ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Hautsch, Nikolaus; Huang, Ruihong

Working Paper

The market impact of a limit order

SFB 649 Discussion Paper, No. 2009,051

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Hautsch, Nikolaus; Huang, Ruihong (2009) : The market impact of a limit order, SFB 649 Discussion Paper, No. 2009,051, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/39308

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

SFB 649 Discussion Paper 2009-051

The Market Impact of a Limit Order

Nikolaus Hautsch*
Ruihong Huang*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

The Market Impact of a Limit Order *

Nikolaus Hautsch†
Humboldt-Universität zu Berlin, CASE, CFS
Quantitative Products Laboratory

Ruihong Huang‡
Quantitative Products Laboratory
Humboldt-Universität zu Berlin

This version: October 2009

Abstract

Despite their importance in modern electronic trading, virtually no systematic empirical evidence on the market impact of incoming orders is existing. We quantify the short-run and long-run price effect of posting a limit order by proposing a high-frequency cointegrated VAR model for ask and bid quotes and several levels of order book depth. Price impacts are estimated by means of appropriate impulse response functions. Analyzing order book data of 30 stocks traded at Euronext Amsterdam, we show that limit orders have significant market impacts and cause a dynamic (and typically asymmetric) rebalancing of the book. The

*For helpful comments and discussions we thank Ulrich Horst, Lada Kyj, Roel Oomen and the participants of the Conference on "Individual Decision Making, High Frequency Econometrics and Limit Order Book Dynamics" in Warwick, 2009, as well as of workshops at the Humboldt-Universität zu Berlin and at the Quantitative Products Laboratory. This research is supported by the Deutsche Bank AG via the Quantitative Products Laboratory and the Deutsche Forschungsgemeinschaft (DFG) via the Collaborative Research Center 649 "Economic Risk".

†Institute for Statistics and Econometrics and Center for Applied Statistics and Economics (CASE), Humboldt-Universität zu Berlin as well as Quantitative Products Laboratory (QPL), Berlin, and Center for Financial Studies (CFS), Frankfurt. Email: nikolaus.hautsch@wiwi.hu-berlin.de. Address: Spandauer Str. 1, D-10178 Berlin, Germany.

‡Quantitative Products Laboratory (QPL), Berlin and Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin. Email: ruihong.huang@db.com. Address: Alexander Str. 5, D-10178 Berlin, Germany.

strength and direction of quote and spread responses depend on the incoming orders' aggressiveness, their size and the state of the book. We show that the effects are qualitatively quite stable across the market. Cross-sectional variations in the magnitudes of price impacts are well explained by the underlying trading frequency and relative tick size.

Keywords: price impact, limit order, impulse response function, cointegration

JEL classification: G14, C32, G17

1 Introduction

It is well known that the revelation of trading intention adversely affects asset prices. Passive order placement through limit orders incurs significant market impact even if the order is not been executed. The risk to "scare" and to ultimately shift the market by limit order placements is well-known in financial practice and is taken into account in trading strategies. As a consequence, liquidity provision through hidden order types (such as iceberg orders or hidden orders) has gained popularity in recent years. However, despite the importance of limit order trading in modern markets, the actual impact of an incoming (visible) limit order on the subsequent price process is still not systematically explored and quantified. While the price impact resulting from a trade has been extensively studied by, e.g., Hasbrouck (1991), Dufour and Engle (2000) and Engle and Patton (2004), empirical evidence on the actual market impact of limit order placements is virtually not existent.

This paper aims to fill this gap in the literature and addresses the following empirical research questions: (i) How strong is the short-run and long-run impact of an incoming limit order in dependence of its position in the book, its size and the state of the book? (ii) Are ask and bid quote responses to incoming limit orders widely symmetric or is there evidence for an asymmetric rebalancing of the book? (iii) How different is the market impact of a limit order compared to that caused by a trade of similar size? (iv) How stable are these effects across the market and do they depend on stock-specific characteristics, such as the underlying trading intensity, minimum tick size and average trade size?

We propose modelling the processes of ask and bid quotes as well as several levels of depth volume on both sides of the market in terms of a cointegrated vector autoregressive (VAR) model. This framework allows us to study the price impact of limit orders by means of impulse response functions. Each limit order is represented by a shock disturbing the multivariate system of quotes and depths and influencing it dynamically

over time. Designing the shock vectors in a specific way allows us to characterize the type of the limit order represented by its size and its position in the order queue as well as the current state of the book.

The motivation for using a cointegrating system stems from the fact that ask and bid quotes are naturally integrated and tend to move in locksteps. Cointegration analysis reveals a stationary linear combination of bid and ask quotes which closely resembles the bid-ask spread. The idea of jointly modelling ask and bid quote dynamics in terms of a cointegrated system originates from Engle and Patton (2004) based on the work of Hasbrouck (1991) and has been used in other approaches, such as Hansen and Lunde (2006) and Escribano and Pascual (2006). Our setting extends and modifies this approach in two major respects: Firstly, we model quotes and depth simultaneously. This yields a novel type of order book model capturing not only quote and depth dynamics but implicitly also dynamics of midquotes, midquote returns, spreads, spread changes as well as order book imbalances. Secondly, we model the system not only on a trade-to-trade basis but exploit the complete order arrival process. Therefore, the model captures all relevant trading characteristics in a limit order book market and thus provides a complete description of the order book in a range close to the best quotes. Hence, the model is particularly useful for liquid assets where most of the market activity is concentrated at the best quote levels. In this sense, the approach complements to the dynamic model for complete order book curves introduced by Härdle, Hautsch, and Mihoci (2009).

