December 1987 A.E. Res. 87-32

YEAR-SPECIFIC ESTIMATION OF OPTIMAL HEDGES FOR CENTRAL ILLINOIS SOYBEAN PRODUCERS

Joy L. Harwood
and
William G. Tomek

Department of Agricultural Economics
Cornell University Agricultural Experiment Station
New York State College of Agriculture and Life Sciences
A Statutory College of the State University
Cornell University, Ithaca, New York 14853-7801

Abstract

This study estimates year-specific optimal hedges for Central Illinois soybean producers. The optimal hedge objective function employed is based on harvesttime cash price and basis levels and their respective variances and covariance, where farmers are assumed to make anticipatory hedges of production decisions. Past studies which use such objective functions have typically assumed that the parameters of price distributions are constants over the sample period. This study, in contrast, addresses the measurement of optimal hedge parameters, permitting both means and variances to change through time.

Cash price and basis equations are specified and estimated initially using ordinary least squares (OLS). These equations are then examined for the presence of non-constant error variances (the property of heteroskedasticity). Specifically, it is hypothesized that increases in use relative to stocks for soybeans may be associated with increasing variability in the equations' errors.

The cash price and basis equations are also modelled as two seemingly unrelated regression systems: one with constant variances and the other with non-constant variances. The dependent variables provide estimates of the cash price and basis levels in the optimal hedge equation. The equations' residuals provide the base for estimating the variance and covariance parameters of the optimal hedge.

The models were fitted with 1965-1984 data. With conventional (constant parameter) models, the optimal (short) futures position for the risk averse hedger is nearly 100 percent of expected production. With the non-constant variance specification, optimal hedges for the risk averse hedger ranged from 75 to 100 percent of production. The non-constant variance specification, however, had a somewhat unexpected result: the heteroskedasticity is related to the Chicago stocks variable in the basis equation and not the use to stocks specification.

All of the hedge levels obtained in this research are considerably larger than the hedges typically made by Central Illinois soybean producers. Thus, the results obtained here do not explain why farmers seemingly hedge at sub-optimal levels.

Foreward

Readers interested in more detail than this report provides are referred to the dissertation, "Year-Specific Estimation of Optimal Hedges: An Econometric Approach." It was completed by Joy L. Harwood at Cornell University in 1987. The research was directed by William G. Tomek. J.V. Meenakshi assisted with computing and Craig Jagger helped with graphics.

Any errors in this report are the sole responsibility of the authors.

Copies of this report may be obtained from:

Office of Publications
Department of Agricultural Economics
Warren Hall
Cornell University
Ithaca, New York 14853-7801

The soybean field shown in the cover photo is part of the Ward and Blanche Harwood farming operation near Crescent City, Illinois.

Table of Contents

List of Tables

List of Figures

Introduction

Past optimal hedge analyses have assumed that price risks are constant over time, and treat the optimal hedge ratio as fixed. Evidence from the price behavior literature, however, indicates that the distribution of supply and demand factors influences the means and variances of price distributions, and suggests that optimal hedges should be measured on a year-specific basis.

In this context, year-specific estimation of optimal hedges is examined for the Central Illinois soybean producer. A conceptual model is developed which permits the mean and variance parameters defining the optimal hedge to change over time. Econometric equations are specified and estimated for the Central Illinois area for the 1965-1984 period, and are used in developing the framework for the optimal hedge analysis. The results illustrate the potentially significant effects of year-specific estimation on the determination of the optimal hedge.

Optimal Hedge Defined

Optimal hedge estimation requires assumptions about measurement and about the specification of the objective function. The objective function used in this paper is based on harvesttime cash price and basis levels, and their respective variances and covariance. The producer is assumed to make anticipatory hedges of production decisions in May, in expectation of the October date when the hedge is lifted and the soybean crop is simultaneously sold. Yield risk is not considered. The mean-variance criterion for utility maximization is given by:[1]

$$\text{Maximize } L = E(U) = E(R) + m \operatorname{Var}(R) \qquad (1)$$

The E(U) term is expected utility; E(R) defines the expected return; Var(R) denotes the variance in returns; and m is the risk aversion parameter. Assumed known, m in the above formulation is negative for the risk averse hedger, increasing in absolute value as the risk aversion level increases.

Because the soybean producer in effect has a long position in the cash market, the futures position is assumed to be short. Therefore, returns (R) and expected returns [E(R)] are defined by:

$$R = QP_t + Q_h(F_{t-i} - P_t - B_t), \text{ and} \qquad (2)$$

$$E(R) = (Q - Q_h)P_t^* + Q_hF_{t-i} - Q_hB_t^* \qquad (3)$$

---

[1]The literature suggests that either a) mean-variance analysis is a good approximation to maximizing expected utility; or b) mean variance analysis is equivalent to maximizing expected utility (Meyer; Levy and Markowitz).

The cash return equals the bushel volume of the cash sale (Q) times the harvesttime cash price (P_t). The futures market return equals the number of bushels hedged (Q_h), multiplied by the difference between the springtime futures price (F_{t-i}) and the levels of the harvesttime cash price and basis ($P_t + B_t$).[2] The springtime futures price (F_{t-i}) is known at the time of hedge placement. The expectations of the harvesttime cash and basis prices, $E(P_t)=P_t^*$ and $E(B_t)=B_t^*$, are forecast the prior spring, at the time the hedge is placed. Because yield risk is ignored, Q is non-stochastic.

The variance in returns, Var(R), is given by:

(4) $$Var(R) = (Q-Q_h)^2 \sigma_p^2 + Q_h^2 \sigma_b^2 - 2Q_h(Q-Q_h) \sigma_{pb}$$

The σ_p^2 and σ_b^2 parameters denote the variances associated with the cash and basis price series, respectively, and σ_{pb} represents the cash price-basis covariance.

Substituting equations (3) and (4) into equation (1), the optimal hedge, Q_h^*, is then derived from the first order condition for a maximum, $dE(U)/dQ_h = 0$:[3]

(5) $$Q_h^* = Q[(\sigma_p^2+\sigma_{pb})/(\sigma_p^2+2\sigma_{pb}+\sigma_b^2)] - [(F_{t-i}-P_t^*-B_t^*)/2m(\sigma_p^2+2\sigma_{pb}+\sigma_b^2)]$$

The first term on the right hand side of equation (5) is the "hedging component," and represents the portfolio solution when m is assumed extremely large and/or the futures price change ($F_{t-i}-P_t^*-B_t^*$) is assumed near zero. The second term in equation (5) is the "speculative component." This component increases in absolute value as the expected futures price change over the hedging interval becomes larger, with the effect becoming more pronounced as the hedger grows less risk averse. As the anticipated futures price change increases, the optimal hedge (Q_h^*) is an increasingly larger fraction of expected production (Q) for producers with low levels of risk aversion.

---

[2]Because the sum of the harvesttime cash and basis prices (P_t+B_t) equals the harvesttime futures price (F_t), the ($F_{t-i}-P_t-B_t$) term can be written as: ($F_{t-i}-F_t$). The latter specification is more commonly used in the objective function than is the derivation explicitly including the basis. [See Ederington; Hayenga and diPietre; Heifner (1972); and Heifner (1973) for examples.]

[3]Peck introduced this specification of the objective function, which has also been employed by Greenhall in an examination of the optimal hedge for selected Western New York corn producers.

Parameter Measurement Issues

The strategies used to measure the variance, covariance, and price level parameters of the optimal hedge differ in their assumptions about price behavior. The oldest measurement strategy is consistent with the random walk theory, and assumes that the distribution of price changes has a zero mean and constant variance over time. Using this approach, the expected variances and covariance (σ_p^2, σ_b^2, and σ_{pb}) in equation (5) are calculated directly from historical price series. The variance of the cash price level, for instance, is given by:

(6) $$\sigma_p^2 = \Sigma(P_t - \overline{P})^2/n-1$$

The P_t term represents the harvesttime cash price in year t. The $\overline{P}$ parameter is the average of the P_t values, typically calculated over a span of 10 to 20 years (n = 10...20), where $\overline{P}$ is implicitly an estimate of P_t^*. Because price distributions are assumed constant over time, the resulting optimal hedge estimate is continuously applied to the anticipatory hedge decision.

A more recent approach to estimation models price forecasting ability. Using this approach, the accuracy of the hedger's historical forecasting ability defines the associated risk-return parameters. Price expectations are modeled using regression equations in which P_t^* and B_t^* are the dependent variables. The variance-covariance parameters (σ_p^2, σ_b^2, and σ_{pb}) are assumed equal to the equations' mean squared errors.[4] The cash price parameters, for instance, are given by:

(7) $$P_t^* = E(P_t) = \beta_1 + \beta_2 X_{t2} + ... + \beta_k X_{tk}$$

(8) $$\sigma_p^2 = \Sigma\, [P_t - E(P_t)]^2/n-k$$

This approach assumes that the hedger has a subjective distribution of harvesttime prices, where the mean of each distribution can change over time (as the P_t^* and B_t^* forecasts change). The variance, however, is treated as a constant, and depends on the hedger's forecasting ability. The variance-covariance parameters (σ_p^2, σ_b^2, and σ_{pb}) are, therefore, determined from regression series spanning n years.

The approach used in this study allows the variance and covariance parameters associated with the hedger's forecasts to change over time. It suggests that heteroskedasticity may exist, and permits the error variances of the cash price and basis equations to change as the price-generating process changes. Thus, the

---

[4]The mean squared error, $E[R-E(R)]^2$, equals the variance in returns if R is an unbiased estimator (if $E(R) = R^* = \mu_r$, the true mean). Consequently, MSE(R) can be interchanged with Var(R) in equation (4), producing the identical Q_h^* as in equation (5). See Peck; Fried for more details.

study assumes that a given price distribution exists at harvest in each year, characterized by a particular mean price and variance.

The fact that the variances of price changes are non-constant is well-documented, but little understood. The majority of price change studies address the variance of futures prices. Samuelson (1965, 1976), presenting the time-to-maturity hypothesis, argues that the rate of information flow into the market increases as maturity nears, and that the futures price variance per unit of time increases as the time to maturity decreases. Anderson and Danthine, proposing the state variable hypothesis, suggest that futures price variances are high (low) in periods when relatively large (small) levels of supply and demand uncertainty are resolved.[5]

Empirical research has been mixed in support of these hypotheses, but indicates that intra-year price variances are typically non-constant. Rutledge, examining the relationship between cash and futures price volatilities and time to maturity for contracts traded from 1969-1971, rejected Samuelson's hypothesis for wheat and soybean oil, but accepted it for silver and cocoa. Anderson examined nine commodities during the 1966-1980 period, supporting the state variable hypothesis, but finding only weak evidence for the time-to-maturity hypothesis. Castelino and Francis examined the volatility of basis levels and futures price spreads, and concluded that volatilities tend to decline with time-to-maturity. Additional evidence supporting the hypotheses of intra-year variance changes is provided by Miller; Taylor; and Castelino and Vora.

These theoretical and empirical results indicate that daily price changes can be more-or-less variable, depending on time-to-maturity and/or seasonal effects. Specific crop years often differ, in addition, in terms of the timing and intensity of information flows. Because annual variances are a function of the aggregation of daily price changes, these crop year differences may cause annual price change variances to differ. Further, the aggregation of price changes calculated over a specific time period in one year can then differ from the associated variance calculated in a different year.

Several authors have empirically documented annual differences in price change variances. Milonas and Vora, examining five agricultural commodities for the 1966-1982 period, indicate that price change volatilities differ both within and among years, and claim that these volatilities depend on underlying economic conditions. Booth, Kaen, and Koveos, in an examination of gold cash prices, similarly support the existence of combined intra- and inter-year effects on price variability.

---

[5]Two different information-flow scenarios are often hypothesized within the state variable theory. In situations where demand uncertainty is the relatively important factor, uncertainty is usually resolved just prior to the expiration of the contract, and the volatility thus increases as time-to-maturity approaches. In a case in which supply uncertainty is the relatively important factor, it may be resolved at earlier stages of the contract life cycle and thus lead to a decreasing volatility as delivery approaches.

The presence of leptokurtosis[6] in price distributions lends support to the theory that non-constant variances exist, although the debate continues. Mann and Heifner provide evidence contradicting the changing variance hypothesis, and indicate that leptokurtosis results from the existence of infinite variances. Telser, however, argues that such studies do not consider supply and demand distributions, and argues that leptokurtosis results from the aggregation of normal distributions which possess changing means and variances (which depend on supply and demand distributions). Gordon's empirical findings support Telser's argument.

