Analyzing intraday price formation - with structural models

Matthias Prinzbach*

June 2007

*matthias.prinzbach@gmx.de

Contents

1 Introduction

It is no secret that real markets are not perfect, as most economic models suggest. Security markets are often singled out as excellent examples of nearly perfect markets. But transaction costs are an important factor in real markets, even in security markets. A well known concept is, that security prices should only change if new information is available. The price should chance immediately to the new information value. This called the efficient market hypotheses[1]. But in practice there are several market frictions and imperfections. An realistic description of intraday security price changes, must capture these frictions. New information may not reach each trader at the same time, this means some traders are better informed than the others. This could occur through better information management or insider information. Another interesting point in intraday price formation is the U-shaped pattern of quoted bid-ask-spreads and of the trading volume over the day [see, e.g., Harris (1986), Jain and Joh (1988), and McInish and Wood (1992)]. This and other microstructure phenomena could not be explained with previous models neither by theoretical [Easley and OHara (1992), Madhaven (1992), Bloomfield] nor by laboratory experiments [Bloomfield (1996), Bloomfield and O'Hara (1996)]. In 1996 two related models were developed, Huang and Stoll[2] and Madhaven, Richardson and Roomans[3] developed their models independently. MRR analyzed the intraday price formation, HS had other objectives, but their model could also be used to analyze the intraday price formation. This two models should shed light into the price formation process and which frictions or imperfections influence it. Of interest is also the volatility of intraday price changes and which role the microstructure effects are playing. The price volatility is higher during trading hours, this is what French and Roll (1986) had found. More recently Hasbrouck (1991 and 1993) analyzed the price volatility using time series technique to decompose the variance of the returns. The Paper is structured as follows. Section 2 outlines the basic market micro structure, including and explaining the relevant market frictions and imperfections. Section 3 presents the Huang and Stoll model to analyze the influence of market microstructure. Chapter 4 introduces a partly different approach to describe the intraday price formation, the model from Madhaven Richardson and Roomans, here the different patterns in active traded stocks

[1] The efficient market hypotheses was introduced by Fama 1970
[2] The Huang and Stoll model is in the following article called HS
[3] In the following article it is called MRR

and the price volatility is analyzed. Chapter 5 will summarize the findings and compare the results of the approaches.

2 Microstructure Theory

Quoted ask and bid prices are provided by liquidity suppliers, they always buy to the posted bid price and sell at the ask. Such liquidity providers could be market makers or limit order traders. The difference between the bid and ask price is called the spread.
This quoted spread is a misleading measure for transaction costs since trades not only occur at the quoted bid or ask prices. In fact trades occur in the spread and outside. This would be a good reason for the effective spread as measure for transaction costs rather than the posted.
The effective bid-ask spread is computed as follows

$$S_e = 2(midquote - executionprice)Q_t \tag{1}$$

Where midquote is the mean of quoted bid and ask price, and Q_t is a trade indicator variable. It signals if trade is buyer or seller initiated. A trade is called buyer initiated if the execution price is bigger than the midquote or seller initiated if smaller. Q_t takes the value +1 for buyer initiated trade and -1 for seller initiated trade.
The spread consists of 3 different kinds of transaction costs, order processing costs, inventory costs and asymmetric information costs. The liquidity provider incurs this costs, so he wants to be compensated for them via the spread.
Order processing costs, are costs for everything needed for processing the orders, for example labor and equipment. Inventory holding costs, are the opportunity costs of capital needed for the transaction, it is a compensation for holding an suboptimal portfolio. For the asymmetric information costs, a longer explanation is needed. Two types of traders are at the market. Liquidity traders, they buy and sell because of liquidity motives, and informed traders, they buy and sell only if they have new information. Liquidity traders trade all the time, their buys and sells are equally likely. Because of this the expected profit of trading with liquidity traders is zero. Informed traders only buy if they expect to macke a profit due to their private information. An better informed trader would only buy, if the execution price is less than the expected asset value. The liquidity provider, sells in this case for a value less than the true value, the liquidity

supplier makes a loss. To compensate this loss, the asymmetric information costs are put into the spread. The spread will widen an the liquidity traders compensate the dealer for the loss in trading with better informed traders.
The models, which are used to explain the intraday price formation, are so called trade indicator models, they use only the indicator for trade direction, not the trade volume data. Trade volume does not provide much information, because large orders are often splitted in several smaller trades. Because of this, trade direction may have more explanatory power than size. Risky securities have an expected fundamental value V_t (V_t may be the discounted value of future dividends). This value varies over time. Changes in believes about the expected asset value appears from two sources. First new public information ε_t appears, the traders change their beliefs, without any trade. Second, order flow, which could be viewed as a signal about the value, from an better informed trader. Therefore a buy is associated with an upward revision, a sell with a downward revision. The unobservable V_t has to be a function of ε_t, as it is in the efficient market hypotheses, but it is also influenced by the order flow, if there are better informed traders in the market, or if the liquidity provider believe there are some. The execution price which the trader has to pay to get the security is not V_t, because of the transaction costs. Liquidity providers pay $V_t - Costs$ or sell for $V_t + costs$. The price for sellers is lower than the price for buyers, due to the dealer costs. For this reason, a trade with an higher price than the one before could be seen as a buyer initiated trade, a lower price for an seller initiated trade.

