on a given expiration date.[1] Because the option price itself is observable, and because the option price is a function of the future price distribution, then, given a theoretical model, the market prices of options can be used to derive the market's expectation of the future price distribution. This method is referred as a "market-based" forecast because it is based entirely on the expectation of participants in the option market. Black-Scholes (1973) option pricing formula has been widely used to back out the distribution estimate implied by the model, given market determined option premia. In commodity markets, Black's extension of the Black-Scholes model to price options written on futures contracts has been widely used to derive distributions of the underlying commodity prices.

[1] A call option gives the holder the right to buy the underlying asset by a certain date for a certain price. A put option gives the holder the right to sell the underlying asset by a certain date for a certain price.

Compared with using historical data, Black's formula has several advantages. First, this method incorporates market information. The implied distribution reflects the expectation of market participants about the future price distribution using all information available on the day the option is traded. Second, this approach has the advantages of simplicity and tractability. In Black's model, all information required to implement the option pricing formulas is directly observable except for the volatility of the underlying asset price.

Even though Black's model has been widely accepted, this approach is inconsistent in the following sense. In order to compute the current implied volatility from the Black model, the volatility is assumed deterministic over time. But resulting implied estimates, together with the historical time series data, suggest that volatilities are in fact time varying and stochastic. Thus the purpose of this research is to develop and apply an alternative method that will impose fewer restrictions than Black's model.

The option pricing model derived in this research assumes that futures price changes follow a mixture of lognormal distributions (MLN). This method imposes a flexible distributional assumption which is capable of capturing the fat-tailed, peaked, and skewed characteristics of the underlying asset price distribution. This general class of densities is shown by Ritchey (1986) to be descriptive of the majority of common stock returns. Furthermore, the density is mathematically tractable as well as economically plausible. In the context of risk neutral valuation[2], this model provides an exact solution.

[2] Note that the probability derived from option prices will be the risk-neutral probabilities, because contingent claims are priced accordingly to the risk-neutral distribution rather than the actual distribution. These two distributions will be equal if commodity price risk is not priced. For the remainder of this paper, the term probability, CDF, PDF, and the expected value will refer to these measures under the risk-neutral distribution.

This paper develops a method to estimate the distribution of futures price movements from traded European options using a flexible mixture of lognormals distributional assumption. More specifically, the objectives are: 1) develop an alternative method for pricing commodity options when excess skewness and fat-tailed distributions in the underlying futures price can be modeled as a mixture of two lognormal processes. 2) apply this method to elicit implied distributions from corn, soybean, and wheat futures option prices and test how well this forecasting model predicts distribution. 3) evaluate the performance and forecasting ability of this alternative method.

The rest of this paper is organized as follows: Section 2 is literature review. Section 3 describes a mixture of lognormals method for estimating an asset's PDF from European option prices. Section 4 discusses the application of the method to the corn, soybean, and wheat futures market, and compares the estimated distribution with those derived from Black's model. Section 5 concludes.

Section 2: Literature Review

In their path-breaking paper, Black and Scholes succeeded in solving a differential equation to obtain exact formulas for pricing of European call and put options. Based on this option pricing formula for non-dividend paying stock, Black derived the option pricing formula for a commodity futures option[3]. In recent years, the Black-Scholes option valuation formula has achieved wide acceptance by both theoreticians and market practitioners. The distribution underlying the formula is that the asset price follows geometric Brownian motion with constant volatility. Of the five variables[4], which are necessary to specify the model and determine the equilibrium market premium, all are directly observable except volatility (the standard derivation of return from the underlying assets), which can be estimated using historical data or other approaches.

Therefore, simultaneous observation of option prices and the price of the underlying asset can be used to estimate volatility that is expected over the remaining life of the option. A large amount of empirical research has been done to use market determined option premia to back out the distribution estimate implied by the option pricing model. Option prices have been used to describe the distribution of stock returns (Ait-Sahalia and Lo 1996, Jackwerth and Rubinstein 1996), commodity prices (Turvey

---

[3] Futures price F of commodity can be related to its spot prices by an expression of the form $F=Se^{\alpha(T-t)}$, where T is maturity date, t is current date and. In the case of financial asset, α is the risk-free rate of interest less the yield on the asset. In the case of commodity, α is the risk-free rate of interest plus the storage costs per dollar per unit time less the convenience yield. It is shown in Black (1976) that a futures price can be treated in the same way as a security paying a continuous dividend yield at risk free interest rate r. Thus, the expected growth rate in a futures price in a risk-neutral world would be zero. The expected gain to the holder of a futures contract in a risk-neutral world must be zero.

[4] The five variables in Black-Scholes model are strike price, current future price, time to maturity, risk-free interest rate and volatility of the underlying asset.

1990, Hauser and Neff 1985), oil prices (Melick and Thomas 1997), exchange rate (Campa, Chang, and Reinder 1997; Malz 1996), and interest rates (Abken 1995, McCauley and Melick 1996). These studies find that the variances implied from market option premium and the Black-Scholes model (or extensions of the Black-Scholes model in different markets) are often better predictors of variances of underlying assets than those obtained from historical data.

Other studies on the information content of implied volatilities from different underlying assets, including commodity prices, yield disappointing results. Canina and Figlewski (1992) find that "implied volatility has virtually no correlation with future volatility, and it does not incorporate the information contained in recent observed volatility." Day and Lewis (1992) find the coefficient of the implied volatility is barely significant in a volatility forecast equation. In addition, some studies reported that market prices of options appear to deviate systematically from theoretical prices. These aberrations are documented in empirical studies by Black (1975), MacBeth and Merville (1980), Rubinstein (1985), and Whaley (1982). A main explanation for these observed discrepancies is that the distribution of the underlying asset's return is not lognormal, as assumed by Black-Scholes model. In the case of commodity futures, the available evidence doesn't support identical normal distributions for the distribution of commodity futures price changes.

Some researchers have tried to overcome this inconsistency by specifying a pricing model that deals rigorously with the stochastic nature of volatility as in Hull and White (1987) or Wiggins (1987). Stochastic volatility models require the investor to forecast the entire joint probability distribution for asset returns including changes in volatility and also the market price of volatility risk. These requirements make these models significantly more difficult to implement than Black-Scholes and other constant-volatility models. Therefore, some researchers have specified deterministic volatility functions that allow volatility to vary deterministically with the asset price or time. For example, Rubinstein (1994) and Jackwerth and Rubinstein (1996) used implied binominal trees to estimate the underlying binomial probability model assumed to be driving the asset price. However, many of these methods involve complex numerical simulation and optimization routines which are somewhat costly because they are difficult to implement and can take a long time to converge. Some researchers have essentially ignored the evolution of daily volatilities over time and began with an assumption about the future distribution of the underlying asset to directly recover the parameters of that distribution (e.g. Fackler and King 1990). Among these future distributional assumptions, mixtures of lognormal distributions have been applied to different markets.

The mixture of lognormals distribution density is simply tractable. This method has a natural economic interpretation – that of multiple alternative regimes. It has been shown to be appropriate for stock (Ritchey, 1990), crude oil futures (Melick and Tomas 1997), and foreign currency (Leahy and Thomas 1996). In Melick and Thomas (1997), for example, three different lognormal distributions for the price of oil correspond to various outcomes of the 1991 Gulf War. In Melick and Thomas's work, they suggested "this methodology should be useful to researchers who wish to impose a minimum of structure and are i) examining other markets during unsettled times, or ii) investigating asset price distribution that are not adequately described by the lognormal distribution (e.g. a distribution leptokurtosis)." Although this method has been used to derive implied volatility for many markets, the performance of this model has not yet been tested for estimating distributions from commodity futures options. Commodity price series tend to exhibit stochastic volatility and non-stationarity features. Furthermore, commodity prices tend to have significantly "peaked" and "fat-tailed" distributions relative to the Gaussian density. The mixture of lognormals distribution could give a flexible shape to the distribution of commodity prices at maturity. The mixture assumption could also accommodate large price changes and be capable of capturing other observed features of commodity prices.

Section 3: Option Pricing and Implied Probability Distributions

In this section, an option pricing formula for commodity futures option is developed to compute the implied mixture of lognormal distribution for the underlying commodity futures prices. First, Black's option pricing model for commodity futures is presented based on the standard lognormal distribution assumption for the underlying futures prices. Second, a risk-neutral option pricing formula for European options will be developed for the case when the underlying futures price can be described by a mixture of two lognormal distributions (MLN). This functional form of the terminal distribution can easily accommodate a wide variety of shapes for the terminal futures price distributions, giving it the advantage of flexibility, parsimony, and generality. Finally, the methods to derive implied distributions of commodity prices from both Black's model and mixture of lognormals method will be discussed.

3.1 Black's Commodity Option Pricing Model

Black's extension of the Black-Scholes formula can be used to value European options[5] on commodity futures contracts (Black 1976). The assumption behind Black's

[5] A "European option' is one that can be exercised only on the maturity date and not earlier. An "American option" is one that can be exercised at any time up to the date the option expires. The price that is paid for the asset when the option is exercised is called the "exercise price" or "strike price." The last day on which the option may be exercised is called the "expiration date" or "maturity date." Most options on commodities are American options, meaning that they can be exercised at any date on or before maturity or at any time within a specific period (e.g. one month) before maturity. Thus, an option's value will depend on the entire stochastic process for future prices, not just the distribution of future prices at the option's expiration. In this paper, a model is developed to value European options, which involve a less costly and simple solution technique. Some studies have found that the early exercise feature of American options on futures contracts are generally small and there is little loss in pricing efficiency by ignoring the early exercise feature (Sgastri and Tandon 1980, Ramaswamy and Sunderesan 1985). Such results suggest that European pricing models can serve as a useful approximation.

model is that the underlying commodity price follows a Geometric Brownian process. It can be expressed as

$$\frac{dF}{F} = \zeta dt + \sigma dZ \tag{3.1}$$

Here F is the futures price; ζ is the instantaneous expected relative change in the futures price; σ is the instantaneous standard deviation; t is time and Z is a Wiener process. Both ζ and σ are assumed to be constant. Here $dZ = \varepsilon\sqrt{dt}$, where ε is a standard variate. Thus the implicit assumption is that percentage price changes dF/F, over the interval dt are independently drawn from a stationary normal distribution.

Black's futures option pricing model is based on an arbitrage relationship between the risk-free rate of return and the return on a portfolio containing the option. A hedge portfolio is adjusted continuously such that the resultant portfolio is riskless and has a return that replicates a risk-free bond. The number of options hedged against the underlying commodity depends on the strike price, the current price of the underlying commodity, the time to expiration, the interest rate on a risk-free bond, and the variance of proportional price changes of the underlying commodity price. When these factors are known, the proper hedging portfolio is known. Black succeeded in solving the partial differential equation to obtain exact formulas for the prices of European call and put options. In the case of futures the solution to the partial differential equation doesn't include the expected return. Thus, the differential equation for option pricing does not involve parameters that are affected by risk preferences. Therefore, Black's equilibrium option premium is invariant to risk preference and to expected change in the underlying commodity price.

A simpler approach to obtain the same solution is risk neutral valuation (Cox and Ross 1976). Cox and Ross developed a general theory of option pricing which 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 returns 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. In a world where investors are risk neutral, the expected return on all securities is the risk-free rate of interest, r.

Hull (1993) says, "When we move from a risk-neutral world to a risk-averse world, two things happen. The expected growth rate in the underlying asset changes and the discount rate that must be used for any payoffs from the derivative security changes. It happens that these two effects always offset each other exactly."

In a risk neutral world, two important restrictions hold. The first restriction is that the option would be priced according to its risk neutral expected value at maturity, discounted back to the current period at the risk-free rate of interest. For commodity options on futures contracts, this implies,

$$P_{call} = e^{-r(T-t)}\hat{E}[\max(F_T - K,0)]$$

$$P_{put} = e^{-r(T-t)}\hat{E}[\max(K - F_T,0)] \qquad (3.2)$$

where $\hat{E}(\cdot)$ denotes expected value in a risk-neutral world; P_{call} is the price of call options; P_{put} is the price of put options; K is the strike price; F_T is price of future contract at maturity; T is expiration date; and r is the risk free interest rate.

The second restriction is that futures prices are unbiased. Current futures price is the expected value of the futures price at maturity. When using the risk-neutral approach for Black's futures option model, the geometric mean, $\ln(F_{t+\Delta t}/F_t)$, where F_t is the future price at time t, is zero. This zero geometric mean of the log-price return comes from Black's assumption of a zero holding cost for a futures contract. It is shown in Black (1976) that a futures price can be treated in the same way as a security paying a continuous dividend yield at rate r. Thus, the expected growth rate in a futures price in a risk neutral world is zero. If the two restrictions mentioned above did not hold, there would be unexploited profit opportunities.

Using the lognormal distribution and risk neutral valuation, Black's formula for European call and put options written on futures are:

$$P_{call} = e^{-r(T-t)}[F \cdot N(d_1) - K \cdot N(d_2)]$$

$$P_{put} = -e^{-r(T-t)}[F \cdot N(-d_1) - K \cdot N(-d_2)]$$

where

$$d_1 = \frac{[\ln F - \ln K + \frac{\sigma^2}{2}(T-t)]}{\sigma\sqrt{(T-t)}}$$

$$d_2 = d_1 - \sigma \cdot \sqrt{(T-t)} \qquad (3.3)$$

where $N(\cdot)$ is cumulative normal density function.

The only parameter in the pricing formula that can not be observed is σ. When the market option price and the four parameters, *K, F, r, T* are known, the only unknown parameter σ can be backed of the formula. The volatility implied by an option price observed in the market is called "implied volatility" and completely describes the lognormal distribution at maturity that is implied by the option premium. Very often, several implied volatilities are obtained simultaneously from different options on the same stock and a composite implied volatility for the price is then calculated by taking a suitable weighted average of the individual implied volatilities.

Black's work provides the foundation for the theory of futures option valuation. The model for valuing futures options leads to a closed form solution which implies only nonnegative prices. In actual applications, however, the model has certain well-known deficiencies as discussed in section 2.

3.2 A Mixture of Lognormal Approach

In order to overcome the deficiencies in the Black model for valuing commodity futures options, an alternative method is proposed here. In choosing a functional form for the terminal distribution, one should try to balance flexibility, parsimony, and ease of interpretation. Here the futures price at the option's expiration is specified to be drawn from a mixture of two lognormal distributions. The benefits of this technique arise from its flexibility, generality, and directness. This functional form for the terminal distribution can easily accommodate a wide variety of shapes. This approach leads to a probability density function (PDF) clearly distinct from the lognormal benchmark, and typically characterized by skewness and leptokurtosis.

With European style options, the value of the call (put) option is simply the risk neutral expected payoff discounted back to the present using an appropriate interest rate. At expiration the call and put option values must be $\max(F_T - K, 0)$ and $\max(K - F_T, 0)$ respectively. Assuming risk neutral valuation, the option price for calls and puts can be expressed as

$$P_{call} = e^{-r(T-t)} \int_{K}^{+\infty} (F_T - K) \cdot g(F_T) dF_T$$

$$P_{put} = e^{-r(T-t)} \int_{0}^{K} (K - F_T) \cdot g(F_T) dF_T \tag{3.4}$$

where $g(F_T)$ is the risk-neutral probability density function for the commodity futures at time T, which is assumed to be a mixture of two lognormal distributions here.

In the mixture of two lognormals method, the change of log futures price at the option's expiration is specified to be drawn from a mixture of two normal distributions. That is,

$$\ln F_T - \ln F_t \sim \lambda N[\mu_1, \sigma_1^2] + (1-\lambda) N[\mu_2, \sigma_2^2] \tag{3.5}$$

Where $N(\mu_i, \sigma_i^2)$ is a normal distribution with mean μ_i and standard deviation σ_i; and λ and $(1-\lambda)$ are the weights on each normal distribution respectively, where $0 \le \lambda \le 1$. Thus, the probability density function of log future prices, $G(\ln F_T)$, is given by:

$$G(\ln F_T) = \lambda \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{(\ln F_T - \mu_1)}{2\sigma_1}} + (1-\lambda) \frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{(\ln F_T - \mu_2)}{2\sigma_2}} \tag{3.6}$$

Because $\ln F_T$ follows a mixture of normal distributions, and F_T follows a mixture of lognormal distributions. Thus, the density function of F_T is,

$$g(F_T) = \frac{1}{F_T} [\lambda \frac{1}{\sqrt{2\pi} \cdot \sigma_1} e^{\frac{(\ln F_T - \mu_1)}{2\sigma_1}} + (1-\lambda) \frac{1}{\sqrt{2\pi} \cdot \sigma_2} e^{\frac{(\ln F_T - \mu_2)}{2\sigma_2}}]$$

$$= \lambda g_1(F_T) + (1-\lambda) g_2(F_T) \tag{3.7}$$

where $g_i(F_T)$ is the single lognormal distribution

$$g_i(F_T) = \frac{1}{\sqrt{2\pi} \cdot \sigma_i \cdot F_T} e^{\frac{(\ln F_T - \mu_i)^2}{2\sigma_i^2}}$$

Here $g(F_T)$ represents the PDF of mixture of two lognormal distributions, and $g_i(F_T)$ represents the PDF of a single lognormal. There are five unknown parameters λ, μ_1, σ_1, μ_2, σ_2 in this equation.

The option pricing formula is derived as follows where the futures prices follows a mixture of two lognormal distributions. First, consider the call option pricing formula for European options. Substituting (3.7) into (3.4) yields

$$P_{call} = e^{-r(T-t)}[\lambda \int_K^{+\infty}(F_T - K)g_1(F_T)dF_T + (1-\lambda)\int_K^{+\infty}(F_T - K)g_2(F_T)dF_T] \quad (3.8)$$

Thus $P_{call} = \lambda P_{call_1} + (1-\lambda)P_{call_2}$ (3.9)

Where $P_{call_i} = e^{-r(T-t)} \int_K^{+\infty} (F_T - K)g_i(F_T)dF_T$ (3.10)

The call option pricing formula for European option with mixture of two lognormal distributions is then (using the standard Black model)[6],

$$\begin{aligned} P_{call} &= \lambda e^{-r(T-t)}[e^{(\mu_1+\frac{\sigma_1^2}{2})} \cdot N(d_{11}) - K \cdot N(d_{12})] \\ &+ (1-\lambda)e^{-r(T-t)}[e^{(\mu_2+\frac{\sigma_2^2}{2})} \cdot N(d_{21}) - K \cdot N(d_{22})] \end{aligned} \quad (3.11)$$

Where $d_{i1} = \dfrac{-\ln K + \mu_i + \sigma_i^2}{\sigma_i}$, $d_{i2} = \dfrac{-\ln K + \mu_i}{\sigma_i}$ i=1,2

By using the same approach for put,

$$\begin{aligned} P_{put} &= -\lambda e^{-r(T-t)}[e^{(\mu_1+\frac{\sigma_1^2}{2})} \cdot N(-d_{11}) - K \cdot N(-d_{12})] \\ &- (1-\lambda)e^{-r(T-t)}[e^{(\mu_2+\frac{\sigma_2}{2})} \cdot N(-d_{21}) - K \cdot N(-d_{22})] \end{aligned} \quad (3.12)$$

[6] The derivation of option pricing formula 3.11 and 3.12 refer to appendix I.

Under the risk neutral assumption, current futures price is an unbiased predictor of the futures price at maturity. That is the expected value of futures price at maturity equals the current futures price:

$$\hat{E}(F_T) = F_T \tag{3.13}$$

Where $\hat{E}(F_T)$ is the risk neutral expectation of F_T conditional on information available at t. In the case of mixture of two lognormal distributions,

$$\hat{E}(F_T) = \lambda \cdot e^{\mu_1 + \frac{\sigma_1^2}{2}} + (1-\lambda) \cdot e^{\mu_2 + \frac{\sigma_2^2}{2}} \tag{3.14}$$

This constraint can be used to substitute out one of the five unknown parameters in the option pricing formula.

3.3 Using Market Traded Option Prices to Generate Implied Distribution Forecasts

The mixture of lognormal formula is an appealing choice, most importantly because it is sufficiently flexible to allow a wide range of stochastic price behavior. Thus, the implied distribution derived from a mixture of lognormals model should be more realistic than that derived from Black's model.

In order to derive the implied distribution, one typically defines a loss function expressed as the sum of squared deviations of the estimated option prices and the actual option prices. In the case of Black's model, there is only one unknown parameter in the option pricing formulas, σ. Let $\hat{P}_t(K,T,r,F,\sigma)$ denote the estimated price at time t using Black's option pricing model for a call or put that has an exercise price K and maturity at T. Then if the option pricing model is correct, $\hat{P}_t(K,T,r,F,\sigma)$ is expected to be close to the observed market option price P.

Therefore, σ can be estimated by solving

$$\min \sum_i Q_i [P_c - \hat{P}_{ct}(K_i, T, r, F; \sigma)]^2$$

$$\min \sum_i Q_i [P_p - \hat{P}_{pt}(K_i, T, r, F, \sigma)]^2 \quad (3.15)$$

Where p_c is the price for call and P_p is the price for put. *i* is the number of call and put options at time t. Equation (3.15) is weighted sum of squared deviations between realized market options and those estimated via Black option pricing formula. Q_i is weight of each option. The summation i is over call and put options with identical maturities but different strike prices.

In the mixture of two lognormals method, let $\hat{P}_t(K_i, T, r, F, \lambda, \mu_1, \mu_2, \sigma_1, \sigma_2)$ denote the price of the option (call or put) at time *t* with strike price of K_i and maturity date of T by using the mixture of lognormal option pricing model conditional on parameters $(\lambda, \mu_1, \mu_2, \sigma_1, \sigma_2)$. This suggests that the mixture of lognormal parameters implied by options can be estimated by solving:

$$\min \sum_i Q_i [P_c - \hat{P}_{ct}(K_i, T, r, F, \lambda, \mu_1, \mu_2, \sigma_1, \sigma_2)]^2$$

$$\min \sum_i Q_i [P_p - \hat{P}_{pt}(K_i, T, r, F, \lambda, \mu_1, \mu_2, \sigma_1, \sigma_2)]^2 \quad (3.16)$$

which is weighted sum of squared deviations between realized market option prices and those estimated via the mixture of normal option pricing model.

In this paper, options with the same maturity date and different strike prices from the same day will be employed to derive the implied mixture of lognormal parameters. In literature, there are different methods to derive the weights *Q*. One approach in Black's model is to put all the weights on a single contract for which the price of the options is more sensitive to changes in the volatility of the underlying commodity. Usually, this is the at-the-money option, but occasionally it is the near-the-money option[7]. The reasons to use at-the-money options are as follows: First, at-the-money option price is sensitive to volatility, it should therefore return the most accurate measure of volatility. Second, at-the-money options have high volume, which indicate they contain more information about underlying asset's volatility than thinly traded options that are well out of money.

However, this method ignores information about volatility that is available in the contracts that are not at-the-money. In order to account for all the information in the option market, a method is chosen in this research that will contain the information of all options. This method recognizes that the true distribution is the same for all the options on a given future contract with the same maturity date, but different strike price, and chooses the single estimate of distribution which is closest to satisfying the option pricing equation for all exercise price. "Closeness" is measured by the mean square deviation between the observed and the theoretical prices, aggregated over all contracts with a given maturity and weighted by Q_i.

