This is to certify that the

thesis entitled

FORECASTING CORN BASIS AND CASH PRICE DISTRIBUTIONS

presented by

MIN ZHUANG

has been accepted towards fulfillment
of the requirements for

M.S. degree in Agricultural Economics

Major professor

Date 11/06/2000

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

6/01 c:/CIRC/DateDue.p65-p.15

FORECASTING CORN BASIS AND CASH PRICE DISTRIBUTIONS

By

Min Zhuang

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Master of Science

Department of Agricultural Economics

2000

ABSTRACT

FORECASTING CORN BASIS AND CASH PRICE DISDRIBUTIONS

By

Min Zhuang

This research estimates the basis distribution of corn using time series techniques and combines the resulting forecasted basis distribution with the market implied futures price distribution to generate a forecast of the cash distribution for corn.

The data used in this study are the storage basis on corn at Saginaw, MI, relative July Futures price in Chicago Board of Trade. This research studies the stochastic properties of the basis and provides evidence that multi-year basis data could be stationary. An autoregressive conditional heteroscedasticity model (ARCH) and a generalized ARCH model are specified to estimate and forecast the basis distributions. The out-of-sample forecast evaluation implies the ARCH model provides a good description of distribution for the Saginaw corn storage basis.

Then black's option pricing model is used to generate the implied volatility of futures price. Using simulation techniques, we combine forecasts of the basis distribution with the market implied futures price distributions to get a forecast of the cash price distribution. This combination gives us a new approach to generate a forecast of commodity cash price distributions. With the predictability features of basis distributions and the available market information reflected in traded commodity option prices, this combination may outperform forecasts of cash distributions either from structural models or pure series models.

To my parents Sun Lihua, Zhuang Jiedun

and my husband Ziqiang Li

ACKNOWLEDGMENTS

I am grateful to Professor Steve Hanson, my major advisor, for his support, encouragement, and efforts to provide me with a complete training in scientific research, and especially for his patience. I have benefited tremendously from Dr. Hanson's love of research, his perseverance, and his extraordinary kindness. I also thank Dr. Robert Myers, who taught me valuable knowledge of econometrics and helped me throughout the research. Thanks Dr. James Hilker for giving me empirical knowledge about basis and many important materials for the research.

Dr. Hanson, Dr. Myers, Dr. Hilker's contribution and their spirits are reflected through this work. Without their help, it is impossible for me to carry out this research. Their help is very important and highly appreciated.

I would like to thank the entire department staff for being so pleasant and helpful, and especially thank Nicole for all her help. I also want to thank the many students in the department for their friendship. They make my study at Michigan State a very memorable period of my life.

All I accomplished is given by my parents, who, even in the most difficult years, taught me the importance of education and knowledge.

Finally, I want to thank my husband, Ziqiang Li. His love has always been my mental support and the source of my happiness.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Commodity prices vary across time as a result of seasonal and daily supply and demand events. These price movements generate price uncertainty for farmers and agribusiness firms. A risk averse decision maker's response to price risk generally focuses on reducing the likelihood of unfavorable outcomes through the use of one or more risk management strategies. One strategy is to use pricing instruments such as forward, futures, or option contracts to offset or "hedge" price risk. Other strategies include such things as diversifying risk by producing multiple commodities, or maintaining flexibility in the selection of inputs and/or input levels.

Futures and option markets are important pricing instruments for many of the major commodities in the United States including corn, soybean, wheat, hogs and cattle. Firms producing or using these commodities may be able to reduce price risk by taking a "hedge" position in one or more of these markets. The essence of hedging is to have a position in a pricing instrument that is "opposite" to the position held in the cash market. Because cash and futures prices for a commodity tend to move in parallel, losses (gains) in the cash market tend to be offset with gains (losses) in the opposite futures position.

A key to effective hedging in these markets is to understand how closely cash and futures price movements move together. Basis is a concise way to represent the relationship between futures and cash prices and is commonly defined as the difference between a cash market price at a specific location and time and the price for a given futures contract at a specific time. Working (1949) defines basis as the futures price minus the cash price at a particular point in time. More formally, basis can be defined as:

(1) $$b_t = f_t(T) - p_t$$

where b_t is the basis at time t; p_t is the cash price at a particular location at time t; $f_t(T)$ is the futures price at t for delivery at time T.

There are many ways to measure basis. "Nearby basis", "new crop basis", "harvest basis" and "storage basis" represent alternative measures of basis that each can have important uses (see Appendix 1). However, the focus of this study is on the storage basis. Storage plays an important function in many commodity markets. Agents store commodities for a variety of purposes. Producers may store grain while waiting for a better price; elevators may profit by providing storage facilities to producers or by storing grain for themselves; and processors may store grain to ensure an adequate input supply. The storage basis is a key determinant of the potential benefit from storing a commodity.

To illustrate the role of the storage basis, consider a storage operator who takes a long position in the cash market and a short position in the futures market in period t and reverses the position in period $t+1$. The operator buys the cash commodity at p_t and sells at p_{t+1}, and sells an equal amount of the futures at $f_t(T)$ and buys it back at $f_{t+1}(T)$. The total return is $(p_{t+1}-p_t)+(f_t(T)-f_{t+1}(T))$; or in terms of the basis defined in (1), the return is given by $b_t - b_{t+1}$. Thus, the change in the storage basis over the life of the hedge determines the operator's gross return from the storage hedge.

Because cash and futures prices tend to move up and down in the same direction, the variation in basis is typically small relative to variation in prices themselves. For commodities with continuous inventories, like corn, new market information that affects cash prices tends to have similar impacts on futures prices. If traders anticipate a decrease in demand for grain, this will tend to depress both distant and nearby futures since the different months are linked through inventories (Tomek, 1981). Therefore, even if cash and futures price levels change greatly over time, their difference may not change much. Thus, compared to cash and futures price levels, basis is reasonably predictable, especially as delivery time approaches on a particular futures contract. Theoretically, the basis narrows as the delivery month approaches reflecting the decreasing cost of storage and the converging supply and demand conditions in cash and futures markets. The relatively high level of predictability of basis results in hedgers typically preferring the "basis risk" associated with offsetting positions in the cash and futures markets to the risk associated with a cash market position alone.

Most previous studies on basis focus on measuring basis changes over time or identifying factors that influence basis levels (Working, Tomek). Only a small number deal with forecasting basis levels, and even less work has been done on modeling or forecasting basis distributions (Tilley, 1988, Liu, 1994).

The predictability and relatively low variability of basis are a key to effective hedging. In addition, these features may facilitate improvements in forecasting cash price distributions. Commodity cash prices observed through time are the result of a complex mixture of changes associated with supply and demand conditions. The time series of cash price may include seasonal, cyclical, trend, and irregular factors (Ender 1995). Economists have attempted to identify empirical regularities in cash prices using a variety of techniques. Unfortunately, it is often difficult to forecast cash prices accurately from historic observations on past prices and other market factors. To complicate matters, data on other market factors influencing supply and demand are often only available after significant time delays.

Futures prices provide a forecast of the expected cash price at the delivery location during the period when the contract matures. In addition, options on futures prices can be used to generate forecasts of the distribution of futures price at contract maturity using implied volatility techniques (Hull 1993, Hilliard and Reis 1999). An implied volatility is the volatility of the underlying futures price implied by market traded option prices. It can be used to monitor the market's opinion about the volatility of a particular futures contract. Because options premiums presumably account for all available market information on the day the options is valued, implied volatility forecasts of futures price distributions may be more reliable and up to date than estimates using only historical data. Furthermore it may be that combining forecasts of the relatively predictable basis distribution using public data with the forecasts of the futures price distribution using information reflected in traded commodity option premiums will result in improved forecasts of the local cash price distribution relative to forecasts based solely on publicly available data.

This study estimates the basis distribution of corn using time series techniques and combines the resulting forecasted basis distribution with the market implied futures price distribution to generate a forecast of the cash distribution for corn. Because the resulting cash price distribution estimate uses information contained in commodity futures option premiums, the resulting price distribution should be superior to forecasts based solely on historical data.

More specifically, the objectives of this research are: 1) generate alternative time series models that characterize the corn basis distribution; 2) use the basis distribution models to develop forecasts of basis the distribution; 3) evaluate the performance and forecasting ability of the alternative models; 4) combine the forecasted basis distribution with a market forecast of the futures price distribution to generate a forecast of the corn cash price distribution. The data we used to estimate the price distribution in this study are storage basis on corn at Saginaw, Michigan relative July Futures price in Chicago Board of Trade (CBOT). Because we only generate a single cash price distribution at one point in time, we do not evaluate the performance of the cash price distribution forecast in this study. Testing whether the method developed and illustrated in this study improves forecasts of cash price distributions is left to future research.

Literature Review

Previous Research on Modeling Basis

We begin by discussing the theory underlying storage basis behavior and then discuss time series approaches that can be used to estimate corn price basis distributions. There are two basic approaches used to model basis. One approach is to model basis directly using a structural model. This has been the approach taken in many of the studies on basis behavior which rely on the storage theory developed by Working (1949) and Brennan (1958). The implication is that the basis for a storable commodity represents the market's estimate of the price of storage where the price of storage is, in turn, determined by the supply and demand for storage. The theory of storage argues that if firms operate in competitive markets, then in long-run equilibrium their expected gains from storage for a given period must equal the storage cost of the firm. Kaldor (1939) and Working (1949) modify the theory by introducing the concept of "*convenience yield*" which is the value generated by having the commodity readily available to meet users needs and will be positive but approaches zero at high stock levels. Convenience yield may offset what appears as a loss from the storage function itself and provides an explanation why storage occurs when the speculative return from storage is below observable costs.

Tomek (1996) applies these arguments to derive a typical structural model for basis that incorporates both observable economic costs as well as convenience yield. The main components of the basis model (price of storage) are the sum of three items: opportunity cost of carrying inventory from one period to the next (usually represented as interest cost); cost of other inputs involved in storage; and convenience yield. In this framework basis can be modeled as:

$$b_t = c(t,T) = p_t\, r(t, T) + s(t, T) - y(t, T) \tag{2}$$

where $c(t, T)$ is the marginal storage cost from t to T; p_t is spot price at time t; $r(t,T)$ is an opportunity cost rate of return from t to T; $s(t, T)$ is physical storage cost from t to T, which could include warehousing charges, cost of insurance, the cost of any processing or transformation services, as well as any anticipated losses from deterioration; and $y(t,T)$ is convenience yield. This general model of storage basis provides a basic insight into the underlying concepts that drive basis behavior for storable commodities. The model can easily be extended to account for a number of additional factors, such as inventory levels, risk premium and transportation cost.

