Working Papers in
AGRICULTURAL ECONOMICS

February 1991 **91-4**

Variability in Soybean Futures Prices: Economic and Econometric Issues

by

Deborah H. Streeter and William G. Tomek

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

Abstract

The variance of price changes for soybean futures contracts is analyzed using recent developments in econometric modeling. The econometric methodology starts with a comprehensive model and uses a systematic approach to simplify the specification. The empirical results suggest that both the characteristics of the flow of new information into the market and the structure of the market are important factors influencing changes in the variance. It is difficult to appraise the value of the econometric procedures, but the final results appear to be relatively more robust than if a traditional, ad hoc approach to modeling had been used.

The authors are, respectively, Assistant Professor and Professor of Agricultural Economics at Cornell University. The authors would like to thank Bernard Stanton for his helpful comments. Funding for this research was provided by USDA Cooperative Agreement 58-3AEK-8-00102 and Hatch Grant No. 447.

Variability in Soybean Futures Prices:
Economic and Econometric Issues

> "To make sense of model design as a positive tool, it is essential to distinguish very clearly between two aspects of the modelling process, namely, the context of discovery and the context of justification. In the former, it is accepted that we do not know the truth and are trying to construct a useful model that helps us understand the gestalt of existing evidence. . . . In the context of justification, however, attention is concentrated on critical evaluation, and the claimed model is subjected to searching examination to locate its strengths and weaknesses; . . ."
>
> -Hendry, p. 37

Most studies of price behavior have focused on changes in levels or means of prices, while ignoring the variance of prices. Only recently have analysts come to understand that measures of volatility, such as the variance of price changes, shift over time in systematic ways. This paper is about the specification and evaluation of models designed to explain price volatility. Understanding the behavior of the variance of price changes for futures contracts is of interest to a variety of decision-makers, including exchange committee members who set margins on futures contracts, market participants who trade options and futures contracts, and regulators.

Two distinct threads exist in the literature on price volatility of futures. One focuses on how volatility is influenced by "state variables," such as the time-to-maturity or seasonal components of information flows (e.g., Anderson and Danthine). The other thread of work on price volatility emphasizes the effects of market structure measures, such as the ratio of speculators to hedgers (e.g., Peck). A need exists, however, to weave these two threads together in a more unified framework. Thus, a primary objective of this paper is to specify a model of the variance of changes in futures prices which includes both state and market structure variables.

While the principal thrust of the paper is concerned with economic issues, a second objective focuses on econometrics issues. The robustness of empirical results has been questioned (e.g., Leamer), and it has become increasingly obvious that models with seemingly "healthy" t-ratios, R^2s, and Durbin-Watson statistics nonetheless may be misspecified (e.g., Spanos 1990). Given the sensitivity of empirical results to changes in specification and sample period, reliance on common statistical measures to guide empirical work belies the considerable difficulty of obtaining precise and consistent results. Previous studies of the changes in the variance of futures prices are vulnerable to this criticism.

This paper uses the prices of soybean futures prices as a case study to explore the economic and econometric issues raised above. A general model is specified, following the principles of Hendry and Richard, and a battery of specification error tests is used to judge the performance of the resulting model. The general model is built using an integrated framework which considers both market structure and information flow effects. Efforts are also made to correctly represent the dynamic structure in the model.

Having started with a relatively complete model, the final model, which passes most of the specification error tests, produces qualitatively similar results to those obtained in the initial model, which passed only a few of the specification tests. In addition, changes to the model resulted only in modest improvements in t-ratios and stability of coefficients. Thus, in this particular case, researchers would not have been grossly misled by considering only standard econometric measures of goodness of fit.

The paper is organized as follows. Following a brief review of previous work, the general model is presented, including subsections on three groupings of variables: market structure, flow and certainty , and current economic information and interaction. In the next two sections, modeling procedures are discussed and results are presented. The summary and conclusions are included in a final section.

Previous Work

Early studies of price volatility of futures (Rutledge, Miller) investigated the time-to-maturity effects, labeled the "Samuelson effect" because Samuelson suggested that the variance of prices should increase as contract maturity approaches. More recently, Anderson and Danthine have argued that the time-to-maturity effects are really a special case of what they call the "state variable hypothesis." This hypothesis suggests that the *ex ante* variance of futures prices depends on the pattern of demand and supply uncertainties, which are resolved with the passage of time. In addition, Kenyon et al. make a distinction between economic variables, such as current production levels, and the state variables. In general, they found that both types of variables influenced price volatilities.

Research focusing on Samuelson or state variable effects has assumed that futures markets are competitive. However, other researchers have responded to concerns that "excess speculation" might increase volatility in futures markets. For example, Peck explored the impact of changes in the level of speculation on wheat, corn, and soybean prices. Her results indicate that in the 1964-78 period, speculation and price variability are inversely related. Accordingly, Peck concluded that allegations of too much speculation were unfounded, and in fact that speculation was inadequate relative to hedging activity, which had strained the liquidity of the markets and increased price variability. In another study of market structure influences, Brorsen and Irwin failed to find support for another popular hypothesis that the growth in volume attributed to futures pools (funds) increases price variability.

The General Model

Despite the fact that earlier studies found no evidence that increased speculation increases futures price volatility, Peck's paper suggests that structure of the market can influence the behavior of futures prices. In addition, as speed and volume of information flows increase with improved communication technology, there is concern over how such flows affect price movement. Thus, the model of price volatility in this paper includes explanatory variables to represent three conceptual categories: flow of information and certainty about this information, current economic information, and market structure.

Price behavior in futures markets depends on the entry of new information into the market and on how the market responds to this information. Truly new information enters randomly; otherwise, it would be predictable and prices would have anticipated the effects of the predictable component. Nonetheless, new information can have seasonal and trend components. For example, precipitation amounts and frequency can be seasonal, but the amount and timing of an individual storm are unpredictable until shortly before its occurrence.

Market structure and current levels of supply and demand provide the economic environment in which prices adjust to new information. Specifically, the importance of hedging use relative to speculation, the liquidity of the market, and the degree of market concentration may affect price adjustments. In addition, price adjustments may be larger or smaller depending on whether current inventories are small or large relative to current demand. Thus, there is good reason to believe that interaction effects may exist between the introduction of new information and the current level of supply and demand.

