MPRA

Munich Personal RePEc Archive

Implications of microstructure theory for empirical research in stock price behavior

Hawawini, Gabriel; Cohen, Kalman; Maier, Steven; Schwartz, Robert and Whitcomb, David

1980

Implications of Microstructure Theory for Empirical Research on Stock Price Behavior

Kalman J. Cohen; Gabriel A. Hawawini; Steven F. Maier; Robert A. Schwartz; David K. Whitcomb

The Journal of Finance is currently published by American Finance Association.

Implications of Microstructure Theory for Empirical Research on Stock Price Behavior

KALMAN J. COHEN, GABRIEL A. HAWAWINI, STEVEN F. MAIER,
ROBERT A. SCHWARTZ and DAVID K. WHITCOMB*

THE MAJOR BODY of literature in financial economics has assumed a frictionless trading process (much as Newtonian mechanics modeled the movement of point masses in a perfect vacuum); such analyses directly address the underlying processes in their pure form. However, to understand better the behavior of markets (or the flight of a feather from the Leaning Tower of Pisa), the effects of friction must be modeled.

The literature on security market microstructure treats the interplay between market participants, trading mechanisms, and the dynamic behavior of security prices in a regime where friction impedes the trading process. As such, it has implications for public policy concerning the design of a trading system, and for empirical research on security price behavior. In this paper, we deal with implications for empirical research.

Recently, several studies have focused on the fact that transaction prices differ from what they would otherwise be in a frictionless environment.[1] Oldfield-Rogalski-Jarrow's (1977) empirical evidence suggests that the arrival of transactions is best described by a (sporadic) jump process rather than by a diffusion process (such as Brownian motion). In Cohen-Maier-Schwartz-Whitcomb [CMSW] (1979b), we argue that such non-continuous trading accounts for spreads in markets composed of many traders posting bid and/or ask quotes. Goldman-Beja's (1979) model of the dynamic behavior of stock prices, based on a distinction between actual price and a theoretical, frictionless equilibrium price, yields implications concerning variance, correlation, and the role of the specialist. Goldman-Sosin (1979) model the manner in which the interaction of risk-neutral speculators and risk averse investors leads to the delayed impounding in market price of new information when information is not freely and instantaneously available. Also, Copeland (1976), Garman (1976), Scholes-Williams (1977), and CMSW (1979a) show that returns can be autocorrelated even though the underlying generation process is random. In addition, Scholes-Williams (1977), Dimson (1979), and Cohen-Hawawini-Maier-Schwartz-Whitcomb [CHMSW] (1979a, 1979b) have shown how delays in the trading process cause beta estimates to be biased.

* Cohen and Maier are at the Graduate School of Business Administration, Duke University; Hawawini at Baruch College, City University of New York; Schwartz at the Graduate School of Business Administration, New York University; and Whitcomb at the Graduate School of Business Administration, Rutgers University.
[1] Early recognition that the mechanics of the trading process affect stock price movements is in Demsetz's (1968) paper on market making and the bid-ask spread and in Fisher (1966).

We list below six interrelated empirical phenomena which have been reported in the literature.[2] Two elements of commonality among most of the phenomena are evident, an intervaling effect, and the impact of a security's "thinness" (or the inverse, its market value or value of trading). By and large, the latter has not been recognized in the literature. The phenomena are:

(1) Weak serial correlation in individual securities' daily returns, with the proportion of securities yielding significant autocorrelations decreasing as the differencing (i.e., returns measurement) interval increases, and with predominantly negative sign for thin securities and positive sign for "thick" (high value) securities.

(2) Positive serial cross-correlations between security returns and market index returns, with a lesser effect as the differencing interval is increased.[3]

(3) Positive serial correlation in market index returns, with the effect smallest for long differencing intervals and those indexes giving the least weight to thin securities.

(4) Autocorrelation of market model residuals which is weakly positive for daily data but which becomes predominantly negative as the differencing interval increases.

(5) Sensitivity of OLSE beta estimates to changes in the differencing interval, with thin security betas rising as the interval is lengthened and very high value security betas falling.

(6) Increase in market model R^2 as the differencing interval is lengthened, with the largest effect for thin securities.

