Index Arbitrage and Nonlinear Dynamics Between the S&P 500 Futures and Cash

Gerald P. Dwyer, Jr., Peter Locke, and Wei Yu

Federal Reserve Bank of Atlanta
Working Paper 95-17
November 1995

Abstract: We use a cost of carry model with nonzero transactions costs to motivate estimation of a nonlinear dynamic relationship between the S&P 500 futures and cash indexes. Discontinuous arbitrage suggests that a threshold error correction mechanism may characterize many aspects of the relationship between the futures and cash indexes. We use minute-by-minute data on the S&P 500 futures and cash indexes. The results indicate that nonlinear dynamics are important and related to arbitrage and suggest that arbitrage is associated with more rapid convergence of the basis to the cost of carry than would be indicated by a linear model.

JEL classification: G13, C32

The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta, the Federal Reserve System, or the Commodity Futures Trading Commission (CFTC). The authors thank Mason Gerety, David B. Gordon, Laura Kodres, Ted Jaditz, Chera Sayers, Fallaw Sowell, the editor Andrew Lo, and a referee for helpful comments. Participants in seminars at the CFTC, the University of Illinois, and the University of Wyoming also provided helpful comments. This work was begun during a short visit at the CFTC by Dwyer, and the final stages were completed while he was a Visiting Scholar at the Federal Reserve Bank of Atlanta. Dwyer thanks these institutions and the National Science Foundation for research support. Any remaining errors are the authors' responsibility.

Please address questions of substance to Gerald P. Dwyer, Jr., Research Department, Federal Reserve Bank of Atlanta, 104 Marietta Street, N.W., Atlanta, Georgia 30303-2713, 404/521-7095, fax 404/521-8956; dwyerg@clemson.edu; Peter Locke, Commodity Futures Trading Commission, Three Lafayette Centre, 1155 21st Street, N.W., Washington, D.C. 20581, 202/418-5260; and Wei Yu, Gerontology Center, Boston University, 53 Bay State Road, Boston, Massachusetts 02215, 617/353-5045, 617-353-5047 (fax), wyu@bu.edu.

Index Arbitrage and Nonlinear Dynamics between the S&P 500 Futures and Cash

Arbitrage is like gravity. We often invoke it and take its effects for granted. From the viewpoint of straightforward economic theory, it is hard to find ill effects of arbitrage. Nonetheless, when arbitrage is observable, for example, for stock index futures or foreign exchange, it is surprisingly controversial. Our analysis in this paper provides a natural empirical framework in which to evaluate the effects of arbitrage on the time paths of prices. In this paper, we examine arbitrage in the specific context of index arbitrage for the S&P 500.

Early research on stock index futures by, among others, Modest and Sundaresan (1983) and Figlewski (1984) focused on persistent mispricing, deviations from the cost of carry. Recent estimates of transactions costs [for example, Klemkosky and Lee (1991); Chung (1991)] indicate that the frequency of mispricing has fallen over time. Brennan and Schwartz (1990) find that the mispricing series calculated by MacKinlay and Ramaswamy (1988) is reasonably consistent with an arbitrage model they develop, which emphasizes position limits and offsets of arbitrage positions before contract expiration.

More recently, some have suggested that the importance of index arbitrage has been overstated. Kawaller (1991) identifies five categories of index futures traders and calculates break-even mispricing for each category based on their relevant transaction costs. Using his costs, he suggests that, "at every futures price, there is at least one market participant who 'should' be using this market" [Kawaller (1991, p. 460)]. Miller, Muthuswamy, and Whaley (1994) show that the S&P 500 basis is mean reverting and argue that much of the mean reversion is explicable by infrequent trading in the cash market. On the other hand, Neal (1992) finds that index arbitrage tends to be associated with mean reversion in the basis, although his sample is relatively small and some arbitrage appears at apparently unprofitable times.

Partially motivated by the effects of arbitrage, another line of research examines the dynamic link between the futures and cash indexes using vector autoregressions. These analyses can be interpreted along the lines of Garbade and Silber's (1983) model in which a continuum of traders induces continuity in mean reversion. Among others, Kawaller, Koch, and Koch (1987), Stoll and Whaley (1990), and K. Chan (1991) have estimated such models. A general finding is that changes in the futures lead the cash market, with less or no feedback from the cash to futures.

Focusing on the implications of infrequent arbitrage, Yadav, Pope, and Paudyal (1994) suggest a nonlinear model for deviations of futures and cash stock indexes. Using hourly data on the FTSE100 index futures traded on the London International Financial Futures Exchange and spot values from the International Stock Exchange, they test whether the basis is better characterized by a self-exciting threshold autoregression (a generalization of an autoregression) than by an autoregression. Their test results suggest that a threshold model for the basis, which they estimate by a heuristic, characterize the data better than an autoregression.

In this paper, we tie together pieces of these somewhat disparate strands of research. We estimate the relationship between the minute-by-minute S&P 500 futures and cash indexes and provide estimates of the effects of arbitrage on the convergence of futures and cash values. A cost of carry model with nonzero transactions costs motivates the empirical work. If there is a group of arbitrageurs with similar transactions costs who enter the market relatively infrequently, then purchases and sales by these agents can affect the dynamic relationship between the futures and cash. To allow for this differential behavior, we estimate a nonlinear econometric model, in particular a threshold error correction mechanism. This generalizes prior estimated vector autoregressions in two ways. First, we allow for the cointegration of the futures and cash indexes by estimating an error correction mechanism. Second, the parameters of the error correction mechanism, and therefore the mean reversion, depend on the level of mispricing. The thresholds are signals for index arbitrage, which can affect the speed of convergence of the basis to its equilibrium value. The use of a threshold autoregression in this context was developed independently by Yadav, Pope, and Paudyal

(1994). The threshold error correction mechanism is a generalization of an autoregression for the basis that allows the futures and cash indexes to adjust at different speeds to mispricing.

Before specifying the relationship between the futures and cash, we provide evidence in Section 1 on possible asymmetry of arbitrage bounds and on mean reversion in the futures and cash indexes and the basis. In Section 2, we motivate our empirical analysis with a simple cost of carry model with transactions costs and describe the threshold error correction mechanism and our estimation strategy. In Section 3, we discuss the estimates in detail.

1. MEAN REVERSION AND SYMMETRY OF THE BASIS

We explicitly introduce transactions costs below, but our preliminary empirical analysis can be understood in the context of arbitrage with zero transactions costs. Let F_t be the futures price of the shares underlying a futures contract at time t that expires in τ days and P_t be the price at t on the cash market for the same shares. Suppose that the interest rate, ρ, is constant to expiration, an assumption that simplifies the algebra.[1] Also suppose that dividends are perfectly predictable and define the dividend rate, δ, to be the ratio of the present value per unit time of dividends received to the cash price at t, P_t. The no-arbitrage condition between the futures and cash prices of stock implied by the cost of carry model with zero transactions costs is

$$(1) \qquad F_t = P_t e^{\rho\tau}(1-\delta\tau).$$

If $\delta\tau$ is small, $(1-\delta\tau)\approx e^{-\delta\tau}$ and equation (1) can be approximated by

$$(2) \qquad F_t = P_t e^{(\rho-\delta)\tau}.$$

Taking logarithms of both sides of (2) and rearranging, we get

$$(3) \qquad f_t - (\rho-\delta)\tau = p_t,$$

where f and p denote logarithms of F and P.

If the logarithms of the futures index adjusted for interest and dividends and the cash index have unit roots, equation (3) implies that the logarithms of the adjusted futures and cash indexes are cointegrated with a coefficient of unity. If the logarithms of the adjusted futures and cash indexes do not have unit roots, equation (3) also implies that the difference does not have a unit root.

A. The Data

Our data are S&P 500 futures and cash time and sales records maintained by the Commodity Futures Trading Commission (CFTC). Our sample period extends from the inception of the S&P 500 futures market in 1982 to the end of 1990. There are four futures contracts expiring each year in March, June, September, and December. Until just before expiration, the next contract to expire is the most frequently traded contract. All of our data are for the final 13 weeks of trading for each contract with a sampling interval of one minute.

The underlying data from the futures market at the Chicago Mercantile Exchange (CME) are the sequence of transactions prices with the time to the second. There can be, and there sometimes are, two or more trades in the same second. Prices are entered into a computer by CME employees in the trading pit at the CME after observing a price change, typically 10 to 20 times per minute. The cash index is calculated less frequently, approximately every 15 seconds, as the weighted sum of the most recent transactions prices of the underlying stocks.

We delete the first 30 minutes of each day from all calculations because the futures and cash indexes have large and continuing deviations from the daily mean for the first 10 to 20 minutes. These deviations presumably reflect delayed openings for stocks on the cash market, the use of call markets to clear the cash at the open, and possibly other factors. The futures and cash markets opened at 10:00 A.M. and closed at 4:00 P.M. and 4:15 P.M. Eastern Time, respectively, from March 1982 to September 30, 1985. Both markets open at 9:30 A.M. since

September 30, 1985. Hence, there are 331 observations for each day before September 30, 1985, and 361 observations thereafter.

We use two samples for our estimates. The first sample contains the last 13 weeks before expiration of the contracts that expired in 1989 and 1990. This sample provides evidence on any patterns over time within a contract. The second sample includes the fifth week before expiration of each contract from June 1982 to December 1990. When the fifth week has a nontrading day, we use the fourth week to increase the number of observations. This second sample provides evidence on any changes over time across contracts since the inception of S&P 500 futures. We refer to this as the "sample across contracts."

We estimate models combining minute by minute data for a week.[2] First, we subtract daily means from the logarithms of the futures and cash indexes. Demeaning the futures removes any constant in the logarithms of the futures due to the constant part of dividends and interest rates for that day. The difference between the demeaned logarithms of the futures and cash indexes is the deviation of the "basis" from its daily mean. If dividend and interest rates are relatively constant during the day, this adjusted basis is an estimate of a mispricing series that does not require other explicit assumptions about expected dividend or interest rates. We create lagged variables using the data only within a day. Then the observations for trading days in a week are combined.

B. Cointegration of the Futures and Cash Indexes

Because the futures index adjusted for interest and dividends is cointegrated with the cash index with a coefficient of unity under (3), an augmented Dickey-Fuller test for each of the weeks in the two samples is a convenient way of testing for cointegration. Below, we find clear evidence of heteroskedastic error terms in more general estimated equations; therefore, we use the Phillips-Perron correction to the standard errors of the estimated coefficients [Phillips and Perron (1988)]. Although sequential tests of lag length generally indicate lag lengths shorter than 10, we present tests based on regressions with 1, 10, and 20 lags to show the insensitivity of our results to the number of lags.

Figure 1 shows test statistics for the nearby weeks of the contracts in 1989 and 1990 for the futures and cash indexes as well as the basis. To assess their sensitivity to lag length, test statistics are presented from regressions with 1, 10, and 20 lags of the left-hand-side variable. The results are quite consistent with unit roots in the futures and cash indexes. Conversely, the results are generally inconsistent with the null hypothesis that the basis has a unit root. With 10 lags, only 7 of the 104 test statistics are consistent with the hypothesis of a unit root in the basis. With 20 lags, 34 of the 104 test statistics are consistent with the null hypothesis, but this is far fewer than would be expected if a unit root in the basis were consistent with the data for all weeks. We conclude that, in 1989 and 1990, the futures and cash indexes have unit roots but the basis does not. That is, cointegration of the futures and cash indexes clearly is supported by the data for 1989 and 1990.

Figure 2 shows the test statistics for the sample across contracts. Again, the test statistics are quite consistent with unit roots in the futures and cash indexes. The null hypothesis of a unit root in the basis, which is inconsistent with cointegration, is more consistent with the data prior to 1987. With 10 lags, the statistics are consistent with a unit root in the basis for 7 of the 19 weeks preceding 1987. The statistics are consistent with a unit root in the basis for only 1 of the remaining 16 weeks from 1987 through 1990. Even though a unit root in the basis often is consistent with the data before 1987, the results are consistent with greater mean reversion of the basis than of the underlying indexes. Even the consistency with a unit root in the basis of 37 percent of the weeks before 1987 is quite unlikely if this null hypothesis is correct for all weeks. If the statistic for each week (13 weeks apart) is independent with a probability of .95 of being consistent with a unit root, the probability of 7 or fewer rejections out of 19 trials is on the order of 10^{-12}. We conclude that, even before 1987, the data do not support the inference that the basis has a unit root.