Storage theory lends support to the hypothesis that price variances can change depending on underlying supply and demand conditions. Figure 1 indicates that the variance in the basis level can change depending on the level of the demand for storage. Because the supply of storage function is non-linear, the basis may be less variable as the demand for storage shifts into the range of larger inventories (I_t ranges from b to c) than it is at smaller inventory levels (I_t ranges from a to b).

These changing (price level) variance relationships depend on the limitations to relative cash and futures price movements. As inventories dwindle, the cash price can rise to an unlimited level above the futures price. In this case, the basis becomes very wide. At large inventory levels, on the other hand, a basis that is wider than the cost of storage provides a large incentive to store, increasing the cash price and reducing the basis. In this case, cash and futures are closely linked, and the basis cannot exceed the cost of storage.

Figure 1 is drawn to suggest that the error variance associated with the structural supply equation may change as a function of I_t. When inventories are low, random events may have a larger impact on the basis than when inventories are more abundant. Heteroskedasticity, as a result, may be a declining function of inventory levels (particularly beyond a critical level, such as "b" in Figure 1). Because the basis is determined from the relative levels of futures and cash prices, a changing basis variance also can imply changing variances in cash and futures price levels.

Econometric equations modelling Central Illinois cash price and basis reduced forms are specified in the following analysis. These equations are used to develop the harvesttime cash price and basis level forecasts (P_t^* and B_t^*), and to estimate the non-constant variance-covariance parameters (σ_p^2, σ_b^2, and σ_{pb} are estimated using the equations' mean squared errors).

Specification searches were used to develop the reduced-form equations for the 1965-1984 sample period. These searches examined: the "best" specification of variables thought to critically affect the cash price and basis levels; the impact of "doubtful" variables on explanatory power (Leamer); functional form issues; and the stability of the models across different sample periods. Because accurate forecasts reduce price risks, the identification of models which forecast well is emphasized.

---

[6] A leptokurtic distribution has a greater concentration of observations in the tails of the distribution than would be expected if the parent population were normal.

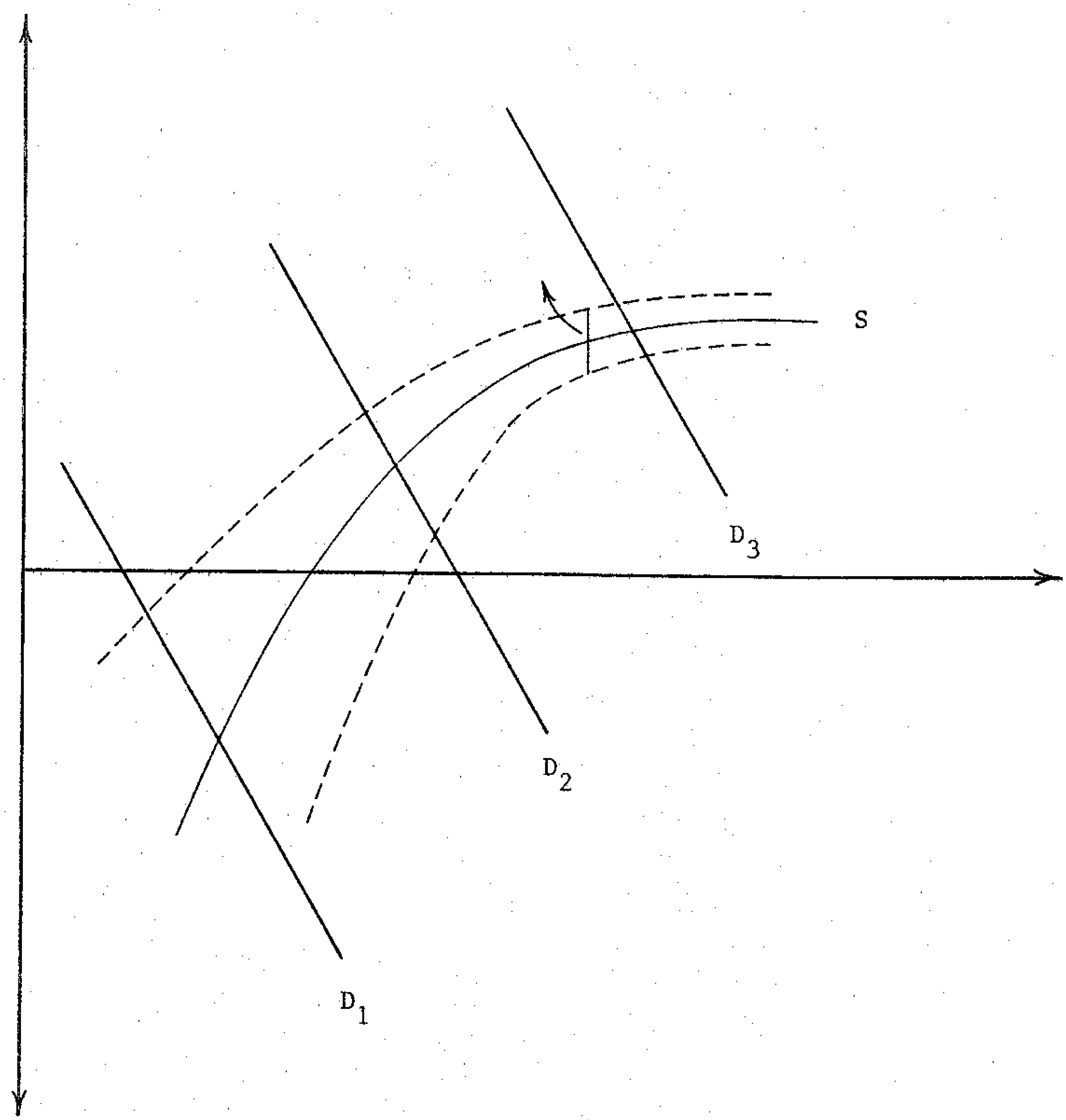

FIGURE 1
CHANGING VARIANCE RELATIONSHIPS AND DETERMINATION OF THE BASIS

Basis Equation

Storage theory is fundamental to understanding the basis for annually produced, storable commodities (Working). The basis is defined as a price for future delivery less a specified cash price, and measures the expected return from storage (the "price of storage") over a given time interval. At non-delivery points, the basis also should reflect characteristics unique to the local market because it measures the value of the commodity both over time and relative to the delivery point (Martin, Groenewegen, and Pidgeon). Empirical basis studies include those by Garcia and Good; Kahl; and Powers and Johnson.

Reduced Form

The equilibrium levels of the basis and the size of current inventory at a point in time are jointly determined by the supply of and demand for inventories, as shown in Figure 1. Hence, a reduced form basis equation can be derived from a two equation system in which the basis and inventory levels are both endogenous. The following analysis uses a two-period setting and assumes that the commodity is storable. (A similar conceptual model was applied to an analysis of soybean prices and inventories at the national level by Helmberger and Akinyosoye.)

In period 1, the inverse demand relationship is given by:

(9) $$P_1 = \alpha_1 + \beta_1 D_1$$

where P_1 is the period 1 cash price and D_1 is the level of demand.

Period 1 demand can be written as a function of production (S_1) and beginning and ending inventories (I_0 and I_1):

(10) $$D_1 = S_1 + I_0 - I_1$$

Production (S_1) is determined by prior plantings and beginning inventories (I_0) are known. Period 1 ending inventories (I_1), as well as P_1 and D_1, are endogenous.

Period 2 demand can be expressed similarly as:

(11) $$F_1 = \alpha_2 + \beta_2 D_2$$

(12) $$\text{and } D_2 = S_2 + I_1 - I_2$$

where F_1 is the period 1 price quote for future delivery in period 2 and D_2 represents expected demand. The S_2 variable denotes expected production in period

2. The I_2 variable, representing expected inventory at the end of period 2, is often given by the mean level: $I_2 = \overline{I}$.

Equations (10) and (12) can be substituted into equations (9) and (11), respectively, and the difference between the transformed equations gives a demand for inventory equation (where it is assumed that $I_2 = \overline{I}$):

(13) $$B_1 = F_1 - P_1 = (\alpha_2 - \alpha_1) + \beta_3 (2I_1 - S_1 - I_0 - \overline{I})$$

To complete the model, a supply of inventory equation is needed:

(14) $$B_1 = F_1 - P_1 = f(I_1)$$

$$\text{or } I_1 = f(B_1)$$

where B_1 is the period 1 basis, defined as the period 1 futures price (F_1) less the period 1 cash price (P_1).

To solve for the reduced form of the basis, equation (13) can be rewritten to express I_1 as a function of B_1. The resulting expression can be substituted in equation (14), and the reduced form of the basis is then solved as:

(15) $$B_1 = \pi_0 + \pi_1(S_1 + I_0) + \pi_2 X_1 + \pi_3 Y_1$$

The reduced form basis equation is a function of production (S_1), beginning inventories (I_0), and demand- and supply-shifters (given by X_1 and Y_1, respectively).

Specification

The Central Illinois area, defined by the Central, East, West Southwest, and East Southeast crop reporting districts (see Figure 2), comprises approximately ten percent of annual U.S. soybean production (Illinois Agricultural Statistics; USDA Agricultural Statistics). Central Illinois soybeans are harvested in late September and early October, and about thirty percent of the crop is sold off the farm at harvest to local elevators (Illinois Agricultural Statistics). The average Central Illinois basis for the first ten trading days in October (B_t) is given by:

(16) $$B_t = \Sigma\, (F_{it} - P_{it})/10$$

where: F_{it} = the November futures price quoted on each of the first ten trading days (i=1,2,...,10) in October in year t;

P_{it} = the Central Illinois cash price quoted on the identical days; t=1,2,...,20 years.

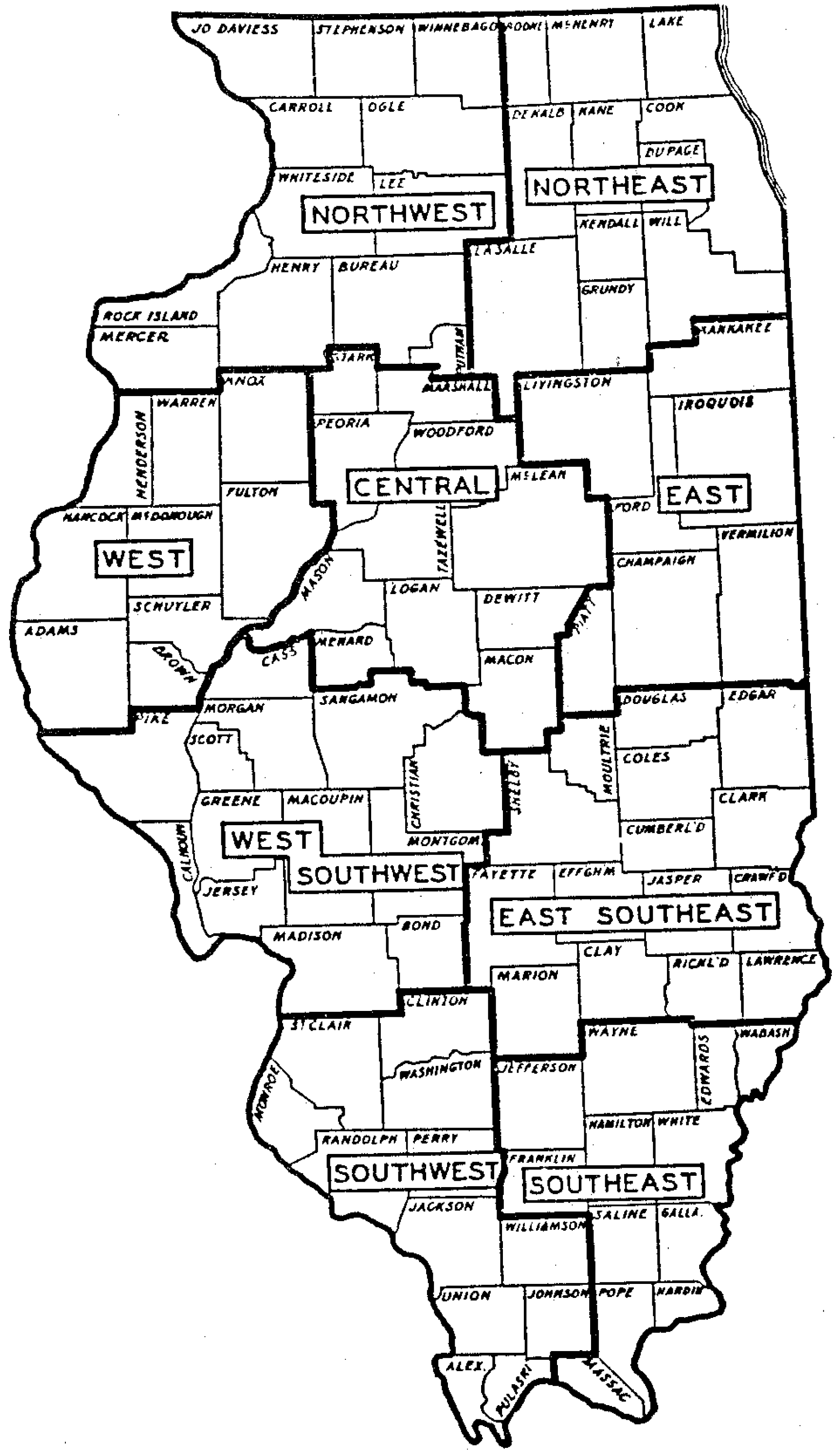

FIGURE 2
ILLINOIS CROP REPORTING DISTRICTS

The futures prices are daily closing prices for the Chicago Board of Trade's November soybean contract, and the cash price series represents elevator cash bid prices for the Central Illinois (Decatur) area (Good).[7] The harvest basis ranges from -8.72 to 31.15 cents per bushel over the 1965-1984 sample period. The mean basis is 14.06 cents per bushel; the standard deviation is 12.13 cents per bushel.