Both purely statistical and microstructure explanations of some of the phenomena are reported in the literature.[4] The statistical approaches have focused on the effect of serial cross-correlation among security returns, while the microstructure studies have also been concerned with friction in the trading process. The discussion has been incomplete, however, in that it has dealt with only a few of the manifestations of friction, it has not attempted to explain all of the observed phenomena, and it has not developed a comprehensive analytical framework.

Following the microstructure approach, we show here that all six phenomena can be attributed to friction in the trading process causing a bid-ask spread and price-adjustment delays that differ systematically across securities. We first present a simple equation that formalizes the impact of price-adjustment delays and the spread on observed returns, and then consider why price-adjustment delays might exist. We next consider the relationship between price-adjustment delays and market size, and then bring together and extend several propositions which appear elsewhere [CMSW (1979a), CHMSW (1979a), and Hawawini (1980)] so as to provide a unified explanation of the phenomena reported above.

---

[2] The relevant citations include: Fisher (1966), Fama (1970), Altman-Jacquillat-Levasseur (1974), Pogue-Solnik (1974), Schwartz-Whitcomb (1974, 1977), Dimson'(1975), Hawawini (1977), Scholes-Williams (1977), Smith (1978), CHMSW (1979b), and Hawawini (1980).

[3] The intervaling effect in serial cross-correlation has not been observed directly, but ppers by Hawawini-Vora (1980) and Francis (1975) report, respectively, substantial lead/lagged beta estimates for daily data and inconsequential lead/lagged betas for monthly intervals. Since, given stationarity, $\beta_j(s) = \rho_{jM}(s)\sigma_j/\sigma_M$ for any s^{th} order, the finding stated above follows.

[4] See Hawawini (1977, 1980) for statistical explanations; the microstructure explanations are cited above.
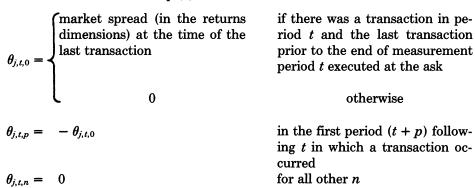
A. The Impact of Friction on Observed Returns

Defining returns as the natural logarithms of price relatives adjusted for dividends and splits, assume that in a frictionless world, a series of "true" returns, r_{jt}, would be generated for security j in successive discrete time periods t. With friction in the trading process, however, we can only obtain a series of "observed" returns, r_{jt}^0. We model the relationship between observed and true returns as:[5]

$$r_{jt}^0 = \sum_{n=0}^{N} \left(\gamma_{j,t-n,n} r_{j,t-n} + \theta_{j,t-n,n} \right) \qquad (1)$$

The random variable $\gamma_{j,t-n,n}$ represents the proportion of the true return of security j generated in period $t - i$ that is actually incorporated in observed returns in period t. Another way of interpreting the γs is that $\gamma_{j,t,0}, \gamma_{j,t,1}, \cdots, \gamma_{j,t,N}$ represent a delay distribution indicating how the true return generated in period t impacts on observed returns during period t and the next N future periods. It is assumed that there exists some finite number N, such that all of the true returns generated at or before period $t - N$ will have either made themselves felt or been forever lost by period t.

The random variable $\theta_{j,t-n,n}$ reflects the direct impact of the bid-ask spread on observed returns. Transaction prices bounce between the bid and the ask, so we arbitrarily take the bid as defining the base level price. The process can be described in the context of eq. (1) as:

$$
\theta_{j,t,0} = \begin{cases}
\begin{array}{l} \text{market spread (in the returns} \\ \text{dimensions) at the time of the} \\ \text{last transaction} \end{array} & \begin{array}{l} \text{if there was a transaction in pe-} \\ \text{riod } t \text{ and the last transaction} \\ \text{prior to the end of measurement} \\ \text{period } t \text{ executed at the ask} \end{array} \\
\\
\qquad\qquad 0 & \qquad\qquad \text{otherwise}
\end{cases}
$$

$$\theta_{j,t,p} = -\theta_{j,t,0} \qquad\qquad \begin{array}{l} \text{in the first period } (t + p) \text{ follow-} \\ \text{ing } t \text{ in which a transaction oc-} \\ \text{curred} \end{array}$$

$$\theta_{j,t,n} = 0 \qquad\qquad\qquad \text{for all other } n$$

B. Price Adjustment Delays

The γs in eq. (1) reflect the impact of various factors which we refer to as "price-adjustment delays." We define security j as having a "strictly greater expected price-adjustment delay" than security k if for all $l < N$ a smaller proportion of the true return generated for j in period t is expected to be reflected in j's observed returns during the interval for periods 0 to l than is the case for k. The various price-adjustment delays include:

