AN ORDERED PROBIT ANALYSIS OF TRANSACTION STOCK PRICES

by

Jerry Hausman, Andrew W. Lo, and A. Craig MacKinlay

Latest Revision: January 1991

Working Paper No. 3234–90–EFA

AN ORDERED PROBIT ANALYSIS OF TRANSACTION STOCK PRICES

Jerry A. Hausman*, Andrew W. Lo**, and A. Craig MacKinlay***

First Draft: May 1990
Latest Revision: January 1991

We estimate the conditional distribution of trade-to-trade price changes using *ordered probit*, a statistical model for discrete dependent variables that possess a natural ordering. Such an approach takes into account the fact that transaction price changes occur in discrete increments, typically eighths of a dollar, and occur at irregularly spaced time intervals. Unlike existing continuous-time/discrete-state models of transaction prices, ordered probit can capture the effects of other economic variables on price changes, such as volume, past price changes, and the time between trades. Using 1988 transactions data for ten randomly chosen U.S. stocks, we estimate the ordered probit model via maximum likelihood and use the parameter estimates to measure several transaction-related quantities, such as the price impact of trades of a given size, the tendency towards price reversals from one transaction to the next, and the empirical significance of price discreteness.

*Department of Economics, Massachusetts Institute of Technology.
**Sloan School of Management, Massachusetts Institute of Technology.
***Wharton School, University of Pennsylvania.

We thank Sarah Fisher, Ayman Hindy, and John Simpson for research assistance, and seminar participants at Harvard, London Business School, M.I.T., and Larry Harris and Whitney Newey for helpful comments. Research support from the Batterymarch Fellowship, the Geewax-Terker Investments Research Fund, the National Science Foundation (SES–8618769, SES–8821583), and the Q Group is gratefully acknowledged.

1. Introduction.

Common to virtually all empirical investigations of the microstructure of securities markets is the need for a statistical model of asset prices that can capture the salient features of price movements from one transaction to the next. For example, because there are several theories of why bid/ask spreads exist, a stochastic model for prices is a prerequisite to empirically decomposing observed spreads into components due to order-processing costs, adverse selection, and specialist market power.[1] The benefits and costs of particular aspects of a market's microstructure, such as margin requirements, the degree of competition faced by dealers, the frequency that orders are cleared, and intraday volatility also depend intimately on the particular specification of price dynamics.[2] In fact, it is difficult to imagine an economically relevant feature of the microstructure problem that does *not* hinge on such price dynamics.

Since stock prices are perhaps the most closely watched economic variables to date, they have been modeled by many competing specifications, beginning with the simple random walk or Brownian motion. The majority of such specifications have been unable to capture at least three aspects of *transactions* prices. First, on most U.S. stock exchanges prices are quoted in increments of eighths of a dollar, a feature not captured by stochastic processes with continuous state spaces. Of course, discreteness is less problematic for coarser-sampled data, which may be well-approximated by a continuous-state process. But discreteness is of paramount importance for intra-daily price movements, since such finely-sampled price changes may take on only five or six distinct values.[3]

Second, another distinguishing feature of transaction prices is their timing, which is irregular and random. Therefore, such prices may be modeled by discrete-time processes only if we are prepared to ignore the information contained in waiting-times for transactions.

Finally, although many have computed correlations between transaction price changes and other economic variables, to date none of the existing models for transaction prices have been able to quantify such effects formally. Such models have focused primarily on the *unconditional* distribution of price changes, whereas what is often of more interest is the *conditional* distribution, conditioned on economic quantities such as volume, time

[1] See, for example, Glosten and Harris (1988), Hasbrouck (1988), Roll (1984), and Stoll (1989).

[2] See Cohen et al. (1986), Harris, Sofianos, and Shapiro (1990), Hasbrouck (1989a), Madhavan and Smidt (1990), and Stoll and Whaley (1989).

[3] The implications discreteness has been considered in many studies. See, for example, Cho and Frees (1988), Gottlieb and Kalay (1985), Harris (1987, 1989a,b), and Petersen (1986).

between trades, and the *sequence* of past price changes. For example, one of the unanswered empirical questions in this literature is what the total costs of immediate execution are, which many take to be a measure of market liquidity. Perhaps the largest component of such costs is the price impact of large trades. Indeed, a floor broker seeking to unload 100,000 shares of stock will generally break up the sale into smaller blocks to minimize the price impact of the trades. How do we measure price impact? Such a question is a question about the conditional distribution of price changes, conditional upon a particular sequence of volume and price changes (i.e. order flow).

In this paper, we propose a specification of transaction price changes that addresses all three of these issues, and yet is still tractable enough to permit estimation via standard techniques. This specification is known as ***ordered probit***, which has been used most frequently in cross-sectional studies of dependent variables that are limited to a finite number of values possessing a natural ordering.[4] Heuristically, ordered probit analysis is a generalization of the linear regression model to cases where the dependent variable is discrete. As such, among the existing models of stock price discreteness,[5] ordered probit is perhaps the only specification that can easily capture the impact of "explanatory" variables on price changes while also accounting for price discreteness and irregular trade times.

Underlying the analysis is a "virtual" regression model with an unobserved continuous dependent variable Z^* whose conditional mean is a linear function of observed "explanatory" variables. Although Z^* is unobserved, it is related to an observable discrete random variable Z, whose realizations are determined by where Z^* lies in its domain or state space. By partitioning the state space into a finite number of distinct regions, Z may be viewed as an indicator function for Z^* over these regions. For example, a discrete random variable Z taking on the values $\{ -\frac{1}{8} , 0 , \frac{1}{8} \}$ may be modeled as an indicator variable that takes on the value $-\frac{1}{8}$ whenever $Z^* \leq \alpha_1$, the value 0 whenever $\alpha_1 < Z^* \leq \alpha_2$, and the value $\frac{1}{8}$ whenever $Z^* > \alpha_2$. Ordered probit analysis consists of estimating α_1, α_2 and the coefficients of the unobserved regression model for Z^*.

Since α_1, α_2 and Z^* may depend on a vector of "regressors" X, ordered probit analysis is considerably more general than its simple structure suggests. In fact, it is well known that ordered probit can fit any arbitrary multinomial distribution. However, because of the underlying linear regression framework, ordered probit can also capture the price effects of

[4] For example, the dependent variable might be the level of education, as measured by three categories: less than high school, high school, and college education. The dependent variable is discrete, and is naturally ordered since college education always follows high school. See Maddala (1983) for further details.

[5] See, for example, Ball (1988), Cho and Frees (1988), Gottlieb and Kalay (1985), and Harris (1987).

many economic variables in a way that models of the unconditional distribution of price changes cannot.

Using ordered probit analysis we investigate several issues specific to transaction prices. First, how does the particular sequence of trades affect the conditional distribution of price changes, and how do these effects differ across stocks? For example, does a sequence of three consecutive buyer-initiated trades ("buys") generate price pressure, so that the next price change is more likely to be positive than if the sequence were three consecutive seller-initiated trades ("sells"), and how does this pressure change from stock to stock? Second, does trade size affect price changes as some theories suggest, and if so, what is the price impact per unit volume of trade from one transaction to the next? Third, does price discreteness matter? In particular, can the conditional distribution of price changes be modeled as a simple linear regression of price changes on explanatory variables without accounting for discreteness?

Using 1988 transactions data from the Institute for the Study of Securities Markets (ISSM) for ten randomly chosen U.S. stocks, we find that the sequence of trades does affect the conditional distribution for price changes, and the effect is greater for larger capitalization and more actively traded securities. Moreover, trade size is also an important factor in the conditional distribution of price changes, with larger trades creating more price pressure, but in a nonlinear fashion. The price impact of a trade depends critically on the *sequence* of past price changes and order flows (buy/sell/buy versus buy/buy/buy). The ordered probit framework allows us to compare the price impact of trading over many different market scenarios, such as trading "with" versus "against" the market, trading in "up and down" markets, etc.. Finally, we show that discreteness does matter, in the sense that the simpler linear regression analysis of price changes cannot capture all the features of transaction price changes evident in the ordered probit estimates, such as the clustering of price changes on even eighths.

In Section 2 we review the ordered probit model, provide a few illustrative examples of its virtuosity, and describe its estimation via maximum likelihood. We describe the data in Section 3 by presenting some summary statistics for our sample of ten securities. In Section 4 we discuss the empirical specification and selection of conditioning or "explanatory" variables. We report reports the maximum likelihood estimates for our sample in Section 5 and we use these parameter estimates to address the three issues mentioned above. We conclude in Section 6.

2. The Ordered Probit Model.

Consider a sequence of transaction prices $P(t_0)$, $P(t_1)$, $P(t_2)$, ..., $P(t_n)$ observed at times t_0, t_1, t_2, ..., t_n, and denote by Z_1, Z_2, ..., Z_n the corresponding price changes, where $Z_k \equiv P(t_k) - P(t_{k-1})$ is assumed to be integer multiples of some divisor called "ticks" (such as an eighth of a dollar). Let Z_k^* denote an unobservable continuous random variable such that:

$$Z_k^* = X_k'\beta + \epsilon_k \quad , \quad E[\epsilon_k|X_k] = 0 \quad , \quad \epsilon_k \text{ i.n.i.d. } N(0,\sigma_k^2) \qquad (2.1)$$

where the term "i.n.i.d." indicates that the $epsilon_k$'s are independently but *not* identically distributed, and X_k is a qx1 vector of predetermined variables that governs the conditional mean of Z_k^*. Note that subscripts are used to denote "transaction" time, whereas time arguments t_k denote calendar or "clock" time, a convention we shall follow throughout.

The essence of the ordered probit model is the assumption that observed price changes Z_k are related to the continuous variable Z_k^* in the following manner:

$$Z_k = \begin{cases} s_1 & \text{if } Z_k^* \in A_{1k} \\ s_2 & \text{if } Z_k^* \in A_{2k} \\ \vdots & \vdots \\ s_m & \text{if } Z_k^* \in A_{mk} \end{cases} \qquad (2.2)$$

where the sets A_{jk} form a *partition* of the state space S^* of Z_k^* (i.e., $S^* = \bigcup_{j=1}^m A_{jk}$, and $A_{ik} \cap A_{jk} = \emptyset$ for $i \neq j$), and the s_j's are the discrete states that comprise the state space S of Z_k. In our application the s_j's are 0, $-\frac{1}{8}$, $+\frac{1}{8}$, $-\frac{2}{8}$, $+\frac{2}{8}$, and so on, and for simplicity we define the state-space partition of S^* to be intervals:

$$A_{1k} \equiv (-\infty \, , \, \alpha_{1k}] \qquad (2.3)$$

$$A_{2k} \equiv (\alpha_{1k} \, , \, \alpha_{2k}] \qquad (2.4)$$

$$\vdots$$

$$A_{ik} \equiv (\alpha_{i-1k} , \alpha_{ik}] \quad (2.5)$$

$$\vdots$$

$$A_{mk} \equiv (\alpha_{m-1k} , \infty) \quad (2.6)$$

where the partition boundaries α_{jk} may also depend on X_k.

Although the observed price change can be any number of ticks, positive or negative, to limit the number of parameters we assume that m in (2.2) is finite. This poses no problems since we may always let some states in S represent a multiple (and possibly uncountable) number of values for the observed price change. For example, in our empirical application we define s_1 to be a price change of -4 ticks *or less*, s_9 to be a price change of $+4$ ticks *or more*, and s_2 to s_8 to be price changes of -3 ticks to $+3$ ticks respectively. This parsimony is obtained at the cost of losing "price resolution" – the ordered probit model does not distinguish between price changes of $+4$ and price changes greater than $+4$ (since the $+4$-tick outcome and the greater than $+4$-tick outcome have been grouped into a common event), and similarly for price changes of -4 ticks versus price changes less than -4. This, however, is rarely a problem in practice since the resolution may be made arbitrarily finer by simply introducing more states, i.e., by increasing m. Therefore, the loss in resolution from a finite m may be made negligible at the cost of computational complexity.[6]

Observe that the ϵ_k's in (2.1) are assumed to be conditionally independently but *not* identically distributed.[7] This allows for clock-time effects, as in the case of an arithmetic Brownian motion where the variance σ_k^2 of price changes is linear in the time between trades. We also allow for conditional heteroskedasticity by letting σ_k^2 depend linearly on other economic variables. The dependence structure of the observed process Z_k is clearly induced by that of Z_k^* and the definitions of the A_{jk}'s, since:

$$\mathrm{P}(Z_k = s_j | Z_{k-1} = s_i) \;=\; \mathrm{P}(Z_k^* \in A_{jk} | Z_{k-1}^* \in A_{ik-1}) \,. \quad (2.7)$$

As a consequence, if the regressors X_k and the partitions A_{ik} are temporally independent, the observed process Z_k is also temporally independent. Of course, these are fairly restrictive assumptions that amount to requiring prices to follow random walks, and are certainly

[6]Moreover, as long as (2.1) is correctly specified, then increasing price resolution will not affect the estimated β's asymptotically. Of course, finite sample properties may differ.

[7]Conditional on the X_k's and other economic quantities influencing the conditional variance σ_k^2. Unless explicitly stated otherwise, all the probabilities we deal with in this study are conditional probabilities, and all inferences and statements concerning these probabilities are conditional, conditioned on these variables.

not necessary for any of the statistical inferences that follow. We require only that the ϵ_k's be conditionally independent, so that all serial dependence is captured by the X_k's. Consequently, the independence of the ϵ_k's does not imply that the Z_k^*'s are independently distributed because we have placed no restrictions on the temporal dependence of the X_k's.

The conditional distribution of observed price changes Z_k, conditioned on the regressors X_k, is determined by the partition boundaries and the particular distribution of ϵ_k. For Gaussian ϵ_k's, the conditional distribution is:

$$P(Z_k = s_i | X_k) \quad = \quad P(Z_k^* \in A_{ik} | X_k) \quad = \quad P(\ X_k'\beta + \epsilon_k \in A_{ik} \mid X_k\ ) \tag{2.8}$$

$$= \begin{cases} P(\ X_k'\beta + \epsilon_k \leq \alpha_{1k} \mid X_k\ ) & \text{if } i = 1 \\ \\ P(\ \alpha_{i-1k} < X_k'\beta + \epsilon_k \leq \alpha_{ik} \mid X_k\ ) & \text{if } 1 < i < m \\ \\ P(\ \alpha_{m-1k} < X_k'\beta + \epsilon_k \mid X_k\ ) & \text{if } i = m \end{cases} \tag{2.9}$$

$$= \begin{cases} \Phi\left(\frac{\alpha_{1k} - X_k'\beta}{\sigma_k}\right) & \text{if } i = 1 \\ \\ \Phi\left(\frac{\alpha_{ik} - X_k'\beta}{\sigma_k}\right) - \Phi\left(\frac{\alpha_{i-1k} - X_k'\beta}{\sigma_k}\right) & \text{if } 1 < i < m \\ \\ 1 - \Phi\left(\frac{\alpha_{m-1k} - X_k'\beta}{\sigma_k}\right) & \text{if } i = m \end{cases} \quad . \tag{2.10}$$

To develop some intuition for the ordered probit model, observe that the probability of any particular observed price change is determined by where the conditional mean lies relative to the partition boundaries. Therefore, for a given conditional mean $X_k'\beta$, shifting the boundaries will alter the probabilities of observing each state [see Figure 1]. In fact, by shifting the boundaries appropriately, ordered probit can fit any arbitrary multinomial distribution. This implies that the assumption of normality underlying ordered probit plays no special role in determining the probabilities of states – a logistic distribution, for example, could have served equally well.[8]

[8] However, it is considerably more difficult to capture conditional heteroskedasticity in the ordered logit model.

Alternatively, given the partition boundaries, a higher conditional mean $X_k'\beta$ implies a higher probability of observing a more extreme state. Of course, the labelling of states is arbitrary, but the *ordered* probit model makes use of the natural ordering of the states. The regressors allow us to separate the effects of various economic factors that influence the likelihood of one state over another. For example, suppose that a large positive value of X_1 usually implies a large negative observed price change and vice-versa. Then the ordered probit coefficient β_1 will be negative in sign and large in magnitude (relative to σ of course).

