THE JOURNAL OF FINANCE • VOL. LIV, NO. 6 • DECEMBER 1999

The Dynamics of Discrete Bid and Ask Quotes

JOEL HASBROUCK*

ABSTRACT

This paper presents an empirical microstructure model of bid and ask quotes that features discreteness, random costs of market making, and ARCH volatility effects. Applied to intraday quotes at 15-minute intervals for Alcoa (a randomly chosen Dow stock), the results show that quote exposure costs contain stochastic components that are persistent and large relative to the deterministic intraday "U" components. Analysis of the filtered estimates of the system suggest that bid and ask costs contain common components, and that these costs reflect risk as proxied by ARCH variance forecasts.

EMPIRICAL MARKET MICROSTRUCTURE SEEKS to characterize the dynamics of prices, spreads and volumes using specifications consistent with economic principles and actual institutional arrangements. Applications to U.S. equities markets, where the data are particularly rich, have been fruitful. The usual empirical methods employed here, however, presume that parameters are either constant in time or exhibit at best deterministic (e.g., intraday) variation. An important alternative line of research, exemplified by the ARCH literature, considers more general models that admit stochastic parameter variation. Yet although ARCH-family models have been estimated for high-frequency data, the specifications do not usually incorporate the structural market features that characterize the microstructure models. Thus, the two approaches have mostly evolved with distinct aims and methodologies.

Integration promises mutual gains. The short-run adaptation of trading arrangements to stochastic changes in risk is, for example, a proper concern of microstructure. From the alternative perspective, volatility specifications may well be improved by conditioning on microstructure variables. The model proposed in this paper takes a step toward this integration, with a latent (unobserved) dynamic structure that involves ARCH price dynamics and stochastic costs of market making.

In constructing a microstructure model with stochastic parameter variation, discreteness quickly arises as a potential concern. In most microstructure models (and in most actual markets), the determinants of security prices

* Professor of Finance, Stern School of Business, New York University. For comments on an earlier draft, I am indebted to James Angel, Darrell Duffie, James Hamilton, Mark Rubinstein, Neil Shephard, Matt Spiegel, and seminar participants at Cornell University, Georgetown University, New York University, Stanford University, the University of California at Berkeley, and the UCSD Conference on High Frequency Data. All errors are my own responsibility.

(e.g., conditional expectations of security value or the costs of market making) are or are likely to be continuous variables. Institutional arrangements, on the other hand, generally constrain prices to a discrete grid. This grid may be coarse relative to quantities of interest, such as the price variation over brief intervals or the economic costs of order submission and execution. The grid may be coarser still relative to time variation (deterministic or stochastic) in the model parameters.

The present application, for example, involves bid and ask quotes collected at 15-minute intervals for a typical NYSE stock (Alcoa, symbol AA) over a sample period when the tick size is one-eighth of a dollar. The changes in the discrete quotes overwhelmingly concentrate on zero, one, or two ticks. This suggests that the tick size is large relative to short-term price movements. Furthermore, in 45.9 percent of the intervals, the closing spread is one-eighth. This suggests that the tick size is frequently binding on the economic spread.

The NYSE, however, recently moved to a tick size of one-sixteenth for most stocks, and plans to move to decimal trading by the year 2000. In light of this change, is the economic and statistical importance of discreteness likely to persist? One answer is that, while the diminution of the tick size mitigates the pressure to model discreteness explicitly, fundamental economic considerations suggest that the tick size in the equilibrium remains large relative to certain trading costs. Harris (1997) points out that in a market that at least partially respects time priority for orders at a given price, the tick is the cost of jumping to the head of the existing queue. An arriving buyer, for example, can step ahead of the existing limit buy orders only by bidding one tick higher than the best prevailing bid. To the extent that market institutions view time priority as a desirable feature in encouraging the supply of liquidity, they are unlikely to be indifferent to arbitrarily small tick sizes.

The present model suggests that bid and ask quotes arise from an implicit efficient price and quote-exposure (market making) costs, all of which are continuous random variables. The implicit efficient price possesses ARCH dynamics, and the quote exposure costs are autoregressive. The rounding transformation generating the discrete bid and ask quotes is asymmetric. The observed bid is the implicit efficient price less the continuous bid exposure cost rounded down to the next tick; the ask is the efficient price plus the continuous ask exposure cost rounded up to the next tick. The model can be cast in nonlinear state-space form and estimated via maximum likelihood.

The principal results are interesting in several respects. First, the quote-exposure costs exhibit stochastic variation larger than the deterministic intraday ("U") variation. The interpretation of this depends on the economic model in which these costs are viewed, but one implication is that market liquidity also exhibits substantial stochastic variation. The quote exposure costs also exhibit short-term persistence. The distribution of the changes in the (latent) efficient price is highly leptokurtic. This implies that the kurtosis found in observed short-term price changes is not solely attributable to discreteness. The ARCH component of the efficient price process also exhibits persistence, although not enough to account for volatility persistence at daily and longer horizons. Finally, although the estimation does not incorporate correlation between volatility and the quote exposure costs, the filtered within-sample estimates do in fact exhibit positive correlation. This suggests an economic connection between stochastic variation in the costs of market making and volatility, and points to the importance of joint modeling.

The paper is organized as follows. The model's clearest precursors belong to the lines of research dealing (mostly separately) with discreteness in stock prices, deterministic patterns in market data, and high-frequency ARCH models. Section I reviews the literature in these areas. The model, and in particular the relation between the continuous latent variables and the observed discrete bid and ask quotes, is presented in Section II. The paper then turns to the problem of inference: How to estimate the underlying model from the observed discrete bid and ask prices. Section III discusses the restrictions imposed on the underlying variables by the discrete observations. Section IV introduces the nonlinear filtering algorithm, the associated maximum likelihood procedure and computational techniques. The full dynamic model, which incorporates stochastic and deterministic time variation in the cost and efficient price volatility, is presented in Section V. The model is estimated for a representative NYSE stock in Section VI. A brief summary concludes the paper in Section VII.

I. Background and Literature Review

The primary roots of the present paper lie in three areas of work: Models of stock price discreteness, analyses of deterministic variation in market parameters, and studies of stochastic variation in volatility. These three lines of inquiry have evolved along lines sufficiently distinct that it is possible to summarize them separately. The following discussion proceeds to do this, but also attempts to draw connections and emphasize points of common interest.

A. Stock Price Discreteness

The classic models of discreteness in transaction prices were motivated primarily by the desire for stock return volatility estimates largely (to put matters in a historical context) for purposes of option valuation. Most of the transaction price models can be placed within the following framework:

$$m_t = m_{t-1} + u_t \tag{1a}$$

$$b_t = m_t - c \tag{1b}$$

$$a_t = m_t + c \tag{1c}$$

$$p_t = \begin{cases} \text{Round}(b_t) & \text{if } \pi_t = -1 \\ \text{Round}(a_t) & \text{if } \pi_t = +1, \end{cases} \tag{1d}$$

where m_t is the implicit efficient ("true") price, c is a market-making cost, b_t and a_t are the bid and ask quotes, π_t is a buy/sell indicator variable, and Round$(\cdot)$ rounds to the nearest integer. (Throughout this paper all variables are assumed scaled so that the tick/grid size is unity.)

Discrete transaction prices were initially modeled as random walk realizations, rounded to the nearest grid point (Gottlieb and Kalay (1985), and Ball (1988)), as if c in equation (1d) were zero. One might conjecture in this model that traders negotiate a continuous price, which is then discretized. Harris (1990a) allows a half-spread, $c > 0$. The bid and ask quotes (implicit in Harris's presentation) lie on a continuum. Glosten and Harris (1988) generalize this to allow for a quote schedule linear in trade size. Dravid (1991) (citing a suggestion of Paul Pfleiderer) allows for asymmetric rounding of the bid and ask quotes: $b_t = \text{Floor}(m_t - c)$ and $a_t = \text{Ceiling}(m_t + c)$, where Floor$(\cdot)$ rounds its argument down to the next lowest grid point and Ceiling$(\cdot)$ rounds its argument up.

The asymmetric rounding used by Dravid (1991) is also a key feature of the model proposed in this paper. Although a full discussion of this property is deferred until Section II, some simple considerations may help motivate the present developments. Most markets impose discreteness on quotes as well as transaction prices. By asymmetrically rounding up on the ask and down on the bid, a market maker avoids the possibility of loss on the incoming trade. If the rounding is symmetric (all prices rounded up, all prices rounded down, or all prices rounded to the nearest integer), then one or both sides of the quotes might be associated with an expected loss. Furthermore, symmetric rounding may imply degenerate quotes (identical bid and ask prices) if c is small.

Depending on cost and rounding methods used, the above models have different (and generally incompatible) implications for the behavior of the transaction price, allowing it to evolve as a rounded random walk, a rounded signal-plus-noise, or a value randomly selected from two rounded random walks (one associated with the bid and the other with the ask). The empirical implementations of the models assume that u_t and π_t are identically and independently distributed (i.i.d.). In Gottlieb and Kalay (1985), Ball (1988), and Dravid (1991), the implied variances and autocovariances of the transaction price changes are used to compute moment estimators. The i.i.d. assumptions are crucial to the tractability of these estimators. These assumptions are, of course, violated in applications where the moments are changing in a deterministic fashion (as in studies of intraday patterns) or in a stochastic fashion (as in the time-varying volatility models).[1] Although Harris (1990a) also invokes an i.i.d. assumption, this is not essential to his iterative max-

[1] These comments on the restrictiveness of the i.i.d. assumptions apply to most moment-based estimators, including the GMM approach used by Madhavan, Richardson, and Roomans (1997) and the vector autoregression approaches used in, for example, Hasbrouck (1991) and surveyed in Hasbrouck (1996). Deterministic intraday patterns in these models are typically investigated by estimating them over subperiods (e.g., the first and last half-hour of trading) during which time the processes are assumed stationary.

imum likelihood approach. His approach is in fact a variant of the nonlinear state space procedure described by Kitagawa (1987), which is also used in the present paper.

B. Deterministic Variation in Market Parameters

A large class of studies considers deterministic variation in microstructure data. Equity markets, for example, typically display elevation of volatility, spreads, and volumes around the beginning and end of trading sessions (see Wood, McInish, and Ord (1985), Harris (1986), Brock and Kleidon (1992), McInish and Wood (1992), Lee, Mucklow, and Ready (1993), and Chan, Christie, and Schultz (1995)). Other studies have characterized intraday periodicities in foreign exchange returns (e.g., Muller et al. (1990), Baillie and Bollerslev (1991), and Bollerslev and Domowitz (1993)).

Intraday patterns are usually assessed, however, from discrete transaction prices or quotes, and the intraday variation may be small relative to the tick size. Lee, Mucklow, and Ready (1993), for example, report a beginning-of-day elevation in the quoted spread for NYSE stocks of roughly 10 percent (relative to the daily mean). For an average spread of 1/4 (arguably on the high side for a large NYSE stock), the 10 percent elevation is only one-fifth of the 1/8 tick size.

In the present model, deterministic variation is important in two respects. First, as emphasized below, it is important for models of stochastic variation to incorporate deterministic components (and so avoid misspecification and misattribution of effects). Secondly, the deterministic variation serves as a well-established benchmark for assessing the magnitude of stochastic effects.

C. Stochastic Variation in Volatility

Another prominent group of studies (beginning with Engle (1982) and surveyed in Bollerslev, Engle, and Nelson (1994)) concentrates on modeling stochastic variation in volatility with ARCH-family models. Although the initial applications were to returns computed at daily or longer horizons, interest subsequently broadened to include high-frequency (intraday) data. One strong motivation for this trend is the fact that volatility estimates and evaluation of volatility forecasts are improved by more frequent sampling (Andersen and Bollerslev (1998b)).

