E C O N O M I C S B U L L E T I N

The impact of trading mechanisms and stock characteristics on order processing and information costs: A panel GMM approach

Gerhard Kling
Utrecht School of Economics

Abstract

My study provides a panel approach to quantify the impact of trading mechanisms and stock characteristics on spread components. Based on the two−way decomposition of Huang and Stoll (1997), a cross−sectional dimension is added. Arrelano and Bover's (1995) dynamic GMM procedure and the Helmert's transformation allow controlling for company specific effects. In line with former research, I confirm higher order processing costs on the NASDAQ. My model identifies the reasons for higher information costs on dealer markets, namely lower market capitalization and less attention of financial analysts. Yet the trading mechanism itself is not responsible for higher information costs.

I thank Thomas Augustin, Jaap Bos, Clemens Kool, and Sven Rady for their helpful comments.
Citation: Kling, Gerhard, (2005) "The impact of trading mechanisms and stock characteristics on order processing and information costs: A panel GMM approach." *Economics Bulletin,* Vol. 7, No. 5 pp. 1−11
Submitted: June 6, 2005. **Accepted:** July 22, 2005.
URL: http://www.economicsbulletin.com/2005/volume7/EB−05G10006A.pdf

1. Introduction

Transactions executed on auction markets, i.e. the NYSE, or on dealer markets like the NASDAQ differ in transaction costs.[1] The properties of these trading systems could affect order processing, inventory, and information costs as well as market liquidity. My paper tries to detect differences in components of transaction costs that stem from trading mechanisms and stock characteristics.

A common procedure to quantify the components of transaction costs is decomposing the bid-ask spread into its components: order processing, inventory, and information costs. These components are widely discussed in the theoretical literature. Besides order processing costs, inventory costs are theoretically justified (Ho and Stoll, 1981, 1983) – but empirically often neglected.[2] Copeland and Galai (1983) as well as Glosten and Milgrom (1985) show that, even if inventory and order processing costs are neglected, the resulting bid-ask spread should be positive due to information costs.

Comparing order processing costs and the scale of information asymmetry between dealer and auction markets is not a new topic; however, one central aspect has not been analyzed thus far, namely the simultaneous influence of trading mechanisms and stock characteristics on spread components. The standard procedure is to isolate the impact of trading mechanisms and stock characteristics by matched sampling. Affleck-Graves et al. (1994), Huang and Stoll (1996), Bessembinder and Kaufman (1997), and Bessembinder (1999) among others use matched samples based on firm size, trading volume, share prices, and other criteria. They obtain a pair of stocks with similar stock characteristics, but traded on different exchanges. If spreads and spread components differ between two matched stocks, trading mechanisms are made responsible for that. Through matching of samples, interesting information is lost. For instance, stocks with low market capitalization usually exhibit higher spreads – but this relation might depend on trading mechanisms. My paper tries to determine such interdependencies among stock characteristics, trading mechanisms, and spread components.

As I try to uncover the impact of stock characteristics on spread components, I have to deal with cross-sectional differences. Henceforth, a panel data approach suggests itself, for time series data (transaction data) are required to estimate spread components and cross-sectional data (individual stocks) enable to investigate cross-sectional differences (i.e. market capitalization differs among stocks). My paper tries to address this issue by applying a panel data approach that is an extension of the two-way decomposition model developed by Huang and Stoll (1997), which is a time series model.[3] The GMM methods by Huang and Stoll (1997) as well as Madhavan et al. (1997) do not allow a panel structure consisting of companies and successive intra-daily transactions. Hence, a panel GMM approach is required, and one has to deal with alleged company specific effects. Using Arrelano and Bover's (1995) dynamic GMM estimation procedure and a Helmert's transformation, company specific effects can be eliminated. At the same time, one obtains GMM estimates for spread components and can reveal partial impacts of stock characteristics and trading mechanisms on these components.