To study these general concepts empirically, the November and March soybean futures contracts were selected for analysis. Soybean prices typically have been above government support levels and appear to have had varying levels of volatility through time. The November and March contracts are characterized by large trading volume and represent two different parts of the marketing year.[1] For the analysis reported in this paper, the dependent variable is the variance of daily changes of the logarithms of prices, estimated by month. The following subsections discuss the explanatory variables.[2] The specific definitions and units of measure for the variables are given in Table 1 for the 1976-1986 sample period.

Market Structure variables

Three variables in the model reflect various facets of market structure: a speculative index, a measure of scalping (reflecting liquidity), and a measure of market concentration. The speculative index, originally developed by Working (see Peck), is intended to measure the adequacy of speculation as an offset to hedging.[3] Peck hypothesized that the speculative index would be inversely related to price volatility. A large index implies that speculation is large relative to short hedging use, presumably providing hedgers with a large quantity of speculation on the opposite side of their trades. However, other observers think that speculation can be too large relative to hedging use. Thus, the expected sign of this variable is unclear.

The speculative index variable reflects the position trading of speculators, and consideration must also be given to scalping activity, in which traders enter and exit the market often but do not have open positions at the end of a trading day. Unfortunately, data are not available on the level of scalping in a market. In the absence of a better variable, the ratio of daily volume to open interest in a given contract month is computed and then averaged for the month as a proxy for scalping activity.[4] Students of futures markets typically expect that the increased liquidity provided by scalpers is associated with reduced price variability, although Peck obtained a positive relationship.

The model also contains variables representing market concentration. While the speculative index and the proxy for the scalping variable focus specifically on speculative influences, the concentration variables are intended to reflect the presence of large positions relative to total open interest, regardless of whether they are hedging or speculative positions. As Paul has pointed out, large hedgers may have more opportunities than large speculators to influence price behavior. The concentration measures are defined as the percent of total open interest in all soybean contracts held by the four largest traders in long and short positions respectively, as reported by the Commodity Futures Trading Commission. Although there is little theory or empirical evidence to suggest how market concentration affects the variance of prices, a popular presumption is that as large players wield their influence, price volatility increases. Thus, one might expect large concentration ratios to be associated with large variances.

Flow and Certainty of Information Variables

There are several components to the flow and certainty of information effects in the soybean market. If the Samuelson effect holds, then *ceteris paribus*, volatility should increase as the time-to-maturity decreases[5] because the amount of information, which affects price, increases as contract expiration approaches. For example, if maturity were five years distant, little or no new, influential information would be available from day to day. On the other hand, if maturity were five months distant, new information available each day would influence prices. By analogy, Samuelson argued that as contract maturity approaches, the amount of information about that maturity month increases. However, in practice, trading in commodity futures contracts only begins when prices are variable and nearby and more distant contracts for the grains are closely linked through inventories. Thus, it is unclear whether a measurable "Samuelson effect" can be found.

Clearly the uncertainty about soybean production has a seasonal component. The hypothesis is that price variability is largest during the growing season and declines as crop prospects become more certain. Harmonic variables are used to reflect seasonality because potentially a smooth seasonal can be well represented with less than 11 variables. The choice of sine and cosine variables is made on an empirical basis.[6]

A price level variable, computed as the average of daily closing prices for each month, is also included. Since prices are influenced by changes in expected economic conditions, they help measure information flow effects. In addition, the lagged dependent variable is used as an independent variable, which is analogous to an ARCH effect (Engle). As explained below, a systematic procedure was used to derive the lag structure of the model.

Current Economic Situation and Interaction Variables

Three measures of supply and use are included to measure the current economic context: annual total supply (production plus carry-in), monthly disappearance, and mill stocks at the beginning of the month. Thus, use is measured relative to stocks currently in the hands of soybean crushers and relative to the initial total supply for the crop year (Table 1).

Interaction terms are included in the models. Information about expected supply and demand becomes available throughout the seasonal cycle and the extent of its impact on prices is conditioned by the size of current supplies. The interaction variables are defined as the products of various supply-type variables with seasonality and time-to-maturity variables. Specific definitions are given in Table 1.

Thus, the economic questions examined by this paper include (1) whether or not the variables included are important influences on futures price variability, (2) whether or not significant interactions exist between categories of information-type variables, and finally, (3) whether or not gains can be obtained by using an inclusive framework as a basis for analysis.

Procedures

The model specification approach is based on the philosophy of Hendry and his colleagues (e.g., Hendry and Richard; McAleer, et al.; Hendry). The procedure consists of a systematic search for a well specified model, starting from a model which includes (hopefully) all potentially important variables, including a complete specification of the lag structure. Then the model is simplified (Hendry and Richard), using a battery of diagnostic tests as a quality control device (e.g., McAleer, et al. p. 299). The intent is to obtain a model that is consistent with theory and the data generating process. The final model is expected to have relatively stable parameter estimates and to encompass alternative models.

The procedure was initially applied to a single equation for the November contract. Then the March equation was fitted, and the two equations were combined into a seemingly unrelated regression (SUR) model. Space limitations do not permit discussion of all details of all models, but we highlight major points.

The single equations are estimated by ordinary least squares, a feasible generalized least squares estimator, and a nonlinear least squares estimator. The latter two estimators are used because common factor tests suggest that a second-order autocorrelation specification for the error terms is appropriate. The nonlinear least squares estimator has the practical virtue that the common factor restrictions are imposed directly on the model. Then the error terms in each equation have the "classical" properties, and the two equations can be fitted as a seemingly unrelated regression system. As it turns out, the feasible GLS and the nonlinear LS estimates of the single equations are almost identical, and only the results from using GLS are reported.[7]

The following subsections emphasize decision-making about model specification, reserving the detailed discussion of the economic content of the empirical findings for the next section. The narration begins with the single equation estimation of the November model and ends with the system estimation of the November and March contracts.

