DAVID J. S. RUTLEDGE*

ESTIMATION OF HEDGING AND SPECULATIVE POSITIONS IN FUTURES MARKETS: AN ALTERNATIVE APPROACH

A topic of perennial interest to students of commodity futures markets is the behavior and motivation of different market participants. Information about speculative behavior is of considerable importance in studies of the price effects of speculation (*1*,*10*). Data on hedging participation in commodity markets are required for an assessment of the commercial value of these markets (*6*,*11*). Furthermore, data on traders' commitments have been employed in studies of the effects of margin requirements on commodity markets and are often used by technical price analysts (*3*, *9*).

To some extent any argument purporting to demonstrate the need for classification of the total open interest into speculation and hedging begs the question as to how these component parts are to be defined. Economists have increasingly turned to the view that speculation and hedging are not discrete concepts but rather that a continuum exists between them (*4*,*8*). Nevertheless, legislation providing for the regulation of commodity markets in the United States requires that these concepts be given operational definitions. Indeed it is this legislation which has given rise to the existing body of data on hedging and speculative trading commitments.[1]

Unfortunately, these published data are deficient in several respects, the most important being their incompleteness. The data distinguish between hedging and speculative positions for "large" traders only. Working (*12*) has proposed a procedure for dealing with this difficulty, and Larson (*2*) developed a different technique making use of a series of episodic full-market surveys.

The present paper describes yet another technique which appears to yield more accurate estimates of speculative and hedging positions than Larson's procedure, relying on theoretical statistical results not available at the time of Working's and Larson's contributions.

LARSON'S PROCEDURE

Before proceeding to a discussion of Larson's method the following notation is introduced, where each variable is expressed as a proportion of the total open interest:

RHL = reported long hedging
RHS = reported short hedging
RSL = reported long speculation
RSS = reported short speculation
RSP = reported spreading
NRL = nonreporting long positions
NRS = nonreporting short positions

Observations of the above variables are obtained from regular (monthly or semimonthly) reports issued by the regulating authority.

Further, the following variables are defined also as a proportion of the total open interest:

HL = total long hedging
HS = total short hedging
SL = total long speculation
SS = total short speculation
M   = total matching positions

Data on these variables are available from full market surveys conducted from time to time.

The following relationships among the above variables hold identically:

$$HL + SL = HS + SS = 1 - M \tag{1}$$

$$RHL + RSL + NRL = RHS + RSS + NRS = 1 - RSP \tag{2}$$

Larson's procedure for allocating the nonreporting positions (NRL and NRS) into hedging, speculative, and matching classes is based on estimated relationships between total and reported figures for hedging, speculative, and matching positions on those occasions when data on both sets of variables are available.

More precisely he estimates the following relationships:

$$\log (HL/RHL) = \lambda_0 + \lambda_1 \ NRL + \epsilon_1 \tag{3}$$

$$\log (HS/RHS) = \lambda_2 + \lambda_3 \ NRS + \epsilon_2 \tag{4}$$

$$\log (SL/RSL) \ = \mu_0 + \mu_1 \ NRL + \epsilon_3 \tag{5}$$

$$\log (SS/RSS) \ \ = \mu_2 + \mu_3 \ NRS + \epsilon_4 \tag{6}$$

$$M = \nu_0 + \nu_1 D + \nu_2 \ RSP + \epsilon_5 \tag{7}$$

where D is a dummy variable to be discussed below and the ϵ_i are random disturbances assumed to be such as to allow application of ordinary least squares (OLS) to Equations (3)-(7). The basic data set employed by Larson is taken from 26 full market surveys in nine commodities during the period 1946 to 1960.

Although Larson reported no standard errors or other summary statistics, he included several charts which show that the data fit the estimated equations tolerably well. Nevertheless Larson's procedure is not without its weaknesses, of which three warrant particular mention.

