Bachelor Thesis

Jort Cuperus, 347474
Supervisor: mr. S. Ozturk

July 1, 2013

Time based analysis of the Hasbrouck measure of price discovery

Abstract

In this paper the matter of price discovery (i.e. the incorporation of new information in the security price) in fragmented markets is addressed. I refer to fragmented markets when the same security is traded in different markets. This paper investigates which market is leading in the process of price discovery over a major time frame, where the trading volume is increasing strongly, and for comparison I analyse simularly for a relative short time frame, where the trading volume is significant lower than first mentioned investigation. Especially, I try to determine whether the increasing market efficienty is reflected in the price discovery. The paper describe the model from Hasbrouck (1995) and the corresponding measure for information shares of the multiple markets. As a new application, I consider the midquotes across the nine markets of IBM from 2002 till 2007 and for comparison I also consider the midquotes across the seven market of Expedia in July 2007. The paper concludes that for both stocks the market leader can be determinated, seasonality and start-up problems do not seem relevant for price discovery.

The number of stock trades and the trading speed have increased exponentially over the past few years. According to a Philippine Stock Exchange [1] study, the total online transaction value rose in 2012 with 18.8 percent to 195.75 billion dollars. This shows that the financial markets have become more open and aggressive to embrace technological advancements in terms of finding alternative ways of participating in the stock market. The number of datapoints in the IBM data I used for this research increased with more than 7.5 times in six years.

As already mentioned, markets are not completely efficient in pricing. The same share can be priced differently across multiple markets. When a market is priced at multiple different markets, the price of the stock can differ, but will not diverge.

The structure of this paper is as follows. Section I presents the model and corresponding information share measure used in Hasbrouck (1995) and some other researches which are worth mentioning. Furthermore, the methods concerning seasonality and start-up problems can also be

[1] Market X, see Table 6 in Appendix A.

found in Section I. I give an application to data of IBM and Expedia in Section II. The results can be found in Section III. To conclude, I give my conclusions in Section IV.

I Hasbrouck method

There has been quite some research about price discovery. First I briefly describe the most important researches about price discovery which are *not* based on Hasbrouck (1995), but give us a nice insight in the fields of price discovery. Remark that most of them are published before 1995. Then, I discuss researches about price discovery, based on the Hasbrouck model, which are discussed, applied and extended in their papers. Finally, I discuss Hasbrouck (1995) extensively.

I.1 Researches about price discovery - not based on Hasbrouck (1995)

In former times, it was quite easy to determine the market leader, but since the number of transactions has grown unprecidented like the last years, we sometimes can not see the wood for the trees. Garbade and Silber (1979) analyzed NYSE and regional exchange trading patterns, and noted that the regional exchanges were not as satellite as NYSE, i.e. regional exchanges have a less advanced data recovery system that ensures less fast data generating. In some regional exchanges, systems automatically change quotes to the NYSE quotes, also known as 'autoquotes'. The NYSE and the regional exchanges are electronically linked, so all trades and quotes are brought together by a central transmission authority.

Shapiro (1993) concluded that the volume of trading in NYSE-listed equities has quietly moved from the NYSE market to alternative, regional markets. Manaster and Mann (1996) presented evidence that exchange locals have informational advantage over the off-exchange traders. Locals have two advantages: they can time their trades more closely and have low execution costs. According to Kurov and Lasser (2002), higher trade volume over the years makes the difference between core (NASDAQ, NYSE, BATS) and regional exchanges vanish. Kurov and Lasser (2002) disprove the hypothesis that trades initiated by exchange locals are more informative than off-exchange initiated trades.

Harris *et al.* (2002) apply the techniques of Gonzalo and Granger (1995) to analyze the common factor weight. The estimation is done from a fully specified error-correction model. From this, the Harris-McInish-Wood (2002) measure follows. The resulting weights can be interpreted, according to Mizrach and Neely (2008), as the changes in the price in relation to the shock vector, when the time horizon goes to infinity. Baillie *et al.* (2002) present theoretical evidence that the Hasbrouck information share performs better than the Harris-McInish-Wood measure, so I will use the Hasbrouck measure in this research.

I.2 Researches about price discovery - based on Hasbrouck (1995)

The Hasbrouck method is often discussed, applied and extended in the financial literature. I present you the most prominent subjects:

Later research showed the presence of more factors which influence price discovery than Hasbrouck (1995) assumes. Mizrach and Neely (2008) illustrate that both transitory factors (daily variation in liquidity, volatility and macro-economic announcements), but also long-term trends (the movement to electronic markets) influence price discovery.

There are introduced some new measures for the contribution of each market to price discovery, related to the Hasbrouck (1995) information shares. I present the most prominent ones:

As already mentioned, the first alternative is the Harris-McInish-Wood measure, based on the Gonzalo and Granger (1995) method, and described in Harris *et al.* (2002), which provides worse results than Hasbrouck (1995). Hasbrouck (2002) uses a simulation model to criticize the Gonzalo-Granger (GG). However, the resulting weights can be interpreted, according to Mizrach and Neely (2008), as the changes in the price in relation to the shock vector, with the time horizon to infinity.

De Jong and Schotman (2010) defined the information shares directly within the unobserved components model, instead of defining the measure within a reduced-form time-series model. de Jong and Schotman (2010) generalize the structural model of Hasbrouck (1993), which Hasbrouck also used in Hasbrouck (1995), to a multi-variate setting. Hasbrouck (1995) suggests to provide upper and lower bounds, acquired by different ordering of the markets, which results in narrow bounds in a two variable system, and wide bounds in multiple variable system. De Jong en Schotman (2010) information share measure does not depend on an arbitrary way to split the correlation over the different markets, and is meaningful in high dimensional settings.

Yan and Zivot (2010) add the component share (CS) to the information share (IS) model. The use of the CS in conjunction with the IS can demount the compounding effects of the two types of shocks, because the use of IS estimations, based on high sampling frequencies, may be incorrect by transitory frictions and may leave important price discovery information unused.

