Statistical Properties of the Roll Serial Covariance Bid/Ask Spread Estimator
Author(s): Lawrence Harris
Source: *The Journal of Finance*, Vol. 45, No. 2 (Jun., 1990), pp. 579-590
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2328671
Accessed: 26/11/2009 07:17

THE JOURNAL OF FINANCE • VOL. XLV, NO. 2 • JUNE 1990

Statistical Properties of the Roll Serial Covariance Bid/Ask Spread Estimator

LAWRENCE HARRIS*

ABSTRACT

Exact small sample population moments of the standard serial covariance and variance estimators are derived under the assumptions of the Roll bid/ask spread model. Noise explains why serial covariance estimates are often positive in annual samples of daily and weekly returns. Small sample estimator bias partially explains why weekly estimates are more negative than daily estimates. Noise causes the Roll spread estimator to be severely biased by Jensen's inequality. The French-Roll adjusted variance estimator is unbiased but noisy. Empirical tests confirm the major implications.

THE ROLL (1984) SPREAD ESTIMATOR is seductively simple. Under ideal conditions, the effective bid/ask spread can be estimated by $2\sqrt{-SCov}$, where "$SCov$" is the first-order serial covariance of price changes. The ideal conditions assume that the underlying stock value (ignoring trading effects) follows a random walk, that buy and sell orders are equally probable and serially independent, and that underlying value is independent of the order flow.[1]

Unfortunately, the serial covariance estimator yields poor empirical results when used to estimate individual security spreads from a year of daily or weekly data. Estimated first-order serial covariances are positive for about half of all securities so that the square root in the estimator is not properly defined. Further, estimated spreads depend on the observation interval. In particular, daily estimates are smaller than weekly estimates. Both problems appear in Roll's empirical results.[2]

Several authors examine violations of the ideal conditions in attempts to explain these problems and to further understand price serial covariance. Choi, Salandro, and Shastri (1988) relax the assumption of serial independence of order type by assuming that the conditional probability of a switch from bid to

* Economic Fellow, Office of Economic Analysis, U.S. Securities and Exchange Commission, and Associate Professor of Finance and Business Economics, School of Business Administration, University of Southern California. The comments and opinions contained in this paper are those of the author only. In particular, the views here expressed do not necessarily reflect those of the U.S. Securities and Exchange Commission or of the author's colleagues on its staff. I first developed the ideas presented here for presentation to a Ph.D. Seminar in Empirical Methods in Finance that I led at the University of Southern California. I wish to thank my students, Carolyn C. Wu Chang and Suh-Pyng Ku, for their participation and patience. Any remaining errors are mine alone.

[1] If the estimator is applied to the serial covariances of returns instead of price changes, percentage spreads will be estimated. This study considers only price changes because all results can be restated with little difficulty in terms of returns.

[2] These results were replicated and extended for 1983–1986 with no significant differences. Similar results also appear in Ohlson and Penman (1985), Harris (1988), and Karpoff and Walkling (1988).

ask (or from ask to bid) is not equal to one-half. Bhattacharya (1986) allows some orders to be executed at the midpoint between the bid and ask prices. Glosten (1987) analyzes serial covariances assuming that the total spread is the sum of an adverse selection component and a gross-profit component. These studies identify potential biases in the Roll estimator, but none provides a completely adequate explanation of why serial covariances are so often positive or why the average weekly serial covariance is more negative than the average daily serial covariance.

The obvious (but not exclusive) explanation for the latter phenomenon is that the underlying value process is positively serially correlated. This proposition, which is recognized by all of the above authors including Roll, is explicitly considered by Amihud and Mendelson (1987), Richardson and Smith (1988), and Hasbrouck and Ho (1987).

This study derives the exact small sample population mean and variance of the serial covariance estimator under Roll's ideal conditions. The results show that the estimator is quite inaccurate in data measured at daily (and longer) intervals for sample sizes typically encountered in financial studies. Its expected error is so great that noise alone can explain why nearly half of the estimated serial covariances are positive. The results also show that the serial covariance estimator has a downward bias in small samples of data measured over long intervals. The bias is due to the estimator adjustment for the mean price change. It can partially explain why serial covariances estimated from weekly data are more negative than those estimated from daily data. Finally, Jensen's inequality severely biases downward individual stock spread estimates computed by taking the square root of serial covariance estimates obtained from daily or weekly data. The expected spread estimator is about half of the true spread in annual samples of daily or weekly data. The bias is due to the large estimator error caused by value innovation variation.