The proposed quote and depth model is estimated by Johansen's (1991) full information maximum likelihood estimator, using high-frequency order book data for 30 stocks traded on Euronext Amsterdam covering a sample period over two months in 2008. We find strong evidence for the existence of a common stochastic component in quotes and corresponding depths resulting in cointegration relationships which significantly deviate from the bid-ask spread. In this sense, our results shed some light on the strength of co-movements in ask and bid prices depending on the underlying depth. Furthermore, we show that incoming limit orders have significant impacts on subsequent ask and bid processes. It turns out that the magnitude and direction of quote adjustments strongly depend on the order's aggressiveness, its (relative) size and the prevailing depth in the book. In particular, we show the following results: (i) Quote adjustments are the stronger and the faster, the closer the incoming order is posted to the market. Most significant effects are reported for orders posted on up to two levels behind the market. For less aggressive orders virtually no effects can be quantified. (ii) Limit orders temporarily narrow the spread. Converse effects are shown for market

orders. In the long-run, these effects are reverted back in an asymmetric way. (iii) Large limit orders posted inside the spread induce severe long-run effects pushing the market in the intended trading direction. In contrast, small limit orders posted inside the spread tend to be picked up quickly inducing adverse price reactions. (iv) The long run market impact of aggressive market orders walking up (or down, respectively) the book is the higher the smaller the prevailing depth behind the market. (v) The effects are qualitatively stable across the market, where the absolute magnitudes of price impacts differ in dependence of underlying stock-specific characteristics. It turns out that approximately 60%-80% of the cross-sectional variation in market impacts can be explained by the trading frequency and the minimum tick size.

The remainder of this paper is structured as follows. In Section 2, we describe the trading structure of Euronext Amsterdam and provide descriptive statistics. The econometric approach is explained in Section 3. Section 4 gives the estimation results and Section 5 provides the quantified price impacts of different types of limit orders. Finally, Section 6 concludes.

2 Data and Market Environment

The Euronext NSC system is a transparent electronic trading system with price and time order precedence rules. During continuous trading between 9:00 and 17:30 CET, a centralized computer system matches market orders against the best (in terms of price) prevailing limit order on the opposite side of the limit order book. If there is insufficient volume to fully execute the incoming order, the remaining part of the order will be consolidated into the book. Euronext supports various order types, such as pure market orders (order execution without a price limit), stop orders (issuing limit orders or pure market orders when a triggered price is reached) and iceberg orders (displaying only a part of the size in the book). Consolidation of these orders results in sequences of limit and market order submissions or cancellations, respectively.

Our dataset is provided by Deutsche Bank and comprises of trades and limit order activities of the 30 most frequently traded stocks at Euronext Amsterdam between August 1st and September 30th, 2008. Every transaction and every change of the order book are recorded in milliseconds. The data contains information on the prevailing market depth (in terms of the number of shares) for the five best quotes on both sides of the market. Preliminary analyses (which are also supported by the results given in Section 5) show that aggressive limit orders queued close to the best ask and bid quotes have the highest market impact while induced price effects significantly decline with

the distance to the market. Accordingly, we focus only on the best three price levels in the book.

Unlike the trade data which is well filtered by built-in filters in the database[1], the order book data is completely raw. We remove observations where (i) the spread is zero or negative, and (ii) ask or bid quotes change by more than 2%.[2] Moreover, to remove effects due to the opening and closing of the market, we discard the data of the first five and last five minutes of the continuous trading period.

Matching of trade and quote data is achieved by a matching algorithm which is described in detail in Appendix A. This algorithm matches a trade with the corresponding order book observation by comparing its price and volume with the resulting changes of quotes and depths in the book within an adaptively chosen time window. This algorithm minimizes the probability of misclassifications and as a by-product provides an estimate of the time asynchroneity between trade and order book records.[3]

To classify the initiation type of trades, we use a hybrid procedure according to Lee and Ready (1991). Firstly, we determine the type of trades which are located in more than one second time distance to previous trades using the mid-quote method. I.e., if a trade occurs with a price greater (less) than the most current mid-quote, it is classified as buy (sell). If the transaction price equals the mid-quote, it is marked as "undetermined". Secondly, "undetermined" trades and trades which follow previous transactions in less than one second time distance are classified by the tick-test method. Accordingly, if the trade price is higher (lower) than the previous one, it is identified as a buy (sell). If it does not change the price, it is categorized as the same type as the previous one. Finally, we identify sub-transactions arising from the execution of a big market order against several (smaller) limit orders if they occur in less than one second after the previous trade and have the same initiation types. All corresponding sub-transactions are consolidated to a single transaction.