Due to the fact that Affleck-Graves et al. (1994), Huang and Stoll (1996), Bessembinder and Kaufman (1997), and Bessembinder (1999) construct matched samples, I do the same in order to allow comparisons between former and new empirical findings. In a time series approach, the matching procedure enables to avoid biased estimates due to uncontrolled stock

[1] Harris (2002) discussed the different trading mechanisms in detail.
[2] Empirical models like the serial covariance model of George et al. (1991) neglect inventory costs.
[3] Consider that the serial covariance models of Stoll (1989) and George et al. (1991) cannot cope with the impact of stock characteristics on spread components. The non-linear relationship between the estimated parameters and the calculated components does not permit a direct inclusion of stock characteristics into the two regression equations of Stoll's (1989) model without causing biases. Consequently, potential influential factors such as market capitalization cannot be embedded into a multivariate regression framework.

characteristics. In a panel, one should expect that stock characteristics could influence spread components regardless of matched or random sampling.

My paper is organized as follows. Section two describes the construction of matched and random samples of companies listed on the NYSE and NASDAQ. Section three introduces the trade indicator model, and section four highlights my empirical findings followed by concluding remarks.

2. Data and method of sampling

The TAQ2 Database provides intra-daily transaction prices, bid, ask quotes, and the number of traded shares for US stock markets. For my empirical method, it is essential working with intra-daily data, as one has to decide whether transactions are buyer- or seller-initiated trades. For that purpose, transaction prices are compared with quoted bid and ask prices offered by market makers.[4] As this is a preliminary study, I choose only one trading day, namely 30th November 2000. Table 1 provides summary statistics and an overview concerning the number of stocks listed on the NYSE and the NASDAQ that fulfill the basic requirement, namely at least 50 transactions per day. The average relative price fluctuation is based on Chiang and Venkatesh (1986).[5] Besides this measure, the volatility of midquote returns serves as indicator for risk. The next step is to select 50 stocks for each exchange randomly and to construct a matched sample. For matched sampling, stocks are classified regarding closing prices, the number of transactions, and volatilities. If stocks listed on different exchanges belong to the same 5% percentiles with regard to these three criteria, they are matched and build up a pair of observations. In line with the random selection, the matched sample contains 50 companies for the NYSE and 50 companies listed on the NASDAQ.

3. Trade indicator model

My trade indicator model is an extension of the Huang and Stoll (1997) model in that stock characteristics and hence cross-sectional differences are incorporated. To capture the influence of stock characteristics, one has to use a panel dataset rather than an individual time series approach.[6] The model can be described by the relation between changes in transaction prices denoted ΔP_{it} and order processing costs K_{it} as well as information costs L_{it}. This specification refers to equation (14) of Huang and Stoll (p. 1003, 1997) and thus is a two-way decomposition of the spread.

$$\Delta P_{it} = K_{it} \cdot \Delta Q_{it} + L_{it} \cdot Q_{it} + e_{it} \qquad (1)$$

The direction of trade labeled Q_{it} of transaction t (t=1;…;50) and stock i (i=1;…;100) is obtained by the classification of trades. If transactions are seller-initiated, Q_{it} takes the value minus one, and plus one if investors try to purchase stocks. The error term e_{it} should exhibit autocorrelation due to inventory costs (Huang and Stoll, 1997), and heteroscedasticity seems to be likely. Following the considerations of Glosten and Harris (1988) and Jennings (1994) that assume a linear relationship between the information costs L_{it} and the number of traded shares Z_{it}, my model (2) permits an impact of trading volume Z_{it} on the degree of information asymmetry. I deviate from Glosten and Harris (1988) and Jennings (1994) in that I use log

---

[4] I use Lee and Ready's (1991) algorithm as modified by Bessembinder (2003) to classify trades.

[5] Affleck-Graves et al. (1994) use this risk measure for constructing their matched sample.

[6] Yet one can think about the following alternative. One could determine spread components for each stock separately and then regress the difference between components on stock characteristics. However, there are several shortcomings: first, one cannot reveal the impact of trading mechanisms and stock characteristics simultaneously; second, one reduces the number of observations to 100 and does not use the advantages of a pooled sample; third, one cannot control for company specific effects as the number of observations is equal to the number of companies. Accordingly, I think that a panel approach is attractive when one wants to uncover cross-sectional differences.

trading volume, as the distribution of $\boldsymbol{Z_{it}}$ is skewed to the right. Accordingly, equation (1) is extended by the interaction term $\boldsymbol{\log Z_{it} Q_{it}}$.

$$\Delta P_{it} = k_0 \cdot \Delta Q_{it} + l_0 Q_{it} + l_1 \cdot \log Z_{it} \cdot Q_{it} + e_{it} \qquad (2)$$

This means that the number of traded shares $\boldsymbol{Z_{it}}$ influences information costs $\boldsymbol{L_{it}}$, whereas order processing costs $\boldsymbol{K_{it}}$ are independent from trading volume.[7] As the relevance of trading volume is hardly disputable and widely accepted in the literature, model (2) is an appropriate reference model before including additional stock characteristics.