The weight is used to influence the relative contribution of each option in the determination of the implied mixture of lognormal model. The weight Q_i may be assumed to be equal. But a more reasonable assumption is to let each option be weighted by some measure of its liquidity such as volume. This weighting mechanism seems reasonable because options that are far from the money have prices that are not as sensitive to volatility, and should therefore get less weight in the forecast than the more sensitive near-the-money options. However, in this research, due to the lack of

[7] This is a loosely defined term that means the strike price and spot price are not too far.

information of the trading volume of option with different strike price, we will give equal weight to each option.

Section 4: Application to Corn, Soybean and Wheat Futures Prices

4.1 Data Sources

In this section, the mixture of lognormals method is applied to corn, soybean, and wheat futures options traded on the Chicago Board of Trade. Soybean futures contracts are available for January, March, May, July, August, September, and November expiration dates; corn futures contracts are available for March, May, July, September, and December expiration dates; and wheat futures contracts are available for March, May, July, September, and December expiration dates. Options are written on the futures prices for each commodity with an expiration date on the last Friday preceding the first notice day[8] of the corresponding futures contract by at least five business days. Each of these grains has a "storage contract" which expires around the end of usual planting period or the beginning of the usual harvest period, and a "harvest contract" which expires around the end of usual harvesting period. The storage contracts for corn, soybean, and wheat futures are the July contract, July contract, and March contract, respectively; and the harvest contracts for corn, soybean and wheat are December contract, November contract, and July contract, respectively. In this section, the implied distributions for each contract are presented at three points in time. For the harvest contracts, implied distributions are presented for a single day around the most active planting date, a single day near the maturity date of the contract, and a single day in the middle of these two days[9]. For the storage contracts, implied distributions are generated for a single day right after the maturity of the harvest contract, a single day near the

[8] Notice day is the seventh business day preceding the last business day of the delivery month.
[9] The reason for this data selection method is that the pattern of changes of distribution can be observed as we get closer to the maturity date and market uncertainty decreases.

maturity date of the contract, and one day in the middle of these two days. That is, data selection starts at the earliest date and moves closer to the maturity date. Data for market option premiums on futures and the corresponding futures prices were taken from the Wall Street Journal for each contract on the dates specified in Table 4.1.

Table 4.1: Data source

Soybean November 2000 contract – harvest contract			
Date:	05/24/00	08/16/00	10/11/00
Time to maturity:	149	65	9
Soybean July2000 contract – storage contract			
Date:	10/13/99	03/15/00	06/14/00
Time to maturity:	248	97	6
Corn December 2000 contract – harvest contract			
Date:	05/10/00	08/16/00	11/15/00
Time to maturity:	194	96	5
Corn July 2000 contract – storage contract			
Date:	11/17/00	03/15/00	06/14/00
Time to maturity:	216	97	6
Wheat July 2000 contract – harvest contract			
Date:	10/13/99	02/16/00	06/14/00
Time to maturity:	251	125	6
Wheat March2000 contract – storage contract			
Date:	06/14/00	11/29/00	02/14/01
Time to maturity:	251	83	6

For each day, closing option premiums for 6 separate strike prices were quoted in the Wall Street Journal for both call and put options. This provides a total of 12 closing option premiums (6 calls and 6 puts)[10]. In addition, the corresponding closing futures price was collected. There are a total of 6 contracts, each priced at 3 different dates for a total of 18 implied distributions estimated.

The only other data needed to estimate the implied distributions is the risk-free interest rate, which is estimated using data collected from the Wall Street Journal on the

[10] In some days, quotations of call or put for certain strike price are not available.

yields to maturity on Treasury bills with the same maturity date as the options used in the estimation.

For each day, the implied distributions at the option maturity date were estimated for the MLN model and the Black model. A total of 18 implied distribution were derived for each model varying by commodity, storage versus harvest contract, and time to maturity.

4.2 Estimating the Implied Distribution

Traded option premia on selected days are employed to generate implied distributions for MLN and Black methods. The optimization of the objective function (3.16) was performed with a trust-region reflective Newton algorithm using MATLAB function *fmincon* to find the minimum of a constrained nonlinear multivariable objective function.[11]

[11] The MATLAB code can be run under any system that has installed MATLAB with the version equal or higher than 5.3.1 R11.1. Bounds for the parameters were set so that, $lnF_t-10*r<\mu_i< lnF_t+10*r$, $0.001<\sigma_i<0.8$. According to Black's model, $\mu=lnF_t-\sigma^2/2$, thus a reasonable searching bound for μ in MLN should center around *lnF*. Here *10r is* chosen because this range is big enough in all our calculation cases. After trying other larger bounds, such as *15r* or *20r*, the optimization results are the same. This means the chosen bound has already included the minimum points. Similar arguments go to σ_i. I have tried other larger upper bounds for σ_i, and the optimization results remain the same. The chosen bounds here are a good tradeoff between calculation cost and accuracy. The chosen bound for this program ensures accurate optimization results and also make it possible to finish the calculation within several minutes using a PC. Because this is a constrained nonlinear optimization problem, in some cases, the fitting parameters are local optimization points. The results are sensitive to the initial searching points selected. Therefore the parameter searching ranges were divided into multiple smaller lattices and the center of each individual lattice was chosen as the initial points. After this kind of global search, the global minimum was chosen from the minimal local minimum. For the chosen searching bounds, each dimension was divided into three segments, the optimization results would converge and further separation will give the same results.

Table 4.2: Estimated parameters for all contracts/days by MLN methods[12]

	λ	μ_1	μ_2	σ_1	σ_2
Soybean November 2000 contract – harvest contract					
05/24/00	0.08585	2.16043	1.64139	0.16473	0.13922
08/16/00	0.00689	1.85784	1.54968	0.59303	0.07689
10/11/00	0.40717	1.58125	1.55150	0.05289	0.02212
Soybean July 2000 contract – storage contract					
10/13/99	0.07779	2.03517	1.61374	0.16767	0.15402
03/15/00	0.42548	1.74122	0.61329	0.17482	0.08058
06/14/00	0.15659	1.60771	1.63390	0.00148	0.04529
Corn December 2000 contract – harvest contract					
05/10/00	0.17366	1.31442	0.85136	0.04287	0.15358
08/16/00	0.04498	0.62658	0.64106	0.36855	0.08662
11/15/00	0.00298	0.44647	0.73009	0.12726	0.01484
Corn July 2000 contract – storage contract					
11/17/99	0.43659	0.85537	0.69627	0.20277	0.10594
03/15/00	0.44671	0.96236	0.81482	0.19403	0.07584
06/14/00	0.01928	0.47976	0.70982	0.54835	0.04340
Wheat July 2000 contract – harvest contract					
10/13/99	0.07973	1.03267	1.03866	0.65019	0.15930
02/16/00	0.07436	1.08811	1.02844	0.66529	0.11043
06/14/00	0.00419	1.18298	0.95861	0.05150	0.03578
Wheat March 2000 contract – storage contract					
06/14/00	0.17352	1.33423	1.02885	0.27324	0.14286
11/29/00	0.41174	1.03822	0.95037	0.11549	0.07028
02/14/01	0.00552	0.89180	0.98443	0.04961	0.00376

Table 4.2 presents the implied mixture of lognormal estimated parameters $(\lambda, \mu_1, \mu_2, \sigma_1, \sigma_2)$ for each contract on each set of data.

In the following three sections, the graphs of distributions and option pricing errors of each commodity will be presented. The comparison of graphs of different commodities and different days to maturity will be given. The contract selected is harvest contract for

[12] λ, μ_1, μ_2, σ_1, σ_2 are the five parameters in the mixture of lognormal density function defined in equation (3.7)

each commodity. The corresponding graphs of storage contract of these commodities are attached in appendix.

Section 5: Soybean Estimation Results

Figure 5.1 and Figure 5.2 depict probability density functions and cumulative density functions taken from both Black and MLN models using data of May 24, 2000, August 16, 2000, and October 11, 2000 respectively. The top panel uses estimates from the November contract of soybean (harvest contract) on May 24, 2000 (149 days to maturity). The middle panel uses estimates of the same contract on August 16, 2000 (65 days to maturity). The bottom panel uses estimates of the same contract on October 11, 2000 (9 days to maturity). Figure 5.3 plots the difference between estimated option prices and actual prices for calls and puts of these three days.

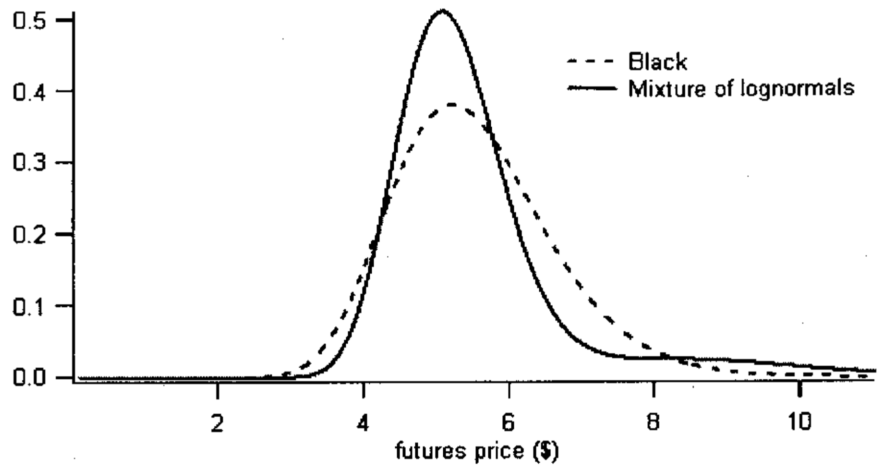
In Figure 5.1 for the data that are far from maturity date, the PDFs are flat and fat, which indicate a bigger variance compared with those from the data that are near the maturity. The PDFs from data that are near the maturity are tall and thin. This is the same for PDFs from both Black and MLN methods. Compared with single lognormal distribution, PDFs of MLN model are always high-peaked and fat-tailed compared with PDFs of Black model.

As shown in Figure 5.2, on the right side, the CDF from MLN model crosses Black CDF at certain point. For Example, in Figure 5.2.c the CDF from MLN model cross the Black CDF at $5, about 4.60% above the current future price ($4.78). To the right of this point, the derived CDF of MLN method lies below that from Black model, indicating greater probability attributed to realization above this level.

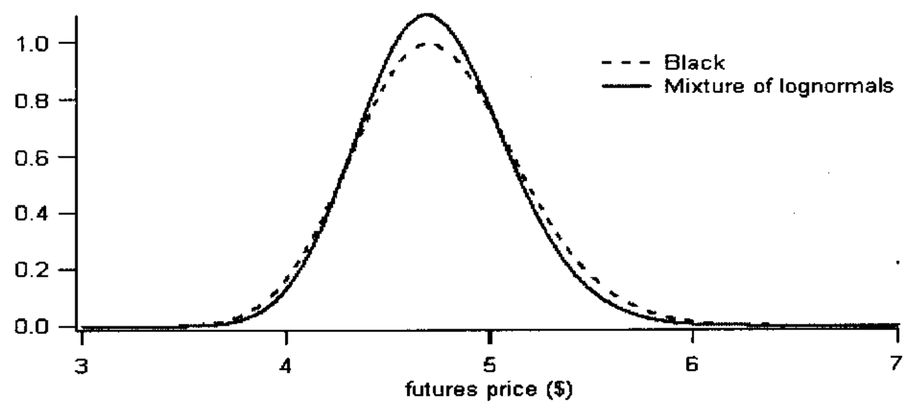
Figure 5.3 depicts call and put option pricing errors. It is clear from the graph that Black model tend to underprice in-the-money call and out-of-the-money put options, and overprice the out-of-the-money call and in-the-money put options, which is consistent with the findings from literature. In Figure 5.3, the option pricing errors from MLN method are in the order of $\$1*10^{-3}$, while the errors from Black model is in the order of $\$1*10^{-2}$. This indicates the MLN method gives better fitting results than Black model.

Figure 5.1: Estimated PDFs -- November (harvest) contract of soybean

a. November Option Maturity as of May 24, 2000



b. November Option Maturity as of August 16, 2000



c. November Option Maturity as of October 11, 2000

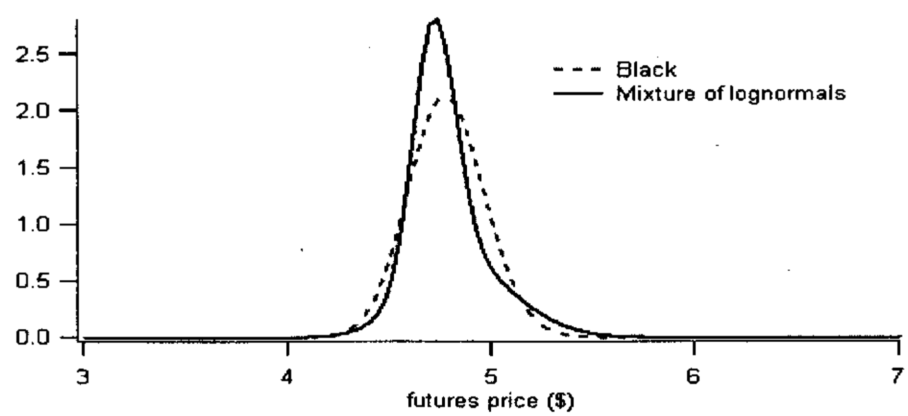
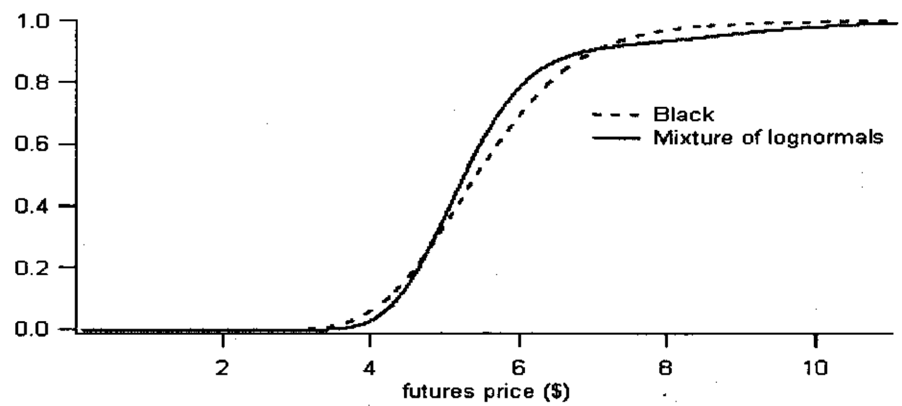
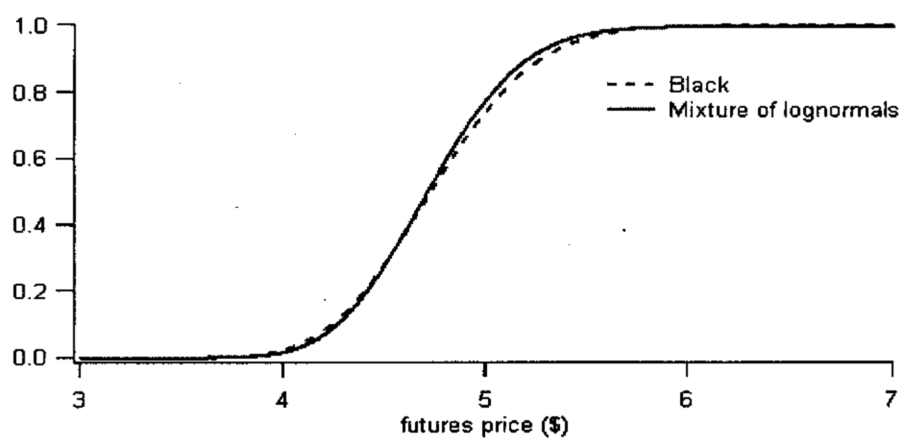


Figure 5.2: Estimated CDFs – November (harvest) contract of soybean

a. November Option Maturity as of May 24, 2000



b. November Option Maturity as of August 16, 2000



c. November Option Maturity as of October 11, 2000

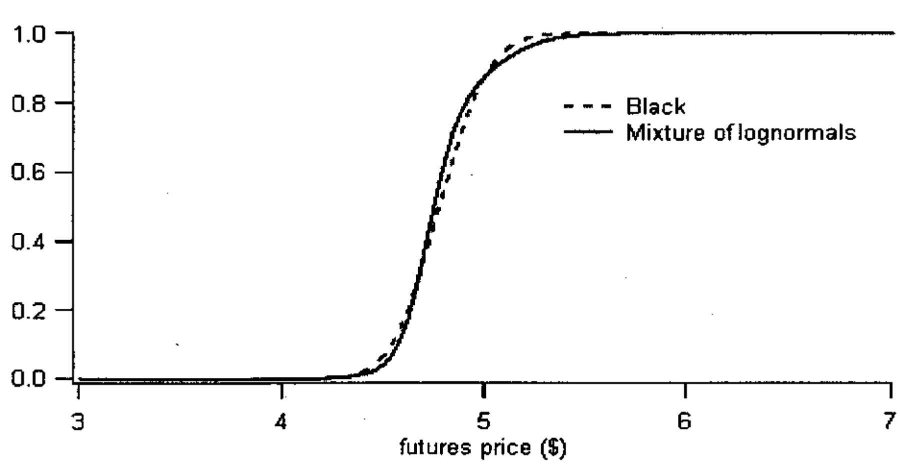
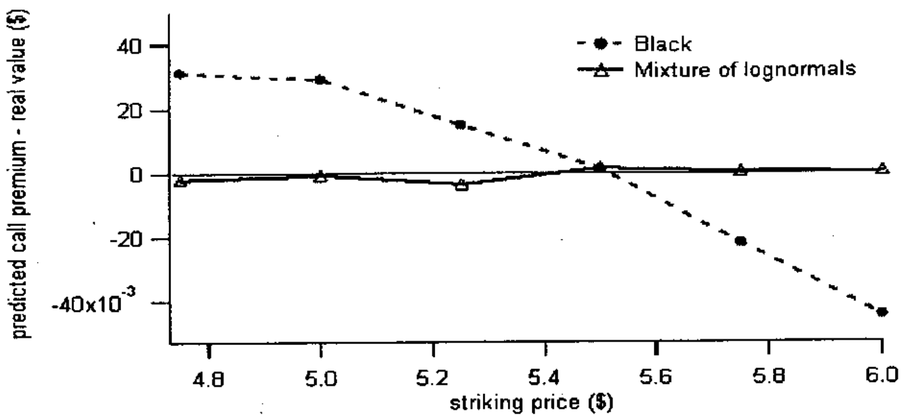
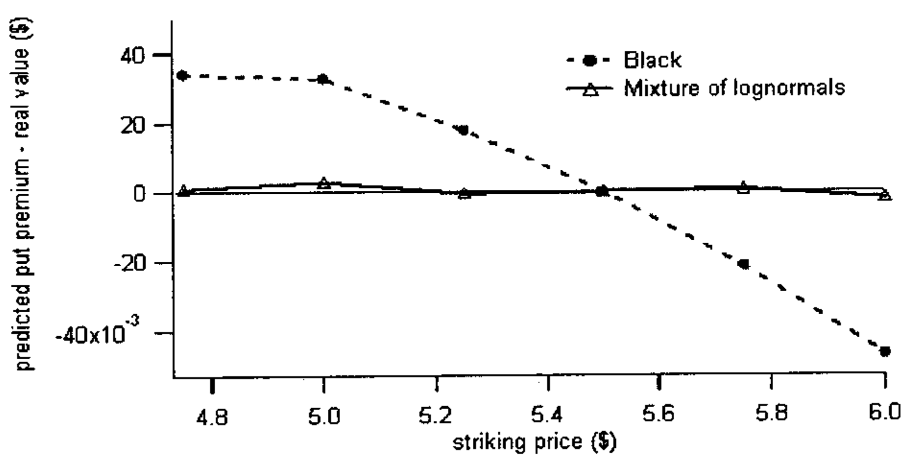


Figure 5.3: Option pricing errors -- November (harvest) contract of soybean

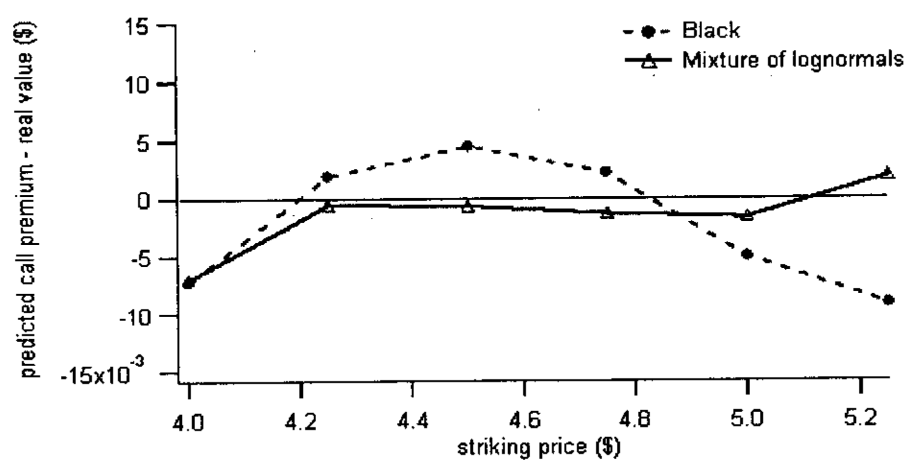
a. Call Pricing Errors of May 24, 2000



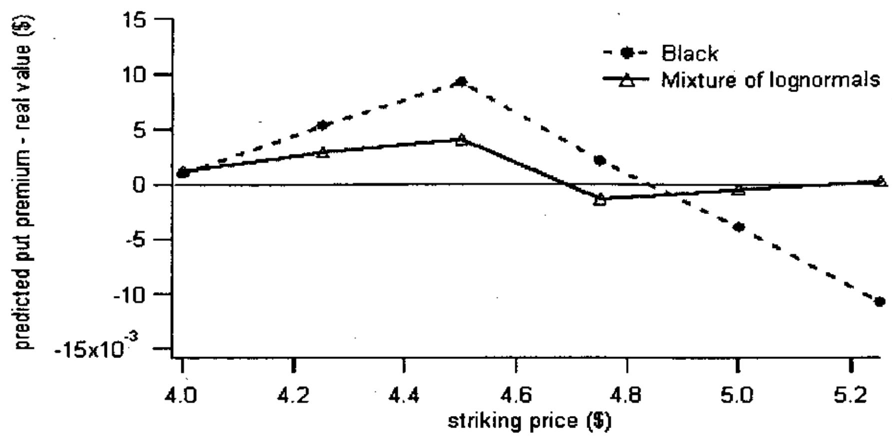
b. Put Pricing Errors of May 24, 2000



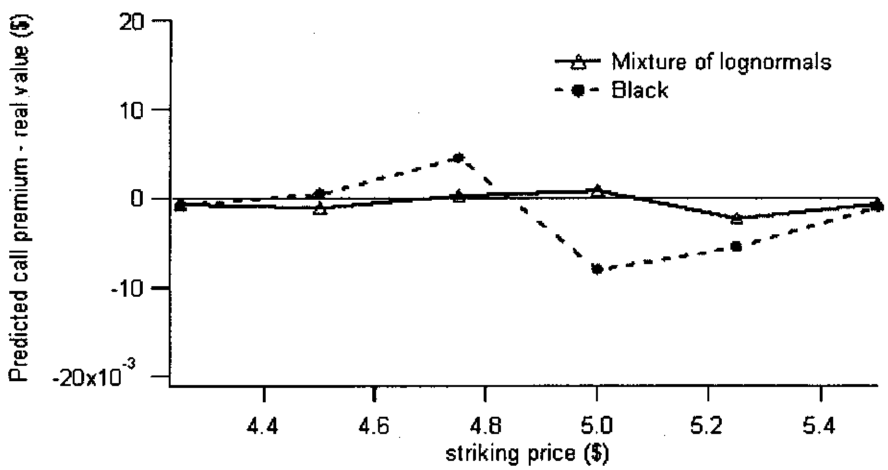
c. Call Pricing Errors of August 16, 2000