Table 1 gives descriptive statistics of the resulting data used in the paper. We observe significantly more limit order activities than market orders. The average bid-ask spread is the higher the less liquid the underlying stock. Moreover, second level

[1]Besides recording errors, block trades and transactions in auction periods are excluded.

[2]In order to limit the volatility, Euronext NSC suspends continuous trading if prices change by more than 2%. This is not exactly the same rule as that implemented here, but it is reasonably mimicked.

[3]Due to technological progress in the last decades, time delay between trade and quote records is nowadays hardly greater than one second. Consequently, the "five-second" rule according to Lee and Ready (1991), which has been commonly used in empirical market microstructure literature is not appropriate anymore for more recent datasets.

stocks	#trades	#LO activ.	ask			bid			Mean of ask depth			Mean of bid depth		
	per day	per day	min	mean	max	min	mean	max	L1	L2	L3	L1	L2	L3
ING	1606.8	66569.1	20.255	21.518	23.290	20.250	21.507	23.275	3.64	3.94	4.12	3.45	3.90	4.14
RDSa	1166.2	48630.6	21.900	22.991	23.935	21.890	22.981	23.930	4.30	5.21	5.80	4.00	5.06	5.59
FOR	1304.6	27574.0	8.770	9.351	10.160	8.760	9.338	10.150	16.78	25.76	25.03	16.35	26.25	24.20
AEGN	982.5	43270.2	7.290	7.909	8.400	7.280	7.902	8.395	5.12	4.99	4.86	4.98	4.98	4.79
UNc	1152.1	46023.7	17.110	18.635	19.670	17.100	18.625	19.660	4.76	5.24	6.44	4.52	5.33	6.49
ISPA	903.1	35708.2	49.990	52.694	56.440	49.910	52.661	56.420	1.85	2.76	3.66	1.97	3.08	3.84
PHG	1108.3	34722.0	20.875	22.381	23.465	20.870	22.368	23.450	2.18	2.36	2.70	1.95	2.19	2.59
AHLN	1119.4	18730.3	7.540	8.510	8.970	7.530	8.502	8.960	7.89	9.80	10.23	8.18	10.64	10.59
TNT	949.7	20412.7	22.040	24.598	27.000	22.030	24.566	26.970	1.57	1.91	2.15	1.51	1.96	2.24
AKZO	960.0	20061.2	35.460	39.571	41.920	35.400	39.541	41.910	0.89	0.96	1.00	0.78	0.90	0.98
KPN	954.0	20733.8	10.915	11.274	11.680	10.905	11.266	11.670	9.61	12.10	12.77	8.79	10.57	11.57
HEIN	927.2	19782.1	29.540	31.796	33.660	29.520	31.767	33.600	0.98	1.10	1.13	0.92	1.00	1.04
ASML	853.8	26249.5	14.290	15.964	17.400	14.280	15.949	17.390	3.80	5.86	6.50	3.48	5.21	6.01
DSMN	826.7	21574.5	36.050	37.919	40.000	36.020	37.886	39.990	0.77	0.87	0.99	0.77	0.88	0.99
SBMO	603.7	18676.3	13.530	14.934	16.700	13.520	14.911	16.680	1.84	2.63	2.99	1.76	2.51	2.79
RAND	543.4	17265.2	17.710	19.432	21.430	17.690	19.397	21.400	1.09	1.56	1.75	1.07	1.47	1.47
ELSN	488.5	29702.2	10.390	11.049	11.510	10.350	11.035	11.500	7.27	11.57	11.96	6.81	11.34	12.44
WLSNc	548.8	16003.6	14.610	15.973	17.020	14.550	15.950	17.000	1.92	1.88	1.96	1.94	1.83	1.89
FUGRc	505.0	8846.5	43.620	47.701	53.200	43.610	47.631	53.180	0.56	0.54	0.52	0.49	0.49	0.47
TOM2	505.3	16822.0	14.340	16.017	17.550	14.300	15.987	17.540	1.31	1.71	2.06	1.25	1.69	1.75
BOSN	419.6	8013.0	32.320	36.323	41.900	32.250	36.247	41.890	0.52	0.52	0.49	0.53	0.51	0.47
BAMN	416.8	6334.1	9.900	10.736	12.220	9.860	10.714	12.200	2.06	2.35	2.38	1.99	2.25	2.19
SR	347.5	6396.6	10.370	11.588	13.200	10.360	11.563	13.180	1.70	1.80	1.76	1.72	1.71	1.48
CSMNc	340.2	7478.4	17.910	20.395	24.260	17.890	20.361	24.240	0.81	0.88	0.92	0.84	0.90	0.91
COR	327.1	12103.2	47.090	49.273	51.210	47.010	49.175	51.140	0.43	0.41	0.37	0.39	0.38	0.34
SMTNc	272.4	7648.8	43.920	52.282	60.440	43.840	52.112	60.300	0.22	0.25	0.22	0.26	0.27	0.26
IMUN	292.7	5735.9	14.300	16.178	17.710	14.280	16.148	17.700	0.92	1.17	1.24	0.85	0.91	0.88
NUTR	256.6	8043.2	41.160	43.275	44.900	41.120	43.192	44.890	0.40	0.36	0.33	0.37	0.38	0.38
USGP	248.5	6342.3	9.670	11.198	12.630	9.650	11.168	12.600	1.47	1.51	1.41	1.59	1.39	1.19
HEIO	181.0	14011.0	27.120	29.854	31.300	27.080	29.809	31.290	0.44	0.53	0.61	0.50	0.64	0.70