November Model

A model for the November delivery month was considered first, and an attempt was made to specify a general model both in terms of conceptual components and dynamics. The initial specification included four lags for every variable except the time-to-maturity, seasonal, and interaction variables. Then, following McAleer et al. nested tests were used to identify possible common factors in the lag structure, which was eventually reduced to a model with a one period lag in the regressors and second-order autocorrelation in the residuals.[8] In addition, at this stage several alternative definitions of particular concepts were explored, and the seasonal and interaction specifications were simplified based on statistical tests. A "final" result for this equation is reported as model N1 in the first column of Table 2.

In general, the model seems to perform well, given the reasonable R^2 and large t-ratios for variables in each of the major classes of variables. In many cases, the signs of coefficients are consistent with expectations, but there are some seemingly illogical results as well. For example, the inverse of current mill stocks has a negative sign, implying that larger inventories are associated with larger variability. Other variables expected to be important, such as the disappearance variable, have relatively low t-values. The economic content of the model is revisited in the next section.

Despite the believable and seemingly usable results from this model, the outcome of the specification testing is discouraging. The test results reported in Table 3 show that Model N1 passes only the autocorrelation test and the trend version of the specification error test. It fails the linearity, normality and heteroscedasticity tests. In addition to the tests presented, the stability of the coefficients was examined, using a recursive estimation procedure[9] and the coefficients for many of the variables changed substantially in the 1978-1983 period. In a number of instances (including the time-to-maturity variable), the coefficients change from negative to positive. (For more detail, see Streeter and Tomek.) Model N1 clearly fails the tests of adequacy, which raises doubts about whether other empirical models reported in the literature could pass similar tests.

Further work was done to improve the performance of the model. First, partial-regression leverage regression plots revealed July 1977 as a significant outlier.[10] Hence, revised models have the relevant data point dummied out.[11] Second, possible nonlinearities in the regressors were considered, and the square of the time-to-maturity variable was added. Based on small t-ratios, the seasonal structure and interaction terms were simplified, and several lag terms were dropped. Results for the revised model (Model N2) are reported in column 2 of Table 2. The results of the specification tests for the N2 (Table 3) show a considerable improvement, as the model passes all but one of the heteroscedasticity tests.

March Contract

The November specification was imposed on the March soybean data (with appropriate adjustments for contract-specific variables) and the results are reported in the third column of Table 2 (Model M1). For some variables, such as those in the category of economic impacts, the November and March models perform in a similar way, but there are some important differences. For example, the flow of information variables are less powerful in explaining price variability in the March contract than they are in the November contract.

Turning to specification considerations, the standard econometric results are reasonable, although the adjusted R^2 is somewhat lower for March than for November. The specification tests, as reported in Table 3, indicate some problems with linearity and heteroscedasticity.

In light of the results for the M1 model, an additional model was specified based on a simpler interaction structure. The resulting model (Model M2) is displayed in the fourth column of Table 2. In general, modest gains were made over the previous specification. The linearity test statistic moved below the critical value (Table 3); t-values were generally improved; and the results are closer to those reported for the November contract.

An example of an improvement achieved by modifying the March model specification is the impact of the seasonality variables. The seasonal effects for Models N2, M1 and M2 are shown in Figure 1. Before the modifications, the M1 seasonal structure differed somewhat from the November contract, especially in the early part of the year, but with the model revision, the seasonal effect is similar in the two models.

Coefficient stability is an important concern. The original specification of the November model resulted in coefficients that are unstable over the sample period. The recursive coefficient plots for both contract months for key variables are displayed in Figure 2. The overall picture is positive, as the sign of each coefficient is unchanged and the coefficient values stabilize at the end of the sample period to the values reported in this paper.[12]

System Estimation, a Seemingly Unrelated Regression Model

The single equation estimates reported for the March and November contracts produced models that have reasonable economic results in terms of economic logic. However, the single equation estimation procedure ignores the likely relationship between the error terms of the two equations, and since some regressors are different, the SUR estimator should provide statistically more efficient estimates. The t-values should increase and the overall performance of the models should improve.[13]

The SUR results are reported in Table 4.[14] In terms of econometric performance, the outcome can be characterized as mixed. The t-ratios did not increase uniformly, although they were sharpened for some important variables (TIME11, SPINDEX, AVMO11, AVMO03). In general the coefficient values were similar to the single equation estimates. Finally, the autocorrelation of residuals appears not to be very significant, as the t-values on the AR terms are less than 2.

To summarize, the procedures followed for evolving the model from a single equation to a system resulted in improvements from an econometric standpoint. The process produced a model which performed well on most of the tests of specification error, and can be considered more robust that the initial model.

Results

The flow of information effects are examined first. The seasonal effects, when examined on a net basis,[15] are consistent across all models; as expected, prices are more volatile during the growing season. In addition, the lagged dependent variable is consistently important, with a coefficient of about 0.4. Apparently, distributed lag effects exist in the response of monthly changes in the variance. As expected, the higher the average price, the higher the volatility, other factors constant. Interactions between the seasonal variables and the quantity variables were less important than expected.

The partial derivative of volatility with respect to TIMEii can be used to evaluate the marginal impact of time-to-maturity:

$$\partial LOGDIFii/\partial t = b_1 + 2b_2t + b_3AVMOii + b_4SUPTOT$$

where: t = TIMEii
b_1 is coefficient on TIMEii
b_2 is coefficient on TIMEii^2
b_3 is coefficient on TIMEii x AVMOii
b_4 is coefficient on TIMEii x SUPTOT

If the Samuelson effect holds, the sign of $2b_2t$ (the slope of the equation) should be negative, indicating that as t grows smaller (approaches maturity), volatility is larger. For all the models which contain the squared term, b_2 is indeed negative, lending support to the Samuelson hypothesis. However, the magnitudes of the estimated b_2's vary considerably, and the results for the shifters, b_3 and b_4, also vary. Although *a priori* expectations were that large supplies would dampen the time-to-maturity effects, this is supported only in the initial March model (M1), while for all other models, total supply shifted the effects of time-to-maturity downward. In addition, for most models, the average monthly price is a downward shifter, indicating that at higher price levels the time-to-maturity effects on volatility are muted. These puzzling results may be caused by problems in measurement of variables, an issue to be discussed further.