3 Huang and Stoll model

3.1 Model description

In the Basic model of Huang and Stoll, the unobservable expected value of the security is modeled as follows

$$V_t = V_{t-1} + \alpha \frac{S}{2} Q_{t-1} + \varepsilon \tag{2}$$

Where S is the traded spread and α is the halfspread part attributable to private information. ε_t is the innovation in beliefs due to public information. V_t is the value prior to the order at t. The expected prior trade value depends on the expected value prior to the last trade, the public information which occurred since the last trade and the private information revealed through the last trade $\alpha\frac{S}{2}Q_{t-1}$. In case of a perfect market (no information asymmetry) α is of value zero, and the expected value is a random walk. V_t is unobservable but we can observe the spread midpoint, denoted M_t. According to inventory theories liquidity providers adjust the midpoint relative to the cumulated inventory at time t, to induce inventory balancing trades. Under this models the midpoint looks like this

$$M_t = V_t + \beta\frac{S}{2}\sum_{i=1}^{t-1} Q_i \tag{3}$$

β is here the halfspread part due to inventory costs. $\sum_{i=1}^{t-1} Q_i$ is the cumulated inventory of the security hold by the market maker, given constant trade sizes. Without inventory holding costs there would be a one-to-one relation of M_t and V_t. Combine equation 2 and 3, and taking first difference yields

$$\Delta M_t = (\alpha + \beta)\frac{S}{2}Q_{t-1} + \varepsilon \tag{4}$$

This leads to the final equation:

$$P_t = M_t + \frac{S}{2}Q_t \tag{5}$$

Here the price consists of the midpoint and the halfspread for a buy order, or less the halfspread for a sell order. In the original paper there is a η which captures rounding errors associated with price discreetness. It was needed, because the ticks were only 1/8 in their data, but nowadays, ticks are small enough, though there shouldnt be rounding errors[4]. Equation (4) and (5) together yields the regression equation,

$$\Delta P_t = \frac{S}{2}(Q_t - Q_{t-1}) + \lambda\frac{S}{2}Q_{t-1} + \varepsilon_t \tag{6}$$

Where $\lambda = \alpha + \beta$, since the model cannot distinguish between them. So this basic model lumps together private information and inventory holding components of the spread. But it can distinguish between them and order processing. The order processing part is given by $(1 - \lambda)\frac{S}{2}$ the remaining spread part.

[4]In Chapter 4.5, I tried to estimate the Variance of this rounding errors, but most of the estimates where insignificant at the 5% level, this is a strong evidence for none rounding errors

3.2 DATA

Investigated are 15 Stocks[5] at the NYSE and the same stocks at the TSX. All firms are Canadian, operating in different sectors, from fertilizer to financial services. The sample is drawn from the 01.01.2003 to the 31.03.2003. It consists of price change, trade indicator with two lags and the time. The data is divided in 4 time intervals, 09:30-10:30, 10:30-12:30, 12:30-14:30, 14:30-16:00. Only Stocks with more than 250 observations in each trading interval over the observation period are included. MDS is skipped because of to few observations (only 81 in the first trading interval over the sampling period). In table 8 it can be seen that the mean of trades per stock between 09:30 and 10:30 is higher in Toronto than in New York, for the remaining day the mean of trades per stock at the NYSE is higher. Trade number at the NYSE has a wider range, the minimum is lower and the maximum is higher at the NYSE. The first trading hour is the most active, between 12:30 and 14:30 number of trades are lowest. Number of trades exhibit U-shape patterns. Some stocks are more often traded in Toronto, some in New York.

3.3 Estimation procedure

Like in the original paper of Huang and Stoll (1997), the Generalized Method of Moments (GMM) is used to estimate the model. This method does not need strong distribution assumptions like Maximumlikelihood. The GMM procedure also accounts for conditional heteroscedasticity and serial correlation in the residuals. To estimate the covariance matrix the Newey-West procedure is used. The model implies these orthogonality conditions

$$E\begin{pmatrix} e_t Q_t \\ e_t Q_{t-1} \end{pmatrix} = 0 \tag{7}$$

The Generalized Method of Moments minimized the criterion function

$$J_t(\beta) = g_t(\beta)' S_T g_t(\beta) \tag{8}$$

Where $\beta = \{S\ \lambda\}$ is the parameter vector, $g_t(\beta)$ is the sample mean of the orthogonality conditions, and S_T is a symmetric weighting matrix. Under weak conditions, the GMM estimator β is consistent and asymptotically normal. The

[5] Detailed Stocklist see Appendix A

summary statistics of estimates for Toronto

		09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
S					
	Mean	4.4697	3.6144	3.5069	3.5700
	Avg.Std.Er.	0.2092	0.1088	0.1176	0.1166
	Std.Dev.	1.9235	1.4997	1.3289	1.4054
	Median	3.6365	2.9455	2.9942	3.3341
λ					
	Mean	0.4087	0.3274	0.3143	0.3239
	Avg.Std.Er.	0.0310	0.0175	0.0197	0.0196
	Std.Dev.	0.1845	0.1274	0.1355	0.1167
	Median	0.3370	0.3296	0.2931	0.2651

summary statistics of estimates for NYSE

		09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
S					
	Mean	1.6326	1.4194	1.3138	1.3233
	Avg.Std.Er.	0.1208	0.0800	0.0787	0.6912
	Std.Dev.	0.6501	0.5308	0.4878	0.5190
	Median	1.4219	1.2591	1.1244	1.0592
λ					
	Mean	0.4652	0.3672	0.3487	0.3873
	Avg.Std.Er.	0.0616	0.0397	0.0436	0.0353
	Std.Dev.	0.0698	0.1035	0.1108	0.1210
	Median	0.4653	0.3841	0.3612	0.3829

Table 1: This table presents the summary statistic over the 15 stocks per stock exchange. The values are the medians, the mean, the average standard error and the standard deviation in each trading interval, S is measured in Cent, λ is the percentage of inventory holding and order processing costs.

model is exactly identified, that means the number of orthogonality conditions equals the number of parameters to be estimated.

