Impacts of trades in an error-correction model of quote prices [☆]

Robert F. Engle[a], Andrew J. Patton[b],*

[a] *Department of Finance, NYU Stern School of Business and the Department of Economics, University of California, San Diego, USA*
[b] *Department of Accounting and Finance, Financial Markets Group, London School of Economics, Houghton Street, London WC2A 2AE, UK*

Abstract

In this paper we analyze and interpret the quote price dynamics of 100 NYSE stocks stratified by trade frequency. We specify an error-correction model for the log difference of the bid and the ask price with the spread acting as the error-correction term, and include as regressors the characteristics of the trades occurring between quote observations, if any. From this model we are also able to extract the implied model for the spread and the mid-quote. We find that short duration and medium volume trades have the largest impacts on quote prices for all one hundred stocks. Further, we find that buys have a greater impact on the ask price than on the bid price, while sells have a greater impact on the bid price than on the ask price. Both buys and sells increase spreads in the short run, but in the absence of further trades, the spreads mean revert. Trades have a greater impact on quotes for the infrequently traded stocks than for the more actively traded stocks.
© 2003 Elsevier Science B.V. All rights reserved.

1. Introduction

The process by which information is incorporated into prices is a central focus of both empirical and theoretical microstructure. An important role in this story is generally ascribed to transactions, which have the potential to reveal private information and order flow. In response to transactions, market makers and limit order suppliers adjust their positions. On the NYSE we see the response of the specialist as he updates his quotes, including bid and ask prices and bid and ask depths. These may, but need not, correspond to the inside prices on the limit order book.

There is a vast literature examining the behavior of bid-ask spreads, return dynamics, and time variation of liquidity. In this paper a simple model is proposed which incorporates the key features revealed by this research and extends the analysis to a carefully stratified set of NYSE stocks to further illuminate pricing behavior. The formulation is a simple error-correction model between bid and ask prices; consequently prices are integrated processes while the bid-ask spread is stationary. Trades have a permanent impact on prices but transitory impacts on spreads. A nice feature of our framework is that we are able to extract the implied vector autoregression (VAR) for the spread and the mid-quote from our model for the bid and ask prices. A similar structure has been independently proposed by Escribano and Pascual (2000).

Dufour and Engle (2000) and Engle and Lange (2001) find that various measures of liquidity deteriorate when trading becomes more active. The price impact of a trade and the spread are greater when trades are closer together, and the depth of the market decreases with increased trading frequency. These results, however, are based on the behavior of frequently traded stocks and are interpreted in terms of the Easley and O'Hara (1992) model of asymmetric information. It is not clear whether similar results hold for infrequently traded stocks and whether the notion of high trade intensity would be the same in these markets.

Hasbrouck (1991), in his seminal paper on measuring the price impact of a trade through the use of a VAR, finds that infrequently traded stocks have greater price impacts than frequently traded stocks. However, this result is estimated assuming impacts to be constant over time whereas Dufour and Engle (2000), using an extension of the structure, find that the impact is especially large when trades are frequent. Hasbrouck (1991) also shows that high volume trades have greater price impacts and that this effect diminishes with size.

Easley et al. (1996), employing a completely different methodology which uses only trade flow information and which is carefully grounded in economic theory, find a similar result. Infrequently traded stocks are subject to greater asymmetric information and therefore will have greater price impacts and spreads than more frequently traded stocks. The result does not extend to the least frequently traded decile they consider. They conclude that the spreads there are best understood by a combination of asymmetric information and inventory models. As their model is estimated under the asymmetric information framework, it is not clear how to interpret these findings. Further, it is likely that additional insights and precision might come from the incorporation of price information.

The present study is closer to the Hasbrouck (1991) formulation, but uses a design similar to Easley et al. (1996). It covers a longer and more recent period with a minimum tick size of 1/16 rather than 1/8. Four different trade intensity deciles are examined to better determine where the effects might stop or change form.

Jang and Venkatesh (1991) point out with clever descriptive statistics that the standard microstructure theories and estimates have an assumption that is wildly inconsistent with the data. Market makers and other liquidity suppliers are assumed to respond to news by adjusting both bid and ask prices. In fact, the most common outcome is no adjustment but after that it is found that typically only one is changed. For example when a trade occurs at the ask (generally classified as a buy) the ask is more likely to be raised than the bid. Thus the dynamics of the spread is heavily influenced by the differential response of bids and asks to buys and sells. This is well known to practitioners but has not previously been incorporated in a model to our knowledge.

Standard microstructure theories based on asymmetric information or inventory costs are formulated as if the specialist were taking the other side of every trade. In fact, most of the trades on the NYSE are executed against the limit order book. While it may be tempting to assume that the specialist simply quotes the depth of the inside orders, Kavajecz (1999) finds that the specialist instead acts in much the same way as the limit order suppliers by decreasing his component of the depth at times of high adverse information. For this purpose, the limit order book can be interpreted as competing market makers.

In this setting, sluggish response by the limit order suppliers can easily explain the Jang and Venkatesh (1991) finding. A trade at the ask may exhaust the orders at that price leading the specialist to post a quote at the next higher price on the book if he does not choose to take the entire ask depth himself at the old price. The bid price would move up as well if the specialist or the buying limit orders instantaneously filled the next higher price on the bid side. In this case the midquote would rise but the spread would remain constant. Any delay in these new orders would lead to a temporary spread increase, which is the Jang and Venkatesh finding.

In fact, it is not clear that even swift buying limit orders would optimally move to the new higher price point. The trade could signal good news, which would motivate the higher bid, but it could also signal a new information event, in which case, a more cautious strategy would be to wait. Although models have not yet been solved where news arrives with some stochastic process and the news is known to some of the agents, it is clear that a trade will convey an increased probability of new information and in this case the spread should widen. This hypothesized extension of Easley and O'Hara (1992) should explain how the asymmetric movements of bids and asks would be natural responses even of a specialist who took all trades.

We propose capturing the way in which information in trades is incorporated into quote prices via an error-correction model for the log-difference of bid and the ask prices with the spread acting as the error-correction term, and including as regressors the characteristics of the trades. This model is similar to that of Kavajecz and Odders-White (2001), who estimate the bid and ask quote price changes and the bid and ask quote depths in a system of simultaneous equations, though they specifically

ignore the possibility that the bid and ask prices are cointegrated (Kavajecz and Odders-White, 2001, p. 683). Further, their model is static in the sense that the full reaction is assumed to occur instantly, and they ignore the serial dependence of these variables. These authors use the TORQ dataset, which enables them to include information from the limit order book in their model, but restricts them to analysing the three-month period from November 1, 1990 to January 31, 1991. In our paper, on the other hand, we use eighteen months of data (from January 1997 to June 1998) on a selection of one hundred stocks with a range of trade intensities. In addition to contrasting the impacts of various types of trades, this data set allows us to contrast the differences in the impacts of trades on the quote prices of infrequently and frequently traded stocks.[1]

Modeling the bid and ask prices in a VAR enables us to analyze any asymmetries in the short-run impacts of trades on the bid or the ask price. For example, the view that quote prices move 'one leg at a time' would suggest that a buy has a greater impact on the ask price than on the bid price, while a sell has a greater impact on the bid price than on the ask price. We are able to test this hypothesis using our model. Further, we are able to extract the implied VAR for the spread and the mid-quote from the model for the bid and ask prices, and analyze the impacts of various trade characteristics on these variables as well. The impact of trades on the spread is of particular interest, as it represents a key measure of the amount of friction in the market. Stoll (2000) suggests that the spread be interpreted as the cost of *immediacy*.

The structure of the remainder of the paper is as follows: In Section 2 we describe the trade and quote data used, and the methods employed in preparing it for analysis. Section 3 presents the model and the variables contained therein. Section 4 relates the empirical results to the testable economic hypotheses relevant to this paper, for both the model of the bid and ask prices, and the implied model for the spread and the mid-quote. Finally, we conclude in Section 5.