In the market structure category, the results are fairly consistent across all models. The positive impact of the concentration variables on volatility coincides with expectations, although the concentration of short traders is far more important than of long traders. The speculative index is a significant negative influence on variability, confirming previous studies which show that high levels of speculation were not associated with increased volatility. Finally, while expectations were that a high degree of liquidity provided by large scalping activity reduces volatility, the scalping variable is positively associated with the dependent variable. Although the sign is the same as in Peck's paper, it is an unexpected result and simply may mean that the measure of scalping is a poor proxy for the concept.[16]

The findings on variables representing economic information did not differ substantially across models. The supply and demand variables (total supply, mill stocks, and inventory) perform poorly, either having no effect or having illogical signs. The average price level is positively related to variability, and since price can be viewed as aggregating all market information, it may make the other economic variables redundant. Unfortunately it is difficult to disaggregate the effects of expectations and the economic context into which new information is introduced.

Finally, the coefficient for dummy variable representing the data point July 1977 is positive with a large t-ratio. This finding is consistent across all models.

A major challenge for future work is to find better proxies for the economic variables. In particular, proxies which vary on a monthly basis are highly desirable.

Conclusions and Implications

The companion goals of the paper are to specify a general, unified model for price variability of soybean futures and to use recent literature in econometrics to conduct a systematic search for a well specified model. Consequently, the results have implications not only for the economic issues of interest, but also for the on-going debates about model-building philosophies.

The results of the study support the value of a unified framework in which the influence of both information flows and market structure considerations are incorporated. In this regard, the model provides evidence on the following points: (1) the Samuelson effect exists in the soybean markets but depends in part on interactions with information variables; (2) seasonality is an important influence with the variance largest in the summer months, and important interactions exist between seasonal variables and other information flows; (3) speculators do not dominate hedgers in a way that creates additional price volatility and indeed when speculation is large relative to hedging, the variance is smaller; (4) the effects of activity by larger traders (especially on the short side) do increase volatility, (6) a clear distributed lag effect exists in the models, and (7) large supplies perhaps reduce variability.

The value of starting with a large model and seeking a more parsimonious one by the use of a battery of specification tests is more difficult to appraise. We have demonstrated that initial results, which have reasonable conventional test statistics, can have highly unstable coefficients and that it is possible to modify the model to obtain more stable coefficients, at least for the given sample period. Nonetheless, only the passage of time will tell whether the results obtained by this procedure are robust in the face of new data points. One encouraging result is that the qualitative nature of the economic interpretation changes relatively little over the alternative models, when one starts with a model which encompasses a number of special cases. A discouraging result, however, is that the magnitudes of certain coefficients differ considerably over alternative models. For example, the coefficient of the time-to-maturity ranges from -.38 to .59.[17] In Leamer's terms, some coefficient estimates are fragile.

We suspect, but cannot show authoritatively, that if we had started with a simple model and tried to build to a more complex model, the results would have been econometrically poorer than those obtained. For comparison purposes, we did fit a model analogous to the one used by Peck, which contains only market structure variables. Even though the qualitative results, such as a negative sign on the speculation variable and a positive sign on the scalping variable, were similar in both models,[18] the Peck model failed nearly all specification tests. Clearly our model encompasses such earlier models, and in this sense represents an improvement.

An important problem in studies of futures prices is potential errors in measuring the explanatory variables. The flow of information into the market is represented by imperfect proxies like time-to-maturity trends and seasonal variables and the structure of the market, which influences the response to new information, is probably not well measured by available data. This paper has emphasized modeling in the context of the imperfect data which are available, and while we probably have not exhausted the alternative models, further progress may be limited by existing data.

Clearly obtaining useful empirical econometric models is proving to be difficult in many areas of study and appears to be particularly difficult for models of price volatility. Nonetheless, the results in this paper suggest that some useful insights can be obtained into the factors affecting the variance. The next steps in research include extending the analysis to other markets, to updated data sets, and, if available, to better data.

References

Anderson, Ronald W., and Jean-Pierre Danthine. "The Time Pattern of Hedging and the Volatility of Futures Prices," Review of Economic Studies 50(1983): 249-266.

Belsely, David A., Edwin Kuh, and Roy E. Welsch. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley and Sons, 1980.

Brorsen, B. Wade, and Scott H. Irwin. "Futures Funds and Price Volatility," The Review of Futures Markets 6(1987): 118-135.

Doran, H.E., and J.J. Quilkey. "Harmonic Analysis of Seasonal Data: Some Important Properties," American Journal of Agricultural Economics 54(1972): 646-651.

Engle, Robert F. "Autoregessive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations," Econometrica 50(1982): 987-1007.

Harvey, A. C. The Econometric Analysis of Time Series. Philip Allan, 1981.

Hendry, David F. "Econometric Methodology: A Personal Perspective," Advances in Econometrics Fifth World Congress, Volume II, ed. Truman F. Bewley, Cambridge: Cambridge University Press, 1987.

Hendry, David F., and Jean-Francois Richard. "On the Formulation of Empirical Models in Dynamic Econometrics." Journal of Econometrics 20(1982): 3-33.

Kenyon, David, et al. "Some Determinants of the Volatility of Futures Prices," The Journal of Futures Markets 7(1987): 73-91.

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

McAleer, Michael, et al. "What Will Take the Con out of Econometrics?" American Economic Review 75(1985): 293-307.

Miller, K.D. "The Relation between Volatility and Maturity in Futures Contracts." in Commodity Markets and Futures Prices. R.M Leuthold, ed. Chicago Mercantile Exchange. 1979. pp. 25-36.

Paul, Allen B. Treatment of Hedging in Commodity Market Regulation. USDA Technical Bulletin No. 1538, April 1976.

Peck, Anne E. "Measures and Price Effects of Changes in Speculation on the Wheat, Corn, and Soybean Futures Markets," in Research on Speculation. Chicago Board of Trade, 1981. pp. 138-149.

Rutledge, David. J. S. "A Note on the Variability of Futures Prices." Review of Economics and Statistics 63(1976): 118-120.

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

Spanos, Ariel. "An Integrated Approach to Linear Regression: Specification, Misspecification, Respecification and Identification." Department of Economics, VPI, 1990 (unpublished).