3.4 Estimation results

First the estimated spread in Toronto is much higher than the spread at the NYSE. The U-shape is much more pronounced at the TSX, but is not an complete U, the upward slope at the end is not as high as the starting point. But there is a minimum spread in the 12:30-14:30 interval at both stock exchanges. The decline from morning to noon, is about 20 percent at both exchanges. The increase from noon to the end of trade time is 2 percent. The NYSE has less transaction costs than the TSX, the minimum is between 12:30 and 14:30. The average standard error declines from beginning to the end of the trading day, for each parameter. The standard deviations of the spreads show an U-shape pattern, at the beginning the value is always highest. The standard deviation of λ is decreasing over the day in Toronto, but in New York it is increasing. The parts of the spread are interesting. At the TSX the λ, which contains the asymmetric information and the inventory part of the spread is 40 percent at the beginning and declines after the opening period to about 31 percent for the rest of the day. This means the sum of asymmetric information and inventory holding costs is highest at the beginning of the day and more or less the same for the rest of the day. The order processing part $(1 - \lambda)$ is smallest at the beginning 60 percent on average and stays at 70 percent during the rest of the day. For the NYSE, the λ is bigger all the day, so the part of the spread due to asymmetric information and inventory holding costs is bigger at the NYSE. But in absolute value the costs are smaller. The order processing part $(1 - \lambda)$ is smaller than the Toronto value. The NYSE is cheaper. The order processing costs are only one third, the other cost part is about half the value at the TSX. Compared with the results in Huang and Stoll (1996), the spread is much smaller. Huang and Stoll estimated an average spread of 12 Cent at the NYSE. Here the highest mean spread is 1,6 Cent at the NYSE in the first trading interval. The λ, is higher, it is between 0.3 and 0.45, in contrast, the λ of Huang and Stoll was only 0.11 on average. The absolute value of this costs λS is not as much smaller as the spread, but it is only half of the value in the original paper. HS do not estimate different trading periods. In no period, the estimated values reach their estimates.

Huang and Stoll also provide a model which decomposes the spread in 3 components, but the problem with this model, is that the asymmetric information parameter got negative, which is impossible from theoretical viewpoint. This happens in the original paper and for the data used in this paper too. Because of this, the model is skipped.

4 MRR model

4.1 Model description

The unobservable value is modeled as follows

$$V_t = V_{t-1} + \theta(Q_t - E[Q_t|Q_{t-1}]) + \varepsilon_t \tag{9}$$

The new value V_t, which is the value immediately after the trade at t, is the past value, the news shock and private information signal from trade. This signal is $(Q_t - E[Q_t|Q_{t-1}])$, wich is the surprise in order flow, this surprise is multiplied with θ the degree of information asymmetry. θ captures the permanent impact of order flow on prices. The bigger θ, the bigger the revision in beliefs due to order flow. Market maker set ex post rational quotes, so the ask is conditioned on buyer initiated trade and the bid for seller initiated. The quotations also reflects market makers compensation for providing liquidity. ϕ is market makers cost per share, it covers inventory and order processing costs. Now the transaction price can be expressed as

$$P_t = V_t + \phi Q_t \tag{10}$$

In the MRR (1996) paper there is an η which covers the rounding errors, induced by price discreteness, but in the sampling period the ticks were small enough, to assume this pricing errors to be zero[6]. V_t is the value given Q_t is observed. Market makers fix their quotes before Q_t is observed. The ask price is conditioned on a buy order, bid on a sell order. From equation 9 and 10 follows

$$P_t = V_{t-1} + \theta(Q_t - E[Q_t|Q_{t-1}]) + \phi Q_t + \varepsilon_t \tag{11}$$

For estimation purposes we need to describe the temporal behavior of order flow, for Q_t a general markov process[7] is assumed. So Q_t is weak stationary. With $E[Q_t|Q_{t-1}] = \rho Q_t$[8] and the first difference, to back out the unobservable V_t. The last equation is

$$P_t - P_{t-1} = (\phi + \theta)Q_t - (\phi + \rho\theta)Q_{t-1} + \varepsilon_t \tag{12}$$

ΔP_t is weak stationary[9] This equation shows, that the intraday price movement is due to public information and market frictions. In opposite to the HS model this model separates the asymmetric information component (θ) from inventory holding which is lumped together with the order processing costs in ϕ. The two parts of the spread are θ and ϕ, so the implied spread could be computed as $S = 2(\theta + \phi)$. The information ratio is $r = \frac{\theta}{\theta+\phi}$.