2. The data

The data used in this paper were taken from the Trades and Quotes (TAQ) dataset, produced by the New York Stock Exchange. In order to obtain a sample of stocks with a range of trade frequencies, we first used data on the total number of trades in the 1997 calendar year on all NYSE listed stocks to determine trade frequency deciles. The average number of trades ranged from 1,090 for Decile 1 to 174,825 for Decile 10. We then randomly selected 25 stocks from each of the second, fourth, sixth and eighth deciles, checking that each stock traded continuously for the entire sample period, and excluding American Depositary Receipts and Real Estate Investment Trusts. The 100 stocks selected via this procedure became the sample for this study. Table 13 in the appendix contains a list of the stocks selected, and Table 1 presents some summary statistics of the data over the sample period, January 1998 to June 1999.

---

[1] The average number of trades per day in our sample ranges from 5.03 for International Aluminum Corp to 450.35 for Donaldson Lufkin Jenrette Inc.

Table 1
Description of the data

		Decile 2	Decile 4	Decile 6	Decile 8
Panel A: *trades*					
Obs:	Min	1,903	4,514	13,818	29,275
	Max	18,203	34,162	54,411	170,232
	Median	3,663	10,549	26,737	56,733
	Mean	4,624	11,543	27,298	60,633
Mean Proportions:	BUYS	0.3872	0.4058	0.4278	0.4388
	SELLS	0.5018	0.4961	0.4654	0.4731
	Mid-quote	0.1110	0.0981	0.1068	0.0880
	V(small)	0.7522	0.8085	0.7704	0.7452
	V(medium)	0.2297	0.1815	0.2112	0.2310
	V(big)	0.0181	0.0100	0.0184	0.0238
	D(short)	0.2589	0.2932	0.3558	0.4744
	D(medium)	0.1303	0.2023	0.3011	0.3425
	D(long)	0.6108	0.5044	0.3430	0.1831
Panel B: *quotes*					
Obs:	Min	3,141	5,083	19,439	33,463
	Max	96,380	42,044	77,100	172,800
	Median	6,427	21,022	43,467	90,465
	Mean	12,981	20,780	44,638	90,676
No price change (%)		0.5601	0.5925	0.4973	0.4519
Panel C: *thinned quotes*					
Obs:	Min	758	2,656	8,893	18,462
	Max	25,589	24,281	39,034	96,817
	Median	3,898	12,934	20,427	37,597
	Mean	6,216	12,264	21,920	40,682
Proportions	Change $< = -2$ ticks	0.1062	0.1343	0.1072	0.0975
	Change $= -1$ tick	0.1760	0.1514	0.1726	0.1801
	Change $= 0$ ticks	0.4401	0.4337	0.4435	0.4483
	Change $= +1$ tick	0.1728	0.1476	0.1699	0.1764
	Change $>= +2$ ticks	0.1049	0.1330	0.1068	0.0977
Median number of trades between quotes		0.7697	0.8624	1.1776	1.4686
Mean number of trades between quotes		1.2137	0.9884	1.2818	1.4397

Note. This table provides an overview of the data set used in this paper. The sample period is January 1998 to June 1999. The top of Panel A reports summary statistics on the number of trades in the sample period for each of the four deciles. For example: the most frequently traded decile 2 stock had 18,203 trades in the sample period. The bottom of Panel A reports the average proportions of different types of trades over the sample period. For example: the average proportion of trades that were designated BUYs across all decile 4 stocks was 40.58%. Panel B reports similar statistics for the full quote data set. Panel C reports similar statistics for the 'thinned' quote data set, which is the full quote data set removing all quotes that did not reflect a change in at least one of the quoted prices.

We remove any trades that occur immediately after the payment of a dividend,[2] or the resumption of trade after a trade halt of some kind. In the case of the former, we do so to remove the impact of the drop in price that occurs following the payment of a dividend, which is essentially a deterministic part of the price dynamics. In the case of the latter, we do so to reflect the fact that the first price after such a halt is probably not generated by the same dynamics as the other prices. As a filter for recording or transcription errors, we exclude any observations that represent a change of more than 50% from the previous observation, if followed by another change of a similar magnitude in the opposite direction.

2.1. Trade data

Although we do not actually model transaction prices in this paper, we do use the characteristics of transactions in our model of quote prices. As such it was necessary to prepare the trade data set as though it were to be modelled. We first removed any trades that occurred with non-standard correction or G127 codes (both of these are fields in the trades data base on the TAQ CDs), such as trades that were cancelled, trades that were recorded out of time sequence, and trades that called for the delivery of the stock at some later date. We cumulated any trades that were recorded with the same time stamp into one trade.[3] To do this we summed the total volume of the trades, attributed it to the first trade, and then removed the other trades from the sample.

Any trades that were recorded to have occurred before 9:30am (the official start of trading on the NYSE) or after 4pm (the official close of trading) were removed. Trade durations were calculated as the difference in seconds between two successive trade time stamps, except for the overnight period. For the overnight period, we calculated the duration as the time between the last trade of the previous day and 4pm plus the time between 9:30am and the time of the first trade on the next day, thus ignoring the entire overnight (and weekend) period.

We constructed BUY and SELL indicator variables for the trades according to a procedure proposed and tested by Lee and Ready (1991) that is now common in the empirical market microstructure literature, see Hasbrouck (1991) and Easley et al. (1996). This procedure identifies the standing quote at a given trade, calculates the mid-quote, and compares the price at which the trade occurred with the mid-quote. If the trade price was higher than the mid-quote, the trade is considered "buyer-initiated". If the trade price was lower than the mid-quote then it is considered

---

[2] As stocks usually go ex-dividend overnight, the trade removed is usually the first trade of the day the stock went ex-dividend.

[3] Simultaneous trading of a stock on two or more exchanges is possible. Such trades are guaranteed to execute at the same price, and thus there is no difficulty in determining the appropriate price to assign to the aggregated trade.

"seller-initiated". Rather than use the 'tick-test' for trades that occurred exactly at the mid-quote, as Lee and Ready (1991) suggest, we instead simply consider these trades to be indeterminate. We also consider all trades that occur before the first quote of the day to be indeterminate. Due to problems in the trade and quote recording process, we follow a suggestion of Lee and Ready and use quotes that are at least five seconds old as the standing quote for a trade.

2.2. Quote data

We use only quotes that were posted at the NYSE in this study, rather than all quotes posted at all exchanges. Blume and Goldstein (1997) find that the NYSE quote, on average, determines or matches the national best quote around 95% of the time, compared with the next best exchange (the Cincinnati exchange) with 11% to 12%. Further, as mentioned in Section 1, Hasbrouck (1995) found that for a sample of 30 stocks the NYSE is responsible for over 92% of the price discovery. The results of these two studies suggest that restricting our attention to NYSE quotes is a reasonable means of reducing the available data to a manageable size. In a similar fashion to the formatting of the trade data, we removed any observations that appeared with non-standard quote modes (a field in the TAQ quote data base) or that were recorded as occurring before 9:30am or after 4pm.

A common feature of microstructure data is the high ratio of the number of quotes in a period to the number of trades. A large proportion of these additional quotes are adjustments to the quote depths at a particular price, and not changes in actual quote prices, see Table 1 for the proportions of quotes with no price changes. For the purposes of the present study, we record a new quote observation whenever one (or both) of the quote prices change.

3. The model

We model the bid and ask prices as a system, similar to Hasbrouck (1995) and Kavajecz and Odders-White (2001), rather than just the mid-quote as is done is some previous studies, see Dufour and Engle (2000) and Hasbrouck (1991). By modeling the two quote prices we are able to derive models for the mid-quote and the spread, and are also able to capture any asymmetries in the dynamics of the two series. It should be noted that although we extend some previous work by modelling both quote prices, we take the intervening trades between quotes to be pre-determined, but do not model them as do Hasbrouck (1991) and Dufour and Engle (2000). One can interpret our model of the bid and ask prices as the first two equations of a larger VAR that includes trades as a third dependent variable. Examining only the first two equations of the larger VAR avoids dealing with the problem of modelling trades with their associated volume and duration characteristics.

Below we present the model used in the analysis of the bid and ask prices. As expected, the log levels of the bid and ask prices were found to be integrated of order

one in most cases,[4] and so the models are specified in terms of log-differences. We estimate the models on the bid and ask series as tenth-order VAR, as below. The variables in the model are described in Table 2.