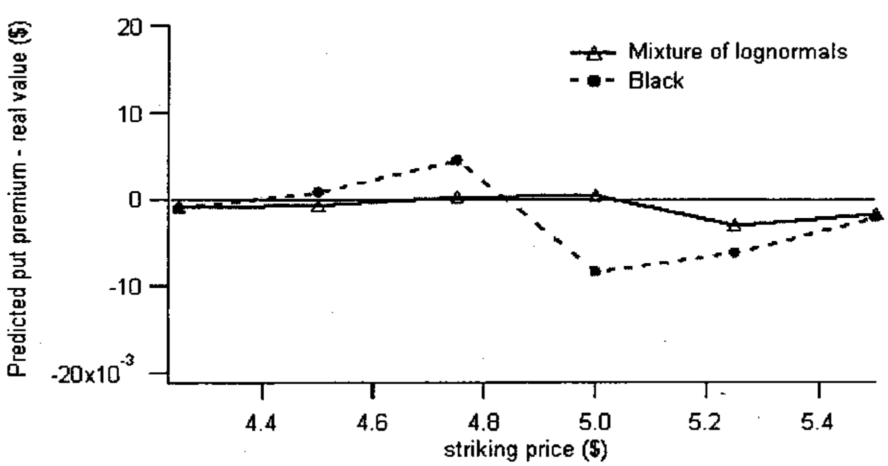
d. Put Pricing Errors of August 16, 2000



e. Call Pricing Errors of October 11, 2000



f. Put pricing Errors of October 11, 2000



Section 6: Corn Estimation Results

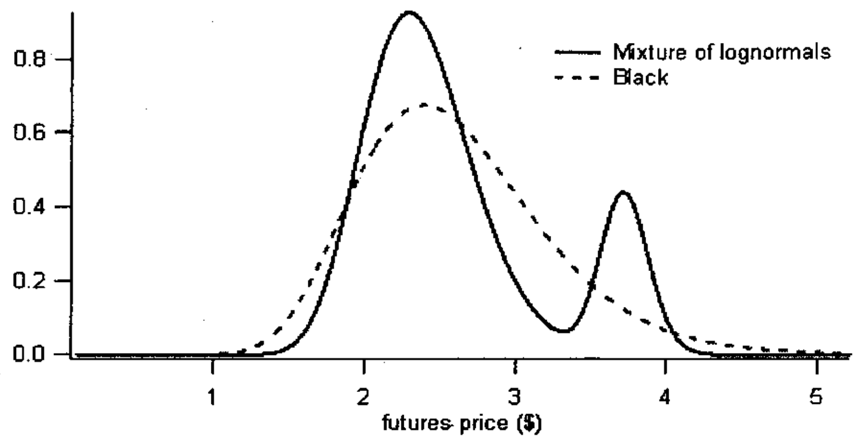
Figure 6.1 and Figure 6.2 depict probability density functions and cumulative density functions taken from both Black and MLN models by using data of May 10, 2000, August 16, 2000, and November 11, 2000 respectively. The top panel uses estimates from the November contract of corn (harvest contract) on May 10, 2000 (194 days to maturity). The middle panel uses estimates of the same contract on August 16, 2000 (96 days to maturity). The bottom panel uses estimates of the same contract on November 15, 2000 (5 days to maturity). Figure 6.3 plots the difference between estimated option prices and actual prices for calls and puts of these three days.

In these graphs, MLN captures key deviations from Black such as a fatter right-hand-side tail, and a mode shifted slightly to the left. A lot of empirical test of commodity price show that variance of commodity price distribution is not constant and commodity price distributions are characterized by high-peak and fat-tail. Thus, the derived distribution from MLN is consistent with the literature.

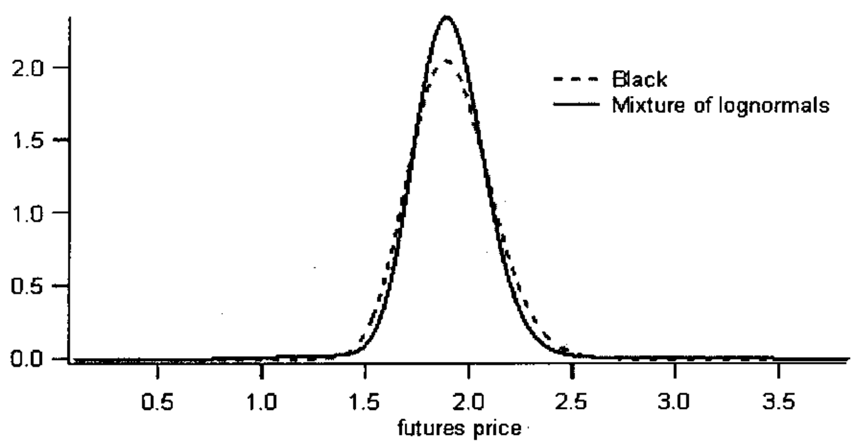
Figure 6.3 depicts call and put option pricing errors. The MLN method tends to give much smaller option pricing errors when the date is far from maturity. When the date is near maturity, there is no significant difference between the option pricing errors from the two methods. When the date is far from maturity, there is more uncertainty in the market. The MLN has a more flexible shape to accommodate the market participants' expectations.

Figure 6.1: Estimated PDFs -- December (harvest) contract of corn

a. December Option Maturity as of May 10, 2000



b. December Option Maturity as of August 16, 2000



c. December Option Maturity as of November 15, 2000

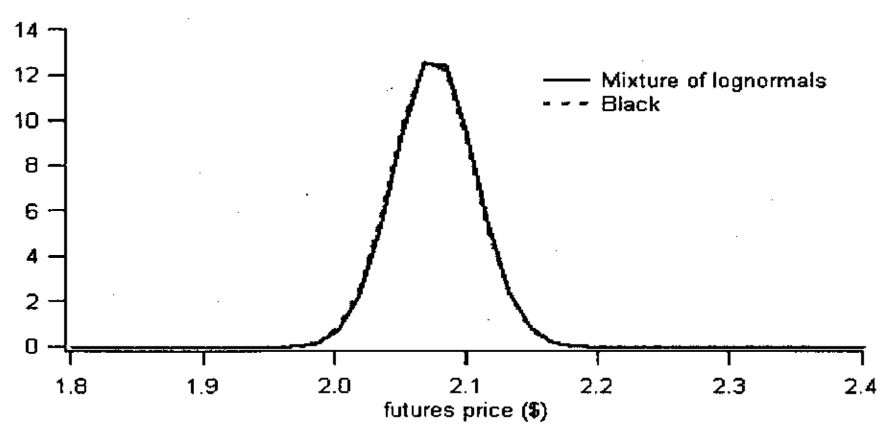
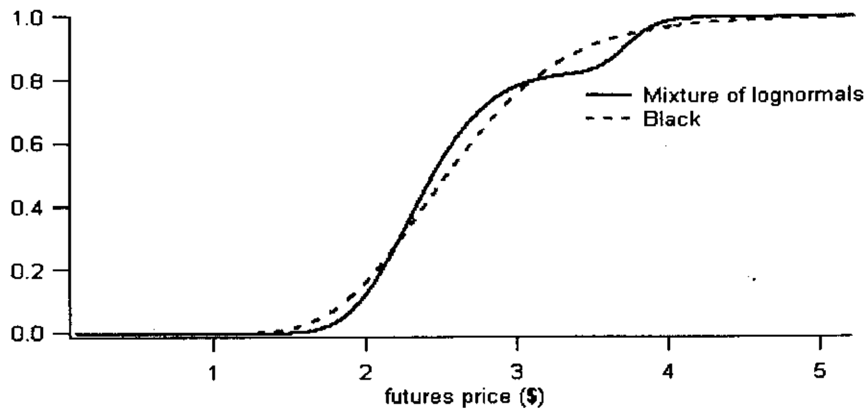
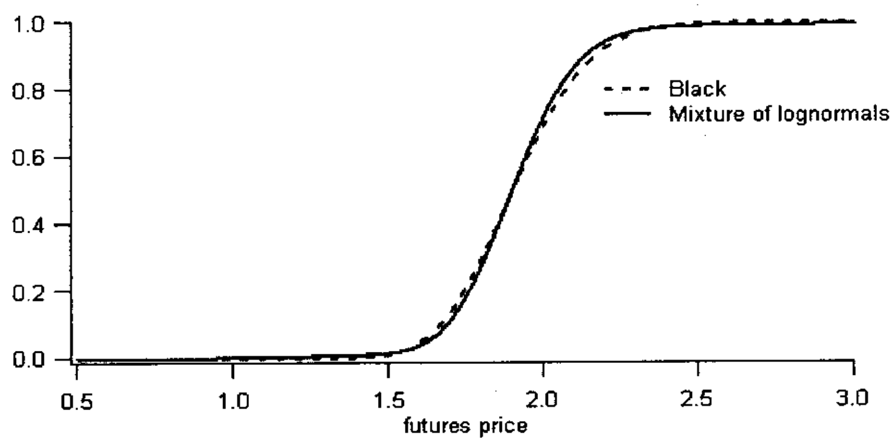


Figure 6.2: Estimated CDFs – December (harvest) contract of corn

a. December Option Maturity as of May 10, 2000



b. December Option Maturity as of August 16, 2000



c. December Option Maturity as of November 15, 2000

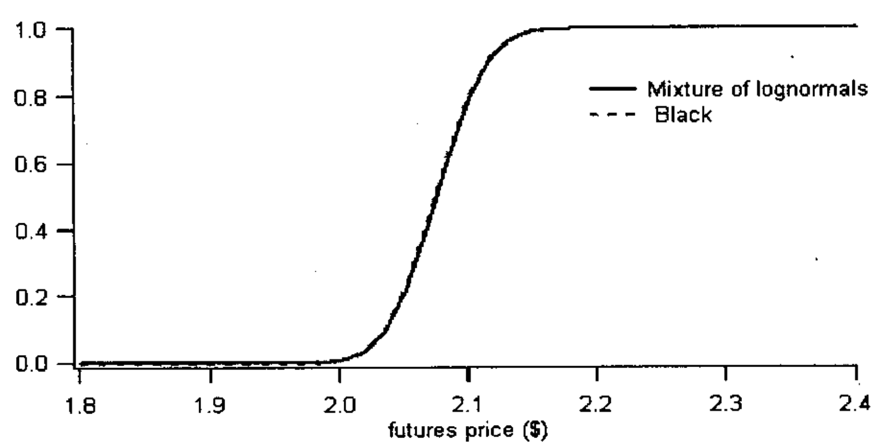
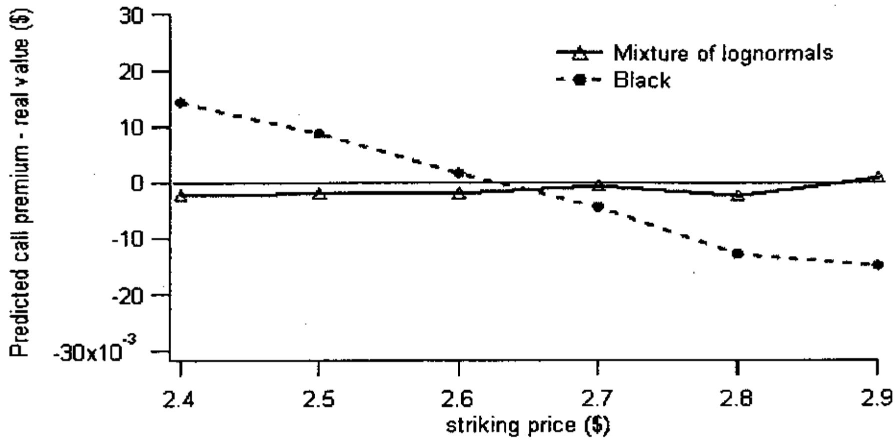
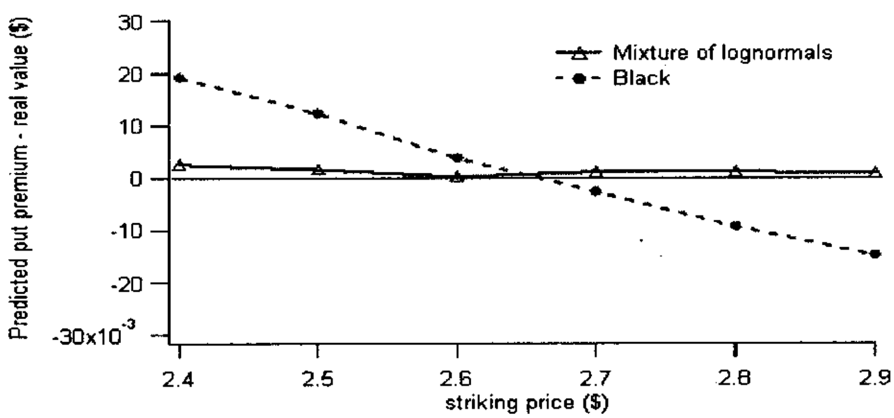


Figure 6.3: Option pricing errors -- December (harvest) contract of corn

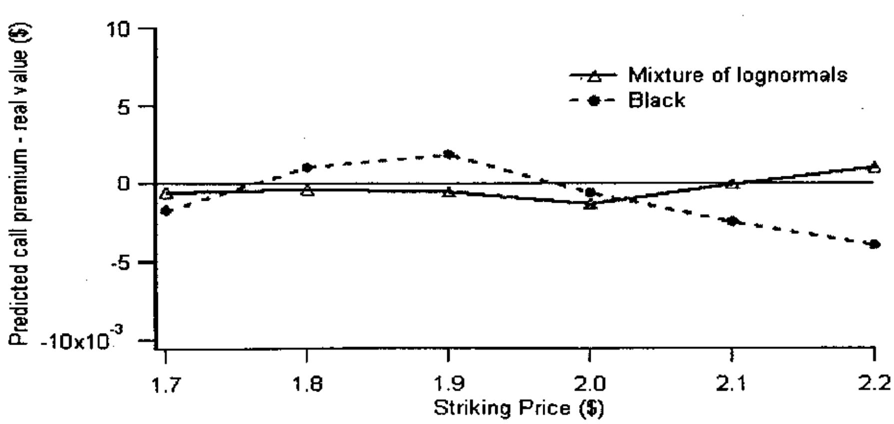
a. Call Pricing Errors of May 10, 2000



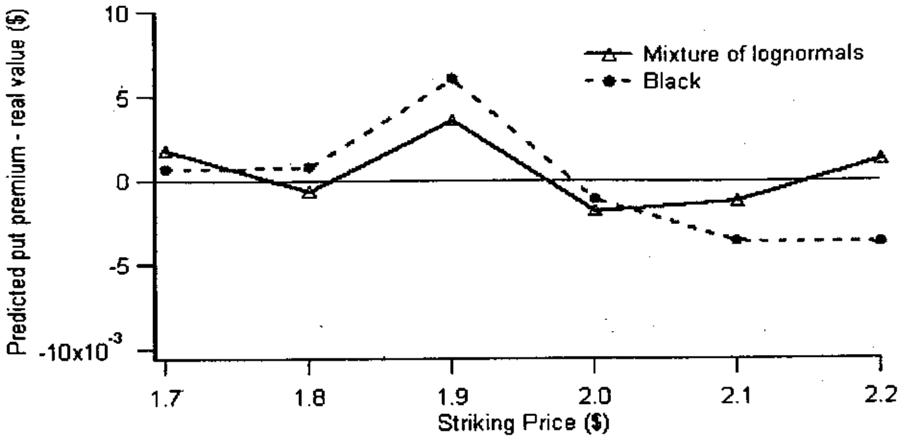
b. Put Pricing Errors of May 10, 2000



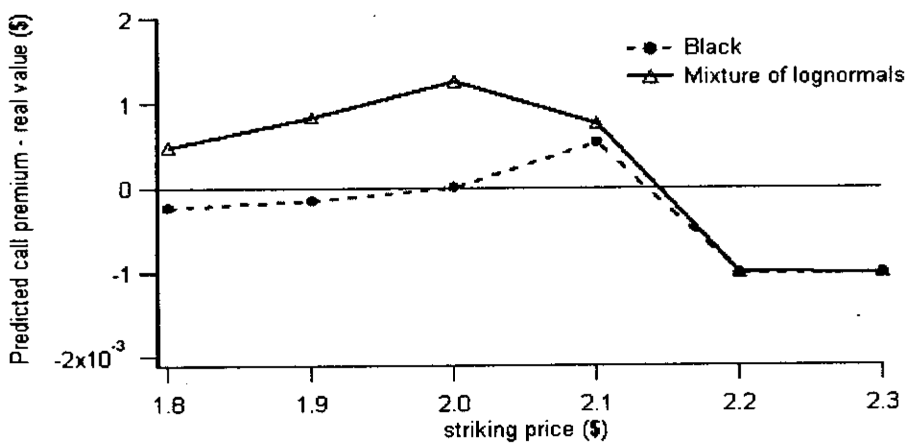
c. Call Pricing Errors of August 16, 2000



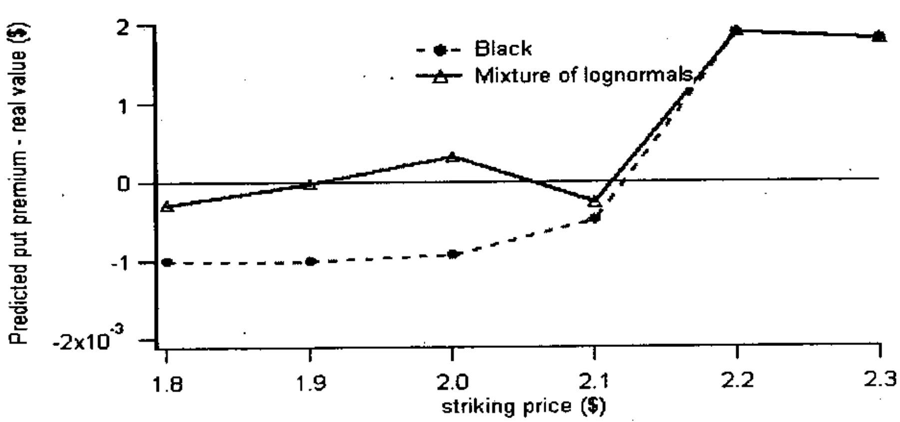
d. Put Pricing Errors of August 16, 2000



e. Call Pricing Errors of November 15, 2000



f. Put pricing Errors of November 15, 2000



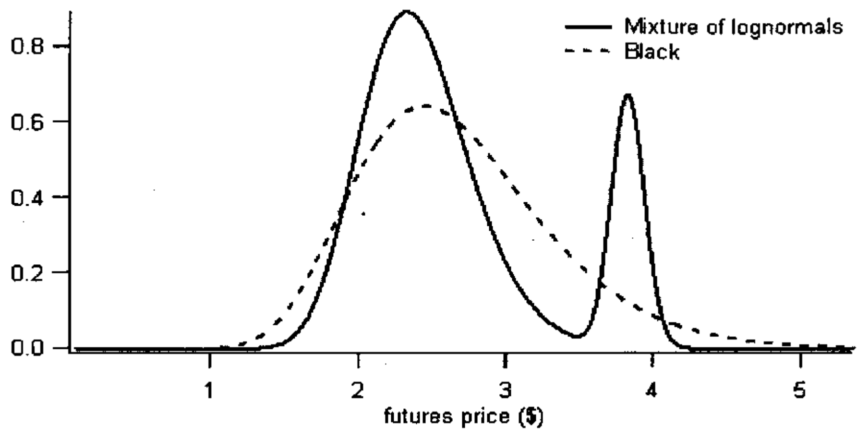
The PDF in Figure 6.1.a turns out as a bi-modal distribution, while Black is unimodal all the time due to the single lognormal assumption. As the data near the maturity date, there is little qualitative difference in the two estimates, while during periods far away from maturity; the estimate from single lognormal method cannot easily accommodate the significant probability mass on the right tail. The bimodal shape could be the result of uncertainty about supply and demand.

To investigate further, Figure 6.5 and Figure 6.6 show three consecutive week's PDFs and CDFs of corn December (2000) contract in the month of May. A trend can be observed when time goes on. The right tail becomes smaller and smaller. The pattern of the graphs could reflect a change in the market situation. When the date is far away from maturity, there is a lot of uncertainty in the market. There might be different expectations of outcomes in the market. The mixture lognormal distribution can accommodate different expectation of market participants. It can also easily accommodate a single lognormal distribution if that would best fit the data. We expect as news hit the market, the relative weighting of the two lognormals might change, as well as the parameters of each of the two lognormals.

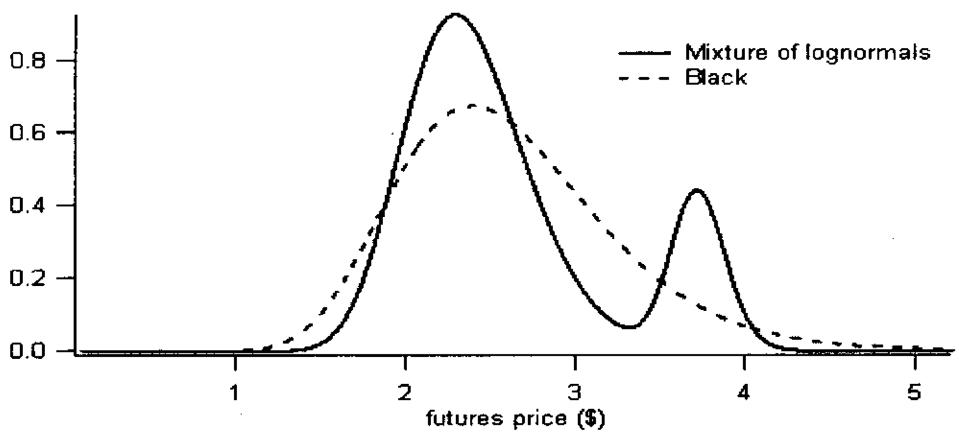
A hypothesis for the following patterns is that there was some news about weather situation, demand or supply conditions. For example, there was an expectation of drought in the beginning of May 2000; thus, there was an expectation of high corn price at maturity if the drought was realized. As May progressed, there was rain and expectation of high corn price became small.