Production and carryover levels for the Central Illinois area and the U.S. should affect the local basis (through futures and cash prices), and are important demand-shifting factors [denoted by S_1 and I_0 in equation (15)]. As Central Illinois soybean inventories increase (relative to the national level), the cash price is expected to decrease relative to the futures price, and the basis is expected to widen. Increases in national (relative to local) inventories are expected to decrease the futures price relative to the local cash price, and narrow the basis. Supply is measured as the sum of production and carryover. The LOCPRO, LOCSUP, NATPROL, and NATSUPL variables, as well as all other variables primary to the analysis, are defined in Table 1.

As contract maturity approaches, the basis also may be influenced by delivery costs. The "short squeeze" is a type of market imperfection which occurs when deliverable Chicago stocks are in short supply relative to open interest, and represents the potentially large costs of making delivery (Paul, 1976). In this situation, short position-holders may prefer to offset their positions at a premium instead of obtaining deliverable stocks and making delivery, thus widening the basis. The regressors measuring the "squeeze potential" include Chicago stocks levels (CHICST), open interest (OPINT), and the ratio of Chicago stocks to open interest.

In practice, the supply of storage function can shift, as represented by the Y_1 variable in equation (15). The opportunity cost of grain storage (OPPCOST) captures the largest and most variable component of the rental price of binspace, and represents an important supply-shift factor. As the rental price of storage increases, fewer soybeans should be held in storage, and larger volumes of soybeans are expected to be for sale in the cash market. The cash price should drop relative to the futures price then, as the rental price increases, and the basis is expected to widen.

Shifts in consumption demand should also contribute to explanation of the basis [represented by the X_1 variable in equation (15)]. Variables measuring changes in the demand for Central Illinois soybeans (relative to the national level) were not available; thus, national demand data (representing expected crush and export levels) were used as proxies (FOSCRUSH and FOSEXP).

Other variables also may be important in explaining the harvesttime basis. The basis from past time periods (the previous springtime or the past year) may reflect distributed lag effects which exist in basis behavior. The ratio of on-farm to off-farm stocks may be important: as on-farm stocks increase relative to off-farm levels, farmers may sell only given higher prices and a narrower level of the basis. The concept of competition for binspace (Paul, 1970) also may be an

---

[7]The elevator cash bid prices are above the prices received by farmers; the difference between the elevator and farm prices should, however, be stable (Good).

TABLE 1
INDEPENDENT VARIABLE DEFINITIONS

Variable Definition	Data Source	Units	$\overline{X}$, σ_X
Inventory Variables (crop year):[a]			
LOCPRO--Local soybean production	Illinois Ag. Statistics	million bushels	166.41, 36.75
LOCSUP--Local soybean supply (production plus carryover)	Illinois Ag. Statistics	million bushels	188.60, 44.09
NATPROL--US (less local) soybean production	USDA and Illinois Ag. Statistics	million bushels	1290.84, 393.13
NATSUPL--US (less local) soybean supply	USDA and Illinois Ag. Statistics	million bushels	1468.06, 437.76
Market Imperfection Variables:			
CHICST--Ave. volume of soybeans in Chicago shipping district and afloat on Lake Michigan, first ten days in October	CBOT's Statistical Annual	million bushels	4.94, 4.77
OPINT--Ave. volume of soybeans held in long or short positions, first ten days in October	CBOT's Statistical Annual	million bushels	156.24, 70.66
Opportunity Cost Variable:			
OPPCOST--Product of ave. cash price and the 3-month Treasury Bill rate (adjusted to 1-month until maturity), first ten trading days in October	U. of Ill.; CEA's Economic Indicators	cents per bushel	3.22, 2.29
Consumption Demand Variables:[b]			
FOSCRUSH--ex ante crop year forecasts of expected US crush	USDA's Fats and Oils (Oil Crops) Situation	million bushels	828.74, 210.08
FOSEXP--ex ante crop year forecasts of expected US exports	USDA's Fats and Oils (Oil Crops) Situation	million bushels	553.84, 222.13

[a] Although production is reported by crop reporting district, carryover is reported at only the state level. Central Illinois carryover is thus calculated assuming that the Central Illinois fraction of state production equals the Central Illinois fraction of state carryover.

[b] Estimates from Crop Production and Grain Stocks in all Positions are used to calculate the inventory variables used in the step-ahed analysis for 1985 and 1986 (discussed later in the text). The Central Illinois fraction of state

important factor affecting the basis, and is represented by the percent of local binspace available for soybeans, and the level of Central Illinois corn production.

Estimation

Although the conceptual model suggests the types of variables that are important, a specification search is necessary to determine the "final" forecasting model. The inventory, market imperfection, and opportunity cost variables, hypothesized as the most important variables influencing the level of the basis, appeared in all equations. The effects of introducing variables thought to be of "doubtful" importance,[8] as well as non-linear functional forms, were examined through pretesting.

The criteria used to evaluate the models include traditional measures such as the logic of signs, Durbin-Watson values, and goodness-of-fit (R^2 and t-ratio values). The potential effects of outliers and the sensitivity of coefficients to alternative sample period lengths were also examined. The "final" pretest estimator may produce misleadingly good results, a point which is discussed later.

Results

Production and carryover, combined into one supply variable, proved to have the greatest explanatory power. Although production and carryover variables are expected to be important factors explaining the basis, the local and national levels are collinear ($r = 0.96$), and the independent information these variables individually contribute to the model is small. National and state supply regressors are, however, somewhat less collinear than the production and stocks variables expressed individually, and appear most important in explaining the basis. The opportunity cost variable also appears consistently important.

Although open interest and the ratio of stocks to open interest appear unimportant, the Chicago stocks variable has a large t-ratio. Low Chicago stocks levels in early October may indicate conditions which frequently continue until contract maturity (harvest delays or low rates of farmer marketing), and which affect the futures price within the delivery period. This hypothesis is supported by empirical evidence: stocks in early October vary considerably among years, but increase only moderately within the month in any given year.

None of the variables hypothesized to be of "doubtful" importance appear important in the exploratory analysis. The coefficients associated with the lagged

---

[8]"Doubtful" variables include the lagged basis, the ratio of on-farm to off-farm stocks, and variables representing grain competition for binspace. Because of the lack of data on Central Illinois grain movements, the consumption demand variables (national crush and export levels) are also considered "doubtful". For a discussion of the distinction between "free" and "doubtful" variables in model specification, see Leamer.

basis variable, the ratio of on-farm to off-farm stocks variable, and the grain competition for binspace variables all had small t-ratios. The consumption demand variables also appeared unimportant.

In this context, quadratic and logarithmic transformations of the supply, opportunity cost, and Chicago stocks regressors were examined. Quadratic specifications of the supply and Chicago stocks variables, shown in equations (17) to (20) in Table 2, improved the fit of the model. Neither logarithmic transformations, nor non-linear transformation of the opportunity cost variable, provided better results than the original linear forms of the variables.

The Chicago stocks coefficients in equations (17) and (18) are noteworthy. The coefficient in equation (17) conforms to the original market imperfection hypothesis. The coefficients in equation (18), on the other hand, indicate that increasing Chicago stocks levels up to the 8.1 million bushel mark are associated with larger basis levels. Beyond 8.1 million bushels, larger Chicago stocks are associated with declining basis levels. This result, of course, contradicts the market imperfection hypothesis.

Indeed, the quadratic specification of the Chicago stocks regressors may be capturing an effect other than the squeeze potential. Chicago stocks are highly correlated with Central Illinois and national stocks, and at low Chicago stocks levels, concerns about deliverable supplies may affect the basis. At large Chicago stocks levels, however, the squeeze potential no longer exists, and the regressors may be a proxy for demand factors that are reflected in inventory changes (such as increasing movements to export).

Partial regression leverage plots lend support to the inventory change concept. They indicate that predicted basis values are about twelve cents per bushel greater than the actual basis for the years 1966, 1969, 1970, 1972, 1973, and 1981. The only factor appearing particular to these years is the optimistic level of the USDA's harvesttime expected use estimates (Fats and Oils Situation). Thus, forecasts of particularly large national use levels may indicate grain movements from the Central Illinois area, and a narrower basis. (These years are also typically associated with fairly large Chicago stocks levels.)

When the national export and crush variables were reexamined as proxies for local shipments, however, they each appeared unimportant. Perhaps the large degree of collinearity between the export, crush, and national supply variables caused this result, or perhaps the observed variables inaccurately captured actual expectations within the industry. Using a dummy variable equal to one in the six years (and zero otherwise) to capture consumption demand effects results in equation (19). Incorporating a linear Chicago stocks variable produces equation (20).

TABLE 2
SELECTED BASIS EQUATION SPECIFICATIONS[a]
(1965-1984 SAMPLE PERIOD)[b]

---

(17) 37.04 - 1.74 LOCSUP + 0.0061 QLOCSUP + 0.17 NATSUPL - 0.000076 QNATSUPL -
(0.75)(-1.44) (1.92) (1.55) (-2.21)

1.00 CHICST + 2.55 OPPCOST
(-1.55) (1.10)

$R^2 = 0.54$ Durbin-Watson = 1.28

(18) 76.84 - 1.95 LOCSUP + 0.0065 QLOCSUP + 0.12 NATSUPL - 0.000059 QNATSUPL +
(2.02)(-2.20) (2.81) (1.47) (-2.31)

5.00 CHICST - 0.31 QCHICST + 3.66 OPPCOST
(2.80) (-3.49) (2.11)

$R^2 = 0.77$ Durbin-Watson = 1.78

(19) 59.53 - 1.87 LOCSUP + 0.0064 QLOCSUP + 0.15 NATSUPL - 0.000071 QNATSUPL
(3.42)(-4.63) (6.10) (3.95) (-6.05)

3.22 CHICST - 0.25 QCHICST + 4.77 OPPCOST - 12.17 DUM
(3.79) (-6.15) (5.93) (-6.88)

$R^2 = 0.96$ Durbin-Watson = 2.36

(20) 25.20 - 1.69 LOCSUP + 0.0061 QLOCSUP + 0.19 NATSUPL - 0.000086 QNATSUPL -
(0.76)(-2.08) (2.86) (2.58) (-3.72)

1.81 CHICST + 4.09 OPPCOST - 14.35 DUM
(-3.80) (2.54) (-4.10)

$R^2 = 0.81$ Durbin-Watson = 1.33

---

a t-ratios are in parentheses.

b LOCSUP=Central Illinois supply; QLOCSUP=$(LOCSUP)^2$; NATSUPL=national supply (less local); QNATSUPL= $(NATSUPL)^2$; CHICST=Chicago stocks; QCHICST=$(CHICST)^2$; OPPCOST=opportunity cost; DUM=1 if 1966, 1969, 1970, 1972, 1973, 1981; DUM=0 if otherwise.

Appraisal

For the quadratic equations, the t-ratio and R^2 values are fairly large, and the Durbin-Watson statistic indicates that autocorrelation is unimportant.[9] With perhaps the exception of the quadratic Chicago stocks coefficients, all coefficients conform to economic logic. When the dummy variable is included in the model, no significant outliers are apparent.

Although the models appear to fit the sample period well, specification error is still an issue. The correct interpretation of the Chicago stocks coefficients is particularly ambiguous, and may be related to the specification issue. Specifically, the superior performance of the quadratic specification may be caused by the omission of variables which measure the effects of consumption demand.

In addition, the coefficient values shown in Table 2 change considerably for a particular variable among the specifications. Moreover, coefficients vary as the sample period is altered. When the specification search is applied to different time periods, however, alternative models rank identically in relative performance to those given in Table 2. Thus, no clear evidence exists for altering the specifications.

