NEW YORK UNIVERSITY
STERN SCHOOL OF BUSINESS
FINANCE DEPARTMENT

Working Paper Series, 1995

The Dynamics of Discrete Bid and Ask Quotes.

Hasbrouck, Joel.

FIN-95-23

The Dynamics of Discrete Bid and Ask Quotes

Joel Hasbrouck

February 16, 1996

Preliminary Draft
Not for Quotation
Comments Welcome

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

Tel: (212) 998-0310
Fax: (212) 995-4901
E-mail: jhasbrou@stern.nyu.edu
Web: http://www.stern.nyu.edu/~jhasbrou

For comments on an earlier draft, I am indebted to James Angel and seminar participants at Georgetown University. All errors are my own responsibility.

The Dynamics of Discrete Bid and Ask Quotes

Abstract

This analysis models discrete quotes as arising from two continuous random variables, the efficient price and a cost of quote exposure (information and processing costs). The former less the latter rounded down to the next tick yields the bid; the former plus the latter rounded up yields the ask. To deal with situations in which the cost of quote exposure possesses both stochastic and deterministic components, the paper proposes a nonlinear state-space estimation method. The method is applied to intraday quotes at fifteen-minute intervals for Alcoa (a randomly chosen Dow stock). The results confirm the existence of deterministic and stochastic components of the cost that are of roughly comparable magnitudes.

1. Introduction

Although most determinants of a security price are likely to be continuous variables, the security price is usually constrained by institutional arrangement to a discrete grid. Since this grid may be coarse relative to the price variation over brief intervals, discreteness becomes more important as the observation interval shrinks. This paper suggests that the bid and ask quotes arise from an implicit efficient price and a quote-exposure cost variable, both of which are continuous random variables. The efficient price less the cost is rounded down to produce the bid; the price plus the cost is rounded up for the ask. The paper proposes a nonlinear state-space procedure for estimation and (in real time applications) online filtering, and applies this model to fifteen-minute bid and ask quotes for a New York Stock Exchange stock.

This paper is most closely related to earlier empirical studies of discreteness in stock prices. The first analyses in this area focused on estimation of long-term stock return variances from transaction prices, a concern motivated by option pricing applications (see Gottlieb and Kalay (1985), Cho and Frees (1988) and Ball(1988)). The emphasis in later studies of transaction prices shifted to microstructure phenomena (see Glosten and Harris (1988), Harris (1990, 1991, 1994), Hausman, Lo and MacKinlay (1992), Angel (1994) and Madhavan, Richardson and Roomans (1994)). Discreteness is often encountered as a "nuisance" effect, a data characteristic that must be addressed en route to confronting more interesting economic hypotheses. But since the minimum tick size may affect trading activity, discreteness is also of theoretical and policy interest (Harris (1990, 1991), Bernhardt and Hughson (1990, 1992), Brown, Laux and Schacter (1991), Glosten (1994), Chordia and Subrahmanyam (1995), Ahn, Cao and Choe (1996)).

The present study seeks to model bid and ask quotes as opposed to transaction prices. Quotes are of particular interest in microstructure studies because they can be updated in the absence of trades to reflect changing information and also because they reflect perceived asymmetric information costs. Discreteness has different effects on quotes and transaction prices. A risk-neutral trader would presumably be indifferent to the fair-game perturbation associated with symmetrically rounding the unobserved continuous price to the nearest tick. A market maker posting bid and offer quotes, on the other hand, must round his bid price down and his offer price up in order avoid the expectation of losing money on the next trade.

The paper is also closely related to studies of discreteness in the bid-ask spread. Harris (1994) and Bollerslev and Melvin (1994) model the discrete spread using ordered qualitative-data approaches. In these studies, the spread is a continuous function of observable variables and a random disturbance that is transformed onto a discrete grid. The spread in the present model, in contrast, is driven by an underlying variable that is unobservable, stochastic and autocorrelated. Furthermore, in modeling the bid and ask separately, the present analysis incorporates a rich specification of the efficient price dynamics.

The paper represents the model in state-space form in which the unobserved efficient price and quote exposure cost are the state variables and the bid and ask prices are the observables. This framework is appealing because the recursive procedure used to compute the likelihood function is a Bayesian updating that mimics agents' inferences. Furthermore, state-space models are natural and convenient tools with which to investigate deterministic and stochastic time variation in parameters. The paper implements several such generalizations.

In comparison with reduced-from vector autoregressive (VAR) microstructure models (e.g., Hasbrouck (1991, 1993), the present design assumes more structure in the form of the probability densities and the discrete-valued functions that map the continuous state variables onto prices. These assumptions suffice to identify a (nonlinear) state-space model. Although tightly structured in its discreteness aspects, this model remains general in other regards, notably those related to time variation in the parameters. The conventional approach to characterizing intraday parameter variation involves estimating fixed-parameter models over intraday subsamples (e.g., as in Hasbrouck (1991), the first hour of trading). In contrast, the present approach admits stochastic and deterministic parameter variation in a comprehensive statistical model.

Structural restrictions usually arise from specific economic models. The analyses of Madhavan and Smidt (1991, 1993), Madhavan, Richardson and Roomans (1994), Easley, Kiefer and O'Hara (1994), for example, estimate specifications derived from detailed characterizations of agents' behavior and/or specifications about the informational features of the market. These assumptions serve to identify the "deep" economic structural parameters. In this respect, the present aims are much more modest.

Thus, although the paper deals primarily with discreteness, the ultimate aim in this line of inquiry is a modeling framework flexible enough to accommodate parameter shifts resulting from the start and finish of trading and random variation in the underlying informational and liquidity determinants of trading activity.

The analysis does not extend to clustering (the affinity of transaction prices and quotes for integers, halves, quarters, etc., in decreasing frequency). Clustering in transaction prices is examined by Niederhoffer (1965, 1966) and Harris (1991), and in quotes by Christie and Shultz (1995a, 1995b). In a dynamic setting, clustering requires specification of a mapping from continuous state variables to discrete observations that is more complicated than the simple rounding functions employed here.

The paper is organized as follows. The next section describes the underlying economic model that generates the bid and ask quotes. The paper then turns to the problem of inference: how to estimate the underlying model from the observed discrete bid and ask prices. Section 3 discusses inference when the successive quote exposure costs are serially independent. Section 4 presents a richer model that incorporates stochastic and deterministic time variation in the cost and efficient price volatility. The details of the estimation procedure are presented in an appendix. The model is estimated for a representative NYSE stock in section 5. Section 6 discusses extensions. A brief summary concludes the paper in section 7.

2. The economic model

Denote by m the implicit efficient price of the security. The agent establishing the quotes is assumed to be subject to a positive cost of quote exposure $c>0$, such that in the absence of discreteness restrictions she would quote a bid price of $m-c$ and an ask (offer) price of $m+c$. This cost is assumed to impound fixed transaction costs and asymmetric information costs. Assuming a one-unit tick size in the market, the actual bid and ask quotes are:

$$\begin{aligned} b &= \text{Floor}[m-c] \\ a &= \text{Ceiling}[m+c] \end{aligned} \tag{1}$$

where Floor[·] rounds its argument down to the next whole integer and Ceiling[·] rounds its argument up to the next whole integer. If the tick size is not unity, all variables may simply be rescaled.

