

# Commonalities in the order book

Helena Beltran

Pierre Giot

Joachim Grammig \*

26th January 2005

## Abstract

Recent contributions to microstructure theory hint at commonalities in the price-depth pairs which constitute the open limit order book. In this paper we provide empirical evidence that indeed a small number of latent factors, two for each side of the book, capture most of the variation in the price-depth pairs. The results also indicate that a heterogeneous trader population is active on the buy and sell sides. The respective latent factors explaining the buy and sell side variation exhibit specific dynamics. When we exploit results from microstructure theory to empirically assess whether the majority of the book variation is due to either informational effects or non-informational fluctuations of liquidity we obtain mixed results.

*Keywords:* limit order book, commonalities, liquidity, market microstructure

*JEL classification:* G10, C32

---

\*Helena Beltran is a FNRS research fellow at CORE, Université catholique de Louvain, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium. Phone: +32 (0) 10 478160. Email: beltran@core.ucl.ac.be. Pierre Giot is Professor of Finance at the Department of Business Administration & CEREFIM, University of Namur, Rempart de la Vierge, 8, 5000 Namur, Belgium and CORE at Université catholique de Louvain. Phone: +32 (0) 81 724887. Email: pierre.giot@fundp.ac.be. Joachim Grammig is Professor of Econometrics at the Faculty of Economics, University of Tübingen, Mohlstrasse 36, D-72074 Tübingen, Germany. Phone: +49 7071 29 76009. Fax: +49 29 5546. Email: joachim.grammig@uni-tuebingen.de. Send all correspondence to pierre.giot@fundp.ac.be. Work supported in part by the European Community's Human Potential Programme under contract HPRN-CT-2002-00232 Microstructure of Financial Markets in Europe. The first author gratefully acknowledges financial support from the Belgian 'Fonds National de la Recherche Scientifique' (FNRS). We are grateful to Deutsche Börse AG for providing access to the limit order data. We thank Michel Baes for helpful comments and Stefan Frey for his collaboration in the reconstruction of the limit order books. Uwe Schweickert offered invaluable expertise regarding the Xetra trading system. The usual disclaimer applies.

During the last decade, automated auction systems have gradually become the most widely used trading structure of equity markets. Nowadays, most exchanges are organized either as pure automated auction markets or hybrid trading systems combining features from automated auction and dealership markets (see e.g. Domowitz (1993) for a taxonomy of trading systems). Examples of the former include Euronext, the Frankfurt Stock Exchange, the Hong-Kong and the Australian stock exchange, and Electronic Crossing Networks such as Island or Instinet. Important market maker and dealer markets, among those the New York Stock Exchange (NYSE) and the London Stock Exchange, have recently adopted one of the key features of automated auction systems, namely the public display of the limit order book.<sup>1</sup> The order book consolidates the committed liquidity supply by displaying the sorted prices and volumes of non-executed limit orders.

Over the same period, academic research on automated auction markets has surged and many papers have investigated the characteristics of such markets. Most of these studies focus on the provision of liquidity and the interaction between market orders and limit orders (Biais, Hillion, and Spatt (1995), Ahn, Bae, and Chan (2001), Bae, Jang, and Park (2003)), the influence of volatility on the dynamics of the trading process (Beltran, Durré, and Giot (2004)), the trading costs of such structures (Degryse (1999), Venkatamaran (2001)), or the information content of the limit order book (Cao, Hansch, and Wang (2004), Irvine, Benston, and Kandel (2000) and Coppejans, Domowitz, and Madhavan (2002)). However, partially due to a lack of suitable data, only a few papers have conducted empirical analyses of the behavior of the limit order book beyond the best prices.

Recent developments in market microstructure theory hint at commonalities in the price-depth pairs displayed in the limit order book. According to Glosten (1994), limit order traders post orders whose characteristics (limit price and volume) depend on the underlying expected value of the stock and protect themselves against adverse price changes in the value of the security (the so-called picking-off risk). One could therefore argue that a single latent factor, the expected value of the stock, affects all price-volume pairs. The papers by Parlour (1998) and Foucault, Kadan, and Kandel (2003) suggest the existence of another latent variable explaining the variation of the order book. This can be motivated as follows. Orders are

---

<sup>1</sup>An important example is the NYSE open book program. Since January 2002, the NYSE has publicly displayed all limit orders

executed according to both time and price priorities. The limit order with the best price is executed first, and for equal prices the order submitted first has precedence. When the book is thick, limit order traders may have to wait for a long time before their order gets executed. If traders bear waiting costs, they may post limit orders closer to the best quote (or undercut it) to get a better position in the queue and shorten the time to execution of the order. Traders thus face a trade-off between the price improvement brought about by a limit order and the quick execution resulting of a market order or an aggressive limit order which jumps ahead of the order queue. Order submission trade-offs may therefore explain part of the order book variation.<sup>2</sup> These considerations suggest that two latent variables affect the state of the order book, one information-driven and one liquidity-driven. The papers cited above assume some degree of symmetry between the buy and sell side of the order book. One should therefore presume that informational and liquidity factors, if they exist, are identical for both sides of the book. However, if heterogenous traders with different liquidity needs and risk aversion place orders on both sides of the book, different factors may explain the variation of the buy and sell sides of the order book.

This paper aims to provide empirical evidence about these issues by addressing the following questions:

1. Do the price-depth pairs in a limit order book feature commonalities? In other words, can liquidity supply displayed in the limit order book be ‘summarized’ with a small number of latent factors?
2. Do both sides of the book feature the same commonalities or are different factors required to explain buy and sell side liquidity (heterogeneity of buy and sell limit order traders)?

---

<sup>2</sup>In Foucault, Kadan, and Kandel (2003), the choice between market and limit orders depends on the maximization of the gain from the trade. As impatient traders incur larger waiting costs, they are more likely to post market orders and thus demand liquidity, whereas patient traders incur smaller waiting costs and are thus ready to provide liquidity by means of limit orders. Consequently, the shape of the book depends first on the mix between patient and impatient traders, and second on the order arrival rate. When the proportion of patient traders increases, the likelihood of an execution decreases as these traders submit mainly limit orders. Therefore, when the proportion of patient traders increases, more aggressive limit orders, and thus smaller inside spreads and larger book spreads, are to be expected.

3. Can the extracted factors be related to the behavior of market participants? If yes, is this behavior compatible with predictions of market microstructure theory?
4. Can we decide whether the majority of the book variation is due to informational effects or non-information related fluctuations in liquidity?

In our empirical analysis, we use Frankfurt Stock Exchange (FSE) order book data from the Xetra trading system for all stocks constituting the DAX30 index.<sup>3</sup> Like the other important continental European trading system, Euronext, Xetra is an automated auction system with an open limit order book. No designated market makers exist for the most frequently traded stocks. In contrast to the NYSE or NASDAQ, the Xetra order book at the FSE faces no significant competitor in the hunt for liquidity. Although a floor trading system (also operated by the FSE) and some local exchanges offer alternatives to the Xetra book, these other trading venues only account for a small amount (2% to 5%) of the total trading activity. Some of the stocks are also cross-listed at foreign trading venues, like the NYSE (DaimlerChrysler is the most prominent example). However, compared to the home market the foreign trading volume is also small.

The data used in the paper come directly from the Xetra production system. They include all necessary information to re-build order books at any chosen time. The FSE is using the data for internal reporting purposes which ensures their high quality and enables us to correctly reconstruct the real time sequence order books for each of the thirty stocks during the three month period. The data present an unique opportunity to analyze the market participants' behavior in a framework close to the one described by theoretical papers focusing on order books, e.g. by Glosten (1994).

The original contribution of the paper is to search for commonalities in limit order book, first separately for each side of the book, then jointly. We employ principal components analysis (PCA) and interpret the extracted factors in light of recent contributions of market microstructure theory of order book markets. In a regression framework using explanatory variables suggested by microstructure theory we associate the extracted principal components with informational and liquidity effects. We also provide a detailed analysis of the sources of

---

<sup>3</sup>The DAX30 is the value weighted index of the thirty largest German stocks and one of the leading European stock indexes, along with the French CAC and the British FTSE. The Xetra system also operates at the Dublin and Vienna stock exchanges and at the European Energy Exchange

intra-day variation of liquidity supply. The PCA is conducted for the visible portion and for the complete order book.<sup>4</sup>

The main results can be summarized as follows. Two principal components already provide an explanatory power of more than 92% in a ‘one-sided’ analysis, i.e. when the PCA is conducted separately for the buy and the sell side of the order book. We explain that the effect of the first principal component is a shift of the supply and demand curves while the second principal component acts on the slope of the book. The distinction of the order book sides seems mandatory. We provide empirical evidence that the bid and ask sides of the book are driven by distinct, although related factors. When we use the PCA on both sides of the book, four factors are needed to achieve a similar explanatory power as in the ‘one-sided’ analysis. We interpret this as evidence that the order book market attracts heterogeneous population of limit order traders who need not be active on both sides of the market at the same time. Furthermore, we show that the visible and hidden portions of the limit order book share common dynamics although there are clear idiosyncracies. In accord with what has been documented for other exchanges, we conclude that the Xetra system is characterized by a relatively poor liquidity at the start of the trading day, which is consistent with the asymmetric information hypothesis. We show that this diurnality is captured by the first extracted factor which exhibits a pronounced L-shaped intra-day-seasonality matching the time-of-day volatility pattern. Averaged across stocks, the first principal component accounts for almost 80 % of the book variation while the second principal component explains on average about 15 % of the book variation. When we exploit results from microstructure theory to empirically assess whether the majority of the book variation (accounted for by the first principal component) is due to informational effects or non-informational fluctuations of liquidity we obtain somewhat ambiguous results. Time series evidence suggests that both first and second principal component are associated with informational effects as well as non-informational fluctuations of liquidity, while cross sectional regressions indicate that the majority of the book variation is due to non-informational liquidity processes.

The remainder of the paper is organized as follows. Section I outlines the market structure and provides a descriptive analysis of the data. In section II we review the empirical methodology. Section III reports the results of the principal components analysis applied to

---

<sup>4</sup>Similar to Euronext, the Xetra trading rules allow to hide part of a limit order

the reconstructed order books. Section IV sheds light on the interpretation of the extracted factors from a market microstructure point of view. We conclude in section V.

## I. Market structure and liquidity measures

### A. The Xetra open order book

Xetra is an open order book system developed and maintained by the German Stock Exchange.<sup>5</sup> It has operated since 1997 as the main trading platform for German blue chip stocks at the Frankfurt Stock Exchange. Between an opening and a closing call auction - and interrupted by another mid-day call auction - trading is based on a continuous double auction mechanism with automatic matching of orders based on the usual rules of price and time priority. During pre- and post-trading hours it is possible to enter, revise and cancel orders, but order executions are not conducted, even if possible. Trading hours extend from 9 a.m. C.E.T to 5.30 p.m. C.E.T. For the DAX30 stocks there are no designated market makers. Traders have access to the full order book, except hidden shares coming from so-called iceberg orders. An iceberg or ‘hidden’ order is similar to a limit order in that it has pre-specified limit price and volume. The difference is that a portion of the volume is kept hidden from the other traders that is not visible in the open book. The visible portion of the iceberg order, called the ‘peak’, enjoys full price and time priorities (as any limit order). The hidden portion only gets price priority. All disclosed volumes are executed first, even if those volumes entered the book after the iceberg order submission. When a market order hits the hidden portion, a new portion of the iceberg order (equal to the peak size) is revealed to the market participants and is granted time priority over subsequent order submissions. Consequently, a trader submitting a market order may receive an unexpected price improvement if her market order is executed against a hidden order. Iceberg orders are convenient for some market participants as they provide a tool to hide investment decisions. Whether those decisions are liquidity-motivated or informed remains an open question. In the realm of limit order book markets, the Xetra system is thus very close to the Euronext trading system (at the time of this writing, the Euronext trading system encompasses the Amsterdam, Brussels, Lisbon and Paris stock exchanges), see e.g. Biais, Hillion, and Spatt (1995). Xetra is also very close to

---

<sup>5</sup>See Deutsche Börse AG (1999) for a detailed description of the Xetra system.

the trading system used at the Hong Kong stock exchange, which has been recently studied by Ahn and Cheung (1999), Brockman and Chung (1999) or Ahn, Bae, and Chan (2001). A Xetra market design feature which makes the data for the purpose of our study preferable is that market orders exceeding the volume at the best quote are allowed to ‘walk up the book’. In other words, market orders are guaranteed immediate full execution, at the cost of incurring a higher price impact on the trades. The liquidity measure used in the present paper and discussed in the next subsection implicitly assumes that a ‘walk up the book’ is possible.