From these observations, it is apparent that the rounding/eighths-barriers models of discreteness in Ball (1988), Cho and Frees (1988), Gottlieb and Kalay (1985), and Harris (1989c) may be re-parameterized as ordered probit models. Consider first the case of a "true" price process that is an arithmetic Brownian motion, with trades occurring only when this continuous-state process crosses an eighths threshold [see Cho and Frees (1988)]. Observed trades from such a process may be fit to an ordered probit model where the partition boundaries are fixed at multiples of eighths and the single regressor is the time interval (or first-passage time) between crossings, which appears in both the conditional mean and variance of Z_k^*. For the rounding models of Ball (1988), Gottlieb and Kalay (1985), and Harris (1989c) which do not make use of waiting-times between trades, define the partition boundaries as the midpoint between eighths [e.g. the observed price change is $\frac{3}{8}$ if the virtual price process lies in the interval $[\frac{5}{16}, \frac{7}{16})$] and omit the waiting time as a regressor in both the conditional mean and variance [see the discussion in Section 5.3 below].

The generality of the ordered probit model comes from the fact that the rounding and eighths-barrier models of discreteness can both be incorporated by appropriate definitions of the partition boundaries. In fact, since the boundaries may be parameterized to be time- and state-dependent, ordered probit allows for more general kinds of rounding and eighths barriers. In addition to fitting any arbitrary multinomial distribution, ordered probit may also accommodate finite-state Markov chains and compound Poisson processes.

Of course, other models of discreteness are not necessarily obsolete, since in several cases the parameters of interest may not be simple functions of the ordered probit parameters. For example, a tedious calculation will show that although Harris's (1989c) rounding model may be represented as an ordered probit, the bid/ask spread parameter c is not easily recoverable from the ordered probit parameters. In such cases, other equivalent specifications may allow more direct estimation of the relevant parameters.

2.1. The Likelihood Function.

Let Y_{ik} be an indicator variable which takes on the value 1 if the realization of the k-th observation Z_k is the i-th state s_i, and zero otherwise. Then the log-likelihood function $\mathcal{L}$ for the vector of price changes $Z = [\ Z_1\ Z_2\ \cdots\ Z_n\ ]'$, conditional on the explanatory variables $X = [\ X_1\ X_2\ \cdots\ X_n\ ]'$, is given by:

$$\mathcal{L}(Z|X) = \sum_{k=1}^{n} \Bigg\{ Y_{1k} \cdot \log \Phi\Big(\frac{\alpha_{1k} - X_k'\beta}{\sigma_k}\Big) +$$

$$\sum_{i=2}^{m-1} Y_{ik} \cdot \log \Bigg[\Phi\Big(\frac{\alpha_{ik} - X_k'\beta}{\sigma_k}\Big) - \Phi\Big(\frac{\alpha_{i-1k} - X_k'\beta}{\sigma_k}\Big) \Bigg] +$$

$$Y_{mk} \cdot \log \Bigg[1 - \Phi\Big(\frac{\alpha_{m-1k} - X_k'\beta}{\sigma_k}\Big) \Bigg] \Bigg\} . \qquad (2.11)$$

Time-varying probabilities of transiting from one state to another may be allowed by letting the partition boundaries be time- and state-dependent, so for example we may let α_{ik} be a linear function of predetermined variables. For simplicity, we assume that the α_{ik}'s are constant in our current application, hence we omit the subscript k and write the partition boundaries as α_i.

Recall that σ_k^2 is a conditional variance, conditioned upon X_k. This allows for conditional heteroscedasticity in the Z_k^*'s, as in the rounding model of Cho and Frees (1988) where the Z_k^*'s are increments of arithmetic Brownian motion with variance proportional to $t_k - t_{k-1}$. For this special case, we have:

$$X_k'\beta = \mu \Delta t_k \qquad (2.12)$$

$$\sigma_k^2 = \gamma^2 \Delta t_k . \qquad (2.13)$$

More generally, we may also let σ_k^2 depend on other economic variables W_k so that:

$$\sigma_k^2 = \gamma_0^2 + \sum_{i=1}^{K_\sigma} \gamma_i^2 W_{ik} . \qquad (2.14)$$

There are, however, some constraints that must be placed on these parameters to achieve identification since, for example, doubling the α's, the β's, and σ_k leaves the likelihood unchanged. We shall return to this issue in Section 4.

3. The Data.

The ISSM transaction database consists of time-stamped trades (to the nearest second), trade size, and bid/ask quotes from the New York and American Stock Exchanges and the consolidated regional exchanges from January 4 to December 29 of 1988. Because of the sheer size of the ISSM transaction database, we focus our attention on only ten randomly selected securities that did not undergo any stock splits during 1988.[9] They are: Abitibi-Price Incorporated (ABY), Quantum Chemical Corporation (CUE), Dow Chemical Corporation (DOW), First Chicago Corporation (FNB), Foster Wheeler Corporation (FWC), Handy and Harmon Company (HNH), Navistar International Corporation (NAV), Reebok International Limited (RBK), Sears Roebuck and Company (S), and American Telephone and Telegraph Incorporated (T). These ten stocks provide a reasonably broad and representative cross-section of U.S. securities in terms of market capitalization, price level, and other characteristics.

We take as our basic time series the *intra-day* price changes from trade to trade, i.e., all overnight price changes are discarded. The first and last trade of each day were also discarded, since those trades may differ systematically from others due to institutional features. Several other screens were imposed to eliminate "problem" trades, yielding sample sizes from 1,515 trades for ABY to 178,813 trades for T.[10]

To obtain a better grasp of this dataset, we report a few summary statistics in Tables 1a and b. To see that our sample of ten stocks contains considerable dispersion, observe that the low stock price ranges from \$3.875 (NAV) to \$77.375 (DOW), whereas the high ranges from \$7.250 (NAV) to \$107.000 (CUE). At \$22 million, HNH has the smallest market capitalization our sample, and T has the largest with a market value of \$30.3 billion.

For our empirical analysis we require some indicator of whether a transaction was a buy or a sell. Following Blume, MacKinlay and Terker (1989), we classify all transaction

[9] We confine our attention to stocks that have not split simply to minimize the effects of large changes in price levels.

[10] Specifically, the following observations were removed from the sample: (1) trades that occur when the "firm quotation obligation" is suspended; (2) trades occurring during "fast trading" conditions; (3) trades immediately following a trading halt due to "news dissemination"; and (4) trades larger than 3,276,000 shares. See the ISSM documentation for further details. Also, because we use three lags of price changes as explanatory variables, and three lags of 5-minute returns on the S&P 500 index futures prices, we do not use the first three price changes or price changes during the first 15 minutes of the day (whichever is greater) as observations of the dependent variable.

prices into three categories using the prevailing bid and ask price quotes: a "buy" if the transaction price is greater than the mean of the bid and ask prices, a "sell" if the transaction price is less than the mean of the bid and ask prices, and "neutral" if the transaction price is equal to the mean of the bid and ask prices. From Tables 1a,b we see that between 20 and 25 percent of each stock's transactions are neutral, and the remaining trades fall almost equally into the two remaining categories. The two exceptions are the two smallest stocks, ABY and HNH. The former has almost twice as many buys as sells, whereas the latter has more than twice as many sells as buys.

The means and standard deviations of other variables to be used in our ordered probit analysis are also given in Tables 1a and b. The precise definitions of these variables will be given below in Section 4, but briefly, Z_k is the price change between transactions $k-1$ and k, Δt_k is the time elapsed between these trades, AB_k is the bid/ask spread prevailing at transaction k, SP500_k is the return on the S&P 500 index futures price over the five-minute period immediately preceding transaction k, IBS_k is the buy/sell indicator described above (1 for a buy, -1 for a sell, and 0 for a neutral), and V_k is the natural logarithm of the dollar volume of transaction k. Note that for the larger stocks, trades occur almost every minute on average, with the exception FNB which has an average Δt_k of about five minutes. The smaller stocks trade less frequently, with ABY trading only once every thirty minutes on average.

Finally, Figure 2 contains histograms for the price change, time between trade, and volume variables. For all ten stocks, the distributions of price changes are remarkably symmetric, whereas the distributions of time between trades are not.

4. The Empirical Specification.

To estimate the parameters of the ordered probit model via maximum likelihood, we must first specify: (i) the partition boundaries α_{ik}; (ii) the number of states m; (iii) the explanatory variables X_k; and (iv) the parametrization of the variance σ_k^2. For simplicity, we assume that the α_{ik}'s are parameters constant through time, hence we drop the k subscript.

In selecting m, we must balance resolution against the practical constraint that an m too large will yield no observations in the extreme states s_1 and s_m. For example, if we set m to 101 and define the states s_1 and s_{101} symmetrically to be price changes of -50 ticks and $+50$ ticks respectively, we would find no Z_k's among our ten stocks falling into these two states. From the histograms in Figure 2, we set $m = 9$ for the larger stocks, implying extreme states of -4 ticks or less and $+4$ ticks or more. For the three smaller stocks, ABY, FWC and HNH, we set $m = 5$ implying extreme states of -2 ticks or less and $+2$ ticks or more.[11]

In selecting the explanatory variables X_k, we seek to capture several aspects of transaction price changes. First, we would like to allow for clock-time effects, since there is currently some dispute over whether trade-to-trade prices are stable in transaction time versus clock time. Second, we would like to account for the effects of the bid/ask spread on price changes since many transactions are merely movements from the bid price to the ask price or vice-versa. If, for example, in sequence of three trades the first and third were buyer-initiated while the second was seller-initiated, the sequence of transaction prices would exhibit reversals due solely to the bid/ask "bounce." Third, we would like to measure how the conditional distribution of price changes shifts in response to a trade of a given volume, i.e., the price impact per unit volume of trade. And fourth, we would like to capture the effects of "systematic" or market-wide movements in prices on the conditional distribution of an individual stock's price changes. To address these four issues, we first construct the following variables:

Δt_k: The time elapsed between transactions $k-1$ and k, in seconds.

AB_{k-1}: The bid/ask spread prevailing at time t_{k-1}, in ticks.

Z_{k-l}: Three lags ($l = 1,\ 2,\ 3$) of the dependent variable Z_k. Recall that for $m = 9$, price changes less then -4 ticks are set equal to -4 ticks (state s_1), and price changes greater than $+4$ ticks are set equal to $+4$ ticks (state s_9), and similarly for $m = 5$.

VOL_{k-l}: Three lags ($l = 1,\ 2,\ 3$) of the natural logarithm of the dollar volume of the $(k-l)$-th transaction, defined as the price of the $(k-l)$-th transaction (in dollars, not ticks) times the number of shares traded (denominated in 100's of shares), hence dollar volume is denominated in \$100's of dollars. All trades greater than 10,000 shares are set equal to 10,000 to reduce the influence of extraordinarily large trades.[12]

SP500_{k-l}: Three lags ($l = 1,\ 2,\ 3$) of 5-minute continuously compounded return

[11] The definition of states need not be symmetric – state s_1 can be -6 ticks or less, implying that state s_9 is $+2$ ticks or more. However, the symmetry of the histogram of price changes in Figures 2 suggests a symmetric definition of the s_j's.

[12] This is motivated by the New York Stock Exchange's classification of all trades greater than 10,000 shares as block trades.

of the Standard and Poor's 500 index futures price, for the contract maturing in the closest month beyond the month in which transaction $k-l$ occurred, where the return is computed with the futures price recorded one minute before the nearest round minute *prior* to t_{k-l} and the price recorded five minutes before this. More formally, we have:

$$\text{SP500}_{k-1} \equiv \log \frac{F(t_{k-1}^{-} - 60)}{F(t_{k-1}^{-} - 360)} \tag{4.1}$$

$$\text{SP500}_{k-2} \equiv \log \frac{F(t_{k-1}^{-} - 360)}{F(t_{k-1}^{-} - 660)} \tag{4.2}$$

$$\text{SP500}_{k-3} \equiv \log \frac{F(t_{k-1}^{-} - 660)}{F(t_{k-1}^{-} - 960)} \tag{4.3}$$

where $F(t^{-})$ is the S&P 500 index futures price at time t^{-} (measured in seconds) for the contract maturing the closest month beyond the month of transaction $k-l$, and t^{-} is the nearest round minute prior to time t (for example, if t is 10:35:47, then t^{-} is 10:35:00).[13]

IBS_{k-l}: Three lags ($l = 1, 2, 3$) of an indicator variable that takes the value 1 if the $(k-l)$-th transaction price is greater than the average of the quoted bid and ask prices at time t_{k-l}, the value −1 if the $(k-l)$-th transaction price is less than the average of the bid and ask prices at time t_{k-1}, and 0 otherwise, i.e.,

$$\text{IBS}_{k-l} \equiv \begin{cases} 1 & \text{if } P_{k-l} > \frac{1}{2}(P_{k-l}^{a} + P_{k-l}^{b}) \\ 0 & \text{if } P_{k-l} = \frac{1}{2}(P_{k-l}^{a} + P_{k-l}^{b}) \\ -1 & \text{if } P_{k-l} < \frac{1}{2}(P_{k-l}^{a} + P_{k-l}^{b}) \end{cases} . \tag{4.4}$$

Whether the $(k-l)$-th transaction price is closer to the ask price or the bid price is one measure of whether the transaction was buyer-initiated ($\text{IBS}_{k-l} = 1$) or seller-initiated ($\text{IBS}_{k-l} = -1$). If the transaction price

[13] This rather convoluted timing for computing SP500_k ensures that there is no temporal overlap between price changes and the returns to the index futures price. In particular, we first construct a minute-by-minute time series for futures prices by assigning to each round minute the nearest futures transaction price occurring *after* that minute but before the next (hence if the first futures transaction after 10:35:00 occurs at 10:35:15, the futures price assigned to 10:35:00 is this one). If no transaction occurs during this minute, the price prevailing at the previous minute is assigned to the current minute. Then for the price change Z_k, we compute SP500_{k-1} using the futures price one minute before the nearest round minute *prior* to t_{k-1}, and the price five minutes before this (hence if t_{k-1} is 10:36:45, we use the futures price assigned to 10:35:00 and 10:30:00 to compute SP500_{k-1}).

is at the midpoint of the bid and ask prices, the indicator is "neutral" ($\text{IBS}_{k-l} = 0$).

Our specification of $X_k'\beta$ is then given by the following expression:

$$X_k'\beta = \beta_1 \Delta t_k + \beta_2 Z_{k-1} + \beta_3 Z_{k-2} + \beta_4 Z_{k-3} + \beta_5 \text{SP500}_{k-1} + \beta_6 \text{SP500}_{k-2} + \beta_7 \text{SP500}_{k-3} + \beta_8 \text{IBS}_{k-1} + \beta_9 \text{IBS}_{k-2} + \beta_{10} \text{IBS}_{k-3} + \beta_{11}(\text{VOL}_{k-1} \cdot \text{IBS}_{k-1}) + \beta_{12}(\text{VOL}_{k-2} \cdot \text{IBS}_{k-2}) + \beta_{13}(\text{VOL}_{k-3} \cdot \text{IBS}_{k-3}) . \tag{4.5}$$

The variable Δt_k is included in X_k to allow for clock-time effects on the conditional mean of Z_k^*. If prices are stable in "transaction" time rather than clock time, this coefficient should be zero. Lagged price changes are included to account for serial dependencies, and lagged returns of the S&P500 index futures price are included to account for market-wide effects on price changes.

To measure the price impact of a trade per unit volume, we include the term VOL_{k-l} interacted with IBS_{k-l}, an indicator of whether the trade was buyer-initiated ($\text{IBS}_k = 1$), seller-initiated ($\text{IBS}_k = -1$), or neutral ($\text{IBS}_k = 0$). A positive β_{11} would imply that buyer-initiated trades tend to push prices up and seller-initiated trades tend to drive prices down. Such a relation is predicted by several information-based models of trading, e.g. Easley and O'Hara (1987). Moreover, the magnitude of β_{11} is the per-unit volume impact on the conditional mean of Z_k^*, which may be readily translated into the impact on the conditional probabilities of observed price changes. The sign and magnitudes of β_{12} and β_{13} measure the persistence of price impact.

