Estimation of Order-Handling Costs on the Nasdaq Stock Exchange

Jorim Schraven

Université Toulouse 1 des Sciences Sociales

June 2002

Abstract

This paper estimates order handling costs in financial markets based on the model by Roll (1984). The empirical innovation is that this study estimates the model completely without making assumptions on the probabilities of sizes and directions of market orders as Roll did in his original paper. This study analyses micro-structure data on 46 Nasdaq 100 stocks over a period of one month. The estimated half traded spreads found by the method employed here are significantly lower than those arrived at by the Roll method.

Introduction[1]:

Order handling costs provide valuable information with regards to the institutional efficiency of financial markets. They contribute to friction involved in security trading. Friction is the price concession paid for an immediate transaction (Demsetz, 1968). Total friction could be seen as a broad indicator of the efficiency of the trading process of a particular institutional configuration. These institutional features can be divided into 'real' frictions, due to the expenditure of real economic resources in the trading process, such as order-handling and inventory costs and market power, and 'informational' frictions arising from adverse selection or the 'option-element' in limit orders. The source of friction is important for asset pricing purposes. Just as inefficiencies and costs reduce profits in normal production processes, real friction must reduce asset prices to generate returns sufficiently high to offset the real cost of trading and holding the asset for the holding period (Stoll, 2000:1483).

The quoted spread, the difference between the best ask and bid price quoted on the limit order book, provides a static measure of friction. It measures what a trader would earn on a round-trip of two trades. This assumes that quotes are not adjusted in response to trades. Informational effects would require adjustments of quotes to trades because trades convey information about the underlying fundamental value of the asset. Moreover, inventory costs may motivate downward quote adjustments in response to a trade at the bid to equilibrate inventories (reducing the expected price-change due to the next trade) and increasing the probability of a following trade at the ask (increasing the expected price-change due to the next trade). Only when both effects balance out and there are no informational effects, does the quoted spread reflect real friction. Otherwise, (still assuming that the inventory effects balance) the traded spread (the difference between average trade prices at the ask and at the bid) provides a more accurate estimate of the real friction because it reflects what the broker actually earns on a round-trip of two trades. The difference between the quoted spread and the traded spread, estimated over a longer time horizon, provides an estimate of informational effects.

[1] The analysis of friction in this section is mainly based on Stoll (2000).

There exists, however, another dynamic approach to estimating the spread caused by real friction due to Roll (1984). Suppose that friction was completely due to real factors and that neither quotes nor the underlying true value adjust in the short term. Then, each trade at the bid is executed at the same price, which is different from the price at the ask. Consequently, serial correlation of prices must be negative (this idea is shown more formally in the next section) due to the ensuing 'bid-ask bounce'. Conversely, Glosten and Milgrom (1985) first showed that if friction is completely informational, there will be no bid-ask bounce because the transaction price would be a martingale, and consequently the serial correlation of prices would be zero. Hence, the negative serial covariance of prices is an indicator of real friction. The next section established its relation to the traded spread.

Theoretical model[2]:

Suppose N risk neutral competitive market makers place schedules of limit orders incurring order handling costs c(q).

Assume:

A1) $c(q) = \frac{c}{2} q^2$

A2) $Fundamental - value(\tilde{v}) : E(\tilde{v}) = \pi$

Then, the individual market maker i has the following objective for price p:

1) $$\max_{q_i} \quad E[(\tilde{v} - p)q_i] - \tfrac{c}{2} q_i^2$$

2) F.O.C.: $q_i = \dfrac{\pi - p}{c}$ S.O.C.: $-c<0$

[2] This model is based on the idea of Roll (1984) and its basic features are set out in Biais, Glosten and Spatt (2001:6).

Impose a market clearing condition for a market order Q:

3) $$Q+\sum_{i=1}^{N} q_i = 0 \Leftrightarrow Q + N\frac{\pi - p}{c} = 0 \Leftrightarrow p(Q) = Q\frac{c}{N} + \pi$$

This entails a spread (S) of:

4) $$S(Q) = 2Q\frac{c}{N}$$

Empirical strategy:

We assume that the fundamental value of the stock follows a martingale and that the market orders are identically and independently distributed. Then we have:

5) $p_t(Q_t) = Q_t\frac{c}{N} + \pi_t$ where $\pi_t = E(\tilde{v}|H_t)$ and H_t is the information set.

