THE JOURNAL OF FINANCE • VOL. LVI, NO. 5 • OCT. 2001

True Spreads and Equilibrium Prices

CLIFFORD A. BALL and TARUN CHORDIA*

ABSTRACT

Stocks and other financial assets are traded at prices that lie on a fixed grid determined by the minimum tick size. Observed prices and quoted spreads do not correspond to the equilibrium prices and true spreads that would exist in a market with no minimum tick size. Using Monte Carlo Markov Chain methods, this paper estimates the equilibrium prices and true spreads. For large stocks, most of the quoted spread is attributable to the rounding of prices and the adverse selection component is small. The true spread and the adverse selection component are greater for mid-sized stocks.

As a wedge between what buyers pay and what sellers receive, the bid-ask spread has long interested students of transaction costs. Apart from order processing costs, the market microstructure literature has focused on two main components of the spread: the inventory and the adverse selection costs of trading. Demsetz (1968), Stoll (1978), and Ho and Stoll (1981, 1983) model the inventory holding costs of market makers whereas Copeland and Galai (1983), Glosten and Milgrom (1985), Kyle (1985), and Easley and O'Hara (1987) model the adverse selection costs. A number of empirical models measure the components of the spread. Roll (1984), Stoll (1989), and George, Kaul, and Nimalendran (1991) make inferences about the spread from the serial covariance properties of transaction prices. Glosten and Harris (1988), Madhavan and Smidt (1991), Huang and Stoll (1997), and Madhavan, Richardson, and Roomans (1997) use a trade indicator model to make inferences about the spread. Hasbrouck (1988, 1991, 1993) models the time series of quotes and trades in a vector autoregression framework. Most of the above models ignore the institutional feature of discretization. In fact, discreteness has often been treated as something to be addressed while examining other

* Clifford A. Ball is from the Owen School of Management, Vanderbilt University and Tarun Chordia is from the Goizueta Business School, Emory University. We thank Hank Bessembinder, Charles Cao, Siddhartha Chib, Bill Christie, Rob Engle, Larry Harris, Joel Hasbrouck, Roger Huang, Bruce Lehmann, Craig Lewis, Marc Lipson, Ron Masulis, Venkatesh Panchpagesan, Paul Schultz, George Sofianos, Matt Spiegel, Hans Stoll, George Tauchen, Dan Weaver, and seminar participants at Emory University, UCLA, Vanderbilt University, Washington University, the NBER market microstructure program meeting, and the WFA meetings for helpful comments and suggestions. We especially thank Siddhartha Chib and John Geweke for providing the Bayesian Analysis, Computation and Communication subroutines. We are grateful to René Stulz (editor) and an anonymous referee whose comments and suggestions have greatly improved the paper. All errors are our own.

hypotheses. It is only recently (Hausman, Lo, and MacKinlay (1992), Manrique and Shephard (1997), Hasbrouck (1999a, 1999b)) that the empirical market microstructure literature has emphasized the effects of discretization.

Nonzero price changes of financial securities have a lower bound called the tick size. Observed prices and quoted spreads are constrained by the tick to lie on a discrete grid. Ball, Torous, and Tschoegl (1985), Harris (1991, 1994), and Chordia and Subrahmanyam (1995) argue that this constraining impacts trading activity. Ball (1988) and Cho and Frees (1988) examine the model estimation bias induced by the rounding of prices. In this paper, we explicitly take into account the rounding of prices and quoted spreads onto a discrete grid, in modeling the dynamic behavior of the spreads. Our model, which is a generalization of Huang and Stoll (1997), allows us to separately estimate the adverse selection component of the spread and the sum of the inventory and order processing components. It also allows us to separate out these costs from the pure effect of the discrete tick size. For the largest stocks, the effect of discretization is shown to be larger than all the other components combined. This result has, heretofore, not been recognized in the literature.

The observed transaction price is a discretization of the sum of (1) the true unobserved, continuous price that evolves as a random walk subject to information shocks and the impact of adverse selection in trades, and (2) the component of the spread due to inventory and order processing costs. Thus, over short horizons, the observed price is a discrete version of the sum of a permanent informational component and the transient components arising out of the trading mechanism. The true spread, which is the continuous spread that would exist in the absence of the tick, is modeled as a transform of a Gaussian autoregressive process with additional structural variables associated with the time of day, the time between trades, and with the size and the depth of the prior trade. Consistent with earlier work, the quoted ask is the true ask rounded up to the nearest grid point and the quoted bid is the true bid rounded down.[1] The resultant model may be represented in a bivariate state space form but the rounding destroys the Gaussian structure and the time series independence of errors, rendering the Kalman filter estimation method inapplicable. We use the Monte Carlo Markov Chain (MCMC) methodology to handle the problems of a non-Gaussian state space model.

Hasbrouck (1999a, 1999b) and Manrique and Shephard (1997) also use the MCMC methodology to estimate a model of bid and ask quotes. These papers model the equilibrium price as a random walk with *independent* bid and ask exposure costs. In contrast, our paper develops a bivariate model of the true spreads and equilibrium prices. In this way, we integrate the responses to information shocks into the spread and price variables simultaneously. Further, we develop a structural model for the bid-ask spread that captures the known regularities in the data.

[1] For example, see Dravid (1991) and Hasbrouck (1999a, 1999b).

We implement the above technology on transactions data for a sample of seven large stocks traded on the New York Stock Exchange during February 1997 and during November 1997, that is, before and after the reduction in tick size from 12.5 cents to 6.25 cents in June 1997. This design allows us to assess the relative effects of discreteness and the robustness of our methodology across two different tick size regimes. Clearly, market pressures and perceived low true spreads forced the tick size lower. The analysis of true spreads within the lower tick environment enables us to assess whether further tick size reduction is advisable.

The strength of our conclusions hinges on adequate model specification. We generate simulation diagnostics that indicate that the sampling is working efficiently. It appears that the MCMC method is well suited to our application. We also ran the same analysis on a selection of midsized stocks for the last quarter of 1997. In this way, we can check whether the results are consistent over different classes of stocks and over varying lengths of time series. Competing models with more general error assumptions were also analyzed. The details are described in the robustness subsection of the empirical section. The results are startling and consistent.

At the transaction level, the minimum tick size is large compared to the noise in the underlying equilibrium price changes. The standard deviation of the increment to the equilibrium price, for the large stocks, is less than 6.2 cents in February 1997 and less than 5 cents in November 1997. This is in comparison to the prevailing minimum tick size of 12.5 cents in February 1997 and 6.25 cents in November 1997. The discrete grid is coarse relative to price changes and the tick size places binding constraints on prices and quoted spreads. This is especially true for our sample of large stocks where, due to the large number of transactions, equilibrium price changes between transactions are small.

In February 1997, for the large stocks, the median true spread that would prevail in the absence of a tick size varies from 11 percent to 24 percent of the actual quoted spread and the adverse selection component of this inferred spread varies from 70 percent to 95 percent. In other words, for an average quoted spread of (say) 20 cents, the true spread varies from 2 cents to 6 cents and the adverse selection component of this spread varies from 1.6 to 4.4 cents. The inventory and order processing costs vary from about 0.10 to 1.6 cents.[2] Concentrating on the large stocks around the change in the tick size, we see that the ratio of true spreads to price is practically the same before and after the change in the tick size. However, the quoted spreads are significantly larger than spreads that would prevail without discreteness. In fact, the results strongly suggest that the effect of rounding on quoted spreads is larger than all the other components of the spread combined.

[2] Similar results obtain for November 1997 and for a sample of large stocks in the first quarter of 1992.

Our results differ from the existing literature in terms of inventory and order processing costs. For example, using a sample of Nasdaq stocks in October, November, and December 1984, Stoll (1989) estimates that the adverse information costs account for about 43 percent of the quoted spread with the remaining 57 percent being attributed to inventory holding and order processing costs. Furthermore, for a sample of 20 Major Market Index stocks in 1992, Huang and Stoll (1997) find an average traded spread of 12.2 cents. Their estimate of the adverse selection and inventory holding costs is an average of 1.4 cents with the remaining 10.8 cents due to order processing costs. Additionally, for a sample of 274 stocks in 1990, Madhavan, Richardson, and Roomans (1997) find adverse selection costs of about 3.1 cents and order processing costs of 4.2 cents. For a sample of 1,544 stocks during March, April, and May of 1993, Cao, Choe, and Hatheway (1997) find adverse selection costs of 3.8 cents and order processing costs of about 10.4 cents. Given variation in the estimated adverse selection costs in the literature, our estimates of these costs are fairly comparable. However, we differ significantly in our estimation of the inventory and order processing costs because we explicitly account for rounding. In fact, using the Huang and Stoll (1997) data, we find an average true spread of about 2.8 cents compared to their 12.2 cents. The prior literature, since it does not explicitly allow for discretization, captures the effect of rounding as a part of the order processing costs and thereby, overstates these costs.

Our results indicate that markets for large stocks are informationally efficient in the sense that the adverse selection component is small and, furthermore, that the biggest component of the quoted spread is the adjustment onto the discrete grid. The fact that the tick size is so large and the adverse selection cost is so low, while new, should not be surprising. The quoted spreads are made wide by the minimum tick size and the institutionally mandated tick size may produce excess market maker profits.[3] These excess profits have presumably encouraged the practices of payment-for-order-flow and internalization. In fact, the move by both NYSE and Nasdaq to decimal trading, the existence of payment-for-order-flow and internalization, and the proliferation of Electronic Communication Networks (ECNs) are all consistent with our story of large tick sizes resulting in large profits being made by market makers. The policy implication is that the tick size for large, heavily traded stocks should be dramatically reduced.

The rest of the paper is organized as follows. The next section develops the model. The transaction data sources and the required filtering are discussed in Section II. Section III presents the MCMC approach. Section IV discusses model selection and the diagnostics. The results are presented in Section V and Section VI concludes.

[3] One measure of these excess profits is the price of seats on the NYSE.

I. Model

This section describes the asymmetric information spread model. It explicitly takes into account the effect of institutional features such as the tick size and the discretization of prices and spreads. Our model is a generalization of Huang and Stoll (1997) in three important respects:

1. We explicitly allow for rounding onto the tick grid. This important feature of the data is ignored by Huang and Stoll (1997). This is important because previous researchers[4] have modeled the first differences of the observed price changes as an autoregressive process with Gaussian errors. Due to discreteness, the first difference of the price changes cannot have Gaussian errors and it is possible that this misspecification results in erroneous conclusions. The ordered probit model of Hausman et al. (1992) and more recently the models of Hasbrouck (1999a, 1999b) and Manrique and Shephard (1997) address this misspecification problem in the literature.
2. We allow the true spread to vary from transaction to transaction. In their basic model, Huang and Stoll (1997) assume a constant spread. Clearly, the spread, in the absence of a minimum tick, will respond to the information in the order flow.
3. We allow the prior trade size and market depth to impact the spreads in a continuous fashion. Huang and Stoll (1997) use dummy variables for three trade size regimes: small (1,000 shares or less), medium (1,000 to 10,000 shares), and large (10,000 shares or more). The true spread in our model is also a function of the time between trades and the intraday seasonalities. Our model choice is determined by the marginal likelihood derived from the Gibbs sampling output as suggested by Chib (1995).