Spanos, Ariel. Statistical Foundations of Econometric Modelling. Cambridge University Press, 1986.

Streeter, Deborah H. and William G. Tomek. Models of the Variability of Futures Prices: Specifications and Evaluation, Cornell Staff Paper 89-14, May 1989.

Endnotes

[1]The time series used in this research are based on the twelve months which run from the prior year's maturity month through the month just prior to maturity. Thus, maturity month observations are ignored, as are observations more than 12 months before maturity. This is justified by the relative thinness of trading in distant months and by possible aberrant observations that sometime occur at expiration. The entire sample consists of 11 years of such data.

[2]Peck and Brorsen and Irwin use measures of variability which are based on the daily range of prices. Accordingly, this study also considered an alternate model, in which the dependent variable was the monthly average of the daily price range, measured from the daily high to the daily low. As results were similar, only the variance model is presented in this paper. For more detail, see Streeter and Tomek.

[3] For the precise formula, see footnote d, Table 1. The components of the index are taken from monthly Commitment of Traders data published by the Commodity Futures Commission. For details on how the authors resolved various problems arising from this data source, see Streeter and Tomek.

[4]This is a slight variation on the procedure followed by Peck, who computed the scalping variable by averaging daily volume during the month and dividing by the open interest observed at the end of the month. We also explored using a variable that aggregated the volume and open interest components for all contract months, but the variable did not perform well and was dropped.

[5]Time-to-maturity is measured as the number of months left to contract expiration.

[6]A full specification of the seasonal effect would make the dependent variable a function of the sum of six sine and six cosine variables (Doran and Quilky). However, one of the variables must be dropped in a linear model to avoid perfect collinearity. It is anticipated that four or five such variables can adequately represent the seasonal pattern.

[7]The estimators were implemented using Micro TSP, and the close agreement of the GLS and nonlinear LS results imply that the computations are accurate.

[8]The intermediate model had third-order autocorrelation, but the third-order autocorrelation coefficient typically had a small t ratios, see Streeter and Tomek.

[9]The coefficients are estimated recursively starting with degrees of freedom equal to the number of independent variables, and adding one

row of data at a time. The update procedure outlined in Harvey (pp. 54-56) is used. Belsely et al argue that degrees of freedom should be such that K/T > .4 (p. 17), where K is the number of parameters and T is the number of observations, so that no individual observation is too influential.

[10] Based on articles appearing in the *Wall Street Journal* during the month of July, 1977, one explanation for the high levels of volatility is that the market was very vulnerable to weather scares. Prices rose in early 1977 as useage was high and carry-in from the previous year had been at record low levels. However, in late April optimism began to erode as it became clear harvested acreage would be higher than the previous year. Between April and July, prices fell by about 40%. On the one hand, a record high crop could quickly cancel out the effects of strong demand. On the other hand, given very low carry-in and brisk demand, any weather damage could easily lead to a bull market. Weather patterns in July of 1977 were sporadic, which may have led to unusually large fluctuations in price.

[11] Before resorting to the dummy variable, the suspect data point was reviewed to eliminate the possibility of a computational error. Creating a zero-one variable in which only one observations equals one is equivalent to deleting that data point. However, it has the advantage of retaining a continuous data set and providing a measure of the importance of the observation through the size of the t-ratio on the coefficient of the dummy variable.

[12]The final ending value on the plot may differ slightly from the coefficient values reported in Table 2, because the recursive estimation procedure does not allow for AR terms.

[13]The motivation for using an SUR framework comes from the belief that there is high correlation between the residuals of the two equations. However, the gains to be made depend in part on how many common regressors are used in the equations within the system. For example, if the regressors are identical, no benefit exists. Therefore, since the equations in this model share many regressors, only modest gains can be expected.

[14]The AR(2) structure is taken into account by imposing these common factor restrictions on the two equations and estimating the system in a nonlinear least squares framework.

[15] For interaction terms including seasonal variables, the other variables are evaluated at their means.

[16]Specifically, volume of trading may be large relative to open interest precisely because prices are variable, and their large volume may or may not represent larger amounts of scalping.

[17]The range in coefficient estimates includes models from the early stages (e.g., the model with no lags, estimated with Ordinary Least Squares) of the specification procedures, as well as models presented in

this paper. The coefficient's stability over the sample period is improved in the final model relative to earlier models.

[18]Peck's model, which used changes in the daily range of prices to measure price volatility, produced similar qualitative results.

