



components arising from the trading mechanism. Whether one regards these latter microstructure components as chief concerns or nuisance effects, cursory consideration of their magnitude suggests that they be addressed in some fashion in the modeling process. This paper proposes a simple structural model for bid and ask quote dynamics that incorporates several key microstructure effects common to most security markets: a stochastic cost of market-making, discreteness (restriction of feasible prices to a grid) and clustering (the tendency of observed prices to lie on 'natural' multiples of the minimum tick). The present paper estimates the model for bid and ask quotes in the US Dollar/Deutsche-mark market, but application to other markets (such as US equities) is a natural extension.

The economic features of the model figure prominently in the microstructure literature. There are strong theoretical arguments and compelling empirical evidence that the fixed and informational costs incurred by dealers are impounded in the bid and ask quotes they expose to the market. O'Hara (1995) surveys theoretical models. Hasbrouck (1996) reviews empirical approaches to measuring and resolving these components. Although equity markets provided the initial impetus for this work, recent analyses have focused on foreign exchange (FX) markets (Lyons, 1995, 1996; Yao, 1997).

Discreteness of security prices became an important empirical consideration when option valuation applications gave rise to a need for estimates of stock price volatility. The earliest papers view transaction prices as symmetrically rounded random walks. Later work incorporates more realistic and complete microstructure effects.<sup>2</sup> Although statistical models incorporating discreteness have become progressively more refined, however, the normative analysis of economic behavior with a discrete price set is not yet fully developed.<sup>3</sup> As noted below, the rounding scheme used in the present paper is consistent with Bertrand competition among dealers in some (but not all) market structures.

Like discreteness, clustering first arose as an empirical concern. Ball et al. (1985) find clustering in the London gold market. Harris (1991) notes that for US equity markets (in a sample that predates the recent switch to sixteenths), prices lying on whole numbers are more common and prices ending in 'odd-eighths'

---

<sup>2</sup> Gottlieb and Kalay (1985) and Ball (1988) advocate the rounded-random-walk paradigm. The analyses of Dravid (1991) and Harris (1990) incorporate a bid/ask spread in a covariance-stationary method-of-moments framework. Probit models have been estimated for discrete transaction price changes (Hausman et al., 1992) and spreads (Bollerslev and Melvin, 1994). Closest in approach to the present paper are the latent-variable state-space models used by Glosten and Harris (1988) and Hasbrouck (1998). Hasbrouck (1998) reviews these and related developments.

<sup>3</sup> Economic analyses of the effect of the tick size on agent behavior and/or the optimal tick size include Ahn et al. (1996), Angel (1997), Anshuman and Kalay (1998), Bernhardt and Hughson (1996), Brown et al. (1991), Chordia and Subrahmanyam (1995), Cordella and Foucault (1996), Glosten (1994), Harris (1990) and Harris (1991).

(e.g., 5/8) are more rare (relative to the unconditionally expected incidence). Prior to recent reforms, quotes in the US Nasdaq equity market exhibited dramatic clustering. (See Christie and Schultz (1994) and Christie et al. (1994). Schwert (1997) reviews these and related papers. Chan et al. (1995), Bessembinder (1994) and Bollerslev and Melvin (1994) note spread clustering in US \$/DM quotes.

From an economic perspective, clustering is often characterized as arising when market participants agree (perhaps implicitly) to a price increment that is coarser than the technically mandated minimum. Harris (1991, 1994, 1997) suggests that such a convention arises from traders seeking to minimize negotiation costs, to avoid extended rounds of bargaining over amounts of diminishing importance. Natural multiples (round numbers) arise as focal points in the negotiation game (see Schelling, 1960). Christie and Schultz suggest that the Nasdaq quote clustering reflected a collusive coordination mechanism, a hypothesis to which subsequently subpoenaed trading room tapes have lent considerable support (U.S Securities and Exchange Commission, 1996).

Bearing these economic considerations in mind, the present model may be summarized as follows. The permanent price component common to the observed bid and ask quotes follows a random walk. The dealer posting a bid or ask quote is subject to a stochastic trading cost that encompasses clearing, asymmetric information and inventory (position management) costs specific to the next trade and some allocation of fixed costs. These costs imply ask and bid reservation prices (the permanent component plus and minus the trading cost), which map in turn onto the posted discrete quotes via a stochastic rounding process.

The rounding process is governed by a stochastic tick/pip size.<sup>4</sup> In the DM/\$ market the minimum tick available on the computer systems is 0.0001 DM/\$. Nevertheless, bid and ask quotes tend to concentrate on multiples of 0.0005. In the present model, the implicit tick size is viewed as a Bernoulli random variable. Conditional on the implicit tick, the rounding is asymmetric. The posted bid quote is the permanent component less the cost of trading rounded down to the next available (implicit) tick multiple. Similarly, the posted ask quote is the permanent component plus the trading cost rounded up.

The structural model is similar to that of Hasbrouck (1998) in its incorporation of stochastic quote exposure costs and asymmetric rounding of bid and ask quotes. The modeling of clustering in the present paper is new. The estimation approaches of the two papers also differ significantly. Whereas Hasbrouck

---

<sup>4</sup> As a matter of terminology, both 'tick' and 'pip' generally refer to the smallest unit in which a price may be quoted. The former term predominates in usage related to most markets (including most equities, futures and options), while the latter is more common in FX markets. As the focus of this paper is on broader issues of discreteness, the more widespread term ('tick') is used.

(1998) relies on maximization of an approximate likelihood function (based on Kitagawa (1987)), the present study uses a Gibbs sampler approach that is more computationally efficient and amenable to extension.

The paper is organized as follows. First, to motivate the model and provide an application, the next section characterizes in a preliminary fashion a sample of daily \$/DM exchange rate quotes. A simple model broadly consistent with the features of the data is presented in Section 3. Section 4 discusses estimation via traditional methods and the alternative MCMC strategy. Results for the \$/DM data are reported in Section 5. Section 6 describes extensions to the simple model. A brief summary concludes the paper in Section 7.

## **2. A preliminary look at the data**

As a representative series, this study examines bid and ask quotes in the US Dollar/Deutschmark market. The structure is a continuous dealer market. Bid and ask quotes are disseminated by Reuters (among other systems), but there is also an active interdealer market for which no public quotes are visible. Bessembinder (1994), Bollerslev and Melvin (1994), Lyons (1995), Peiers (1997) and Yao (1997) discuss the structure of this market and certain features of the data relevant to the present study.

The sample consists of bid and ask quotes prevailing on the Deutschmark on regular UK business days in 1996, quoted in DM per US dollar. The data are Reuters quotes from the Olsen Associates HFDF II data set. The frequency of this data set is half-hourly, but most detailed results in the present study are based on an extracted daily series (values as of noon GMT, which is roughly the middle of the London trading day). A representative model is estimated, however, for the half-hourly data.

The Reuters quotes are indicative quotes entered by participating dealers. The bid and ask quotes used here are simply those most recently entered (as of noon GMT). Although they are the most timely quotes, they are not necessarily 'best', since the bid or ask (or both) might be dominated by prices submitted earlier that are still considered active by the dealers who posted them. These quotes are not, in any event, firm. Furthermore, they are often dominated by bids and asks available in the interdealer market (Lyons, 1995) and on electronic trading systems such as the Reuters D2000-2 system.<sup>5</sup> Prices are quoted to four decimal places, e.g., 1.4123 DM/\$. The tick size imposed by the information systems is thus 0.0001 DM/\$. For convenience, all prices in this paper will be reported in ticks, e.g., 14,123.

---

<sup>5</sup> I am indebted to Charles Goodhart and Michael Melvin for their comments on the limitations of the Reuters quotes.

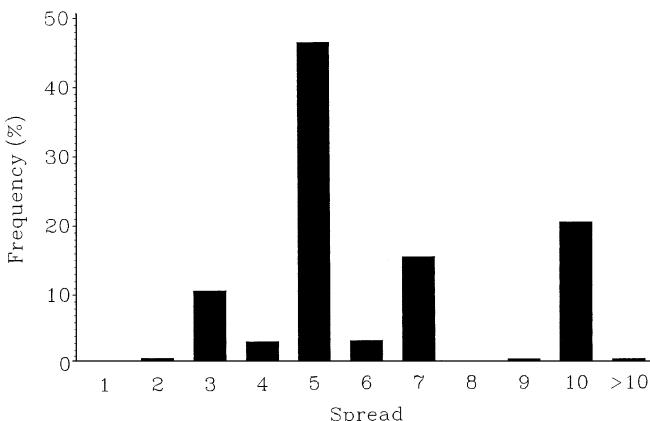


Fig. 1. Distribution of Bid–Ask Spread (Note: Frequency (in DM 0.0001 ticks) of bid–ask spreads for \$/DM exchange rates as of noon, GMT, 1996).

Fig. 1 depicts the distribution of the spread (the ask less the bid). The spread exhibits considerable variation, but is almost always ten ticks or less. (In only 0.4% of the cases was the spread greater than ten ticks.) This suggests that the resolution of the spread is generally less than ten percent of its magnitude. Even in light of the discreteness mandated by the tick size, however, the distribution of the spread can by no means be considered smooth. The distribution is sharply bimodal, with peaks at five and ten ticks. Bessembinder (1994) and Bollerslev and Melvin (1994) note spread clustering in the Reuters indicative quotes. Goodhart et al. (1996) note, however, that clustering is not observed in the D2000-2 quotes.