To complete our specification we must parametrize the conditional variance $\sigma_k^2 \equiv \gamma_0^2 + \sum \gamma_i^2 W_{ik}$. To allow for clock-time effects we include Δt_k, and since there is some evidence linking bid/ask spreads to the information content and volatility of price changes,[14] we also include the lagged spread AB_{k-1}. Finally, recall from Section 2.1 that the parameters α, β, and γ are unidentified without additional restrictions, hence we make the identification assumption that $\gamma_0^2 = 1$. Our variance parametrization is then:

$$\sigma_k^2 \equiv 1 + \gamma_1^2 \Delta t_k + \gamma_2^2 \text{AB}_{k-1} . \tag{4.6}$$

[14] See, for example, Glosten (1987), Hasbrouck (1988, 1989a,b), and Petersen and Umlauf (1990).

In summary, our specification requires the estimation of 23 parameters, the partition boundaries $\alpha_1, \ldots, \alpha_8$, the variance parameters γ_1 and γ_2, and the coefficients of the explanatory variables $\beta_1, \ldots, \beta_{13}$.

5. The Maximum Likelihood Estimates.

We compute the maximum likelihood (ML) estimators numerically using the algorithm proposed by Berndt, Hall, Hall, and Hausman (1974), hereafter BHHH. The advantage of BHHH over other search algorithms is its reliance on only first derivatives, an important computational consideration for sample sizes such as ours.

In Tables 2a,b we report ML estimates of the ordered probit model for our ten stocks. Entries in the columns labelled with ticker symbols are the parameter estimates, and to the immediate right of each entry is the corresponding z-statistic, which is asymptotically distributed as a standard normal variate under the null hypothesis that the coefficient is zero, i.e., it is the parameter estimate divided by its asymptotic standard error.

Tables 2a,b show that the partition boundaries are estimated with high precision for all stocks. As expected, the z-statistics are much larger for those stocks with many more observations. The parameters for σ_k^2 are also statistically significant, hence homoskedasticity may be rejected at conventional significance levels. Larger bid/ask spreads and longer time intervals both increase the conditional volatility of the disturbance.

The conditional means of the Z_k^*'s for all stocks are only marginally affected by Δt. Moreover, the z-statistics are minuscule, especially in light of the large sample sizes. However, as mentioned above, Δt does enter into the σ_k^2 expression significantly, hence clock-time is important for conditional variances, but not for conditional means.

More striking is the significance and sign of the lagged price change coefficients $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$ – they are negative for all stocks, implying a tendency towards price reversals. For example, if the past three price changes were each 1 tick, the conditional mean of Z_k^* changes by $\hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4$. However, if the sequence of price changes was 1/–1/1, then the effect on the conditional mean is $\hat{\beta}_2 - \hat{\beta}_3 + \hat{\beta}_4$, a quantity closer to zero for each of the security's parameter estimates.[15]

Note that these coefficients measure reversal tendencies beyond that induced by the

[15] In an earlier specification, in place of lagged price changes we included separate indicator variables for eight of the nine states of each lagged price change. But because the coefficients of the indicator variables increased monotonically from the –4 state to the +4 state (state 0 was omitted) in almost exact proportion to the tick-change, we chose the more parsimonious specification of including the actual lagged price change.

presence of a constant bid/ask spread, as in Roll (1984). The effect of this "bid/ask bounce" on the conditional mean should be captured by the indicator variables IBS_{k-1}, IBS_{k-2},and IBS_{k-3}. In the absence of all other information (such as market movements, past price changes, etc.), these variables pick up any price effects that buys and sells might have on the conditional mean. As expected, the estimated coefficients are generally negative, indicating the presence of reversals due to movements from bid to ask or ask to bid prices.

More importantly, for each stock the coefficients on the three lagged price change variables are different, implying that the conditional mean of price changes is *path dependent* on past price changes. That is, a sequence of price changes of 1/–1/1 will have a different effect on the conditional mean than the sequence –1/1/1 even though both sequences yield the same total price change over the three trades. Similarly, the coefficients of the three lagged volume variables $V_{t-k}\mathrm{IBS}_{t-k}$ are also different. Taken together, these two findings lend support to Easley and O'Hara's (1987) prediction that information-based trading can lead to path dependent price changes, so that "To calculate the distribution of the next trade price, p_{t+1}, therefore, we need to know not only the current price p_t, but also how the market got to the current price."

The lagged S&P 500 returns are also significant, but has a more persistent affect on some securities. For example, the coefficient for the first lag of the S&P is large and significant for DOW, but the coefficients for the second and third are small and insignificant. However, for the less actively traded stocks such as CUE, all three coefficients are significant and of the same order of magnitude. As a measure of how quickly market-wide information is impounded into prices, these coefficients confirm the common intuition that smaller stocks react more slowly than larger stocks, and is consistent with the lead/lag effects uncovered by Lo and MacKinlay (1990a).

5.1. Measuring Price Impact Per Unit Volume of Trade.

By price impact we mean the effect of a sequence of trades on the conditional distribution of the *next* price change. As such, the coefficients of the variables $\mathrm{VOL}_{k-1}\cdot\mathrm{IBS}_{k-1}$, $\mathrm{VOL}_{k-2}\cdot\mathrm{IBS}_{k-2}$, and $\mathrm{VOL}_{k-3}\cdot\mathrm{IBS}_{k-3}$ measure the price impact of trades per unit volume. More precisely, recall that our definition of the volume variable is the logarithm of actual dollar volume divided by 100, hence the coefficient β_{11} is the contribution to the conditional mean $X_k'\beta$ that results from a \$271.828 trade (since $\log(271.828/100) = 1$).

Therefore, the impact of an \$M trade at time $k-1$ on $X_k'\beta$ is simply $\beta_{11}\log(M/100)$. The estimated coefficients in Tables 2a,b are generally positive and significant for all stocks, with the most recent trade having the largest impact. However, this is not the impact we seek since $X_k'\beta$ is the conditional mean of the unobserved variable Z_k^*, not the observed price change Z_k. In particular, since $X_k'\beta$ is scaled by σ_k in (2.10), it is difficult to make meaningful comparisons of these coefficients across stocks.

To obtain a measure of a trade's price impact that we *can* compare across stocks, we must translate the impact on $X_k'\beta$ into an impact on the conditional distribution of the Z_k's, conditioned on the trade size and other quantities. Since we have already established that the conditional distribution of price changes is path-dependent, we must condition on a specific *path* for past price changes and trade sizes. We do this by substituting our parameter estimates into (2.10), choosing particular values for the explanatory variables X_k, and computing the probabilities explicitly. In particular, we set Δt_k and AB_{k-1} to their sample means for each stock, and set the following variables to the same values across all stocks,

$$
\begin{aligned}
\mathrm{V}_{k-2} &= 5.298 \\
\mathrm{V}_{k-3} &= 5.298 \\
\mathrm{SP500}_{k-1} &= 0.001 \\
\mathrm{SP500}_{k-2} &= 0.001 \\
\mathrm{SP500}_{k-3} &= 0.001 \\
\mathrm{IBS}_{k-1} &= 1 \\
\mathrm{IBS}_{k-2} &= 1 \\
\mathrm{IBS}_{k-3} &= 1\,.
\end{aligned}
$$

Specifying values for these variables is equivalent to specifying the market conditions that we wish to measure price impact under. These particular values correspond to a scenario in which the most recent three trades are buys, where the sizes of the two earlier trades are \$20,000 each [since $5.298 = \log(\$20,000/100)$], and where the market index return is at its sample average during these trades. We then evaluate the probabilities in (2.10) for different values of V_{k-1}, Z_{k-1}, Z_{k-2}, and Z_{k-3}.

For brevity, we focus only on the means of these conditional distributions, which we report in Tables 3 and 4 for the ten stocks. The entries in Tables 3a and b are computed under the assumption that $Z_{k-1} = Z_{k-2} = Z_{k-3} = +1$, whereas those in Tables 4a and b are computed under the assumption that $Z_{k-1} = Z_{k-2} = Z_{k-3} = 0$. The first entry in the "ABY" column in Table 3a, –0.395, is the expected price change (in ticks) of the next transaction of ABY stock following a $5,000 buy. The seemingly counterintuitive sign of this conditional mean is the result of the "bid/ask bounce" – since the past three trades were assumed to be buys, the parameter estimates reflect the empirical fact that the next transaction can be a sell, in which case the transaction price change will often be negative since the price will go from ask to bid. To account for this effect, we would need to include a *contemporaneous* buy/sell indicator, IBS_k, in X'_k and condition on this variable as well. But such a variable is clearly endogenous to Z_k and our parameter estimates would suffer from the familiar simultaneous-equations biases.

However, to measure price impact we can "net out" the effect of the bid/ask spread by computing the *change* in the conditional mean for trade sizes larger than our base case $5,000 buy. As long as the bid/ask spread remains relatively constant, the change in the conditional mean induced by larger trades will give us a measure of price impact that is independent of it since it doesn't change as we vary the trade size. For example, the second entry in the "ABY" column of Table 3a shows that purchasing an additional $5,000 of ABY ($10,000 total) increases the conditional mean by 0.027 ticks. However, purchasing an additional $495,000 of ABY ($500,000 total) increases the conditional mean by 0.182 ticks. As expected, in all cases trading a larger quantity yields a larger price impact. Moreover, the values of the increases yield useful information: they determine how to break up larger trades into smaller ones so as to minimize overall price impact.

A comparison across columns in Tables 3a,b shows that large trades have higher price impact for CUE than for the other nine stocks. However, such a comparison ignores the fact that these stocks trade at different price levels, hence a price impact of 0.425 ticks for $500,000 of CUE may not be as large a percentage of price as a price impact of 0.088 ticks for $500,000 of NAV. The second panels of Tables 3a,b reports the price impact as percentages of the average of the high and low prices of each stock, and between CUE and NAV a trade of $500,000 does have a higher percentage price impact for NAV [0.197 percent versus 0.061 percent] even though it is considerably smaller when measured in ticks. Interestingly, even as a percentage, price impact increases with dollar volume.

In Tables 4a,b where price impact values have been computed under the alternative assumption that $Z_{k-1} = Z_{k-2} = Z_{k-3} = 0$, the conditional means $\mathrm{E}[Z_k]$ are closer to zero for the \$5,000 buy. For example, the expected price change of NAV is now –0.211 ticks, whereas in Table 3a it was –1.543 ticks. Since now we are conditioning on a scenario in which the three most recent transactions are buys that have no impact on prices, the empirical estimates imply more probability in the right tail of the conditional distribution of the next price change.

That the conditional mean is still negative may signal the continued importance of the bid/ask spread, nevertheless the price impact measure $\Delta\mathrm{E}[Z_k]$ does increase with dollar volume. Moreover, these values are similar in magnitude to those in Tables 3a,b – in percentage terms the price impact is virtually identical in both tables for CUE, DOW and FNB. For these stocks, price impact seems less sensitive to the path of past prices. An implication of this finding is that, whereas the timing of trades does depend on the sequence of past prices [since the conditional mean does change between Tables 3a and b], the decision to break up a large buy into several smaller orders *need not*. This, however, is an empirical feature not shared by NAV and RBK, since the price impact for those stocks differ considerably between Tables 3b and 4b. This suggests that price impact must measured security by security.

Of course, there is no reason to focus solely on the mean of the conditional distribution of Z_k since we have at our disposal an estimate of the entire distribution. Under the scenarios of Tables 3 and 4 we have also computed the standard deviation of conditional distribution, but since it is quite stable across the two scenarios we have omitted them from the tables for the sake of brevity. However, to get a sense of their sensitivity to the conditioning variables, we have plotted in Figure 3 the estimated conditional probabilities for the ten stocks under both scenarios. In each graph, the lightly cross-hatched bars represent the conditional distribution for the sequence of three buys with a +1 tick price change at each trade, for a fixed trade size of \$50,000 each. The dark-shaded bars represent the conditional distribution for the same sequence of three buys but with zero price change for each of the three transactions, also for a fixed trade size of \$20,000 each. The conditional distribution is clearly shifted more to the right under the first scenario than under the second, as the conditional means in Tables 3 and 4 foreshadowed. However, the general shape of the distribution seems rather well-preserved – changing the path of past price changes seems to *translate* the conditional distribution without greatly altering the tail probabilities.

As an aside, note that the conditional distributions of CUE and HNH display a striking pattern – the probabilities of even ticks are higher than odd ticks. This is evidence of price changes "clustering" at the even eighths, so that price changes of +2 ticks tend to be more likely than price changes of +1 tick. This finding differs from those in the extant literature in at least two ways. First, we find evidence of *price change clustering* whereas others such as Harris (1989a) focus on the clustering of price levels. And second, the evidence of clustering in Harris (1989a) is based on simple frequency counts of prices falling on eighths, quarters, etc., hence they are estimates of *unconditional* probabilities. Our finding is based on *conditional* probabilities which control for other effects such as market-wide shocks, past volume, order flow, etc.

As a final summary of price impact for these securities, we plot "price response" functions in Figure 4 for the ten stocks, which gives the percentage price impact as a function of dollar volume. These graphs show that the percentage price impact increases with volume, and that it increases at a decreasing rate. This, of course, is a feature of our log-specification for the volume explanatory variable – a plot of the percentage price impact against the logarithm of dollar volume would yield nearly linear relations. We are currently investigating more flexible functional forms for the volume variable, such as the Box and Cox (1964) transformation, where the particular shape is estimated and not imposed.[16] The price response function may be used to capture several features of the market microstructure. For example, market liquidity is often defined as the ability to trade any volume with little or no price impact. For such markets, the price response function is constant at zero, hence a direct measure of liquidity is how far the empirical price response function is from the x-axis. Since price response functions are defined in terms of percentage price impact, cross-stock comparisons of liquidity can also be made. Figure 4 shows that NAV and RBK are considerably less liquid than the four other stocks, with percentage price impacts more than triple those of the others. This is partly due to the low price ranges that NAV and RBK traded in [see Table 1a,b] – although RBK and S have comparable price impacts when measured in ticks [see Table 3a], RBK looks much less liquid when impact is measured as a percentage of price since its share price traded between $10.375 and $17.500 whereas S traded between $32.250 and $46.000 during 1988. Not surprisingly, CUE and DOW have the lowest percentage price impacts since their price ranges are the highest in the sample.

The shape of the price response function measures whether there are any economies

[16] The Box-Cox transformation $f(x)$ is given by $f(x) = (x^\lambda - 1)/\lambda$, where λ is a fixed parameter between 0 and 1. Observe that our logarithmic transformation is included as the special case when $\lambda = 0$.

of scale in trading. For example, a flat response function implies that the percentage price impact is not affected by the size of the trade. However, an upward sloping curve implies dis-economies of scale – larger dollar volume trades will yield higher percentage price impact. As such, the slope may be one measure of "market depth." For example, if the market for a security is "deep," this is usually taken to mean that large volumes may be traded before much of a price impact is observed. In such cases, the price response function may be downward sloping.

5.2. Endogeneity of Δt_k and IBS_k.

Our inferences in the preceding sections are based on the implicit assumption that the explanatory variables X_k are all exogenous or predetermined with respect to the dependent variable Z_k. However, the variable Δt_k is contemporaneous to Z_k and deserves further discussion.

Recall that Z_k is the price change between trades at time t_{k-1} and time t_k. Since Δt_k is simply $t_k - t_{k-1}$, it may well be that Δt_k and Z_k are determined simultaneously, in which case our parameter estimates are generally inconsistent. In fact, there are several plausible arguments for the endogeneity of Δt_k.[17] One such argument turns on the tendency of floor brokers to break up large trades into smaller ones, and time the executions carefully during the course of the day or several days. By "working" the order, the floor broker can minimize the price impact of his trades and obtain more favorable execution prices for his clients. But by selecting the times between his trades based on current market conditions, which include information also affecting price changes, the floor broker is creating endogenous trade times.