[6]In Chapter 4.5, I tried to estimate the Variance of this rounding errors, but most of the estimates where insignificant at the 5% level, this is a strong evidence for none rounding errors

[7]The autocorrelation vanishes after some lags

[8]Derivation in Appendix B

[9]see Appendix B

4.2 Estimation procedure

There are 3 Parameters (θ, ϕ, ρ) describing the behavior of transaction prices, in a linear function. MRR used the Generalized Method of Moments to estimation. This process is chosen because it does not need strong assumptions about the data generating process and it allows adjustment for general forms of autocorrelation and conditional heteroscedasticity. GMM chooses the parameters to minimize the criterion function based on the moment Conditions.

$$E \begin{pmatrix} e_t \\ e_t Q_t \\ e_t Q_{t-1} \\ Q_t Q_{t-1} - Q_t^2 \rho \end{pmatrix} = 0 \tag{13}$$

The number of Parameters to be estimated, is less the number of orthogonality conditions, this is called over identified. Now GMM takes the Parameter values which approximate the underlying population moments. The estimates are consistent and asymptotically normal. The covariance matrix of the errors is estimated by the Newey-West procedure, to obtain heteroscedasticity and serial correlation consistent covariance matrix. The parameters are estimated for each exchange, stock and time interval separately. This yields 112 estimates.

4.3 Estimation results

For estimating this model the same data as in the Huang and Stoll model is used. Table 2 shows the estimated parameters. For most stocks the parameters are significantly different from zero at the 1% significance level. Only Ten of the 112[10] estimated parameters are not significant[11].
First the inventory and orderprocesing costs, ϕ this is the part of the price, due to this costs. At the exchange in Toronto the mean declines from the beginning to the end of trading day. The only exception is a small decrease in the last period. This pattern may reflect the increasing risk to carry inventory over night. The median follows a surprising pattern. The all time low is in the second period, after this an increase followed by a small decrease in the last period follows. The mean at the NYSE is different, it has a smaller value, it is approximately a third. The pattern is different too. After the opening period there is a decrease, for the next to periods it stays at the same value and drops

[10] 2 exchanges, 4 time intervals, 14 Stocks
[11] see Appendix Table 7

in the last period. The median has an pattern similar to the mean in Toronto, the increase is from morning to noon, and a small drop at the last period. The pattern in the first three periods can be explained, by the overnight inventory holding risk, but the drop in the last period does not follow this explanation.

Now the asymetric information parameter θ, the expected loss of the liquidity provider to a better informed trader, per trade. At both exchanges the mean pattern is similar, but the value is once more higher in Toronto. A sharp drop after the opening period, a decrease until 14:30 in the last period. At the NYSE there is a decrease in the last period too. The median in Toronto has the same pattern as the mean, but is less in value. In New York the median decreases steadily over the day. The information asymmetry is highest at the opening period, the traders learn most about the fundamental value in the first period. The information asymmetry grows in the last period, this result is consistent with the estimates in the MRR paper, but here the growth is higher, they do not provide an explanation. The standard errors of both parameters, at both exchanges has the same pattern. The highest value is at the opening period, the value in the remaining three periods is more or less the same. The standard deviation in Toronto has the same pattern. But in New York its different, the standard deviation of ϕ is increasing up to the third period, and drops after this. The standard deviation of θ has a U shape pattern. The autocorrelation is in the first to periods more or less the same an decreases over the two last periods. The Median is over the first three periods the same and decreases in the last period. The average standard errors are very small, so the estimation is very precise. The standard deviation is very small too, it seems most of the stocks have the same autocorrelation.

Compared with the results in MRR (1996) which was estimated with NYSE data, all the parameters are much smaller. ϕ the order processing and inventory holding costs are onlay 1/10 of the estimates in MRR (1996), the same for the information parameter θ. This smaller coefficients could be due to the fact that MRR estimated the coefficient only for trades which are executed outside the quoted spread. Their trade indicator for trades inside the spread was zero, this means the trades inside the spread were not used to estimate this parameters. This could be a good explanation for the smaller values. The autocorrelation is 0.1 smaller here, this could also be influenced by the trades inside the spread. Positive autocorrelation is due to large orders which are broken up to some smaller orders, because of easier execution. The results could be interpreted in the way that this broken-up trades are more often outside the

summary statistics of estimates for Toronto

		09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
ϕ					
	Mean	0.9346	1.0186	1.0322	1.0313
	Avg.Std.Er.	0.1367	0.0625	0.0646	0.0646
	Std.Dev.	0.8209	0.6560	0.6000	0.5519
	Median	0.9321	0.8853	0.9551	0.9485
θ					
	Mean	1.2910	0.7772	0.7115	0.7460
	Avg.Std.Er.	0.1503	0.0533	0.0579	0.0571
	Std.Dev.	1.0218	0.4427	0.4084	0.3974
	Median	1.0370	0.7154	0.6382	0.6823
ρ					
	Mean	0.2501	0.2505	0.2426	0.2249
	Avg.Std.Er.	0.0162	0.0135	0.0153	0.0148
	Std.Dev.	0.0085	0.0069	0.0076	0.0075
	Median	0.2618	0.2493	0.2462	0.2208

summary statistics of estimates for NYSE

		09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
ϕ					
	Mean	0.2660	0.3419	0.3423	0.2874
	Avg.Std.Er.	0.0793	0.0508	0.0494	0.0425
	Std.Dev.	0.1076	0.1682	0.1904	0.0927
	Median	0.24015	0.2648	0.2663	0.2611
θ					
	Mean	0.5476	0.3660	0.3126	0.3732
	Avg.Std.Er.	0.0930	0.0572	0.0540	0.0491
	Std.Dev.	0.2743	0.1838	0.1649	0.2695
	Median	0.4502	0.2881	0.2612	0.2593
ρ					
	Mean	0.2981	0.3040	0.2973	0.2741
	Avg.Std.Er.	0.0191	0.0143	0.0156	0.0151
	Std.Dev.	0.0523	0.0431	0.0461	0.0475
	Median	0.2966	0.3019	0.3100	0.2812