This construct can be motivated by most simple models of dealer behavior. In the framework of Glosten and Milgrom (1985) quote setters face a population of informed and uninformed traders. m is the expectation of the final value of the security conditional on all public information (including the transaction price history). The quote exposure cost c is defined implicitly by the condition that $m-c$ and $m+c$ are the quotes that ensure the quote-setter zero expected profits and no ex post regret. This outcome arises from Bertrand competition. By rounding up on the ask and down on the bid, the market maker ensures a positive expected profit on each trade.

The rounding in this model is asymmetric: the bid is rounded down, while the ask is rounded up. If the rounding were 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. For example, if the efficient price is 5 and the cost is 1.1, nearest-integer rounding yields a bid of 4 and an ask of 6, both of which yield expected losses. Furthermore, symmetric rounding may in some instances lead to identical bid and ask prices (if c is small).

Although this model allows for randomness in both c and m, the discreteness aspect of the model arises from a nonstochastic transformation. There is no discreteness "error" or disturbance that is required to impound the effect of discreteness.

As noted in the introduction, most studies of discreteness in security markets have focused on transaction prices. Quotes and transaction prices are obviously related, however, and the transaction price models therefore offer useful points of comparison. In this connection, there is at the outset one obvious incompatibility. If transactions arise as uncorrelated equiprobable realizations of the bid and ask quotes determined by equation (1), these prices cannot be described as a symmetrically-rounded random-walk.

The models actually estimated in the paper will allow this cost to exhibit both deterministic and stochastic dynamic behavior. As a preliminary, however, it is useful to point out that even when c is constant, random variation in m suffices to induce randomness in the spread. For example, if c=1/4, then the spread is one tick as long as the fractional part of m is between 1/4 and 3/4; and the spread is two ticks otherwise. Therefore, variability in the discrete spread may be an erroneous proxy for variability in the spread's continuous determinants. In addition, price transitions will sometimes be marked by quotes that appear to move "one leg at a time". (Consider the quotes associated with the m_t sequence {0.4, 0.9, 1.3}.) U.S. stock quotes often exhibit this behavior.

Finally, this model does not explicitly capture any particular agent's optimization problem. The framework is implicitly one in which the agent's solution to a continuous optimization problem (c) is then subjected to a transformation to yield discrete strategies. Fundamentally, however, a discrete price and quantity grid gives rise to an integer programming problem. There is no obvious necessary reason to assume that the optimal strategies of all potential quote setters can be characterized as rounded continuous solutions. A trader contemplating the submission of a limit order, for example, must balance costs (including those associated with execution failure), value and execution probability (see Harris (1994)).

3. Inference from observed bid and ask quotes.

Viewed as a transformation of continuous random inputs (m and c) into discrete bid and ask prices, the model described by (1) is a very simple one. From the perspective of the econometrician (and that of many market participants), however, the observed bid and ask prices are given, and inference focuses on the unobserved inputs. Viewed in this direction, the model is more complex.

Taking the bid and ask quotes as given, the feasible region for (m, c) consistent with model (1) is the region

$$Q(b,a) = \{(m,c): c > 0, b \le m - c < b + 1 \text{ and } a - 1 < m + c \le a\} \tag{2}$$

Figure 1 depicts $Q(b=0, a=1)$ (a one-tick spread), $Q(b=0, a=2)$ (a two-tick spread), and $Q(b=0, a=3)$ (a three-tick spread). The diamond shape of the region $Q(b=0, a=2)$, for example, can be viewed as 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 $a=1$ is a small neighborhood about one. When c is 1/2, m can range from 1/2 to 3/2.

Most interesting applications will involve situations where the quote exposure cost is random. This randomness can be viewed as arising from several sources. Along the lines of the Glosten-Milgrom model, there may be random time variation in the determinants of this cost, such as perceived exposure to adverse information or holding costs. In this view all dealers and potential dealers are subject to the same cost. 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 cost is the lowest c in the group.

The only firm requirement on the latent cost variable is nonnegativity. A simple density function that satisfies this requirement is the lognormal: Ln(c) is assumed to be distributed normally with mean μ and standard deviation σ.

In the distributional assumptions on the latent efficient price, the model admits a wide generality. The full model presented in the next section incorporates an EGARCH process for m. Initially, however, it is useful to characterize m by a uniform diffuse prior, i.e., a probability density that is constant over some suitably large region. Formally it suffices to take

$$f(m) = \begin{cases} \kappa^{-1} \text{ for } m \in (0,\kappa) \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

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 $f(m, c)=f(m)f(c)$.

The posterior densities

I now turn to the construction of the density of (m, c) conditional on observing that the bid and ask quotes are b and a. The joint conditional density is:

$$f(m,c|b,a)=\begin{cases}\dfrac{f(m,c)}{\Pr(b,a)} & \text{if } (m,c)\in Q(b,a)\\ 0, & \text{otherwise}\end{cases} \tag{4}$$

where $\Pr(b, a)$ is the probability of observing discrete bid and ask quotes b and a:

$$\Pr(b,a)=\int_{(m,c)\in Q(b,a)} f(m,c)\,dm\,dc \tag{5}$$

(The m density parameter κ drops out of $f(m, c|b, a)$ in equation (4) because it appears as a factor in both the numerator and denominator of the fraction on the right-hand-side.)

Since this conditioning imposes a truncation on the ranges of the variables, it might seem that the conditional densities would be simple truncated versions of the priors. The truncations defined by $Q(a,b)$, however, apply to linear combinations of the variables, not the variables themselves. The shape of $Q(a,b)$ effectively forces a nonlinear transformation on the priors.

As an illustration, consider the marginal densities $f(c|b, a)$ and $f(m|b, a)$ assuming that that the parameters of the lognormal density for the quote exposure cost c are $\mu=-1$ and $\sigma = 0.6$. This implies (from equation (5)) that $\Pr[a-b=1] = 0.29$, $\Pr[a-b=2] = 0.58$, $\Pr[a-b=3] = 0.11$, and $\Pr[a-b>3] = 0.03$. Figure 2 depicts the prior and conditional density functions conditional on 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 the allowable range of m $(1/2<m<3/2)$. If m is near an endpoint, the range of feasible c values is a small one, 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. (The peak in the conditional cost density arises from the "corners" in the diamond $Q(a=0, b=2)$ in Figure 1.)

Estimation

Suppose now that the quote generation process occurs over a sequence of time periods $t=1, \ldots, T$ with a realization $\{m_t, c_t\}$. It will be assumed initially that the $\{c_t\}$ are identically and independently distributed, and that they are also distributed independently of the m_t. For simplicity, the assumption of a diffuse uniform prior for all m_t is maintained.