(1) *Transaction price adjustments lag quotation price adjustments.* Returns are typically measured using prices of the last transactions that occurred before

---

[5] The specific set of conditions that the γs and θ in eq. (1) are assumed to satisfy is presented in the Appendix (available upon request).

the ends of successive measurement intervals of fixed length (e.g., a day). However, quotes can be updated without an accompanying transaction, and the "last transaction" generally occurs at a random point in time before the end of the trading day. Hence, closing transaction prices typically reflect out-of-date information. This is the price-adjustment delay cited by Scholes-Williams (1977) and CMSW (1979a).

(2) *Specialists/dealers impede quotation price adjustments.* Specialist intervention to make an "orderly" market by satisfying exchange continuity requirements directly impedes the adjustment of prices to a new equilibrium level. In addition, the models of Amihud-Mendelson (1979) and Ho-Stoll (1979) show that the quotes of any specialist/dealer will depend on his inventory position. Thus if a specialist/dealer develops an unbalanced inventory position by, e.g., being a net buyer, he will subsequently set his quotes to redress the imbalance by being, on expectation, a net seller. Hence, if the specialist/dealer accumulates inventory while handling the trades generated by news in one period, he will work off the imbalance in subsequent periods; price adjustments attributable to the initial period's news will continue until the specialist/dealer's inventory is rebalanced.

(3) *Quotation price adjustment lags for individual traders.* Given transaction costs, it is optimal for an investor or speculator to accumulate "news bits" for periodic review rather than continuously to assess the import of each news bit as it arrives; this has been demonstrated in Goldman-Sosin (1979). Further, when assessments are made, CMSW (1979b) have shown that with transaction costs, an investor may find a "do-nothing" strategy to be optimal even if a trade would have been sought in a frictionless environment. With an accumulation of news being required before a trade is sought, new orders transmitted to the market in part reflect the impact of "old" information. In addition, with continuous monitoring of the market being prohibitively expensive, limit orders left with specialists/dealers can go "stale" without being withdrawn. Thus, even if some investor were to submit his order immediately upon the receipt of information, if a stale limit order is hit, the transaction would occur at a price that reflects out-of-date information. It is also possible that, when any trade is desired, a large trader might feel his order would have an unduly large impact on price. For this reason, large traders sometimes break up their orders for execution over time.

Note that some of the price-adjustment delays might be relatively brief: transaction price changes lagging behind quotation price changes, and those caused by specialists seeking to make an orderly market, by stale limit orders being left on the book, and by some traders breaking up their large orders. But other delays might be quite protracted: those caused by specialists/dealers attempting to redress their inventory imbalances, and those caused by individuals seeking to trade only when information, decision, and transaction costs are sufficiently compensated for.

The impact of any price-adjustment delay will clearly diminish as the differencing interval is increased, when the differencing interval is greater than the delay. Since delays are finite, this suggests that the six empirical phenomena will be less apparent in longer differencing interval data, as indeed, they are. Alternatively stated, the price-adjustment delays introduce a pattern of auto- and

cross-serial correlations into the data and, with finite delays, these correlations eventually decay as the differencing interval is lengthened.[6] This comprehends most of the intervaling effect phenomena noted, and we amplify further only when necessary.

C. The Relationship Between Price-Adjustment Delays and Market Size

We expect security analysis to be more frequent and more intensive for larger issues.[7] When security analysis is undertaken more frequently, orders to trade will clearly reflect more up-to-date information (effectively, each news bit will remain in "inventory" for a shorter time before it is acted upon). The connection between the intensity of the analysis and the magnitude of price delays is less obvious. To the extent that security analysis generates useful information, more intensive analysis ought to increase the homogeneity of investor expectations; and we suggest that, with expectations being more nearly homogeneous, price-adjustment delays will be shorter. The reason is two-fold: (a) any trader who (ex post) would "agree" with the market assessment of security values would be less apt to seek a trade after he eventually does reassess his portfolio holding in a security, and (b) any given trade will have a smaller impact on security price the more elastic the market's demand for the security.[8,9]

D. The Correlation Structure in Returns

We next consider implications of the preceding discussion for the existence for non-zero autocorrelation and serial cross-correlation patterns in returns. In the Appendix we prove the following:[10]

[6] For a statistical analysis of the differencing interval-autocorrelation relationship, see Hawawini (1978).