Figure 1. Net Seasonal Effects For Various Models

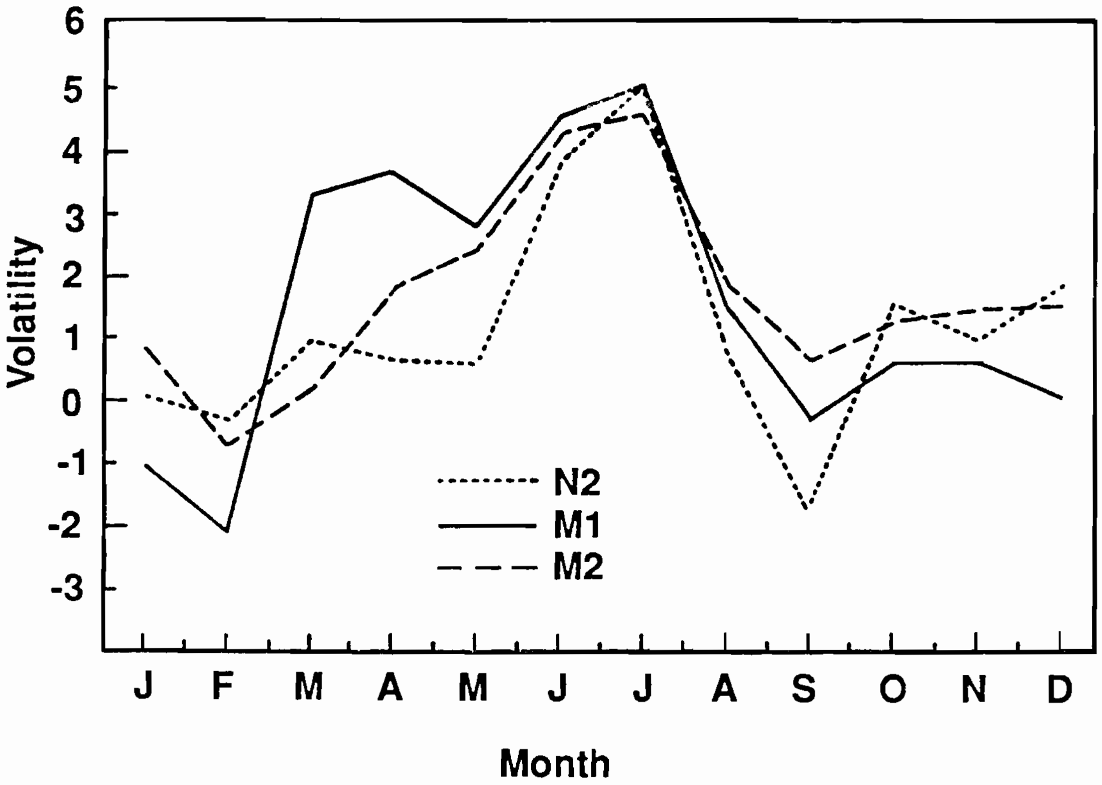

Figure 2. Recursive Coefficient Estimates for Selected Variables (1978-1986)

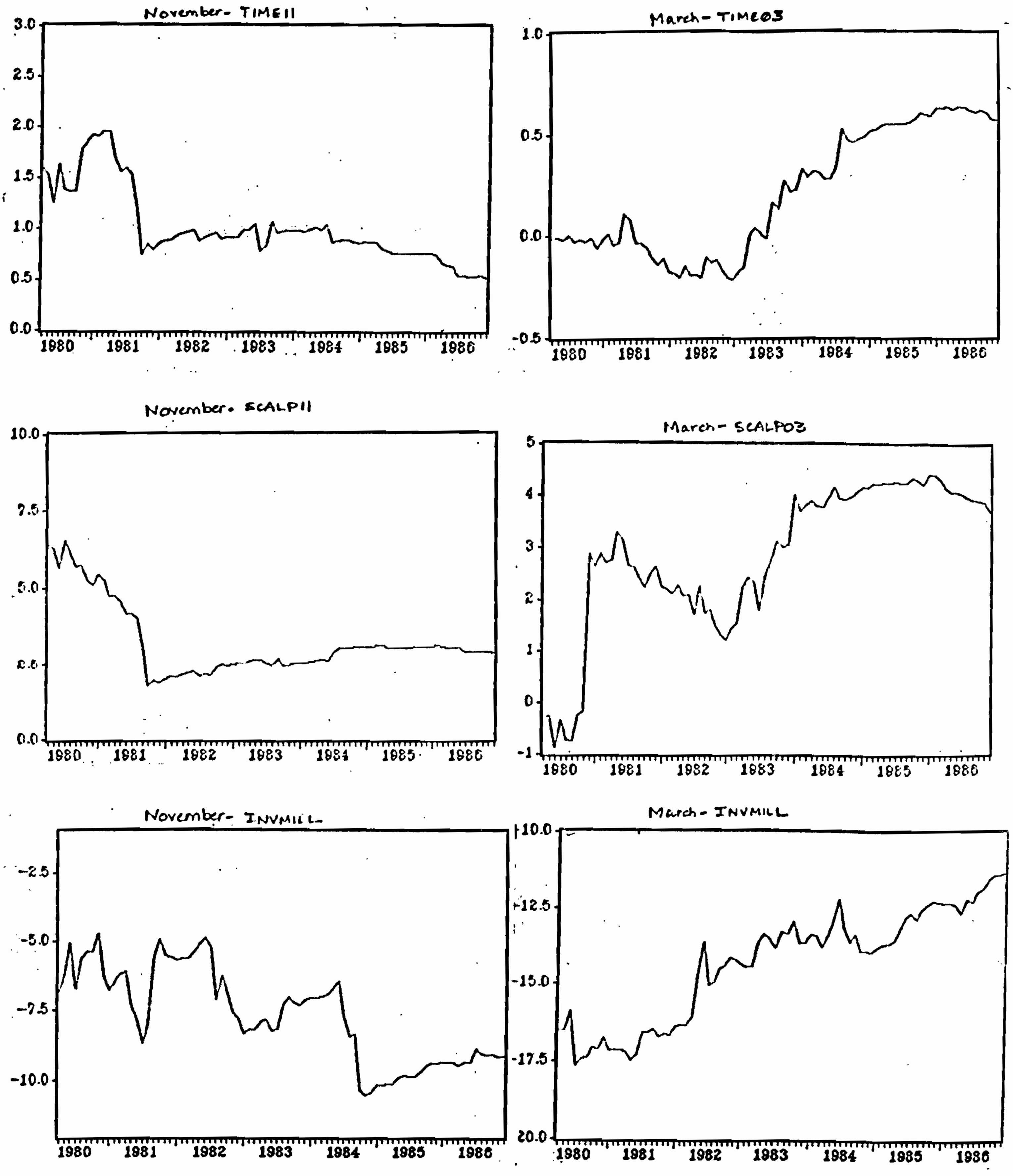

Table 1. Definition of Variables in Models of Variance of Price Changes, Soybean Futures

Name	Description	Mean[a]	S.D.
AVMO03	Monthly Average of Daily Close (March Contract)	6.65	1.06
AVMO11	Monthly Average of Daily Close (November Contract)	6.49	.98
DARANG	Monthly Average of Daily Price Range (November Contract)	.10	.05
DUMMY	Dummy Variable Where =1 if 1977.07	--	--
INVMILL	Inverse of the MILL variable (see below) x 10,000[b]	.01	.01
LOGDIF03	Monthly Variance of Daily Differences in Log Prices x 10,000 (March Contract)	1.79	1.35
LOGDIF11	Monthly Variance of Daily Differences in Log Prices x 10,000 (November Contract)	1.68	1.38
LONG4	Monthly Percentage of Long Open Interest Held by[c] Four Largest Traders	9.4	3.9
MILL	USDA Estimate of Mill Stocks (Million Bushels)	89.6	39
SCALP03	Monthly Average of Daily Ratio of Volume to Open Interest (March contract)	.22	.18
SCALP11	Monthly Average of Daily Ratio of Volume to Open Interest (November Contract)	.37	.21
SHORT4	Monthly Percentage of Short Open Interest Held by Four Largest Traders	13	4.7
SPINDEX	Speculative Index[d]	1.52	.15
SUPTOT	Annual Total Supply Estimate of USDA (Million Bu.)[b]	2109	284
TIME03	Number of Months to Expiration of March Contract (Mar=12, Apr=11, May=10, etc.)	--	--
TIME032	TIME03 Squared	--	--