Table 1: Summary of synchronized trade and order book data. The stocks are sorted according to their trading frequencies. Market depth is measured in thousand shares. Trading at Euronext Amsterdam in August 2008. L1-L3 denote the order book level one to three.

market depth is higher than first level depth while it is approximately equal to the third level.

3 Econometric Modelling

3.1 A Cointegrated VAR Model for Quotes and Depths

Denote t as a (business) time index, indicating all order book activities, i.e., incoming limit or market orders as well as limit order cancellations. Furthermore, p_t^a and p_t^b define the best log ask and bid quotes instantaneously after the t-th order activity. Moreover, $v_t^{a,j}$ and $v_t^{b,j}$ for $j = 1, \ldots, k$, denote the log depth on the j-th best observed quote level on the ask and bid side, respectively.

To capture the high-frequency dynamics in quotes and depths we define a $K(= 2 + 2\times k)$-dimensional vector of endogenous variables $y_t := [p_t^a, p_t^b, v_t^{a,1}, \ldots, v_t^{a,k}, v^{b,1}, \ldots, v_t^{b,k}]'$. The quote levels associated with $v_t^{a,j}$ and $v_t^{b,j}$ are not observed on a *fixed* grid at and behind the best quotes. Hence, their price distance to p_t^a and p_t^b is not necessarily exactly $j-1$ ticks but might be higher if there are no limit orders on all possible price levels behind the market. Consequently, we only exploit the information that $v_t^{a,j}$ and $v_t^{b,j}$ are the depths of the currently observed j-th best price level and ignore information about their actual price distance to p_t^a and p_t^b. Two reasons justify this proceeding: Firstly, for liquid assets, gaps in the price grids around the best quotes do not occur very often and are negligible. Hence in this case, level j mostly corresponds to a distance of $j-1$ ticks to the corresponding best quote. Secondly, incorporating not only the market depth on the individual levels but also the corresponding price information would significantly increase the dimension of the underlying system and would complicate our analysis without providing substantial additional insights.

Modelling log volumes instead of plain volumes is a common practice in many empirical studies to reduce the impact of extraordinarily large volumes. This is also suggested by Potters and Bouchaud (2003) studying the statistical properties of market impacts of trades. Moreover, using logs implies that changes in market depth can be interpreted as *relative* changes with respect to the current depth level.

Hence, we model log quotes and log depths as a cointegrated VAR(p) model augmented by s lags of exogenous variables – henceforth VARX(p, s) model – with the vector error correction (VEC) form

$$\Delta y_t = \mu + \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \sum_{j=1}^{s} B_j x_{t-j} + u_t, \tag{1}$$

where μ is a constant, α and β denote the $K \times r$ loading and cointegrating matrices with $r < K$, and Γ_i, $i = 1, \ldots, p-1$, is a $K \times K$ parameter matrix. The vector $x_t = [BUY_t, SELL_t]'$ denotes a 2×1 vector of dummy variables indicating the occurrence of a buy or sell trade, respectively, with corresponding parameter vector B_j, $j = 1, \ldots, s$. The inclusion of x_t is necessary in order to be able to distinguish between the effects caused by a market order and that induced by a cancellation. Both events remove volume from the book, however, presumably have quite different long run market impacts.

Note that we endogenize only quotes and depths but not order choice decisions themselves. Including the latter would significantly increase the complexity of the model and would make the cointegration analysis more difficult without yielding significantly more insights given the objective of our study. Hence, the model can be seen as a reduced form description of the dynamics of quotes and depths caused by an arriving order. Consequently, we treat x_t as a weakly exogenous variable.

The noise term u_t is assumed to be serially uncorrelated with zero mean and covariance Σ_u. Since limit orders placed inside of the spread and large market orders "walking down" or "up" the order book imply a simultaneous change of both quotes and depths, the covariance matrix Σ_u is obviously not diagonal. Table 2 summarizes the definition of these variables.

Variable	Description
p_t^a	Log ask quote after the arrival of the t-th order.
p_t^b	Log bid quote after the arrival of the t-th order.
$v_t^{a,l}$	Log depth at the l-th best ask price after the arrival of the t-th order.
$v_t^{b,l}$	Log depth at the l-th best bid price after the arrival of the t-th order.
BUY_t	Buy dummy, equal to one if the t-th order is a buy trade.
$SELL_t$	Sell dummy, equal to one if the t-th order is a sell trade.