The change in the test statistics for the basis in 1987 reflects an increase in mean reversion in the basis since 1987. Figure 3 shows over time a summary measure of the basis's mean reversion: one minus the sum of the estimated coefficients in an autoregression for the level of the basis. The larger the value of this measure, the greater the mean reversion. While mean reversion is greatest for June 1987 through March 1988, this evidence suggests that mean reversion has been consistently greater since 1987.

We conclude that the evidence supports cointegration of the futures and cash indexes. This means that vector autoregressions using differences of the indexes, such as those estimated by Kawaller, Kocn, and Koch (1987) and Stoll and Whaley (1990), are misspecified. Indeed, the mean reversion may well explain the long estimates of lags in their vector autoregressions: long lag distributions of first differences with equal, small coefficients can integrate first differences and approximate a relationship in levels.

C. Symmetry of the Distribution of the Basis

If the transactions costs of arbitrage are asymmetric and arbitrage affects the path of the basis as Brennan and Schwartz (1990) suggest, such asymmetry is likely to be reflected in the distribution of the basis.[3] To examine the importance of such asymmetry, we test the null hypothesis of symmetry of the distribution of residuals of basis autoregressions with 10 lags. An autoregression is a linear filter that, at least asymptotically, reduces dependence in the basis and does not affect the symmetry of the distribution. We use parametric and nonparametric tests of symmetry.

The null hypothesis of the parametric test is that the standardized skewness is zero [Gupta (1967)]. This test assumes that the data are independent with finite moments up to the sixth. At the 5 percent significance level, the null is consistent with the data for 98 of the 104 weeks in the 1989 and 1990 sample and for 29 of the 35 weeks in the sample across contracts.

The nonparametric test has two parts [Bradley (1968); Gibbons (1985)]. First, we use a sign test to examine whether the mean equals the median, an implication of symmetry. Then, we use the Wilcoxon signed-rank test to examine whether, conditional on the hypothesis that the median equals the mean, the distribution is symmetric about the mean. At the 5 percent significance level, the null that the mean equals the median is consistent with the data for 102 of the 104 weeks in the 1989 and 1990 sample and for 34 of the 35 weeks in the sample across contracts. At the 5 percent significance level, the null that the distribution is symmetric is consistent with the data for 103 of the 104 weeks in 1989 and 1990 and for all 35 weeks in the sample across contracts.

There is virtually no evidence of asymmetry in the residuals from basis autoregressions. We conclude that, at a minimum, symmetric transactions cost bounds are not obviously inconsistent with the data.

2. THE MODEL

A. Cost of Carry with Transactions Costs

Transactions costs in a cost of carry model provide the motivation for nonlinearity in the basis. With nonzero transactions costs, suppose that the futures price is above the future value of stock determined by the current cash price and the cost of carry. If the position is held to maturity, an arbitrageur will buy in the cash market and sell in the futures market if

$$F_t - P_t e^{(\rho-\delta)\tau} > C, \tag{4}$$

where C is the transactions cost per index contract for a round trip in the futures and cash markets. This inequality can be rearranged as

$$F_t > P_t e^{(\rho-\delta)\tau} [1+c(P_t)], \tag{5}$$

where $c(P_t)=CP_t^{-1}e^{-(\rho-\delta)\tau}$. Taking logarithms of both sides of (5) yields

$$(6) \qquad f_t > p_t + (\rho-\delta)\tau + \ln[1+c(P_t)].$$

Because $e^{-(\rho-\delta)\tau}$ is close to unity and CP_t^{-1} is small [less than one-half percent for member firms and institutions according to Klemkosky and Lee (1991)], $\ln[1+c(P_t)]\approx c(P_t)$. If $c(P_t)$ is approximately constant, the condition for arbitrage with a long position in the cash index and short the futures contract is

$$(7) \qquad b_t - (\rho-\delta)\tau > c,$$

where $b_t=f_t-p_t$ and $c(P_t)=c$. The corresponding condition for being long futures and short cash is

$$(8) \qquad b_t - (\rho-\delta)\tau < -c.$$

For estimation, we make one adjustment to (7) and (8). We introduce a positive delay, d, because we do not expect arbitrage to occur and affect the futures and cash indexes in the same minute that an arbitrage opportunity appears. A delay implies that conditions (7) and (8) can be rewritten as

$$(9) \qquad \begin{aligned} b_{t-d} - (\rho-\delta)\tau &> c, \\ b_{t-d} - (\rho-\delta)\tau &< -c. \end{aligned}$$

The inequalities in (9) suggest that the value of the basis can be interpreted as a trigger for arbitrage. Our statistical analysis is based on the supposition that, when either inequality in (9) is satisfied, the dynamics of the basis are different. In particular, we examine the hypothesis that, because of arbitrage, any mean reversion in the basis is stronger when either inequality in (9) is satisfied. It even is possible for the basis to be mean reverting outside the arbitrage bounds but not within them, although this result would be surprising from the viewpoint of Kawaller's analysis (1991).

The inequalities in (9) are based on the supposition that the arbitrageur expects to hold the position until maturity. It is common, however, for index arbitrageurs to unwind positions before expiration [Neal (1992); Sofianos (1993)]. For example, if an arbitrageur takes a long cash position and the cash price rises relative to the futures so that short arbitrage is profitable, there is some reason to unwind the prior position to make the arbitrage. The marginal cost of unwinding the prior position is the bid-ask spread on the cash market, which is less than the marginal cost of establishing a new position.[4] Brennan and Schwartz (1988, 1990) point out that establishing an arbitrage position can be interpreted as making the arbitrage and acquiring an option to unwind the position when there are arbitrage profits from unwinding it. Sofianos (1993) finds that an S&P 500 arbitrageur actually is quite likely to be able to unwind a position in a day with an arbitrage profit. The expected arbitrage profits when unwinding the position reduce the deviation of the futures and cash necessary to trigger arbitrage by roughly one-half of the round-trip transactions costs. This suggests that the trigger for index arbitrage is about one-half of the round-trip transactions costs, and the estimated threshold value, c, should be about one-half of round-trip transaction costs relative to the index value.

B. The Econometric Model

Equation (9) suggests three regimes for arbitrage: two with arbitrage profitable and one with arbitrage not profitable. The tests above for a unit root in the basis indicate that the logarithms of the futures and cash indexes are cointegrated with a coefficient of unity. This cointegration of the indexes implies that an error correction mechanism characterizes the relationship between them [Engle and Granger (1987)]. If arbitrage activity affects the size of the responses of the futures and cash indexes to lagged variables, the values of the parameters in the error correction mechanism depend on the regime, which we index by i. Together, transactions costs of arbitrage and cointegration suggest a generalization of the relationship to a threshold error correction mechanism (TECM),

$$(10)\qquad \begin{aligned}\Delta f_t &= \beta_i^f b_{t-1} + \gamma_i^f(L)\Delta f_{t-1} + \delta_i^f(L)\Delta p_{t-1} + \varepsilon_t^f\\ \Delta p_t &= \beta_i^p b_{t-1} + \gamma_i^p(L)\Delta f_{t-1} + \delta_i^p(L)\Delta p_{t-1} + \varepsilon_t^p\end{aligned}$$

where β_i^f and β_i^p are the lagged error correction terms, $\gamma_i^f(L)$, $\delta_i^f(L)$, $\gamma_i^p(L)$ and $\delta_i^p(L)$ are polynomials in the lag operator L, and ε_t^f and ε_t^p are zero mean, serially uncorrelated error terms that can be contemporaneously correlated. There are no constant terms in (10) because f, p, and b are deviations from daily means. The regimes are determined by

$$(11)\qquad \begin{array}{lll} i = 1 & \text{if} & c < b_{t-d}, \\ i = 2 & \text{if} & -c \le b_{t-d} \le c, \\ i = 3 & \text{if} & b_{t-d} < -c, \end{array}$$

where the parameters in (11) to be estimated are c, the threshold, and d, the delay. In addition to delayed impacts of arbitrage, lagged changes in the cash index in (10) allow for effects of nonsynchronous trading on the cash index that are reflected in serial correlation of changes in the index.

We reduce the computational burden of estimation by collapsing the equation in the TECM, (10), into a single equation for the basis. This reduces the number of parameters estimated by almost half and the number of observations on left-hand-side variables by half. Subtracting the second equation in (10) from the first and rearranging, we get

$$(12)\qquad b_t = \beta_i(L)b_{t-1} + \delta_i(L)\Delta p_{t-1} + \varepsilon_t,$$

where

$$\begin{aligned}\beta_{i,1} &= [1 + (\beta_i^f - \beta_i^p)] + [\gamma_{i,1}^f - \gamma_{i,1}^p],\\ \beta_{i,j} &= [\gamma_{i,j}^f - \gamma_{i,j}^p] - [\gamma_{i,j-1}^f - \gamma_{i,j-1}^p], \qquad j=2,\ldots,k,\\ \beta_{i,k+1} &= -[\gamma_{i,k}^f - \gamma_{i,k}^p],\\ \delta_i(L) &= \gamma_i^f(L) - \gamma_i^p(L) + \delta_i^f(L) - \delta_i^p(L),\\ \varepsilon_t &= \varepsilon_t^f - \varepsilon_t^p.\end{aligned}$$

With (11), equation (12) is a self-exciting threshold autoregression for the basis with predetermined variables. In (12), the parameters in (10) are not identified, but the threshold and delay parameters still are identified.

C. The Estimation Strategy

We use (11) and (12) to estimate the threshold parameter, c, and the delay, d. Conditional on these values of c and d, we then can estimate the parameters in the equations, (10). With some necessary changes, we use the estimation strategy suggested by Tong (1983; 1990), who also discusses the ergodicity of the model and the consistency of the estimator.

As a preliminary matter, we estimated (11) and (12) for subsamples and examined the results. The details are presented in Yu (1992). These results suggested two modifications in the estimation. First, in part because autoregressive conditional heteroskedasticity (ARCH) is a clear alternative to models nonlinear in the mean, we tested for ARCH in the residuals from estimates of (12) and could reject a null hypothesis of no ARCH.[5] All estimates in this paper allow for ARCH, which has substantial implications for the estimation procedure. Second, Tong (1990) suggests searching for the best lag length simultaneously with the other parameters. For these samples, estimating the lag length has virtually no effect on the estimates of other parameters. Because sequential F-tests on versions of (12) with no threshold indicate that fixing the lag length in differences at 10 is quite consistent with the data and doing so dramatically reduces the search space, we fix the lag length of changes in cash values at 10 and of levels of the basis at 11. While these lag lengths generally are too long, any efficiency loss with 1750 observations is trivial.

With ARCH, the statistical model is (11) and (12) supplemented by an equation for the conditional variance of ε_t, h_t,

(13) $$h_t = \theta_0 + \theta(L)\varepsilon_t^2,$$

where $\theta(L)$ is a polynomial in the lag operator L and the unconditional variance of the errors is $\theta_0/(1-\theta(1))$.[6] Partly because of the practical problem of losing observations every time the basis crosses a regime boundary, we assume that the process (13) is the same for all observations. Preliminary estimates suggest that five lagged squared residuals are more than adequate to characterize the conditional variance. If ε_t is normally distributed with zero mean and conditional variance h_t, the joint log likelihood for the innovations conditional on initial values is

$$(14) \qquad l(\varepsilon) \propto -\sum \ln(h_t) - \sum \varepsilon_t^2/h_t.$$

With sample values replacing population values, we estimate the parameters by maximizing this log likelihood with respect to the parameters.

We use a grid search to estimate the delay, d, and the threshold parameter, c, that maximize the likelihood function for the threshold autoregression (Tong 1983; 1990).[7] Our grid search has two steps. In the first coarse step, we allow the delay to vary from 1 to 5 minutes and the threshold values to vary from .25 to 1.6 standard deviations of the basis with a step width of .15. This range for the basis is based on preliminary estimates. If the estimated threshold that maximizes the likelihood function is equal to the limiting value, we increase the search range. As it turns out, increasing the search range never increases the likelihood function. In the second step, the delay is fixed at the estimate from the first step, and we refine the estimates of the threshold by searching the neighborhood of the best preliminary estimates of the threshold. With step sizes of .02 standard deviations of the basis, we calculate the likelihood function for seven threshold values on each side of the preliminary estimate and use the threshold that maximizes the likelihood. In a few cases, the next-lower value of the likelihood function is relatively close to the one that maximizes the likelihood function and the estimates of the delay or threshold are quite different. In these cases, we do a refined search on this next-lower value as well and report the estimates with the highest value of the likelihood function from either refinement.