Table 2: This table presents the summary statistic over the 15 stocks per stock exchange. The values are the medians, the mean, the average standard error and the standard deviation in each trading interval, ϕ and θ are measured in Cent.

summary statistics of estimates for Toronto

		09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
S_i					
	Mean	4.4512	3.5916	3.4873	3.5546
	Avg.Std.Er.	0.2083	0.1089	0.1176	0.1176
	Std.Dev.	1.9235	1.4997	1.3289	1.4054
	Median	3.6397	2.9454	2.9942	3.3341
r_i					
	Mean	0.4550	0.4343	0.4117	0.4187
	Avg.Std.Er.	0.0452	0.0256	0.0284	0.0275
	Std.Dev.	0.2816	0.1949	0.1972	0.1788
	Median	0.4708	0.4473	0.4151	0.3517

summary statistics of estimates for NYSE

		09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
S_i					
	Mean	1.6272	1.4158	1.3097	1.3213
	Avg.Std.Er.	0.1209	0..080	0.0787	0.0691
	Std.Dev.	0.6501	0.5308	0.4877	0.5190
	Median	1.4149	1.2555	1.1240	1.0564
r_i					
	Mean	0.6555	0.5184	0.4873	0.5229
	Avg.Std.Er.	0.0930	0.0606	0.0637	0.0505
	Std.Dev.	0.1125	0.1381	0.1579	0.1551
	Median	0.6360	0.5283	0.5035	0.5222

Table 3: This table presents the summary statistic over the 15 stocks per stock exchange. The values are the medians, the mean, the average standard error and the standard deviation in each trading interval, the implied spread is measured in Cent, info part is the percentage of the asymmetric information component at the spread.

quoted spread than inside. This may be one possible explanation. With the estimated parameters, we can compute the spread implied by the model and the ratio of information component to the spread. The implied spread S_i and the information ratio r_i are computed as follows,

$$S_i = 2(\phi + \theta) \tag{14}$$

$$r_i = \frac{\theta}{\phi + \theta} \tag{15}$$

The implied spread has similar values than the estimated spread in the Huang and Stoll model. At both stock exchanges the spread has the U-shaped pattern, more pronounced at the TSX but also at the NYSE. Here the ratio between information part and the spread is declining through the day, at the last time interval a small increase. But this is consistent with the estimates in the original paper, the information ratio declines and ϕ, the orderprocessing and inventory holding costs grow from the beginning to the end of the day. Following MRR the decrease in θ and the increase in ϕ explain the U-shape in the spread, but θ increases in the last period, this leads to the increase of the spread in the last period. Especially since the order processing and inventory holding cost component increases in the last period. The standard errors are computed using the delta method.

$$\sigma_P = \frac{\partial P}{\partial \beta} Cov_\beta \frac{\partial P}{\partial \beta} \tag{16}$$

The variance of the parameter is the variance/covariance of the underlying structural parameters times their weight in the parameter to the square. Of course the implied spread is smaller than the one in the MRR paper, because it is a function of the underlying structural parameters. The asymmetric information ratio is higher, its from 0.48 to 0.65 in the MRR paper, from 0.36 to 0.51 here. The data MRR used was drawn in 1990, the data in this paper is from 2003. The spread could have declined through this 13 years. This and the inside spread estimation could explain the much smaller parameters and the spread. A interesting question is, how well does the modell fit, this could be expressed as how close is the objective function to zero. In case of an exactly identified model all sample moments are zero, but in case of an overidentified model, not all moments are exactly zero. But the closer they are to zero, the better the model fits. Hansen (1982) provides a test for overidentified restrictions. The J-Test, the smaller the J-statistic the better themodel fit. Here 112 parameter are estimated, for 108 the null, that the moments are zero, could not be rejected at the 1% significance level. Because of the wide range of trades per stock[12], it may be interesting to look at microstructure effects for more or less active stocks.

The estimation results are divided in 2 trade categories, one for often traded, onefor less often traded. Now there are seven often traded and seven less often traded stocks. There is a huge difference between often and seldomly traded stocks, not only θ the information parameter is higher in seldomly traded stocks[13]. Also ϕ and ρ are bigger. But the intraday pattern of the parameters