Table 2: Variable definitions

For the impulse-response analysis below, it turns out to be more convenient to work with the reduced VARX form of model (1)

$$y_t = \mu + \sum_{i=1}^{p} A_i y_{t-i} + \sum_{j=1}^{s} B_j x_{t-j} + u_t, \tag{2}$$

where $A_1 := I_K + \alpha\beta' + \Gamma_1$ with I_K denoting a $K \times K$ identity matrix, $A_i := \Gamma_i - \Gamma_{i-1}$ with $1 < i < p$ and $A_p := -\Gamma_{p-1}$.

While representation (1) is a model for (relative) *changes* in quotes and depths, specification (2) is a model for quote and depth *levels*. Obviously, model (2) can be further rotated in order to represent dynamics in spreads, relative spread changes, midquotes, midquote returns as well as (ask-bid) depth imbalances. Hence, the model is sufficiently flexible to capture the high-frequency dynamics of all relevant trading variables. In models involving only quote dynamics (see, e.g., Engle and Patton 2004) or spread dynamics (see, e.g., Lo and Sapp 2006), the error correction term $\beta' y_t$ is typically assumed to be equal to the spread implying a linear restriction $R'\beta = 0$ with $R' = [1, 1, 0, \ldots, 0]$. Note that we do *not* impose this assumption here. As depth contains information on the equilibrium (long run) state of the order book as well, we expect the existence of stationary processes which are linear combinations of both quotes *and* depths.

Model (1) is estimated by the Full Information Maximum Likelihood (FIML) estimator proposed by Johansen (1991) and Johansen and Juselius (1990). Let $z_{0t} := \Delta y_t$, and $z_{1t} := y_{t-1}$. Further let z_{2t} be the vector of stacked variables,

$$z_{2t} := (\Delta y_{t-1}, \cdots, \Delta y_{t-p+1}, x_{t-1}, \cdots, x_{t-s}, 1)'$$

with corresponding parameter vector $\Gamma = (\Gamma_1, \ldots, \Gamma_{p-1}, B_1, \ldots, B_s, \mu)$. Define the product moment matrices

$$M_{ij} := T^{-1} \sum_{t=1}^{T} z_{it} z_{jt}', \quad i, j = 0, 1, 2,$$

where T is the number of observations. Moreover, let

$$S_{ij} := M_{ij} - M_{i2} M_{22}^{-1} M_{2j}.$$

We then solve the generalized eigenvalue problem

$$\left| \lambda S_{11} - S_{10} S_{00}^{-1} S_{01} \right| = 0$$

for the eigenvalues $1 > \hat{\lambda}_1 > \cdots > \hat{\lambda}_K > 0$ and corresponding eigenvector $\hat{V} = (\hat{v}_1, \cdots, \hat{v}_K)$ which is normalized by $\hat{V} S_{11} \hat{V} = I_K$. Johansen's (1991) trace test or maximum eigenvalue test can be used to determine the underlying cointegration rank r. Under the hypothesis that there exist r cointegration relationships, the $K \times r$ cointegration matrix β is estimated by

$$\hat{\beta} = (\hat{v}_1, \ldots, \hat{v}_r)$$

with corresponding maximized log-likelihood function

$$l_{\max}(\hat{\beta}) = -\frac{T}{2}\left(\ln|S_{00}| + \sum_{i=1}^{r}\ln(1-\hat{\lambda}_i)\right). \tag{3}$$

The magnitude of $\hat{\lambda}_i$ can be interpreted as a measure of the "stationarity" of the product $\hat{\beta}_i' y_t$. The larger $\hat{\lambda}_i$, the closer the stochastic properties of the underlying relationship to that of a stationary process. The parameters α and Γ are estimated by OLS after inserting $\hat{\beta}$ into equation (1) and computing $\hat{\Sigma}_u$ as $\hat{\Sigma}_u = S_{00} - \hat{\alpha}\hat{\alpha}'$.

Following Lütkepohl and Reimers (1992), the parameters of equation (1) can be easily transformed to equation (2). In this context, we define a transformation matrix

$$D = \underbrace{\left[\begin{array}{cccccc|ccccc} I_K & 0 & 0 & \cdots & 0 & 0 & & & & & \\ I_K & -I_K & 0 & \cdots & 0 & 0 & & & & & \\ 0 & I_K & -I_K & \ddots & 0 & 0 & & & & & \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & & & 0 & & \\ 0 & 0 & 0 & \ddots & 0 & 0 & & & & & \\ 0 & 0 & 0 & \cdots & -I_K & 0 & & & & & \\ 0 & 0 & 0 & \cdots & I_K & -I_K & & & & & \\ \hline & & & & & & I_2 & 0 & \cdots & 0 & 0 \\ & & & 0 & & & 0 & I_2 & & 0 & 0 \\ & & & & & & \vdots & & \ddots & 0 & 0 \\ & & & & & & 0 & 0 & \cdots & I_2 & 0 \\ & & & & & & 0 & 0 & \cdots & 0 & 1 \end{array}\right]}_{(KP+2s+1)\times(KP+2s+1)}$$

such that

$$[A_1, \cdots, A_p, B_1, \cdots, B_s, \mu] = [\alpha\beta', \Gamma]D + J^*, \tag{4}$$

where $J^* := [I_K : 0 : \cdots : 0]$ is a $K \times (Kp+2s+1)$ matrix. The theorem below provides a consistent estimator of A and B:

