Anne E. Peck*

ESTIMATION OF HEDGING AND SPECULATIVE POSITIONS IN FUTURES MARKETS REVISITED†

With the spectacular growth in trading in futures contracts in the United States in the past two decades has emerged a concomitant, if not commensurate, growth in research interest. Since 1960, total volume of trading on all domestic futures markets has increased nearly twenty-five fold. Traditional (mainly agricultural) futures markets came out of the doldrums of a surplus era into the storms of export demand, new markets emerged in the livestock segment, currencies and precious metals were released from official restraints and allowed to trade, and a host of new futures markets in financial instruments were successfully launched.

Research into the performance of these markets can proceed from the solid foundation of earlier analyses of the agricultural markets (notably those of Holbrook Working, 1977), and is more favored by an available data base than much economic research. Trading volume, open interest, and prices on these markets are published on a daily basis, by delivery month, and for some markets a very long historical record of such data is available. Much of this information is conveniently accessible on computer tapes covering the past 20 years.

One continuing data series, published by the Commodity Futures Trading Commission (CFTC), in continuation of a series published by its predecessor agency, the Commodity Exchange Authority (CEA), is potentially a veritable gold mine of information. It never quite lives up to its potential, however, because of certain limitations. The major limitation—which this paper may help to overcome—is that what it purports to show, viz., the classification of traders' positions as hedging, speculation, or spreading, is shown for only *large* traders, arbitrarily defined for each market. The "large" trader component, as a proportion of the total, varies among markets and over time, and the

* Associate Professor, Food Research Institute.

† The support of the Graduate School of Business Administration, Harvard University, is gratefully acknowledged as are the helpful comments of Roger Gray, William Jones, and Todd Petzel.

Food Research Institute Studies, Vol. XVIII, No. 2, 1982

temptation is great to estimate the breakdown of the total from the known breakdown of the partial.

A possible secondary limitation is inaccuracies in the reported data. In addition to the usual sources of reporting error, these data may also be subject to error either because of misunderstanding of the measuring of the classification or, more important, because of certain incentives to misreporting. Limits apply to speculative positions in most markets, providing incentives to report these as hedging if they exceed the limits.[1] Tax treatment also differs between hedging and speculative positions and may provide positive incentives to misreport in either direction to conform with misrepresentation for tax purposes. It is doubtful however whether generalized misreporting vitiates the overall usefulness of the reporting traders' classifications. As a partial correction, the level of the reporting and speculative limits are included in the following analyses.

The classification-of-positions data have been used to examine questions as diverse as seasonal patterns of hedging use (Peck, 1979–80), the hedging pressure hypothesis (Gray, 1979; Cootner, 1967; Telser, 1967), profit flows among the participants (Houthakker, 1960; Rockwell, 1967), and changes in speculative indices over time (Working, 1960; Nathan, 1967; Peck, 1981). In the first two types of analyses, the authors implicitly assume that the distribution of total positions is similar to that reported for large traders. In the profit-flow studies, the authors calculate profits to the nonreporting, small traders separately. In discussing the results, however, they assume that the nonreporting traders are primarily speculators. Calculation of speculative indices requires knowledge of the complete distribution of positions, and these studies have used formal allocation schemes to distribute the positions of the small traders. Two such schemes have been proposed by Arnold Larson (1961) and David Rutledge (1978). Both rely on the relatively infrequent occasions when the total distribution of positions is known, and both use data available prior to 1960. However, market compositions have changed markedly since 1960, and neither scheme can be used satisfactorily with current market data. This paper compares various previously used allocations and proposes an alternative method that would make the reporting data more useful. As will be seen, none of the solutions completely solves the problems with these data.

THE PROBLEM

Data from the corn futures market illustrate the difficulties of using and interpreting the commitments-of-traders data. At the end of January 1970, the CEA reported the commitments of corn traders on the Chicago Board of Trade, expressed as percentages of the total open interest, as follows:[2]

[1] However, if a trader reports a position as a hedge, then complementary forms must be submitted describing the cash positions which are being hedged.

[2] Formally, reported spreading includes positions which are taken in the same commodity, whether or not they are in the same market. For example, reported spreading in the corn market would include spreads between the Chicago Board of Trade and the Mid-America Exchange contracts. As the data here are market-specific, all unmatched spreading on a market has been added to reported speculation (long or short, as appropriate), and the remainder is labeled "matching."