It might be hypothesized that this clustering simply reflects a bimodal cost structure, two regimes in which the costs of market making are near five and ten ticks. This hypothesis can be investigated by considering the distribution of bid and ask prices separately. Fig. 2 plots the distribution of the last digit of the bid and ask prices. If the clustering of the spread were arising from clustering in the underlying costs, one would still expect to observe a flat distribution of price last-digits. Fig. 2 suggests that this is not the case. The bid and ask prices considered separately exhibit considerable clustering.

Not only do we observe a higher-than-expected occurrence of five- and ten-tick bids and offers, but if the bid lies on a five-tick multiple, there is a higher probability that the ask also lies on a five-tick multiple. A simple classification test soundly rejects the null hypothesis of independence. It should also be noted that the data are drawn from a time of the day (noon) when the market is extremely active. Clustering is more pronounced at other times.

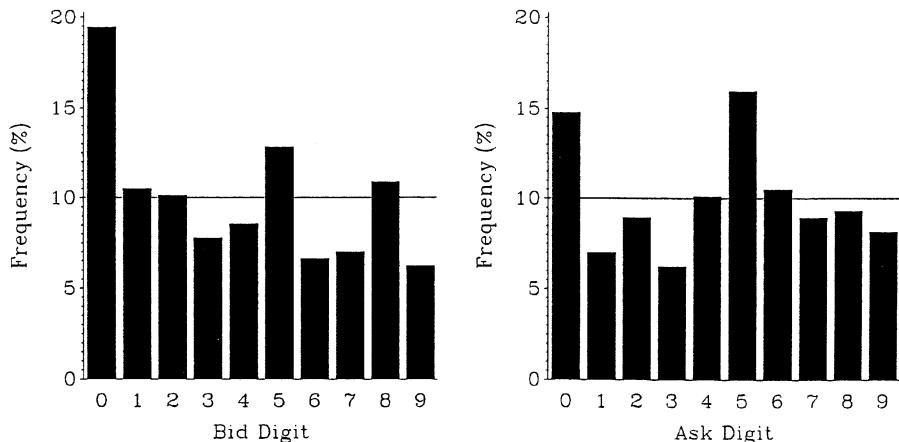


Fig. 2. Distributions of last digits in bid and ask quotes (Note: Frequency (in DM 0.0001 ticks) of the last digit in the bid and ask quotes for all trading days in 1996 as of noon GMT).

Table 1  
Preliminary descriptive statistics for \$/DM quotes

Mean of bid-ask spread (ticks)	6.1
Standard deviation of <i>bid-ask</i> spread (ticks)	2.3
Variance of daily log quote midpoint price changes	$2.03 \times 10^{-5}$
Percentage of quotes in which both the <i>bid</i> and <i>ask</i> lie on a five-tick multiple.	23.4%

Note: \$/DM quotes as of noon (GMT) for all trading days in 1996 (257 observations). (Source: Olsen Associates HFDF II).

These features of the data are broadly consistent with a model that distinguishes between the minimum tick feasible with the market's technology and regulations and the implicit tick that arises as a convention among traders. From a data modeling perspective, we might hypothesize a tick size that switches between one and five, with the former being somewhat more likely.

The model described in the next section is parameterized by the mean and variance of a quote exposure cost that approximates the 'half-spread', the variance of the daily random walk component of the price and the probability that the implicit tick size at a given time is five. Table 1 summarizes related quantities computed from the discrete data.

### 3. The basic model

This section describes key features of the model. No claim is made that the specification is in any sense optimal. Rather, the intent is to establish a

parsimonious specification that is comprehensive enough to capture the features of the representative quote series discussed in the last section: the bid–ask spread, discreteness and clustering.

The latent price variable in the market is an implicit efficient price,  $M_t$ , that is economically interpreted as the expectation of end-of-trading value of the security conditional on all public information. It is assumed to follow a lognormal random walk:

$$m_t = m_{t-1} + u_t \quad (1)$$

where  $m_t = \log(M_t)$  and the  $u_t \stackrel{\text{iid}}{\sim} N(0, \sigma_u^2)$  reflect updates to public information.

Dealers are assumed subject to a nonnegative cost of market making or quote-exposure, denoted  $C_t$ , that impounds fixed and marginal (per trade) costs of their operations. Stochastic variation in the determinants of these costs (particularly volatility and asymmetric information) is presumed to induce variation in the overall cost. Furthermore, the FX quotes modeled in the present paper represent a random draw from a population of market-makers. The quote exposure cost will also therefore impound the effects of cross-sectional variation. Although some of the cost determinants (particularly volatility) are known to be serially correlated, the cost is assumed here to be i.i.d.. More precisely, it is assumed lognormal:  $\log(C_t) = c_t \stackrel{\text{iid}}{\sim} N(\mu_c, \sigma_c^2)$ .

I assume that in the absence of tick restrictions, a market maker would bid  $M_t - C_t$  and offer (ask)  $M_t + C_t$ . The mapping of these continuous analogs of the quotes onto the discrete quotes is a rounding transformation. Without loss of generality, let the prices be scaled so that the mandated tick size is unity, i.e., that quotes are constrained to the set  $\{1, 2, \dots\}$ . In the absence of clustering, the rounding transformation is assumed to be:

$$q_t = \begin{bmatrix} \text{bid}_t \\ \text{ask}_t \end{bmatrix} = \begin{bmatrix} \text{Floor}(M_t - C_t) \\ \text{Ceiling}(M_t + C_t) \end{bmatrix} \quad (2)$$

where the floor and ceiling functions round down (and up, respectively) to the next integer.

The asymmetric rounding is motivated by models of Bertrand competition in which no competitor is willing to make a market at an expected loss. This view is compatible with most models that incorporate discreteness (Glosten, 1994; Chordia and Subrahmanyam, 1995; Kandel and Marx, 1997). On the other hand, Bernhardt and Hughson (1996) show that in a market where absolute time priority is enforced (first-come, first-served at a given price), strategic considerations may cause dealers to post bid and ask quotes that are less aggressive than the present rounding model would suggest. Time priority is not observed, however, in most dealer markets (such as the FX market). Even in equity markets time priority is enforced at best locally.

Clustering is introduced by specifying the discrete price grid as  $\{K_t, 2K_t, 3K_t, \dots\}$  where  $K_t$  is a positive integer random variable. Intuitively,

$K_t$  may be viewed as the tick size used by the quote-setter at time  $t$ . When  $K_t = 1$ , the price grid is the set of positive integers (as above);  $K_t = 5$  gives the set of five-multiples  $\{5, 10, 15, \dots\}$ . The modified rounding transformation is

$$q_t = \begin{bmatrix} \text{bid}_t \\ \text{ask}_t \end{bmatrix} = \begin{bmatrix} \text{Floor}(M_t - C_t, K_t) \\ \text{Ceiling}(M_t + C_t, K_t) \end{bmatrix} \quad (3)$$

where the *Floor* and *Ceiling* functions round to the next lower and higher multiples of  $K_t$ . Thus, when  $K_t = 1$ , all positive integer prices are observable. When  $K_t = 5$ , only multiples of five are permitted. The clustering variable is a latent structural construct, rather than an observed datum. The distinction may be illustrated by example. Observing a bid and offer of 100 and 105 (both of which are multiples of five) does not imply that  $K_t = 5$ . It might be the case that  $K_t = 1$  and the values of  $M_t$  and  $C_t$  happen to round (via Eq. (3)) to the observed values. The implicit tick size is assumed to be distributed as an i.i.d. Bernoulli variable:

$$K_t = \begin{cases} 1 & \text{w probability } (1 - k), \\ 5 & \text{w probability } k. \end{cases} \quad (4)$$

The parameter  $k$  has a natural interpretation as the clustering probability or intensity.

Harris (1991) suggests that clustering depends (positively) on the dispersion in traders reservation prices, which is itself likely to depend (positively) on volatility and extent of information asymmetries. Stochastic variation in the latter can therefore lead to randomness in the clustering tendency. Clustering theories based on market structure (e.g., Dutta and Madhavan, 1997 and Kandel and Marx, 1997) do not directly predict variation in clustering. They are not, however, incompatible with variation in clustering, since underlying factors (such as volatility and information asymmetry) are likely to affect the costs and benefits of forming and sustaining coalitions and collusive agreements.

Although the market-structure studies address the pre-reform Nasdaq market, the latter resembles in certain respects the FX market. The FX markets and pre-reform Nasdaq are/were both dealer markets with non-anonymous quotes and a minimum tick size. Both markets possess/possessed an interdealer market in which dealers could lay off positions at prices that are not available to outside customers.<sup>6</sup> In both markets there are/were arrangements or affinities that bind individual customers to particular dealers.<sup>7</sup>

---

<sup>6</sup> Nasdaq has an interdealer system (SelectNet) on which dealers can post quotes visible only to other dealers. The FX market, too, has interdealer brokers. The institution seems to be a regular feature of dealer markets, being present as well in the US government bond market and the London Stock Exchange.

<sup>7</sup> The Nasdaq market features preferencing and payment for order flow. In the FX market, it is the usual practice for a customer to check quotes only at banks with whom they already have a relationship of some sort.