First, the procedure is highly ad hoc in nature. In particular the functional form was chosen simply on the basis of providing the best fit to the available data. In addition, in estimating Equation (7) Larson employed a dummy variable to allow for a shift in the intercept term but provided no rationale for its definition. There is no strong reason for supposing that new observations will conform to the same configuration as the sample data so that prediction (the eventual aim of the exercise) may be highly unreliable. Second, there is no guarantee that predicted values from Equations (3)-(7) will satisfy the internal consistency conditions corresponding to Equation (1); indeed this in general will *not* be the case. Failure to meet the consistency condition means that predictions obtained from these equations must be subjected to further manipulation before being used. Finally, use of OLS overlooks the possibility that there may be some degree of cross-correlation among the error terms of different equations.

AN ALTERNATIVE APPROACH

Rather than basing the statistical analysis solely on "data mining," it may be useful to recognize the problem at hand explicitly as an allocation problem. Data are available on nonreporting long positions and nonreporting short positions, each of which are to be *allocated* among the hedging, speculative, and matching classes. This formulation can take advantage of statistical methods designed specifically to deal with allocation problems. One such procedure has been considered by Powell (5) who discusses the estimation of the following system of "allocation equations:"

$$Y_i = X_0 \beta_{0i} + X \beta_i + Z_i \lambda + \mu_i \qquad (i = 1, 2, \ldots m) \qquad (8)$$

where $X_{0t} = \sum_{i=1}^{m} Y_{it}$ is an aggregate variable to be allocated among its m components Y_{it}. X is an observation matrix on k variables assumed to be common to every allocation equation and Z_i is an observation matrix on r variables assumed to be peculiar to the ith equation. The problem is to estimate the parameters of (8) subject to the "adding up" constraints:

$$\sum_{i=1}^{m} \beta_{0i} = 1 \sum_{i=1}^{m} \beta_{ji} = 0 \ (j = 1, 2, \ldots k)$$

Powell shows that the solution to this problem is obtained by omitting any one of the m equations (8) and estimating the remaining m-1 by Aitken's principle. The solution is invariant under the choice of equation to be omitted. In the special case where all right-hand variables appear in every equation (i.e., r=0), the solution is obtained by omitting any one equation and applying OLS to the remaining m-1 equations.

In the present context $NRHL(=HL-RHL)$, $NRHS(=HS-RHS)$, $NRSL(=SL-RSL)$, $NRSS(=SS-RSS)$, and $NRM(=M-RSP)$ are defined as the variables corresponding to the Y_i of Equation (8). For the following allocation equations, Equations (9) and (10) correspond to the allocation system (8) for nonreported long positions in which the "matching" equation has been omitted.

$$NRHL = \alpha_{00} + \alpha_{01}\, NRL + \alpha_{11}\, NRL^2 + \alpha_{21}\, RHL + \alpha_{31}\, RSL + \mu_1 \quad (9)$$

$$NRSL = \alpha_{10} + \alpha_{02}\, NRL + \alpha_{12}\, NRL^2 + \alpha_{22}\, RHL + \alpha_{32}\, RSL + \mu_2 \quad (10)$$

$$NRHS = \beta_{00} + \beta_{01}\, NRS + \beta_{11}\, NRS^2 + \beta_{21}\, RHS + \beta_{31}\, RSS + \nu_1 \quad (11)$$

$$NRSS = \beta_{10} + \beta_{02}\, NRS + \beta_{12}\, NRS^2 + \beta_{22}\, RHS + \beta_{32}\, RSS + \nu_2 \quad (12)$$

NRL is the aggregate variable to be allocated into its component parts NRHL and NRSL. As the set of explanatory variables is the same in both equations this system corresponds to the case $r=0$ described above. The choice of right-hand variables, other than NRL, requires justification. NRL^2 is included so as not to constrain the proportion of nonreported long positions classified as either hedging or speculation to be constant regardless of the size of that position. In the absence of any additional information it seems reasonable to assume that nonreported hedging and speculative positions respond to the same economic forces as do the reported quantitites. Consequently, as a first pass, nonreported long hedging might be expected to be larger when reported long hedging is larger and when reported long speculation is smaller. Hence RHL and RSL are included in Equations (9) and (10).

