

NEW YORK UNIVERSITY
STERN SCHOOL OF BUSINESS
FINANCE DEPARTMENT

Working Paper Series, 1995

Modeling Market Microstructure Time Series.

Hasbrouck, Joel.

FIN-95-24

Modeling Market Microstructure Time Series.

Joel Hasbrouck

March 9, 1995

This draft: February 16, 1996

Preliminary Draft
Comments Welcome

Department of Finance
Stern School of Business
New York University
Suite 9-190
44 West 4th St.
New York, NY 10012-1126

E-mail: jhasbrou@stern.nyu.edu
Web: <http://www.stern.nyu.edu/~jhasbrou>

Copyright 1996, Joel Hasbrouck, All rights reserved.

All errors are my own responsibility.

Modeling Market Microstructure Time Series

Abstract

Microstructure data typically consist of trades and bid and offer quotes for financial securities that are collected at fine sampling intervals (often within the day). This paper reviews approaches taken to modeling these data. The emphasis is on techniques of stationary multivariate time series analysis: autoregressive and moving average representations of standard microstructure models, vector autoregressive estimation, random-walk decompositions and cointegration. The paper also discusses the challenges posed by irregular observation frequencies, discreteness and nonlinearity.

Modeling Market Microstructure Time Series

Contents

1. INTRODUCTION.....	1
2. SIMPLE UNIVARIATE MODELS OF PRICES.....	6
3. SIMPLE BIVARIATE MODELS OF PRICES AND TRADES.	14
4. GENERAL SPECIFICATIONS.	26
5. TIME.	34
6. DISCRETENESS.....	39
7. NONLINEARITY.	42
8. MULTIPLE MECHANISMS AND MARKETS.....	43
9. SUMMARY AND DIRECTIONS FOR FURTHER WORK.....	50
10. REFERENCES.....	53

1. Introduction

Market microstructure is the area of financial economics that focuses on the trading process. Factors both practical and academic are motivating research here. On the practical side, innovation in financial markets has resulted in increased trading volume in standard securities (stocks, bonds, etc.), creation of new types of securities, and greater experimentation with alternative trading mechanisms. From the academic perspective comes a fuller understanding of the role played by trading in the incorporation of new information into security prices. Empirical work in the area has also benefited from the increasing availability of detailed transaction data.

Microstructure research seeks to address two sorts of questions. The first belong to the study of markets narrowly defined: how should transaction costs be estimated; what are the optimal trading strategies; and, how should markets be organized? The second and broader set of questions arises from the role that the market plays in price discovery (the incorporation of new information into the security price): how can we characterize the determinants of security value that we loosely refer to as public and private information? Ultimately these two types of questions are related. The organization of a market may affect the transactions costs, and therefore the net return to an investor, the valuation of the asset and the allocation of real resources (Amihud and Mendelson (1986)). Conversely, the characteristics of an asset (risk, return, homogeneity, divisibility) may favor certain holding patterns among investors and certain market structures (Grossman and Miller (1988)).

Empirical microstructure analyses draw on three areas of knowledge. The first is comprised by the formal economic models of individual behavior that offer substantive predictions about how observable variables should behave. The second area is statistical time series analysis. The third area concerns the institutional realities: the actual procedures by which individuals and automated systems work to accomplish trades in a particular market.

The theoretical work in market microstructure has centered around several reasonably well-defined paradigms that serve as a common basis for variations. The

evolution of thought on security transaction price behavior has passed from basic martingale models, to noninformational cost models (order processing and inventory control paradigms), and finally to models that incorporate the distinctly informational and strategic aspects of trading. Although this paper will describe the intuitions behind these models, it does not present a rigorous discussion. O'Hara (1994) provides a comprehensive textbook discussion that establishes much of the economic background for this paper.

Present empirical work in microstructure is characterized by a wide diversity of techniques. Market data exhibit a panoply of features that are hostile to statistical modeling: complex dynamics, nonlinearities, nonstationarities, and irregular timing to name a few. The impracticality of modeling all of these features jointly, in a specification that can also potentially resolve alternative economic hypotheses, leads to a multitude of more modest models that simply try to capture one or two phenomena relevant to the problem at hand.

To establish a common footing, however, the models considered in this paper are cast in the framework of linear multivariate time series analysis. Most of the statistical techniques discussed here were originally developed and applied to macroeconomic time series. (Lutkepohl (1993) and Hamilton (1994) are excellent textbook presentations.) The reader approaching the present paper from a macro perspective will find most of the time series results familiar. But time series analysis is not a mechanical procedure, and the application of any technique to a new problem involves some reflection on the economics of the situation and the nature of the data. Some issues that cause great difficulty in macro applications are conveniently absent in microstructure data: microstructure observations are exceedingly numerous and the fine time intervals over which the data are collected greatly mitigate the simultaneity induced by time aggregation. On the other hand, microstructure data often exhibit troublesome properties such as discreteness that rarely arise in macro analyses.

Except as necessary to motivate the economic or statistical material, this paper does not discuss the institutional details of particular markets. For reasons of data

availability, however, most empirical work has focused on U.S. equity markets, particularly the New York Stock Exchange (NYSE). Hasbrouck, Sofianos and Sosebee (1993) discuss the NYSE in detail. The NYSE and other U.S. and non-U.S. equity markets are described in Schwartz (1988 and 1991).

In contemplating the various empirical approaches to microstructure modeling, it is useful to bear in mind two dichotomies or principles of differentiation. The first dichotomy arises from the issues to which microstructure analysis is commonly addressed: the narrowly defined questions of market design and operational market performance vs. the broader informational and security valuation issues. From an economic perspective, the actual security price in many microstructure models can be interpreted as an idealized “informationally efficient” price, corrupted by perturbations attributable to the frictions of the trading process. From an empirical viewpoint, the distinction can loosely be viewed as one based on time horizon. New information imparts a permanent revision to the expectation of a security’s value, while microstructure effects are short-lived and transient. The first principle, then, is the dichotomy of security price variations into permanent (informational) and transitory (market-friction-related) components.

The second dichotomy addresses the *source* of the price variations, as to whether or not they are trading-related, i.e., attributable to one or more transactions. This distinction is more subtle than the first, because while the difference between permanent and transitory components arises frequently in economic analysis, the preoccupation with the role of trades *per se* in price determination is largely peculiar to microstructure studies.

For the present purpose, the most important aspects of a trade are the fact and time of its occurrence, the price and volume (quantity), and whether the trade was initiated by the buyer or the seller. This last characteristic may require some elaboration. Academic economists have long reacted to lay statements like, “Heavy buying drove stock prices higher today,” with retorts along the lines of, “So, there were no sellers?” Certainly there must be a seller for every buyer. At a fine level of observation, however, it is often sensible to identify the active and passive sides of the transaction. The active transactor can be viewed (in the sense of Demsetz (1968)) as the agent who seeks to trade

immediately, and is willing to pay a price to do so. The passive transactor is the supplier of immediacy. In many security markets, for example, the passive traders are those who post bid and offer quotes (indicated prices at which they are willing to buy or sell), and wait. The traders who impatiently demand an immediate trade, and accept one of the quotes (hitting the bid or lifting the offer) are active.

A trade can affect both the permanent and transitory components of the price. The permanent effect is informational. In asymmetric information models, the informational impact of a trade is attributed to market's estimate of the private information content of the trade. The price rises in response to a buyer-initiated trade, for example, in accordance with the market's assessment of the chances that the trade was initiated by positive information known to the buyer, but not to the public. The portion of the permanent price movements that can be attributed to trades is therefore related to the degree of information asymmetry concerning the firm's value. From a statistical viewpoint, it may be measured by the explanatory power of trade-related variables in accounting for price changes.

The transitory price effect of a trade is a perturbation induced by the trade that drives the current (and possibly subsequent) transaction prices away from the corresponding informationally accurate (permanent component) prices. For a particular trade, this divergence may sometimes be interpreted as a trading cost. In simple bid-ask spread models, for example, the divergence corresponds to a cost paid by the active trader to the passive trader. More generally, the trade-related transitory effect will reflect influences such as price discreteness and inventory control (position management) by dealers.

For the sake of completeness it should be mentioned that both permanent and transitory price components may be due to considerations not directly related to trades. Security prices (or indicated prices) react to public information, such as news releases. The permanent effect of a public news release is informational. Any lagged adjustment toward to new permanent price would constitute a transitory component.

The principal dichotomies of permanent vs. transitory and trade-related vs. trade-unrelated are summarized in Table 1. For each combination, the table gives economic examples and also considerations useful in empirical resolution. These will be discussed at length in the following sections.

Although these distinctions are useful for classification and exposition, this simplicity comes at the cost of neglecting economic considerations that cross over these dichotomies. As noted earlier, the operational features of a security market may affect the informational characteristics of a security and vice versa. However, many useful analyses can proceed under plausible *ceteris paribus* assumptions. Assuming that market structure stays fixed, one may want to examine shifts in information characteristics surrounding corporate announcements. Alternatively, assuming that the informational structure stays fixed, one might want to examine the effect of a change in the tick size (minimum price increment). The literature contains examples of both sorts of analyses.

While an overview of any sort requires the imposition of some classification scheme, the particular perspective adopted here follows from a personal preoccupation with the dynamic properties of microstructure data. One could organize a survey historically or from the perspective of different market participants, perhaps with equal justification. Nor is the perspective adopted here is an exhaustive one. I attempt to point the reader to approaches that lie outside of this framework, but cannot claim to do justice to these studies.¹

The organization of the paper is as follows. The next two sections describe the basic economic paradigms of market microstructure using simple structural models. Section 4 presents a general statistical framework in which the diverse microstructure effects can be accommodated while maintaining the two distinctions described above. The next sections address particular characteristics of microstructure data that lie beyond (or at least at the fringes of) conventional techniques: irregular timing of market events such as

¹ A recent survey by Goodhart and O'Hara (1995) provides more background on volatility modeling and non-equity market applications.

trades (Section 5); price discreteness (Section 6); nonlinearities in the trade-price relation (Section 7); and multiple security / multiple market situations (Section 8). A summary concludes the paper in Section 9.

2. Simple univariate models of prices.

2.1. *Martingales and the random-walk model.*

The efficient markets hypothesis of financial economics generally implies that a security price (perhaps normalized to reflect an expected return) behaves as a martingale, a stochastic process with unforecastable changes (Samuelson (1965) and Fama (1970)). A special case useful for empirical work is the homoskedastic random walk, wherein the evolution of the security price p_t is given by

$$p_t = p_{t-1} + w_t \quad (2.1)$$

Where the w_t are disturbances with $E w_t = 0$, $E w_t^2 = \sigma_w^2$ and $E w_t w_\tau = 0$ for $t \neq \tau$. These unforecastable increments derive from updates to the market's information set (cf. Table 1). This model is often generalized to include an unconditional expected price change or return, but for reasons both expositional and practical (described below) this component is omitted in the present discussion.

The martingale property typically arises because the fundamental security valuation in many models is characterized as a conditional expectation of the security's terminal (liquidation) cash flow. A sequence of conditional expectations is a martingale (Karlin and Taylor (1975, p. 246)). For the actual security price to behave as a martingale, however, additional structure must be imposed. The hypothesis that transaction prices behave as a random walk rests on assumptions (most importantly, the absence of transaction costs) that do not hold even approximately at the level of the microstructure phenomena considered in this paper.

The random-walk model is nevertheless a useful point of departure. Even if the (martingale) conditional expectation does not completely determine the security price, it certainly constitutes a component that is large and economically important. Accordingly, even for models in which actual transaction price processes exhibit complicated

dependencies, examination of the random-walk component of the price will illuminate the informational structure of the market. Furthermore, the departure of actual prices from the implicit martingale component may be used to illuminate the costs of transacting in the market.

In embedding the random-walk model in microstructure frameworks, however, one should bear in mind the importance of the conditioning information. A price p_t is said to be a martingale with respect to a (possibly vector-valued) information process Φ_t if $E[p_{t+1} | \Phi_0, \Phi_1, \dots, \Phi_t] = p_t$. If the conditioning information includes the price ($p_t \subset \Phi_t$), then $E[p_{t+1} | p_0, p_1, \dots, p_t] = p_t$. This ensures that the increments w_t in (2.1) are unforecastable.

The assertion that $p_t \subset \Phi_t$ is frequently supported by institutional fact. Most of the early theoretical and empirical work on market efficiency focused on U.S. equity markets, for which transaction prices are promptly reported and widely disseminated. Many markets, however, such as the U.S. government securities market, do not enforce trade reporting, or, as in the case of the London equities market, permit delayed reporting of certain trades (Naik, Neuberger and Viswanathan (1994)).

In the absence of prompt trade reporting, the fallback justification of (2.1) is that the transaction price is redundant, i.e., that it contains no new information beyond that available in the public information set. This view is unattractive because current economic thought accords great significance to the role played by prices as aggregators or signals of *private* information. In summary then, the random-walk model, which is a component of most of the specifications discussed in this paper, is only appropriate in markets with prompt transaction reporting. Absent this disclosure, other approaches must be used. Instead of using transaction prices that may not be widely disseminated, for example, it may be preferable to use dealer bid and offer quotes.

Correct specification of the conditioning information at the transaction level may be exceedingly difficult because knowledge will often differ in a subtle fashion across participants by reason of proximity to the market and cost. For example, the contents of the book (pending orders) on the Tokyo Stock Exchange are publicly available in the

sense that anyone may obtain the information from his or her broker. But the data are electronically transmitted only in response to an inquiry and only to the broker's lead office (Hamao and Hasbrouck (1995), Lehmann and Modest (1994)). Costs of information acquisition that are small at long time lags may become large over microstructure time frames. Daily closing security prices are available for the price of a newspaper, for example, while immediate updates require expensive real-time data feeds.

The preceding remarks are intended to heighten the reader's sensitivity to informational issues that are often suppressed (in the interests of tractability) in the formal models. When aspects of these models are incorporated into specifications and estimated for real market data, these considerations usually warrant at least some qualification of the conclusions.

Equation (2.1) is specified in terms of price levels. It is often useful to interpret p_t as the natural logarithm of the price, in which case the first difference is a continuously compounded rate of return. This is particularly convenient when the analysis covers multiple securities spanning a wide range of prices, and in many applications does not affect the conclusions. It should be borne in mind, however, that most of the formal models are constructed using price levels. Furthermore, certain microstructure phenomena (discreteness, in particular) depend fundamentally on the price level.

Many tests have been proposed and applied to the problem of determining whether stock prices follow a random walk over daily or longer intervals (Fama (1970) and Lo and MacKinlay (1988)). At the level of transaction prices, however, the random-walk conjecture is a straw man, a hypothesis that is very easy to reject in most markets even in small data samples. In microstructure, the question is not "whether" transaction prices diverge from a random walk, but rather "how much?" and "why?" For the present, however, it is useful to discuss several aspects of estimation in random-walk models that will also apply in more realistic situations.

Microstructure data sets typically contain by large numbers of observations (often in the thousands for each security) over a relatively brief period of calendar time (such as a few months). To the econometrician seeking to estimate the parameters of a

microstructure model, the abundance of observations appear to hold out the promise of high precision. Unfortunately, when the number of observations is a consequence of fine sampling (rather than a long span of calendar time), the increase in precision is partially illusory. In particular, Merton (1980) shows that while precision of the estimate of variance per unit time increases, that of the mean estimate does not. In view of the large estimation errors for the mean, Merton suggests estimating the variance using the noncentral sample moment.