For a rogue dealer who is contemplating breaking the established implicit tick convention, both of these features decrease the expected benefits and increase the expected costs. The strong relationships between customers and their dealers suggest that more aggressive pricing will not be rewarded by increased market share. If the aggressive pricing leads to the dealer being shunned in the inter-dealer market, their inventory management costs will increase.

Similarities notwithstanding, there are also significant differences between the two markets. While the quotes set by pre-reform Nasdaq dealers established actual transaction prices for a large volume of retail trades, customers in the FX market are primarily institutions who have more leverage in negotiation. As noted in Section 2, the FX market comprises alternative trading mechanisms that may offer bids and asks that are better than the indicative quotes used here. Thus, transaction prices (which are not reported in the FX market) may reflect an effective spread that is substantially narrower than that publicly quoted. Furthermore, the clustering behavior on the pre-reform Nasdaq market was a fairly stable phenomenon. Transitions from eighth quoting to quarter quoting and vice versa were rare. In contrast, clustering in the DM/\$ market is much more transient, with no obvious regime breaks.

The three stochastic variables in the model (efficient price, dealer cost and clustering intensity) are assumed independent. This independence is not used as an identifying restriction in the estimation. It is nevertheless clear that commonality of underlying determinants is likely to induce dependence. For example, both clustering intensity and market-making cost are likely to depend on volatility. Allowance for correlations among the variables is therefore a logical and desirable extension.

The model represents a simple structural approach to dealing with security prices confined to a discrete support. Standard techniques for modeling discrete data suggest alternative approaches. Hausman et al. (1992) use an ordered probit model to characterize short-term transaction price transitions for NYSE stocks. Bollerslev and Melvin (1994) estimate an ordered probit model for the spread in the \$/DM market. Discrete transitions could also be modeled as draws from a multinomial density.

These alternative approaches are reduced form models, capable of modeling the spread or quote transitions in a highly flexible fashion. It may be more difficult, however, to attach economic interpretation to these specifications. The simple model described in this section, for example, does not possess a conventional ordered probit representation. Furthermore, although one could easily compute multinomial transition probabilities from the model, the reverse inference (from transition probabilities to model parameters) is extremely difficult. Problems of interpretation are not limited to highly structured models. Most microstructure models of price evolution, for example, reflect a distinction between a martingale (information-related) component and transient (market- or liquidity-related) components. Yet even this general

characterization is difficult to extract from a probit or multinomial specification. As is usually the case, the choice between reduced-form or structural models turns on whether the objective is fitting/forecasting the observed data, or investigating latent economic features of interest.

## 4. Estimation and testing

### 4.1. Estimation

At each time  $t$ , the unobserved state vector of the model is  $s_t = (m_t, c_t, K_t)$ . Based on a sample of observed quotes  $q = \{q_1, q_2, \dots, q_T\}$  we wish to make inferences about the parameter set  $\Theta = (\sigma_u^2, \mu_c, \sigma_c^2, k)$ .

The log likelihood function of the observations is

$$\log f(q; \Theta) = \sum \log f(q_{t+1} | q_t, q_{t-1}, \dots, q_1; \Theta). \quad (5)$$

The summands on the right-hand side are prediction densities. Given  $f(s_t | q_t, q_{t-1}, \dots, q_1)$ ,

$$f(q_{t+1} | q_t, q_{t-1}, \dots, q_1) = \int_{s_{t+1} \in q_{t+1}} \int f(s_{t+1} | s_t) f(s_t | q_t, q_{t-1}, \dots, q_1) ds_t ds_{t+1} \quad (6)$$

where  $f(s_{t+1} | s_t)$  is the transition density of the state variables (and  $\Theta$  has been suppressed for notational convenience). The  $s_{t+1} \in q_{t+1}$  notation indicates that the integration is over all values of the state variables that map to the observed quotes. (This restriction is redundant in the inner integral because  $f(s_t | q_t, \dots) = 0$  for all  $s_t \notin q_t$ .)

The transition density for this model is relatively simple:

$$f(s_{t+1} | s_t) = f(m_{t+1}, c_{t+1}, K_{t+1} | m_t, c_t, K_t) = f(m_{t+1} | m_t) f(c_{t+1}) \Pr(K_t) \quad (7)$$

where the three r.h.s. densities are respectively normal, lognormal and Bernoulli. Due to the truncated integration regions in Eq. (6), however, the prediction densities do not possess convenient closed-form representations. In state-space models of this sort (nonlinear and non-Gaussian), Kitagawa (1987) suggests approximating the conditional densities by step functions on a fine grid and performing the integrations numerically. Estimation then proceeds via maximum likelihood. Hasbrouck (1998) uses this technique to estimate a variant of the model proposed here.

The present paper departs from this approach by casting the estimation in a Bayesian framework and using a Gibbs sampler to compute the posterior densities. To the reader who has previously encountered Bayesian analysis

solely as a tool for incorporation of prior beliefs, it should be emphasized that the priors used in the present problem are diffuse, and their content is therefore low. The principal motivation is instead the conceptual and computational simplicity afforded by the modern Bayesian perspective in dealing with latent data (such as the state variables in the present problem). The conceptual simplicities arise from treating unobserved data in the same manner as the unobserved parameters. The computational efficiencies result from characterizing analytically intractable densities using simulation, invoking ‘... the essential duality between a sample and the density (distribution) from which it is generated’ (Smith and Gelfand, 1992).

The estimation is formulated in a Bayesian framework as follows. Given a prior distribution on the parameter set  $f(\Theta)$  and observed quotes  $q$ , we seek a posterior  $f(\Theta | q)$  to use in summary analysis or tabulation. This posterior is constructed as the marginal of the joint conditional density  $f(\Theta, s | q)$  where  $s = \{s_1, s_2, \dots, s_T\}$  is the set of unobserved states. The density  $f(\Theta, s | q)$  is joint over both parameters and unobserved data and lies at the heart of the analysis. Although it is analytically intractable, it may be characterized by a large sample of random draws  $\Theta^{(i)}$  and  $s^{(i)}$  for  $i = 1, \dots, N$ . To analyze the parameter posterior, for example, we essentially ‘discard’ the  $s^{(i)}$  draws and plot the histograms (or kernel density estimates) of the  $\Theta^{(i)}$  draws.

The  $\Theta^{(i)}$  and  $s^{(i)}$  draws are generated using a Gibbs sampler, a technique that produces joint draws by iterating over conditional draws. Specifically:

- Step 0. Initialize  $s$ .
- Step 1. Draw new  $\Theta$  from  $f(\Theta | s, q) = f(\Theta | s)$ .
- Step 2. Draw new  $s$  from  $f(s | \Theta, q)$
- Step 3. Repeat steps 1 and 2.

Each iteration of this process is called a scan or sweep. Under mild regularity conditions, in the limit (as the number of scans goes to infinity), successive draws of  $s$  and  $\Theta$  can be viewed as (dependent) draws from the target distribution  $f(\Theta, s | q)$ . In practice, the draws are monitored to assess the convergence. When the stream of simulated values is judged to be sufficiently large and well mixed, posterior parameter distributions may be constructed from the simulated parameter draws.

Excellent general introductions to MCMC analysis include Casella and George (1992), Chib and Greenberg (1996), Gilks et al. (1996), Smith and Gelfand (1992) and Tanner (1996). Chib and Greenberg (1996) survey applications in econometrics. Jacquier et al. (1994) present an application to stochastic volatility modeling (see in addition the comments and response for this paper in the same volume). Kim et al. (1998) discuss further aspects of stochastic volatility modeling. Key references in the MCMC analysis of state space models include Carter and Kohn (1994), De Jong and Shephard (1995), Shephard and Pitt (1997b) and Shephard and Pitt (1997a). Manrique and Shephard (1998) and

Manrique (1997) discuss general issues in the application of MCMC methods to limited dependent variable time series analysis (such as the present application). Closest to the present paper, Manrique and Shephard (1997) present an MCMC estimation of a variant of the model suggested in Hasbrouck (1998).

As  $N$  gets large, the drawn sample of  $\Theta^{(t)}$  tends to exactly characterize the *small sample* posterior  $f(\Theta | q)$  for a given set of observations. ( $N$  is not limited by the size of the original data sample,  $T$ .) In the present application the posteriors are data-dominated (i.e., not highly sensitive to prior specification). The posterior modes therefore approximate the frequentist maximum likelihood estimates and possess the usual large-sample properties (consistency and asymptotic normality). With at least 250 observations, the present sample sizes would normally be considered sufficiently large to justify the asymptotic approximation, but the appeal to asymptotic behavior is not necessary for Bayesian estimation.

The above algorithm describes the implementation of the Gibbs sampler at the ‘top level’ of the problem, successively conditioning on  $\Theta$  and  $s$ . In what follows, Gibbs samplers are also implemented within steps 1 and 2, to draw from  $f(\Theta | s)$  and  $f(s | \Theta, q)$ . It is important to note that these latter ‘lower level’ samplers need not be iterated to convergence within each of the steps 1 and 2. Rather the entire process can be viewed as a single Gibbs sampler with many steps, each of which involves a single draw of one parameter or latent variable conditional on all the remaining parameters and variables. We now turn to the specifics of the draws.