This study also derives the exact small sample population mean and variance of the standard variance estimator, and its covariance with the serial covariance estimator, under Roll's ideal conditions. These moments are used to characterize the small sample distribution of the French and Roll (1986) adjusted variance estimator ($Var + 2\ SCov$). The results show that the expected mean squared error of the adjusted estimator typically exceeds that of the unadjusted estimator in yearly samples of daily data.

The remainder of the paper is divided into three sections. Section I introduces the small sample population means and variances of the standard serial covariance and variance estimators and comments upon their significance for empirical studies. Section II presents additional empirical results obtained from actual stock market data that confirm the important implications of the model. The final section contains a summary and some additional perspectives.

I. Estimator Distributions under the Roll Bid/Ask Spread Model

A. Roll's Bid/Ask Spread Model

Roll's model can be stated formally as

$$V_t = V_{t-1} + e_t, \tag{1}$$

$$P_t = V_t + s/2\ Q_t \tag{2}$$

or more concisely as

$$\Delta P_t = s/2\ \Delta Q_t + e_t \tag{3}$$

where V_t is the underlying value at time t, e_t is a random innovation to underlying value, P_t is the actual observed price at time t, s is the total effective spread, and Q_t is a random indicator of whether the tth transaction took place at the bid or ask. The value innovation, e_t, has mean μ, standard deviation σ, and excess kurtosis η. The order flow indicator, Q_t, takes the values of -1 and 1 with equal probability. The distributions of $\{e_t\}$ and $\{Q_t\}$ are serially independent and independent of each other.

Under these assumptions, the implied variance and serial covariance of observed price changes are

$$\text{var}(\Delta P_t) = \sigma^2 + s^2/2, \tag{4}$$

$$\text{cov}(\Delta P_t, \Delta P_{t-1}) = -s^2/4. \tag{5}$$

Roll's effective bid/ask spread estimator is obtained by substituting a serial covariance estimate for the population serial covariance in (5) and solving for s.

B. The Serial Covariance Estimator

The standard serial covariance estimator used is the mean cross-product:

$$SCov = \frac{1}{n} \textstyle\sum_{t=1}^{n} \Delta P_t \Delta P_{t-1} - \overline{\Delta P}^2, \tag{6}$$

where $\overline{\Delta P}$ is the sample mean of $\{\Delta P_t\}$. The exact finite sample population mean and variance of this estimator are derived and presented in Harris (1989b). The results have terms of several orders of $1/n$. The zero and first-order terms are

$$E(SCov) \approx -s^2/4 - \sigma^2/n, \tag{7}$$

$$\text{var}(SCov) \approx (5/16\ s^4 + s^2\sigma^2 + \sigma^4)/n. \tag{8}$$

The population distribution of $SCov$ is asymptotically normal as n increases. For small n, the square of the sample mean in (6) causes the population distribution to be left-skewed because the sample mean is approximately normally distributed.

B.1. Serial Covariance Estimator Error

The ratio of the estimator mean to its standard deviation is a measure of how far the mean is from zero. It is approximately

$$t = E(SCov)/STD(SCov) \approx -\ (r^2 + 4/n)/((5r^2 + 16r^2 + 16)/n)^{1/2}, \tag{9}$$

where r is s/σ. A tabulation of the exact expression for t (Table I, Panel A) shows that serial covariance estimates will often be near zero when computed from a year of daily price observations on a stock with a daily return standard deviation of 2 percent and a spread equal to 1 percent of its price (\$1/8 on a \$12.5 stock; \$1/4 on a \$25 stock). The expected serial covariance estimate corresponding to $n = 252$ and $r = 0.5$ is only a fraction of -0.94 of its theoretical standard error. The prior probability of observing a positive serial covariance for this stock is 17.4

percent, assuming that the asymptotic normal sampling distribution is a good approximation to the actual finite sampling distribution (Table I, Panel B). A sample size of about 850 daily observations—more than three years—is necessary to be 95 percent confident that the serial covariance estimate for this stock is positive. When observations on prices are available every five minutes, only 15 observations are necessary.[3] If the bid/ask spread is only half as large, there is a 38.1 percent prior chance of observing positive serial covariances in daily data and more than 10,000 observations are necessary to be 95 percent confident of estimating a positive serial covariance.

