Soybean Market Forecast Errors

by

Arthur Havenner and Marlene Cerchi

SOYBEAN MARKET FORECAST ERRORS

Arthur Havenner and Marlene Cerchi*

Two conditions are required for profitable transactions in financial markets: the market forecast must be wrong, and the bettor's forecast must be right. Of course, all market participants are trying to make money. If their actions aggregate into prices that always "fully reflect" available information, then the market meets the ideal (for resource allocation) that Fama has called "efficient," and there will be no profit[1] to be made on average. In this case, the price revisions would be completely random and any modeling attempt would be in vain. One of Fama's sufficient conditions--costless information--turns out to be necessary as well, however,[2] so that we should not immediately grant the market omniscience in all cases.

This paper examines the particular case of soybean, soybean meal, and soybean oil prices to determine whether or not the costs associated with gathering information on the myriad of international forces affecting these prices leave a small expected profit for a sophisticated

*Arthur Havenner and Marlene Cerchi from the University of California at Davis.

forecaster. After extending the notion of market efficiency to include the types of nonstationary stochastic processes required for agricultural commodities in the second section, five series of revisions in the aggregated forecasts of market participants--each corresponding to Department of Agriculture announcements--are introduced in the third section, and a multivariate time series model is fitted to these revisions in the fourth section. Analysis of the model parameter estimates and forecasts brings a conclusion on soybean market efficiency.

Nonstationarity and Market Efficiency

At any given time, the market prices of soybeans and related products are a complicated dollar-weighted function of market participants' estimates of the impacts of a variety of exogenous influences. Consider, for example, the March future for December soybean oil, a price that depends on estimates of everything from farmer's plantings and eventual yields to storage costs and processing capacity, even including hedging behavior (as processors sell oil forward to hedge earlier bean purchases) and the forecast price of complements (the joint product meal); the quantitative and judgmental analysis of the factors necessary to make an informed estimate of these prices is difficult in the extreme. The market does, however, produce forecasts incorporating available information on these factors each time a future price evolves. If market participants can be characterized as risk neutral on the margin, then a (stochastic) arbitrate equilibrium condition must hold between the forecasts of all future futures (including the zero future, the spot) for delivery in a given period. Suppose, for example, that a contract for delivery in period τ is entered in period t; if either party to the transaction expected that later future prices for period τ would be more advantageous to them, they would wait until that period to act, thus changing supply (driving current futures up and expected future futures down) or demand (moving current futures down and expected future futures up) until the market-weighted expectations are all the same. Thus, future prices for period τ formed in period t, ${}_tF_\tau$, necessarily contain forecasts of all future futures for that terminal period:

$$ {}_tF_\tau = {}_{t+1}\hat{F}_\tau = \ldots = {}_{\tau-1}\hat{F}_\tau = {}_\tau\hat{F}_\tau . $$

If expectations are rational, these will even be optimal forecasts, either minimum mean squared error, or at least minimum variance unbiased. Of course, these forecasts are not at all the same as forecasts for other terminal periods. Everyone may expect soybeans to be expensive in July and cheap in October, so that ${}_{\text{March}}F_{\text{July}} = {}_{\text{July}}\hat{F}_{\text{July}}$, and ${}_{\text{March}}F_{\text{October}} = {}_{\text{July}}\hat{F}_{\text{October}}$, but no one would expect ${}_{\text{October}}\hat{F}_{\text{October}} = {}_{\text{March}}\hat{F}_{\text{March}}$ for example, or even ${}_{\text{March}}\hat{F}_{\text{October}} = {}_{\text{March}}F_{\text{March}}$; thus, the price distribution is nonstationary in that, at a minimum, its mean exhibits seasonal shifts. [If these were General Motors stock prices rather than soybean prices, it is possible to construct a model such that the optimal forecast of any future price is the current price (the venerable random walk on Wall Street), so that in March both

${}_{\text{October}}\hat{F}_{\text{October}} = {}_{\text{March}}F_{\text{March}}$ and ${}_{\text{July}}\hat{F}_{\text{October}} = {}_{\text{March}}F_{\text{March}}$. In this case differences of the historical spot prices are forecast errors and will be serially uncorrelated given the efficient markets assumptions.] When the mean of the price distribution depends on exogenous variables, as is the case with agricultural commodities, forecasting price in different periods is not sufficient for profitable transactions, since changes in prices are not market forecast errors--everybody knows strawberries are expensive in December. On the other hand, differences of successively revised futures with the same terminal period, ${}_{t+1}F_{\tau} - {}_{t}F_{\tau}$, ${}_{t+2}F_{\tau} - {}_{t+1}F_{\tau}$, ..., ${}_{\tau}F_{\tau} - {}_{\tau-1}F_{\tau}$, do represent a sequence of market forecast errors, because each future is a forecast of all future futures for that terminal period. Thus, for example, the forecast error ${}_{t+1}F_{\tau} - {}_{t+1}\hat{F}_{\tau}$ is ${}_{t+1}F_{\tau} - {}_{t}F_{\tau}$, since ${}_{t}F_{\tau}$ is the market forecast (the conditional expectation given the information available at time t, if expectations are rational) of ${}_{t+1}F_{\tau}$.