We proceed to estimate the order handling costs c by the serial covariance strategy outlined above:

6) $$Cov(\Delta P_{t+1}, \Delta P_t) = Cov[(\pi_{t+1} - \pi_t) + \tfrac{c}{N}(Q_{t+1} - Q_t), (\pi_t - \pi_{t-1}) + \tfrac{c}{N}(Q_t - Q_{t-1})]$$

We can eliminate the (π_{t+1} –π_t) and (π_t –π_{t-1}) elements because the fundamental value follows a random walk. Similarly, Cov (Q_{t+1}, Q_t)= Cov (Q_{t+}, Q_{t-1})= Cov (Q_t, Q_{t-1})=0 because of the assumption that market orders are i.i.d. Thus, we have:

7) $$Cov(\Delta P_{t+1}, \Delta P_t) = Cov(-\tfrac{c}{N}Q_t, \tfrac{c}{N}Q_t) = -\tfrac{c}{N}^2 Var(Q_t)$$

Hence, to estimate order handling costs we must estimate:

8) $$\frac{c}{N} = \sqrt{\frac{-Cov(\Delta P_{t+1}, \Delta P_t)}{Var(Q_t)}}$$

Our estimated traded half spread is:

9) $$Avg(Q) * \frac{c}{N} = Avg(Q) * \sqrt{\frac{-Cov(\Delta P_{t+1}, \Delta P_t)}{Var(Q_t)}}$$

And finally, the percentage average cost is estimated as:

10) $$\frac{\frac{Avg(Q)}{N} * \frac{c}{2}}{Avg(P)} = \frac{\frac{Avg(Q)}{2} * \sqrt{\frac{-Cov(\Delta P_{t+1}, \Delta P_t)}{Var(Q_t)}}}{Avg(P)}$$

Previous research: findings of Roll (1984) and Stoll (2000):

Based on a study on 1,706 NYSE/AMSE stocks and 2,184 Nasdaq stocks, Stoll (2000:1493) finds that 100% of Nasdaq stocks in each size category bar one exhibit negative serial covariances, whereas for the NSYE/AMSE stocks no size category has fewer than 96% of stocks with a negative serial covariance. This is a very strong result and is consistent with the model set out above.

Roll did not have data on the quantities of each order but only data on daily price changes. Given these restrictions, he assumed that respective orders at the ask and bid came with equal probabilities of 0.5 and in one unit quantities. This yields $Var(Q_t)=1$ and $S = -2\sqrt{Cov}$, and allows the estimation of order handling costs based on the available data. Using this method he finds spreads of 0.298% for daily data and 1.74% for weekly data. However, the importance of the bid-ask bounce in generating serial covariance on data with such a large time horizon is likely to be low, yielding an inaccurate estimator. Stoll (2000:1493 footnote), does use data on all trades in the day, which explains why he obtains a much higher incidence of negative covariance than Roll did. Curiously, Stoll does not use the information on the variance of market order sizes but relies on Roll's assumptions and subsequent finding of $S = -2\sqrt{Cov}$ instead. This study estimates the model completely without making such assumptions.

Data:

Data are obtained from the NASTRAQ data set distributed by the Nasdaq. The period covered is April 1, 2000, to April 30, 2000, comprising 19 trading days. In addition data has been extracted from Datastream and obtained from Equity Analysis Ltd. The procedure was to obtain serial covariances of price changes and variances of market order sizes for each stock for each day. The initial sample of stocks comprised the NASDAQ 100, which were all traded on each trading day in the sample. See table I in appendix I for a listing of the eliminated stocks, and the reasons for which they were eliminated. The final sample consists of 46 stocks or 874 daily observations.