If c_t is observed directly, the parameters of the lognormal cost density function, μ and σ, can be estimated easily. In the usual fashion, the log likelihood function for a single observation is given by $l(c_t;\theta) = Log[f(c_t;\theta)]$, where $\theta=[\mu\ \sigma]'$ is the parameter vector and f is the lognormal density function. In the usual fashion, a maximum likelihood estimate (MLE) can be constructed by maximizing the sample log likelihood $\sum_{t=1}^{T} l(c_t;\theta)$. This estimate is asymptotically normally distributed with mean θ and covariance matrix $[T\ I(\theta)]^{-1}$, where $I(\theta)$ is the information matrix $I(\theta) = E[(\partial l/\partial\theta)(\partial l/\partial\theta')]$. Usually, however, the observations will consist of the series of discrete bid and ask prices $\{b_t, a_t\}$. In this case, the log likelihood is $l(b_t, a_t;\theta) = Log[\Pr(b_t, a_t;\theta)]$. The sample log likelihood function and information matrices may computed using this functional form.

From an econometric viewpoint, discreteness in this model causes problems similar to those generally stemming from grouped data. Grouping is tantamount to a loss of information, and it therefore imposes an efficiency penalty. This penalty may be assessed in the present application by comparing the information matrices of the MLE based on the log likelihood with continuous observations $l(c_t;\theta)$ and that of the MLE based on $l(b_t,a_t;\theta)$, the discrete-data counterpart.

With the same numerical parameter values used to illustrate the conditional density calculation ($\mu=-1.0$ and $\sigma=0.6$), the inverse information matrix for the continuous MLE is $\begin{bmatrix} 0.36 & 0. \\ 0. & 0.18 \end{bmatrix}$. In a sample of size n=100, for example, the standard error of estimate of μ would be approximately $(0.36/100)^{1/2} = 0.06$ (which could also be obtained as $\sigma/\sqrt{n}$). The absence of correlation between the parameter estimates is a consequence of normality.

The inverse information matrix for the discrete MLE is $\begin{bmatrix} 0.88 & -0.37 \\ -0.37 & 0.62 \end{bmatrix}$. The loss of information from discreteness is reflected in the increased magnitudes of all entries. To obtain the same standard error of estimate for μ that was achieved with 100 continuous observations, for example, roughly 244 discrete observations would be required (implied by $0.06=(0.88/n)^{1/2}$). Furthermore, the discrete estimates of μ and σ are negatively correlated.

4. The full dynamic model

Overview

The analysis to this point has suppressed dynamic considerations in the assumption of a diffuse prior to the efficient price m_t and the i.i.d. assumption on the quote exposure cost c_t. It is useful to model the efficient price for two reasons. First, this is the quantity that will be of primary interest in many applications. Second, it is apparent from the shape of the feasible regions (the Q in Figure 1) that knowledge about the location of m_t is useful in determining the location of c_t: by incorporating the dynamics of m_t we expect to obtain sharper estimates of the parameters of the c_t process. As for the dynamics of c_t, realistic applications would seem to require allowance for deterministic time effects (such as the intraday "U" patterns) and also for autocorrelated stochastic effects.

The general model is:

$$m_t = m_{t-1} + u_t \tag{6}$$

$$\ln(c_t) = \mu_t^c + \phi^c\left(\ln(c_{t-1}) - \mu_{t-1}^c\right) + v_t \tag{7}$$

The increments to the efficient price u_t are assumed to follow an EGARCH (exponential generalized autoregressive conditional heteroskedastic) process based on the GED (generalized error density) function. From earlier studies, it appears desirable to allow for

"U" shapes in both the efficient price volatilities and the cost means. I now turn to a discussion of the EGARCH and deterministic components of the model.

The EGARCH component

The heteroskedasticity modeling in this analysis follows Nelson (1991). To allow for leptokurtosis and time-varying volatility in the efficient price increments, I assume that after standardizing by its standard deviation, the increment is drawn from a standard generalized error distribution (GED) with parameter ν: $z_t \equiv u_t/\sigma_t \overset{d}{\sim} GED(\nu)$. This distribution is given by:

$$f_{GED}(z;\nu) = \frac{\nu \exp\left[-\left(\frac{1}{2}\right)\left|\frac{z}{\lambda}\right|^{\nu}\right]}{\lambda 2^{(1+1/\nu)}\Gamma(1/\nu)}, \text{ where } \lambda \equiv \left[2^{(-2/\nu)}\Gamma(1/\nu)/\Gamma(3/\nu)\right]^{1/2} \tag{8}$$

In the case where ν=2, this reduces to the standard normal density.

A standard EGARCH specification models time-varying variances as:

$$\ln(\sigma_t^2) = \mu_t^{\sigma} + \phi^{\sigma}\left(\ln(\sigma_{t-1}^2) - \mu_{t-1}^{\sigma}\right) + \gamma\left(|z_{t-1}| - \mathrm{E}|z_{t-1}|\right) \tag{9}$$

The leading term on the right-hand-side, μ_t^{σ}, is a deterministic function that impounds the time-of-day behavior in variances. The second term is a first-order autoregressive component. The last component is a disturbance component driven by the prior period's shock. The expected absolute value is unconditional and time-invariant, depending only on the tail-thickness parameter. It is given by $E|z| = \lambda 2^{(1/\nu)}\Gamma(2/\nu)/\Gamma(1/\nu)$. (The asymmetry term suggested by Nelson is omitted.)

In the present application, a problem arises from the fact that the m_t (and therefore u_t and $z_t=u_t/\sigma_t$) are not observable. Since knowledge of the bid and offer quote history is insufficient to compute equation (9), σ_t must be carried as an unobservable state variable. This is not computationally feasible. As a more tractable alternative, I assume that the variance process is driven by the conditional expectation of the absolute efficient price increment. That is, $|z_{t-1}|$ in (9) is replaced by its conditional expectation $\mathrm{E}_{t-1}[|z_{t-1}|] = \mathrm{E}_{t-1}[|z_{t-1}|]/\sigma_{t-1}$, yielding

$$\ln(\sigma_t^2) = \mu_t^{\sigma} + \phi^{\sigma}\left(\ln(\sigma_{t-1}^2) - \mu_{t-1}^{\sigma}\right) + \gamma\left(\mathrm{E}_{t-1}[|u_{t-1}|]/\sigma_{t-1} - \mathrm{E}|z_{t-1}|\right) \tag{10}$$

where $\mathrm{E}_{t-1}|u_{t-1}| = \mathrm{E}\left[|u_{t-1}| \,|\, q_{t-1}, q_{t-2}, \ldots\right]$. This quantity is easily computed in the course of the iterative update.