#### 4.1.1. Sampling from $f(\Theta | s)$

The distribution  $f(\Theta | s)$  is proportional to the product  $f(s | \Theta)g(\Theta)$ , where  $f(s | \Theta)$  is the usual model likelihood and  $g(\Theta)$  is the prior distribution. (Were the states observed,  $f(\Theta | s)$  would be the posterior parameter distribution.) Although  $f(\Theta | s)$  possesses no convenient closed form, random draws from this distribution can be generated using a Gibbs sampler. That is, we may make sequential draws from  $f(\theta_i | s, \Theta_{/\theta_i})$  where  $\Theta_{/\theta_i}$  is the parameter vector with  $\theta_i$  deleted.

The motivation for use of a Bayesian approach in this application is computational simplicity, not the desire to incorporate prior beliefs. This suggests use of noninformative priors. The standard noninformative prior densities are improper, however. Verification that the posterior derived from an improper prior is itself proper is usually accomplished by direct inspection of the posterior. This is not feasible in this (and most) Gibbs sampler situations. Despite this limitation, for simplicity of exposition, the present analysis employs noninformative (improper) priors. The numerical results were virtually unchanged, however, when diffuse proper priors were employed.

The distribution of the quote exposure cost,  $\log(C_t) = c_t \stackrel{\text{iid}}{\sim} N(\mu_c, \sigma_c^2)$ , has two parameters. The noninformative prior for  $\mu_c$  is flat over the real line. The

posterior,  $f(\mu_c | s, \Theta_{/\mu_c}) = f(\mu_c | c_1, \dots, c_T, \sigma_c^2)$  is normal with mean  $\bar{c} = (\sum c_t)/T$  and variance  $\sigma_c^2/T$ . The noninformative prior for  $\sigma_c^2$  is  $f(\sigma_c^2) \propto 1/\sigma_c^2$ . The posterior  $f(\sigma_c^2 | s, \Theta_{/\sigma_c^2}) = f(\sigma_c^2 | c_1, \dots, c_T, \mu_c)$  is an inverse-chi squared distribution, denoted  $(\sum \tilde{c}_t^2) \chi_{df=T}^{-2}$ , where  $\tilde{c}_t = c_t - \mu_c$ . Similarly for  $\sigma_u^2$ , with a prior  $f(\sigma_u^2) \propto 1/\sigma_u^2$ , the posterior is  $(\sum \tilde{u}_t^2) \chi_{df=T-1}^{-2}$  where  $\tilde{u}_t = m_t - m_{t-1}$ . The noninformative prior for  $k$  is the beta distribution  $B(\alpha = 1/2, \beta = 1/2)$ . The posterior is  $B(\alpha = 1/2 + n, \beta = 1/2 + T - n)$  where  $n$  is the number of ‘hits’ (occurrences of ‘5’ in the set  $K = \{K_1, K_2, \dots, K_T\}$ ). (These are textbook results discussed in Carlin and Louis (1996), Press (1989) and Tanner (1996), among other sources.)

#### 4.1.2. Sampling from $f(s | \Theta, q)$

We may sample from  $f(s | \Theta, q)$  by successively drawing from  $f(s_t | s_{/t}, \Theta, q)$  where  $s_{/t} = \{s_i : i = 1, \dots, T \text{ and } i \neq t\}$ , i.e., a Gibbs sampler over  $t$ . At each time  $t$ , furthermore, we may draw from  $f(s_t | s_{/t}, \Theta, q)$  by successively drawing from the components of  $s_t$ , from  $f(C_t | m_t, K_t, s_{/t}, \Theta, q), f(m_t | C_t, K_t, s_{/t}, \Theta, q)$  and  $f(K_t | m_t, C_t, s_{/t}, \Theta, q)$ , i.e., a Gibbs sampler within each  $t$ .

To sample from  $f(C_t | m_t, K_t, s_{/t}, \Theta, q) = f(C_t | m_t, K_t, \mu_c, \sigma_c^2, q_t)$ , note first that the effect of conditioning on  $m_t, K_t$  and  $q_t$  is to impose a truncation consistent with Eq. (3). That is, Eq. (3) implies

$$\begin{aligned} bid < M - C < bid + K, \\ ask - K < M + C < ask, \end{aligned} \tag{8}$$

where the time subscript is suppressed and  $M \equiv e^m$ . Holding the other variables constant, the implied bounds on  $C$  are

$$\text{Max}[0, M - bid - K, ask - M - K] < C < \text{Min}[M - bid, ask - M]. \tag{9}$$

We may therefore sample  $C_t$  from the lognormal density  $f(C_t; \mu_c, \sigma_c^2)$  subject to this truncation.

To sample from  $f(m_t | C_t, K_t, s_{/t}, \Theta, q) = f(m_t | C_t, K_t, m_{t-1}, m_{t+1}, \sigma_u^2, q_t)$  for  $t = 2, \dots, T-1$ , note that  $f(m_t | m_{t-1}, m_{t+1}, \sigma_u^2)$  is normal with mean  $(m_{t-1} + m_{t+1})/2$  and variance  $\sigma_u^2/2$ . We may sample from this distribution subject to the bounds implied by Eq. (8):

$$\text{Max}[bid + C, ask - K - C] < M < \text{Min}[bid + C + K, ask - C]. \tag{10}$$

The endpoint modification for  $t = 1$  and  $t = T$  is direct.

To sample from  $f(K_t | m_t, C_t, s_{/t}, \Theta, q) = f(K_t | m_t, C_t, k, q_t)$ , note that a straightforward application of Bayes Law gives the conditional probability

$$\Pr(K = 5 | M, C, k, q) = \frac{k \Pr(q | M, C, K = 5)}{k \Pr(q | M, C, K = 5) + (1 - k) \Pr(q | M, C, K = 1)}. \tag{11}$$

Given that the mapping from state variables to quotes is nonstochastic, the probabilities on the r.h.s. are merely indicator variables that are equal to one if the conditioning state variables are compatible with the observed quotes and zero otherwise.  $K_t$  may then be drawn as a Bernoulli random variable with the conditional probability given in Eq. (11).<sup>8</sup>

#### 4.2. Model evaluation

In the absence of a convenient and widely accepted goodness-of-fit measure applicable to the present class of models, this paper will instead examine the spread distribution implied by the model. This is motivated by a desire to examine how well the model fits with respect to the discreteness and clustering features that are the model's main points of distinction, as well as by simplicity. This is not to assert that the model is likely to be adequate in all other respects (such as the assumed homoscedastic lognormality of the efficient price innovations), only that the discreteness and clustering features are the most interesting.<sup>9</sup>

The model may also be compared to alternatives in which the clustering mechanism is suppressed and/or the rounding transformation is assumed to be symmetric. These estimates are presented below along with numerical likelihoods, which could in principle be used in comparison procedures based on Bayesian or Akaike information criteria (discussed, for example, in Carlin and Louis (1996)).

Numerical likelihoods are computed using the auxiliary particle filter suggested by Shephard and Pitt (1997a). In this approach  $f(s_t | q_t, q_{t-1}, \dots, q_1)$  is approximated by a 'cloud' of (conditionally simulated) points  $s_t^i$  and associated probabilities  $\pi_t^i$  for  $i = 1, \dots, M$ . The number of simulated points,  $M$ , determines the quality of the approximation. The prediction density needed for likelihood computation is approximated (cf. Eq. (6)) by

$$f(q_{t+1} | q_t, q_{t-1}, \dots, q_1) = \sum_{i=1}^M \pi_t^i \int_{s_{t+1} \in q_{t+1}} f(s_{t+1} | s_t^i) ds_{t+1} \quad (12)$$

<sup>8</sup> The actual sequencing of the Gibbs sampler used here deviates slightly from that described above. Since the draws for  $C_t$  and  $K_t$  do not depend on the state variables at any other time, computational efficiency is enhanced by separately drawing  $C_t$  and  $K_t$  for  $t = 1, \dots, T$  in two (vectorized) steps. The draw of  $m_t$ , on the other hand, depends on  $m_{t-1}$ , and  $m_{t+1}$  however, and we must therefore iterate over  $t$ .

<sup>9</sup> Were the observed  $q_t$  in this model continuously distributed random variables, the sequence of conditional distribution function values  $F(q_1 | q_{t-1}, \dots, q_1) \equiv \Pr[\tilde{q}_t < q_1 | q_{t-1}, \dots, q_1]$  would be i.i.d. uniform on  $[0, 1]$  (Rosenblatt, 1952). Shephard and Pitt (1997a) (among others) describe how sample analogs to this sequence can form the basis for goodness-of-fit tests. This approach does not carry over directly to discrete observations.

The integrated conditional probabilities are computed numerically (using Gaussian quadrature). (They are also used to weight the auxiliary variable draws, as described in Shephard and Pitt (1997a).)

The appearance of a likelihood computation at this point in the discussion may cause the reader to question the paper's earlier claim that the Gibbs sampler approach achieved a computational advantage over maximum likelihood estimation. Indeed, Shephard and Pitt (1997a) discuss maximum likelihood estimation as one of the uses of the auxiliary particle filter. In the present application, however, although the accuracy of the simulated likelihood function approximation is sufficient to warrant use in model comparisons, it is not sufficiently well-behaved to merit use as a maximization objective function. This deficiency is fundamental (and also suggests infeasibility of *EM* and simulated *EM* approaches). The state-sampling (draws of the  $s_t^i$ ) is performed using rejection, a procedure that is not smooth in model parameters. A small parameter variation (such as that associated with the computation of a numerical derivative or search near the optimum) might 'move' a previously-acceptable  $s_t^i$  outside of the region that maps to the observed quote. The drop in  $f(q_t | s_t^i)$  from one to zero will induce a discontinuity in the computed likelihood function.