Theorem 1 (Lütkepohl and Reimers, 1992)**.** *Let $\hat{\alpha}$, $\hat{\beta}$, $\hat{\Gamma}$ and $\hat{\Sigma}_u$ denote the FIML estimates of the parameters of model* (1)*. Moreover, $\hat{A}_1, \cdots, \hat{A}_p, \hat{B}_1, \cdots, \hat{B}_s$ are computed by the transformation in* (4)*. Then,*

$$\sqrt{T}\left[\text{vec}(\hat{A}_1, \cdots, \hat{A}_p, \hat{B}_1, \cdots, \hat{B}_s, \hat{\mu}) - \text{vec}(A_1, \cdots, A_p, B_1, \cdots, B_s, \mu)\right] \xrightarrow{d} \mathcal{N}(0, \Sigma_{AB}), \tag{5}$$

where

$$\Sigma_{AB} = D' \begin{bmatrix} \beta & 0 \\ 0 & I_{K(p-1)+2s+1} \end{bmatrix} \Omega^{-1} \begin{bmatrix} \beta' & 0 \\ 0 & I_{K(p-1)+2s+1} \end{bmatrix} D \otimes \Sigma_u,$$

$$\Omega = \operatorname{plim} \frac{1}{T} \begin{bmatrix} \beta' M_{11} \beta & \beta' M_{12} \\ M_{21} \beta & M_{22} \end{bmatrix}$$

are consistently estimated by

$$\hat{\Sigma}_{AB} = D' \begin{bmatrix} \hat{\beta} & 0 \\ 0 & I_{K(p-1)+2s+1} \end{bmatrix} \hat{\Omega}^{-1} \begin{bmatrix} \hat{\beta}' & 0 \\ 0 & I_{K(p-1)+2s+1} \end{bmatrix} D \otimes \hat{\Sigma}_u$$

$$\hat{\Omega} = \begin{bmatrix} \hat{\beta}' M_{11} \hat{\beta} & \hat{\beta}' M_{12} \\ M_{21} \hat{\beta} & M_{22} \end{bmatrix}.$$

Proof. See Lütkepohl and Reimers (1992) by noting that their proof still holds with additional exogenous variables. □

Linear restrictions on β can be tested by the likelihood ratio test proposed by Johansen (1991). Consider, for instance, the restriction $R'\beta = 0$ with $R' = [1, 1, 0, \ldots, 0]$ implying the bid-ask spread as cointegration relationship. By defining

$$H = R_\perp = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ -1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ddots & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}}_{K \times (K-1)},$$

where $a_\perp$ denotes the basis of the null space of a', the restriction becomes

$$\mathcal{H}_0 : \ \beta = H\varphi,$$

with the $(K-1) \times r$ matrix φ denoting the parameter vector. Under this hypothesis, φ can be estimated by solving

$$\left| \lambda^* H' S_{11} H - H' S_{10} S_{00}^{-1} S_{01} H \right| = 0$$

and collecting eigenvectors associated with the first r largest eigenvalues $\hat{\lambda}_1^* > \cdots > \hat{\lambda}_r^*$. The corresponding likelihood ratio test statistic is given by

$$T \sum_{i=1}^{r} \ln \left((1 - \hat{\lambda}_i^*) / (1 - \hat{\lambda}_i) \right),$$

which is asymptotically χ^2-distributed with r degrees of freedom.

3.2 Limit Orders as Shocks to the System

In this section, we illustrate how to represent incoming orders as shocks to the system specified in equation (2). Whenever an order enters the order book, it (i) will change the depth in the book, (ii) may change the best quotes depending on which position in the queue it is placed, and (iii) will change the trade indication dummy in case of a market order. We represent these changes in terms of an impulse vector $\delta := [\delta_v', \delta_p', \delta_x']'$ with δ_v being a $2k \times 1$ vector associated with shocks to the depths, δ_p denoting a 2×1 vector consisting of shocks to the quotes and δ_x being a 2×1 vector representing shocks to the trade indication dummy.

Note that in some situations, one side of the order book may be completely "shifted" by an incoming order. For example, a bid limit order posted inside the spread improves the bid quote and thus establishes a new best price level. As a consequence, all volumes on the bid side are simultaneously shifted by one quote level.[4]

We design impulse vectors associated with five scenarios commonly faced by market participants. As graphically illustrated by Figures 1 to 4, a three-level order book is initialized by the best ask quote $p_t^a = 1002$, best bid quote $p_t^b = 1000$, second best ask quote 1003, second best bid quote 999, and levels of depths on the bid side $V_t^{b,1} = 1$, $V_t^{b,2} = 1.5$, $V_t^{b,3} = V_t^{b,4} = 1.4$. The following scenarios are considered:[5]

Scenario 1a (normal limit order): Arrival of a bid limit order with price 1000 and size 0.5 to be placed *at the market*, i.e. posted at the best bid quote. As shown in Figure 1, this order will be consolidated at the best bid without changing the prevailing quotes. Because the initial depth on the first level is assumed to be 1.0, the change of the log depth is $\ln(1.5) \approx 0.4$. Correspondingly, the shock vectors are given by $\delta_v = [0, 0, 0, 0.4, 0, 0]'$, $\delta_p = \delta_x = [0, 0]'$.