The observed price, P_t is modeled as follows:

$$P_t \equiv [p_t^{NR}]_{Round} = [m_t + (1-\lambda)s_t Q_t/2]_{Round}, \qquad (1)$$

where p_t^{NR} is the nonrounded price; m_t is the true price of the security at time t, immediately after a trade; Q_t is a trade indicator for buyer/seller classification of trades and is $+1$ if the trade is buyer initiated, -1 if the trade is seller initiated, and 0 if we are unable to sign the trade; λ is the adverse selection component of the spread; and s_t is the spread that would obtain in a market with prices quoted on a continuous scale. The notation $[.]_{Round}$ indicates rounding onto the tick grid. Thus, the observed price is a rounding or a discretization of the sum of the true price and the inventory and order processing cost component of the spread. Note, that in the pres-

[4] See also Brennan and Subrahmanyam (1995).

ence of rounding the disturbances in observed price changes are not Gaussian. Most market microstructure models ignore rounding and, thus, are unlikely to be correctly specified, especially if the rounding is severe.

The true price, m_t, is assumed to evolve as follows:

$$m_t = m_{t-1} + \lambda \frac{s_t Q_t}{2} + u_t, \tag{2}$$

where $\{u_t, t = 1,2\ldots T\}$ are *i.i.d.* $N(0,\sigma_u^2)$ and represent information shocks. The second term in equation (2) is the fraction of the half spread attributable to adverse selection. The parameter λ is interpreted as measuring the adverse selection impact on the true price.[5]

In a world without ticks, let a_t (b_t) be the true ask (bid) price and so $s_t = a_t - b_t$. Let A_t (B_t) be the quoted ask (bid). For the market maker's trading profits to be nonnegative, the quoted ask, A_t, cannot be less than true ask, a_t, and the quoted bid, B_t, cannot be greater than the true bid, b_t. Thus, the quoted ask A_t is the true ask a_t rounded up to the nearest tick and the quoted bid B_t is the true bid b_t rounded down to the nearest tick.

The rounding rule is economically sound for several reasons. The specialist faces strong competition from a number of liquidity providers including the floor brokers, the limit order book, other exchanges, and off-floor market makers. Internalization and payment-for-order flow have led to a situation where any large spread orders would be sucked away from the floor of the NYSE. In fact, the specialist participates in less than 20 percent (10 percent for the largest stocks) of the trades even though he has privileged information about the limit order book. Given the competitive environment, it is extremely unlikely that any one of the above liquidity providers could sustain losses on some trades with the hope of more than recovery on other trades. Additionally, for the largest stocks that we consider, the tick size is binding in a large fraction of the transactions. For instance, for GE in November 1997, the tick size is binding for over 63 percent of the quotes. Furthermore, due to transactions within the spread, the tick size is binding for over 73 percent of the trades. The fact that the tick size is binding and that internalization and payment-for-order-flow occur mainly in the large stocks strongly suggests that the true spread is far smaller than the tick size and that our rounding rule is justified. Finally, we note that there is a precedence in the literature for the rounding rule that we adopt. Dravid (1991), Manrique and Shephard (1997), and Hasbrouck (1999a, 1999b) use the same rounding rule, and other rounding rules, such as randomization, could lead to zero or negative spreads that would be economically indefensible.

[5] While λ is assumed to be constant in the model, it is likely that company-specific or industry announcements may create shocks. However, given the huge number of transactions involved, we believe that such an effect is likely to be minimal.

The observed spread is $S_t = A_t - B_t$. Since $A_t \geq a_t$ and $B_t \leq b_t$, the true spread cannot be larger than the quoted spread. The true spread will not be constant; rather it will vary according to a stochastic process adjusting to micro information flows and possibly responding to large information-laden trades. We model the true spread, s_t, as a first-order logarithmic autoregression with additional structural variables as follows:

$$\ln(s_t) = \alpha + \beta \ln(s_{t-1}) + \delta \ln \frac{V_{t-1}}{D_{t-1}} + \tau TIME_{t-1} + d_1 BEG_t + d_2 END_t + e_t, \tag{3}$$

where $\{e_t, t = 1,2\ldots T\}$ are *i.i.d.* $N(0,\sigma_e^2)$, V_{t-1} is the volume of stock transacted at the previous trade, and D_{t-1} is the corresponding depth of trade for which the then prevailing bid-ask quote held. $TIME_{t-1}$ is the time, in seconds, between the last trade and the one before it and *BEG* (*END*) represent dummies to denote the first (last) hour of the trading day.

We allow the relative size of trade to depth on the previous transaction to possibly impact the ensuing spread at the current transaction.[6] The dummy variables capture the intraday seasonalities and the use of the lagged time between trades is motivated by Easley and O'Hara (1992), who suggest that the absence of trade may provide information about the occurrence of information events.[7] Our final choice of the variables in (3) will be driven by a comparison of the marginal likelihoods across different functional forms. Note that, due to rounding, the quoted spread cannot be modeled as an autoregressive process with Gaussian errors. However, the (log) true spread lies on the real line and may be modeled as in equation (3).

Lastly, for algebraic convenience define

$$x_t = m_t - \lambda s_t Q_t/2. \tag{4}$$

Combining the above equations we have the following econometric model:

$$P_t \equiv [p_t^{NR}]_{Round} = [x_t + s_t Q_t/2]_{Round}, \tag{5}$$

$$x_t = x_{t-1} + \lambda s_{t-1} Q_{t-1}/2 + \sigma_u \eta_{1t}, \tag{6}$$

$$\gamma_t = \alpha + \beta\gamma_{t-1} + \delta \ln \frac{V_{t-1}}{D_{t-1}} + \tau TIME_{t-1} + d_1 BEG_t + d_2 END_t + \sigma_e \eta_{2t}, \tag{7}$$

[6] Using just the prior trade (and not the depth) did not change the underlying conclusions of the paper that the tick size is too large and the true spreads too small when compared to the quoted spreads. However, we choose to use the ratio of trade size to depth since Chordia, Roll, and Subrahmanyam (2001) show that the depth had decreased in July 1997 when the tick size was reduced from 12.5 cents to 6.25 cents.

[7] See also the autoregressive conditional duration model of Engle and Russell (1998), where they study the dynamics of the spacing of financial market activities.

where $\gamma_t \equiv \ln(s_t)$ and $\{\eta_{1t}, \eta_{2t},\ t = 1,2\ldots T\}$ are *i.i.d.* bivariate standard normals with correlation ρ. The adjusted equilibrium price, x_t, includes a spread component, s_t, so we allow correlation between the errors in our model.

We adhere to the following convention for the discretization process:[8]

1. If the observed price, P_t, is at the ask (bid) then we assume that the nonrounded price, p_t^{NR}, has been rounded up (down) to the nearest tick. Furthermore, the bid (ask) price is assumed to have been rounded down (up). Thus, for a trade at the ask, $x_t + (s_t/2) \in [P_t - tick, P_t]$ and $x_t - (s_t/2) \in [B_t, B_t + tick]$. Similarly, for a trade at the bid, $x_t - (s_t/2) \in [P_t, P_t + tick]$ and $x_t + (s_t/2) \in [A_t - tick, A_t]$.
2. If the trade is a customer buy, $Q_t = +1$, the price is not the same as the ask, $P_t \neq A_t$, and if the effective spread $P_t - B_t \neq tick$, then the nonrounded price is assumed to have been rounded up, that is, $x_t + (s_t/2) \in [P_t - tick, P_t]$. On the other hand, if $Q_t = +1$, $P_t \neq A_t$, and $P_t - B_t = tick$, then $x_t + (s_t/2) \in [P_t - tick, P_t]$ and $x_t - (s_t/2) \in [B_t, B_t + tick]$. In other words, for trades not at the ask and when the quotes are larger than a tick, we use information from one side of the market only. Using information from both sides of the market did not change the results.
3. If the trade is a customer sell, $Q_t = -1$, the price is not the same as the bid, $P_t \neq B_t$, and if the effective spread is $A_t - P_t \neq tick$, then the nonrounded price is assumed to have been rounded down, that is, $x_t - (s_t/2) \in [P_t, P_t + tick]$. On the other hand, if $Q_t = -1$, $P_t \neq B_t$, and $A_t - P_t = tick$, then $x_t - (s_t/2) \in [P_t, P_t + tick]$ and $x_t + (s_t/2) \in [A_t - tick, A_t]$. Once again, for trades not at the bid and when the quotes are larger than a tick, we use information from one side of the market only.
4. If the direction of the trade is indeterminable, $Q_t = 0$, then the nonrounded price is assumed to have been rounded around the observed price, that is, $x_t \in [P_t - tick/2, P_t + tick/2]$.

Thus, at each time point t, we have the following information:

$$\begin{aligned} x_t + s_t/2 &\in I_{1,t}, \\ x_t - s_t/2 &\in I_{2,t}, \end{aligned} \tag{8}$$

where, as shown in the discussion of the discretization process above, $I_{1,t}$ and $I_{2,t}$ indicate the intervals, of length the tick size, that each linear functional of the state variable must lie within. In other words, the observed information places the adjusted equilibrium price plus the half-spread in one interval of length tick size and places the adjusted equilibrium price

[8] We have checked for robustness of results by changing some of the discretization assumptions. For instance, all trades with $Q_t = 0$ were eliminated from the sample, and in another instance, all trades not at the bid or the ask were eliminated. The results were essentially the same.

minus the half-spread in another interval of length tick size. We summarize this information at t with the bivariate observation y_t. Define $Y_t \equiv \{y_1, y_2, \ldots, y_t\}$, the history of the observation vector through time t. The observed discrete bivariate y_t gives some but not complete information on the unobserved latent variables x_t and s_t. The goal is to extract the information in Y_T in order to make inferences on the unobservable equilibrium prices and spreads. Given a model of state variable evolution, we also wish to estimate the parameters that characterize the model.

The econometric model in equations (6), (7), and (8) has been cast into the state space framework with (6) and (7) as the Transition equations and (8) as the Measurement equation. Note that the rounding mechanism embedded in the Measurement equation destroys the Gaussian structure and the time-series independence of the errors, rendering standard estimation methods invalid. We have a nonlinear, non-Gaussian state space model, the estimation of which is done using the Monte Carlo Markov Chain (MCMC) estimation approach. Before discussing the MCMC approach we first present our data.