There are two practical implications of this for transaction-level analyses. First, if we are willing to accept a small bias in our estimates, the precision of these estimates is enhanced by ignoring the unconditional expected return (suppressing the intercept in price-change specifications). The discussions that follow do this as a matter of routine, although it is usually a simple matter to add a non-zero expected return. Second, tests of economic hypotheses that are based on second moments (variances and covariances) are likely to be more powerful than those that rely on first moments.

2.2. *Models with random pricing errors.*

It is useful to generalize the random-walk model by allowing the security price to reflect a stationary disturbance in addition to the random-walk component. The general structural model is:

$$\begin{aligned} m_t &= m_{t-1} + w_t \\ p_t &= m_t + s_t \end{aligned} \tag{2.2}$$

Here, the random-walk term is m_t , which may be interpreted as an implicit efficient price, where (as in (2.1)) the w_t are unforecastable increments arising from updates to the conditional expectation of the security's terminal value. The second component in the price equation (s_t) is a stationary component that for the moment can be viewed in an ad hoc fashion as a residual or perturbation that drives the transaction price away from the implicit efficient price.

Model (2.1) establishes the first of the principal dichotomies alluded to in the introduction (cf. Table 1). The informational aspects of a model may be characterized by

analysis of the m_t or the w_t . The noninformational features show up in the s_t . Since the dichotomy is not observable, some additional structure must be imposed on the problem in order to make substantive statements. It is often useful to estimate the w_t and the s_t at a point in time (as a function of various sets of conditioning information), to estimate the variances σ_w^2 and σ_s^2 , and to ascertain the components of these variances. In a sense, most of this paper is devoted to consideration of the full generality of (2.1).

The motivation for and interpretation of w_t are essentially the same as in the random walk model. The new feature that has been introduced is the stationary pricing error. The terminology stems from its role as a discrepancy between the implicit efficient price and the actual transaction price. If $s_t > 0$, then there is a sense in which the buyer lost (paid in excess of the efficient price) and the seller gained. Aggregating over the buyer and seller, s_t is a zero-sum game. If s_t were randomly distributed over trades and traders, then one would be tempted to argue its irrelevance by the law of large numbers. Equality of traders in real markets, however, is a poor assumption. Agents' characteristics (small trader, large trader or dealer) have a large effect on the sort of prices they give and take, and it is therefore likely that the pricing error will induce systematic distributional effects.

2.3. *The simple bid-ask spread model.*

A useful special case of the preceding model arises from the following trading process. The implicit efficient price is common knowledge to all participants. A market-maker or dealer in the security posts a price at which he is willing to buy (the bid price) and a price at which he is willing to sell (the offer or ask price). These bid and ask quotes will be denoted q_t^b and q_t^a , and the difference between them is termed the spread, $S_t = q_t^a - q_t^b$. In economic terms, this spread can be viewed as a consequence of the dealer's need to recover fixed transaction costs and a normal profit (Tinic (1972)). Alternatively, the spread may arise endogenously from the choices of traders deciding between market (active) and limit (passive) orders, as in Cohen, Maier, Schwartz and Whitcomb (1981). These are noninformational spread models; other alternatives will be considered below.

Assuming that the spread is constant at S , that the bid and ask quotes are set to bracket symmetrically the implicit efficient price ($q_t^b = m_t - S/2$ and $q_t^a = m_t + S/2$), and that at each time point, an agent arrives at the dealer and either buys (at price q_t^a) or sells a single unit of the security (at q_t^b). The full model is now

$$\begin{aligned} m_t &= m_{t-1} + w_t \\ p_t &= m_t + c_t \\ c_t &= \pm S/2 \end{aligned} \tag{2.3}$$

The vacillations of c_t are sometimes called “bid-ask bounce”.

The market mechanics imply that c_t in (2.3) is a stationary random process with the following properties: $E c_t = 0$; $E c_t^2 = \sigma_c^2$; $E c_t c_\tau = 0$ for $t \neq \tau$ and $E c_t w_\tau = 0$ for all t, τ .

The first three properties establish c_t as a zero-mean homoskedastic random variable with no serial correlation. The fourth property asserts that it is uncorrelated with the information process, i.e., that the increments in the implicit efficient price are not trade-related. By comparing this model with (2.2) it is apparent that $c_t = s_t$, the pricing error. The variance of the pricing error is a useful summary measure of how close actual transaction prices track the implicit efficient price. In this model, $\sigma_s^2 = \sigma_c^2 = S^2/4$.

In this model s_t is clearly driven by the incoming trade (buy or sell).. In modern microstructure data sets, these trades (or convenient proxies) are often observable, and it is possible to model them directly. Representative bivariate price and trade models will be discussed extensively below. Many older historical data sets, however, are limited to transaction prices. We therefore consider inference based only on these prices.

We are in effect attempting to make inferences about the two unobserved components of the transaction price, m_t and $s_t (= c_t)$. The price changes are:

$$\Delta p_t = p_t - p_{t-1} = w_t + s_t - s_{t-1}. \tag{2.4}$$

with first and second-order autocovariances given by $\gamma_0 = E \Delta p_t^2 = \sigma_w^2 + 2\sigma_s^2$ and $\gamma_1 = E \Delta p_t \Delta p_{t-1} = -\sigma_s^2$. The autocovariances at higher orders are zero. From these first two autocovariances (or estimates thereof), we may solve for σ_s^2 and σ_w^2 . Most importantly, the spread is given by

$$S = 2\sigma_c = 2\sigma_s = 2\sqrt{-\gamma_1}. \quad (2.5)$$

The last expression is commonly known as Roll's (1984) estimate of the spread. This obviously requires $\gamma_1 \leq 0$. Harris (1990) discusses the statistical properties of this estimator.

Another useful characterization of this model is the innovations or moving average form. A process that possesses zero autocovariances beyond the first lag may be characterized as a first order moving average (MA(1)) process:

$$\Delta p_t = \varepsilon_t + \theta \varepsilon_{t-1}. \quad (2.6)$$

where the ε_t are serially uncorrelated homoskedastic increments. By equating the price change autocovariances implied by (2.4) and (2.6), the correspondence between the two sets of parameters may be established. In the one direction, $\sigma_w^2 = (1 + \theta)^2 \sigma_\varepsilon^2$ and $\sigma_s^2 = -\theta \sigma_\varepsilon^2$.

There is a useful intuition behind the expression for σ_w^2 . The impulse response function of a time series model specifies how the variables react to particular initial shocks. Suppose in the present case that the lagged innovations $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ are zero. If the innovation at time t is nonzero, the expected current and subsequent price changes implied by equation (2.6) are $E[\Delta p_t | \varepsilon_t] = \varepsilon_t$, $E[\Delta p_{t+1} | \varepsilon_t] = \theta \varepsilon_t$, and $E[\Delta p_{t+k} | \varepsilon_t] = 0$ for $k > 0$. The cumulative expected price change is therefore

$$E[\Delta p_t + \Delta p_{t+1} + \Delta p_{t+2} + \dots | \varepsilon_t] = (1 + \theta) \varepsilon_t. \quad (2.7)$$

This is the long-run expected price impact of an innovation, i.e., the informational impact of the innovation. This implies $w_t = (1 + \theta) \varepsilon_t$, from which the expression for σ_w^2 follows immediately. In the discussions that follow, impulse response functions are often used to characterize the dynamic properties of structural models.

While many economic hypotheses of interest can be addressed by considering the variances of the random-walk and pricing error components, it is often desirable to know w_t and s_t at a particular time. On the basis of the transaction prices these quantities are not identified in this model (even if we condition on prices subsequent to t), although filtered estimates are attainable.

2.4. Lagged price adjustment.

The simple bid-ask model predicts that the price change will exhibit a negative first-order autocovariance. This is in fact usually the case in transaction price data. The model may be generalized to permit price change dependencies at orders higher than one by introducing lagged price adjustment. Goldman and Beja (1979) suggest that security dealers do not instantaneously adjust their quotes to new information, but do so gradually.

More generally, lagged adjustment can arise from lagged dissemination of information, price smoothing by market makers and discreteness. Other analyses that feature lagged adjustment are Amihud and Mendelson (1987), Beja and Goldman (1980), Damodaran (1992) and Hasbrouck and Ho (1987).

A simple lagged-adjustment model is given by:

$$\begin{aligned} m_t &= m_{t-1} + w_t \\ p_t &= p_{t-1} + \alpha(m_t - p_{t-1}), \end{aligned} \quad (2.8)$$

where α is an adjustment speed parameter. (The spread is suppressed here in order to focus on the lagged adjustment.) The price dynamics implied by this model may be illustrated with an impulse response function. Figure 1 depicts the price subsequent to a one-unit shock in the efficient price ($w_0 = 1$), assuming an adjustment parameter of $\alpha=0.5$. At each step, half of the remaining adjustment is made toward the efficient price. If $0 < \alpha < 1$, this adjustment is monotonic.

By substitution from (2.8), it is seen that price changes are generated as the first-order autoregressive process: $\Delta p_t = (1 - \alpha)\Delta p_{t-1} + \alpha w_t$. If the estimated model is $\Delta p_t = \phi \Delta p_{t-1} + \varepsilon_t$, the structural parameters may be computed as: $\sigma_w^2 = \sigma_\varepsilon^2 / (1 - \phi)^2$, $\alpha = 1 - \phi$. As in the simple bid-ask spread model, σ_w^2 has an impulse response interpretation. The random-walk innovation may be computed as $w_t = (1 + \phi + \phi^2 + \dots)\varepsilon_t = (1 - \phi)^{-1}\varepsilon_t$, which effectively sums each period's contribution to price subsequent to the initial disturbance. The pricing error is $s_t = p_t - m_t$, which implies $s_t = (1 - \alpha)s_{t-1} - (1 - \alpha)w_t = \phi s_{t-1} - \phi w_t$ and $\sigma_s^2 = [\phi^2 \sigma_\varepsilon^2] / [(1 - \phi^2)(1 - \phi)^2]$.

Since there is one disturbance driving this model (w_t), both w_t and s_t can be recovered from the price record. This is a stronger result than obtained in the simple bid-

ask spread model. From a time-series perspective, this is due to the fact that the stationary component in the present model is an exact linear function of past w 's. In the simple bid-ask model, whether the trade took place at the bid or the ask (i.e., the value of s_t) is independent of w_t .

3. Simple bivariate models of prices and trades.

The univariate price models described above are capable of exhibiting dynamics that reflect microstructure phenomena and can also capture the first dichotomy mentioned in the introduction, that between permanent (informational) and transient (market) effects. The models described in this section encompass trades as well, with a view toward establishing the second important distinction, that between trade-related and -unrelated sources of price variation.

3.1. Inventory models.

Buyers and sellers in the simple bid-ask spread model are assumed to arrive independently and with equal probability. Let x_t denote the signed trade quantity, positive if the arriving trader buys from the dealer and negative if the trader sells. The cumulative quantity from time zero through time t is $\sum_{i=0}^t x_i$. In the paper that introduced the term "microstructure", Garman (1976) pointed out that as t increased, this sum would diverge, implying that the dealer bought or sold (net) an infinite amount. Real-world dealers face capital constraints, however, and would in any event avoid large positions due to risk-aversion. This motivates the need for some sort of inventory control or position management.

The inventory control problem in classical microeconomics is one of specifying a restocking strategy subject to order and stock-out costs. The security market dealer, on the other hand has traditionally been supposed to achieve inventory control by shifting the quotes to elicit an imbalance of buy and sell orders. Formal models of this effect include Amihud and Mendelson (1980), Ho and Stoll (1981), O'Hara and Oldfield (1986) and Stoll (1978).

As an illustration, consider a generalization of the simple bid-ask spread model in which quote-setting is depends on the dealer's inventory position and incoming order flow depends on the quotes:

$$\begin{aligned}
 m_t &= m_{t-1} + w_t \\
 q_t &= m_t - bI_{t-1} \\
 I_t &= I_{t-1} - x_t \\
 x_t &= -\alpha(q_t - m_t) + v_t \\
 p_t &= q_t + cx_t
 \end{aligned} \tag{3.1}$$

The first equation describes the random-walk evolution of the efficient price. The quotes are summarized by the quote midpoint (the average of the bid and ask quotes), q_t . This is equal to the efficient price plus an inventory control component, where I_t is the dealer's inventory at the close of period t . Without loss of generality, the dealer's target inventory is assumed to be zero. The quote-midpoint equation specifies that with $b > 0$, the dealer lowers his price if he has a long position. The net demand, x_t , is driven by a price sensitive component ($\alpha > 0$) and a random component. The usefulness of the quote position as an inventory-management tool is based on the demand price elasticity.

Since the dealer is assumed to be the counterparty to all trades, the change in inventory is equal to the negative of the net demand. The transaction price is equal to the quote midpoint, plus a cost component cx_t . This cost is proportional to trade size: rather than quoting a bid and offer price, the dealer quotes a linear bid and offer schedule. A trader wanting to buy an amount $|x_t|$ will be quoted an ask price of $q_t^a = q_t + c|x_t|$, and a trader wanting to sell will be quoted a bid price of $q_t^b = q_t - c|x_t|$. The trade innovation v_t is assumed to be serially correlated, and uncorrelated at all leads and lags with w_t .

The essential features of this model can be illustrated by examining the impulse response function for a particular set of parameter values. Let $\alpha=0.8$, $b=0.04$ and $c=0.5$, and consider the paths of price and inventory subsequent to a trade shock at time zero of $v_0=1$, i.e., a purchase of one unit from the dealer. These paths are graphed in Figure 2. The buy is associated with an immediate price jump due to the cost component. Reversion is not immediate, however. Subsequent to the trade, the dealer has a inventory

shortfall and must raise his quotes to elicit an incoming sell order. As the sell orders arrive (in expectation), the dealer resets the quotes to the initial level. The inventory path reflects the initial depletion caused by the purchase (*from* the dealer) and the subsequent sales (*to* the dealer). At the end of the adjustment process, both price and inventory have completely reverted. There is no permanent price impact of a trade in this model because trades are independent of information.

The permanent component of the price change is w_t , which is due entirely to public information. The pricing error is:

$$s_t = p_t - m_t = cx_t - bI_{t-1} \quad (3.2)$$

This is entirely trade-driven. As in the simple bid-ask model, the buyer pays the half-spread cx_t . The second term depends on the dealer's previous inventory position. If the dealer happened to have an inventory surplus, the buyer's cost would be reduced.

If both p_t and I_t are observable, the model may be written as:

$\Delta p_t = -cI_t + (2c - b)I_{t-1} + (b - c)I_{t-2} + w_t$ and $I_t = (1 - ab)I_{t-1} - v_t$. Formally, this is a bivariate vector autoregressive (VAR) model, with a contemporaneous recursive structure, which may be estimated directly by least squares. There is sufficient structure here to recover both w_t and s_t from current and past observations.

Among the various sorts of microstructure data available, however, dealer inventory data are about the rarest. Implicit in these data are the dealer's trading strategies and trading profits, both of which are usually kept private. If I_t is not known, then inference must proceed solely from prices. On the basis of the univariate time-series representation of the price changes, the structural model is underidentified. Two important structural parameters are identified, however: the variances of the random-walk and pricing error components.

Due to the paucity of inventory data, there are few analyses of pure inventory control models. In a U.S. S.E.C. (1971) study, Smidt presents some results for NYSE stock specialists based on daily positions and price changes. Ho and Macris (1984) estimate a transaction level model for an American Stock Exchange options specialist.