Korenok *et al.* (2011) compare the four alternative measures: Hasbrouck, Harris-McInish-Wood, de Jong-Schotman and Yan-Zivot. The paper describes analytically the problems with the measures, for example negative information shares, non-uniqueness and potential violations of market efficiency. Korenok *et al.* (2011) show their findings with some simulation evidence.

Some other measures are designed, based on the Hasbrouck (1995) information shares measure, by incorporated the criticism against Hasbrouck (1995) into their, abovementioned, new measures. The most prominent paper that criticism Hasbrouck is Grammig and Peter (2010), who resolve the main drawback of the widely used Hasbrouck (1995) methodology and demonstrate those main drawbacks by some simulations.

To sum up, the Hasbrouck information share method has brought some nice insight to price discovery, but is, like most measures, not without limitations.

I.3 Theoretical background Hasbrouck

In the literature of price discovery, common factor models can be used to measure the contribution of the multiple markets, in which the same stock is traded, to the price discovering process. Plentyful articles in the financial literature of the past several decades provide an interesting debate on these common factor models. Especially, the Hasbrouck (1995) and Gonzalo and Granger (1995) common factor models are discussed and applied. Like Baillie *et al.* (2002) argue, these two models complement each other and provide different views of the price discovery process between markets.

Hasbrouck (1995) introduce an econometric approach based on an implicit unobservable efficient price common to all markets. This approach relies on cointegration, i.e. the feature that while two prices may be non-stationary, they do not diverge from each other without bound. Statistically, the common factor is defined as the random walk component of the prices from the different markets. The innovation variance in this random walk is a measure of the information intensity of the efficient price process. Hasbrouck (1995) defines the information share as the proportion of the innovation variance of this random walk that can be attributed to a certain market. Furthermore, Hasbrouck (1995) introduces upper and lower bounds by permutating the groups in the covariance matrix. The spread between the two bounds is positively related to the degree of correlation. Baillie *et al.* (2002) argue that the average of the upper and lower bounds provides a sensible estimate of the contribution of the markets to the impounding of new information in the efficient price.

The most prominent research of the abovementioned is Hasbrouck (1995). He used common factor models, elucidated later in this paper, to determine which market is leading in price discovery. Nowadays the market has become much more efficient, by, for example, automatic trading and higher internet speed. Therefore, in this paper I will review whether the method of Hasbrouck (1995) is losing value over time, because of the exponential growth of the stock trading volume and not just because of the increase of the factors noise trading and temporary order imbalances.

My research considers the impounding of new information into the price (i.e. price discovery) and, more specifically, where this new information is incorporated. Especially, I want to investigate the possible changes in clarity of determining a market leader over time, according to the Hasbrouck information share method.

In my research I provide an application of the Hasbrouck (1995) information share measures, where the focus will lie on whether or not the presence of a market leader. So, the main questions are, whether it is still possible to determine the market leader when the stock trading volume increases, like it did over the last years and if there is a difference in clearness by determining the market leader between a stock with relative high trading volume and a stock with relative low trading volume. So, it is important to note that I will use the Expedia data to analyze the price discovery based on the difference in trading quantity, since Expedia has less trading volume than IBM. More about the used data can be found in Section II.

In section I.4 till I.6 I describe the main model I use for my empirical analysis in Section III. Section I.4 presents a basic microstructure model to get a better understanding of the underlying principles. The Hasbrouck (1995) model is presented in section I.5, followed by the Hasbrouck information share measures in section I.6.

I.4 A basic microstructure model

Hasbrouck (1995) uses a simple microstructure model to show the cointegration principles of stocks, which we shall see in later analyses. Consider a security that is traded across two different markets, with following prices:

$$Y_{1,t} = Y_{1,t-1} + w_t, \tag{1}$$
$$Y_{2,t} = Y_{1,t-2} + \epsilon_t,$$

where $Y_{1,t}$ and $Y_{2,t}$ are empirical price variables such as quotes (e.g. the midquotes, like I will use in the empirical analysis in Section III). The ω_t and ϵ_t are i.i.d. distributed variables. I define Y_t as $(Y_{1,t}, Y_{2,t})'$.

Because the prices are non-stationary, I take the first differences of the prices and the price changes are stationary. This leads to the following:

$$Y_{1,t} - Y_{2,t} = Y_{1,t} - (Y_{1,t-2} + \epsilon_t) = \omega_t + \omega_{t-1} - \epsilon_t. \tag{2}$$

I realize a linear combination of integrated and stationary variables. In this case, the variables are cointegrated.

I.5 Hasbrouck (1995) model

Baillie *et al.* (2002) state that the Hasbrouck (1995) model starts from the following Vector Error Correction Model (VECM):

$$\Delta Y_t = \delta\beta' Y_{t-1} + \sum_{j=1}^{k} A_j \Delta Y_{t-j} + \epsilon_t. \tag{3}$$

In this representation, δ is the error correction vector, β the cointegration vector and ϵ is a zero-mean vector of serially uncorrelated innovations with covariance matrix Ω and k implies the number of lags included in the model.

It is important to recall that the equation for the VECM consists of two parts. The first part, $\delta\beta' Y_{t-1}$, is the long-run dynamics between the prices across the different markets, and the second part, $\sum_{j=1}^{k} A_j \Delta Y_{t-j}$, is the short-run dynamics, sometimes seen as due to market imperfections.

Starting from the VECM, Hasbrouck (1995) expresses the price differences by a Vector Moving Average (VMA) model:

$$\Delta Y_{1,t} = w_t, \tag{4}$$
$$\Delta Y_{2,t} = Y_{1,t-2} + \epsilon_t - \epsilon_{t-1}.$$

In this representation, the price changes are exclusively expressed in terms of current and lagged incorporation of information in the price. The more general form is:

$$\Delta Y_t = \Psi(L)e_t, \tag{5}$$

and its integrated form:

$$Y_t = \Psi(1)\left(\sum_{s=1}^{t} e_s\right) + \Psi^*(L)e_t, \tag{6}$$

with $\Psi(L)$ and $\Psi^*(L)$ as polynomials in the lag operator L.