[12] Appendix table 8

[13] The higher information part for seldom traded stocks was analysed by Easily, Kiefer, O'Hara and Paperman (1996)

are much more interesting, the order processing and inventory holding costs do not follow a clear pattern. This has an reasonable economic interpretation. The risk to carry inventory over night is not very big for stocks often traded. The intraday pattern of ϕ for less often traded stocks is upward sloped over the day, at the NYSE there is a drop in the last period. But the value is still higher than in the opening period. This growing over the day can bevexplained, by growing risk for carrying the inventory over night. This risk is of course higher for seldomly traded stocks because it is not as likely to sell them as for active traded stocks. The pattern of θ, the asymetric information component, is highest in the first period and decreases over the day, because traders learn from the trading process. In Toronto θ is faster decreasing for seldom traded stocks, but it stays at a higher level because the starting value is twice the value for often traded stocks. In the last period the value increases, there is a higher information asymmetry in the end of the day. At the NYSE the information asymetry decreases over the day for often as well as for seldomly traded stocks. In opposite to the TSX the information asymmetry parameter for the seldom traded stocks reaches the value of the often traded stocks. This means the NYSE can reduce the information asymmetry of seldom traded stocks to the value of often traded stocks during the day. But for seldom traded stocks θ increases in the last period, this can not be explained. The autocorrelation of order flow is higher in New York so there are more splitted large orders in New York than in Toronto. At both exchanges the autocorrelation is higher for seldom traded stocks. This may be due to the fact that informed traders give large orders to maximize their expected profit. The higher share of informed traders at seldom traded stocks does explain the autocorrelation.

4.4 Price volatility

Price changes are driven by public news, asymmetric information and transaction costs. These factors also determine the price volatility. Using equation (12) the variance of stock price changes could be computed.

$$Var(\Delta p_t) = \sigma^2 + 2\phi^2(1-\rho) + \theta^2(1-\rho^2) + 2\theta\phi(1-\rho^2) \tag{17}$$

Where $\pi = var(\varepsilon)$ is the variance of public news shocks. This equation shows that, the variance consists of four parts (i) σ^2 the proportion attributable to the variance of public news (ii) $2\phi^2(1-\rho)$ the part due to transaction costs (iii)

$\theta^2(1-\rho^2)$ the part attributable to asymmetric information and (iiii) $2\theta\phi(1-\rho^2)$ an interaction part of private information and transaction costs. The bigger the autocorrelation in order flow the smaller the part attributable to microstructure noise. Now the ratios from microstructure noise to the whole variance can be computed as follows

$$\pi = \frac{2\phi^2(1-\rho) + \theta^2(1-\rho^2) + 2\theta\phi(1-\rho^2)}{\sigma^2 + 2\phi^2(1-\rho) + \theta^2(1-\rho^2) + 2\theta\phi(1-\rho^2)} \tag{18}$$

This is highly interesting, since this equation shows how much of the price volatility is due to the trading process. For π an additional parameter is needed, σ the variance of public news shocks. For estimating σ another orthogonality condition is needed.

$$E[\varepsilon^2 - \sigma^2] = 0 \tag{19}$$

This together with the other four moment condition leads to the following estimates. Estimating σ together with the other parameters or not, does not influence the other parameters, because the moment in equation (19) is independent of the other parameters. σ^2 is only the variance of the estimation residual. In table 4 the intraday pattern of the parts of price volatility are shown. The biggest part is the variation of public news shocks, it is about 75 percent of the variation of price changes. The proportion decreases 1% from the beginning to the end of the trading day. The absolute value is highest in the opening period and drops sharply the remaining three periods shows the familiar U-shape. The variation proportion of asymmetric information starts at nine percent and decreases throughout the day, the absolute value has the same pattern. The transaction cost part starts at 8 percent and jumps to about 13 percent for the rest of the day. The interaction part starts at 6 percent and increases to 9 percent at the end of the day.

In table 5 the intraday patterns at the NYSE are shown. The public information part starts at 78 percent, drops 1 percent and shows a U-shape for the last three periods. The absolute value decreases through the day. The asymmetric information part starts at 10 % drops to 5 % for two periods and increases to 8 percent in the last period. The transaction cost part decreases from the first to the third period and drops in the last. The interaction part proportion is more or less the same all the day, the absolute value decreases. The high public information variation in the morning reflects information which cumulated over night and uncertainty about the interpretation of the news. The decline over the statistics of volatility parts for Toronto

			09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
	var ε	vol.share	0.759	0.750	0.752	0.746
		Mean	19.919	12.901	12.157	12.626
		Std.Dev.	3.702	2.250	1.766	1.977
		Median	13.242	8.678	8.963	11.062
	Asym.Info	vol.share	0.092	0.042	0.038	0.039
		Mean	2.402	0.717	0.610	0.335
		Std.Dev.	3.885	0.771	0.701	0.638
		Median	1.006	0.485	0.372	0.638
	Trans.Costs	vol.share	0.088	0.126	0.131	0.127
		Mean	2.299	2.170	2.115	2.155
		Std.Dev.	3.169	3.446	2.819	2.776
		Median	1.313	1.267	1.367	1.448
	Interaction	vol.share	0.061	0.082	0.080	0.088
		Mean	1.582	1.415	1.294	1.581
		Std.Dev.	2.643	1.128	0.923	1.285
		Median	1.154	0.849	0.923	1.139