Figure 6.4 Estimated PDFs – December (harvest) contract of corn – 2000

a. December Option Maturity as of May 3, 2000



b. December Option Maturity as of May 10, 2000



c. December Option Maturity as of May 24, 2000

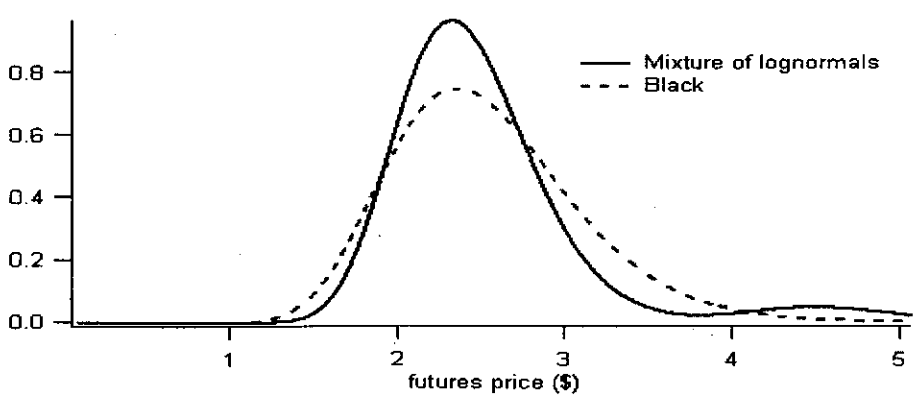
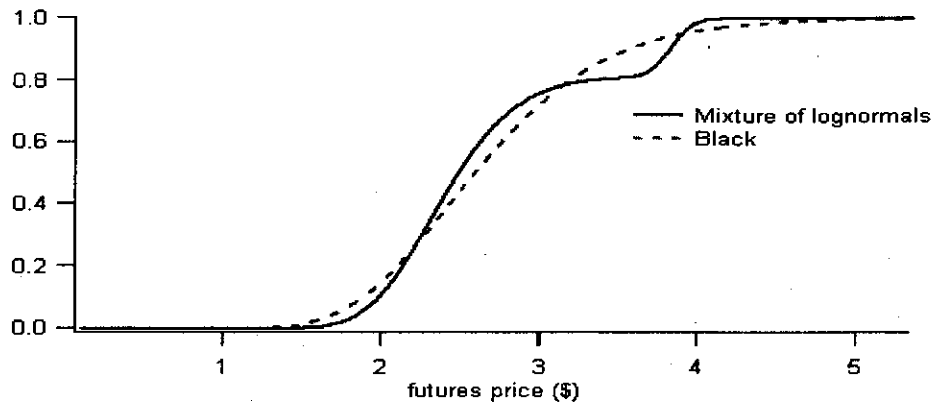
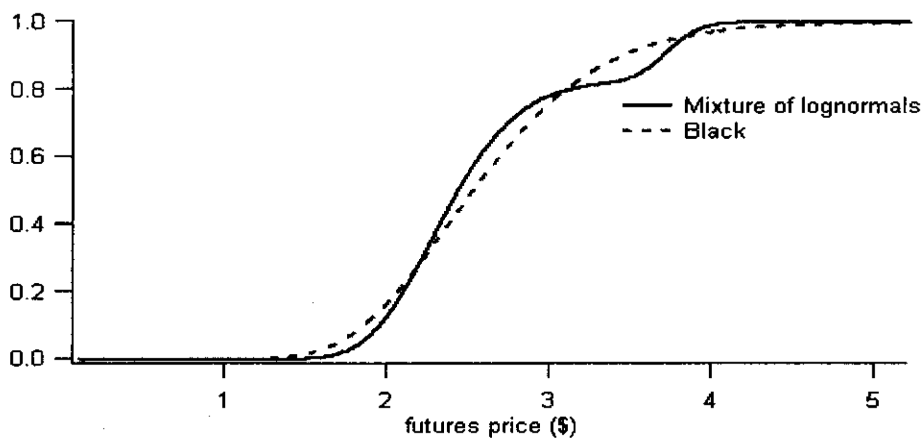


Figure 6.5 Estimated CDFs – December (harvest) contract of corn – 2000

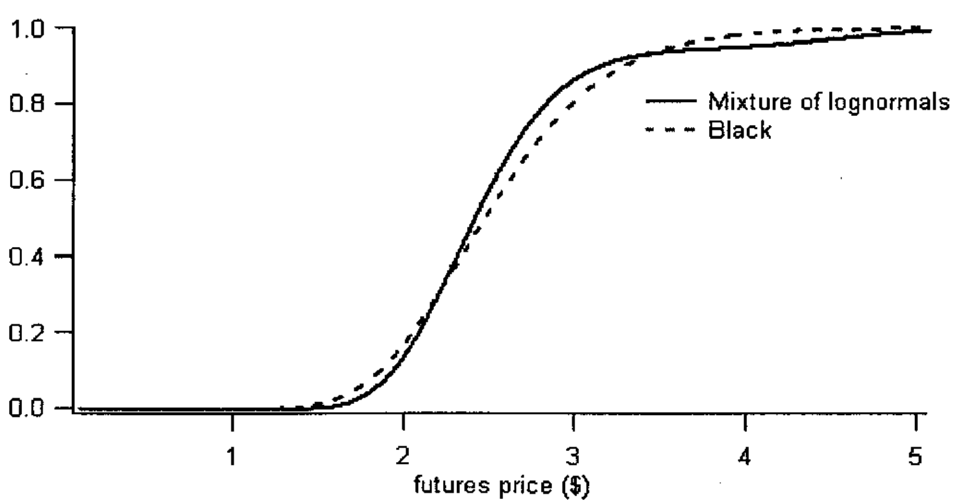
a. December Option Maturity as of May 3, 2000



b. December Option Maturity as of May 10, 2000



c. December Option maturity as of May 23, 2000

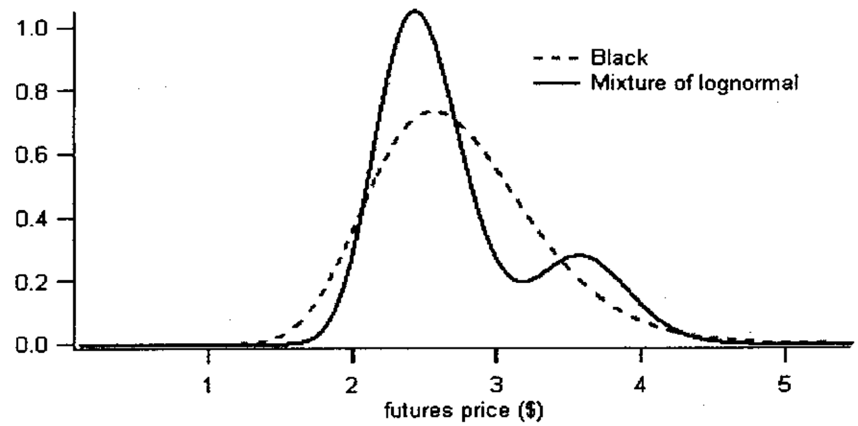


Figures 6.7- 6.8 and Figures 6.9- 6.10 show three consecutive week's PDFs and CDFs of December contract of corn in the year of 1998 and year 1999 respectively. Figure 6.7 and 6.8 depict probability density functions and cumulative density function taken from both Black and MLN models by using data of May 6, 1998, May 13, 1998 and May 20, 1998 respectively. Figure 6.9 and 6.10 depict probability density functions and cumulative density function taken from both Black and MLN models by using data of May 5, 1999, May 12, 1999 and May 19, 1999 respectively.

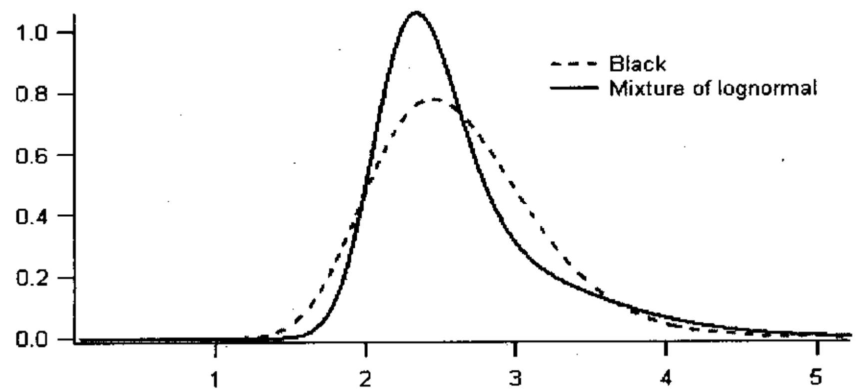
The PDF in Figure 6.7.a turns out as a bi-modal distribution, while this bi-modal pattern doesn't apply to the other graphs in the year of 1998 and year 1999. The results suggest that the market perceptions of the harvest futures price distribution are quite different during the three weeks in the year 1998 and 1999. Again, this may be due to the news about weather or supply and demand situations during that period. More analysis is needed to test these hypothesis.

Figure 6.6 Estimated PDFs – December (harvest) contract of corn -- 1998

a. December Option Maturity as of May 6, 1998



b. December Option Maturity as of May 13, 1998



c. December Option Maturity as of May 20, 1998

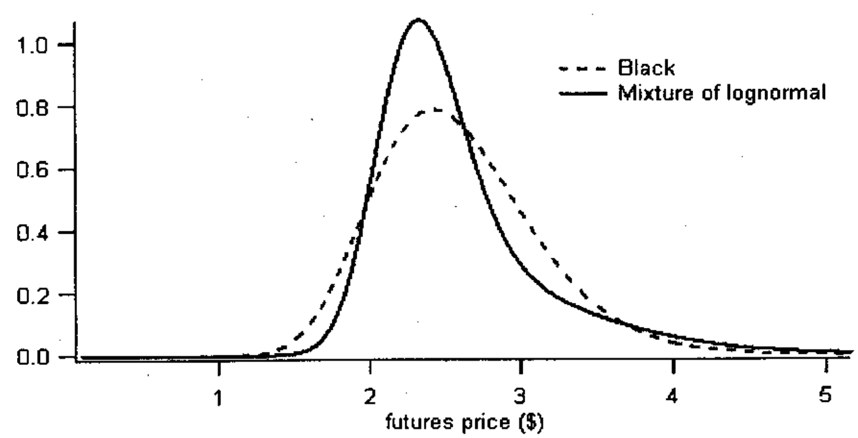
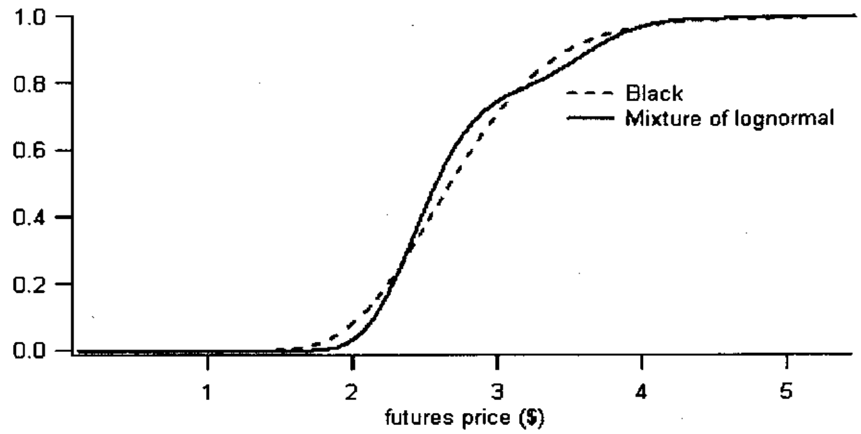
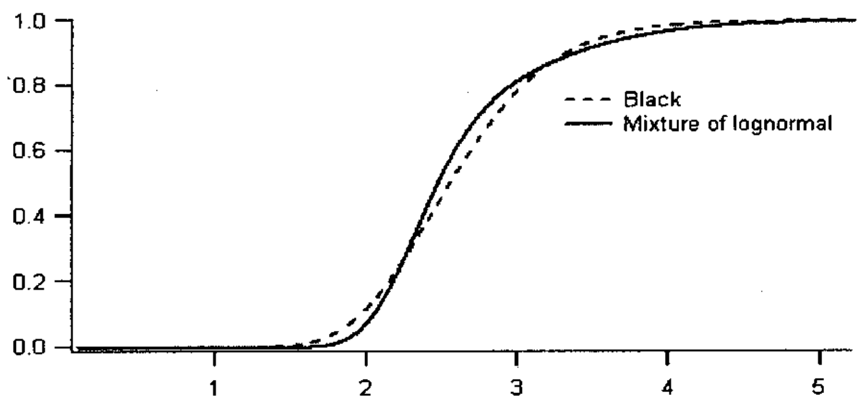


Figure 6.7 Estimated CDFs – December (harvest) contract of corn -- 1998

a. December Option Maturity as of May 6, 1998



b. December Option Maturity as of May 13, 1998



c. December Option Maturity as of May 20, 1998

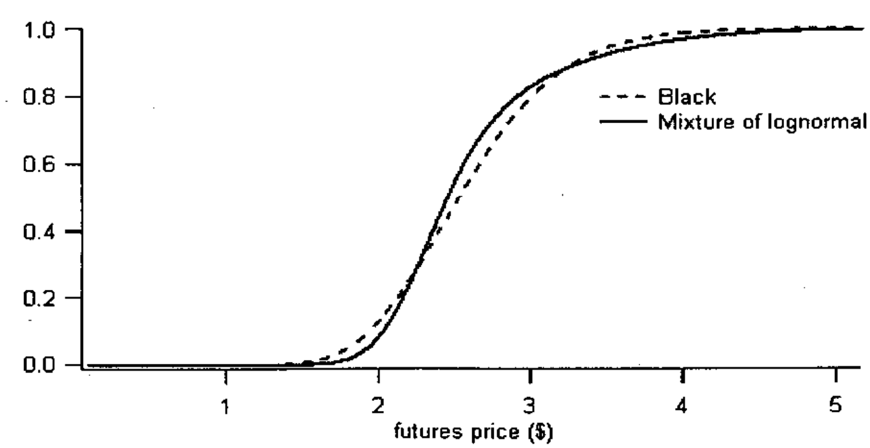
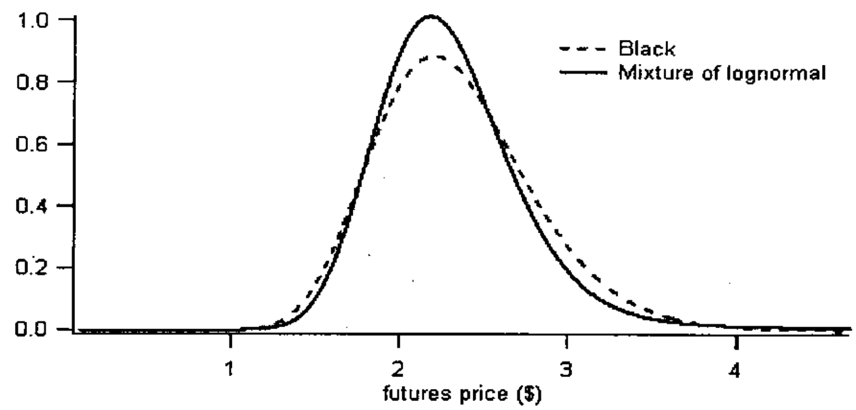
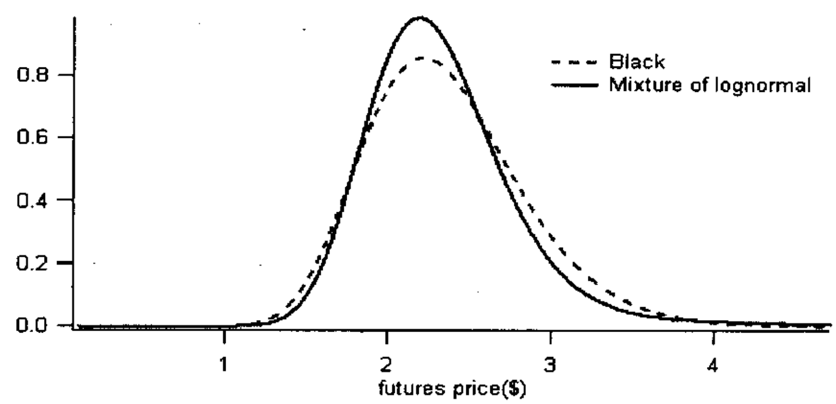


Figure 6.8: Estimated PDFs – December (harvest) contract of corn -- 1999

a. December Option Maturity as of May 5, 1999



b. December Option Maturity as of May 12, 1999



c. December Option Maturity as of May 19, 1999

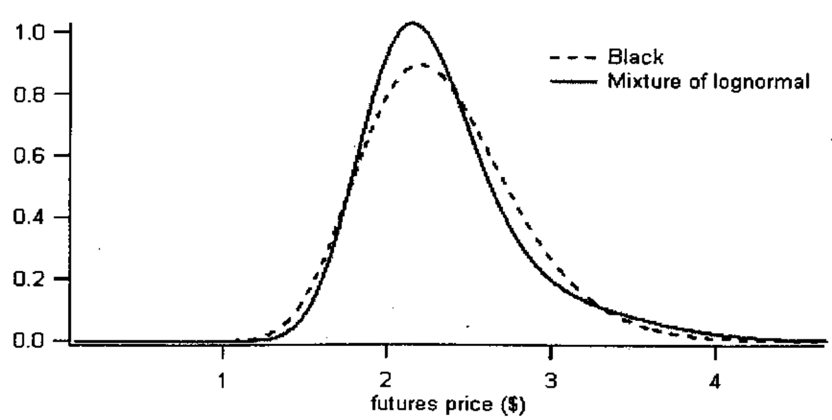
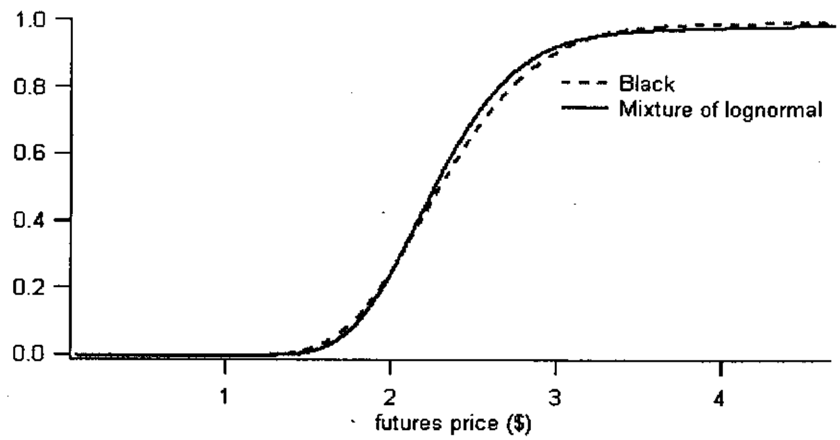
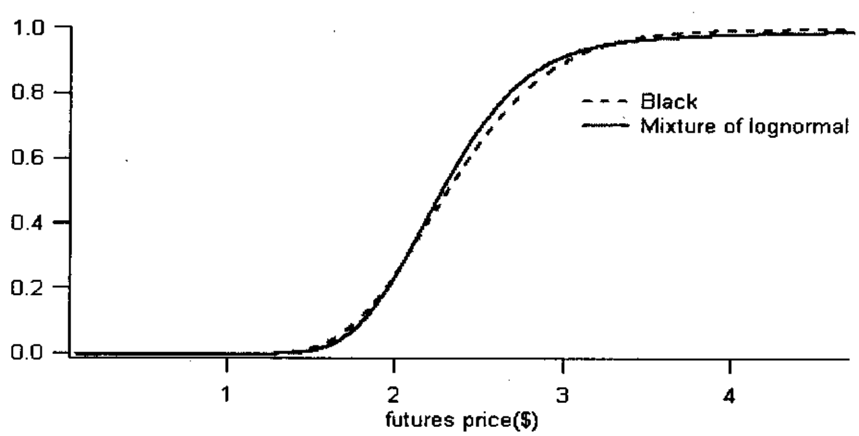


Figure 6.9: Estimated CDFs – December (harvest) contract of corn -- 1999

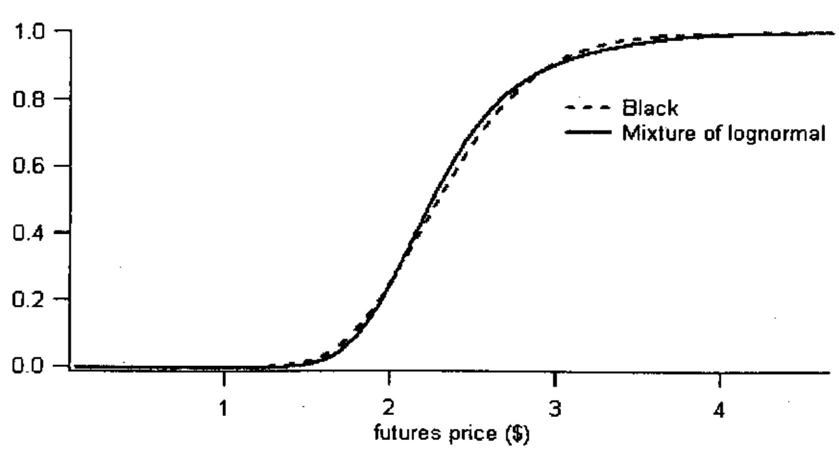
a. December Option Maturity as of May 5, 1999



b. December Option Maturity as of May 12, 1999



c. December Option Maturity as of May 19, 1999



Section 7: Further Comparisons of Enlarged graphs of Black and MLN Models

From Figure 6.1, it is not easy to tell the difference of tails intuitively, especially for those distributions that are near maturity. Figure 7.11 is an enlarged PDF for the December contract of corn on August 16, 2000. It gives intuitive sense of the difference of tails between MLN and Black methods. The numerical integration[13] results in Table 6.4 give a better explanation about the difference of Black and MLN methods.

The reason that MLN's PDF has a larger standard error can be illustrated in Figure 7.11 From the figures, it is obvious that around 1.4 and 2.6, the two PDFs have the same density value. Numerical integration can be done region by region to examine their individual contribution.

From Table 7.4, the major reason why MLN has a larger standard error is that it has a longer and thicker tail (from 2.6 to 38). This can also be shown by the kurtosis values. MLN has a much larger kurtosis (0.56127) than that of the Black's (0.262851). This means MLN has a much larger tail and this affects the standard error in a larger degree than what is seen intuitively from the graph of PDF.