When I assume $\psi = (\psi_1, \psi_2)$ to be equal to the common row vector in $\Psi(1)$, the sum of the moving average coefficients, the integrated form can be presented as:

$$Y_t = \iota\psi\left(\sum_{s=1}^{t} e_s\right) + \Psi^*(L)e_t, \tag{7}$$

I.6 Hasbrouck (1995) information share measure

For determining which market is leading, I use the Hasbrouck (1995) measure derived for calculating the information share, i.e. the market stock variance of a certain market relative to the total variance (S_j). The shares can be interpreted as a measurement of the proportion of variance that is explained by shocks of the certain market. In this research, the market with an higher share than the other markets is called the leading market.

If there is no correlation between the markets, i.e. Ω is a diagonal matrix, Hasbrouck (1995) defines the information share for market j by:

$$S_j = \frac{\psi_j^2\sigma_j^2}{\psi\Omega\psi'}. \tag{8}$$

Unfortunately, this is not quite often the case, so in the case of significant correlation between the price innovations accross markets, Hasbrouck (1995) applies the Cholesky factorization of Ω. By this factorisation, we get $\Omega = MM'$, with M a lower triangular matrix. The information shares are given by:

$$S_j = \frac{([\psi M]_j)^2}{\psi\Omega\psi'}, \tag{9}$$

where $[\psi M]_j$ is the the j^{th} element of the vector ψM.

The factorization implies a hierarchy that gives a higher information share for the first price and a lower information share for the last price, in most cases. Hasbrouck (1995) derives upper and lower bounds for the information share of a market by performing some permutations in the covariance matrix Ω. Hasbrouck (1995) changes the order of the rows in *n!* different ways, with n defined as the number of groups. In my application I performed six permutations, because as described in Section III, I decided to distribute the markets into *three* groups in both stocks.

I use this measure for six years of IBM shares and a month of Expedia shares. I average the upper and lower bounds of the information shares for every month and analyse over years (weighted average of the months in a certain year) and seasonality (weighted average of the month in the sample periode). For analysing the start-up problems I average the upper and lower bounds of the shares for the days before closing of the financial markets and days after closing of the financial markets for the whole sample period, what will be in the most cases Mondays and Fridays, respectively.

II Application: six years IBM and one month Expedia

This section introduces the background and data for my applications. In Section II.1 I will mention some details about the company IBM and several remarkable characteristics of the price pattern between January 2002 and December 2007. This section will not be too specific about the Expedia data, because I analyze Expedia only for a month to compare the influence of trading volume to the information shares of the IBM stock. An introduction of my data of IBM and Expedia is provided in Section II.2.

II.1 Development stock IBM

International Business Machines Corporation, or IBM, is an American multinational technology and consulting corporation and has become big by offering infrastructure, hosting and consulting services in areas ranging from mainframe computers to nanotechnology. For example, IBM designed chips are currently used in PlayStation 3, Xbox 360, and Wii game consoles.

The company has undergone several organizational changes since its inception, where acquiring organization PwC (PricewaterhouseCoopers) consulting business in 2002 was a important one. At the 4^{th} of Januari the price of one IBM stock was \$125.60 and fell down to \$56.60 in exactly nine months, which implies a decrease of 55%. After purchasing PwC the value of the stock roses with 54% the next two months to \$87.10 per share. The share price development of IBM between 2002 and 2007 is shown in Figure 1.

Figure 1: The development of the IBM stock price (January 2, 2002 - December 31, 2007)

II.2 Data

For this research I work with midquotes for the six years IBM data, i.e. from the 2^{nd} of January 2002 till the 31^{th} of December 2007 and one month Expedia data, i.e. from the 2^{nd} July 2007 till the 31^{th} July 2007. The exchanges are opened from 9.30 a.m. till 4.00 p.m., except for weekends and Holidays. I have corrected my data for outliers. I defined an outlier as an ask or bid quotes which is four times the daily standard deviation away from the previous and following ask / bid quotes, or quotes with a bid-ask spread higher than 3.50 dollars for the IBM data and 0.30 dollars for the Expedia data. This gap can be explained by the difference in the stock price: an IBM share on 20 july 2007 was worth approximately 115 dollars, and an Expedia share was worth only

27.50 dollars on the same day. Furthermore, I deleted quotes with a zero bid and / or ask, and quotes with bid prices higher than ask prices.

IBM IBM stocks are traded on seven, eight and nine American exchanges in 2002 till 2005, 2006 and 2007 respectively, see Table 1. I allocated the data into three different groups. The first group consists markets P and N (NYSE), the second group consists of markets B, T and X (NASDAQ) and the third group consists the rest of the exchanges. Table 6, which can be found in Appendix A, provides the corresponding market with the letter. The share of transactions for NYSE, NASDAQ and REST are around 50%, 35% and 15%, respectively. It is a remarkable fact that the proportion of the transactions of the NYSE-group declines and the share of transactions of the REST-group increase over time. Abovementioned sets of markets are the best possible distribution of exchanges over (meaningful) groups.

Table 1: Trades distributed over the 9 markets IBM is enlisted on

		% trades					
Group	**Code**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
NASDAQ	B	4,6	1,9	3,4	2,5	1,4	0,0
REST	C	3,0	1,7	1,5	1,6	4,4	14,0
REST	D	-	-	-	-	29,8	2,1
REST	I	-	-	-	-	-	10,8
REST	M	7,1	6,8	7,3	0,1	0,1	0,0
NYSE	N	35,6	23,6	28,6	26,3	19,4	23,9
NYSE	P	13,7	29,7	39,3	32,4	18,8	20,3
NASDAQ	T	28,0	32,9	16,1	32,3	21,6	24,9
NASDAQ	X	8,0	3,3	3,8	4,9	4,6	4,0

In Table 2, the auto-covariances between the certain years are shown. The NASDAQ group has more autocorrelation than the other two groups, but definitely not significant.

Table 2: Auto-covariances of data based on the log prices (one-second sampling), multiplied by 1000.