For the purpose of this paper the German Stock Exchange granted access to a database containing complete information about Xetra order book events (entries, cancellations, revisions, expirations, partial-fills and full-fills of market, limit, and iceberg orders) that occurred during the first three months of 2004 (January, 2nd - March, 31st). The data encompasses the 30 stocks belonging to the DAX30 index. Based on the event histories we perform a real time reconstruction of the order book sequences. Starting from an initial state of the order book, we track each change in the order book implied by entry, partial or full fill, cancellation and expiration of market, limit, and iceberg orders. This is done by implementing the rules of the Xetra trading protocol in a GAUSS program. Although running an exhaustive battery of consistency checks no errors occur during the reconstruction process. From the resulting real-time sequences of order books, snapshots at 5-minute intervals during the continuous trading hours were taken (102 observations per trading day). For each snapshot, the order book entries were sorted on the bid (ask) side in price descending (price ascending) order. As iceberg orders are not fully disclosed to market’s participants, we reconstruct, at each point in time, two limit order books. First, we have the visible order book, which contains all limit orders as well as the visible portion of the iceberg orders. Second, we study the complete order book that features all orders, including the hidden portions.

Table I presents some characteristics of the DAX30 stocks. Liquidity supply for these stocks is indeed quite active, as on average 13,000 (11,000) non-marketable limit orders per stock are submitted (cancelled) each day. Implicit transaction costs are relatively small with relative spreads ranging from 0.04 to 0.14%. On average, 2,100 trades per stock are executed each day and on average 15.2% of these walk up the book (i.e. they are matched by standing limit orders beyond the best bid and ask prices). We refer to those events as ‘aggressive trades’.

## B. Measuring liquidity beyond the best quotes: the price impacts

Automated auction markets enforce price and time priority rules which govern the trading process. This implies that the price impact of a buy (sell) market order is an increasing (decreasing) function of the trade size. The available pre-trade liquidity of the book can be characterized by the unit price for selling  $v$  shares at time  $t$ :

$$b_t(v) = \frac{\sum_k b_{k,t} v_{k,t}}{v} \quad (1)$$

where  $v$  is the volume executed at  $k$  different unique bid prices  $b_{k,t}$  with corresponding volumes  $v_{k,t}$  standing in the limit order book at time  $t$ . The unit price  $a_t(v)$  of a buy of size  $v$  at time  $t$  can be computed analogously. The unit prices  $b_t(v)$  and  $a_t(v)$  can be computed for an arbitrary wide range of possible volumes  $v$ . The liquidity supply in an order book can thus be characterized at any given point at time  $t$  by a grid of unit ask and bid prices conditional on the traded volume  $v$ . The bi-dimensionality of liquidity (price and volume) calls for measures such as the unit price to capture liquidity provision.<sup>6</sup>

Unit price series are not stationary. As PCA requires stationary data input we run the analysis after an appropriate data transformation which produces stationary ‘price impacts’. Price impacts capture the premium paid by traders when the transaction is executed against standing limit orders beyond the best quotes. Bid price impacts are defined as

$$bp_t(v) = \frac{b_t(1) - b_t(v)}{b_t(1)} \cdot 100. \quad (2)$$

An equivalent formula is used to obtain the ask price impact,  $ap_t(v)$ . To ensure comparability between the thirty DAX stocks, we do not express the trade size  $v$  in number of shares per trade, but use a grid of trade sizes with  $v$  expressed in euros.

By construction, the larger the transaction (i.e. the larger  $v$ ), the higher the price paid as the market order hits more and more limit orders and is likely to walk up further in the book. A 10% bid price impact for a trade size of, say, 100,000 euros means that the price the seller receives for each share sold is 10% lower than the price she would get by trading quantities smaller than the depth at the best price. When  $v$  does not consume all the share available at the best price, the price impact is thus zero. Of course, in a perfectly liquid market,  $ap_t$  and

---

<sup>6</sup>see Irvine, Benston, and Kandel (2000) and Gomber, Schweickert, and Theissen (2004) for a similar measure of book liquidity.



$bp_t$  do not depend on  $v$  and are equal to zero. The gap between  $ap_t(v)$  and zero (respectively zero and  $bp_t(v)$ ) thus measures the pre-trade available committed liquidity. By focusing on price impacts, we restrict our analysis to the modelling of the slope of the order book, and we do not consider the variation in unit ask and bid prices expressed as nominal amounts. This implies that the econometric modelling can be split into three separate parts, a model for the midquote dynamics (e.g. modelling the marginal utility weighted price of the security as a martingale process), a model for the inside spread (allowing for time-of-day effects and serial persistence), and a model for the ‘shape’ of the order book reflected by the price impacts. This latter is the focus of the present paper.

We compute the price impacts at both sides of the book for a grid of trade sizes  $v$  ranging from 50,000 euros to two million euros, with an increment of 50,000 euros. This produces a total of 40 price impacts for both the ask and bid sides. Although marketable limit orders of 2 million euros are very rare, this grid neatly summarizes the book and includes unit bid and ask prices quite far from the best prices.<sup>7</sup> During the reconstruction process we store two order book sequences. While both feature the same grid of volumes, the first solely contains the shares visible to traders (the ‘visible’ order book), whereas the second also includes the hidden shares (the ‘complete order book’). Consequently, three types of price impacts can be computed. We refer to *visible* price impacts as those that results from an application of Equation (1) using only the visible book depth, while *complete* price impacts are computed using the complete book, including hidden volumes. The latter are by definition always smaller or equal (if there is no hidden volume) than the former. Third, we compute the price improvements brought about by the hidden orders as the difference between the price impact implied by the complete order book and the price impact implied by the visible order book.

### C. Price impacts: descriptive analysis

A first important empirical result is that the time series properties of the price impacts are remarkably similar for the 30 stocks in our sample. Hence we shall present our results as holding ‘on average for all stocks’. Whenever suitable we present averages across stocks and also the confidence bounds indicating the distribution across stocks. We will also illustrate

---

<sup>7</sup>On average for the stocks in our sample, 20 different price levels are necessary to trade 2 million euros. The results presented in the paper are quite robust to changes in the trade size grid.

the analysis using the results on a single stock. For no particular reason, except that it can be considered a ‘typical’ stock, we chose Bayer (ticker symbol BAY), a company specializing in life sciences. Figure 1 presents an example of timepaths for 50,000, 250,000, 1 million and 2 million euros ask price impacts for Bayer. The price impacts were computed using the complete order book. The time series plots are similar for the bid side and for the price impacts computed with the visible book only. As indicated in Figure 1, small volume price impacts are highly volatile and exhibit little persistence. Price impacts for large volumes are much more correlated as evidenced in the bottom right panel of each figure.

These patterns suggest that the available liquidity is far from being ‘static’. This is not surprising, given the evidence provided in earlier papers such as e.g. Ahn, Bae, and Chan (2001) and Biais, Hillion, and Spatt (1995) for the Hong-Kong and Paris Bourse stock exchanges, respectively. Figure 2, which gives the autocorrelations, tells the same story. For relatively small volumes, the price impacts exhibit only small serial correlation, while the 2 million euros price impacts are clearly autocorrelated. All series exhibit a strong intra-day seasonality, although this pattern is less pronounced for the 50,000 euros price impact.

We use spectrum analysis to decompose the variance of a series into different frequency bands. The area below the spectrum is the variance of the series, and the area between two frequencies corresponds to the portion of the variance that is due to this frequency band. If the area between two frequencies is large, the frequency band explains a large fraction of the variance. Low frequencies (close to 0) correspond to long-term evolutions, and high frequencies (close to  $\pi$ ) to intra-day variations. Figure 3 displays the spectrum for the four price impacts analyzed before. These patterns are in agreement with our previous results: the larger the hypothetical trade size, the more important the long-run frequencies. The spectrum of the smallest trade size (50,000 euros) is flat, i.e. quite close to the spectrum of a white-noise process. Table II decomposes the variance according to selected frequency bands. The table shows that, for all price impacts, 80% of the variance is due to intra-day variations. When we consider the smallest trade size, 76.5% of the variance is governed by one-hour dynamics. This percentage decreases to 24.5% for the largest trade size. Weekly variations (excluding intra-day variations) account for 19.3% of the variance far in the book, and only 3.2% close to the best prices. To summarize, price impacts for small trade sizes exhibit a low autocorrelation and are highly volatile; price impacts for large trade sizes are

strongly correlated and persistent. Persistence implies that the traders' standing positions far in the book are less frequently updated. On the other hand, the lack of autocorrelation for the small trade size price impacts shows that traders cancel/(re)submit orders very actively close to the best quotes. Thus the state of the book close to the best quotes is hardly correlated with its own state 5 minutes before. This also implies that trading opportunities such as abnormal depths at the best quotes are probably short-lived. Finally, the cross-correlations functions (Table III) show that, as conjectured in this paper, price impacts for different trade sizes display some common dynamics. We delve more deeply in those issues in the next two sections of the paper.

## II. Modeling the order book using principal components analysis

The main objective of the paper is to uncover and study commonalities in the limit order book. Several statistical methods are at hand for this task. In this paper we employ principal components analysis, or PCA in short.<sup>8</sup> In the following we will briefly sketch the application of the methodology in the context of order book modelling. The illustration focusses on the ask side of the book, the methodology for the bid side and the joint analysis of both sides is analogous.

The data input for the PCA are the relative price impacts  $ap_t(v)$ , computed according to Equations (1) and (2). We compute the price impacts for a grid of  $N$  trading volumes and for a sequence of five-minute order book snapshots during the three-month sample period. The price impacts are collected in a  $T \times N$  matrix  $X$  such that each column of  $X$  contains a time series sequence of price impacts computed for a given volume  $v$ . Columns are sorted by increasing  $v$ . The time  $t$  grid-point  $i$  element of the price impact is denoted by  $x_{t,i}$ . The index  $t = 1, \dots, T$  is used to identify the time of the order book snapshots, and the index is used  $i = 1, \dots, N$  for the volume grid. Since PCA requires standardized stationary data input (otherwise the variables with largest sample variance tend to receive too much weight)

---

<sup>8</sup>Gouriéroux, Le Fol, and Meyer (1998) present the first application of PCA to order book data of French Euronext-traded stocks. Other applications of PCA in finance are discussed in Alexander (2001).

we compute standardized price impacts,

$$z_{t,i} = (x_{t,i} - x_i) / sd x_i, \quad (3)$$

where  $x_i$  and  $sd x_i$  denote sample mean and standard deviation of column  $i$ , respectively. The standardized price impacts are stored in a  $T \times N$  matrix  $Z$ .

The basic idea of applying PCA in the context of this paper is to express the standardized price impacts as a linear combination of  $N$  orthogonal principal components,

$$Z = P W', \quad (4)$$

where  $P = (F1, F2, \dots, FN)$  is a  $T \times N$  matrix containing the orthogonal principal components and  $W$  is a  $N \times N$  matrix of weights. The  $i$ -th ( $i=1, \dots, N$ ) column of  $P$  will be referred to as the  $i$ -th principal component (or simply ‘factor’) and denoted  $Fi = (Fi_1, Fi_2, \dots, Fi_T)'$ .<sup>9</sup> The extraction of orthogonal principal components is achieved by computing the eigenvalues and associated eigenvectors of the sample correlation matrix of standardized price impacts,  $\frac{1}{T} Z' Z$ . The columns of the weight matrix  $W$  containing these eigenvectors are arranged in descending order of the associated eigenvalues. Having computed and ordered the eigenvectors in the described way, the principal components can be computed as  $P = Z W$ . The weights/eigenvectors in  $W$  provide the mapping between the principal components and the price impact series that are the inputs of the analysis. Below we will show how an analysis of the weight matrix  $W$  helps to understand how the principal components ‘explain’ the variation of the price impacts. The ordering by largest eigenvalue implies that the first principal component,  $F1$ , will be the one with the largest variance that explains most of the variation in the price impacts. The second principal component,  $F2$ , is the one with the second largest variance and explanatory power and so forth. The explanatory power (referred to as cumulative  $R^2$ ) of the first  $n$  principal components  $F1, F2, \dots, Fn$  can be computed by dividing the sum of the  $n$  first eigenvalues by the total sum of all  $N$  eigenvalues.

If the price impact series share common components (commonalities), then the first  $n \ll N$  principal components will have a considerable explanatory power. The first  $n$  principal components, collected in the  $T \times n$  matrix  $P^{(n)} = (F1, F2, \dots, Fn)$ , combined with weight

---

<sup>9</sup>When we distinguish the ask and the bid side of the book we will denote the principal components as  $FAi$  and  $FBi$ , respectively

matrix  $W^{(n)}$ , containing the first  $n$  columns and rows of  $W$ , will then contain sufficient information to ‘rebuild’ the whole state of the order book at time  $t$ . For this purpose we compute  $\widehat{Z} = P^{(n)}W^{(n)}$ , and add back to the elements of  $\widehat{Z}$  the standard deviation and mean of each column using the inverse transformation of Equation (3). This produces  $\widehat{X}$ , the reconstructed price impacts matrix using only the information provided by the first  $n$  principal components. If the price impacts feature important commonalities the reconstructed price impact series will be highly correlated with the observed price impacts. Provided that  $n$  is small, the PCA can deliver a formidable reduction of complexity. The variation of the order book can then be explained by a small number of orthogonal factors.