Table 4: This table presents the statistics of volatility parts over the 14 stocks per stock exchange. The values are the share of the price variation, the medians, the mean and the standard deviation in each trading interval,

statistics of volatility parts for NYSE

			09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
	var ε	vol.share	0.783	0.775	0.767	0.774
		Mean	2.658	2.005	1.717	1.742
		Std.Dev.	0.422	0.282	0.238	0.266
		Median	2.041	1.575	1.717	1.117
	Asym.Info	vol.share	0.099	0.058	0.050	0.084
		Mean	0.335	0.151	0.113	0.188
		Std.Dev.	0.350	0.150	0.110	0.298
		Median	0.177	0.076	0.060	0.062
	Trans.Costs	vol.share	0.035	0.075	0.096	0.058
		Mean	0.118	0.195	0.214	0.131
		Std.Dev.	0.107	0.189	0.257	0.075
		Median	0.082	0.101	0.104	0.098
	Interaction	vol.share	0.084	0.091	0.087	0.085
		Mean	0.284	0.235	0.194	0.191
		Std.Dev.	0.223	0.170	0.166	0.131
		Median	0.189	0.140	0.101	0.107

Table 5: This table presents the statistics of volatility parts over the 14 stocks per stock exchange. The values are the share of the price variation, the medians, the mean and the standard deviation in each trading interval,

day may be because of more frequent occurrences of public information early in the day. The asymmetric information volatility decreases, because traders learn from order flow and the private information decreases. The volatility part of public information is twice the value of the MRR (1996) estimated.

5 Conclusion

The model of Huang and Stoll does not deliver as interesting results as the MRR model. The problem is, the 3 way decomposition model does not work. The implied spread from the MRR model is equal to the estimated spread in the HS model. But the λ, the sum of inventory and asymmetric information costs in the HS model is smaller than the estimated asymmetric information component in the MRR model. This may be explained by the different influence of the transaction costs. In the HS model the midpoint and the price is influenced by the cumulated inventory of the dealer. In the MRR model only the inventory costs of the actual trade matters. The MRR model has also some problems. the first is the estimates do not support the theory of explaining the U-shape spread pattern by the decrease of information asymmetry and increasing inventory holding costs. The problem is mostly the last period the estimates do rise if they should fall or vice versa. This is the case for both trade activity categories, the pattern seems to change after the third period. But there is no possible explanation for an decreasing order processing/inventory holding cost component. The asymmetric information seems to be inversely related to the number of trades per hour. The number is decreasing from beginning to the third period, and increasing in the last. The information ratios reflect this, there is a small increase in the last period. The U-shape pattern is not pronounced in the implied spread, there is a small increase in the last period. The different patterns for the activity categories seems interesting, but the last period is once more the problem. This will need a new explanation. The volatility decomposition is easier to interpret. The public news volatility is dominating the volatility of transaction price changes. The microstructure generated volatility stays at about 25 percent over the day.

References

[1] Bloomfield, R., 1996, "Quotes, Prices, and Estimates in a Laboratory Market," Journal of Finance, 51, 1791-1808.

[2] Bloomfield, R., and M.O'Hara, 1996, "Market Transparency: Who wins and who loses?," working paper, Cornell University.

[3] Easily, D., and M. O'Hara, 1992, "Time and the Process of Security Price Adjustment," Journal of Finance, 47, 577-606.

[4] Easily, D., Kiefer, N., O'Hara, M., and J. Paperman 1996, "Liquidity, Information, and Infrequently Traded Stocks," Journal of Finance, 4, 1405-1436.

[5] French, K., and R. Roll, 1986, "Stock Return Variances: The Arrival of Information and the Reaction of Traders," Jorunal of Financial Economics, 17, 5-26.

[6] Harris L. 1986 "A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns", Journal of Financial Economics, 16, 99-118

[7] Hasbrouck, J., 1991, "The Summary Informativeness of Stock Trades: An Econometric Analysis," Review of financial studies, 4, 571-595.

[8] Hasbrouck, J., 1993, "Assessing the Quality of a Security Market: A New Approach to Transaction Cost Measurement," Reviw of Financial Studies, 6, 191-212.

[9] Huang, R., and Stoll, H., 1996, "The Components of the Bid-Ask Spread: A General Approach", Review of financial studies 1997, 10, 4, 995-1034

[10] Jain, P. and G. Joh 1988, "The Components of the Bid-Ask Spread: A General Approach", Review of Financial Studies, 10, 995-1034.

[11] Madhaven A., 1992, "Trading Mechanisms in Securities Markets," Journal of Finance, 47, 607-642.

[12] MCInish, T., and R. Wood, 1992, "An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks," Journal ofFinance, 47, 753-764.

[13] Madhaven, A., Richardson, M., Roomans, M., 1996, "Why do security prices change? A Transaction Level Analysis of NYSE Stocks", Review of financial studies 1997, 10, 4, 1035-1064

A Huang and Stoll model

Stocklist:

AGNICO-EAGLE MINES LTD
AGRIUM INC.
ALCAN INC
BARRICK GOLD CORP
BCE INC
GLAMIS GOLD LTD
IPSCO INC.
KINGSWAY FINANCIAL SVCS INC
MANULIFE FINANCIAL CORP
MDS INC
MOORE WALLACE INCORPORATED
NEXEN INC
NORANDA INC F
PETRO-CANADA COM
PETROKAZAKHSTAN INC