At shorter return horizons, however, market microstructure effects in general (and discreteness effects in particular) may become significant relative to changes in the latent security price. Thus, most of these high-frequency ARCH applications involve foreign exchange (FX) or stock index returns, for which (as will be shown) discreteness is negligible (e.g., Engle, Ito, and Lin (1990), Hamao, Masulis, and Ng (1990), Baillie and Bollerslev (1991), Chan, Chan, and Karolyi (1991), and Andersen and Bollerslev (1997a, 1997b, 1998b)). Although ARCH models have been applied to returns for an individual stock over daily and longer horizons (e.g., Cheung and Ng (1992) and Duffee (1995)), the application in this paper constitutes the first estimation (to my knowledge) of a high-frequency ARCH model for an individual stock.

A useful starting point is the exponential generalized autoregressive conditional heteroskedasticity model (EGARCH) proposed for security returns by Nelson (1991):

$$m_t = m_{t-1} + u_t \tag{2a}$$

$$\sigma_t^2 = \text{Var}(u_t) \tag{2b}$$

$$\ln(\sigma_t^2) = \eta + \varphi(\ln(\sigma_{t-1}^2) - \eta) + \gamma(|\zeta_{t-1}| - E|\zeta_{t-1}|) \tag{2c}$$

where $\zeta_t \equiv u_t/\sigma_t$ is the standardized increment. The terms on the right-hand side reflect a mean (η), an autoregressive adjustment (at rate φ) toward the mean, and a disturbance component with coefficient γ driven by the prior period's shock. (The asymmetry term used by Nelson is omitted.) Nelson suggests that the standardized increment ζ_t be distributed in accordance with the generalized error distribution (GED), denoted $f_{GED}(\zeta_t;\nu)$, where ν is the tail-thickness parameter. When $\nu = 2$, the GED reduces to the standard normal density. The expected absolute value $E|\zeta_{t-1}|$ in equation (2c) is unconditional and time-invariant, depending only on the tail-thickness parameter. Nelson applies this model to the return series computed from daily closing S&P prices.

The ARCH increments are continuous random variables. The practical importance of the misspecification associated with applying this model to return data computed from discrete prices varies considerably. For an S&P futures contract, the tick is roughly 0.01 percent of the price. (The current index level is approximately 1,000; the current minimum price variation is 0.1.) Andersen and Bollerslev (1997b) estimate the standard deviation of five-minute futures returns to be about 0.1 percent. Similarly, in the DM/$ market the smallest price unit conventionally displayed is 0.0001(DM/$), about 0.005 percent of a representative price of two DM/$. Andersen and Bollerslev estimate the standard deviation of five-minute returns to be 0.047 percent.[2] In both cases, the standard deviation is roughly an order of magnitude larger than the tick size, and one would not normally presume discreteness to be an issue.

The relative tick size is more significant in the U.S. equities market. As an illustration, consider a stock with an annual log return standard deviation of 0.30 ("30 percent"). Assuming that this variance is realized entirely within 250 $6\frac{1}{2}$-hour trading days, the implied five-minute return standard deviation is $\sqrt{0.3^2/(250 \times 6.5 \times 12)} = 0.002$ ("0.2 percent"). For a $25 share, the current $1/16 tick is 0.25 percent, which is roughly comparable.

The data used in the present paper also indicate a striking consequence of ignoring discreteness. When the conventional continuous EGARCH/GED model is applied to the discrete quote-midpoint changes, likelihood minimi-

[2] Andersen and Bollerslev also apply their volatility models to prices that are averaged (over the bid and ask sides of the market, and over the two quotes whose time stamps straddle the five-minute mark). This further mitigates the importance of discreteness.

zation with various algorithms and starting points generally fails to converge. Simulations suggest that the problem lies in the interaction of discreteness with the GED distribution. As the tail-thickness parameter drops below two (the normal case), the GED becomes sharply peaked at zero. When zero-mean continuous data are rounded, the observations near zero collapse to a high peak at zero. This feature of the sample often dominates the minimization, driving the value of ν inexorably downward.

Although the models of deterministic volatility components have evolved (for the most part) separately from those of stochastic volatility components, recent developments mark something of a convergence. The reasons for this (discussed more completely in Andersen and Bollerslev (1997a, 1997b)) are relevant for the present study. When ARCH models are estimated for returns computed over varying horizons, theoretical temporal aggregation results break down at high frequencies. Typically, the volatility persistence implied by an ARCH model of intraday returns understates the persistence of daily volatility. Andersen and Bollerslev suggest that incorporation of deterministic volatility patterns in the intraday ARCH estimations can greatly mitigate this inconsistency. In the present paper, this accommodation is made in the mean parameter of equation (2c). Specifically, I allow $\eta = \eta_t$ (a deterministic time-of-day function). While this approach allows intraday periodicities, it does not (unlike that of Andersen and Bollerslev) allow for interaction between deterministic and stochastic components.

D. Other Related Studies

Econometric models of ordered discrete variables often employ ordered probit models. Probit analyses relevant to the present paper include Hausman, Lo, and MacKinlay (1992) and Bollerslev and Melvin (1994). Broadly speaking, probit models are reduced-form specifications: Robust and flexible in modeling data, but difficult in some respects to interpret structurally.

Effects in the standard probit model flow from observable predetermined variables to an unobservable continuous variable (via a linear specification), and thence to the observed discrete analog of the continuous variable (via a mapping function defined by a set of breakpoints). In Hausman et al. (1992), observable variables (volume, the futures return, the bid-ask spread, etc.) plus a random disturbance determine a latent continuous price change, which then maps into the observed discrete price change.

The requirement that the latent continuous variable depend only on observable data rules out an attractive and basic feature of the discreteness models discussed above, namely, the dependence of the latent unobservable price on its prior value. In the discreteness models cited in Section A, for example, the latent unobservable continuous price m_t depends on m_{t-1}, not on any observable discretized transform of m_{t-1}. Viewed as a reduced form model, a probit specification of observed price changes may accord well with observed data, but there are no obvious candidates for the underlying structural model of continuous prices.

The vagueness of the underlying structural model may be tolerable in many situations. As an example, note that in a probit model the mapping from the latent continuous variable to the discrete observable may distribute or smooth out any sharp peaks in the data. As mentioned above, the sharp peak of the GED distribution leads to computational difficulties in fitting EGARCH models to discrete fat-tailed data. A probit-like modification to the model may ameliorate these problems.

Nevertheless many interesting microstructure models may involve structural features that the probit approach cannot easily resolve. In Harris (1990a), for example, the discrete quotes reflect a continuous variable (m_t) and a continuous parameter (c). The present paper expands this framework in numerous respects, in particular allowing c to follow a persistent stochastic process. When both m_t and c are dynamic and unobservable, structural inference from a reduced-form probit approach is likely to be even more challenging. The state-space approach advocated here, in contrast, starts from the specification of processes for the unobservable continuous variables. Inference concerning the parameters of these processes follows from the (numerical) computation of the sample likelihood of the observed discrete data.

Although the models for discrete transaction prices given in the first part of this section do not possess simple probit reduced forms, probit models are more general in certain respects. Most notably, the mapping from the latent continuous variable to the observable discrete variable in a probit model is described by a set of breakpoints that need only be ordered. This allows for a degree of flexibility in the implicit "rounding" function that is not available in the structural models (both the ones described above and the one proposed in the present paper).

The model proposed by Bollerslev and Melvin (1994) consists of a GARCH specification for deutsche mark/dollar quotes, the output of which (conditional volatility forecasts) feeds into a probit model for the discrete spread. They find that the spread depends positively on the forecasted volatility. Though their model and the present one both allow the market-making costs (cf. c in the above models) to vary stochastically, there are several structural differences.

The Bollerslev–Melvin (1994) ARCH-type component is a continuous model applied directly to the discrete data, and the ARCH-type component of the present model is specified in terms of the continuous unobservable variable. The use of a continuous likelihood for discrete data is feasible in foreign exchange data because the tick size is relatively small. As noted above, this practice is not computationally feasible in the present data set (U.S. equity data), where the tick size is larger.

A more fundamental difference concerns the manner in which discreteness is imposed. In Bollerslev and Melvin, the latent cost of market making is transformed via the probit mapping into a discrete spread. In Dravid's (1991) model and the one advanced here, the bid and ask quotes are discretized separately. The present model also allows for different costs on the bid and ask sides of the market.

In allowing the cost of market making to reflect the volatility forecast, however, the market characterization in the Bollerslev–Melvin analysis is richer than the present model. The present analysis assumes that the implicit efficient price and the costs of market making evolve independently. This simplification is mandated by present computational limitations, however, and is not fundamental to the overall approach.

Another group of analyses relevant to the present paper focuses on modeling time passage and time deformation in market processes. The connection to price discreteness arises from the view that the discrete price grid defines a set of permeable barriers in the space where the continuous latent price evolves. If transactions are assumed to occur when the latent price crosses one of these barriers, the observable crossing times can be used to infer the parameters of the latent price process. This approach is advocated by Marsh and Rosenfeld (1986) and Cho and Frees (1988). More recent studies of time passage/deformation, however, stress characterization of the underlying implicit information process rather than discreteness (Engle and Russell (1998) and Engle (1996)).

II. The Model

As in the preceding section let m be the implicit efficient price of the security in the usual sense of the expectation of the security's terminal value, conditional on all public information. To develop the intuition behind the model, suppose that the agent establishing the bid quote is assumed to be subject to a nonnegative cost of quote exposure $\beta \geq 0$ for incoming sell orders, such that in the absence of discreteness restrictions she would quote a bid price of $m - \beta$. (For ease of exposition, the time subscripts are suppressed in this section.) This cost may be provisionally assumed to reflect fixed transaction costs and asymmetric information costs. Its interpretation and assumed nonnegativity are discussed more fully below. As in the Dravid (1991) model, she is assumed to quote a discrete bid price that is rounded down: $b = \text{Floor}(m - \beta)$. Similarly, the agent establishing the ask quote is assumed to be driven by a quote exposure cost $\alpha \geq 0$ (also for small trades), such that in the absence of discreteness restrictions he would quote an ask price of $m + \alpha$. Constrained by discreteness, he rounds up to a discrete ask quote of $a = \text{Ceiling}(m + \alpha)$. In summary, the bid and ask prices are given by

$$b = \text{Floor}(m - \beta) \tag{3a}$$

$$a = \text{Ceiling}(m + \alpha). \tag{3b}$$

In general, since α and β are related to the costs and benefits of providing liquidity, the interpretation of these parameters depends on the model- or context-determined definition of liquidity. Perhaps the simplest interpretation arises in the sequential trade model of Glosten and Milgrom (1985). The quote setters here are dealers who face a population of informed and uninformed traders. The quote exposure costs are defined implicitly by the conditions $m - \beta$ and $m + \alpha$. The Bertrand competitive outcome ensures the dealers zero expected profits and no ex post regret (net of informational and fixed transaction costs). Due to the asymmetric rounding, trades will be profitable (both ex ante and ex post). This need not lead to competitive price cutting because the discreteness restriction ensures that any such action, if feasible, will result in a loss.

If the trade size is not fixed (a generalization given by Easley and O'Hara (1987)), then α and β reflect the cost of quote exposure for small trades. The interpretation of this depends on the nature of the quoted price schedule. If the quote setter can condition on the total size of the order (e.g., if she is the market's sole supplier of liquidity), then α and β are the marginal costs of small trades. If the quote setter cannot condition thus (e.g., if she has placed a limit order that now holds priority), then α and β are the (expected) marginal costs of any trade that causes the quote to be filled at the quoted price. In this latter case, the conditioning reflects the possibility that the quote is filled in the course of execution of a large incoming order that proceeds to hit other orders further away from the market (see Glosten (1994) and Rock (1998)).