For each threshold and delay in the grid search, we use one step of the method of scoring to estimate the parameters in (12) and (13) [Greene (1993, pp. 438-40)]. The usual maximum likelihood estimation of the parameters with ARCH is infeasible because it would require solving a nonlinear maximization problem at each point in the grid search. We get initial consistent estimates of (12) and (13) by ordinary least squares. Because the conditional expectation of the Hessian is block diagonal [Engle (1982)], we update the estimates of (13) and (12) sequentially. We first update the estimates of the parameters in (13) by one Gauss-Newton step. Then, we similarly update the estimates of the parameters in (12). This estimator is not the same as the maximum likelihood estimator, but the estimator has the same asymptotic normal distribution as the maximum likelihood estimator.

3. THE ESTIMATES

In this section, we examine the empirical evidence concerning the effects of arbitrage on the time paths of the futures and cash indexes. First, we test whether the threshold error correction mechanism (TECM) fits better than a simple ECM. A better fit is necessary but not sufficient for arbitrage to have interesting effects on the paths of the futures and cash indexes. We then examine the estimated thresholds for their consistency with independent estimates of transactions costs of arbitrage and the occurrence of arbitrage. Taken together, these results shed light on the usefulness of the TECM for characterizing the response of the futures and cash indexes to arbitrage.

A. The Fit of the Estimates

Does the threshold error correction mechanism (TECM) fit better than a simple error correction mechanism (ECM) without thresholds? We use a Monte Carlo procedure to examine whether the estimated TECMs fit better than ECMs. Even without ARCH, the distribution of the usual "likelihood ratio test statistic" (-2 times the logarithm of the likelihood ratio) is nonstandard and practically impossible to determine analytically [K. S. Chan (1991)]. Because the consistent procedure that we use is not the maximum likelihood estimator and there is no proof that the distribution of the test statistic is invariant to ARCH, we cannot use Chan's tabulated distribution. As a result, we estimate the distribution of the likelihood ratio test statistic for our estimates by Monte Carlo estimation.

In the test, the ECM is the null hypothesis and the TECM is the alternative hypothesis.[8] We compare the estimated likelihood ratio test statistics for 1989 and 1990 with an estimate of the distribution from 250 samples drawn under the null hypothesis of an ECM. For the Monte Carlo estimate of the distribution, we calculate the average values of the parameters of estimated ECMs for the 104 nearby weeks in 1989 and 1990 and use these averages as estimates of the true parameters. For each point in the Monte Carlo estimate of the distribution of the likelihood ratio test statistic, we generate 2000 standard normal deviates and use the average parameters to generate a sequence of futures and cash indexes. Using the last 1750 observations, we estimate an ECM, the correct model under the null hypothesis, and a TECM using our estimation procedure, the alternative hypothesis. We then calculate the likelihood ratio test statistic using the relative values of the likelihood function (14). To estimate the distribution of the TECM's likelihood ratio statistic under the null, we repeat this procedure 250 times.

The resulting estimated distribution of the likelihood ratio test statistic under the null hypothesis of an ECM is presented in the top graph in Figure 4. The distribution of the test statistics for the 1989 and 1990 sample is presented in the bottom panel of Figure 4. This distribution also is summarized by contract and week to maturity in Table 1. Under the null hypothesis, 4.8 percent of the likelihood ratio test statistics are greater than or equal to 80.99.

Forty-eight of the 104 test statistics are greater than or equal to 80.99, a quite unlikely result if there are no threshold effects for any week. If the probability of not rejecting the null is .952 and the draws are independent, the probability of only 56 or fewer of 104 test statistics consistent with the null is on the order of 10^{-8}. Hence, we conclude that the null of no threshold effects is inconsistent with the data.

We consider two interpretations of almost half of 104 estimates being in the upper 5 percent tail of the distribution. These interpretations depend on the alternative hypothesis considered relative to the null hypothesis of no thresholds in any week. One alternative hypothesis is that some weeks are consistent with the null hypothesis and others are not. The second alternative is that all weeks have threshold effects. The relative appeal of these interpretations depends on whether there are any patterns in the deviations from the null.

We see no clear pattern in the values of the likelihood ratio test statistics in Table 1. One way of summarizing the results is in terms of a regression of the likelihood ratio test statistics on dummy variables for each contract and for weeks to expiration. In such a regression, neither a set of dummy variables for each contract nor a set of dummy variables for weeks to expiration is statistically significant at even the 20 percent significance level.[9] We see no evidence of systematic patterns in the test statistics and conclude that there is no reason to single out some weeks as better characterized by a TECM with ARCH than an ECM with ARCH.

Figure 5 shows the estimated likelihood ratio test statistics for the sample across contracts. This figure suggests a possible upward trend over time.

B. The Estimated Thresholds and Delays

If the TECM reflects transactions costs and arbitrage, the estimated thresholds should be related to the transaction costs of index arbitrage. Estimates of transaction costs generally range from 0.3 percent to 0.6 percent of the index value [Neal (1992)]. Using data from

Klemkosky and Lee (1991) and assumptions in Kawaller (1991), we estimate that the transaction costs of index arbitrage are about 0.25 percent of the index value for New York Stock Exchange (NYSE) member firms and 0.38 percent of the index value for institutional investors [Yu (1992, pp. 24-29)]. These estimates are based on values used by Kawaller (1991): an index value of 335, not much different than the average value in 1989 and 1990 of 328; and an average value-weighted stock price of $67. Because index arbitrageurs unwind positions before maturity, one-half of these estimates of the round-trip transactions costs are rough estimates of the arbitrage bounds. If the thresholds are related to transactions costs, these estimates of transactions costs indicate that the estimated thresholds should be around 0.125 and 0.19 percent, or .00125 to .00190.

The estimated thresholds largely are consistent with these independent estimates of the transactions costs. For the sample across contracts, the median value of the estimated thresholds is 0.00163, which is greater than half the transactions costs for NYSE member firms and less than for institutional investors. The median value of the estimated thresholds for the 104 weeks in 1989 and 1990 of 0.00102 is below the estimated lower bound of .00125, half the transactions costs for member firms, but this could be due to imprecision or errors in the estimates of the transactions costs, the median threshold, or both. We conclude that our estimated thresholds generally are consistent with transactions costs of index arbitrage.

Data on index arbitrage trades provided by the Securities and Exchange Commission provide additional evidence about the relationship between our estimated thresholds and arbitrage. These data are quantities and approximate times of S&P 500 index arbitrage trades reported by NYSE member firms to the NYSE. Table 2 shows summary statistics on the proportions of the total time in the tail regimes and the proportions of index arbitrage trades and volume in the tail regimes for the first 50 weeks in 1989 (excluding the mini-crash in the week of October 9). To allow for some mistiming, we pad our estimates of the tail regimes by

0, 1, and 2 minutes around each side of a minute in the estimated tail regimes. For example, for one minute of padding, if a minute is in a tail regime, we include the preceding minute in the regime whether or not the additional minute is in the tail regime; similarly for the succeeding minute. We did not expect all of the times with index arbitrage to correspond exactly to times in the estimated tail regimes, and Table 2 shows that result. With no padding, about 43 percent of trades and volume are in the tail regimes; with 2 minutes of padding, about 67 percent of trades and volume are in the tail regimes. Whether these percentages are large or small in and of themselves is a matter of judgment.

If our interpretation of the estimated thresholds is correct, however, there should be more trades in the tail regimes than would be expected by chance. If all the mean reversion in the basis were due to nonsynchronous trading or traders shopping for the best price, there would be a "random" proportion of index arbitrage in the tail regimes. We examine this hypothesis by testing whether the proportion of trades and volume in the tail regimes is the same as the proportion of minutes in the tail regimes. The normally distributed test statistics in Table 2 are inconsistent with this null hypothesis. The failure of volume to be even more concentrated outside the estimated thresholds suggests that our estimates are missing a systematic component of the data. Nonetheless, the concentration of arbitrage outside the thresholds lends further credence to interpreting the thresholds as estimates of arbitrage bounds.

The estimated thresholds for the sample across contracts, shown in Figure 6, suggest that the arbitrage bounds have decreased over time, but the evidence is not overwhelming. If the first contract is ignored because the estimated threshold is not statistically significant at the 5 percent level, there is some support for decreasing bounds in the early period. Personally, we see a clearer suggestion that the variation of our estimates is higher in early years.

We see no clear pattern in the estimated thresholds as expiration approaches. Figure 7 shows the estimated thresholds for 1989 and 1990 and the time remaining to expiration.

Perhaps the strongest suggestion in the figure is a decrease in the variance of the estimated thresholds in the last weeks before expiration.

There is clearer evidence that delays in responses have decreased over time. Figure 8 shows the estimated delays for the sample across contracts. In the first half of the sample across contracts, 41 percent of the weeks have estimated delays of 3 minutes or less; in the second half of the sample, 89 percent of the weeks have estimated delays of 3 minutes or less. The estimates for 1989 and 1990, shown in Figure 9, provide additional support for shorter delay in more recent years. All but 8 of these 104 weeks have delays of 3 minutes or less. This evidence is not necessarily compelling, but we conclude that delay has decreased over time.

C. The Estimated Equations

Table 3 summarizes the equations estimated for the sample where we expect more homogeneity, 1989 and 1990. We summarize the distribution of estimates by presenting the median of the estimates.[10] A summary of the distribution is inevitable, given the large number of parameters estimated. We present medians in part because they are less sensitive than other averages to extreme values. Coefficients are estimated for three regimes: an upper regime in which the deviation of the basis from its mean is greater than the upper threshold, c; a middle regime in which the deviation of the basis is between the thresholds; and a lower regime in which the deviation of the basis is less than the lower threshold, -c. The ARCH process is assumed to have the same parameters in the three regimes. It is clear that the equations have too many lags. Nonetheless, the sets of all estimated coefficients generally are statistically significant.

The median estimated equations for the S&P 500 cash index are statistically significant at less than the 10^{-6} significance level in all three regimes. The futures equations have median p-values less than commonly used significance levels in the middle regime but not the tail regimes. At least in part, this reflects the larger number of observations in the middle regime than in the tail regimes. The R^2s of the equations actually are greater in the tail regimes. Generally, there are about 1700 observations for a week, with 150 observations in each tail regime and 1400 observations in the middle regime. The R^2s in Table 3 are weighted to reflect the heteroskedasticity and the variation in the variables is for the weighted variables.[11] Both tail regimes have median R^2s for the futures of about .16 and the middle regime has an R^2 of only .04. This pattern of greater R^2s in the tail regimes is repeated for the cash index, with median R^2s of about .48 in the tail regimes and only .14 in the middle regime.

General indications of the interrelationships between variables are provided by tests that coefficients of variables other than the dependent variable equal zero. For the cash equations in all three regimes, the median p-values for test statistics indicate that there are nonzero coefficients of the lagged basis and lagged changes in the futures equations. On the other hand, for the futures equations, we find a corresponding result for the middle regime but not the tail regimes. For the futures equations in the tail regimes, the median p-values are consistent at the 5 percent significance level with zero coefficients for the lagged basis and lagged changes in cash indexes.[12] As for the statistical significance of the overall equations, this may reflect the numbers of observations in the tail regimes. It is tempting to examine differences in regimes by looking at individual coefficients, but complex interactions among coefficients can make such an exercise quite misleading.