Forecasting

Accurate forecasts reduce basis risk and can enhance hedging returns. In the following discussion, the predictive abilities of three naive models are compared with the performance of equations (17) through (20). The naive models employed in this analysis are often used to forecast the basis, and include the sample mean of the regression series, the previous harvest (lagged) basis, and a three-year moving average of past harvest bases. Tracking ability (Type I and II error), the root mean squared error (RMSE), and Theil's U_2 coefficient (U_2) are used to evaluate

---

[9]Positive autocorrelation is suggested in equations (17) and (20). GLS estimates of the models produce little change in coefficient values, and are not reported.

the sample period simulations.[10] Step-ahead forecasts for 1985 and 1986 are then examined.

Within-Sample Results

The within-sample results, given in Table 3, indicate that the regression forecasts are better than the naive forecasts. Among the regression models, the quadratic specifications that include the dummy variable provide forecasts which are better than those equations that exclude the dummy. Of the econometric models, equation (19) performs the best in all categories and is followed in performance by equation (20). The superiority of the econometric specifications may, however, result from over-fitting the model to the sample period. Thus, out-of-sample performance is important in appraising the results.

Out-of-Sample Results

The step-ahead forecast results are shown in Table 4 for the 1985 and 1986 years. The 1985 forecasts are based on the 1965-1984 regression coefficients and 1985 harvesttime data. The models were then re-estimated for the 1965-1985 sample period, and 1986 harvesttime data were used to develop the 1986 forecasts.

The forecast errors, on average, are similar for the regression and naive models. The three naive forecasts misjudge the actual basis by an average of 9.7 and 6.2 cents per bushel, respectively, in 1985 and 1986. The regression models produce an average error of 6.4 cents per bushel in 1985 and 10.0 cents per bushel in 1986.

---

[10]Turning point errors are often used to assess the ability of the model to predict directional changes (tracking ability). A turning point occurs when the sign of $(A_t - A_{t-1})$ does not equal the sign of $(A_{t-1} - A_{t-2})$. (A_t is the current period actual price, A_{t-1} is the previous period actual price, and A_{t-2} is the actual price two periods past.) A turning point is predicted when $(F_t - A_{t-1})$ has a different sign than $(A_{t-1} - A_{t-2})$; no turning point is predicted when the signs of $(F_t - A_{t-1})$ and $(A_{t-1} - A_{t-2})$ are equal. (F_t is the current period forecast.) Thus, a turning point error exists when either a turning point occurs but is not predicted or a predicted turning point is not realized.

The RMSE is calculated as: $\text{RMSE} = \sqrt{[(1/n)\ \Sigma(F_i - A_i)^2]}$, where F_i equals the current period forecast (F_t) less the previous period actual value (A_{t-1}); A_i equals the current period actual value (A_t) less the previous period actual (A_{t-1}); and n is the number of periods over which the forecast is evaluated.

Theil's U_2 coefficient is the ratio of the square root of the mean squared error to the square root of the average squared actual change in values: $U_2 = \sqrt{[(1/n)\Sigma(F_i - A_i)^2]} / \sqrt{[(1/n)\Sigma(A_i)^2]}$. The U_2 coefficient is equal to one when the forecast is measured as the previous period's actual value; if U_2 is less than one, then the model is better than the naive forecast.

TABLE 3
WITHIN-SAMPLE PREDICTIVE ABILITY OF
ALTERNATIVE BASIS MODELS
(1965-1984 SAMPLE PERIOD)

Model	Turning Point Errors		Root Mean Squared Error	Theil's U_2 Coefficient
	% Type I[a]	% Type II[b]		
Regression Model:				
(17)	0.11	0.11	8.140	0.879
(18)	0.11	0.11	5.557	0.600
(19)	0.06	0.00	2.432	0.263
(20)	0.06	0.06	5.307	0.573
Naive Model:				
(1) sample mean	0.22	0.17	12.039	1.301
(2) lagged basis	0.00	100.00	9.257	1.000
(3) 3-year average	0.39	0.06	11.145	1.204

a A turning point is forecast, but no turning point occurs.
b No turning point is forecast, but a turning point occurs.

TABLE 4
ONE-STEP AHEAD PREDICTIVE ABILITY OF
ALTERNATIVE BASIS MODELS

Model	Forecast	Actual Basis	Error
	cents per bushel	cents per bushel	cents per bushel
		1985 Forecasts	
Regression Model:[a]			
(17)	0.52	2.43	-1.91
(18)	14.14	2.43	11.71
(19)	7.84	2.43	5.41
(20)	-3.97	2.43	-6.40
Naive Model:			
(1) sample mean	14.06	2.43	11.63
(2) lagged basis	-8.72	2.43	-11.15
(3) 3-year average	-3.90	2.43	-6.33
		1986 Forecasts	
Regression Model:[b]			
(17)	0.99	2.03	-1.04
(18)	7.01	2.03	4.98
(19)	-11.31	2.03	-13.34
(20)	-18.61	2.03	-20.64
Naive Model:			
(1) sample mean	13.51	2.03	11.48
(2) lagged basis	2.43	2.03	0.40
(3) 3-year average	-4.76	2.03	-6.79

a See Table 2.
b Estimates for 1965-1985 sample not shown.

Of the reduced form models considered, equation (17) performs the best in the out-of-sample analysis, with forecast errors of -1.91 and -1.04 cents per bushel in 1985 and 1986, respectively. This performance is considerably better than the errors associated with the 1985 naive forecasts, but because of the small change in the actual basis from 1985 to 1986, the lagged basis was the best individual forecast for 1986.

The superior out-of-sample performance of equation (17) appears to be a function of local and national supply levels, which are near the maximum values of the sample range in 1985 and outside the sample in 1986. In this case, the simpler specification (excluding the dummy variable and using the linear Chicago stocks variable), provides the best forecasts. In contrast, equations (19) and (20), which provided better forecasts than equation (17) in the in-sample analysis, produce 1986 forecasts which significantly underestimate the basis, reflecting the impact of the quadratic supply variables when the dummy variable is included in the model.

The results illustrate the potential benefits and the limitations from using econometric models in making marketing decisions. Equations (19) and (20) appear best from the standpoint of conventional test statistics (t-ratios, R^2 values, and Durbin-Watson values within the sample period), and perhaps provide superior out-of-sample forecasts when the data lie within the bounds of the sample. The simpler specification, equation (17), seemingly provides the best out-of-sample forecasts when the data lie outside the sample range. The best-fitting equations for the sample period, obtained by pretesting, clearly need not be the best out-of-sample predictors.

Equations (19) and (20), providing the best results over the sample range, are used in the optimal hedge analysis. From the 1986 perspective, observations on the production, stocks, and opportunity cost variables appear to be declining from the high levels of the mid-1980's. Thus, models performing well when these variables are at neither extremely large nor small levels, such as equation (19) and (20), appear best for short-run forecasting analysis.

Cash Price Equation

Several empirical models have examined the price determination process within the soybean complex. The Houck-Mann model uses a 15-equation dynamic block recursive system which emphasizes the multiple market and joint product relationships characterizing the demand-side of the market. Mathews, Womack, and Hoffman specified a model which recursively links six regional acreage equations to the Houck-Mann demand block. Meilke and Griffith developed a simultaneous system of the world soybean/rapeseed market. Various other models have been specified and used to examine world linkages within the oil-meal complex (Vandenborre; and Knipscheer and Hill).

Reduced Form

The number of exogenous factors influencing the harvesttime cash price is extremely large. In the Houck-Mann model, for instance, over twenty exogenous variables are treated as important in the price determination process. For small samples (such as the Central Illinois data set), however, the number of independent variables must be narrowed, and should probably not exceed 40% of the number of observations (Belsley, Kuh, and Welsch).

The price of the November futures contract in May combines the many exogenous price influencing factors into a single variable that the hedger can use as a forecast of the harvesttime cash price (Tomek and Gray; Kofi). One alternative for forecasting the harvesttime cash price level is then:

(21) $$P_1 = f(F_0)$$

where P_1 is the forecast of the harvesttime cash price and F_0 is a specified level of the May futures price for November delivery.

With the passage of time, however, expectations change and the springtime futures price and the maturity cash price typically differ. In addition, the producer's expectations when the hedge is placed may differ significantly from market expectations.

For these reasons, harvesttime estimates of inventory and use levels are included in the equation to allow modification of the futures price estimate. The reduced form for the cash price level is given by:

(22) $$P_1 = \pi_0 + \pi_1 F_0 + \pi_2 I_1 + \pi_3 Y_1$$

As expressed in equation (22), Central Illinois inventories (I_1), rather than the national inventory level, are expected to influence the local cash price. Because local consumption data were not available, national use estimates were used as proxies (Y_1).

Specification

The harvesttime cash price (P_t) is measured as the average Central Illinois cash price (P_{it}) during the first ten (i) trading days in October for t = 1...20 years:

(23) $$P_t = \Sigma P_{it}/10$$

Cash prices are representative for the Decatur area and are identical to the cash prices used in calculating the harvesttime basis. The harvesttime cash price ranges from 228.6 to 878.50 cents per bushel over the 1965-1984 sample period. The mean cash price is 503.00 cents per bushel, with a standard deviation of 217.40 cents per bushel.

The May futures quote for the November contract (MAYNF), measured in cents per bushel, is the average price quoted during the first ten trading days in May. Central Illinois inventories are defined by the production (LOCPRO) and carryover variables discussed in the basis estimation section, with the sum of production and carryover reflecting a measure of supply. Expected use is measured by the identical crush and export variables used in the basis analysis (FOSCRUSH and FOSEXP). These crush and export variables are based on U.S.D.A. crop year forecasts published in the Fats and Oils (Oil Crops) Situation during the August/September period, with the total expected use variable representing the sum of crush and export levels (TOTUSE). All inventory and use variables are measured in million bushels. (See Table 1 for additional information.)

Estimation

While the reduced form suggests the types of variables that are important, a specification search is used to determine the "final" forecasting model. The futures price, inventory, and expected use variables were included in all equations. The effects of alternative specification of these variables (such as production instead of supply), as well as non-linear functional forms, were examined through pretesting.

The criteria used to evaluate the models were identical to the criteria discussed in the harvesttime basis analysis, and include traditional measures such as the logic of signs, Durbin-Watson values, and goodness of fit (R^2 and t-ratio values). The potential effects of outliers and the sensitivity of coefficients to alternative sample period lengths were also examined. As mentioned previously, the "final" pretest estimator may produce misleadingly good results, a point which is discussed later.

Results

The alternative supply specifications indicate that the production regressors explain the largest proportion of the harvesttime cash price for each given use specification. The production variables consistently produce the largest t-ratio and R^2 values in comparison with the supply (production and carryover) specifications. When the supply variables are disaggregated into production and carryover components, the carryover variable is statistically unimportant.

These results are likely related to the information captured by the May futures price regressor. Because uncertainties about yields can cause dramatic changes in production estimates between May and October, the futures price quote in May can differ significantly from the October cash price. In contrast, knowledge of October stock levels are typically reflected in the May futures quote because

May and October stock levels are closely correlated. As a result, the production regressors contribute more additional information to the model than do the carryover or supply regressors.

The preliminary regressions, each including the production variable, are shown in equations (24) through (27) in Table 5. The export and crush variables each appear unimportant in equation (24), perhaps because of the large intercorrelations between these regressors (r = 0.95). Deleting the export and crush regressors in turn produces equations (25) and (26); combining the export and crush variables into a total expected use regressor results in equation (27).

Quadratic and logarithmic transformations of the production and consumption demand variables were then examined. Neither logarithmic transformations, nor quadratic transformations of the production variables, improved the fit of the models. Quadratic specifications of the consumption demand variables, shown in equations (28) through (30), did improve the fit of the models: the t-ratios for nearly all variables are larger than when the consumption demand variables are expressed in linear form, and the Durbin-Watson statistic has increased from near 0.90 to the 1.20 to 1.50 range.

The negative quadratic use coefficients in equations (28) to (30) are noteworthy, indicating that the rate of increase in the cash price declines as consumption demand increases. At the largest demand values within the sample range, the cash price level declines in absolute value as expected use increases. The quadratic total use variable, for instance, indicates that the price flexibility with respect to total use is 2.43 at a use level of 900 million bushels, and is 0.60 at a use level of 1,700 million bushels. This result is somewhat counterintuitive because large consumption levels are more often associated with greater price variability than are low demand levels.

Appraisal

For the quadratic equations, the t-ratio and R^2 values are fairly large, and the Durbin-Watson statistics indicate that autocorrelation is less important than when the linear forms are specified. All variable signs and magnitudes appear logical, perhaps with the exception of the quadratic consumption demand variables.