More generally, β is the quote setter's marginal cost on the bid side of the market at a particular time. Some of the components of this cost may be negative. An example of this arises in the context of inventory control. Suppose that the cost of clearing a trade is 0.5 (ticks). A dealer who is short (relative to her desired holdings) may nevertheless bid as if $\beta = 0.2$, reflecting a greater propensity to accumulate a position. There is an implicit benefit of accumulation that may be viewed as a negative cost of -0.3. If the same dealer is also offering the security, we may also expect her α to be high relative to the clearing cost, reflecting her reluctance to accommodate further sales. Analyses of NYSE data suggest that the effect of specialist inventories on quotes is negligible (Hasbrouck and Sofianos (1993) and Madhavan and Smidt (1993)). On the other hand, Zhou (1996) and Lyons (1995, 1996) suggest that inventory effects are important in FX markets.

Negative cost components may also arise in the case of quotes established by public limit order traders. Their principal alternative to a limit order is a market order. They are not seeking to realize a dealer's profit on average, but merely to reduce their costs of trading (Harris and Hasbrouck (1996)). Furthermore, although the dealer (quote setter) in most microstructure models is assumed to be uninformed, the quote exposure costs may reflect the dealer's private information by including a displacement $m - m^*$, where m^* is the dealer's (informed) estimate of the security's value.

In the light of these considerations the nonnegativity restrictions on α and β must be viewed as less than ideal approximations. The requirements are strongly motivated by tractability. From a modeling viewpoint, nonnegative α and β serve to prevent the bid and ask implied by model (3) from coinciding or crossing. Nonnegativity is a simple way to enforce this (realistic) condition.[3] More generally, the bid and ask quotes prevailing at a point in time reflect entered orders that have been subjected to a matching procedure according to the rules of the market. In most markets, if the situation were such that the value of m and the most aggressive (lowest) α and β would result in crossed quotes, a transaction between the two parties would result, leaving the quotes to be determined by the next lowest α and β. The assumptions of a common m and nonnegativity of α and β are expedients that avoid the necessity of explicitly modeling this matching/censoring process.

The view that α and β are the costs of supplying liquidity implicitly assumes competition among quote setters. A referee notes that under imperfect competition α and β may impound rents. Imperfect competition may arise in the usual ways, such as from institutional or regulatory barriers to entry, but it may also occur as a transient phenomenon. For rents to be dissipated by competition among new arrivals, potential competitors must become aware of the rents and react. If (as seems reasonable) limit order traders arrive at varying rates, the limit order book may not always lie in the static equilibrium described by Glosten (1994). Early submitters of limit orders may enjoy rents at the expense of market orders that arrive before the book has reached equilibrium.

Most interesting applications involve situations where the quote exposure costs are random. In line with the preceding discussion, this randomness may reflect variation in costs (such as perceived exposure to adverse information or holding costs) or rents. Alternatively, we may view the quote setter as an agent drawn from a population of traders with random cost functions. If more than one such agent is active at an instant, then the relevant costs are the minimum α's and β's in the set.

Although this model allows for randomness in α, β, and m, the discreteness aspect of the model arises from a nonstochastic transformation. There is no error or disturbance that impounds the effect of discreteness.

There is no assumption that all trades take place at the posted quotes. Following Rock (1998), the posted quotes modeled here are viewed as the best available prices without knowledge of the full size of the incoming order. A trader (such as a specialist or floor trader) who can bid or offer conditional on the incoming order size may better the posted quotes. In practice, such agents bid or offer after the order has been received, and these implicit quotes do not prevail after the transaction has occurred.

In the present model the quote setter's solution to an implicit continuous optimization problem (α or β) is subjected to a transformation to yield discrete quote placement strategies. This must be viewed as an approximation to a decision process in which discreteness is more fundamentally incorporated into the calculation—that is, an integer programming problem. Models along these

[3] In the absence of discreteness noncrossing quotes could be ensured by requiring $m + \alpha > m - \beta \Leftrightarrow \alpha + \beta > 0$, but this does not suffice if the quotes are discrete. Requiring $\alpha + \beta > 1$ (the tick size) will preclude quote crossing, but at the cost of also eliminating one-tick spreads.

lines include those of Anshuman and Kalay (1998), Glosten (1994), Chordia and Subrahmanyam (1995), Bernhardt and Hughson (1996), and Cordella and Foucault (1996). These models are stylized in numerous respects (typically allowing a restricted set of traders and permissible interactions) and focus almost exclusively on information costs. In these characterizations, continuous "prerounding" cost constructs (such as the present α and β) do not explicitly arise. It could nevertheless be argued that such quantities exist implicitly, and that they impound the costs of quote-setting mentioned above (although they would also incorporate discreteness effects).

The present model takes the tick size as an exogenous institutional constraint. In a broader context, of course, the tick size is endogenous. Analyses of optimal tick size include Ahn, Cao, and Choe (1996), Angel (1997), Anshuman and Kalay (1998), Bernhardt and Hughson (1996), Brown, Laux, and Schachter (1991), Chordia and Subrahmanyam (1995), Cordella and Foucault (1999), Glosten (1994), and Harris (1990a, 1990b, 1991)). Furthermore, the present model does not account for clustering (Niederhoffer (1965, 1966), Harris (1991, 1994); Christie and Schultz (1994); Christie, Harris, and Schultz (1994), Hasbrouck (1998)). In a dynamic setting, clustering requires specification of a stochastic mapping from continuous state variables to discrete observations that is more complicated than the simple rounding functions employed here.

III. Inference from Observed Bid and Ask Quotes

Viewed as a transformation of continuous random inputs (m, α, and β) into discrete bid and ask prices, model (3) is a very simple one. From the perspective of the econometrician (and that of many market participants), however, these inputs are sought from the observed bid and ask prices, a more complex inference.

As a function of the observed bid and ask quotes (b, a), the feasible region for (m, α, β) consistent with model (3b) is:

$$Q(b,a) = \{(m,\alpha,\beta)\colon \alpha > 0, \beta > 0, b \leq m - \beta < b + 1 \text{ and } a - 1 < m + \alpha \leq a\}. \tag{4}$$

The inequalities define a convex polytope (geometric solid) of up to six faces.

Although the estimations in this paper are based on equation (4), it is easier to visualize the special case in which the quote exposure costs are the same on both bid and offer sides. In this equal-cost model, $c = \alpha = \beta$, and the feasible region is

$$Q(b,a) = \{(m,c)\colon c > 0, b \leq m - c < b + 1 \text{ and } a - 1 < m + c \leq a\}. \tag{5}$$

These inequalities define a two-dimensional region. Figure 1 depicts these feasible regions for representative one-, two-, and three-tick spreads. The diamond shape of the region $Q(b = 0, a = 2)$, for example, can be viewed as

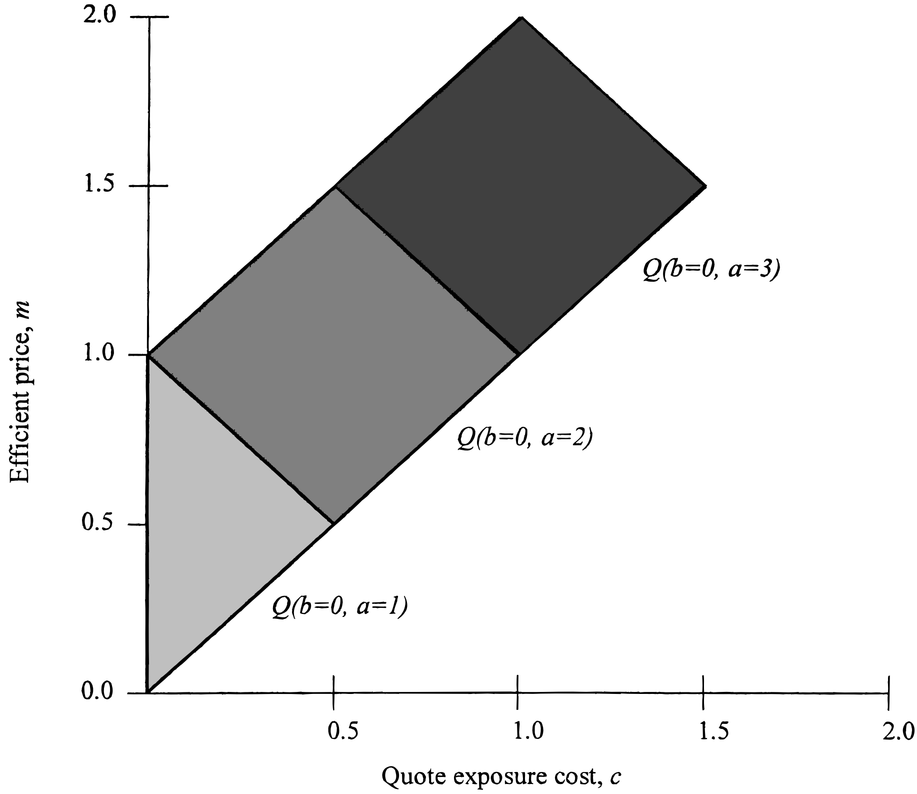

Figure 1. Feasible regions for the equal cost model. As a function of the efficient price m and quote exposure cost c, the discrete bid and ask quotes are given by $b = \text{Floor}(m - c)$ and $a = \text{Ceiling}(m + c)$. Given bid and ask quotes b and a, the region of feasible m and c is $Q(b,a) = \{(m,c)\colon c > 0,\ b \leq m - c < b + 1 \text{ and } a - 1 < m + c \leq a\}$. The figure depicts the regions $Q(0, 1)$, $Q(0, 2)$, and $Q(0, 3)$.

arising in the following way. When c is just slightly greater than zero or slightly less than one, the range of m consistent with $b = 0$ and $a = 2$ is a small neighborhood about one. When $c = 1/2$, $1/2 < m < 3/2$.

Given a prior probability density function $f(m,c)$, the posterior density conditional on observing bid and ask quotes b and a is:

$$f(m,c|b,a) = \begin{cases} \dfrac{f(m,c)}{\Pr(b,a)} & \text{if } (m,c) \in Q(b,a) \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

where $\Pr(b,a) = \int_{Q(b,a)} f(m,c)\,dm\,dc$ is the probability of observing discrete bid and ask quotes b and a. Since this conditioning imposes a truncation on the ranges of the variables, it may seem that the conditional densities are simply truncated versions of the priors. The truncations defined by $Q(b,a)$, however, apply to linear combinations of the variables, not the variables themselves. The shape of $Q(b,a)$ effectively forces a nonlinear transformation on the priors.

The nature of this distortion can be illustrated by supposing flat priors—that is, by assuming that m and c are each independently and uniformly distributed over the positive real line (or, more properly, over a finite region of the positive real line large enough to encompass any values of potential interest). It may be supposed that flat priors would suffice to generate uniform posteriors. On observing a bid of zero and an ask of two, and being unwilling or unable to model c, for example, one may first conjecture simply that the efficient price is uniformly distributed between zero and two. In fact, by integrating over c (horizontally) in the diamond-shaped feasible region defined by $Q(b = 0, a = 2)$ in Figure 1, the correct posterior for m is triangular between 0.5 and 1.5 with a mode at one. Thus, the model imposes structure even under the assumption of flat priors.

The geometry of the feasible region also carries implications for estimation. Still assuming a flat prior for m, the posterior for m conditional on observing $b = 0$, $a = 2$, and in addition knowing c, the posterior for m is always symmetric triangular about one for all c in the feasible range. This suggests that knowledge about c will be less informative about the location of m, but more informative about the dispersion of m. Since the location of m is not sensitive to our prior for c, the cost of assuming a flat prior on c when we are estimating the dynamics of m may be acceptably low. It is worth reemphasizing that even in this case, however, the model imposes structure. Similar arguments can be made for the distribution of m conditional on c.

Similar remarks apply to the more general model with distinct bid- and ask-side exposure costs. By integrating over the α and β axes in the feasible region defined in equation (4), it can be shown that under flat priors on all the variables, the posterior for m conditional on observing bid b and ask a is in general trapezoidal over (b, a). The location-invariance property of the symmetric cost case does not hold here, however. That is, given $b = 0$ and $a = 2$, the location of the posterior for m depends on α and β.