However, any given sequence of trades in our dataset does not necessarily correspond to consecutive transactions of any single individual (other than the specialist of course), but is the result of many buyers and sellers interacting with the specialist. For example, even if a floor broker were working a large order, in between his orders might be purchases and sales from other floor brokers, market orders, and triggered limit orders. Therefore, the Δt_k's also reflect these trades, which are not necessarily information-motivated.

Another more intriguing reason that Δt_k may be exogenous is that floor brokers have an economic incentive to minimize the correlation between Δt_k and virtually all other exogenous and predetermined variables. To see this, suppose the floor broker timed

[17] See, for example, Admati and Pfleiderer (1988, 1989) and Easley and O'Hara (1990)

his trades in response to some exogenous variable also affecting price changes, call it "weather." Suppose that price changes tend to be positive in good weather and negative in bad weather. Knowing this, the floor broker will wait until bad weather prevails before buying, hence trade times and price changes are simultaneously determined by weather. However, if other traders are also aware of these relations, they can garner information about the floor broker's intent by watching his trades and by recording the weather, and trade against him successfully. To prevent this, the floor broker must trade to deliberately minimize the correlation between his trade times and the weather. As such, the floor broker has an economic incentive to reduce simultaneous equations bias! Moreover, this argument applies to any other economic variable that can be used to jointly forecast trade times and price changes. For these two reasons, we assume that Δt_k is exogenous.[18]

However, endogeneity does matter for the inclusion of the contemporaneous buy/sell indicator IBS_k. or a sell. Without the contemporaneous indicator IBS_k, we are conditioning only on whether the *past* trades were buys or sells. By adding IBS_k as a regressor, we obtain an alternative to the price impact measure constructed in Section 5.1. With such a specification, we can now ask how large the *next* price change is likely to be conditional on the *next* trade being a buy or a sell. But there are few circumstances in which the contemporaneous buy/sell indicator IBS_k may be considered exogenous, since simple economic intuition suggests that factors affecting price changes must also enter the decision to buy or sell. Indeed, limit orders are explicit functions of the current price. Therefore, if IBS_k is to be included as an explanatory variable in X_k, its endogeneity must be taken into account. Unfortunately, the standard estimation techniques such as two-stage or three-stage least squares do not apply here because of our discrete dependent variable. Moreover, techniques that allow for discrete dependent variables also cannot be applied because the endogenous regressor IBS_k is also discrete. In principle, it may be possible to derive consistent estimators by considering a joint ordered probit model for both variables, but this is beyond the scope of the current paper. For the present, we restrict our specification to include only lags of IBS_k.

[18] We have also explored some adjustments for the endogeneity of Δt_k along the lines of Hausman (1978) and Newey (1985), and our preliminary estimates show that although exogeneity of Δt may be rejected at conventional significance levels (recall our sample sizes), the estimates do not change much once endogeneity is accounted for by an instrumental variables estimation procedure.

5.3. Does Discreteness Matter?

Despite the elegance and generality with which the ordered probit framework accounts for price discreteness, irregular trading intervals, and the influence of explanatory variables, the complexity of the estimation procedure raises the question of whether these features can be satisfactorily addressed by a simpler model. Since ordered probit may be viewed as a generalization of the linear regression model to discrete dependent variables, it is not surprising that the latter may share many of the advantages of the former, price discreteness aside. However, linear regression is considerably easier to implement. Therefore, what is gained by ordered probit? For example, suppose we ignore the fact that price changes Z_k are discrete, estimate the following simple regression model via ordinary least squares:

$$Z_k \;=\; X_k'\beta \,+\, \epsilon_k \tag{5.1}$$

and then compute the conditional distribution of Z_k by "rounding," thus:

$$\Pr\left(Z_k = \frac{j}{8} \right) \;=\; \Pr\left(\frac{j-1}{8} \le X_k'\beta \,+\, \epsilon_k < \frac{j}{8} \right) . \tag{5.2}$$

With suitable restrictions on the ϵ_k's, the regression model (5.1) is known as the "linear probability" model. The problems associated with applying ordinary least squares to (5.1) are well-known [see for example Judge et al. (1985, Ch. 18.2.1)], and numerous extensions have been developed to account for such problems. However, implementing such extensions is at least as involved as maximum likelihood estimation of the ordered probit model, therefore the comparison is of less immediate interest. In spite of these problems, we may still ask whether the OLS estimates of (5.1) and (5.2) yield an adequate "approximation" to a more formal model of price discreteness. Specifically, how different are the probabilities in (5.2) from those of the ordered probit model? If the differences are small, then the linear regression model (5.1) may be an adequate substitute to ordered probit.

Under the assumption of i.i.d. Gaussian ϵ_k's, we evaluate the conditional probabilities in (5.2) using the OLS parameter estimates and the same values for the X_k's as in Section 5.1, and graph them and the corresponding ordered probit probabilities in Figure 5. These graphs show that the two models yield very different conditional probabilities. All of the

OLS conditional distributions are unimodal and have little weight in the tails, in sharp contrast to the much more varied conditional distributions generated by ordered probit. For example, the OLS conditional probabilities show no evidence of the clustering that is readily apparent from the ordered probit probabilities of CUE. This is not surprising given the extra degrees of freedom that the ordered probit model has to fit the conditional distribution of price changes. Because the ordered probit partition boundaries $\{\alpha_i\}$ are determined by the data, the tail probabilities of the conditional distribution of price changes may be large or small relative to the probabilities of more central observations, unlike those of (5.1) which are dictated by the (Gaussian) distribution function of ϵ_k. Moreover, it is unlikely that using another distribution function will provide as much flexibility as ordered probit, for the simple reason that (5.1) constrains the state probabilities to be *linear* in the X_k's (hence the term "linear probability model"), whereas ordered probit allows for *nonlinear* effects by letting the data determine the partition boundaries $\{\alpha_i\}$.

A more direct test of the difference between ordered probit and the simple "rounded" linear regression model is to consider the special case of ordered probit in which all the partition boundaries $\{\alpha_i\}$ are equally spaced and fall on sixteenths. That is, let the observed discrete price change Z_k is related to the unobserved continuous random variable Z_k^* in the following manner:

$$Z_k = \begin{cases} -\frac{4}{8} \text{ or less} & \text{if } Z_k^* \in (-\infty\,,\, -\frac{4}{8}+\frac{1}{16}) \\ \\ \frac{j}{8} & \text{if } Z_k^* \in [\,\frac{j}{8}-\frac{1}{16}\,,\, \frac{j}{8}+\frac{1}{16}\,),\ j=-3,\ldots,3\ . \\ \\ \frac{4}{8} \text{ or more} & \text{if } Z_k^* \in [\,\frac{4}{8}-\frac{1}{16}\,,\,\infty\,) \end{cases} \tag{5.3}$$

This follows the spirit of Ball (1988), in which there exists a "virtual" or "true" price change Z_k^* linked to the observed price change Z_k by rounding Z_k^* to the nearest multiple of eighths of a dollar. A testable implication of (5.3) is that the partition boundaries $\{\alpha_i\}$ are equally-spaced, i.e.,

$$\alpha_2 - \alpha_1 \;=\; \alpha_3 - \alpha_2 \;=\; \cdots \;=\; \alpha_{m-1} - \alpha_{m-2} \tag{5.4}$$

where m is the number of states in our ordered probit model. We can re-write (5.4) as a linear hypothesis for the $(m-1 \times 1)$-vector of α's in the following way:

$$\mathrm{H}: \quad A\alpha \;=\; 0 \tag{5.5}$$

$$\text{where} \quad \underset{(m-3\times m-1)}{A} \;\equiv\; \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{pmatrix} . \tag{5.6}$$

Since the asymptotic distribution of the maximum likelihood estimator $\hat{\alpha}$ is given by:

$$\sqrt{T}(\hat{\alpha} - \alpha) \;\overset{a}{\sim}\; N(0, \Sigma) \tag{5.7}$$

where Σ is the appropriate sub-matrix of the inverse of the information matrix corresponding to the likelihood function (2.11), the "delta method" yields the asymptotic distribution of the following statistic θ under the null hypothesis H:

$$\mathrm{H}: \quad \theta \;\equiv\; T\hat{\alpha}'A'(A\Sigma A')^{-1}A\hat{\alpha} \;\overset{a}{\sim}\; \chi^2_{m-3} \,. \tag{5.8}$$

Table 5 reports the θ's for our sample of ten stocks, and since the 1 percent critical values of the χ^2_2 and χ^2_6 are 9.21 and 16.8 respectively, we can easily reject the null hypothesis H for each of the ten stocks. However, because our sample sizes are so large, large χ^2 statistics need not signal important *economic* departures from the null hypothesis. Nevertheless the point estimates of the α's in Tables 2a,b show that they do differ in economically important ways from the simpler rounding model (5.3). With CUE, for example, $\hat{\alpha}_3 - \hat{\alpha}_2$ is 2.890 but

$\hat{\alpha}_4 - \hat{\alpha}_3$ is 1.122. Such a difference captures the empirical fact that (conditioned on the X_k's) -1 tick changes are rarer than -2 tick changes, rarer than predicted by the simple linear probability model. Discreteness does matter.

6. Conclusion.

We conclude by discussing several extensions that we hope to pursue in ongoing research. Because of the flexibility and robustness of the ordered probit framework, we suffer from an embarrassment of riches in that there are too many empirical issues that may be investigated, even with the small group of ten stocks we have chosen. For example, we hope to see how sensitive the price response functions are to specific sequences of trades, which may be viewed as a measure of the path dependence predicted by Easley and O'Hara (1987). We also plan to construct a formal measure of market liquidity using the slope of the price response function and compare liquidity across stocks, across time, and over various price ranges. Beyond the current sample of stocks are many others with considerably different characteristics, and we hope to broaden our sample to obtain truly representative cross-section.

Finally, diagnostics for the "residuals" $\hat{\epsilon}_k$ are called for, such as simple tests for autocorrelation. If the disturbances are autocorrelated but independent of the regressors, then our parameter estimates are consistent but the standard errors are not. In this case, we may still obtain consistent standard errors by a simple extension of the results in Levine (1983) and Poirier and Ruud (1988). However, autocorrelation in ϵ_k may be due to omitted variables, in which case our parameter estimates are inconsistent. In such circumstances, economic theory must guide our selection of additional regressors, and we hope to stimulate the development of such theories with these findings.

References

Admati, A. and P. Pfleiderer, 1988, "A Theory of Intraday Patterns: Volume and Price Variability," *Review of Financial Studies* 1, 3–40.

Admati, A. and P. Pfleiderer, 1989, "Divide and Conquer: A Theory of Intrady and Day-of-the-Week Mean Effects," *Review of Financial Studies* 2, 189–224.

Ball, C., 1988, "Estimation Bias Induced by Discrete Security Prices," *Journal of Finance* 43, 841–865.

Berndt, E., Hall, B., Hall, R. and J. Hausman, 1974, "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement* 3, 653–65.

Blume, M., MacKinlay, C. and B. Terker, 1989, "Order Imbalances and Stock Price Movements on October 19 and 20, 1987," *Journal of Finance* 44, 827–848.

Box, G. and D. Cox, 1964, "An Analysis of Transformations," *Journal of the Royal Statistical Society, Series B*, 26, 211–243.

Cho, D. and E. Frees, 1988, "Estimating the Volatility of Discrete Stock Prices," *Journal of Finance* 43, 451–466.

Cohen, K., Maier, S., Schwartz, R. and D. Whitcomb, 1986, *The Microstructure of Securities Markets.* Englewood Cliffs: Prentice-Hall.

Easley, D. and M. O'Hara, 1987, "Price, Trade Size, and Information in Securities Markets," *Journal of Financial Economics* 19, 69–90.

Easley, D. and M. O'Hara, 1990, "The Process of Price Adjustment in Securities Markets," Working Paper, Johnson Graduate School of Management, Cornell University.

Glosten, L., 1987, "Components of the Bid–Ask Spread and the Statistical Properties of Transaction Prices," *Journal of Finance* 42, 1293–1307.

Glosten, L. and L. Harris, 1988, "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economics* 21, 123–142.

Gottlieb, G. and A. Kalay, 1985, "Implications of the Discreteness of Observed Stock Prices," *Journal of Finance* 40, 135–154.

Harris, L., 1987, "Estimation of 'True' Stock Price Variances and Bid–Ask Spreads from Discrete Observations," unpublished working paper, School of Business Administration, University of Southern California.

Harris, L., 1989a, "Stock Price Clustering, Discreteness Regulation, and Bid/Ask Spreads," New York Stock Exchange Working Paper #89–01.

Harris, L., 1989b, "Estimation of Stock Variances and Serial Covariances from Discrete Observations," Working Paper, University of Southern California.

Harris, L., Sofianos, G. and J. Shapiro, 1990, "Program Trading and Intraday Volatility," New York Stock Exchange Working Paper #90–03.

Hasbrouck, J., 1988, "Trades, Quotes, Inventories, and Information," *Journal of Financial Economics* 22, 229–252.

Hasbrouck, J., 1989a, "Inferring the Extent of Informational Asymmetries from Trades and the Variance of Efficient Prices: An Econometric Analysis," unpublished working paper.

Hasbrouck, J., 1989b, "Measuring the Information Content of Stock Trades," unpublished working paper.

Hasbrouck, J. and T. Ho, 1987, "Order Arrival, Quote Behavior, and the Return-Generating Process," *Journal of Finance* 42, 1035–1048.

Hausman, J., 1978, "Specification Tests in Econometrics," *Econometrica* 46, 1251–1271.

Judge, G., Griffiths, W., Hill, C., Lütkepohl H. and T. Lee, 1985, *The Theory and Practice of Econometrics.* New York: John Wiley and Sons.

Levine, D., 1983, "A Remark on Serial Correlation in Maximum Likelihood," *Journal of Econometrics* 23, 337–342.

Lo, A. and C. MacKinlay, 1990a, "When Are Contrarian Profits Due To Stock Market Overreaction?" *Review of Financial Studies* 3, 175–205.

Lo, A. and C. MacKinlay, 1990b, "Data-Snooping Biases in Tests of Financial Asset Pricing Models," *Review of Financial Studies* 3, 431–468.

Maddala, G. 1983, *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

Madhavan, A. , and S. Smidt, 1990, "A Bayesian Model of Intraday Specialist Pricing," Working Paper, University of Pennsylvania.

Newey, W., 1985, "Semiparametric Estimation of Limited Dependent Variable Models With Endogenous Explanatory Variables," *Annales de L'Insee* 59/60, 219–237.

Petersen, M., 1986, *Testing the Efficient Market Hypothesis: Information Lags, the Spread, and the Role of the Market Makers*, Undergraduate Thesis, Princeton University.

Petersen, M. and S. Umlauf, 1990, "An Empirical Examination of the Intraday Behavior of the NYSE Specialist," Working Paper, M.I.T.

Poirier, D. and P. Ruud, 1988, "Probit with Dependent Observations," *Review of Economic Studies* 55, 593–614.

Roll, R., 1984, "A Simple Implicit Measure of the Effective Bid–Ask Spread in an Efficient Market," *Journal of Finance* 39, 1127–1139.

Stoll, H., 1989, "Inferring the Components of the Bid–Ask Spread: Theory and Empirical Tests," *Journal of Finance* 44, 115–134.

Stoll, H. and R. Whaley, 1989, "Stock Market Structure and Volatility," Working Paper No. 89–14, Owen Graduate School of Management, Vanderbilt University.

Table 1a

Summary statistics for transactions prices of Abitibi-Price Incorporated (ABY – 1,515 trades), Quantum Chemical Corporation (CUE – 27,141 trades), Dow Chemical Company (DOW – 81,916 trades), First Chicago Corporation (FNB – 17,915 trades), and Foster Wheeler Corporation (FWC – 18,460 trades), for the sample period from 4 January 1988 to 29 December 1988. Note: Market values are computed at the beginning of the year.