II. Data

Transaction data for seven large stocks were obtained from the TAQ database for the months of February 1997 and November 1997. On June 24, 1997, the NYSE reduced the tick size from an eighth to a sixteenth. Thus, using data from before and after the change in the tick size regime allows us to assess the consistency and generality of the model specification.[9] We also study seven mid-cap stocks in the last quarter of 1997. Although we have certain expectations regarding the relative size of the adverse selection component and inventory costs for the smaller stocks, we expect that, in a well-specified model, the results should be broadly consistent over different times and different classes of stock. With such a large number of data observations, it is important to exclude obvious errors and misreported information. Below, we systematically list the set of filter rules applied to the data:

1. Opening batch trades are excluded since the trading mechanism at the open is different from that during the rest of the day.
2. Trades reported out of sequence, those with special settlement conditions and those following the daily close are not considered.
3. Since price discovery takes place mainly on the NYSE, only NYSE quotes that are eligible for the Best-Bid-or-Offer (BBO) calculation are used as reference quotes.[10]

[9] For the same seven large stocks, we used data from the first quarter of 1992. This is part of the data used by Huang and Stoll (1997) whom we thank for providing their data. For space considerations, we note that the results for 1992 are consistent with results found in the more recent data.

[10] See Hasbrouck (1995).

4. In accordance with the convention used by Lee and Ready (1991), any quote in the five seconds preceding the trade is ignored in favor of the previous quote.
5. The last trade of any day and the first trade of any day after the opening batch trade are also excluded from the analysis so as to alleviate the impact of any overnight price movement.
6. Any obvious data errors such as negative prices, negative bid or ask quotes, and negative spreads are deleted. Further, the following records are also deleted: (i) quoted spreads greater than five dollars, (ii) ratio of effective spreads to quoted spreads greater than four, (iii) ratio of relative effective spreads to relative quoted spreads greater than four, and (iv) ratio of quoted spread to transaction price greater than 0.4.

Except for deleting the last and the first trade of any day, these filters are exactly the same as those used by Chordia et al. (2001). These filters remove fewer than 0.02 percent of all transaction records. The filtering criteria are designed to remove the obvious errors, including clearly erroneous large price changes such as those shown in Table I of Hasbrouck (1995).

The trade indicator, Q_t, is determined as follows. If the transaction occurs above the quote midpoint, it is regarded as buyer initiated and if the transaction occurs below the quote midpoint it is classified as seller initiated. If a transaction occurs at the quote midpoint, it is signed using the tick-test, which assigns a positive (negative) sign to the trade if the price increases (decreases) from the previous transaction price. If the observed price, P_t, is the same as the previous transaction price, P_{t-1}, the test is applied to P_{t-2}. The test is applied through price P_{t-5}. If no sign is so indicated, we set $Q_t = 0$.

Table I shows the summary statistics for the data. Given the large sample sizes, our analysis is restricted to the lesser of the sample size and the first 14,000 trades for each stock during each sample period. Panels A and B present the summary statistics for the seven large stocks over two reported periods of study. Panel C provides summary statistics for the sample of seven mid-cap stocks. The average quoted spread, number of transactions, and the average price are obtained from TAQ and ISSM. The market capitalization as of the end of January 1997 (Panel A), October 1997 (Panel B), and September 1997 (Panel C) is obtained from the Center for Research on Security Prices (CRSP). Consistent with Chordia et al. (2001) we find that the quoted spreads decline after the decrease in the tick size. There is substantial variation in the stock prices and market capitalizations. For instance, in 1997, GE has the largest market capitalization and DOW has the lowest for both months. In Panels A and B, DOW is by far the smallest stock as confirmed by the number of transactions. Stock prices also exhibit considerable cross-sectional variation. For instance, in February 1997, the average price varies from a high of 149.85 for IBM to a low of 57.54 for GM. The quoted spreads do not exhibit the same variation across stocks suggesting that the tick size is binding for a large fraction of the trades.

Table I

Sample Description

This table describes the sample of seven, large, well-known stocks (Panels A and B) and seven mid-cap stocks (Panel C). *N* represents the number of transactions in the sample period; *QSPR* is the average quoted spread; *PRICE* is the average price over the sample; *SIZE* denotes the market capitalization in billions of dollars as of the last trading day of January 1997 (Panel A), October 1997 (Panel B), and September 1997 (Panel C) . The sample size consists of all transactions and matching NYSE, best-bid-and-offer (BBO) eligible quotes. The data is obtained from TAQ. The analysis is limited to the minimum of the sample size or the first 14,000 transactions in the sample period.

	N	*QSPR* ($)	*PRICE* ($)	*SIZE* ($bn)
Panel A: February 1997				
Citicorp	14,000	0.231	120.68	54.75
Dow	9,312	0.195	79.79	18.74
Exxon	14,000	0.192	102.81	128.68
GE	14,000	0.209	104.28	170.40
GM	14,000	0.201	57.54	44.61
IBM	14,000	0.257	149.85	81.19
Merck	14,000	0.203	92.60	109.24
Panel B: November 1997				
Citicorp	14,000	0.142	124.29	57.29
Dow	8,162	0.132	94.83	20.66
Exxon	14,000	0.116	60.88	152.02
GE	14,000	0.094	68.05	211.50
GM	14,000	0.097	64.13	46.23
IBM	14,000	0.126	100.76	96.75
Merck	14,000	0.126	88.92	107.80
Panel C: October–December 1997				
BRR (Barrett Resources)	4,934	0.156	29.63	0.78
BWA (Borq Warner)	3,759	0.183	47.13	0.78
CBC (Centura Bank)	2,721	0.165	59.81	0.79
LEE (Lee Enterprises)	2,413	0.211	27.31	0.80
BNK (CNB Bankshares)	2,096	0.189	42.06	0.80
NAP (National Processing)	1,290	0.145	10.63	0.81
IHC (Interstate Hotels)	3,064	0.164	37.72	0.81

Notice that the mid-cap stocks in Panel C are approximately 100 times smaller than our seven large stocks in terms of market capitalization. Not only are they much smaller but they also trade far less frequently. The panel lists the number of transactions for the last quarter of 1997. The number varies from 1,290 transactions for NAP to 4,934 for BRR. For the larger stocks, in most cases, the data limit of 14,000 observations was reached in the first week or two of trades. Thus, for the largest stocks, we have snapshots of data over the two time periods. Alternatively, for each of the mid-cap stocks, we have a three-month-long time series of observations. Despite the enormous differences in the character of the stock issues, both sets of stocks must trade under the same minimum tick-size rule. It is our contention that the same minimum tick size cannot be appropriate for both sets of stocks, as well as for other far smaller stocks.

III. The Monte Carlo Markov Chain Approach

We have observations $\{y_t\}$, which, due to discretization, contain limited information on $\{x_t\}$ and $\{s_t\}$ as described in (8). The type of transaction—buy, sell, or cross trade—is known and captured through $Q_t = +1, -1, 0$, respectively. We also know the prior trade size, V_{t-1}, and the prior depth, D_{t-1}, the time between trades, $TIME_{t-1}$, and whether the transaction occurred at the beginning or end of the trading day. We do not know the latent variables, the adjusted equilibrium price x_t, or the true spread s_t.

From a traditional statistical standpoint, the parameters of the model are given by:

$$\Theta = \{\lambda, \alpha, \beta, \delta, \tau, d_1, d_2, \sigma_e, \sigma_u, \rho\}.$$

The latent or state variables of the model are x_t and s_t. Absent discreteness and rounding of prices and under identification of the state variables, the model can be expressed in a generalized multivariate regression framework and, under Gaussian error assumptions, may be estimated by means of maximum likelihood or, in a Bayesian framework, as a seemingly unrelated regression (SUR) (see Zellner (1971)). The discreteness of the measurement equation is handled using the MCMC approach.

The MCMC approach is Bayesian and simulation-based. These simulation methods have gained popularity in the statistical literature as the power of computers has increased.[11] The basic idea behind these models is to expand the parameter space by the time series of latent variables, place priors on the expanded parameter space, and estimate the posterior distribution of the parameters given the priors and the observed data. In general, the calculation of the resultant joint posterior distribution is daunting. However, all we really need is the *marginal* posterior distribution of the original parameters, Θ. The natural idea of integrating out the state variable is computationally impractical, especially in this bivariate case. The Gibbs sampler provides a solution to the problem. It generates the marginal distribution of Θ by working with conditional distributions of the parameters given the state variables and observed data. The sampler provides a simulation rule

[11] See Chib and Greenberg (1996) for a discussion of MCMC methods, and see Casella and George (1992) for a review of the Gibbs sampler.

to draw from a Markov chain, whose limiting distribution has the desired marginal distribution of the parameters. By repeating draws from this chain, we obtain accurate estimates of the required marginal posterior distribution. Care is required in determining the size of the simulation, assessing the resultant accuracy of the simulation, and devising the means to sample efficiently from the Markov Chain. Key ingredients of this method are the Markov structure of the latent variables and the exact information provided by the discrete prices and quotes about the equilibrium prices and quotes.

To formalize notation let

$$s = \{s_1, s_2, \ldots, s_T\},$$

$$x = \{x_1, x_2, \ldots, x_T\},$$

$$Y_T = \{y_1, y_2, \ldots, y_T\}.$$

Represent the joint distribution of the expanded parameter space given the vector of observations Y_T by $f(\Theta, x, s | Y_T)$. We seek the marginal posterior distribution $f(\Theta | Y_T)$. The Gibbs sampler draws first from $f(x, s | Y_T, \Theta)$, and then from $f(\Theta | x, s, Y_T)$ and repeats. That this repeated conditional sampling generates the appropriate limiting distribution is the essence of the Gibbs sampler. See Geman and Geman (1984) for the proof and Gelfand and Smith (1990, 1992) for relevant applications. Of course, for the method to work efficiently, it is essential that the conditional distributions be drawn efficiently. We now outline the scheme.

The method makes repeated use of conditional sampling from an element of a vector given all other elements. We use the notation $x_{\sim t}$ to indicate all elements of the vector x except x_t.

1. Initialize x, s, and Θ and place priors on Θ. We choose near diffuse priors as we explain in Section IV.C. We set s_t equal to half the quoted spread, $x_t = P_t - s_t Q_t / 2$. We initially set $\lambda = 0.4$, $\alpha = 0.03$, $\beta = 0.85$, $\delta = 0.005$, $\tau = 0.001$, $d_1 = 0.001$, $d_2 = 0.001$, $\rho = -0.20$, $\sigma_u = 0.08$, and $\sigma_e = 0.02$. The final properties of the sampler do not depend on the initial settings. We simply record this information for completeness.
2. Sample with replacement $x_t, s_t | x_{\sim t}, s_{\sim t}, \Theta, Y_T$.
3. Sample $\alpha, \beta, \delta, \lambda | x, s, \sigma_u, \sigma_e, \rho, Y_T$.
4. Sample $\sigma_u, \sigma_e, \rho | x, s, \alpha, \beta, \delta, \lambda, Y_T$.
5. Go to step 2.