A better way of looking at the differences in the regimes is to examine responses to shocks from the futures and cash markets. For this analysis, we use the estimated TECMs with ARCH for the sample of 104 weeks in 1989 and 1990. We make no pretense of having a structural model with the indexes jointly dependent. Any model in which one index value is exogenous to the other would be little different, though, because the correlations of the innovations generally are not larger than 0.10 in magnitude. We use unit shocks to the futures and cash equations, respectively, to generate responses of the futures and cash indexes and the basis for each week. The dynamic response in each regime is indicated by the median response of the basis for the 104 weeks.[18]

For each regime, Figure 10 shows the basis's responses to shocks from the futures and cash markets. These graphs do not show the actual responses in the tail regimes. Estimating the actual responses in the tail regimes would require estimating or simulating the points at which the basis crosses into the middle regime. Instead, the graphs show the response, normalized to one, to a shock in each of the three regimes. These graphs can be used to estimate the actual response to any shock in the tail regimes, including its crossing into the middle regime.

Figure 10 indicates that, in response to shocks to the futures index, the basis converges to zero more rapidly in the tail regimes than in the middle regime. The mean reversion evident in the middle regime is consistent with Kawaller's (1991) argument that traders continually shop for the best price. One way of characterizing the importance of the faster mean reversion in the tail regimes is in terms of the length of time for the basis to return from the tail regimes to the middle regime. One estimate of large deviations of the basis above and below the thresholds for 1989 and 1990 is the median value of the largest weekly deviations from the daily means. The median value of the maximum weekly deviation of the basis in 1989 and 1990 is 0.00273, which can be compared with the threshold estimate of 0.00102. When the weekly deviation of the basis is this large, the basis returns from the upper regime to the middle regime when it falls to about 0.3736 from a normalized value of one. Figure 10 indicates that, in regime one, the basis falls this far in five minutes. In the middle regime, this large a fall takes fourteen minutes. Most of this response to the initial shock is an increase in the cash index. From initial shocked values of one for the futures index and zero for the cash index, the futures index falls to about 0.92 in five minutes and the cash index rises to 0.56.

The response to a shock in the futures index that lowers the basis is similar. The median weekly minimum in 1989 and 1990 is -0.00297. Given the estimated lower threshold of -0.00102, the basis returns from the lower regime to the middle regime when it rises to about -0.3434 from an initial value of minus one. This mean reversion occurs in seven minutes when the basis is below the lower threshold and fifteen minutes when the basis is between the two thresholds. Again, most of the adjustment is in the cash index. In seven minutes, the futures index rises from minus one to -0.94 and the cash index falls from zero to -0.59.

If there are differences across regimes in responses to shocks from the cash market, such shocks appear to result in greater initial mean reversion in the basis in the middle regime, with mean reversion greater in the tail regimes only after fifteen minutes. In response to positive or negative shocks to the cash index, the basis returns to values inside the transactions cost bounds only after more than 20 minutes. Information flows from the futures to the cash indexes and nonsynchronous trading are plausible initial explanations of the different responses to shocks from the futures and cash markets. We have no evidence, however, to support or reject any particular explanation. Overall, the evidence suggests that index arbitrage substantially increases the speed of response to shocks from the futures market.

4. SUMMARY

Mean reversion in the S&P 500 basis is empirically important. This implies that a finite-lag vector autoregression for the first differences of futures and cash indexes is a misspecification. We also find virtually no trace of asymmetry in the basis, which suggests that symmetric models are adequate to characterize the basis.

Using minute-by-minute data, we estimate an error correction mechanism for the S&P 500 futures and cash indexes that allows for nonlinearity suggested by arbitrage with transaction costs. This model is a threshold error correction mechanism (TECM). This TECM fits significantly better than an error correction mechanism. In addition, the estimated thresholds are reasonably close to independent estimates of arbitrageurs' transactions costs, and the estimated periods outside the arbitrage bounds are associated with reported index arbitrage activity. In response to shocks from the futures market, the basis converges as much in five to seven minutes when arbitrage is profitable as it converges in fifteen minutes when arbitrage is unprofitable.

We conclude that nonlinearity over and above persistent volatility is important for adequately characterizing the relationship between the S&P 500 futures and cash indexes. While nonlinear models with sensitive dependence on initial conditions may or may not be useful for understanding crashes in asset markets, we are somewhat more optimistic than Hsieh (1991) about the usefulness of nonlinear models other than ARCH for understanding financial markets.

NOTES

1 The equivalence of futures and forward prices implicit in equation (1) requires that the interest rate and dividend rate be deterministic, not constant.

2 Preliminary estimates for each day have too few observations for reliable estimates of parameters in the tail regimes. Furthermore, whenever we ran tests for individual days and for weeks, the results for individual days are in the same direction as those for weeks, but the variability of the results across weeks is dramatically lower. This is consistent with an increase in the power of the tests without a substantial increase in heterogeneity.

3 Bendat (1990) shows the effects of some asymmetric functions on normally distributed inputs.

4 Commissions for buying and selling the futures are paid when executing the first transaction. Because delivery is in dollars, not shares of stock, the brokerage costs for buying and selling the stock are not avoidable once the position is taken. The only avoidable cost is the bid-ask spread on the second transaction, which can be avoided since the June 1987 contract by selling the shares at the opening prices on expiration day, which determine the expiration price of the futures contract.

5 At the 5 percent significance level for the standard Lagrange multiplier test, the null hypothesis of no ARCH in (12) is inconsistent with the data for 96 of the 104 weeks in the 1989 and 1990 sample and for 29 of the 35 weeks in the sample across contracts.

6 For simplicity, we also estimate simple univariate equations such as (13) for (10), even though the exact relationship of the equations (10) and (12) does not follow for the variance functions.

7 Standard nonlinear routines cannot be used because delay is an integer (and hence the likelihood function is not a continuous function of the delay) and the threshold cannot be estimated more finely than the difference between the observations at the threshold. In

addition, the likelihood function can be far from unimodal. For our data, the likelihood functions quite commonly have three or more local peaks.

8 Under both hypotheses, we estimate the same number of parameters to characterize ARCH, thereby holding constant the number of parameters in this part of the specification.

9 We exclude the likelihood ratio of 6,438 from all regressions. The results are the same whether the likelihood ratios of 218 and 353 are included or excluded.

10 Estimates of all parameters for individual weeks are available on request from Dwyer.

11 Because of heteroskedastic residuals, the usual R^2 is not an adequate measure of goodness of fit. We present R^2s that are the fraction of variation explained of the weighted variables around the weighted means [Judge et al., (1985 p. 32)]. This measure keeps most of the usual properties: in particular, it is bounded above by one. Our estimation procedure does not seek to maximize a transformation of this R^2, and, therefore, the estimated R^2 can be negative. The patterns we emphasize are reflected in unweighted R^2 as well.

12 We include the coefficient on the lagged basis in both tests because the basis includes both variables.

13 The absolute value of the largest eigenvalue is greater than one in 20 weeks for regime 1, in 2 weeks for regime 2, and in 20 weeks in regime 3. Twelve weeks have absolute values of eigenvalues greater than one in both regime 1 and in regime 3. The greater frequency of eigenvalues greater than one in the tail regimes may reflect the inadequacy of delay alone to capture divergence and then convergence occurring when the basis is outside the bounds.

REFERENCES

Bendat, J. S., 1990, Nonlinear System Analysis and Identification, John Wiley and Sons, New York.

Bradley, J. V., 1968, Distribution Free Statistical Tests, Prentice-Hall, Englewood Cliffs, New Jersey.

Brennan, M. J., and E. S. Schwartz, 1988, "Optimal Arbitrage Strategies Under Basis Variability," Studies in Banking and Finance, 5, 167-80.

Brennan, M. J., and E. S. Schwartz, 1990, "Arbitrage in Stock Index Futures," Journal of Business, 63, S7-S31.

Chan, K., 1991, "A Further Analysis of the Lead Lag Relationship between the Cash Market and Index Futures Market," Review of Financial Studies, 5, 123-51.

Chan, K. S., 1991, "Percentage Points of Likelihood Ratio Tests for Threshold Autoregression," Journal of the Royal Statistical Society, Series B, 53, 691-96.

Chung, Y. P., 1991, "A Transactions Data Test of Stock Index Futures Market Efficiency and Index Arbitrage Profitability," Journal of Finance, 46, 1791-1809.

Dickey, D. A., and W. A. Fuller, 1979, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," Journal of the American Statistical Association, 74, 427-31.

Engle, R. F., 1982, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," Econometrica, 50, 987-1007.

Engle, R. F., and C. W. J. Granger, 1987, "Co-integration and Error Correction: Representation, Estimation, and Testing," Econometrica, 55, 251-76.

Figlewski, S., 1984, "Hedging Performance and Basis Risk in Stock Index Futures," Journal of Finance, 39, 657-70.

Fuller, W. A., 1976, Introduction to Statistical Time Series, John Wiley and Sons, Inc., New York.

Garbade, K. D., and W. L. Silber, 1983, "Price Movements and Price Discovery in Futures and Cash Markets," Review of Economics and Statistics, 65, 289-97.

Gibbons, J. D., 1985, Nonparametric Statistical Inference, 2nd edition, M. Dekker, New York.

Greene, W. H., 1993, Econometric Analysis, Second edition, Macmillan Publishing Company, New York.

Gupta, M. K., 1967, "An Asymptotically Nonparametric Test of Symmetry," Annals of Mathematical Statistics, 38, 849-66.

Hsieh, D. A., 1991, "Chaos and Nonlinear Dynamics: Applications to Financial Markets," Journal of Finance, 46, 1839-77.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T. Lee, 1985, The Theory and Practice of Econometrics, 2nd edition, John Wiley and Sons, New York.

Kawaller, I. G., 1991, "Determining the Relevant Fair Value(s) of S&P 500 Futures: A Case Study Approach," Journal of Futures Markets, 11, 453-60.

Kawaller, I. G., P. D. Koch, and T. W. Koch, 1987, "The Temporal Price Relationship between the S&P 500 Futures and the S&P 500 Index," Journal of Finance, 42, 1309-29.

Klemkosky, R. C., and J. H. Lee, 1991, "The Intraday Ex Post and Ex Ante Profitability of Index Arbitrage," Journal of Futures Markets, 11, 291-311.

MacKinlay, A. C., and K. Ramaswamy, 1988, "Index-Futures Arbitrage and the Behavior of Stock Index Futures Prices," Review of Financial Studies, 1, 137-58.

Miller, M. H., J. Muthuswamy, and R. E. Whaley, 1994, "Mean Reversion of S&P 500 Index Basis Changes: Arbitrage-Induced or Statistical Illusion?" Journal of Finance, 49, 479-513.

Modest, D. M., and M. Sundaresan, 1983, "The Relationship Between Spot and Futures Prices in Stock Index Futures Markets: Some Preliminary Evidence," Journal of Futures Markets, 3, 15-41.

Neal, R., 1992, "Direct Tests of Index Arbitrage Models," working paper, University of Washington.

Phillips, P. C. B, and P. Perron, 1988, "Testing for a Unit Root in Time Series Regression," Biometrika, 75, 335-46.

Said, E., and D. A. Dickey, 1984, "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," Biometrika, 71, 599-607.

Sofianos, G., 1993, "Index Arbitrage Profitability," Journal of Derivatives, 1, 6-20.

Stoll, H. R., and R. E. Whaley, 1987, "Program Trading and Expiration-day Effects," Financial Analysts Journal, 43, 16-28.

Stoll, H. R., and R. E. Whaley, 1990, "The Dynamics of Stock Index and Stock Index Futures Returns," Journal of Financial and Quantitative Analysis, 25, 441-68.

Tong, H., 1983, Threshold Models in Non-linear Time Series Analysis, Springer-Verlag, New York.

Tong, H., 1990, Non-linear Time Series: A Dynamical Systems Approach, Clarendon Press, Oxford.

Yadav, P. K., P. F. Pope, and K. Paudyal, 1994, "Threshold Autoregressive Modeling in Finance: The Price Differences of Equivalent Assets," Mathematical Finance, 4, 205-21.

Yu, W., 1992, "The Dynamics of Cash and Futures Markets for Stock," Unpublished Ph.D. dissertation, Clemson University.