The implied positive serial covariance estimate frequencies given by this model are of the same order of magnitude as those observed by Roll, but they differ substantially. Roll reports 51.4 percent positive in daily data and 34.8 percent positive in weekly data. Assuming that a typical stock has a daily return standard deviation of 2 percent and a spread equal to 1 percent of price, the corresponding model implications are 17.4 percent and 40.9 percent. They are 38.1 percent and 43.6 percent if the spread is 0.5 percent. Although it may be possible to account for some of the differences by more accurately characterizing the "typical stock" or by computing separately a prediction for each stock and then aggregating, the improvement cannot be very great. The population probability distribution implies that the frequency of positive serial correlations in the daily data is smaller than that observed in the weekly data, but Roll finds the opposite.

As noted in the introduction, one explanation may be that underlying values are positively serially correlated in daily data. Unfortunately, empirical efforts to separately identify this phenomenon and the bid/ask process will be frustrated by the large ratio of variation due to underlying value changes to variation due to bid/ask bounce in daily data.

The ratio of the variance of five-day returns (expressed as an average daily rate) to that of one-day returns provides some simple evidence of positively correlated underlying value changes. High values indicate positive serial correlation in the data. Roll's model implies that this variance ratio should be equal to $(5\sigma^2 + s^2/2)/5 \div (\sigma^2 + s^2/2) = (10 + r^2)/(10 + 5r^2)$, where r is s/σ. (The implied ratio is less than one because the bid/ask process induces first-order serial correlation into price changes.) This expression is equal to 0.911 for a stock with spread equal to one-half of its daily standard deviation, and 0.976 if the spread is one-fourth of the standard deviation. For 22 of the 25 years examined in this study (1963–1986), the cross-sectional mean variance ratio (computed from nonoverlapping returns) is significantly (t greater than 2) greater than 0.911. For 15 years, it is significantly greater than 0.976. Perhaps most notable, the correlation coefficient between the annual variance ratio mean and the mean annual difference between the daily and weekly positive serial covariance frequencies is

[3] This calculation assumes that the transaction data conform to the ideal conditions and that the underlying daily return variance is 78 (6.5 hours × 12 intervals per hour) times the five-minute variance. Given the Hasbrouck and Ho (1987) serial covariance results, evidence on jump processes in intraday data presented by Oldfield, Rogalski, and Jarrow (1977), and the fact that some of the daily return variance is due to overnight returns, neither assumption is particularly attractive.

Table I

Serial Covariance Estimator Mean/Standard Error Ratios and Implied Positive Frequencies

The serial covariance estimator is $SCov = \frac{1}{n} \sum_{t=1}^{n} \Delta P_t \Delta P_{t-1} - \overline{\Delta P}^2$, where P_t is the actual observed price at time t, $\overline{\Delta P}$ is the sample mean of (ΔP_t). The moments are obtained from the model: $\Delta P_t = s/2\, \Delta Q_t + e_t$, where e_t is a normally distributed random innovation to underlying value having standard deviation σ, s is the total effective spread, and Q_t is a random indicator of whether the transaction at t took place at the bid or ask. The order flow indicator, Q_t, takes the values of -1 and 1 with equal probability. The distributions of (e_t) and (Q_t) are serially independent and independent of each other. An approximate expression ratio of mean to the standard deviation of the estimator is given by $-(r^2 + 4/n)/(5r^2 + 16r^2 + 16)/n)^{1/2}$. The ratio computed from the exact small sample moments derived in Harris (1989b) is evaluated here.