Deterministic time variation

Both the quote exposure cost function in (6) and the variance specification in (10) allow for deterministic effects. At a bare minimum it appears necessary to allow for the "U" shapes that characterize market data. A parsimonious function that permits end-point elevation can be built from exponential decay functions. The deterministic component of the cost process is:

$$\mu_t^c = \mu_0^c + \alpha^{c,open} \exp\left(-\lambda^{c,open} \tau_t^{open}\right) + \alpha^{c,close} \exp\left(-\lambda^{c,close} \tau_t^{close}\right) \tag{11}$$

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 deterministic component of the variance is:

$$\mu_t^\sigma = \begin{cases} \mu_0^\sigma + \alpha^{\sigma,open} \exp\left(-\lambda^{\sigma,open} \tau_t^{open}\right) + \alpha^{\sigma,close} \exp\left(-\lambda^{\sigma,close} \tau_t^{close}\right), \\ \quad \text{if } t \text{ is not the daily close} \\ \mu_{overnight}^\sigma, \ \text{if } t \text{ is the daily close} \end{cases} \tag{12}$$

Estimation

For purposes of estimation, the system may be viewed as a nonlinear state-space model. The state variables are m_t and c_t, with associated transition equations (6) and (7). The observations consist of the bid and ask quotes, with observation equation (1).

The estimation procedure follows Kitagawa's (1987) approach. The appendix contains a summary of this method as it applies to the present situation. Briefly, the approach involves representing the conditional densities of the state variables as numerical grids. Updating these densities at each time period requires numerical integration over the state variables. Relative to the "usual" nonlinear filtering problem, these integrations are simplified somewhat because the integration regions are the diamond and half-diamond shapes in Figure 1, rather than the entire two-dimensional plane. Nevertheless, the process remains computationally intensive.

Alternative specifications

The full model described above is a joint description of the quote exposure cost and the efficient price. Given its complexity and computational burdens, one might reasonably ask if univariate models might yield similar results. That is, if only the cost dynamics are of interest, is it sufficient to estimate a simpler model that assumes a diffuse prior for the efficient price (cf. the development in section 3)? This variant, consisting of the cost equation from (7) and equation(11), 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. The resulting estimates can be viewed as being derived from the history of the spread, $a_t - b_t$.

Alternatively, if only the efficient price is of concern, one might estimate the first equation in (6), together with the EGARCH specification described in equations (10) and (12), under the assumption of a diffuse prior for the quote exposure cost over the positive real line. This variant is termed the "discrete EGARCH model".

5. Estimation

Data

I estimate the specifications described in the last section to NYSE bid and ask quotes for Alcoa (ticker symbol AA) for all trading days in 1994. Alcoa is the first Dow Stock (in alphabetical ordering) 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 points). There are 6527 observations.[1]

Estimates of the full model

Table 1 reports parameter estimates. For purposes of exposition, these may be grouped as cost- and variance- related. The deterministic cost parameters depict the usual

[1] With 28 integration points used in the half-diamond region and 56 in the diamond region, convergence required about two days for the full model, and six hours for the cost- and discrete-EGARCH models. Computations were programmed in GAUSS and executed on a 100 MHz Pentium system.

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^c are strongly positive. The autoregressive parameter suggests that about 30% of the excess log cost persists at the subsequent time point (fifteen minutes later).

Figure 4 depicts the relative importance of the deterministic and stochastic sources of variation in the quote exposure cost. The solid line graphs the intraday variation in the deterministic component implied by the model estimates. (For example, the constant parameter in the log quote exposure cost is $\mu_0^c = -1.499$, which implies a cost of $e^{-1.499} = 0.2234$ *ticks*, i.e., 0.2238/8=\$0.028.)

The stochastic portion of the log quote exposure cost is a first-order autoregressive process, which possesses a standard deviation of $\left(\sigma_v^2 / \left(1-\left(\phi^c\right)^2\right)\right)$ = $\left(0.705^2/\left(1-0.300^2\right)\right)^{1/2} = 0.739$. The dotted lines in Figure 4 define the plus-or-minus two-standard deviation range on the cost. This range may be interpreted as a "confidence interval" containing the cost at a given time on a randomly selected day. (This calculation uses the model point estimates only, and so does not reflect estimation errors. This range is not symmetric about the deterministic component due to the log transform.)

Consideration of the deterministic component (solid line in Figure 4) suggests a quote exposure cost that is typically about six cents at the open and three cents thereafter. (The elevation at the close is small). When the stochastic component is added in (as indicated by the dashed lines), the range appears to be large relative to the deterministic intraday variation.

Turning to the variance component of the model, the deterministic component suggests end-point elevation and a sharply higher overnight variance. The autoregressive parameter ϕ^σ also suggests strong persistence of the variance. The tail-thickness parameter estimate of $\nu = 0.855$ implies a sharply peaked, but otherwise fairly broad distribution. For comparison purposes, Figure 3 graphs the GED density associated with this value against the standard normal.

Estimates of the cost model.

The cost model is a computationally simpler subcase of the full model that follows from an assumption of a diffuse prior on the efficient price. The estimates are given in the "cost" model column of Table 1. Not only are the estimates virtually identical to those obtained for the full model, but so are the estimated standard errors. Although one might have hoped that by using the price level information the full specification would result in more precise estimates, this does not appear to be the case.

Several considerations could account for this failure. One possibility is simply general model misspecification. But it is also possible that even in a correctly specified model, the information about c contributed by m is small. There is no strong economic presumption supporting correlation between the two variables. Accordingly when we are in a full diamond region of Figure 1, knowledge of m is more informative about the dispersion of c than the expected value.

Estimates of the discrete-EGARCH model.

This model follows from an assumption of a uniform diffuse prior for the quote exposure cost. The estimates are reported in the "EGARCH" column of Table 1. The point estimates of the parameters are quite similar to those found for the full model. The standard errors are slightly worse, suggesting that the full model possesses an efficiency advantage.

The usefulness of c for determining the dynamics of m (but not the reverse) can be explained by two factors. In the first place, the lognormal density assumed for c implies a steeper gradient (at the tails) than the fat-tailed density of m. Secondly (and again by reference to Figure 1) knowledge of c is informative about the dispersion of m. But in this case, it is a quantity closely related to this dispersion (the expected absolute value) that drives the EGARCH process.

6. Extension to different quote exposure costs for the bid and ask.

The model described in equation (1) and used throughout this paper is one in which the same quote exposure cost applies to the bid and ask side of the market. There are practical reasons for questioning this symmetry. In many markets (including the

NYSE), the bid and ask quotes may be placed by different agents, who are subject to different quote exposure costs. It would seem to be a simple matter to generalize (1) to allow for a bid cost β>0, an ask cost α>0, and rounding rules:

$$\begin{aligned} b &= \text{Floor}[m-\beta] \\ a &= \text{Ceiling}[m+\alpha] \end{aligned} \tag{13}$$

Given a realization of discrete bid and ask quotes, the feasible region for (m, α, β) consistent with model (13) is the polytrope:

$$Q(b,a) = \{(m,\alpha,\beta): \alpha, \beta > 0, b < m-\beta < b+1 \text{ and } a-1 < m+\alpha < a\} \tag{14}$$

Figure 5 presents the region $Q(0,1)$ (a one-tick spread), along with several rotated perspectives. The corresponding region in the symmetric cost case is the half-diamond in Figure 1, a subset of the region in Figure 5 that lies on the β=α plane (not depicted). Relative to the symmetric cost case, this model allows the efficient price m to lie anywhere between the observed discrete bid and ask quote.