Statistic	ABY	CUE	DOW	FNB	FWC
Low Price	15.500	66.000	77.375	19.500	11.500
High Price	21.500	107.000	93.000	35.125	16.875
Market Value ($\times\$10^9$)	0.144	0.223	17.738	1.125	0.501
% Trades at Prices:					
> Mean of Bid/Ask	50.10	42.29	39.46	37.58	37.64
= Mean of Bid/Ask	20.59	20.07	24.81	24.66	27.00
< Mean of Bid/Ask	29.31	37.64	35.73	37.76	35.36
Means and SD's:					
Mean(Z_k)	0.0238	0.0012	0.0003	0.0011	–0.0005
SD(Z_k)	0.7356	1.2332	0.7841	0.6835	0.6395
Mean(Δt_k)	1815.4271	203.7964	68.6545	307.0260	296.9080
SD(Δt_k)	2400.9813	282.8014	84.6750	459.4382	416.5058
Mean(AB_k)	2.1373	3.2922	2.3265	2.3764	2.0740
SD(AB_k)	1.2946	1.6236	1.3034	1.4659	1.1467
Mean($SP500_k$)	0.0036	–0.0014	0.0010	–0.0025	–0.0056
SD($SP500_k$)	0.1058	0.1237	0.1371	0.1494	0.1498
Mean(IBS_k)	0.2079	0.0466	0.0373	–0.0018	–0.0228
SD(IBS_k)	0.8668	0.8928	0.8663	0.8680	0.8541
Mean($V_k \times IBS_k$)	0.8859	0.3285	0.3224	0.0192	–0.0712
SD($V_k \times IBS_k$)	3.8407	5.5935	5.4837	4.4660	3.7294

Table 1b

Summary statistics for transactions prices of Handy and Harmon Company (HNH – 3,621 trades), Navistar International Corporation (NAV – 90,212 trades), Reebok International Limited (RBK – 62,512 trades), Sears Roebuck and Company (S – 90,262 trades), and American Telephone and Telegraph Company (T – 178,813 trades), for the sample period from 4 January 1988 to 29 December 1988. Note: Market values are computed at the beginning of the year.

Statistic	HNH	NAV	RBK	S	T
Low Price	14.500	3.875	10.375	32.250	24.250
High Price	18.375	7.250	17.500	46.000	30.000
Market Value ($\times\$10^9$)	0.022	1.057	1.209	13.390	30.332
% Trades at Prices:					
> Mean of Bid/Ask	22.31	40.79	38.24	38.23	41.51
= Mean of Bid/Ask	28.82	17.88	25.70	23.50	25.96
< Mean of Bid/Ask	50.87	41.33	36.06	38.27	32.53
Means and S.D.'s of:					
Mean(Z_k)	0.0000	0.0032	0.0033	0.0047	0.0050
SD(Z_k)	0.7551	0.6434	0.6322	0.6766	0.6445
Mean(Δt_k)	1158.4952	59.0604	89.6373	59.2988	31.0025
SD(Δt_k)	1520.8643	79.4750	131.1619	76.4844	35.3834
Mean(AB_k)	2.4134	1.4884	1.7917	2.2179	1.6616
SD(AB_k)	0.9155	0.7693	1.2887	1.2356	0.7990
Mean(SP500_k)	–0.0026	–0.0011	–0.0030	–0.0011	0.0004
SD(SP500_k)	0.1152	0.1224	0.1278	0.1180	0.1204
Mean(IBS_k)	–0.2856	–0.0054	0.0219	–0.0003	–0.0898
SD(IBS_k)	0.8065	0.9062	0.8617	0.8747	0.8557
Mean($\text{V}_k \times \text{IBS}_k$)	–1.1957	–0.0023	0.0947	0.0196	–0.3188
SD($\text{V}_k \times \text{IBS}_k$)	3.5037	3.3163	3.6875	4.6903	4.1509

Table 2a

Maximum likelihood estimates of the ordered probit model for transactions price changes of Abitibi-Price Incorporated (ABY – 1,515 trades), Quantum Chemical Corporation (CUE – 27,141 trades), Dow Chemical Company (DOW – 81,916 trades), First Chicago Corporation (FNB – 17,915 trades), and Foster Wheeler Corporation (FWC – 18,460 trades), for the sample period from 4 January 1988 to 29 December 1988. Each z-statistic is asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. Note: the orderd probit specification for ABY and FWC contains only 5 states (–2 ticks or less, –1, 0, +1, +2 ticks or more), hence only four α's were required.

Parameter	ABY	z	CUE	z	DOW	z	FNB	z	FWC	z
α_1	-5.408	-4.25	-6.770	-17.46	-6.571	-54.53	-5.563	-27.99	-5.012	-18.943
α_2	-2.533	-4.23	-5.928	-17.46	-5.759	-54.42	-5.171	-31.74	-1.958	-19.462
α_3	1.925	3.92	-3.038	-17.53	-3.791	-56.08	-3.449	-35.28	1.909	19.663
α_4	5.610	3.88	-1.916	-17.39	-1.516	-54.40	-1.399	-34.58	4.936	19.144
α_5	—	—	1.749	17.22	1.487	53.83	1.345	34.38	—	—
α_6	—	—	3.024	17.44	3.856	55.72	3.429	35.51	—	—
α_7	—	—	6.013	17.47	5.827	54.26	5.109	28.51	—	—
α_8	—	—	6.723	17.36	6.540	52.48	5.727	25.42	—	—
$\gamma_1 : \Delta t/100$	0.248	3.34	0.541	11.17	0.405	12.94	0.252	12.91	0.304	9.910
$\gamma_2 : AB_{-1}$	1.335	3.26	1.227	14.57	0.828	35.82	0.550	16.25	0.907	12.829
$\beta_1 : \Delta t/100$	-0.001	-0.16	-0.014	-1.98	-0.017	-2.10	-0.007	-2.24	-0.015	-3.525
$\beta_2 : Z_{-1}$	-0.568	-3.37	-0.339	-12.41	-1.084	-51.88	-0.846	-29.79	-1.472	-19.239
$\beta_3 : Z_{-2}$	-0.369	-2.29	-0.001	-0.03	-0.478	-33.96	-0.381	-16.08	-0.687	-14.367
$\beta_4 : Z_{-3}$	-0.040	-0.38	-0.017	-1.16	-0.177	-18.62	-0.127	-6.71	-0.230	-8.295
$\beta_5 : SP500_{-1}$	0.754	1.06	2.478	12.94	1.678	32.66	1.166	13.81	1.549	12.307
$\beta_6 : SP500_{-2}$	0.544	0.90	1.495	9.43	0.044	1.07	0.546	6.06	0.385	3.277
$\beta_7 : SP500_{-3}$	1.194	1.79	0.765	5.34	-0.017	-0.39	0.429	4.80	0.260	2.218
$\beta_8 : IBS_{-1}$	-1.879	-3.26	-2.193	-14.59	-1.539	-34.24	-0.768	-11.95	-1.066	-11.413
$\beta_9 : IBS_{-2}$	-0.122	-0.41	-0.286	-3.15	-0.403	-11.04	-0.074	-1.30	-0.230	-3.231
$\beta_{10} : IBS_{-3}$	0.277	0.93	0.105	1.18	-0.176	-4.94	-0.047	-0.84	-0.230	-3.323
$\beta_{11} : V_{-1}IBS_{-1}$	0.188	2.23	0.245	12.40	0.170	27.39	0.085	7.43	0.110	6.821
$\beta_{12} : V_{-2}IBS_{-2}$	0.051	0.77	0.039	2.78	0.049	8.87	0.013	1.23	0.024	1.637
$\beta_{13} : V_{-3}IBS_{-3}$	-0.056	-0.85	0.003	0.25	0.022	4.08	0.004	0.33	0.034	2.306

Table 2b

Maximum likelihood estimates of the ordered probit model for transactions price changes of Handy and Harmon Company (HNH – 3,621 trades), Navistar International Corporation (NAV – 90,212 trades), Reebok International Limited (RBK – 62,512 trades), Sears Roebuck and Company (S – 90,262 trades), and American Telephone and Telegraph Company (T – 178,813 trades), for the sample period from 4 January 1988 to 29 December 1988. Each z-statistic is asymptotically standard normal under the null hypothesis that the corresponding coefficient is zero. Note: the orderd probit specification for HNH contains only 5 states (–2 ticks or less, –1, 0, +1, +2 ticks or more), hence only four α's were required.

Parameter	HNH	z	NAV	z	RBK	z	S	z	T	z
α_1	–6.032	–3.52	–6.916	–42.59	–6.258	–52.56	–6.185	–61.37	–8.217	–52.878
α_2	–2.350	–3.52	–6.644	–40.64	–5.975	–50.16	–5.690	–64.96	–7.451	–56.901
α_3	2.703	3.60	–5.864	–43.05	–4.601	–54.21	–4.102	–71.39	–5.673	–56.622
α_4	6.158	3.55	–1.892	–40.84	–1.756	–52.99	–1.588	–71.52	–1.968	–55.410
α_5	—	—	1.803	40.73	1.659	52.48	1.492	71.22	2.014	56.593
α_6	—	—	5.777	44.05	4.552	53.24	4.025	72.94	5.476	56.634
α_7	—	—	6.493	46.16	5.857	46.46	5.481	69.19	7.445	51.784
α_8	—	—	6.620	47.29	6.050	48.15	5.849	66.11	8.438	46.843
$\gamma_1 : \Delta t/100$	0.260	2.99	0.390	10.04	0.242	7.80	0.295	10.09	0.344	5.689
$\gamma_2 : AB_{-1}$	1.612	3.06	0.778	21.68	0.655	25.47	0.596	36.16	0.914	35.822
$\beta_1 : \Delta t/100$	–0.011	–2.02	–0.038	–4.82	–0.011	–2.06	–0.036	–4.93	–0.133	–9.824
$\beta_2 : Z_{-1}$	–0.898	–3.46	–2.221	–44.25	–1.715	–55.16	–1.634	–70.74	–2.225	–56.965
$\beta_3 : Z_{-2}$	–0.372	–2.74	–1.269	–38.58	–0.987	–43.90	–0.959	–58.49	–1.286	–50.684
$\beta_4 : Z_{-3}$	–0.138	–1.71	–0.479	–33.58	–0.409	–32.53	–0.393	–40.23	–0.476	–43.323
$\beta_5 : SP500_{-1}$	0.772	1.68	0.406	8.17	0.709	14.00	0.877	20.83	0.577	15.365
$\beta_6 : SP500_{-2}$	0.778	1.54	0.129	2.60	0.362	7.07	0.369	8.60	0.196	5.330
$\beta_7 : SP500_{-3}$	0.787	1.73	0.154	3.09	0.214	4.17	0.144	3.26	0.139	3.751
$\beta_8 : IBS_{-1}$	–1.533	–3.13	–0.651	–22.25	–0.904	–26.29	–0.828	–30.61	–1.070	–37.432
$\beta_9 : IBS_{-2}$	–0.387	–1.54	–0.359	–15.25	–0.314	–10.51	–0.318	–12.59	–0.387	–17.576
$\beta_{10} : IBS_{-3}$	0.262	1.11	–0.238	–10.68	–0.267	–9.10	–0.196	–7.95	–0.254	–11.960
$\beta_{11} : V_{-1}IBS_{-1}$	0.109	1.83	0.040	7.78	0.096	14.67	0.096	21.31	0.100	24.040
$\beta_{12} : V_{-2}IBS_{-2}$	0.086	1.56	0.035	6.69	0.045	7.02	0.044	10.08	0.037	9.339
$\beta_{13} : V_{-3}IBS_{-3}$	–0.062	–1.17	0.019	3.58	0.039	6.09	0.022	5.00	0.013	3.341

Table 3a

Price impact of trades as measured by the change in conditional mean of Z_k, or $\Delta E[Z_k]$, when trade sizes are increased incrementally above the base case of a \$5,000 trade. These changes are computed from the ordered probit probabilities, conditional on the three most recent trades being buyer-initiated, and the three most recent price changes being +1 tick each, for Abitibi-Price Incorporated (ABY – 1,515 trades), Quantum Chemical Corporation (CUE – 27,141 trades), Dow Chemical Company (DOW – 81,916 trades), First Chicago Corporation (FNB – 17,915 trades), and Foster Wheeler Corporation (FWC – 18,460 trades), for the sample period from 4 January 1988 to 29 December 1988, for the sample period from 4 January 1988 to 29 December 1988. Percentage price impact is computed as a percentage of the average of the high and low prices.

\$ Volume	ABY	CUE	DOW	FNB	FWC
(Ticks)					
$E[Z_k]$: 5,000	−0.395	−0.584	−1.206	−0.807	−0.892
$\Delta E[Z_k]$: 10,000	0.027	0.063	0.054	0.028	0.020
$\Delta E[Z_k]$: 20,000	0.054	0.127	0.107	0.057	0.041
$\Delta E[Z_k]$: 50,000	0.091	0.212	0.177	0.094	0.068
$\Delta E[Z_k]$: 100,000	0.118	0.276	0.230	0.121	0.089
$\Delta E[Z_k]$: 250,000	0.154	0.360	0.299	0.158	0.117
$\Delta E[Z_k]$: 500,000	0.182	0.425	0.351	0.185	0.137
(% of Price)					
$E[Z_k]$: 5,000	−0.275	−0.084	−0.177	−0.369	−0.786
$\Delta E[Z_k]$: 10,000	0.019	0.009	0.008	0.013	0.018
$\Delta E[Z_k]$: 20,000	0.038	0.018	0.016	0.026	0.036
$\Delta E[Z_k]$: 50,000	0.063	0.031	0.026	0.043	0.060
$\Delta E[Z_k]$: 100,000	0.082	0.040	0.034	0.056	0.078
$\Delta E[Z_k]$: 250,000	0.107	0.052	0.044	0.072	0.103
$\Delta E[Z_k]$: 500,000	0.126	0.061	0.051	0.085	0.121

Table 3b

Price impact of trades as measured by the change in conditional mean of Z_k, or $\Delta E[Z_k]$, when trade sizes are increased incrementally above the base case of a $5,000 trade. These changes are computed from the ordered probit probabilities, conditional on the three most recent trades being buyer-initiated, and the three most recent price changes being +1 tick each, for Handy and Harmon Company (HNH – 3,621 trades), Navistar International Corporation (NAV – 90,212 trades), Reebok International Limited (RBK – 62,512 trades), Sears Roebuck and Company (S – 90,262 trades), and American Telephone and Telegraph Company (T – 178,813 trades), for the sample period from 4 January 1988 to 29 December 1988. Percentage price impact is computed as a percentage of the average of the high and low prices.

$ Volume	HNH	NAV	RBK	S	T
(Ticks)					
$E[Z_k]$: 5,000	−0.460	−1.543	−1.394	−1.473	−1.565
$\Delta E[Z_k]$: 10,000	0.012	0.013	0.034	0.035	0.028
$\Delta E[Z_k]$: 20,000	0.025	0.027	0.067	0.070	0.056
$\Delta E[Z_k]$: 50,000	0.041	0.044	0.110	0.115	0.092
$\Delta E[Z_k]$: 100,000	0.053	0.058	0.143	0.149	0.120
$\Delta E[Z_k]$: 250,000	0.069	0.075	0.185	0.193	0.155
$\Delta E[Z_k]$: 500,000	0.082	0.088	0.216	0.226	0.182
(% of Price)					
$E[Z_k]$: 5,000	−0.350	−3.468	−1.250	−0.470	−0.721
$\Delta E[Z_k]$: 10,000	0.009	0.030	0.030	0.011	0.013
$\Delta E[Z_k]$: 20,000	0.019	0.060	0.060	0.022	0.026
$\Delta E[Z_k]$: 50,000	0.031	0.100	0.099	0.037	0.043
$\Delta E[Z_k]$: 100,000	0.040	0.129	0.128	0.048	0.055
$\Delta E[Z_k]$: 250,000	0.053	0.168	0.166	0.062	0.072
$\Delta E[Z_k]$: 500,000	0.062	0.197	0.194	0.072	0.084

Table 4a

Price impact of trades as measured by the change in conditional mean of Z_k, or $\Delta E[Z_k]$, when trade sizes are increased incrementally above the base case of a $5,000 trade. These changes are computed from the ordered probit probabilities, conditional on the three most recent trades being buyer-initiated, and the three most recent price changes being 0 tick each, for Abitibi-Price Incorporated (ABY – 1,515 trades), Quantum Chemical Corporation (CUE – 27,141 trades), Dow Chemical Company (DOW – 81,916 trades), First Chicago Corporation (FNB – 17,915 trades), and Foster Wheeler Corporation (FWC – 18,460 trades), for the sample period from 4 January 1988 to 29 December 1988, for the sample period from 4 January 1988 to 29 December 1988. Percentage price impact is computed as a percentage of the average of the high and low prices.