Table 1
Likelihood Ratio Statistics by Contract and Weeks to Expiration
1989 and 1990

Weeks to expiration	March 1989	June 1989	September 1989	December 1989	March 1990	June 1990	September 1990	December 1990
13	72.54	65.43	71.54	62.19	133.99	99.92	68.09	84.78
12	88.02	73.57	71.81	80.58	100.56	72.30	77.01	84.03
11	60.62	81.87	72.05	108.35	65.20	61.59	76.85	94.94
10	94.37	91.20	58.95	6437.52[a]	57.08	85.71	91.71	94.51
9	83.24	59.98	77.99	126.82	65.56	71.87	73.92	67.84
8	82.77	91.13	70.69	80.89	76.84	65.78	76.69	78.80
7	66.04	112.00	73.44	63.22	69.64	93.41	217.73	60.17
6	79.15	68.15	66.82	89.79	81.21	82.70	49.54	88.69
5	89.92	85.69	68.53	86.03	75.50	97.77	83.10	78.59
4	55.28	77.56	104.41	63.60	70.96	87.44	87.86	77.32
3	78.84	62.21	78.17	84.94	78.28	93.53	81.98	59.36
2	105.29	83.88	97.63	79.53	86.07	76.66	81.52	76.80
1	80.60	69.93	104.37	111.51	89.51	115.39	85.98	353.47

a. This week includes the mini-crash in the stock market on October 13, 1989.

Table 2
Fractions of Time and Index Arbitrage in the Tail Regimes
First 50 weeks in 1989

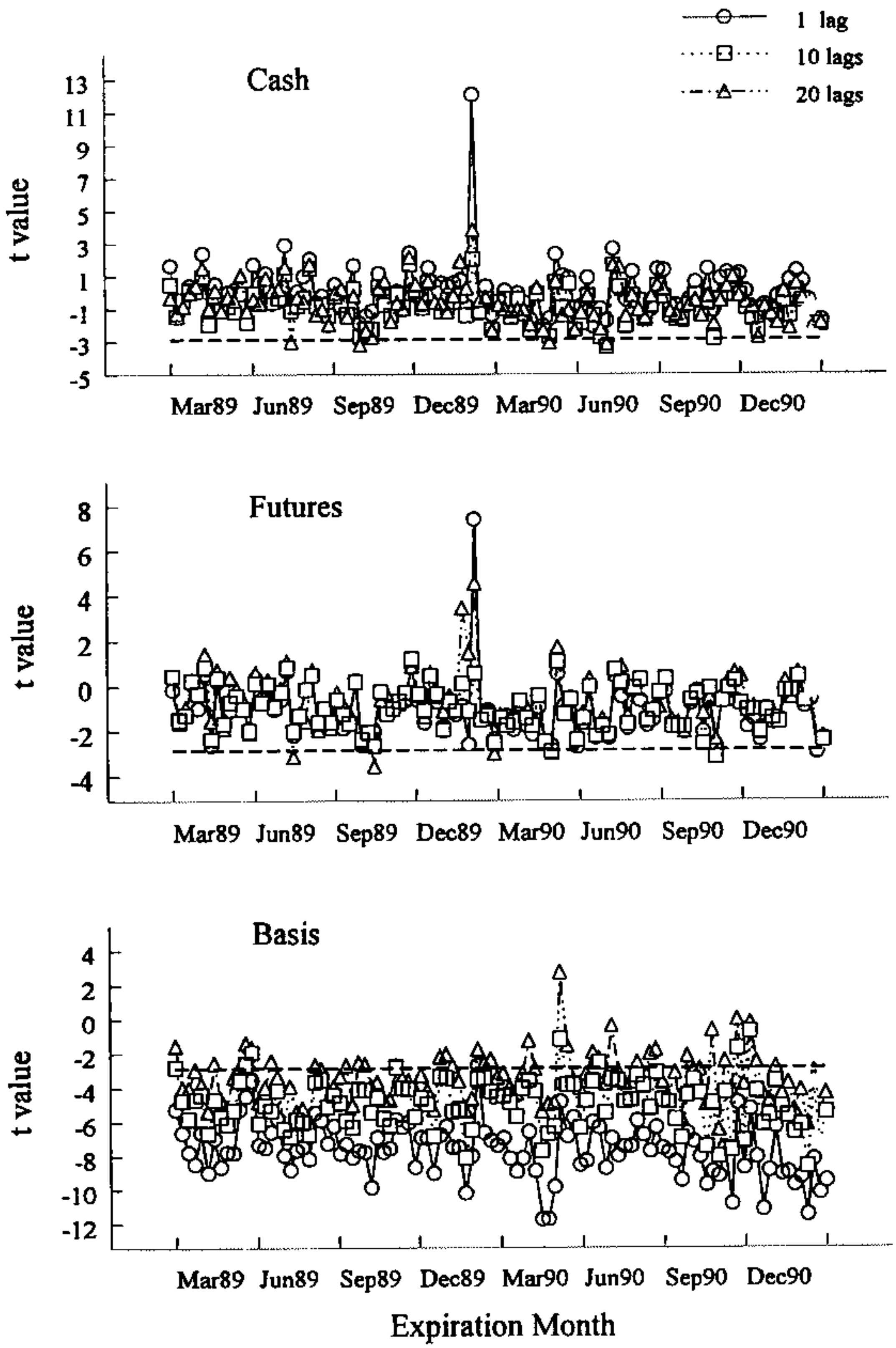
Padding (Minutes)	Fractions in tail regimes			Test statistics	
	Fraction of time	Fraction of arbitrage trades	Fraction of arbitrage volume	Fraction of trades equals fraction of time	Fraction of volume equals fraction of time
0	.215	.427	.430	37.79	40.28
1	.324	.592	.586	42.34	42.79
2	.405	.672	.666	40.51	40.37

The test statistics have asymptotic normal distributions. The test statistics are quite inconsistent with hypotheses that the fraction of arbitrage trades and arbitrage volume in the tail regimes simply reflect the fraction of time in the regimes. Padding the tail regimes to allow for imprecision in our point estimates has no effect on our conclusion. The week of October 9, 1989, which includes a mini-crash, is excluded from the calculations.

Table 3
Estimated Equations
Median Values of Statistics
1989 and 1990

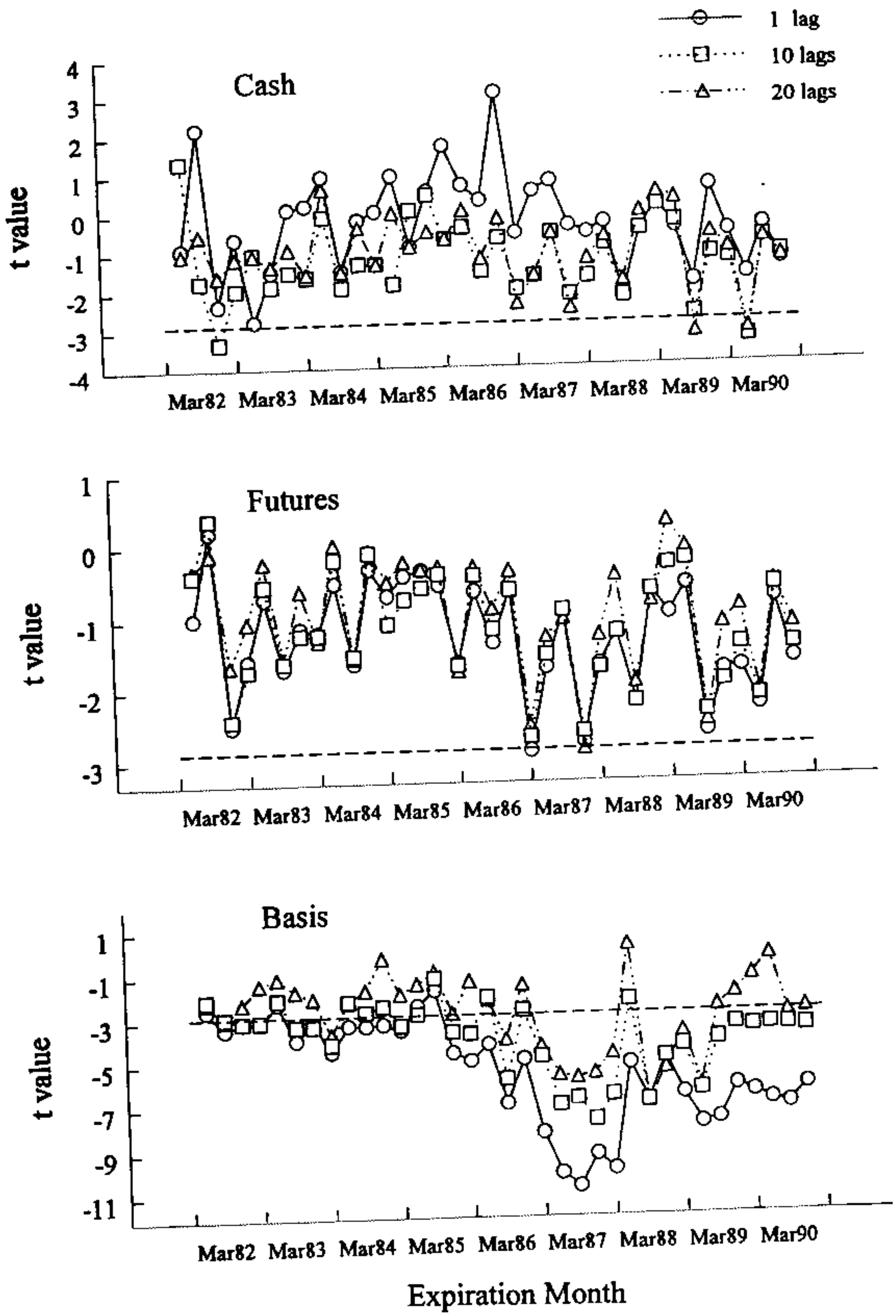
	Regime 1 Basis greater than upper threshold Left-hand-side variable		Regime 2 Basis between thresholds Left-hand-side variable		Regime 3 Basis less than lower threshold Left-hand-side variable	
	Δf	Δp	Δf	Δp	Δf	Δp
Right-hand-side variable			Estimated coefficient (Estimated standard deviation)			
b_{-1}	-0.013	0.041	-0.027	0.035	-0.029	0.024
	(0.043)	(0.015)	(0.040)	(0.014)	(0.021)	(0.008)
Δf_{-1}	-0.063	0.079	-0.041	0.056	-0.020	0.025
	(0.094)	(0.032)	(0.090)	(0.029)	(0.036)	(0.013)
Δf_{-2}	-0.009	0.127	-0.025	0.100	0.019	0.056
	(0.094)	(0.031)	(0.089)	(0.030)	(0.035)	(0.012)
Δf_{-3}	-0.042	0.060	-0.024	0.060	0.003	0.045
	(0.094)	(0.032)	(0.091)	(0.030)	(0.034)	(0.012)
Δf_{-4}	-0.015	0.047	-0.008	0.046	0.002	0.033
	(0.095)	(0.032)	(0.096)	(0.031)	(0.034)	(0.012)
Δf_{-5}	-0.017	0.024	-0.010	0.021	-0.004	0.024
	(0.095)	(0.033)	(0.095)	(0.031)	(0.033)	(0.012)
Δf_{-6}	-0.001	0.030	-0.004	0.030	-0.002	0.020
	(0.091)	(0.032)	(0.092)	(0.031)	(0.032)	(0.012)
Δf_{-7}	-0.006	0.020	-0.008	0.021	0.001	0.016
	(0.090)	(0.033)	(0.089)	(0.031)	(0.031)	(0.011)
Δf_{-8}	-0.011	0.009	-0.007	0.013	-0.008	0.011
	(0.091)	(0.032)	(0.089)	(0.031)	(0.031)	(0.011)
Δf_{-9}	0.012	0.007	-0.010	-0.000	-0.002	0.009
	(0.086)	(0.031)	(0.087)	(0.030)	(0.030)	(0.011)
Δf_{-10}	-0.003	0.007	-0.021	-0.002	0.004	0.006
	(0.086)	(0.030)	(0.086)	(0.029)	(0.029)	(0.011)
Δp_{-1}	0.203	0.236	0.233	0.209	0.301	0.072
	(0.167)	(0.072)	(0.174)	(0.075)	(0.085)	(0.035)
Δp_{-2}	-0.044	-0.052	0.013	-0.054	0.000	0.015
	(0.183)	(0.073)	(0.193)	(0.074)	(0.082)	(0.031)
Δp_{-3}	-0.024	-0.016	0.011	-0.002	-0.031	0.003
	(0.195)	(0.074)	(0.203)	(0.076)	(0.077)	(0.029)
Δp_{-4}	0.010	-0.019	-0.021	-0.012	0.013	-.000
	(0.199)	(0.076)	(0.209)	(0.077)	(0.075)	(0.029)
Δp_{-5}	-0.052	-0.000	-0.000	0.012	0.006	-.003
	(0.207)	(0.079)	(0.214)	(0.075)	(0.073)	(0.028)
Δp_{-6}	-0.014	0.008	-0.014	-0.006	-0.028	-.0018
	(0.205)	(0.075)	(0.217)	(0.078)	(0.072)	(0.027)
Δp_{-7}	-0.003	0.028	0.010	0.012	-0.016	-.004
	(0.206)	(0.079)	(0.220)	(0.076)	(0.071)	(0.026)
Δp_{-8}	-0.066	-0.009	-0.027	-0.011	-0.017	0.000
	(0.213)	(0.079)	(0.219)	(0.079)	(0.070)	(0.026)
Δp_{-9}	-0.001	-0.028	-0.001	0.013	-0.020	-.002
	(0.214)	(0.079)	(0.220)	(0.076)	(0.069)	(0.025)
Δp_{-10}	0.020	0.015	-0.007	0.022	-0.006	-.001
	(0.215)	(0.073)	(0.065)	(0.077)	(0.205)	(0.024)
Threshold	1.023×10^{-4}					
Delay	2					
ARCH terms	Δf	Δp				
θ_0	8.702×10^{-8}	1.254×10^{-8}				
	(5.806×10^{-9})	(7.159×10^{-10})				
ϵ^2_{t-1}	0.097	0.156				
	(0.028)	(0.028)				
ϵ^2_{t-2}	0.062	0.036				
	(0.025)	(0.019)				
ϵ^2_{t-3}	0.038	0.020				
	(0.023)	(0.016)				
ϵ^2_{t-4}	0.043	0.020				
	(0.023)	(0.017)				
ϵ^2_{t-5}	0.046	0.019				
	(0.024)	(0.017)				
R^2-weighted	0.168	0.484	0.036	0.139	0.164	0.488
All coefficients zero -- p-value	0.069	$<1.\times10^{-6}$	.001	$<1.\times10^{-6}$	0.099	$<1.\times10^{-6}$
Other variables' coefficients zero -- p-value	0.100	$<1.\times10^{-6}$	.001	$<1.\times10^{-6}$	0.054	$<1.\times10^{-6}$