Quote observations are indexed $t = 1, 2, ..., T$, while trades are indexed according to the quote they precede: $\tau(t) - j$ indexes the jth most recent trade to quote observation t. As the equation below indicates, we include information on the five most recent trades as exogenous regressors in our model. The function $k(t)$ counts the number of trades occurring between quote $t-1$ and quote t. The bottom two rows of Table 1 show the median and mean of $k(t)$ for the four deciles.

$$
\begin{bmatrix} \Delta\log(ASK_t) \\ \Delta\log(BID_t) \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} + \sum_{i=1}^{10} \begin{bmatrix} \alpha_i & \alpha_{i+10} \\ \beta_i & \beta_{i+10} \end{bmatrix} \cdot \begin{bmatrix} \Delta\log(ASK_{t-i}) \\ \Delta\log(BID_{t-i}) \end{bmatrix}
$$

$$
+ \begin{bmatrix} \alpha_{21} \\ \beta_{21} \end{bmatrix} \cdot SPR_{t-1} + \begin{bmatrix} \alpha_{22} \\ \beta_{22} \end{bmatrix} \cdot DPTH\_DIFF_{t-1}
$$

$$
+ \sum_{j=1}^{5} (\Psi \cdot BUY_{\tau(t)-j} + \Phi \cdot SELL_{\tau(t)-j}) \cdot \begin{bmatrix} 1 \\ V^{\mathrm{med}}_{\tau(t)-j} \\ V^{\mathrm{big}}_{\tau(t)-j} \\ D^{\mathrm{sht}}_{\tau(t)-j} \\ D^{\mathrm{lng}}_{\tau(t)-j} \end{bmatrix}
$$

$$
+ \begin{bmatrix} \alpha_{73} & \alpha_{74} \\ \beta_{73} & \beta_{74} \end{bmatrix} \cdot \begin{bmatrix} \sum_{j=1}^{k(t)} BUY_{\tau(t)-j} \\ \sum_{j=1}^{k(t)} SELL_{\tau(t)-j} \end{bmatrix}
$$

$$
+ \sum_{d=1}^{8} \begin{bmatrix} \alpha_{74+d} \\ \beta_{74+d} \end{bmatrix} \cdot DIURN_t^d \tag{1}
$$

where

$$
\Psi = \begin{bmatrix} \alpha_{22+j} & \alpha_{27+j} & \alpha_{32+j} & \alpha_{37+j} & \alpha_{42+j} \\ \beta_{22+j} & \beta_{27+j} & \beta_{32+j} & \beta_{37+j} & \beta_{42+j} \end{bmatrix},
$$

$$
\Phi = \begin{bmatrix} \alpha_{47+j} & \alpha_{52+j} & \alpha_{57+j} & \alpha_{62+j} & \alpha_{67+j} \\ \beta_{47+j} & \beta_{52+j} & \beta_{57+j} & \beta_{562+j} & \beta_{67+j} \end{bmatrix}.
$$

The equation above was estimated using ordinary least squares. We anticipated the presence of heteroskedasticity in the residuals and so computed the

---

[4] We used an Augmented Dickey-Fuller test to determine whether we could reject the null hypothesis of a unit root in the log level of the quote prices, using 10 lags of the change in the variable to remove any serial correlation. For 92 bid series and 91 ask series we could not reject the null hypothesis of the presence of a unit root, using an alpha level of 1%.

variance-covariance matrix using White's (1980) method. This method provides standard errors that are robust to a wide variety of forms of heteroskedasticity.

We are able to derive the coefficient estimates and standard errors of the implied model for the spread and mid-quote by applying the appropriate rotation to these equations. Consider the following simplified example:

$$
\begin{bmatrix} \Delta\log(ASK_t) \\ \Delta\log(BID_t) \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{bmatrix} \cdot \begin{bmatrix} \Delta\log(ASK_{t-1}) \\ \Delta\log(BID_{t-1}) \end{bmatrix}
$$
$$
+ \begin{bmatrix} \alpha_3 \\ \beta_3 \end{bmatrix} \cdot SPR_{t-1} + \begin{bmatrix} \alpha_4 \\ \beta_4 \end{bmatrix} \cdot X_{t-1} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}. \tag{2}
$$

We can obtain a model for the difference in the log-spread, ΔSPR_t, and the log difference in the mid-quote, $\Delta\log(MQ_t)$, by applying the rotation:

$$
\begin{bmatrix} 1 & -1 \\ 0.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} \Delta\log(ASK_t) \\ \Delta\log(BID_t) \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix}
$$
$$
+ \begin{bmatrix} 1 & -1 \\ 0.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{bmatrix} \cdot \begin{bmatrix} \Delta\log(ASK_{t-1}) \\ \Delta\log(BID_{t-1}) \end{bmatrix}
$$
$$
+ \begin{bmatrix} 1 & -1 \\ 0.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} \alpha_3 \\ \beta_3 \end{bmatrix} \cdot SPR_{t-1}
$$
$$
+ \begin{bmatrix} 1 & -1 \\ 0.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} \alpha_4 \\ \beta_4 \end{bmatrix} \cdot X_{t-1}
$$
$$
+ \begin{bmatrix} 1 & -1 \\ 0.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \tag{3}
$$

$$
\begin{bmatrix} \Delta SPR_t \\ \Delta\log(MQ_t) \end{bmatrix} = \begin{bmatrix} \alpha_0 - \beta_0 \\ 0.5(\alpha_0 + \beta_0) \end{bmatrix}
$$
$$
+ \begin{bmatrix} \alpha_1 - \beta_1 & \alpha_2 - \beta_2 \\ 0.5(\alpha_1 + \beta_1) & 0.5(\alpha_2 + \beta_2) \end{bmatrix} \cdot \begin{bmatrix} \Delta\log(ASK_{t-1}) \\ \Delta\log(BID_{t-1}) \end{bmatrix}
$$
$$
+ \begin{bmatrix} \alpha_3 - \beta_3 \\ 0.5(\alpha_3 + \beta_3) \end{bmatrix} \cdot SPR_{t-1}
$$
$$
+ \begin{bmatrix} \alpha_4 - \beta_4 \\ 0.5(\alpha_4 + \beta_4) \end{bmatrix} \cdot X_{t-1} + \begin{bmatrix} \varepsilon_t - \eta_t \\ 0.5(\varepsilon_t + \eta_t) \end{bmatrix}. \tag{4}
$$

We can further manipulate the above expression so as to obtain a model of the *level* of the log-spread, which is more desirable given that this variable is stationary, and the log difference of the mid-quote. We can also re-write the combinations of $\Delta\log(ASK_{t-1})$, $\Delta\log(BID_{t-1})$ and SPR_{t-1} that appear on the right-hand side of the above equation as combinations of $\Delta\log(MQ_{t-1})$, $\Delta\log(MQ_{t-2})$, SPR_{t-1}

Table 2
Description of the variables

Variable	Description
Quote variables	
$\Delta\log(ASK_t)$	Log difference of the ask price between quote t and quote $t-1$.
$\Delta\log(BID_t)$	Log difference of the bid price between quote t and quote $t-1$.
SPR_t	The spread in logs: $\log(ASK_t) - \log(BID_t)$.
$DEPTH\_DIFF_t$	The difference between the log of the quote depth at the ask price and the log of the quote depth at the bid price at quote t.
Trade variables	
$k(t)$	The number of trades between quote t and quote $t-1$.
$\tau(t)-j$	Denotes the jth most recent trade at quote t.
$BUY_{\tau(t)-j}$	Buy indicator: equals 1 if $k(t) \geqslant j$ and the jth most recent trade at quote t was identified as a buy, else this variable equals 0.
$SELL_{\tau(t)-j}$	Sell indicator: equals 1 if $k(t) \geqslant j$ and the jth most recent trade at quote t was identified as a sell, else this variable equals 0.
$V^{\text{med}}_{\tau(t)-j}$	Medium volume trade indicator: equals 1 if the jth most recent trade at quote t had volume between 1,000 and 10,000 shares, else this variable equals 0.
$V^{\text{big}}_{\tau(t)-j}$	Large volume trade indicator: equals 1 if the jth most recent trade at quote t had volume of over 10,000 shares, else this variable equals 0.
$D^{\text{sht}}_{\tau(t)-j}$	Short duration trade indicator: equals 1 if the jth most recent trade at quote t had a duration less than 60 seconds, else this variable equals 0.
$D^{\text{lng}}_{\tau(t)-j}$	Long duration trade indicator: : equals 1 if the jth most recent trade at quote t had duration longer than 5 minutes, else this variable equals 0.
Deterministic variables	
$DIURN^d_t$	Diurnal adjustment variable: the value of the dth diurnal indicator variable at quote t.