Scenario 1b (passive limit order): Arrival of a bid limit order with price 999 and size 0.5 to be posted *behind the market*, i.e. its limit price is smaller than the current best bid quote. As in the scenario above, it does not change the prevailing quotes and only affects the depth. Because the initial depth on the second level is 1.5, the log depth change is $\ln(1 + 0.5/1.5) \approx 0.29$. Consequently, we have $\delta_v = [0, 0, 0, 0, 0.29, 0]'$, $\delta_p = \delta_x = [0, 0]'$.

[4] An exception occurs whenever the depth of the order book is uniformly distributed. In this case, the incoming order only "shocks" the depth at the best quote. However, this scenario is quite unrealistic.

[5] For sake of brevity, the scenarios are only characterized for the bid side. For ask orders, the setting is correspondingly shifted to the other side of the market.

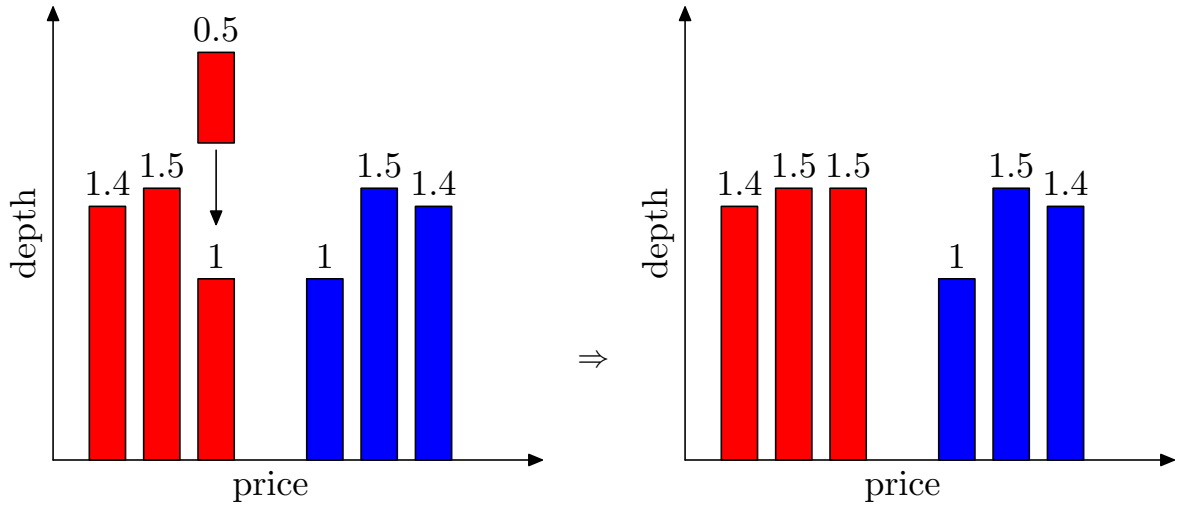


Figure 1 (Scenario 1a (normal limit order)): An incoming bid limit order with price 1000 and size 0.5. It affects only the depth at the best bid without changing the prevailing quotes or resulting in a trade. The underlying shock vectors are $\delta_v = [0, 0, 0, 0.4, 0, 0]'$ and $\delta_p = \delta_x = [0, 0]'$.