	Reporting traders			Small trader positions
	Speculation	Hedging	Matching	
Long	2.8	46.0	17.0	34.2
Short	5.6	54.6	17.0	22.8

By the end of January 1976, traders' commitments were:

	Reporting traders			Small trader positions
	Speculation	Hedging	Matching	
Long	3.0	74.3	7.3	15.4
Short	8.3	61.0	7.3	23.4

A common assumption among users of these data is that most small traders are speculators (e.g. Rockwell, 1967). A convenient extension of this assumption is that all small traders are speculators, that is, that hedging and matching are fully reported. Applying this assumption gives the distributions shown in Section I of Table 1. Long speculation appears to have declined markedly over the six-year period, a direct reflection of the relative decline in the long positions of the small traders. Though extreme, this assumption is often useful because it provides an estimate of the maximum speculative distribution of traders. Speculation was not more than 37 percent of the long or 28 percent of the short open interest in the January 1970 corn market. Unfortunately, this assumption does not also provide an upper bound to changes in speculation. However, it remains a benchmark against which to consider other alternatives.

A second approach is to assume that hedging, matching and speculative positions are equally fully reported. Small traders' unreported positions are thus distributed proportionally to the large traders' reported positions. In 1970, matching positions averaged 23.9 percent of all reported positions. Nonreported long and short positions averaged 28.5 percent of the open interest. If matching positions made up the same share of nonreported as of reported positions, they would have comprised 6.8 percent (.238 × .285) of the open interest. Long hedging averaged 94.3 percent of the nonmatching reported long positions and nonreported long hedging would be 25.8 percent (.943 × .274) of the open interest. The residual, nonreported long speculation would be 1.6 percent. The resulting total distribution is shown in Section II of Table 1.

According to this estimate, hedging positions in 1976 were at least 80 percent of the open interest on both sides of the market, and matching positions had declined by more than 60 percent. Because hedging is the largest of the reported categories, most of the nonreported positions are assigned to hedging, and the resulting allocation is nearly the opposite extreme of the previous one. This will not always be true, however, since hedging is not always the largest of the reported positions.

A compromise assumption would recognize that, while not all of the small positions are speculative, neither are positions equally fully reported. Matching positions are allocated as in the second method. Then, the excess of long

TABLE 1.—ALTERNATIVE DISTRIBUTIONS OF THE OPEN TOTAL INTEREST IN THE CORN FUTURES MARKET ON TWO REPORTING DATES
(Percentage of open interest)

Assumption date	Long positions			Short positions	
	Speculative	Hedging	Matching	Speculative	Hedging
I. All small trader positions are speculative					
1/70	37.0	46.0	17.0	28.4	54.6
1/76	18.4	74.3	7.3	31.5	61.0
II. All positions are equally fully reported					
1/70	4.4	71.8	23.8	7.1	69.1
1/76	3.5	87.3	9.1	10.9	80.0
III. Net small trader position is speculative, remainder is proportional to the ratio of long to short large trader positions					
1/70	18.0	58.2	23.8	7.1	69.1
1/76	3.5	87.0	9.1	19.1	71.4
IV. Larson's method					
1/70	17.1	46.2	36.7	6.5	56.8
1/76	25.7	50.7	23.6	12.5	63.9
V. Rutledge's method					
1/70	26.7	54.0	19.3	12.9	67.4
1/76	0.6	83.4	16.0	14.9	68.9

Source: Total distributions calculated as described using the positions reports in the text. Long (short) positions, speculation and hedging, plus matching equal 100 percent of the open interest.

(short) positions over short (long) positions among the nonreported positions is assumed to be long (short) speculation. The residual, now balanced, small trader positions are then allocated between hedging and speculation according to their representation in the known distribution.[3] This leads to the distribution shown in Section III.

Other assumptions could be considered, but these demonstrate the problem: the conclusions from analyses that use these data are likely to be highly sensitive to the assumption made about the distribution of nonreported positions. Without additional information about the actual distribution of positions, it is difficult to evaluate which of the possible assumptions is most reasonable. Occasionally, additional data are available in the form of full-market surveys of the open interest. These surveys are rare, with only 36 having been undertaken in more than three decades since World War II. An additional difficulty with these data is that the surveys were normally conducted as part of broader

[3] Specifically, the ratios of long to short hedging and speculation are assumed to be similar among the remaining small positions and the reallocated large positions, both of which are now balanced long and short. Working's analysis (1960) suggests that constant ratios may be a more reasonable assumption than constant percentage distributions.

investigations of market activity during periods of "unusual" price behavior. However, they do provide a means to compare the allocations from the differing assumptions.