If markets are efficient by Fama's definition, the forecasts are optimal (incorporating all available information) and the future price for any period is a sufficient statistic, i.e., $E_t({}_{t+j}F_{\tau} | {}_{t}F_{\tau}) = E_t({}_{t+j}F_{\tau} | \Omega_t)$, where Ω_t is all other data at time t. In this case, there can be no pattern to the market forecast revisions that is exploitable _a priori_, else it is contained in Ω_t and thus in ${}_{t}F_{\tau}$.

Suppose, however, that agents do not enjoy this game of stochastic arbitrage for itself, and must be compensated for continually monitoring the market. If they cannot earn a return for their efforts, they will not engage in the arbitrage necessary to police the market, in which case they could profitably arbitrage, which, when they do, guarantees there can be no profit in arbitrage. This is a paradox due to Grossman and Stiglitz, who note that costless information is not merely a sufficient condition for market efficiency, but a necessary condition as well. The sheer quantity and variety of information to be incorporated in soybean price forecasts suggests that the assumption of costless information may be inappropriate so that revisions in forward prices may not be without pattern.

Revisions in Market Forecasts

Daily observations from 1967 to 1978 on Chicago Board of Trade spot and future prices on soybean, soybean oil, and soybean meal contracts have been used to construct five series of changes in the market forecasts corresponding to USDA Crop Reporting Board Announcements of intended acreage (March), planted acreage (July), and stocks on hand (January, April, and October). For each of five maturity months (March, May, July, August, and September), closing prices for the last working day prior to the announcement have been subtracted from closing prices for the first working day after the announcements to provide five series of market forecast revisions (at each of the announcement dates) for the five maturity months for each commodity (beans, meal, and oil). Each series is a set of revisions in implicit market price forecasts for all maturities for all soybean products, as opposed to a price series for each commodity and maturity. Thus, each series represents a pooled time series (for the years 1967-1978) of cross-sections (for the five maturity months) of cross-sections (for the three commodities--beans, meal, and oil); see Table 1. Pooling the observations in this way is based on the assumption that autocorrelations--although not the product of autocorrelations and past revisions, i.e., forecasts--are constant over maturities and commodities for each series of revisions, since they all depend on the same innovation, the announcement information. The revisions can realistically be modelled as stationary stochastic processes, even though the prices themselves could not, on the assumption that market forecasts have already incorporated the mean-shifting nonstationarities (such as seasonality, trade agreements, weather, etc.), leaving only the stochastic components--a design that avoids reproducing the market wisdom and concentrates exclusively on improving it.

The Multivariate Time Series Model and Forecasts

Fitting a multivariate time series model[3] to the five series of market forecast revisions provides a basis for tests of market efficiency and forecasts of market price revisions. Jointly estimating all five series allows more efficient parameter estimates and forecasts when there is correlation between the error terms of the various series. The model is

$$\Phi(B)\underline{Z}_t = \underline{\delta} + \Theta(B)\underline{u}_t$$

Table 1. Revisions in Market Forecasts of Maturity Months for the Year and Commodity Listed

Year	Commodity	Series				
		January revisions	March revisions	April revisions	July revisions	October revisions
1968						
	Beans	March	March	. . .		March
		May	.			.
		July	.			.
		August	.			.
		September	September			September
	Meal	March	March	. . .		March
		May	.			.
		July	.			.
		August	.			.
		September	September			September
	Oil	March	March	. . .		March
		May	.			.
		July	.			.
		August	.			.
		September	September			September
1977						
	Beans	March	March	. . .		March
		May	.			.
		July	.			.
		August	.			.
		September	September			September
	Meal	March	March	. . .		March
		May	.			.
		July	.			.
		August	.			.
		September	September			September
	Oil	March	March	. . .		March
		May	.			.
		July	.			.
		August	.			.
		September	September			September

where $\underline{\delta}$ is a vector of intercepts and $\Phi(B)$ and $\Theta(B)$ are matrix polynomials in the backshift operator B of the form

$$\Phi(B) = I - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_p B^p$$

and

$$\Theta(B) = I - \Theta_1 B - \Theta_2 B^2 - \ldots - \Theta_q B^q.$$

The (5.15 x 1) vector $\underline{Z}_t$ is composed of observations on all five maturities for all three commodities (the 15) for each of the five series of revisions in the t^{th} year, and $\underline{u}_t$ is the analogous vector of white noise errors. The additional assumptions of normality and constant covariances Σ between series for all time-series and cross-sectional units complete the specification.