$ Volume	ABY	CUE	DOW	FNB	FWC
(Ticks)					
$E[Z_k]$: 5,000	–0.190	–0.451	–0.447	–0.196	–0.227
$\Delta E[Z_k]$: 10,000	0.028	0.064	0.049	0.025	0.022
$\Delta E[Z_k]$: 20,000	0.055	0.128	0.098	0.050	0.043
$\Delta E[Z_k]$: 50,000	0.092	0.213	0.162	0.083	0.072
$\Delta E[Z_k]$: 100,000	0.120	0.277	0.211	0.108	0.093
$\Delta E[Z_k]$: 250,000	0.156	0.362	0.274	0.141	0.122
$\Delta E[Z_k]$: 500,000	0.184	0.426	0.323	0.166	0.143
(% of Price)					
$E[Z_k]$: 5,000	–0.132	–0.065	–0.066	–0.090	–0.200
$\Delta E[Z_k]$: 10,000	0.019	0.009	0.007	0.012	0.019
$\Delta E[Z_k]$: 20,000	0.038	0.018	0.014	0.023	0.038
$\Delta E[Z_k]$: 50,000	0.064	0.031	0.024	0.038	0.063
$\Delta E[Z_k]$: 100,000	0.083	0.040	0.031	0.050	0.082
$\Delta E[Z_k]$: 250,000	0.109	0.052	0.040	0.065	0.107
$\Delta E[Z_k]$: 500,000	0.128	0.062	0.047	0.076	0.126

Table 4b

Price impact of trades as measured by the change in conditional mean of Z_k, or $\Delta E[Z_k]$, when trade sizes are increased incrementally above the base case of a \$5,000 trade. These changes are computed from the ordered probit probabilities, conditional on the three most recent trades being buyer-initiated, and the three most recent price changes being 0 tick each, for Handy and Harmon Company (HNH – 3,621 trades), Navistar International Corporation (NAV – 90,212 trades), Reebok International Limited (RBK – 62,512 trades), Sears Roebuck and Company (S – 90,262 trades), and American Telephone and Telegraph Company (T – 178,813 trades), for the sample period from 4 January 1988 to 29 December 1988. Percentage price impact is computed as a percentage of the average of the high and low prices.

\$ Volume	HNH	NAV	RBK	S	T
(Ticks)					
$E[Z_k]$: 5,000	-0.228	-0.211	-0.192	-0.207	-0.292
$\Delta E[Z_k]$: 10,000	0.013	0.008	0.021	0.024	0.018
$\Delta E[Z_k]$: 20,000	0.025	0.015	0.041	0.047	0.037
$\Delta E[Z_k]$: 50,000	0.042	0.025	0.068	0.078	0.061
$\Delta E[Z_k]$: 100,000	0.055	0.032	0.089	0.101	0.079
$\Delta E[Z_k]$: 250,000	0.071	0.042	0.116	0.132	0.104
$\Delta E[Z_k]$: 500,000	0.084	0.050	0.136	0.156	0.122
(% of Price)					
$E[Z_k]$: 5,000	-0.173	-0.474	-0.172	-0.066	-0.134
$\Delta E[Z_k]$: 10,000	0.010	0.017	0.019	0.008	0.008
$\Delta E[Z_k]$: 20,000	0.019	0.034	0.037	0.015	0.017
$\Delta E[Z_k]$: 50,000	0.032	0.056	0.061	0.025	0.028
$\Delta E[Z_k]$: 100,000	0.042	0.073	0.080	0.032	0.037
$\Delta E[Z_k]$: 250,000	0.054	0.095	0.104	0.042	0.048
$\Delta E[Z_k]$: 500,000	0.064	0.112	0.122	0.050	0.056

Table 5

Tests of equally spaced partition boundaries $\{\alpha_i\}$ from the ordered probit model for Abitibi-Price Incorporated (ABY), Quantum Chemical Corporation (CUE), Dow Chemical Company (DOW), First Chicago Corporation (FNB), and Foster Wheeler Corporation (FWC), Handy and Harmon Company (HNH), Navistar International Corporation (NAV), Reebok International Limited (RBK), Sears Roebuck and Company (S), and American Telephone and Telegraph Company (T), for the sample period from 4 January 1988 to 29 December 1988. Entries in the column labelled "m" denote the number of states in the ordered probit specification. The 5 and 1 percent critical values of a χ_2^2 random variate are 5.99 and 9.21 respectively. The 5 and 1 percent critical values of a χ_6^2 random variate are 12.6 and 16.8 respectively.

Stock	Sample Size	$\theta \overset{a}{\sim} \chi^2_{m-3}$	m
ABY	1,515	12.91	5
CUE	27,151	306.01	9
DOW	81,916	1,559.07	9
FNB	17,915	462.07	9
FWC	18,460	157.35	5
HNH	3,261	12.60	5
NAV	90,212	1,588.72	9
RBK	62,512	1,907.53	9
S	90,262	2,711.75	9
T	178,813	1,667.73	9

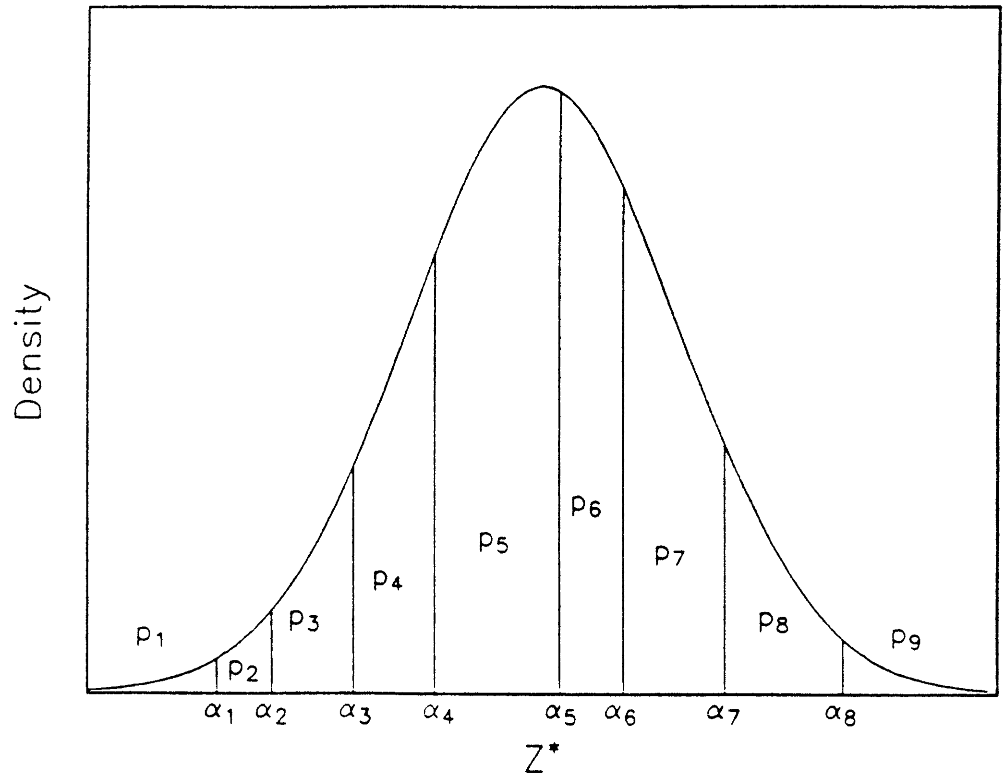

Figure 1.

Illustration of ordered probit probabilities p_i which are determined by the α_i's and distribution of Z_k^*. In particular, $p_i \equiv \text{Prob}(Z = s_i) = \text{Prob}(\alpha_{i-1} \leq Z^* < \alpha_i)$, $i = 1, \ldots, 9$ where, for notational simplicity, we define $\alpha_0 \equiv -\infty$ and $\alpha_9 \equiv +\infty$.

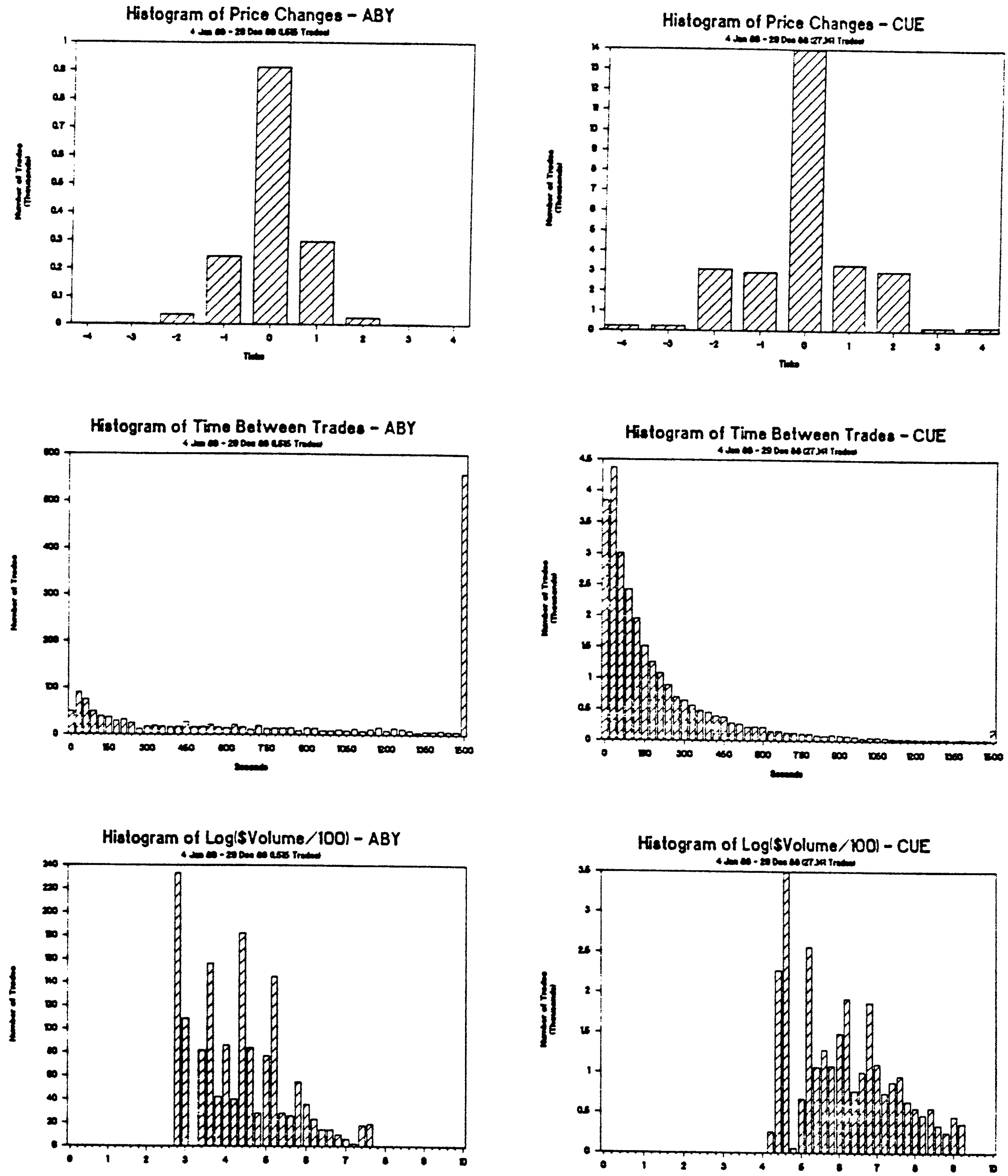
Histogram of Price Changes - ABY
Histogram of Price Changes - CUE
Histogram of Time Between Trades - ABY
Histogram of Time Between Trades - CUE
Histogram of Log($Volume/100) - ABY
Histogram of Log($Volume/100) - CUE

Figure 2.

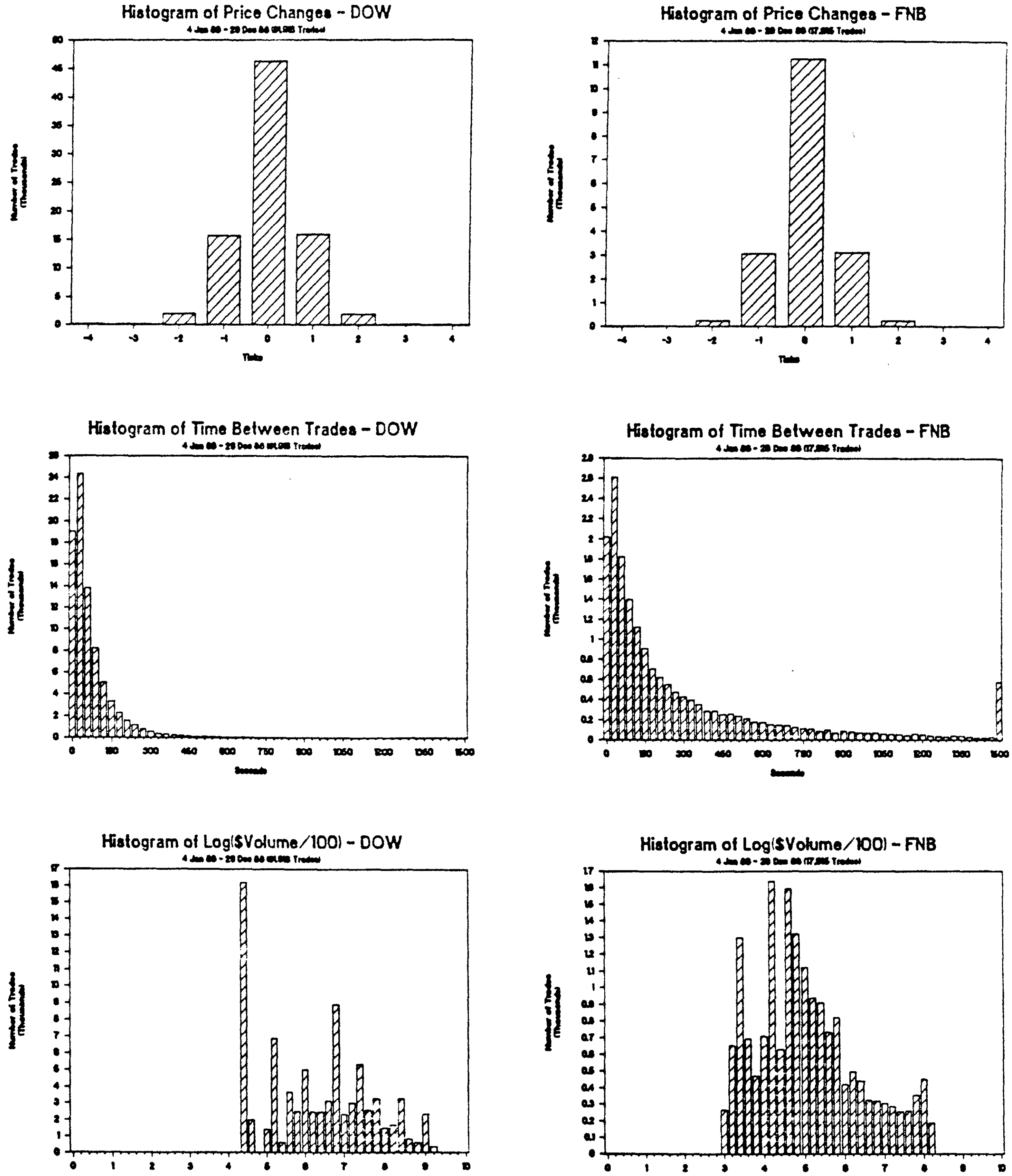

Figure 2 (Continued).