Note. This table describes the variable names used in the specification of the model in Eq. (1).

and SPR_{t-2}.

$$
\begin{bmatrix} SPR_t \\ \Delta\log(MQ_t) \end{bmatrix}
$$

$$
= \begin{bmatrix} \alpha_0 - \beta_0 \\ 0.5(\alpha_0 + \beta_0) \end{bmatrix} + \begin{bmatrix} \alpha_4 - \beta_4 \\ 0.5(\alpha_4 + \beta_4) \end{bmatrix} \cdot X_{t-1}
$$

$$
+ \begin{bmatrix} 1 + 0.5(\alpha_1 - \beta_1 - \alpha_2 + \beta_2) + \alpha_3 - \beta_3 & \alpha_1 - \beta_1 + \alpha_2 - \beta_2 \\ 0.25(\alpha_1 + \beta_1 - \alpha_2 - \beta_2) + 0.5(\alpha_3 + \beta_3) & 0.5(\alpha_1 + \beta_1 + \alpha_2 + \beta_2) \end{bmatrix}
$$

$$
\cdot \begin{bmatrix} SPR_{t-1} \\ \Delta\log(MQ_{t-1}) \end{bmatrix}
$$

$$
+ \begin{bmatrix} -0.5(\alpha_1 - \beta_1 - \alpha_2 + \beta_2) & 0 \\ -0.25(\alpha_1 + \beta_1 - \alpha_2 - \beta_2) & 0 \end{bmatrix} \cdot \begin{bmatrix} SPR_{t-2} \\ \Delta\log(MQ_{t-2}) \end{bmatrix}
$$

$$
+ \begin{bmatrix} \varepsilon_t - \eta_t \\ 0.5(\varepsilon_t + \eta_t) \end{bmatrix}. \tag{5}
$$

Thus our simplified VAR(1) model for the log-difference of the bid and ask prices is 'rotated' into a VAR(2) model for the log-level of the spread and the log-difference of the mid-quote. The coefficients of the implied VAR(2) model of the spread and the mid-quote are simply linear combinations of the coefficients in the model for the bid and the ask prices and so we can obtain the means and standard deviations of the implied coefficients quite easily. We apply the above reasoning to the model given in Eq. (1) to obtain the model for the spread and the mid-quote implied by the results for the model of the bid and ask prices. The model for the spread is particularly interesting as the spread is a common measure of friction in the market. A large spread indicates a high cost to a trader wishing to execute a trade quickly; a high cost of *immediacy* in Stoll's (2000) terminology.

4. Empirical results

The model presented above was estimated on all 100 stocks. Space constraints prevent us from presenting the complete estimation results here; the reader is referred to Engle and Patton (2000) for the complete results on four of the one hundred stocks. The complete results on all one hundred stocks are available upon request. The results for a particular variable for the entire sample of stocks are presented in the form of the median coefficient for each decile, along with a count of the number of times the coefficient was positive and significant, and negative and significant. This format enables us to draw general cross-sectional conclusions as to the significance and magnitude of coefficients across deciles.

4.1. Specification tests

The regressions were all significant, with all F-statistics (testing for the joint significance of the variables in the model) being large. The R^2 statistics ranged from 0.15 to 0.54, and averaged 0.30 over the 200 equations. We tested for serial correlation using the Ljung and Box (1978). This test is asymptotically equivalent to the more common LM test, but is simpler to compute. For all but one stock, we find no evidence of serial correlation in the residuals at the 1% alpha level, using ten lags. This indicates that the inclusion of ten lags of the dependent variables and five lags of the trade indicator variables was sufficient to capture the commonly observed negative serial correlation in microstructure data.[5] We use White's (1980) robust standard errors as these series all exhibit heteroskedasticity.

4.2. Evidence of error-correction behavior

It is generally accepted that the log-levels of price series are integrated, and that the log-spread should be stationary—we know that the spread cannot take negative

---

[5] We previously estimated the same model with only one lag of the trade indicator variables, and found substantial evidence of serial correlation. Stoll (2000) also finds evidence of negative serial correlation in quote price changes.

values, and we also expect it to remain finite. These facts combined suggest that the log bid and ask prices of a stock on a given market are cointegrated.[6] The lagged spread is the natural, though not unique, object to consider as the error-correction term. We thus expect a large spread at the previous quote to lead to a rise in the bid price and a fall in the ask price at the following quote, to restore the spread to its long-run equilibrium value. A small spread should do the opposite.

Easley and O'Hara (1992) suggest something similar to this, in their second proposition. They assert that the bid and ask price will converge toward the equilibrium stock price in periods with no trades, which implies that the equilibrium spread in the absence of trades is zero. Jang and Venkatesh (1991) also allude to error-correcting behavior, when they write that a decrease in the spread is more likely than an increase when the spread is greater than some threshold (three-eighths of a dollar), and more likely to increase when it is below some threshold (one-quarter of a dollar).

The empirical results show that, as postulated, a high spread leads to a decrease in the ask price and an increase in the bid price, moving the spread toward its equilibrium value: in the entire sample of 100 stocks (and thus 200 quote price series) we find that that the coefficient on the spread is significant and of the hypothesized sign 187 times. This is very strong support for the importance of this variable in describing quote price dynamics. This result complements the descriptive results presented in Jang and Venkatesh (1991), and supports Easley and O'Hara's (1992) proposition.

The magnitudes of the coefficients are approximately the same across stocks and deciles, as can be seen in Table 3 above. The similar magnitudes of these coefficients across the deciles indicate that the speed (in event time) of adjustment of quote prices from a large or small spread is not affected by average trade frequency. However, as fewer events happen per day for the lower deciles, the calendar time taken for the quotes to adjust is greater for infrequently traded stocks than frequently traded stocks.

4.3. Asymmetric impact of BUYs and SELLs

Economic theory suggests that a BUY should have a positive impact on both the bid and the ask prices, while a SELL should have a negative impact. However, there are conflicting theories as to relative impacts of BUYs or SELLs on the bid and ask prices: Glosten and Milgrom (1985) find that a BUY (for example) causes the mid-quote to rise, and the spread to fall, as the uncertainty regarding the information is partially reduced by the trade. This suggests that both the bid and the ask prices move up, and that the bid price increases by *more* than the ask price. Alternatively, one could imagine a situation where a BUY leads to an *increase* in uncertainty as the market maker becomes unsure whether an information event has occurred and that

---

[6]Augmented Dickey-Fuller tests on the log spread, including 10 lags of the change in the spread to account for serial correlation, indicate that we can reject the null hypothesis that the spread is I(1) for all 100 stocks at the 1% alpha level.

Table 3
Coefficients on the SPR_{t-1} variable

	Decile 2		Decile 4		Decile 6		Decile 8	
	Ask	*Bid*	*Ask*	*Bid*	*Ask*	*Bid*	*Ask*	*Bid*
Median	−0.1333	0.1220	−0.1342	0.1640	−0.1446	0.1444	−0.1569	0.1359
Signif pos	0	22	0	23	0	25	0	25
Signif neg	19	0	23	0	25	0	25	0

Note. *Signif pos* and *Signif neg* count the number of times this variable was significant at the 1% level, and of the indicated sign. The maximum possible is 25.

Table 4
Coefficients on the $BUY_{\tau(t)-1}$ and $SELL_{\tau(t)-1}$ variables

$BUY_{\tau(t)-1}$	ASK				BID			
	Dec 2	*Dec 4*	*Dec 6*	*Dec 8*	*Dec 2*	*Dec 4*	*Dec 6*	*Dec 8*
Median	0.2146	0.1179	0.0904	0.0624	0.1192	0.0637	0.0431	0.0290
Signif pos	21	24	25	25	18	22	24	25
Signif neg	0	0	0	0	0	0	0	0
$SELL_{\tau(t)-1}$								
Median	−0.0844	−0.0504	−0.0399	−0.0297	−0.1822	−0.0940	−0.0743	−0.0525
Signif pos	0	0	0	0	0	0	0	0
Signif neg	18	22	24	25	17	24	25	25

Note. *Signif pos* and *Signif neg* count the number of times this variable was significant at the 1% level, and of the indicated sign. The maximum possible is 25.

the current price is below the new equilibrium price, or that the trade was merely a liquidity trade. In such a situation we would expect to see both the mid-quote and the spread initially increase, implying that the ask price would rise, and that either the bid price would rise by less than the ask price, or even possibly fall.