As a numerical example of a nontrivial prior on c, consider the case where $\ln(c)$ is assumed normal with mean $\mu = -1$ and standard deviation $\sigma = 0.6$. This implies $\Pr(a - b = 1) = 0.29$, $\Pr(a - b = 2) = 0.58$, $\Pr(a - b = 3) = 0.11$, and $\Pr(a - b > 3) = 0.03$, that is, frequencies of one-, two-, and three-tick spreads that may be observed for a typical NYSE stock. Assume a flat prior for m—that is, m is uniformly distributed on $(0, \kappa)$ where κ is "large" (but not infinite). The choice of κ is arbitrary; it integrates out of all calculations. The cost parameter is assumed to be independent of the price level, which implies that the prior density of the latent variables may be written as $f(m,c) = f(m)f(c)$.

Figure 2 depicts the prior and conditional density functions conditional on observing bid and ask quotes $b = 0$ and $a = 2$. The prior for m is drawn as a flat line of height κ^{-1}. The conditional density for m is not uniform over

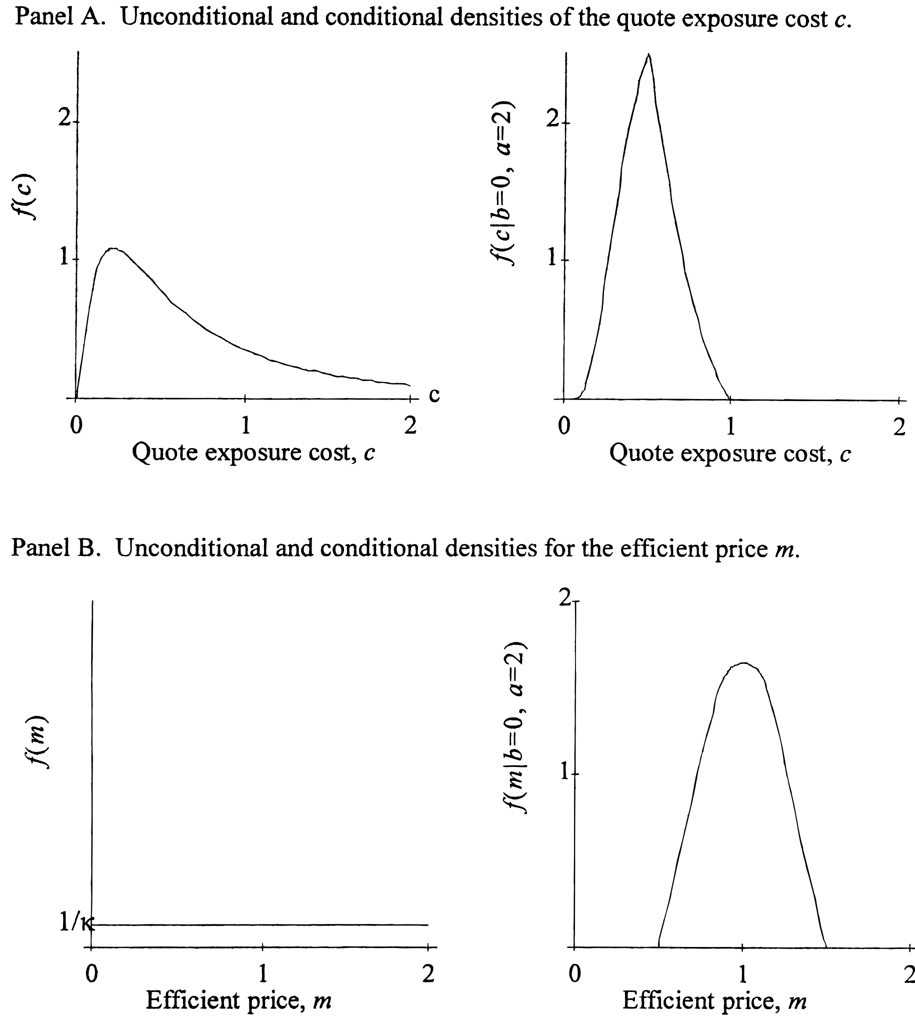

Figure 2. Prior and posterior distributions with symmetric quote exposure costs. The figure depicts prior and posterior probability densities for the efficient price m and quote exposure cost c. The prior density of c is lognormal: $\ln(c) \sim N(\mu = -1, \sigma = 0.6)$. The prior density for m is a uniform prior on the interval $(0, \kappa)$, where κ is an arbitrary positive constant (and does not appear in the posterior densities). The posterior densities are conditional on observing bid and ask quotes of $b = 0$ and $a = 2$.

the allowable range of m $(1/2 < m < 3/2)$. If m is near an endpoint, the range of feasible c values is small, with correspondingly low probability. If m is near the center of the range, the feasible set for c is larger. Similarly the conditional density for c is not simply a truncated log normal, but slopes down gradually to the boundary defined by $c = 1$. When c lies on this boundary, the set of m values consistent with the observed quotes is a single point (of probability zero, given the continuous prior assumed for m). As we move inward from this constraint, the set of feasible m becomes larger.

IV. Maximum Likelihood Estimation

Suppose now that the quote generation process occurs over time periods $t = 1,\ldots,T$ with state variable realizations $z_t = (m_t, \alpha_t, \beta_t)$ and corresponding observed bid and ask quotes $q_t = (b_t, a_t)$. In most applications the state variables are not i.i.d. Typically m_t follows a random walk with non-i.i.d. increments, and the latent cost variables exhibit serial correlation. The model in such cases is neither linear nor Gaussian. The general estimation approach follows Hamilton (1994b) and Harvey (1991). The numerical technique is due to Kitagawa (1987), which is summarized in Hamilton (1994a). (Glosten and Harris (1988) and Harris (1990a) employ another variant of this method.)

The essence of the procedure is a recursive likelihood calculation. Suppose that the probability density function of the current state variables conditional on current and past observed quotes, $f(z_t|q_t,q_{t-1},\ldots)$, is known for some time t. Looking ahead to $t + 1$,

$$f(z_{t+1}|q_t,q_{t-1},\ldots) = \int f(z_{t+1}|z_t)f(z_t|q_t,q_{t-1},\ldots)dz_t, \tag{7}$$

where $f(z_{t+1}|z_t)$ is the (possibly time-dependent) state transition density function. The joint density of state variables and observed quotes is

$$f(z_{t+1},q_{t+1}|q_t,q_{t-1},\ldots) = \begin{cases} f(z_{t+1}|q_1,q_{t-1},\ldots), & \text{if } z_{t+1} \in Q_{t+1} \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

The conditional probability of observing q_{t+1} is

$$\Pr(q_{t+1}|q_t,q_{t-1},\ldots) = \int_{z_{t+1}\in Q_{t+1}} f(z_{t+1}|q_t,q_{t-1},\ldots)dz_{t+1}, \tag{9}$$

where $Q_{t+1} \equiv Q(b_{t+1},a_{t+1})$ as defined in equation (4). Continuing,

$$\begin{aligned} f(z_{t+1}|q_{t+1},q_t,q_{t-1},\ldots) &= \frac{f(z_{t+1},q_{t+1}|q_t,q_{t-1},\ldots)}{\Pr(q_{t+1}|q_t,q_{t-1},\ldots)} \\ &= \begin{cases} \dfrac{f(z_{t+1}|q_t,q_{t-1},\ldots)}{\Pr(q_{t+1}|q_t,q_{t-1},\ldots)} & \text{if } z_{t+1} \in Q_{t+1} \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{10}$$

This completes the update.

In an estimation context the state transition density is assumed to depend on a parameter vector θ: $f(z_t|z_{t-1}) = f(z_t|z_{t-1};\theta)$. Maximum likelihood estimation proceeds by maximizing the sum of the log of the conditional probabilities $\Pr(q_{t+1}|q_t,q_{t-1},\ldots;\theta)$. Estimated residuals, in the usual sense of the difference between actual and predicted values, do not exist in this framework. Insight into the behavior of the model can alternatively be obtained from the filtered estimates of the state variables, defined as $\hat{z}_t \equiv E[z_t|q_t,q_{t-1},\ldots] = \int z_t f(z_t|q_t,q_{t-1},\ldots)dz_t$.

The model possesses two essential features that simplify the implementation. The mapping from the state variables to the observations is (although nonlinear) nonstochastic. This means that the mapping given in equation (8) and the integration in equation (9) do not involve observational errors. Furthermore, the observed quotes sharply bound the possible values of the state variables: Q_t defines a small region of $\mathbf{R}^3$. In an estimation context, therefore, $f(z_{t+1}|q_t,q_{t-1},\ldots)$ in equation (7) need only be evaluated for points that map to observed quotes $z_{t+1} \in Q_{t+1}$, and the integration need only cover those points $z_t \in Q_t$.

Although straightforward in principle, this update requires the evaluation of two integrals for which closed-form solutions are not readily available. In the standard Kalman filter, all joint, marginal, and conditional densities are normal, and the results of the integrations are summarized by update formulas for the conditional means and variances. In the present case, successive updates involve computation of nested, truncated densities of increasing dimension. These are calculated using a grid approximation. Appendix A discusses the computational details. Appendix B presents some simulation results.

V. The Full Dynamic Model

The full model estimated here combines the EGARCH/GED model (equation (2c)) for the efficient price dynamics with the quote exposure cost model from Section II, allowing intraday periodicities in both components.

The bid and ask quote exposure costs are assumed to evolve as:

$$\ln(\alpha_t) = \mu_t + \phi(\ln(\alpha_{t-1}) - \mu_{t-1}) + \nu_t^\alpha \tag{11a}$$

$$\ln(\beta_t) = \mu_t + \phi(\ln(\beta_{t-1}) - \mu_{t-1}) + \nu_t^\beta, \tag{11b}$$

where ν_t^α and ν_t^β are independently distributed as $N(0,\sigma_\nu^2)$. This specification allows for deterministic time variation in the mean, and a persistent stochastic component. To allow for an intraday “U” effect, the deterministic component of the cost process is specified as a combination of exponential decay functions:

$$\mu_t = k_1 + k_2^{open}\exp(-k_3^{open}\,\tau_t^{open}) + k_2^{close}\exp(-k_3^{close}\,\tau_t^{close}), \tag{12}$$

where τ_t^{open} is the elapsed time since the opening quote of the day (in hours) and τ_t^{close} is the time remaining before the scheduled market close (in hours).

The present implementation assumes that the efficient price evolves independently of the quote exposure costs. In the present application, however, $|\zeta_{t-1}|$ in equation (2c) is not observed. As a more tractable alternative to carrying σ_t as a (fourth) unobservable state variable, I assume for purposes of estimation that the variance process is driven by the conditional expectation of the absolute efficient price increment. That is, $|\zeta_{t-1}|$ in equation (2c) is replaced by its conditional expectation $E[|\zeta_{t-1}||q_{t\text{-}2},q_{t\text{-}3},\ldots]$. This is a filtered estimate, which is easily computed in the course of the iterative update. The deterministic component of the variance is modeled as:

$$\eta_t = \begin{cases} l_1 + l_2^{open}\exp(-l_3^{open}\,\tau_t^{open}) + l_2^{close}, & \text{if } t \text{ is an intraday interval} \\ \eta^{overnight}, & \text{if } t \text{ is an overnight interval.} \end{cases} \tag{13}$$

The intraday component is formally similar to that used for the costs in equation (12). Since the end-of-day variance elevation is found to be essentially concentrated in the last 15 minutes, however, equation (13) uses an end-of-day dummy variable (instead of the exponential function used in equation (12)). The exponential functions used here are designed to capture the well-known "U" shapes found in intraday U.S. equity data. Andersen and Bollerslev (1997a, 1997b, 1998) employ a more general flexible Fourier form.