How can one interpret this basic regression equation? If the transaction is buyer-initiated, transaction prices should go up due to information costs. As informed trading might motivate this transaction, one could assume that private information is partly conveyed in the order stream. Hence, share prices should increase when someone buys a large number of shares. In contrast, order processing costs do not have any persistent influence on transaction prices. They are modeled by the bid-ask bounce $\boldsymbol{\Delta Q_{it}}$.

In order to test for differences between the two trading systems, the model is slightly modified, by inserting the dummy variable $\boldsymbol{D_{NYSE}}$ that takes the value one if the stock is traded on the NYSE.

$$\Delta P_{it} = k_0 \Delta Q_{it} + k_1 D_{\text{NYSE}} \Delta Q_{it} + l_0 Q_{it} + l_1 \log Z_{it} Q_{it} + l_2 D_{\text{NYSE}} Q_{it} + l_3 D_{\text{NYSE}} \log Z_{it} Q_{it} + e_{it} \qquad (3)$$

The interaction term accounts for the fact that the impact of trading volume on information costs might differ between trading mechanisms.

4. Empirical results

Running regression (3) provides a first overview concerning spread components for the two trading systems and the relevance of trading volume (see table 2). Before interpreting these results, it is worth mentioning that the autoregression of residuals uncovers autocorrelation that makes GLS or an autocorrelation resistant estimation procedure of the covariance matrix necessary to obtain unbiased p-values.[8] A Breusch-Pagan test reveals heteroscedasticity. The standard Huber-White Sandwich estimator is only robust in the presence of heteroscedasticity – but not if serial dependency among successive transactions plays a role. Serial dependency can be regarded as dependency within a cluster defined by the respective stock – the cross-sectional unit. Applying a modified sandwich estimator avoids the problem of within-cluster correlation and yields robust p-values. Obviously, this modified Sandwich estimation only corrects p-values – but OLS estimates of coefficients might be inconsistent due to an endogenity bias. GMM can cope with serial dependencies and a potential endogenity bias.

Due to the high correlation coefficient of 0.8065 in the random and 0.8450 in the matched sample between the variables $\boldsymbol{\log Z_{it} Q_{it}}$ and $\boldsymbol{Q_{it}}$, table 2 reports the regression results with and without $\boldsymbol{\log Z_{it} Q_{it}}$. To check whether one can exclude the variable $\boldsymbol{\log Z_{it} Q_{it}}$ without creating an omitted variable bias, I apply Ramsey Reset tests that confirmed that the model is not misspecified. Obviously, stocks traded on the NASDAQ exhibit higher order processing costs, as the coefficient $\boldsymbol{D_{NYSE} \Delta Q_{it}}$ is in all models and for both samples significantly different from zero and negative. Interestingly, the NYSE has higher information costs indicated by the significant coefficient of $\boldsymbol{D_{NYSE} Q_{it}}$. This effect is offset by the significantly negative impact of the interaction term $\boldsymbol{D_{NYSE} \log Z_{it} Q_{it}}$. The interaction term captures the impact of the trading mechanism on liquidity. Consider that this coefficient $\boldsymbol{l_3}$ can be regarded as a measure for inverse liquidity. Liquidity is defined as the price movement caused by a transaction with a specific trading volume. The inverse liquidity is defined as the reciprocal liquidity that is equivalent to the partial derivative of the price change $\boldsymbol{\Delta P_t}$ with respect to $\boldsymbol{D_{NYSE} \log Z_{it} Q_{it}}$. This is captured by the magnitude of the coefficient $\boldsymbol{l_3}$. Thus, one might suspect that liquidity is

---

[7] If one supposes that trading volume is able to influence order processing costs, regressions do not confirm such an impact.

[8] An AR(1) process for residuals uncovers p-values of about 0.000 for the lagged residual.

higher on the NYSE, so trades with a high trading volume should be better executed on an auction market, as prices are only slightly affected.[9] Accordingly, the impact of trading volume on spreads is not negligible when one wants to compare the two trading mechanisms. The following paragraphs deal with the problem of inserting more stock characteristics and estimating their partial impact on spread components.