Although simple in principle, this modification presents practical problems. Consider first the situation corresponding to the case discussed in section 3. Assume that the α and β variates are lognormally distributed with common mean μ, common standard deviation σ and correlation coefficient ρ, and that these variates are not serially correlated. Under the assumption of a diffuse prior for m, we may repeat the information matrix calculations in the earlier section to assess sample size requirements in a representative situation. When this calculation is attempted however (with numeric values similar to those used earlier), the information matrix for the discrete case is nearly singular (and therefore noninvertible).

Technically, given the probability structure of the problem, this matrix is not singular, but this calculation (and any likely application) is limited by the precision of the numerical integrations. As a practical matter, it seems that assumption of a diffuse prior on m renders identification of the three parameters of the cost density impossible.

Of course, there remains the possibility that stronger identification may exist when the model incorporates a dynamic structure for m. In Figure 5, for example, if m is more likely to lie near 1 (the ask price), it is more likely that the ask cost variable α is low. In the symmetric cost framework, incorporation of a model for m has virtually no impact on the cost estimates or their precision, however. A final difficulty arises from the dimensionality of the state space. There are now three state variables, and the shape of the integration region is more complicated than in the symmetric cost case. The numerical difficulties are likely to be substantial.

Further generality would allow for negative quote-exposure cost. A limit-order buyer, for example, might well decide to bid above the efficient price of a security (if his alternative were lifting a very high ask quote). Once negative costs are allowed, however, additional structure must be added to the model to preclude the crossing of bids and asks (e.g. suppose $m = 10$, $\beta = 0.2$ and $\alpha = -0.1$). Moving further in this direction suggests generalizing the quote exposure cost to a trading propensity variable that can range over positive and negative values. Transactions occur when the discretized trading propensities of a buyer and seller cross, which implies that the prevailing bid and ask quotes are generated by a complex censoring process.

7. Conclusion

This paper has presented a dynamic model of discrete bid and ask quotes. The discrete quotes are rounded transformations of a continuous efficient price and a continuous quote exposure cost. The full statistical model is a rich one, allowing for stochastic and deterministic time variation in the efficient price volatility and the quote exposure cost. The model may be estimated by maximum likelihood using a nonlinear filtering approach due to Kitagawa (1987).

This specification is estimated for NYSE bid and ask quotes collected at the end of 15-minute intervals for Alcoa over 1994. The estimates confirm the existence of deterministic "U" shapes in the quote cost and efficient price volatility. More importantly, however, the estimates confirm the existence of a persistent stochastic component of the quote exposure cost. The magnitude of this component is roughly comparable to the variation associated with the "U" shapes.

In extending the model to incorporate other aspects of the market process, there are several guidelines. It is relatively easy to incorporate deterministic effects and observed exogenous variables into either the cost or efficient price specifications. Such developments usually require additional parameters in the likelihood function, which does not significantly affect the time required for the numerical calculation of this function (although it will probably increase the number of iterations required for convergence).

It is more difficult to add endogenous variables, such as quote sizes (number of shares at the bid and ask) or trades that are determined in part by prevailing quotes. These developments require an expansion of the set of state variables and a large accompanying increase in the computational burden. One might also want to specify a model for the quote exposure cost that is more complicated than the first-order autoregressive process employed here, by including additional autoregressive or moving average terms. These modifications also require additional state variables.

Expansion of the state variable set runs into the "curse of dimensionality" because of the requirement that the integration of the conditional probabilities be computed numerically over all variables. The remains the possibility, however, of approximating these densities (or a subset of them) by simpler functions for which closed-form integrals exist. Work is in progress along these lines.

Appendix:

Maximum Likelihood Estimation via Nonlinear Filtering

This appendix describes the maximum likelihood procedure used to estimate the model described in Section 4. The analysis is a special case of the general state-space approach discussed in Hamilton (1994a, 1994b) and Harvey (1991). The present model is neither linear nor Gaussian. The estimation method used here is due to Kitagawa (1987), which is summarized in Hamilton (1994b).

The essential construct is the probability density function of the current state variables conditional on current and past observations. The crux of the calculation is the updating of these beliefs to incorporate the next observation. The details of this updating process are as follows. Suppose that we possess the density function for the current state variables (m_t and c_t) conditional on the current and previous observations (quotes q_t), $f(m_t, c_t | q_t, q_{t-1}, \ldots)$. Looking ahead to time t+1 (prior to any new observation),

$$f(m_{t+1}, c_{t+1} | q_t, q_{t-1}, \ldots) = \int_{(m_t, c_t) \in Q_t} f(m_{t+1}, c_{t+1} | m_t, c_t) f(m_t, c_t | q_t, q_{t-1}, \ldots) dm_t dc_t \qquad \text{(A.1)}$$

where $Q_t \equiv Q(b_t, a_t)$. The transition density $f(m_{t+1}, c_{t+1} | m_t, c_t)$ may be computed from (6) and (7).

The range of integration in (A.1) is a distinctive feature of the present problem. In "typical" filtering applications, the integration range covers the entire real line (or multidimensional analog): if the observation history suggests there is a trivial probability of finding the state variable in a given neighborhood, the value of the conditional density is close to zero. In the present application, however, the quotes serve to bound the possible values of the state variables. Rather than integrate over the real half-plane defined by $m \in (-\infty, +\infty)$, $c \in (0, +\infty)$, the integration is over *one* of the diamond or half-diamond regions illustrated in Figure 1. Since the integral will be computed numerically, this is a welcome simplification.

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

$$\Pr(q_{t+1} | q_t, q_{t-1}, \ldots) = \int_{(m_{t+1}, c_{t+1}) \in Q_{t+1}} f(m_{t+1}, c_{t+1} | q_t, q_{t-1}, \ldots) dm_{t+1} dc_{t+1} \qquad \text{(A.2)}$$

In an online application we would be interested in computing this probability for a number of possible realizations of q_{t+1}. In an estimation situation, however, we need only compute the probability for the value of q_{t+1} that actually occurred in the sample.

Next, note that the joint density of next period's state variables and quotes is:

$$f\left(m_{t+1}, c_{t+1}, q_{t+1} \middle| q_t, q_{t-1}, \ldots\right) = \begin{cases} f\left(m_{t+1}, c_{t+1} \middle| q_t, q_{t-1}, \ldots\right), \text{ if } \left(m_{t+1}, c_{t+1}\right) \in Q_{t+1} \\ 0, \text{ otherwise} \end{cases} \quad \text{(A.3)}$$

This too reflects a simplification peculiar to the present problem: computation of the right-hand-side density usually involves integration over a density function of the observational errors. Here, the observations (the quotes) are a deterministic function of the state variables. Therefore

$$f\left(m_{t+1}, c_{t+1} \middle| q_{t+1}, q_t, q_{t-1}, \ldots\right) = \begin{cases} \dfrac{f\left(m_{t+1}, c_{t+1}, q_{t+1} \middle| q_t, q_{t-1}, \ldots\right)}{\Pr\left(q_{t+1} \middle| q_t, q_{t-1}, \ldots\right)}, \text{ if } \left(m_{t+1}, c_{t+1}\right) \in Q_{t+1} \\ 0, \text{ otherwise} \end{cases} \quad \text{(A.4)}$$

This completes the update.