	IBM					
	2002	**2003**	**2004**	**2005**	**2006**	**2007**
2002	33,19	-3,39	6,41	9,93	1,72	-7,68
2003	-3,39	2,44	-0,51	-1,09	-0,02	1,12
2004	6,41	-0,51	2,83	2,58	-0,33	-0,74
2005	9,93	-1,09	2,58	5,96	0,17	-2,84
2006	1,72	-0,02	-0,33	0,17	1,69	-1,34
2007	-7,68	1,12	-0,74	-2,84	-1,34	3,51
σ^2 = 0,1488						

In my analysis, I use high frequency tick data sampled on a one-second frequenty, because, according to Goodhart and O'Hara (1997), the Hasbrouck method generates no significant results for lower frequenties than a few seconds. If there were given multiple prices in a second, I took the last observed price. This an arbitrary choice, because taking the first observed price of that second would obviously lead to nearly the same results. If there was a second where no price was observed or where I deleted the price by our filter criteria, I filled this with the price of the closest observation on a daily basis.

Finally, it is important to recall and show that the speed of information transmission has grown over the years. In Figure 2 the upward trend of datapoints is noticeable, especially the growth since 2005 looks exponential. According to Figure 2, there were on average 1.17 price updates per second[2] in 2002 and in 2007 that number was on average close to tenfold: 10.18 price updates per second.[3] Furthermore, it is worth mentioning that the number of erroneous price updates, which I deleted by my filters, increases proportionally over the years.

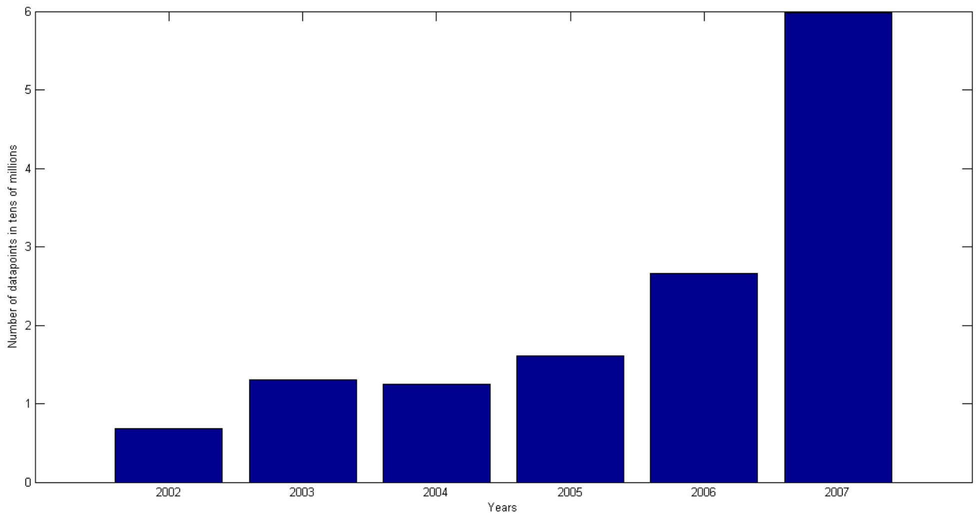

Figure 2: The development of the number of price updates of the stock IBM (2002 - 2007)

Expedia Expedia stocks are traded on seven American exchanges in July 2007. Data for our second application consist of midquotes for Expedia sampled at the one-second frequency for the 21 trading days in July 2007. Quotes are taken from the TAQ database and split in three series depending on the origin of the quote, according to de Jong and Schotman (2010). The first group consists market P (NYSE), the second group consists market D and T (NASDAQ) and the third group consists the rest of the markets. Again, in Appendix A, Table 6, can be found which letter implies which market. The overall fraction of quotes issued by these three groups is around 47% for NASDAQ, 18% for NYSE and the remaining 35% for the other markets.

[2]based on 252 trading days with 6.5 trading hours.
[3]based on 251 trading days with 6.5 trading hours.

III Results

This Section gives an overview of the results from my analysis on the IBM and Expedia data. I start with the calculated information shares by the Hasbrouck (1995) measure over the six years of IBM data in Section III.1. In Section III.2 I analyse the IBM data for seasonality and Section III.3 analyses the IBM data for start-up problems in financial markets. Finally, the information shares for the Expedia data for July 2007 will be presented in Section III.4.

III.1 Information share analyse over years (IBM)

For obtaining the Hasbrouck (1995) information shares for six years, I calculate the midquotes, defined as the mean of the bid and ask quotes. By permuting the different groups (NYSE, NASDAQ and REST) in six different ways, as described in Section I.6, I deduce the upper and lower bounds. These bounds are calculated as the minimum and maximum of the information shares per group.

I estimated the parameters of the VECM, as presented in Equation (3), by using OLS, including 300 lags for our one-second sampling data. I need to transform our VECM to a VMA model to calculate the information share. I do this with a so-called intermediate model, namely the Vector Auto Regressive (VAR) model:

$$\Delta Y_t = \beta + \sum_{j=1}^{k} A_j \Delta Y_{t-j} + \epsilon_t. \tag{10}$$

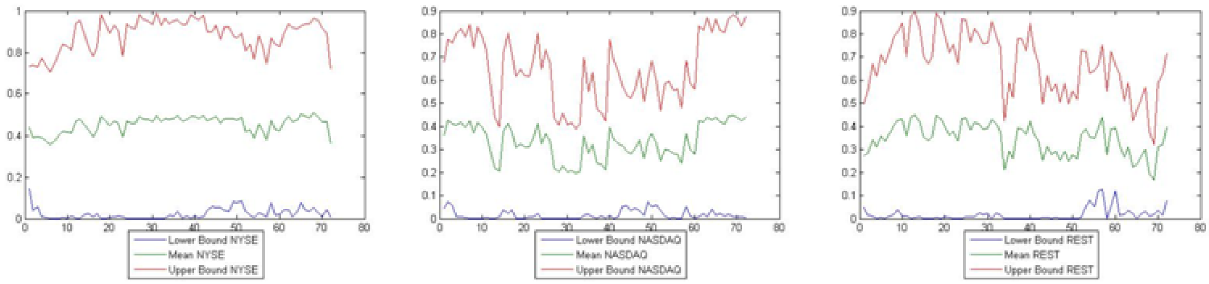

Figure 3: The information shares for the different groups, calculated with Hasbrouck (1995) on one-second sampling midquotes, 300 lags, 72 months (January 2002 till December 2007)

Figure 3 presents graphs of the lower and upper bounds and its mean for the three different groups, for the midquote data. For this, I use the one-second sampling quotes in a model with 300 lags, which I average per month. These results can be found in Table 7, which is included in Appendix A. Table 3 shows the information share results for one-second sampling data in a model with 300 lags, averaged per year.