Figure 1
Test Statistics for One Unit Root
Estimated Autoregressions with 1, 10, and 20 Lags
1989 and 1990



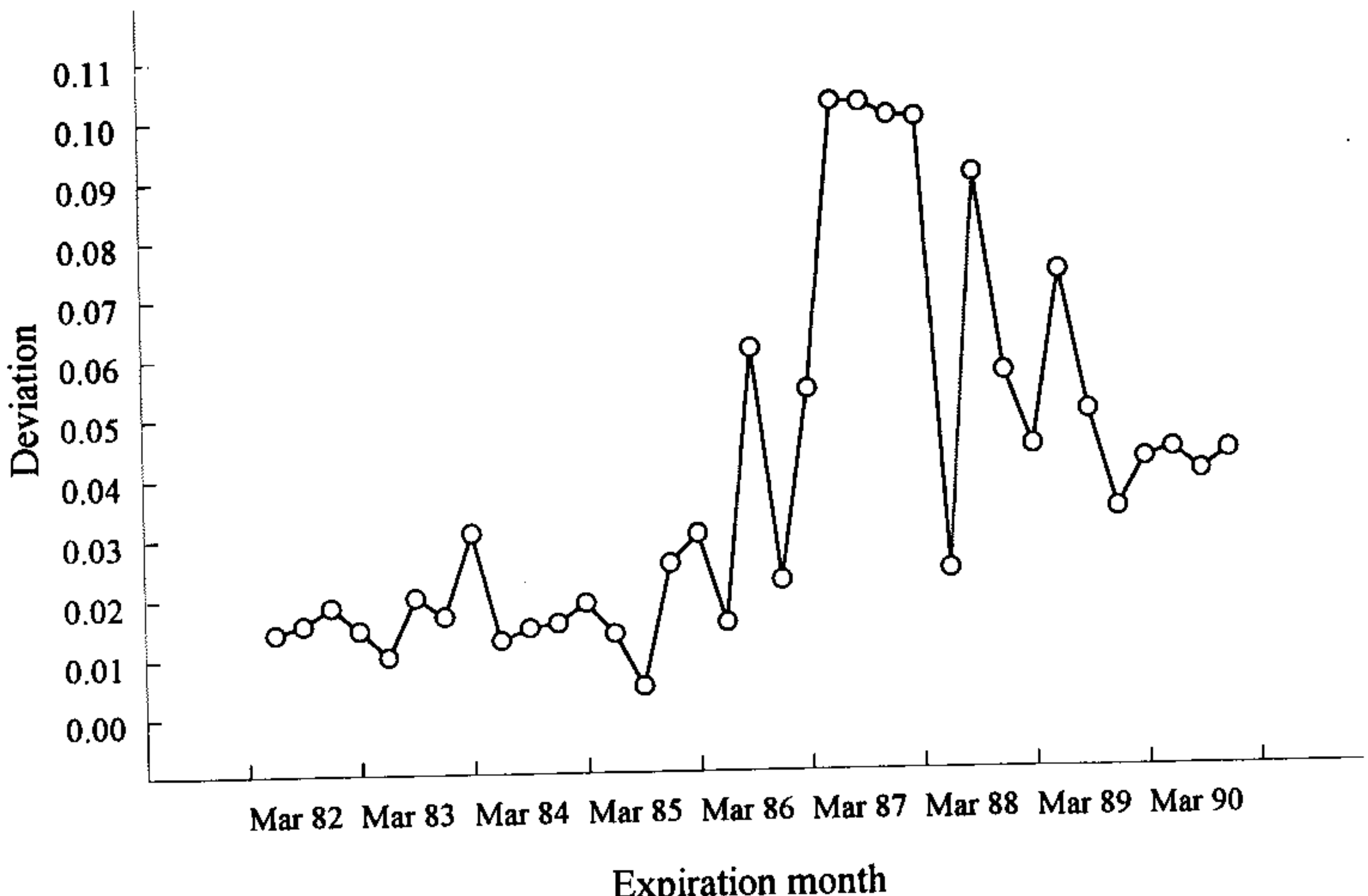
The vertical axis in each panel is the value of the t-statistic for the Dickey-Fuller test for a unit root [Dickey and Fuller (1979); Said and Dickey (1984)]. The horizontal axis is the set of weeks in 1989 and 1990. The t-statistics are estimated with 1, 10, and 20 lags of the left-hand-side variable. The horizontal dashed line at -2.88 is the t-statistic from Fuller (1976) at the 5 percent significance level. Values less than -2.88 are inconsistent with the null hypothesis of a unit root. Greater mean reversion of the basis than the cash or futures is evident.

Figure 2
Test Statistics for One Unit Root
Estimated Autoregressions with 1, 10, and 20 Lags
Sample across Contracts



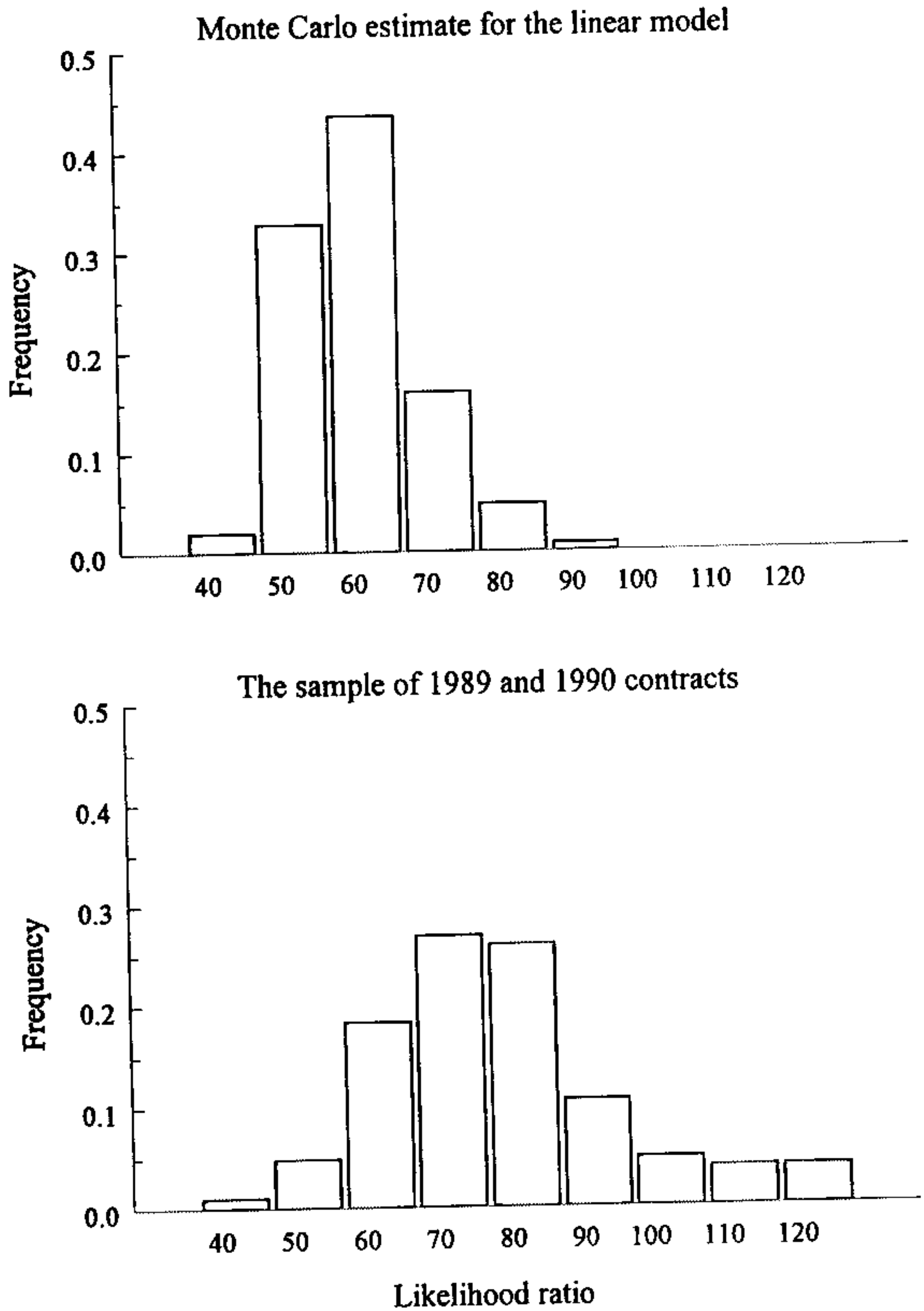
Other than the horizontal axis being the set of contracts from inception through 1990, this figure is the same as Figure 1. Greater mean reversion of the basis than the cash or futures also is evident in this figure. The statistics suggest the possibility of greater mean reversion of the basis since 1987.