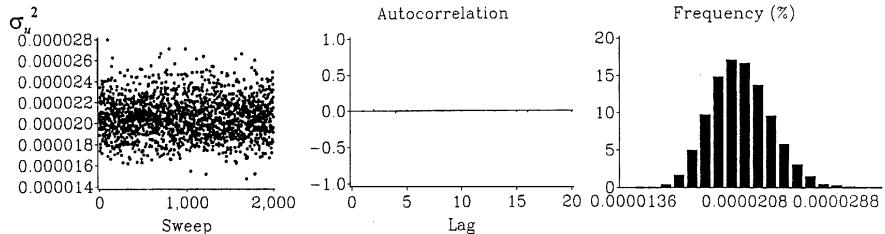
## 5. Estimation of the \$/DM series

### 5.1. Analysis of the model with clustering and asymmetric rounding

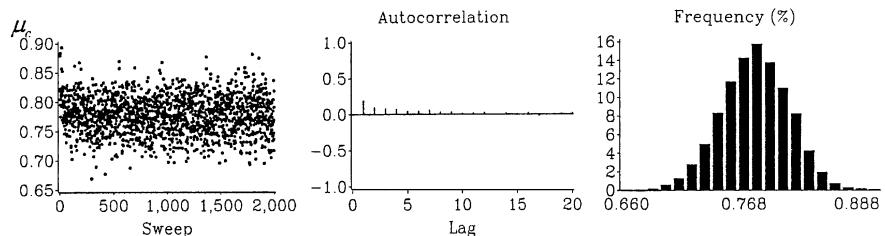
To start the sampler, the model parameters were initialized as  $\mu_c = 10$ ,  $\sigma_c^2 = 10$ ,  $\sigma_u^2 = 10$  and  $k = 0$ . The choice of  $k$  is convenient because it forces all  $K_t$  to be equal to one, a value that is compatible with all observed quotes. The other initializations were chosen (after preliminary testing of the sampler) as 'bad' starting values, i.e., values sufficiently removed from the final parameter estimates that convergence might be challenged. The  $m_t$  were all initialized to the log midpoint of the observed bid and ask. All of the  $K_t$  were initialized to one. I computed 11,000 iterations of the sampler. The program was implemented in the matrix language OX (Doornik, 1996) and took roughly seven minutes on a 200 mHz Pentium Pro.

Fig. 3 plots for each parameter the first 2000 draws of the sampler, as well as correlograms and histograms for the last 10,000 draws (i.e., after discarding the initial 1000). Visual inspection of the draws suggests that the sampler converges quickly. The  $\mu_c$  and  $k$  parameters exhibit the slowest convergence, but even here the effects of the starting values appear to die out after roughly one hundred draws. A Gibbs sampler does not produce independent draws. Since it is difficult to discern serial dependence from plots of the draws, it is useful to examine the correlograms. The autocorrelations are generally modest, with the largest

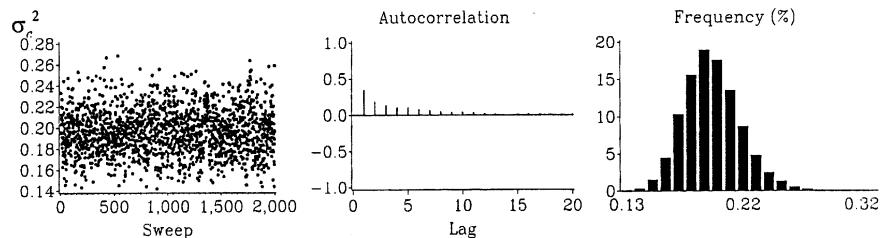
## Panel A.



## Panel B



## Panel C



## Panel D

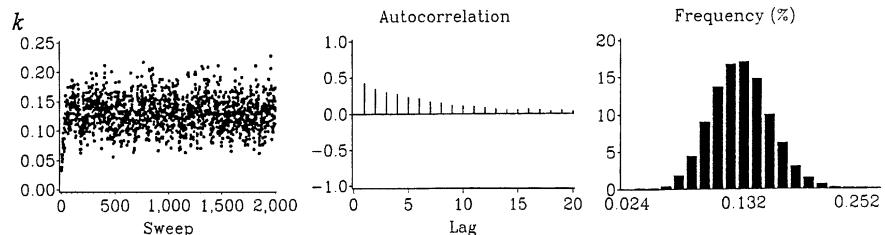


Fig. 3. Simulated Parameters. Figure is based on a single Gibbs sampler run of 11,000 iterations as described in Section 4. For each of the model parameters, the leftmost graph depicts the first 1000 draws. The middle graph shows the autocorrelations of the last 10,000 draws (after the first 1000 were discarded as burn-in period). The rightmost graph is the histogram of the last 10,000 draws.

Table 2

## Summary of Posterior Parameter Distributions

Summary statistics for simulated posterior distributions of the parameters for the a model of discrete bid and ask quotes with clustering and asymmetric rounding. The quote exposure cost is distributed as  $\log(C) \sim N(\mu_c, \sigma_c^2)$ ;  $\sigma_u^2$  is the variance of the implicit efficient price changes;  $k$  is the clustering intensity (the probability of rounding to five-tick multiples). The model was estimated for \$/DM quotes as of noon (GMT) for all trading days in 1996 (257 observations). The posteriors were constructed using a Gibbs sampler over 11 000 iterations, with the first 1000 discarded. The mode is a smoothed estimate. The standard error of the mean estimates is a spectral estimate (using a Parzen window) that corrects for autocorrelation in the sample draws.

Parameter	Mean	Mode	Std. Error of Mean	Std. Dev.
$\mu_c$	0.779	0.779	0.00062	0.031
$\sigma_c^2$	0.196	0.193	0.00050	0.022
$\sigma_u^2$	$2.06 \times 10^{-5}$	$2.15 \times 10^{-5}$	$2.5 \times 10^{-8}$	$1.8 \times 10^{-6}$
$k$	0.130	0.126	0.00080	0.027

arising for the clustering probability  $k$ . The rightmost graph for each parameter gives the histogram of the simulated parameters. This is an approximation to the exact small-sample posterior distribution for the parameters. The approximation is exact in the limit as the number of sweeps goes to infinity, holding the number of observations fixed.

Table 2 reports summary statistics for these simulated posteriors. At the posterior mean values, the cost parameter  $c_t \sim N(\mu_c = 0.779, \sigma_c^2 = 0.196)$ . Restated to level terms using the lognormal transformation, the mean and standard deviation of  $C_t = e^{c_t}$  are 2.40 ticks and 1.12 ticks, respectively. These quantities are roughly the same magnitude as the (one-tick) resolution of the discrete spread. Both figures are slightly lower than the corresponding values computed on the basis of the observed discrete spreads (cf. 6.1/2 and 2.3/2 from Table 1).

The mean estimate of  $\sigma_u^2$  is quite close to the corresponding value estimated from the discrete quote midpoints ( $2.06 \times 10^{-5}$  vs.  $2.03 \times 10^{-5}$ ). This is a consequence of the fact that at a daily frequency, the typical sizes of quote midpoint changes are large relative to the tick size. This may not be true, of course, for shorter intervals. Finally, the mean estimate of  $k$  implies that the implicit tick size is five roughly 13% of the time. In the discrete data, the bid and ask were found to lie on five-tick multiples 23% of the time. From the perspective of the model, however, these events include states where the implicit tick size is one, but the pre-rounded quantities just happen to lie close to a five-tick multiple.

The standard error of the mean estimate in Table 2 is a spectral estimate that corrects for autocorrelation. For example, the correlograms suggest that the

Table 3  
Representative transition probabilities

Bid	Ask	Probability
15,000	15,001	$2 \times 10^{-7}$
15,000	15,002	$610^{-5}$
15,000	15,003	0.0005
15,000	15,004	0.0010
15,000	15,005	0.0018
		...
15,000	15,009	0.0002
15,000	15,010	0.0027
15,000	15,011	$8 \times 10^{-5}$
		...
15,000	15,014	$2 \times 10^{-5}$
15,000	15,015	0.0005
15,000	15,016	$6 \times 10^{-6}$
15,001	15,006	0.0011
15,001	15,011	0.0001

Note: Given that the market is initially in the steady state with  $bid = 15,000$  (DM/\$  $\times 10,000$ ) and  $ask = 15,002$ , the table reports the estimated transition probabilities associated with moving to the indicated bid and ask one day later.

most pronounced autocorrelation is found in the estimates of the  $k$  parameter. With a sample standard deviation of 0.027 and 10,000 observations (11,000 draws less the 1000 discarded initial draws), the usual standard error of the mean under the independence assumption would be  $0.027/\sqrt{10,000} = 0.00027$ . The spectral estimate (0.00080) is over twice as large.

Some insight into the estimated model may be obtained by considering representative implied transition probabilities. As a convenient starting point, suppose that we are initially in a steady state defined by  $bid = 15,000$  (DM/\$  $\times 10,000$ ) and  $ask = 15,002$ . (That is, these are the current noontime quotes, and they are unchanged from recent days.) Table 3 describes several outcomes (observed quotes as of noon on the next day) and the associated transition probabilities. The probabilities are all relatively low, because the total probability is not concentrated on a small number of states. The effect of clustering is nevertheless highly visible from the probabilities of the clustered outcomes relative to nearby non-clustered outcomes. For example, the clustered outcome ( $bid = 15,000$ ,  $ask = 15,010$ ) has a probability of 0.0027, while the non-clustered outcome ( $bid = 15,001$ ,  $ask = 15,011$ ) has a probability of 0.0001, even though the spread in both cases is ten ticks.