	Number of Observations								
$r = s/\sigma$	10	25	50	100	250	500	1000	2500	5000
	Panel A: $t = E(SCov)/STD(SCov)$								
0	−0.35	−0.21	−0.14	−0.10	−0.06	−0.04	−0.03	−0.02	−0.01
0.005	−0.35	−0.21	−0.14	−0.10	−0.06	−0.04	−0.03	−0.02	−0.01
0.01	−0.35	−0.21	−0.14	−0.10	−0.06	−0.05	−0.03	−0.02	−0.02
0.025	−0.35	−0.21	−0.15	−0.10	−0.07	−0.05	−0.04	−0.03	−0.03
0.05	−0.35	−0.21	−0.15	−0.11	−0.07	−0.06	−0.05	−0.05	−0.06
0.1	−0.35	−0.22	−0.16	−0.13	−0.10	−0.10	−0.11	−0.14	−0.19
0.25	−0.39	−0.28	−0.25	−0.25	−0.30	−0.38	−0.51	−0.78	−1.09
0.5	−0.49	−0.47	−0.53	−0.65	−0.94	−1.28	−1.78	−2.79	−3.94
1	−0.77	−0.97	−1.27	−1.72	−2.65	−3.71	−5.22	−8.23	−11.64
2.5	−1.23	−1.83	−2.55	−3.57	−5.62	−7.93	−11.21	−17.72	−25.05
5	−1.40	−2.14	−2.99	−4.22	−6.65	−9.40	−13.29	−21.01	−29.71
	Panel B: Predicted Positive Serial Covariance Estimate Frequencies								
0	36.5%	41.8	44.3	46.0	47.5	48.2	48.7	49.2	49.4
0.005	36.5	41.8	44.3	46.0	47.5	48.2	48.7	49.2	49.4
0.01	36.5	41.8	44.3	46.0	47.5	48.2	48.7	49.2	49.4
0.025	36.5	41.7	44.2	45.9	47.4	48.1	48.5	48.9	49.0
0.05	36.4	41.7	44.1	45.7	47.1	47.7	48.0	48.0	47.7
0.1	36.3	41.3	43.6	45.0	45.9	46.0	45.6	44.3	42.5
0.25	35.0	39.0	40.2	40.1	38.1	35.1	30.5	21.9	13.9
0.5	31.2	31.9	29.9	25.8	17.5	10.0	3.7	0.3	0.0
1	22.2	16.5	10.2	4.3	0.4	0.0	0.0	0.0	0.0
2.5	10.9	3.3	0.5	0.0	0.0	0.0	0.0	0.0	0.0
5	8.1	1.6	0.1	0.0	0.0	0.0	0.0	0.0	0.0

0.725. These results all suggest that there is positive serial correlation in the underlying value innovation process.

B.2. Serial Covariance Estimator Bias

The small sample expression for the expected serial covariance found in equation (7) provides a partial explanation for why the mean weekly serial covariance estimate is more negative than the mean daily serial covariance estimate. The second term of this expression (the bias in the serial covariance estimator), $-\sigma^2/n$, is due to noise in the estimator of the mean price change. The mean estimator appears in the serial covariance estimator to adjust it to eliminate bias that results if μ, the mean value innovation, is not equal to zero.[4] Unlike the first term, this term depends on the observation interval length since σ^2 increases with the observation interval and since, for a given sample period, n decreases inversely with the observation interval. The average daily and weekly return variances in the 25-year sample are $9.25(\%^2)$ and $40.06(\%^2)$. Combining these figures gives an estimate of the implied difference in the mean daily and weekly serial covariance estimates. It is $-9.25(\%^2)/252 - -40.06(\%^2)/52 = 0.73(\%^2)$. This is about thirty percent of the actual mean difference of $2.45(\%^2)$. The remainder is probably due to positive serial correlation in the underlying value innovation process.

B.3. Biases in the Serial Covariance Spread Estimator

The large variance of the serial covariance estimator suggests that Jensen's inequality significantly biases individual spread estimates computed by taking the square root of serial covariance estimates obtained from daily and weekly data. The ratio of the expected value of Roll's spread estimator to the value of Roll's estimator evaluated at the expected value of the serial covariance estimator is tabulated in Table II, Panel A.[5] This ratio is less than one because noise in the covariance estimator causes a significant fraction of its distribution to range over negative values. The bias is virtually nonexistent in transaction data for which the ratio of the spread to the underlying value innovation standard deviation is large.

For a stock with spread equal to one-half of its standard deviation, the implied ratio of mean spread estimator to the square root transform of the mean serial covariance estimator is 0.78 for daily data and 0.41 for weekly data. (These ratios are 0.47 and 0.35 if the spread is one-fourth of the standard deviation.) The actual ratios of the cross-security mean spread estimate to the average spread estimate obtained by transforming the 25-year mean serial covariance estimates are 0.15 (0.23/1.56) for daily data and 0.47 (1.64/3.50) for weekly data. Both realized ratios are lower than the implied ratios. This is due to unaccounted-for cross-security variation in spreads since Jensen's inequality holds without regard to whether the source of variation in the serial covariance estimates is due to estimator error or cross-sectional variation. Also, the implied ratio would be smaller than indicated in Table II if the underlying value process were positively correlated, as appears to be true for daily data.