### III. Order book commonalities

Separate PCAs are conducted to analyze commonalities in visible, complete and hidden price impacts. We can immediately point out that the weight structure and the explanatory power of the first four principal components are remarkably similar for the 30 stocks in the sample.<sup>10</sup> The results discussed thereafter should be viewed as being meaningful for all stocks. In the same vein, the numbers in tables and figures that we refer to are averages across the thirty stocks. Whenever useful, we also report standard deviation, minimum and maximum as well as confidence bounds computed from the cross section of stocks.

#### A. One-sided analysis of commonalities in the order book

We start with the visible order book data and run the PCA on the price impact series using the grid of forty volumes from 50,000 to 2 million euros. The results given in Table IV show that the first two principal components explain more than 92% of the variation in visible price impacts (cumulative  $R^2$ ). The estimation results lead us to select  $n = 2$  principal components. The choice of two factors stems from the low explanatory power of the third principal component (less than 5%). There are only minor differences between the bid and ask sides. Figure 4 shows the weights of the first two principal components for the bid and the ask side. The weights of the first principal component (F1) are all positive and approximately equal for volumes ranging from 500,000 to 2,000,000 euros. Weights decrease by about 50

---

<sup>10</sup>Note that this does not imply that the order book of the 30 stocks are generated by the same factors. Indeed, the factors for the 30 stocks are different, although they may share some common dynamics.

% when  $v$  decreases from 500,000 euros to 50,000 euros. An increase in the first principal component would therefore induce an upward shift in the whole structure of the price impacts, with price impacts for volumes smaller than 500,000 euros being less affected. This result is illustrated in the top panel of Figure 5 where we perform simulations on the first principal component using the Bayer ask side order book as an example. Starting from a base case, we multiply the value of the first principal component, F1, by .25, 0.5, 2 and 4, respectively, compute the predicted change in the whole range of price impacts and plot all five cases. The figure shows how the whole ask side of the order book shifts upwards (downwards) when F1 is increased (decreased) and how the price impacts of large volumes  $v$  are more affected than the price impact of small volumes. Hereinafter, the first factor will be referred to as ‘shift factor’ since it moves all price impacts move in the same direction (albeit in a non-linear fashion).

The weights of the second principal component, F2, are negative for volumes up to about 900,000 euro and positive for larger  $v$ . An increase in F2 will thus reduce the price impacts for smaller volumes, and increase those for large volumes. This effect is illustrated in the bottom panel of Figure 5 where again the ask side of the Bayer order book is used as an example. As in the simulation invoking the first principal component we start from a base case, increase and decrease the value of the second factor and compute the resulting price impacts. Figure 5 shows that as F2 is increased, the whole ask side of the book rotates counterclockwise, around a volume of about 900,000 euros. It is therefore natural to refer to the second principal component as a ‘rotation factor’.

Figure 4 also shows that the weight pattern is remarkably similar for both sides of the book. We come back to this issue below when we investigate whether the ask and bid sides are subject to the same dynamics. Results for the PCA on the complete order book are presented in Table IV. The reported figures for the explanatory power and the pattern of the weight structures (not given) are extremely close to what we find for the visible order book.

## **B. Joint analysis of the ask and bid sides**

In this section we analyze the commonalities of the ask and bid sides of the order book simultaneously. This is tantamount to running the PCA on a dataset made up of the juxtaposition of the ask and bid side price impacts. The ‘one-sided’ analysis in the previous subsection

showed that the first two principal components (shift and rotation factor) had a high explanatory power for both sides of the order book and that the weight structures are quite similar, too. One could therefore expect that we would also need only two factors, one for shift and one for rotation of the price impacts, when running a joint PCA on the both sides of the book. However, the empirical results reported below contradict this conjecture and point to asymmetries in the order book.

The bottom panel of table IV gives the explanatory power of the first four principal components in the joint PCA. Averaged across stocks the cumulative  $R^2$  of the first two principal components is only 0.79, a considerable decrease compared to 0.92 in the one-sided analysis. The explanatory power of the three first principal components amounts to 0.87 and with four principal components we reach 0.93. In other words, four principal components are needed to achieve a explanatory power comparable to the two factors in the one-sided analysis. We conclude that the variation of the bid and ask side price impacts cannot be explained by the same factors. Unlike for the one-sided analysis, the interpretation of the four ‘joint’ principal components is not clear-cut and warrants some further investigation. Table IV shows that the explanatory power of the first two joint factors is quite similar to the explanatory power of the first principal component (shift factor) from the one-sided analysis (around 0.7). This suggests that the one-sided shift factor captures the same dynamics as the first two principal components from the joint PCA. If this bears out, then a bid or ask price impact series reconstructed from the shift factor should be closely correlated with a price impact series reconstructed from the first two joint principal components. An analogous reasoning can be applied with respect to the one-sided rotation factor and the third and fourth principal component from the joint PCA. The empirical results support these conjectures. Indeed, the correlations between the price impacts reconstructed from the shift factor and the price impacts reconstructed from the first two joint factors are larger than 0.98 for all price impacts and all stocks, both for the bid or the ask side. This shows that the first two joint factors from the joint PCA capture the same order book variation as the shift factors from the one-sided analysis. Analogous results are achieved regarding the relation of the one-sided rotation factor with the third and fourth principal component.

Further evidence on the asymmetries between bid and ask side of the book is provided in Table III where we report the correlations between the original price impacts at each side of

the book for different trade size. All correlations between the bid and ask side price impacts of the (visible) book are smaller than 30%. Furthermore, the correlations between shift factors (buy and sell side) and rotation factors (also buy and sell side) obtained from the one-sided analyses are close to zero. An exception is the significant correlation between the ask shift factor and bid shift factor (average across stocks: 0.22).

Our results regarding the asymmetries of the buy and sell sides of the order book are not quite surprising if one takes into account the fact that in most limit order markets there is no binding reason for limit order traders to be active on both sides of the market at all times. A open limit order market is in this respect quite different from a specialist market, where the market maker has the obligation to supply liquidity for both sides of the market to maintain an orderly market.<sup>11</sup> By contrast, liquidity supply in an automated auction market like Xetra is entirely voluntary. The results indicate that this trading design attracts a multitude of heterogenous agents, who need not be active on both sides of the book.<sup>12</sup>

### C. Commonalities between the visible book and hidden liquidity

A distinct feature of our data is that it allows a complete view of the visible and hidden liquidity. As such, the next step in our analysis aims at characterizing the commonalities between the visible and hidden liquidity of the book. For this purpose we conduct a PCA on the dataset made up of the visible and the hidden ask price impacts. The hidden price impacts capture the price improvements brought about by iceberg orders. They are computed as the difference of the price impacts computed from the complete and the visible book. The analysis is performed one-sided, i.e. separately for the buy and the sell side.

Table IV shows that four principal components are needed to achieve an explanatory power comparable to the one-sided analysis using visible impacts. Whilst in the latter application two principal components (identified as shift and rotation factor) achieved a cumulative  $R^2$  of about 0.92, the cumulative explanatory power of the first two principal components in the joint visible-hidden PCA is only .77 (average across stocks). We have to include also the

---

<sup>11</sup>For example, the classical inventory model of Stoll (1978) suggests rather symmetrical ask and bid side variations. In the inventory model, the dealer shifts his bid and ask prices simultaneously to manage his inventory position.

<sup>12</sup>This is also pointed out in Danielsson and Payne (2001) for the Reuters D2000-2 order book trading system)



third and fourth principal component to increase the cumulative  $R^2$  to 0.93. Conducting a separate PCA for the hidden price impacts only, the results are qualitatively similar to the results on the visible book.<sup>13</sup> The first two principal components explain more than 90% of the variation of the hidden price impacts. Furthermore, the interpretation of the first principal component as ‘shift factor’ and the second as ‘rotation factor’ also applies. Similar to the previous section, the variation of the hidden part of the book is explained by different factors than the visible part. The asymmetries between the buy and the sell sides of the book thus extend to the provision of visible and hidden liquidity, again indicating the presence of a heterogeneous population of limit order traders.

However, the visible and hidden portions on the same side of the book exhibit more commonalities than the (visible) buy and sell sides. Indeed, the correlation between the hidden book shift factor (i.e. the first principal component extracted from the hidden price impacts) and the shift factor extracted from the visible book is about 0.3. The correlation between visible and hidden rotation factors is of the same size. Cross-correlations between shift and rotation factors are smaller, although non negligible. They amount to about .09 (visible shift factor and hidden rotation factor) and 0.04 (visible rotation factor and hidden shift factor), respectively.

## IV. Interpreting the factors in light of market microstructure theory

### A. Diurnal variation of the factors

The diurnal or intra-day variation of liquidity variables such as the inside spread or the depth has been actively studied in the empirical literature ever since intra-day data of trades and quotes were made available by the exchanges. Regarding the spread, the evidence seems to indicate that it follows either a ‘L’ or ‘U’ pattern, depending on the trading protocol of the respective exchange. At the New York Stock Exchange for example, both volume and spread are much higher shortly after the opening at 9h30 and before the close at 16h than during the trading day. Brock and Kleidon (1992) focus on the quasi-monopolistic behavior of the specialist at these times to explain this U-shaped pattern. More precisely, because there is an

---

<sup>13</sup>We do not report the detailed results of this analysis. They are available upon request.

increased and less elastic demand to trade near the open and close, the monopolistic specialist can charge a higher rent at these times, i.e. increase his quoted spread. Using a database indicating the origin of the quotes (specialist or limit order), Chung, Van Ness, and Van Ness (1999) show that most quotes originate from limit orders and that the U-shaped pattern in observed spreads can be largely explained by the intra-day pattern of order placement and execution of limit orders. Specialists quote more actively at the start of the day, i.e. providing liquidity when needed. Furthermore, the spread of the specialist is higher than the spread of limit orders and is also considerably larger at the start of the day. This is compatible with adverse selection models of the spread. The mechanism of competing market makers at the NASDAQ has been studied by Chan, Christie, and Schultz (1995), who show that it gives rise to a different, L-shaped intraday pattern. They argue that a decreasing inside spread at the end of the day is observed because of the competition between market makers. For the Hong Kong stock exchange, Brockman and Chung (1998) and Ahn and Cheung (1999) study the intraday pattern of the spread, while Brockman and Chung (1999) focus on the diurnal variation in the inside depth. These studies show that both the spread and the inside depth follow a U-shaped pattern, quite similar to what is known for the NYSE. They interpret the higher spread and poorer depth at the open as evidence in favor of the asymmetric information hypothesis. However, they would have expected increasing liquidity at the end of the trading day as the order book involves many market participants who are keen on balancing out their positions, i.e. an end-of-day behavior similar to what is known for the NASDAQ.

Figure 6 plots the diurnal variation of the first two principal components (one-sided, visible order book). It can be seen that the first principal component (shift factor) exhibits intra-day seasonality while the second principal component (rotation factor) does not. The diurnal variation of the shift factor is distinctively L-shaped. Combined with the weight structure attached to the first principal component (all weights are positive and increase with the traded amount), this implies that the book liquidity is poor at the start of the day and improves thereafter and stays at a constant level throughout the day, including near the close of trading. These results are consistent with an asymmetric information hypothesis for the start of the trading day (traders are cautious at those times and fear informed trading). Madhavan (1992) considers a model in which information asymmetry is gradually resolved throughout the trading day implying higher spreads at the opening. In the well-known Foster

and Viswanathan (1994) model, competition between informed traders leads to high return volatility and spreads at the start of trading. In contrast to what has been reported for the Hong Kong stock exchange, our results hint at inventory management at the close of the trading day, which would explain the good book depth at that time.

## B. Factors and volatility

New information triggers stock price changes. Thus, a factor that captures expected changes in the value of the security may increase volatility. On the other hand, large depths and thus small price impacts make aggressive trades less likely. Transitory price movements due to orders walking up the book should be less frequent when the book is thick (see Ahn, Bae, and Chan (2001) for a discussion of this issue at the Hong-Kong stock exchange). Therefore, a factor that captures increased liquidity provision close to the best quotes should decrease volatility.

To interpret the shift and rotation factors, we regress a volatility measure which is computed as the sum of five successive one-minute squared midquote returns, denoted  $\epsilon_t^2$ , on the first two principal components extracted from the respective one-sided PCAs:

$$\epsilon_t^2 = \alpha + \beta_a FA1_{t-1} + \gamma_a FA2_{t-1} + \beta_b FB1_{t-1} + \gamma_b FB2_{t-1} + u_t \quad (5)$$

$FA1$  and  $FA2$  ( $FB1$  and  $FB2$ ) stand for the shift and rotation factor for the ask (bid) side respectively. Table V reports the estimation results. Coefficient  $\beta_a$  is positive and significant for 29 out of 30 stocks;  $\beta_b$  is also positive and significant for 28 out of 30 stocks. Accordingly, when the shift factor increases (that is, when liquidity decreases throughout the book), volatility increases. This is consistent with a book that has been depleted because traders anticipate incoming information, e.g. through large, aggressive trades. On the other hand, the  $\gamma$  coefficients are negative and less significant. Moreover, the amplitude of the impact is smaller (the  $\beta$ 's are larger than the  $\gamma$ 's in absolute terms for 27 out of 30 stocks). When the book slope changes because of changes in the rotation factor, it has a small or insignificant impact on volatility. This is consistent with a rotation factor that is not linked to information arrivals. On the other hand, the shift factors do forecast squared price changes which means that the order book contains information regarding future volatility, since the factors reconstruct most of the book slope.