It is also useful to consider how the structure of the model affects what we might infer from observing particular quotes. Suppose that we observe

$bid = 15,000$  and  $ask = 15,005$ . Suppressing all considerations of discreteness, but maintaining the usual assumption that the bid and ask quotes symmetrically bracket the ‘true’ (efficient) price, we would exactly infer that the efficient price is 15,002.5 (the midpoint) and that the cost of market making is 2.5 (the half-spread). From the present perspective, however, the discrete quotes arise from rounding. This introduces indeterminacy in our beliefs.

Fig. 4 illustrates the nature of this inference. Suppose that the model structure and parameters are known. (For illustration purposes, I fix them at their estimated values.) Our priors (before observing the quotes) are depicted in the left-hand graphs. Our beliefs about  $C$  are simply those given by the lognormal unconditional distribution. Our beliefs about  $M$  are ‘diffuse’, with a probability density represented in the figure as a flat line of ‘infinitesimal’ height. Our prior belief that the quotes are clustered is  $k = 0.13$  (‘13 percent’).

Given the observed quotes ( $bid = 15,000$ ;  $ask = 15,005$ ), the right-most graphs depict the posteriors. (The posterior estimate of the clustering probability is  $k = 0.40$ .) The posteriors for  $C$  and  $M$  are a blend of two component posteriors (center graphs), reflecting beliefs with and without clustering. The posterior for  $M$  is centered at the midpoint of the bid and ask, but there is also substantial probability mass away from this midpoint. The posterior for  $C$  shows that the actual quote exposure cost is almost certainly lower than 2.5, and may well be lower than 1.0. The peakedness of the posteriors reflects the truncation imposed by conditioning on the observed quotes (cf. Eq. (3)).

## 5.2. Model comparison and evaluation

As noted above, the model is most distinctive with respect to its treatment of clustering and rounding. One obvious alternative model is one in which clustering is suppressed, i.e., in which  $K_t = 1$  for all  $t$ . A simple sensible alternative to the asymmetric rounding assumed in Eq. (3) is symmetric rounding:

$$q_t = \begin{bmatrix} bid_t \\ ask_t \end{bmatrix} = \begin{bmatrix} Round(M_t - C_t, K_t) \\ Round(M_t + C_t, K_t) \end{bmatrix} \quad (13)$$

where  $Round(\cdot, K)$  rounds to the nearest  $K$ -multiple. This model is a priori suspect in that it may generate crossed quotes (states of equal bid and ask).

Table 4 presents log likelihood values for the four models arising from the clustering and rounding possibilities. (The model described in the previous section, with asymmetric rounding and clustering, is model  $C$ .) The rounding variants without clustering ( $A$  and  $B$ ) are virtually indistinguishable. Clustering, however, is strongly supported. By either the Bayesian or Akaike Information Criteria, the evidence favoring model  $C$  (clustered) over  $A$  (not clustered) is overwhelming. Interestingly, there is also strong evidence favoring  $D$  (symmetric rounding) over  $C$  (asymmetric rounding).

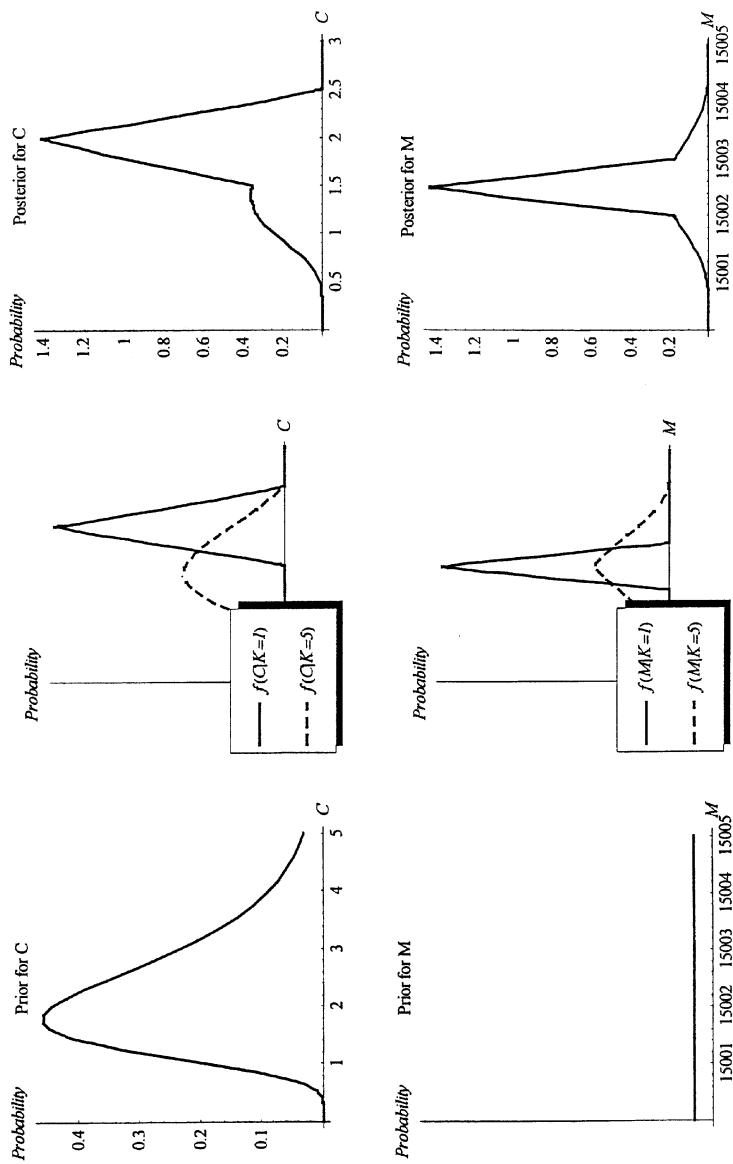


Fig. 4. Prior and posterior probability distributions implied by estimated model. The graphs below depict priors and posteriors for the estimated model. The top row corresponds to the quote exposure cost ( $C$ ); the bottom row to the efficient price ( $M$ ). Leftmost graphs reflect agent beliefs before observing the bid and ask. (The improper prior for  $M$  is depicted.) The remaining graphs are conditional on observing  $bid = 15,000$  and  $ask = 15,005$ . Middle graphs reflect the posteriors conditional on no clustering ( $K = 1$ ) and clustering ( $K = 5$ ). Rightmost graphs are the posteriors.

Table 4  
Comparison of alternative models

	Model:			
	A: asymmetric rounding; no clustering	B: symmetric rounding; no clustering	C: asymmetric rounding, clustering	D: symmetric rounding, clustering
Log likelihood	– 2000.87	– 2001.10	– 1968.62	– 1949.33
Number of parameters	3	3	4	

Note: Log likelihoods for four alternative models of bid and ask quotes, estimated for \$/DM quotes as of noon (GMT) for all trading days in 1996 (257 observations). Log likelihoods are estimated at modal parameter values using an auxiliary particle filter

As this last finding runs counter to the economic hypotheses underlying the model, it deserves fuller illumination. Fig. 5 presents the spread distributions implied by the four models, along with that of the actual data (from Fig. 1). All models tend to overpredict the incidence of large (above ten-pip) spreads. This suggests that the lognorral distribution assumed for the underlying exposure costs is too fat-tailed. Manrique and Shephard (1997) report a similar finding for the spread on an NYSE stock.

Neither of the nonclustered models (*A* and *B*) closely capture the peaks at five and ten pips. Both clustered models (*C* and *D*) perform better in this respect, but the model *D* (symmetric rounding) achieves a better fit at the five-pip peak. This is apparently driving the likelihood magnitudes.

From an economic viewpoint, symmetric rounding of the quotes has two drawbacks. The first is that a bid or ask formed in this fashion might be associated with an expected loss. The quotes in this market, however, are merely indicative, not binding. The market maker may refuse to honor a previously posted quote, albeit at some cost in reputation. The second drawback is that with symmetric rounding, the discrete bid and ask may be equal, a state which is in most senses incompatible with most microstructure models.<sup>10</sup> This problem may be remedied by positing some sort of censoring process. For example, the observed bid and ask may be viewed as arising from a probability model in which, once  $M$  is determined, a market maker is drawn from a large population. If her value of  $C$  is such that her quotes would cross, she declines to participate. Market makers are drawn until a set of noncrossed quotes arises.

<sup>10</sup> In actual markets, crossed quotes can arise from communication failures, credit problems (one side refuses to trade with the other) and pricing conventions (e.g., only the aggressor in a trade ‘pays’ a commission), among other things.