Although simple 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 formulae for the conditional means and variances. In the present case, successive updates would involve computation of nested, truncated densities of increasing dimension. This is not computationally feasible.

The present analysis follows Kitagawa (1987) in approximating the conditional density $f\left(m_t, c_t \middle| q_t, q_{t-1}, \ldots\right)$ by a numerical grid. The integrations in (A.1) and (A.2) are computed using Gaussian quadrature. The log likelihood function is $\sum \ln\left(\Pr\left(q_t \middle| q_{t-1}, q_{t-2}, \ldots\right)\right)$.

References

Ahn, H.-J, C. Q. Cao, and H. Choe, 1996, Tick size, spread and volume, Journal of Financial Intermediation, 5.

Angel, J., 1994, Tick size, share prices and stock splits, working paper, Georgetown University.

Ball, C. A., 1988, Estimation bias introduced by discrete security prices, Journal of Finance, 43, 841-865.

Bollerslev, T., and M. Melvin, 1994, Bid-ask spreads in the foreign exchange market: an empirical analysis, Journal of International Economics.

Bernhardt, D. and E. Hughson, 1990, Discrete pricing and dealer competition, Working Paper, California Institute of Technology.

Bernhardt, D. and E. Hughson, 1992, Discrete pricing and institutional design of dealership markets, Working Paper, California Institute of Technology.

Bollerslev, T., R. F. Engle and D. B. Nelson, 1994, ARCH models, in R. F. Engle and D. L. McFadden, eds., Handbook of Econometrics, Volume IV, (Elsevier).

Brown, S., P. Laux and B. Schacter, 1991, On the existence of an optimal tick size, Review of Futures Markets, 10, 50-72.

Cho, D. C. and E. W. Frees, 1988, Estimating the volatility of discrete stock prices, Journal of Finance, 43, 451-466.

Chordia, T. and A. Subrahmanyam, 1995, Market making, the tick size, and payment-for-order flow: theory and evidence, Journal of Business, 68, 543-575.

Christie, W. G. and P. H. Schultz, 1994a, Why do NASDAQ market makers avoid odd-eighth quotes?, Journal of Finance, 49, 1813-40.

Christie, W. G. and P. H. Schultz, 1994b, Why did NASDAQ market makers stop avoiding odd-eighth quotes?, Journal of Finance, 49, 1841-60.

Glosten, L. R. and L. E. Harris, 1988, Estimating the components of the bid-ask spread, Journal of Financial Economics, 21, 123-42.

Easley, D., N. Kiefer, and M. O'Hara, 1994, A day in the life of a very common stock, working paper, Cornell University Graduate School of Management.

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

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

Gottlieb, G. and A. Kalay, 1985, Implications of the discreteness of observed stock prices, Journal of Finance, 40, 135-153.

Hamilton, J. D., 1994, "State-Space Models," in R. F. Engle and D. L. McFadden, eds., Handbook of Econometrics, Volume IV, (Elsevier).

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

Harris, L. E., 1990, Estimation of stock variances and serial covariances from discrete observations, Journal of Financial and Quantitative Analysis, 25, 291-306.

Harris, L. E., 1990, Liquidity, trading rules and electronic trading systems, Salomon Center Monograph Series in Finance and Economics 1990-4 (Stern School of Business, New York University).

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

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

Harris, L. E., 1994, Limit Orders, working paper, University of Southern California.

Harvey, A. C., 1991, Forecasting, Structural Time Series Models and the Kalman Filter, (Cambridge).

Hasbrouck, J., 1991, The summary informativeness of stock trades: an econometric investigation, Review of Financial Studies, 4, 571-95.

Hasbrouck, J., 1993, Assessing the quality of a security market: a new approach to measuring transaction costs, Review of Financial Studies, 6, 191-212.

Hausman, J. A., A. W. Lo and A. C. MacKinlay, 1992, "An Ordered Probit Analysis of Transaction Stock Prices," Journal of Financial Economics, 31, 319-330.

Kitagawa, G., 1987, "Non-Gaussian State-Space Modeling of Nonstationary Time Series," Journal of the American Statistical Association, 82, 1032-1041.

Madhavan, A., M. and S. Smidt, 1991, A Bayesian Model of Intraday Specialist Pricing, Journal of Financial Economics, 30, 99-134.

Madhavan, A., M. and S. Smidt, 1993, An Analysis of Daily Changes in Specialist's Inventories and Quotations, Journal of Finance, 48, 1595-1628.

Madhavan, A., M. Richardson and M. Roomans, 1994, Why do security prices change? A transaction level analysis of NYSE stocks, Working Paper, Wharton School.

Nelson, D. B., 1991, Conditional Heteroskedasticity in asset returns: a new approach, Econometrica, 59, 347-370.

Niederhoffer, V., 1965, Clustering of stock prices, Operations Research, 13, 258-265.

Niederhoffer, V., 1966, A new look at clustering of stock prices, Journal of Business, 39, 309-313.

Table 1.

The state variables in the model are the implicit efficient price, m_t, and the quote exposure cost c_t, where t indexes 15-minute intraday intervals (plus the overnight period). The dynamics of the state variables are:

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

$$\ln(c_t) = \mu_t^c + \phi^c\left(\ln(c_{t-1}) - \mu_{t-1}^c\right) + v_t$$

The cost disturbance $v_t \overset{d}{\sim} \mathrm{N}(0, \sigma_v)$. The efficient price disturbance, u_t, has standard deviation σ_t and after standardization is distributed as a generalized error distribution variate with tail-thickness parameter ν:

$$z_t \equiv u_t / \sigma_t \overset{d}{\sim} GED(\nu)$$

The efficient price variance follows a modified EGARCH process:

$$\ln(\sigma_t^2) = \mu_t^\sigma + \phi^\sigma\left(\ln(\sigma_{t-1}^2) - \mu_{t-1}^\sigma\right) + \gamma\left(\mathrm{E}_{t-1}\left[|u_{t-1}|\right]/\sigma_{t-1} - \mathrm{E}|z_{t-1}|\right)$$

where $\mathrm{E}_{t-1}|u_{t-1}|$ is the filtered estimate conditional on the bid and ask prices through t-1.