### C. Factors and trader behaviour

In the one-sided PCA we have identified a shift and a rotation factor which explain a considerable part of the order book variation beyond the best quotes. We have seen that an increase of the shift factor decreases the book liquidity at all price levels, while a downward movement of the rotation factor decreases liquidity close to the best quotes and increases liquidity deeper in the book. In this section we study the effect of predetermined variables suggested by microstructure theory on the shift and rotation factor. Our objective is to associate the two principal components with informational and liquidity effects. For this purpose we estimate regressions where the extracted shift and rotation factors serve as dependent variables and indicators suggested by microstructure theory as regressors. By construction, variables with a significant and positive coefficient when regressed on the shift factor increase the price impacts at all trade sizes  $v$  (decrease liquidity throughout the book). Variables with a significant and negative impact on the rotation factor cause a drain of liquidity close to the inside market, while decreasing price impacts for larger  $v$ . Using the principal components instead of the price impacts (or other measures of book liquidity) as dependent variables thus allows us to disentangle shifts in liquidity from slope rotations.

In the following, we discuss the selection of meaningful regressors rooted in microstructure theory. There are few papers with testable results regarding book depth. In Glosten (1994), the marginal price paid (received) by a buy (sell) market order trader provides information about the true value of the asset. Accordingly, trade size is informative as large trades are more likely to walk up the book. In a similar setting, Seppi (1997) shows that large trades, if informative, decrease depths in the book, while non-informative large trades attract limit orders and increase liquidity. In the regression framework this has the following implications. If large trades increase the shift factor then this would provide evidence that the first principal component is associated with informational effects. On the other hand, if the effect of large trades on the rotation factor is negative, then it is the second principal component that would be associated with informational effects. Recall that a decrease in the rotation factor implies that price impacts close to the inside market increase while they decrease deeper into the book. However, if the effect of large trades on the shift (rotation) factor is negative (positive), then this would be evidence for the interpretation that the respective factors are associated with liquidity effects.

Seppi (1997) stresses the importance of the disequilibrium between buyers and sellers in that a larger proportion of buy than sell trades increase the likelihood of execution for sell limit orders.<sup>14</sup> For example, more buy than sell trades should exert a positive impact on a liquidity factor explaining the variation of the sell side of the book. If the rotation factor qualifies as a liquidity factor then we would expect a positive effect of a properly constructed order imbalance indicator. If the shift factor qualifies as a liquidity factor then we would expect a negative effect of order imbalances. The same reasoning holds for the buy side of the book.

A second strand of the literature suggests that the factors should be influenced by liquidity-driven variables. Foucault, Kadan, and Kandel (2003) put forward a framework that does not rely on (informationally driven) changes in the value of the stock. They focus on liquidity effects to explain the shape of the book. In their framework, limit orders are more aggressive (and liquidity high and price impacts close to the inside market low) when either the proportion of patient traders increases (assuming a constant rate of order submission) or when more orders are submitted (assuming a constant proportion of patient traders). This motivates the inclusion of three additional variables which should explain the behavior of a liquidity factor, the number of submissions, the percentage of limit orders among all submissions (a proxy for traders' impatience assuming that only impatient traders submit market orders), and an interaction variable equal to the product of the two first variables. If the effect of these variables on the rotation factor is positive, then this would be an indication that this principal component captures liquidity effects.

We also consider volatility measures as explanatory variables for the principal components. Like for trade size, the impact of volatility on book depth is ambiguous. Ahn, Bae, and Chan (2001) point out that transitory (non-informational) variations in the best prices improve liquidity as limit order executions become more likely which attracts aggressive limit orders. However, if volatility is associated with informational effects we would expect a negative effect on liquidity as picking-off risk increases. Similar to the effect of trade size discussed above, sign and significance of the parameter estimates in the regression of the shift and rotation factor on volatility measure can provide further evidence regarding the interpretation of the first two principal components. We distinguish bid and ask side realized volatility to

---

<sup>14</sup>Chordia, Roll, and Subrahmanyam (2002) also stress the role of order imbalances

capture possible asymmetric sources of risk on each side of the market. The bid (ask) realized volatility are computed as the sum, over 5-minute intervals, of 1-minute squared bid (ask) price returns. Finally, Foucault (1999) suggest that the size of the inside spread influences the order submission process and thus the liquidity provision.

These considerations lead us to specify the following regression equations. For the ask side we have

$$\begin{aligned}
FAi_t &= \beta_0 + \alpha FAi_{t-1} + \beta_1 nbtr\_agg\_bid_t + \beta_2 trade\_size\_bid_t + \beta_3 vola\_bid_t + \beta_4 vola\_ask_t \\
&+ \beta_5 nbsub\_bid_t + \beta_6 per\_lo\_bid_t + \beta_7 nbsub\_bid * per\_lo\_bid_t \\
&+ \beta_8 time\_ex\_ask_t + \beta_9 dese\_bid\_ask_t + \beta_{10} spread_t + u_t,
\end{aligned} \tag{6}$$

where  $i = 1, 2$ .  $FA1_t$  is the shift factor (first principal component) extracted from the ask side PCA.<sup>15</sup>  $FA2_t$  is the ask side rotation factor. For the bid side we have, analogously:

$$\begin{aligned}
FBi_t &= \beta_0 + \alpha FBi_{t-1} + \beta_1 nbtr\_agg\_ask_t + \beta_2 trade\_size\_ask_t + \beta_3 vola\_bid_t + \beta_4 vola\_ask_t \\
&+ \beta_5 nbsub\_ask_t + \beta_6 per\_lo\_ask_t + \beta_7 nbsub\_ask * per\_lo\_ask_t \\
&+ \beta_8 time\_ex\_bid_t + \beta_9 dese\_bid\_ask_t + \beta_{10} spread_t + u_t.
\end{aligned} \tag{7}$$

where  $i = 1, 2$ .  $FB1_t$  denotes the bid side shift factor and  $FB2_t$  the bid side rotation factor.

The regression equations include a first-order autoregressive term as the autocorrelation in the price impacts may have spilled over to the factors. The regressors are designed to capture the effect of trade size, order imbalances, composition of order flow, volatility and spread discussed above. They are constructed as follows: *nbtr\_agg\_bid* (*nbtr\_agg\_ask*) stands for the number of trades that walk up the book on the ask (bid) side, i.e. aggressive buy (sell) trades; *trade\_size\_bid* (*trade\_size\_ask*) is the average trade size for buy (sell) trades; *vola\_bid* and *vola\_ask* are the volatility measures computed on the best bid and ask prices respectively; *nbsub\_bid* (*nbsub\_ask*) stands for the number of orders submitted on the bid (ask) side; *per\_lo\_bid* and *per\_lo\_ask*, respectively, denotes the percentage of these orders that are limit orders; *time\_ex\_ask* (*time\_ex\_bid*) is the time to execution for limit orders standing on the ask (bid) side; *deseq\_bid\_ask* is the difference of buy minus sell market orders; *spread* is

---

<sup>15</sup>We focus on the visible book factors, but the results for the complete book are qualitatively identical.

the bid-ask spread. All variables except the spread are computed over the 5-minute interval **prior** to the observation of the factor at time  $t$ .

Regression results for the thirty stocks are given in Table VI. As conjectured, the autoregressive coefficient is indeed significant, i.e. the factors are autocorrelated. For both sides of the book the same explanatory variables are significant. We therefore focus on the results for the ask side of the book.

Few variables in the rotation factor (F2) equation are significant, and the sign of the coefficients varies across the 30 stocks. The most often significant variables (see the ‘sign’ column in the table) are trade size, spread and trade aggressiveness. It seems that, when buy orders are aggressive, that is these trades walk up the book, traders on the ask side provide cheap liquidity to the market. This supports the interpretation that the rotation factor captures liquidity effects. On the other hand, large trades deter liquidity provision at the best quotes as trade size tends to have negative impact on the rotation factor. This would rather associate the rotation factor with informational effects. The cross stock evidence regarding the effect of volatility on the rotation factor is mixed: For most stocks a higher volatility deters limit order submission close to the best prices, while for others stocks it attracts liquidity. Overall, the results provide mixed evidence regarding the interpretation of the rotation factor as capturing liquidity or informational effects.

The effect of the spread on the rotation factor is mostly positive, which implies that liquidity close to the inside market increases and decreases deeper in the book. Traders submit more limit orders when market orders are expensive due to large spreads, and they submit their orders at or close to the best quotes. Larger spreads lead to deeper books. This implies that liquidity measured by the inside spread and liquidity as measured by price impacts are not the same.

Trade aggressiveness and to some extent trade size exert a negative effect on the shift factor (F1). This is at odds with predictions of the papers by Glosten (1994) and Seppi (1997) which assume that trade size and trade aggressiveness are informative and thus should decrease liquidity. This result supports an interpretation of the shift factor as capturing liquidity rather than informational effects. On the other hand both bid and ask realized volatility measures increase the shift factor and consequently the size of the price impacts. This is consistent with the hypothesis that the shift factor captures the information effects.

Finally, the book tends to become more liquid at the ask side when there are more order submissions on the bid side. This effect is muted if the proportion of patient traders increases. This is consistent with the model of Foucault, Kadan, and Kandel (2003) which predicts that an increase in orders submitted to the book increases the likelihood of a trade and thus limit order traders ask for a smaller reward for the provision of liquidity. This effect is less pronounced if, at the same time, traders become more patient and submit more limit orders. Again this is consistent with the interpretation of the shift factor as being liquidity driven.

To summarize, although we can clearly identify the role of the first and second principal components as factors that shift and rotate the price impact profiles of the order book, the empirical evidence regarding their interpretation is mixed. Both factors seem to be linked to informational effects and non-informationally linked liquidity processes.

#### **D. Cross sectional evidence**

We have outlined above that the PCA delivers a stock specific measure of the importance of the shift and the rotation factor (increments of the cumulative  $R^2$ s). Table IV shows that although for all stocks the explanatory power of the shift factor exceeds that of the rotation factor there exists considerable cross sectional variation. Take as an example the ask side of the visible book. The explanatory power of the first principal components ranges from 0.57 to 0.85, while the explanatory power of the second principal component ranges from 0.11 to 0.3.

This data is used to further investigate the ambiguous interpretations of shift and rotation factor. For this purpose we run cross sectional regressions where the stock specific explanatory powers of the shift and rotation factor are used as dependent variables and regressed on a stock specific measure of the informational content of the order flow and a transaction costs measure that is purged of informational effects. There are several ways to produce these indicators. One could compute spread decompositions (Glosten and Harris (1988), Huang and Stoll (1997), Madhavan, Richardson, and Roomans (1997)) or adopt the approach by Easley, Kiefer, O'Hara, and Paperman (1996) to estimate the probability of informed trading for each stock. We employ a robust, model free method and compute stock specific average effective and realized spreads. The difference of effective and realized spread is a measure for the informational order flow, and the realized spread can be interpreted as transaction costs



net of trader information (see e.g. Boehmer (2004) and SEC (2001)). The average effective spread is computed by taking two times the absolute difference of the transaction price of a trade and the prevailing midquote and averaging over all trades of a stock. The average realized spread is computed similarly, but instead of taking the prevailing midquote, we use the midquote five minutes after the trade.<sup>16</sup>

To provide cross sectional evidence on the interpretation of the shift and rotation factors, the average realized spread ( $\overline{RS}$ ) and the difference of average effective spread and average realized spread ( $\overline{ES} - \overline{RS}$ ) are regressed on the explanatory power of the shift factor ( $F1R^2$ ) and the rotation factor ( $F2R^2$ ), respectively. Ask and bid sides of the book are pooled so that we have 60 observations.