Because of the difficulty of making ancillary forecasts of explanatory variables in structural models, basis forecasts often have been made from time series models. Autoregressive Integrated Moving Average process (ARIMA) and autoregressive conditional heteroscedasticity (ARCH) models, and Generalized ARCH (GARCH) models are popular time series methods to model and forecast commodity prices distributions.

Generally, ARIMA models perform well relative to alternative methods when forecasting over short time horizons, but performance deteriorates over longer time horizons. The existence of stochastic trends commonly found in time series of commodity prices can easily be accommodated in ARIMA models. These models are typically reduced form equations utilizing the information in past commodity prices, but may also take account of other available information that might impact the price levels.

Commodity price distributions often exhibit time-varying volatility in commodity prices that can be observed as autocorrelation patterns in the conditional variance of price innovations (Baillie and Myers 1992, Yang and Brorsen 1992). Although volatility clustering has a long history as an empirical regularity characterizing high-frequency commodity prices, it was not until recently that applied researchers recognized the importance of explicitly modeling time-varying second-order moments. Two common ways of accounting for time-varying volatility in commodity prices are the ARCH model introduced by Engle (1982) and the GARCH model developed by Bollerslev (1986). ARCH and GARCH models are highly nonlinear and maximum likelihood is the usual estimation approach. ARCH and GARCH models also can capture some of the observed non-normalities in the empirical distributions of commodity prices.

While time varying volatility in commodity cash and futures prices have been well documented, the behavior of basis volatility has received relatively little attention. Since basis represents the relationship between futures and cash price, one can generate basis distributions from the estimates of futures price $f_t(T)$ and cash price p_t distributions.

This can be done in a variety of ways ranging for estimating supply and demand systems for a commodity to estimating reduced-form equations for futures and cash prices. Baillie and Myers (1991) use bivariate GARCH models to estimate joint cash and futures price distributions for six commodities. While the study did not specifically examine basis relationships, the authors' approach provides an implicit model of the basis distribution for each commodity and could also be used to generate forecasts of basis distributions using simulation techniques.

Previous Research on Modeling Commodity Cash Price Distributions

Methods used to estimate commodity price distributions are very similar to those used to estimate basis distribution models. Structural models for cash prices are specified according to economic theory and market conditions. Since prices of agricultural products are often related, multiple-market partial equilibrium models may be specified. Forecasts can then be generated from either single equation or from the unrestricted or restricted reduced-form models. Depending on the research problem and commodity, the model may include endogenous variables, lagged endogenous variables and exogenous variables. Streeter and Tomek (1992) find 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. However, as discussed in the previous section, one important issue in generating forecasts from structural models is the timing of when information is known. The analyst must not incorporate variables into the model whose values will not be known at the time that actual forecasts are made. If there is such a variable in the model, then its value must be forecasted independently and, hence, treated as exogenous in the model. Most of the research on commodity price distributions use time series and non-parametric approaches, such as ARCH, GARCH, Poisson Jump, time-independent mixture-of-normal distribution model (Kim and Kon 1994) and exponentially weighted moving average models (Venkateswaran and Meenakshi 1993). These models do a good job of capturing the time-varying volatility of commodity prices. However, all these methods only involve historical price data and do not include all available market information. Thus, the forecasted price distributions are backward rather than forward looking. This study will incorporate available market information into forecasts of cash price distributions by combining the forecasts of basis distribution using historical time series data with forecasts of the corresponding futures price distribution using the information contained in traded option premiums.

Empirical Models of Basis

Data

The primary focus in this study is on basis distributions for intra-year storage. Basis is measured relative to the July futures contract because it is nearest the end of the crop year (storage season). Basis is calculated using local cash prices for corn at Saginaw, Michigan, and July Futures price at the Chicago Board of Trade (CBOT). Cash and futures prices are each quoted as the closing price on Wednesday each week. So the basis data are daily prices sampled at weekly intervals. The sample data contains 1,218 basis observations starting in the first week of August 1974 and ending in the third week in Dec. 1997. To construct the continuous series when the current year's futures contract matures and is no longer traded, basis is calculated by switching to the following year's futures price.

Typically, the bulk of Saginaw's corn harvest is in November. After harvest there is usually a large amount of corn available for storage. Although supply is at its peak in the fall and declines as the crop year progresses, the demand for corn is relatively constant across time. Theoretically, the cash prices will typically be at their lowest point at the peak of harvest time and then rise gradually until crop year ends. Because the futures price is a forecast of cash price at a future delivery time, we would not expect any seasonal component in futures price levels. Thus we would expect basis levels to exhibit a seasonal component, which reaches its highest point at November harvest and then decreases gradually through the storage season. However, basis may not follow this pattern in the last few weeks of the crop year since futures prices are often highly variable close to maturity. Thus there may be a type of structural change each year in the data series at the contract switching period because the two prices represent different supply and demand conditions. The graph of the basis realizations are shown in figure 1:

By inspection, the graph of storage basis for corn at Saginaw shows a fairly consistent pattern across each year, although some variations are notable. The expected seasonality and contract jump can be detected in the graph. As expected, the basis level is generally wide from August to November (harvest time) and then narrows during storage. The obvious outliers at end of the 1995 crop year (July 1996) are due to relative high futures prices at contract maturity time, reflecting the expectation for a short supply situation for next crop year. Visual examination of the basis series also suggests a possible shift in the average level of the series, implying a potential structural change

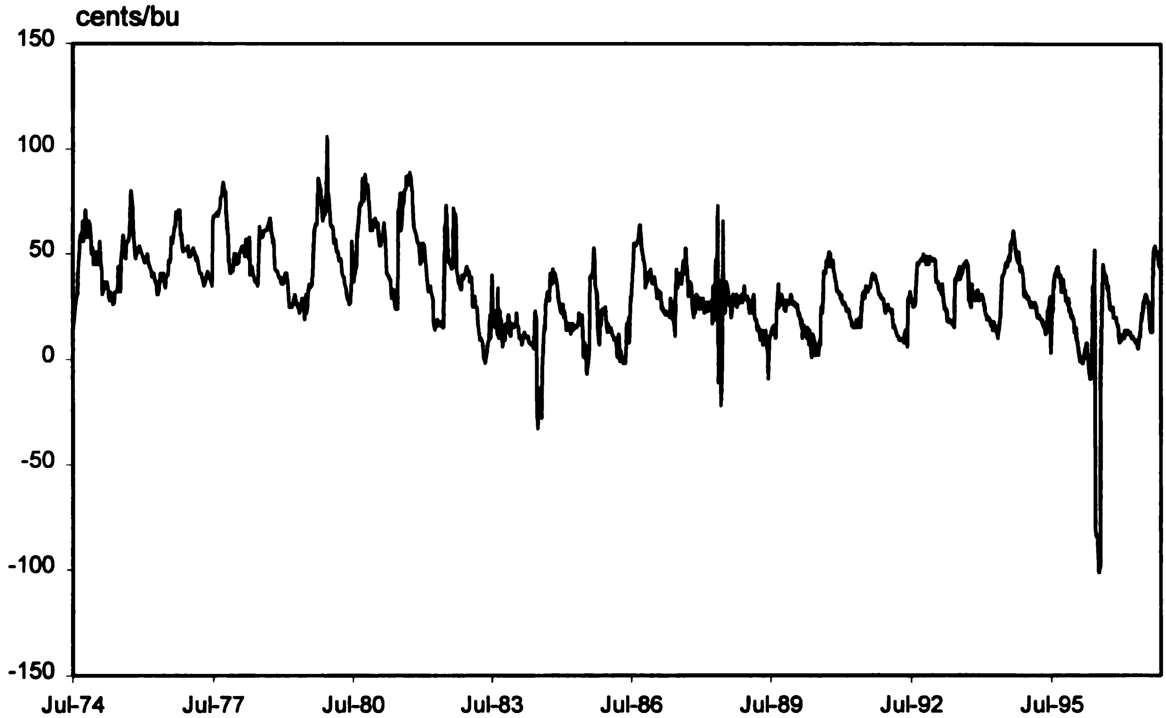


Figure1 Weekly Corn Basis Levels from 1974 to 1997

occurred around the 1982 crop year. However, visual examination at the series shows little evidence of any random drift. Visual examination of the data always is the first step to identify the data generating process. It indicates that the basis series generating process may include seasonal components, contract jump effects and possible structural shift. The information is very helpful in studying stochastic properties of the data and estimating an appropriate basis model specification.

Stochastic Properties of Basis Data

An important question when using time series approaches is whether or not the data is characterized by a stochastic trend. There is empirical evidence suggesting most high frequency (daily, weekly, even monthly) commodity prices are non-stationary. If the time series of interest is non-stationary, then conventional econometric techniques and statistical inferences are no longer valid.

There are numerous studies that have examined stochastic properties of commodity prices. Baillie and Myers (1991) test for unit roots in the logarithm of daily cash and futures price levels of six commodities and find a unit root for each commodity price over a one-year period. Goodwin and Schroeder (1991) finds unit roots in weekly beef prices for eleven U.S. regional markets; and Ardeni (1989) find unit roots for a group of nonstorable and storable commodities using monthly data. Goodwin (1992) tests monthly wheat prices covering the period from January 1978 through December 1989 for five important international wheat markets and finds evidence of nonstationary.

Despite the detailed analysis of stochastic properties of commodity prices, there has been relatively little attention given to analyzing stochastic properties of basis. Some studies examine cointegration relationships between futures prices and cash prices. These studies attempt to determine whether the two markets are influenced similarly by pricing information by investigating the properties of the price differences between the two markets. Cointegration tests of futures and cash price relationships are measures of the extent to whether basis is "stationary" over time. Note that a necessary condition for cointegration to exist is that each of the cointegrated variables must contain a unit root.

Baillie and Myers (1991) suggest that basis are non-stationary since basis is function of the interest rate which has been found to be nonstationary. Thus, a unit root in interest rate will indicate a unit root in basis. This suggests there may be good reasons for expecting not to find cointegration between cash prices and the prices on a particular futures contract with a fixed maturity. They conduct cointegration tests on daily futures and cash prices over one year and find no evidence of cointegration for any commodity. However, they do not preclude some other linear combination of cash and futures prices being stationary.