TIME11	Number of Months to Expiration of November Contract (Nov=12, Dec=11, Jan=10, etc.)	--	--
TIME112	TIME11 Squared	--	--
TOTDIS	Monthly USDA Estimates of Soybean Disappearance (Million Bushels)	142,274	32,197
<u>Seasonal Variables</u>			
COS1	Cosine of First Harmonic Wave [$\cos(\lambda_k t)$, where $\lambda_k=2\pi\kappa/12$, and t=2]	.0075	.7097
COS2	Cosine of Second Harmonic Wave	.0075	.7097
COS4	Cosine of Fourth Harmonic Wave	.0075	.7097
SIN2	Sine of Second Harmonic Wave	0	.7044
SIN3	Sine of Third Harmonic Wave	0	.7044
Sin4	Sine of Fourth Harmonic Wave	0	.7044

a. Sample period is 1975.12 - 1986.12 with 1977.07 omitted.

b. Source: USDA.

c. Source: Monthly Commitment of Traders Report, CFTC.

d. The speculative index is:

$$1 + \frac{SS}{HS + HL} \text{ when } HS > HL, \text{ and } 1 + \frac{SL}{HL + HS} \text{ when } HL > HS$$

where SS = speculation short positions
SL = speculation long positions
HL = hedging long positions
HS = hedging short positions

and unbalanced matching trades are allocated to short or long.

Table 2. Estimated Variance of Price Change Equations, Soybean Futures, Using GLS

	November Contract		March Contract	
	Model N1	Model N2	Model M1	Model M2
Flow of Information Variables				
1. LOGDIFii(-1)[b]	.56	.47	.49	.40
	(5.98)[c]	(7.43)	(5.21)	(3.88)
2. TIMEii[a]	-.43	.57	.52	.56
	(-2.05)	(1.69)	(1.38)	(1.42)
3 TIMEii^2	---	-.07	-.01	-.04
	---	(-2.92)	(-.55)	(-1.70)
4. COS1	-.47	.78	-1.12	-1.26
	(-1.99)	(2.16)	(-3.00)	(-2.90)
5. COS2	2.53	2.18	.87	1.11
	(4.02)	(3.38)	(1.02)	(2.23)
6. SIN2	-.12	---	---	---
	(-.22)	---	---	---
7. SIN3	-2.44	-1.76	-.9	-.44
	(-3.25)	(-2.37)	(1.00)	(-2.79)
8. SIN4	.31	.31	.57	.48
	(1.56)	(2.45)	(4.00)	(3.42)
9. AVMOii	.96	1.24	1.18	.88
	(3.96)	(6.09)	(3.96)	(3.65)
10. AVMOii(-1)	-.72	-.87	-.81	-.72
	(-3.38)	(-4.87)	(-3.48)	(-3.10)
Interaction Terms				
11. COS1xINVMILL.	---	---	---	2.99
	---	---	---	(1.02)
12. COS2xINVMILL	-7.16	-7.79	-10	-8.46
	(-2.92)	(-3.17)	(-3.09)	(-2.41)
13. COS2xSUPTOT	-.0007	-.0004	.0003	---
	(-2.42)	(-1.25)	(.76)	---
14. SIN2xAVMOii	.06	---	---	---
	(.75)	---	---	---
15. SIN3xSUPTOT	.001	.001	.0002	---
	(3.05)	(1.96)	(.57)	---
16 SIN4xLOGDIFii	.1	---	---	---
	(1.3)	---	---	---
17 TIMEiixAVMOii	-.02	-.03	-.03	---
	(-.92)	(-1.91)	(-1.22)	---
18. TIMEiixSUPTOT	.0003	.0002	-.0001	---
	(3.51)	(3.25)	(-1.51)	---

Table 2 Continued

Market Structure Variables					
19.	SPINDEX	-.71	-.89	-.68	-.63
		(-1.38)	(-2.03)	(-1.16)	(-1.00)
20.	SCALPii	2.74	2.27	2.71	3.71
		(3.96)	(5.17)	(1.84)	(2.51)
21.	SCALPii(-1)	-1.05	---	---	---
		(-1.58)	---	---	---
22.	LONG4	.03	.04	.009	.004
		(1.89)	(2.62)	(.47)	(.18)
23.	SHORT4	.03	.03	.03	.03
		(2.27)	(2.33)	(1.54)	(1.21)
Economic Variables					
24.	SUPTOT	-.001	0	.001	.001
		(-.55)	(0)	(.9)	(.71)
25.	SUPTOT(-1)	-.001	-.001	-.0003	-.001
		(-.8)	(-1.09)	(-.26)	(-.7)
26	INVMILL	-10.89	-10.3	-11.46	-11.51
		(-4.62)	(-4.53)	(-4.50)	(-4.36)
27.	INVMILL(-1)	3.27	4.47	4.84	3.03
		(1.87)	(2.41)	(2.45)	(1.47)
28.	TOTDIS	-.02	-.07	-.67	-.08
		(-.32)	(-1.56)	(-1.14)	(-1.28)
29.	TOTDIS(-1)	-.11	-.09	-.09	-.08
		(-1.97)	(-1.72)	(-1.53)	(-1.21)
Other Variables					
30.	Constant	5.34	2.19	-.4	1.83
		(2.79)	(1.31)	(-.21)	(1.31)
31.	Dummy	---	4.44	4.36	4.55
	(1977.07=1, otherwise 0)	---	(5.52)	(4.40)	(4.80)
32.	AR(1)	-.40	-.35	-.15	-.04
		(-2.91)	(-3.01)	(-1.04)	(-.28)
33	AR(2)	-.06	-.16	-.1	-.08
		(-.45)	(-1.47)	(-.84)	(-.75)
34.	AR(3)	-.12	---	---	---
		(-1.97)	---	---	---

Table 2 Continued

Adjusted R^2	.75	.81	.70	.70
DW	2.0	2.0	2	2
N=133 (1975.12-1986.12)				

a The designation of ii indicates the appropriate contract month, that is either 11 or 03.
b One month lag designated by (-1).
c. Ratio of coefficient to its standard deviation.