Most recent studies allow for the possibility of asymmetric information in addition to inventory control, and these are discussed below.

3.2. *Asymmetric information.*

The models considered to this point have assumed that all market participants possess the same information. This sort of public information may be thought of as instantaneous news releases, in response to which bid and offer quotes would adjust with no necessity of trading. The most important recent developments in theoretical microstructure, however, have been models that allow for heterogeneously informed traders. If a trade might be motivated by superior information, the occurrence of a trade (a public event in most models) will communicate to the market something about this private information. Some studies that initially addressed this phenomenon in microstructure settings are Bagehot (1971), Copeland and Galai (1983), Glosten and Milgrom (1985), Kyle (1985) and Easley and O'Hara (1987). O'Hara (1994, Ch. 3) provides an overview.

A simple model of private information with fixed transaction costs can be given as:

$$\begin{aligned}
 m_t &= m_{t-1} + w_t \\
 w_t &= u_t + gx_t \\
 q_t &= m_{t-1} + u_t \\
 p_t &= q_t + cx_t
 \end{aligned}
 \tag{3.3}$$

Relative to the earlier models, the novelty here is in the random-walk innovation, w_t . It is now composed of two components. The first, u_t , is assumed to reflect updates to the public information set. The second, gx_t , with $g > 0$, reflects the market's estimate of the information contained in the trade. For this component to be serially uncorrelated, it must be the case that x_t is serially uncorrelated, i.e., we are back to assuming that buy and sell orders arrive randomly. This model is a variant of one suggested by Glosten (1987).

Actual transaction prices are subject to a bid-ask spread related to the direction of the trade. There are two ways of interpreting the cx_t term in the price specification. First, if the magnitude of the trade is fixed, say $x_t \in \{-1, +1\}$, then c is one-half the bid-ask spread ($S/2$), with transactions occurring at the bid and offer prices ($q_t^b = q_t - S/2$ and

$q_i^a = q_i + S/2$). Alternatively, if trade size is continuous, then c gives the slope of the dealer's linear bid and offer schedule.

The dynamic behavior of prices and trades may be illustrated by the impulse response function based on parameter values $c=0.5$ and $g=0.2$, subsequent to an initial buy order of one unit ($x_0=1$). These are graphed in Figure 3. The initial price jump simply reflects the bid-ask bounce, but in contrast with the inventory control model, the reversion is not total. Of the initial 0.5 price jump, 0.2 is the inferred information content, which remains permanently impounded in the stock price. By assumption there are no serial dependencies in trades: the initial purchase engenders no subsequent order flow effects.

The evolution in the efficient price now reflects both public and private information components, so

$$\sigma_w^2 = \sigma_u^2 + g^2 \sigma_x^2, \quad (3.4)$$

which isolates the non-trade and trade-related components of the efficient price change. A useful summary measure of the relative importance of trades in explaining movements in the efficient price is the proportion

$$R_{w,x}^2 = g^2 \sigma_x^2 / \sigma_w^2 \quad (3.5)$$

The R^2 notation denotes the usual "proportion of total variance explained." This measure generalizes beyond the present model, and is a useful proxy for the extent of asymmetric information.

The private information effects in this model reflect the market's beliefs about the probabilistic structure of the private information, not the actual level of private information. That is, the price impact of a particular trade depends only on the market's general beliefs about extent and nature of private information, and not directly on the actual information possessed by the trader. A model of this sort cannot be used to identify, for example, illegal insider trades in a sample of data.

The pricing error is

$$s_i = p_i - m_i = (c - g)x_i \quad (3.6)$$

The pricing error is entirely trade driven. Relative to the simple bid-ask model with no private information, however, s_t is reduced by the information content of the trade, gx_t . It is generally assumed that $c > g$ because the dealer is setting the half-spread to recover both information costs g and additional order processing costs.

The return series is given by:

$$\Delta p_t = p_t - p_{t-1} = u_t + cx_t - (c - g)x_{t-1} \quad (3.7)$$

If trades and prices are observable, this may be estimated directly. Early transaction-based estimations of trade impacts on price are Marsh and Rock (1986), Glosten and Harris (1988), and Hasbrouck (1988).

When trades are not observed, however, the inference must proceed solely on the basis of transaction prices. This model superficially resembles the simple bid-ask model considered in section 2.3. Like the earlier model, it possesses an MA(1) representation of the form (2.6). Here, however, the two parameters of the MA model $\{\sigma_s^2, \theta\}$ are insufficient to identify the four parameters of the structural model $\{c, g, \sigma_u^2, \sigma_x^2\}$. The random walk variance is identified as before: $\sigma_w^2 = (1 + \theta)^2 \sigma_s^2 = \sigma_u^2 + g^2 \sigma_x^2$. In contrast with the earlier model, however, we cannot assume that the pricing error is uncorrelated with the increment to the efficient price.

The connection to the simple model may be illustrated by considering the estimate of the spread given in equation (2.5). Suppose that $x_t \in \{-1, +1\}$, $\sigma_x^2 = 1$ (from the assumption of equiprobable buy and sell orders) and that c is the half-spread $S/2$. From (3.6) the pricing error variance is $\sigma_s^2 = (c - g)^2$. The estimate of the spread implied by the simple bid-ask model will generally be biased downward. In the present model, the first-order autocovariance is $\gamma_1 = -c(c - g)\sigma_x^2 = -c(c - g)$. For example, if $c = g$, i.e., if the spread is entirely information-based, then the transaction price changes will exhibit no autocorrelation, and the simple estimate of the spread will be zero.

From a statistical viewpoint, the pricing error in the simple model is uncorrelated with w_t (the increment in the efficient price). In the present model since $s_t = (c - g)x_t$ and $w_t = u_t + gx_t$, the two are correlated due to the shared influence of trades. This correlation

will not be perfect, except in the special case where $\sigma_u^2 = 0$, i.e., where there is no nontrade public information. Although this case is not attractive from an economic viewpoint, the value of σ_s^2 implied by this restriction possesses the useful property that it establishes a lower bound for σ_s^2 (over all correlations between w_t and s_t , holding constant the parameters of the observed return model $\{\sigma_\varepsilon^2, \theta\}$).

In terms of the moving average representation (2.6), the assumption of perfect correlation implies that both s_t and w_t are proportional to ε_t . Equating w_t to the cumulative effect of a disturbance (cf. the discussion following equation (2.7)) gives $w_t = (1 + \theta)\varepsilon_t$. From (2.2), $\Delta p_t = \varepsilon_t + \theta\varepsilon_{t-1} = (1 + \theta)\varepsilon_t + s_t - s_{t-1}$, which implies by inspection that $s_t = -\theta\varepsilon_t$, and $\sigma_{s, \text{lower bound}}^2 = \theta^2 \sigma_\varepsilon^2$. Since $-1 < \theta < 0$, this is obviously less than or equal to the estimate of σ_s^2 implied by the simple model, $-\theta\sigma_\varepsilon^2$. This lower bound is generalized in section 4.

In summary, based on knowing the parameters of the return process for this model (autocovariances or, equivalently, the moving average parameters), we can compute the random-walk (implicit efficient price) variance. Neither the pricing error variance nor derived measures such as the spread, however, are identified in the absence of further restrictions. Unfortunately, neither of the two identification restrictions considered above is particularly attractive, as they involve a choice between suppressing all public information or alternatively all private information.

3.3. *Models with both asymmetric information and inventory control.*

The following model combines inventory control and asymmetric information in an additive fashion:

$$\begin{aligned}
 m_t &= m_{t-1} + w_t \\
 w_t &= u_t + g v_t \\
 q_t &= m_{t-1} + u_t - b I_{t-1} \\
 x_t &= -a(q_t - (m_{t-1} + u_t)) + v_t \\
 I_t &= I_{t-1} - x_t \\
 p_t &= q_t + c x_t
 \end{aligned} \tag{3.8}$$

The m_t and w_t expressions are the same as in the asymmetric information model of the last section. The quote-midpoint expression includes an inventory control component. When information is entering the model from two sources, one must pay particular attention to the timing. At time t , public information (u_t) arrives, quotes are set (q_t), net demand is realized (x_t), which leads to a transaction at price p_t . Finally, the new efficient price m_t is set to reflect the information contained in the trade. The increment to the efficient price is driven by the trade innovation v_t and not simply the total trade. (Any new information imputed to the trade should come from the trade innovation.) The quote midpoint is set to reflect the current public information (u_t) and the inventory imbalance, but not the private information inferred from the time- t trade (which is not known at the time the quote is set). The incoming net demand reflects the difference between the current quote and the efficient price inclusive of public information.

The essential features of this model are illustrated by the impulse response function. The same parameter values are used as for the pure inventory control case in Figure 2, with $g=0.2$. Figure 3 depicts the time path subsequent to a one-unit innovation in the demand ($v_0 = 1$, a one-unit purchase from the dealer). The essential difference between this and Figure 2 is that the price reversion is incomplete. There is a permanent price effect of the buy order innovation, equal to $gv_t=0.2(1)$.

The pricing error is

$$s_t = p_t - m_t = cx_t - gv_t - bI_{t-1} \quad (3.9)$$

The $cx_t - gv_t$ term is analogous to the $(c - g)x_t$ expression for the pricing error in the pure asymmetric information model (3.6). Note, however, that the half-spread c is paid on the full trade, while the information update is driven solely by the trade innovation. The role of the $-bI_{t-1}$ term is the same as in the inventory control model (cf. equation (3.2)). Both terms are trade-driven.

The joint specification for returns and inventory levels may be written as a bivariate VAR in which all structural parameters are identified. If only transaction prices are available, only the random-walk variance (not the pricing error variance) may be identified from the reduced form.

By comparing the price impulse responses for the inventory control model (Figure 2), the asymmetric information model (Figure 3) and the combined model (Figure 4), it is apparent that the short-run price effects implied by the inventory and asymmetric information effects are very similar. In the pure inventory control model, the price rises in response to a buy because the dealer now has an inventory deficit and must attract more selling interest. In the asymmetric information model, the price rise reflects the new information revealed by the trade.

The similarity of the short-run price responses engendered by the inventory and information effects makes resolution of the two very difficult. Since the inventory control paradigm arose first, it was natural for early studies detecting a positive impact of trades on prices to affirm the existence inventory effects. Empirical tests of (more recent) asymmetric information models tended to attribute the initial price rise to the information content of a trade.

In practice, the two mechanisms can be resolved only by a dynamic analysis of both short and long-run effects. Studies of dealer (specialist) trading in equities on the NYSE suggest that inventory control is indeed practiced. However, the mechanism is considerably more complicated than that allowed for by the simple models considered here. The hypothetical impulse response functions discussed here depict a rapid inventory adjustment process, spanning a dozen trades at most. Trades are hypothetically negatively autocorrelated: a purchase should (in expectation) be followed in short order by sales. In actuality, however, trades exhibit strong positive autocorrelation in the short run (Hasbrouck and Ho (1987) and Hasbrouck (1988)). Furthermore, NYSE specialist positions appear to possess large long-run components (on the order of weeks or months). The ability of the available data samples to support reliable identification of transient inventory-control quote effects at these horizons is poor. See Hasbrouck and Sofianos (1993) and Madhavan and Smidt (1991 and 1993).

As noted above, this simple model combines inventory and asymmetric information effects in an additive fashion. The demand of an informed trader (and the market's estimate of the information content of a trade), however, will in principle depend on the

prevailing bid and offer quotes, which are also determined by the dealer's inventory position. The Madhavan and Smidt models illuminate these interactions.

3.4. *Prices, inventories and trades.*

The preceding analyses suggest that in the presence of asymmetric information or some combination of asymmetric information and inventory control, the results available from reduced-form price-change specifications are meager: σ_w^2 is identified, but σ_s^2 is not. It was also noted, however, that data sets that include dealer inventory data are rare. (There are presently none to my knowledge that exist in the public domain.)

It is often possible, however, to obtain good proxies for the trade series, x_t . A common practice when trade prices and volumes are reported and bid and ask quotes are available is to construct the proxy

$$x_t = \begin{cases} +(volume)_t, & \text{if } p_t > q_t. \\ 0, & \text{if } p_t = q_t. \\ -(volume)_t, & \text{if } p_t < q_t. \end{cases} \quad (3.10)$$

where q_t is the quote midpoint prevailing at the time the trade occurred. In the pure asymmetric information model of section 3.2, this proxy is sufficient.

When inventory control is present, however, matters become more complicated. By construction in the models discussed to this point, the dealer inventory is related to the trade by $I_t = I_{t-1} - x_t$. Because trades convey information only about the inventory changes, but not about the levels, they are generally inadequate proxies. From a statistical viewpoint, the problem is one of overdifferencing. When a variable such as a security price contains a random walk component, it is common to specify a stationary model in terms of the first difference (the price change, as we have done here). If one takes the first difference of a variable that is already stationary, however, the first difference will still be stationary, but it will not possess a convergent autoregressive representation. The overdifferenced variable is said to be noninvertible. The general role of the invertibility assumption in microstructure models will be discussed in section 4.1. But the

consequences for the specification of inventory control models can be illustrated with the simple models considered here.

In the pure inventory control model of section 3.1, the specification given in equations (3.1) may be reworked to give a univariate representation for the inventory level: $I_t = (1 - ab)I_{t-1} - v_t$, a simple first-order autoregression that is easily estimated. The trade series obtained by taking the (negative of) the first difference of the inventory is $x_t = -(I_t - I_{t-1}) = (1 - ab)x_{t-1} + v_t - v_{t-1}$, a mixed autoregressive-moving average (ARMA) form. No recursive substitution will yield an autoregressive representation for x_t with declining coefficients. The dilemma is not solved by adding the price change: there does not exist a convergent vector autoregressive representation for $\{\Delta p_t, x_t\}$. Nor is it generally convenient to estimate the ARMA specification given for x_t directly, since most techniques assume invertibility. (Exceptions are those based on exact maximum-likelihood Kalman filter methods. See Hamilton (1994).)

Despite this cautionary note, there are many situations in which models based on trades will in fact be invertible. The noninvertibility of the trade specifications arises from the fact that the trade series is the (negative) first-difference of the (presumably stationary) inventory series. In some data sets this is indeed the case: transactions are identified as to sign (buy or sell) and counterparty (e.g., the London Stock Exchange data used by Neuberger (1992) or the computerized trade reconstruction (CTR) data used by Manaster and Mann (1992)). The trade series composed of all the buys and sells to and from a particular dealer is, by construction, the first difference of the dealer inventory and it is implausible to assume invertibility.

In many markets, however, the dealer is not invariably the counterparty to the outside order. On the NYSE, for example, the dealer (specialist) participates in a relatively small portion of the trades. Often the bid and ask quotes represent non-specialist orders. There is a strong presumption of mean reversion in dealer inventories. But the other traders effectively placing bid and ask quotes represent a large, diverse and changing population of agents. There is little reason to suspect that the aggregate trades

of this group integrate up to a stationary series, and therefore little concern that trades will constitute an overdifferenced and noninvertible time series.

As an example, consider the following ad hoc model designed to capture many of the essential features of the inventory and asymmetric information model, but specified without direct reference to inventories:

$$\begin{aligned}
 m_t &= m_{t-1} + w_t \\
 w_t &= u_t + g v_t \\
 q_t &= m_{t-1} + u_t + d(q_{t-1} - (m_{t-2} + u_{t-1})) + b x_t \\
 x_t &= -a(q_t - (m_{t-1} + u_t)) + v_t \\
 p_t &= q_t + c x_t
 \end{aligned} \tag{3.11}$$

The essential difference between this and (3.8) is in the quote midpoint equation. The inventory dependence has been replaced by an explicit mean-reversion component that mimics the behavior associated with inventory control. This model was originally suggested by Lawrence Glosten, and is discussed in Hasbrouck (1991).