When having a look at the results, I can conclude that the information shares of all groups fluctuate heavily over time. When averaging the information shares on monthly basis, this reduces the fluctuation properly. However, in the NASDAQ group and from 2005 in the REST group the information share fluctuations are still present to a certain degree. This latter is probably due to the fact that the IBM stocks have been traded on more markets from 2005, which I added to the REST group. [4] The reason why the information share in the NASDAQ group is not stable, like it is the case with the NYSE group, is not clear. Moreover, what immediatelly strikes is the enormeous wide of the spread between the upper and lower bounds. This is due to the fact of high contemporaneous correlations between the different groups on one-second level.

Table 3: Hasbrouck (1995) information shares, calculated on a daily basis (using a one-second frequency with 300 lags), averaged on yearly basis.

		MIDQUOTE	
		Lower	Upper
	BTX	0,0170	0,7612
2002	PN	0,0240	0,7825
	REST	0,0147	0,7093
	BTX	0,0134	0,6276
2003	PN	0,0110	0,8931
	REST	0,0037	0,7839
	BTX	0,0038	0,5145
2004	PN	0,0058	0,9494
	REST	0,0113	0,7234
	BTX	0,0237	0,5688
2005	PN	0,0278	0,9335
	REST	0,0020	0,6559
	BTX	0,0267	0,5702
2006	PN	0,0402	0,8363
	REST	0,0496	0,6335
	BTX	0,0151	0,8470
2007	PN	0,0370	0,9065
	REST	0,0258	0,5132

Based on Table 3 and Figure 3, the leading market by using the Hasbrouck (1995) information shares is quite clear. Table 3 established that the largest information share always belongs to the NYSE group. The years 2002 and 2007 might not be entirely convincing, but the other four years are. On average, the NYSE group contains around 45% of the information of the IBM security price, the NASDAQ group and the REST group both around 32.5%. Figure 3 shows us that the NYSE group has the larges information shares quite often and is also the most stable group of markets. Remarkable is the fact that the information share of NYSE decline in the end of 2007 heavily, while the NASDAQ group information share rises violently in the begin of 2007, so that might be an interesting topic as an addition to this research. Note that over the whole sample periode it is possible to designate the market leader, even with the growth of the stock trading volume over the years.

III.2 Information Share seasonality analyse (IBM)

Seasonality is a characteristic of a dataset in which the data experiences regular and predictable changes which occur and repeats every calendar year. It is important to recall that traders who understand the seasonality of certain companies can time trades and make other decisions to co-

[4] See Table 1.

incide with the expected seasonality.

Some research has been done about the seasonality of stock price changes. Figure 4 demonstrates the average price change in every month of the last 33 / 34 years, [5] and the corresponding 95%-confident interval, so this provides a general insight in the seasonality of IBM. In this section I investigate seasonality of the Hasbrouck information share measure, applied to the IBM shares.

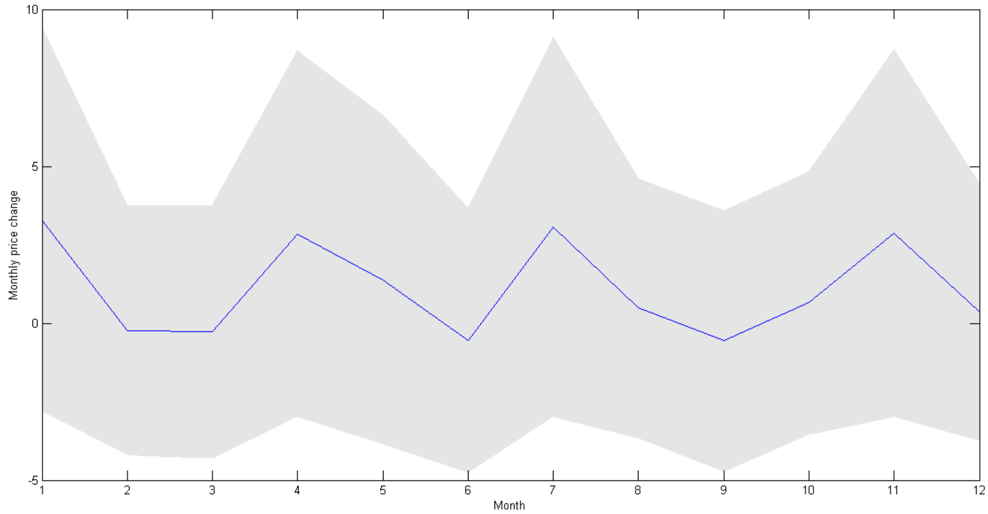

Figure 4: 95%-confident interval of the monthly price change of shares IBM over the period January 1980 till May 2013

I averaged the information shares of all the days in the certain month over the whole sample period. The extensive table of this, Table 8, can be found in Appendix A. Unfortunatelly, I can not remark any significant results. Every single month (over the years) shows an information share for NYSE, NASDAQ and REST around 45, 32 and 33%, not much different than the results in Section III.1.

III.3 Information Share start-up problems analyse (IBM)

A considerable number of papers are written about start-up problems of financial markets. Most of them focus on the intra-day start-up problems, also known as the U-shaped curve of trading volumes. At the start of each day overnight information is incorporated in the asset price, resulting in higher trading volumes. At the end of the day trading volumes increase as well, due to high-frequency traders and day-traders who are closing their positions to avoid overnight trading risk.