If stocks for which the informational content of the order flow (measured as  $\overline{ES} - \overline{RS}$ ) is large tend to have a higher importance of shift or rotation factor, then this would support the hypothesis that the respective factor is associated with informational effects. Furthermore, we would assume that for stocks with average realized spreads non-informative fluctuations of liquidity play a more important role in explaining the variation of the book. Accordingly, if the average realized spread is positively correlated with the explanatory power of one of the two principal components then this would suggest that the respective factor is liquidity-linked. We estimate the regressions based on the visible and the complete book data and obtain the following results (robust t-statistics are reported in the parentheses):

### Visible book

$$\begin{aligned}
 F1R^2 &= 0.90 - 192.0 (\overline{ES} - \overline{RS}) + 180.6 \overline{RS} + \hat{\epsilon}_1 & R_{adj}^2 &= 0.47 \\
 &(40.3) \quad (-5.2) & & (4.3) \\
 F2R^2 &= 0.09 + 81.9 (\overline{ES} - \overline{RS}) - 53.6 \overline{RS} + \hat{\epsilon}_2 & R_{adj}^2 &= 0.23 \\
 &(6.3) \quad (3.4) & & (-2.0)
 \end{aligned}$$

---

<sup>16</sup>To ensure comparability across stocks, we compute effective and realized spreads relative to the midquote prevailing at the time of the trade. The difference of effective and realized spread is often referred to as price impact (see Boehmer (2004)). We don't use this notion in order to avoid confusion with our definition of price impacts in Equation (2)

### Complete book

$$F1R^2 = 0.89 - 166.9 (\overline{ES} - \overline{RS}) + 145.0 \overline{RS} + \hat{\epsilon}_3 \quad R_{adj}^2 = 0.41$$

(42.2) (-4.8) (3.8)

$$F2R^2 = 0.09 + 80.3 (\overline{ES} - \overline{RS}) - 56.6 \overline{RS} + \hat{\epsilon}_4 \quad R_{adj}^2 = 0.24$$

(6.5) (3.5) (-2.3)

These results imply that for stocks for which we measure a higher informational content of the order flow, more of the variation of the order book is explained by the rotation factor and less by the shift factor. On the other hand for stocks for with higher transaction costs (net of informational effects) the shift factor explains more of the variation of the book. These results support the interpretation that the rotation factor is associated with informational effects and that the shift factor is linked to non-information induced liquidity variations.

## V. Conclusions

This paper looks at the possible commonalities exhibited by a limit order book market. The empirical application deals with the Xetra trading system, which is a pure automated auction market similar to Euronext or the Hong Kong stock exchange. Our dataset is quite extensive, as we have the full reconstructed limit order book (in ‘real time’) for the 30 stocks that make up the DAX index over a 3-month time frame. Moreover we also have the so-called visible and complete order books, i.e. we reconstruct the order book with and without the hidden orders.

The search for and modelling of commonalities relies on the use of principal components analysis. A first important result is that very few factors are needed to model the full (either visible or complete) order book. Actually, two factors (interpreted as a shift and rotation factor) already provide a total explanatory power of more than 92% for a grid of traded volumes that range from 50,000 to 2 million euros (bid side or ask side of the book). This allows for a formidable reduction in the complexity of the modelling as we can therefore focus on these two factors and their associated set of weights. A by-product of our analysis is that we can show that the bid and ask sides do feature separate dynamics. Indeed, a joint analysis of both sides shows that a 4-factor structure is needed when both sides are to be modelled simultaneously. From a finance point of view, we interpret this evidence as indicating that order book markets involve heterogeneous agents who need not be active on both sides of the

market. We also provide a time-of-day analysis of the two principal components. Given the set of weights and the diurnal variation displayed by the principal components, we conclude that the Xetra system is characterized by a rather poor liquidity at the start of the trading day, but also that liquidity remains good up to the close of trading. In other words, we document a L-shaped pattern for the intraday liquidity, quite similar to what is known for NASDAQ. In the final part of the paper, we interpret the factors in light of recent market microstructure theory. First, in a time series framework, we regress the factors identified previously on a selection of market variables which summarize the trade flow and order placement strategies of the market participants. Secondly, in a cross section framework, we run regressions where the stock specific explanatory powers of the shift and rotation factor are regressed on a stock specific measure of the informational content of the order flow and a transaction costs measure that is purged of informational effects. Both methodologies tend to show that the shift and rotation factors seem to be linked to informational effects and liquidity effects, although the cross sectional analysis favors the interpretation according to which the rotation factor is associated with informational effects and the shift factor is linked to non-information induced liquidity variations.

## References

- AHN, H., K. BAE, AND K. CHAN (2001): “Limit orders, depth and volatility: evidence from the stock exchange of Hong Kong,” *Journal of Finance*, 56, 767–788.
- AHN, H., AND Y. CHEUNG (1999): “The intraday patterns of the spread and depth in a market with market makers: the Stock Exchange of Hong Kong,” *Pacific-Basin Finance Journal*, 7, 539–556.
- ALEXANDER, C. (2001): *Market models*. Wiley, New York.
- BAE, K., H. JANG, AND S. PARK (2003): “Traders’ choice between limit and market orders: evidence from NYSE stocks,” *Journal of Financial Markets*, 6, 517–538.
- BELTRAN, H., A. DURRÉ, AND P. GIOT (2004): “How does liquidity react to stress periods in a limit order market?,” National Bank of Belgium Working Paper 49.
- BIAIS, B., P. HILLION, AND C. SPATT (1995): “An empirical analysis of the limit order book and the order flow in the Paris Bourse,” *Journal of Finance*, 50, 1655–1689.
- BOEHMER, E. (2004): “Dimensions of execution quality: Recent evidence for U.S. equity markets,” Working Paper, Texas A&M University.
- BROCK, W., AND A. KLEIDON (1992): “Periodic market closure and trading volume: a model of intraday bids and asks,” *Journal of Economic Dynamics and Control*, 16, 451–489.
- BROCKMAN, P., AND D. CHUNG (1998): “Inter- and intra-day liquidity patterns on the Stock Exchange of Hong Kong,” *Journal of International Financial Markets, Institutions and Money*, 8, 277–298.
- (1999): “An analysis of depth behavior in an electronic, order-driven environment,” *Journal of Banking and Finance*, 23, 1861–1886.
- CAO, C., O. HANSCH, AND X. WANG (2004): “The informational content of an open limit order book,” .
- CHAN, K., W. CHRISTIE, AND P. SCHULTZ (1995): “Market structure and the intraday evolution of bid-ask spreads for NASDAQ securities,” *Journal of Business*, 68, 35–60.

- CHORDIA, T., R. ROLL, AND A. SUBRAHMANYAM (2002): “Order imbalance, liquidity, and market returns,” *Journal of Financial Economics*, 65, 111–130.
- CHUNG, K., B. VAN NESS, AND R. VAN NESS (1999): “Limit orders and the bid-ask spread,” *Journal of Financial Economics*, 53, 255–287.
- COPPEJANS, M., I. DOMOWITZ, AND A. MADHAVAN (2002): “Liquidity in an automated auction,” Mimeo.
- DANIELSSON, J., AND R. PAYNE (2001): “Measuring and explaining liquidity on an electronic limit order book: evidence from Reuters D2000-2,” Financial Markets Group, London School of Economics.
- DEGRYSE, H. (1999): “The total cost of trading Belgian shares: Brussels versus London,” *Journal of Banking and Finance*, 23, 1331–1355.
- DEUTSCHE BÖRSE AG (1999): “Xetra Market Model Release 3 Stock Trading,” Technical Report.
- DOMOWITZ, I. (1993): “A taxonomy of automated trading systems,” *Journal of International Money and Finance*, 12, 607–631.
- EASLEY, D., N. KIEFER, M. O’HARA, AND J. PAPERMAN (1996): “Liquidity, Information, and Less-Frequently Traded Stocks,” *Journal of Finance*, 51, 1405–1436.
- FOSTER, F., AND S. VISWANATHAN (1994): “Strategic trading with asymmetrically informed investors and long-lived information,” *Journal of Financial and Quantitative Analysis*, 29, 499–518.
- FOUCAULT, T. (1999): “Order flow composition and trading costs in a dynamic limit order market,” *Journal of Financial Markets*, 2, 99–134.
- FOUCAULT, T., O. KADAN, AND E. KANDEL (2003): “Limit order book as a market for liquidity,” *Forthcoming, Review of Financial Studies*.
- GLOSTEN, L. (1994): “Is the electronic open limit order book inevitable?,” *Journal of Finance*, 49, 1127–1161.

- GLOSTEN, L., AND L. HARRIS (1988): “Estimating the Components of the Bid-Ask Spread,” *Journal of Financial Economics*, 21, 123–142.
- GOMBER, P., U. SCHWEICKERT, AND E. THEISSEN (2004): “Zooming in on Liquidity,” EFA 2004 Working Paper, University of Bonn.
- GOURIÉROUX, C., G. LE FOL, AND B. MEYER (1998): “Analyse du carnet d’ordres,” *Banque et Marchés*, 36, 5–20.
- HUANG, R., AND H. STOLL (1997): “The Components of the Bid-Ask Spread: A General Approach,” *Review of Financial Studies*, 10(4), 995–1034.
- IRVINE, P., G. BENSTON, AND E. KANDEL (2000): “Liquidity beyond the inside spread: measuring and using information in the limit order book,” Mimeo, Goizueta Business School, Emory University, Atlanta.
- MADHAVAN, A. (1992): “Trading mechanisms in securities markets,” *Journal of Finance*, 47, 607–642.
- MADHAVAN, A., M. RICHARDSON, AND M. ROOMANS (1997): “Why do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks,” *Review of Financial Studies*, (10), 1035–1064.
- PARLOUR, C. A. (1998): “Price Dynamics in Limit Order Markets,” *Review of Financial Studies*, 11, 789–816.
- SEC (2001): “Report on the comparison of order execution across equity market structures,” U.S. Securities and Exchange Commission, Washington.
- SEPPI, D. J. (1997): “Liquidity provision with limit orders and a strategic specialist,” *Review of Financial Studies*, 10, 103–150.
- STOLL, H. (1978): “The supply of dealer services in securities markets,” *Journal of Finance*, 33, 1133–1151.
- VENKATAMARAN, K. (2001): “Automated versus floor trading: an analysis of execution costs on the Paris and New York Stock Exchange,” *Journal of Finance*, 56, 1445–1485.

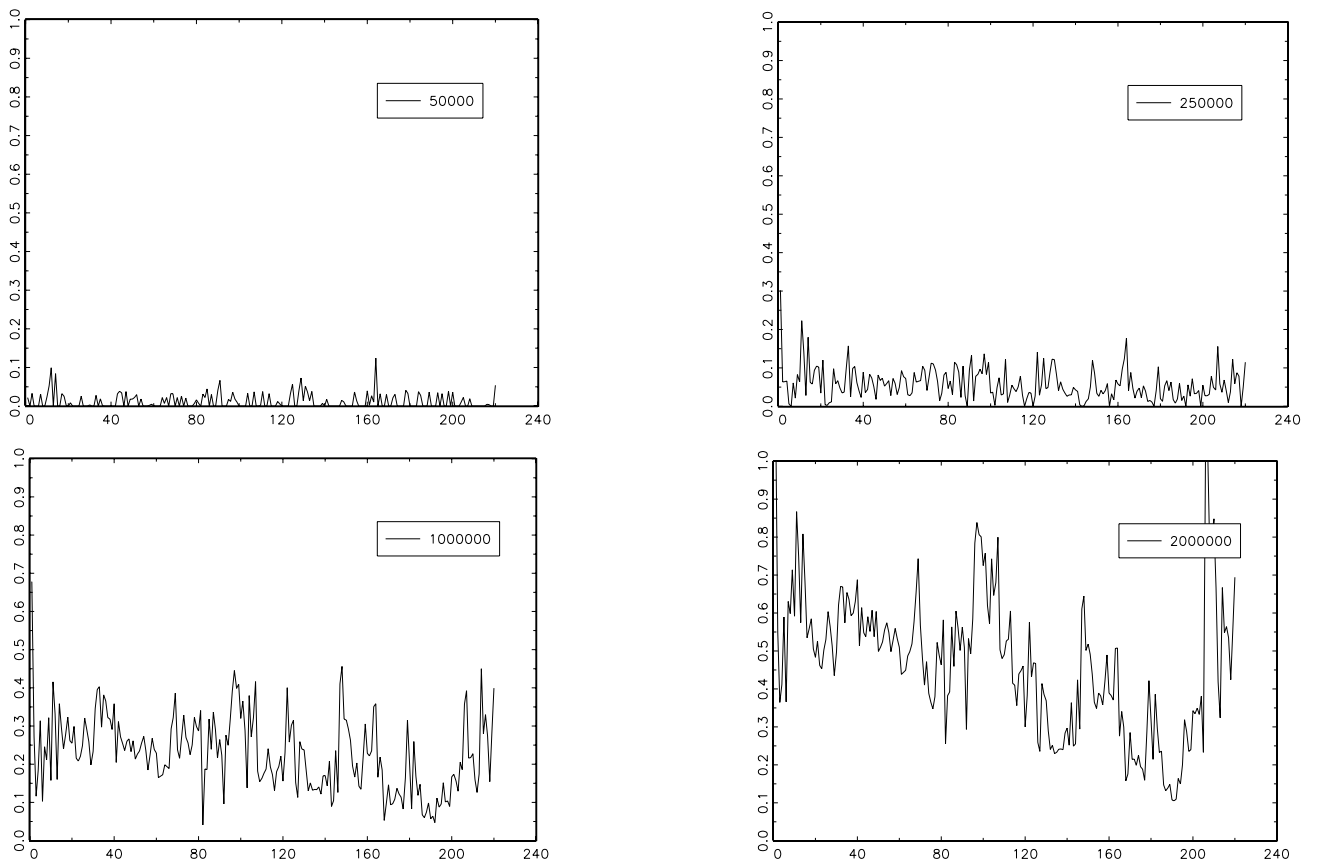


Figure 1: Time series paths for  $v$ -euros ask price impacts, complete book, BAYER. The first 240 observations at a 5-minute sampling time are plotted (102 observations per trading day).