[7] The reason is because any news bit that would have a given absolute impact measured in terms of percentage price impact would have a greater total monetary impact for securities of larger total market value.

[8] This is consistent with the argument we presented in CMSW (1978a) that the demand curve to hold a security is less elastic the more heterogeneous are expectations, and that the less elastic the demand curve, the greater will be the price impact of a buy or sell order of any given size.

[9] Another link between price-adjustment delays and market size is provided by the fact that the bid-ask spread for a security is negatively related to its trading activity [see CMSW (1979b)]. Hence, larger issues should have smaller spreads, and thus trades should occur more frequently in response to informational change since a new order is more apt to trigger a trade when the spread is small. This would keep transaction prices more up to date for the larger issues. It is unlikely for NYSE securities, however, that trading delays of this type would account for the correlation structure that appears to be present in the data.

[10] Proofs of this Proposition and Nos. 2–5 (below) using eq. (1) are given in the Appendix (available on request). Propositions 1–4 are also derived in CMSW (1979a) using a model which encompasses only those price-adjustment delays due to the fact that transactions lag quotation adjustments. Propositions 6 and 7 are proved in CHMSW (1979a) and two methods of adjusting beta for the intervaling effect bias are presented there. A model of the intervaling effect bias based on the transaction price lag and an "errors in variables" beta adjustment procedure are presented by Scholes-Williams (1977).

1. Given zero autocorrelation in the true returns series for a security, its observed returns may exhibit non-zero autocorrelation.

The $\theta_n s$ cause returns to be negatively serially correlated as transactions bounce randomly between the bid and the ask. The sign of serial correlation introduced by γ_n for $n > 0$ depends on their specific structure. Goldman-Sosin (1979) show that both overshooting and undershooting are possible as new information is impounded in market price; hence the γs may not be a monotonic decay function.

Goldman-Beja (1979) have demonstrated that non-clairvoyant specialists who attempt to absorb a fraction of all excess demand will introduce negative auto-correlation into observed returns (when there is nonzero drift in true returns), but that "over-interference" by such specialists could result in observed returns being positively serially correlated. As discussed in Schwartz-Whitcomb (1976), negative autocorrelation might also result from specialist intervention to satisfy a depth requirement that on the NYSE delimits the size of permissible price changes.

Positive serial correlation would result from specialist intervention to provide price continuity, from "clairvoyant" specialist absorption of the "random" component in current demand [see Goldman-Beja (1979)], from specialist/dealer inventory rebalancing, and from the breaking up of large orders. In addition, delays in updating limit orders on the book can lead to positive serial correlation as those traders who respond quickly enough to new information have an opportunity to hit orders that have not yet been revised (and in so doing, to generate a trade at an "intermediate" price).[11]

Thus, while Proposition 1 predicts non-zero autocorrelation, the sign and relationship to a security's thinness are indeterminate. The empirical evidence appears to suggest a positive relationship between autocorrelation and market size, with thinner issues exhibiting predominantly negative autocorrelation, and the thicker issues exhibiting predominantly positive autocorrelation.

We next consider how price-adjustment delays result in serial cross-correlations between security returns, which in turn generates autocorrelation in market indexes and biased estimates of beta. With price-adjustment delays that differ across securities, observed returns will not reflect properly matched responses to change in the information set. We expect mismatching because, as discussed above, price-adjustment lags should be systematically greater for thinner issues.

For asynchronous data, and assuming that observed returns are related to true returns according to eq. (1) and that true returns for each security are generated by

$$r_{jt} = \alpha_j + \beta_j r_{Mt} + e_{jt} \qquad (2)$$

---

[11] One might further question whether a succession of trades coming in over time due to delayed decision making would introduce serial correlation (of first or higher order) into returns. This would not necessarily be the case since late assessers will be late traders only if they disagree with the market's reassessment to that point, and if they do disagree, they could with equal probability believe the market has over- or under-adjusted. That is, we know of no reason for expecting an association between a trader's optimism and when he assesses, and therefore presume that early traders make an unbiased assessment of news.

with zero autocorrelation in the r_{Mt} and e_{jt}, and zero cross-correlation between the e_{jt} and $e_{k\tau}$ for all t,[1] the following propositions can be demonstrated:

2. The first-order serial cross-correlation in the observed returns of securities j and k is nonzero and has the sign of $\beta_j \beta_k$.