These results concerning the Jensen inequality suggest that the Roll estimator is not reliable for quantitative estimates of individual firm spreads from daily and weekly data. When an average spread estimate is desired, however, the Roll

[4] Researchers who need to estimate serial covariances from small samples of volatile data with means a priori known to be small may wish to consider whether the gains associated with the mean adjustment are worth the additional estimator noise and the new bias that the adjustment introduces. This question can be easily answered with the methods used in this study.

[5] Numeric integration over the asymptotic normal distribution of *SCov* is used to compute the expected spread estimator. The integral is computed by assigning negative spread values for positive serial covariances, as did Roll. Monte Carlo simulations (not presented) demonstrate that normality provides an adequate description of the estimator distribution for $n = 10$.

Table II

Biases in the Roll Serial Covariance Spread Estimator

The expected serial covariance spread estimator is computed numerically assuming that the serial covariance estimator, *SCov*, is normally distributed, with mean and variance given by the formulas derived in Harris (1989b) and given approximately by $E(SCov) \approx -s^2/4 - \sigma^2 n$ and $\text{var}(SCov) \approx (5/16\ s^4 + s^2\sigma^2 + \sigma^4)/n$, where s is the total effective spread, σ is the standard deviation of the underlying price change process, and n is the number of sample observations. The actual small sample distribution of *SCov* is slightly skewed to the left. If the skewness were modeled, the values in this table would be very slightly smaller. However, even for $n = 10$, simulation studies demonstrate that the normal approximation yields results accurate to at least two digits.

	Number of Observations								
$r = s/\sigma$	10	25	50	100	250	500	1000	2500	5000
	Panel A: Jensen's Inequality Bias								
	$E(Spread)/2\sqrt{-E(SCov)}$, where $Spread = \begin{cases} 2\sqrt{-SCov} & \text{if } SCov \le 0 \\ -2\sqrt{SCov} & \text{if } SCov > 0 \end{cases}$								
0	0.50	0.39	0.33	0.27	0.22	0.18	0.15	0.12	0.10
0.005	0.50	0.39	0.33	0.27	0.22	0.18	0.15	0.12	0.10
0.01	0.50	0.39	0.33	0.27	0.22	0.18	0.16	0.13	0.11
0.025	0.50	0.39	0.33	0.28	0.22	0.19	0.16	0.14	0.14
0.05	0.50	0.39	0.33	0.28	0.23	0.21	0.20	0.19	0.21
0.1	0.50	0.40	0.34	0.30	0.28	0.27	0.29	0.33	0.37
0.25	0.53	0.45	0.43	0.43	0.47	0.53	0.60	0.72	0.82
0.5	0.59	0.58	0.61	0.67	0.78	0.86	0.93	0.98	0.99
1	0.72	0.79	0.86	0.93	0.98	0.99	1.00	1.00	1.00
2.5	0.85	0.94	0.97	0.99	1.00	1.00	1.00	1.00	1.00
5	0.88	0.96	0.98	0.99	1.00	1.00	1.00	1.00	1.00
	Panel B: Total Bias in the Roll Serial Covariance Spread Estimator $E(Spread)/s$								
0.005	63.28	31.25	18.45	10.94	5.49	3.27	1.95	0.99	0.60
0.01	31.64	15.63	9.23	5.48	2.76	1.65	0.99	0.52	0.32
0.025	12.67	6.27	3.72	2.22	1.14	0.70	0.45	0.27	0.21
0.05	6.36	3.17	1.90	1.16	0.63	0.43	0.31	0.25	0.24
0.1	3.23	1.65	1.03	0.68	0.44	0.37	0.34	0.35	0.39
0.25	1.44	0.85	0.64	0.55	0.53	0.56	0.62	0.73	0.82
0.5	0.96	0.74	0.70	0.72	0.80	0.88	0.94	0.98	0.99
1	0.86	0.85	0.89	0.95	0.98	0.99	1.00	1.00	1.00
2.5	0.89	0.95	0.98	0.99	1.00	1.00	1.00	1.00	1.00
5	0.90	0.96	0.99	0.99	1.00	1.00	1.00	1.00	1.00

estimator may be useful if applied to the mean serial covariance estimate. This statistic is unbiased in a large cross-sectional sample if all stocks have equal spreads. Otherwise, it estimates the root mean squared spread, which is greater than the mean spread. Individual spread estimates may be used as noisy qualitative proxies for the spread, but it is not obvious why they would be better used for this purpose than untransformed serial covariance estimates.