---

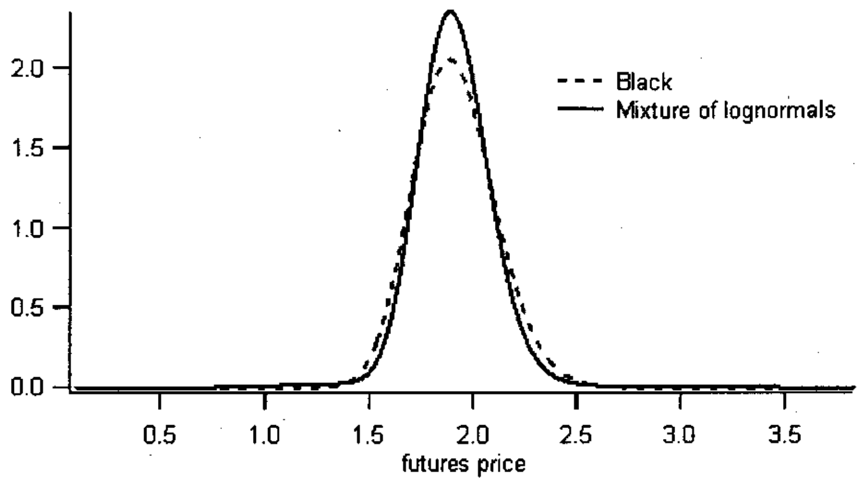
[13] In order to calculate (1) and (2) numerically, the approximate equations for expected value and variance are respectively: $\bar{x} \approx \sum_i x_i \cdot g(x_i) \cdot \Delta x_i$; $Var \approx \sum_i (x_i - \bar{x})^2 \cdot g(x_i) \cdot \Delta x_i$

where x is the futures price and g(x) is the PDF of the futures price derived from both MLN and Black model. Δx_i is the small step used to divide the integration region, and x_i is the value of x at grid point i.

Numerical integration have been done from 0.0001 to 37.5787 with the step size of 0.00037578. Here numerical integration range is not from 0 to positive infinity, because outside of the calculation region, the PDF's value is almost zero and that contribution to the integration can be neglected. For MLN's PDF, the calculated expected value is 1.91, and the standard error is 0.22972. For Black's PDF, the calculated expected value is 1.91, and the standard error is 0.19696. This is almost the same as those from the analytical results presented in Table 4.3.

Figure 7.1: Enlarged PDF of Black and MLN models – December contract of corn

a. December Option maturity as of August 16, 2000



b. Enlarged PDF of the same PDF as above

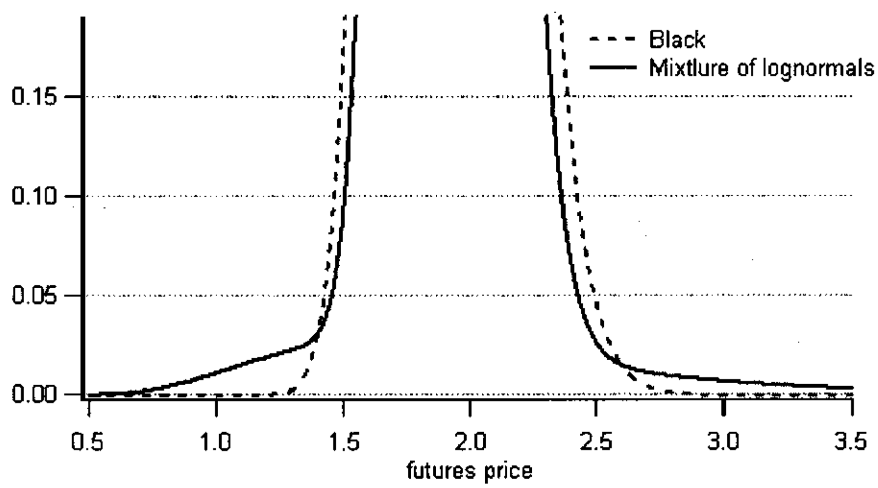


Table 7.1: Contribution to variance by different range of futures price – by Black and MLN models

	Integration from .0001 to 1.4	Integration from 1.4 to 1.91	Integration from 1.91 to 2.6	Integration from 2.6 to 38	Total Variance	Standard Error
MLN	0.0059713	0.013848	0.015066	0.017886	0.0527713	0.22972
Black	0.00045408	0.017353	0.020308	0.00067893	0.03879401	0.19696
MLN-Black	0.00551722	-0.003505	-0.005242	0.01720707	0.01397729	0.03276

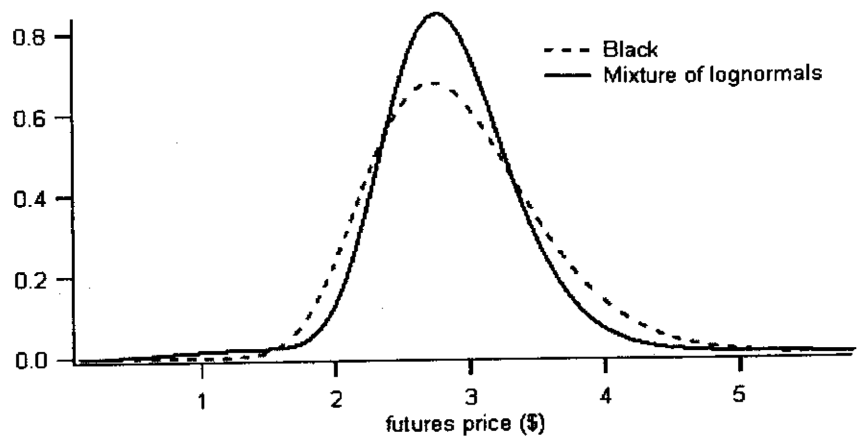
Section 8: Wheat Estimation Results

Figure 8.1 and Figure 8.2 depict probability density functions and cumulative density functions taken from both Black and MLN models by using data of October 13, 1999, February 16, 2000, and June 14, 2000 respectively. The top panel uses estimates from the November contract of soybean (harvest contract) on October 13, 1999 (251 days to maturity). The middle panel uses estimates of the same contract on February 16, 2000 (125 days to maturity). The bottom panel uses estimates of the same contract on June 14, 2000 (6 days to maturity). Figure 8.3 plots the difference between estimated option prices and actual prices for calls and puts of these three days.

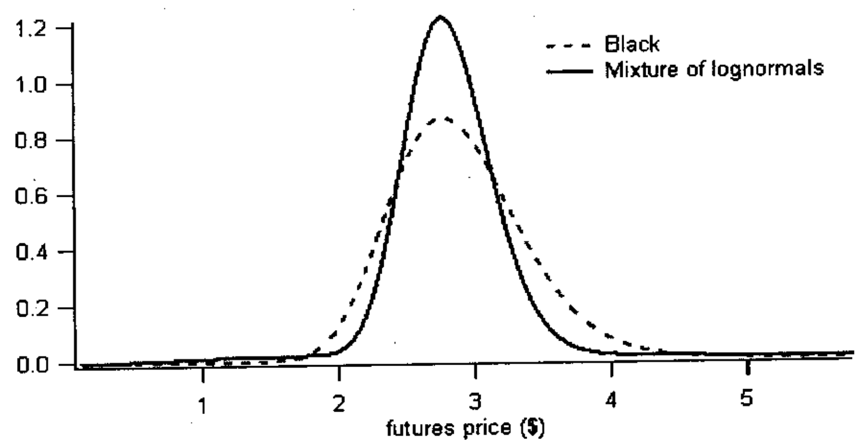
The derived distributions of wheat futures prices have the same characteristics as those of soybean and corn. PDFs for wheat are characterized by high peak and fat-tail. The option pricing errors from MLN methods are much smaller than that from Black model. The differences of these two models are obvious for those days that are far from maturity.

Figure 8.1: Estimated PDFs -- July (harvest) contract of wheat

a. July Option Maturity as of October 13, 1999



b. July Option Maturity as of February 16, 2000



c. July Option Maturity as of June 14, 2000

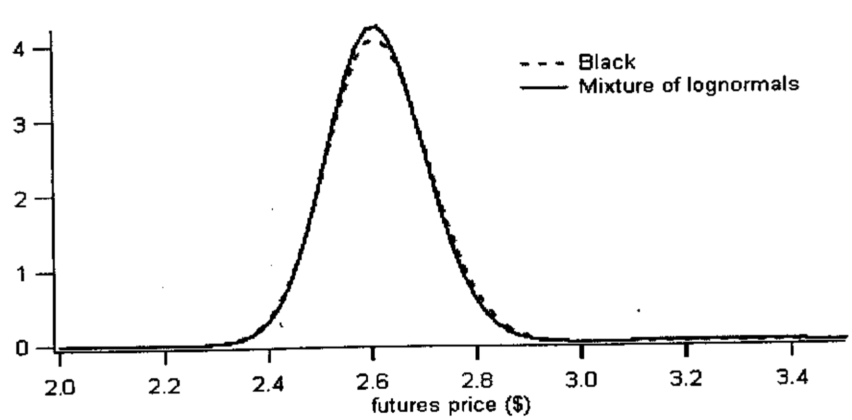
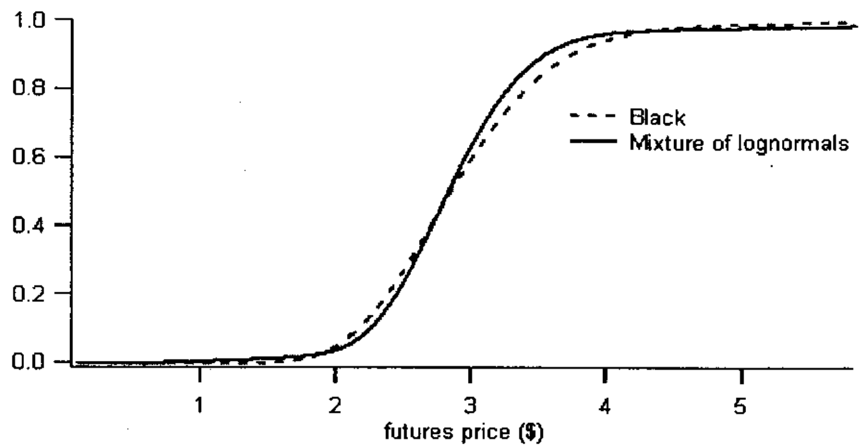
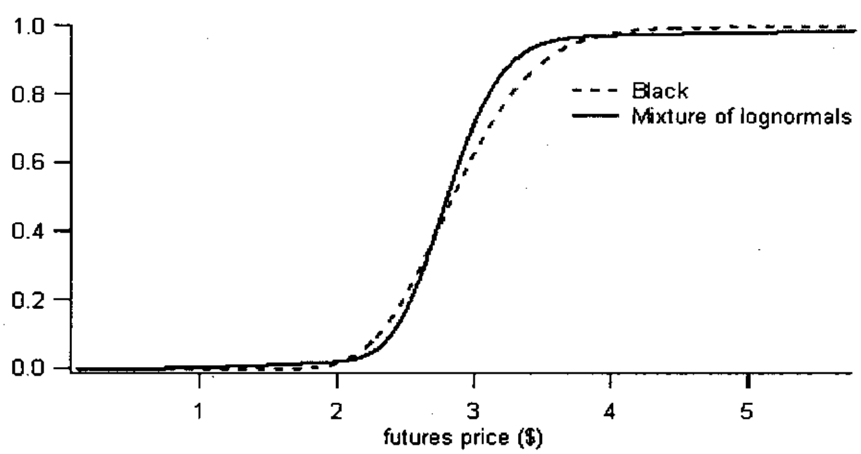


Figure 8.2: Estimated CDFs – July (harvest) contract of wheat

a. July Option Maturity as of October 13, 2000



b. July Option Maturity as of February 16, 2000



c. July Option Maturity as of June 14, 2000

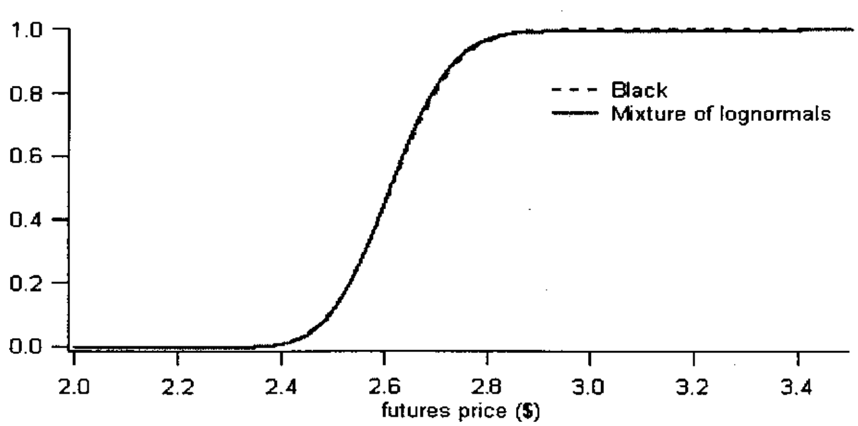
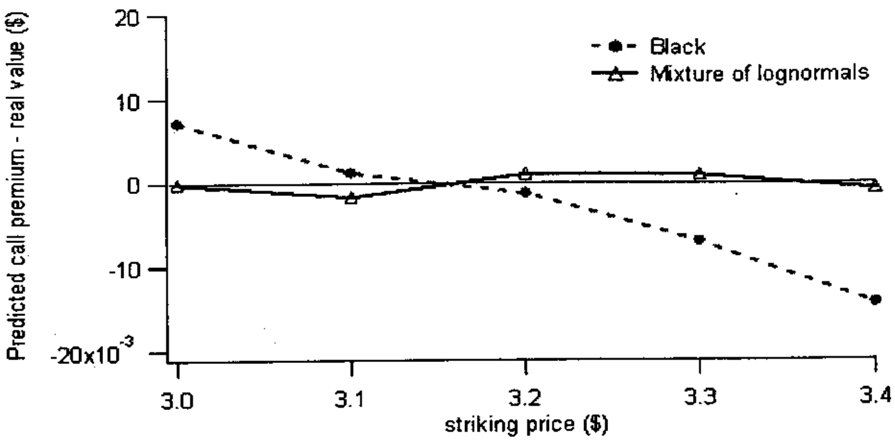
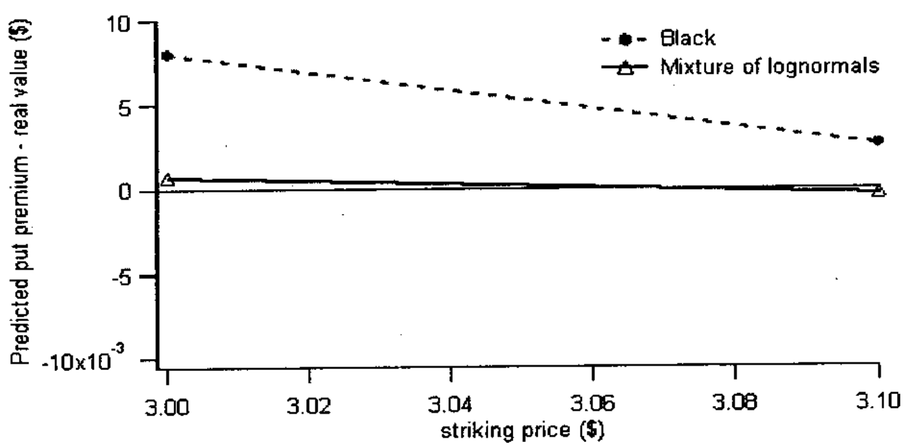


Figure 8.3: Option pricing errors – July (harvest) contract of wheat

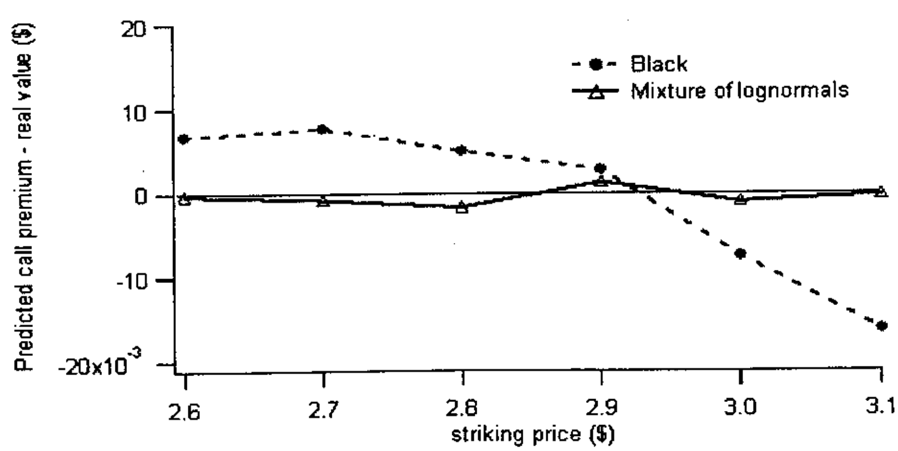
a. Call Pricing Errors of October 13, 2000



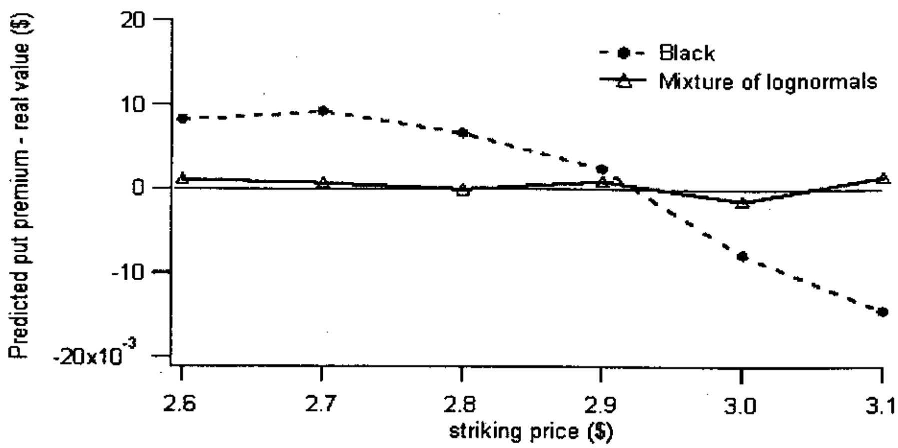
b. Put Pricing Errors of October 13, 2000



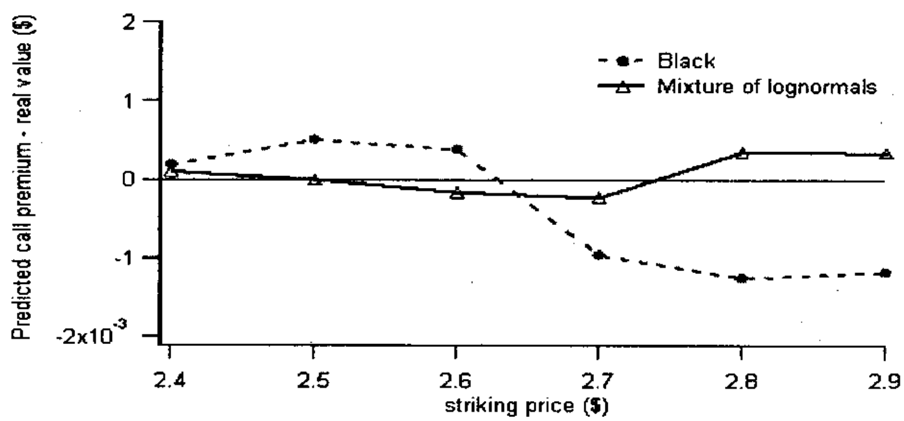
c. Call Pricing Errors of February 16, 2000



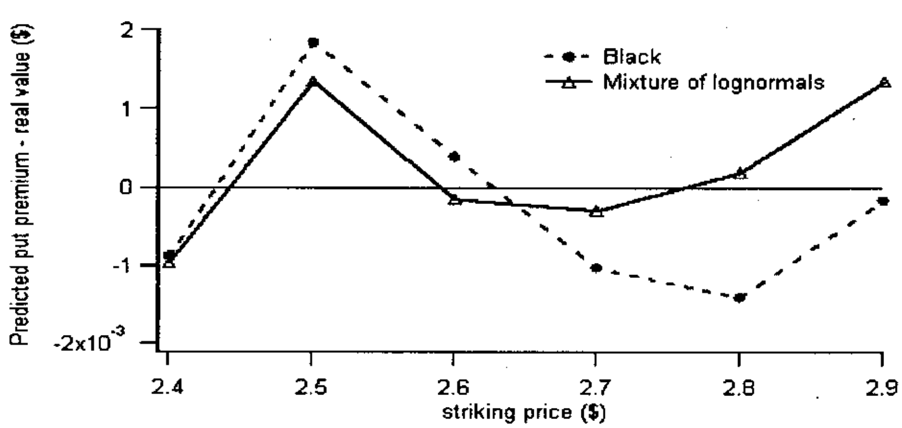
d. Put Pricing Errors of February 16, 2000



e. Call Pricing Errors of June 14, 2000



f. Put pricing Errors of June 14, 2000



Section 9: Conclusion

9.1 Comparison of the two models

Compared with the single lognormal distribution, the mixture of two lognormal distributions clearly reflects some additional information embedded in option data. The empirical results show that MLN model has a stronger ability to account for the market information of soybean, corn and wheat futures prices.

Table 9.3 reports the mean, standard deviation, skewness, and kurtosis of corn, soybean, wheat prices and mean square deviation from objective function. The mean of the risk-neutral distribution always equals the current futures price. From the estimated results above, the standard errors, skewness and kurtosis from MLN methods are always greater than that from Black model.

Skewness reflects simply a distribution's asymmetry. Positive skewness indicates that the right tail is "fatter" (high probability) and usually extends out further than the left tail; large positive realizations are more likely than large negative realizations. Kurtosis is a measure of the thickness of the tails of the distribution. Table 9.3 further suggests the skewness from commodity price distribution is often systematically one-sided. And the bigger kurtosis of distribution of MLN is an indication of fatter tail than the Black's.

Mean square deviation is the minimum value of the objective function defined in Section 3. For those days that are far from maturity, the differences between these two methods are large.

Table 9.1: Moments of the implied distribution and mean square deviation of objective function for all contracts/days[14]

	Expected Value	Standard Error	Skewness	Kurtosis	Mean Square Deviation
Soybean November contract—harvest contract					
05/24/00					
Black	5.520	1.099	0.666	1.520	8.067e-004
MLN	5.520	1.294	2.754	2.250	2.696e-006
08/16/00					
Black	4.744	0.404	0.103	0.537	3.659e-005
MLN	4.744	0.599	9.781	3.032	7.164e-006
10/11/00					
Black	4.780	0.188	0.022	0.248	2.071e-005
MLN	4.780	0.197	0.187	0.294	1.735e-005
Soybean July contract –storage contract					
10/13/99					
Black	5.290	1.006	0.581	1.386	1.742e-004
MLN	5.290	1.105	1.720	1.807	1.686e-006
03/15/00					
Black	5.3575	0.781	0.344	1.056	5.304e-004
MLN	5.3575	0.823	1.110	1.277	2.916e-006
06/14/00					
Black	5.1075	0.213	0.027	0.281	1.413e-005
MLN	5.1075	0.219	0.086	0.297	1.292e-006
Corn December contract—harvest contract					
05/10/00					
Black	2.606	0.635	0.474	0.899	1.284e-004
MLN	2.606	0.615	0.498	0.764	2.697e-006
08/16/00					
Black	1.910	0.197	0.061	0.263	7.909e-006
MLN	1.910	0.230	0.610	0.561	2.230e-006
11/15/00					
Black	2.074	0.031	1.39e-003	0.041	1.028e-006
MLN	2.074	0.043	-0.296	0.138	1.025e-006

[14] The third moment of the distribution, or its skewness is defined mathematically as $E[(F_T-\mu)^3]$, where μ is the mean of the distribution and F_T is the futures price at maturity. The fourth moment of the distribution, or its kurtosis is defined as $E[(F_T-\mu)^4]$.