In 1961, Larson proposed that the data from these surveys could be used directly to estimate the relationships between the complete distribution of positions and that available in the reporting data. Larson had available 26 observations on complete and reported distributions. In 1979, Rutledge proposed a revised estimation procedure and contrasted it with Larson's for nine survey dates which were not included in the estimates of either author's model.

Larson's model consists of three equations: a linear relationship between total matching and reported matching positions; a semilog relationshp between the ratio of total to reported long hedging and all nonreported long positions; and a semilog relationship between the ratio of total to reported short hedging and all nonreported short positions.[4] The 26 observations from the full-market surveys combined with the reported positions on the same dates are used to estimate the coefficients of the above equations. These estimated equations are then used to allocate positions when only the regular reports are available. The application of Larson's equations to the corn data results in the positions shown in Section IV of Table 1. In contrast to earlier results, total speculation, both long and short, shows a significant increase over the six-year period, and total hedging remains nearly constant. Note, however, that Larson's method estimates total long hedging to be only 50.7 percent of the open interest in 1976, while reported long hedging was 74.3 percent of the open interest. The difficulty is that Larson's estimation techniques did not constrain the fitted values to fall within reasonable bounds. For example, any time the nonreported long positions are less than 32 percent of the open interest, Larson's model will estimate total long hedging to be less than reported long hedging.

Rutledge uses the more formal allocation techniques to estimate relationships between the complete surveys and the reporting positions. In addition, he focuses directly upon the nonreported positions rather than the total positions. These are allocated into hedging, speculative, and matching positions and then added to their respective reported positions. His procedure estimates four equations, two for the long and two for the short positions. The third position on each side may be solved for from an identity, and the technique assures that this choice is irrelevant.

Application of Rutledge's technique to the corn data results in the total positions shown in Section V of Table 1. As is clear with the 1976 data, this technique also suffers a forecasting problem. A nonreported speculative long position of −33.3 percent was initially estimated. The second problem with this technique is more fundamental. The technique allocates long and short positions separately and leads to different estimates of the matching positions in each. In the 1976 data, the estimated nonreported matching positions from the long side were 39.6 percent of the open interest (larger than the total nonreported long position) and only 8.7 percent when estimated from the

[4] Larson estimated separately the relationships to explain speculative positions. These did not fit as well as the hedging equations. Since only one set is required (the omitted ones are derived from market identities) discussion here uses only the hedging equations.

short side. In the figures reported in Table 1, the smaller figure is taken as the estimate of nonreported matching positions and the difference is added to speculation giving an estimated total long speculation of 0.6 percent (3.0 − 33.3 + 30.9) of the open interest.

Since both Larson's and Rutledge's approaches have the advantage of employing the available data to determine the relationships among total and reported positions, an alternative is proposed below which constrains the forecasts to permissible values. Then all the methods are compared as to their allocative power.

AN ALTERNATIVE MODEL

The solution to the Larson-Rutledge problems is to constrain the estimating equations to provide forecasts that fall within the desired intervals, that is, to employ logit- or probit-estimating techniques. In Larson's short hedging equation, the variable to be explained is (the log of) the ratio of the total short hedging (THS) to reported short hedging (HS), where all variables throughout these derivations are taken relative to the total open interest. Assume temporarily that the allocation of the matching positions has already occurred. Total short hedging (THS) must be at least as great as reported short hedging (HS) but no larger than the open interest remaining to be allocated: HS $\leqslant$ THS $\leqslant$ 1 − SS − TM. SS is reported short speculation, and TM is total matching, equal to reported matching (M) plus nonreported matching (NRM). All variables are taken relative to the total open interest. Thus, the ratio of total to reported short hedging is constrained to be:

$$1 \leqslant \frac{THS}{HS} \leqslant \frac{1 - SS - M - NRM}{HS} .$$

With the identity, 1 ≡ HS + SS + M + NRS, where NRS are the nonreported short positions, the above inequality can be manipulated to give:

$$1 \leqslant \frac{THS}{HS} \leqslant 1 + \frac{NRS - NRM}{HS} \qquad \text{or} \qquad 0 \leqslant \frac{NRHS}{XNRS} \leqslant 1.$$

The ratio of total to reported short hedging must be at least equal to 1, as total hedging cannot be less than reported hedging. At the other extreme, the hedging ratio cannot be larger than 1 plus the nonreported positions remaining to be allocated relative to short hedging. Alternatively, nonreported short hedging (NRHS) as a percent of the nonreported positions remaining to be allocated (XNRS = NRS − NRM) must fall between 0 and 1. Thus, when appropriately constrained, Larson's approach, which focuses on total relative to reported positions, is equivalent to Rutledge's, which focuses on the nonreported positions.