Fortenbery and Zapata (1993) analyze the relationship between cash and nearby futures prices for corn and soybeans. Unlike previous studies, they investigate a much longer time series, employing daily data from crop year 1980 to 1990. They conducted cointegration tests on the daily basis data year by year, as well as for the aggregate period (multiple-year). Year-by-year tests revealed basis in some years is cointegrated while in other years it is not. When years were aggregated, evidence of cointegration was detected for all cash and futures market pairs considered, that is, the multiple-year basis is stationary. Though these studies do not account for contract switching in multiple-year basis, they give an indication that the one-year period basis may have different stochastic properties than the multiple-year basis.

Understanding the stochastic properties of the basis data can lead to improved model specification for estimating and forecasting basis. Based on the cash and futures price properties, July corn storage basis tends to be large in November and gradually decrease as the maturity date approaches, reflecting the decrease of storage costs within the crop year. The basis on the delivery date approximates transportation costs between the cash and futures markets (from Saginaw to Chicago).

For basis series over a one-year period (daily or weekly data), the intra-year basis may be nonstationary based on the theory that basis is a function of the interest rate. For a multiple-year basis series, the basis generally reverts to approximately the level of transportation costs at the contract maturity date. Because shocks to a stationary time series are temporary, the effects of the shocks will dissipate over time and the series will revert to its long-run mean level. For multiple-year basis series, if transportation costs are stationary, the basis data may converge to this unconditional mean after eliminating any deterministic trend. Thus the multiple-year basis data may be a stationary process even if the intra-year basis is non-stationary.

Since our basis data are multiple-year data, the observations on dates when futures contracts switch may cause the unit root test results to be biased. To eliminate the contract switching effects, we conduct the unit root tests on multiple-year basis data after removing the observations immediately before and after the contract switch. So the first differencing data don't include the differencing basis between different contracts. The Augmented Dickey-Fuller tests are used on the data series after removing deterministic components, such as time trend and seasonal effects (see Appendix 2 for detail). The test results reject the null hypothesis that the basis data contains a unit root.[1] The unit root test results for the multi-year basis data suggest we can work with basis levels directly. It provides the evidence that multi-year basis data can be stationary. So stochastic trend that dominate in high frequency movement might be countered over by structural breaks (contract switching effects).

Estimate Basis Generating Process

In this section we present the results of the estimated basis generation process. We begin by constructing a general ARMA model for the basis level. The correlogram plot of basis data including autocorrelation function (AC) and partial autocorrelation function

[1] We also conduct the Dickey-Fuller unit root test for intra-year basis data. The results of crop year 74, 75, 76, 77, 78, 79, 96 show the null hypothesis of unit root can be rejected at 5% significant level.

(PAC) can help us decide the proper lags for AR and MA process. The correlogram plot of basis shows that PAC cut off after lag2 and get higher around lag 7, AC declines geometrically. The correlogram also shows distinct seasonal spikes in AC. The sine and cosine functions are used to capture seasonal effects. So we try ARMA(2, 7) model with seasonal variables and time trend first. However, the Q-tests of the residuals show that there are autocorrelations in the residuals at 5% significance level. After adjust, we get ARMA [(1,6), (1, 5, 6, 8)] and don't find autocorrelation in the residuals.

Using Ljung-Box Q-statistics to test autocorrelation in the squared residuals of the ARMA model, we find ARCH effects in the squared residuals suggesting the volatility of the series is not constant and that there is serial correlation between periods of large (small) volatility. Therefore univariate ARCH and GARCH models are used to represent the basis generating process for corn. ARCH (GARCH) models can be used to produce both point forecasts and volatility forecasts. The volatility forecasts can be measured by the variance of the forecast error conditional on the current information Ω_t used to construct the forecast, $\mathrm{var}(e_{t,m}|\Omega_t)$. When ARCH (GARCH) is present, $\mathrm{var}(e_{t,m}|\Omega_t)$ will depend on the elements of Ω_t and will, in general, be time varying. Again, the correlogram plot of squared residuals can help us identify the proper conditional variance equation. The AC and PAC show there are autocorrelation in lag1, 3, 5 and 6 in squared residuals suggesting an ARCH (6) model for the conditional variance.

The t-statistics of the estimated parameters suggest that both basis mean and variance have seasonality. Based on the visual examination of the data and the preliminary diagnostic tests, the following simple univariate ARMA-ARCH specification is first used to describe corn basis:

(3) $$b_t = \mu + \sum_{i=1}^{p} \alpha_i b_{t-i} + \sum_{j=1}^{q} \beta_j e_{t-j} + S_t + e_t$$

$$e_t | \Omega_{t-1} \sim D(0, h_t)$$
$$h_t = \omega + \alpha_1 e_{t-1}^2 + \cdots + \alpha_i e_{t-i}^2 + S_t$$

where $S_t = \sum_{j=1}^{2} [\varphi_j \sin(\frac{2\pi w_t}{52}) + \phi_j \cos(\frac{2\pi w_t}{52})]$

(p=6, q=8)

The degree of frequency in seasonality variables can be determine by log-likelihood ratio test, AIC and SIC criterions.

ARMA and ARCH parts are simultaneously estimated by numerically maximizing the log likelihood using the Newton Remdt algorithm. We use the Ljung-Box Q statistics to check the autocorrelation of the residuals from the ARMA part and the squared standardized residuals from the ARCH part to get a "best-fitting" model specification. If autocorrelation is present, we need to adjust the ARCH and ARMA models to eliminate the autocorrelation in the residuals and squared standardized residuals. The estimated results are shown in table 1.

The Q test results suggest there is no remaining autocorrelation in the residual series or squared standardized residuals of the ARCH model at a 5% significance level. Thus the model appears well specified in the sense that the models do a good job of capturing all the autocorrelations. However, the kurtosis of this model is larger than the normal distribution indicating a relative peak distribution. Many studies on empirical

Table 1. ARCH -normal model specification

Coefficient	Estimate	t-value	p-value
μ	1.171167	3.255497	0.001
φ_1	1.570692	3.369696	0.001
ϕ_1	1.712171	8.205744	0.000
φ_2	1.226470	5.928530	0.000
ϕ_2	0.422059	1.814145	0.070
α_1	0.908873	23.693460	0.000
α_6	0.060739	1.918030	0.055
β_1	-0.153154	-2.868962	0.004
β_5	-0.054035	-0.632912	0.527
β_6	-0.112358	-2.870888	0.004
β_8	-0.060424	-2.505826	0.012
	Variance Level		
ω	21.782927	0.000790	0.000
φ_1	16.155443	0.000346	0.000
ϕ_1	18.584877	0.000458	0.000
φ_2	10.758560	6.41E-06	0.000
ϕ_2	1.223889	0.540180	0.589
α_1	0.321646	0.169086	0.000
α_2	0.092008	0.057418	0.395
α_3	0.169329	0.052172	0.085
α_4	0.059913	1.976094	0.048
α_5	0.013637	0.122843	0.902
α_6	0.014503	0.559427	0.576

Q(1) = 0.0002	$Q^2(1)$ = 0.0100
Q(5) = 1.8650	$Q^2(5)$ = 6.8321
Q(10) = 15.9750	$Q^2(10)$ = 9.2115
Q(15) = 22.4445	$Q^2(15)$ = 12.6287
Q(20) = 23.0954	$Q^2(20)$ = 13.9002
Number of Observation: 1218	Skewness 0.22895251
Log likelihood -3780.7238	Kurtosis 8.78171906

Note: Q (k) is a Q test for k degree autocorrelation in the residuals; Q^2(k) is a Q test for k degree autocorrelation in the squared standardized residuals.

commodity price distribution have shown prices are not independent identical normally distributed. The tails of price distribution often appear to be much fatter than that of normal, indicating excess kurtosis in commodity price distributions. ARCH and GARCH models can lead to a partial solution of the excess kurtosis problem, but may not capture all of the excess kurtosis in commodity price distributions. One solution is to assume the conditional distribution of price innovations in ARCH (GARCH) models follow a *t*-distribution with degree of freedom treated as a parameter to be estimated (Bollerslev 1987, Baillie and Myers 1991, Yang and Brorsen 1992). Unfortunately, we were unable to fit ARCH (or GARCH) -*t* model for this basis sample data. The introduce of *t*-distribution in the ARCH model causes autocorrelations in residuals and squared standardized residuals which are difficult to eliminate.

The estimated model is a simple specification that does not include any specific corrections for the potential structural change issues discussed earlier. Next we introduce structural change variables into the basis model to see whether we can improve the specification and forecasting performance of the basis model. To account for the annual contract switches and apparent structural change during the 1982 crop year, we introduce dummy variables to measure contract jumps and structural change. Including the dummy variables may also change the specification of the trend, seasonality and volatility components of the model. After restricting the model with the dummy variables for structural change and contract switching we obtain a form of GARCH-in-Mean (GARCH-M) model (Engle, Lilien, Robins, 1987) which contains the conditional standard deviation in the mean equation. The GARCH-M basis model is:

(4) $$b_t = \mu + \theta_1 d1 + \theta_2 d2 + \sum_{i=1}^{p} \alpha_i b_{t-i} + \sum_{j=1}^{q} \beta_j e_{t-j} + S_t + \eta\sqrt{h_t} + e_t$$

$$e_t | \Omega_{t-1} \sim D(0, h_t)$$

$$h_t = \omega + \alpha_1 e_{t-1}^2 + \cdots + \alpha_i e_{t-i}^2 + \beta_1 h_{t-1} + \cdots + \beta_i h_{t-j} + S_t$$

where $S_t = \sum_{j=1}^{2} [\varphi_j \sin(\frac{2\pi w_t}{52}) + \phi_j \cos(\frac{2\pi w_t}{52})]$

($p=6$, $q=8$)

d1 represents contract jump dummy; *d2* represents structural change dummy variable where structural change happened in May 1983. The estimated model parameters are shown in table 2.

The GARCH-M model is often used in financial applications where the expected return on an asset is related to the expected asset risk. The estimated coefficient on the expected risk is a measure of the risk-return tradeoff. In this basis GARCH-M model, the basis risk is measured by the conditional variance h_t. Remember the storage basis definition, $b_t - b_{t+1}$ represents the return of a storage operator who takes a long position in the cash market and a short position in the futures market in period t and reverses the position in period $t+1$. The negative GARCH-M term η indicates that larger basis variance in period $t+1$ are associated with smaller basis level in period $t+1$. Thus the higher basis risk at $t+1$ will lead to higher expected storage return. This is consistent with standard risk-return arguments.