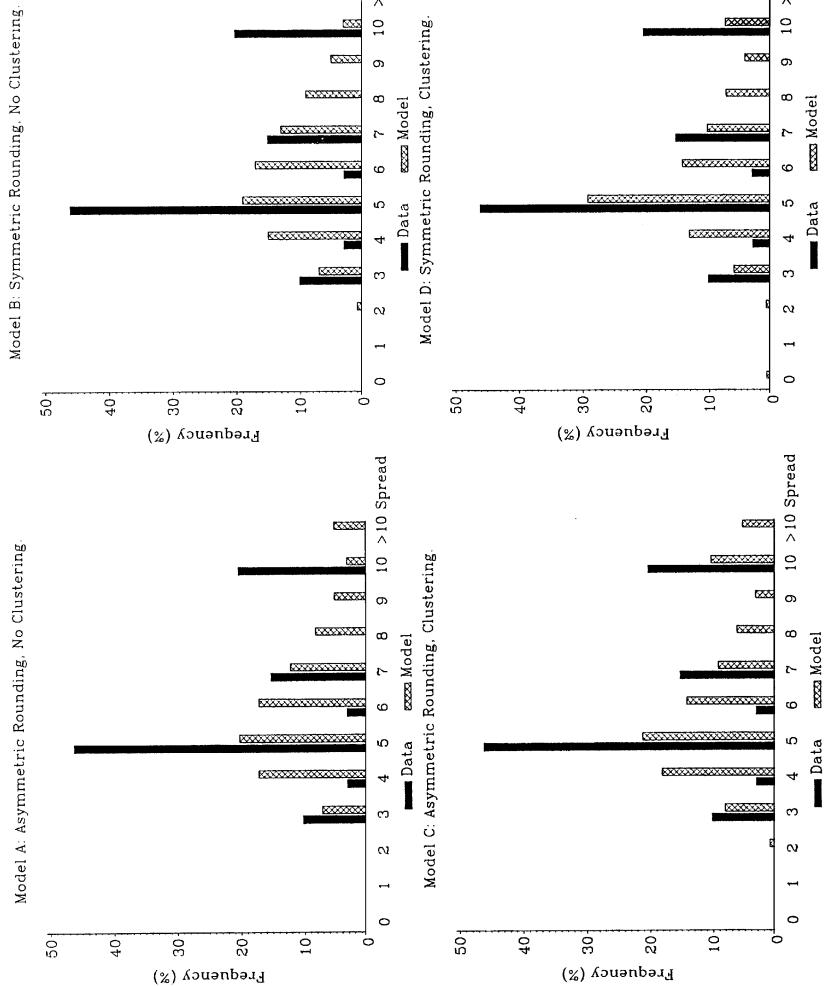


Fig. 5. Spread distributions: actual and implied by alternative models. (Note: distributions of bid–ask spreads predicted by each of four alternative models, and the actual distribution of the spreads in the sample (\$/DM exchange rates as of noon, GMT, 1996; 257 observations).

### 5.3. Intraday patterns

To investigate time-of-day patterns in market behavior, Model C (clustering, asymmetric rounding) was estimated allowing for different parameter values for each of the forty-eight half-hour periods. The estimated parameter values (means) for each period are plotted against time-of-day in Fig. 6. (Active trading hours in this market are roughly 8:00 to 20:00, with London dominating in the early hours and New York in the later times.)

The efficient price volatility parameter  $\sigma_u^2$  exhibits the usual elevation during active trading. The spread parameters also change markedly during the trading day. Away from the active hours, the mean cost ( $\mu_c$ ) is lower; the cost volatility ( $\sigma_c^2$ ) is higher; and the clustering probability  $k$  is higher. The latter finding suggests that more activity (and competition) among market makers leads to more aggressive (and less clustered) quoting practices. The intraday negative

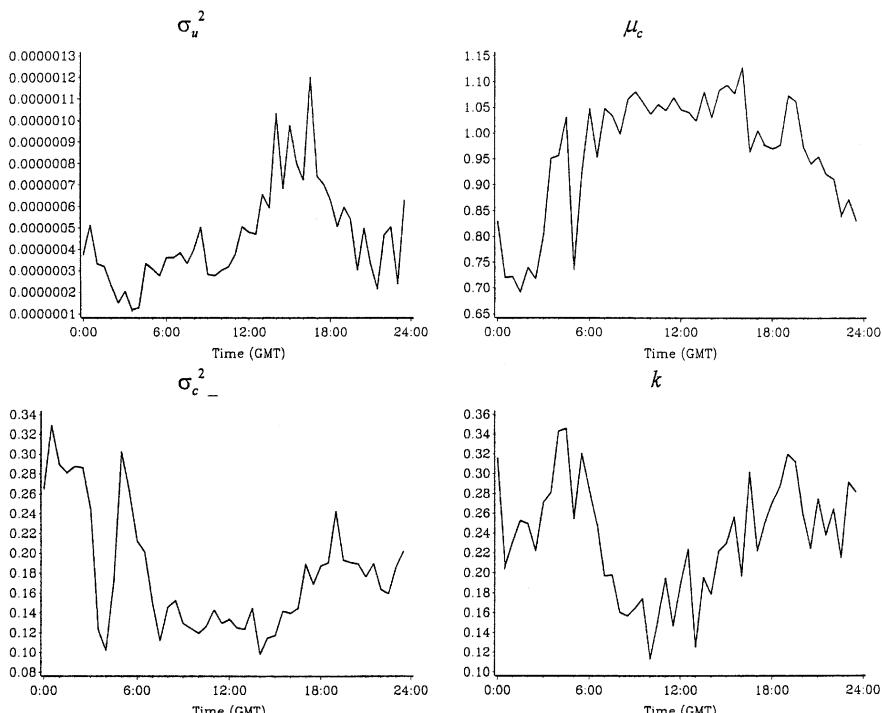


Fig. 6. Intraday patterns in model parameters. (Note: Results for a model with symmetric rounding of the *bid* and *ask* quotes and clustering, estimated for *bid* and *ask* quotes in the \$/DM market at half-hourly intervals for 1996.  $\sigma_u^2$  is the (half-hourly) variance of the efficient price innovations;  $\mu_c$  and  $\sigma_c^2$  are the mean and variance of the (log) quote cost;  $k$  is the clustering intensity).

relation between clustering and volatility is consistent with that found by Harris (1991) and Ball et al. (1985) in other markets.

The simple model's suitability over horizons as short or shorter than a half-hour, however, is open to question. Most studies of security returns suggest that deviations from normality are more pronounced at higher frequencies. It is also more likely that the cost variable would exhibit serial correlation at higher frequencies. Approaches to these and other complications are discussed below.

## 6. Extensions

The discussion in this section turns to refinements that may make the simple model implemented here more useful and appealing. In general, the easiest extensions involve deterministic patterns in the parameters (e.g., the intraday analysis presented above). These generally require minimal modifications to the parameter and state simulations. It is also relatively easy incorporate additional sources of randomness that are serially independent. These will usually involve only an additional step in the Gibbs sampler.

Extensions that incorporate serial dependencies in the state variables are more challenging. If these dependencies can be modeled as transitions of Gaussian state variables, Shephard (1994) among others suggests that simulation is still quite practical. It is important to emphasize that the model need not be fully Gaussian, only conditionally so. The framework for dealing with limited dependent variable processes suggested by Manrique (1997) and Manrique and Shephard (1998) are particularly useful in this respect.

Perhaps most importantly, the MCMC approach does not suffer from the curse of dimensionality. Incorporation of additional state variables or parameters requires only that these quantities be simulated in each sweep. The incremental computational burden of this simulation does not (to a first approximation) depend on the number of variables or parameters already present in the model.

### 6.1. Time-varying volatility

The intraday analysis discussed above allowed for deterministic time-of-day effects in  $\sigma_u^2$ . Volatility that changes in a random fashion is a little more involved. There are two main approaches here. In ARCH-family models,  $\sigma_{u,t}^2$  depends solely on prior disturbances  $u_t$ . In stochastic volatility models,  $\sigma_{u,t}^2$  follows a random process with increments that are independent of  $u_t$  at all leads and lags. ARCH-type models predominate in the empirical literature, probably in part due to the ease of estimation. Stochastic volatility models predominate in the MCMC literature, however, at least in part for the reason that the MCMC paradigm provides a computationally efficient likelihood-based approach to

estimation.<sup>11</sup> Either ARCH or stochastic volatility extensions could be implemented here. These are obvious and desirable refinements.

### 6.2. *Time-varying and serially correlated trading costs*

In many markets discrete spreads are found to exhibit periodicities at daily frequencies. Spreads are typically elevated at the start and end of trading sessions. This may be incorporated (as in the intraday analysis, above) by allowing the quote exposure cost parameters to be indexed by time-of-day. It is also reasonable (in high-frequency data) to allow the cost to have persistent stochastic components (arising perhaps from persistence in volatility or asymmetric information intensity). Hasbrouck (1998) suggests an autoregressive dependence. Manrique and Shephard (1997) implement this (and asymmetric bid and ask costs) in an MCMC framework.

### 6.3. *Asymmetric bid and ask costs*

In the present model, the quote exposure cost is the same on both the bid and ask sides of the market. This is a reasonable assumption if the same quote-setter is active on both sides of the market, and if the quotes are not sensitive to current inventory positions. In many markets, however, the bid and ask quotes may reflect different individual limit order traders, presumably subject to different trading costs. Hasbrouck (1998) captures this by modeling trading costs as two processes that evolve independently of each other.

### 6.4. *Alternative clustering models*

The implicit tick size in the simple model is a Bernoulli random variable. In some applications, a more refined multinomial specification may be desirable. Harris (1991), for example, suggests that the implicit tick size may be a function of volatility and trading costs.