The remaining difficulty is in allocating the matching positions. Rutledge's procedure leads to contradictions—matching positions derived from the nonreported long positions differ significantly from those derived from the short positions. A two-step procedure, as in Larson's method, is used to circumvent this problem. Matching positions are allocated first, and the residual is then allocated between hedging and speculation. Each step may be viewed as a probability model, and standard methods may be applied. Empirical work with both logit and probit models showed the former to be more satisfactory and only estimates from the logit models will be reported here.

Specifically, the relationships to be estimated are:

$$\ln \frac{\text{NRM}}{\text{NR} - \text{NRM}} = \alpha_0 + \alpha_1\text{DS} + \alpha_2\text{M} + \alpha_3\text{NRL} + \alpha_4\text{NRS} + \alpha_5\text{LIM} + \varepsilon_1 \quad (1)$$

$$\ln \frac{\text{NRSL}}{\text{NRHL}} = \gamma_0 + \gamma_1\text{DS} + \gamma_2\text{XNRL} + \gamma_3\text{SL} + \gamma_4\text{HL} + \gamma_5\text{LIM} + \varepsilon_3 \quad (2)$$

$$\ln \frac{\text{NRSS}}{\text{NRHS}} = \beta_0 + \beta_1\text{DS} + \beta_2\text{XNRS} + \beta_3\text{SS} + \beta_4\text{HS} + \beta_5\text{LIM} + \varepsilon_2 \quad (3)$$

All variables are expressed relative to the total open interest.

The dependent variables are defined above except for NR in Equation (1), which is the smaller of the two nonreporting categories, NRS and NRL.[5] The first equation apportions the nonreported positions between matching and nonmatching positions. The second and third equations apportion the nonmatching, nonreported long and short positions between speculation and hedging. Note that these specific dependent variables are derived from the required ratios:

$$\ln \left[\frac{\text{NRM/NR}}{\text{NRNM/NR}}\right], \quad \ln \left[\frac{\text{NRSS/XNRS}}{1 - \dfrac{\text{NRSS}}{\text{XNRS}}}\right] \quad \text{and} \quad \ln \left[\frac{\text{NRSL/XNRL}}{\dfrac{\text{NRSL}}{\text{XNRL}}}\right],$$

where NRNM is nonreported, nonmatching small trader positions.

In addition to the variable being allocated, the independent variables include the relevant reported positions, taken from the regular reports coinciding with the full-market surveys. Two additional variables have been included, DS and LIM. DS is a binary variable which is 0 for storable commodities and 1 for perishable commodities. Larson's work found the relationship between total and reported matching to be significantly different as between the egg, onion, and potato markets and the storable commodity markets.[6] LIM is a variable designed to reflect the varying regulatory conditions under which these data were collected. Both reporting levels and speculative limits differ among commodities and for one commodity with time. The reporting level is the minimum number of contracts an individual trader must own to become a reporting trader and hence enter the "known" distribution of positions. All markets included in the analysis were subject to reporting limits. Thus, as a reflection of basic differences in regulatory constraints among these markets, the reporting level is the most obvious choice. The speculative limit is the maximum number of contracts an individual speculator may own. If speculative limits are low, then there are strong incentives to classify trades as hedges and supply the supplemental documentation to the CFTC which justifies the hedging classification. If limits are high, then the classification would not affect the size of individual traders' positions and the decision classifying the trade may be made on other grounds. Estimates were made using each variable. Those

[5] The smaller of the two is used to keep the forecast of nonreporting matching positions smaller than the total nonreporting positions.

[6] Rutledge found that commodity-specific binary variables were not significant.