The driving disturbances (ν_t^{α}, ν_t^{β}, and u_t) are assumed to be independent. This allows the state transition density $f(m_t,\alpha_t,\beta_t|m_{t-1},\alpha_{t-1},\beta_{t-1})$ to be factored as $f(m_t|m_{t-1})f(\alpha_t|\alpha_{t-1})f(\beta_t|\beta_{t-1})$. Although this appreciably enhances the speed of the computations, it is less desirable from a modeling viewpoint. The assumption of independence between α and β is most appropriate to a market setting in which the bid and ask quotes reflect limit orders originating from different idiosyncratic liquidity traders. In the case where the quotes reflect the interests of a single dealer or derive from common influences, it is more appropriate to allow for positive correlation between the two costs. One may also expect (following Bollerslev and Melvin (1994)) correlation between the quote exposure costs and volatility in m. With one exception, these generalizations are not at present computationally feasible, although they certainly provide directions for future research. The exception involves the equal-cost model defined by $\alpha = \beta = c$, in which case the state space is two-dimensional (m and c). Nevertheless, although independence of the components is assumed in the present model, it is not enforced by the estimation procedure. It is therefore possible to investigate the independence using the filtered estimates of the state variables.

The specification described above jointly models the bid and ask quote exposure costs and the efficient price. Given the computational complexity of the likelihood calculation, however, it is also useful to consider simpler models. For example, if only the quote exposure costs are of interest, the model may be estimated assuming a flat prior on the efficient price (at each

Table I

Descriptive Statistics for 15-minute Bid Changes, Alcoa, 1994

The bid changes (in 1/8 dollar ticks) are computed for Alcoa for all trading days in 1994 (plus the overnight change).

	Intraday	Overnight
N	6,528	251
Min (ticks)	−10	−15
Max (ticks)	11	19
Mean (ticks)	0.03	−0.21
Std. dev. (ticks)	1.58	3.70
Distribution		
% with no change	39%	18%
% with 1-tick change	37%	27%
% with >1-tick change	25%	47%

point in time). This variant, consisting solely of the cost equations (11b) and (12), is termed the cost model. When m is eliminated as a state variable, the numerical grid approach is still necessary due to the stochastic variation in the cost, but the reduction in dimension speeds computation. Alternatively, if the efficient price dynamics are the sole concern, one may estimate the EGARCH/GED component of the model under the assumption of flat priors on the quote exposure costs. This variant is termed the EGARCH/GED model. Both of the cost and EGARCH/GED submodels are in principle misspecified, but as a practical matter the cost of the misspecification may be small relative to the computational gains.

VI. Estimation

A. Data

I estimate the specifications described in the preceding section for NYSE bid and ask quotes for Alcoa (ticker symbol AA) for all trading days in 1994. Alcoa is alphabetically the first Dow Stock and is viewed as a representative high-activity security. Bid and ask quotes are those prevailing at the close of 15-minute intervals. The first observation of a day generally corresponds to 9:45, the last to 16:00 (26 daily observations). There are 6,780 observations (reflecting delayed openings and early closings). The 15-minute interval is chosen as a frequency that is high enough to be of microstructure interest, yet still yield a computationally tractable number of observations.

Table I reports descriptive statistics for the absolute value of the bid first-differences. (Results for ask first-differences are virtually identical.) The proportion of intervals for which the bid change is zero is 39 percent (intraday) and 18 percent (overnight). Not reported in the table is the additional finding that in 24 percent of the intraday intervals and in eight percent of the overnight intervals, there is no change in either the bid or the ask quote. If

Table II

Descriptive Statistics for Bid-Ask Spread at 15-Minute Intervals, Alcoa, 1994

Spreads (in 1/8 dollar ticks) are computed for Alcoa at 15-minute intervals during the trading day, for all trading days in 1994.

N	6,780
Min	1
Max	5
Mean	1.65
Std. dev.	0.67
Distribution	
1-tick	45.9%
2-tick	43.5%
3-tick	10.3%
4 or more ticks	0.3%

the underlying changes in the efficient price are viewed as arising from a continuous distribution of modest leptokurtosis, these figures suggest that the efficient price changes are not large relative to the tick size. The extreme values in the sample lie roughly seven standard deviations from the mean for the intraday intervals and five standard deviations from the mean for the overnight intervals.[4]

Table II reports descriptive statistics for the bid-ask spread. There is clear variation in the spread. In a sense, one purpose of the present model is the allocation of this variation to deterministic and stochastic effects.

B. Estimates of the Full Model

Table III reports parameter estimates. For purposes of exposition, these may be grouped as cost- and EGARCH (variance)-related. The EGARCH-related parameter estimates suggest a strong persistent stochastic component of the return variance. The GED tail-thickness parameters of $\nu^{day} = 0.86$ and $\nu^{overnight} = 1.02$. For comparison purposes, Figure 3 graphs the GED density with $\nu = 0.86$ against the standard normal.

A GED with $\nu = 0.86$ has a standardized kurtosis in excess of the normal of roughly 4.6. To place this in perspective, the intraday 15-minute bid changes for the present sample (standardized by time-of-day standard deviation) exhibit an excess kurtosis of 3.84. It is recognized, however, that both discreteness and ARCH effects acting separately can affect kurtosis (see Gottlieb and Kalay (1985) for the former; Bollerslev and Melvin (1994) for the latter). The present analysis therefore suggests that the kurtosis in the bid price

[4] Under the assumption of homoskedastic normality, such extreme values possess implausibly low probability. For a normally distributed variate, the probability of observing an extreme value seven standard deviations from the mean in a sample of 6,528 observations is approximately 1×10^{-10}; that of an extreme value five standard deviations from the mean in a sample of 251 observations is approximately 1×10^{-7}.

Table III

Model Estimates

The state variables in the model are $z_t = (m_t, \alpha_t, \beta_t)$ where m_t is the implicit efficient price, α_t is the quote exposure cost on the ask or offer side of the market, and β_t is the quote exposure cost on the bid side of the market. t indexes 15-minute intraday intervals (plus the overnight period). The dynamics of the state variables are

$$m_t = m_{t-1} + u_t$$

$$\ln(\alpha_t) = \mu_t + \phi(\ln(\alpha_{t-1}) - \mu_{t-1}) + \nu_t^{\alpha}$$

$$\ln(\beta_t) = \mu_t + \phi(\ln(\beta_{t-1}) - \mu_{t-1}) + \nu_t^{\beta},$$

where ν_t^{α} and ν_t^{β} are independently distributed as $N(0,\sigma_\nu^2)$. The efficient price disturbance, u_t, has standard deviation σ_t. Its standardized value $\zeta_t = u_t/\sigma_t$ is distributed as $f_{GED}(\zeta_t;\nu)$ where ν is the tail-thickness parameter. The efficient price variance follows a modified EGARCH process:

$$\ln(\sigma_t^2) = \eta_t + \varphi(\ln(\sigma_{t-1}^2) - \eta_{t-1}) + \gamma(E_{t-1}|\zeta_{t-1}| - E|\zeta_{t-1}|),$$

where $E_{t-1}|\zeta_{t-1}| = E[|u_{t-1}|/\sigma_{t-1}|q_{t-1},q_{t\text{-}2},\ldots]$ is the filtered estimate conditional on the bid and ask prices through $t-1$. The k and l values parameterize the exponential functions characterizing the intraday deterministic components of the quote exposures costs and efficient price variance (μ_t and η_t). The observations are the discrete bid and ask quotes: $b_t = \text{Floor}(m_t - \beta_t)$ and $a_t = \text{Ceiling}(m_t + \alpha_t)$. In the full model, all parameters are jointly estimated; the cost and EGARCH estimates are based on submodels. The models are estimated for Alcoa over all trading days in 1994, with t indexing 15-minute intervals within the day (and the overnight interval). Standard errors are reported in parentheses.

	Model		
	Full	Cost	EGARCH
Quote exposure cost parameters			
k_1	−1.67	−1.68	
	(0.03)	(0.03)	
k_2^{open}	0.45	0.46	
	(0.06)	(0.06)	
k_3^{open}	2.42	2.40	
	(0.72)	(0.69)	
k_2^{close}	0.21	0.19	
	(0.07)	(0.07)	
k_3^{close}	3.48	3.50	
	(2.50)	(2.68)	
ϕ	0.37	0.39	
	(0.03)	(0.03)	
σ_ν	0.86	0.86	
	(0.03)	(0.02)	
EGARCH parameters			
l_1	0.39		0.35
	(0.06)		(0.07)
l_2^{open}	1.73		1.76
	(0.23)		(0.24)
l_3^{open}	1.11		1.11
	(0.23)		(0.23)
l_2^{close}	0.59		0.61
	(0.14)		(0.15)
$\eta^{overnight}$	2.73		2.73
	(0.13)		(0.14)
φ	0.88		0.90
	(0.02)		(0.02)
γ	0.29		0.28
	(0.03)		(0.03)
ν^{day}	0.86		0.81
	(0.02)		(0.02)
$\nu^{overnight}$	1.02		1.01
	(0.12)		(0.11)

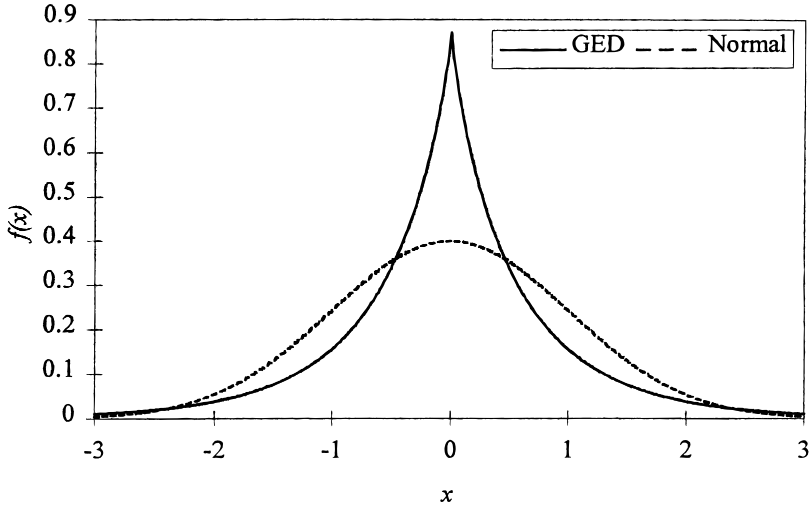

Figure 3. Normal and GED probability densities. The figure depicts the probability density functions for the standard normal and standard GED with tail-thickness parameter $\nu = 0.86$.

changes is similar in magnitude to that estimated for the underlying efficient price process (the driving GED). This implies that kurtosis is fundamental and not an artifact of discreteness in the data.

For daily CRSP index returns, Nelson (1991) found $\nu = 1.6$. The present estimates imply a more pronounced leptokurtosis, consistent with a "lumpy" intraday information arrival process for individual stocks. This kurtosis is mitigated in the daily index returns due to aggregation over firms and time.

Turning now to the quote-exposure cost estimates, the deterministic parameters depict the usual U-shaped intraday pattern, although the standard errors of the decay rates are large. Of more interest is the characterization of the stochastic component. Both the disturbance variance σ_v and the autoregressive parameter ϕ are strongly positive. The autoregressive parameter suggests that 37 percent of the excess log cost persists at the subsequent time point (15 minutes later).

The relative importance of the deterministic and stochastic sources of variation in the quote exposure cost can be ascertained from simulations of the model using the parameter estimates. For a simulation of 2,500 days, Figure 4 depicts the intraday values of the 10^{th}, 50^{th}, and 90^{th} percentiles of the cost expressed in dollars per share. The 50^{th} percentile (the median) displays the "U" shape characteristically found in spreads. The median is roughly four cents per share at the open and two cents thereafter, rising slightly at the close. Most importantly, the elevation associated with the beginning and end of trading is modest compared with the stochastic variation implied by the 10^{th} and 90^{th} percentile bands. This suggests that the stochastic component is relatively large.