Several factors may be at work to require this simple view to be modified. The nonreported positions may not be homogeneous in the sense that they may comprise many very small positions and many speculative positions just below the reporting level. Furthermore, the data set on which these equations are to be estimated consists of a quite disparate collection of commodities, including wheat, corn, cotton, wool, eggs, and potatoes. Reporting requirements vary widely and somewhat arbitrarily from one commodity to another, and this may blur any reflection of reported positions in the nonreported data. All this suggests that the results of estimating Equations (9)-(12) should be interpreted with care. However, it is worth noting also that these difficulties are no less severe in estimating Equations (3)-(7), and that the allocation Equations (9)-(12) permit more information about hedging and speculative positions to be used than do Equations (3)-(7).

Equations (11) and (12) correspond to the allocation system (8) for nonreporting short positions in which the "matching" equation has been omitted. Here the shift variables are NRS^2, RHS, and RSS.

Equations (9)-(12) represent two allocation systems, one for NRL and the other for NRS. Powell's analysis establishes that OLS should be used on a single allocation system if no Z variables are present, as in this case. However, the possibility remains that errors in one allocation system may be correlated with those in the other. Accordingly Equations (9)-(12) are estimated using Zellner's "seemingly unrelated regression" technique (13) as well as by OLS.

COMPARISON OF PROCEDURES

In this section the procedure described in the previous section, labelled below as the "revised" procedure, is compared with Larson's procedure. First, Larson's equations are reestimated using very slight modifications to his data set. The estimated equations so obtained are reported in Table 1.[2] Second, the same data is used to estimate Equations (9)-(12) by OLS and the results are shown in Table 2. The results obtained using the seemingly unrelated regression method are so close to those reported in Table 2 as not to warrant publication of a separate table. The equations were initially estimated with dummy variables on the intercept term for each commodity. These variables taken together were not significant and have been omitted from the equations reported here. The coefficients in Table 2 appear to be sensible, although the signs of RSL in Equation (9) and RSS in Equation (10) may seem puzzling at first glance. To some extent these signs reflect the difficulties discussed in the preceding section. A further contributing factor is the large negative correlation between RHL and RSL and between RHS and RSS.

The R^2 figures in Table 1 should not be compared directly with those of Table 2 as the dependent variables are different. A more accurate notion of the relative performance of the two sets of equations in terms of their goodness-of-fit can be obtained from Table 3, which shows that except for long hedging, the revised procedure performs better than Larson's procedure.

Goodness-of-fit to sample data may not be the best criterion of model selection, especially if the goal of the exercise is prediction as in the present case. To consider the predictive performance of the two models the estimated form of each model based on Larson's data is used to allocate the nonreporting positions for 10 subsequent occasions on which complete market survey data are available. The results are shown in Table 4, which shows Theil's U statistic (7, p. 32) for both the Larson and revised procedures. The smaller the value of U obtained the better is the predictive performance of the method. The procedures appear to give approximately equally accurate forecasts of the long hedging position, but the revised method yields clearly superior forecasts of the other three categories.

TABLE 1.—ESTIMATION OF EQUATIONS (3)-(6): LARSON'S METHOD

Dependent variable / Independent variable	Log (HL/RHL)	Log (HS/RHS)	Log (SL/RSL)	Log (SS/RSS)
Constant	-0.836	-0.172	-0.448	0.132
	(0.269)	(0.063)	(0.219)	(0.228)
NRL	2.368		2.732	
	(0.420)		(0.369)	
NRS		1.085		2.576
		(0.158)		(0.567)
R^2	0.602	0.663	0.723	0.548

[2] The differences between the estimates reported in Table 1 and those reported by Larson are due in part to minor differences in the data but mainly to Larson's use of logarithms to base 10 rather than natural logarithms.