Although the models appear to fit the sample period well, the coefficients (particularly those associated with the May futures price and production regressors) appear quite sensitive to sample period length. This result is perhaps related to the presence of multicollinearity relative to the number of observations. If additional observations are less collinear than the existing data, as is true for the mid-1980's observations, the effect of new information can be large. When specification searches are applied to different time periods, however, alternative models rank identically in performance to those given in Table 5.

TABLE 5
SELECTED CASH PRICE EQUATION SPECIFICATIONS[a]
(1965-1984 SAMPLE PERIOD)b

(24) 167.86 - 3.68 LOCPRO + 0.62 FOSCRUSH - 0.081 FOSEXP + 0.98 MAYNF
(0.65) (-3.06) (0.98) (-0.13) (3.35)

$R^2 = 0.81$ Durbin-Watson = 0.90

(25) 195.95 - 3.70 LOCPRO + 0.54 FOSCRUSH + 0.96 MAYNF
(1.43) (-3.21) (2.37) (3.68)

$R^2 = 0.81$ Durbin-Watson = 0.90

(26) 389.79 - 3.73 LOCPRO + 0.49 FOSEXP + 0.94 MAYNF
(3.10) (-3.11) (2.08) (3.25)

$R^2 = 0.80$ Durbin-Watson = 1.01

(27) 295.90 - 3.75 LOCPRO + 0.27 TOTUSE + 0.93 MAYNF
(2.39) (-3.20) (2.28) (3.39)

$R^2 = 0.81$ Durbin-Watson = 0.93

(28) -445.59 - 4.18 LOCPRO + 2.26 FOSCRUSH - 0.00091 QFOSCRUSH + 0.88 MAYNF
(-0.96) (-3.59) (1.87) (-1.44) (3.40)

$R^2 = 0.84$ Durbin-Watson = 1.17

(29) 130.16 - 4.20 LOCPRO + 1.87 FOSEXP - 0.0011 QFOSEXP + 0.84 MAYNF
(0.76) (-3.74) (2.61) (-2.02) (3.14)

$R^2 = 0.84$ Durbin-Watson = 1.54

(30) -164.66 - 4.25 LOCPRO + 1.08 TOTUSE - 0.00026 QTOTUSE + 0.84 MAYNF
(-0.57) (-3.73) (2.25) (-1.74) (3.19)

$R^2 = 0.84$ Durbin-Watson = 1.33

---

a t-ratios are in parentheses.

b LOCPRO=Central Illinois production; FOSCRUSH=forecast crop year crush; QFOSCRUSH=$(FOSCRUSH)^2$; FOSEXP=forecast exports; QFOSEXP= $(FOSEXP)^2$; MAYNF=May quote for November futures; TOTUSE=FOSCRUSH + FOSEXP;

Forecasting

This section examines the forecasting abilities of equations (24) through (30), and compares the results with those obtained from three naive forecasting methods. The naive models used in the analysis are analogous to those used in the basis forecasting comparison, and include: 1) the sample mean of the cash price series as predictor; 2) the previous year's harvesttime cash price as predictor; and 3) a three-year moving average of past harvesttime cash prices as predictor.

Evaluation methods are also identical to those used in the basis forecasting analysis. The within sample criteria, explained earlier, include: the tracking abilities of the models (as measured by turning point errors); the root mean squared error (RMSE); and Theil's U_2 coefficient (U_2). The out-of-sample criteria are based on an analysis of step-ahead forecasts for the 1985 and 1986 years.

Within-Sample Results

The within-sample comparisons of predictive ability are presented in Table 6, and indicate that the regression forecasts are better than the naive model forecasts. Within the regression category, the quadratic models are superior to those of a linear nature: each quadratic model has a lower percentage of forecast errors (using all criteria) than do any of the linear forms. Of the regression models, equation (29) performs the best in each evaluation category, followed by equation (30).

The superiority of the quadratic functional forms may result from over-fitting the model to the sample period. As in the basis analysis, the search procedure may have resulted in equations which fit the sample period, but that do not accurately capture the systematic factors that affect the price determination process. The out-of-sample analysis, discussed in the following section, is a more objective approach than is the in-sample procedure.

Out-of-Sample Results

Step-ahead forecasts are shown in Table 7 for the years 1985 and 1986. The 1985 forecasts are based on the 1965-1984 regression coefficients and 1985 harvesttime data. The models were then re-estimated for the 1965-1985 sample period, and 1986 harvesttime data were used to develop the 1986 forecasts.

The results indicate that the out-of-sample performance for the naive models is poorer, on average, than for the regression models. The average regression forecasts for 1985 and 1986 respectively misjudge the harvest cash price by 33.5 and 11.5 cents per bushel. For the naive models, the average forecast error is

TABLE 6
WITHIN-SAMPLE PREDICTIVE ABILITY OF
ALTERNATIVE CASH PRICE MODELS
(1965-1984 SAMPLE PERIOD)

Model	Turning Point Errors		Root Mean Squared Error	Theil's U_2 Coefficient
	% Type I[a]	% Type II[b]		
Regression Model:				
(24)	0.17	0.06	93.515	0.551
(25)	0.17	0.06	93.538	0.551
(26)	0.11	0.06	96.034	0.566
(27)	0.11	0.06	94.323	0.556
(28)	0.06	0.06	88.032	0.519
(29)	0.06	0.06	85.910	0.506
(30)	0.06	0.06	86.542	0.510
Naive Model:				
(1) sample mean	0.28	0.22	208.751	1.230
(2) lagged price	0.00	100.00	169.734	1.000
(3) 3-year average	0.50	0.06	160.694	0.947

[a] A turning point is forecast, but no turning point occurs.
[b] No turning point is forecast, but a turning point occurs.

TABLE 7

OUT-OF-SAMPLE PREDICTIVE ABILITY OF ALTERNATIVE CASH PRICE MODELS (1965-1984 SAMPLE PERIOD)

Model	Forecast	Actual Price	Error
	cents per bushel	cents per bushel	cents per bushel
		1985 Forecasts	
Regression Model:[a]			
(24)	557.78	503.85	53.93
(25)	548.96	503.85	45.11
(26)	490.01	503.85	-13.84
(27)	518.14	503.85	14.29
(28)	558.61	503.85	54.76
(29)	521.55	503.85	17.70
(30)	538.40	503.85	34.55
Naive Model:			
(1) sample mean	503.00	503.85	-0.85
(2) lagged price	607.80	503.85	103.95
(3) 3-year average	665.02	503.85	161.17
		1986 Forecasts	
Regression Model:[b]			
(24)	486.71	474.00	12.71
(25)	487.20	474.00	13.20
(26)	471.13	474.00	-2.87
(27)	481.87	474.00	7.87
(28)	478.69	474.00	4.69
(29)	492.01	474.00	18.01
(30)	495.23	474.00	21.23
Naive Model:			
(1) sample mean	503.04	474.00	29.04
(2) lagged price	503.85	474.00	29.85
(3) 3-year average	657.27	474.00	183.27

[a] See Table 5.
[b] Estimates for 1965-1985 sample are not shown.

considerably larger: in 1985, the error is 88.7 cents per bushel, and in 1986, the average error is 80.7 cents per bushel.[11]

In an analysis of individual performance, equations (26) and (27) provides the most accurate out-of-sample forecasts, followed by equation (29). Equation (29) is chosen, however, for use in the optimal hedge analysis because it provides the best in-sample forecasts and among the best step-ahead forecasts in the out-of-sample analysis. This specification expresses the importance of export movements, which typically peak at harvest. In addition, forecasting the level of the export regressors may be simpler in the ancillary analysis than forecasting total use (composed of exports and crush), which is expressed in equation (27).

In the following sections, the cash price equation [equation (29)] and the basis equations [equations (19) and (20)] are used to develop constant and non-constant variance estimates of optimal hedges. These equations are first tested for the presence of heteroskedasticity (non-constant variances). Knowledge of the error structure then is used to estimate the Central Illinois anticipatory hedge for soybeans.

Heteroskedasticity Tests

The Koenker test for heteroskedasticity is chosen for use in the analysis. This test can be used when the error variance is related to either a single variable or a linear combination of variables (unlike the Goldfeld-Quandt and Szroeter's class of tests, which require that the heteroskedasticity is related to only one variable). In addition, the Koenker test is less sensitive to error deviations from normality than are either the Breusch-Pagan or White tests (particularly when the sample size is small).[12]

The Koenker test statistic is developed from the regression of the (normalized) OLS errors on the variables thought to affect the error variance. This regression, estimated by OLS, is given by:

(31) $$(\hat{u}_t^2 - \sigma^2) = f\,(z_t'a)$$

where:
- $\hat{u}_t^2$ = the (TX1) vector of squared OLS residuals;
- σ^2 = the average of the $\hat{u}_t^2$;
- z_t = a (TXK) matrix of variables thought to influence the error variance, with first element equal to unity;
- a = a (KX1) vector of unknown coefficients.

---

[11]Although the regression models on average out-perform the naive models, the sample mean and lagged price forecasts are surprisingly close to the actual 1985 and 1986 prices. This result is specific to the forecast years analyzed; it is a peculiarity that rarely occurs in other years.

[12]For discussions concerning the assumptions and procedures associated with these heteroskedasticity tests, see Judge, et al. and Kmenta.

The test statistic, N^*, is given by:

(32) $$N^* = TR^2$$

where: T = the number of observations in the equation given by (31);
R^2 = the coefficient of determination obtained from the regression in (31).

Only one or two variables (in addition to the intercept) appeared in each error variance regression. Variables were specified in the z_t matrix using the functional forms which appeared in basis equations (19) and (20) and in cash price equation (29). That is, variables that appeared as quadratics in the original regressions were specified as quadratics in equation (31). One variable was used if the linear form of the variable appeared in the original regression.

Because the equations are quite sensitive to sample period length, the Koenker test also may be quite sensitive to the number of observations. Five series of errors were thus used in equation (31); these series were obtained from the regressions given in equations (19), (20), and (29), estimated from the 1965-1981 period through the 1965-1985 period.

The test results, shown in Table 8, indicate that heteroskedasticity is present only in equation (20), and that it is a function of the level of Chicago stocks. This result appears only in the 1965-1983 through 1965-1985 sample periods. When the Chicago stocks variable is ordered sequentially from smallest to largest values in these sample periods and the corresponding error terms are examined, the error variance appears to increase as the Chicago stocks level increases.

This result contradicts the hypothesized relationship that the error variance increases as stocks decline, and may be caused by misspecification of the functional form. The heteroskedasticity associated with the linear Chicago stocks specification may indicate that equation (20) does not capture variable(s) that affect the basis (and that possess changing variance relationships reflected in the error). The quadratic Chicago stocks specification, equation (19), although fitting the data better and resulting in the absence of heteroskedasticity, produces an unexpected coefficient sign (unlike the linear form, which has expected signs).

The trade-offs existing between equation (19)--with the quadratic Chicago stocks variables, and equation (20)--with the linear specification, make it difficult to determine which, if either, is the appropriate specification. Certainly, it is possible that the quadratic Chicago stocks variable is "picking up" unmeasurable effects related to consumption demand or other omitted variables that are not explicitly included in the linear specification. Thus, neither equation may correctly measure the factors that affect basis determination. These factors should be kept in mind throughout the remainder of the analysis.

The heteroskedasticity results indicate that two systems, with different error variance-covariance structures, can be composed from equations (19), (20), and (29). One system [containing equations (19) and (29)] can be used to obtain constant variance estimates of the optimal hedge parameters. The other system [composed of

TABLE 8

KOENKER TEST VALUES FOR HETEROSKEDASTICITY

	LOCSUP*	NATSUPL*	CHICST*	OPPCOST**
Equation 19:				
1965-1981	0.97	1.61	2.85	1.82
1965-1982	1.44	1.26	2.16	1.98
1965-1983	0.95	1.33	3.04	4.18
1965-1984	1.80	1.80	0.16	0.80
1965-1985	1.85	2.14	0.09	1.23

	LOCSUP*	NATSUPL*	CHICST**	OPPCOST**
Equation 20:				
1965-1981	1.51	1.72	0.03	1.79
1965-1982	0.36	0.36	0.002	1.26
1965-1983	0.76	1.14	8.74	3.04
1965-1984	0.40	0.80	5.60	2.20
1965-1985	0.40	0.85	6.52	2.23

	LOCPRO**	FOSEXP*	MAYNF**
Equation 29:			
1965-1981	1.50	3.20	2.15
1965-1982	1.57	3.42	2.08
1965-1983	1.52	3.61	1.35
1965-1984	0.38	3.15	2.78
1965-1985	0.17	2.50	2.47

Note: For the quadratic specifications, the degrees of freedom equal (K-1) = 3-1 = 2 (where K includes the intercept). For the linear specifications, the degrees of freedom equal (K-1) = 2-1 = 1.