An often-used hypothesis regarding the impact of market capitalization on spread components is that market capitalization is negatively related to information costs. Market capitalization is a measure for firm size; hence, the interest of analysts should be higher if the company is large. The distribution of market capitalization $\boldsymbol{M_i}$ is skewed to the right; therefore, it seems to be appropriate to transform the variable logarithmically. Regression (2) is extended to account for an influence of market capitalization $\boldsymbol{M_i}$ on information costs $\boldsymbol{L_{it}}$.

$$\Delta P_{it} = k_0 \cdot \Delta Q_{it} + l_0 Q_{it} + l_1 \cdot \log Z_{it} \cdot Q_{it} + l_2 \cdot \log M_i \cdot Q_{it} + e_{it} \tag{4}$$

The volatility of midquote returns labeled $\boldsymbol{\sigma^2_i}$ serves as a measure for risk. A reasonable hypothesis would be that price fluctuations represent the advantage of informed traders. High volatility indicates an abnormal degree of uncertainty with respect to the true value of the stock. In a volatile market, an informed agent, who has superior knowledge about the true value, has meaningful advantages. Therefore, one can suggest that higher volatility makes dealers more cautious. Cautious means that dealers react more sensitively to high trading volumes, and they adapt their expectations about the true value stronger than on normal days. Consequently, an order of a given size causes higher price movements and liquidity decreases. The following specification could test this hypothesis.

$$\Delta P_{it} = k_0 \cdot \Delta Q_{it} + l_0 Q_{it} + l_1 \cdot \log Z_{it} \cdot Q_{it} + l_2 \cdot \log Z_{it} \cdot Q_{it} \sigma_i^2 + e_{it} \tag{6}$$

An additional selection criterion used by Affleck-Graves et al. (1994) and for my matched sample is the share price. Hence, one assumes that share prices might affect spread components. To test this assertion, middle prices, namely the average between the highest and lowest transaction price, are calculated and denoted $\boldsymbol{P_i}$. Using middle prices avoids possible biases caused by relying on closing prices. As – from my point of view – there exists no convincing hypothesis about the expected influence of share prices on spread components, all possible relationships are tested.

Thus far, OLS is applied to estimate the coefficients of model (3), and a modified sandwich estimator determines the covariance matrix. Nevertheless, recent literature (Huang and Stoll, 1997, and Madhavan et al., 1997) stress the advantages of GMM procedures in estimating spread decomposition models. Especially, the usually observed negative autocorrelation of successive returns due to inventory costs and the bid-ask bounce can be corrected by GMM procedures. Without any doubts, these models are useful when applied to individual time series of successive transactions. To reveal cross-sectional differences in spread components related to stock characteristics, a panel data approach is required. However, former GMM procedures cannot be easily applied to panel data. Fortunately, the literature on dynamic panel data estimation provides useful solutions. The seminal paper of Arrelano and Bover (1995) paves the ground for this still developing field. They derive a GMM procedure that can be applied to dynamic panel data. In addition, they propose future mean differencing – the so-called Helmert's transformation – to control for company specific effects. Consequently, individual effects denoted $\boldsymbol{f_i}$ are inserted into model (3), and the hypotheses (3), (4) and (6) are combined into one regression framework. Lagged values of the dependent variable up to lag p are considered to account for serial dependency. Using the Arrelano-Bond test, I can set p equal to four.

[9] Fixed effects or random effects are not relevant. Joint hypothesis tests (F-tests) indicate that all coefficients of company specific dummy variables are not significantly different from zero. Likelihood ratio tests cannot reject the null hypotheses that the variance of a company specific error term is equal to zero.

$$\Delta P_{it} = k_0 \Delta Q_{it} + k_1 D_{\text{NYSE}} \Delta Q_{it} + k_2 P_i \Delta Q_{it} + l_0 Q_{it} + l_1 \log Z_{it} Q_{it} + l_2 D_{\text{NYSE}} Q_{it} \quad (7)$$
$$+ l_3 D_{\text{NYSE}} \log Z_{it} Q_{it} + l_4 \log M_i Q_{it} + l_5 P_i Q_{it} + l_6 \log Z_{it} Q_{it} \sigma_i^2 + f_i + \sum_{j=1}^{p} \Delta P_{it-j} + e_{it}$$