The total bias in Roll's spread estimator is caused both by the bias in the serial covariance estimator and by the bias due to Jensen's inequality. The ratio of the expected spread estimator to the true spread is tabulated in Table II, Panel B.

When the spread is small relative to the value innovation standard deviation, the negative bias in the serial covariance estimator dominates and the spread is overestimated. For large values of r, the Jensen's inequality bias dominates and the spread is underestimated. The expected spread estimator in a year of daily data is 80 percent of the true spread for stock whose spread is one-half of its daily value innovation standard deviation ($r = 0.5$). This ratio is 66 percent in a year of weekly data and is 53 percent and 95 percent, respectively, if the stock spread is one-fourth of its daily standard deviation.

C. The French and Roll Adjusted Variance Estimator

French and Roll (1986) propose that first-order serial covariance estimates be used to remove trading effects from return variance estimates. Their estimator of the underlying return variation, $Var + 2\ SCov$, is unbiased in large samples, but the adjustment increases its sampling error. Since in some applications, such as option pricing for an individual security, accuracy may be more important than unbiasedness, it is important to understand the population distribution of the adjusted estimator.

The exact small sample population moments of the standard unadjusted variance estimator,

$$Var = \textstyle\sum_{t=1}^{n} (\Delta P_t - \overline{\Delta P})^2/(n-1), \tag{10}$$

are derived in Harris (1989b). The zero and first-order terms of its mean, variance, and covariance with $SCov$ are

$$E(Var) = \sigma^2 + s^2/2 + s^2/2n, \tag{11}$$

$$\mathrm{var}(Var) \approx (s^4/4 + 2s^2\sigma^2 + 2\sigma^4 + \eta\sigma^4)\ n/(n-1)^2, \tag{12}$$

$$\mathrm{cov}(SCov, Var) \approx -(s^4/4 + s^2\sigma^2)/(n-1). \tag{13}$$

Manipulation of these moments yields zero and first-order terms of the mean and variance of the French and Roll estimator:

$$E(Var + 2\ SCov) \approx \sigma^2 + (s^2/2 - 2\sigma^2)/n, \tag{14}$$

$$\mathrm{var}(Var + 2\ SCov) \approx (s^4/2 + 2s^2\sigma^2 + 6\sigma^4 + \eta\sigma^4)\ n/(n-1)^2. \tag{15}$$

The zero and first-order terms of the mean squared error of the adjusted and unadjusted variance estimators about the underlying population value innovation variance, σ^2, are

$$MSE(Var + 2\ SCov) \approx (s^4/2 + 2s^2\sigma^2 + 6\sigma^4 + \eta\sigma^4)\ n/(n-1)^2, \tag{16}$$

$$MSE(Var) \approx s^4/4\ n^2/(n-1)^2 + (s^4/4 + 2s^2\sigma^2 + 2\sigma^4 + \eta\sigma^4)\ n/(n-1)^2. \tag{17}$$

These imply that the unadjusted estimator mean squared error is less than that of the French and Roll estimator approximately when

$$n < 1 + 16/r^4. \tag{18}$$

For a stock with a spread equal to one-half of its daily value innovation standard deviation ($r = 0.5$), the unadjusted variance has a lower mean squared error for

n less than 257, or about a year of data. If the spread is half as large, equality is obtained for n equal to 4097, or about 16 years. In transaction data, the adjustment is much more important. The critical value of n in hourly data (for which the variance is assumed to be one-sixth of the daily variance) is only about 9.5 (using a higher order approximation) for a stock with $r = 0.5$.

Although the French and Roll adjusted variance estimator does not yield much increase in accuracy when applied to individual securities, the bias adjustment is important and appropriate if inference is to be made from an average of variances across stocks. In this application, the noise in the serial covariance estimator is reduced by averaging so that, in a large enough cross-sectional sample, the results will be reliable. Event studies of volatility and trading time variance studies such as conducted by French and Roll (1986) are examples of these applications.

II. Additional Empirical Tests

This section examines serial covariance estimates computed over varying sample sizes to demonstrate empirically that estimator error and bias depend on sample size. Three replications of the demonstration are undertaken. Each replication uses eight years of CRSP daily returns data to compute serial covariances over sample lengths equivalent to one, two, four, and eight years. Every stock on the CRSP tape that traded continuously over the entire eight-year sample period and that traded an average of at least 200 days per year is included in the sample.