The deterministic component of the cost process is:

$$\mu_t^c = \mu_0^c + \alpha^{c,open}\exp\left(-\lambda^{c,open}\tau_t^{open}\right) + \alpha^{c,close}\exp\left(-\lambda^{c,close}\tau_t^{close}\right)$$

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 deterministic component of the variance is:

$$\mu_t^\sigma = \begin{cases} \mu_0^\sigma + \alpha^{\sigma,open}\exp\left(-\lambda^{\sigma,open}\tau_t^{open}\right) + \alpha^{\sigma,close}\exp\left(-\lambda^{\sigma,close}\tau_t^{close}\right), \\ \quad \text{if } t \text{ is not the daily close} \\ \mu_{overnight}^\sigma, \text{ if } t \text{ is the daily close} \end{cases}$$

The observations are the quotes, which comprise a bid and ask price, $q_t = \{b_t, a_t\}$. These are functions of the state variables:

$$b_t = \mathrm{Floor}[m_t - c_t]$$
$$a_t = \mathrm{Ceiling}[m_t + c_t]$$

The column corresponding to the "full" model gives parameter estimates based the Kitagawa nonlinear filtering procedure. The "cost" estimates reflect an estimate of the cost-related parameters assuming a diffuse prior for the efficient price (also using the Kitagawa procedure). The "EGARCH" estimates refer to maximum likelihood estimation of a discretized EGARCH specification.

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.[JH1]

		Model		
		Full	Cost	EGARCH
Quote exposure cost parameters:	μ_0^c	−1.499 (0.037)	−1.498 (0.037)	
	$\alpha^{c,open}$	0.696 (0.137)	0.699 (0.135)	
	$\lambda^{c,open}$	2.319 (0.750)	2.317 (0.734)	
	$\alpha^{c,close}$	0.218 (0.062)	0.217 (0.061)	
	$\lambda^{c,close}$	1.055 (0.715)	1.057 (0.714)	
	ϕ^c	0.300 (0.024)	0.317 (0.024)	
	σ_v	0.705 (0.017)	0.702 (0.017)	
EGARCH parameters:	μ_0^σ	−0.706 (1.286)		−0.593 (1.325)
	$\alpha^{\sigma,open}$	2.641 (1.085)		2.749 (1.109)
	$\lambda^{\sigma,open}$	0.432 (0.244)		0.458 (0.267)
	$\alpha^{\sigma,close}$	1.901 (0.896)		1.949 (0.936)
	$\lambda^{\sigma,close}$	0.567 (0.379)		0.587 (0.422)
	$\mu_{overnight}^\sigma$	2.908 (0.157)		3.084 (0.187)
	ϕ^σ	0.877 (0.020)		0.875 (0.022)
	γ	0.299 (0.031)		0.348 (0.038)
	ν	0.855 (0.021)		0.719 (0.019)

Figure 1

As a function of the efficient price *m* and quote exposure cost *c*, the discrete bid and ask quotes are given by *b*=Floor[*m*−*c*] and *a*=Ceiling[*m*+*c*]. Given bid and ask quotes *a* and *b*, the region of feasible *m* and *c* is:

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

The figure depicts the regions *Q*(0, 1), *Q*(0, 2) and *Q*(0, 3).

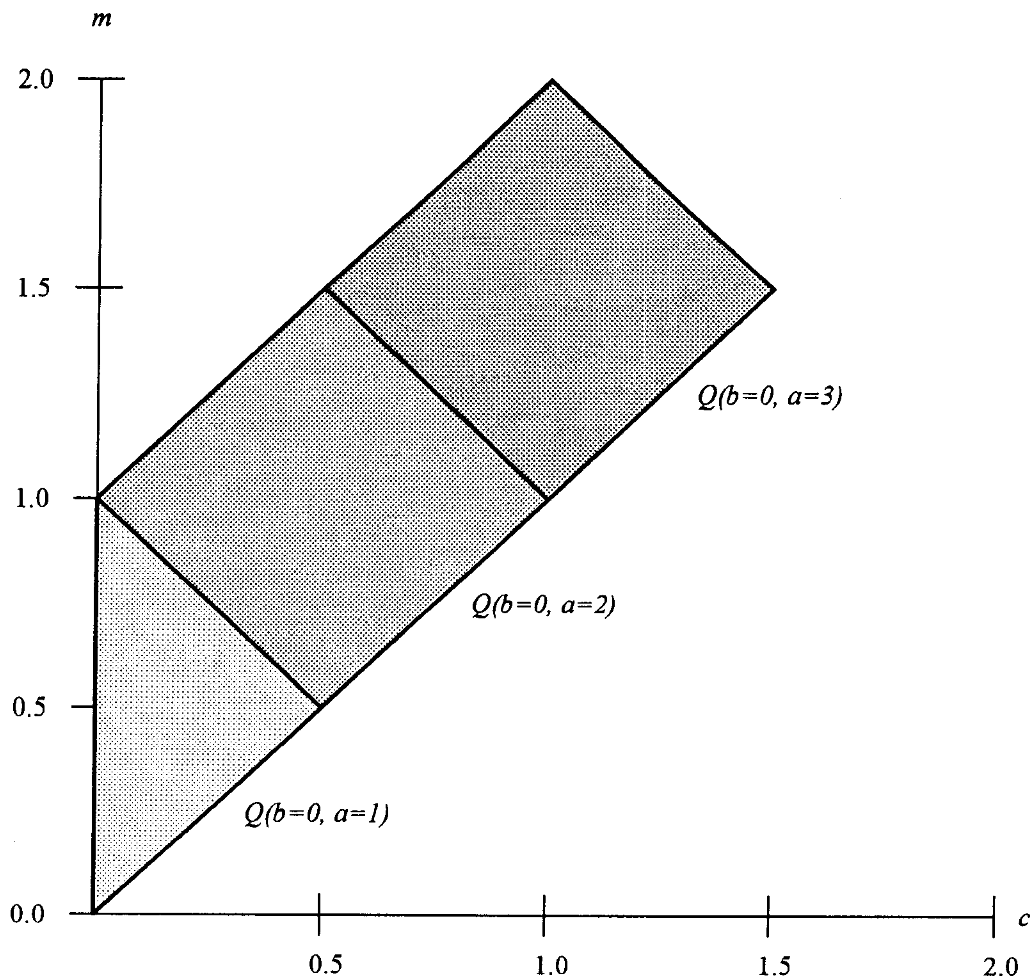
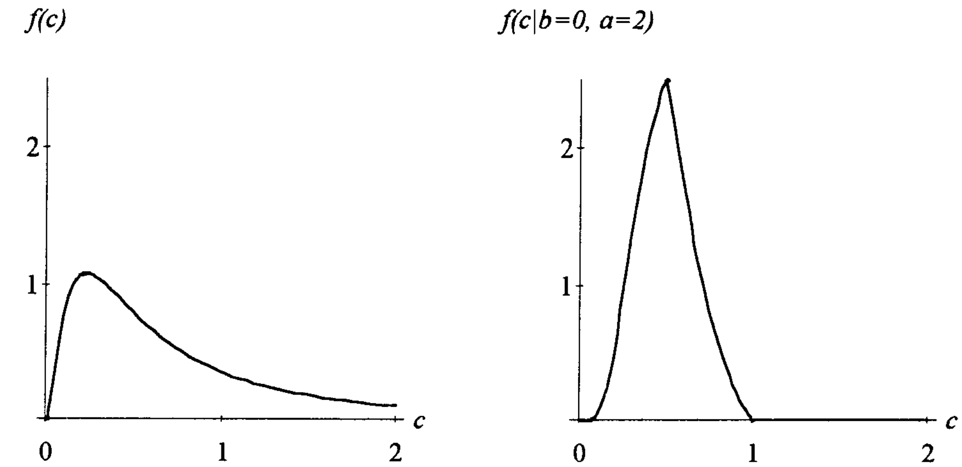