Table I: Characteristics of the DAX30 constituent stocks.

| Ticker  | Company Name              | Daily     |          | Market cap<br>(mill.) | Trade<br>size | % agg<br>trades | Daily nb<br>trades | Daily nb<br>subm. | Daily nb<br>cancel. | Price | Spread | Spread<br>% |
|---------|---------------------------|-----------|----------|-----------------------|---------------|-----------------|--------------------|-------------------|---------------------|-------|--------|-------------|
|         |                           | Turnover  | Turnover |                       |               |                 |                    |                   |                     |       |        |             |
| DTE     | DT.TELEKOM AG NA          | 350627866 | 34858    | 78884                 | 5.0           | 4445            | 14498              | 11009             | 15.7                | 0.01  | 0.07   |             |
| SIE     | SIEMENS AG NA             | 321704299 | 52893    | 72831                 | 16.7          | 4418            | 23659              | 19920             | 64.0                | 0.03  | 0.05   |             |
| DBK     | DEUTSCHE BANK AG NA O.N.  | 309282831 | 38228    | 78083                 | 19.3          | 3961            | 23169              | 19772             | 67.2                | 0.03  | 0.05   |             |
| ALV     | ALLIANZ AG VNA O.N.       | 289980556 | 33805    | 64114                 | 21.4          | 4523            | 29791              | 25882             | 100.1               | 0.05  | 0.05   |             |
| MUV2    | MUENCH.RUECKVERS.VNA O.N. | 207353230 | 16396    | 60534                 | 20.7          | 3425            | 20154              | 16894             | 93.9                | 0.06  | 0.06   |             |
| DCX     | DAIMLERCHRYSLER AG NA O.N | 187737846 | 30316    | 56736                 | 14.5          | 3309            | 18722              | 15919             | 36.4                | 0.02  | 0.06   |             |
| SAP     | SAP AG ST O.N.            | 184628162 | 27412    | 65795                 | 21.9          | 2806            | 19733              | 17095             | 131.5               | 0.08  | 0.06   |             |
| EOA     | E.ON AG O.N.              | 160625983 | 33753    | 55950                 | 13.6          | 2871            | 18899              | 16468             | 52.5                | 0.03  | 0.06   |             |
| IFX     | INFINEON TECH.AG NA O.N.  | 146462315 | 4790     | 52331                 | 8.6           | 2799            | 10320              | 7744              | 11.6                | 0.01  | 0.10   |             |
| BAS     | BASF AG O.N.              | 124434537 | 25425    | 48236                 | 13.8          | 2580            | 18211              | 15898             | 43.3                | 0.03  | 0.06   |             |
| VOW     | VOLKSWAGEN AG ST O.N.     | 104249843 | 9688     | 40963                 | 16.0          | 2545            | 13474              | 11273             | 39.2                | 0.03  | 0.07   |             |
| HVM     | BAY.HYPO-VEREINSBK.O.N.   | 98351090  | 6629     | 50783                 | 15.0          | 1937            | 10204              | 8293              | 18.7                | 0.02  | 0.11   |             |
| RWE     | RWE AG ST O.N.            | 97655566  | 12653    | 42203                 | 13.0          | 2314            | 14438              | 12355             | 33.8                | 0.03  | 0.08   |             |
| BAY     | BAYER AG O.N.             | 88776121  | 15911    | 36994                 | 12.4          | 2400            | 15258              | 12988             | 23.1                | 0.02  | 0.08   |             |
| BMW     | BAY.MOTOREN WERKE AG ST   | 87854358  | 12211    | 41639                 | 14.4          | 2110            | 14736              | 12764             | 34.7                | 0.02  | 0.07   |             |
| CBK     | COMMERZBANK AG O.N.       | 53171668  | 7569     | 36659                 | 12.6          | 1450            | 11922              | 10476             | 15.4                | 0.02  | 0.11   |             |
| SCH     | SCHERING AG O.N.          | 51413053  | 7055     | 33756                 | 16.2          | 1523            | 9111               | 7669              | 40.8                | 0.04  | 0.09   |             |
| LHA     | LUFTHANSA AG VNA O.N.     | 43946809  | 4548     | 32504                 | 11.9          | 1352            | 8079               | 6780              | 14.2                | 0.02  | 0.12   |             |
| DPW     | DEUTSCHE POST AG NA O.N.  | 43836617  | 6806     | 33330                 | 11.0          | 1315            | 6861               | 5666              | 18.2                | 0.02  | 0.11   |             |
| MEO     | METRO AG ST O.N.          | 38874669  | 5018     | 31480                 | 15.7          | 1235            | 7975               | 6702              | 35.0                | 0.04  | 0.12   |             |
| TKA     | THYSSENKRUPP AG O.N.      | 37892493  | 6450     | 30017                 | 11.3          | 1262            | 7864               | 6672              | 15.9                | 0.02  | 0.13   |             |
| DBI     | DEUTSCHE BOERSE NA O.N.   | 35696903  | 4847     | 36359                 | 18.4          | 982             | 6598               | 5698              | 46.9                | 0.04  | 0.10   |             |
| ADS     | ADIDAS-SALOMON AG O.N.    | 31976047  | 4104     | 32635                 | 20.1          | 980             | 8057               | 7105              | 92.6                | 0.08  | 0.09   |             |
| ALT     | ALTANA AG O.N.            | 30985416  | 3338     | 28310                 | 18.9          | 1095            | 7718               | 6609              | 48.6                | 0.05  | 0.10   |             |
| MAN     | MAN AG ST O.N.            | 27685031  | 2434     | 26189                 | 13.0          | 1057            | 7214               | 6235              | 27.7                | 0.03  | 0.12   |             |
| TUI     | TUI AG O.N.               | 26281175  | 2025     | 24723                 | 17.6          | 1063            | 6767               | 5714              | 18.7                | 0.03  | 0.14   |             |
| CONT    | CONTINENTAL AG O.N.       | 25627638  | 4060     | 25574                 | 13.5          | 1002            | 8036               | 7052              | 31.6                | 0.04  | 0.11   |             |
| LIN     | LINDE AG O.N.             | 22378772  | 3448     | 24971                 | 15.8          | 896             | 8342               | 7454              | 43.6                | 0.05  | 0.11   |             |
| HEN3    | HENKEL KGAA VZO O.N.      | 18174548  | 3682     | 25904                 | 16.6          | 702             | 7989               | 7306              | 65.9                | 0.07  | 0.10   |             |
| FME     | FRESEN.MED.CARE AG O.N.   | 12850947  | 1944     | 20680                 | 16.7          | 621             | 5764               | 5195              | 54.0                | 0.07  | 0.13   |             |
| Average |                           | 108683880 | 14076    | 42972                 | 15.2          | 2099            | 12785              | 10887             | 44.5                | 0.04  | 0.09   |             |

The table reports some characteristics of the stocks belonging to the DAX30 index. For each stock, *Market cap* is the market capitalization in million euros at the end of December 2003, *Trade size* is the average trade size over the 3-month sample, and *%agg trades* is the percentage of total trading volume that has not been executed at the best prices (that is, the order was aggressive and walked up the book). The remaining statistics are daily averages over the 3 months: *Daily turnover* is the total amount of shares traded in euros per trading day, *Daily nb trades* is the daily number of trades, *Daily nb subm.* is the daily number of order submissions, market orders excluded, and *Daily nb cancel.* is the daily number of order cancellations. Finally, *Price*, *Spread* and *Spread*, % refer to the average midquote, spread and relative spread over the 3 months. The sample ranges from January 2, 2003 to March 31, 2003.



Table II: Fraction of the variance of the price impacts explained by given frequency bands

|           | one hour | intraday | week | month | any |
|-----------|----------|----------|------|-------|-----|
| 50,000    | 76.5     | 19.4     | 3.2  | 0.6   | 0.2 |
| 250,000   | 60.3     | 30.4     | 7.4  | 1.5   | 0.4 |
| 1,000,000 | 35.2     | 45.6     | 15.2 | 3.1   | 0.9 |
| 2,000,000 | 24.5     | 51.0     | 19.3 | 4.0   | 1.2 |

This table shows how much each frequency band contributes to the total variance of the price impacts. Each row corresponds to a complete book ask price impact series, for a hypothetical trade size of 50,000, 250,000, 1 million and 2 million euros. Each column pertains to a frequency band: *one hour* for frequencies above one hour, *intraday* for frequencies larger than the daily frequency but lower than the hourly frequency, *week* for frequencies between the daily and the weekly frequencies, *month* for frequencies between the weekly and the monthly frequencies, and *any* for all frequencies lower than the monthly frequency. The table reads as such: 19.4% of the variance of the price impact of 50,000-euro trades is due to intraday variations, variations within the hour not included; if we include all intraday frequencies, the percentage rises to 95.9%.

Table III: Correlations between price impact series

| variables             | avg   | std  | min   | max   |
|-----------------------|-------|------|-------|-------|
| PIA 3 — PIA 21        | 0.41  | 0.10 | 0.01  | 0.51  |
| PIA 3 — PIA 41        | 0.27  | 0.10 | -0.01 | 0.38  |
| PIA 3 — PIB 3         | 0.10  | 0.06 | 0.00  | 0.23  |
| PIA 3 — PIB 21        | 0.15  | 0.06 | 0.05  | 0.31  |
| PIA 3 — PIB 41        | 0.14  | 0.05 | 0.06  | 0.24  |
| DIFF1 PIA 3 — PIA 21  | -0.16 | 0.06 | -0.29 | -0.02 |
| DIFF1 PIA 3 — PIA 41  | -0.08 | 0.04 | -0.17 | 0.01  |
| DIFF1 PIA 3 — PIB 3   | 0.00  | 0.02 | -0.02 | 0.06  |
| DIFF1 PIA 3 — PIB 21  | -0.01 | 0.02 | -0.03 | 0.03  |
| DIFF1 PIA 3 — PIB 41  | -0.01 | 0.02 | -0.05 | 0.01  |
| PIA 21 — PIA 3        | 0.41  | 0.10 | 0.01  | 0.51  |
| PIA 21 — PIA 41       | 0.77  | 0.20 | 0.00  | 0.91  |
| PIA 21 — PIB 3        | 0.15  | 0.06 | 0.00  | 0.25  |
| PIA 21 — PIB 21       | 0.25  | 0.11 | 0.00  | 0.58  |
| PIA 21 — PIB 41       | 0.23  | 0.11 | 0.02  | 0.48  |
| DIFF1 PIA 21 — PIA 3  | -0.27 | 0.07 | -0.37 | -0.01 |
| DIFF1 PIA 21 — PIA 41 | -0.21 | 0.09 | -0.40 | -0.01 |
| DIFF1 PIA 21 — PIB 3  | 0.00  | 0.02 | -0.03 | 0.05  |
| DIFF1 PIA 21 — PIB 21 | -0.02 | 0.02 | -0.05 | 0.01  |
| DIFF1 PIA 21 — PIB 41 | -0.03 | 0.02 | -0.06 | 0.01  |
| PIA 41 — PIA 3        | 0.27  | 0.10 | -0.01 | 0.38  |
| PIA 41 — PIA 21       | 0.77  | 0.20 | 0.00  | 0.91  |
| PIA 41 — PIB 3        | 0.12  | 0.06 | -0.01 | 0.23  |
| PIA 41 — PIB 21       | 0.25  | 0.10 | 0.00  | 0.43  |
| PIA 41 — PIB 41       | 0.25  | 0.13 | 0.04  | 0.54  |
| DIFF1 PIA 41 — PIA 3  | -0.18 | 0.07 | -0.29 | 0.01  |
| DIFF1 PIA 41 — PIA 21 | -0.27 | 0.12 | -0.48 | -0.01 |
| DIFF1 PIA 41 — PIB 3  | -0.01 | 0.02 | -0.05 | 0.04  |
| DIFF1 PIA 41 — PIB 21 | -0.03 | 0.02 | -0.10 | 0.00  |
| DIFF1 PIA 41 — PIB 41 | -0.04 | 0.05 | -0.30 | 0.01  |

This table reports the correlations between the visible price impact series at the bid and ask side. PIA stands for ask price impacts, PIB stands for bid price impacts. PIA3 corresponds to an ask price impact for a trade size of 100,000 euros, PIA21 1 million, and PIA41 2 million euros (the same holds for the bid price impacts PIB3, PIB21, PIB41). *DIFF1* means that the variable is differentiated as e.g.  $DIFF1PIA3 = PIA3_{t+1} - PIA3_t$ . Variables not differentiated are observed at time  $t$ . *Avg* is the average correlation across the 30 stocks, *Std* its standard deviation, and *Min* and *Max* the smallest and largest correlations across the 30 stocks. For example, an entry reads as follows: the correlation between the ask price impact for a trade size of 100,000 euros (PIA3) and the ask price impact for a trade of 1 million euros (PIA21) is on average 41%.