Table 2.—Estimates of the Logit Model to Allocate the Nonreporting Positions

Dependent variable	Intercept	Perishable commodity shifter[a]	Total nonreported positions[b]: Long	Total nonreported positions[b]: Short	Total reported positions[c]: Matching	Total reported positions[c]: Speculative	Total reported positions[c]: Hedging	Reporting limit	Adjusted R^2
I. Logit model estimated with data from 35 full-market surveys									
$\text{Ln}\left[\frac{NRM}{NR-NRM}\right]$	– 2.7	– 1.1	3.1	– 1.1	6.9			– 45.2	.43
	(– 1.7)	(– 2.6)	(1.9)	(– 0.9)	(1.9)			(– 0.9)	
$\text{Ln}\left[\frac{NRSL}{NRHL}\right]$	3.9	0.6	– 1.8			– 3.2	– 9.1	– 30.7	.50
	(3.7)	(1.3)	(– 1.0)			(– 1.4)	(– 4.8)	(– 0.7)	
$\text{Ln}\left[\frac{NRSS}{NRHS}\right]$	4.5	0.6		– 5.2		– 1.3)	– 7.1	11.8	.62
	(4.9)	(1.7)		(– 3.6)		(– 0.5)	(– 4.8)	(0.4)	
II. Logit model estimated with subset of 20 surveys for commodities still actively traded									
$\text{Ln}\left[\frac{NRM}{NR-NRH}\right]$	– 1.6	– 1.4	4.4	1.1	– 2.2			– 498.5	.81
	(1.2)	(– 3.7)	(2.9)	(1.1)	(– 0.7)			(4.8)	
$\text{Ln}\left[\frac{NRSL}{RHL}\right]$	5.5	0.4	– 5.4			– 5.8	– 9.3	139.0	.72
	(6.9)	(1.2)	(– 3.3)			(– 2.7)	(– 7.1)	(1.8)	
$\text{Ln}\left[\frac{NRSS}{NRHS}\right]$	6.0	1.1		– 9.5		– 6.5	– 8.8	121.8	.63
	(4.2)	(1.8)		(– 3.3)		(– 1.2)	(– 4.0)	(1.0)	

Source: Based on data from full-market surveys and the corresponding regular positions reports. Tables of the data and their sources are available from the author. T-statistics appear in parenthesis. See text for a description of the logit model, its application, and definitions of the dependent variables.

[a]A binary variable which is 1 if the commodity is perishable (onions, potatoes, eggs, live cattle) and zero elsewhere.

[b]In Equations (2) and (3), the nonreported positions which have been estimated as matching in Equation (1) are subtracted from the total nonreported long or short positions.

[c]In Equation (2), the long reported speculative and hedging positions are used. In Equation (3) the short positions are used.

using the reporting limit were consistently more satisfactory and only these are reported.

With the comparative luxury of 36 full-market surveys, attention was given to their representativeness. On this ground, the wheat survey of February 1947 was omitted; it covered only the Chicago market, whereas the regular reports were for positions in all wheat futures markets. In addition, the data set included observations from the wool, wool tops, refrigerator eggs, and onions markets—markets which no longer exist or are very inactive. Deleting those, a sample of 20 surveys remained, including the corn, wheat, soybeans, cotton, potatoes, fresh eggs, live cattle, pork bellies, and orange juice markets. Estimates from both samples (35 and 20) are presented.

THE RESULTS

Estimates of the logit models developed above are presented in Table 2. The logit model fitted to the subsample of only 20 surveys appears to perform best. The R^2's, adjusted for degrees of freedom, are markedly higher in two of the three equations and marginally higher in the remaining one. However, a test for significant differences between the estimated equations produced no significant F-statistics. With few exceptions, the results are consistent between the two models and with expectations. In the matching equations, the perishability variable is significant; there is significantly less nonreported spreading in the perishable commodities, all other factors held constant. Spreading in these markets is apparently done mainly by large, professional traders. Spread relationships in these markets are not well understood and fluctuate more widely than in storable commodity markets. Of the two nonreported position variables (long and short), only the long positions are significant. Nonreported long positions tended to be larger and hence more likely to shift the relationship between reported and nonreported matching positions. The reported positions were significant and positive, as expected, in the first results. However, in the smaller sample, significance disappears and the sign changes. Finally, the reporting limit is negatively associated with the amount of nonreported positions that are matching. Markets with larger relative reporting limits had smaller amounts of nonreported matching. The variable is significant only in the results from the smaller sample. Ten of these 20 surveys were taken in the 1960s and were not included in the previous estimates, suggesting that regulatory differences between markets have only become significant in more recent periods.

Comparison of the two models for Equations (2) and (3) reveals close agreement in signs and significance of the explanatory variables. Relatively, nonreported speculation (both long and short) increases in markets for perishable commodities. It is inversely related to the size of the nonreported positions remaining to be allocated and to both reported speculation and reported hedging. Coefficients for the reporting limit are again insignificant in the first estimates. The limit is significant, but positively associated with the amount of nonreported long speculation in the second estimates. That is, markets that have more restrictive regulatory constraints (smaller limits relative to market size) tend to have less nonreported long speculation and more long hedging, other allocations held constant, contrary to expectations.