The critical Chi-squared values are therefore:

* At $\alpha = 0.05$, $X_2^2 = 5.99$; at $\alpha = 0.10$, $X_2^2 = 4.61$.
** At $\alpha = 0.05$, $X_1^2 = 3.84$; at $\alpha = 0.10$, $X_1^2 = 2.76$.

equations (20) and (29)] can be transformed to obtain non-constant, year-specific estimates.

The cash price and basis equations in each system contain similar (although not identical) regressors, and are interrelated through the cash price component. Because it is also likely that the equations share common information in the error terms (the across-equation covariances are non-zero), a seemingly unrelated regression (SUR) framework is used to obtain final estimates of the coefficients.

<u>Constant Variance System</u>

The basis and cash price equations used in the SUR estimation of the constant variance system can be defined in stacked notation by:

(33)
$$\underset{(2TX1)}{\begin{bmatrix} B \\ P \end{bmatrix}} = \underset{(2TX[K_1+K_2])}{\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}} \underset{([K_1+K_2]X1)}{\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}} + \underset{(2TX1)}{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}$$

Each equation contains T observations on K_1 (or K_2) regressors (including the intercept). The error variance structure associated with this system, E[ee'], is given by:

(34)
$$\Omega = \left[\begin{array}{cccc|cccc} \sigma_b^2 & & 0 & & \sigma_{bp} & & 0 & \\ & \sigma_b^2 & & & & \sigma_{bp} & & \\ 0 & & \sigma_b^2 & & 0 & & \sigma_{bp} & \\ & & & \cdot & & & & \cdot \\ \hline \sigma_{pb} & & 0 & & \sigma_p^2 & & 0 & \\ & \sigma_{pb} & & & & \sigma_p^2 & & \\ 0 & & \sigma_{pb} & & 0 & & \sigma_p^2 & \\ & & & \cdot & & & & \cdot \end{array}\right]$$

(2TX2T)

The elements in the second and fourth quadrants reflect the variances of the basis (σ_b^2) and the cash price (σ_p^2) equations, respectively, and the elements in the first and third quadrants represent the across-equation error covariance (where $\sigma_{pb}=\sigma_{bp}$). The error variance (or covariance) within a partition is assumed constant over time.

The OLS approach is inefficient when across-equation correlations are non-zero ($\sigma_{pb}, \sigma_{bp} \neq 0$): OLS, a single equation approach, does not consider across-equation information. Zellner's seemingly unrelated regression (SUR) framework, using generalized least squares (GLS) estimation, considers the entire structure of the model (Zellner). The GLS estimator, using the inverse of the Ω matrix presented above, is given by:

(35) $$b = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y$$

When the true variances and covariances among equations are known, GLS estimation is more efficient than OLS. The gain in efficiency, however, depends on: a) the degree of correlation among the errors; and b) the correlation among the regressors in the different equations. As the error intercorrelations become larger and the correlations among regressors become smaller, the efficiency of GLS estimation (relative to OLS) increases (Judge, et al.; Kmenta).

In practice, an estimated GLS (EGLS) approach is required because the true variances and covariances used in the Ω^{-1} matrix are unknown. The most common EGLS approach employs Zellner's two-step procedure; it uses the OLS errors to estimate the variance-covariance elements of the Ω^{-1} matrix and then $\hat{b}$ is based on $\hat{\Omega}^{-1}$. The iterative procedure, an alternative to the two-step method, is used in this analysis; it re-estimates the EGLS errors until the error estimates converge (Judge, et al.; Kmenta).

Both the two-step and iterative approaches produce asymptotically efficient estimates of the Ω^{-1} and B matrices. In addition, Monte Carlo studies suggest that these EGLS approaches are more efficient than OLS even in small samples (Kmenta).

The SYSNLIN systems subprogram of the Statistical Analysis System (SAS) package required ten iterations for the SUR-GLS estimation of the constant variance system. The SUR regression estimates, shown in Table 9, are similar to the least squares estimates, reflecting unexpectedly low across-equation error correlations and the large across-equation correlations among regressors. The error variance-covariance values for the system are: $\sigma_p^2 = 9{,}432.71$; $\sigma_{pb} = 162.80$; and $\sigma_b^2 = 12.25$.

Optimal hedge estimates based on the constant variance system are shown in Table 10. Risk aversion levels of -infinity, -0.1, and -0.01 are employed, and it is assumed that Q=1 (that the results can be interpreted as a percent of production).[13] Recall that the optimal hedge formula is given by:

(38) $$Q_h^* = Q[(\sigma_p^2+\sigma_{pb})/(\sigma_p^2+2\sigma_{pb}+\sigma_b^2)] - [(F_{t-i}-P_t^*-B_t^*)/2m(\sigma_p^2+2\sigma_{pb}+\sigma_b^2)]$$

---

[13]The risk aversion levels given by m = -0.1 and m = -0.01 are commonly used in optimal hedge analyses (Peck). The assumption that Q = 1 is also common (Peck).

TABLE 9

SUR ESTIMATES OF THE CONSTANT VARIANCE MODEL[a] (1965-1984 SAMPLE PERIOD)[b]

Basis Equation

(36)

$$\underset{(3.48)}{58.14} - \underset{(-4.96)}{1.89}\ LOCSUP + \underset{(6.47)}{0.0064}\ QLOCSUP + \underset{(4.33)}{0.15}\ NATSUPL -$$

$$\underset{(-6.46)}{0.000072}\ QNATSUPL + \underset{(3.55)}{2.84}\ CHICST - \underset{(-6.06)}{0.24}\ QCHICST + \underset{(5.73)}{4.37}\ OPPCOST -$$

$$\underset{(-6.63)}{11.14}\ DUM$$

Cash Price Equation

(37)

$$\underset{(0.93)}{157.09} - \underset{(-3.93)}{4.30}\ LOCPRO + \underset{(2.54)}{1.74}\ FOSEXP - \underset{(-1.98)}{0.0010}\ QFOSEXP + \underset{(3.66)}{0.92}\ MAYNF$$

---

a t-ratios are in parentheses.
b Variables are defined in Tables 2 and 5.

TABLE 10

CONSTANT VARIANCE ESTIMATES OF THE OPTIMAL HEDGE

Year	Price Change $(F_{t-i}-P_t^*-B_t^*)$	Hedging Component	Total Hedge (Hedging and Speculative Components)	
	cents per bushel	% of production	% of production m=-0.1	m=-0.01
1965	12.6	98	99	105
1966	-47.0	98	96	74
1967	-51.8	98	96	72
1968	3.5	98	98	100
1969	52.3	98	101	125
1970	-148.8	98	91	22
1971	-39.0	98	96	78
1972	-59.6	98	95	68
1973	-57.2	98	95	69
1974	-245.7	98	86	-27
1975	57.7	98	101	128
1976	-114.3	98	92	40
1977	75.2	98	102	137
1978	-23.9	98	97	86
1979	117.8	98	104	158
1980	-61.3	98	95	67
1981	54.2	98	101	126
1982	109.7	98	104	154
1983	-136.1	98	91	29
1984	-76.7	98	94	59

For the extremely risk averse hedger (m=-infinity), the speculative component of the optimal hedge is zero. The hedge ratio therefore depends solely on the hedging component, which indicates that the short futures position should equal 98 percent of expected production:

$$Q_h^*/Q = (\sigma_p^2 + \sigma_{pb})/(\sigma_p^2 + 2\sigma_{pb} + \sigma_b^2) = 0.98 \quad (39)$$

For the less risk averse producer, (m=-0.1 or m=-.01), Table 10 indicates that the speculative component of equation (38) can have a large impact on the optimal hedge. For the moderately risk averse hedger, (m=-0.1), the hedge ratio varies from 86 to 104 percent of expected production. With forecasts of an increase in the futures price, ($F_{t-i}-P_t^*-B_t^* < 0$), the hedge ratio drops below the 98 percent level defined by the hedging component of the equation. If a decline in the futures price is forecast, ($F_{t-i}-P_t^*-B_t^* > 0$), the hedge increases to over 98 percent of expected production.

For the less risk averse producer, (m=-0.01), the speculative component dominates the optimal hedge calculation, and the hedge ratio varies from a 158 percent short hedge of expected production (when a price decline of 118 cents per bushel is anticipated), to a 27 percent long hedge (when a price rise of 246 cents per bushel is expected).

In summary, the optimal hedge for the extremely risk averse producer (m= -infinity) is estimated at 98 percent of expected production. This result occurs because the cash price variance (σ_p^2) dominates the hedging component calculation relative to the σ_b^2 and σ_{pb} parameters. For the less risk averse hedger (m=-0.1 or m=-0.01), however, the optimal hedge can differ significantly from the 98 percent level as harvesttime price forecasts increasingly diverge from market expectations, and the speculative component becomes increasingly important.

Non-Constant Variance System

Estimation of the non-constant variance system is based on equations (20) and (29). Because the non-constant variance nature of the basis equation produces coefficient estimates which are inefficient, a weighted GLS procedure is used within the SUR-GLS estimation of the system.[14] The weighted regression results then provide a framework for the year-specific estimation of the non-constant variance parameters.

The weighting approach captures the relationship between the error variance and the Chicago stocks variable. The functional form of the relationship is determined by examining the effects of various weighting assumptions on the coefficient and t-ratio values of the OLS basis equation. The two weighting

[14]Least squares estimation is inefficient because it weights large and small errors equally. The weighted GLS procedure assigns larger weights to high variance observations than to other observations, providing minimum variance estimates.

assumptions used in the analysis are commonly encountered when heteroskedasticity is present.[15] The first assumption models the error variance as a linear function of the Chicago stocks level in each year, $(CHICST)_t$:

$$(40) \qquad \sigma_{bt}^2 = \sigma_b \, (CHICST)_t$$

The second employs a quadratic functional form:

$$(41) \qquad \sigma_{bt}^2 = \sigma_b \, (CHICST)_t^2$$

In performing the weighting procedure, all observations on each variable in the OLS basis equation are weighted by the associated Chicago stocks level in each year. If the structure in equation (40) is assumed, for instance, each row of the basis data matrix (including observations for both dependent and independent variables) is multiplied (weighted) by the corresponding $1/\sqrt{CHICST_t}$ value for that year.

The weighted basis equation estimates, shown in Table 11, are similar to the unweighted OLS estimates. Because the quadratic form used in estimating equation (43) produces an unexpected (positive) Chicago stocks coefficient, the linear weight [used in estimating equation (42)] is chosen for use in the SUR analysis. The signs associated with the linear weight are logical, but the t-ratios are lower than when the quadratic weight is used.

Expanding the analysis to the SUR system, the heteroskedasticity present in the (unweighted) basis equation also affects the across-equation error covariance. The non-constant variance relationships, using the linear Chicago stocks functional form suggested by equation (42), are given by:

$$(44) \qquad \sigma_{bt}^2 = \sigma_b^2 \, (CHICST)_t$$

$$(45) \qquad \sigma_{pbt} = \sigma_{pb} \, (CHICST)_t$$

The variance-covariance matrix for the system, in year-specific form, is given by:

---

[15]Because the heteroskedastic structure is estimated empirically, the weighted least squares estimator is only asymptotically efficient. As for small sample properties, recent theoretical results indicate that weighted least squares estimation is unbiased and that the efficiency loss from having to estimate the heteroskedastic structure is relatively small.