Going through the cycle from step 2 to step 4 represents a complete sweep of the Gibbs sampler. Step 2 is the bivariate state variable sampling. Given the state variables, x and s, steps 3 and 4 update the parameters. We discuss steps 2, 3, and 4 in more detail in the following subsections.

A. Implementation of the Bivariate State Variable Sweep

We express the model in matrix notation:[12]

$$z_t = A_0 + A_t z_{t-1} + C_1 \epsilon_1, \tag{9}$$

where z_t represents the column vector $(x_t, \gamma_t)'$, A_0 is a column vector of constants, A_t is a matrix of coefficients, C_1 is the Cholesky factorization of the covariance matrix of the errors, and ϵ_1 is a column vector of (two) independent standard normals.

The main problem we have is in updating the distribution of z_t given $z_{\sim t}$ and price and spread information at the current time, t.[13] The data place constraints on the bivariate z_t process. As discussed in our discretization assumptions, the data imply that $x_t + s_t/2$ lies in one band determined by the tick size and that $x_t - s_t/2$ lies in another such band.

Since z_t possesses a (bivariate) Markov structure, the distribution of z_t given $z_{\sim t}$ depends only on z_{t+1} and z_{t-1}. Thus we seek $f(z_t|z_{t+1}, z_{t-1})$. Since by assumption z is multivariate normal, the conditional distribution we seek is also multivariate normal.[14]

The problem of drawing bivariate normal samples subject to constraints like these has been explored by Geweke (1999). The Bayesian Analysis, Computation and Communication (BACC) Web page, maintained by Geweke, contains an efficient code written in Fortran and involving an embedded Gibbs sampler to draw realizations from a given multivariate normal distribution with linear constraints such as this.[15] We employed this subroutine to draw z_t given all parameters, observed values of data, and all values of z except the current ones. This procedure is then repeated through the t index to draw values from the full bivariate sequence. Observe that the discreteness condition destroys the multivariate normal structure for the drawn series and so renders full Kalman filter multisweeps (see, e.g., Kim, Shephard, and Chib (1998) and Mahieu and Schotman (1998)) nonimplementable. However, the one-step-ahead updating is quite satisfactory.

B. Implementation of the Parameter Updates

Alternative approaches for estimating a standard SUR regression problem are documented in the literature (Zellner (1971)) and computer programs to implement this approach are available (see BACC). We review the basic Gibbs sampler approach to this problem and highlight appropriate modifications for rounded and missing data.

[12] We linearize $\exp(\gamma_t)$ by means of a first-order linear approximation to express in this form. In Appendix C, we present the Metropolis–Hastings algorithm to provide exact sampling without the linear approximation.

[13] By definition, $z_{\sim t} \equiv \{z_1, \ldots, z_{t-1}, z_{t+1}, \ldots, z_T\}$.

[14] The resultant moments may be calculated using standard results from multivariate normal distribution theory.

[15] http://www.econ.umn.edu/geweke/.

First, given the error covariance structure and a multivariate normal prior on the vector of parameters, we may invoke standard theory to infer a multivariate normal posterior. By means of a Cholesky decomposition, it is straightforward to draw from this distribution. We refer the reader to Geweke (1999) for details.

Second, given these parameter values, we can now sample the covariance matrix. Following Anderson (1984), we place an inverted Wishart prior on the covariance structure.[16] This specification produces a conjugate prior system and the posterior distribution of the error covariance is also inverted Wishart with appropriately modified parameters. We must now draw from the inverted Wishart to obtain samples from the posterior distribution. Unfortunately, we are sampling from an inverted Wishart with some 14,000 underlying observations. To overcome this difficulty, we apply a multivariate central limit theorem and invoke approximate multivariate normality. It is known from theory that such a central limit theorem applies in this case (Anderson (1984)); however the rate of convergence to multivariate normality is relatively slow. Fortunately, given the large sample size, the approximation will prove accurate.

After running each sweep of the MCMC sampler, we have drawings from Θ. We completed 6,000 sweeps, excluding the first 1,000 sweeps and providing summary results on the remaining 5,000 sweeps.[17]

C. Robustness Methods

In addition to the econometric model estimated above, we also consider models with a more general class of error assumptions. Under Gaussian assumptions for the error distributions in the innovations of spread and price, we limit the possibilities of extreme shifts in state variables. The possibility of large shifts in price might alter the specification of adverse selection and inventory costs. To accommodate this possibility we generalize the current model to allow a mixture of bivariate normals for the error distribution. With fixed probability π, we model the errors to be drawn from one bivariate normal distribution and under an alternative regime with complementary probability. The details of implementing this methodology are described in Appendix B. We ran a number of diagnostic checks and the method works very well.

In Section III.A, we described the linearization of the model to preserve multivariate normality. Strictly speaking, this is an approximation. Appendix C introduces the Metropolis–Hastings selection algorithm to refine the MCMC methodology to provide exact sampling without the linear approximation. In fact, the differences from the simpler linear approach turn out to be quite small. We report results from applying the linear approximation to

[16] The details are in Appendix A.

[17] To ensure that there were no programming errors, we ran our programs on simulated data, where we had prespecified the parameter values. The program returned the prespecified parameter values and the state variables.

the basic econometric model. The results from running the Metropolis–Hastings algorithm in conjunction with the mixture of bivariate normal error assumptions with $\pi = 0.9$ are available upon request. Results with alternative values of π were quite similar. The methodologies produce remarkably similar economic conclusions confirming the robustness of the model. While the more general econometric model appears to fit well, the simpler approach provides satisfactory conclusions.

IV. Model Selection

Equations (6), (7), and (8) describe the state space econometric model. We consider various alternative specifications of the spread latent variable evolution while retaining the same measurement equation and the same dynamics for the adjusted equilibrium price variables. The base model posits that the log true spread, γ_t, is an autoregressive process with a single structural variable, $\ln(V_{t-1}/D_{t-1})$, where V is the transaction size and D is the depth. Intuitively, the greater the prior volume per unit depth, the more likely it is that the spread will be widened. Of course there are other factors that may affect the spread. We capture the time-of-day effect with two dummies, BEG_t, which indicates when a trade takes place in the first 60 minutes of the trading day, and END_t, which indicates when a trade takes place in the last 60 minutes of a trading day. We include the variable $TIME_{t-1}$, the length of time in seconds between the last trade and the one prior to that. Researchers have found significant time-of-day effects and time between trades effects with quoted spreads and we allow for this possibility for the equilibrium spreads in this model.[18] We consider four models:

M1: The basic model with a single structural variable;

M2: The basic model plus the Time variable;

M3: The basic model plus the two time-of-day dummies;

M4: The basic model plus the Time variable and the two time-of-day-dummies;

and we compare them through a marginal likelihood approach.

Suppose $f(Y_T|\Theta_K, M_K)$ is the probability function of the data (Y_T) under the model M_K given the model-specific parameter vector Θ_K.[19,20] Let the prior density of Θ_K under model M_K be denoted by $\pi(\Theta_K|M_K)$. Then the marginal likelihood of (Y_T) under model M_K is defined by

$$m(Y_T|M_K) = \int f(Y_T|\Theta_K, M_K)\pi(\Theta_K|M_K)\, d\Theta_K. \tag{10}$$

[18] Clearly, a number of other variables could have been tried. However, we wanted to restrict ourselves to variables identified by prior research. We should point out that our methodology is general enough to handle any additional variables that are in the agents' information set.

[19] Since the data is discrete, the probability function of the data is also discrete.

[20] Recall that the history of the data through time point t is denoted Y_t.

The model with the highest marginal likelihood is preferred. Commonly, Bayes factors, the log of the ratio of the marginal likelihoods for competing models, are computed.

Using the Basic Marginal Likelihood Identity (BMI), we may decompose the marginal likelihood into

$$m(Y_T) = \frac{f(Y_T|\Theta)\pi(\Theta)}{\pi(\Theta|Y_T)}, \tag{11}$$

for any value of Θ.[21] Now taking logs and evaluating at a particular Θ, say Θ^*, we have

$$\ln(m(Y_T)) = \ln(f(Y_T|\Theta^*)) + \ln(\pi(\Theta^*)) - \ln(\pi(\Theta^*|Y_T)). \tag{12}$$

Thus, the log marginal likelihood is the sum of the log likelihood at a given point, the log prior at the same given point, minus the log posterior density at the same point. Chib (1995) argues that given the MCMC sampler, the log posterior at a point can be evaluated by repeated runnings of the sampler. We implemented his approach in this context. The log prior can be easily calculated once it is specified while the major difficulty lies with the calculation of the log likelihood.

This presents a formidable computational problem. Fortunately, we need only evaluate the log likelihood at a single point. The best choice is the mode of the posterior density derived from the MCMC sampler.[22] We elected to compute the likelihood numerically using a nonlinear filter along the lines suggested by Kitagawa (1987). This exercise took approximately 15 hours of processing time. It is obviously impractical to run a full maximum likelihood estimation for problems like this with some ten parameters.

We selected multivariate independent normals for the elements of Θ_1 each with mean zero and variance 100. For the Inverted Wishart distribution on Θ_2, using notation consistent with Anderson (1984), we selected $m = 10$ and Φ to be 0.1 times the identity matrix. The conditional multivariate normal and conditional inverted Wishart generate conjugate prior structures, rendering steps 3 and 4 of the MCMC sampler computationally straightforward. Notice that we are using quite diffuse priors and we have some 14,000 points in the time series.

The results of running the marginal likelihood analysis for GE in November 1997 are presented in Table II. While there are some effects due to the number of parameters in the model, the evaluation of the prior and, to a lesser extent, the posterior distribution vary only slightly across models. The

[21] This analysis follows from Chib (1995).

[22] Chib (1995) argues that the accuracy of the marginal analysis is highest at a parameter point with the highest posterior density.