Table 2. GARCH -M-normal model specification

Coefficient	Estimate	t-value	p-value
μ	4.211778	4.570104	0.000
θ_1	8.769474	5.117442	0.000
θ_2	-1.497059	-4.230295	0.000
φ_1	2.285101	5.103892	0.000
ϕ_1	1.187119	4.448481	0.000
φ_2	1.349246	5.632816	0.000
ϕ_2	-0.144758	-0.508309	0.611
α_1	0.884268	32.110248	0.000
α_6	0.042354	1.970483	0.049
β_1	-0.150849	-2.799777	0.005
β_5	-0.049928	-1.477248	0.140
β_6	-0.096190	-2.899947	0.004
β_8	-0.047532	-2.018285	0.044
η	-0.156100	-1.916877	0.055
	Variance Level		
ω	7.833631	4.275425	0.000
φ_1	5.589322	3.411699	0.001
ϕ_1	7.242676	4.658292	0.000
φ_2	4.323185	3.772534	0.000
ϕ_2	1.121909	2.114415	0.035
α_1	0.332767	6.167886	0.000
β_1	0.492811	4.264462	0.000
β_2	0.251041	2.050258	0.041
β_3	-0.168414	-2.191059	0.029

$Q(1) = 0.1109$	$Q^2(1) = 0.0603$
$Q(5) = 1.9655$	$Q^2(5) = 7.7843$
$Q(10) = 14.763$	$Q^2(10) = 11.218$
$Q(15) = 21.079$	$Q^2(15) = 11.984$
$Q(20) = 21.720$	$Q^2(20) = 13.600$
Number of Observation: 1218	Skewness -0.115
Log likelihood −3754.2458	Kurtosis 9.04

The *t*-statistic of θ_1 and θ_2 indicate that both contract jump dummy and structural change dummy variables are statistic significant. It is worth nothing that the kurtosis level in this model is slightly higher than in the previous model in terms of the parameter estimation of the new added variable (*d1* and *d2*) seem to provide us additional information on the basis generating process. The estimate of the contract jump dummy θ_l shows that the first basis observation following each contract switch is average 8.8 cents/bushel higher than the last observation of the previous contract. The estimate of structural dummy θ_2 provides evidence that a structural change occurred sometime around the crop year 1982, lowering the average corn basis level by 1.5 cents/bu from previous period, perhaps due to the change of transportation costs and interest rate.

In general, both basis models appear well specified in terms of capturing all the autocorrelations in residual series and squared standard residuals. The GARCH-M model with the structural change and contract switching dummies appears to provide additional information to explain the basis behavior. However, because the purpose of the models is to forecast basis distributions, we evaluate these basis models by comparing their out-of-sample forecasting performance.

Simulating the Corn Basis Distribution

We can use the basis generating process to forecast the basis distribution *m* periods from the current period *t* to terminal period *T*. Let b_t be the initial basis price on Dec. 17, 1997, which is assumed to be the post-harvest time of corn. The terminal date is the second week of July 1998, which is the assumed maturity time of July corn futures. The simulation procedure for the basis distribution is as follows:

1. Simulate *h* by iterating the variance equation forward one period.

$$h_{t+1} = \omega + \alpha_1 e_t^2 + \cdots + \alpha_i e_{t-i}^2 + \beta_1 h_t + \cdots + \beta_j h_{t-j} + S_{t+1}$$

$$S_{t+1} = \sum_{j=1}^{J} [\varphi_j \sin(\frac{2\pi w_{t+1}}{52}) + \phi_j \cos(\frac{2\pi w_{t+1}}{52})]$$

Calculate conditional variance h_t based on previous conditional variance and realized values of previous squared innovation e_{t-i}^2 ($\beta = 0$ for ARCH model).

2. Generate random draw (v) from a normal distribution with zero mean and unit variance so that ($\sqrt{h_{t+1}}v$) are independently distribution normal vector with mean 0 and variance h_{t+1}.

3. Calculate the error one step ahead forecast as:

$$e_{t+1} = \sqrt{h_{t+1}}v$$

where the error e_{t+1} is now sampled from a normal distribution with mean zero and variance h_{t+1}.

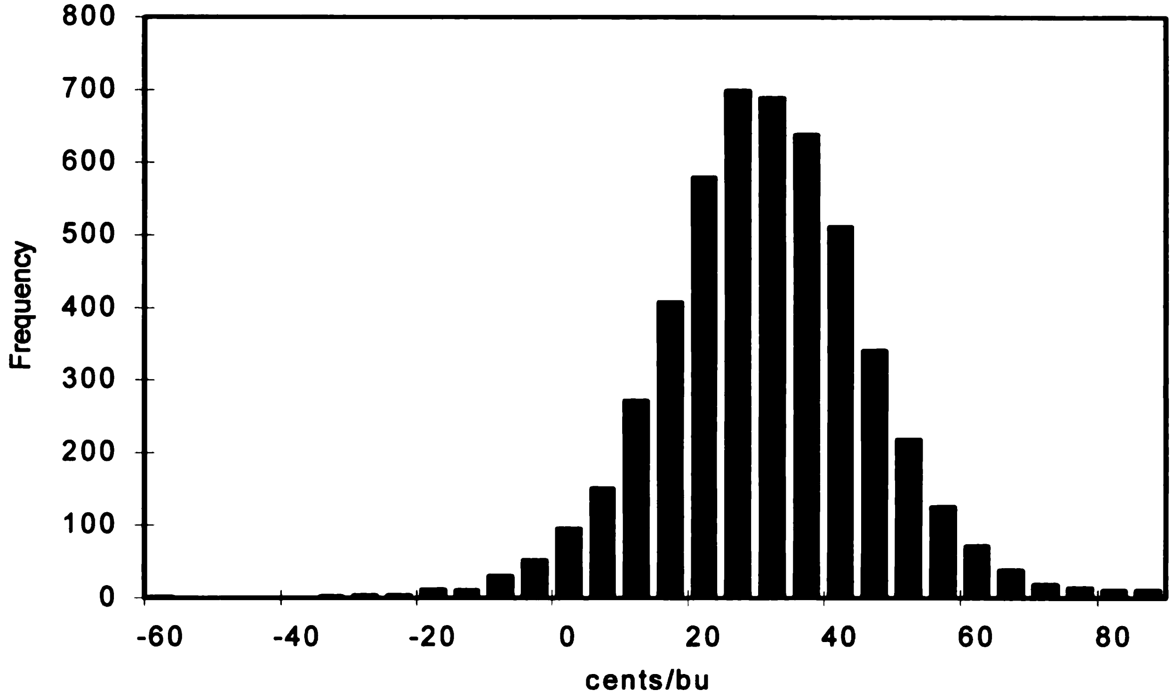
4. Calculate basis price at $t+1$ by using the errors calculate from 3.

$$b_{t+1} = \mu + \sum_{i=1}^{p} \alpha_i b_{t+1-i} + \sum_{j=1}^{q} \beta_j e_{t+1-j} + S_{t+1} + e_{t+1}$$

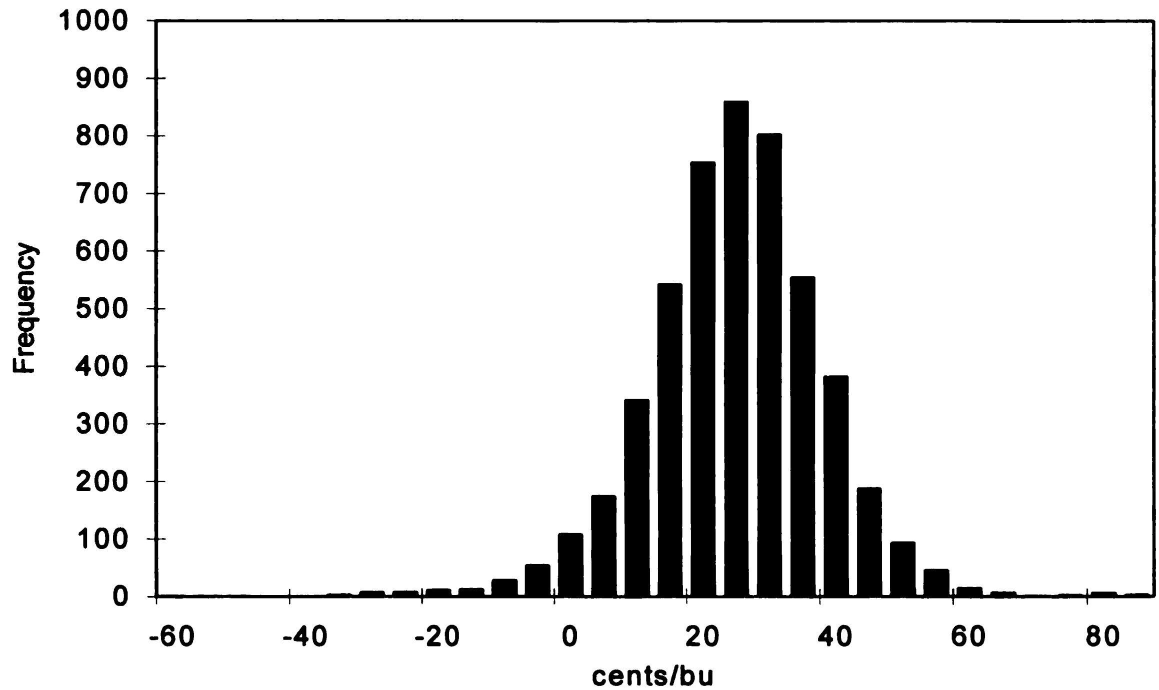
5. Let $t=t+1$ and repeat step 1 – 4 to calculate basis for next period until one realization of b_T is obtained.

6. Repeat step 1-5 n times to generate an n-size sample of the basis price distribution at maturity time, where n is 5000 in this study.

The histogram of the simulated distributions of basis prices at the maturity time (m = 30) from ARCH and GARCH-M models are shown in figure 2. The first four moments of the basis distribution are calculated and listed in table 3.