The iterative time-series identification procedures of Maravall and Tiao et al. suggest a vector autoregressive model of order three; the full information maximum likelihood parameter estimates in Table 2 are based on 105 effective observations, after dropping 45 observations per series for each of the three startup years.

Autocorrelations of the residuals and mean squared errors of predicted values in and out of sample provide measures of model validity. Means, standard deviations, and cross-autocorrelations of the in-sample residuals are presented in Table 3. There were no significant correlations beyond lag four, and even without the additional correlation matrices of lag five and beyond, the six significant correlations out of the hundred to lag four are approximately what would be

Table 2. Model Parameter Estimates and Standard Errors

$$Jan^{Z}{}_{\tau} = \underset{(.47)}{2.0} + \underset{(.077)}{.33}\ Mar^{Z}{}_{\tau-1} + \underset{(.044)}{.49}\ Oct^{Z}{}_{\tau-1} + \hat{u}_{\tau}$$

$$Mar^{Z}{}_{\tau} = \underset{(.082)}{-.14}\ Jan^{Z}{}_{\tau-1} - \underset{(.06)}{.30}\ Apr^{A}{}_{\tau-1} - \underset{(.097)}{.38}\ Mar^{Z}{}_{\tau-2} - \underset{(.088)}{.097}\ Oct^{Z}{}_{\tau-2} + \underset{(.14)}{.44}$$

$$Jan^{Z}{}_{\tau-3} + \underset{(.19)}{.30}\ Mar^{Z}{}_{\tau-3} - \underset{(.22)}{.88}\ Oct^{Z}{}_{\tau-3} + \hat{u}_{\tau}$$

$$Apr^{Z}{}_{\tau} = \underset{(.74)}{3.98} + \underset{(.13)}{.69}\ Jan^{Z}{}_{\tau-1} + \underset{(.17)}{.54}\ Mar^{Z}{}_{\tau-1} - \underset{(.11)}{.62}\ Apr^{Z}{}_{\tau-2} - \underset{(0.96)}{.12}\ Oct^{Z}{}_{\tau-2}$$

$$+ \underset{(.28)}{1.68}\ Oct^{Z}{}_{\tau-3} + \hat{u}_{\tau}$$

$$Jul^{Z}{}_{\tau} = \underset{(1.46)}{3.54} + \underset{(.24)}{.46}\ Jan^{Z}{}_{\tau-1} - \underset{(.11)}{.42}\ Jul^{Z}{}_{\tau-1} + \underset{(.18)}{.24}\ Oct^{Z}{}_{\tau-2} + \hat{u}_{\tau}$$

$$Oct^{Z}{}_{\tau} = \underset{(.095)}{-1.38}\ Jan^{Z}{}_{\tau-1} - \underset{(.046)}{.072}\ Jul^{Z}{}_{\tau-1} + \underset{(.074)}{.50}\ Oct^{Z}{}_{\tau-1} + \underset{(.20)}{.78}\ Oct^{Z}{}_{\tau-3} + \hat{u}_{\tau}$$

Error covariance matrix:

$Jan^{u}{}_{\tau}$	26.12				
$Mar^{u}{}_{\tau}$	-5.14	14.40			
$Apr^{u}{}_{\tau}$	-1.84	-8.71	41.90		
$Jul^{u}{}_{\tau}$	9.44	-8.49	-1.46	201.14	
$Oct^{u}{}_{\tau}$	20.52	-2.68	-15.72	13.57	55.20
	$Jan^{u}{}_{\tau}$	$Mar^{u}{}_{\tau}$	$Apr^{u}{}_{\tau}$	$Jul^{u}{}_{\tau}$	$Oct^{u}{}_{\tau}$

Table 3. Residual Summary Statistics

		$\text{Jan}^u\tau$	$\text{Mar}^u\tau$	$\text{Apr}^u\tau$	$\text{Jul}^u\tau$	$\text{Oct}^u\tau$
Mean		0.19	-0.07	-0.11	0.14	0.49
Standard deviation		5.11	3.79	6.47	14.18	7.41
Correlations[a]						
Lag 1:	Jan^u	*	*	*	*	*
	Mar^u	*	*	*	*	2.35
	Apr^u	*	*	*	*	*
	Jul^u	*	*	*	*	*
	Oct^u					
Lag 2:	Jan^u	*	*	*	*	-2.35
	Mar^u	*	*	*	*	-2.67
	Apr^u	*	*	*	*	*
	Jul^u	*	*	*	*	*
	Oct^u	*	*	*	*	*
Lag 3:	Jan^u	*	*	*	*	*
	Mar^u	*	*	*	*	-2.67
	Apr^u	*	*	*	*	-2.77
	Jul^u	*	*	*	*	*
	Oct^u	*	*	*	*	*
Lag 4:	Jan^u	*	*	*	*	*
	Mar^u	*	*	*	*	*
	Apr^u	*	*	*	*	2.56
	Jul^u	*	*	*	*	*
	Oct^u	*	*	*	*	*

[a]* Indicates residual correlations insignificant at the 5 percent level; the numbers are t-statistics.

expected by chance alone: by this measure the model appears to be correctly specified.