That the model exhibits characteristics of both inventory control and asymmetric information models can be seen from the impulse response functions (Figure 5) subsequent to a one-unit purchase innovation. The cumulative trade series is plotted as an analog to the (negative) inventory level. The parameter values are $\alpha=0.8$, $b=0.4$, $c=0.5$, $g=0.2$ and $d=0.5$. Like the basic inventory control model, there is a decaying reversion in the transaction price. Like the asymmetric information model, the reversion is not complete.

3.5. *Summary remarks on the simple models*

This section and the one preceding have illustrated the basic economic paradigms that underlie modern microstructure. The results may be summarized as follows. The bid-ask spread reflects fixed-cost and asymmetric information factors. The cost effect introduces a short-run transient “bounce” in price movements, while the asymmetric information effect is associated with a relatively rapid and permanent impact of a trade on the security price. Neither effect should necessarily induce any particular behavior in subsequent trades. Lagged price adjustment and inventory control create transients of longer duration. The price transients caused by the former, however, tend to smooth

informational responses, while those induced by inventory control induce price reversals. Inventory control should furthermore be associated with endogenous effects on the incoming trades.

4. General specifications.

The last section introduced basic microstructure concepts using simple structural models. These models are useful for calibrating the economist's intuition, but they are generally not good candidates for direct estimation. Key variables (such as the dealer's inventory) are often unobserved; the mechanisms are often more complicated than the stylized models suggest; the effects are often operating in concert; and finally, they are complicated by a host of other (primarily institutional) considerations discussed below. While it is always preferable to base a statistical model on a well-specified theoretical model, these considerations impose limitations on what can be achieved.

The models discussed in this section are in contrast nonrestrictive statistical models of microstructure data. The perspective here is one of foregoing precise estimates of structural parameters in hopes of achieving a characterization of microstructure effects that is both broad and robust. Most importantly, it is still possible under minimal assumptions to characterize the permanent/transient and trade-related/-unrelated dichotomies set forth in the introduction.

4.1. Vector Autoregressions (VARs)

A vector autoregression is a linear regression specification in which current values of all variables are regressed against lagged values of all variables. The inventory and asymmetric information models discussed in the last section, for example, can be specified as bivariate vector autoregressions. More general and flexible models can be obtained by extending the number of lags in estimation. VARs are relatively easy to estimate (least squares usually suffices) and interpret (via the impulse response functions or other transformations considered below). Their value in microstructure studies also rests, however, on their ability to characterize very general time series models. It is useful at this point to outline the assumptions underlying this generality, and also the ways in which they might be violated in microstructure applications.

The broad applicability of VARs ultimately rests on the Wold theorem. A zero-mean vector time series y_t is said to be weakly stationary (covariance stationary) if the autocovariances do not depend on t , $E y_t y_{t-j}' = \Gamma_j$. The Wold theorem states that a zero-mean weakly stationary nondeterministic process can be written as a convergent vector moving average (VMA) process (possibly of infinite order):

$$y_t = e_t + B_1 e_{t-1} + B_2 e_{t-2} + \dots = B(L)e_t, \quad (4.1)$$

where the e_t are serially uncorrelated homoskedastic increments with covariance matrix Ω and L is the backshift operator, $L(\cdot)_t = (\cdot)_{t-1}$ (Hamilton (1994) and Sargent (1987)). This is nothing more than the innovations representation of the process. This section assumes that the conditions of the Wold theorem are satisfied. The stationarity assumption will be examined in greater detail in section 5.

Suppose that we are working with price changes and trades (as in the model of section 3.4), so that the state vector is

$$y_t = \begin{bmatrix} \Delta p_t \\ x_t \end{bmatrix} \text{ and } e_t = \begin{bmatrix} u_t \\ v_t \end{bmatrix}; \text{Var}(e_t) = \Omega = \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} \quad (4.2)$$

The orthogonality of the residuals is based on the economic assumption that contemporaneous causality flows from trade to the transaction price. This characterized all of the simple structural models discussed in the last section. It is easy to contemplate market structures in which this assumption might be violated, but in many settings it is a reasonable approximation.

If all of the roots of the polynomial equation $\det(B(z)) = 0$ lie outside of the unit circle, then the VMA representation is said to be invertible, that is, it may be reworked to give a (possibly infinite) convergent VAR representation:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \dots + e_t = A(L)y_t + e_t. \quad (4.3)$$

In microstructure applications, the invertibility assumption is commonly violated by overdifferencing or cointegration. As noted in section 3.4, overdifferencing is a real possibility when the model involves inventories, but the data contain only trades (the first difference of the inventory). Cointegration arises when the state vector includes two or

more price variables for the same security (like the bid and ask quotes, or the transaction price and either quote), and is discussed further in section 8. All of the simple models discussed in the preceding sections may be represented in the form (4.3).

A minor inconvenience arises because all of the bivariate VAR models in the last section include a contemporaneous term on the right hand side:

$y_t = A_0^* y_t + A_1^* y_{t-1} + A_2^* y_{t-2} + \dots + e_t^*$. It is easy to rework this into the form (4.3) by noting $y_t = (I - A_0^*)^{-1} A_1^* y_{t-1} + (I - A_0^*)^{-1} A_2^* y_{t-2} + \dots + (I - A_0^*)^{-1} e_t^*$. Estimating the model in the form that includes the contemporaneous term is a convenient way of forcing orthogonality on the estimated residuals. Most econometric texts, however, employ the form (4.3), and this will be used here as well. There are several ways of computing the VMA (4.1) from the VAR. Conceptually, the simplest procedure involves simulating the behavior of the system subsequent to one-unit initial shocks (Hamilton (1974)).

4.2. *Random-walk decompositions.*

In the simple models the distinction between permanent and transitory price changes was expressed by equation (2.2). In the earlier sections, the specification of s_t was implicitly given by the structural form of the model. In this section, we take a more frankly statistical perspective, defining m_t and s_t in terms of their time series properties. Formally, the model is equation (2.2), but with the additional *statistical* assumptions that:

1. m_t follows a homoskedastic random walk: $E w_t = 0$, $E w_t^2 = \sigma_w^2$ and $E w_t w_\tau = 0$ for $t \neq \tau$.
2. s_t is a covariance stationary stochastic process.

It is worth emphasizing that the pricing error is not assumed to be serially uncorrelated or uncorrelated with w_t .

To establish the connection between the random walk decomposition (2.2) and the VAR described in (4.3), we will be working with the component of the VMA representation that corresponds to the price changes:

$$\Delta p_t = b(L) e_t \quad (4.4)$$

where $b(L)$ is the first row of the $B(L)$ matrix in (4.1). We assume that the pricing error can be written as a linear combination of current and lagged e_t , plus (to allow for other sources of variation) current and lagged η_t where η_t is a scalar disturbance uncorrelated with e_t :

$$s_t = c(L)e_t + d(L)\eta_t \quad (4.5)$$

In terms of the random-walk decomposition model, the price changes can be written as:

$$\Delta p_t = (1-L)m_t + (1-L)s_t = w_t + (1-L)s_t \quad (4.6)$$

The autocovariance generating function for a vector process y_t is

$$h_y(z) = \cdots \Gamma_{-2}z^{-2} + \Gamma_{-1}z^{-1} + \Gamma_0 + \Gamma_1z^1 + \Gamma_2z^2 + \cdots, \quad (4.7)$$

where z is a complex scalar (Hamilton (1994) p. 266). For a VMA process such as (4.1), $h_y(z) = B(z)\Omega B(z^{-1})'$. Equations (4.4) and (4.6) lead to two alternative representations for the autocovariance generating function of Δp_t :

$$h_{\Delta p}(z) = b(z)\Omega b(z^{-1})' = \sigma_w^2 + (1-z)h_s(z)(1-z^{-1}) \quad (4.8)$$

where $h_{\Delta p}(z)$ and $h_s(z)$ are the autocovariance generating functions for Δp and s . By setting $z=1$, we obtain:

$$\sigma_w^2 = b(1)\Omega b(1)' \quad (4.9)$$

This expression for the random-walk variance depends only on the parameters of the observed model, and hence is always identified. For example, the bid-ask model (with or without asymmetric information) can be represented as a first-order moving average model given by equation (2.6). In this case, $b(L) = 1 + \theta L$ and $\Omega = \sigma_\varepsilon^2$, which implies $\sigma_w^2 = (1 + \theta)^2 \sigma_\varepsilon^2$.

Returning to the bivariate case with price changes and trades, let $b(L)$ be partitioned as $b(L) = \begin{bmatrix} b_{\Delta p}(L) & b_x(L) \end{bmatrix}$. Given the diagonal structure of Ω , the random-walk variance can be decomposed as:

$$\sigma_w^2 = [b_{\Delta p}(1)]^2 \sigma_u^2 + [b_x(1)]^2 \sigma_v^2 \quad (4.10)$$

The two variance terms correspond to the non-trade and trade-related contributions to the efficient price variance. The R^2 measure introduced in (3.5) as a summary of the extent of asymmetric information can be generalized as:

$$R_{w,x}^2 = [b_{\Delta p}(1)]^2 \sigma_u^2 / \sigma_w^2 \quad (4.11)$$

Turning to the pricing error, we find that most results require further structure. If it is assumed that the pricing error is driven entirely by e_t , then we may eliminate the $d(L)\eta_t$ term in (4.5). This yields $b(L)e_t = w_t + (1-L)c(L)e_t$, which implies $w_t = [b(L) - (1-L)c(L)]e_t$. A solution for this is $w_t = b(1)e_t$, which is obviously consistent with the random-walk variance described above. By solving $b(L) = b(1) + (1-L)c(L)$, the coefficients of the $c(L)$ polynomial are found to be: $c_i = -\sum_{j=i+1}^{\infty} b_j$. Once the $c(L)$ coefficients are obtained, we may compute the value for s_t at a point in time, the unconditional variance of the pricing error, and also the trade- and nontrade-related components of this error. Given the diagonality of the innovation covariance matrix, these may be partitioned into trade-related and -unrelated components following the same procedure used in the analysis of σ_w^2 above. The restriction that $d(L)\eta_t = 0$ was originally suggested by Beveridge and Nelson (1981).

If the pricing error is assumed to be orthogonal to the random-walk increment, then the $c(L)e_t$ term in (4.5) vanishes. In this case, the coefficients of the $d(L)$ polynomial must be found by factoring the autocovariance generating function. The autocovariance generating function for s_t is $h_s(z) = d(z)\sigma_\eta^2 d(z^{-1})$ with d_0 normalized to unity. This may be substituted into (4.8) and the $d(L)$ coefficients found by factorization. This identification restriction is due to Watson (1986).

Watson also establishes some filtering results that are very useful in microstructure applications. We are assumed to possess a VMA for the observed processes (equation (4.1)) and wish to establish a correspondence to an unobserved components model (equations (2.2) with pricing error given by (4.5)). Watson shows that the best one-sided linear estimate (i.e., linear function of current and past observables) of the stationary component (pricing error) is the one associated with the Beveridge-Nelson identification

restriction. (Since η_t in (4.5) is orthogonal to the e_t , the best one-sided projection involves only the e_t .) This one-sided projection, denoted \hat{s}_t , is:

$$\hat{s}_t = E^*[s_t | e_t, e_{t-1}, \dots] = c(L)e_t \quad (4.12)$$

where the $c(L)$ coefficients are given above.

It is noted in Hasbrouck (1993) that the variance of the error in the one-sided projection is: $E(s_t - \hat{s}_t)^2 = E s_t^2 - E \hat{s}_t^2 \geq 0$ where the equality follows from the fact that the projection errors are uncorrelated with the projection: $E(s_t - \hat{s}_t)\hat{s}_t = 0$. This implies $E s_t^2 \geq E \hat{s}_t^2$: the variance of the one-sided (Beveridge-Nelson) projection establishes a lower bound on the variance of the pricing error. A related result is discussed in Eckbo and Liu (1993).

The tightness of the lower bound for the pricing error variance depends on the nature of the unobserved components model and also on the available data. In the asymmetric information model of section 3.2, the lower bound is exact (coincides with the true pricing error variance) if the model is estimated using both prices and trades. The actual variance exceeds the computed lower bound, however, if the model is estimated solely on the basis of prices. Hasbrouck (1993) discusses implementation considerations.

4.3. *Model order.*

The VAR and VMA representations discussed above are possibly infinite in length. In most applications these will be approximated by truncated specifications. This raises the question of how many lags should be included in the specification.

It is tempting here to rely on the usual statistical tests for model order (see Lutkepohl (1993), Ch. 4). In macroeconomic applications these tests usually (and conveniently) lead to models of modest order. This may be a consequence, however, of the low power of these tests to identify weak long term dependencies in typical macroeconomic data sets. In contrast, the large number of observations in microstructure applications is often sufficient to suggest statistical significance of weak dependencies at lags that would drive the number of model parameters beyond the capacity of most computer programs.

Many empirical and theoretical considerations do in fact militate in favor of extremely long lags. A number of studies, for example, have documented stock return dependencies over horizons on the order of five or ten years. A correct specification for stock price changes at the transaction level should in principle also account for observed behavior over longer horizons as well. It would therefore appear that estimations limited to, say, the five or ten most recent transactions are seriously misspecified.

If the concern is the behavior of stock returns over annual and longer cycles, however, it can be argued that the misspecification in short-run transaction studies is both economically irrelevant (for microstructure) and small in magnitude. The long-term swings in stock prices are generally held to reflect changes in expected returns. These are presumably due to business cycle factors in the real economy that have little connection to the short-run trading characteristics. Microstructure phenomena are almost by definition confined to short horizons. A truncated transaction-level model may not achieve an accurate resolution of transitory and permanent effects, but it may nevertheless still satisfactorily resolve microstructure and non-microstructure effects.

It must be acknowledged, however, that between horizons that are clearly microstructure-related (five transactions) and those that are clearly macroeconomic (five years) lie hourly or daily horizons over which microstructure phenomena might be important but difficult to detect. It was noted that dealer inventories often exhibit long-term components. Furthermore, traders sometimes employ strategies that spread order placement over many days. Such effects may not be detected in short-run transaction studies. This point is particularly imprint when the variable set includes nonpublic data, as discussed below.

4.4. *Expanding the variable set.*

Since the models discussed in sections 2 and 3 involve only prices and trades or inventories, the discussion has been limited to bivariate VARs. It is not difficult, however, to imagine hypotheses that would involve additional variables. For example, Huang and Stoll (1994) incorporate futures market variables into stock return specifications; Hasbrouck (1994) includes order flow; and Laux and Furbush (1994) examine program

trades. Such studies typically attempt to test hypotheses concerning the informational content of particular data that are usually associated with the trading process. While the details of these models lie beyond the present discussion, it is appropriate here to raise certain issues of modeling philosophy.

In contemplating the addition of a variable to a stock price specification, perhaps the most important question is whether or not or in what sense it is public knowledge. Given the complexities of the trading process, the usual situation is a murky one in which the data are known by a subset of agents (see section 2.1). Transaction-level microstructure VAR's typically reflect the explanatory or predictive power of a variable over a relatively short time horizon. If the variable does not enter the public information set within the horizon, however, then its information content will be not be measured correctly.