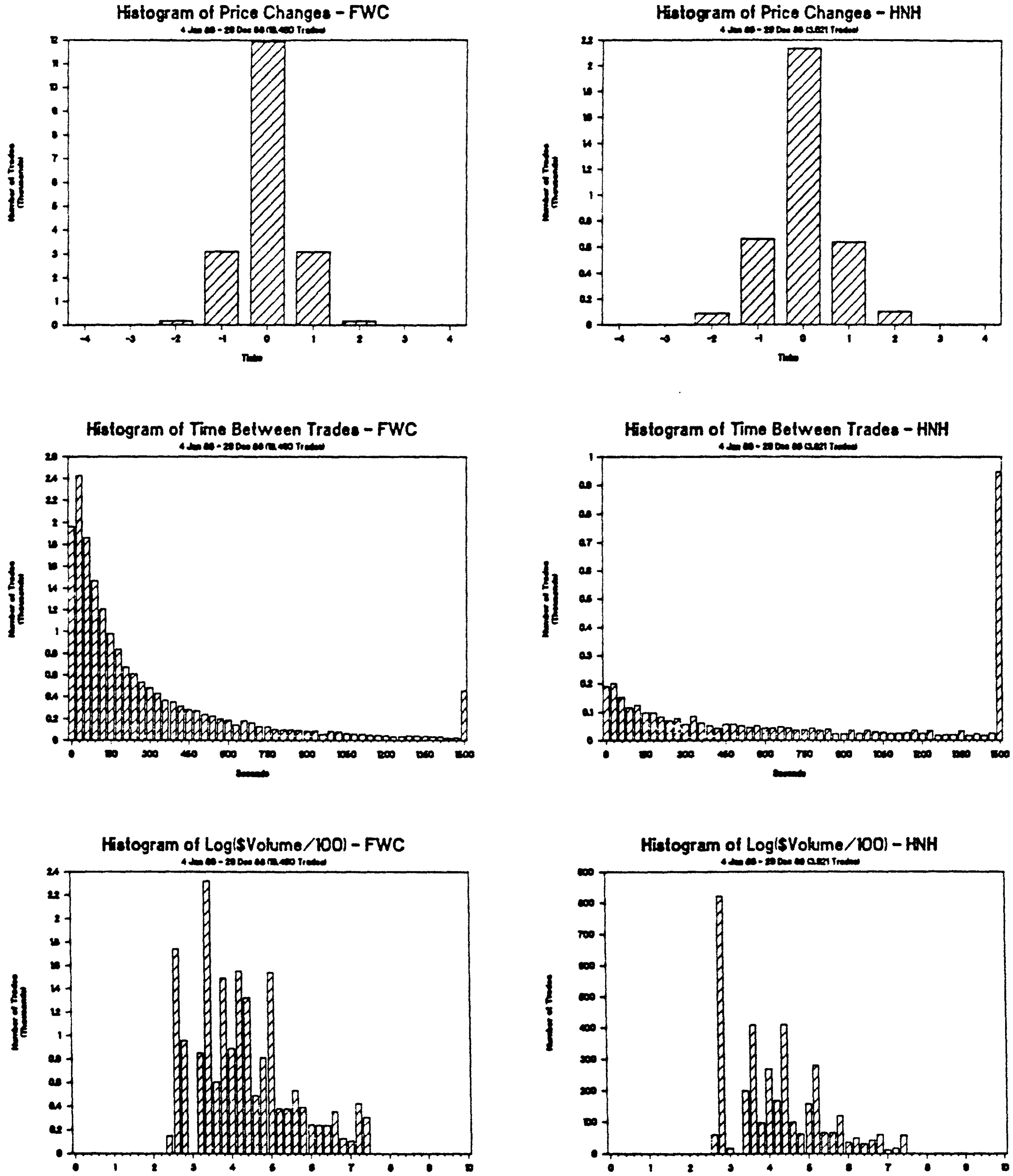

Figure 2 (Continued).

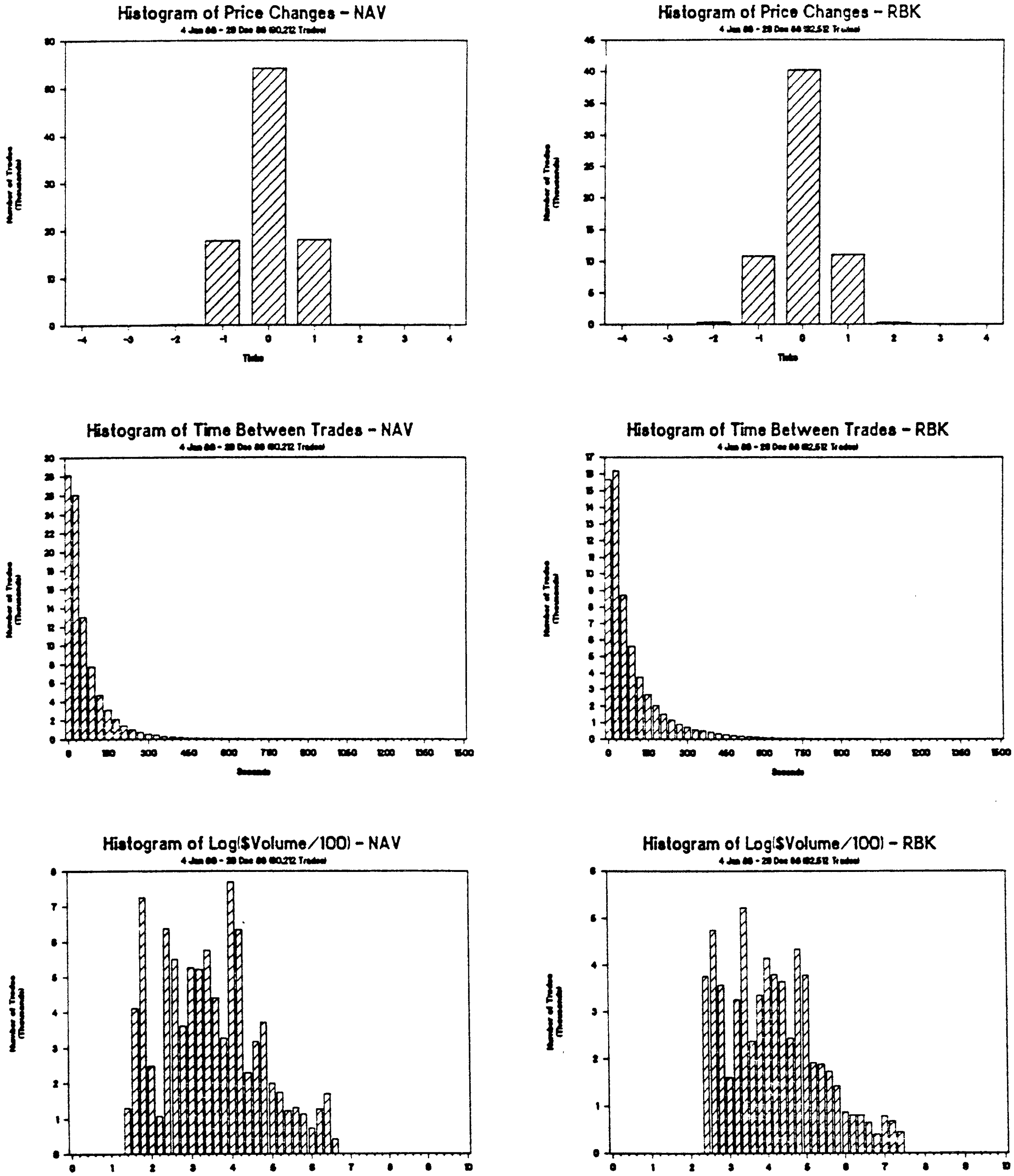

Figure 2 (Continued).

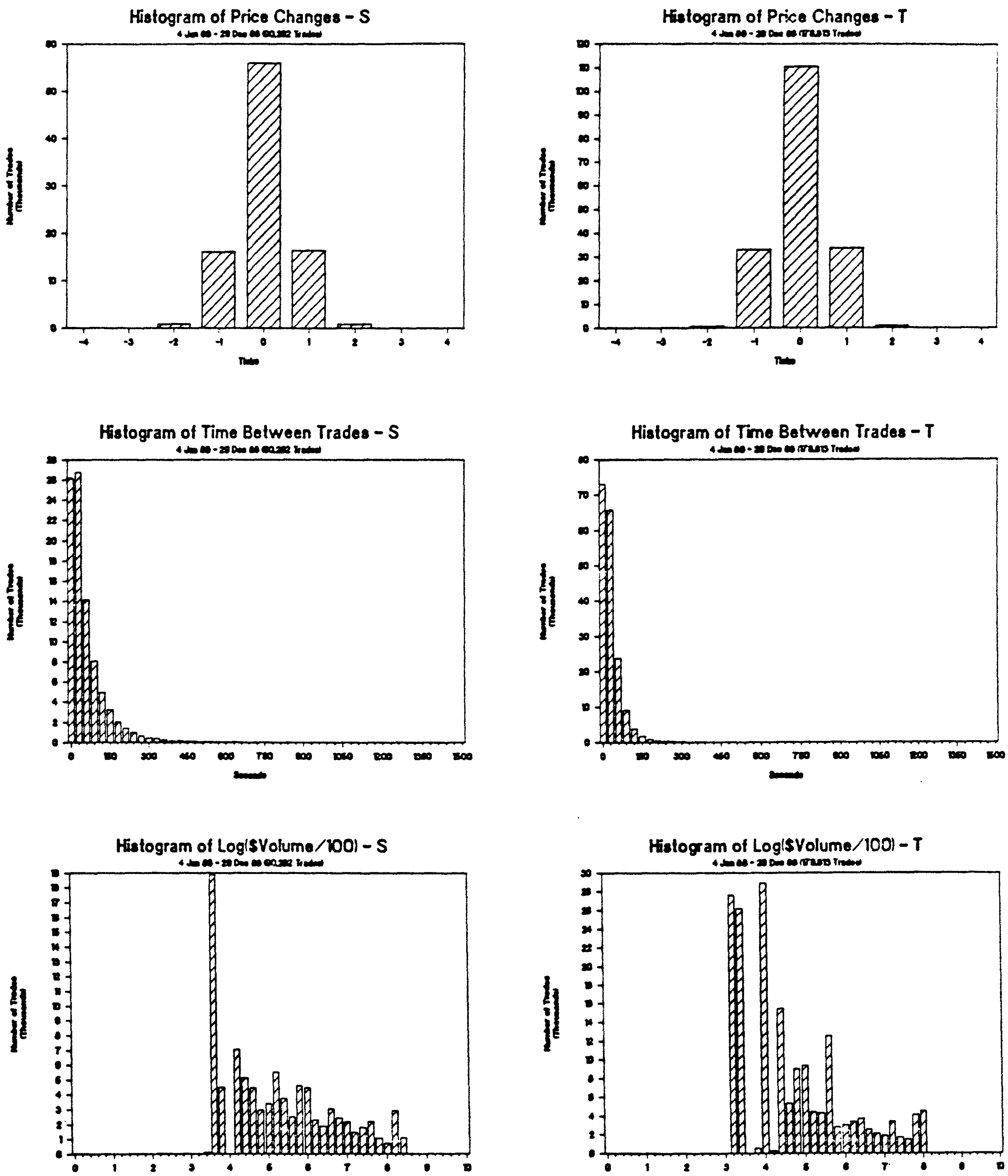

Figure 2 (Continued).

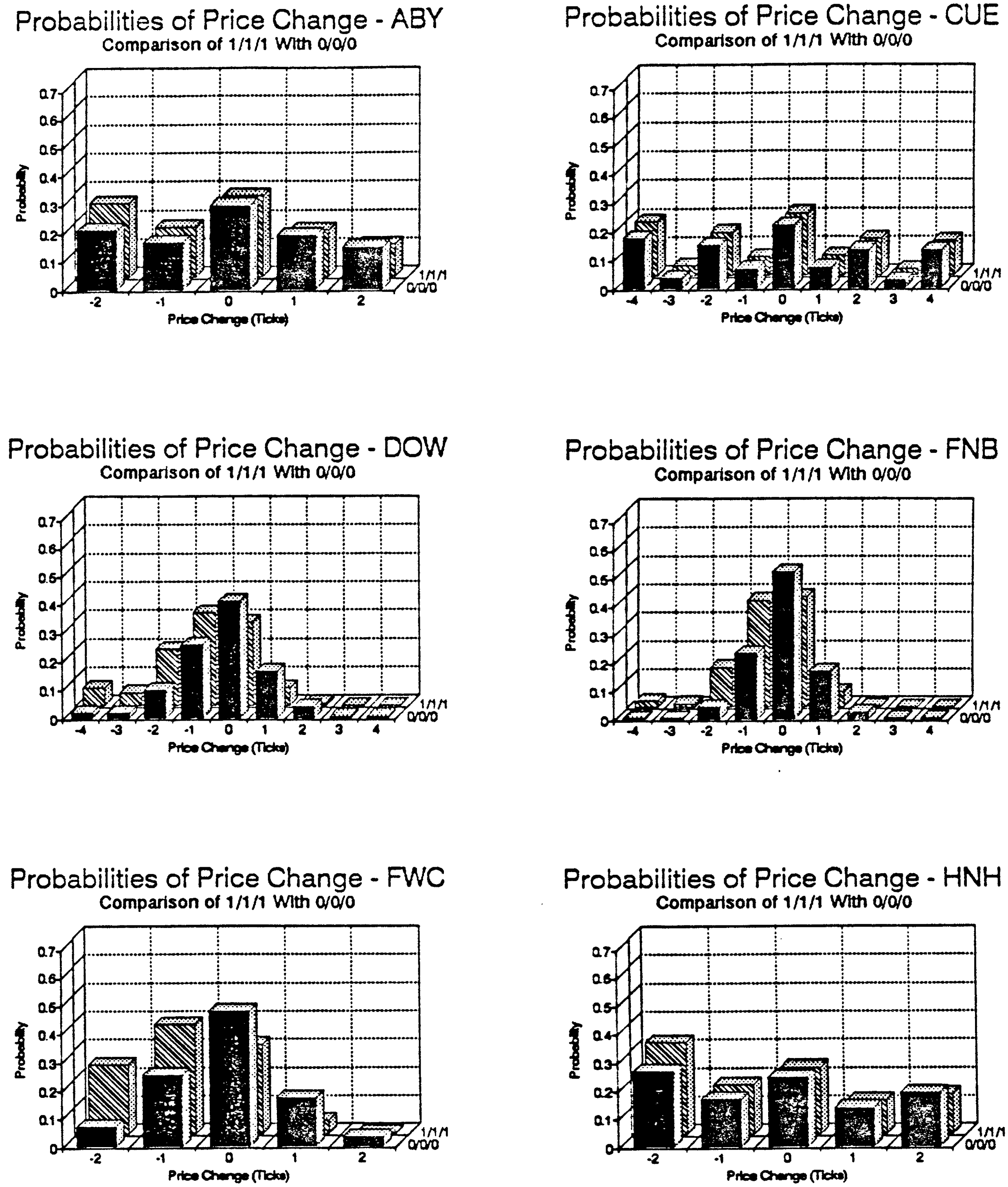
Probabilities of Price Change - ABY
Comparison of 1/1/1 With 0/0/0
Probability
Price Change (Ticks)
1/1/1
0/0/0
Probabilities of Price Change - CUE
Comparison of 1/1/1 With 0/0/0
Probability
Price Change (Ticks)
1/1/1
0/0/0
Probabilities of Price Change - DOW
Comparison of 1/1/1 With 0/0/0
Probability
Price Change (Ticks)
1/1/1
0/0/0
Probabilities of Price Change - FNB
Comparison of 1/1/1 With 0/0/0
Probability
Price Change (Ticks)
1/1/1
0/0/0
Probabilities of Price Change - FWC
Comparison of 1/1/1 With 0/0/0
Probability
Price Change (Ticks)
1/1/1
0/0/0
Probabilities of Price Change - HNH
Comparison of 1/1/1 With 0/0/0
Probability
Price Change (Ticks)
1/1/1
0/0/0

Figure 3.