Table II

Model Selection Using Log Marginal Likelihoods

This table presents the marginal likelihoods for four different models. The four models are variants of the following system:

$$P_t = [m_t + (1-\lambda)s_t Q_t/2]_{Round},$$

$$m_t = m_{t-1} + \lambda s_t Q_t/2 + u_t,$$

$$\ln(s_t) = \alpha + \beta\ln(s_{t-1}) + \delta\ln(V_{t-1}/D_{t-1}) + \tau\, TIME_{t-1} + d_1 BEG + d_2 END + e_t.$$

Here, P_t is the observed transaction price; $m_t(s_t)$ is the unobservable true price (spread) at transaction time t; Q_t is the trade indicator; V_t is the size of the transaction in number of shares; D_t is the corresponding depth of trade for which the then prevailing bid-ask quote held; *TIME* denotes the time between trades in seconds, and *BEG* and *END* are dummy variables indicating the first hour and the last hour of each trading day; and the notation $[.]_{Round}$ indicates rounding onto the discrete grid. In the first model, (M_1), $d_1 = d_2 = \tau = 0$. In the second model, (M_2), $d_1 = d_2 = 0$. In the third model, (M_3), $\tau = 0$, and in the last model, (M_4), the parameters are unrestricted. The log marginal likelihood is the sum of the log likelihood and the log prior minus the log posterior density evaluated at the posterior modes, Θ^*:

$$\ln(m(Y_T)) = \ln(f(Y_T|\Theta^*)) + \ln(\pi(\Theta^*)) - \ln(\pi(\Theta^*|Y_T)),$$

where Y_T denotes the data and $\Theta^* = \{\Theta_1^*, \Theta_2^*\}$; Θ_1^* denotes the slope parameters of the above model and Θ_2^* denotes the parameters of the variance–covariance matrix. The sample consists of the first 14,000 transactions for GE in November 1997.

			Log Posterior		
	Log Likelihood	Log Prior	Slope	Covariance Matrix	Log Marginal Likelihood
M_1	−22,584	−54	17	43	−22,578
M_2	−22,503	−57	25	43	−22,492
M_3	−22,384	−60	24	43	−22,377
M_4	−22,236	−63	33	43	−22,223

bulk of the difference in marginal likelihoods is attributable to the contribution of the log likelihood. Given the large number of observations involved, this is not surprising. In fact, for future research with such large sample sizes, there is likely to be no significant contribution beyond the likelihood and a recommended approach is to simply compare log likelihoods evaluated at posterior means or modes. The results of this exercise clearly point to using the full model. In fact, we can clearly order the models based on Bayes factors: M_4 dominates M_3 dominates M_2 dominates M_1. We have strong evidence that equilibrium spreads are affected by time of day and time between trades as well as by relative depth of the prior trade. Spreads are higher at the beginning and end of the day and decrease with the time between trades. Based on the log marginal likelihood analysis for GE, we run the full model for all other stocks as well.

Before presenting the empirical results we note that we generated a variety of diagnostics to assess the performance of the MCMC methodology in this application. We examined the autocorrelation in the parameter time series and adjusted for it accordingly. We also generated Monte Carlo standard errors and inefficiencies (see Kim et al. (1998) for definitions). The diagnostics confirm that the methodology is working satisfactorily.

V. Empirical Results

Table III presents the parameter estimates of the model and Table IV presents the true spreads. In both tables, Panels A and B present the basic results and Panel C presents results using the mixture of normals and the Metropolis–Hastings sampling for a sample of mid-cap stocks. We first discuss the basic results in Panels A and B and then present the robustness checks.

Table III presents the parameter estimates and the standard deviations of the marginal posterior distributions. The standard deviations of the parameter estimates are low, suggesting tight posterior distributions. The standard deviation of the natural log of the spread, σ_e, is small and in all cases is less than one. On a transaction-by-transaction basis, spreads do not vary widely. The standard deviation of the change in adjusted equilibrium price, σ_u, is consistently estimated to be very small. The largest mean value (excluding IBM) in February 1997 is 5.6 cents for Citicorp and the lowest mean value is 4 cents for GM.[23]

In November 1997, the largest mean value of σ_u is 5 cents for Citicorp and the lowest mean value is 2.3 cents for both GM and GE. This means that, on a transaction-by-transaction time scale, the noise associated with changes in the equilibrium price is small compared to the minimum tick size and the rounding mechanism has a significant impact on observed prices. Earlier models that failed to take this effect into account are grossly misspecified and are likely to draw inappropriate inferences.

The correlations between innovations in true spread and adjusted equilibrium price are small. An exception is the February 1997 GM correlation, which is material with a $\rho = 0.17$. Due to the implicit correlation in the model, it is important to allow for the possibility of correlated errors in the empirical estimation.

The prior size of the trade, adjusted for depth, has a positive impact on spread as evidenced by the significantly positive values of δ throughout. Intuitively, the market is adjusting to larger trades per unit depth by in-

[23] In February 1997, IBM has higher values of σ_u, quoted spreads, and true spreads than any other stock. This was a surprising result. Upon examining the Dow Jones News Retrieval Service, we found that on January 28, 1997, the IBM board of directors had declared a two-for-one common share split. On the announcement of the split, the stock gained \$5 to close at \$150.75. On October 28, 1997 IBM announced that it had allocated \$3.5 billion for stock repurchases. Thus, during both months, February and November 1997, we expect to find higher than normal quoted and true spreads for IBM.

Table III

Monte Carlo Markov Chain Parameter Estimates

This table presents parameter estimates for the following model:

$$P_t = [m_t + (1-\lambda)s_t Q_t/2]_{Round},$$

$$m_t = m_{t-1} + \lambda s_t Q_t/2 + u_t,$$

$$\ln(s_t) = \alpha + \beta \ln(s_{t-1}) + \delta \ln(V_{t-1}/D_{t-1}) + \tau\, TIME_{t-1} + d_1 BEG + d_2 END + e_t.$$

Here, P_t is the observed transaction price; $m_t(s_t)$ is the unobservable true price (spread) at transaction time t; Q_t is the trade indicator; V_t is the size of the transaction in number of shares; D_t is the corresponding depth of trade for which the then prevailing bid-ask quote held; *TIME* denotes the time between trades, and *BEG* and *END* are dummy variables denoting the first hour and the last hour of each trading day; and the notation $[.]_{Round}$ indicates rounding onto the discrete grid. The table reports the mean and the standard deviation (Std. Dev.) for each variable. The sample is described in Table I.

Panel A: February 1997

	CITICORP		DOW		EXXON		GE		GM		IBM		MERCK	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
λ	0.842	0.015	0.945	0.022	0.809	0.022	0.707	0.017	0.790	0.018	0.728	0.015	0.702	0.020
α	−0.882	0.032	−1.029	0.051	−1.045	0.044	−0.916	0.038	−0.782	0.038	−0.653	0.028	−0.771	0.035
β	0.735	0.010	0.701	0.014	0.724	0.012	0.734	0.011	0.787	0.010	0.773	0.009	0.784	0.010
δ	0.065	0.006	0.085	0.008	0.071	0.007	0.066	0.006	0.050	0.004	0.047	0.005	0.059	0.005
σ_u	0.056	0.001	0.042	0.001	0.045	0.001	0.046	0.001	0.040	0.000	0.062	0.001	0.045	0.001
σ_e	0.649	0.011	0.659	0.016	0.757	0.015	0.705	0.013	0.614	0.013	0.582	0.01	0.666	0.013
ρ	−0.141	0.018	−0.012	0.025	−0.002	0.019	−0.023	0.017	0.170	0.019	−0.023	0.018	−0.060	0.018
d_1	0.036	0.017	0.054	0.022	0.067	0.020	0.078	0.019	0.057	0.016	0.010	0.014	0.027	0.017
d_2	0.084	0.017	0.025	0.023	0.075	0.021	0.106	0.020	0.068	0.017	0.044	0.015	0.030	0.018
τ	−0.255	0.131	−0.077	0.141	−0.135	0.171	−1.099	0.355	−0.353	0.129	−1.560	0.368	0.000	0.175

Panel B: November 1997

	CITICORP		DOW		EXXON		GE		GM		IBM		MERCK	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
λ	0.608	0.010	0.704	0.015	0.550	0.012	0.593	0.029	0.548	0.015	0.449	0.011	0.630	0.011
α	−0.445	0.019	−0.518	0.029	−0.323	0.021	−0.328	0.025	−0.527	0.031	−0.283	0.017	−0.298	0.019
β	0.852	0.005	0.825	0.008	0.897	0.005	0.919	0.005	0.860	0.007	0.910	0.004	0.907	0.005
δ	0.050	0.004	0.057	0.007	0.046	0.005	0.027	0.005	0.060	0.005	0.030	0.004	0.035	0.004
σ_u	0.050	0.000	0.047	0.001	0.034	0.000	0.023	0.000	0.023	0.000	0.035	0.000	0.036	0.000
σ_e	0.591	0.008	0.644	0.011	0.508	0.010	0.522	0.012	0.580	0.011	0.460	0.007	0.503	0.009
ρ	−0.083	0.012	0.036	0.016	−0.037	0.014	−0.076	0.017	0.010	0.016	0.007	0.015	−0.047	0.014
d_1	0.033	0.014	0.020	0.019	0.058	0.012	0.032	0.011	0.068	0.014	0.038	0.010	0.051	0.012
d_2	0.052	0.014	0.004	0.020	0.052	0.012	0.022	0.014	0.026	0.015	0.063	0.011	0.041	0.012
τ	−0.261	0.143	−0.106	0.086	−0.333	0.153	−0.171	0.163	0.177	0.120	−0.360	0.196	−0.055	0.198

Panel C: Mid-cap stocks—October to December 1997

	BRR		BWA		CBC		LEE		BNK		NAP		IHC	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
λ	0.678	0.020	0.768	0.024	0.849	0.021	0.533	0.024	0.497	0.028	0.794	0.050	0.790	0.025
α	−0.723	0.039	−0.807	0.043	−0.872	0.053	−0.599	0.044	−0.396	0.046	−1.107	0.098	−0.608	0.046
β	0.749	0.012	0.696	0.015	0.687	0.019	0.727	0.018	0.802	0.016	0.640	0.032	0.791	0.014
δ	0.071	0.009	0.083	0.011	0.093	0.015	0.097	0.016	0.112	0.015	0.071	0.021	0.087	0.011
$\sigma_u^{(1)}$	0.048	0.001	0.069	0.001	0.058	0.001	0.082	0.002	0.067	0.002	0.041	0.001	0.051	0.001
$\sigma_e^{(1)}$	0.733	0.015	0.700	0.016	0.734	0.021	0.672	0.019	0.689	0.022	0.709	0.029	0.690	0.018
$\rho^{(1)}$	−0.046	0.025	−0.037	0.027	−0.058	0.034	−0.027	0.037	−0.038	0.043	0.054	0.053	−0.050	0.031
$\sigma_u^{(2)}$	0.156	0.006	0.203	0.009	0.151	0.008	0.199	0.011	0.202	0.011	0.124	0.013	0.161	0.008
$\sigma_e^{(2)}$	1.092	0.052	1.170	0.061	1.346	0.080	1.038	0.074	0.967	0.070	1.047	0.103	1.030	0.058
$\rho^{(2)}$	−0.047	0.053	−0.028	0.061	0.043	0.070	0.094	0.080	−0.224	0.079	0.086	0.130	−0.063	0.065
d_1	0.023	0.034	0.132	0.036	0.054	0.050	0.082	0.055	0.028	0.045	0.098	0.073	0.094	0.039
d_2	0.021	0.031	0.006	0.033	0.043	0.048	0.048	0.042	0.002	0.054	0.133	0.075	0.024	0.042
τ	−0.034	0.032	0.0004	0.027	0.013	0.000	0.022	0.022	−0.058	0.021	−0.007	0.022	0.030	0.022

Table IV

Estimates of the True Spread

This table presents the means and the medians of the quoted spreads, the true spreads, the adverse selection, and the inventory and order processing components of the true spread, the ratio of the true to the quoted spreads, and the ratio of the true spreads to prices for seven large stocks in Panels A and B, and for seven mid-cap stocks in Panel C. The true spread is a state variable in the model. The sample is described in Table I.