The information content of a trade, for example, can plausibly be assessed by short-run analyses because in most markets trades are reported quickly. But suppose the econometrician possesses a series of trades that has been identified (some months after the fact) as originating from corporate insiders illegally trading on advance knowledge of earnings announcements. If the insiders trade a week in advance of the public announcement, then the association between an insider purchase and the price rise occurring a week later will not be detected in a short-run microstructure VAR. The VAR will pick up the information content of a purchase, but not the additional informational content of an *insider* purchase.

Addition of other variables may cloud attribution of information effects in another respect. The simple models were constructed with explicit timing assumptions that generally sufficed to impose a recursive structure on the disturbances. In each time interval for the asymmetric information model, for example, the quote is revised to reflect public information, then a trade arrives, and then expectations are updated. This recursive economic structure gives rise to the statistical property that trade innovations are uncorrelated with public information, which in turn supports a clear resolution of trade and non-trade information effects. Often, however, particularly when the data are

collected from diverse sources, the time-stamps may not be clear enough to establish a recursive structure. The econometrician's imposition of a particular choice may exaggerate the informational content of variables appearing early in the assumed recursion.

In such situations, the behavior of the model may be investigated by examining alternative recursion assumptions. It is often possible, for example, to establish bounds on the variance decomposition components in expressions such as (4.10) using Cholesky factorizations of the innovation covariance matrix. Hamilton (1994) discusses general principles; Hasbrouck (1995) presents a microstructure application.

5. Time.

The microstructure models studied in the earlier sections were implicitly cast in real time, sometimes referred to as “calendar time” by macro econometricians or “wall-clock time” by microstructure students. In the interest of simplicity we implicitly took the time subscript t in the usual sense, as an index of equally-spaced points in real time. The stationarity assumptions necessary to support inference were assumed to hold with respect to this time index.

Timing considerations in actual markets, however, are considerably more involved. Markets do not usually operate continuously. The few that are in principle open twenty-four hours per day exhibit strong concentration of activity. Furthermore, trades usually take place at random times throughout the market session. This section discusses ways in which more realistic notions of time can be incorporated into statistical models.

5.1. *Deterministic time considerations*

Some of the time properties of markets appear to be deterministic, like the regular or predictable seasonalities encountered in macro time series. Two related examples in microstructure data are market closures and intraday patterns.

In most markets, trading takes place continuously during organized trading sessions. In between are periods of nontrading, typically over a lunch break, overnight, or over a weekend or holiday. If we are interested only in the behavior of the market during a trading session, we may drop from the sample all observations that span trading sessions, e.g., we might ignore an overnight return. If the aim of the analysis is a comprehensive

model of the market evolution during periods of trading and nontrading, however, the econometrician must first take a position on whether or not the market evolution is time homogeneous, i.e., whether prices (security values) behave in the same way during trading and nontrading periods. If homogeneity is assumed, then we are taking the view that the timing of the observations in our sample is merely an artifact of some sampling process that is not related to the behavior of the system. Obviously for models in which trading plays a central role (such as those involving asymmetric information), time homogeneity is not an attractive assumption. In testing less refined hypotheses, however, the conjecture might be a workable approximation. This motivates consideration of how time homogeneity is empirically examined.

Most of what we know about the role of time in microstructure data derives from the analysis of price-change variances (rather than means). This reliance on second moment properties characterizes not only the analysis of trading vs. nontrading periods, but also most of the work done on intra-trading session evolution. The reasons for this emphasis are the ones raised in Section 2.1: if the price follows a random walk, the precision of variance estimates is improved by more frequent sampling, the precision of mean estimates is not.

In U.S. equity markets, at least, the hypothesis that the return variance per unit time is constant over trading and nontrading periods is easily rejected (Fama (1965), Granger and Morgenstern (1970), Oldfield and Rogalski (1980) and Christie (1981)). Based on an analysis of returns computed using daily closing prices, French and Roll (1986) estimate that the return variance per unit time is at least an order of magnitude higher when the market is open than when it is closed. This is due in part to the fact that production of public information (such as news releases) is more likely to occur during normal business hours, but it is also due to the role of trading itself in the price discovery process.

Having rejected time homogeneity in the large, that is over trading and non-trading periods, might we still provisionally assume that it holds during trading sessions, at least well enough to support intraday analysis? There is considerable evidence to the contrary.

As a general rule, microstructure data exhibit distinctive behavior at the beginning and end of trading sessions. Most notably, return variances per unit time exhibit “U”-shapes, i.e., elevations at the session endpoints. Marked intraday patterns are also found in measures of trading activity such as transaction frequency, trading volume rates and bid-ask spreads (Jain and Joh (1988), McInish and Wood (1990), McInish and Wood (1992) and Wood, McInish and Ord (1985)).

5.2. *Stochastic time effects*

Although trading processes unfold in continuous time, they are marked by discrete events (e.g., trades or quote revisions). The determination of these occurrence times is at least in part random. Ideally, then, how should these processes be modeled from a purely statistical perspective? Furthermore, what is the economic significance of the occurrence times?

Specification of continuous-time models that allow for random intervals between events is difficult. There is a well-established literature on the analysis of irregularly spaced time series. (See Parzen (1984), Jones (1985), and the references therein.) It is commonly assumed in these models that the irregularity is a property of the observational process per se, i.e., that the underlying process evolves homogeneously in real time, and that the irregular observation times are either fixed or are at least exogenous to the evolution of the process. In microstructure applications both of these assumptions are problematic, the former on account of intraday volatility patterns and the latter for reasons yet to be discussed. Nevertheless, this approach does achieve an appealing unity in capturing the discrete and continuous time aspects of a simple model. Furthermore, the techniques used to specify and estimate these models may yet be generalized to more complicated and realistic situations.

Garbade and Lieber (1976) specify a variant on the simple bid-ask model in which the implicit random-walk variance per unit time is constant and the random-walk variance over a transaction interval is scaled by the intertransaction time. It is also necessary to assume that the intertransaction times are identically and independently distributed exponential random variables (i.e., a Poisson trade arrival process). Garbade and Lieber

find that the model performs well in a study of transaction data for IBM and Potlatch over ten trading days. The data suggest, however, more clustering of trades (over intervals shorter than approximately ten minutes) than is consistent with the hypothesized Poisson arrival process. In a more recent and comprehensive study of stock transaction data, Engle and Russel (1994) also find clustering and suggest an autoregressive duration model.

Although the GL model predated the advent of the inventory control and asymmetric information models, it could easily be adapted to incorporate these effects. The principal limitation of the approach from a current perspective is the assumed independence of the observation (“transaction generation”) process. The model implies, for example, that the probability that a trade will occur is independent of the size of the innovation in the security value, i.e., that we would be no more likely to witness a trade in the one minute following the close of a major press conference than we would in the middle of an uneventful August afternoon. This independence is not realistic.

Alternative approaches to the transaction occurrence problem have been employed in multiple security settings. The principle that (for a random walk) precision of variance estimates is enhanced by refinement of the observation interval also applies to estimates of covariances and betas, both of which are central to the standard portfolio problem. In addition, portfolio groupings are often employed to reduce measurement errors in certain applications, particularly the estimation of the return autocorrelations. Yet as the use of daily closing prices has become common, it has also been recognized that trading and reporting practices can induce significant estimation error in betas and significant autocorrelation in measured portfolio returns.

Campbell, Lo and MacKinlay (1993) provide an overview of these developments. Applications with asynchronous trading and last-trade reporting have historically attracted the most attention. Fisher (1966) discusses implications for stock index construction and interpretation. Analyses focusing on beta and covariance estimations are given in Scholes and Williams (1977), Dimson (1979), Cohen, Hawawini, Maier, Schwartz and Whitcomb (1983a, b), Shanken (1987). Studies emphasizing the effects on portfolio return

autocorrelations include Atchison, Butler and Simonds (1987), Boudoukh, Richardson and Whitelaw (1994), Cohen, Maier, Schwartz and Whitcomb (1986), Conrad and Kaul (1989), Conrad, Kaul and Nimalendran (1991), Lo and MacKinlay (1988a, b, 1990a, b), McInish and Wood (1991) and Mech (1993).

Traders sometimes characterize a market at a given time as being “slow” or “fast”. The description extends beyond the speed of price changes. Prices do tend to move quickly in a fast market, but the frequency of order arrival and transaction occurrence is also higher. It is as if “an hour’s worth of trading is packed into five minutes.” From a modeling viewpoint, this is more than figurative speech. It is calling attention to the distinction between real time and operational time, the time scale over which the process evolves at a constant rate. Stock (1988) describes this as time deformation.

Time deformation themes have been advanced in many empirical microstructure studies (not always using this terminology). Although the asymmetric information link between trades and prices has been formalized relatively recently, the idea that price variance is related to trading activity is older. Clark (1973) suggests that stock prices follow a subordinated stochastic process, one in which the “clock” of the process is trades. A number of studies find that over fixed real time intervals (such as a day or hour), the variance of equity price changes is positively related to the number of transactions and/or the trading volume (Harris (1987), Tauchen and Pitts (1992)). McInish and Wood (1991) and Jones, Kaul and Lipson (1994)) suggest that the association between return variance and trade frequency is higher than that between return variance and trade volume.

From an economic perspective, time deformation in market data is usually assumed to result from variation in the “information intensity” of the market, the rate at which the informational primitives (public and private signals) evolve. This is difficult to operationalize because these primitives, with the exception of sharply defined events like press conferences, are rarely observed. Also, in most theoretical models, the informational primitives are exogenous, implying that the resulting time deformation would also be exogenous.

Other economic considerations, however, strongly suggest endogenous time effects. A market-maker, for example, might diminish the frequency of incoming order arrival simply by widening the bid-ask spread. This sometimes occurs in response to a particularly significant informational announcement. In this instance, the econometrician relying on trade frequency as a proxy for informational intensity will draw exactly the wrong inference. Easley and O'Hara (1992), Easley, Kiefer and O'Hara (1993, 1994) and Easley, O'Hara and Paperman (1995) discuss these effects and suggest empirical tests. Strategic quote-setting behavior that can also lead to trade frequency effects is discussed by Leach and Madhavan (1992, 1993).

5.3. *Recommendations.*

Incorporating realistic time effects into microstructure models is a difficult task that is likely to call forth more and better research efforts. But if time *per se* is not the focus of a particular analysis, the econometrician needs to match the method to the immediate problem and the data. For investigating broad hypotheses about intraday patterns in market data and associations in these patterns, it appears sufficient to rely on data aggregated over fixed time intervals (e.g., hours). For investigating causal relations (such as trade price impacts) that would be obscured by aggregation, the econometrician should lean toward modeling the data purely in event time, i.e., where t indexes trades, quote revisions, etc. This is generally preferable to real-time modeling because it mitigates the effect of intraday patterns, and it incorporates some of the intuition of the formal time deformation approach: the "clock" of the process is assumed to be events.

6. **Discreteness.**

Although the models discussed to this point have assumed that both prices and quantities are continuous random variables, both are in fact discrete. Of course, most economic data are discrete in the sense that they are collected and reported subject to rounding or truncation errors. Market data are different, though, firstly because the discreteness is not merely an artifact of the observational process and secondly because the discreteness is economically significant. On the NYSE, for example, the standard

transaction size is a “round lot” of 100 shares. Deviations from multiples of this transaction size may lead to more difficulty in completing the trade and higher proportional transaction costs. Also, a stock priced at \$5 or more per share trades in ticks of 1/8 dollar (12.5 cents). By way of comparison, the per share commission on an institutional trade is roughly five cents per share.

Inability to smoothly adjust prices and quantities plays havoc with the intuition behind the simple models discussed earlier. Discreteness effectively transforms the decisions faced by agents from relatively tractable continuous optimization problems to complicated integer programming problems. In the simple asymmetric information model of section 3.2, for example, it might be conjectured that a dealer contemplating a one-tick quote increase would wait until a sequence of buy orders had occurred. It appears to be all but inevitable that discreteness will induce dynamic effects. Economic models that incorporate these and other aspects of discreteness include Bernhardt and Hughson (1990, 1992), Harris (1991, 1994), Chordia and Subrahmanyam (1992) and Glosten (1994).

6.1. The statistical modeling of discreteness.

Although investigation of the economic aspects of discreteness is coming into its own as an important subject for inquiry, its status in empirical models has traditionally been that of a nuisance effect. Discreteness is often viewed as a feature of market data that needs to be addressed or controlled for in some fashion while investigating other hypotheses. Most of the initial work on discreteness arose in response to the need to estimate return variances for purposes of option valuation. From a statistical viewpoint it is most convenient to model discreteness as a rounding disturbance (possibly to a floor or ceiling) (Ball (1990), Cho and Frees (1988), Gottlieb and Kalay (1985) and Harris (1990)).

At first glance, discreteness would seem to cause intractable problems for the simple models of Section 3 and the generalized VAR models of Section 4, for the reasons usually given in econometrics texts regarding the estimation of limited dependent variable models using linear specifications. Consistency of least squares estimation does not require that the residuals be independent of the explanatory variables, however, only that

they be uncorrelated. In many situations, absence of correlation can be motivated by appeal to the Wold Theorem, which is not contingent on an assumption that the variables are continuous. If the assumption of joint covariance stationarity is tenable in the time scale used to specify the model (usually either wall-clock time or transaction time), then there is no particular reason why discreteness should pose problems for estimating general VAR microstructure models and related constructs such as impulse response functions and variance decompositions. For many purposes, this approach will suffice.

The characterization of the market obtained in this fashion, however, is incomplete. The implied impulse response functions, for example, represent the continuous paths of the expected evolution of the market, which will look quite different from the sample paths that arise in discrete data. Furthermore this perspective is ill-suited for examining hypotheses in which discreteness parameters (such as the tick size) are of interest.

Hausman, Lo and MacKinlay (1992) present an ordered probit model of price changes. This is a single equation model in which trades and other explanatory variables (notably including the time between trades) drive a latent continuous price variable, which is in turn mapped onto the set of discrete prices using ordered breakpoints (that are estimated). Conditional on particular values of the explanatory variables, the predictions from this sort of model are given as probabilities of prespecified discrete price changes.

6.2. *Clustering.*

Market prices have an affinity for whole numbers that is difficult to justify on economic grounds. In most economic and statistical models, discreteness is specified as a grid on which strategies and outcomes must lie, but no distinctive properties are attributed to particular points on the grid. In a discrete random walk with $1/8$ ticks, for example, the price change is equally likely to be $+1/8$ or $1/8$. If the current stock price is $50 \frac{1}{8}$, it is equally likely that the next price will be 50 or $50 \frac{1}{4}$. Yet, as Harris (1991) notes, “Stock prices cluster on round fractions. Integers are more common than halves; halves are more common than odd quarters; odd quarters are more common than odd eighths; other fractions are rarely observed. This phenomenon is remarkably persistent across stocks.”

Similar effects are found in NYSE limit order prices (Neiderhoffer (1965, 1966)), NYSE quotes (Harris (1994), and (to a striking degree) in U.S. National Market System quotes (Christie and Shultz (1994a, 1994b)). Clustering suggests the existence of an implicit price grid that is coarser than the one mandated by the market rules. The economics of why these trading conventions arise and persist are not well understood.