Corn July contract –storage contract					
11/17/99					
Black	2.185	0.396	0.218	0.543	7.891e-005
MLN	2.185	0.409	0.545	0.633	8.579e-007
03/15/00					
Black	2.445	0.391	0.193	0.538	1.719e-004
MLN	2.445	0.422	0.633	0.666	1.159e-006
06/14/00					
Black	2.0325	0.113	0.019	0.150	3.984e-005
MLN	2.0325	0.179	1.302	0.732	2.338e-005
Wheat July contract—harvest contract					
10/13/00					
Black	2.91	0.628	0.413	0.875	5.420e-005
MLN	2.91	0.852	5.687	2.797	8.456e-007
02/16/00					
Black	2.88	0.478	0.241	0.652	7.745e-005
MLN	2.88	0.844	7.864	3.104	1.063e-006
06/14/00					
Black	2.6125	0.098	0.011	0.129	9.718e-007
MLN	2.6125	0.103	0.135	0.191	4.191e-007
Wheat March contract –storage contract					
06/14/00					
Black	3.020	0.640	0.413	0.890	2.511e-004
MLN	3.020	0.723	1.622	1.326	6.924e-007
11/29/00					
Black	2.696	0.272	0.083	0.363	2.605e-005
MLN	2.696	0.282	0.255	0.408	2.139e-006
02/14/01					
Black	2.675	2.40e-003	6.36e-006	3.13e-003	4.470e-005
MLN	2.675	0.022	-0.258	0.082	4.466e-005

9.2 Conclusion

This paper develops a method for using options prices to estimate the market's probability distribution for the underlying asset's price. The mixture of two lognormal method developed in this paper overcomes the deficiencies in standard Black option pricing model for commodity futures option. Black's model assumes proportional changes in the underlying futures price follow identical lognormal distribution. Yet empirical evidence suggests that variance of commodity futures price are not constant and the tails of empirical price distributions appear to be much fatter than the lognormal benchmark, indicating excess kurtosis in commodity price changes.

Since deriving price distributions has a strong motivation from both theoretical and practical perspective, it is very important to derive a option pricing formula with a more realistic distribution assumption which can be used to back out the implied distribution of commodity prices.

The mixture of lognormals method is quite general, allowing the standard lognormal distribution to be replaced by any distribution from within a wide class. As the focus is only on the asset's distribution, minimum structure is placed on the stochastic process governing movements in the asset price over time. This lack of structure is appealing since we generally have little priori information about the stochastic process that market participants have assumes.

In the application to the soybean, corn and wheat futures market, MLN model is able to capture more information embedded in option data. Empirical results show that MLN model accounts for the true properties of commodity price distribution, which characterized by peaked, fat-tailed, and frequently positively skewed patten relative to a Gusssian density.

The mean square deviation generated from MLN is always smaller than those from Black's model, which indicates smaller option pricing errors by MLN method. Black model frequently overprices in-the-money call and out-of-the-money put options, and underprice out-of-the-money call and in-the-money puts options. MLN formula overcomes this deficiency by reducing the levels of errors in option pricing. The in-sample option pricing errors from MLN are usually at a level of $\$1*10^{-3}$ or even smaller, which has a better performance than that from Black model.

Extension of this research could go in a number of directions. First, statistical comparison of Black's model and mixture of lognormal distribution could be conducted to test whether the results from MLN is statistically significant. Second, out of sample forecast of option prices could be conducted by using the estimated parameters to check whether the results form MLN are stable.

Appendix I: Derivation of formulas for MLN method

1.1 Derivation of the density function for lognormal distribution

Theorem: $G(\ln F_T)d(\ln F_T) = g(F_T)dF_T$ (1.1.1)

Here $G(\cdot)$ is the probability density function of $\ln F_T$; $g(\cdot)$ is the probability density function of F_T.

Thus, a relationship equation can be derived,

$$
\begin{aligned}
&G(\ln F_T)d(\ln F_T) = g(F_T)dF_T \\
&\Rightarrow \frac{1}{F_T}G(\ln F_T)dF_T = g(F_T)dF_T \qquad (1.1.2)\\
&\Rightarrow \frac{1}{F_T}G(\ln F_T) = g(F_T)
\end{aligned}
$$

This relationship equation is applied in the case of normal and lognormal distributions. Assume $\ln F_T$ follows normal distribution; and F_T follows lognormal distribution. (2.1.1) can be expanded as

$$
\begin{aligned}
&G(\ln F_T)d\ln F_T = N(\mu,\sigma)d\ln F_T = N(\mu,\sigma)\frac{1}{F_T}dF_T \\
&= \frac{1}{F_T}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\ln F_T-\mu)^2}{2\sigma^2}}dF_T \qquad (1.1.3)\\
&= g(F_T)dF_T
\end{aligned}
$$

Thus, $g(F_T) = \frac{1}{F_T}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\ln F_T-\mu)^2}{2\sigma^2}}$ (1.1.4)

This is the density function for single lognormal distribution. Thus, the density function for mixture of two lognormal distributions is

$$g(F_T)=\frac{1}{F_T}(\lambda\frac{1}{\sqrt{2\pi}\cdot\sigma_1}e^{\frac{(\ln F_T-\mu_1)}{2\sigma_1}}(1-\lambda)\frac{1}{\sqrt{2\pi}\cdot\sigma_2}e^{\frac{(\ln F_T-\mu_2)}{2\sigma_2}})$$

$$=\lambda g_1(F_T)+(1-\lambda)g_2(F_T) \tag{1.1.5}$$

1.2 Derivation of call option pricing formula for mixture of two lognormal distributions

The option prices for call is

$$P_{call}=e^{-r(T-t)}\int_K^{+\infty}(F_T-K)g(F_T)dF_T$$

$$=e^{-r(T-t)}[\lambda\int_K^{+\infty}(F_T-K)g_1(F_T)dF_T+(1-\lambda)\int_K^{+\infty}(F_T-K)g_2(F_T)dF_T] \tag{1.2.1}$$

Thus $P_{call}=\lambda P_{call_1}+(1-\lambda)P_{call_2}$ (1.2.2)

$$P_{call_i}=e^{-r(T-t)}\int_K^{+\infty}(F_T-K)g_i(F_T)dF_T \tag{1.2.3}$$

The following steps derive the formula for $P_{call\,i}$

$$P_{call}=e^{-r(T-t)}\int_K^{+\infty}(F_T-K)g(F_T)dF_T$$

$$=e^{-r(T-t)}[\int_K^{+\infty}F_Tg(F_T)dF_T-K\int_K^{+\infty}g(F_T)dF_T] \tag{1.2.4}$$

where $g(F_T)=\frac{1}{\sqrt{2\pi}\cdot\sigma\cdot F_T}e^{-\frac{(\ln F_T-\mu)^2}{2\sigma^2}}$

It contains two integrations. The calculation will be done one by one.

$$\int_K^{+\infty} F_T g(F_T) dF_T$$

$$= \int_K^{+\infty} F_T \cdot \frac{1}{F_T} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln F_T - \mu)^2}{2\sigma^2}} dF_T \qquad (1.2.5)$$

Let $x = lnF_T$, then $e^x dx = dF_T$

Thus, (1.2.5) is equal to

$$= \int_{\ln K}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= e^{\frac{(\mu+\sigma^2)^2-\mu^2}{2\sigma^2}} \int_{\ln K}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-(\mu+\sigma^2))^2}{2\sigma^2}} dx \qquad (1.2.6)$$

Let $y = \frac{x-(\mu+\sigma^2)}{\sigma}$, then (1.2.6) is equal to

$$= e^{(\mu+\frac{\sigma^2}{2})} \int_{\frac{\ln K-(\mu+\sigma^2)}{\sigma}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

$$= e^{(\mu+\frac{\sigma^2}{2})} \int_{-\infty}^{\frac{-\ln K+(\mu+\sigma^2)}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \qquad (1.2.7)$$

$$= e^{(\mu+\frac{\sigma^2}{2})} \cdot cdfn(\frac{-\ln K + \mu + \sigma^2}{\sigma})$$

Next go to calculate the second integration.

$$
\begin{aligned}
& K\int_{K}^{+\infty} g(F_T)dF_T \\
&= K\int_{K}^{+\infty} \frac{1}{F_T}\cdot\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln F_T-\mu)^2}{2\sigma^2}} dF_T \\
&= K\int_{K}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln F_T-\mu)^2}{2\sigma^2}} d\ln F_T
\end{aligned}
\qquad (1.2.8)
$$

First use x = lnF_T, then (1.2.8) is equal to

$$
= K\int_{\ln K}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} dx \qquad (1.2.9)
$$

Then we use $y = \frac{x-\mu}{\sigma}$. Thus (1.2.9) is equal to

$$
\begin{aligned}
&= K\int_{\frac{\ln K-\mu}{\sigma}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\
&= K\int_{-\infty}^{\frac{-\ln K+\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\
&= K\cdot cdfn(\frac{-\ln K+\mu}{\sigma})
\end{aligned}
\qquad (1.2.10)
$$

Therefore, the expected call option price under single lognormal distribution could be written as

$$
P_{call} = e^{-r(T-t)}[e^{(\mu+\frac{\sigma^2}{2})}\cdot cdfn(\frac{-\ln K+\mu+\sigma^2}{\sigma}) - K\cdot cdfn(\frac{-\ln K+\mu}{\sigma})] \qquad (1.2.11)
$$

Thus, the option pricing formula for mixture of two lognormal distributions is

$$P_{call} = \lambda e^{-r(T-t)}[e^{(\mu_1+\frac{\sigma_1^2}{2})} \cdot cdfn(d_{11}) - K \cdot cdfn(d_{21})]$$ (1.2.12)

$$+(1-\lambda)e^{-r(T-t)}[e^{(\mu_2+\frac{\sigma_2^2}{2})} \cdot cdfn(d_{12}) - K \cdot cdfn(d_{22})]$$

Where $d_{1i} = \dfrac{-\ln K + \mu_i + \sigma_i^2}{\sigma_i}$, $d_{2i} = \dfrac{-\ln K + \mu_i}{\sigma_i}$ i=1,2

1.3 Derivation of put option pricing formula for mixture of two lognormal distributions

$$P_{put} = \lambda P_{put_1} + (1-\lambda)P_{put_2} \tag{1.3.1}$$

where $P_{put_i} = e^{-r(T-t)}\int_0^K (K-F_T)g_i(F_T)dF_T$ (1.3.2)

The following steps derive P_{put_i}

$$P_{put} = e^{-r(T-t)}\int_0^K (K-F_T)g(F_T)dF_T$$ (1.3.3)

$$= e^{-r(T-t)}[K\int_0^K g(F_T)dF_T - \int_K^{+\infty} F_T g(F_T)dF_T]$$

where $g(F_T) = \dfrac{1}{\sqrt{2\pi}\cdot\sigma\cdot F_T} e^{-\frac{(\ln F_T - \mu)^2}{2\sigma^2}}$

So it contains two integrations. The calculation will be done one by one.

$$
\begin{aligned}
&K\int_0^K g(F_T)dF_T \\
&= K\int_0^K \frac{1}{F_T}\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\ln F_T-\mu)^2}{2\sigma^2}}dF_T \\
&= K\int_0^K \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\ln F_T-\mu)^2}{2\sigma^2}}d\ln F_T \\
&= K\int_{-\infty}^{\ln K} \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\
&= K\int_{-\infty}^{\frac{\ln K-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy \\
&= K\cdot cdfn(\frac{\ln K-\mu}{\sigma})
\end{aligned}
\qquad (1.3.4)
$$

During the integration, first let $x = \ln F_T$, then let $y = \frac{x-\mu}{\sigma}$ to get the derivation.

The second integration could be calculated as:

$$
\begin{aligned}
&\int_0^K F_T g(F_T)dF_T \\
&= \int_0^K F_T\cdot\frac{1}{F_T}\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\ln F_T-\mu)^2}{2\sigma^2}}dF_T \\
&= \int_0^K \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\ln F_T-\mu)^2}{2\sigma^2}}dF_T \\
&= \int_{-\infty}^{\ln K} \frac{1}{\sigma\sqrt{2\pi}}e^{x}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\
&= e^{\frac{(\mu+\sigma^2)^2-\mu^2}{2\sigma^2}}\int_{-\infty}^{\ln K} \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-(\mu+\sigma^2))}{2\sigma^2}}dx \\
&= e^{(\mu+\frac{\sigma^2}{2})}\int_{-\infty}^{\frac{\ln K-(\mu+\sigma^2)}{\sigma}} \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy \\
&= e^{(\mu+\frac{\sigma^2}{2})}\cdot cdfn(\frac{\ln K-\mu-\sigma^2}{\sigma})
\end{aligned}
\qquad (1.3.5)
$$

During the integration, we first let $x = \ln F_T$, then let $y = \dfrac{x-(\mu+\sigma^2)}{\sigma}$ to get the derivation.

Therefore, the expected put option price under single lognormal distribution could be written as

$$P_{put} = e^{-r(T-t)}[K \cdot cdfn(\frac{\ln K - \mu}{\sigma}) - e^{(\mu+\frac{\sigma^2}{2})} \cdot cdfn(\frac{\ln K - \mu - \sigma^2}{\sigma})] \qquad (1.3.6)$$

Thus, the option pricing formula for mixture of two lognormal distributions is

$$\begin{aligned} P_{put} &= -\lambda e^{-r(T-t)}[e^{(\mu_1+\frac{\sigma_1^2}{2})} \cdot cdfn(-d_{11}) - K \cdot cdfn(-d_{21})] \\ &- (1-\lambda)e^{-r(T-t)}[e^{(\mu_2+\frac{\sigma_2}{2})} \cdot cdfn(-d_{12}) - K \cdot cdfn(-d_{22})] \end{aligned} \qquad (1.3.7)$$

$$Where\ d_{1i} = \frac{-\ln K + \mu_i + \sigma_i^2}{\sigma_i}, \quad d_{2i} = \frac{-\ln K + \mu_i}{\sigma_i} \quad i=1,2$$

1.4 Derivation of the relationship between μ_1, μ_2 and F_T

Expected future price F_T can be calculated as

$$\begin{aligned} \hat{E}(F_T) &= \int_{-\infty}^{+\infty} F_T g(F_T) dF_T) \\ &= \int_0^{+\infty} F_T(\lambda g_1(F_T) + (1-\lambda) g_2(F_T)) dF_T \\ &= \lambda \int_0^{+\infty} F_T g_1(F_T) dF_T + (1-\lambda) \int_0^{+\infty} F_T g_2(F_T) dF_T \\ &= \lambda \hat{E}_1(F_T) + (1-\lambda)\hat{E}_2(F_T) \end{aligned} \qquad (1.4.1)$$

where $\hat{E}_i(F_T)$ is the expected value of future price with a single lognormal distributions in the risk neutral world. (In order to make it easy, I will not use the subscript in the derivation).

$$\hat{E}_i(F_T) = \int_{-\infty}^{+\infty} F_T g_i(F_T) dF_T$$

$$= \int_0^{+\infty} F_T \cdot \frac{1}{F_T} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln F_T - \mu)^2}{2\sigma^2}} dF_T$$

$$= \int_0^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln F_T - \mu)^2}{2\sigma^2}} dF_T \qquad (1.4.2)$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= e^{\frac{(\mu+\sigma^2)^2-\mu^2}{2\sigma^2}} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-(\mu+\sigma^2))^2}{2\sigma^2}} dx$$

$$= e^{\mu+\frac{\sigma^2}{2}}$$

During this integration, let $x = \ln F_T$.

Thus, $\hat{E}(F_T) = \pi \cdot e^{\mu_1+\frac{\sigma_1^2}{2}} + (1-\pi) \cdot e^{\mu_2+\frac{\sigma_2^2}{2}}$ (1.4.3)

1.5 Derivation of formulas for Expected Value, Variance, Skewness, and Kurtosis of mixture of two lognormal distributions

The single lognormal distribution can be expressed as

$$g(F_T) = \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot F_T} e^{-\frac{(\ln F_T - \mu)^2}{2\sigma^2}} \qquad (1.5.1)$$

Thus

$$E(F_T{}^n) = \int_0^{+\infty} F_T{}^n \cdot g(F_T) dF_T = \int_0^{+\infty} F_T{}^n \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot F_T} e^{-\frac{(\ln F_T - \mu)^2}{2\sigma^2}} dF_T \qquad (1.5.2)$$

Let $\ln F_T = X$, then $F_T = e^X$, and $dX = \frac{dF_T}{F_T}$.

$$E(F_T{}^n) = \int_0^{+\infty} F_T{}^n \cdot g(F_T) dF_T = \int_0^{+\infty} F_T{}^n \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot F_T} e^{-\frac{(\ln F_T - \mu)^2}{2\sigma^2}} dF_T$$

$$= \int_{-\infty}^{+\infty} \frac{e^{nX}}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}} dX = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2} + nX} dX$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(X-(\mu+n\sigma))^2 - 2n\mu\sigma^2 - n^2\sigma^4}{2\sigma^2}} dX$$

$$= e^{n\mu + \frac{n^2\sigma^2}{2}} \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(X-(\mu+n\sigma))^2}{2\sigma^2}} dX$$

$$= e^{n\mu + \frac{n^2\sigma^2}{2}}$$

Thus the following relationship exist.

$$E(F_T) = e^{\mu + \frac{\sigma^2}{2}} \qquad (\text{n} = 1) \tag{1.5.4}$$

$$E(F_T{}^2) = e^{2\mu + 2\sigma^2} \qquad (\text{n} = 2) \tag{1.5.5}$$

$$E(F_T{}^3) = e^{3\mu + \frac{9\sigma^2}{2}} \qquad (\text{n} = 3) \tag{1.5.6}$$

$$E(F_T{}^4) = e^{4\mu + 8\sigma^2} \qquad (\text{n} = 4) \tag{1.5.7}$$

The above formula are for single lognormal distribution. In mixture of two lognormal distributions, $E(F_T{}^n)$ is a linear combination of $E_1(F_T{}^n)$ and $E_2(F_T{}^n)$ with the weight of λ and 1-λ respectively. That is,

$$E_1(F_T{}^n) = \lambda E_1(F_T{}^n) + (1-\lambda) E_2(F_T{}^n) \tag{1.5.8}$$

The definitions of expected value, variance, skewness and kurtosis are as follows:

$E_1(F_T)$ is expected value.

For variance:

$$Var(F_T) = E[(F_T - E(F_T))^2] = E[F_T{}^2 - 2 \cdot F_T \cdot E(F_T) + E(F_T)^2]$$
$$= Exp(F_T{}^2) - 2 \cdot E(F_T) \cdot E(F_T) + E(F_T)^2 \qquad (1.5.9)$$
$$= E(F_T{}^2) - E(F_T)^2$$

For skewness:

$$E[(F_T - E(F_T))^3] = E[F_T{}^3 - 3F_T{}^2 E(F_t) + 3F_T E(F_T)^2 - E(F_T)^3]$$
$$= E(F_T{}^3) - 3E(F_T{}^2)E(F_T) + 2(E(F_T))^3 \qquad (1.5.10)$$

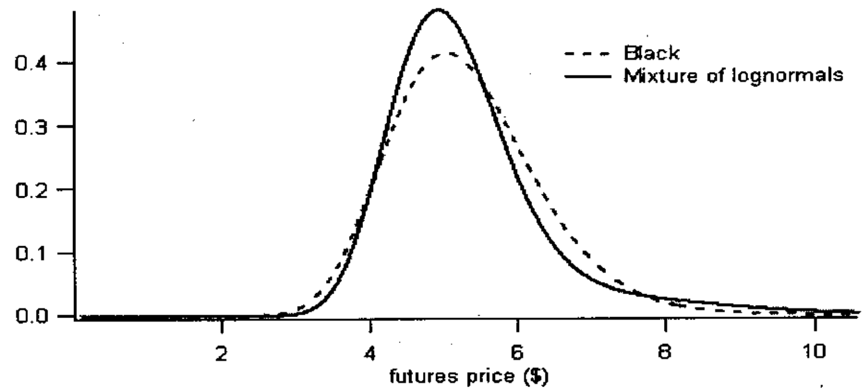
For kurtosis:

$$E[(F_t - E(F_T))^4]$$
$$= E[F_T{}^4 + 4F_T{}^2(E(F_T))^2 + (E(F_T))^4$$
$$- 4F_T{}^3(E(F_T)) + 2F_T{}^2(E(F_T))^2 - 4F_T(E(F_T))^3] \qquad (1.5.11)$$
$$= E(F_T{}^4) + 6E(F_T{}^2)(E(F_T))^2 - 4E(F_T{}^3)(E(F_T)) - 3(E(F_T))^4$$

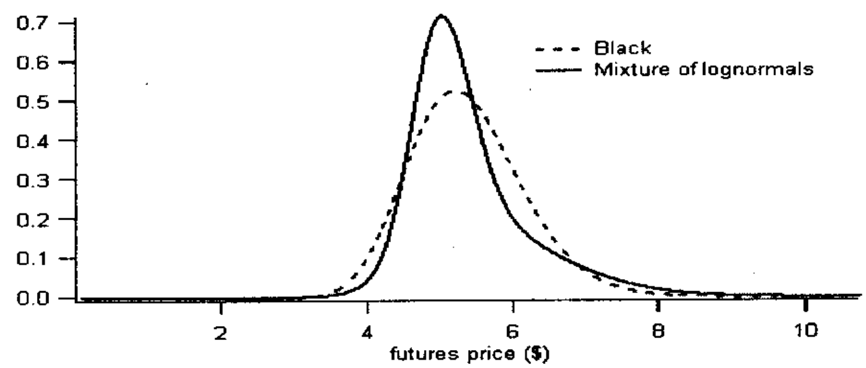
Appendix II: PDFs, CDFs and option pricing errors of storage contracts of soybean, corn and wheat