In order to calculate the variance of market order sizes, each trade was associated with the nearest inside quote preceding the trade by at least 5 seconds. According to Lee and Ready (1991) this method performs relatively well[3]. When a trade price was inside the inside quote it was attributed a sell (buy) if closer to the bid (ask). When exactly at the middle, sell or buy status was attributed randomly according to the prevailing proportion of buy and sell orders for that stock on that particular day. In order to avoid double associations, whenever there were two inside quotes for one stock with exactly the same time-stamp one was eliminated randomly. Only trades executed and inside quotes reported during trading hours between 9:30 and 16:00 were considered. In addition, trades were eliminated for the following reasons: late trade indicators (such trades presumably carry different order handling costs); Executed at another market than the Nasdaq (namely Chicago, for the concern of inaccurate time-stamps); and where the difference between the execution and reporting time exceeded 10 seconds (because this is an indication that it may not have been a market order). Finally, trades associated to inside quotes with a negative quoted bid-ask spread were eliminated. See the Appendix II for details on how these eliminations were carried out exactly.

[3] Alternatively the 'tick-test' could have been applied, using information about the price changes in subsequent trades to attribute buy or sale status to trades at the mid-quote.

Results:

This paper set out to estimate the order-handling costs. There are two principal ways in which to measure these, the traded-spread and the percentage average cost. The estimated traded spread provides a measure of how much it costs to conduct a 'round-trip' trade. Half-traded spreads are reported in dollar-cents for each stock in table II. These results can be compared to those in Stoll (2000:1494), who reports half-traded spreads for ten market value deciles. The most appropriate comparison is with the largest two deciles, considering that the Nasdaq 100 includes the larger companies on the Nasdaq. The estimated half-traded spreads are lower but of the same order of magnitude as those found by Stoll. The same applies to the half-quoted spreads reported. In fact, the estimated half-traded spread is larger than the reported half-quoted spread for many of the stocks in this study but Stoll finds the same for the top two market value deciles.

Even though we have relatively few observations per stock, 42 out of 46 stocks have significant t-values[4] at 5%. The variance of the half-traded spreads over all stocks, although much lower than the variance of serial covariances of price changes, is sufficiently high to make the average value insignificant. Note that the variance of serial covariances of price changes is much higher than the variance of half-traded spreads, which have been calculated using in addition the variance of market order quantities. This suggests that the current method is superior. More data is needed to confirm this. Moreover, considering that Stoll bases his estimates on the serial covariances of price changes, calculated in much the same way as in this study, and that he finds the average half-traded spreads over stocks to be highly significant, we can expect that more data would certainly yield significant averages for the current method also.

There may be another way to reduce the variance of the results, which is suggested by the fact that the variance of percentage average order handling costs is much lower than the variance of half-traded spreads. The direct suggestion would be that the order handling costs may depend on the price of the stock as well. More likely is that the price is correlated with the market size and trade volume associated with a stock and that these affect order handling

[4] Critical t-values have been extracted from the student t-distribution

costs. These effects are not encompassed by the model under investigation and hence not explicitly estimated. In order to test these possibilities, sufficient data is needed to divide stocks into size and volume categories. Moreover, one can question the validity of the theoretical model, which for instance implies that average order handling costs increase with the size of the order. Stronger empirical results may be obtained with other theoretical specifications with more intuitive implications. Finally, the variance in the results may be reduced by deleting outliers. This approach has not been taken in this paper because particularly large orders do belong to the prevailing order handling structure and the author had no objective means or justification for deleting such orders.

To summarise, both for the values of estimated half-traded spreads and of percentage average costs, 45 out of 46 are significantly different from zero. This contrasts with 34 out of 46 for the serial covariance of price changes. The average value overall of half-traded spreads is 3.08 cents but is not significantly different from zero because of the heterogeneity of stocks in the Nasdaq 100 and the relatively low number of data points. Moreover, considering the crude nature of statistical testing employed here -simply based on the number of covariance estimates in each sample- we must be careful in putting too much trust into these t-values. The average serial covariances of price changes are negative for all stocks in the sample but the t-value on the overall average is not significant. The associated average Roll price is with 8.74 about three times as high as the estimated half-traded spread. Additionally, the overall percentage average cost is found to be 0.049% and is significantly different from zero.