7. **Nonlinearity.**

The models in Sections 2-4 express current variables as linear functions of past variables and disturbances. Although one can construct theoretical models for which linearity is appropriate, such a requirement is uncomfortably restrictive in applications to actual markets. This section discusses the motivation and approaches for nonlinear generalizations.

Among all of the aspects of microstructure modeling which we have examined so far, the one in which accurate functional specification is most important is the relation linking trades and price changes. Implicit in this relation are both the mapping from trades to inferred private information content and also the mapping from trades to trading costs. These mappings are determinants of individual agents' order placement strategies: how much to trade and whether to split the total quantity across different orders. From a social viewpoint, these mappings may admit or reject the possibility of market manipulation.

Most of the structural models that allow for nonlinearity in the trade/price impact mapping are single-equation specifications of price changes in which trades are assumed exogenous and the dynamic aspects of the market are not explicitly modeled. One standard model of this sort is due to Glosten and Harris (1988). Their specification can be viewed as a generalization of the asymmetric information model of section 3.2 in which there is an implied intercept in the cost and information functions. Variations of this model include George, Kaul and Nimalendran (1991), Neuberger and Roell (1991), Huang and Stoll (1994) and Madhavan, Richardson and Roomans (1994).

Intercepts and other nonlinearities can be incorporated into the general VAR models of section 4 in an ad hoc fashion. If price changes and signed trades are jointly stationary, then any transformations of price changes and signed trades are also jointly

stationary. This suggests that the dynamic VAR models can be generalized by expanding the state vector to include nonlinear transformations. Hasbrouck (1991a, 1991b, 1993) employs polynomial functions. Although a continuous function of a real variable can generally be approximated by a polynomial of sufficiently high degree, however, there is no assurance, that the approximation is a parsimonious one, an important consideration in practical applications.

This motivates consideration of more flexible characterizations of the trade-price change relation, of the sort provided by nonparametric analysis. Algert (1992) applies locally weighted regression to NYSE price and trade data, and concludes that the price change maps most closely to a low fractional power of the trade, suggesting that a square root transformation is preferable to the quadratic. Further applications of nonparametric and semiparametric methods in characterizing microstructure relations are likely to be illuminating.

Related studies focus primarily on the price impact of large (block) trades in the U.S. equity market: Holthausen, Leftwich and Mayers (1987), Barclay and Warner (1993). Such trades are of interest not only because of their size, but also due to their trade mechanism, as discussed in the next section.

8. Multiple mechanisms and markets

The basic market paradigm used in this paper is one in which patient or passive traders (including dealers) post bid and offer quotes in some centralized venue like a stock exchange. Trades occur when impatient active traders arrive and hit these quotes. While this is the most common mechanism, actual markets exhibit considerable diversity. It is in fact rare for a security to trade solely in one market setting using one procedure. Most continuous equity markets, for example, employ a batching procedure to open a trading session or to handle large order imbalances. There may be special mechanisms to handle large trades. Finally, multiple markets in the same security may simply operate in parallel, with varying degrees of formal integration. The important economic issues in these situations concern the merits of alternative market structures and the nature of the competition between markets (see, for example, Chowdhry and Nanda (1991)). The

empirical challenges involve the building of specifications general enough to handle the diverse trading mechanisms while retaining enough structure to address the economic hypotheses of interest. We consider in this section some common situations.

8.1. Call auctions.

A call auction is a procedure that approximates the Walrasian auction often used as a conceptual device to explain price determination in an idealized competitive market. Over some order entry period, traders submit supply and demand schedules specifying how much they intend to buy or sell at a particular price. At some clearing time, orders are crossed at the price given by the intersection of the aggregate supply and demand curves. Although conceptually simple, the practical aspects of implementation are decidedly nontrivial, ranging from how much information to display before clearing to the pricing of order entry and exchange services.

There is much current interest in the economic analysis of call and continuous markets. This is perhaps a consequence of the realization that with current communications technology, a call auction simultaneously involving large numbers of geographically dispersed participants is, for the first time, feasible. Advocates of call auctions argue that pricing errors will be minimized because the aggregate supply and demand schedules will reduce (by the law of large numbers) the impact of idiosyncratic randomness in individual demands and arrivals (Mendelson (1982), Schwartz and Economides (1995) and Schwartz (1996)). Advocates of continuous markets place a high value on the availability of immediate execution, which is of particular importance in hedging and dynamic portfolio strategies.

At the NYSE, a call is used to open continuous trading, and also to reopen continuous trading after a trading halt. A call (itayose) is also used to initiate continuous trading on the Tokyo Stock Exchange (Lehmann and Modest (1994), Hamao and Hasbrouck (1995)). The Frankfurt Bourse runs a noon call, at which time most of the retail orders for German equities are traded.

If the primary aim of a study is characterization of the continuous trading mechanism (which usually accounts for the bulk of the trading activity and most of the

price change variance), then one commonly drops the opening price (and the overnight price change) from the analysis. For hypotheses that specify the joint behavior of the two mechanisms, however, other methods are required.

It is rare in empirical studies for the two mechanisms to be modeled jointly with fully specified models of both mechanisms. Instead, the merits are usually investigated by comparing opening call prices with one or more prices from the continuous session.

Suppose that the time index $t=1, 2, \dots$ is constructed so that the odd times $t=1, 3, 5, \dots$ correspond to market opening times, and the even times $t=2, 4, \dots$ correspond to market closing prices (or some other price taken from the continuous trading session). Using the basic random walk decomposition model from section 2.2, a two-period price change may be written as $\Delta p_t^{[2]} = (w_t + w_{t-1}) + s_t - s_{t-2}$. Assuming that the w_t and s_t are mutually and serially uncorrelated, the variance of the two-period price change is

$$\text{Var}(\Delta p_t^{[2]}) = \text{Var}(w_t) + \text{Var}(w_{t-1}) + \text{Var}(s_t) + \text{Var}(s_{t-2}) \quad (8.1)$$

We now consider how this variance depends on whether t is odd (an open-to-open price change) or even (close-to-close). There are two random walk terms. Whether or not t is even, one of the pair t and $t-1$ is even and the other is odd. Therefore $\text{Var}(w_t) + \text{Var}(w_{t-1})$ does not depend on whether t is even. It is the variance of the 24-hour innovation in the efficient price. The pricing error time subscripts, on the other hand will be both even or both odd. We may therefore write:

$$\begin{aligned} \text{Var}(\Delta p_t^{open}) &= \text{Var}(w_t) + \text{Var}(w_{t-1}) + 2 \text{Var}(s_t^{open}) \\ \text{Var}(\Delta p_t^{close}) &= \text{Var}(w_t) + \text{Var}(w_{t-1}) + 2 \text{Var}(s_t^{close}) \end{aligned} \quad (8.2)$$

The difference between these two variances is therefore twice the difference in variances of the opening and closing pricing errors. If the variance of the opening pricing error is greater than that of the closing pricing error, this difference is positive.

Alternatively, the variance ratio of the first variance to the second is greater than one.

Amihud and Mendelson (1987) and Stoll and Whalley (1990) find that on average for NYSE stocks this ratio is indeed greater than one (larger variance of pricing error at the opening call). These results have not settled the mechanism debate. It has been

argued that the elevated opening variance at the NYSE is due to particular features of the NYSE call (selective ability of traders to “recontract”, the last-move advantage of the specialist, etc.). It may also be that the period of overnight market closure is associated with transient opening effects that are not associated with the call mechanism per se. The Tokyo Stock Exchange trading day is broken into morning and afternoon sessions, both of which begin with a call. Amihud and Mendelson (1991) find that while the variance of the morning open is elevated (consistent with U.S. findings), the variance of the afternoon call is not. Related studies include Amihud, Mendelson and Murgia (1990) (Italy), Gerety and Mulherin (1994) (long-run U.S.) and Masulis and Ng (1991) (London). Smith (1994) and Ronen (1994) discuss the general statistical properties of variance ratio estimates in these applications. Lee, Ready and Seguin (1994) discuss calls subsequent to trading halts.

More general variance ratios of another type arise in microstructure studies as a summary measure of the extent to which a price series deviates from a random walk. It is a property of a homoskedastic random walk that the variance of the increments is a linear function of the time interval over which the increment is computed. That is, in simple random-walk model (section 2.1) the variance of the one-period price change is $\text{Var}(\Delta p_t) = \text{Var}(p_t - p_{t-1}) = \sigma_w^2$; that of the two-period change is $\text{Var}(\Delta p_t^{[2]}) = \text{Var}(p_t - p_{t-2}) = 2\sigma_w^2$ and so on. The ratio of these two variances scaled by the time intervals is $(\text{Var}[\Delta p_t^{[2]}/2])/\text{Var}(\Delta p_t)$ is equal to unity. More generally, the variance ratio formed from the n -period price change (relative to the one-period change) is

$$V_n = \frac{\text{Var}(\Delta p_t^{[n]})}{n\text{Var}(\Delta p_t)} \quad (8.3)$$

For a random-walk, $V_n=1$ for all n . The extent to which this ratio deviates from unity is sometimes taken as a measure of how much the process deviates from a random walk.

A useful alternative form for V_n is obtained by expanding $\text{Var}(\Delta p_t^{[n]})$ in terms of the price-change autocovariances, and dividing through by $\text{Var}(\Delta p_t)$, yielding $V_n = 1 + 2\sum_{i=1}^{n-1} \rho_i$ where ρ_i is the price-change autocorrelation at lag i . Written in this

fashion, it becomes apparent that for the simple bid-ask model of section 2.3, the only non-zero autocorrelation is $\rho_1 < 0$, which will in turn drive V_n below unity. On the other hand, positive autocorrelation (induced perhaps by lagged adjustment) can lead to variance ratios above one. A mixed pattern of positive and negative autocorrelations can lead to a variance ratio equal to unity for a price-change process that is distinctly different from a random walk.

An early application of variance ratios to stock return data is Barnea (1974), who interprets the nine-day/one-day variance ratio as a performance measure for New York Stock Exchange specialists (designated dealers). Hasbrouck and Schwartz (1988) estimate variance ratios using transaction data for stocks traded on the New York, American and National Market System (“over-the-counter”) exchanges. Kaul and Nimalendran (1994) use variance ratios to resolve bid-ask and overreaction effects. Lo and MacKinlay (1988) employ variance ratios to examine the random walk hypothesis in weekly stock return data, and describe the asymptotic properties of the variance ratio and related estimates under the null (random walk) hypothesis. Their paper also contains citations to other occurrences of variance ratios in the statistical and economics literature.

8.2. *Large trade mechanisms.*

Trade cost is related to trade size. When a trader is contemplating a transaction that is much larger than the normal trade size for a market, this cost might be reduced by breaking the order into smaller pieces brought to the market over time. For traders demanding immediacy in large size, however, alternative trading procedures have often evolved. On the NYSE, for example, large (block) trades are typically negotiated in the “upstairs” market, and then formally transacted (“crossed”) on the exchange and reported to the transaction tape. Economic issues are considered by Burdett and O’Hara (1987), Grossman (1992), Seppi (1990, 1992).

The last section cited studies of the price impact of block trades. As in the case of different opening mechanisms, there are no analyses employing fully realized joint specifications of the regular (“downstairs”) and upstairs markets. In fact it is not possible to infer from the public quote and transaction record which trades were negotiated in the

upstairs market. Accordingly, most empirical studies simply treat block trades as “large” trades, ignoring the details of the negotiation process.

8.3. *Parallel markets.*

It is convenient to view opening call auctions and block trades (at least in the U.S. equities markets) as alternative mechanisms functioning as close adjuncts to regular trading in a single market. When the alternative trading mechanisms for a security diverge greatly with respect to their clientele, locations or procedures, it may be more natural to view the alternatives as distinctly different markets.

For example, equities listed on the NYSE also trade on the U.S. regional exchanges. Although there are electronic links among the exchanges, trading and quote-setting may vary considerably across venues. As a second example, while the Paris Bourse accounts for much of the trading volume in French equities, large trades are frequently done on the London Stock Exchange. There is no formal integration of the two, although it is likely that someone contemplating a trade would check the prices in both markets (de Jong, Nijman and Roell (1993)). Grunbichler, Longstaff and Schwartz (1992) discuss multiple markets in German equities. The current trend toward increased dispersal of trading activity is termed “fragmentation”.

It might be hoped that with market data on a single security trading in or more markets, one could estimate the market dynamics jointly, simply by “stacking” the market data to combine them in a single estimation. If these data include two or more price series for the security, however, specification becomes tricky. The complexities can be illustrated in a simple model of a single security trading in two markets, with imperfect flows of information. The implicit efficient price follows a random walk, but with increments that are “revealed” to each market separately:

$$\begin{aligned}
 m_t &= m_{t-1} + w_t \\
 w_t &= u_{1,t} + u_{2,t} \\
 p_{1,t} &= m_{t-1} + u_{1,t} + (1 - a_1)u_{2,t} = m_t - a_1 u_{2,t} \\
 p_{2,t} &= m_{t-1} + u_{2,t} + (1 - a_2)u_{1,t} = m_t - a_2 u_{1,t}
 \end{aligned}
 \tag{8.4}$$

The price equations are consistent with lagged adjustment to information originating in the other market. The price in the first market, for example, reflects only $(1 - \alpha_1)$ of the contemporaneous innovation in the second market. The remaining portion is reflected in the subsequent time period. If the u_i are uncorrelated, the total variance of the implicit efficient price changes is $\sigma_w^2 = \text{Var}(u_{1,t}) + \text{Var}(u_{2,t})$. The proportion of information contributed by the i th market, termed the “information share” in Hasbrouck (1995), is $\text{Var}(u_{i,t})/\sigma_w^2$

It may be shown that although a VMA representation for the price changes exists in this model it is not invertible: a convergent VAR representation for the price changes does not exist. This is not a consequence of the stylized nature of the model. It is rather a reflection of the fact that even though both price series possess random-walk components (formally, possess unit roots), the difference between the prices is stationary. Such systems are said to be cointegrated. (See Davidson, Hendry, Srba and Yeo (1978), Engle and Granger (1987), and, at a textbook level, Hamilton (1994) and Banerjee, Dolado, Galbraith and Hendry (1994).)

Cointegrated systems can often be represented in numerous alternative ways, some of which are more useful for interpretation and others for estimation. Of particular importance in the present application is the Stock-Watson common trends representation. If two prices are cointegrated, they may be written:

$$\begin{bmatrix} p_{1,t} \\ p_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} m_t + \begin{bmatrix} s_{1,t} \\ s_{2,t} \end{bmatrix} \quad (8.5)$$

This is a multivariate generalization of the basic dichotomy between permanent and transitory components. It is important to note that the two prices share the same permanent component.

In a cointegrated system, a convergent VAR representation for the price changes will never exist. One generally has more success with a slightly modified specification, the so-called error correction model (ECM). For a two-price model, a typical ECM is:

$$\Delta p_t = \alpha(p_{1,t-1} - p_{2,t-1}) + A_1 \Delta p_{t-1} + A_2 \Delta p_{t-2} + \dots + u_t \quad (8.6)$$

where the A_i are (2x2) coefficient matrices and α is a (2x1) vector of coefficients. From (8.6) a VMA representation for the price changes may be recovered. This in turn will support computation of market information contributions described above (see Hasbrouck (1995)). Although ECMs are frequently employed as general reduced form specifications, their existence is not guaranteed. If $\alpha_1 = \alpha_2 = 1$, the model given in equation (8.4) will not possess a convergent ECM representation, although state-space estimation may remain feasible.