In this section I investigate the effect of these start-up problems on the information shares, and how these problems relate to the extremely high costs of market closure. I will not analyse this

[5] 34 years for [January - May], 33 years for [June - December].

problem on an intra-day basis, but I compare days where the market is closed on the next day with days where the market was closed on the previous day. If a day meets both criterea, I do not use that day. In general, this will mostly contain comparisons between Mondays and Fridays, but in weeks with Holidays this includes also other days. It could be possible that the U-shaped volume pattern on a intra-day basis, the different levels of volatility and spread size could influence the price discovery patterns on days before closing and after closing of the financial markets.

Table 4: Hasbrouck (1995) information shares, calculated on a daily basis (using a one-second frequency with 300 lags), averaged for the days *before* closing of the financial markets and days *after* closing of the financial markets.

		MIDQUOTE	
		Lower	Upper
BEFORE	BTX	0,0167	0,6476
	PN	0,0204	0,8785
	REST	0,0229	0,6667
AFTER	BTX	0,0164	0,6267
	PN	0,0240	0,8816
	REST	0,0144	0,6834

According to Table 4, the information shares for both types of days is not significant different. This could imply that each market experienced the same quantity start-up problems or that the markets simply do not face start-up problems.

III.4 Information shares analyse (Expedia)

Table 5: Hasbrouck (1995) information shares for Expedia and IBM, calculated on a daily basis (using a one-second frequency with 300 lags), averaged for July 2007.

		MIDQUOTE								
		Expedia				**IBM**				
		Lower	Upper	Mean	Scaled Mean	Lower	Upper	Mean	Scaled Mean	
July 2007	NASDAQ	0,0111	0,4004	0,2058	0,2145	0,0502	0,5938	0,3220	0,3054	
	NYSE	0,0483	0,7995	0,4239	0,4420	0,0373	0,9058	0,4716	0,4472	
	REST	0,0233	0,6357	0,3295	0,3435	0,0057	0,5161	0,2609	0,2474	+
				0,9592	1,0000			1,0545	1,0000	

Table 5 shows that for the Expedia stock NYSE, despite the fact that the NYSE group takes account for only 18% of the market trades, is the unchallenged market leader. The range of the upper and lower bound is in both stocks simular. When having a look at the difference between the information share of Expedia and IBM for July 2007, I can not present any results, based on the lower and upper bounds. I calculate the mean of the upper and lower bounds, and scale them. Based on the scaled means of the upper and lower bounds of the information shares of both stocks, the determination of the market leader is more convincing at the IBM stock than the Expedia stock.

Based on Figure 5, in comparison with Figure 3, the information shares of Expedia fluctuated more heavily than the information shares of IBM. This is though incorrect, because Figure 3 shows the information shares on monthly basis. In fact, on daily basis the information shares of Expedia fluctuate significantly less than the information shares of IBM. The possible explanation for this might be concealed in the fact that the trading volume of IBM is much higher than the trading volume of Expedia. For comparison, the Expedia data provided in July 2007 1.4 million datapoints and the IBM data 6.3 million in the same time frame.

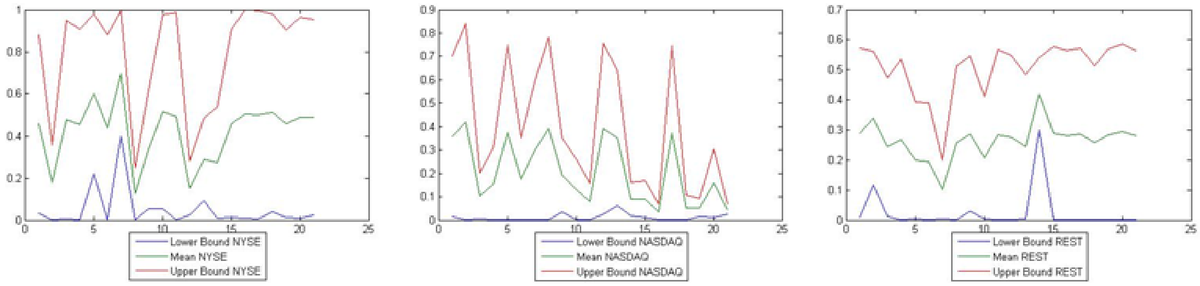

Figure 5: The information shares for the different groups of the Expedia stock, calculated with Hasbrouck (1995) on a one-second sampling midquotes, 300 lags (July 2007)

IV Conclusions

In this paper the method of Hasbrouck (1995) is used to calculate the information shares of three different groups of markets on which the stock IBM is enlisted on from 2002 till 2007. As a second application I calculate the information shares for the Expedia stock in July 2007.

Hasbrouck (1995) uses a Vector Error Correction Model (VECM), in which the current log-price movement of a stock is regressed on the past stock price, and the residuals in the previous periods. The VECM boils down to a Vector Moving Average (VMA) model. With the use of permutations, the lower and upper bounds for the different information shares were calculated.

According to de Jong and Schotman (2010), I have shown that the method of Hasbrouck (1995) results in a large spread between the lower and upper bounds for the different groups. I determine the leading market NYSE for both stocks, by using the methods of Hasbrouck (1995). Based on the information share analyse on the IBM stock, although the trading volume rose violently, the leading market was still clearly identifiable. When having a look at the Expedia analyse, we can not conclude that the determination of the market leader is more convincing at the Expedia stock, although that was lying in my expectations, because Expedia had less trading volume than IBM. However, the information shares of IBM fluctuate more heavily than the information shares of Expedia, which actually was lying in my expectations for the abovementioned reasoning. Unfortunatelly, I was not able to find some breakthroughs in terms of seasonality. I also showed that whether the day is before closure of the market or after closure of the market, it does not affect the Hasbrouck information shares. An intra-day analyse of the price discovering would be a nice topic for further research and an addition to my research if using the IBM data from 2002 till 2007.