Figure 2

Figure depicts unconditional and condition probability densities for the efficient price *m* and quote exposure cost *c*. The unconditional density of *c* is lognormal: Log[*c*] is normally distributed with mean −1.0 and standard deviation 0.6. The unconditional density for *m* is a uniform diffuse prior on the interval (0, κ), where κ is an arbitrary positive constant (and does not appear in the conditional densities). The conditional densities are conditional on observing bid and ask quotes of *b*=0 and *a*=2.

Panel A. Unconditional and conditional densities of the quote exposure cost *c*.



Panel B. Unconditional and conditional densities for the efficient price *m*.

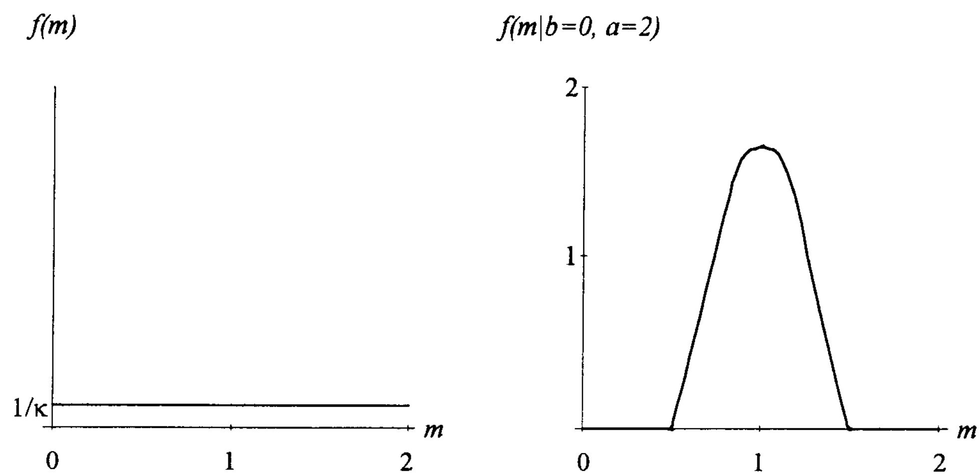


Figure 3

Figure depicts the probability density functions for the standard normal and standard GED with tail-thickness parameter ν=0.855.

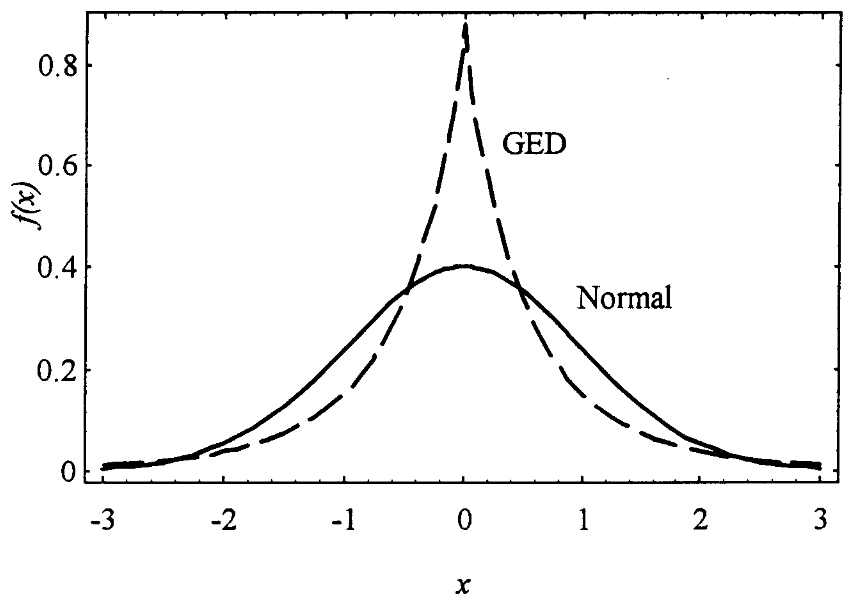


Figure 4

Figure depicts the deterministic component of the quote exposure cost (solid line). The dashed lines define the implied plus-or-minus two-standard deviation limits of the stochastic component of the quote exposure cost. (NB: these are not estimation confidence intervals.)

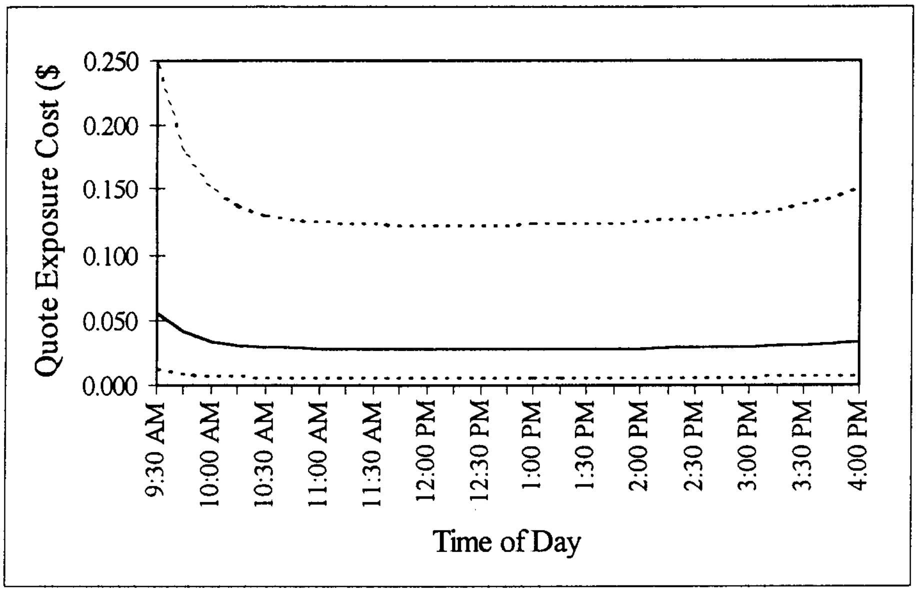


Figure 5

Figure depicts the polytrope associated with the feasible set for bid quote exposure cost β, ask quote exposure cost α, and efficient price m for observed discrete bid b=0 and ask a=1. This region is $Q(b,a)=\{(m,\alpha,\beta):\alpha,\beta>0, b<m-\beta<b+1 \text{ and } a-1<m+\alpha<a\}$. The figure shows a detailed view and rotated perspectives.