From the median results across each decile presented in Table 4, we observe that both the BUY variables and the SELL variables all have the a priori expected signs: the BUY variables all have positive signs, indicating that a buyer-initiated trade raises both the bid and the ask prices, while a seller-initiated trade lowers the quote prices. This is in accordance with the models of Glosten and Milgrom (1985) and Huang and Stoll (1994), *inter alia*. Further, we can see that the coefficients on the BUY variables are more significant and greater in magnitude that the SELL variables in the models for the ask price. The reverse is true for the models for the bid price. This asymmetry in the impacts of BUYs and SELLs on the ask and bid prices would not have been as easily detectable in a model for the spread and mid-quote. This result suggests that a trade increases the uncertainty in the market: a BUY leads to a larger increase in the ask price than in the bid price, for example, so the mid-quote and the spread both rise. This contradicts the theoretical implications of the model of Glosten and Milgrom (1985): our results suggest that observing an agent

Table 5
Coefficients on $BUY_{\tau(t)-1}$ and $SELL_{\tau(t)-1}$ variables in the models for the spread and mid-quote

Spread	BUY				SELL			
	Dec 2	Dec 4	Dec 6	Dec 8	Dec 2	Dec 4	Dec 6	Dec 8
Median	0.0782	0.0589	0.0480	0.0325	0.0376	0.0481	0.0352	0.0236
Signif pos	9	22	24	25	7	17	24	23
Signif neg	0	0	0	0	2	0	0	0
Mid-quote								
Median	0.1639	0.0917	0.0668	0.0468	−0.1473	−0.0687	−0.0566	−0.0403
Signif pos	21	24	25	25	0	0	0	0
Signif neg	0	0	0	0	20	24	25	25

Note. *Signif pos* and *Signif neg* count the number of times this variable was significant at the 1% level, and of the indicated sign. The maximum possible is 25.

willing to trade at the current quoted prices increases the uncertainty about the true price of the stock, leading to a wider spread.

There is a little evidence of asymmetry in the impacts of BUYs and SELLs on the spread and the change in the mid-quote, see Table 5 above. In the models we find some weak evidence that a BUY has a larger short-run impact than a SELL. For the other trade variables (those including indicator variables for the volume or the duration of the trade) we find no evidence of asymmetry in the impacts of BUYs or SELLs on either the spread or the mid-quote.

These results are inconsistent with the theoretical models of Glosten and Milgrom (1985) and Easley and O'Hara (1992) that assume a single information event per trade day, and so the only uncertainty the market maker faces is what type of information event occurred. In these models, trades reduce uncertainty. Our results suggest that there may be many information events per day, and that the market maker faces two types of uncertainty: the first relates to *whether* an information event has occurred since he last adjusted his quote prices, and the second relates to the *type* of information event that has occurred, given that an information event has indeed occurred.

Kavajecz and Odders-White (2001) identify their model by restricting BUY variables to appear only in the ask price equation, and SELL variables to appear only in the bid price equation. We do not impose this restriction in our model, and find that although the magnitude of the impact of a SELL in the ask equation is smaller than that of a BUY, for example, it is still significant for all but the least frequently traded stocks. A similar result is found for the impact of a BUY on the bid price. We do find that the SELL variables that include indicator variables for the volume or the duration of the trade are generally not significant in the ask price model and similarly for the BUY variables in the bid equation. In presenting the results for the remaining trade coefficients we will present only those multiplied by a BUY in the ask price model, and only those multiplied by a SELL in the bid price model, in the interests of parsimony.

4.4. Does trade size matter?

In order to answer this question we introduce two indicator variables designed to reflect the size of the trade. The V^{med} variable is one if the trade volume was between 1,000 and 10,000 shares and zero otherwise, while the V^{big} variable is one if the trade volume is greater than 10,000 shares and zero otherwise. The literature is not unanimous on the effect of trade size on price impacts. Hasbrouck (1991) finds that larger trade volumes increase the spread more than smaller volumes, while Barclay and Warner (1993) and Keim and Madhavan (1996) suggest that a quadratic relation may be more appropriate. Barclay and Warner (1993) find that medium volume trades (defined as those with volume between 500 and 9900 shares; very similar to our definition) drive most of the cumulative stock price movements, and suggest that informed traders break up their trades so as to remain less conspicuous. Keim and Madhavan provide evidence of information leakage as block trades are 'shopped' in the upstairs market.[7] Information leakage occurs when the price of the stock rises or falls prior to the execution of the block trade as the broker 'shops' the block trade, thus revealing some of the information the block trade carries before it is actually executed.

Our empirical results indicate that trade size does matter, but only up to a point. We find that the coefficient on the V^{med} variable is generally significant, when multiplied by the appropriate trade indicator (BUY or SELL) variable; see Table 6 below. This indicates that a trade with volume between 1,000 shares and 10,000 shares carries more quote price relevant information than a trade of less than 1,000 shares. This finding is consistent with many previous studies: Easley and O'Hara (1987), Hasbrouck (1991), and Barclay and Warner (1993), amongst others, though it contradicts the findings of Easley et al. (1997). Notice that the magnitude of the coefficient on the medium volume variable decreases as we increase the trade frequency, indicating increased liquidity in the market for more frequently traded stocks.

Table 7 below presents the results for the coefficients on the large volume trade indicator. The coefficients on this variable are less often significant than those on the medium volume trade indicator, particularly for the less frequently traded stocks. Recall that the proportion of trades with very large volumes is only slightly larger for deciles 6 and 8, as reported in Table 1, but that the total number of trades for stocks in these deciles is also larger. Thus there are many more large volume trades for the frequently traded stocks than for the infrequently traded stocks. This larger number provides a possible reason for the increased significance of the V^{big} variable in the higher deciles: the coefficient may in truth be different from zero in the lower deciles also, but we do not have enough observations of such large trades to estimate it accurately.[8]

---

[7] The 'downstairs' market is the regular stock market, where trades are accomplished via buying or selling through the market maker. The upstairs market, on the other hand, is a market for very high volume trades (minimum 10,000 shares) where the transaction is carried out by a broker or intermediary, who locates (potentially many) counter-parties to the trade.

[8] In fact, for one stock (Fort Dearborn Income Securities) we did not observe *any* SELLs with volume greater than 10,000 in the sample period, meaning that for this stock we had to force the impact of a large trade to be symmetric—we used the variable (BUY-SELL)$*V^{big}$ rather than each variable separately.

Table 6
Coefficients on the V^{med} variable

	Ask models: $BUY * V^{\text{med}}$				Bid models: $SELL * V^{\text{med}}$			
	Dec 2	Dec 4	Dec 6	Dec 8	Dec 2	Dec 4	Dec 6	Dec 8
Median	0.1049	0.0757	0.0515	0.0465	−0.1650	−0.0759	−0.0651	−0.0454
Signif pos	15	24	25	25	0	0	0	0
Signif neg	0	0	0	0	20	25	25	25

Note. *Signif pos* and *Signif neg* count the number of times this variable was significant at the 1% level, and of the indicated sign. The maximum possible is 25.

Table 7
Coefficients on the V^{big} variable

	Ask models: $BUY * V^{\text{big}}$				Bid models: $SELL * V^{\text{big}}$			
	Dec 2	Dec 4	Dec 6	Dec 8	Dec 2	Dec 4	Dec 6	Dec 8
Median	0.1028	0.0827	0.0615	0.0587	−0.1126	−0.0670	−0.0601	−0.0526
Signif pos	7	10	19	25	0	0	0	0
Signif neg	0	0	0	0	7	9	18	25

Note. *Signif pos* and *Signif neg* count the number of times this variable was significant at the 1% level, and of the indicated sign. The maximum possible is 25 for all deciles except Decile 2, which has a potential maximum of 24 (see footnote 8).

The magnitudes of the coefficients on the medium and large volume trade indicators are roughly equal; certainly close enough that we cannot detect a significant statistical difference between the two. Thus we can reject the hypothesis that quote price impacts are monotonic in trade volume.