TABLE 2.—ESTIMATES OF EQUATIONS (9)-(12): REVISED METHOD

Dependent variable Independent variable	NRHL	NRHS	NRSL	NRSS
Constant	-0.670	-0.574	-0.451	0.252
	(0.316)	(0.164)	(0.314)	(0.162)
NRL	1.438		1.644	
	(0.792)		(0.756)	
NRL^2	-0.628		-0.564	
	(0.542)		(0.545)	
RHL	0.708		-0.183	
	(0.227)		(0.281)	
RSL	0.217		0.460	
	(0.183)		(0.269)	
NRS		1.422		-0.721
		(0.444)		(0.438)
NRS^2		-0.816		1.413
		(0.579)		(0.571)
RHS		0.541		-0.194
		(0.199)		(0.196)
RSS		0.428		0.318
		(0.230)		(0.227)
R^2	0.546	0.528	0.770	0.798

TABLE 3.—COMPARISON OF PROCEDURES: GOODNESS-OF FIT (r^2)[a]

Variable Method	NRHL	NRHS	NRSL	NRSS
Larson	0.672	0.506	0.252	0.491
Revised	0.546	0.528	0.770	0.798

[a] r^2 is square of simple correlation coefficient between actual and fitted value.

TABLE 4.—COMPARISON OF PROCEDURES: PREDICTIVE ABILITY (U)[a]

Variable Method	NRHL	NRHS	NRSL	NRSS
Larson	0.646	0.411	0.223⁻	0.422
Revised	0.668	0.342	0.109	0.156

[a] U is Theil's measure of forecast accuracy.

CONCLUSION

This note has proposed a procedure for providing consistent estimates of hedging, speculative, and matching positions which can be readily computed from monthly reports presently issued by the Commodity Futures Trading Commission. The estimates so obtained may be useful in studying speculative or hedging participation in commodity markets. There remains little doubt, however, that the estimates reported here could be further improved by the inclusion of new survey data. The need for new surveys to be carried out has been given further impetus by the merging importance of futures trading in financial instruments where patterns of commercial use presumably differ somewhat from the more traditionally traded agricultural commodities. Additional data will be required to determine the extent to which the results reported in this paper can be usefully applied to any regularly published data for these markets.

CITATIONS

1 Roger W. Gray, "Price Effects of a Lack of Speculation," *Food Research Institute Studies*, VII, Supplement, 1967

2 A. B. Larson, "Estimation of Hedging and Speculative Positions in Futures Markets," *Food Research Institute Studies*, II, 3, November 1961.

3 W. G. McCain, "An Empirical Investigation into the Effects of Margin Requirements in Organized Commodity Futures Markets," Ph.D. Dissertation, Stanford University, June 1969.

4 Allen B. Paul, "Treatment of Hedging in Commodity Market Regulation," U.S. Department of Agriculture, Economic Research Service, Technical Bulletin, No. 1538, Washington, D.C., April 1976.

5 A. A. Powell, "Aitken Estimators as a Tool in Allocating Predetermined Aggregates," *Journal of the American Statistical Association*, September 1969.

6 David J. S. Rutledge, "Hedgers' Demand for Futures Contracts: A Theoretical Framework with Applications to the U.S. Soybean Complex," *Food Research Institute Studies*, XI, 3, 1972.

7 H. Theil, *Economic Forecasts and Policy*, Second Edition (Revised), North-Holland Publishing Company, Amsterdam, 1965.

8 R. W. Ward and L. B. Fletcher, "From Hedging to Pure Speculation: A Micro Model of Optimal Futures and Cash Market Positions," *American Journal of Agricultural Economics*, 53, 1, February 1971.

9 U.S. Department of Agriculture, Economic Research Service, *Margins, Speculation and Prices in Grain Futures Markets*, Washington, D.C., December 1967.

10 U.S. General Accounting Office, *Improvements Needed in Regulation of Commodity Futures Trading*, Washington, D.C., June 24, 1975.

11 Holbrook Working, "Hedging Reconsidered," *Journal of Farm Economics*, 35, 1953.

12 ———, "Speculation on Hedging Markets," *Food Research Institute Studies*, 1, 2, May 1960.

13 A. Zellner, "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 56, 298, June 1962.