However, this expectation is derived from experience in a single market that is growing over time. Market growth is normally associated with increased commercial use and hence with relative increases in hedging positions. Hedging would tend to become more fully reported, and thus all the nonreporting positions could be expected to be speculative. Evidently, the variation in relative reporting limits among markets was not similar to that expected over time and most of the expected variation over time is being captured by the relative sizes of the nonreported positions, which are inversely related to relative speculation.

MODEL PERFORMANCE

Since the estimated relationships from the two samples are not significantly different, a choice between them cannot be made simply on the basis of their adjusted $\bar{R}^2$. Further, their allocative performance needs to be evaluated in the context of the alternatives, both formal models and informal assumptions.

The logit model must first be solved for the actual market allocations. The reported positions, commodity perishability and reporting limits for the market are assumed to be known. Let $\hat{Y}_1$ be the fitted value from the nonreported matching allocation [Equation (1)]. Then,

$$\text{Exp}(\hat{Y}_1) = \frac{\text{NRM/NR}}{\text{NRNM/NR}}$$

which gives

$$\text{NRM} = \left[\frac{\text{Exp}(\hat{Y}_1)}{1 + \text{Exp}(\hat{Y}_1)}\right] \text{NR},$$

where NR is the smaller of the long and short nonreported open interests (which prevents the estimated nonreported matching positions from exceeding the total nonreported positions).

The remaining nonreported positions (XNRL and XNRS) are allocated as follows:

$$\text{NRSL} = \left[\frac{\text{Exp}(\hat{Y}_2)}{1 + \text{Exp}(\hat{Y}_2)}\right] \text{XNRL} \quad \text{and} \quad \text{NRSS} = \left[\frac{\text{Exp}(\hat{Y}_3)}{1 + \text{Exp}(\hat{Y}_3)}\right] \text{XNRS}$$

where $\hat{Y}_2$ and $\hat{Y}_3$ are the fitted values from Equations (2) and (3), respectively. Finally, the hedging positions are derived as residuals from the identities.

Comparisons can now be made with the allocations derived from alternative methods. The first comparisons focus on the 35 observations for which the actual distributions are known. Using the various assumed distributions and formal models, the reporting data are converted to complete market descriptions. These are compared to the actual distribution via the squared correlation coefficient and reported in Table 3.

Taken as a group, the three naive allocations provide the worst explanations of the actual distributions. The most interesting result here is the relative ease

TABLE 3.—GOODNESS OF FIT (R^2) OF ALTERNATIVE ALLOCATION SCHEMES[a]

Method	Long positions			Short positions	
Item	Speculative	Hedging	Matching	Speculative	Hedging
I. All small trader positions are speculative					
Nonreported	.5382	0	0	.5794	0
Total	.6654	.6517	.6928	.6929	.7775
II. All positions are equally fully reported					
Nonreported	.5098	.3892	.4178	.5496	.6188
Total	.7290	.7937	.7945	.7726	.8888
III. Net small trader position is speculative, remainder is proportional to the ratio of long to short large trader positions					
Nonreported	.2178	.1275	.4178	.5655	.3616
Total	.6647	.6598	.7945	.8452	.8795
IV. Larson's method					
Nonreported	.7302	.4458	.2045	.7554	.5111
Total	.8225	.7873	.6760	.8833	.9345
V. Rutledge's method					
Nonreported	.7262	.2990	.2504	.8133	.4581
Total	.8161	.7246	.7002	.8869	.9075
VI. Logit model fitted over 35 observations					
Nonreported	.7529	.4762	.5610	.7833	.6576
Total	.8475	.8233	.8258	.8852	.9259
VII. Logit model fitted over 20 observations					
Nonreported	.6861	.6620	.7409	.7907	.6983
Total	.8285	.8597	.8945	.8941	.9390

[a]Comparison of assumed (or fitted) distribution of positions with actual distribution for the 35 instances when the latter is known. The first item—nonreported—compares the nonreported allocation with its actual allocation. The second item—total—reports the fits for the total allocation.

in allocating short hedging. When all the nonreported positions are assumed to be speculative, reported short hedging is the total short hedging, and it explains 78 percent of the variation in actual total short hedging. Each of the alternative naive allocations improves this percentage somewhat. Short hedging is mainly inventory hedging, so that individual positions are determined by the size of stocks, and these tend to be large and thus reportable.