The root-mean-squared-errors (RMSE's) are presented in Table 4, along with Theil's proportionate decomposition of the mean squared errors into bias, variation, and correlation effects:

$$[MSE/MSE] = 1 = [(\overline{F}-\overline{A})^2 + s_F-S_A)^2 + (1-\rho)s_F s_A]/MSE$$

where, respectively, $\overline{F}$ and $\overline{A}$ are the means of the forecasts and the actual values, s_F and s_A are the standard deviations of the forecasts and the actual values, and ρ is the correlation coefficient between forecast and actual values. None of the forecasts exhibit either serious bias or mismatch of variance, indicating basic agreement of the first two moments of the model forecasts and the data.

Table 4. Forecast Statistics

	Series				
	January	March	April	July	October
RMSE	5.29	12.27	8.72	14.99	7.71
Proportion of MSE attributed to:					
Bias	.06	.07	.08	.01	.02
Correlation	.24	.08	.00	.42	.10
Correlation	.70	.85	.92	.57	.88

Conclusion

Given the unique form of the data, i.e., revisions in market forecasts, the very existence of the multivariate time-series model implies that the market for soybeans and related products is not efficient in Fama's sense. An efficient market necessarily incorporates all available information in the futures market necessarily incorporates all available information in the futures prices, arbitraging away the cross-autocorrelations underlying the time-series model. The significant coefficients in Table 2 contradict this, however, indicating that there is exploitable information in the past values of the series--the revisions are not serially uncorrelated white noise, but rather are, at least partially, predictable given past market revisions.

In addition to coefficient significance, the forecast values for the market revisions can be tested against zero. Nonzero market revisions are sufficient but not necessary to contradict market efficiency, implying that not only that there is information in past values (the significant coefficients), but also that it is important enough to produce mean forecast revisions a considerable distance from the efficient markets null hypothesis of zero. Table 5 shows that mean revisions in March and April, with test statistics of -4.79 and 2.64, respectively, are significantly different from zero.

The two sets of tests above rely on the assumption that the multivariate time-series representation of the stochastic process generating the market forecast revisions is correct. While the tests of model

Table 5. Tests of Revisions Against Zero (F = Forecast, A = Actual)

	January	March	April	July	October
$\bar{Z}_F$	0.94	-4.62	2.23	.19	-1.45
s_F	5.46	9.98	8.64	6.50	10.74
$\bar{Z}_F \sqrt{105}/s_F$	-1.76	-4.79	2.64	.30	-1.38
$\bar{Z}_A$	.37	-1.26	4.63	2.02	-.39
s_A	8.04	6.44	9.20	16.19	13.13
$\bar{Z}_A \sqrt{105}/s_A$	.47	-2.00	5.16	1.28	-.30

validity support this assumption, it is also possible to form a model-free test of market efficiency by testing the mean actual market revision against zero--analogous to the second test above but on the actual market revisions, rather than the model forecast market revisions. Again, nonzero mean revisions are sufficient but not necessary to reject the null hypothesis of market efficiency. The second section of Table 5 reports the results, with significantly nonzero mean revisions in March and April, in agreement with the same test based on the model forecasts.

The design of the data series in this research was intended to produce statistically powerful tests of market efficiency. While all three categories of tests rejected efficiency as a description of the markets for soybeans and products, it should be noted that the assumption of risk neutrality is crucial--the investor would truly need an iron constitution to reap the expected profits.

Footnotes

[1]Samuelson, p. 41: "In competitive markets there is a buyer for every seller. If one could be sure that a price will rise, it would have already risen."

[2]See Grossman and Stiglitz.

[3]The multivariate time series form may be considered to be a direct estimate of the final form of a complicated but unspecified structural model.

References

Fama, E. F. "Efficient Capital Markets: A Review of Theory and Empirical Work." Journal of Finance, Papers and Proceedings, (1970):383-417.

Grossman, S. and J. Stiglitz. "Information and Competitive Price Systems." American Economic Review, Papers and Proceedings, (1976):246-253.

Maravall, Agustin. "A Note on Identification on Multivariate Time-Series Models." Journal of Econometrics 16(1981):237-247.

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

Theil, H. Applied Economic Forecasting. Chicago, IL: Rand McNally & Co., 1966.

Tiao, G. C., G. E. P. Box, M. R. Gripe, G. B. Hudak, W. R. Bell, and I. Chang. "The Wisconsin Multiple Time Series Program: A Preliminary Guide." Unpublished, 1981.