Due to the likely correlation between individual effects f_i with the lagged values of the dependent variable, fixed effects models are inappropriate to control for company specific effects. Thus, one has to apply the Helmert's transformation as defined in equation (8). Hereby, T indicates the total number of observations, and z_{it}^* represents the transformed series, whereas z_{it} is the original series.

$$z_{it}^* = \left(\frac{T-t}{T-t+1} \right)^{0.5} \left(z_{it} - \frac{1}{T-t} \sum_{j=1}^{T} z_{i(t+j)} \right) \quad (8)$$

The Helmert's procedure transforms the time series in levels by subtracting the future expected value from the current value of the variable. Obviously, using these transformed variables in regression (7) violates the assumption of weak exogenity because the variables incorporate future information. Thus, transformed variables are not predetermined. To estimate the regression with modified series, one has to apply the GMM procedure as thoroughly discussed by Arellano and Bover (1995). In particular, the non-transformed lagged variables serve as instruments for the modified variables. Table 3 summarizes the results of model (7) for the random and the matched sample. I estimate regression (7) with and without company specific effects using GMM. After transforming the individual series by Helmert's transformation and GMM estimation with the non-transformed variables as instruments, the results are to some extend affected. Wald statistics indicate an improvement of the model fit caused by accounting for company specific effects.

The results once again indicate that the NYSE has lower order processing costs. The coefficient for the partial impact of stock prices on order processing costs is highly significant in the case of the matched sample. Contrarily, the random sample reveals a negative influence of market capitalization on information costs, which is predicted by theoretical considerations that analysts monitor larger companies better. To illustrate my empirical findings, table 4 summarizes the estimated spread components based on the GMM estimates with individual effects for the matched and unmatched sample. Calculating the spread components due to trading mechanisms and stock characteristic refers to an average stock listed on the NYSE and NASDAQ. Table 4 contains the different components of the spread and the importance of trading mechanisms and stock characteristics for the respective component. Consider that this table only reports partial impacts that are significant on the 10% level of significance.

Focusing on the results for the random sample, one can state that order processing costs and information asymmetry costs are smaller on the NYSE. Despite the fact that the GMM model uncovers a coefficient of 0.2685 of the variable Q_{it} for both exchanges, information costs differ due to higher market capitalization and higher share prices on the NYSE. Consequently, this model cannot only determine the magnitude of spread components – but also the underlying reasons for the components. The size of a company as measured by market capitalization matters in that it reduces information costs. This empirical finding is in line with theoretical considerations that larger companies attract the interest of analysts and are hence better monitored, which lowers the degree of information asymmetry. Furthermore, the total spread is considerably higher on the NASDAQ mainly due to higher order processing costs.[10]

Shifting the attention to the matched sample, one should consider that stock characteristics are not random. In the random sample, the average share price on the NYSE is 15.57US$ compared to 20.38US$ on the NYSE. Caused by the matching of stocks, average share prices for the matched sample are 26.29US$ for NASDAQ stocks and 25.39US$ for stocks listed on

[10] Note that the total spread is two times the sum of the order processing and the information costs.

the NYSE. The gap with respect to share prices is closed by the matching procedure; hence, companies on the NASDAQ exhibit `artificially' higher share prices than in the case of a random sample. This fact could explain the pronounced impact of share prices on order processing costs that is not observable when stocks are selected randomly. Obviously, matched sampling affects the partial impact of stock characteristics. Nevertheless, the estimates of the spread components and the total spread are similar in comparison to random sampling. Although matched sampling has considerable advantages when applied in a time series analysis, it cannot control for cross-section differences that arise in a panel dataset. Yet this is not the task of matching.

5. Conclusion

My analysis provides evidence that order processing costs are higher on a dealer market in comparison to an auction market, as on auction markets like the NYSE a considerable part of the order stream is executed directly through the order book. This empirical finding is in line with Affleck-Graves et al. (1994), Huang and Stoll (1996), Bessembinder and Kaufman (1997), and Bessembinder (1999) among others.

I show that information asymmetry, measured by the information costs component of the spread, is slightly higher on the NASDAQ. This finding is due to lower share prices – but companies are smaller on the NASDAQ, which partly offsets the first impact. This finding emphasizes that large companies exhibit a higher analysts coverage and hence a lower degree of information asymmetry. Accordingly, my model allows uncovering the reasons for higher information costs on the NASDAQ, namely lower market capitalization and less attention of financial analysts. The trading mechanism, consequently, is not responsible for higher information costs on the NASDAQ, which is an essential finding.