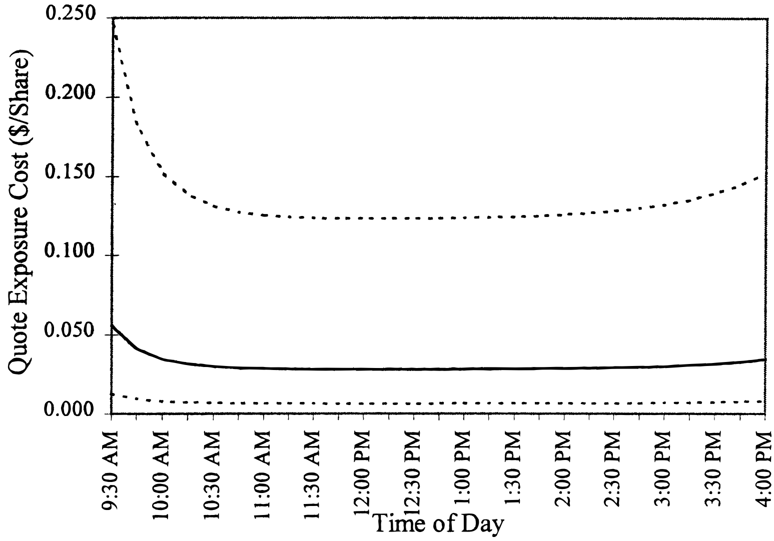

Figure 4. Intraday patterns. The figure depicts the time of day patterns in the quote exposure cost for ticker symbol AA implied by the model and estimates given in Table III. The solid line is the 50^{th} percentile of the cost. The upper and lower dashed lines are the 10^{th} and 90^{th} percentiles (respectively). (Note that these are not estimation confidence intervals.)

The finding of a mean-reverting persistent component in the cost of quote exposure is consistent with several models of economic behavior. It may reflect mean-reversion in the underlying cost determinants (such as asymmetric information exposure costs or inventory holding costs related to risk) that are common across all actual and potential quote-setters. It may also reflect, however, the arrivals and departures of individual traders with differing costs. A buyer anxious to trade, for example, might enter a limit order that betters the prevailing bid. At some point this limit order is likely to be removed, either because it is hit or else because the trader withdraws it and replaces it with a market buy order.

C. Properties of the Filtered Estimates

As noted in Section IV the estimation approach readily generates filtered estimates of the state variables $\hat{z}_t \equiv E[z_t|q_t, q_{t-1}, \ldots]$. Let the components of $\hat{z}_t$ be denoted $\hat{m}_t$, $\hat{\alpha}_t$, and $\hat{\beta}_t$, and let $\hat{\sigma}^2_{t+1}$ denote the forecasted volatility from the EGARCH component of the model (conditional on quotes observed through time t). The disturbance in the α_t transition equation (11b) may be isolated as $\nu_t^\alpha = \ln(\alpha_t) - \mu_t - \phi(\ln(\alpha_{t-1}) - \mu_{t-1})$. The corresponding estimate is $\hat{\nu}_t^\alpha = \ln(\hat{\alpha}_t) - \mu_t - \phi(\ln(\hat{\alpha}_{t-1}) - \mu_{t-1})$ where parameter estimates are used for ϕ and μ. (Note that since the conditioning information sets for $\hat{\alpha}_t$ and $\hat{\alpha}_{t-1}$ differ, $\hat{\nu}_t^\alpha$ is not, strictly speaking, a filtered estimate.) We may construct $\hat{\nu}_t^\beta$ in a similar fashion, and let $\hat{u}_t = \hat{m}_t - \hat{m}_{t-1}$.

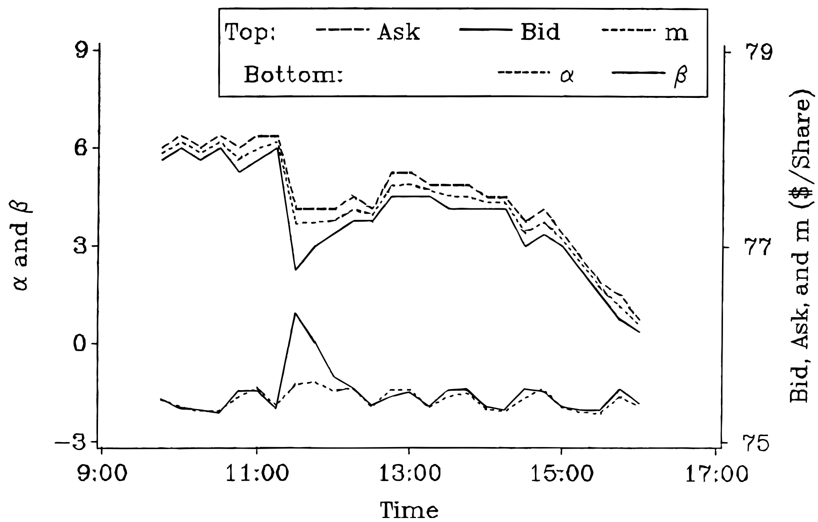

Figure 5. Bids, asks, and filtered variables for Alcoa on February 4, 1994. The figure depicts bid and ask quotes on February 4, 1994 on the right-hand scale. α and β are the filtered estimates of the model quote exposure cost parameters on the ask and bid sides of the market (left-hand scale); m is the filtered estimate of the implicit efficient price (right-hand scale). All data are as of the end of 15-minute intervals.

The behavior of the filtered estimates may be illustrated by considering the late morning of Friday, February 4, 1994, when Alcoa experienced large price movements. These movements were market-related: The major indices were down sharply for the day, there were no unusually large trades in Alcoa, and a media search turned up no major announcements. Figure 5 depicts the quotes and filtered estimates of the model variables. Between 11:15 and 11:30, both bid and ask quotes drop dramatically and the spread widens. Yet the model-implied filtered estimates do not simply impute a symmetric increase in the costs of quote exposure on the bid and ask sides. Instead the model attributes the bulk of the spread increase to the bid-side (β). Over the next two intervals this shock decays and the spread narrows. It is important to note, however, that this narrowing is not symmetric—it is achieved solely by successive increases in the bid. An alternative view of this path may be gleaned from the path of the (filtered) implicit efficient price. This price does not simply track the midpoint of the bid and ask quotes. Instead it behaves asymmetrically, lying above the quote midpoint after the drop, suggesting that the plunge in the bid is in large part a transient liquidity effect.

Although the model specification assumes that all disturbances are uncorrelated, this property is not forced in the estimation. The bivariate correlations between estimated disturbances are reported in the upper left of Table IV. Most importantly, $Corr(\hat{\nu}_t^{\alpha}, \hat{\nu}_t^{\beta}) = 0.81$. This implies that (contrary

Table IV

Correlations between Filtered Estimates

Based on the model and parameter estimates given in Table III, I compute filtered estimates (i.e., expectations conditional on current and preceding observations) for $\hat{\nu}_t^\alpha$ and $\hat{\nu}_t^\beta$ (the cost disturbances on the ask and bid sides of the market); $\hat{u}_t/\hat{\sigma}_t$ (the standardized increment in the efficient price); $\hat{\alpha}_t - \mu_t$ and $\hat{\beta}_t - \mu_t$ (the quote exposure costs on ask and bid sides less the time-of-day mean); and $\ln(\hat{\sigma}_{t+1}^2) - \eta_{t+1}$ (the efficient price variance implied by the EGARCH component of the model, less the corresponding time-of-day mean).

	$\hat{\nu}_t^\alpha$	$\hat{\nu}_t^\beta$	$\hat{u}_t/\hat{\sigma}_t$	$\hat{\alpha}_t - \mu_t$	$\hat{\beta}_t - \mu_t$	$\ln(\hat{\sigma}_{t+1}^2) - \eta_{t+1}$
$\hat{\nu}_t^\alpha$	1.0					
$\hat{\nu}_t^\beta$	0.81	1.0				
$\hat{u}_t/\hat{\sigma}_t$	0.16	−0.16	1.0			
$\hat{\alpha}_t - \mu_t$	0.93	0.75	0.16	1.0		
$\hat{\beta}_t - \mu_t$	0.75	0.93	−0.15	0.81	1.0	
$\ln(\hat{\sigma}_{t+1}^2) - \eta_{t+1}$	0.13	0.12	0.05	0.18	0.17	1.0

to the modeled specification) there is substantial correlation between the random components of the quote exposure costs on the bid and ask side. This suggests that underlying economic determinants such as order processing and asymmetric information costs also positively covary on the bid and ask sides. The positive correlation cannot be attributed to inventory control. A quote-setter attempting to attract an unbalanced order flow sets quotes asymmetrically about m, implying a negative correlation between α and β. The standardized efficient price innovation $\hat{u}_t/\hat{\sigma}_t$ exhibits a modest positive correlation (0.16) with $\hat{\nu}_t^\alpha$ and a modest negative correlation (−0.16) with $\hat{\nu}_t^\beta$, apparently a consequence of mean reversion in the bid and ask.

The filtered estimates also offer some insights into quote exposure costs and volatility. The lower right portion of Table IV reports correlations between these variables with intraday components removed. The correlation between $ln(\hat{\sigma}_{t+1}^2) - \eta_{t+1}$ *and* $\hat{\alpha}_t - \mu_t$ is 0.18, and that between $ln(\hat{\sigma}_{t+1}^2) - \eta_{t+1}$ and $\hat{\beta}_t - \mu_t$ is 0.17. This suggests that quote exposure costs reflect expected volatility.

The nonzero correlations found between pairs of estimated disturbances may partially arise from estimation errors. It is, as noted above however, economically reasonable to expect positive correlation between costs on both sides of the market, and between these costs and expected volatility. Although certainly desirable, it is not computationally feasible with the technology used in this paper to allow for such dependencies. It is nevertheless possible to address several points regarding the likely consequences of model misspecification.

First, as demonstrated in the next section, estimates of cost parameters of the model are not sensitive to particulars of the efficient price modeling, and vice versa. This suggests that the impact of misspecification is largely confined to the misspecified component, and in particular that the estimates are not sensitive to the assumed independence between the efficient price

Table V

Correlations between Filtered Estimates and Market Activity

This table presents bivariate correlations between filtered estimates of the costs of quote exposure on the bid and ask side less time-of-day means ($\hat{\alpha}_t - \mu_t$ and $\hat{\beta}_t - \mu_t$) and volume measures cumulated over the preceding 15-minute interval. $\sqrt{\$Volume}$ is the square root of the total dollar volume. The remaining measures are based on transaction value signed positively if the price is above the prevailing quote midpoint and negatively if below. *Signed* $\sqrt{\$Volume}$ is the cumulation over the interval; $(\ )^+$ and $(\ )^-$ indicate the positive and negative parts. All volume variables are standardized by time-of-day.

	$\hat{\alpha}_t - \mu_t$	$\hat{\beta}_t - \mu_t$	$\sqrt{\$Volume}$	$\frac{Signed}{\sqrt{\$Volume}}$	$\left(\frac{Signed}{\sqrt{\$Volume}}\right)^+$	$\left(\frac{Signed}{\sqrt{\$Volume}}\right)^-$
$\hat{\alpha}_t - \mu_t$	1.0					
$\hat{\beta}_t - \mu_t$	0.81	1.0				
$\sqrt{\$Volume}$	0.13	0.09	1.0			
$\frac{Signed}{\sqrt{\$Volume}}$	0.14	−0.08	0.23	1.0		
$\left(\frac{Signed}{\sqrt{\$Volume}}\right)^+$	0.15	−0.02	0.52	0.85	1.0	
$\left(\frac{Signed}{\sqrt{\$Volume}}\right)^-$	−0.07	0.11	0.21	−0.78	−0.33	1.0

increments and the quote exposure costs. Furthermore, as to the assumed independence between α and β, it is likely that any increase in $Corr(\nu_t^{\alpha}, \nu_t^{\beta})$ (above zero) would be accompanied by a decline in σ_ν^2.[5]

Finally, although the model does not incorporate trade variables, it is useful to examine the joint behavior of trade-related series and the filtered estimates. For each 15-minute interval, I construct the total dollar trading volume and signed dollar volume. The latter variable is constructed in the usual fashion (see, e.g., Hasbrouck (1991)), with the dollar value of the trade being signed positively if the trade price lies above the prevailing quote midpoint and negatively if it lies below. To mitigate the influence of outliers, I apply a (signed) square-root transformation. To gauge the impact of nonlinearities and asymmetries, I examine the positive and negative parts of the signed order flow measures. All variables are normalized to remove time-of-day patterns.