## 7. Conclusions

Structural models of security prices over short intervals usually reflect micro-structure effects arising from the manner in which the security is traded. The model in this paper features stochastic costs of market-making. Discreteness in

---

<sup>11</sup> Exceptions to this alignment in the literature include Kim et al. (1998), who estimate a GARCH model in an MCMC framework, and Fridman and Harris (1998), who estimate a stochastic volatility model using a numerical-integration-based likelihood approach.

the quotes is enforced by an explicit rounding mechanism. Clustering in the quotes arises from a random implicit tick convention. Despite the simplicity of the model, however, there are no convenient closed-form representations for moments or conditional probabilities. Inference using minimum distance estimators (maximum likelihood or method of moments) is therefore beset with computational difficulties. Inference in a Bayesian framework, however, is greatly simplified by the Gibbs sampler.

An application to \$/DM exchange rate bid and ask quotes illustrates the usefulness of the method. Analysis of alternative specifications supports the superiority of a model that allows for clustering. Nevertheless, the clustering in the quote data appears to be even more pronounced than the present models are capable of capturing. A more surprising finding concerns the underlying rounding process. Economic considerations suggest that the quotes should be rounded asymmetrically, i.e., that the bid should be rounded down and the ask should be rounded up (to preclude expected market making losses). In these data, however, a clustered model in which the (implied continuous) bid and ask quotes are simply rounded to the nearest tick outperforms a model in which the rounding is asymmetric. For none of the models investigated here, however, does the implied spread distribution closely match that found in the data.

The models and estimation approaches discussed here do not by any means establish the limits of what is technically and computationally feasible. Markov chain Monte Carlo procedures (including the Gibbs sampler used here) tend to grow but linearly in complexity and computational demands with the number of latent state variables and parameters. The procedures are, furthermore, modular. For example, once a given set of implicit efficient prices ( $M_t$ ) has been simulated in the framework of the present paper, these prices could be used as the input to an MCMC stochastic volatility estimation. These approaches therefore open the door to the investigation of broad classes of structural, and (ideally) more realistic models of price evolution.

## References

- Ahn, H.J., Cao, C.Q., Choe, H., 1996. Tick size, spread and volume. *Journal of Financial Intermediation* 5, 2–22.
- Angel, J.J., 1997. Tick size, share prices, and stock splits. *Journal of Finance* 52, 655–681.
- Anshuman, V.R., Kalay, A., 1998. Market-making with discrete prices. *Review of Financial Studies* 11, 81–109.
- Ball, C.A., 1988. Estimation bias introduced by discrete security prices. *Journal of Finance* 43, 841–865.
- Ball, C.A., Torus, W.A., Tschoegl, A.E., 1985. The degree of price resolution: the case of the gold market. *Journal of Futures Markets* 5, 29–43.
- Bernhardt, D., Hughson, E., 1996. Discrete pricing and the design of dealership markets. *Journal of Economic Theory* 71, 148–182.

- Bessembinder, H., 1994. Bid–ask spreads in the interbank foreign exchange markets. *Journal of Financial Economics* 35, 317–348.
- Bollerslev, T., Melvin, M., 1994. Bid–ask spreads and volatility in the foreign exchange market: an empirical analysis. *Journal of International Economics* 36, 355–372.
- Brown, S., Laux, P., Schachter, B., 1991. On the existence of an optimal tick size. *Review of Futures Markets* 10, 50–72.
- Carlin, B.P., Louis, T.A., 1996. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, London.
- Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. *Biometrika* 81, 541–553.
- Casella, G., George, E.I., 1992. Explaining the Gibbs sampler. *The American Statistician* 46, 167–190.
- Chan, K.C., Christie, W.G., Schultz, P.H., 1995. Market structure and the intraday pattern of bid–ask spreads for NASDAQ securities. *Journal of Business* 68, 35–60.
- Chib, S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* 12, 409–431.
- Chordia, T., Subrahmanyam, A., 1995. Market making, the tick size, and payment-for-order flow: theory and evidence. *Journal of Business* 68, 543–575.
- Christie, W.G., Harris, J.H., Schultz, P.H., 1994. Why did NASDAQ market makers stop avoiding odd-eighth quotes? *Journal of Finance* 49, 1841–1860.
- Christie, W.G., Schultz, P.H., 1994. Why do NASDAQ market makers avoid odd-eighth quotes? *Journal of Finance* 49, 1813–1840.
- Cordella, T., Foucault, T., 1996. Minimum price variations, time priority and quote dynamics, Working Paper. Universtat Pompeu Fabra, Catalonia, Spain.
- De Jong, P., Shephard, N., 1995. The simulation smoother for time series models. *Biometrika* 82, 229–350.
- Doornik, J.A., 1996. *Object-Oriented Matrix Programming using OX*. International Thomson Business Press, London.
- David, A.R., 1991. Effects of bid–ask spreads and price discreteness on stock returns. Working Paper. Rodney L. White Center, Wharton School.
- Dutta, P.K., Madhavan, A.M., 1997. Competition and collusion in dealer markets. *Journal of Finance* 52, 245–276.
- Fridman, M., Harris, L.E., 1998. A maximum likelihood approach for stochastic volatility models. *Journal of Business and Economic Statistics*, forthcoming.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Introducing Markov Chain Monte Carlo. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
- Glosten, L.R., 1994. Is the electronic open limit order book inevitable? *Journal of Finance* 49, 1127–1161.
- Glosten, L.R., Harris, L.E., 1988. Estimating the components of the bid/ask spread. *Journal of Financial Economics* 21, 123–142.
- Goodhart, C., Ito, T., Payne, R., 1996. One Day in June 1993: a Study of the Working of the Reuters 2000–2 Electronic Foreign Exchange Trading System. In: Frankel, J.A., Galli, G., Giovannini, A. (Eds.), *The Microstructure of Foreign Exchange Markets*. University of Chicago Press, Chicago.
- Gottlieb, G., Kalay, A., 1985. Implications of the discreteness of observed stock prices. *Journal of Finance* 40, 135–153.
- Harris, L.E., 1990. Estimation of stock price variances and serial covariances from discrete observations. *Journal of Financial and Quantitative Analysis* 25, 291–306.
- 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., 1997. Decimalization: a review of the arguments and evidence. Working Paper. University of Southern California.

- Hasbrouck, J., 1996. Modeling microstructure time series. In: Maddala, G.S., Rao, C.R. (Eds.), *Handbook of Statistics 14: Statistical Methods in Finance*. North Holland, Amsterdam, Elsevier.
- Hasbrouck, J., 1998. The dynamics of discrete bid and ask quotes. *Journal of Finance*, forthcoming.
- Hausman, J.A., Lo, A.W., MacKinlay, A.C., 1992. An ordered probit analysis of transaction stock prices. *Journal of Financial Economics* 31, 319–379.
- Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12, 371–389.
- Kandel, E., Marx, L.M., 1997. NASDAQ market structure and spread patterns. *Journal of Financial Economics* 45, 61–89.
- Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, forthcoming.
- Kitagawa, G., 1987. Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association* 82, 1032–1041.
- Lyons, R.K., 1995. Tests of microstructural hypotheses in the foreign exchange market. *Journal of Financial Economics* 39, 321–351.
- Lyons, R.K., 1996. Optimal transparency in a dealer market with an application to foreign exchange. *Journal of Financial Intermediation* 5, 225–254.
- Manrique, A., 1997. Econometric analysis of limited dependent time series. Ph.D. Dissertation, Nuffield College, Oxford University.
- Manrique, A., Shephard, N., 1997. Likelihood analysis of a discrete bid/ask price model for a common stock. Working Paper. Nuffield College, Oxford University.
- Manrique, A., Shephard, N., 1998. Likelihood inference for limited dependent processes. *Econometrics Journal*.
- O'Hara, M., 1995. *Market Microstructure Theory*. Blackwell, Cambridge, USA.
- Peiers, B., 1997. Informed traders, intervention and price leadership: a deeper view of the microstructure of the foreign exchange market. *Journal of Finance* 52, 1589–1614.
- Press, S.J., 1989. *Bayesian Statistics: Principles, Models, and Applications*. John Wiley, New York.
- Rosenblatt, M., 1952. Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23, 470–472.
- Schelling, T.C., 1960. *The Strategy of Conflict*. Harvard University Press, Cambridge, MA.
- Schwert, G.W., 1997. Symposium on market microstructure: focus on NASDAQ. *Journal of Financial Economics* 45, 1–8.
- Shephard, N., 1994. Partial non-Gaussian state space. *Biometrika* 81 (1), 115–131.
- Shephard, N., Pitt, M.K., 1997a. Filtering via simulation: auxiliary particle filters. Working Paper. Nuffield College, Oxford University.
- Shephard, N., Pitt, M.K., 1997b. Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 83, 653–667.
- Smith, A.F.M., Gelfand, A.E., 1992. Bayesian statistics without tears: a sampling–resampling perspective. *The American Statistician* 46, 84–88.
- Tanner, M.A., 1996. *Tools for Statistical Inference*. 3rd ed., Springer, New York.
- U.S. Securities and Exchange Commission. 1996. Report Pursuant to Section 21(a) of the Securities Exchange Act of 1934 Regarding NASD and the NASDAQ Market. Web page. Available at [www.sec.gov](http://www.sec.gov).
- Yao, J., 1997. Market making in the interbank foreign exchange market. Working Paper. Stern School of Business, New York University.