The econometric contribution of my paper is the extension of Huang and Stoll's (1997) time series approach by allowing a cross-sectional dimension. This is required to reveal cross-sectional differences due to stock characteristics. Applying a dynamic panel data estimator constructed by Arellano and Bover (1995) to my panel decomposition model overcomes the shortcoming of a panel OLS estimation. Controlling for individual effects by Helmert's transformation and GMM estimation with the lagged untransformed variables as instruments yields different outcomes; hence, company specific effects are relevant.

Applying my model framework, several additional topics could be studied, which is hardly possible using time series approaches. For instance, the impact of regulatory changes on spread components can be investigated within a very short period. Time series models usually use data for one month for an individual stock; thus, short-term market reactions cannot be revealed. Due to the advantages of pooling data, one can choose a smaller interval. Consequently, this panel approach has various potential applications for future research.

References

Affleck-Graves, J., Hedge, S. P., and R.E. Miller (1994) "Trading mechanisms and the components of the bid-ask spread" *Journal of Finance* **49**, 1471-1488.

Arellano, M., and O. Bover (1995) "Another look at the instrumental variable estimation of error-component models" *Journal of Econometrics* **10**, 29-51.

Bessembinder, H. (2003) "Issues in assessing trade execution costs" *Journal of Financial Markets* **6**, 233-257.

Bessembinder, H. (1999) "Trade execution costs on NASDAQ and the NYSE: A post-reform comparison" *Journal of Financial and Quantitative Analysis* **34**, 387-407.

Bessembinder, H., and H. M. Kaufman (1997) "A comparison of trade execution costs for NYSE and NASDAQ-listed stocks" *Journal of Financial and Quantitative Analysis* **32**, 387-407.

Chiang, R., and P.C. Venkatesh (1986) "Information asymmetry and the dealer's bid-ask spread : A case study of earnings and dividends announcements" *Journal of Finance* **41**, 1089-1102.

Copeland, T. C., and D. Galai (1983) "Information effects on the bid-ask spread" *Journal of Finance* **38**, 1457-1469.

George, T. J., Kaul, G., and M. Nimalendran (1991) "Estimation of the bid-ask spread and its components: A new approach" *Review of Financial Studies* **4**, 623-656.

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

Harris, L. (2002) *Trading and exchanges: Market microstructure for practitioners*, Oxford University Press.

Ho, T., and H. R. Stoll (1981) "Optimal dealer pricing under transactions and return uncertainty" *Journal of Financial Economics* **9**, 47-73.

Ho, T., and H. R. Stoll (1983) "The dynamics of dealer markets under competition" *Journal of Finance* **38,** 1053-1074.

Huang, R. D., and H. R. Stoll (1997) "The components of the bid-ask spread: A general approach" *Review of Financial Studies* **10**, 955-1034.

Huang, R. D., and H. R. Stoll (1996) "Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE" *Journal of Financial Economics* **41**, 313-357.

Jennings, R. (1994) "Intraday changes in target firms` share price and bid-ask quotes around takeover announcements" *Journal of Financial Research* **17**, 255-270.

Lee, C., and M. Ready (1991) "Inferring trade direction from intraday data" *Journal of Finance* **46**, 733-746.

Madhavan, A., Richardson, M., and M. Roomans (1997) "Do security prices change? A transaction-level analysis of NYSE stocks" *Review of Financial Studies* **10**, 1035-1064.

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

Table 1: Descriptive statistics for all companies on both exchanges (November 2000)

	NYSE	NASDAQ
Number of listed companies in the dataset	3340	4692
Number of companies >50 transactions	1356	2544
Average number of transactions	401.37	1392.84
Average number of traded shares	1,078,394	1,046,898
Average daily volume in million US$	41.2	35.0
Average daily-low price	79.25	13.70
Average closing price	81.83	14.67
Average daily-high price	82.66	15.51
Average relative price fluctuation in %	5.11%	16.20%