Table IV: Principal components analysis

| Visible price impacts            |      |      |      |      |          |      |      |      |      |
|----------------------------------|------|------|------|------|----------|------|------|------|------|
| Ask                              |      |      |      |      | Bid      |      |      |      |      |
|                                  | Avg  | Stdv | Min  | Max  |          | Avg  | Stdv | Min  | Max  |
| F1                               | 0.78 | 0.08 | 0.57 | 0.85 | F1       | 0.79 | 0.05 | 0.66 | 0.85 |
| F2                               | 0.14 | 0.04 | 0.11 | 0.30 | F2       | 0.14 | 0.02 | 0.10 | 0.20 |
| F3                               | 0.04 | 0.02 | 0.03 | 0.11 | F3       | 0.04 | 0.02 | 0.03 | 0.10 |
| F4                               | 0.02 | 0.01 | 0.01 | 0.05 | F4       | 0.02 | 0.01 | 0.01 | 0.03 |
| Complete price impacts           |      |      |      |      |          |      |      |      |      |
| Ask                              |      |      |      |      | Bid      |      |      |      |      |
|                                  | Avg  | Stdv | Min  | Max  |          | Avg  | Stdv | Min  | Max  |
| F1                               | 0.79 | 0.07 | 0.57 | 0.85 | F1       | 0.79 | 0.05 | 0.67 | 0.85 |
| F2                               | 0.14 | 0.04 | 0.11 | 0.29 | F2       | 0.13 | 0.02 | 0.11 | 0.19 |
| F3                               | 0.04 | 0.01 | 0.03 | 0.08 | F3       | 0.04 | 0.02 | 0.03 | 0.10 |
| F4                               | 0.02 | 0.01 | 0.01 | 0.05 | F4       | 0.02 | 0.01 | 0.01 | 0.03 |
| Visible and hidden price impacts |      |      |      |      |          |      |      |      |      |
| Ask                              |      |      |      |      | Bid      |      |      |      |      |
|                                  | Avg  | Stdv | Min  | Max  |          | Avg  | Stdv | Min  | Max  |
| F1                               | 0.50 | 0.06 | 0.38 | 0.59 | F1       | 0.49 | 0.05 | 0.36 | 0.58 |
| F2                               | 0.27 | 0.04 | 0.18 | 0.34 | F2       | 0.28 | 0.04 | 0.19 | 0.33 |
| F3                               | 0.10 | 0.02 | 0.07 | 0.15 | F3       | 0.09 | 0.02 | 0.07 | 0.14 |
| F4                               | 0.05 | 0.01 | 0.03 | 0.11 | F4       | 0.05 | 0.01 | 0.04 | 0.07 |
| Bid and ask price impacts        |      |      |      |      |          |      |      |      |      |
| Visible                          |      |      |      |      | Complete |      |      |      |      |
|                                  | Avg  | Stdv | Min  | Max  |          | Avg  | Stdv | Min  | Max  |
| F1                               | 0.51 | 0.05 | 0.39 | 0.59 | F1       | 0.50 | 0.05 | 0.40 | 0.59 |
| F2                               | 0.28 | 0.04 | 0.17 | 0.37 | F2       | 0.30 | 0.04 | 0.17 | 0.37 |
| F3                               | 0.08 | 0.02 | 0.05 | 0.13 | F3       | 0.08 | 0.02 | 0.05 | 0.13 |
| F4                               | 0.06 | 0.01 | 0.05 | 0.08 | F4       | 0.06 | 0.01 | 0.05 | 0.08 |

Explanatory power for the first 4 principal components in the PCA. The first panel (starting from the top) reports the explanatory power of a PCA of the visible ask price impacts (left side) and of a PCA of the visible bid price impacts (right side). F1 to F4 refer to the explanatory power for the first to fourth principal components respectively. The second panel displays similar tables except that the PCA is applied to the complete price impacts. The third panel reports the explanatory powers for a PCA on the visible and hidden price impacts, and the bottom panel for a PCA on both sides of the book. *Avg* is the average explanatory power across the 30 stocks, *Std* its standard deviation, and *Min* and *Max* the smallest and largest explanatory power across the 30 stocks.

Table V: Factors and volatility

| Realized volatility |                  |
|---------------------|------------------|
| $\beta_a$           | 0.13<br>[29,0]   |
| $\beta_b$           | 0.08<br>[28,0]   |
| $\gamma_a$          | -0.01<br>[12,12] |
| $\gamma_b$          | -0.02<br>[2,13]  |

The table reports averages of OLS-estimated coefficients across the 30 stocks. Figures in brackets are the number of significant and positive coefficients and number of significant and negative coefficients (at the 5% level). The  $R^2$  (average over the 30 stocks) is 4.2%.

Table VI: OLS results

|                      | Ask side |      |    | Bid side |    |      |                      |       |    |    |       |    |    |
|----------------------|----------|------|----|----------|----|------|----------------------|-------|----|----|-------|----|----|
|                      | F1       | Sign | F2 | Sign     | F1 | Sign | F2                   | Sign  |    |    |       |    |    |
| $\alpha$             | 0.67     | 30   | 0  | 0.46     | 30 | 0    | $\alpha$             | 0.67  | 30 | 0  | 0.46  | 30 | 0  |
| nbtr_agg_bid         | -0.04    | 1    | 23 | 0.01     | 11 | 1    | nbtr_agg_ask         | -0.04 | 1  | 23 | 0.01  | 13 | 4  |
| trade_size_bid       | -0.01    | 4    | 7  | -0.02    | 0  | 14   | trade_size_ask       | 0.00  | 8  | 7  | -0.03 | 0  | 19 |
| vola_bid             | 0.05     | 23   | 1  | -0.02    | 4  | 13   | vola_bid             | 0.07  | 24 | 2  | -0.02 | 4  | 11 |
| vola_ask             | 0.09     | 25   | 0  | -0.01    | 5  | 11   | vola_ask             | 0.04  | 19 | 1  | -0.02 | 4  | 13 |
| nbsub_bid            | -0.11    | 1    | 10 | -0.06    | 2  | 4    | nbsub_ask            | -0.14 | 2  | 10 | -0.02 | 1  | 3  |
| per_lo_bid           | -0.01    | 0    | 3  | 0.00     | 3  | 2    | per_lo_ask           | -0.02 | 0  | 10 | 0.01  | 6  | 0  |
| nbsub_bid*per_lo_bid | 0.11     | 8    | 2  | 0.07     | 5  | 2    | nbsub_ask*per_lo_ask | 0.14  | 12 | 2  | 0.01  | 2  | 1  |
| time_ex_ask          | 0.01     | 6    | 1  | -0.01    | 2  | 8    | time_ex_bid          | 0.02  | 13 | 0  | 0.01  | 11 | 2  |
| deseq_bid_ask        | 0.01     | 12   | 4  | 0.01     | 10 | 4    | deseq_bid_ask        | -0.03 | 1  | 22 | -0.01 | 2  | 7  |
| spread               | -0.01    | 3    | 8  | 0.09     | 28 | 1    | spread               | -0.01 | 6  | 6  | 0.10  | 29 | 0  |

Variable *nbtr\_agg\_bid* stands for the number of trades that walk up the book on the ask side, that is, buy aggressive trades; *trade\_size\_bid* is the average trade size for buy trades; *vola\_bid* and *vola\_ask* are the realized volatility for the bid and ask prices respectively; *nbsub\_bid* is the number of orders submitted on the bid side; *per\_lo\_bid* is the percentage of these orders that are limit orders; *time\_ex\_ask* is the time to execution for limit orders standing on the ask side; *deseq\_bid\_ask* is the difference of buy minus sell market orders; *spread* is the bid-ask spread. Coefficient  $\alpha$  is the autoregressive coefficient. All the variables but the spread are computed over the 5-minute interval prior to the observation of the factor at time  $t$ . The same definitions hold for the bid side. The equation is estimated using standard OLS. The column header 'F1' contains the average (across the 30 stocks) OLS coefficient when the dependent variable is the shift factor, 'F2' is for the rotation factor. *Sign* gives the number of stocks with a significant coefficient at the 5% level, + for positive and significant, - for negative and significant. On average for the bid (ask) side, the adjusted R-squared is 0.52 (0.52) for the shift factor equation and 0.27 (0.27) for the rotation factor equation.

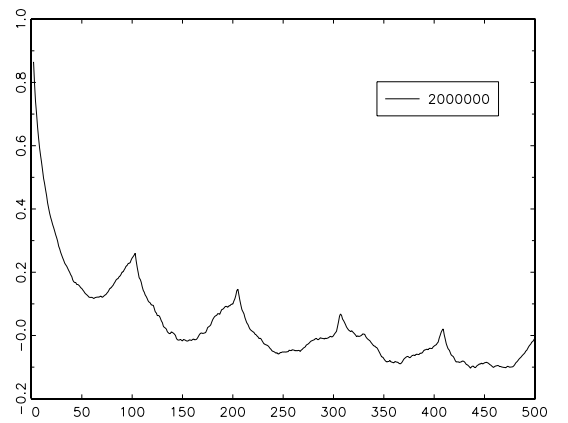
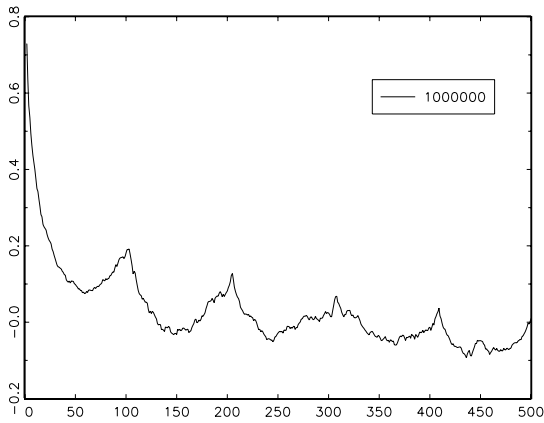
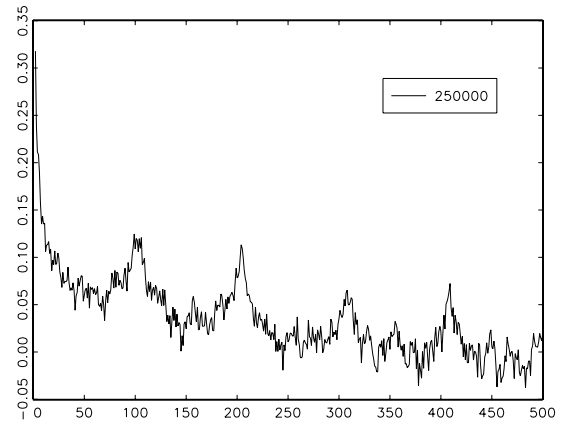
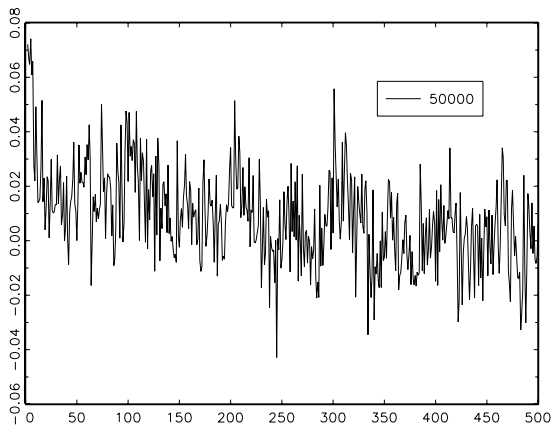


Figure 2: Autocorrelograms for  $v$ -euros ask price impacts (500 lags), complete book, BAYER.  
 One trading day has 102 observations.

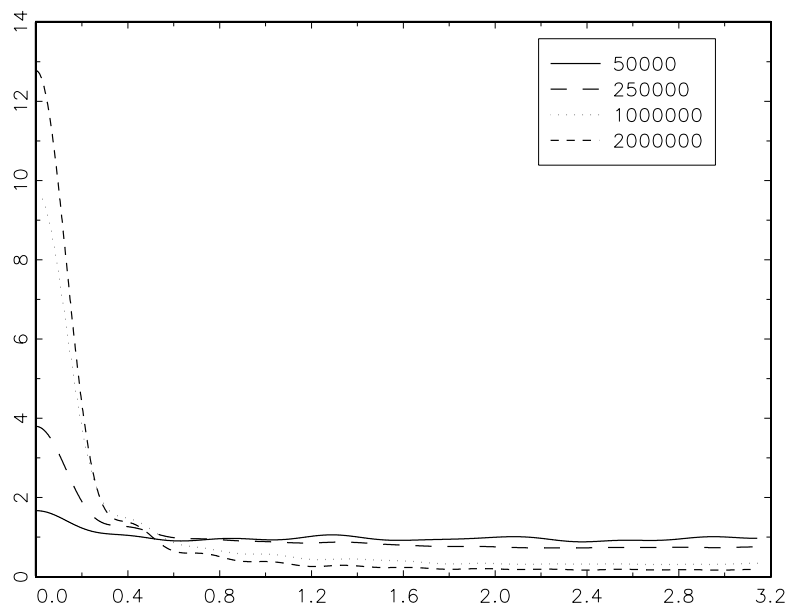


Figure 3: Spectra for  $v$ -euros ask price impacts, complete book, BAYER. The x-axis is the frequency, ranging from 0 to  $\pi$ .