3. A market index constructed from observed returns for a large number of securities will have positive autocorrelation.

4. A value-weighted market index constructed from observed returns will have smaller autocorrelation than a similarly constructed equally-weighted market index.

5. For observed returns of security j, autocorrelation of market model residuals is non-zero with sign and magnitude depending on the signs and magnitudes of autocorrelation of j's returns, autocorrelation of market index returns, and contemporaneous and serial cross-correlations between j's returns and those of the market index.

6. The OLSE beta calculated from observed returns for a finite differencing interval asymptotically approaches true beta as the length of the differencing interval is increased without bound.

7. For any security with positive true beta, and for any finite differencing interval, the greater (less) is the expected trading delay of the security relative to the weighted average trading delay in the market index, the more will the OLSE beta underestimate (overestimate) true beta.

Proposition 2 is consistent with phenomenon 2. When k is the market, $\beta_k = 1$ and the first order serial cross-correlations (lead and lag) between the returns of security j and of the market will be nonzero with sign of β_j. Since most securities have positive betas, first order serial cross-correlation coefficients in observed returns between securities and the market index should be positive.

Proposition 3 predicts phenomenon 3. Index autocorrelation is shown to be a function of the autocorrelation of its component securities as well as the cross-serial correlations between these securities' returns. For an index made up of a large number of securities, positive serial cross-correlations will dominate the individual securities' autocorrelations, thereby inducing positive autocorrelation in the index regardless of the signs of the individual security autocorrelations.

Proposition 4 considers the effect of relative thinness on market index autocorrelation. A value weighted index will give more weight to thicker issues. Since these issues display weaker serial cross-correlations among themselves (due to smaller price-adjustment delays), a value weighted index will exhibit smaller positive autocorrelation than a similarly constructed equally weighted market index.

Proposition 5 sheds light on phenomenon 4. It is demonstrated that unless security autocorrelation, market index autocorrelation, and serial cross-correlations are all zero, market model residuals will be autocorrelated.

Propositions 6 and 7 explain the intervaling and thinness effects on estimated OLSE beta and its resulting bias. It is demonstrated that, as the differencing interval is lengthened, issues that are thinner (thicker) than the market average will have OLSE betas that rise (fall): thinner (thicker) issues approach their true

betas asymptotically from below (above), implying that their OLSE betas are underestimated (overestimated).

Finally, consider phenomenon 6. Hawawini (1980) shows that the sign of the intervaling effect on R^2 is equal to the sign of the intervaling effect on beta plus the difference between market index autocorrelation and security autocorrelation. For thick issues the sign of the former is negative (Propositions 6 and 7) and the sign of the latter is generally positive (market index autocorrelation is positive and security autocorrelation is generally smaller than that of the market index). For thin issues, both signs are positive and the sign of the intervaling effect on R^2 will be positive. For thicker issues, the sign becomes ambigious.

E. Conclusion

A growing literature has focused on serial correlation in returns and on bias in the measurement of a security's systematic risk. We have shown how these phenomena may be attributed to friction which results in price-adjustment lags and the bid-ask spread. A simple equation relating observed to true returns provides the framework for a comprehensive explanation of the observed phenomena. The analysis has also considered how the magnitudes of the various effects are related to a security's market value, and to the length of the differencing interval over which returns are measured.

Friction in the trading process clearly has an intricate and pervasive impact on the returns generation process. We expect that continuing developments in the microstructure literature will shed further light on the issue.