		Quoted Spreads (cents)	True Spreads (cents)	Adverse Selection Component of True Spreads (cents)	Inventory & Order Processing Component (cents)	Ratio of True to Quoted Spreads (%)	Ratio of True Spreads to Prices (%)
				Panel A: February 1997			
Citicorp	Mean	23.13	5.99	5.04	0.95	25.90	0.05
	Median	25.00	3.84	3.23	0.61	15.36	0.03
Dow	Mean	19.49	3.73	3.53	0.20	19.14	0.05
	Median	12.50	2.49	2.35	0.14	19.92	0.03
Exxon	Mean	19.22	3.51	2.84	0.67	18.26	0.03
	Median	12.50	2.03	1.64	0.39	16.24	0.02
GE	Mean	20.94	5.34	3.77	1.57	25.50	0.05
	Median	25.00	2.66	1.88	0.78	10.64	0.03
GM	Mean	20.10	3.62	2.86	0.76	18.01	0.06
	Median	12.50	2.28	1.80	0.48	18.24	0.04
IBM	Mean	25.70	8.18	5.95	2.23	31.83	0.05
	Median	25.00	5.99	4.36	1.63	23.96	0.04
Merck	Mean	20.26	5.29	3.72	1.57	26.11	0.06
	Median	12.50	2.48	1.74	0.74	19.84	0.03

Panel B: November 1997							
Citicorp	Mean	14.21	7.36	4.48	2.88	51.79	0.06
	Median	12.50	4.63	2.82	1.81	37.04	0.04
Dow	Mean	13.21	5.95	4.45	1.50	45.04	0.06
	Median	12.50	3.65	2.73	0.92	29.20	0.04
Exxon	Mean	11.63	4.39	2.41	1.98	37.75	0.07
	Median	12.50	2.46	1.35	1.11	19.68	0.04
GE	Mean	9.38	2.34	1.39	0.95	24.95	0.03
	Median	6.25	1.09	0.65	0.44	17.44	0.02
GM	Mean	9.74	2.35	1.29	1.06	24.13	0.04
	Median	6.25	1.29	0.71	0.58	20.64	0.02
IBM	Mean	12.57	5.46	2.45	3.01	43.44	0.05
	Median	12.50	3.34	1.50	1.84	26.72	0.03
Merck	Mean	12.63	4.97	3.13	1.84	39.35	0.06
	Median	12.50	2.59	1.63	0.96	20.72	0.03
Panel C: Mid-cap Stocks—October to December 1997							
BRR	Mean	15.63	9.46	6.41	3.05	60.52	0.63
	Median	12.50	5.62	3.81	1.81	44.96	0.37
BWA	Mean	18.33	13.36	10.26	3.10	72.89	0.24
	Median	18.75	8.64	6.64	2.00	46.08	0.15
CBC	Mean	16.52	12.01	10.20	1.81	72.70	0.21
	Median	12.50	7.11	6.04	1.07	56.88	0.12
LEE	Mean	21.07	17.82	9.50	8.32	84.58	0.23
	Median	18.75	13.40	7.14	6.26	71.47	0.17
BNK	Mean	18.92	14.24	7.08	7.16	75.26	0.41
	Median	18.75	11.34	5.64	5.70	60.48	0.32
NAP	Mean	14.45	7.63	6.06	1.57	52.80	0.08
	Median	12.50	5.10	4.05	1.05	40.80	0.06
IHC	Mean	16.36	10.10	7.98	2.12	61.74	0.07
	Median	12.50	5.85	4.62	1.23	46.80	0.04

creasing subsequent spreads. The impact wears off over time and spreads revert to their average levels. In effect, there is evidence that spreads rise after large trades so market makers are protecting themselves against informed traders, a form of adverse selection adjustment.

The coefficients for the time-of-day dummies are significantly positive, both in February and November 1997, indicating that the true spreads are higher during the first and the last hour of the trading day. This result is consistent with observations on quoted spreads. The coefficient on TIME, τ, is statistically significantly negative for all stocks in February 1997 except for Merck, for which it is statistically insignificant. In November 1997, τ is once again statistically negative for all stocks except GM. Overall the results suggest that true spreads decline with time between trades. We stress that τ is small even though the TIME variable is measured in seconds so the coefficient picks up the marginal effect of increasing the time between trades by one second. For the heavily traded stocks in our sample, the time between trades is only a matter of seconds and does not vary significantly. For a sample of large stocks, we do not expect this variable to be highly economically significant.

The intercept term α is significantly negative and the β coefficient is positive and less than unity.[24] The α and β coefficients determine the autoregressive structure of the log latent spread when all other drift coefficients are zero, that is, in the middle of the day when the time between trades goes to zero and when the prior volume is at the depth. The autoregressive slope parameter β captures the degree of persistence in the spread and $\alpha/(1-\beta)$ gives the limiting mean of the log spread under the hypothesized conditions. The β parameter is consistently around 0.7 in February 1997 but increases to above 0.8 in November 1997.[25] However the α parameter becomes less negative in November 1997. Interestingly, the average log spread remains fairly constant.

We also incorporate structural variables in assessing the log spread, equation (7). The MCMC sampler makes draws from the distribution of true spreads as it iterates towards the joint posterior distribution of the parameters. As a consequence, by averaging over the draws from the true spread process at each time point, we develop estimates of true spread that take into account the autoregressive structure of the spreads and the impact of the posited structural variables.

To capture the general level of spreads, Table IV lists the means and the medians of the estimated true spread across the stocks in our sample for all periods of study. We may break down the true spread into the adverse selection component and the inventory and order processing components. While we assume that the fraction of the spread due to adverse selection remains constant during the month of the study, we allow for this fraction to change

[24] Clearly, we do not expect the true spread series to be nonstationary.

[25] There may be fluctuation in persistence over time, and further, we see no economic effect that may explain the change in persistence over this period.

from February 1997 to November 1997. Summary statistics are reported for these slices of the spread also. For comparison purposes we also quote the equivalent statistics for the quoted spreads. Table IV calculates the ratio of the estimated true spreads to the quoted spreads and to the observed transaction price.

The minimum value of the quoted spread is the prevailing tick size. For the large stocks, the quoted spread represents an upper bound on what the true spread would be in a market without a minimum tick size and discrete prices. The tick size is binding when the quoted spread equals the tick size. When the median quoted spread equals the tick size, we know that the tick size is binding in over 50 percent of the transactions, which is often the case. In all cases, the institutionally mandated tick size is binding for over a third of the trades.

There is some variability in the ratio of the true spread to the quoted spread. Not surprisingly, the average value of the ratio was much larger in November 1997 when the minimum tick size was 6.25 cents than in February 1997 when the minimum tick size was 12.5 cents.

In February 1997, the average of the median true spread is 2.6 cents and the corresponding statistic for quoted spreads is 20.5 cents.[26] In November 1997, the average median true spread is also 2.6 cents, whereas the average for the quoted spread is 11.8 cents. While the quoted spread decreased with the reduction in the tick size, the true spread remained essentially unchanged. The median ratio of the true spreads to prices also remained constant across the two periods of study. The reduction in the minimum tick size has a big impact on quoted spreads but very little impact on true spreads. The fact that the true spreads do not change with the tick size provides strong support for our model.

Given that the true spreads do not change much, we examine whether the components of the spreads are altered by the change in market policy. In February 1997, the median adverse selection component of the true spreads varies from 1.6 cents for Exxon to 3.2 cents for Citicorp, and in November 1997, the median adverse selection component varies from 0.7 cents for GE to 2.8 cents for Citicorp. There is some evidence that the adverse selection component decreased with the lower minimum tick size. The inventory and order processing component is lower still. The adverse selection component, in our sample, is far smaller than the tick size, suggesting that rounding to the nearest tick provides market makers with ample compensation for the risk of trading with an informed investor.

A. Robustness

To explore whether our striking findings are general and not sample-specific, we perform a number of robustness checks:

[26] Note that the true spreads have a lognormal distribution due to the parameterization in equation (3). Thus, due to skewness, we will focus on the medians of the true spreads.

1. We first redo the above analysis with a Metropolis–Hastings sampling procedure (for details, see Appendix C) and with a mixture of bivariate normals error specification for innovations in the state variables (for details, see Appendix B). The Metropolis–Hastings algorithm resolves the issue of nonlinearity introduced by the log spread transformation, $\gamma_t = \ln(s_t)$. Due to the nonlinearity we do not have a multivariate normal system. This has so far been addressed by a linear approximation of the exponential function. We found that the average acceptance probabilities with the Metropolis–Hastings algorithm were over 98 percent, thus supporting the adequacy of the linear approximation. The mixture of normals allows us to handle shocks to the system by allowing a fat-tailed innovation distribution. The main conclusions are not altered by using this more general methodology. These results confirm that while the more general model may provide a better econometric fit to the data, the basic conclusions and parameter estimates are robust to these types of model misspecification.
2. We also use part of the sample used by Huang and Stoll (1997). Our results continue to hold for this comparison period. The econometric model that we propose includes Huang and Stoll as a special case. Our results highlight the significant impact of rounding and we find much lower levels of inventory and order processing costs.
3. Finally, we use a sample of mid-cap stocks from the last quarter of 1997 to see how our model performs on stocks that are some 100 times smaller and that have far fewer transactions than the earlier sample of the largest stocks. This sample has data over a three-month period and, thus, demonstrates the consistency of the model over longer time spans of data with fewer observations within the span.

We have repeated the earlier analysis (results available upon request) in Panels A and B of Tables III and IV with the Metropolis–Hastings sampling procedure and with a mixture of normals to allow for fat-tailed distributions. The component probability was $\pi = 0.9$ for these runs. The results (not presented for brevity) hardly changed when different values of π were used. The parameters now include two sets of estimates of the variance–covariance matrix for the system. The parameters $\sigma_u^{(1)}$, $\sigma_e^{(1)}$, and $\rho^{(1)}$ represent one set of estimates for 90 percent of the sample and $\sigma_u^{(2)}$, $\sigma_e^{(2)}$, and $\rho^{(2)}$ represent another set of estimates for the remaining 10 percent of the sample. The true spread measures are almost unchanged, strongly suggesting that for the largest stocks, the Metropolis–Hastings sampling algorithm and the mixture of normals error specification do not have a large impact on the results. The level of the spread and the volatility of the spread remain consistent across the two types of estimation method.