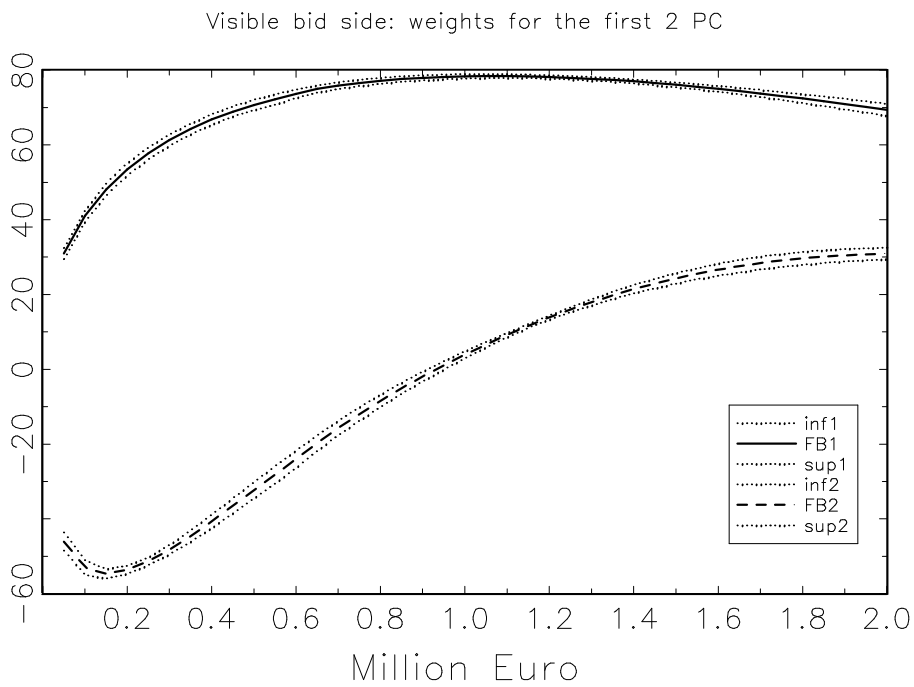
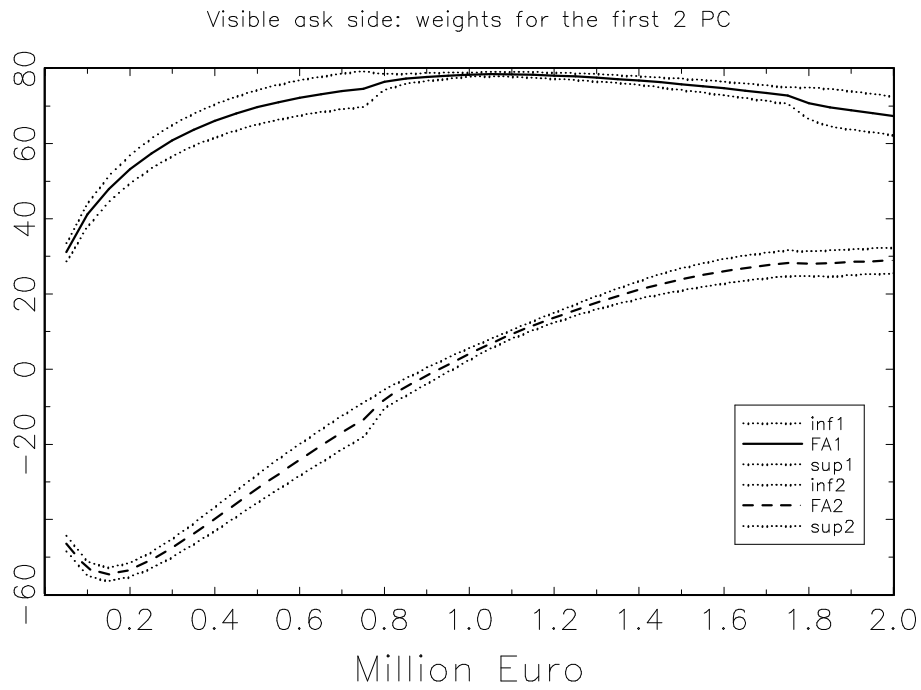


Figure 4: First and second principal components weights for the visible ask side (top of figure) and visible bid side (bottom of figure). Averages for the 30 stocks in our sample. The dotted line are the confidence bands at 95%.



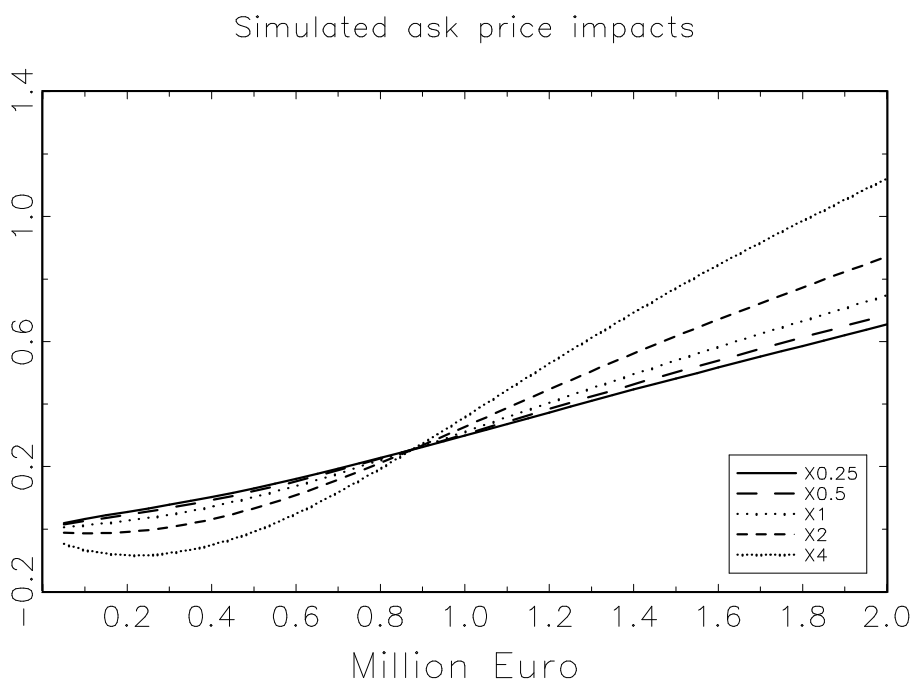
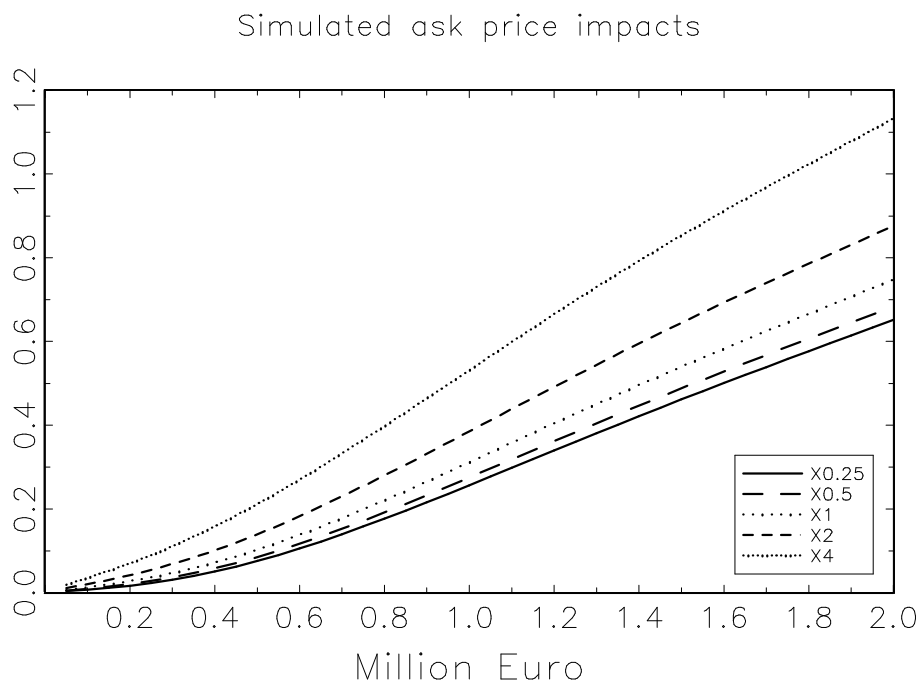
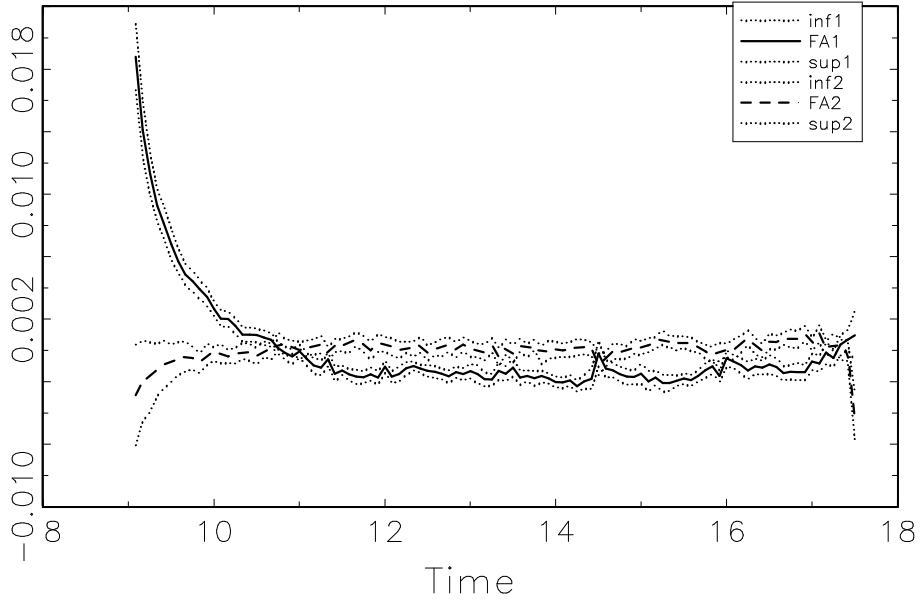


Figure 5: Simulations of the ask side of the book (Bayer stock) when the first and second principal components are individually increased (X2, X4) or decreased (X0.5, X0.25). The initial state is characterized by X1. From top to bottom, the panels present successively the individual effects of the increase/decrease of the first and second principal components.

Visible ask side: diurnal variation for the first 2 PC



Visible bid side: diurnal variation for the first 2 PC

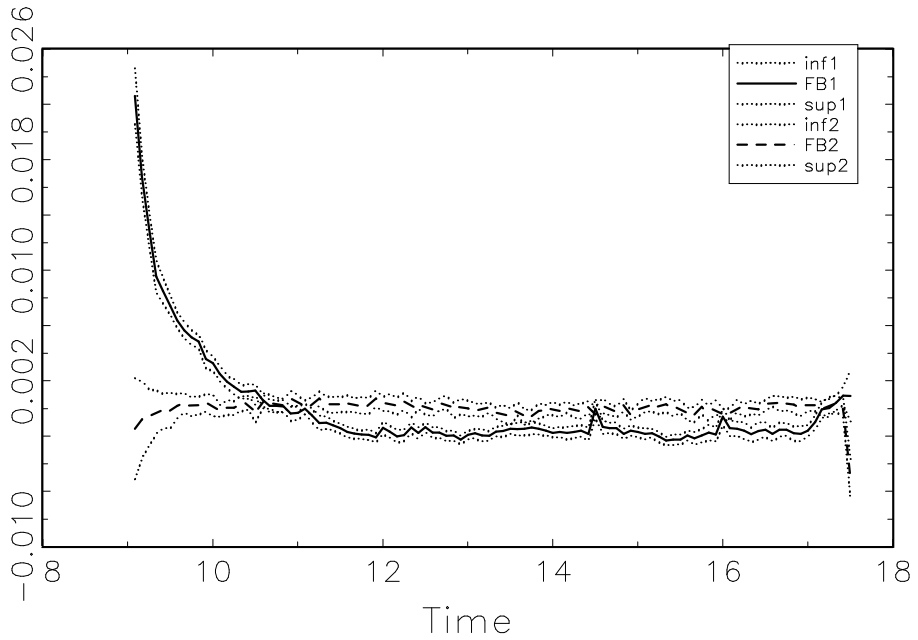


Figure 6: Diurnal variation of the first two principal components for the visible ask side (top of figure) and visible bid side (bottom of figure). Averages for the 30 stocks in our sample. The dotted lines are the confidence bands at 95%.