REFERENCES

Altman, E., B. Jacquillat, and M. Levasseur, "Comparative Market Model Analysis: France and the United States," *Journal of Finance*, 29 (December, 1974).
Amihud, Y. and H. Mendelson, "A Dealership Market: . . ." "Dealership Market: Market-Making with Inventory," *Journal of Financial Economics*, (March, 1980), in Press.
Cohen, K., G. Hawawini, S. Maier, R. Schwartz, and D. Whitcomb, "Friction in the Trading Process and the Estimation of Systematic Risk," Working Paper, 1979a.
———, "Estimating and Adjusting for the Delayed Price-Adjustment Bias in Beta," Working Paper, 1979b.
Cohen, K., S. Maier, R. Schwartz, and D. Whitcomb, "The Returns Generation Process, Returns Variance, and The Effect of Thinness in Securities Markets," *Journal of Finance*, 33 (March, 1978a), 149–167.
———, "Limit Orders, Market Structure, and the Returns Generation Process," *Journal of Finance*, 33 (June, 1978b), 723–738.
———, "On the Existence of Serial Correlation in an Efficient Securities Market," *TIMS Studies in the Management Sciehces*, 11, (1979a), 151–168.
———, "Transaction Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread," Working Paper, 1979b.
Copeland, T., "A Model of Asset Trading Under the Assumption of Sequential Information Arrival," *Journal of Finance*, 31 (September, 1976), 1149–1168.
Demsetz, H., "The Cost of Transacting," *Quarterly Journal of Economics*, 82 (February, 1968), 33–53.
Dimson, E., "Risk Measurement When Shares are Subject to Infrequent Trading," *Journal of Financial Economics*, 7 (June, 1979), 197–226.

Fama, E., "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25 (May, 1970), 383–417.

Fisher, L., "Some New Stock Market Indexes," *Journal of Business*, 39 (January, 1966), 191–225.

Francis, J., "Intertemporal Differences in Systematic Stock Price Movements," *Journal of Financial and Quantitative Analysis*, 9 (June, 1975), 205–220.

Garman, M. "Market Microstructure," *Journal of Financial Economics*, 3 (June, 1976), 33–53.

Goldman, M. and H. Sosin, "Information Dissemination, Market Efficiency and the Frequency of Transactions," *Journal of Financial Economics*, 7 (March, 1979), 29–61.

Goldman, M. and A. Beja, "Market Prices vs. Equilibrium Prices: Returns Variance, Serial Correlation and the Role of the Specialist," *Journal of Finance*, 34 (June, 1979), 595–607.

Hawawini, G., "On the Time Behavior of Financial Parameters: An Investigation of the Intervaling Effect," Ph.D. dissertation, New York University, 1977.

——, "A Note on Temporal Aggregation and Serial Correlation," *Economic Letters*, 1 (1978), 237–242.

——, "Intertemporal Cross Dependence in Securities Daily Returns and the Short-Run Intervaling Effect on Beta," *Journal of Financial and Quantitative Analysis*, 14 (March 1980).

——, and A. Vora, "Evidence of Intertemporal Systematic Risks in the Daily Price Movements of NYSE and AMEX Common Stocks," *Journal of Financial and Quantitative Analysis*, 14 (June 1980).

Ho, T. and H. Stoll, "Optimal Dealer Pricing Under Transactions and Return Uncertainty," Working Paper, 1979.

Oldfield, G., R. Rogalski, and R. Jarrow, "An Autoregressive Jump Process for Common Stock Returns," *Journal of Financial Economics*, 5 (December, 1977), 389–418.

Pogue, G. and B. Solnik, "The Market Model Applied to European Common Stocks: Some Empirical Results," *Journal of Financial and Quantitative Analysis*, 9 (December, 1974).

Scholes, M. and J. Williams, "Estimating Betas from Nonsynchronous Data," *Journal of Financial Economics*, 5 (December, 1977), 309–327.

Schwartz, R. and D. Whitcomb, "Can Betas Be Estimated from Daily Data? Presented at the *Proceedings of the Seminar on the Analysis of Security Prices*, 19 (May, 1974), Center for Research in Security Prices, University of Chicago.

——, "Comment: Assessing the Impact of Stock Exchange Specialists on Stock Volatility," *Journal of Financial and Quantitative Analysis*, 12 (December, 1976).

——, "The Time-Variance Relationship: Evidence on Autocorrelation in Common Stock Returns," *Journal of Finance*, 32 (March, 1977a) 41–55.

——, "Evidence on the Presence and Causes of Serial Correlation in Market Model Residuals," *Journal of Financial and Quantitative Analysis*, 11 (June, 1977b).

Smidt, S., "Continuous vs. Intermittent Trading on Auction Markets," Working Paper, 1979.

Smith, K., "The Effect of Intervaling on Estimating Parameters of the Capital Asset Pricing Model," *Journal of Financial and Quantitative Analysis*, 13 (1978), 313–344.