Conclusion:

This study finds the order handling costs of stock trading based on a simple order handling model. It builds on previous research but goes further in estimating the model completely without making strong assumptions on the sizes and directions of trades. The half traded spreads estimated according to this method suggest that order handling costs are two thirds lower than what had been estimated based on previously employed methods. In addition, the current method achieves higher levels of significance. Nevertheless, this study is limited in scope and extent compared to earlier research and more extensive analysis of a larger dataset is needed to firmly confirm its findings. Moreover, the data suggests that, potentially, order handling costs depend on a wider set of variables. Neither the theoretical underpinnings nor the empirical strategy of this study allow for such a possibility, hence its validity could not be confirmed.

Bibliography:

BIAIS, Bruno, (1993), “Price Formation And The Equilibrium Liquidity In Fragmented And Centralized Markets”, *Journal Of Finance* 48, 157–186.

BIAIS, B., Glosten, L. and Psatt, C. (2001), “The Microstructure of Stock Markets”, Mimeo.

DEMSETZ, H. (1968), “The Cost Of Transacting”, *Quarterly Journal of Economics*, Vol. 82:33-53.

LEE, Charles M. C., and Mark J. Ready, (1991), “Inferring trade direction from intraday data”, *Journal of Finance* 46, 733–746.

GLOSTEN, Lawrence R., and Paul R. Milgrom, (1985), “Bid, Ask And Transaction Prices In A Specialist Market With Heterogeneously Informed Traders”, *Journal of Financial Economics* 14, 71–100.

ROLL, Richard, 1984, “A Simple Implicit Measure Of The Effective Bid-Ask Spread In An Efficient Market”, *Journal of Finance* 39, 1127–1139.

STOLL, H. (2000), “Friction”, *Journal of Finance*, Vol. 55(4), 1479-1514.

Appendix I: Tables

Table I:

Change in	number of	outstanding	stocks[5]:	Undocumented[6]:	Stock split[7]:	Dividends[8]:
ADBE	CEFT	GENZ	QTRN	ADLAC	MCLD	
AMAT	CMGI	IMNX	RNWK	ATHM	MFNX	
AMCC	CMVT	ITWO	SEBL	BMCS		
AMGN	CNET	JDSU	SIAL	GBLX		
AMZN	COMS	MLHR	SSCC	LGTOE		
APCC	CTAS	NXTL	USAI	NETA		
APOL	DELL	ORCL	VRTS	NSOL		
BGEN	DISH	PHSY	WCOM	NTLI		
BMET	EBAY	PSFT	YHOO	NXLK		
BVSN	FISV	QCOM		SDLI		
				VISX		
				VSTR		
				VTSS		

[5] Source: Datastream (a change in the number of outstanding stocks may indicate a stock split, stock buy back or stock issue, each of which would affect the price of the stock and bias the results by affecting the mean price change)

[6] For these stocks there was uncertainty as to whether they had stock issues, splits or dividends.

[7] Source: Equity Analysis Ltd.: http://www.e-analytics.com/split200/stsp0428.htm

[8] Source: Equity Analysis Ltd.: http://www.e-analytics.com/ none of the stocks, which had not already been eliminated by the other methods was reported to have paid dividends over the sample period.

Table II:

Stock name	Ticker	Quoted spread	Effective spread	Traded spread	t-value	% avg cost	t-value	Cov	t-value	Roll price
APPLE COMPUTERS	AAPL	12.47	8.90	3.56	3.52	0.0292	3.25	-0.00954	-2.78	9.77
ADC TELECOM.	ADCT	9.90	6.48	2.40	4.75	0.0463	5.27	-0.00505	-2.09	7.11
ADAPTEC	ADPT	9.85	5.59	1.73	2.93	0.0531	3.36	-0.00324	-1.15	5.69
ALTERA	ALTR	10.12	7.56	2.48	5.01	0.0277	4.49	-0.00618	-3.20	7.86
ATMEL	ATML	9.24	6.28	2.25	5.34	0.0451	4.71	-0.00425	-3.05	6.52
BED BATH & BEYOND	BBBY	9.76	5.50	1.66	3.63	0.0423	3.65	-0.00269	-3.40	5.19
CHIRON CORP	CHIR	12.36	7.53	2.90	3.16	0.0647	3.29	-0.00464	-2.91	6.81
CIENA	CIEN	15.76	12.98	5.61	3.70	0.0538	2.92	-0.02304	-2.44	15.18
COMCAST SPECIAL 'A'	CMCSK	7.45	4.10	1.08	3.67	0.0275	3.95	-0.00171	-1.60	4.13
CONEXANT SYS.	CNXT	11.06	8.16	2.47	2.46	0.0409	2.01	-0.00781	-2.36	8.84
COSTCO WHSL.	COST	9.37	5.54	1.63	3.64	0.0305	3.22	-0.00337	-3.13	5.81
COMPUWARE	CPWR	6.49	3.43	0.76	3.32	0.0511	5.20	-0.00077	-3.15	2.78
CISCO SYSTEMS	CSCO	6.48	4.74	1.28	2.71	0.0188	2.54	-0.00276	-1.06	5.25
CITRIX SYS.	CTXS	12.87	8.92	3.30	3.49	0.0498	3.75	-0.01099	-2.53	10.48
DOLLAR TREE STORES	DLTR	14.74	8.32	2.66	3.51	0.0515	3.24	-0.00522	-3.04	7.23
ERICSSON LM	ERICY	10.91	6.46	2.21	4.23	0.0267	3.59	-0.00539	-2.32	7.34
ELECTRONIC ARTS	ERTS	14.10	8.81	3.15	5.51	0.0522	5.12	-0.00919	-3.22	9.59
GEMSTAR TV GUIDE INTL.	GMST	12.06	7.93	2.78	3.25	0.0541	3.26	-0.00915	-2.35	9.56
INTEL	INTC	7.21	6.06	1.85	4.91	0.0147	4.11	-0.00431	-1.82	6.56
INTUIT	INTU	10.08	6.57	1.96	2.07	0.0500	2.18	-0.00475	-2.17	6.89
KLA TENCOR	KLAC	10.70	8.19	2.88	4.62	0.0383	3.78	-0.00801	-2.26	8.95
LYCOS	LCOS	12.43	8.14	2.70	3.85	0.0581	4.24	-0.00639	-3.22	7.99
LINEAR TECH.	LLTC	11.10	7.12	2.78	5.15	0.0571	4.68	-0.00553	-3.44	7.44
LEVEL 3 COMMS.	LVLT	16.63	11.04	4.04	3.50	0.0487	3.64	-0.01484	-2.60	12.18
MICROCHIP TECHNOLOGY	MCHP	16.07	10.09	4.05	6.36	0.0639	6.49	-0.00961	-2.55	9.80
MEDIMMUNE	MEDI	37.64	24.01	11.24	4.47	0.0719	4.87	-0.07483	-3.03	27.36
MOLEX	MOLX	15.28	8.26	3.42	2.48	0.0645	2.38	-0.00581	-2.52	7.62
MICROSOFT	MSFT	6.69	4.14	0.96	4.59	0.0121	5.16	-0.00182	-2.85	4.26
MAXIM INTEG.PRODUCTS	MXIM	12.88	8.33	3.17	2.95	0.0504	3.12	-0.00771	-1.84	8.78
NOVELL	NOVL	7.18	4.15	1.15	4.07	0.0527	4.14	-0.00151	-2.36	3.89
NETWORK APPLIANCE	NTAP	14.66	10.72	4.19	4.34	0.0654	4.34	-0.01710	-2.41	13.08
NORTHWEST AIRLINES	NWAC	9.43	4.95	1.32	3.26	0.0565	3.34	-0.00151	-2.57	3.89
PAYCHEX	PAYX	11.90	7.23	2.27	3.67	0.0445	3.61	-0.00434	-3.23	6.59
PACCAR	PCAR	14.95	7.92	3.41	3.27	0.0725	2.96	-0.00490	-1.55	7.00
PMC-SIERRA	PMCS	22.78	20.75	9.03	3.31	0.0554	3.01	-0.05667	-2.01	23.80
PARAMETRIC TECH.	PMTC	5.04	2.77	0.51	4.20	0.0544	4.92	-0.00052	-2.87	2.29
QLOGIC	QLGC	27.79	18.52	8.89	2.97	0.0971	4.10	-0.04841	-1.92	22.00
RF MICRO DEVICES	RFMD	28.44	19.28	8.90	2.56	0.0885	2.59	-0.04842	-1.59	22.01
SANMINA-SCI	SANM	13.41	8.48	3.31	2.99	0.0581	3.27	-0.00833	-2.54	9.12
STARBUCKS	SBUX	8.39	4.92	1.62	3.09	0.0433	3.14	-0.00205	-3.35	4.53
SYNOPSYS	SNPS	11.11	6.28	2.24	2.86	0.0530	2.70	-0.00364	-3.56	6.03
STAPLES	SPLS	6.95	3.66	0.81	3.06	0.0421	3.21	-0.00087	-4.28	2.95
PANAMSAT NEW	SPOT	19.15	10.96	4.12	3.15	0.0904	3.20	-0.00902	-2.21	9.50
SUN MICROSYSTEMS	SUNW	7.48	6.24	2.12	3.80	0.0240	3.42	-0.00455	-1.88	6.75
TELLABS	TLAB	8.57	5.86	1.66	2.64	0.0322	2.58	-0.00430	-1.52	6.56
XILINX	XLNX	10.57	8.08	3.04	3.93	0.0422	3.65	-0.00805	-2.75	8.97