TABLE 11

WEIGHTED REGRESSION ESTIMATES FOR THE BASIS EQUATION[a] (1965-1984 SAMPLE PERIOD)[b]

<u>Weight = √(Chicst)$_t$</u>

(42) 25.91 - 1.46 LOCSUP + 0.0053 QLOCSUP + 0.16 NATSUPL -
(1.17)(-2.86) (3.81) (3.37)

0.000073 QNATSUPL + 3.81 OPPCOST - 0.87 CHICST - 12.94 DUM
(-4.68) (3.65) (-1.53) (-4.65)

<u>Weight = (Chicst)$_t$</u>

(43) 44.04 - 1.43 LOCSUP + 0.0052 QLOCSUP + 0.12 NATSUPL -
(2.75)(-4.45) (5.69) (4.19)

0.000061 QNATSUPL + 4.20 OPPCOST + 0.61 CHICST - 11.52 DUM
(-6.06) (5.63) (0.85) (-5.63)

---

a t-ratios are in parentheses.
b Variables are defined in Table 2.

$$
(46) \quad \Omega = \left[\begin{array}{ccc|ccc}
\sigma_b^2(CHICST)_t & 0 & & \sigma_{bp}(CHICST)_t & 0 & \\
 & \sigma_b^2(CHICST)_t & & & \sigma_{bp}(CHICST)_t & \\
 & & \sigma_b^2(CHICST)_t & & & \sigma_{bp}(CHICST)_t \\
0 & & \cdot & 0 & & \cdot \\
\hline
\sigma_{pb}(CHICST)_t & 0 & & \sigma_p^2 & & 0 \\
 & \sigma_{pb}(CHICST)_t & & & \sigma_p^2 & \\
0 & & \sigma_{pb}(CHICST)_t & 0 & & \sigma_p^2 \\
 & & \cdot & & & \cdot
\end{array}\right]
$$

The optimal hedge equation, combining equations (44) through (46), can be rewritten to reflect the changing nature of the parameter estimates. It is now given by:

$$
(47) \quad Q_h^* = Q[(\sigma_p^2+\sigma_{pbt})/(\sigma_p^2+2\sigma_{pbt}+\sigma_{bt}^2)] - [(F_{t-i}-P_t^*-B_t^*)/2m(\sigma_p^2+2\sigma_{pbt}+\sigma_{bt}^2)]
$$

Within the SUR system, the weighted basis equation provides a method for obtaining year-specific estimates of the optimal hedge parameters of equation (47). The variance-covariance matrix from the weighted SUR system provides estimates of the constant scale parameters, σ_b^2 and σ_{pb}. To obtain year-specific estimates of σ_{bt}^2 and σ_{pbt}, these scale parameters are multiplied by the Chicago stocks level in each year, $(CHICST)_t$. Yearly forecasts of the P_t^* and B_t^* levels are also developed using the weighted system estimates. The homoskedastic cash price equation is not weighted within the system, of course, and the SUR estimate of σ_p^2 is constant over time.

The SYSNLIN procedure required eight iterations for the SUR estimation of the non-constant variance system. The resulting SUR estimates of the cash price and weighted basis equations are presented in Table 12, and are similar to the OLS estimates. The associated scale error variance-covariance estimates are: σ_p^2 = 10,153.98; σ_b^2 = 11.18; and σ_{pb} = 211.55. These estimates are the constant variance parameters used to adjust the optimal hedge to the year-specific level.

The year-specific levels of the non-constant variance optimal hedge, calculated using equation (47), are shown in Table 13. For the extremely risk averse hedger (m=-infinity), the speculative component of the equation is zero. The optimal hedge, defined by the hedging component, varies from 98 percent of expected production in 1981 (with Chicago stocks at 0.846 million bushels) to a minimum 77 percent in 1983 (with stocks at 18.632 million bushels).

The results indicate that the optimal hedge declines as the Chicago stocks level increases, illustrating the impact of the Chicago stocks variable on the covariance calculation. The basis variance, with a scale parameter of 11.18, is not greatly affected by changing Chicago stocks levels. The Chicago stocks level has a large impact, however, on the year-specific covariance estimate because the

TABLE 12

SUR ESTIMATES OF THE NON-CONSTANT VARIANCE MODEL[a] (1965-1984 SAMPLE PERIOD)[b]

Basis Equation

(48) 26.68 - 1.36 LOCSUP + 0.0049 QLOCSUP + 0.14 NATSUPL -
(1.30)(-3.03) (4.03) (3.46)

0.000066 QNATSUPL - 1.10 CHICST + 3.17 OPPCOST - 10.24 DUM
(-4.72) (-2.06) (3.30) (-4.14)

Cash Price Equation

(49) 276.08 - 4.34 LOCPRO + 1.34 FOSEXP - 0.00056 QFOSEXP + 0.81 MAYNF
(1.67) (-4.05) (2.02) (-1.16) (3.34)

---

a t-ratios are in parentheses.
b Variables are defined in Tables 2 and 5.

TABLE 13

NON-CONSTANT VARIANCE ESTIMATES OF THE HEDGING COMPONENT

Year	Chicago Stocks	Hedge Ratio
	million bushels	% of production
1965	0.904	98
1966	1.331	97
1967	1.211	97
1968	2.279	95
1969	4.311	92
1970	5.865	90
1971	5.716	90
1972	2.485	95
1973	3.479	93
1974	5.391	90
1975	3.786	93
1976	6.891	88
1977	4.560	92
1978	3.353	94
1979	4.830	91
1980	16.223	79
1981	0.846	98
1982	1.908	96
1983	18.632	77
1984	2.555	95

associated scale parameter, at 211.55, is much larger. The denominator of the hedging component contains twice the year-specific covariance magnitude, and the optimal hedge declines quickly with increasing Chicago stocks (and year-specific covariance) levels.

Table 14 indicates that the risk aversion level often has a larger impact on the optimal hedge estimate than does the introduction of non-constant variance estimation. As an example, the 1974 optimal hedge for the extremely risk averse hedger is 90 percent of expected production. When the expected futures price increase of 203 cents per bushel is included in the analysis, $(F_{t-i}-P_t^*-B_t^* = 203)$, the optimal hedge drops to 82 percent (with m=-0.1), or 9 percent (with m=-.01). Similar impacts also appear in the years 1970, 1977, 1979, 1980, and 1983.[16]

An Alternative Approach to Estimation

The in-sample approach may not accurately reflect the risks confronting Central Illinois soybean producers. The regressions capture the sample period data well, with the basis equation, in particular, possessing a small mean squared error. The basis equation error variance is at such a low level, in fact, that the optimal hedge is near 100 percent of expected production for the extremely risk averse hedger. Because producers typically hedge at levels much lower than this, the basis equation forecasts may over-represent forecasting ability (and thus underestimate risk).

The out-of-sample approach used in this section is based on the sequential estimation of the constant variance system (equations (36) and (37)) for the time periods from 1965-1980 to 1965-1985. The system is estimated first for the 1965-1980 time period; the 1981 basis and cash price forecasts are made (using 1981 harvest data); and the 1981 forecast errors are calculated. The system is then updated to the 1965-1981 sample period, the 1982 forecasts and forecast errors are calculated, and the process is continued. There are six forecast errors in total, for the years 1981 through 1986 inclusive, on which the variance-covariance parameters are calculated.

Because there are only six forecast errors, the variance-covariance estimates may be quite sensitive to the addition or deletion of a year in the estimation process. Thus, the errors from four different time period lengths are used in the calculation: the 1981-1983 errors (three errors); the 1981-1984 errors (four errors); through the 1981-1986 errors (six errors). In application, the optimal hedge calculated for the 1981-1983 period would determine the hedge for 1984, and so on.

The results for the extremely risk averse hedger (m=-infinity) are shown in Table 15. They indicate that the out-of-sample hedge level ranges from 97 to 99 percent of expected production in each of the four series of years. These levels are near the 98 percent level suggested by the in-sample approach for the ex-

---

[16]Note that for the less risk-averse hedger (m=-0.1 or m=-0.01), non-constant variance estimation has a smaller effect on optimal hedges than does constant variance estimation.

TABLE 14

NON-CONSTANT VARIANCE ESTIMATES OF THE OPTIMAL HEDGE

Year	Price Change $(F_{t-i}-P_t^*-B_t^*)$	Hedging Component	Total Hedge (Hedging and Speculative Components)	
	cents per bushel	% of production	% of production	
			m=-0.1	m=-0.01
1965	-7.3	98	98	95
1966	-59.7	97	94	69
1967	-57.4	97	94	70
1968	-2.6	95	95	94
1969	48.3	92	94	112
1970	-139.0	90	85	35
1971	-25.6	90	89	80
1972	-50.5	95	93	73
1973	-32.6	93	92	79
1974	-203.3	90	82	9
1975	92.8	93	97	132
1976	-75.7	88	85	59
1977	129.6	92	97	145
1978	-10.7	94	94	89
1979	130.1	91	96	144
1980	-114.4	79	76	46
1981	61.0	98	101	127
1982	66.3	96	99	126
1983	-185.4	77	72	26
1984	-83.2	95	91	58

TABLE 15

OUT-OF-SAMPLE ESTIMATES OF THE OPTIMAL HEDGE RATIO USING EX POST FORECASTS

Period for Error Calculation	σ_p^2	σ_b^2	σ_{pb}	Optimal Hedge (% of production)
1981-1983	13,009	92	378	97
1981-1984	23,939	82	712	97
1981-1985	19,185	71	583	97
1981-1986	17,458	796	-555	99

tremely risk averse hedger; they are also within the 77 to 98 percent range indicated by the in-sample non-constant variance approach.

The previous analysis was repeated using springtime ancillary forecasts of the independent variables. Using this approach, the system is first estimated for the 1965-1980 period, and the 1981 cash price and basis forecasts are estimated (using ancillary forecasts of the independent variables developed from 1965-1980 regressions). The 1981 forecast errors are then computed as the difference between the realized prices and the forecast prices; a series of these errors are used to calculate the variance-covariance parameters. The estimation procedure continues until the 1986 forecasts are made.

The ancillary forecasts used in this process are given for the 1965-1984 period in Table 16. Each equation was determined from a search procedure that analyzed: the explanatory power of the dependent variable measured in May; variables showing a large correlation with the dependent variable; and a trend variable.

The regressors estimated by the ancillary analysis include: local production (LOCPRO); local supply (LOCSUP); national supply less local (NATSUPL); Chicago stocks (CHICST); opportunity cost (OPPCOST); and expected exports (FOSEXP). The production and supply estimates are the only forecasts which were not obtained directly from the ancillary equations. They are estimated as:

$$\text{NATPRO} = (\text{USHA})(\text{USYIELD}) \tag{59}$$
$$\text{NATSUPL} = \text{NATPRO} + \text{TNATSTO} - \text{LOCSUP} \tag{60}$$
$$\text{LOCPRO} = (\text{ILHA})(\text{IYIELD}) \tag{61}$$
$$\text{LOCSUP} = \text{LOCPRO} + \text{TLOCSTO} \tag{62}$$

National production (NATPRO) is the product of the average U.S. yield for a particular year and U.S. harvested acreage for that year [equation (59)]. National supply less local supply (NATSUPL) is the sum of national production and total stocks less the Central Illinois level of supply [equation (60)]. Central Illinois production and supply levels (LOCPRO and LOCSUP) are calculated in a similar manner, according to equations (61) and (62).

The results obtained from the ancillary approach, shown in Table 17, indicate that the optimal hedge varies from 93 to 100 percent of expected production in each sample period. Although the variance-covariance estimates are larger and more variable than before, the cash price variance continues to dominate the optimal hedge calculation. As a result, the optimal hedge for the extremely risk averse producer (m=-infinity) is near the 100 percent level of previous analyses.

Conclusions

Year-specific measurement of the optimal hedge indicates that the hedge level for the Central Illinois soybean producer can vary importantly from year to year. Specifically, non-constant variance estimation can reduce the optimal hedge level for

TABLE 16

ANCILLARY FORECASTS FOR THE BASIS AND CASH PRICE EQUATIONS
(1965-1984 SAMPLE PERIOD)

	Ancillary Equation	R^2
(50)	USYIELD[a] = 10.82 + 0.22 TRD	0.33
(51)	USHA = 1.09 + 0.96 MUSPA	0.97
(52)	TNATSTO = -86.22 + 0.36 MTNATSTO	0.76
(53)	IYIELD = 13.66 + 0.27 TRD	0.17
(54)	ILHA = 0.73 + 0.87 MILPA	0.93
(55)	TLOCSTO = -10.99 + 0.36 MTLOCSTO	0.72
(56)	CHICST = -4.27 + 0.012 MTNATSTO	0.43
(57)	OPPCOST = -20.45 + 0.32 TRD	0.69
(58)	FOSEXP = -2278.05 + 38.19 TRD	0.92

a USYIELD = U.S. soybean yield; TRD = trend; USHA = U.S. harvested soybean acreage; MUSPA = U.S. planted soybean acreage in May; TNATSTO = total U.S. stocks at harvest; MTNATSTO = total U.S. stocks in May.

IYIELD = Illinois soybean yield; ILHA = Illinois harvested soybean acreage; MILPA = Illinois planted soybean acreage in May; TLOCSTO = local Central Illinois stocks at harvest; MTLOCSTO = total Central Illinois stocks in May.

CHICST = Chicago stocks at harvest; OPPCOST = harvest opportunity cost; FOSEXP = predicted crop-year exports at harvest.

TABLE 17

OUT-OF-SAMPLE ESTIMATES OF THE OPTIMAL HEDGE RATIO USING ANCILLARY EX ANTE FORECASTS

Period for Error Calculation	σ_p^2	σ_b^2	σ_{pb}	Optimal Hedge (% of production)
1981-1983	29,309	453	-745	101
1981-1984	25,116	658	440	96
1981-1985	21,241	750	859	93
1981-1986	20,204	627	781	94

the extremely risk averse producer from 98 percent (the constant variance level) to approximately 77 percent of expected production. For the less risk averse hedger, both the constant and non-constant variance optimal hedge levels are affected to an even larger extent. Depending on producer price expectations, the optimal hedge estimates range from a short hedge of 158 percent to a long hedge of 27 percent of expected production.