The Larson and Rutledge models provide generally improved fits, with the exception of the matching allocations. These should not be taken too seriously, however, since the calculation of total matching positions used here was slightly different from that employed by Larson and Rutledge. Their models were not reestimated with the altered matching data and hence are not expected to "fit" particularly well. The redefinition of matching also influences the hedging variable; thus the reported fits may understate the models' relative performances here as well. A more just comparison would have been to reestimate both models with the present data set. Reestimation would not solve the forecasting difficulties, however, and ultimately would not make them more useful.

Finally, the two logit models provide only marginally better explanations of the actual distributions, particularly if allowance is made for the data difficulties in the Larson and Rutledge evaluations. Interestingly, the second logit model which was fitted to the smaller sample of data performs better than the first logit model over the entire range of the 35 observations. This seeming contradiction is a result of the nonlinear nature of the estimated models and the linearity of the comparisons in Table 3. In any event, the differences are only slight, certainly not enough to permit a choice.

A second more relevant measure of model performance is evaluation of the predicted allocations in situations where the reported data differ markedly from those used in the estimations. The data from the corn market used earlier provide a convenient illustration. The logit models' allocations are as follows:

	TSL	THL	TM	TSS	THS
Corn 1/70					
I	10.9	64.9	24.1	12.4	63.5
II	14.0	60.6	24.4	11.6	64.0
Corn 1/76					
I	3.5	87.4	9.1	13.8	77.1
II	4.1	84.6	11.3	11.5	77.2

when I and II distinguish between estimates using 35 and 20 observations, respectively. These estimates are quite similar and appear to fall near the middle of the estimates presented in Table 1. For example, estimates of total long speculation in 1970 in Table 1 ranged from 4.4 percent to 37.0 percent, as compared with estimates of 10.9 or 14.0 percent from the logit models.

When the logit models are applied to reporting data from numerous other markets and dates, the estimates are generally quite similar. When hedging dominates the reported positions (e.g., Kansas City wheat) then most of the nonreported positions are allocated to hedging. When speculation dominates the reported data (e.g., pork bellies), then much of the nonreported open interest is allocated to speculation. While not identical, the estimates from the two logit models are similar.

A significant difference in allocations appears only when the models are applied to data from the silver market, where reported spreading is orders of magnitude larger than the other reporting categories.[7] For example, at the end of August 1979 reported spreading was 50.8 percent of the total open interest on the Chicago market. Long speculation and hedging were only 3.5 and 2.4 percent, respectively. The first logit model allocated the remaining, nonreported long open interest as follows: 33.0 percent to matching, 10.0 percent to speculation, and 0.3 percent to hedging. The second logit model allocated

[7] Reported positions in the silver market are not part of the regular Commodity Futures Trading Commission market reports. They are collected, however, and data from selected dates were reported in the Senate Hearings (1980) on the silver market.

15.1 percent to matching, 27.4 percent to speculation, and 0.8 percent to hedging. The difference in the estimates reflects the changed signs of the reported matching variable in the two estimates of the matching equation. Since a reversal of the reported and nonreported matching categories is not expected, the estimates from the first model (using all 35 observations) are clearly preferable.

In sum, all of the estimated models are significant improvements over the simple assumptions. The logit models are preferred over the other estimated models because they constrain (by construction) the distributions to be within acceptable ranges. And, as between the two logit estimates, the one using all the data was preferred in the only instance where the forecasted distributions were significantly different. However, in all cases, the estimates really reflect the underlying, reported positions, whether or not these imply sensible total distributions.

IMPLICATIONS

The regular reports of positions of traders in commodity futures markets form an established data series that is used frequently in market behavior research. Its length, dating back to the 1920s in the grains, is unusual in economic statistics. Its information, the positions of large traders—commercial firms and speculators—has been used in a variety of interesting research on the composition of markets and on returns to trading groups.

In spite of their usefulness to date, these data suffer from a major shortcoming in that they describe only a varying fraction of traders' positions. The reporting levels relative to the size of individual's positions within a market determine how complete the description is. To complete the information about traders, the researcher must resort to some allocation of the nonreporting, small traders' position. The analysis above has contrasted several alternatives—both informal assumptions and formal models—and investigated a model consistent in a forecasting sense.

Comparative evaluations showed none of the alternatives to be particularly exciting. The three informal assumptions that have been examined do not explain well the small traders' positions. The formal models improve the allocations measurably, though they rely heavily on the reporting positions to impute a distribution to the unknown. If reported hedging is a relatively large fraction of the open interest, then a relatively large fraction of the nonreported positions will be estimated to be hedging positions as well.