2(a) Forecasted July basis distribution from ARCH model



2(b) Forecasted July basis distribution from GARCH-M model

Figure 2 Simulated July Basis Distributions

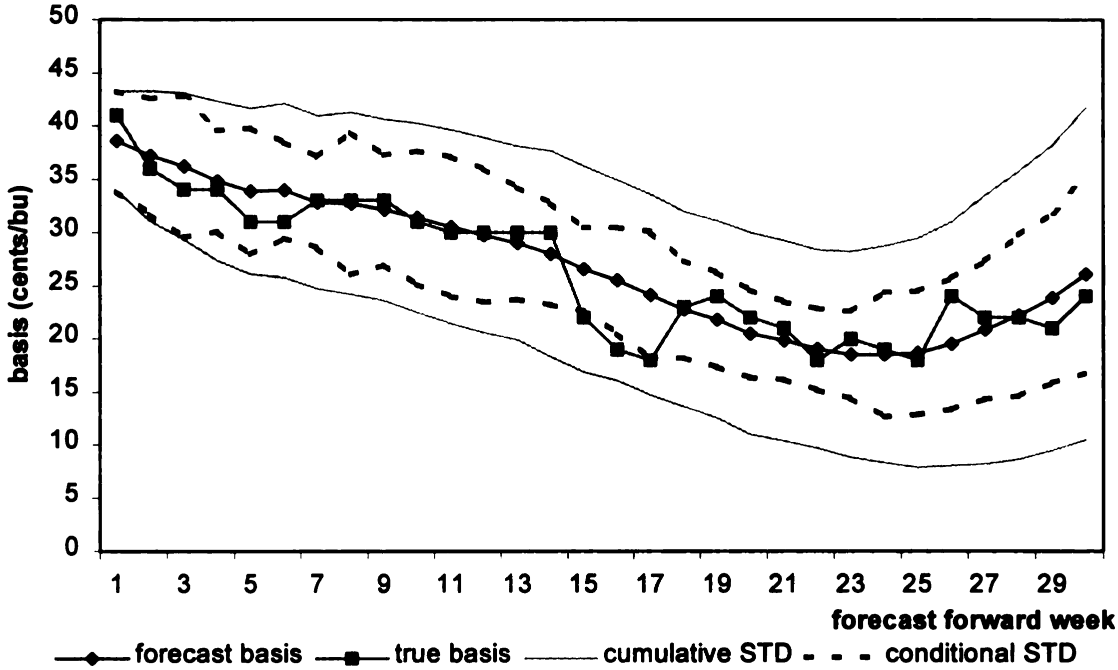
Table 3 Statistics of the Simulated Basis Distributions

Cents/bu	Mean	Standard Deviation	Skewness	Kurtosis
ARCH(6)	26.2763	9.1036	-0.1447	4.9679
GARCH(1,3)-M	22.3223	13.9754	-0.3525	6.7353

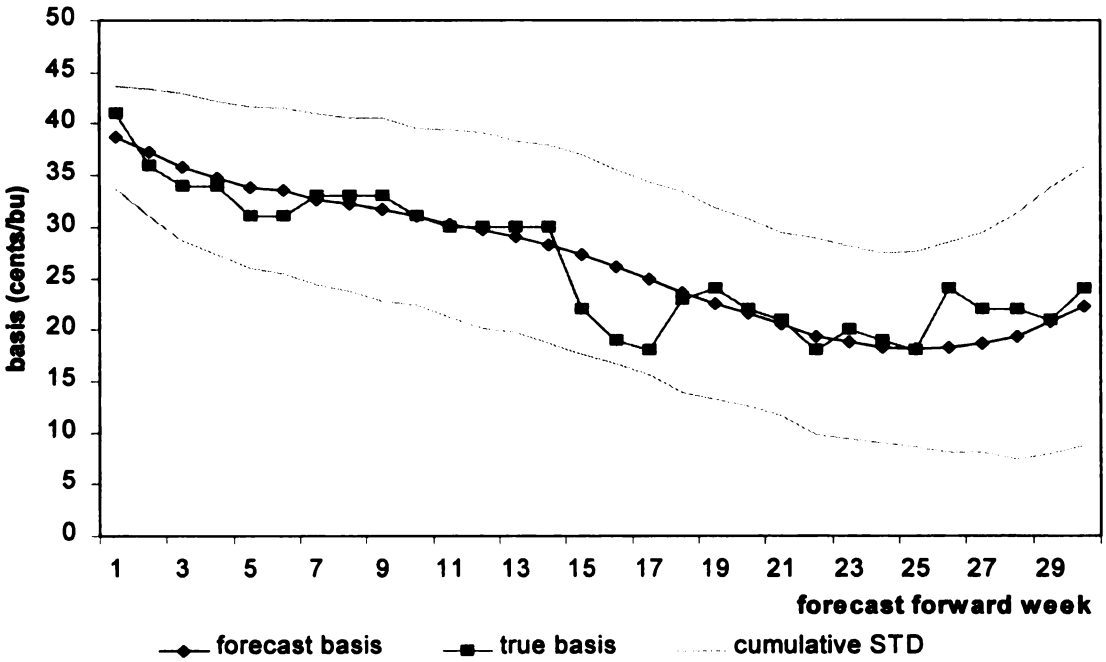
For July 1998, the simulated basis distributions have mean level of 26.2763 and 22.3223 cents/bu from ARCH and GARCH-M models respectively. The realized basis value is 24 cents/bushel. As expected from the estimated basis process, the simulated basis distribution at maturity time is not normally distributed. Both simulated distributions are skewed slightly to the left and have excess kurtosis.

The in-sample forecasts from both models appear to track the basis realizations well. The true test of the models is to forecast beyond the period of fit, or out-of-sample. Also, in-sample forecasts should necessarily perform well. Figure 3 shows out-of sample forecasts of basis.

Note the forecasts of weekly basis from ARCH and GARCH-M models are very similar. Both forecasts seem to do a good job capturing the trend of basis series except during weeks 15 - 17. We can see that forecasted basis narrows from December to mid June (week1 to 26) and then begin to widen after mid June. This is because corn harvest reaches a peak at early November and cash prices usually fall to their lowest point of the season during November and then rise gradually during the storage season.



3(a) ARCH Model



3(b) GARCH-M model

Figure 3 Dynamic Forecasts of weekly Corn Basis

Figure 3(a) also shows the conditional standard deviation of the basis forecasts using a dotted line. The conditional standard deviation measures the basis risk for the corresponding forecasted basis levels and it varies over time. The relatively larger conditional standard deviation at the last few weeks of forecasts implies larger volatility or larger basis risk at delivery, partially reflecting that futures prices usually are quite variable at contract maturity time. From week 27 to 30, the forecasts of basis level from both models widen slightly, also reflecting the variability of the storage basis at maturity time.

Unlike the conditional standard deviation, the cumulative standard deviations for the basis forecasts increase as we forecast further into the future reflecting an increase in the size of the confidence intervals of the forecasted mean as we forecast further into the future.

Evaluate Different Basis Models

Although the out-of-sample forecasts are similar for these two basis models, we can use statistics to help evaluate which does a better job of forecasting during the period under consideration. We use the root mean squared forecast errors (RMSFE) to evaluate the two basis models. The function of RMSFE is calculated as follows:

$$\text{RMSFE} = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(b_t^f - b_t^r)^2}$$

where b_t^f are the forecasted basis and b_t^r are the realized basis; m is the number of forecasted observations. The b_t^f are derived from the two alternative models using only information available before the out-of-sample forecast period (December 17, 1997).

RMSFE for the ARCH model is 2.4747 cents and RMSFE for the GARCH-M model is 2.6967 cents.

The RMSFE results show that ARCH basis model has smaller forecast error, thus performs better in terms of forecasting basis levels out-of-sample. It's important to note that this evaluation only represents one storage season realization. The results could differ if a more extensive performance evaluation was conducted. However based on these results, we choose the ARCH model to forecast the basis distribution.

Forecasting Cash Price Distribution

Black's Model and Implied Volatility

As mentioned earlier, commodity option models can be used to generate implied volatility of futures price. Implied volatility estimation is based on the fact that options are forward looking assets and therefore contain information on futures distributions. The implied volatility reflects the expectation of market participants from the estimation day to the options expiration day. Conditional on an option model, the volatility of the futures distribution is implied from market traded option prices instead of historical time-series data of the futures prices. Thus with the predictability and relatively low variability features of basis, combining implied futures price distributions with our forecasted basis distribution may outperform the forecasts of cash distributions either from structural models or historical time series models.

Black uses arbitrage arguments to obtain an exact formula for pricing options on futures contracts. One property of this Black's differential equation is that the equation does not involve any variables that are affected by the risk preference of investors. The variables that appear in the equation are the current futures price, strike price of options, time to maturity, futures price volatility at maturity, and the risk-free rate of interest. All are independent of risk preferences (J. C. Hull, 1993). If we know these variables, we can compute the corresponding call option price and put option price. The Black's formulas for both European call and put options are presented below:

$$(5)\quad \begin{aligned} P_C &= e^{-rT} * [S * N(d_1) - X_C * N(d_2)] \\ P_P &= -e^{rT} * [S * N(-d_1) - X_P * N(-d_2)] \end{aligned}$$

Where: $d_1 = [\ln(S / X_i) + (r + 0.5 * \sigma^2) * T]/(\sigma * \sqrt{T})$

$d_2 = d_1 - (\sigma * \sqrt{T})$

P_i = value of the option; where P_C for call options and P_P for put options

S = price of the current futures

X_i = strike price of the option; where X_C for strike price of the call option and X_P for strike price of the put option

T = time to expiration in terms of proportion of a year

r = annual interest rate on low risk securities

σ = annualized standard deviation

$N(d)$ = cumulative normal probability values of d

A useful insight for our purpose is that if we know the price of the option and we can compute which value of σ will produce that price, that is, implied volatility of the futures price without adding more exogenous variables (Hull 199). We invert (3) to find a formula that expresses the implied volatility as a function of the option price and remaining needed parameters, that is $\sigma(P_i, S, X_i, T, r)$. We can use maximum likelihood method to calculate the implied volatility σ by minimizing the sum of the squared pricing errors of all options in the sample given by:

$$\min_{\sigma}[(P_C - \hat{P}_C(S, X_C, T, r; \sigma))^2 + (P_P - \hat{P}_P(S, X_P, T, r; \sigma))^2]$$

where $\hat{P}_C(\bullet)$ *and* $\hat{P}_P(\bullet)$ are Black option pricing formulas in (5). Pricing errors are the difference between the actual option prices (P_i) and the option prices estimated from the implied volatility. The data we used are July corn futures option price and option strike price with largest trading volume on Dec. 17, 1997 and interest rate for 52 week-term treasury bill. The estimate result of variance for July corn futures are shown below:

$$\sigma(P_C, P_P, S, X_C, X_P, T, r) = 0.3165$$

where P_C =15 cents, P_P =13.25cents, S =286 cents/bu, X_C =290 cents/bu, X_P=280cents/bu, T =180/365, r = 5.468%.

A fundamental assumption in Black's model is that the volatility of a futures price is exactly the same at each instant. That is, the variance of the futures price is constant over time. Therefore, the volatility of July's futures would always be σ^2 from harvest time to maturity time. In addition, Black's model assumes that futures distribution is lognormal.