Although the impact of non-constant variance estimation is evident, several limitations within the study must be noted. That larger Chicago stocks levels should increase basis variability (and reduce the short hedge as a percent of production) is not intuitive.[17] Other important qualifications involve data availability, the presence of multicollinearity, and uncertainties as to correct model specification. With a longer (and less collinear) data series, the equations may have been specified differently or they may have been more stable. In addition, the limited availability of data (particularly consumption demand data) may have resulted in omitted variable bias.

The specification of the objective function also limits implementation of the non-constant variance approach. Although the objective function used in this study is one of the most widely-used specifications, it has resulted in optimal hedges which are significantly larger than actual farmer hedging levels. This study suggests an optimal hedge of near 100 percent for the risk averse producer, but recent survey evidence indicates that less than 10 percent of Central Illinois farmers hedge, and typically at suboptimal levels (Olmstead).

One explanation for this behavior is that the costs of hedging are not accurately reflected in the objective function. Variation margin often represents an important cash flow constraint when futures prices move against the hedger (for instance, when the hedger is short and the futures price increases). Yield risk is another consideration: the farmer may be forced into a short speculative position (because of smaller quantities produced) at a time when futures prices are increasing. The farmer would experience a significant loss both in the cash market (through the decline in crop size) and in the futures market (through variation margin). The lumpiness of the contract is also important: because few farmers produce soybeans in multiples of the 5,000-bushel contract size, the producer again may be exposed to a substantial speculative position.

Refinements of the year-specific approach used in this study would create a broader understanding of the optimal hedge decision. One refinement involves improving data availability, particularly regarding local Central Illinois consumption demand. Other potential extensions involve analysis of different producer locations and the study of individual production/hedge levels. The examination of non-constant variance estimation of the optimal hedge for alternative commodities would also be useful.

Another potential contribution involves improved specifications of the objective function. Introducing yield risk and the uncertainty of margin calls is particularly important, along with the year-specific analysis of optimal hedge

---

[17]Indeed, the original hypothesis suggested that the error variance should increase, in both the cash price and basis equations, as inventories decline.

sensitivity to these alternative specifications. The non-constant variance approach could also be analyzed through a time-series framework, or by the use of implied options volatilities. Bootstrapping, involving Monte Carlo re-samplings of the empirical errors, provides an alternative approach to variance estimation.

Perhaps the primary conclusion is that the changing variance assumption has potential importance for year-specific estimation of the optimal hedge. In this paper, however, data problems and possible misspecification of the equations do not provide definitive conclusions about optimal hedges for Central Illinois soybean producers.

References

Anderson, Ronald W. "Some Determinants of the Volatility of Futures Prices." Journal of Futures Markets 5 (Fall 1985): 331-348.

Anderson, Ronald W. and Danthine, Jean-Pierre. The Time Pattern of Hedging, Volatility of Futures, and the Resolution of Uncertainty. Columbia University. Graduate School of Business. Center for the Study of Futures Markets. Working Paper Series No. CSFM-7. New York: Columbia University, April 1981.

Belsley, David A.; Kuh, Edwin; and Welsch, Roy E. Regression Diagnostics, first edition. New York: John Wiley and Sons, 1980.

Booth, G.G.; Kaen, F.R.; and Koveos, P.E. "Persistent Dependence in Gold Prices." Journal of Financial Research (Spring 1982): 85-93.

Castelino, Mark and Francis, Jack C. "Basis Speculation in Commodity Futures: The Maturity Effect." Journal of Futures Markets (Summer 1982): 195-206.

Castelino, Mark and Vora, Ashok. "Spread Volatility in Commodity Futures: The Length Effect." Journal of Futures Markets 4 (Spring 1984): 39-46.

Chicago Board of Trade. Statistical Annual. Selected issues. Chicago, Illinois: Chicago Board of Trade, 1965-1986.

Council of Economic Advisors. Economic Indicators. Washington, D.C.: U.S. Government Printing Office, October 1964-October 1986.

Ederington, Louis H. " The Hedging Performance of the New Futures Markets." The Journal of Finance 34 (March 1979): 157-170.

Fried, Joel. "Forecasting and Probability Distributions for Models of Portfolio Selection." The Journal of Finance 25 (June 1970): 539-554.

Garcia, Philip and Good, Darrel. " An Analysis of the Factors Influencing the Illinois Corn Basis, 1971-1982." In Applied Commodity Price Analysis, Forecasting, and Market Risk Management: Proceedings of the NCR-134 Conference, April 28-29, 1983. Des Moines, Iowa: April, 1983.

Good, Darrel. University of Illinois, Urbana, Ill. Personal Communication, Summer 1986 and Fall 1986.

Gordon, J. Douglas. The Distribution of Daily Changes in Commodity Futures Prices. United States Department of Agriculture. Economic Research Service. Technical Bulletin No. 1702. Washington, D.C.: U.S. Government Printing Office, July 1985.

Greenhall, Lawrence J. "Optimal Hedging Levels for Western New York Corn Producers: The Effects of Basis and Yield Risks." M.S. Thesis, Cornell University, 1984.

Hayenga, Marvin L. and DiPietre, Dennis D. "Hedging Wholesale Meat Prices: Analysis of Basis Risk." Journal of Futures Markets 2 (Summer 1982): 131-140.

Heifner, Richard. Hedging Potential in Grain Storage and Livestock Feeding. United States Department of Agriculture. Economic Research Service. Agricultural Economic Report No. 238. Washington, D.C.: U.S. Government Printing Office, 1973.

Heifner, Richard. "Optimal Hedging Levels and Hedging Effectiveness in Cattle Feeding." Agricultural Economics Research 24 (April 1972): 25-30.

Helmberger, Peter G. and Akinyosoye, Vincent. "Competitive Pricing and Storage Uncertainty with an Application to the U.S. Soybean Market." American Journal of Agricultural Economics 66 (May 1984): 119-130.

Houck, James P. and Mann, Jitendar S. An Analysis of Domestic and Foreign Demand for U.S. Soybeans and Soybean Products. The University of Minnesota Agricultural Experiment Station. Technical Bulletin No. 256. St. Paul: The University of Minnesota, 1968.

Illinois Co-Operative Crop Reporting Service. Illinois Department of Agriculture and United States Department of Agriculture Co-Operating. Illinois Agricultural Statistics. Bulletins 64-1 - 86-1. Springfield, Ill.: Illinois Co-Operative Crop Reporting Service, 1964-1985.

Judge, George G.; Griffiths, William E.; Hill, R. Carter; and Lee, Tsoung-Chao. The Theory and Practice of Econometrics, first edition. New York: John Wiley and Sons, 1980.

Kahl, Kandice. "Changes in the Chicago Corn Basis, 1960-1975." Agricultural Economics Research 34 (January 1982): 25-29.

Kmenta, Jan. Elements of Econometrics, second edition. New York: MacMillan Publishing Company, 1986.

Knipscheer, Hendrick C. and Hill, Lowell D. The Demand for Soybean Meal by the European Community: An Econometric Model. The University of Illinois. Department of Agricultural Economics. Agricultural Economics Research Report No. 186. Urbana: The University of Illinois, September 1982.

Kofi, Tetteh A. "A Framework for Comparing the Efficiency of Futures Markets." American Journal of Agricultural Economics 55 (November 1973): 584-594.

Leamer, E.E. "Let's Take the Con Out of Econometrics." American Economic Review 73 (1983): 31-43.

Levy, H. and Markowitz, H.M. "Approximating Expected Utility by a Function of Mean and Variance." American Economic Review 69 (June 1979): 308-317.

Mann, Jitendar and Heifner, Richard. The Distribution of Short-Run Commodity Price Movements. United States Department of Agriculture. Economic Research Service. Technical Bulletin No. 1536. Washington, D.C.: U.S. Government Printing Office, March 1976.

Martin, Larry; Groenewegen, John L.; and Pidgeon, Edward. "Factors Affecting Corn Basis in Southwestern Ontario." American Journal of Agricultural Economics 62 (1980): 107-112.

Mathews, J.L.; Womack, A.W.; and Hoffman, R.G. "Formulation of Market Forecasts for the U.S. Soybean Economy with an Econometric Model." United States Department of Agriculture. Economic Research Service. Fats and Oils Situation. No. 260. Washington, D.C.: U.S. Government Printing Office, November 1971.

Meilke, Karl D. and Griffith, Garry R. "Incorporating Policy Variables in a Model of the World Soybean/Rapeseed Market." American Journal of Agricultural Economics 65 (February 1983): 65-73.

Meyer, Jack. "Two-Moment Decision Models and Expected Utility Maximization." American Economic Review 77 (June 1987): 421-430.

Miller, Katherine Dusak. "The Relation Between Volatility and Maturity in Futures Contracts." In Commodity Markets and Futures Prices. Ed. Raymond M. Leuthold. Chicago: Chicago Mercantile Exchange, 1979.

Milonas, Nikolaos T. and Vora, Ashok. "Sources of Non-Stationarity in Cash and Futures Prices." In Review of Research in Futures Markets Vol. 4, No. 3. Ed. A.E. Peck. Chicago: Chicago Board of Trade, 1984.

Olmstead, Craig. (Illinois State University). "Grain and Soybean Producers' Use of Minimum Price Contracts." Presentation given at the Symposium on Options, Futures, and Agricultural Commodity Programs, Arlington Hotel, Arlington, Virginia, May 27-28, 1987.

Paul, Allen. "The Pricing of Binspace--A Contribution to the Theory of Storage." American Journal of Agricultural Economics 52 (February 1970): 1-12.

Paul, Allen. Treatment of Hedging in Commodity Markets. United States Department of Agriculture. Economic Research Service. Technical Bulletin No. 1538. Washington, D.C.: U.S. Government Printing Office, April 1976.

Peck, Anne E. "Hedging and Income Stability: Concepts, Implications and an Example." American Journal of Agricultural Economics 57 (August 1975): 410-420.

Powers, Nicholas and Johnson, Aaron Jr. "Forecasting the Storage-Season Wisconsin Basis for Corn." In Applied Commodity Price Analysis, Forecasting and Market Risk Management: Proceedings of the NCR-134 Conference, April 28-29, 1983. Des Moines, Iowa: April, 1983.

Rutledge, D.J.S. "A Note on the Variability of Futures Prices." Review of Economics and Statistics 58 (February 1976): 120-123.

Samuelson, Paul A. "Is Real World Price a Tale Told by the Idiot of Chance?" Review of Economics and Statistics 58 (February 1976): 120-123.

Samuelson, Paul A. "Proof that Properly Anticipated Futures Prices Fluctuate Randomly." Industrial Management Review 6 (Spring 1965): 41-49.

Taylor, Stephen J. "The Behavior of Futures Prices Over Time." Applied Economics 17 (August 1985): 713-734.

Telser, Lester G. "Reasons for Having an Organized Futures Market." In Livestock Futures Research Symposium. Ed. Raymond M. Leuthold and Parry Dixon. Chicago: Chicago Mercantile Exchange, 1979.

Tomek, William G. and Gray, Roger W. "Temporal Relationships Among Prices on Commodity Futures Markets: Their Allocative and Stabilizing Roles." American Journal of Agricultural Economics 52 (August 1970): 372-380.

United States Department of Agriculture. Agricultural Statistics. Washington, D.C.: U.S. Government Printing Office, 1965-1985.

United States Department of Agriculture. Crop Reporting Board. Crop Production. Selected issues. Washington, D.C.: U.S. Government Printing Office, 1985-1986.

United States Department of Agriculture. Crop Reporting Board. Grain Stocks in All Positions. Selected issues. Washington, D.C.: U.S. Government Printing Office, 1985-1986.

United States Department of Agriculture. Economic Research Service. Fats and Oils Situation. Selected issues. Washington, D.C.: U.S. Government Printing Office, 1964-1983.

United States Department of Agriculture. Economic Research Service. Oil Crops Outlook and Situation Report. Selected issues. Washington, D.C.: U.S. Government Printing Office, 1983-1986.

Vandenborre, R.J. An Econometric Analysis of the Market for Soybean Oil and Soybean Meal. The University of Illinois. Department of Agricultural Economics. Agricultural Economics Research Report No. 106. Urbana: The University of Illinois, July 1970.

Working, Holbrook. "The Theory of the Price of Storage." American Economic Review 39 (December 1949): 1254-1262.

Zellner, Arnold. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." Journal of the American Statistical Association 57 (1962): 348-368.