In macroeconomic applications, the presence of cointegration and the coefficients of the cointegrating vectors (or a linear basis for these vectors) are often problematic. Matters are usually simpler in microstructure settings. When the cointegration involves two or more prices associated with same security (such as the price in different markets or the bid and ask quote in the same market), a basis for the cointegrating vectors can plausibly be specified a priori. If there are n price variables, there are $n-1$ linearly independent price differences. Rejection of this set of cointegrating vectors is tantamount to asserting that two or more prices will tend over time to diverge without bound. This is not plausible if the prices all pertain to the same security. Harris, McInish, Shoesmith and Wood (1992) and Hasbrouck (1995) discuss these issues and describe applications to the U.S. equities markets.

A similar situation exists when the multiple prices apply not to the same security, but instead to the security and a derivative such as a futures or options contract. Here it is often the case that arbitrage relationships between the derivative and the underlying will lead to cointegration between the price of the underlying and some function of the price of the derivative. Cointegration is likely to arise therefore, in studies of spot and forward prices and stock and option prices.

9. Summary and directions for further work.

This paper has attempted to provide an overview of the various approaches to modeling microstructure time series. Rather than recapitulate these developments, it is perhaps more useful to return to the questions that motivated them. It was claimed in the introduction that microstructure models can potentially examine both narrow questions of

trading behavior and market organization and also broader issues of valuation and the nature of information. The present paper has focused, however, almost exclusively on the former. This emphasis can be justified on the grounds that any study using market transaction data must employ methods that reflect the market realities. But as a practical matter the economic importance of security valuation and the implications for the allocation of real assets almost certainly outweighs the welfare improvements that might result from modest changes in the trading mechanisms for most securities. It is therefore appropriate to briefly indicate some of the ways in which microstructure studies can illuminate aspects of corporate finance.

The classic event study measures the impact of a public information event by the associated change in the security price. The insight of the asymmetric information models is that when the “event” is a trade, the price reaction summarizes the market’s estimation of the private information behind the trade. Studies of the price impact of trades, the spread (under certain assumptions), or the summary $R_{w,x}^2$ measure introduced in section 3.2 thereby broadly characterize the market’s beliefs about the magnitude of information asymmetries. Since these beliefs cannot usually be measured directly, the window offered by microstructure data may well be the only vantage point. Recent studies that explore asymmetric information in the vicinity of corporate announcements include Foster and Viswanathan (1995) (takeover announcements) and Lee, Mucklow and Ready (1993) (earnings announcements). Neal and Wheatley (1994) discuss the asymmetric information characteristics of closed-end mutual funds.

We now return to the narrower microstructure issues. From a statistical perspective, the current state of the art falls considerably short of a plausible comprehensive model of transactions data. The reader who has skimmed over the discussion of time, discreteness nonlinearities and multiple markets in the earlier sections can hardly avoid getting a sense of the tentativeness that marks modeling efforts in these areas, and the need for further work. But statistical models in this area must be ultimately judged by their implications for the economic questions.

From an economic perspective, the standing questions are those of how information enters market prices, how traders should behave (private welfare) and how markets should be organized (social welfare). Studies of trade-price behavior have yielded a modest understanding of the first issue. It is an empirical fact that trades seem to explain part but not all of price changes. This confirms the existence of private information and establishes the importance of trading for the revelation or incorporation of this information.

Answers to the other two fundamental questions, however, remain elusive. Trading strategy in most markets remains the province of human judgment, guided by experience and intuition, beyond the limits of existing normative models, even outside the realm of most ex post performance measurement excepting that of the roughest sort (“Did our investment strategy make money, net of trading costs?”). Nor have academic efforts to define economically efficient trading arrangements been particularly successful. While we have garnered greater insights into the workings of existing markets, we have yet to create yardsticks capable of ranking potential alternative arrangements. No consensus on these questions among academics, practitioners and regulators has yet emerged. It is certainly to be hoped that improved econometric models will provide useful insights.

10. References

- Eckbo, B.E. and J. Liu, 1993, "Temporary Components of Stock Prices: New Univariate Results," *Journal of Financial and Quantitative Analysis*, 28, 161-176.
- Algert, P., 1992, Estimates of nonlinearity in the response of stock prices to order imbalances, Working Paper, Graduate School of Management, University of California at Davis.
- Amihud, Y. and H. Mendelson, 1980, Dealership market: market making with inventory, *Journal of Financial Economics*, 8, 31-53.
- Amihud, Y. and H. Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics*, 17, 223-49.
- Amihud, Y. and H. Mendelson, 1987, Trading mechanisms and stock returns, *Journal of Finance*, 42, 533-53.
- Amihud, Y. and H. Mendelson, 1991, Volatility, efficiency and trading: evidence from the Japanese stock market, *Journal of Finance*, 46, 1765-89.
- Amihud, Y., H. Mendelson and M. Murgia, 1990, Stock market microstructure and return volatility, *Journal of Banking and Finance*, 14, 423-40.
- Atchison, M., K. Butler and R. Simonds, 1987, Nonsynchronous security trading and market index autocorrelation, *Journal of Finance* 42, 533-53.
- Banerjee, A., J. Dolado, J. W. Galbraith and D. F. Hendry, 1994, Co-integration, Error-correction, and the Econometric Analysis of Non-stationary Data, (Oxford University Press, London).
- Barclay, M. J. and J. B. Warner, 1993, Stealth trading and volatility: Which trades move prices, *Journal of Financial Economics*, 34, 281-306.
- Barnea, A., 1974, Performance evaluation of New York Stock Exchange specialists. *Journal of Financial and Quantitative Analysis*, 9, 511-535.
- Glosten, L. R. and L. E. Harris, 1988, Estimating the components of the bid-ask spread, *Journal of Financial Economics*, 21, 123-42.
- Beja, A. and M. Goldman, 1980, On the dynamics of behavior of prices in disequilibrium, *Journal of Finance*, 35, 235-48.
- Bernhardt, D. and E. Hughson, 1990, Discrete pricing and dealer competition, Working Paper, California Institute of Technology.

- Bernhardt, D. and E. Hughson, 1992, Discrete pricing and institutional design of dealership markets, Working Paper, California Institute of Technology.
- Beveridge, S. and C. R. Nelson, 1981, A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to the measurement of the 'business cycle', *Journal of Monetary Economics*, 7, 151-174.
- Blume, M. and M. Goldstein, 1992, Displayed and effective spreads by market, Working paper, University of Pennsylvania.
- Boudoukh, J., M. P. Richardson and R. F. Whitelaw, 1994, A tale of three schools: insights on the autocorrelations of short-horizon stock returns, *Review of Financial Studies* 7, 539-73.
- Burdett, K. and M. O'Hara, 1987, Building blocks: an introduction to block trading, *Journal of Banking and Finance*, 11, 193-212.
- Campbell, J. Y., A. W. Lo and A. C. MacKinlay, *The Econometrics of Financial Markets Chapter 3: Aspects of market microstructure*, Working Paper No. RPCF-1013-93, Research Program in Computational Finance, Sloan School of Management, Massachusetts Institute of Technology.
- Cheng, M. and A. Madhavan, 1994, In search of liquidity: block trades in the upstairs and downstairs markets, Working Paper, New York Stock Exchange.
- Cho, D. C. and E. W. Frees, 1988, Estimating the volatility of discrete stock prices, *Journal of Finance*, 43, 451-466.
- Chordia, T. and A. Subrahmanyam, 1992, Off-floor market-making, payment-for-order-flow and the tick size, Working Paper, UCLA.
- Chowdhry, B. and V. Nanda, 1991, Multimarket trading and market liquidity, *Review of Financial Studies*, 4, 483-512.
- Christie, A. A., 1981, On efficient estimation and intra-week behavior of common stock variances, Working paper, University of Rochester.
- Christie, W. G. and P. H. Schultz, 1994a, Why did NASDAQ market makers stop avoiding odd-eighth quotes?, *Journal of Finance*, 49, 1841-60.
- Christie, W. G. and P. H. Schultz, 1994b, Why do NASDAQ market makers avoid odd-eighth quotes?, *Journal of Finance*, 49, 1813-40.
- Clark, P. K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica*, 41, 135-159.

Cohen, K., D. Maier, R. Schwartz and D. Whitcomb, 1981, Transaction costs, order placement strategy and the existence of the bid-ask spread, *Journal of Political Economy*, 89, 287-305.

Cohen, K., D. Maier, R. Schwartz and D. Whitcomb, 1986, *The Microstructure of Security Markets*, (Prentice-Hall: Englewood Cliffs, NJ).

Cohen, K., G. Hawawini, S. Maier, R. Schwartz and D. Whitcomb, 1983a, Friction in the trading process and the estimation of systematic risk, *Journal of Financial Economics*, 29, 135-148

Cohen, K., G. Hawawini, S. Maier, R. Schwartz and D. Whitcomb, 1983b, Estimating and adjusting for the intervalling-effect bias in beta, *Management Science*, 29, 135-148.

Conrad, J. and G. Kaul, 1989, Mean reversion in short-horizon expected returns, *Review of Financial Studies*, 2, 225-40.

Conrad, J., G. Kaul and M. Nimalendran, 1991, Components of short-horizon individual security returns, *Journal of Financial Economics*, 29, 365-84.

Copeland, T. and D. Galai, 1983, Information effects and the bid-ask spread, *Journal of Finance*, 38, 1457-1469.

Damodaran, A., 1993, A simple measure of price adjustment coefficients, *Journal of Finance*, 48, 387-400.

Davidson, J. E. H., D. F. Hendry, F. Srba and S. Yeo, 1978, Econometric modeling of the aggregate time series relationship between consumers' expenditure and income in the United Kingdom, *Economic Journal*, 88, 661-92.

de Jong, F., T. Nijman and A. Roell, 1993, A comparison of the cost of trading French shares on the Paris Bourse and on SEAQ International, London School of Economics Discussion Paper No. 169.

Dimson, E., 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics*, 7, 197.

Easley, D. and M. O'Hara, 1987, Price, size and information in securities markets, *Journal of Financial Economics*, 19, 69-90.

Easley, D. and M. O'Hara, 1991, Order form and information in securities markets, *Journal of Finance*, 46, 905-927

Easley, D. and M. O'Hara, 1992, Time and the process of security price adjustment, *Journal of Finance*, 47, 577-606.

- Easley, D., N. M. Kiefer and M. O'Hara, 1993, One day in the life of a very common stock, Working Paper, Cornell University.
- Easley, D., N. M. Kiefer and M. O'Hara, 1994, Sequential trading in continuous time, Working Paper, Cornell University.
- Easley, D., N. M. Kiefer, M. O'Hara and J. B. Paperman, 1995, Liquidity, information and infrequently traded stocks, Working Paper, Cornell University.
- Engle, R. F. and C. W. J. Granger, 1987, Co-integration and error correction: representation, estimation and testing, *Econometrica*, 55, 251-76.
- Engle, R. F., and J. R. Russell, 1994, Forecasting transaction rates: the autoregressive conditional duration model. Working Paper No. 4966, National Bureau of Economic Research, (Cambridge, MA).
- Fama, E. F., 1965, The behavior of stock market prices, *Journal of Business*, 38, 34-105.
- Fama, E., 1970, Efficient capital markets: a review of theory and empirical work, *Journal of Finance*.
- Fisher, L., 1966, Some new stock market indexes, *Journal of Business*, 39, 191-225.
- Foster, F. D. and S. Viswanathan, 1990, A Theory of the interday variations in volumes, variances and trading costs in securities markets, *Review of Financial Studies*, 3, 593-624.
- Foster, F. D. and S. Viswanathan, 1995, Trading costs of target firms and corporate takeovers, in *Advances in Financial Economics*, JAI Press.
- French, K. R. and R. Roll, 1986, Stock return variances: the arrival of information and the reaction of traders, *Journal of Financial Economics*, 17, 5-26
- Garbade, K., and Z. Lieber, 1977, On the independence of transactions on the New York Stock Exchange, *Journal of Banking and Finance*, 1, 151-172.
- Garman, M., 1976, Market microstructure, *Journal of Financial Economics*, 3, 257-275.
- George, T. J., G. Kaul and M. Nimalendran, 1991, Estimation of the bid-ask spread and its components: a new approach, *Review of Financial Studies*, 4, 623-656.
- Gerety, M. S. and J. H. Mulherin, 1994, Price formation on the stock exchanges: the evolution of trading within the day, *Review of Financial Studies*, 7, 609-29.
- Glosten, L., 1987, Components of the bid-ask spread and the statistical properties of transaction prices, *Journal of Finance*, 42, 1293-1307.

Glosten, L., 1994, Is the electronic open limit order book inevitable?, *Journal of Finance*, 49, 1127-1161.

Glosten, L. and L. Harris, 1988, Estimating the components of the bid-ask spread, *Journal of Financial Economics*, 21, 123-142.

Glosten, L. R. and P. R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics*, 14, 71-100.

Goldman, M. and A. Beja, 1979, Market prices vs. equilibrium prices: return variances, serial correlation and the role of the specialist, *Journal of Finance*, 34, 595-607.

Goodhart, C. A. E. and M. O'Hara, 1995, High frequency data in financial markets: issues and applications, Working Paper, London School of Economics.

Granger, C. W. J. and O. Morgenstern, 1970, Predictability of stock market prices, (Heath-Lexington, Lexington, MA).

Grossman, S. J. and M. H. Miller, 1988, Liquidity and market structure, *Journal of Finance*, 43, 617-33.

Grossman, S. J., 1992, The informational role of upstairs and downstairs trading, *Journal of Business*, 65, 509-28.

Grunbichler, A., F. A. Longstaff, E. Schwartz, 1992, Electronic screen trading and the transmission of information: an empirical examination, Working Paper, UCLA.

Hamao, Y. and J. Hasbrouck, 1995, Securities trading in the absence of dealers: trades and quotes on the Tokyo Stock Exchange, *Review of Financial Studies*, forthcoming.

Hamilton, J. D., 1994, *Time Series Analysis*, (Princeton University Press, Princeton).

Harris, F. H. deB., T. H. McInish, G. L. Shoesmith and R. A. Wood, 1992, "Cointegration, Error Correction, and Price Discovery on the New York, Philadelphia and Midwest Stock Exchanges," Working paper, Fogelman College of Business and Economics.

Harris, L., 1990, Statistical properties of the Roll serial covariance bid/ask spread estimator, *Journal of Finance*, 45, 579-90.

Harris, L., 1991, Stock price clustering and discreteness, *Review of Financial Studies*, 4, 389-415.

Harris, L., 1994, Minimum price variations, discrete bid-ask spreads and quotation sizes, *Review of Financial Studies*, 7, 149-178.