The serial covariance estimates are constructed to insure that the results for the various sample lengths are as comparable as possible. The eight-year estimate is computed from the entire time series. The two four-year estimates are computed from the two nonoverlapping sets of every second first-order cross-product. The two-year and one-year estimates are likewise computed from the nonoverlapping sets of every fourth and every eighth first-order cross-product. This procedure insures that every serial covariance is computed over the entire sample period and that every security-day is included once and only once in the set of covariance estimates for a given sample length. In particular, this method minimizes variation among the various short sample length estimates for a given stock that would result if they were computed sequentially from a stock whose spread determinants changed over the eight-year sample.

This experimental design requires that the serial covariance estimator be slightly modified to properly model the mean return. This is accomplished by estimating the mean return from the average of the mean current return and the mean lagged return. Since these two sets of returns do not overlap (except for the eight-year estimates), this method produces more accurate means than would be produced from contiguous data. Accordingly, the small sample bias will be smaller than would otherwise be observed.

The results (Table III) confirm the theoretical implications of the Roll model. Standard deviations decrease monotonically as sample length increases for all sample periods in both daily and weekly data. The positive serial covariance frequencies decrease monotonically in the weekly data as sample length increases. In the daily data, they increase very slightly. This is probably because the true serial covariances in daily data are positive, or at least close to zero. Finally, the

Table III

Cross-Sectional Return Serial Covariance Estimate Distributions

A security is included in one of the three samples if it traded at least 200 times in each of the eight years spanned by the sample. For each security, one eight-year, two four-year, four two-year, and eight one-year serial covariance estimates are computed. The eight-year estimate is computed from the entire time series. The two four-year estimates are computed from the two nonoverlapping sets of every second first-order cross-product. The two-year and one-year estimates are likewise computed from the nonoverlapping sets of every fourth and every eighth first-order cross-product. This procedure ensures that every serial covariance is computed over the entire sample period and that every security-day is included once and only once in each of the means for the different sample lengths.

	Observation Interval and Sample Length in Years							
	Daily				Weekly			
Sample Period	1	2	4	8	1	2	4	8
Standard Deviation								
1963–1970	2.07	1.83	1.68	1.62	8.82	6.72	5.28	4.42
1971–1978	2.75	2.55	2.45	2.41	10.27	7.92	6.43	5.57
1979–1986	4.81	3.64	2.90	2.51	9.97	7.41	5.42	4.63
All	3.40	2.79	2.42	2.25	9.77	7.42	5.81	4.97
Positive Serial Covariance Frequency								
1963–1970	53.1	55.2	56.6	58.6	38.2	35.4	30.7	23.0
1971–1978	51.3	52.4	53.3	54.1	38.4	36.4	33.5	28.1
1979–1986	56.3	58.5	61.0	61.4	38.5	35.6	30.9	25.8
All	53.4	55.2	56.7	57.8	38.4	35.8	31.8	25.9
Mean Serial Covariance								
1963–1970	−0.30	−0.29	−0.29	−0.29	−2.51	−2.37	−2.29	−2.29
1971–1978	−0.67	−0.66	−0.66	−0.66	−2.82	−2.64	−2.54	−2.54
1979–1986	−0.27	−0.27	−0.26	−0.26	−1.87	−1.74	−1.65	−1.65
All	−0.43	−0.43	−0.42	−0.42	−2.43	−2.27	−2.18	−2.18
Number of Observations								
1963–1970	11168	5584	2792	1396	11168	5584	2792	1396
1971–1978	14542	7272	3636	1818	14547	7271	3634	1817
1979–1986	12000	6003	3002	1501	12021	6009	3004	1502
All	37710	1885	9430	4715	37736	18864	9430	4715

mean serial covariances in the daily and the weekly data all become monotonically less negative as sample size increases. The effect is strongest in the weekly data as the theory implies.

III. Conclusion

A. Summary

This paper presents the exact small sample moments of the serial covariance estimator and of the standard variance estimator for price changes generated by the Roll spread model. The serial covariance estimator is very noisy in daily data and is biased downward in small samples. The Jensen's inequality bias in Roll's spread estimator is very serious in daily and weekly data.