In reality, however, if we compute volatilities for futures price, we find that volatilities change across time. Why do we use Black's model to estimate the futures price distribution? One compensating factor is that the volatility that makes the Black's value agrees with the market price include all available market information. So if the market participants have recognized the true distribution of the future price, the market traded value of the options will reflect of some correction for the model misspecification.

Hopefully Black's model is a close representation of the true volatility of futures price in the market (J. C. Hull, 1993).

There are other pricing models to generate the implied volatility of futures price. For example, Hilliard and Reis (1999) suggest that a jump-diffusion option pricing model may be more appropriate by allowing non-normal distribution. However, Black's model is still a widely used option pricing model and we will use it to generate forecasts of the futures price distributions.

Correlation between Basis and Futures Prices

A critical issue when we combine the distribution of basis with distribution of futures is the possibility of correlation between the basis distribution and the futures distribution. If they are independent, then these two distributions can simply be linearly summed to generate the cash price.

To begin, let b_m, f_m and c_m represent the basis, futures and cash price at futures contract maturity time, respectively. Then the covariance between basis and futures prices is:

$$\operatorname{cov}(b_m, f_m) = \operatorname{cov}(f_m - c_m, f_m) = \operatorname{var}(f_m) - \operatorname{cov}(f_m, c_m)$$

At futures contract maturity time, if $f_m = c_m$, then

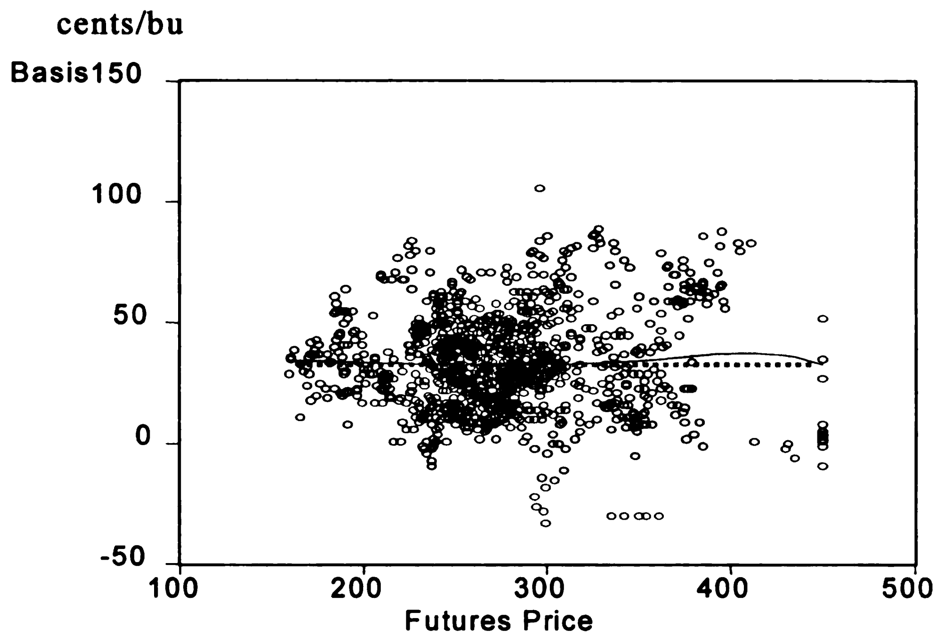
$$\operatorname{cov}(b_m, f_m) = \operatorname{var}(f_m) - \operatorname{cov}(f_m, f_m) = 0$$

Theoretically, if futures price and cash prices converge at contract maturity time to a constant (e.g. transportation costs) then the basis distribution and futures distribution may be independent at maturity. To investigate the independence between futures distribution and cash distribution further, we conduct several statistical tests.

First, a formal independence test for two time series based on cross correlation theory was conducted. According to the theorem, cross correlation of two stationary time series equal to 0 indicate the independence of the two series. Based on the method developed by Brockwell and Davis (1996), we can test independence by testing cross correlation. The test result shows the futures price does not have much influence on basis series (see Appendix 3).

Second, we use Kernel function to estimate conditional mean of basis condition on futures price. Kernel is the function used to weight the observations in each regression of the first series Y on second series X in the group. Figure 4 shows scatter plot of basis and futures price data and Kernel fit plot. The Kernel fit are non-parametric regressions for basis, at each value of futures, by choosing the parameters to minimized the weighted sum-of-squared residuals. The plot of kernel fit represents the conditional mean of basis condition on futures prices. From Figure 4, we can see the conditional mean line (solid curve) is close to the unconditional mean (dotted line), which means the basis prices are not influenced by futures prices.

These test results indicate that the basis distribution and the futures distribution have no strong impact on each other. The assumption of independence between basis and futures price makes the simulating procedure of cash distribution relatively simple and based on the above analysis should not have a serious impact on forecasting results.



... Unconditional mean of basis — Kernel fit plot of conditional mean of basis

Figure 4 Kernel Fit Plot of Basis Condition on Futures Price

Simulating the Corn Cash Price Distributions

With the implied volatility of futures price generated from Black's option model, we can forecast the futures distribution using available market information. Then we combine forecast of basis distribution with futures price distribution to get a forecast of the cash price distribution.

Again, let b_t and f_t be the initial basis and futures price respectively on Dec. 17, 1997. The terminal date is the third week of June 1998, which is the assumed maturity time of option for July corn futures.

Since basis and futures prices are assumed to be independent, we can simulate the cash price distribution as follow:

I. Generate log futures price at $t+m$ simply by:

$$\lg f_{t+m} = \lg f_t + \lg \sigma \sqrt{T} * u$$

where f_t is current futures price; σ is the annualized standard deviation that get from Black's model; T is time to expiration in terms of proportion of a year (number of days to maturity/365); $\sigma * \sqrt{T}$ can adjust standard deviation to the period of forecasting; u are random draws from independent standard normal distribution, so that ($\sigma * \sqrt{T} * u$) are independently distribution normal vector with mean 0 and variance $\sigma^2 T$.

II. Take the anti log to obtain one futures price realization at period m.

III. Follow steps 1 – 5 to generate a basis realization (b_{t+m}).

IV. Calculate one realization of the cash price at $t+m$ as:

$$C_{t+m} = f_{t+m} - b_{t+m}$$

V. Repeat step I to IV n times to generate an n-size sample of the cash price distribution at maturity time, where n is 5000 in this study.

Using the above simulation approach, we can obtain a terminal distribution of cash price conditional on information available at the time of the decision problem by combine basis distribution with futures distribution.

The histograms of the simulated distributions of futures and corn cash prices for the maturity time (m = 26) are shown in figure 5 and figure 6, respectively. The first four moments of the futures and corn cash distribution are calculated and listed in table 4.

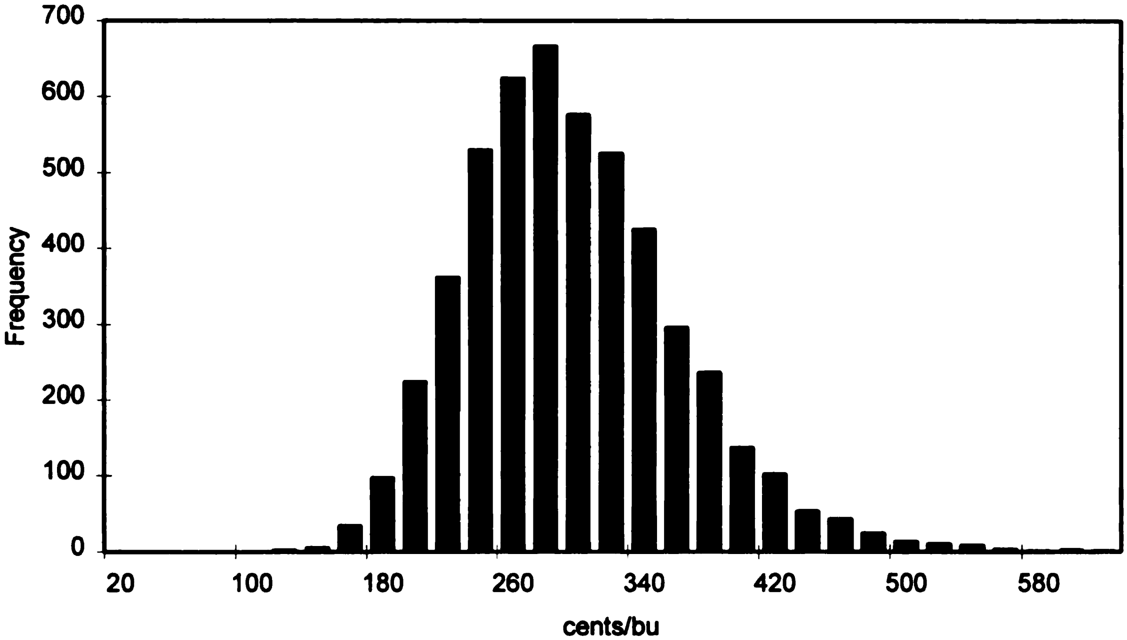


Figure 5 Simulated 1997 Corn Futures Price Distribution at Options Maturity Time

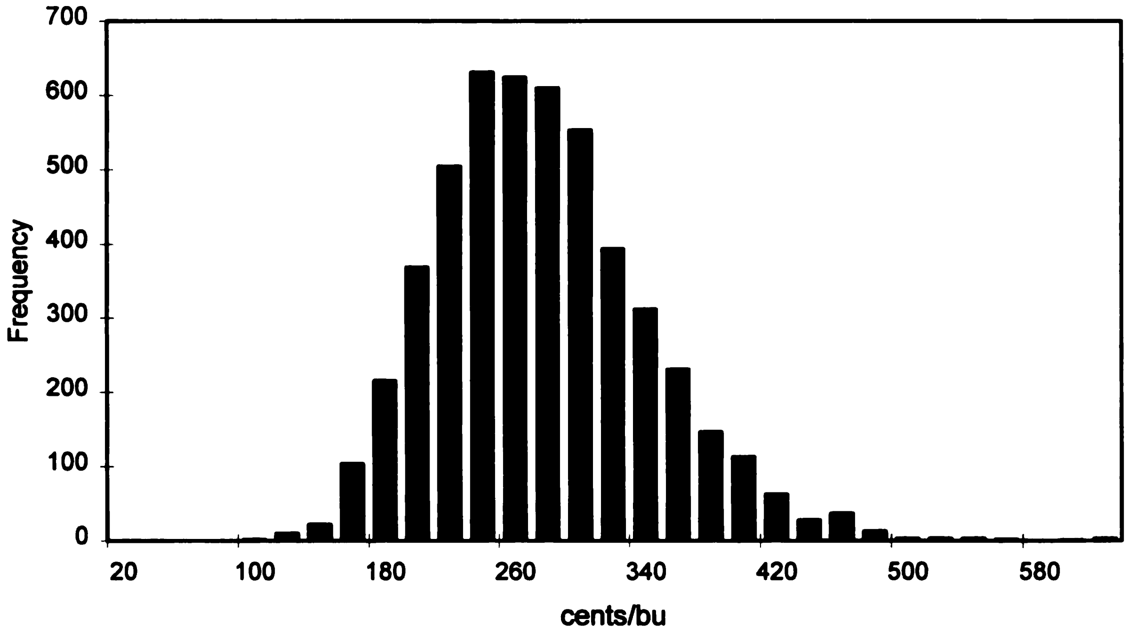


Figure 6 Simulated 1997 Corn Cash Price Distribution at Options Maturity Time

Table 4 Statistics of Simulated Futures and Cash Price Distribution at Option Maturity Time

Cents/bu	Mean	Standard Deviation	Skewness	Kurtosis
Futures	286.2972	65.9286	0.7085	3.9368
Cash	266.9008	67.3788	0.5973	3.6559

The mean level of futures distribution at options maturity time is 286.3 cents/bu, consistent with the futures price that we used in the Black's function (286 cents/bu). As expected, the futures price distribution follows a lognormal distribution.