The results for the implied models for both the spread and the mid-quote indicate that trade size is important for both. Again we find that medium volume trades generally have significant coefficients in all deciles, with the expected signs, see Table 8. The significance increases as we increase the average trade frequency, a feature that seems common to all of our results. Table 8 further shows that the decrease in the magnitude and increase in significance of the coefficients on medium volume trades are also present in the model for the mid-quote.

The coefficients on the large volume trades, presented in summary form in Table 9, were only rarely significant in the lower two deciles, and only marginally significant in decile 6. The magnitudes of the coefficients on large volume trades indicate that large volume trades have roughly the same short-run impact as medium volume trades on both the spread and the mid-quote.

4.5. What kind of news is no news?

In addition to considering the volume of a trade, we also examine the impact of the duration of a trade, defined as the time since the preceding trade. Theoretical

Table 8
Coefficients on the $BUY*V^{\mathrm{med}}$ and $SELL*V^{\mathrm{med}}$ variable in the model for the spread and mid-quote

Spread	BUY				SELL			
	Dec 2	Dec 4	Dec 6	Dec 8	Dec 2	Dec 4	Dec 6	Dec 8
Median	0.0743	0.0503	0.0382	0.0349	0.1148	0.0484	0.0459	0.0390
Signif pos	9	18	25	25	16	20	25	25
Signif neg	1	0	0	0	0	0	0	0
Mid-quote								
Median	0.0677	0.0469	0.0398	0.0280	−0.1076	−0.0515	−0.0399	−0.0259
Signif pos	13	22	25	25	0	0	0	0
Signif neg	0	0	0	0	19	24	25	25

Note. *Signif pos* and *Signif neg* count the number of times this variable was significant at the 1% level, and of the indicated sign. The maximum possible is 25.

Table 9
Coefficients on the $BUY*V^{\mathrm{big}}$ and $SELL*V^{\mathrm{big}}$ variable in the models for the spread and mid-quote

Spread	BUY				SELL			
	Dec 2	Dec 4	Dec 6	Dec 8	Dec 2	Dec 4	Dec 6	Dec 8
Median	0.1127	0.0444	0.0319	0.0462	0.0120	0.0362	0.0269	0.0382
Signif pos	6	6	14	22	6	4	12	23
Signif neg	1	0	0	0	0	0	0	0
Mid-quote								
Median	0.0705	0.0578	0.0398	0.0387	−0.0980	−0.0402	−0.0409	−0.0317
Signif pos	5	4	18	24	0	0	0	0
Signif neg	0	0	0	0	7	7	18	23

Note. *Signif pos* and *Signif neg* count the number of times this variable was significant at the 1% level, and of the indicated sign. The maximum possible is 25 for all deciles except Decile 2, which has a potential maximum of 24 (see footnote 8).

studies by Diamond and Verrecchia (1987) and Easley and O'Hara (1992) suggest that the time between trades conveys information about the type of news a trade carries. Diamond and Verrecchia suggest that long times between trades reflect bad news, as short selling restrictions prevent traders from selling a stock on the basis of bad news. Easley and O'Hara, on the other hand, propose that long durations signal neither bad nor good news. Empirical studies by Dufour and Engle (2000) and Engle (2000) have found evidence to support the Easley and O'Hara hypothesis that long durations imply simply no news. We include a short duration indicator variable for durations of less than 60 seconds and a long duration indicator for durations longer than 5 minutes to capture the effects documented in the above studies.

Our empirical results suggest that no trades imply no news. We find that the long duration variable is significant only 14 times out of a potential 200. The few times

that we do find it significant, the coefficient is usually of the opposite sign to the coefficient on the lagged BUY and lagged SELL variables. This indicates that a trade occurring after a period longer than five minutes with no trades has either the same impact (coefficient not significant) or slightly less impact (coefficient significant and of the opposite sign to the lagged BUY/SELL variable) than a medium duration trade. These findings add further support to the theoretical assertion of Easley and O'Hara (1992) that long durations mean no news, and so a trade that occurs after a long duration is likely to be a liquidity trade rather than one based on valuable information.

Short durations, on the other hand, are significant in most of the stocks' results when multiplied by the appropriate trade indicator variable, see Table 10. In all cases we find that the sign of the coefficient is such that a short duration trade has a larger impact than a medium or long duration trade, consistent with the findings of Dufour and Engle (2000).

As the above table shows, the coefficient on the short duration variable (multiplied by the appropriate BUY/SELL indicator) decreases in magnitude, but increases in significance, as we increase the average trade frequency. Short durations are more rare in the lower deciles than in the higher deciles, by construction, and so we would expect trades with short durations to have a larger impact on the price of infrequently traded stocks than they do on frequently traded stocks.

An alternative explanation of the increased magnitudes of the coefficients on short duration trades in the lower deciles is provided by Easley et al. (1996) who find evidence that the risk of trading with an informed agent is higher for infrequently traded stocks than it is for frequently traded stocks. Such an increase would explain why trades have a larger quote price impact for less frequently traded stocks.

For the mid-quote and the spread we again find that no trades mean no news; only a handful of the coefficients on the long duration indicator are significant in both the model for the spread and the model for the mid-quote. This implies that a trade with duration longer than 5 minutes has the same impact on the spread and the mid-quote as a trade with duration of between 60 seconds and 5 minutes.

Short duration trades exhibit the same trends that we observed in the models for the bid and the ask: the magnitude of the coefficients decreases as we increase the average trade frequency, but the significance increases (Table 11).

Table 10
Coefficients on the D^{sht} variable

	Ask models: $BUY * D^{\text{sht}}$				Bid models: $SELL * D^{\text{sht}}$			
	Dec 2	Dec 4	Dec 6	Dec 8	Dec 2	Dec 4	Dec 6	Dec 8
Median	0.0660	0.0308	0.0206	0.0112	−0.0850	−0.0457	−0.0239	−0.0125
Signif pos	6	13	19	23	0	0	0	0
Signif neg	0	0	0	0	12	16	20	20

Note. *Signif pos* and *Signif neg* count the number of times this variable was significant at the 1% level, and of the indicated sign. The maximum possible is 25.

Table 11
Coefficients on the $BUY * D^{sht}$ and $SELL * D^{sht}$ variable in the models for the spread and mid-quote

Spread	BUY				SELL			
	Dec 2	Dec 4	Dec 6	Dec 8	Dec 2	Dec 4	Dec 6	Dec 8
Median	0.0433	0.0199	0.0158	0.0145	0.0776	0.0317	0.0189	0.0131
Signif pos	4	7	15	23	6	12	15	21
Signif neg	0	0	0	0	0	0	0	0
Mid-quote								
Median	0.0352	0.0245	0.0122	0.0070	−0.0450	−0.0284	−0.0127	−0.0056
Signif pos	4	10	12	18	0	0	0	0
Signif neg	0	0	0	0	7	16	12	15

Note. *Signif pos* and *Signif neg* count the number of times this variable was significant at the 1% level, and of the indicated sign. The maximum possible is 25.

4.6. Inventory balance vs. asymmetric information effects

In our model above we include a variable first suggested by Huang and Stoll (1994) to determine which of the inventory balance and asymmetric information effects is the strongest. This variable is the difference between the log of the quote depth at the ask price and the log of the quote depth at the bid price. The inventory effect asserts that a market maker with excess inventory will simultaneously lower the ask price and raise the ask depth, in order to attract buyers. At the next quote the ask price is raised back to its previous level. A similar argument for the bid price holds in the case when the market maker has too small an inventory. Thus the inventory effect suggests that the coefficient on the difference between the ask depth and the bid depth will be positive: excess depth at the ask will lead to a rise in the ask price, while excess depth at the bid will lead to a fall. The asymmetric information effect, on the other hand, asserts that the impact on quote prices of an excess of supply at the ask over that at the bid is negative. High depth at the ask price potentially indicates a number of sellers on the limit order book, suggesting to the market that the stock is overpriced. Huang and Stoll also note that the presence of the barrier effect would have the same impact as that of the asymmetric information effect.[9] Higher depth at the ask price than at the bid price means less trade volume is required before a downward movement than an upward movement. Thus the *barrier* to a downward movement is weaker than that to an upward movement, making a downward movement more likely. The same logic applies when the depth at the bid price is greater than that at the ask price.