B. Additional Perspectives

Some problems associated with daily data serial covariance spread estimates may be due to unusual statistical properties of closing prices. Amihud and

Mendelson (1987) show that the serial correlation of closing prices differs from that of opening prices. Daily close-to-close returns have more negative serial correlation than open-to-open returns. Richardson and Smith (1988) generalize the result by showing that daily returns measured from different hours during the day display increasing (less negative) serial correlation the later in the day the returns are measured. These results are generally interpreted as evidence of increasing price resolution during the day. The serial covariance spread estimator may therefore estimate the noise in prices as well as the spread. Also, Harris (1989a), McInish and Wood (1989), Porter (1988), and Terry (1988) show that closing prices tend to be at the ask quote more frequently than earlier prices. This suggests that there may be less bid/ask bounce in close-to-close price changes than in day-to-day differences of earlier prices.

The results of this paper generally suggest that the serial covariance spread estimator is well behaved under ideal conditions in intraday data. Unfortunately, these ideal conditions may not exist. Order flows and/or underlying values innovations may be serially correlated (Hasbrouck and Ho (1987) and Choi, Salandro, and Shastri (1988)), orders may cross at spread midpoints (Bhattacharya (1986)), and price discreteness may induce negative serial covariance in price change series (Harris (1988)). These phenomena can all bias the serial covariance spread estimator.

Finally, although Roll's serial covariance spread estimator is noisy in daily data, it is very nearly the best serial covariance spread estimator available. Harris (1988) provides a full information maximum likelihood estimator of the spread parameter in Roll's model, but it yields little improvement over Roll's estimator. Intuition for this result can be developed by noting that, if the order flow indicator Q_t were normally distributed instead of discretely distributed on $[-1,1]$ and if the value innovation were also normally distributed, the sufficient (and therefore full information) statistic for estimating serial covariance would be the summed cross-products. Any gains in efficiency that are associated with the maximum likelihood method are due only to proper modeling of the distributional shape of Q_t. It is unlikely that this would be a great source of efficiency in daily data for which the fraction of total price variation due to changes in Q_t is small relative to value innovation variance.

REFERENCES

Amihud, Yakov and Haim Mendelson, 1987, Trading mechanisms and stock returns: An empirical investigation, *Journal of Finance* 42, 533–553.

Bhattacharya, Mihir, 1986, Direct tests of bid/ask models, Working Paper, Graduate School of Business, University of Michigan, Ann Arbor.

Choi, J. Y., Dan Salandro, and Kuldeep Shastri, 1988, On the estimation of bid-ask spreads: Theory and evidence, *Journal of Financial and Quantitative Analysis* 23, 219–230.

French, Kenneth R. and Richard Roll, 1986, Stock return variances: The arrival of information and reaction of traders, *Journal of Financial Economics* 17, 5–26.

Glosten, Lawrence, 1987, Components of the bid-ask spread and the statistical properties of transaction prices, *Journal of Finance* 42, 1293–1307.

Harris, Lawrence, 1988, Estimation of 'true' stock price variances and bid-ask spreads from discrete observations, Working Paper, University of Southern California, Los Angeles.

———, 1989a, A day-end transaction price anomaly, 1989a, *Journal of Financial and Quantitative Analysis* 24, 29–45.

———, 1989b, The power of the Roll serial covariance bid/ask spread estimator, Working Paper, University of Southern California, Los Angeles.

Hasbrouck, Joel and Thomas S. Y. Ho, 1987, Order arrival, quote behavior, and the return-generating process, *Journal of Finance* 42, 1035–1048.

Karpoff, Jon M. and Ralph Walkling, 1988, Short-term trading around ex-dividend days: Additional evidence, *Journal of Financial Economics* 21, 291–298.

McInish, Thomas H. and Robert A. Wood, 1988, An analysis of transactions data for the Toronto Stock Exchange: Return patterns and the end-of-day effect, Working Paper, University of Texas, Arlington.

Ohlson, James A. and Stephen H. Penman, 1985, Volatility increases subsequent to stock splits: An empirical aberration, *Journal of Financial Economics* 14, 251–266.

Oldfield, George S., Richard J. Rogalski, and Robert A. Jarrow, 1977, An autoregressive jump process for common stock returns, *Journal of Financial Economics* 5, 389–418.

Porter, David C., 1988, Trading within the spread: An examination of interday and intraday behavior, Working Paper, Marquette University, Milwaukee.

Richardson, Matthew and Tom Smith, 1988, On tests for serial dependency of returns, Working Paper, Stanford University, Stanford.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, 1127–1139.

Terry, Eric, 1988, End of the day returns and the bid-ask spread, Working Paper, Stanford University, Stanford.