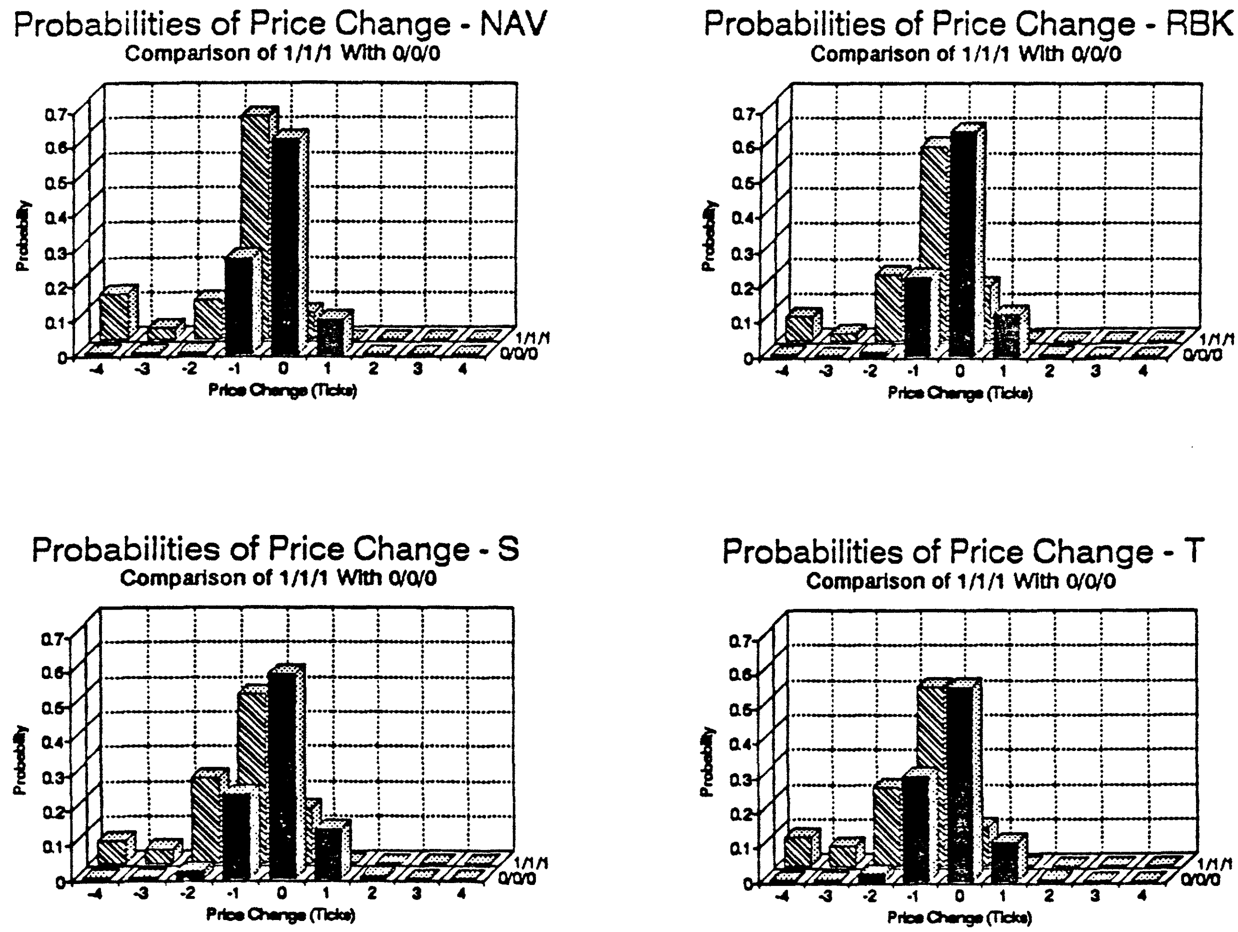

Figure 3 (Continued).

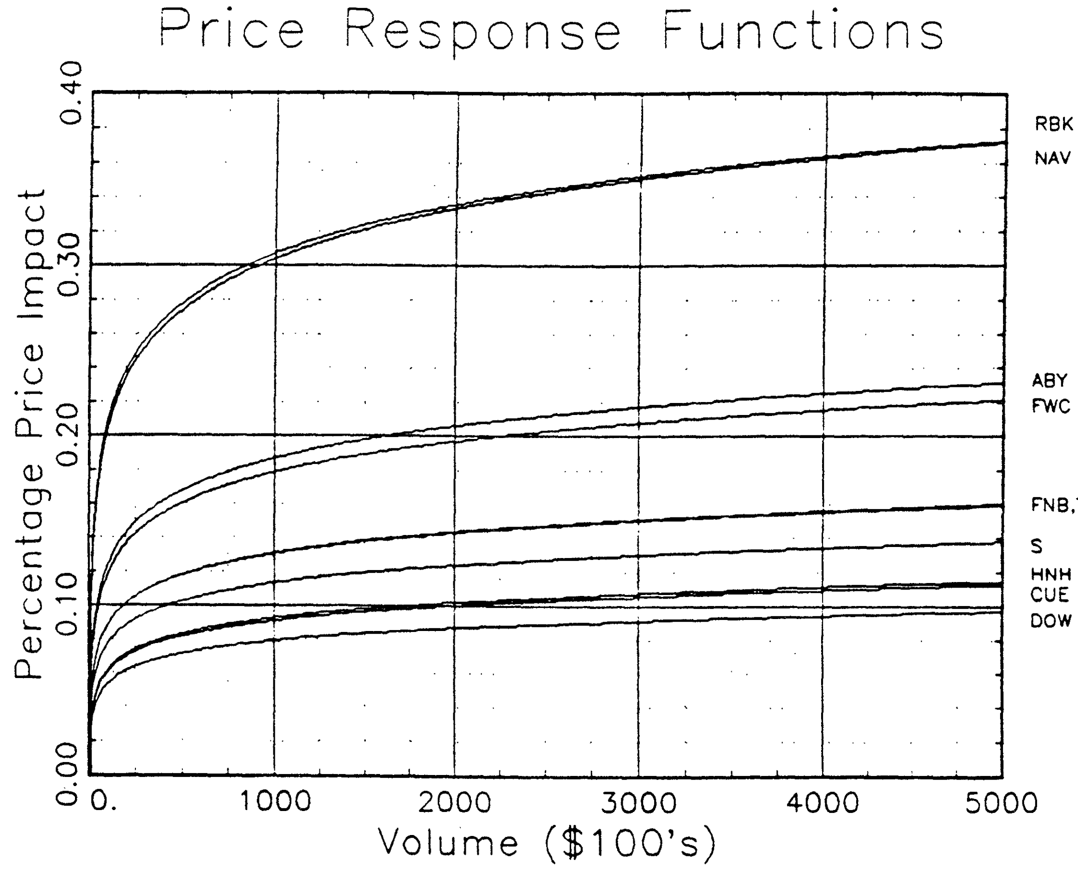

Figure 4.

Percentage price impact as a function of dollar volume computed from ordered probit probabilities, conditional on the three most recent trades being buyer-initiated, and the three most recent price changes being +1 tick each, for Abitibi-Price Incorporated (ABY – 1,515 trades), Quantum Chemical Corporation (CUE – 27,141 trades), Dow Chemical Company (DOW – 81,916 trades), First Chicago Corporation (FNB – 17,915 trades), Foster Wheeler Corporation (FWC – 18,460 trades), Handy and Harmon Company (HNH – 3,621 trades), Navistar International Corporation (NAV – 90,212 trades), Reebok International Limited (RBK – 62,512 trades), Sears Roebuck and Company (S – 90,262 trades), and American Telephone and Telegraph Company (T – 178,813 trades), for the sample period from 4 January 1988 to 29 December 1988. Percentage price impact is computed as a percentage of the average of the high and low prices.

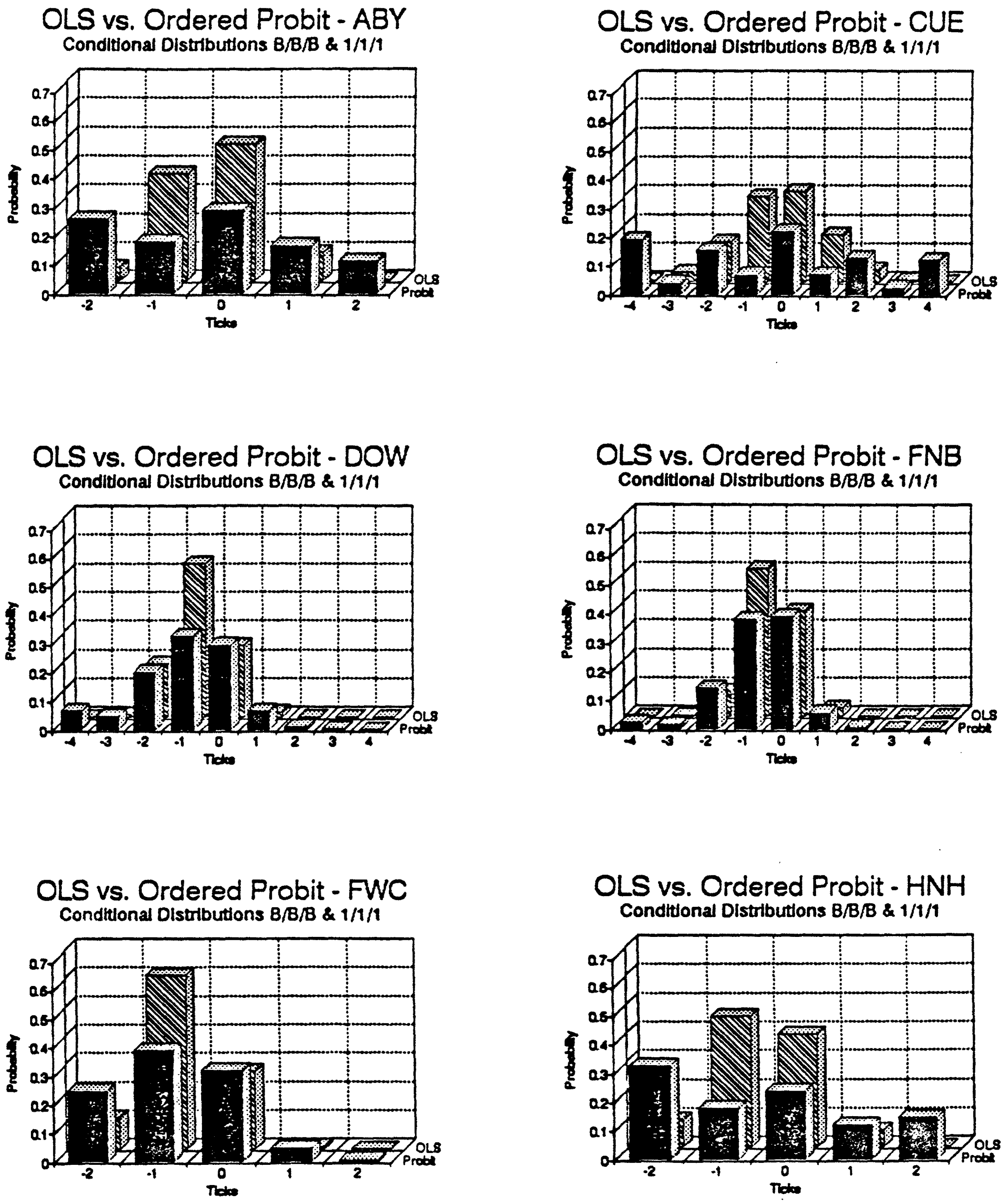
OLS vs. Ordered Probit - ABY
Conditional Distributions B/B/B & 1/1/1
OLS vs. Ordered Probit - CUE
Conditional Distributions B/B/B & 1/1/1
OLS vs. Ordered Probit - DOW
Conditional Distributions B/B/B & 1/1/1
OLS vs. Ordered Probit - FNB
Conditional Distributions B/B/B & 1/1/1
OLS vs. Ordered Probit - FWC
Conditional Distributions B/B/B & 1/1/1
OLS vs. Ordered Probit - HNH
Conditional Distributions B/B/B & 1/1/1
Probability
Ticks
OLS
Probit

Figure 5.

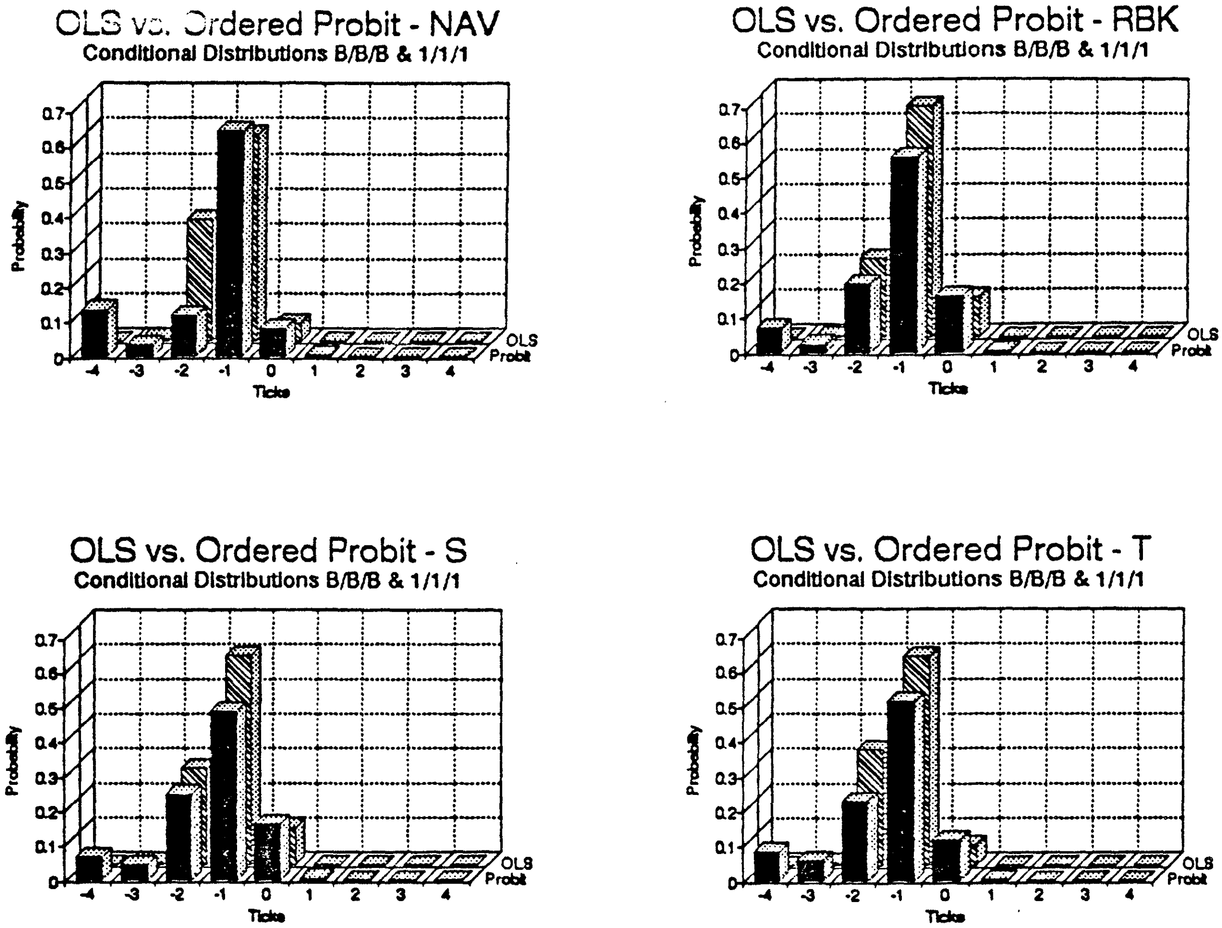

Figure 5 (Continued).