Regrettably, I can not conclude that the increasing market efficienty is reflected in the price discovery. I sadly could not demonstrat by the comparison between IBM and Expedia the positive correlation between the number of trades and the degree of establishing a market leader. Furthermore, the method of Hasbrouck (1995) is not losing value over time, despite of the growth of the stock trading volume. So it is still possible to determine the market leader. These results do not agree with Kurov and Laser (2002), who claims that higher trade volume over the years makes the difference between core (NASDAQ, NYSE) and regional exchanges vanish. This should imply that trades initiated by exchange locals are not always more informative than off-exchange initiated trades.

A Appendix

- Table 6: Overview of the 10 exchanges IBM and Expedia are enlisted on
- Tabel 7: Hasbrouck Information Shares (annual) from 2002 - 2007
- Table 8: Hasbrouck Information Shares (seasonal) from 2002 - 2007

Table 6: Overview of the 10 exchanges IBM and Expedia are enlisted on

Code	**Description**
B	Boston Stock Exchange
C	National Cincinanati Stock Exchange
D	Direct Edge D Stock Exchange
I	International Stock Exchange
M	Chicago Stock Exchange
N	American Arka Exchange
P	Pacific Exchange
T	NASDAQ Stock Exchange
W	Chicago Board Option Exchange
X	Philadelphia Stock Exchange

IBM is listed on the markets B, C, D, I, M, N, P, T and X.

Expedia is listed on the markets B, C, D, P, T, W and X.

Table 7: Hasbrouck (1995) information shares, calculated on a daily basis (using a one-second frequency), averaged monthly for the whole sample period (January 2, 2002 till December 31, 2007)

300 lags

		2002			2003			2004		
		BTX	PN	REST	BTX	PN	REST	BTX	PN	REST
January	Lower	0,04549	0,14656	0,05070	0,00241	0,00268	0,00416	0,00069	0,00089	0,00678
	Upper	0,67714	0,73495	0,49924	0,43712	0,95287	0,89843	0,73346	0,91968	0,86252
February	Lower	0,07310	0,04117	0,01449	0,00978	0,02035	0,00917	0,00063	0,00143	0,00916
	Upper	0,77625	0,73642	0,55565	0,39854	0,88478	0,83185	0,67009	0,91461	0,76438
March	Lower	0,05653	0,05833	0,01152	0,04045	0,02626	0,00085	0,00018	0,00076	0,00792
	Upper	0,76397	0,73078	0,67003	0,71372	0,82811	0,70601	0,43621	0,98178	0,82216
April	Lower	0,00872	0,00799	0,00143	0,02157	0,00929	0,00155	0,00019	0,00043	0,02434
	Upper	0,80173	0,77082	0,61756	0,80326	0,78202	0,67217	0,40593	0,95577	0,79471
May	Lower	0,00806	0,00728	0,00576	0,04045	0,02626	0,00085	0,00058	0,00072	0,01929
	Upper	0,82235	0,73374	0,71235	0,71372	0,82811	0,70601	0,45514	0,95321	0,76039
June	Lower	0,00450	0,00298	0,00150	0,00022	0,00035	0,00436	0,00039	0,00041	0,02188
	Upper	0,78672	0,70859	0,67017	0,61771	0,98379	0,89192	0,40495	0,93348	0,76369
July	Lower	0,00332	0,00260	0,01051	0,00036	0,00045	0,00239	0,00018	0,00021	0,00268
	Upper	0,83954	0,74818	0,73317	0,64668	0,93920	0,86326	0,41536	0,99159	0,85360
August	Lower	0,00163	0,00281	0,01737	0,00181	0,00455	0,00793	0,00019	0,00022	0,02568
	Upper	0,73926	0,79850	0,78559	0,61958	0,89082	0,79715	0,38994	0,93017	0,78965
September	Lower	0,00250	0,00350	0,02723	0,00750	0,00931	0,00519	0,00013	0,00014	0,01398
	Upper	0,77956	0,81995	0,79869	0,61713	0,92986	0,71996	0,41183	0,96304	0,74270
Oktober	Lower	0,00230	0,00348	0,01286	0,00969	0,01195	0,00078	0,01577	0,01552	0,00190
	Upper	0,79000	0,82841	0,84856	0,68326	0,90785	0,75421	0,69959	0,93625	0,42342
November	Lower	0,00414	0,00884	0,01274	0,00749	0,00955	0,00085	0,01882	0,01267	0,00125
	Upper	0,75919	0,81740	0,77796	0,72739	0,84724	0,76794	0,54908	0,95213	0,58760
December	Lower	0,00047	0,00207	0,00248	0,00149	0,00466	0,00109	0,00779	0,03464	0,00100
	Upper	0,56987	0,94008	0,87829	0,60286	0,93255	0,85728	0,64141	0,95281	0,52203