- Harvey, A. C., 1990, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.
- Hasbrouck, J. and G. Sofianos, 1993, The trades of market makers: an empirical analysis of NYSE specialists, *Journal of Finance*, 48, 1565-1593.
- Hasbrouck, J. and T. S. Y. Ho, 1987, Order arrival, quote behavior and the return-generating process, *Journal of Finance*, 42, 1035-1048.
- Hasbrouck, J., 1988, Trades, quotes, inventories and information, *Journal of Financial Economics*, 22, 229-252.
- Hasbrouck, J., 1991a, Measuring the information content of stock trades, *Journal of Finance*, 46, 179-207.
- Hasbrouck, J., 1991b, The summary informativeness of stock trades: an econometric investigation, *Review of Financial Studies*, 4, 571-95.
- Hasbrouck, J., 1993, Assessing the quality of a security market: a new approach to measuring transaction costs, *Review of Financial Studies*, 6, 191-212.
- Hasbrouck, J., 1995, Order characteristics and stock price evolution: an application to program trading, *Journal of Financial Economics*, forthcoming.
- Hasbrouck, J., 1995, One security, many markets: determining the contributions to price discovery, *Journal of Finance*, forthcoming.
- Hasbrouck, J., G. Sofianos, and D. Sosebee, 1993, Orders, trades, reports and quotes at the New York Stock Exchange, (NYSE working paper, Research and Planning Section).
- Hausman, J., A. Lo and A. C. MacKinlay, 1992, An ordered probit analysis of stock transaction prices, *Journal of Financial Economics*, 31, 319-379.
- Ho, T. S. Y and H. R. Stoll, 1981, Optimal dealer pricing under transactions and returns uncertainty, *Journal of Finance*, 28, 1053-1074.
- Holthausen, R. W., R. W. Leftwich and D. Mayers, 1987, The effect of large block transactions on security prices, *Journal of Financial Economics*, 19, 237-67.
- Huang, R. D. and H. R. Stoll, 1994a, Market microstructure and stock return predictions. *Review of Financial Studies*, 7, 179-213.
- Huang, R. D. and H. R. Stoll, 1994b, The components of the bid-ask spread: a general approach, Working Paper 94-33, Owen Graduate School of Management, Vanderbilt University.

- Jain, P. C. and G. H. Joh, 1988, The dependence between hourly prices and trading volume, *Journal of Financial and Quantitative Analysis*, 23, 269-83
- Jones, R. H., 1985, Time series analysis with unequally spaced data, in E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds., *Handbook of Statistics, Volume 5, Time Series in the Time Domain*, (Elsevier Science Publishers, Amsterdam).
- Karlin, S., and H. M. Taylor, 1975, *A First Course in Stochastic Processes*, (Academic Press, New York).
- Kaul, G., and M. Nimalendran, 1990, Price reversals: bid-ask errors or market overreaction, *Journal of Financial Economics*, 28, 67-93.
- Kyle, A. S., 1985, Continuous auctions and insider trading, *Econometrica*, 53, 1315-1336.
- Laux, P. and D. Furbush, 1994, Price Formation, Liquidity, and Volatility of Individual Stocks around Index Arbitrage, working paper, Case Western Reserve University.
- Leach, J. C. and A. N. Madhavan, 1992, Intertemporal discovery by market makers, *Journal of Financial Intermediation*, 2, 207-235.
- Leach, J. C. and A. N. Madhavan, 1993, Price experimentation and security market structure, *Review of Financial Studies*, 6, 375-404.
- Lee, C. M. C. and M. Ready, 1991, Inferring trade direction from intradaily data, *Journal of Finance*, 46, 733-746.
- Lee, C. M. C., B. Mucklow and M. J. Ready, 1993, Spreads, depths and the impact of earnings information: an intraday analysis, *Review of Financial Studies*, 6, 345-374.
- Lee, C. M. C., M. J. Ready and P. J. Seguin, 1994, Volume, volatility and New York Stock Exchange trading halts, *Journal of Finance*, 49, 183-214
- Lehmann, B. and D. Modest, 1994, Trading and liquidity on the Tokyo Stock Exchange: a bird's eye view, *Journal of Finance*, 44, 951-84.
- Lo, A. and A. C. MacKinlay, 1988a, Stock prices do not follow random walks: evidence from a simple specification test, *Review of Financial Studies*, 1, 41-66.
- Lo, A. and A. C. MacKinlay, 1988b, Notes on a Markov model of nonsynchronous trading, Working Paper, Sloan School of Management, Massachusetts Institute of Technology.
- Lo, A. and A. C. MacKinlay, 1990a, An econometric analysis of nonsynchronous trading, *Journal of Econometrics*, 45, 181-212.

- Lo, A. and A. C. MacKinlay, 1990b, When are contrarian profits due to stock market overreaction, *Review of Financial Studies* 3, 175-205.
- Lo, A. and A. C. MacKinlay, 1990c, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431-468.
- Madhavan, A. and S. Smidt, 1991, A Bayesian model of intraday specialist pricing, *Journal of Financial Economics*, 30, 99-134.
- Madhavan, A. and S. Smidt, 1993, An analysis of changes in specialist inventories and quotations, *Journal of Finance*, 48, 1595-1628.
- Madhavan, A., M. Richardson and M. Roomans, 1994, Why do security prices change? A transaction level analysis of NYSE stocks, Working Paper, Wharton School.
- Manaster, S. and S. Mann, 1992, Life in the pits: competitive market making and inventory control, Working Paper, University of Utah.
- Marsh, T. and K. Rock, 1986, The transactions process and rational stock price dynamics, Working Paper, University of California at Berkeley.
- Masulis, R. W. and V. K. Ng, 1991, Stock return dynamics over intra-day trading and non-trading periods in the London stock market, Working Paper No. 91-33, Mitsui Life Financial Research Center, University of Michigan.
- McInish, T. H. and R. A. Wood, 1990, A Transactions Data Analysis of the Variability of Common Stock Returns during 1980-1984, *Journal of Banking and Finance*, 14, 99-112
- McInish, T. H. and R. A. Wood, 1991a, Hourly returns, volume, trade size, and number of trades. *Journal of Financial Research*, 14, 303-15.
- McInish, T. H. and R. A. Wood, 1991b, Autocorrelation of daily index returns: intraday-to-intraday vs. close-to-close intervals, *Journal of Banking and Finance* 15, 193-206.
- McInish, T. H. and R. A. Wood, 1992. An analysis of intraday patterns in bid/ask spreads for NYSE stocks. *Journal of Finance*, 47, 753-64.
- Mech, T., 1993, Portfolio return autocorrelation, *Journal of Financial Economics* 34, 307-44.
- Mendelson, H., 1982, Market behavior in a clearing house, *Econometrica*, 50, 1505-24.
- Merton, R., 1980, Estimating the expected rate of return, *Journal of Financial Economics*, 8, 323-62.

- Naik, N., A. Neuberger and S. Viswanathan, 1994, Disclosure regulation in competitive dealership markets: analysis of the London Stock Exchange, Working Paper, London Business School.
- Neal, R., and S. Wheatley, 1994, How reliable are adverse selection models of the bid-ask spread, Working Paper, Federal Reserve Bank of Kansas City.
- Neuberger, A. J. and A. Roell, 1991, Components of the bid-ask spread: a Glosten-Harris approach, Working Paper, Working Paper, London Business School.
- Neuberger, A. J., 1992, An empirical examination of market maker profits on the London Stock Exchange, *Journal of Financial Services Research*, 343-372.
- Niederhoffer, V. and M. F. M. Osborne, 1966, Market making and reversals on the stock exchange, *Journal of the American Statistical Association*, 61, 897-916.
- Niederhoffer, V., 1965, Clustering of stock prices, *Operations Research*, 13, 258-262.
- Niederhoffer, V., 1966, A new look at clustering of stock prices, *Journal of Business*, 39, 309-313.
- O'Hara, M. and G. S. Oldfield, 1986, The microeconomics of market making, *Journal of Financial and Quantitative Analysis*, 21, 361-76.
- Oldfield, G. S. and R. J. Rogalski, 1980, A theory of common stock returns over trading and non-trading periods, *Journal of Finance*, 37, 857-870.
- Parzen, E., editor, 1984, *Time Series Analysis of Irregularly Observed Data*, (Springer-Verlag, New York).
- Petersen, M. and S. Umlauf, 1991, An empirical examination of intraday quote revisions on the New York Stock Exchange, Working paper, Graduate School of Business, University of Chicago.
- Roll, R., 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance*, 39, 1127-1139.
- Ronen, T., 1994, *Essays in Market Microstructure: Variance Ratios and Trading Structures*, unpub. Ph.D. Dissertation, New York University.
- Samuelson, P., 1965, Proof that properly anticipated prices fluctuate randomly, *Industrial Management Review*.
- Sargent, T. J., 1987, *Macroeconomic Theory*, second edition, (Academic Press: Boston).
- Scholes, M. and J. Williams, 1977, Estimating betas from nonsynchronous data, *Journal of Financial Economics*, 5, 309.

Schwartz, R. A. and N. Economides, 1995, Making the trade: equity trading practices and market structure, *Journal of Portfolio Management*, forthcoming.

Schwartz, R. A., 1988, *Equity Markets: Structure, Trading and Performance*, (Harper and Row, New York).

Schwartz, R. A., 1991, *Reshaping the Equity Markets*, (Harper Business, New York).

Schwartz, R. A., 1996, *Electronic Call Market Trading* (Symposium Proceeding), Irwin Professional.

Seppi, D. J., 1990, Equilibrium block trading and asymmetric information, *Journal of Finance*, 45, 73-94.

Seppi, D. J., 1992, Block trading and information revelation around quarterly earnings announcements, *Review of Financial Studies*, 5, 281-305.

Shanken, J., 1987, Nonsynchronous data and the covariance-factor structure of returns, *Journal of Finance*, 42, 221-232.

Stock, J., 1988, Estimating continuous time processes subject to time deformation, *Journal of the American Statistical Association*, 83, 77-85.

Stock, J. H., and M. W. Watson, 1988, Testing for common trends, *Journal of the American Statistical Association*, 83, 1097-1107.

Smith, T., 1994, Econometrics of financial models and market microstructure effects, *Journal of Financial and Quantitative Analysis*, 29, 519-540.

Stoll, H. R., 1978, The supply of dealer services in securities markets, *Journal of Finance*, 33, 1133-1151.

Stoll, H. R., 1989, Inferring the components of the bid-ask spread: theory and empirical tests, *Journal of Finance*, 44, 115-34.

Tinic, S., 1972, The economics of liquidity services, *Quarterly Journal of Economics*, 86, 79-93.

U.S. Securities and Exchange Commission, 1971, *Institutional Investor Study Report* (Arno Press, New York).

Watson, M. W., 1986, Univariate detrending methods with stochastic trends, *Journal of Monetary Economics*, 18, 49-75.

Wood, R. A., T. H. McInish and J. K. Ord, 1985, An investigation of transactions data for NYSE stocks, *Journal of Finance*, 40, 723-39.

Table 1. A Classification of Microstructure Effects.

		Type of price change	
		Permanent (informational)	Transient (market related)
Source of price change	Trade-induced (attributable to an actively initiated transaction)	<p>Economic: Market's assessment of the information content of the trade (asymmetric information)</p> <p>Statistical: Random-walk component of price attributable to trade variables</p>	<p>Economic: Non-informational spread effects, transaction costs, dealer inventory control effects, price discreteness.</p> <p>Statistical: Stationary component of price attributable to trade variables.</p>
	Not trade-induced	<p>Economic: Public information</p> <p>Statistical: Random-walk component of price change not attributable to trade variables.</p>	<p>Economic: Lagged adjustment to public information, price discreteness</p> <p>Statistical: Stationary component of price not explained by trades</p>

Figure 1.**The Impulse Response Function for the Lagged Price Adjustment Model**

The adjustment of the transaction price (p) subsequent to an initial shock of +1 in the efficient price. The model is the lagged price adjustment model given in equation (2.8), with parameter $\alpha=0.5$.

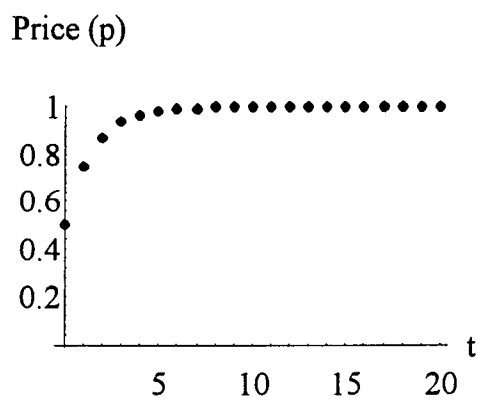


Figure 2.

The Impulse Response Function for the Inventory Model

The adjustment of the transaction price (p) and the dealer's inventory (I) subsequent to an initial purchase of one unit. The model is the inventory control model given in equation (3.1) with parameters $\alpha=0.8$, $b=0.04$ and $c=0.5$.

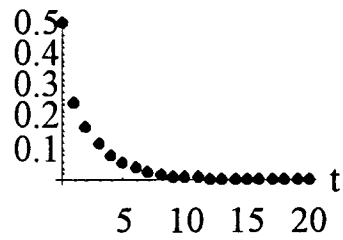
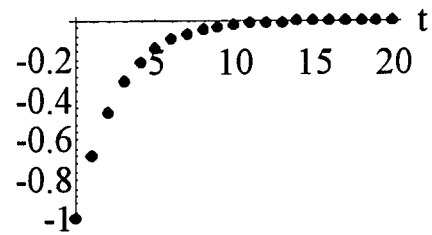
Price (p)Inventory (I)

Figure 3.

The Impulse Response Function for the Asymmetric Information Model

The adjustment of the transaction price (p) and the incoming trade (x) subsequent to an initial purchase of one unit. The model is the asymmetric information model given in equation (3.3) with parameters $c=0.5$ and $g=0.2$.

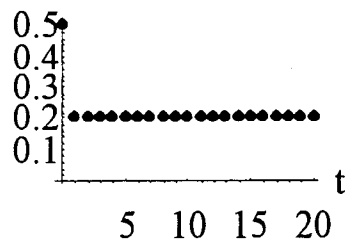
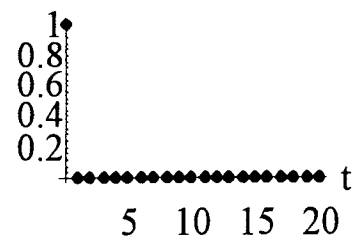
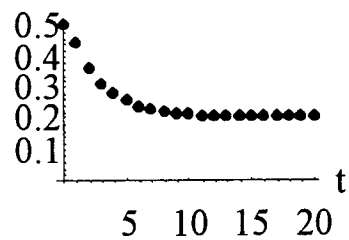
Price (p)Trade (x)

Figure 4.
The Impulse Response Function
for the Inventory Control/Asymmetric Information Model

The adjustment of the transaction price (p) and inventory (I) subsequent to an initial purchase of one unit. The model is the inventory control/asymmetric information model given in equation (3.8) with parameters $\alpha=0.8$, $b=0.04$, $c=0.5$, and $g=0.2$.

Price (p)



Inventory (I)

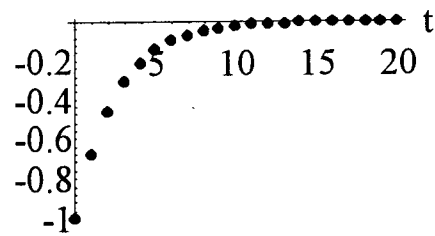
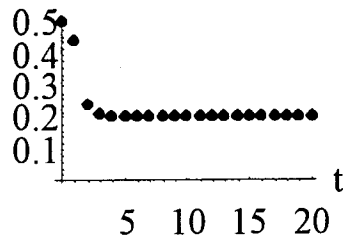


Figure 5.
The Impulse Response Function
for the Asymmetric Information/Trade Model

The adjustment of the transaction price (p) and cumulative trades (Σx) subsequent to an initial purchase of one unit. The model is the asymmetric information/trade model given in equation (3.11) with parameters $\alpha=0.8$, $b=0.4$, $c=0.5$, $g=0.2$ and $d=0.5$.

Price (p)Cumulative Trade (Sum of x)