Figure 3
Deviations of Sums of Estimated Coefficients from One
Estimated Autoregression for Basis with 10 Lags
Sample across Contracts



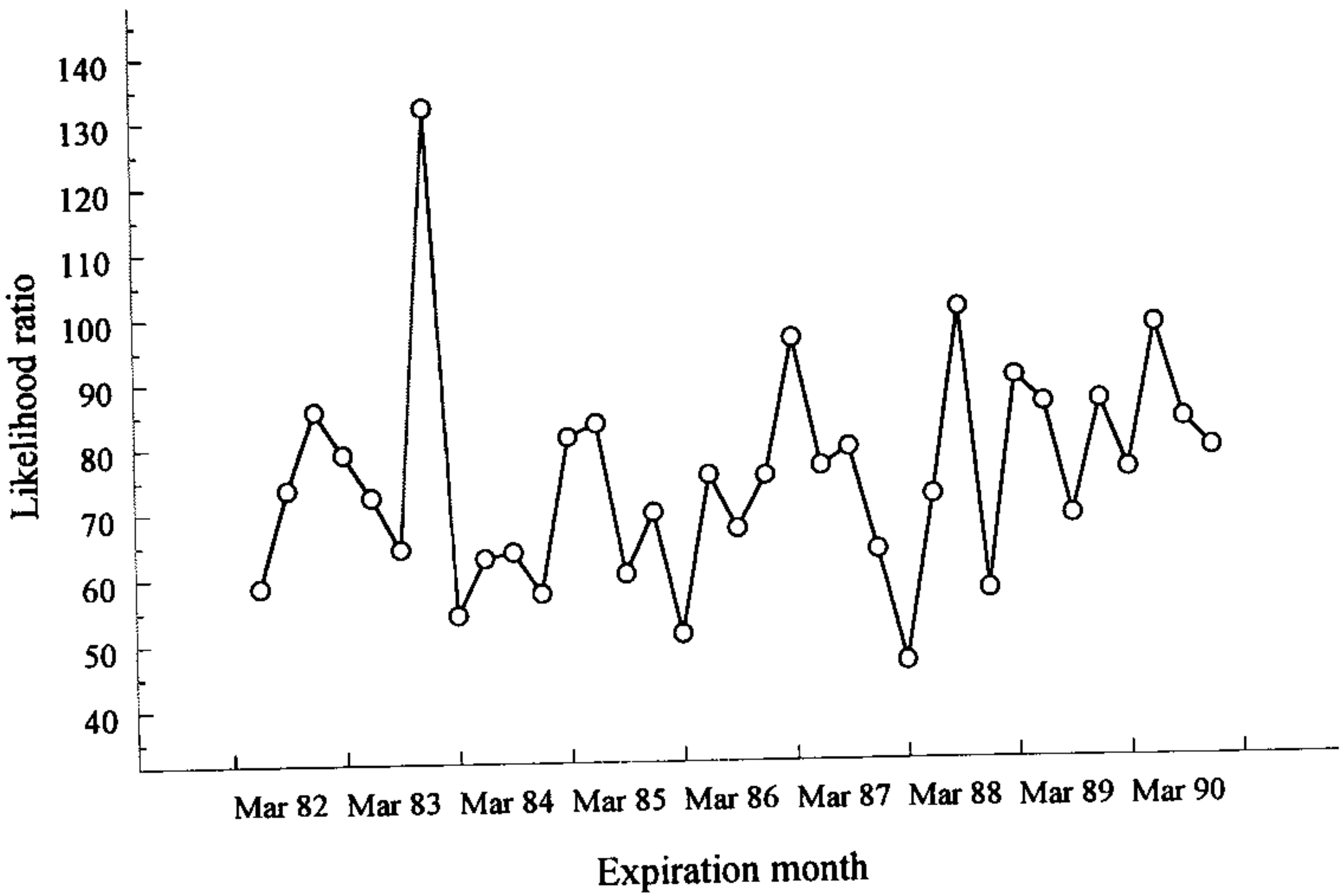
These are the estimates of one minus the sum of the estimated coefficients in autoregressions for the level of the basis. The vertical axis is one minus the sum. The horizontal axis is the set of contracts from inception through 1990. The greater the value in the figure, the greater is the mean reversion of the basis. These estimates are consistent with greater mean reversion since 1987.

Figure 4
Likelihood Ratio Distribution
1989 and 1990



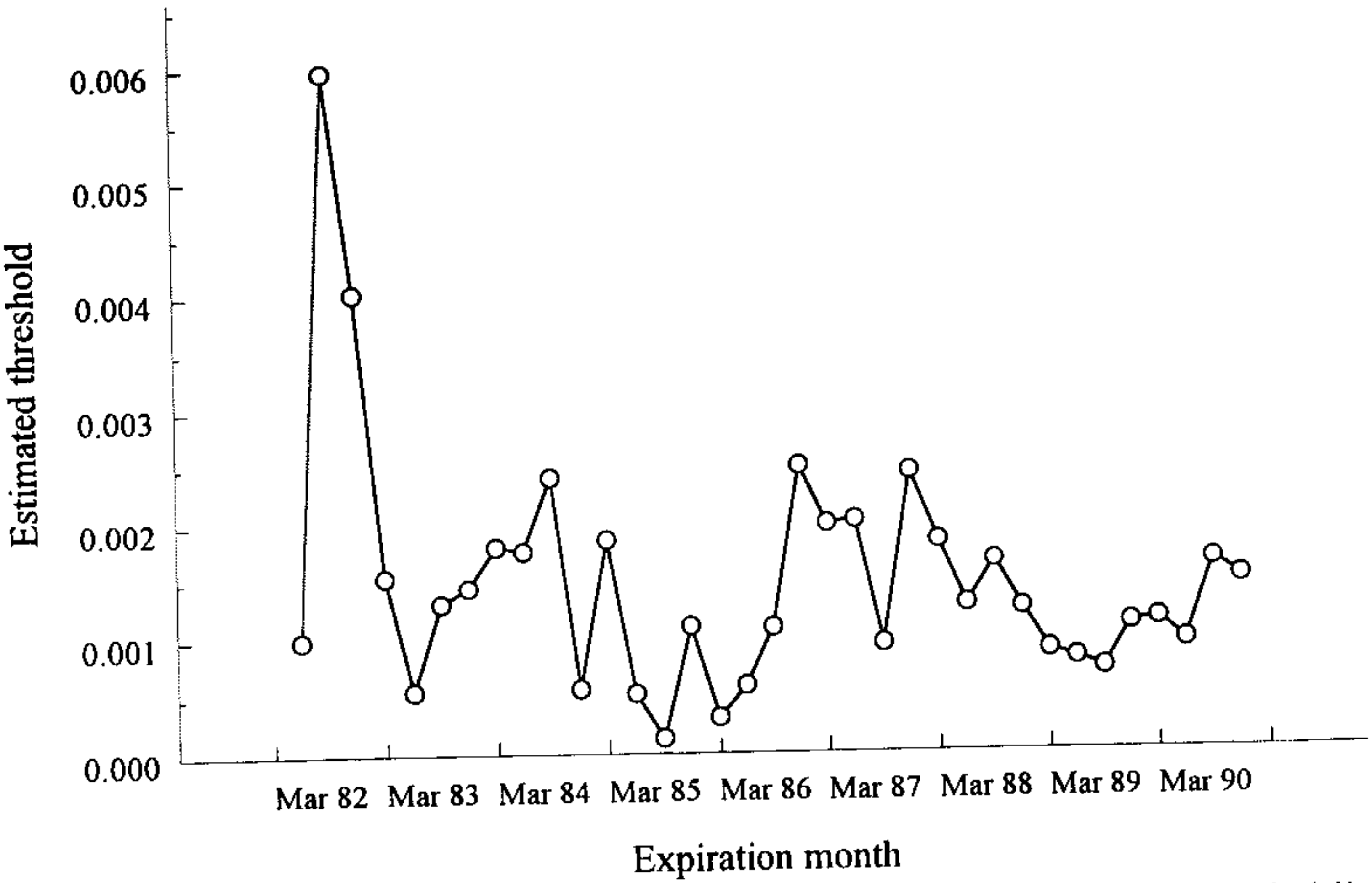
The panels show histograms of the values of the usual likelihood ratio test statistics, -2 times the logarithm of the likelihood ratio, for testing the null hypothesis that the estimated nonlinear TECMS with ARCH fit no better than linear ECMs with ARCH. The vertical axis is the fraction of the estimated values in a class. The top panel shows the estimated distribution under a null hypothesis of an ECM with ARCH with the median values of the estimated parameters by week in 1989 and 1990. The bottom panel shows the actual distribution of the statistics from estimates by week in 1989 and 1990.

Figure 5
Likelihood Ratios
Sample across Contracts



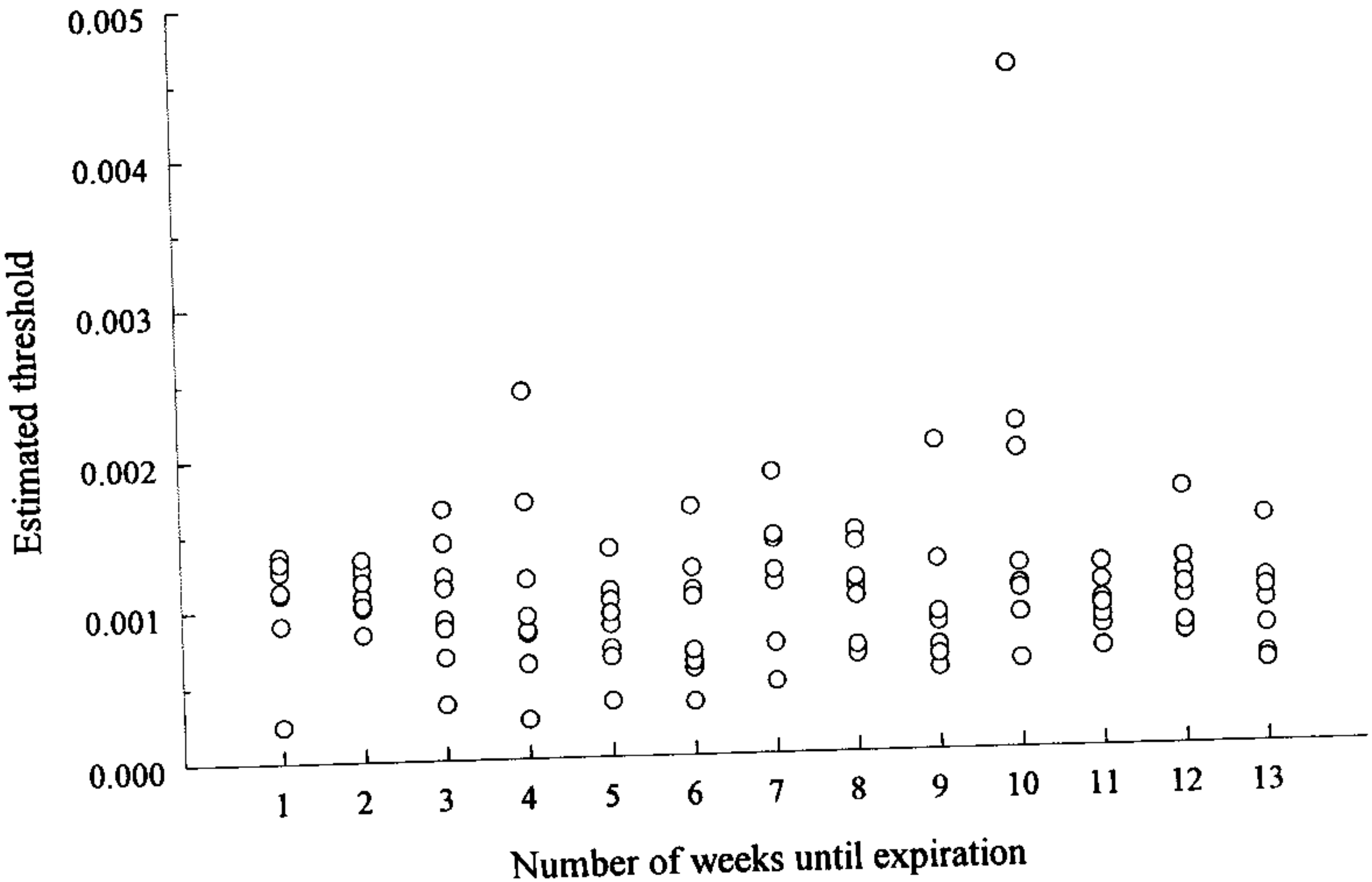
These are the estimates of the usual likelihood ratio test statistics for the sample across contracts. The vertical axis is the value of the statistic and the horizontal axis is the set of contracts from inception through 1990.