Appendix II:

Microsoft Access 2000 SQL-queries:

<u>Generate clean trades trades table:</u>

```
SELECT [trcd11].[rptdate] AS rptdate, [trcd11].[tsec] AS
tsec, [trcd11].[mid] AS mid, [trcd11].[extime] AS extime,
[trcd11].[ent_vol] AS ent_vole, [trcd11].[entprice] AS
entprice INTO tr24
FROM trcd11
WHERE (((([trcd11].[extime])>TimeSerial(9,30,0) And
([trcd11].[extime])<TimeSerial(16,0,0)) And
((Abs([rpttime]-[extime]))<TimeSerial(0,0,10)) And
(([trcd11].[rptmod])="") And
(([trcd11].[marketcd])<>"M"));
```

<u>Generate clean inside quote table[9]:</u>

```
INSERT INTO iq27 ( rptdate, tsec, mid, qtime, bid, ask ) IN
'd:db12a.mdb'
SELECT iqcd10.rptdate, iqcd10.tsec, iqcd10.mid,
iqcd10.qtime, iqcd10.bid, iqcd10.ask
FROM iqcd10
WHERE (((iqcd10.qtime)>=TimeSerial(9,30,0) And
(iqcd10.qtime)<TimeSerial(16,0,0)));
```

<u>Associating trades with inside quotes[10]:</u>

[9] In addition double records with the same stock, date, and time-stamp are eliminated by putting a primary key on the stock, date, and time-stamp fields and copying the data into this format.

[10] Note that this is a particularly computationally demanding procedure, when replicating this type of research one would be advised to use different (faster) software.

```
SELECT Min([extime]-[qtime]) AS mindif, tr.mid, tr.extime,
tr.ent_vol, tr.entprice
FROM iq, tr
WHERE (((([extime]-[qtime])>=TimeSerial(0,0,5)) And
((tr.tsec)=[iq].[tsec]))
GROUP BY tr.mid, tr.extime, tr.ent_vol, tr.entprice;

SELECT tr.mid, tr.rptdate, tr.tsec, tr.extime, tr.ent_vol,
tr.entprice, [iq].[bid], [iq].[ask], [iq].[qtime],
IIf([min].entprice-[bid]>=[ask]-
[min].entprice,[min].ent_vol,-1*[min].ent_vol) AS Quantity
INTO trades24 IN 'd:\trades24.mdb'
FROM iq, [Min] INNER JOIN tr ON ([Min].mid=tr.mid) AND
([Min].extime=tr.extime) AND ([Min].ent_vol=tr.ent_vol)
AND ([Min].entprice=tr.entprice)
WHERE (((tr.tsec)=[iq].[tsec]) And (([min].extime-
[qtime])=[mindif]));
```

Stata 6.0 do-files:

Reading data and performing basic calculations:

```
set mem 100m

set more off

* reading data

insheet id date stock time vol price bid ask qtime quantity
using E:trades.txt, t

* generating relevant variables

sort date stock time

by date stock: gen dprice=price-price[_n-1]

by date stock: gen lagdp=dprice[_n-1]

by date stock: gen dask=ask-ask[_n-1]
```

```
by date stock: gen lagdask=dask[_n-1]

by date stock: gen dbid=bid-bid[_n-1]

by date stock: gen lagdbid=dbid[_n-1]

by date stock: gen spread=ask-bid

by date stock: gen efspread=abs((ask+bid)/2-price)

* dropping unacceptable spreads

drop if spread<0


* distributing trades at the mid-quote according to the distributions of sales and buys of this day

gen sale=1 if quantity<0 & price~=(ask+bid)/2

gen buy=1 if quantity>0 & price~=(ask+bid)/2

replace quantity = -1*quantity if price==(bid+ask)/2 & uniform()<sum(sale)/(sum(buy)+sum(sale))

drop buy sale


* generate quoted and effective spread data

log using e:\spread.txt, append

by date stock: sum spread

by date stock: sum efspread

sum spread

sum efspread

log close


* results by stock

log using e:\paskbid.txt, append

by date stock: cor lagdp dprice, c mean

by date stock: cor dbid lagdbid, c mean

by date stock: cor dask lagdask, c mean

by date stock: cor quantity quantity, c mean
```

log close

Statistical analysis of the daily averaged data:[11]

Generate averages and standard deviations of price and volume:

```
INSERT INTO avgvol ( rptdate, tsec, GemVanbid, GemVanask, GemVanentprice, GemVanent_vol, StDevVanbid, StDevVanask, StDevVanentprice, StDevVanent_vol, AantalVanent_vol ) IN 'd:\varcov.mdb'

SELECT trades.rptdate, trades.tsec, Avg(trades.bid) AS GemVanbid, Avg(trades.ask) AS GemVanask, Avg(trades.entprice) AS GemVanentprice, Avg(trades.ent_vol) AS GemVanent_vol, StDev(trades.bid) AS StDevVanbid, StDev(trades.ask) AS StDevVanask, StDev(trades.entprice) AS StDevVanentprice, StDev(trades.ent_vol) AS StDevVanent_vol, Count(trades.ent_vol) AS AantalVanent_vol
FROM trades
WHERE ((([ask]-[bid])>=0))
GROUP BY trades.rptdate, trades.tsec;
```

Generate mean traded spread and mean percentage average cost:

```
SELECT avgvol.rptdate, avgvol.tsec, 100*Sqr(-[covp]/[varq])*[gemvanent_vol] AS mtspread, fulldata1.covp, fulldata1.obs, avgvol.GemVanentprice, avgvol.GemVanent_vol, Sqr(-
```

[11] The data generated by the previous do-file is saved in log files, which are then read into Microsoft Excel 2000 and reformatted into usable database format to be re-fed into Stata using an automated macro (code available upon request). Clearly, these procedures would be much easier implemented through a single, more flexible, software program such as SAS.

```
[covp]/[varq])*[gemvanent_vol]*74 AS avgcost, 100*Sqr(-
[covp]/[varq])*[gemvanent_vol]/[gemvanentprice] AS
pctavgcost
FROM fulldata1 INNER JOIN avgvol ON (fulldata1.stock =
avgvol.tsec) AND (fulldata1.date = avgvol.rptdate)
WHERE (((fulldata1.covp)<=0));
```

<u>Generate table and t-values:</u>

```
SELECT tradedspreads.tsec, Avg(tradedspreads.mtspread)
AS GemVanmtspread, Avg([mtspread])/StDev([mtspread])
AS tmtspread, Avg(tradedspreads.pctavgcost) AS
GemVanpctavgcost, Avg([pctavgcost])/StDev([pctavgcost])
AS tpctavgcost, Avg(tradedspreads.covp) AS GemVancovp,
Avg([covp])/StDev([covp]) AS tcovp,
StDev(tradedspreads.mtspread) AS StDevVanmtspread,
StDev(tradedspreads.covp) AS StDevVancovp,
StDev(tradedspreads.pctavgcost) AS StDevVanpctavgcost
FROM tradedspreads
GROUP BY tradedspreads.tsec;
```