2.1 Estimated PDFs – July (storage) contract of soybean

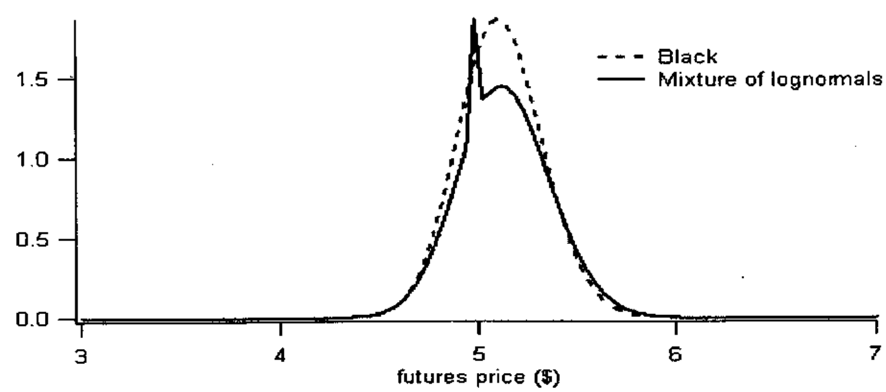
a. July Option Maturity as of October 13, 1999



b. July Option Maturity as of March 15, 2000

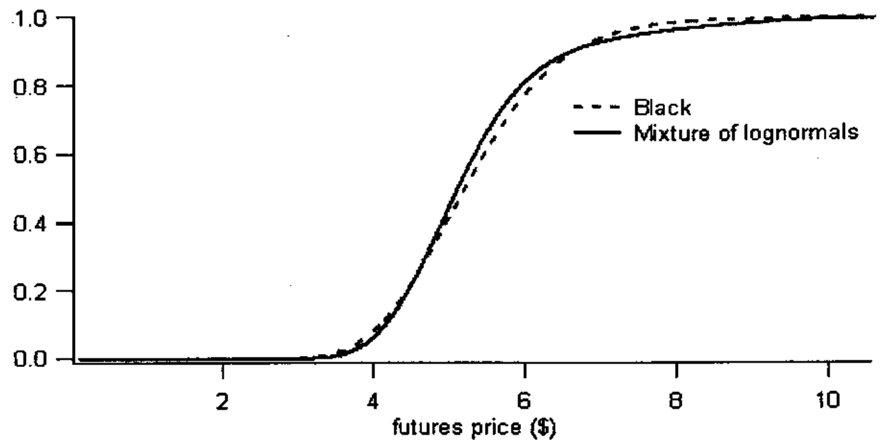


c. July Option Maturity as of June 14, 2000

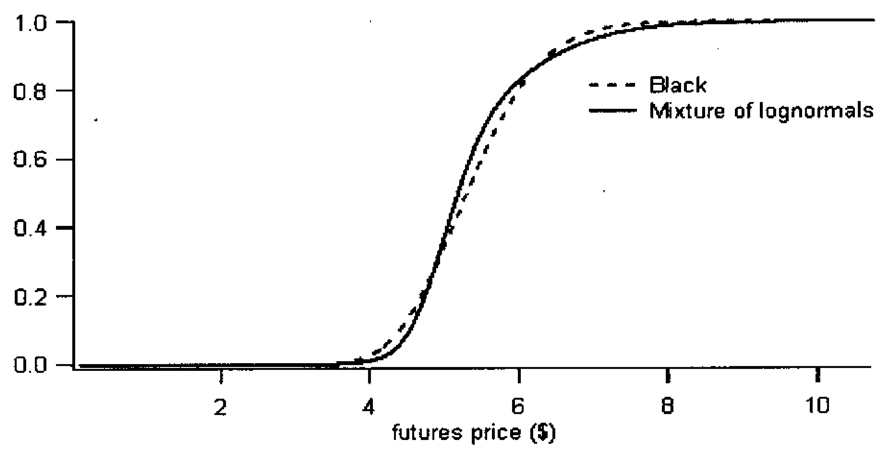


2.2 Estimated CDFs – July (storage) contract of soybean

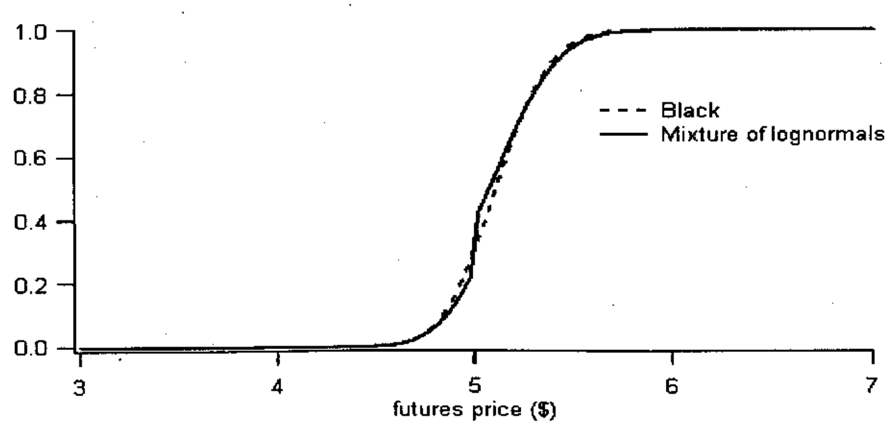
a. July Option Maturity as of October 13, 1999



b. July Option Maturity as of March 15, 2000

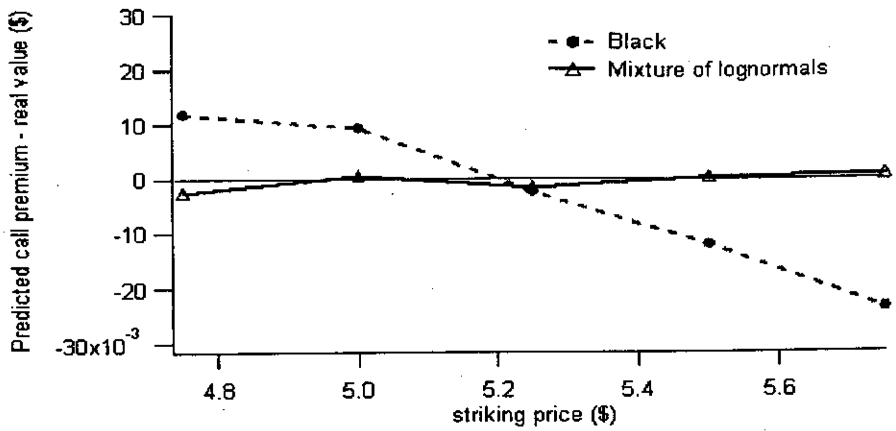


c. July Option Maturity as of June 14, 2000

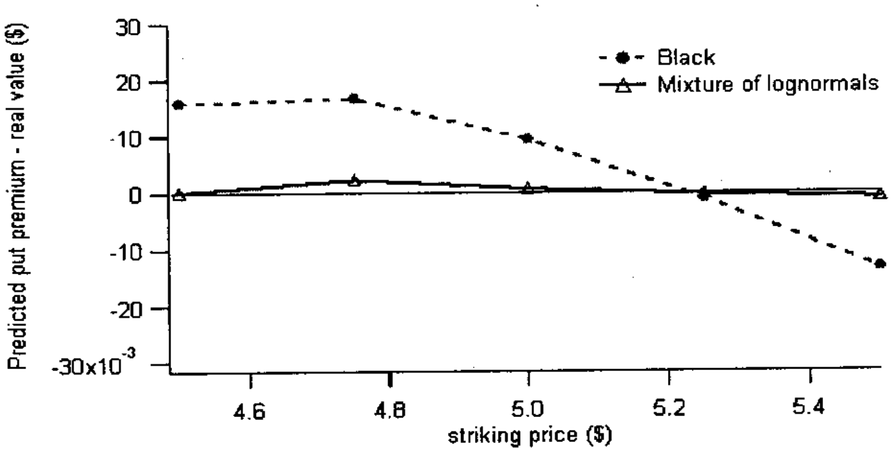


2.3 Option pricing errors – July (storage) contract of soybean

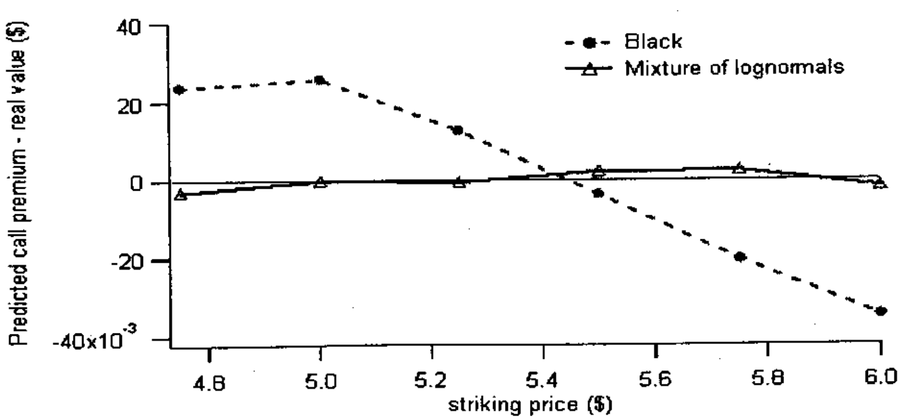
a. Call Pricing Errors of October 13, 1999



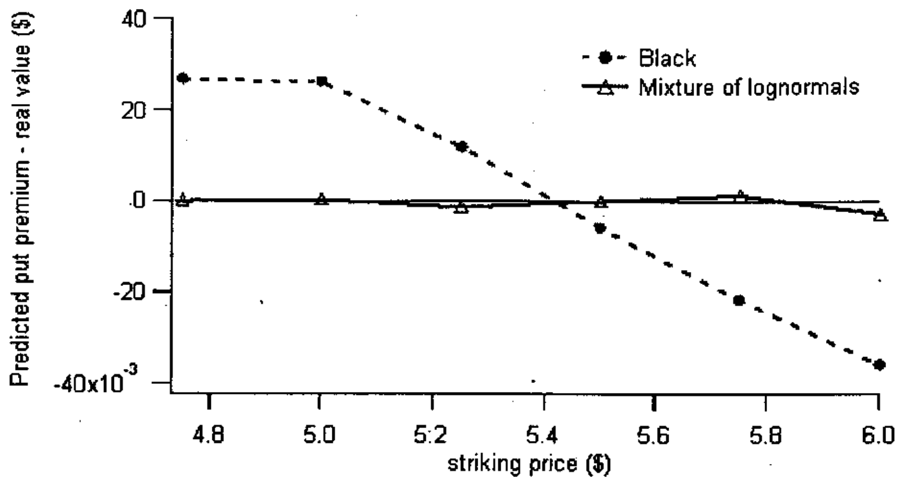
b. Put Pricing Errors of October 13, 1999



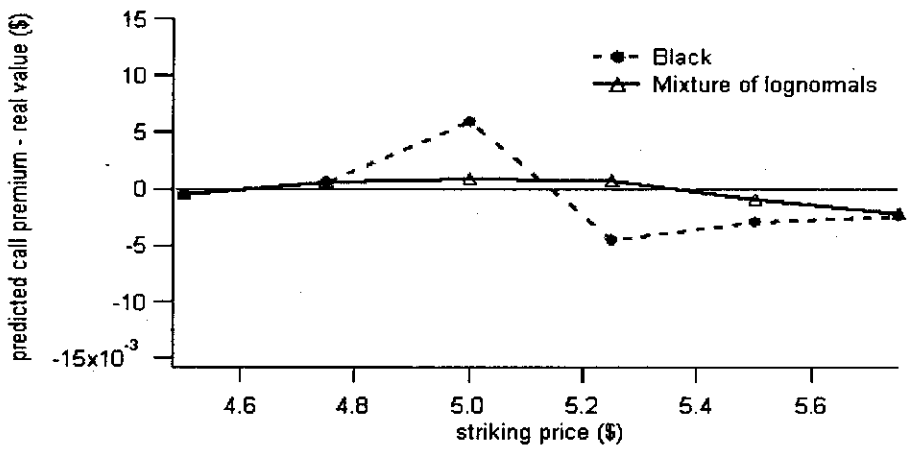
c. Call pricing Errors of March 15, 2000



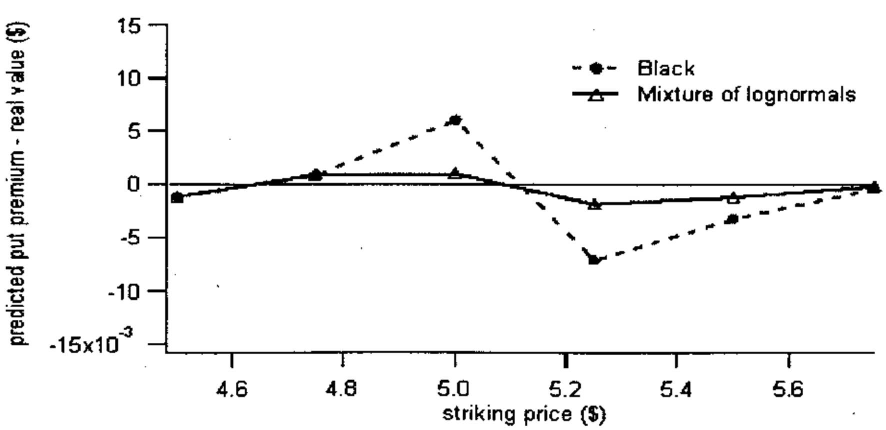
d. Put Pricing Errors of March 15, 2000



e. Call Pricing Errors of June 14, 2000

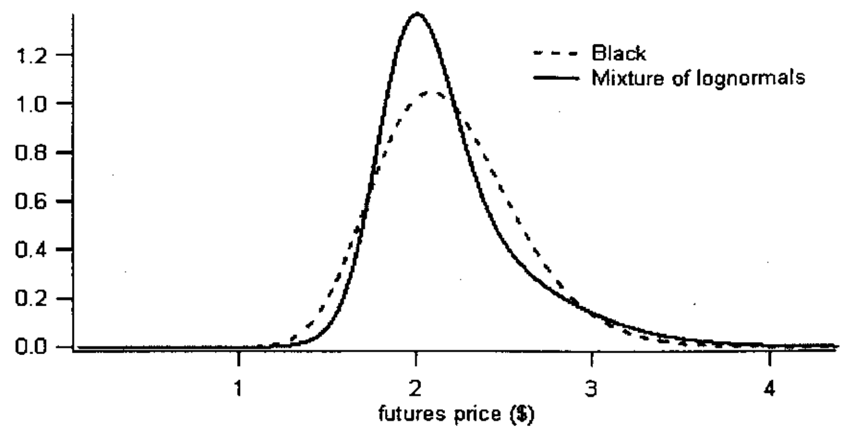


f. Put Pricing Errors of June 14, 2000

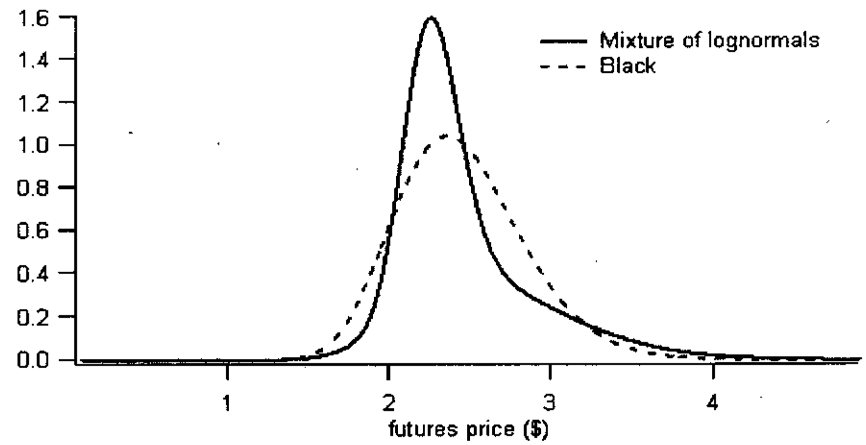


2.4 Estimated PDFs – July (storage) contract of corn

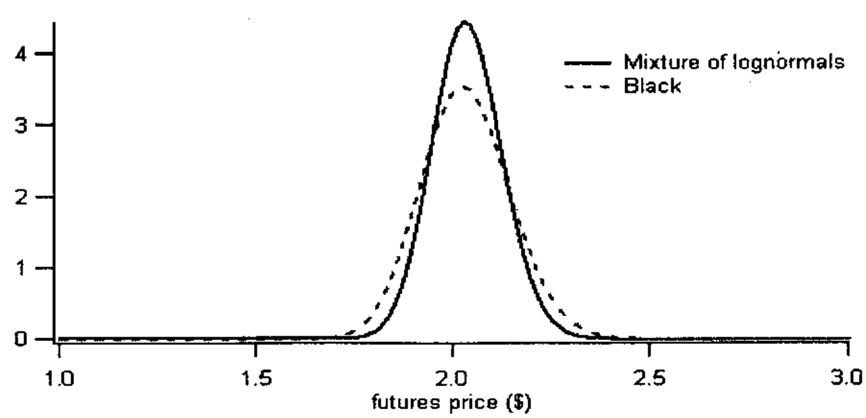
a. July Option Maturity as of November 17, 1999



b. July Option Maturity as of March 15, 2000

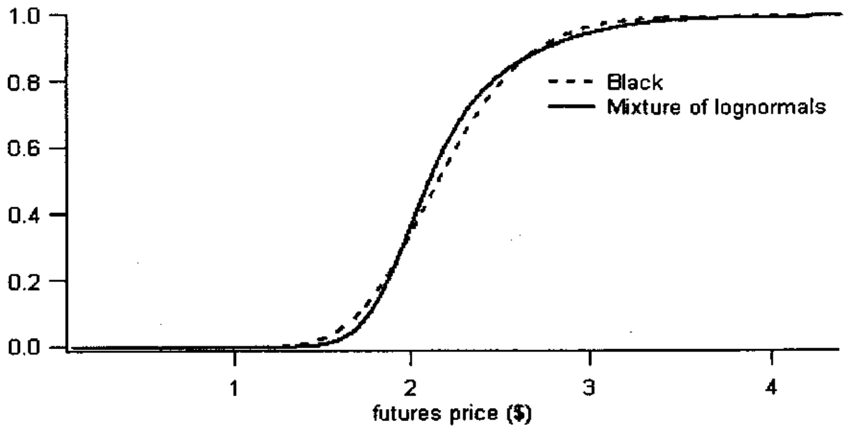


c. July Option Maturity as of June 14, 2000

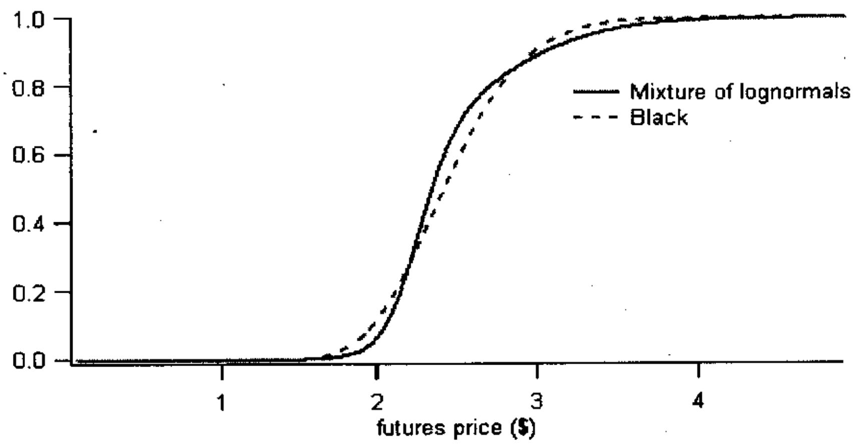


2.5 Estimated CDFs – July (storage) contract of corn

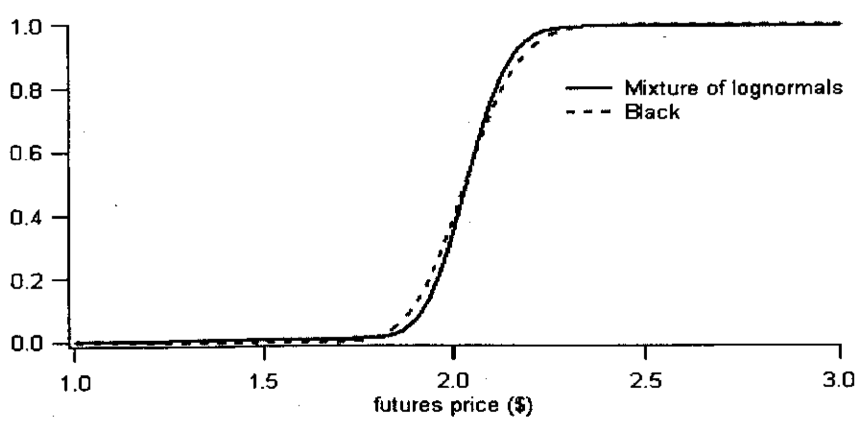
a. July Option Maturity as of November 17, 1999



b. July Option Maturity as of March 15, 2000

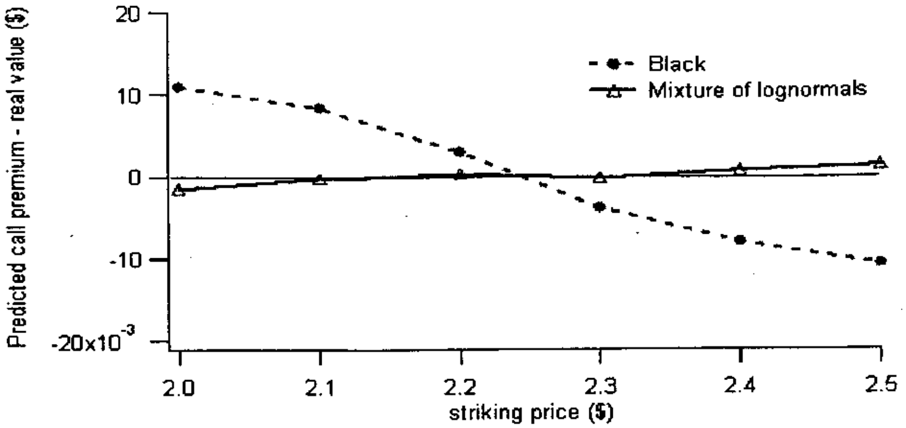


c. July Option Maturity as of June 14, 2000

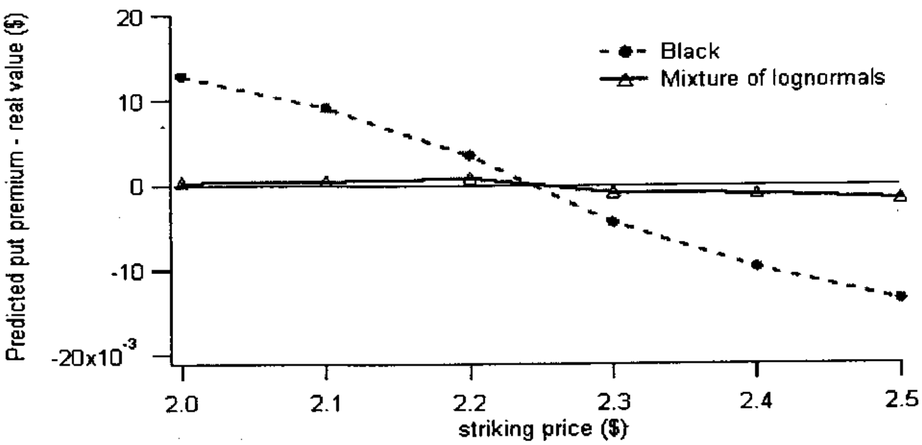


2.6 Option Pricing errors – July (storage) contract of corn

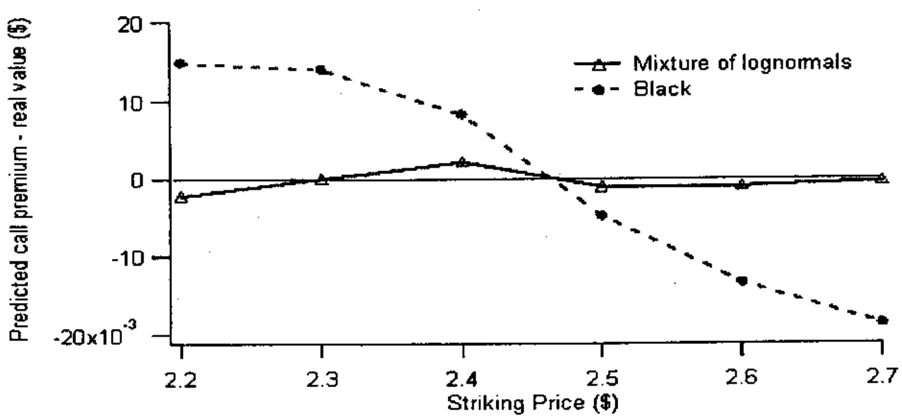
a. Call pricing errors of November 17, 1999