The differences that do exist between the two approaches appear confined to the reversion of the spread to its long-term mean. For example, for GE in November 1997, using the mixture of normals and the Metropolis–Hastings algorithm, we find $\beta = 0.843$ and $\alpha = -0.665$, while using the linear approach, $\beta = 0.919$ and $\alpha = -0.328$. However, the median true spreads are estimated to be 0.93 cents with the mixture of normals and the Metropolis–Hastings algorithm and 1.09 cents under the simpler linear approach. It appears that under the mixture of normals model, we estimate lower levels of estimated β with the possibility of higher levels of transition volatility, but both methods produce consistent long-term levels of the true spread. In sum, the mixture model permits higher levels of one period volatility of spread and indicates less persistency. However, the impact on the true spread and the adverse selection component is minimal.

We now turn our attention to the Huang and Stoll (1997) sample for our largest stocks in the first quarter of 1992. The results (not presented but available upon request) indicate that the median true spreads are once again far smaller than the tick size, ranging from 0.79 cents for GE to 2.87 cents for DOW. These results are in sharp contrast to those obtained by Huang and Stoll (1997) and underscore the effect of rounding. We find a significant component of spread attributable to the discrete grid size, a feature ignored in the Huang and Stoll analysis.

Finally, Panel C in Table III presents the parameter estimates and Panel C in Table IV presents the true spreads for the sample of mid-cap stocks in the last quarter of 1997. The median true spread varies from 5.1 cents for NAP to a high of 13.4 cents for LEE. The average median true spread is 8.15 cents when the minimum tick size was 6.25 cents. Thus, for these stocks, the tick size may no longer be binding in general. However, even for these mid-cap stocks, the tick size may be binding for some trades. Also, a lower tick size would allow for pricing on a finer grid. To summarize, the basic results survive a number of robustness checks. The dramatic findings appear to be pervasive.

B. Summary and Discussion

What do market makers do to earn fair profits and to protect themselves from more informed traders? They can choose where to set the quotes, they can choose to widen the spreads, or they can agree to select a minimum tick size that ensures a lower bound on spreads. All of these represent forms of protection. Formally, we refer to the location of spreads response as adverse selection. In sum, there is a small but detectable price adjustment for adverse selection and a small but detectable increase in spread in response to prior size of trade, but the lion's share of the adjustment is due to the practice of rounding in the market maker's favor onto the discrete price grid. The level of tick size during the periods of this study is well able to protect the market maker against informed traders, particularly before June 24, 1997, and indeed provides substantial profit. The presence of payment for order flow or, in a sense, a reduction in spreads offered by competitive sources reinforces the evidence of excess profits built into the system.

The results in this paper point to dramatic differences in information flow and efficiency for very large stocks as compared to mid-cap stocks. The estimates of adverse selection and inventory costs for the very largest stocks traded on the New York Stock Exchange are very low and the minimum tick size currently in force is far too high for these stocks. The evidence from the mid-cap stocks indicates higher levels of estimated true spread.

VI. Conclusions

Empirical studies that use transaction data within-the-day need to recognize the effects of discrete pricing. For large stocks traded on the New York Stock Exchange in February 1997 with up to 1,000 trades per day, the effects of rounding to 1/8 of a dollar are highly significant. In November, 1997, with a tick size of 1/16 of a dollar, the effects of rounding remain commanding. The effects of rounding are not so severe in the case of mid-cap or smaller stocks.

We have put forward an economic model that directly incorporates the economic gains to rounding and decomposes the market spread into adverse selection and inventory and order processing cost components. The goal of the approach is to identify the equilibrium prices of assets and to ascertain market spreads in a world without discrete quotes. A byproduct of this endeavor highlights the relative size of true spreads and quoted spreads. The resulting model may be investigated empirically when set up in a state-space framework. We investigate various competing statistical models that capture important features of the price setting process such as time-of-day effects, size-of-trade effects, and time-between-trades effects. Unfortunately, due to the rounding process, the framework necessarily involves non-Gaussian and correlated error components that render standard linear Gaussian state-space methods invalid. To overcome these formidable econometric difficulties, we employ a Bayesian approach and calculate the appropriate marginal posterior distributions by means of Monte Carlo Markov Chain simulation methods. With one or two notable exceptions, the problems induced by rounding have been ignored by the market microstructure literature.

For the large, heavily traded stocks we find low levels of adverse selection and tight equilibrium spreads on the order of three to four cents. In one respect, the market for large stocks on the New York Stock Exchange is highly efficient with small spread increases for protection against potentially informed traders. However, the quoting of prices to the nearest tick in the market maker's favor produces significant market-maker profits that dwarf the equilibrium spreads. The practice of payment for order flow and internalization highlights the excess profits due to rounding.

The characteristics of the very largest stocks are much different from the mid-cap stocks and the small stocks, yet all classes of stock currently have the same minimum tick size. For the very largest stocks, quoting prices to a narrower grid would produce much tighter market maker spreads and consequent lower costs to retail customers. The caveat is that quoted depths might decrease.

Appendix A. Sampling the Covariance Matrix: Wishart Distribution

Suppose $x_1, x_2, \ldots, x_n$ are distributed independent p-dimensional multivariate normal $MVN[0, \Sigma]$. Define

$$V = \Sigma_{i=1}^{n} x_i x_i'. \tag{A1}$$

Then the distribution law for V is $W(\Sigma, p, n)$, a Wishart distribution with density

$$p(v) = c\,\frac{\mathrm{Det}(V)^{(n-p-1)/2}}{\mathrm{Det}(\Sigma)^{-n/2}}\exp\left[-\frac{1}{2}\,\mathrm{tr}(\Sigma^{-1}V)\right], \tag{A2}$$

where $c = 1/(2^{0.5pn}\pi^{0.25p(p-1)})$. To sample V from $W[\Sigma, p, n]$, we must compute $\Sigma_{i=1}^{n} x_i x_i'$ and this involves $n * p$ random normals.

Let

$$A(n) = \Sigma_{i=1}^{n} x_i x_i'. \tag{A3}$$

Anderson (1984), Theorem 3.4.4. gives the limiting distribution of $A(n)$. Define

$$B(n) = n^{(-1/2)}[A(n) - n\Sigma], \tag{A4}$$

$$B(n) \sim MVN[0, COV_B], \tag{A5}$$

where COV_B is a function of Σ. In fact, for each i, j, k, l, we have

$$E[B_{i,j}(n)B_{k,l}(n)] = \Sigma_{i,k}\Sigma_{j,l} + \Sigma_{i,l}\Sigma_{j,k}. \tag{A6}$$

For large n we simply draw $B(n)$ from the appropriate multivariate normal distribution using the Cholesky decomposition of COV_B and then generate $A(n)$ by linear transformation.

If A is $W[\Sigma, n, p]$, then A^{-1} is inverted Wishart $W^{-1}[\Sigma^{-1}, n, p]$. Suppose Σ has a prior $W^{-1}[\Phi, m, p]$ where Φ and m are Bayesian hyperparameters. Anderson (1984) establishes the conjugate prior system. If A is n times the sample covariance matrix for the residuals from the SUR model with known coefficients, then the posterior distribution of $\Sigma|A$ is also inverse Wishart with parameters:

$$\Sigma|A \sim W^{-1}[A + \Phi, n + m, p], \tag{A7}$$

or equivalently,

$$\Sigma^{-1}|A \sim W[(A + \Phi)^{-1}, n + m, p]. \tag{A8}$$

We may now use the multivariate normal approximation to the Wishart for large n to draw samples from the inverted Wishart posterior distribution for the covariance.

Appendix B. Mixture of Bivariate Normals Approach

Recall the basic economic framework:

$$x_t = x_{t-1} + \lambda s_{t-1} Q_{t-1}/2 + \sigma_u \eta_{1t}, \tag{B1}$$

$$\gamma_t = \alpha + \beta \gamma_{t-1} + \delta Obs_{t-1} + \sigma_e \eta_{2t}, \tag{B2}$$

where $\gamma_t \equiv \ln(s_t)$ and $\{\eta_{1t}, \eta_{2t},\ t = 1,2\ldots T\}$ are *i.i.d.* bivariate standard normals with correlation ρ. We use the term Obs_{t-1} to represent time $t-1$ observations and fixed parameter values used in any one of the models considered. The bivariate normal assumption for the evolution of the state variables limits the risks to market makers. It is possible that if there were a chance of larger innovations, then both the inventory and adverse selection costs would be larger. To accommodate this possibility we introduce a more sophisticated model for the random error components: a mixture of bivariate normals. In this way, with a small probability, there may be much larger innovations in spreads and equilibrium prices and market makers may take this into account when setting bids and asks. An extension of the MCMC methodology handles this more general specification and allows us to assess the impact of nonnormal error innovations.

We work with a fixed mixture of two bivariate normals.[27] Suppose the error distribution is given by Σ_1 with probability π and Σ_2 with complementary probability. Define a new vector of latent variables, $\{Ind\}$, so that $Ind_t = 1$ if mixture 1 obtains at time t, and $Ind_t = 2$ otherwise. In this way, *given the value of* Ind_t at time t, the distribution of the errors is bivariate normal. Of course, the unconditional distribution of the errors not knowing Ind_t is the prescribed mixture of bivariate normals. The new latent variable Ind needs to be incorporated into the MCMC method. Step 2 of the MCMC is modified to

Draw $x_t, \gamma_t | x_{\sim t}, \gamma_{\sim t}, \{Ind_t\}$.

Knowledge of Ind_t for each t fixes the variance covariance structure and essentially reduces the problem to the previous step 2. However, we need an extra step in the MCMC scheme:

Draw $Ind_t | x_t \gamma_t, Ind_{\sim t}$.

[27] This idea has been used by Mathieu and Schotman (1998) and Kim et al. (1998) for one-dimensional mixtures for application to stochastic volatility estimation.

That is, given the residuals from the model and knowing the values of the time t state variables, we need to draw from the conditional distribution of Ind_t. Let ϵ_1, ϵ_2 be the time t residuals.

Let the density of the residuals given $Ind_t = 1$ be $\phi_1(\epsilon_1, \epsilon_2)$, a bivariate normal with covariance Σ_1 and given $Ind_t = 2$ be $\phi_2(\epsilon_1, \epsilon_2)$, a bivariate normal with covariance Σ_1.