Combining the forecast basis distribution from basis model and forecast futures distribution from Black's option model, the simulated cash distribution at maturity time has a mean of 266.9 cents/bu, a larger cumulative standard deviation, less skewness and less kurtosis than the forecasted futures price distribution. Because this combined cash price distribution includes historic price information from the basis model and available market information from Black option model, it could outperform the cash distribution forecasted from a structural model or a pure time series model.

Conclusion

Basis is a key element of effective hedging strategies in futures markets. The relatively predictability and stability of basis also can contribute to forecasting cash price distributions. This study has developed time series models to estimate and forecast the storage basis distribution for corn. A forecast of the cash distribution is then generated by combining the forecast of the basis distribution with a market implied forecast of the futures price distribution. The data used to parameterize the model are representative of a corn farm in Saginaw, Michigan from 1974 to 1997 crop year.

Previous studies on basis primarily concentrate on the basis behavior; a few studies focus on forecasting basis level, but little work has been done on modeling and forecasting basis distribution. This research studies the stochastic properties of the basis and provides evidence that multiple-year basis data is stationary. Based on the stochastic properties of the basis data, a type of autoregressive moving average processes with an autoregressive conditional heteroscedasticity model (ARCH) is specified to estimate and forecast the basis distributions. The forecasted basis distribution can be used to improve farmers' hedging decisions.

Compared to cash prices, basis is reasonably predictable and has smaller volatility. Most previous studies on modeling and forecasting commodity cash price distributions use time series methods. These studies typically use historical price to estimate these models. However, implied futures prices volatilities generated from option pricing models contain available market information at the time the option is priced. Using simulation techniques, we can combine forecasts of the basis distribution with the implied futures price distributions to get a forecast of the cash price distribution. Combining basis distributions with market implied futures price distributions results in a new approach to generate a forecast of the commodity cash price distribution. With the predictability features of basis distributions and the available market information reflected in traded commodity option prices, this combination may outperform the forecasts of cash price distributions either from structural models or pure time series models. Thus this study develops and illustrates a method that may improve the precision of forecasted cash price distributions.

This research considers only one crop; however it provides some basic results on storage basis behavior and modeling which can be generalized to other commodities. This approach to forecasting cash price distributions includes available market information could outperform alternative structural or time series models. Future research could focus on testing the performance of the approach improving the implied volatility forecast of futures price by using alternative option pricing models.

APPENDICES

APPENDIX 1

Basis Definitions

The "**nearby basis**" (or **continuation basis**) is simply the cash price on a given day subtracted from that day's futures price for the next available maturity contract month. In term of (1), T corresponds to the nearest futures maturity after t and continually roles forward to the next contract month whenever $t = T$. Continuation basis provides a continuous measure of the "nearby" basis across time. It is a popular measurement in technical analysis. However, continuation basis is somewhat difficult to interpret across time because the underlying futures contract changes (i.e. the underlying asset being priced is different).

The "**new crop basis**" is the futures price nearest to and after harvest less the cash forward contract price, that is, $f_t(T)$ - $E_t(p_T)$. Here, $E_t(p_T)$ is expected harvest cash price at time T conditional on information available at t, often represented by the forward contract price for the new crop. T corresponds to the contract month nearest and after harvest. New crop basis provides a market estimate of the expected basis and provides useful information to aid in making pricing decisions by allowing users to calculate the local expected cash price for a commodity. Grain buyers can adjust the size of the new crop basis to control grain flows. Buyers who wish to encourage delivery of additional grain reduce the size of the new crop basis by increasing the relative forward price.

Producers (grain buyers) who wish to hedge future production sell (buy) futures contracts before harvest and buy (sell) them back at harvest time in effort to offset price changes in their long (short) position in cash market. In this case, the "**harvest basis**"

$f_T(T)$ - p_T plays a key role in pre-harvest hedging decision. Here the futures contract month T is at harvest. This is an important basis measure for farmers hedging their production or buyers hedging input costs at harvest.

The "**storage basis**" is the current cash price subtracted from the futures price for some month during the storage season. In (1), T can correspond to the any contract month during the storage season. The storage basis represents the return to storing the commodity from t to T and transporting it from the local region where cash price p_t is offered to the delivery point on the futures where $f_t(T)$ is offered. When the contract matures ($t = T$), there will be no additional storage return and b_t represents transportation costs. For storage operators hedging in the futures market, knowledge of the storage basis can help them design effective hedging strategies. The harvest basis and the storage basis are most important concepts for hedgers.

APPENDIX 2

Unit Root Tests on Basis Data

Dickey and Fuller (1981) have developed unit root tests involving the OLS regression technique using standard t and F statistics that follow non-standard distributions to test for the present of a stochastic trend (unit root). Augmented Dickey-Fuller (ADF) tests improved original Dickey-Fuller tests by including lagged differences of the dependent variables as additional explanatory variables to account for autocorrelation in the residuals from Dickey-Fuller regressions (Dickey and Fuller, 1981).

The Augmented Dickey-Fuller unit root test is conducted on the multi-year basis sample. Since a trend can have deterministic and stochastic components, the testing procedure is confounded by the presence of the deterministic regressors. We begin by removing the deterministic component from the data. The characteristics of storage commodity suggest the presence of a seasonal pattern in the storage basis. We use sine and cosine function to capture the seasonal effects. With the following regression:

$$(1\text{a})\, b_t = \omega + \alpha t + \sum_{j=1}^{J}[\phi_i \sin(2\pi i w/52) + \psi_i \cos(2\pi i w/52)] + e_t$$

where b_t represents basis at time t, J is the degree of frequency in seasonality variables, and w is week, e_t is the residual at time t. The order of J is determined by using likelihood ratio (LR) test.

The log-likelihood ratio test states that the twice of the difference between the log-likelihood of the unrestricted model and the restricted model is X^2 distributed with the number of restrictions as the degree of freedom. The test hypothesis is:

H_0: $\phi_i = \psi_i = 0$

H_1: $\phi_i \neq 0$ or $\psi_i \neq 0$

Calculate the likelihood Ratio of the two model specifications:

$LR = 2(L_U - L_R) \sim X^2(2)$

where L_U and L_R are the log-likelihood of the unrestricted and restricted model, respectively. The log-likelihood and LR tests are listed in table below:

LR tests for degree of frequencies in seasonality

J	Log-likelihood	LR	H_0
5	501.1062	1.2996	$\phi_5 = \psi_5 = 0$
4	500.4564	7.5686	$\phi_4 = \psi_4 = 0$
3	496.6721	14.561	$\phi_3 = \psi_3 = 0$
2	489.3916	42.6166	$\phi_2 = \psi_2 = 0$
1	468.0833		

Critical value for $X^2(2)$5%: 5.99147; 1%: 6.6349

The null hypothesis of no seasonality versus the alternative of 4 degree of seasonality can rejected at 1% significant level.

The test results provide some evidence that the seasonal factor may be characterized by fourth order frequency of sine and cosine variables. So we get $J = 4$.

Using the residuals from (1a), ADF test can be conducted using the following regression:

$$(1b)\ \Delta e_t = \beta e_{t-1} + \sum_i^k \gamma_i \Delta e_{t-i} + residual_t$$

H_0: β=0 unit root process

H_1: β<0 stationary process

Since we use the residuals after regressing the deterministic seasonality from the basis data, ADF test doesn't include constant variable again.

In order to eliminate the contract switch effects, we conduct the unit root tests on multiple-year basis data by taking out the difference (Δe_t) occurred between the two contracts from the data process. This technique only allows the data differencing in the contract and thus eliminates the contract switches effects. However, it causes 23 observations lost each time we include one more lag of the dependent variable.

An additional difficulty is that the true order of the autoregressive process is usually unknown, so that the problem is to select the appropriate lag length. We can start with a relatively long lag length and pare down the model by the usual *t*-test and/or *F*-tests. Repeat the process until the lag is significant different from zero. Moreover, the correlogram of the residuals should appear to be white noise. We start with 53. The Ljung-Box *Q*-test of first 100 lags shows no autocorrelation in the residuals at 5% significant level. Then we decrease lag one by one. The *Q*-tests show autocorrelations in residuals at 1% significant level until we decrease to lag 14. So we include lag 15 in the regression.

Then use OLS to estimate value of β and it's associated t-statistic. Comparing the resulting t-statistic with the critical value reported by Dickey and Fuller allows us to determine whether to accept or reject the null hypothesis $\beta = 0$. Based on the unit root regression equation, we use τ_t critical value at 5% significant level (-3.41) for regression (1b). The t-statistic for β in the test is –6.8747, less than –3.41 (τ_t). We reject the null hypothesis that the process is unit root. Previous studies find one may fail to reject the null hypothesis of a unit root because of a mis-specification concerning the deterministic part of the unit root regression. If the null hypothesis of a unit root is rejected, there is no need to proceed (Ender 1995). Thus, price level will be evaluated throughout this study.

APPENDIX 3

Independent Test for Basis and Futures Price Distributions

H_0: b_t and f_t are independent series;

H_1: b_t and f_t are not independent series.