Table V reports the correlations. The model quote-exposure-cost variables (end-of-period estimates) are positively correlated with total volume over the period. This is consistent with the hypothesis that both variables are jointly (positively) driven by the extent of information asymmetries over the

[5] The following intuition may be useful. If quotes are not confined to the discrete grid, the spread is $\alpha + \beta$, possessing expected value $E[\alpha] + E[\beta]$ (with $E[\alpha] = E[\beta]$) and variance $Var(\alpha) + Var(\beta) + 2Cov(\alpha,\beta)$ (with $Var(\alpha) = Var(\beta)$). Taking these quantities as given (as if we are trying to match a sample mean and variance), any increase (from zero) in $Corr(\alpha,\beta)$ must be offset by a decline in $Var(\alpha) = Var(\beta)$.

period. The correlation between the cost variables and the signed volume measures also appears to work in this direction. High signed volume ("an influx of buy orders") is associated with an increase in the ask cost; low signed volume ("sales") is associated with an increase in the bid cost. This pattern is amplified by the correlations involving the positive and negative parts of the signed order flow.

D. Estimates of the Cost and EGARCH/GED Models

Both the cost and the EGARCH/GED models are computationally simpler subcases of the full model. The first follows from an ongoing assumption of a diffuse prior for the efficient price; the second assumes a diffuse prior on the quote exposure costs. The resulting estimates are given in the last two columns of Table III. Not only are the estimates virtually identical to those obtained for the full model, but so are the estimated standard errors. Although one may have suspected that the full specification would result in more precise estimates, this does not appear to be the case.

Several considerations could account for this. One possibility is simply general model misspecification. But even in a correctly specified model, the information about α and β contributed by m (and vice versa) might be small. Section III points out that the model imposes structure even under the assumption of flat priors. With flat priors in the equal-cost model defined by $\alpha = \beta = c$, it is shown that c is likely to be more informative about the dispersion of m than the location. In estimates of the equal-cost model (not reported), the parameters governing the dynamics of m are essentially similar to those of the full model reported here.

The equal-cost and full models may be thought of as resulting from polar assumptions about the correlation (one and zero, respectively) between α and β. The similarity of the estimates may stem from the inability of the data to identify the correlation. Ignoring discreteness issues for the moment (which compounds the difficulties), the statistical situation is similar to one in which we are attempting to infer the joint distributional properties of two unobserved variables (α and β) on the basis of observing their sum (the spread). The decreased informativeness of the sum taxes both the data and the model specification.

Implications for Volatility Persistence

The autoregressive parameter estimate in the EGARCH component of the model ($\varphi = 0.88$) implies a half-life of about six (15-minute) periods. To put this in perspective, I estimate an EGARCH specification analogous to equation (2c) for AA's daily returns (CRSP, 1985–1995):

$$\ln(\sigma_t^2) = \underset{(0.114)}{-\,8.168} + \underset{(0.007)}{0.987}\left(\ln(\sigma_{t-1}^2) - \underset{(0.114)}{8.168}\right) + \underset{(0.020)}{0.090}(|\zeta_{t-1}| - E|\zeta_{t-1}|). \tag{14}$$

(Standard errors are given in parentheses; the estimated tail-thickness parameter of the GED distribution is $\nu = 1.317$ (0.046).) The autoregressive parameter in the daily estimation is much larger than would be suggested by aggregation of the 15-minute specification. This inconsistency frequently arises when ARCH models are estimated over different time scales. In particular, it is similar to the findings of Andersen and Bollerslev (1997b) for S&P futures and DM/$ returns. Addressing this phenomenon, Andersen and Bollerslev (1997a, 1998a) propose models involving mixing of information processes with varying decay rates and differentiation between public and trade-induced information.

It is noted in the discussion of the filtered estimates that the quote exposure costs (α and β) exhibits modest positive correlation with the fitted volatility forecasts in the intraday analysis. The correlation between these variables and the fitted *daily* volatility forecasts is also positive: The correlation between the filtered end-of-day α and the fitted daily $\ln(\sigma_t^2)$ is 0.080, with an associated p-value of 0.21; the corresponding correlation for the filtered end-of-day β is 0.150 (with $p = 0.02$). Although these are weaker than the corresponding intraday results, it should be noted that there are only 252 daily observations. These results accordingly provide inconclusive support for the hypothesis that spreads impound information about long-term volatilities.

VII. Conclusion

This paper presents a dynamic model of discrete bid and ask quotes. The discrete quotes are rounded transformations of a continuous efficient price and continuous quote exposure costs. The latter are presumed to capture most of the costs usually associated with market making or limit order placement, such as fixed transaction costs and asymmetric information costs. A referee suggests that they may also impound rents to supplying liquidity. The full statistical model is a rich one, allowing for stochastic and deterministic time variation in the efficient price volatility and the quote exposure costs. The model is estimated by maximum likelihood using a nonlinear state-space filtering approach due to Kitagawa (1987).

This model is estimated for NYSE bid and ask quotes collected at the end of 15-minute intervals for Alcoa over 1994. The estimates imply a stochastic component to quote exposure costs that is persistent and also large relative to the deterministic intraday “U” shape. This may be interpreted as reflecting variation in the underlying components of the costs (e.g., inventory or asymmetric information) or in the rents to supplying liquidity.

For market participants, the magnitude of the stochastic component emphasizes the importance for trading strategies of ascertaining current market conditions, rather than simply conditioning on time of day. For economists, this finding challenges theorists developing normative trading strategies under randomly varying market conditions. It also indicates the desirability of accommodating stochastic microstructure effects in statistical models of security price dynamics.

The ARCH component of the model captures persistent intraday stochastic volatility in the efficient price. The in-sample filtered estimates of the system exhibit positive correlation between quote exposure costs on the bid and ask sides of the market, suggesting cost determinants that are common to both sides. Moreover, there is positive correlation between the filtered quote exposure costs and the ARCH volatility forecasts. This is consistent with the economically reasonable connection between market-making costs and risk, and suggests the potential usefulness of bid and ask models in forming volatility forecasts.

Investigation of more realistic models may require, however, methodological advances. The Kitagawa (1987) approach used here relies on multidimensional numerical integration. The approximations necessary for computational feasibility are likely to be highly model-specific (like those in the present paper). A promising alternative technology involves casting the problem in a Bayesian framework and estimating parameters via Markov chain Monte Carlo (MCMC) methods. Manrique and Shephard (1997) and Hasbrouck (1999) discuss MCMC variants of the present model. MCMC computations are approximately *linear* in the number of parameters or latent variables. As such, these methods hold considerable promise for estimating more complicated models.

Appendix A. Approximation Methods

The present analysis follows Kitagawa (1987) in approximating the conditional densities required in Section IV by step functions defined over a lattice. The three-dimensional state variable is $z_t = (m_t, \alpha_t, \beta_t) \in \mathbf{R} \times \mathbf{R}^+ \times \mathbf{R}^+$. The space $\mathbf{R} \times \mathbf{R}^+ \times \mathbf{R}^+$ is assumed to be partitioned into a set of lattice cells $\{C_t^1, C_t^2, \ldots\}$ where each $C_t^i \in \mathbf{R} \times \mathbf{R}^+ \times \mathbf{R}^+$ is a rectangular solid. The conditional density $f(z_t | q_t, q_{t-1}, \ldots)$ is approximated as

$$f(z_t | q_t, q_{t-1}, \ldots) \approx \sum_{i=1}^{\infty} \delta_t^i I(z_t \in C_t^i), \tag{A1}$$

where I is an indicator variable and the δ's are parameters. The state transition densities $f(z_{t+1} | z_t)$ are replaced by the discrete transition probabilities $Pr(C_{t+1}^j | C_t^i)$, and the integration in equation (7) becomes the summation

$$\Pr(C_{t+1}^j | q_t, q_{t-1}, \ldots) = \sum_{i=1}^{\infty} \Pr(C_{t+1}^j | C_t^i) \Pr(C_t^i | q_t, q_{t-1}, \ldots). \tag{A2}$$

The integration in equation (9) becomes

$$\Pr(q_{t+1} | q_t, q_{t-1}, \ldots) = \sum_{j=1}^{\infty} \Pr(C_{t+1}^j | q_t, q_{t-1}) \frac{\mathrm{Vol}(C_{t+1}^j \cap \mathbf{Q}_{t+1})}{\mathrm{Vol}(C_{t+1}^j)}, \tag{A3}$$

where Vol(·) is the volume (in cubic ticks) of its argument. When the lattice cell lies entirely within the feasible region Q_{t+1}, the summand in (A3) is simply $\Pr(C_{t+1}^{j}|q_t,q_{t-1},\ldots)$. But when only a portion of the cell lies within the feasible region, the probability is weighted by the ratio of the feasible volume to the total. The calculation of the intersection volume is performed using the computational geometry algorithms and software routines discussed in O'Rourke (1994).

As with the integration in equation (9), computation of the summation in equation (A3) is facilitated by the restrictions implied by Q_{t+1}. The intersection $C_{t+1}^{j} \cap Q_{t+1}$ is empty for virtually all of cells in the z_t space, and it is easy to specify the small set of nonempty cells. Furthermore, if the purpose of the calculation is "off-line" estimation (rather than real-time forecasting), we can economize on the calculation of the $\Pr(C_{t+1}^{j}|q_t,q_{t-1},\ldots)$ in equation (A3) by computing only the values used in the subsequent probability calculation. These simplifications greatly reduce the computational burden.

In approximating the discrete transition probabilities $\Pr(C_{t+1}^{j}|C_t^{i})$, I generally take

$$\Pr(C_{t+1}^{j}|C_t^{i}) \approx \mathrm{Vol}(C_{t+1}^{j}) f(z_{t+1}^{j}|z_t^{i}), \tag{A4}$$

where z_{t+1}^{j} and z_t^{i} are the cell midpoints. This should be a good approximation when the density function is relatively constant over C_{t+1}^{j}. When the transition from C_t^{i} to C_{t+1}^{j} includes the point of no change in the efficient price, however, concerns over the sharp peak in the GED distribution lead me to use

$$\Pr(C_{t+1}^{j}|C_t^{i}) \approx \int_{z_{t+1}\in C_{t+1}^{j}} f(z_{t+1}|z_t^{i})\,dz_{t+1}, \tag{A5}$$

where the integration calculation uses the midpoint approximation (as in(A4)) for α_t and β_t, and Chebyshev integral approximation along the m axis.

At time t, the set of lattice cells (C_t^{i}) was computed as the cross product of three one-dimensional lattices, one for each of the state variables α, β, and m. The breakpoints for the α lattice were (in ticks): 0.0, 0.01, 0.2, 0.04, 0.07, 0.13, 0.24, 0.46, 0.88, 1.67, 3.16, and 6 (the maximum spread in the analysis). These breakpoints approximate fixed intervals in $\ln(\alpha)$. The lattice for m ranged from the lowest bid in the sample to the highest ask, in 0.2-tick increments.