By Bayes Rule

$$P[Ind_t = 1|\epsilon_1, \epsilon_2] = \frac{\pi\phi_1(\epsilon_1, \epsilon_2)}{\pi\phi_1(\epsilon_1, \epsilon_2) + (1-\pi)\phi_2(\epsilon_1, \epsilon_2)}. \tag{B3}$$

Thus, the conditional distribution of Ind_t is Bernoulli with probabilities as defined above. The sweep through the extended MCMC chain is quite simple.

In this extended model, care is needed in drawing from the distributions of the revised parameter vector β and the two variance covariance matrices, σ_1 and σ_2. Significantly, at each run we know the value of the indicator variable so that we know which data points correspond to regime 1 and which to regime 2. Given the β parameter, we use the same inverted Wishart conjugate prior system that we used in the single regime case and we draw separately from the revised Σ_1 and Σ_2 distributions. For each simulation we know when Σ_1 obtains and when Σ_2 obtains. The draw from the β posterior is Gaussian and two-step in nature. We have a normal prior on beta and we use the Σ_1 data to generate the posterior normal distribution for beta. This resulting distribution acts as the new prior on β for the Σ_2 data. The final posterior is again normal. Appendix A describes the computational methods used to efficiently draw from the posterior in this case.

This method can be extended to a higher number of mixtures quite easily in principle. In this paper we used a mixture of two bivariate normals and we fixed $\pi = 0.9$. We also ran the method with different values of π and the results were in very close proximity.

Appendix C. Metropolis–Hastings Algorithm

Recall the economic framework in equations (B1) and (B2). Since $s_t = \exp(\gamma_t)$, due to the nonlinearity, we do not have a multivariate normal system. If we approximate the exponential function by a linear function, then the resulting approximation preserves multivariate normality.

This Appendix describes how to draw $x_t, s_t | x_{\sim t}, s_{\sim t}, \Theta, y_t$ exactly. The x_t, s_t process is bivariate Markov, and so all we need is $x_t, s_t | x_{t+1}, s_{t+1}, x_{t-1}, s_{t-1}, \Theta, y_t$ or equivalently $x_t, \gamma_t | x_{t+1}, \gamma_{t+1}, x_{t-1}, \gamma_{t-1}, \Theta, y_t$. For space considerations we suppress the dependence on the $t-1$ state variables and the fixed parameter values Θ for the rest of the argument. Fixing x_{t-1} and γ_{t-1}, let $f(x_t, \gamma_t | y_t)$ represent the joint bivariate truncated normal distribution of $x_t, \gamma_t | x_{t-1}, \gamma_{t-1}, y_t$.

Let $p(.) = p(x_t, \gamma_t | x_{t+1}, \gamma_{t+1}, y_t)$ be the joint distribution of x_t and γ_t given y_t and $t+1$ state variables under the exact exponential model and let $f(.) = f(x_t, \gamma_t | x_{t+1}, \gamma_{t+1}, y_t)$ be the corresponding bivariate normal approximation (with truncation implied by y_t). By applying Bayes Theorem, we have for $x_t, \gamma_t \in y_t$

$$p(x_t, \gamma_t | x_{t+1}, \gamma_{t+1}, y_t) = p(x_{t+1}, \gamma_{t+1} | x_t, \gamma_t) f(x_t, \gamma_t)/C_p, \tag{C1}$$

where the normalizing constant C_p is given by

$$C_p = \iint_{x_t, \gamma_t \in y_t} p(x_{t+1}, \gamma_{t+1} | x_t, \gamma_t) f(x_t, \gamma_t)\, dx_t\, d\gamma_t \tag{C2}$$

and similarly

$$f(x_t, \gamma_t | x_{t+1}, \gamma_{t+1}, y_t) = f(x_{t+1}, \gamma_{t+1} | x_t, \gamma_t) f(x_t, \gamma_t)/C_f \tag{C3}$$

where the normalizing constant C_f is given by

$$C_f = \iint_{x_t, \gamma_t \in y_t} f(x_{t+1}, \gamma_{t+1} | x_t, \gamma_t) f(x_t, \gamma_t)\, dx_t\, d\gamma_t. \tag{C4}$$

Of course, if x_t, γ_t are not consistent with y_t, the conditional densities take on the value zero. Drawing from the above conditional distribution is straightforward since it has a well-defined truncated bivariate normal structure.

The Metropolis–Hastings algorithm allows us to make drawings from the $p(.)$ distribution by first making a draw from the $f(.)$ distribution as a candidate approximation and then choosing the selected draw with a probability that depends on the draw. If the draw is not selected, the values from the previous sweep are retained. (See Chib and Greenberg (1995) or Geweke (1999) for details of the method.) To fix notation for the time t draw, we condition on x_{t+1}, x_{t-1} $\gamma_{t+1}, \gamma_{t-1}$ and we have the current values, x_t, γ_t. We use the normal approximation with truncation to draw x_t^*, γ_t^* from $f(.)$. The Metropolis–Hastings algorithm will accept x_t^*, γ_t^* with probability $q(x_t^*, \gamma_t^*)$ and retain x_t, γ_t, the values from the prior sweep, with the complementary probability where

$$q(x_t^*, \gamma_t^*) = \min(h(x_t^*, \gamma_t^*), 1), \tag{C5}$$

where

$$h(x_t^*, \gamma_t^*) = \frac{p(x_t^*, \gamma_t^* | x_{t+1}, \gamma_{t+1}, y_t)/f(x_t^*, \gamma_t^* | x_{t+1}, \gamma_{t+1}, y_t)}{p(x_t, \gamma_t | x_{t+1}, \gamma_{t+1}, y_t)/f(x_t, \gamma_t | x_{t+1}, \gamma_{t+1}, y_t)}. \tag{C6}$$

The normalizing constants C_p and C_f do not depend on time t values of the state variables and hence cancel out for both the $p(.)$ and $f(.)$ distributions allowing for considerable simplification. Using (C1) and (C3) we see that

$$h(x_t^*, \gamma_t^*) = \frac{p(x_{t+1}, \gamma_{t+1}|x_t^*, \gamma_t^*)/f(x_{t+1}, \gamma_{t+1}|x_t^*, \gamma_t^*)}{p(x_{t+1}, \gamma_{t+1}|x_t, \gamma_t)/f(x_{t+1}, \gamma_{t+1}|x_t, \gamma_t)}. \tag{C7}$$

All resultant conditional densities are bivariate normal and correspond to the one-step transition densities for the exact process with the exponential log spread ($p(.)$) and to the linear approximation ($f(.)$). As a result, the calculation of the acceptance probability is extremely quick. Empirically, we calculated the average acceptance probabilities over all sweeps and found it to be around 98 percent. For most cases in this study, the linear approximation without the Metropolis–Hastings step will provide satisfactory results.

REFERENCES

Anderson, Theodore W., 1984, *An Introduction to Multivariate Statistical Analysis*, 2nd Edition, (John Wiley and Sons, New York, NY).

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

Ball, Clifford, Walter Torous, and Adrian E. Tschoegl, 1985, The degree of price resolution: The case of the gold market, *Journal of Futures Markets* 5, 29–43.

Brennan, Michael and Avanidhar Subrahmanyam, 1995, Investment analysis and price formation in securities markets, *Journal of Financial Economics* 38, 361–381.

Cao, Charles, Hyuk Choe, and Frank Hatheway, 1997, Does the specialist matter? Differential execution costs and intersecurity subsidization on the New York Stock Exchange, *Journal of Finance* 52, 1615–1640.

Casella, George, and Edward George, 1992, Explaining the Gibbs sampler, *The American Statistician* 46, 167–174.

Chib, Siddhartha, 1995, Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association* 90, 1313–1321.

Chib, Siddhartha, and Edward Greenberg, 1996, Markov Chain Monte Carlo simulation methods in econometrics, *Econometric Theory* 12, 409–431.

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

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2001, Market liquidity and trading activity, *Journal of Finance* 56, 501–530.

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

Copeland, Thomas, and Dan Galai, 1983, Information effects on the bid-ask spreads, *Journal of Finance* 38, 1457–1469.

Demsetz, Harold, 1968, The cost of transacting, *Quarterly Journal of Economics* 82, 33–53.

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

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

Easley, David, and Maureen O'Hara, 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 577–605.

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

Gelfand, Alan E., and Adrian F. M. Smith, 1990, Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* 85, 398–409.

Gelfand, Alan E., and Adrian F. M. Smith, 1992, Bayesian statistics without tears: A sampling-resampling perspective, *American Statistician* 46, 84–88.

Geman, Stuart, and Donald Geman, 1984, Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images, *IEEE: Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.

George, Thomas, Gautam Kaul, and M. Nimalendran, 1991, Estimation of the bid-ask spread and its components: A new approach, *Review of Financial Studies* 4, 623–656.

Geweke, John, 1999, Using simulation methods for Bayesian econometric models: Inference, development and communication, *Econometric Reviews* 18, 1–126.

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

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

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

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

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

Hasbrouck, Joel, 1988, Trades, quotes, inventories and information, *Journal of Financial Economics* 22, 229–252.

Hasbrouck, Joel, 1991, The summary informativeness of stock trades: An econometric investigation, *Review of Financial Studies* 4, 571–595.

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

Hasbrouck, Joel, 1995, One security, many markets: Determining the contributions to price discovery, *Journal of Finance* 50, 1175–1199.

Hasbrouck, Joel, 1999a, The dynamics of discrete bid and ask quotes, *Journal of Finance* 54, 2109–2142.

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

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

Ho, Thomas, and Hans Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial Economics* 9, 47–73.

Ho, Thomas, and Hans Stoll, 1983, The dynamics of dealer markets under competition, *Journal of Finance* 38, 1053–1074.

Huang, Roger, and Hans Stoll, 1997, The components of the bid-ask spread: A general approach, *Review of Financial Studies* 10, 995–1034.

Kim Sangjoon, Neil Shephard, and Siddhartha Chib, 1998, Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies* 65, 361–393.

Kitagawa, Genshiro, 1987, Non-Gaussian state-space modelling of non-stationary time series, *Journal of the American Statistical Association* 82, 1032–1063.

Kyle, Albert, 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1336.

Lee, Charles, and Mark Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–747.

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

Madhavan, Ananth, and Seymour Smidt, 1991, A Bayesian model of intraday specialist pricing, *Journal of Financial Economics* 30, 99–134.

Mahieu, Ronald J., and Peter C. Schotman, 1998, An empirical application of stochastic volatility models, *Journal of Applied Econometrics* 13, 333–360.

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

Roll, Richard, 1984, A simple measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, 1127–1139.

Stoll, Hans, 1978, The supply of dealer services in securities markets, *Journal of Finance* 33, 1133–1151.

Stoll, Hans, 1989, Inferring the components of the bid-ask spread: Theory and empirical tests, *Journal of Finance* 44, 115–134.

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics*, (John Wiley and Sons, New York, NY).