We combine the results of the bid price models and the ask price models, as they are very similar. For 170 out of the 200 quote price models estimated, and for 86 of the 100 implied models for the mid-quote, we find that the coefficient on this variable is negative and significant. This is consistent with the asymmetric information

---

[9] See O'Hara (1995) for more detail on these three market microstructure theories.

hypothesis dominating the inventory balance effect. While the sign and significance of this variable is consistent across the deciles studied, the magnitude of the variable exhibits some variation. Table 12 below presents some summary statistics on this coefficient across the deciles.

The decrease in the absolute magnitude of this coefficient as we move upwards through the trade frequency deciles suggests that the strength of the asymmetric information (or barrier effect) relative to the inventory balance effect is greater for less frequently traded stocks. This is consistent with the Easley et al. (1996) evidence suggesting that the risk of trading with an informed trader is higher for infrequently traded stocks.

4.7. Deterministic time-of-day effects

In a manner similar to seasonality in macroeconomic variables, intra-daily data may exhibit "intra-daily seasonality", which is more accurately called *diurnality*. In an effort to capture any deterministic component of the intra-day dynamics of the variables under analysis, we use piece-wise linear splines to reflect the time of trade day that the observation appeared. Previous studies, see Engle and Russell (1998) and Dufour and Engle (2000), have also used a similar method of diurnal adjustment. The nodes of the splines are the start of the trade day, 9:30am, and then 10am, 11am, mid-day, 1pm, 2pm, 3pm, 3:30pm and the close of the trade day, 4pm.

Chan et al. (1995) report that the spread on frequently traded NYSE stocks display a U-shape, that is, that spreads are larger at the beginning and end of the trade day than they are in the middle. We can determine whether this holds when controlling for the regressors described in Table 2, and also look for any systematic patterns that emerge between stocks with different average trade frequencies. The significance of the diurnal variables in our model will indicate whether time-of-day effects are important for quote price revisions.

We find substantial evidence of increased spreads at the beginning of the trade day (all but four stocks have a significant negative coefficient on the first diurnal adjustment variable) but we find no evidence of an increase in average spreads towards the end of the trade day. The spline coefficient corresponding to the opening thirty minutes of trading was generally significant in the models of the bid and ask prices (132 times out of 200) but very few others, suggesting that a deterministic

Table 12
Coefficients on the *DPTH_DIFF* variable

	Decile 2	Decile 4	Decile 6	Decile 8
Median	−0.0670	−0.0402	−0.0289	−0.0238
Signif pos	4	8	10	6
Signif neg	45	41	40	44

Note. *Signif pos* and *Signif neg* count the number of times this variable was significant at the 1% level, and of the indicated sign. The maximum possible is 50.

time-of-day effect is not an important source of variation in the quote price series. Similar results were found for the mid-quote models. Dufour and Engle (2000) and Hasbrouck (1999) conclude similarly that only the beginning of the trade day displays a significant deterministic component.

5. Conclusion

In this paper we conducted an empirical investigation of the quote price dynamics of a range of NYSE-listed stocks. We used TAQ data from January 1998 through June 1999 on 100 stocks sorted by trade frequency deciles, based on 1997 data. We estimated a vector autoregression on the log-difference of the bid and the ask prices with the lagged log spread as the error-correction term, and included as regressors the lagged difference in the depths posted at the ask and bid prices and indicator variables for the volume and duration and direction of the most recent trades. Using this framework we were also able to extract the implied model for the spread and the mid-quote, via a simple rotation.

We found that the lagged (log) spread was a very significant error-correction term for the bid and the ask prices in all of the stocks' models. It had the appropriate sign indicating that large spreads lead to falls in the ask price and rises in the bid price, while conversely, small spreads lead to increases in the ask price and falls in the bid price. The event-time speed of adjustment of the spread and mid-quote was comparable across average trade frequencies, though in calendar time the more illiquid stocks took much longer to recover from a trade than the liquid stocks. Although the idea of cointegration analysis in empirical market microstructure is not new, we believe this to be the first study to exploit role of the spread as an error-correction term in a VAR of quote prices.

Strong evidence of differential short-run impacts of buyer-initiated trades versus seller-initiated trades on the bid and the ask prices was found. A BUY increased the mid-quote while a SELL decreased the mid-quote, just as in virtually all studies and theoretical models. We also found that a trade of either type temporarily increased the spread, which stands in contrast to the theoretical results of Glosten and Milgrom (1985). The result suggests the possibility that there are potentially multiple information events per day. The overall pattern suggested by these findings supports the common observation that quote prices move "one leg at a time".

We constructed indicator variables for trade direction, duration and volume, breaking them into short, medium and long, and small, medium and large, respectively. Short duration and medium volume trades were found to be the most significant, and to generally have the largest short-run impacts, for all the deciles. Large volume trades had a significant impact on the bid and ask prices of the frequently traded stocks, but not on the infrequently traded stocks. Long durations were not found to have a significantly different impact than medium durations. We consistently found that while the magnitudes of the coefficients in the lower deciles tended to be greater, the significance of these coefficients was reduced. The greater magnitudes of the coefficients in the illiquid stocks' models was taken to be

indicative of a greater risk of informational asymmetry in the market for these stocks, as suggested by Easley et al. (1996).

We allowed for deterministic time-of-day effects in quoted prices, through the inclusion of eight diurnal adjustment indicator variables. These variables were generally not significant in the models for the bid, ask and mid-quote, though some significance was found for them in the model for the spread. Specifically, spreads seem to be higher at the start of the trade day.

Appendix A

Names of the companies are given in Table 13.

Table 13
Companies included in the sample

Ticker	Company name	Mean number of trades/day:	
		Jan 1997 to Dec 1997	Jan 1998 to June 1999
Panel A: decile 2			
HTD	HUNTINGDON LIFE SCIENCE GP	8.38	7.04
BE	BENGUET CORP	11.54	15.68
PDC	PRESLEY COMPANIES	8.45	14.53
ABG	GROUPE AB S A ADS	11.55	5.62
LSB	LSB INDUSTRIES INC	8.52	8.47
GBX	GREENBRIER COMPANIES INC	9.89	14.20
STC	STEWART INFORMATION SVCS CORP	8.00	48.16
DTC	DOMTAR INC	10.17	11.34
PCZ	PETRO-CANADA VARIABLE VTG SHS	12.35	12.89
TGN	TRIGEN ENERGY CORP COMMON	7.29	8.78
SGD	SCOTTS LIQUID GOLD INC	12.52	11.31
OFG	ORIENTAL FINL GRP HOLD CO.	10.93	19.69
JNS	CHIC BY H.I.S. INC	9.45	10.74
JAX	J ALEXANDER S CORP.	7.55	7.46
HUN	HUNT CORP	11.29	16.13
TPR	TRANSPRO INC.	11.79	8.40
VHI	VALHI INC	10.08	9.47
MIG	MEADOWBROOK INSURANCE GRP INC	10.30	14.33
PIC	PICCADILLY CAFETERIAS INC	12.19	9.42
GOT	GOTTSCHALKS INC	12.05	9.63
GSE	GUNDLE/SLT ENVIROMENTAL INC.	8.91	5.32
CSS	CSS INDUSTRIES INC	9.42	9.68
SAJ	ST JOSEPH LIGHT POWER CO	10.81	12.84
FTD	FORT DEARBORN INCOME SECS	10.44	9.69
IAL	INTERNATIONAL ALUMINUM CORP	7.27	5.03
Panel B: decile 4			
CNE	CONNECTICUT ENERGY CORP	21.04	27.46
TEC	COMMERCIAL INTERTECH CORP	25.69	37.02

Table 13 (*continued*)