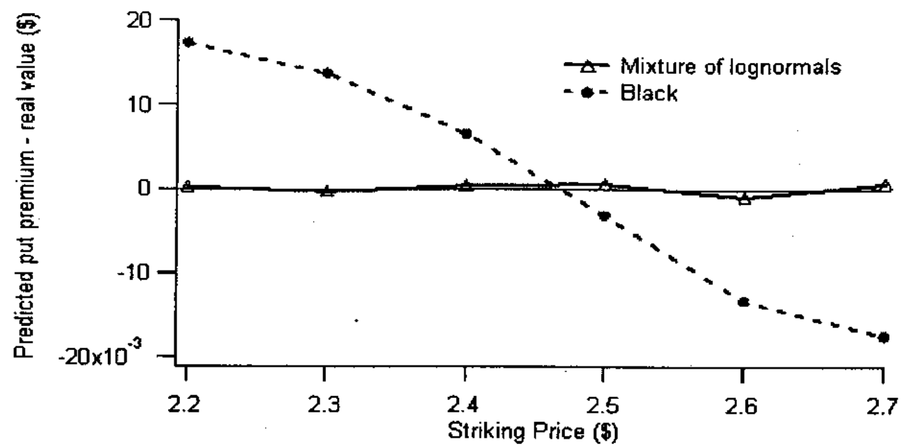
b. Put pricing errors of November 17, 1999



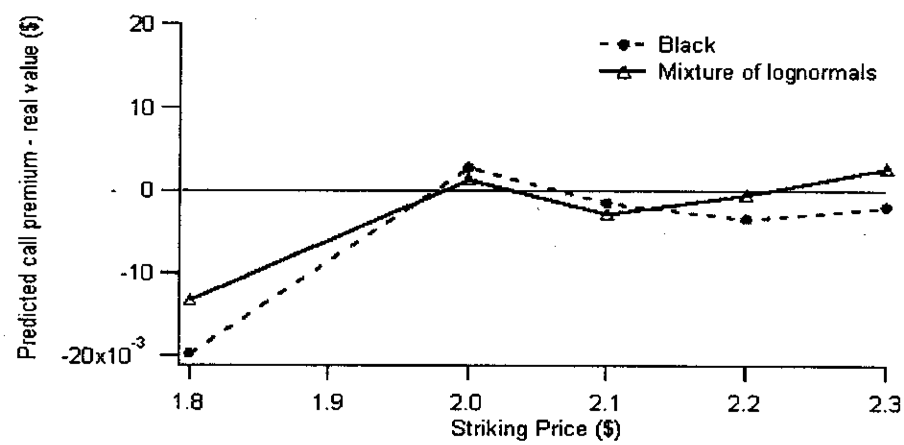
c. Call pricing errors of March 15, 2000



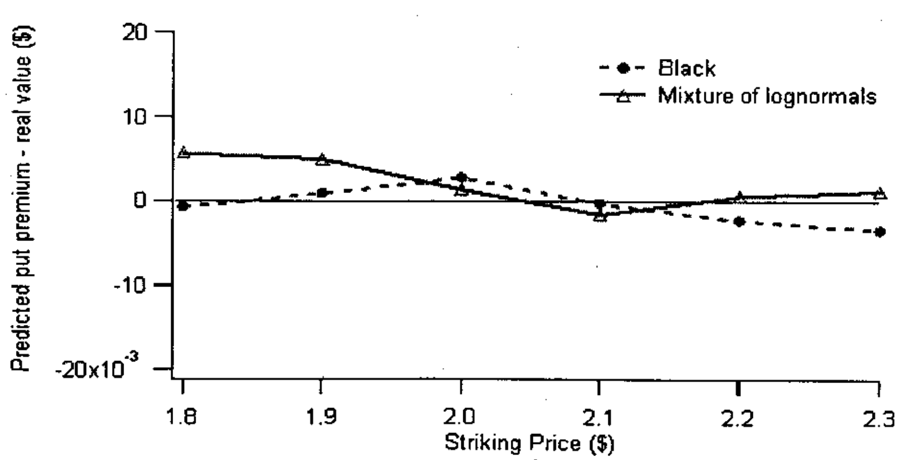
d. Put pricing errors of March 15, 2000



e. Call pricing errors of June 14, 2000

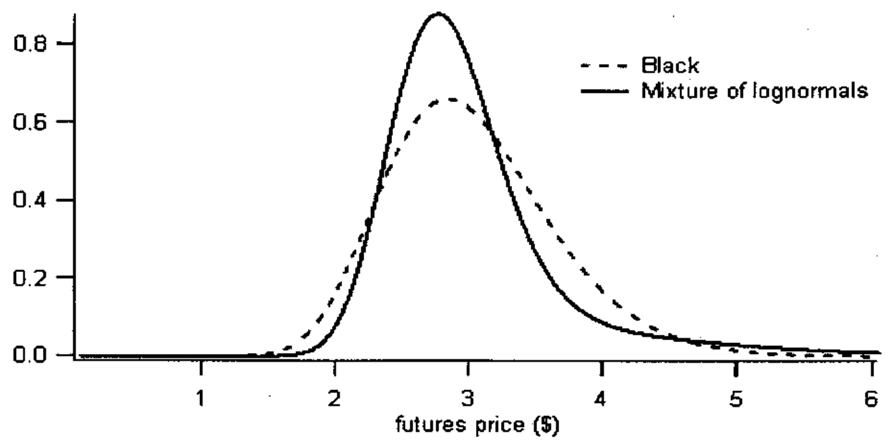


f. Put pricing errors of June 14, 2000

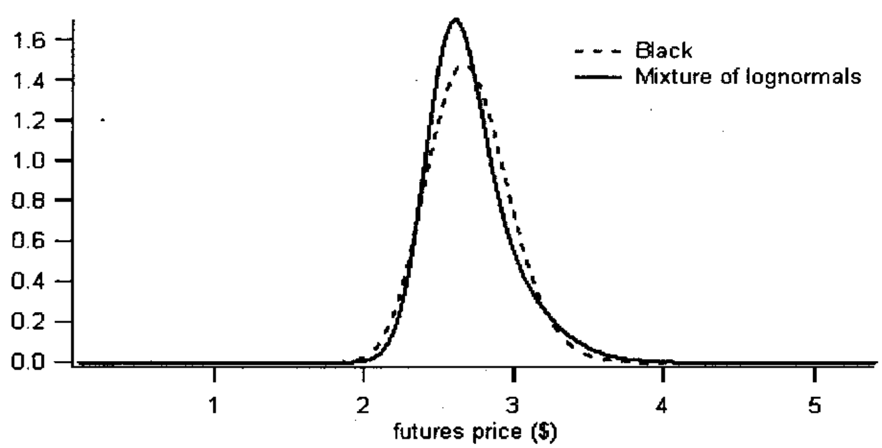


2.7 Estimated PDFs – March (storage) contract of wheat

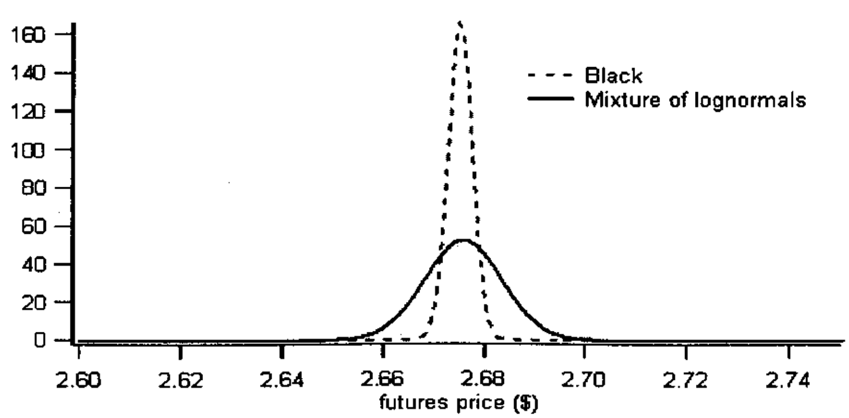
a. March Option Maturity as of June 14, 2000



b. March option Maturity as of November 29, 2000

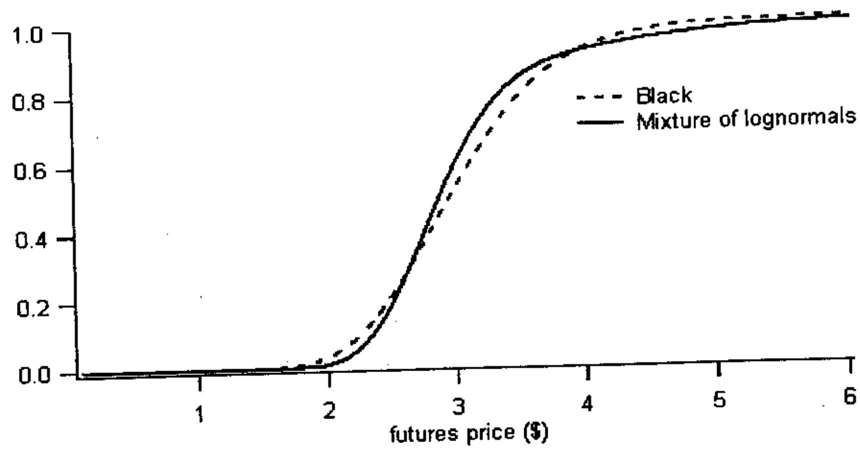


c. March Option Maturity as of February 14, 2000

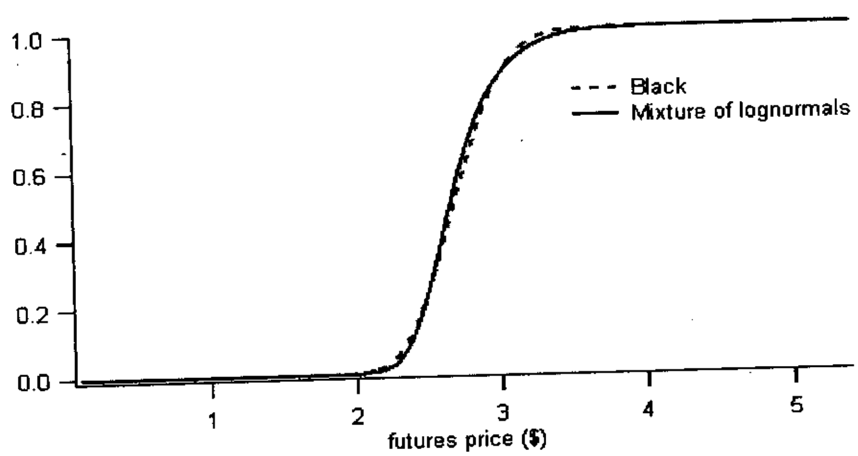


2.8 Estimated CDFs – March (storage) contract of wheat

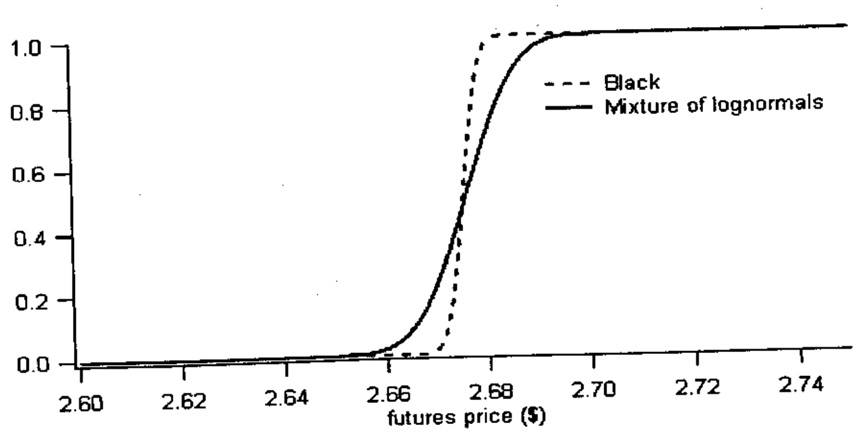
a. March option maturity as of June 14, 2000



b. March Option Maturity as of November 29, 2000

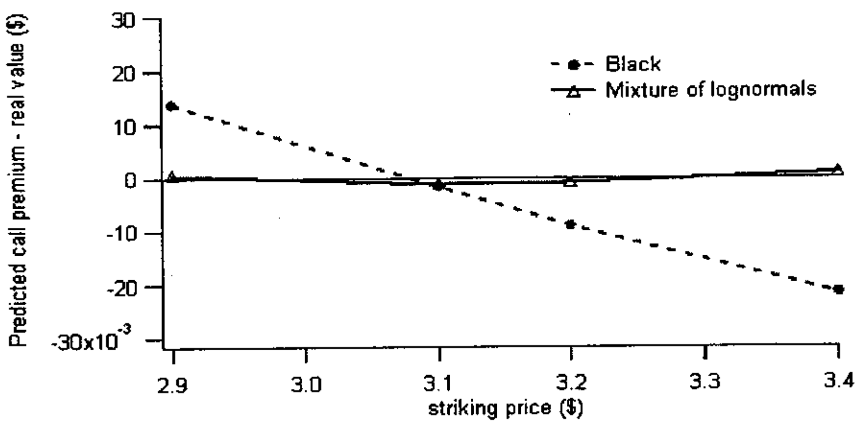


c. March Option Maturity as of February 14, 2000

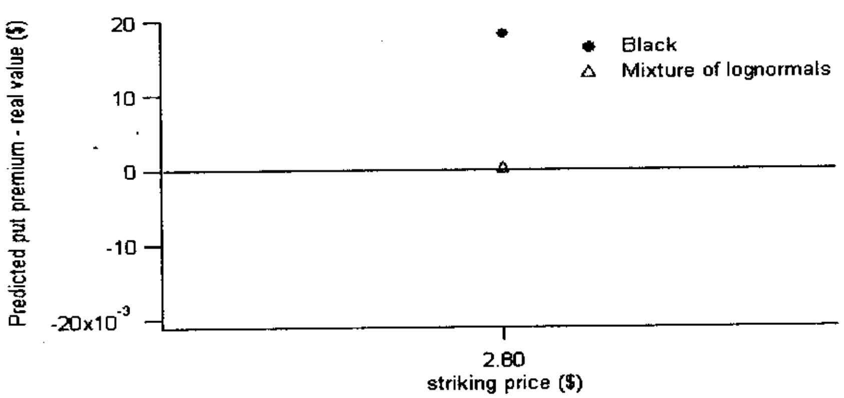


2.9 Option pricing errors – March (storage) contract of wheat

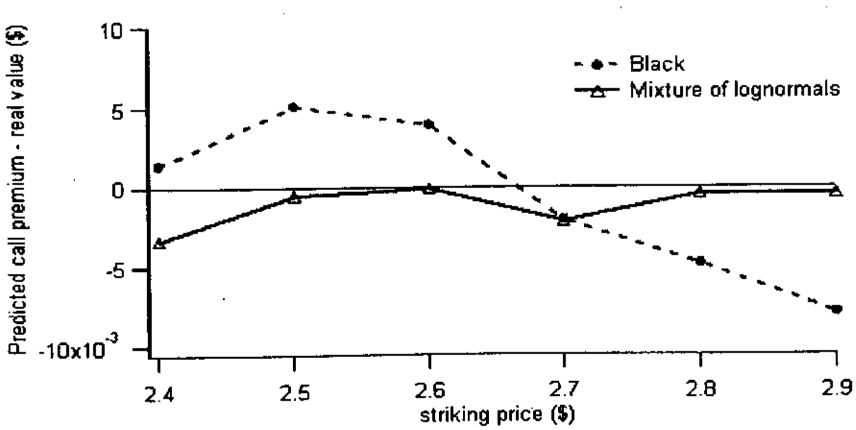
a. Call pricing errors of June 14, 2000



b. Put Pricing Errors of June 14, 2000[15]

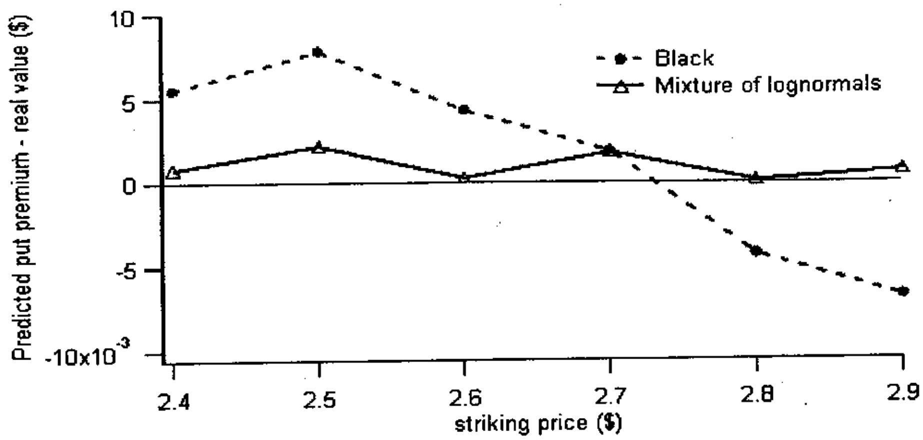


c. Call Pricing Error of November 29, 2000

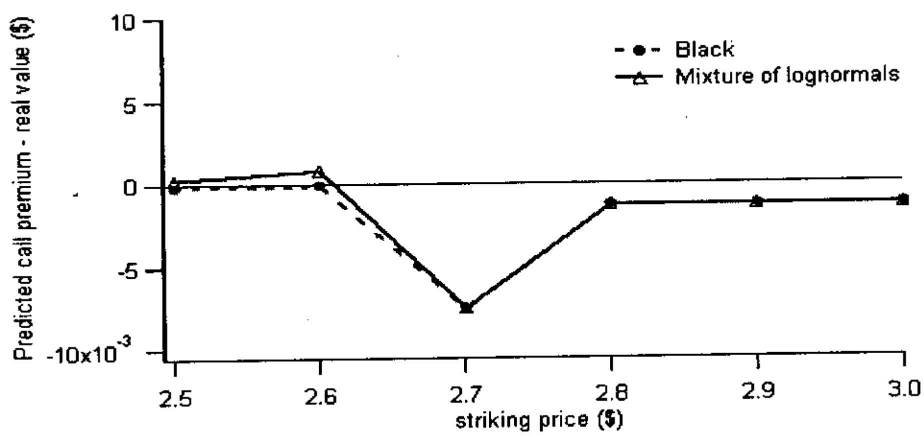


[15] There is only one set of data for put options on June 14, 2000.

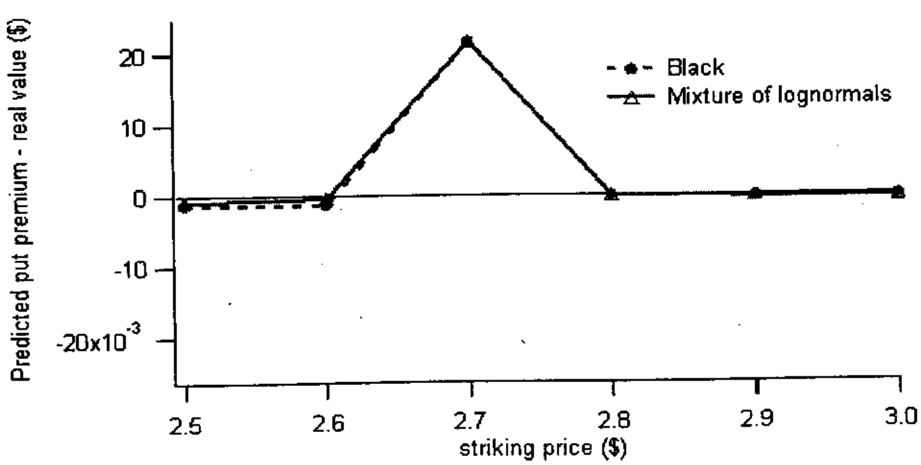
d. Put Pricing Errors of November 29, 2000



e. Call Pricing Errors of February 14, 2000



f. Put Pricing Errors of February 14, 2000



Reference

Black, F. (1976) The pricing of commodity contract. *Journal of Financial Economics* 3, 167-179

Christopher G. Lamoureux and William D. Lastrapes (1993) Forecasting stock-return variance: toward an understanding of stochastic implied volatilities. *The Review of Financial Studies* 6(2), 293-326.

Chriss, N. A. *Black-Scholes and Beyond: option pricing Models* 1997 McGraw-Hill. P327-360.

Cox, T.C. and S.A. Ross (1976) The valuation of options for alternative stochastic process. *Journal of Financial Economics* 3, 145-166.

Day, T. and C. Lewis (1992) Stock market volatility and the information content of stock index options. *Journal of Econometrics* 52, 267-287.

Fackler, P.L and R. P. King (1990) Calibration of option-based probability assessments in agricultural commodity markets. *American Journal of Agricultural Economics* 72(2), 73-83.

Greence William H. (1990) *Econometric Analysis* Macmillan publishing Company

Hauser J. Robert and Andersen K. Dane (1987) Hedging with options under variance uncertainty: an illustration of pricing new-crop soybeans. *American Journal of Agricultural Economics*, 38-45.

Hauser, R. J. and D. Neff (1985) Pricing options on agricultural futures: departures from traditional theory. The Journal of Futures Markets 5, 538-577

Henry A. Latane and Richard J. Rendleman (1976) Standard deviation of stock price ratios implied in option prices. *The Journal of Finance* XXXI (2), 369-381

Hull, J. C. *Options, Futures, and Other Derivative Securities*, Second edition 1993 Prentice-hall, Inc. p224-237.

Jimmy E. Hilliard and Jorge A. Reis (1999) Jump process in commodity futures prices and options pricing. *American Journal of Agricultural Economics* 81(May), 273-286.

Jose M. Campa, P.H. Kevin Chang and Robert L. Reider (1997) Implied exchange rate distributions: evidence from OTC option markets. NBER Working Paper Series, no. 6179.

Lamoureux, Christopher G. and William D. Lastrapes (1993) Forecasting stock-return variance: toward an understanding of stochastic implied volatility. *Review of Financial Studies* 6, 293-326.

Melick, William R. and Charles P. Thomas (1997) Recovering an asset's implied pdf from options prices: an application to crude oil during the Gulf crisis. *Journal of Financial and Quantitative Analysis* 32(1), 91-115.

Myers J. Robert and Hanson D. Steven (1993) Pricing commodity options when the underlying futures price exhibit time-varying volatility. *American Journal of Agricultural Economics* 75(2), 121-130.

Ritchey, R. J. (1990) Call option valuation for discrete normal mixtures. Journal of Financial Research 13(4), 285-296.

Tomek, W. G. (1996) Commodity futures price as forecasts. Review of Agricultural Economics, 19(1), 23-44.