Figure 6
Estimated Thresholds for the Basis
Sample across Contracts



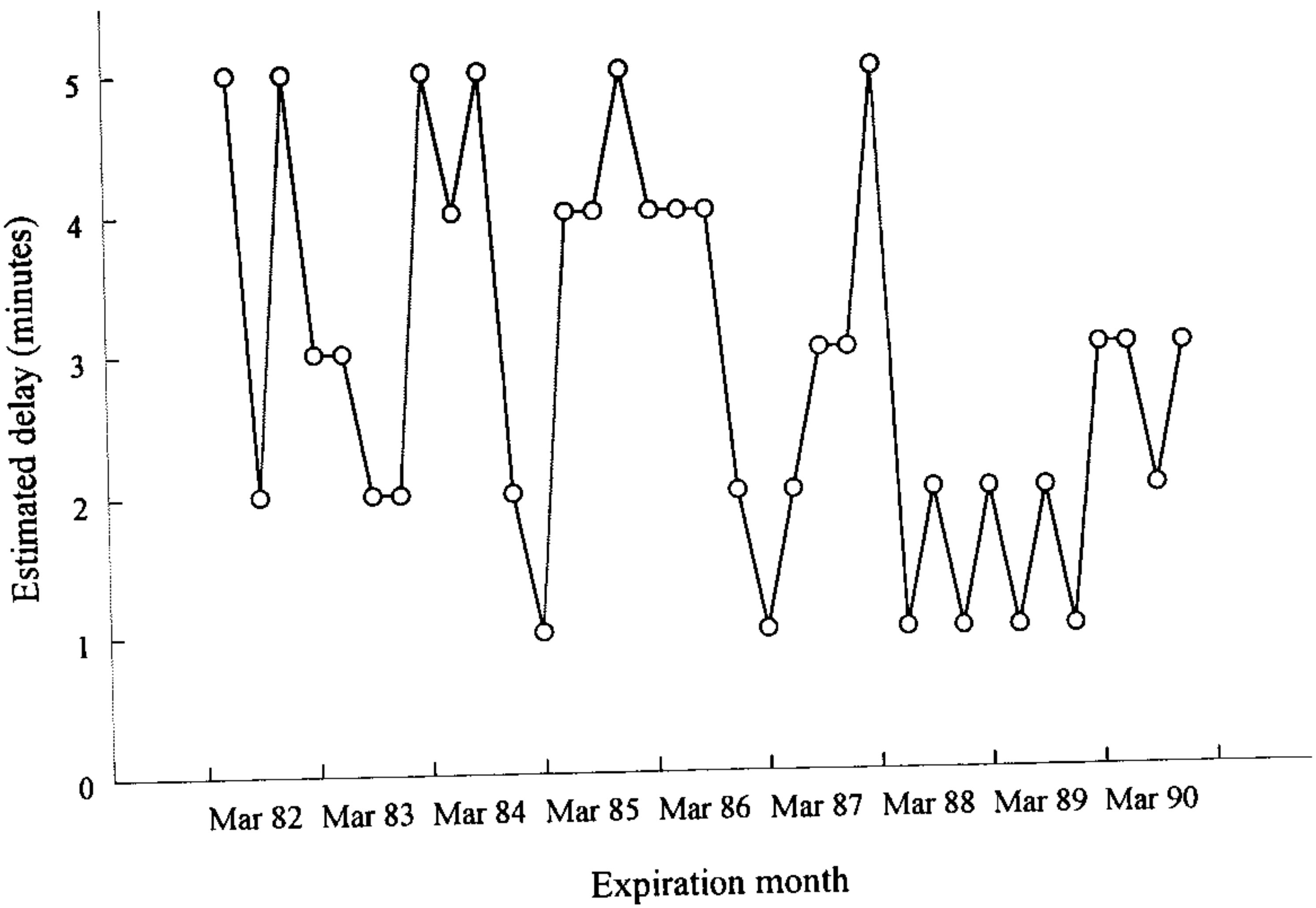
The vertical axis is the value of the estimated threshold and the horizontal axis is the set of contracts from inception through 1990. The thresholds are related to the "basis," the logarithm of the futures minus the logarithm of the cash adjusted for daily means. The estimated thresholds can be interpreted as approximately proportional deviations of the futures and cash indexes from their daily mean difference.

Figure 7
Estimated Thresholds for the Basis
1989 and 1990



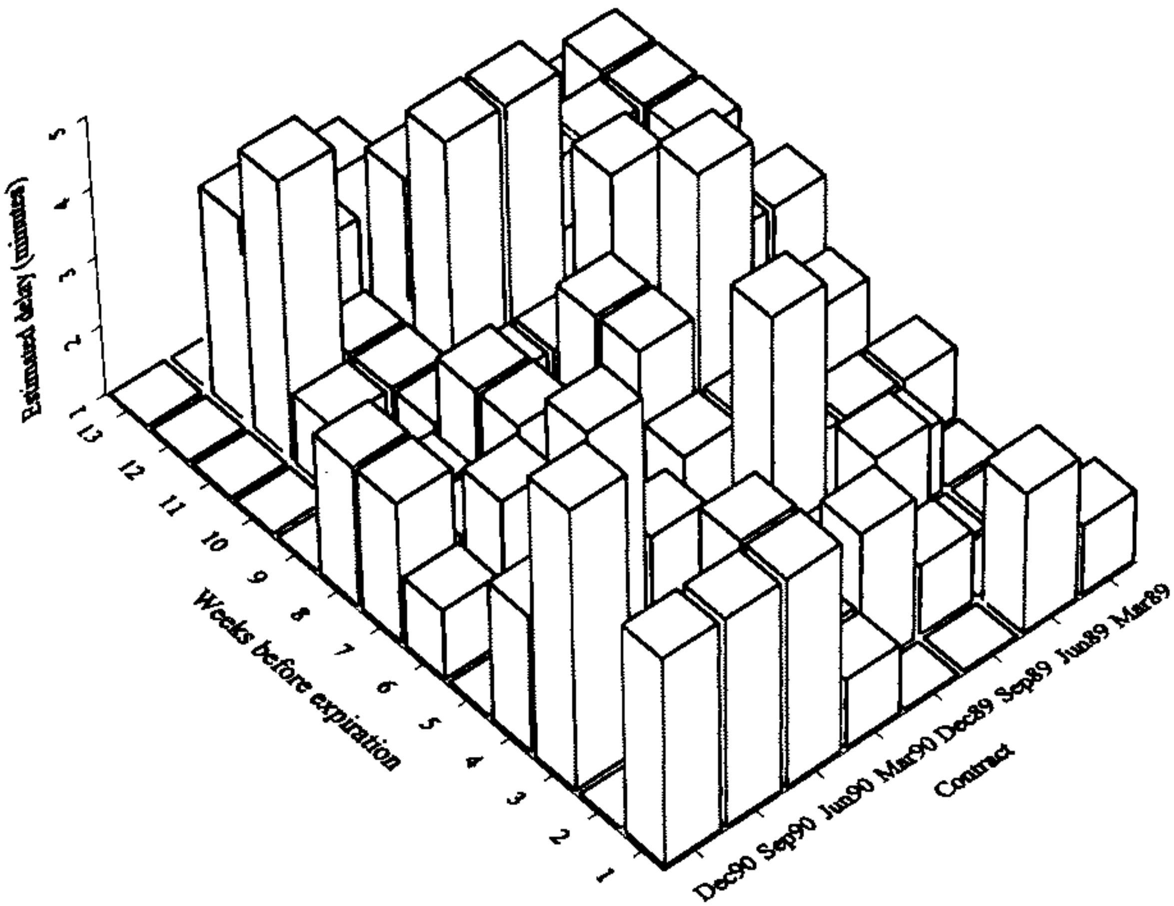
The vertical axis is the estimated threshold and the horizontal axis is the number of weeks to expiration in our sample of the eight contracts in 1989 and 1990.

Figure 8
Estimated Delays
Sample across Contracts



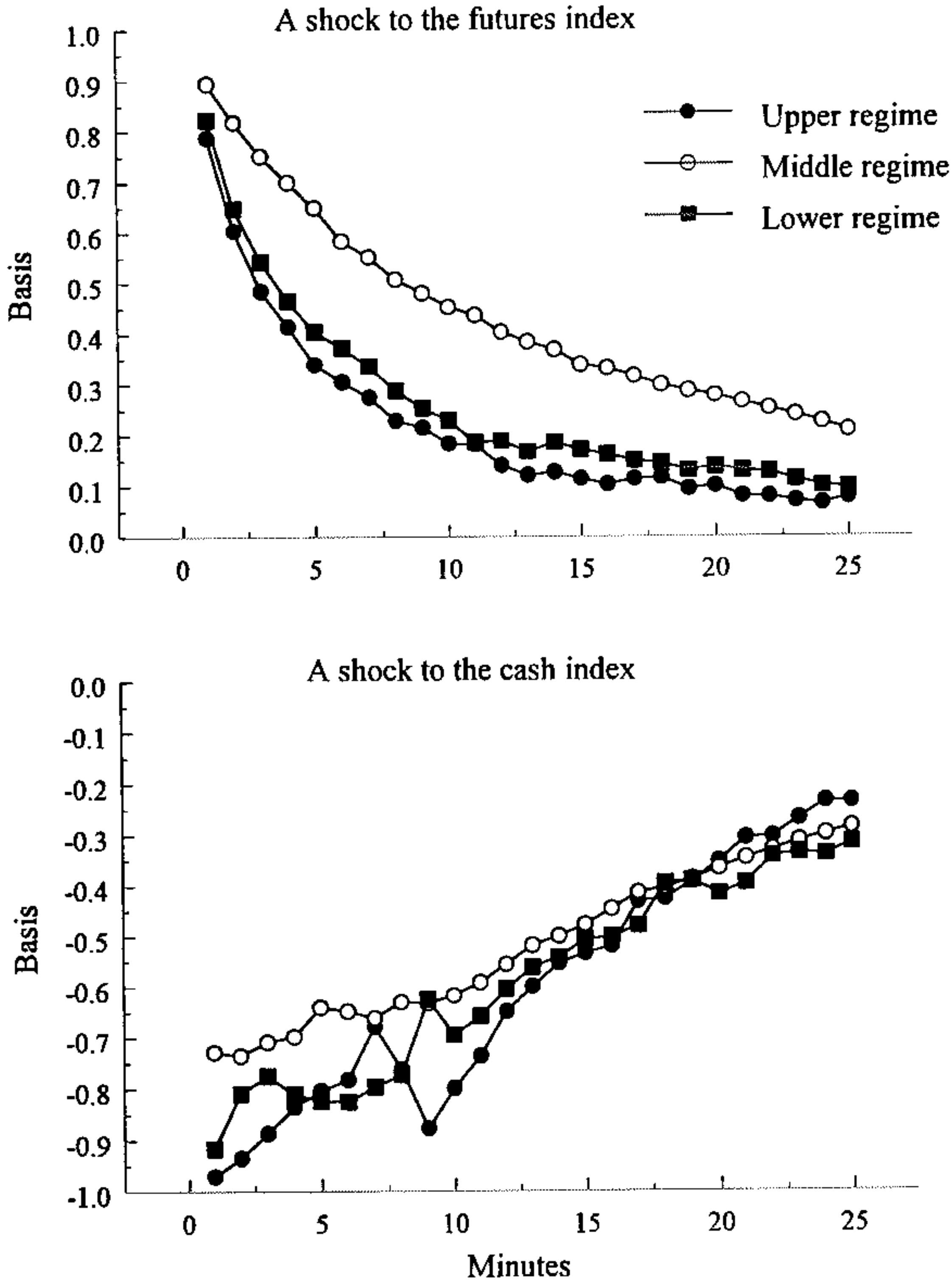
The vertical axis is the estimated delay and the horizontal axis is the set of contracts from inception through 1990. The estimates suggest shorter delays in the later years.

Figure 9
Estimated Delays
1989 and 1990



These are the estimated delays for 1989 and 1990 by contract and number of weeks to expiration. Neither this view of the relationships, nor any other, suggests to us a pattern by weeks to expiration or contract in 1989 and 1990.

Figure 10
Median of the Impulse Response Functions
1989 and 1990



These are medians of the impulse response functions for the "basis," the logarithm of the futures less the logarithm of the cash adjusted for daily means. The underlying responses are calculated from the TECM estimates by week for 1989 and 1990 for unit shocks. The responses are shown for shocks when the basis is more than the estimated threshold (the upper regime), when the basis is less than the negative of the estimated threshold (the lower regime) and when the basis is between the estimated thresholds (the middle regime). These impulse response functions do not include estimates of crossings from the tail regimes into the middle regime, which is part of the basis's actual response to a shock. The figure shows the estimated functions' behavior in the different regimes. In conjunction with the size of a shock, these functions determine the basis's actual response to a shock.