Ticker	Company name	Mean number of trades/day:	
		Jan 1997 to Dec 1997	Jan 1998 to June 1999
JC	JENNY CRAIG INC	24.07	16.67
NRD	NORD RESOURCES CORP	20.36	17.28
LSH	LASALLE RE HOLDINGS LTD	20.09	33.54
PCU	SOUTHERN PERU COPPER CORP	23.75	27.91
BOR	BORG WARNER SECURITIES CORP	19.47	26.93
FC	FRANKLIN COVEY CO.	20.93	38.63
CHP	C D TECHNOLOGIES INC.	20.88	34.25
CGI	COMMERCE GROUP INC	19.80	31.71
XTR	XTRA CORP	21.44	34.52
FIC	FAIR ISAAC AND CO INC	23.13	46.62
FED	FIRSTFED FINANCIAL CORP	19.65	39.73
OSG	OVERSEAS SHIPHOLDING GROUP	26.86	33.58
BKE	BUCKLE INC	19.63	90.38
BNK	CNB BANCSHARES INC	20.64	56.11
UAH	UNITED AMER HEALTHCARE CORP	24.61	13.12
RGC	REPUBLIC GROUP INC	25.29	26.62
RDO	RDO EQUIPMENT CO	19.79	17.64
OXM	OXFORD INDUSTRIES INC	25.48	21.30
FMN	F M NATIONAL CORP	24.09	31.68
FEP	FRANKLIN ELECTRONIC PUBLISHER	19.56	11.94
PTC	PAR TECHNOLOGY CORP	25.63	15.81
TSY	TECH SYM CORP	20.02	13.44
BBR	BUTLER MANUFACTURING CO	22.58	19.56
Panel C: decile 6			
TBY	TCBY ENTERPRISE INC	50.12	65.69
OMM	OMI CORPORATION NEW	39.50	41.27
FUN	CEDAR FAIR DEP R L.P.	51.73	63.46
GRO	MISSISSIPPI CHEMICAL CORP.	50.53	38.43
DGX	QUEST DIAGNOSTICS INC.	48.44	36.56
WSO	WATSCO INC	41.11	47.00
ASL	ASHANTI GOLDFLDS	48.50	89.55
FA	FAIRCHILD CORP CL	46.21	54.12
MPP	GENERAL CIGAR HOLDINGS CL	55.33	57.34
CDI	C D I CORP	44.04	52.00
IEI	INDIANA ENERGY INC HLDG CO	49.79	51.53
LUK	LEUCADIA NATIONAL CORP	45.63	77.59
CWC	CARIBINER INTERNATIONAL INC	43.85	81.12
RDK	RUDDICK CORP	51.71	70.73
WRC	WORLD COLOR PRESS INC	46.54	76.68
BUR	BURLINGTON INDS INC	48.82	83.88
CSL	CARLISLE COMPANIES INC	51.63	91.04
OC	ORION CAPITAL CORP	40.04	73.84
PNM	PUBLIC SERVICE NEW MEXICO	49.12	69.56
CNA	CNA FINANCIAL CORP	39.26	70.38
PNR	PENTAIR INC	39.09	77.13
ZLC	ZALE CORP	55.55	102.04
RYN	RAYONIER INC	49.24	73.17

Table 13 (*continued*)

Ticker	Company name	Mean number of trades/day:	
		Jan 1997 to Dec 1997	Jan 1998 to June 1999
LIN	LINENS N THINGS INC.	44.01	143.94
PMS	POLICY MANAGEMENT SYSTEMS CORP	41.46	117.37
Panel D: decile 8			
RGR	STURM RUGER CO INC	83.23	77.47
AVX	AVX CORP	96.50	77.45
WNC	WABASH NATIONAL CORP	86.97	86.39
SEI	SEITEL INC	113.66	97.97
ARG	AIRGAS INC	111.91	93.31
FLM	FLEMING COS INC	83.88	93.11
BRR	BARRETT RESOURCES CORP	84.71	85.16
LTV	LTV CORP NEW	113.90	131.87
PMI	PREMARK INTERNATIONAL INC	84.12	95.89
TRN	TRINITY INDUSTRIES	111.73	152.31
ASD	AMERICAN STANDARD COS INC	119.90	120.76
TCB	TCF FINANCIAL CORP	83.14	150.09
R	RYDER SYSTEM INC	117.41	173.02
VTS	VERITAS DGC INC.	122.57	232.06
SNC	SNYDER COMMUNICATIONS INC.	82.61	197.24
GAS	NICOR INCORPORATED	86.63	112.22
AVT	AVNET INC	110.48	159.55
GAP	GREAT ATLANTIC PAC TEA	92.77	127.89
DLP	DELTA AND PINE LAND COMPANY	91.20	176.24
COX	COX COMMUNICATIONS INC	82.94	245.96
FMO	FEDERAL-MOGUL CORP	86.08	191.18
CTX	CENTEX CORP	107.34	212.75
CP	CANADIAN PACIFIC LTD ORD NEW	123.38	194.47
CNS	CONSOLIDATED STORES CORP	106.44	275.42
DLJ	DONALDSON LUF JENRETTE INC	93.45	450.35

Note. This table presents the 'ticker' code and full name of the 100 stocks that were used in this paper. Also presented are the average trade frequencies of each of these stocks during 1997 (the period used to form the average trade frequency deciles, from which we randomly selected 25 Decile 2, 4, 6 and 8 stocks) and the average trade frequencies of the stocks during the sample period, January 1998 to June 1999.

References

Barclay, M.J., Warner, J., 1993. Stealth trading and volatility. Journal of Financial Economics 34 (3), 281–305.

Blume, M.E., Goldstein, M.A., 1997. Quotes, order flow, and price discovery. Journal of Finance 52 (1), 221–244.

Chan, K., Chung, Y.P., Johnson, H., 1995. The intraday behavior of bid-ask spreads for NYSE stocks and CBOE options. Journal of Financial and Quantitative Analysis 30 (3), 329–346.

Diamond, D.W., Verrecchia, R.E., 1987. Constraints on short-selling and asset price adjustment to private information. Journal of Financial Economics 18, 277–311.

Dufour, A., Engle, R.F., 2000. Time and the price impact of a trade. Journal of Finance 55, 2467–2498.

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

Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. Journal of Finance 47 (2), 577–605.

Easley, D., Kiefer, N.M., O'Hara, M., Paperman, J.P., 1996. Liquidity, information and infrequently traded stocks. Journal of Finance 51 (4), 1405–1436.

Easley, D., Kiefer, N.M., O'Hara, M., 1997. One day in the life of a very common stock. Review of Financial Studies 10 (3), 805–835.

Engle, R.F., 2000. The econometrics of ultra high frequency data. Econometrica 68, 1–22.

Engle, R.F., Lange, J., 2001. Predicting VNET: a model of the dynamics of market depth. Journal of Financial Markets 4 (2), 113–142.

Engle, R.F., Patton, A.J., 2000. Impacts of Trades in an Error-Correction Model of Quote Prices. Discussion Paper 2000-26, Department of Economics, University of California, San Diego.

Engle, R.F., Russell, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66 (5), 1127–1162.

Escribano, A., Pascual, R., 2000. Asymmetries in bid-ask responses to innovations in the trading process, manuscript, Departamento de Economia, Universidad Carlos III de Madrid.

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

Hasbrouck, J., 1991. Measuring the information content of stock trades. Journal of Finance 46 (1), 179–207.

Hasbrouck, J., 1995. One security, many markets: determining the contributions to price discovery. Journal of Finance 50 (4), 1175–1198.

Hasbrouck, J., 1999. The dynamics of discrete bid and ask quotes. Journal of Finance 54 (6), 2109–2142.

Huang, R.D., Stoll, H.R., 1994. Market microstructure and stock return predictions. Review of Financial Studies 7 (1), 179–213.

Jang, H., Venkatesh, P.C., 1991. Consistency between predicted and actual bid-ask quote revisions. Journal of Finance 46 (1), 433–446.

Kavajecz, K.A., 1999. A specialist's quoted depth and the limit order book. Journal of Finance 2, 747–771.

Kavajecz, K.A., Odders-White, E.R., 2001. An examination of changes in specialists' posted price schedules. Review of Financial Studies 14 (3), 681–704.

Keim, D.B., Madhavan, A., 1996. The upstairs market for large-block transactions: analysis and measurement of price effects. Review of Financial Studies 9 (1), 1–36.

Lee, C.M., Ready, M.J., 1991. Inferring trade direction from intraday data. Journal of Finance 46 (2), 733–746.

Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65 (2), 297–303.

O'Hara, M., 1995. Market Microstructure Theory. Blackwell Publishers Inc., MA.

Stoll, H.R., 2000. Friction. Journal of Finance 55 (4), 1479–1514.

White, H., 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 48, 817–838.