		2005			2006			2007		
		BTX	PN	REST	BTX	PN	REST	BTX	PN	REST
January	Lower	0,00279	0,00252	0,00100	0,07358	0,08156	0,00110	0,02385	0,02495	0,02707
	Upper	0,47367	0,93501	0,78087	0,60198	0,88114	0,49860	0,72905	0,88419	0,54919
February	Lower	0,01168	0,01159	0,00034	0,05370	0,07552	0,00309	0,03050	0,03205	0,02519
	Upper	0,45916	0,93468	0,78028	0,68456	0,87130	0,55298	0,78175	0,86206	0,50793
March	Lower	0,00013	0,00064	0,00017	0,05855	0,08634	0,00136	0,02746	0,01827	0,01829
	Upper	0,42293	0,98467	0,72710	0,59190	0,89199	0,51890	0,73757	0,92195	0,47183
April	Lower	0,01464	0,01436	0,00030	0,03615	0,03556	0,00410	0,04154	0,03198	0,02259
	Upper	0,77739	0,96767	0,84669	0,46531	0,80945	0,72930	0,80758	0,91566	0,42381
May	Lower	0,00192	0,00523	0,00094	0,01300	0,01705	0,05196	0,01204	0,06669	0,00020
	Upper	0,68790	0,95830	0,73084	0,58655	0,83951	0,72387	0,87174	0,93763	0,46410
June	Lower	0,00766	0,01122	0,00021	0,00064	0,00690	0,08049	0,02194	0,04908	0,04545
	Upper	0,63815	0,98006	0,66878	0,59447	0,76975	0,62915	0,77655	0,91135	0,51850
July	Lower	0,04858	0,03628	0,00540	0,00532	0,02878	0,05118	0,02093	0,03281	0,01679
	Upper	0,60380	0,90915	0,50538	0,55688	0,88296	0,64361	0,83383	0,93774	0,57861
August	Lower	0,05473	0,05810	0,00503	0,00328	0,02405	0,07731	0,01542	0,04556	0,01107
	Upper	0,53538	0,90092	0,62124	0,55177	0,85732	0,66619	0,85275	0,96500	0,45401
September	Lower	0,03721	0,05608	0,00239	0,00279	0,02059	0,11863	0,00752	0,03020	0,01496
	Upper	0,51999	0,92974	0,55120	0,54577	0,83033	0,68740	0,88959	0,95984	0,33691
Oktober	Lower	0,04714	0,05415	0,00504	0,00046	0,01227	0,12586	0,00858	0,02644	0,03320
	Upper	0,56516	0,86076	0,58405	0,50447	0,75637	0,71867	0,85045	0,90948	0,56839
November	Lower	0,03674	0,04123	0,00138	0,00136	0,01692	0,10945	0,00802	0,02674	0,01765
	Upper	0,64822	0,92321	0,50350	0,51968	0,78173	0,71375	0,84941	0,89286	0,69093
December	Lower	0,01559	0,03636	0,00134	0,00184	0,02049	0,10382	0,00671	0,00902	0,07978
	Upper	0,50937	0,92323	0,58743	0,58093	0,79490	0,70692	0,86774	0,70377	0,70322

Table 8: Hasbrouck (1995) information shares, calculated on a daily basis (using a one-second frequency), averaged every month of all the years in the sample period

		MIDQUOTE	
		Lower	Upper
January	BTX	0,025	0,609
	PN	0,043	0,885
	REST	0,015	0,681
February	BTX	0,030	0,628
	PN	0,030	0,867
	REST	0,010	0,666
March	BTX	0,031	0,611
	PN	0,032	0,890
	REST	0,007	0,653
April	BTX	0,020	0,677
	PN	0,017	0,867
	REST	0,009	0,681
May	BTX	0,013	0,690
	PN	0,021	0,875
	REST	0,013	0,683
June	BTX	0,006	0,636
	PN	0,012	0,881
	REST	0,026	0,690
July	BTX	0,013	0,649
	PN	0,017	0,901
	REST	0,015	0,696
August	BTX	0,013	0,615
	PN	0,023	0,890
	REST	0,024	0,686
September	BTX	0,010	0,627
	PN	0,020	0,905
	REST	0,030	0,639
October	BTX	0,014	0,682
	PN	0,021	0,867
	REST	0,030	0,650
November	BTX	0,013	0,675
	PN	0,019	0,869
	REST	0,024	0,674
December	BTX	0,006	0,629
	PN	0,018	0,875
	REST	0,032	0,709

References

[1] Baillie, R.T., Booth, G.G., Tse, Y. and Zobotina, T. , 2002. *Price discovery and common factor models.* Journal of Financial Markets 5, pp. 309 - 322.

[2] Gonzalo, J., Granger, C.W.J., 1995. *Estimation of common long-memory components in cointegrated systems.* Journal of Business & Economics Statistics 13, pp. 27 - 36.

[3] Garbade, K.D. and Silber, W.L., 1979. *Structural Organization of Secondary Markets: Clearing Frequency, Dealer Activity and Liquidity Risk* The Journal of Finance, Vol. 34, No. 3, pp. 577 - 593.

[4] Goodhart, C.A.E. and O'Hara, M., 1997. *High frequency data in financial markets: Issues and applications.* Journal of Empirical Finance, Vol. 4, No. 23, pp. 73 114.

[5] Grammig, J., Peter, Franziska J., 2010. *International Price Discovery in Stock Markets - A Unique Intensity Based Information Share.*

[6] Heij, C., Boer, P.M.C. de, Franses, P.H.B.F., Kloek, T. and Dijk, H.K. van, 2004. *Econometric Methods with Applications in Business and Economics.* Oxford University Press.

[7] Harris, F.H.B., McInish, T.H. and Wood, R.A., 2002. *Security price adjustment across exchanges: an investigation of common factor components for Dow stocks.* Journal of Financial Markets 5, pp. 277 - 308.

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

[9] Hasbrouck, J., 1995. *One Security, Many Markets: Determining the Contributions to Price Discovery.* The Journal of Finance 50, pp. 1175 - 1199.

[10] Hasbrouck, J., 2002. *Stalking the 'efficient price' in market microstructure specifications: an overview.* Journal of Financial Markets, Vol. 5, No.3, pp. 329 339.

[11] De Jong, F. and Schotman, P.C., 2010. *Price Discovery in Fragmented Markets.* Journal of Financial Econometrics 8, pp. 1 - 28.

[12] Korenok, O., Mizrach, B., Sachs, G., and Radchenko, S., February 2010. *A Structural Approach To Information Shares.* Journal of Financial Markets, Vol. 13, No. 1, pp. 1 - 19.

[13] Kurov, A., Lasser, D.J., March 2002. *The effect of the introduction of Cubes on the Nasdaq-100 index spot-futures pricing relationship.* Journal of Futures Markets, Vol. 22, No. 3, pp. 197 - 218.

[14] Manaster, S. and Mann, S.C., 1996. *Life in the pits: competitive market making and inventory control.* Rev. Financial Studies, pp. 953 - 975.

[15] Mizrach, B. and Neely, C., 2008. *Information shares in the US Treasury market.* Journal of Banking and Finance, pp. 1221 - 1233.

[16] Shapiro, J.E., 1993. *Recent competitive developments in US equity markets* The Journal of Finance, Vol. 23, No. 1, pp. 201 - 234.

[17] Yan, B. and Zivot, E., February 2010. *A structural analysis of price discovery measures.* Journal of Financial Markets, Vol. 13, No. 1, pp. 1 - 19.