Table 3. Specification Tests,[a] November and March Equations

Test for	November N1	November N2	March M1	March M2	Critical Values
1 Specification Error					
a. Linearity	19	1.22	3.19	3.06	3.09
b. Trend	1.18	3.82	2.36	2.26	3.09
2. Normality	35	.8	.96	5.72	5.99
3. Heteroscedasticity					
a. Breusch-Pagan test	56	65	53.2	53	22.4
b. Arch-Type test	12	3.99	1.33	1.33	9.49
4. Autocorrelation					
a. P1	.04	-.06	-.63	-.25	1.98
b. P2	.02	-.37	-.74	-.38	1.98

a. All but one of the tests used in the study are described in Spanos (1986). The linearity version of the specification error test (p.460) used the squared and cubed residuals in the auxiliary regression of a Lagrange multiplier-type test. Thus, the test can be interpreted as a test for linearity in the relation between the regressors and the dependent variable (or more generally as a test for omitted variables). In addition, another version of the auxiliary regression was run using a trend variable and its squared and cubed terms. The test for normality is described on p. 453, the Breusch-Pagan test for heteroscedasticity on p. 469, and the test for autocorrelation on p. 542. The ARCH-type test for heteroscedasticity is described in Engle (p. 1000), and used four lags of the squared residuals.

Table 4. Seemingly Unrelated Regression Model for Variance of Price Changes, Soybean Futures

		November N3	March M3
Flow of Information Variables			
1.	LOGDIFii(-1)	.43	.41
		(6.69)	(5.24)
2.	TIMEii	.59	-.10
		(2.64)	(-.43)
3.	TIMEii2	-.04	-.004
		(-2.23)	(.26)
4.	COS1	.2	-.75
		(.81)	(-3.09)
5.	COS2	1.82	1.64
		(4.4)	(4.15)
6.	SIN3	-.43	-.40
		(-1.33)	(-2.96)
7.	SIN4	.38	.47
		(3.37)	(3.78)
8.	AVMOii	.84	.47
		(6.89)	(3.68)
9.	AVMOii(-1)	-.33	-.12
		(-2.96)	(-.99)
Interaction Terms			
10.	COS1xINVMILL	---	.99
		---	(.08)
11.	COS2xINVMILL	0	-10.7
		(.35)	(-3.86)
12.	COS2xSUPTOT	-.0001	---
		(-.99)	---
13.	SIN3xSUPTOT	0	---
		(.19)	---
14.	TIMEiixAVMOii	-.03	---
		(-4.31)	---
15.	TIMEiixSUPTOT	0	---
		(.35)	---

Table 4 Continued

Market Structure Variables			
16.	SPINDEX	-.98 (2.16)	-1.38 (-2.54)
17.	$SCALP_{ii}$	.73 (2.92)	1.52 (2.57)
18.	LONG4	.02 (1.40)	-.001 (-.05)
19.	SHORT4	.04 (3.20)	.04 (2.73)
Economic Variables			
20.	SUPTOT	.0003 (.33)	-.0004 (-.32)
21.	SUPTOT(-1)	0 (-.01)	.001 (.71)
22.	INVMILL	-8.4 (-4.11)	-9.94 (4.36)
23.	INVMILL(-1)	2.97 (1.92)	5.08 (3.03)
24.	TOTDIS	-.007 (-1.62)	-.01 (-2.24)
25.	TOTDIS(-1)	-.01 (-2.47)	-.01 (-1.75)
Other Variables			
26.	CONSTANT	.36 (.30)	2.95 (2.86)
27.	DUMMY (1977.07=1, otherwise 0)	4.9 (6.36)	4.44 (5.12)
28.	AR(1)	-.18 (-1.99)	-.15 (-1.52)
29.	AR(2)	-.1 (-1.31)	-.02 (-.3)
Adjusted R^2		.74	.66
DW		1.86	1.8
N = 133 (1975.12-1986.12)			

Other Agricultural Economics Working Papers

No. 90-8	National Factor Markets and the Macroeconomic Context for Environmental Destruction in the Brazilian Amazon	Steven C. Kyle Aercio S. Cunha
No. 90-9	Pitfalls in the Measurement of Real Exchange Rate Effects on Agriculture	Steven C. Kyle
No. 90-10	Policy Analysis in an Imperfectly Competitive Market: A Conjectural Variations Model for the Food Manufacturing Industry	Leo Maier
No. 90-11	Control in a Dynamic Economy: Modeling the Behavior of the Chinese Village Leader	Scott Rozelle Richard N. Boisvert
No. 90-12	Demographic and Income Distribution Factors in Demand Systems for Food	Deborah Peterson Timothy Mount
No. 90-13	Welfare Impacts of Generic Certificates on U.S. Corn Producers	Douglas J. Miller Jason A. Karszes
No. 90-14	A Bioeconomic Model of the Pacific Whiting	Jon M. Conrad
No. 90-15	Optimal Depletion of Non-Renewable Resources under Linear Homogeneous Stock-Dependent Cost	Robert W. Jefferson
No. 90-16	The Economics of a Stock Pollutant: Aldicarb on Long Island	Jon M. Conrad Lars J. Olson
No. 90-17	Productive and Predatory Public Policies: Research Expenditures and Producer Subsidies in Agriculture	Harry de Gorter David J. Nielson Gordon C. Rausser
No. 91-1	Departmental Rankings by AJAE Publication Rates	Deborah H. Streeter Harry M. Kaiser
No. 91-2	Stock Pollutants and Risky Accumulation	Jon Conrad
No. 91-3	Influence of Agricultural Lending Policies on Commercial Bank Loan Volume	Eddy L. LaDue John Thurgood