Trade numbers

B MRR model

Prof of Stationarity & Ergodicity

$E[\Delta p_t] = E[\Delta p_{t-k}] = 0$ $cov(\Delta p_t, \Delta p_{t-j}) = cov(\Delta p_{t-k}, \Delta p_{t-k-j})$ $cov(\Delta P_t, \Delta P_{t-k}) =$
$E[((\phi + \theta)Q_t - (\phi + \rho\theta)Q_{t-1} + \varepsilon_t + \eta_t + \eta_{t-1})$
$((\phi + \theta)Q_{t-k} - (\phi + \rho\theta)Q_{t-k-1} + \varepsilon_{t-k} + \eta_{t-k} + \eta_{t-k-1})$
for $k > 1$
$cov(\Delta P_t, \Delta P_{t-k}) = ((\phi + \theta)Q_t - (\phi + \rho\theta)Q_{t-1})$
$((\phi + \theta)Q_{t-k} - (\phi + \rho\theta)Q_{t-k-1})$
Q_t is a Markov process of order p
for $k > p \Rightarrow cov(\Delta P_t, \Delta P_{t-k}) = 0$
$\sum_{i=0}^{\infty} \rho(i) = \sum_{i=0}^{k} \rho(i) < \infty$

descriptive volume statistics for Toronto an NYSE

		09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
Toronto					
	Mean	7809	11388	8842	9522
	Median	5313	7434	5565	6237
	Std.Dev.	5837	8765	7068	7414
	Minimum	941	1097	916	1149
	Maximum	18058	29154	24394	24862
NYSE					
	Mean	7166	12081	9983	10713
	Median	4847	9148	8802	9105
	Std.Dev	6952	11476	9097	9569
	Median	22165	37965	30722	30415
	Minimum	268	569	489	527
	Maximum	22165	37965	30722	30415

Table 6: This table presents the number of trades over the 15 stocks per exchange.

Derivation of $E[Q_t|Q_{t-1}] = \rho Q_t$

$\gamma = P[Q_t = Q_{t-1}]$

$E[Q_t] = 0$

$var[Q_t] = 1^2 P[Q_t = 1] + (-1)^2 P[Q_t = -1] = 1$

$E[Q_t Q_{t-1}] = 2(1-\gamma)$

$\rho = \frac{E[Q_t Q_{t-1}]}{var[Q_t]}$

$E[Q_t|Q_{t-1} = 1] = \gamma - (1-\gamma)$

$E[Q_t|Q_{t-1} = -1] = -\gamma + (1-\gamma)$

$\Rightarrow E[Q_t|Q_{t-1}] = \rho Q_t$

Insignificant parameters

Activity categories

Insignificant parameters for the MRR model

Stockade	Exchange	Daytime	Parameter	Value	std.Dev.	p-Value
IPS	NYSE	09:30-10:30	ϕ	0.0731	0.3947	0.85
IPS	NYSE	09:30-10:30	θ	0.6704	0.4609	0.16
IPS	NYSE	10:30-12:30	ϕ	0.7033	0.2785	0.01
IPS	NYSE	10:30-12:30	θ	0.2107	0.3233	0.51
IPS	NYSE	12:30-14:30	θ	0.0096	0.3187	0.97
IPS	NYSE	14:30-16:00	ϕ	0.1009	0.2262	0.66
KFS	NYSE	10:30-12:30	θ	0.1531	0.0644	0.02
KFS	NYSE	14:30-16:00	θ	0.1143	0.0616	0.06
MWI	TSX	09:30-10:30	ϕ	-0.6297	0.7704	0.41
NRD	NYSE	09:30-10:30	ϕ	0.2020	0.0951	0.03
PKZ	TSX	09:30-10:30	ϕ	0.2192	0.2056	0.29
NXY	NYSE	09:30-10:30	ϕ	0.4068	0.1630	0.01

Table 7: All insignificant parameters in the MRR model

summary statistics of estimates for Toronto

			09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
often	ϕ	Mean	0.8476	0.8373	0.8729	0.8253
		Median	0.8963	0.8224	0.7924	0.7676
	θ	Mean	0.8592	0.6369	0.5757	0.6133
		Median	0.9064	0.6430	0.5564	0.5782
	ρ	Mean	0.2225	0.2226	0.2230	0.2102
		Median	0.2396	0.2428	0.2396	0.2075
seldom	ϕ	Mean	1.0216	1.1999	1.1914	1.2373
		Median	0.9546	0.9290	1.0144	1.0266
	θ	Mean	1.7228	0.9176	0.8472	0.8787
		Median	1.4221	0.7878	0.6784	0.9421
	ρ	Mean	0.2777	0.2784	0.2622	0.2396
		Median	0.2852	0.2596	0.2591	0.2427

summary statistics of estimates for NYSE

			09:30-10:30	10:30-12:30	12:30-14:30	14:30-16:00
often	ϕ	Mean	0.2959	0.2730	0.2616	0.2731
		Median	0.2421	0.2161	0.2219	0.2537
	θ	Mean	0.4175	0.3368	0.3070	0.3010
		Median	0.4060	0.3235	0.2608	0.2383
	ρ	Mean	0.2733	0.2908	0.2889	0.2683
		Median	0.2805	0.2943	0.3053	0.2661
seldom	ϕ	Mean	0.2361	0.4107	0.4229	0.3017
		Median	0.2020	0.3732	0.3785	0.3061
	θ	Mean	0.6777	0.3953	0.3182	0.4455
		Median	0.6704	0.2527	0.2616	0.4259
	ρ	Mean	0.3228	0.3212	0.3057	0.2799
		Median	0.3183	0.3072	0.3301	0.2826

Table 8: This table present the estimation results for often and seldom traded stocks. Often means more than the median.