According to the theorem 7.3.1 in Brockwell and Davis (1996), cross correlation of two stationary time series equal to 0 indicate the independence of the two series. If one of the two processes is white noise, then the cross correlation between the two processes is approximately normally distributed with mean 0 and variance $1/n$, that is cross correlation $\rho_{12} \sim N(0, 1/n)$. This make the independence test straightforward to test the hypothesis of $\rho_{12} = 0$. Since the basis and futures prices are both stationary series by previous unit root tests, we can develop a method to test for independence of them by testing cross correlation.

Any test for independence of the two component series cannot be based solely on estimated values of cross correlation without taking into account the nature of the two component series. In practice, the difficulty can be solved by transforming the two series to white noise by application of ARMA method. The estimations for basis and futures prices are shown in Table A and B:

Table A: Estimation of ARMA model for Basis

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.316979	0.064708	4.898580	0.0000
SIN1	0.110637	0.013186	8.390291	0.0000
COS1	-0.073650	0.011493	-6.408183	0.0000
SIN2	0.013617	0.010696	1.273139	0.2032
COS2	-0.041254	0.009520	-4.333391	0.0000
AR(1)	0.920142	0.049804	18.47513	0.0000
AR(6)	0.065791	0.049334	1.333583	0.1826
MA(1)	-0.200366	0.092397	-2.168548	0.0303
MA(5)	-0.100803	0.070443	-1.430987	0.1527
MA(6)	-0.245944	0.098230	-2.503760	0.0124
MA(8)	-0.090963	0.041591	-2.187081	0.0289

Table B: Estimation of ARMA model for Futures Prices

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.738024	0.141283	19.37966	0.0000
AR(1)	1.011541	0.016388	61.72516	0.0000
AR(6)	-0.030709	0.015820	-1.941176	0.0525
MA(3)	-0.113554	0.033199	-3.420431	0.0006
MA(5)	-0.050144	0.033635	-1.490823	0.1363
MA(7)	0.100556	0.029205	3.443086	0.0006
MA(8)	0.076515	0.029407	2.601954	0.0094

Check the residuals from the ARMA models for basis and futures, respectively (w_{t1} and w_{t2}). There is no autocorrelation at residuals. They are white noise sequences. Then an approximately test of independence between basis and futures series is independence test of w_{t1} and w_{t2}:

H_0: w_{t1} and w_{t2} are independent series;

H_1: w_{t1} and w_{t2} are not independent series.

This test equal to:

H_0: Cross correlations between w_{t1} and w_{t2} (ρ_{12}) = 0;

H_1: Cross correlations between w_{t1} and w_{t2} (ρ_{12}) $\neq 0$.

At 5% significant level, compare the absolute values of cross correlation between w_{t1} and w_{t2} with $1.96n^{-\frac{1}{2}}$ (n =1218). Under H_0, for any fixed $\rho_{12}(h), h = 0, \pm 1, \cdots$, falls between the bound ± 0.0566 with a probability less than 95%. When we calculate $\rho_{12}(h), h = 0, \pm 1, \cdots \pm 150$, all except eleven of the correlations are between the bounds of ± 0.0566. $11/301 = 0.0365$ < 5%, suggest that the two residual series are uncorrelated. None hypothesis of independence is fail to reject. Hence we can conclude that the basis and futures price series are independent.

REFERENCES

References

Ardeni, Pier G. "Does the Law of One Price Really Hold for Commodity Prices?" *American Journal of Agricultural Economics,* v71(August 1989), p661-669

Baillie, R. T. and R. J. Myers, "Bivariate GARCH Estimation of The Optimal Commodity Futures Hedge" *Journal of Applied Econometrics*, v6(1991): p109-124

Bollerslev, T. "Generalized Autoregressive Conditional Heteroskedasticity" *Journal of Econometrics* v31(1986): p307-327

Bollerslev, T "A conditional Heteroskedastic time series model for speculative prices and rates of return", *Review of Economics and Statistics*, v69(1987): p542-547

Bollerslev, T., R. Y.Chou and K. F. Kronoer "ARCH modeling in finance -A review of the theory and empirical evidence" *Journal of Econometrics* v2(1992): p5-59

Brennan, M. J. "The Supply of Storage" *American Economics Review* v48(1958): 195-206

Brockwell, P. J. and R. A. Davis *Introduction of Time Series and Forecasting* 1996 Springer-Verlag New York, Inc. p230-232

Chriss, N. A. *Black-Scholes and Beyond: Option Pricing Models* 1997 McGraw-Hill, p327-360

Dickey, D.A. and W. A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series With A Unit Root", *Econometrica* v49, n5(1981): p1057-1072

Ender, W. *Time Series Model,* New York: Wiley 1995

Engle, R. F. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation" *Econometrica*, v50, n4 (July, 1982)

Fama, E. F. and K. R. French "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage" *Journal of Business*, v60, n1(1987)

Ferris, J. N. *Lecture Material for Commodity and Futures Marketing*, Department of Agricultural Economic, Michigan State University, 1997, p17-91

French, K. R. "Detecting Spot Price Forecasts in Futures Price" *Journal of Business*, v59, n2,(1986)

Fortenbery, T. R., and H. O. Zapata , "An Examination of Cointegration between Futures and Local Grain Markets" *The Journal of Futures Markets*, v13(1993): p921-932

Galloway, T. M. and R. Kolb "Futures Price and the Maturity Effect" *The Journal of Futures Market*, v16(1996): p809-828

Garcia, P., and R. M. Leuthold and M. E. Sarhan "Basis Risk: Measurement and Analysis of Basis Fluctuations for Selected Livestock Markets" *American Journal of Agricultural Economics* v66 (1984): p499-504

Garcia, P. and D. R. Sanders "Ex Ante Basis Risk in the Live Hog Futures Contract: Has Hedges' Risk Increased?" *The Journal of Futures Market* v16, N4 (1996), p421-440

Goodwin, B.K. and T. C. Schroeder, "Cointegration Tests and Spatial Price Linkages in Regional Cattle Markets", *American Journal of Agricultural Economics* v73, n2(1991), 452-464

Goodwin, B.K. (1992), "Multivariate Cointegration Tests and the Law of One Price in International Wheat Markets", *Review if Agricultural Economics* v14, n1(1992), 117-124

Gordon, J. D. "The Distribution of Daily Changes in Commodity Futures Pieces" *Technical bulletin NO.1702* ERS USDA (1985)

Gujarati, D. N. (1995), *Basis Econometrics,* p709-729, McGraw-Hill, Inc. Third Edition

Hilliard, J. E. and J. A Reis "Jump Processes in Commodity Futures Prices and Options Pricing" *Aamerican Journal of Agricultural Economics* v81, n2(1999): p273-286

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

Kaldor, N. "Speculation and Economic Statility" Review of Economic Studies, v7(1939): p1-27

Kahl, K. H. and C. E. Curtis " A Comparative Analysis of the Corn Basis in Feed Grain Deficit and Surplus Areas" *Review of Research in Futures Markets* v5, n3(1986): p220-232

Kamara, A. "Issue in Futures Markets: A Survey" *Journal of Futures Market* v2(1982): p261-294

Kenyon, D., E. Jones and A. McGuirk "Forecasting Performance of Corn and soybean Harvest Futures Contracts" *American Journal of Agricultural Economics* v75(1993): p399-407

Khoury, N. T. and J. Martel "A Supply of Storage Theory with Asymmetric Information" *Journal of Futures Market* v9, n6 (1989): p573-581

Kim, D. and S. J. Kon "Alternative models for the conditional heteroscedasticity of stock returns" *The Journal of Business* v67, n4(1994): p536-563

Lence S. H., M. L. Hayenga and M.D. Patterson "Storage Perfotability and Hedge Ratio Estimation" *The Journal of Futures Markets* v16, n6(1996): p655-676

Liu S., B. W. Brorsen, C. M. Oellermann and P. L. Farris " Forecasting the Nearby Basis of Live Cattle" *The Journal of Futures Markets* v.14, n3(1994): 259-273

Mann, J. S. and R. G. Heifner " The Distribution of Short-run Commodity Price Movement" *Technical Bulletin NO. 1536* ERS USDA (1976)

Martin, L., J. L. Grotnewegen and E. Pidgeon "Factors Affecting Corn Basis in Southwestern Ontario" *American Journal of Agricultural Economics* v62(1980): p107-112

Myers, R. J. "Time Series Econometrics and Commodity Price analysis: A Review" *Review of Marketing and Agricultural Economics* v62, n2 (1994)

Myers, R. J. and S. D. Hanson "Optimal Dynamic Hedging in Unbiased Futures Markets" *American Journal of Agricultural Economics*, v78, n1(1996): p13-20

Oxley, L., D. A.R. George, C. J. Roberts and S. Sayer *Surveys in Econometrics* Blackwell Publishers 1995 p215-273

Streeter, D. H. and W. Tomek "Variability in Soybean Futures Prices: An Integrted Framework" *Journal of Futures Markets* v12 (1992): p705-728

Tilley, Daniel and S. K. Campbell "Performance of the Weekly Gulf-Kanasas City Hard-Red Winter Wheat Basis" *American Journal of Agricultural Economics* v70 (1988): p929-935

Tomek, W. G. & Gray R.W. "Temporal Relations among Prices on Commodity Futures Market: Their Allocative and Stabilizing Roles." *American Journal of Agricultural Economics,* v152 (1970): p372-380

Tomek, W. G. and K. L. Robinson *Agricultural Product Prices Cornell University Press* (1984) 224-230

Tomek, W. G. and R. J. Myers " Empirical Analysis of Agricultural Commodity Prices: A Viewpoint" *Review of Agricultural Economics*, v15, n1 (Jan. 1993): p181-202

Tomek, W. G. "Commodity Futures Price as Forecasts", *Review of Agricultural Economics* v19, n1(1996): p23-44

Venkateswaran, M. and B. W. Brorsen and J. A. Hall "The Distribution of Standardized Futures Price Changes" *The journal of Futures Markets*, v13(1993): 279-298

Wang, H. "Farmer Risk Management Behavior and Welfare under Alternative Portfolios of Risk Instruments" Ph.D. Dissertation, Department of Agricultural Economics, Michigan State University, 1996, p56-108

Working, H. "The Theory of Price of Storage" *American Journal of Agricultural Economics* v39(1949): p1254-1262

Yang, S. and B. W. Brorsen "Nonlinear Dynamics of Daily Cash Prices", *American Journal of Agricultural Economics*, v74(1992), p706-715