Table 2: Outcomes of the pooled regression (3) for the random and matched sample

Explanatory variables	Matched sample with $\log Z_{it}Q_{it}$	Matched sample	Random sample with $\log Z_{it}Q_{it}$	Random sample
ΔQ_{it}	0.0970	0.0941	0.0917	0.0891
	(0.000)	(0.000)	(0.000)	(0.000)
Q_{it}	0.0008	0.0042	0.0007	0.0045
	(0.822)	(0.002)	(0.858)	(0.000)
$\log Z_{it}Q_{it}$	0.0006	-	0.0007	-
	(0.240)		(0.248)	
$D_{NYSE}\Delta Q_{it}$	-0.0474	-0.0445	-0.0474	-0.0447
	(0.005)	(0.008)	(0.003)	(0.005)
$D_{NYSE}\ Q_{it}$	0.0190	0.0156	0.0090	0.0051
	(0.007)	(0.014)	(0.043)	(0.050)
$D_{NYSE}\log Z_{it}Q_{it}$	-0.0026	-0.0020	-0.0019	-0.0012
	(0.014)	(0.031)	(0.011)	(0.004)
Ramsey RESET (F-test statistic)	0.31	0.41	0.27	0.29
	(0.816)	(0.746)	(0.849)	(0.835)
Breusch-Pagan (Chi^2 test statistic)	0.39	0.05	14.50	14.88
	(0.535)	(0.827)	(0.000)	(0.000)
Fixed effects F-test (p-value)	0.55	0.55	0.43	0.44
	(0.999)	(0.999)	(1.000)	(1.000)
LR test for random effects (p-value)	0.00	0.00	0.00	0.00
	(1.000)	(1.000)	(1.000)	(1.000)
Observations	4900	4900	4900	4900
Adjusted R^2	0.04	0.04	0.12	0.12

(Corrected p-values applying the modified Sandwich estimator are in parentheses)

Table 3: GMM estimates of regression (7) for the random and matched sample

	Random sample		Matched sample	
Explanatory Variables	GMM without individual and time effects	GMM with individual and time effects	GMM without individual and time effects	GMM with individual and time effects
ΔQ_{it}	0.0638	0.0709	0.0376	0.0476
	(0.000)	(0.000)	(0.012)	(0.007)
$D_{NYSE}\Delta Q_{it}$	-0.0386	-0.0475	-0.0473	-0.0499
	(0.002)	(0.000)	(0.000)	(0.000)
$P_i\Delta Q_{it}$	-0.0001	-0.0003	0.0018	0.0013
	(0.918)	(0.724)	(0.005)	(0.091)
Q_{it}	0.2373	0.2685	-0.0019	0.1585
	(0.000)	(0.000)	(0.980)	(0.159)
$D_{NYSE}Q_{it}$	-0.0025	0.0169	0.0093	0.0084
	(0.847)	(0.242)	(0.396)	(0.564)
$D_{NYSE}\log Z_{it}Q_{it}$	-0.0021	-0.0022	-0.0017	-0.0034
	(0.040)	(0.069)	(0.124)	(0.007)
$\log M_iQ_{it}$	-0.0159	-0.0180	0.0025	-0.0097
	(0.000)	(0.000)	(0.666)	(0.255)
P_iQ_{it}	0.0007	0.0011	-0.0008	0.0013
	(0.214)	(0.072)	(0.239)	(0.086)
$\sigma^2_i \log Z_{it}Q_{it}$	-0.0007	-0.0007	-0.0000	-0.0001
	(0.279)	(0.224)	(0.938)	(0.901)
Observations	4400	4200	4400	4200
Wald Chi^2	412.49	434.29	198.74	240.67

(P-values are reported in parentheses. Estimated autocorrelation coefficients are not reported)

Table 4: Estimated spread components and the impact of trading mechanisms

	Random sample		Matched sample	
	NASDAQ	NYSE	NASDAQ	NYSE
Order processing costs	0.0709	0.0234	0.0476	-0.0023
Impact of price on order processing costs	-	-	0.0342	0.0330
Total order processing costs	0.0709	0.0234	0.0818	0.0307
Information costs	0.2685	0.2685	-	-
Impact of volume on information costs	-	-0.0137	-	-0.0214
Impact of firm size on information costs	-0.2590	-0.2596	-	-
Impact of price on information costs	0.0171	0.0224	0.0342	0.0330
Total information costs	0.0266	0.0176	0.0342	0.0116
Total spread	0.1950	0.0820	0.2320	0.0846

(All values are in US$)