The number of cells necessary to cover a given quote region $Q(a,b)$ depends only on the spread $a - b$. For spread sizes of one through six ticks, the corresponding cell counts are: 282, 352, 218, 218, 198, 162 (for the full model); 5, 10, 15, 25, 30 (for the restricted EGARCH-only model); and 74, 95, 42, 40, 40, 25 (for the restricted cost-only model).

Table B.I

Simulation Results

The table reports the summary statistics for 25 simulations and estimations of the model described in Appendix B with 1,000 observations per simulation.

	Parameter and True Value						
Model	μ −1.7	σ_ν 0.9	ϕ 0.4	η 1.0	φ 0.9	γ 0.3	ν 1.0
Full							
Mean estimate	−1.724	0.871	0.419	0.982	0.884	0.321	1.026
Std. dev. of estimates	0.062	0.046	0.053	0.110	0.044	0.072	0.066
Mean of estimated SEs	0.061	0.046	0.061	0.135	0.045	0.069	0.073
EGARCH/GED							
Mean estimate				0.964	0.886	0.337	0.987
Std. dev. of estimates				0.120	0.046	0.077	0.059
Mean of estimated SEs				0.142	0.044	0.073	0.075
Cost							
Mean estimate	−1.722	0.872	0.421				
Std. dev. of estimates	0.061	0.045	0.050				
Mean of estimated SEs	0.061	0.046	0.062				

Appendix B. Simulations

The estimation technique employed here involves numerical approximations and computations of moderate complexity. To assess the adequacy of these approximations and gauge performance of the software, I conduct a series of simulations. The simulated model consists of equation (2c) for the EGARCH/GED evolution of the efficient price and equation (11b) for the evolution of the quote exposure costs. The parameter values are $\eta = 1$, $\varphi = 0.9$, $\gamma = 0.3$, $\nu = 1$, $\mu = -1.7$, $\phi = 0.4$ and $\sigma_\nu = 0.9$, chosen to approximate the values estimated from the sample. The model is essentially identical to the one applied to the actual sample data, with the exception that I do not introduce deterministic (intraday) patterns. This is done to reduce the dimensionality of the optimizations, and should not materially affect the parameter estimates for the stochastic components of the model.

I generate 25 samples of 1,000 observations (vs. approximately 6,000 in the actual sample), and estimate parameters and standard errors (with the same approximation grids used over the actual sample). Table B.I summarizes the results. Overall, convergence to population values is good, and the estimated standard errors agree with the dispersion found in the sample estimates. As in the actual sample data, the results for the full model and two submodels are in agreement. On a 200 mHz Pentium Pro, the average convergence times for the likelihood optimization are 22:42 (minutes:seconds) for the full model, 00:40 for the EGARCH/GED model, and 00:41 for the cost model.

REFERENCES

Ahn, Hee J., Charles Q. Cao, and Hyuk Choe, 1996, Tick size, spread and volume, *Journal of Financial Intermediation* 5, 2–22.

Andersen, Torben G., and Tim Bollerslev, 1997a, Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high-frequency returns, *Journal of Finance* 52, 975–1005.

Andersen, Torben G., and Tim Bollerslev, 1997b, Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance* 4, 115–158.

Andersen, Torben G., and Tim Bollerslev, 1998a, DM-Dollar volatility: Intraday activity patterns, macroeconomic announcements and longer-run dependencies, *Journal of Finance* 53, 219–265.

Andersen, Torben G., and Tim Bollerslev, 1998b, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review* 39, 885–905.

Angel, James J., 1997, Tick size, share prices, and stock splits, *Journal of Finance* 52, 655–681.

Anshuman, V. R., and Avner Kalay, 1998, Market-making with discrete prices, *Review of Financial Studies* 11, 81–109.

Baillie, Richard T., and Tim Bollerslev, 1991, Intra-day and inter-market volatility in foreign exchange rates, *Review of Economic Studies* 58, 565–585.

Ball, Clifford A., 1988, Estimation bias introduced by discrete security prices, *Journal of Finance* 43, 841–865.

Bernhardt, Daniel, and Eric Hughson, 1996, Discrete pricing and the design of dealership markets, *Journal of Economic Theory* 71, 148–182.

Bollerslev, Tim, and Ian Domowitz, 1993, Trading patterns and prices in the interbank foreign exchange market, *Journal of Finance* 48, 1421–1443.

Bollerslev, Tim, Robert F. Engle, and Daniel B. Nelson, 1994, ARCH models, in Robert F. Engle and Daniel L. McFadden, eds.: *Handbook of Econometrics IV* (Elsevier, Amsterdam).

Bollerslev, Tim, and Michael Melvin, 1994, Bid-ask spreads and volatility in the foreign exchange market: An empirical analysis, *Journal of International Economics* 36, 355–372.

Brock, William, and Alan Kleidon, 1992, Periodic market closure and trading volume: A model of intraday bids and asks, *Journal of Economic Dynamics and Control* 16, 451–489.

Brown, Sharon, Paul Laux, and Barry Schachter, 1991, On the existence of an optimal tick size, *Review of Futures Markets* 10, 50–72.

Chan, K., K. C. Chan, and G. A. Karolyi, 1991, Intraday volatility in the stock index and stock index futures market, *Review of Financial Studies* 4, 657–684.

Chan, K. C., William G. Christie, and Paul H. Schultz, 1995, Market structure and the intraday pattern of bid-ask spreads for NASDAQ securities, *Journal of Business* 68, 35–60.

Cheung, Yin-Wong, and Lilian Ng, 1992, Stock price dynamics and firm size: An empirical investigation, *Journal of Finance* 47, 1985–1997.

Cho, David, and Edward W. Frees, 1988, Estimating the volatility of discrete stock prices, *Journal of Finance* 43, 451–466.

Chordia, Tarun, and Avanidhar Subrahmanyam, 1995, Market making, the tick size, and payment-for-order flow: Theory and evidence, *Journal of Business* 68, 543–575.

Christie, William G., Jeffrey H. Harris, and Paul H. Schultz, 1994, Why did NASDAQ market makers stop avoiding odd-eighth quotes?, *Journal of Finance* 49, 1841–160.

Christie, William G., and Paul H. Schultz, 1994, Why do NASDAQ market makers avoid odd-eighth quotes?, *Journal of Finance* 49, 1813–1840.

Cordella, Tito, and Thierry Foucault, 1999, Minimum price variations, time priority and quote dynamics, *Journal of Financial Intermediation* 8, 141–173.

Dravid, A. R., 1991, Effects of bid-ask spreads and price discreteness on stock returns, Working paper, Wharton School.

Duffee, Gregory R., 1995, Stock returns and volatility: A firm-level analysis, *Journal of Financial Economics* 37, 399–420.

Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69–90.

Engle, Robert F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987–1007.

Engle, Robert F., 1996, The econometrics of ultra-high-frequency data, Working paper, University of California at San Diego.

Engle, Robert F., Takatoshi Ito, and Wen L. Lin, 1990, Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market, *Econometrica* 58, 525–542.

Engle, Robert F., and Jeffrey E. Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* 66, 1127–1162.

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

Glosten, Lawrence R., and Lawrence E. Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* 21, 123–142.

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

Gottlieb, Gary, and Avner Kalay, 1985, Implications of the discreteness of observed stock prices, *Journal of Finance* 40, 135–153.

Hamao, Yasushi, Ronald W. Masulis, and Victor Ng, 1990, Correlations in price changes and volatility across international stock markets, *Review of Financial Studies* 3, 281–307.

Hamilton, James D., 1994a, State-space models; in Robert F. Engle and Daniel L. McFadden, eds.: *Handbook of Econometrics IV* (Elsevier, Amsterdam).

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

Harris, Lawrence E., 1986, A transaction data study of weekly and intradaily patterns in stock returns, *Journal of Financial Economics* 16, 99–117.

Harris, Lawrence E., 1990a, Estimation of stock price variances and serial covariances from discrete observations, *Journal of Financial and Quantitative Analysis* 25, 291–306.

Harris, Lawrence E., 1990b, Liquidity, trading rules and electronic trading systems, *Salomon Center Monograph Series in Finance and Economics* 4 (New York University, New York).

Harris, Lawrence E., 1991, Stock price clustering and discreteness, *Review of Financial Studies* 4, 389–415.

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

Harris, Lawrence E., 1997, Decimalization: A review of the arguments and evidence, Working paper, University of Southern California.

Harris, Lawrence E., and Joel Hasbrouck, 1996, Market vs. limit orders: The SuperDOT evidence on order submission strategy, *Journal of Financial and Quantitative Analysis* 31, 213–231.

Harvey, Andrew C., 1991, *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, Cambridge).

Hasbrouck, Joel, 1991, Measuring the information content of stock trades, *Journal of Finance* 46, 179–207.

Hasbrouck, Joel, 1996, Modeling microstructure time series; in G. S. Maddala and C. R. Rao, eds.: *Handbook of Statistics 14: Statistical Methods in Finance* (Elsevier North Holland, Amsterdam).

Hasbrouck, Joel, 1999, Security bid/ask dynamics with discreteness and clustering: Simple strategies for modeling and estimation, *Journal of Financial Markets* 2, 1–28.

Hasbrouck, Joel, and George Sofianos, 1993, The trades of market makers: An empirical examination of NYSE specialists, *Journal of Finance* 48, 1565–1593.

Hausman, Jerry A., Andrew W. Lo, and A. C. MacKinlay, 1992, An ordered probit analysis of transaction stock prices, *Journal of Financial Economics* 31, 319–379.

Kitagawa, Genshiro, 1987, Non-Gaussian state-space modeling of nonstationary time series, *Journal of the American Statistical Association* 82, 1032–1041.

Lee, Charles M., Brenda Mucklow, and Mark J. Ready, 1993, Spreads, depths and the impact of earnings information: An intraday analysis, *Review of Financial Studies* 6, 345–374.

Lyons, Richard K., 1995, Tests of microstructural hypotheses in the foreign exchange market, *Journal of Financial Economics* 39, 321–351.

Lyons, Richard K., 1996, Optimal transparency in a dealer market with an application to foreign exchange, *Journal of Financial Intermediation* 5, 225–254.

Madhavan, Ananth, Matthew Richardson, and Mark Roomans, 1997, Why do security prices change?, *Review of Financial Studies* 10, 1035–1064.

Madhavan, Ananth, and Seymour Smidt, 1993, An analysis of changes in specialist inventories and quotations, *Journal of Finance* 48, 1595–1628.

Manrique, Aurora, and Shephard, Neil, 1997, Likelihood analysis of a discrete bid/ask price model for a common stock , Working paper, Oxford University.

Marsh, Terry, and Eric Rosenfeld, 1986, Non-trading, market making and estimates of stock price volatility, *Journal of Financial Economics* 15, 359–372.

McInish, Thomas H., and Robert A. Wood, 1992, An analysis of intraday patterns in bid/ask spreads for NYSE stocks, *Journal of Finance* 47, 753–764.

Muller, Ulrich A., Michel M. Dacarogna, Richard B. Olsen, Olivier B. Pictet, Matthias Schwarz, and Claude Morgenegg, 1990, A statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis, *Journal of Banking and Finance* 14, 1189–1208.

Nelson, Daniel B., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.

Niederhoffer, Victor, 1965, Clustering of stock prices, *Operations Research* 13, 258–265.

Niederhoffer, Victor, 1966, A new look at the clustering of stock prices, *Journal of Business* 39, 309–313.

O'Rourke, Joseph, 1994, *Computational Geometry in C* (Cambridge University Press, Cambridge).

Rock, Kevin, 1998, The specialist's order book and price anomalies, *Review of Financial Studies*, forthcoming.

Wood, Robert A., Thomas H. McInish, and J. K. Ord, 1985, An investigation of transaction data for NYSE stocks, *Journal of Finance* 25, 723-739.

Zhou, Bin, 1996, High frequency data and volatility in foreign exchange rates, *Journal of Business and Economic Statistics* 14, 45–52.