The formal models are more acceptable statistically, but they are really not much different from the simple, often-used assumptions about the relationships between reported and nonreported positions, relationships that seem tenuous when applied to current market data. For example, total open interest in the corn market has grown from 403 million bushels in 1971 to 1,315 million bushels in 1980. Over this period, individual speculative positions remained limited to a maximum of 3 million bushels and total hedging use (measured by reported hedging) grew much more rapidly than the open interest. In addition, reporting levels have changed twice, from 200,000 to 500,000 bushels in August 1976, and to 1,000,000 bushels in March 1981. Under these conditions, it seems reasonable to believe that, while in 1970 some hedging may not have been reported as such or been too small to be reportable, virtually all hedging was reportable for most of the decade. With the recent change, this may no longer be true. Thus, one suspects that virtually all of the "small traders" in the corn market were speculators for most of the 1970s, but that this was not true either in 1970 or after March 1981. None of the above procedures would show this change. With the exception of the all-speculative assumption, they would show a *larger* fraction of the nonreported positions to be hedging during most of the decade than at its beginning or, again, in the 1981 data.

All of this underlies the pressing need for more data. The most recent complete market survey of a grain market was that conducted of the wheat market in 1967. Markets have changed significantly in the composition of the reportable positions and, undoubtedly, in the composition of the nonreported positions as well. A temporary solution is always to use at least two alternative allocations. A base is provided by the all-speculative assumption. Speculation cannot be larger than the total of reported speculation and all nonreporting positions. The allocations from the logit model (using all 35 observations) estimated here provides a second alternative. These two allocations provide reasonable expectation of spanning the "real" distribution. All analyses should be done using both distributions to investigate the sensitivity of the research results to the position allocations.

A more far-reaching proposal is to have the regulatory authorities provide the complete distributions regularly. Continuous full-market surveys, interviewing all participants on designated dates, are not the solution. Their tremendous detail, although interesting, implies unacceptably high costs. Since large traders report directly to the CFTC, the difficulty is in distinguishing commercial from speculative traders among small traders. However, small trader hedging is identified for a variety of other purposes. Hedging positions are entitled to smaller margins, and brokerage firms routinely require letters of intent from these customers. In addition, profits from properly identified hedging positions are entitled to differential tax treatment. Thus, complete information may not be impossibly difficult to collect on a regular basis, at least in summary form, directly from the brokerage firms. The advent of a National Futures Association may make this proposal more realistic since many practices will become standardized within the industry.

CITATIONS

Paul Cootner (1967), "Speculation and Hedging," *Food Research Institute Studies*, Supplement to Vol. 7, pp. 65–106.

Roger Gray (1979), "The Emergence of Short Speculation," *International Research Seminar Proceedings*, Vol. 6, Chicago Board of Trade.

Hendrik Houthakker (1957), "Can Speculators Forecast Prices?" *Review of Economics and Statistics*, May.

Arnold Larson (1960), "The Estimation of Hedging and Speculative Positions," *Food Research Institute Studies*, Vol. 1, pp. 313–24.

Robert A. Nathan Associates (1967), *Margins, Speculation and Prices in Grain Futures Markets*, U.S. Department of Agriculture, ERS Report.

Anne E. Peck (1979–80), "The Influence of Hedging on Futures Market Activity," *Food Research Institute Studies,* Vol. 17, No. 3.

________ (1981), "Measures and Price Effects of Speculation in Wheat, Corn and Soybean Futures Markets," *Seminar on Speculation, Proceedings,* Chicago Board of Trade.

Charles S. Rockwell (1967), "Normal Backwardation, Forecasting, and the Returns to Commodity Futures Traders," *Food Research Institute Studies,* Supplement to Vol. 7, pp. 107–30.

David Rutledge (1977–78), "Estimation of Hedging and Speculative Positions: An Alternative Approach," *Food Research Institute Studies,* Vol. 16, No. 3.

Lester Telser (1967), "The Supply of Speculative Services in Wheat, Corn, and Soybeans," *Food Research Institute Studies,* Supplement to Vol. 7, pp. 131–76.

United States Senate (1980), 96th Congress, 2d Session, *Price Volatility on the Silver Futures Market,* Hearings, Committee on Agriculture, Nutrition, and Forestry, May 1 and 2.

Holbrook Working (1977), *Readings in Futures Markets, I: Selected Writings of Holbrook Working,* Chicago Board of Trade.

________ (1960), "Speculation on Hedging Markets," *Food Research Institute Studies,* Vol. 1, No. 2.