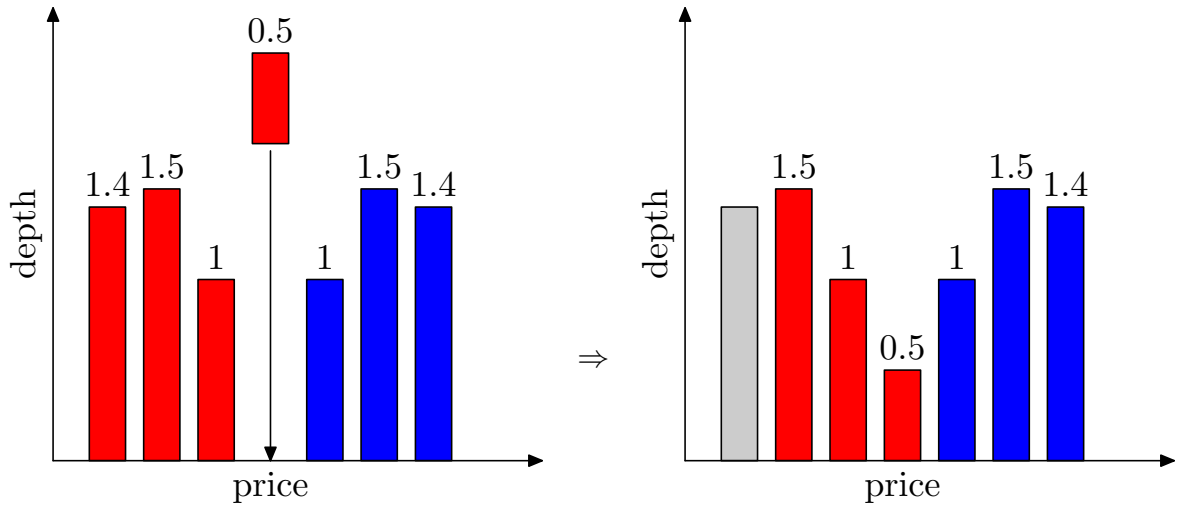


Figure 2 (Scenario 2 (aggressive limit order)): An incoming bid limit order with price 1001 and size 0.5 improving the bid quote and changing all depth levels on the bid side of the order book. The underlying shock vectors are $\delta_v = [0, 0, 0, -0.69, -0.4, 0.07]'$, $\delta_p = [0, 0.001]'$ and $\delta_x = [0, 0]'$.

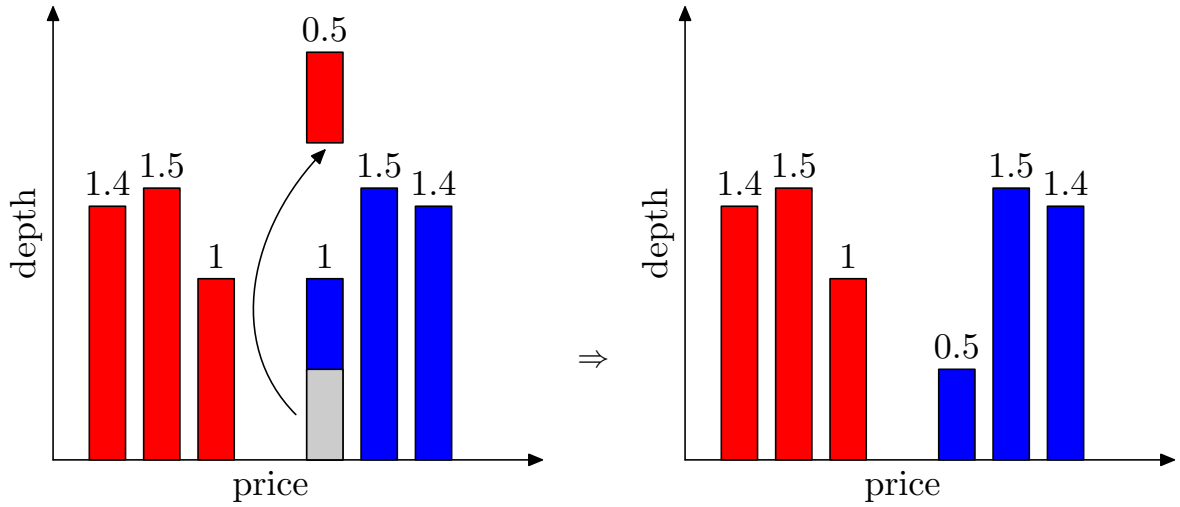


Figure 3 (Scenario 3 (normal market order)): An incoming bid (buy) market order with price 1002 and size 0.5 which results in a buy transaction. The underlying shock vectors are $\delta_v = [-0.69, 0, 0, 0, 0, 0]'$, $\delta_p = [0, 0]'$ and $\delta_x = [1, 0]'$.

Scenario 2 (aggressive limit order): Arrival of a bid limit order with price 1001 and size 0.5 to be posted inside of the current spread. Figure 2 shows that it improves the best bid by 0.1% and accordingly shifts all depth levels on the bid side. The resulting shock vector is given by $\delta_v = [0, 0, 0, (\log(0.5) \approx -0.69), (\ln(1/1.5) \approx -0.4), (\ln(1.5/1.4) \approx 0.07)]'$, $\delta_p = [0, 0.001]'$ and $\delta_x = [0, 0]'$.

Scenario 3 (normal market order): Arrival of a bid order with price 1002 and size 0.5. This order will be executed immediately against pending limit orders at the best ask and thus results in a buy market order. Because it absorbs liquidity from the book, it shocks the corresponding depth levels negatively. Figure 3 depicts the corresponding changes of the order book as represented by $\delta_v = [\ln(0.5) \approx -0.69, 0, 0, 0, 0, 0]'$, $\delta_p = [0, 0]'$ and $\delta_x = [1, 0]'$.

Scenario 4 (aggressive market order): Arrival of a bid order with price 1003 and size 1.2. We refer this to an "aggressive" market order because it "walks up" the order book. Correspondingly, the best ask quote and all depth levels are simultaneously shifted resulting in the shock vector $\delta_v = [(\ln(1.3) \approx 0.26), (\ln(1.4/1.5) \approx -0.07), 0, 0, 0, 0]'$, $\delta_p = [(1/1002) \approx 0.001, 0]'$ and $\delta_x = [1, 0]'$.

Table 3 summarizes the shock vectors implied by the different scenarios.

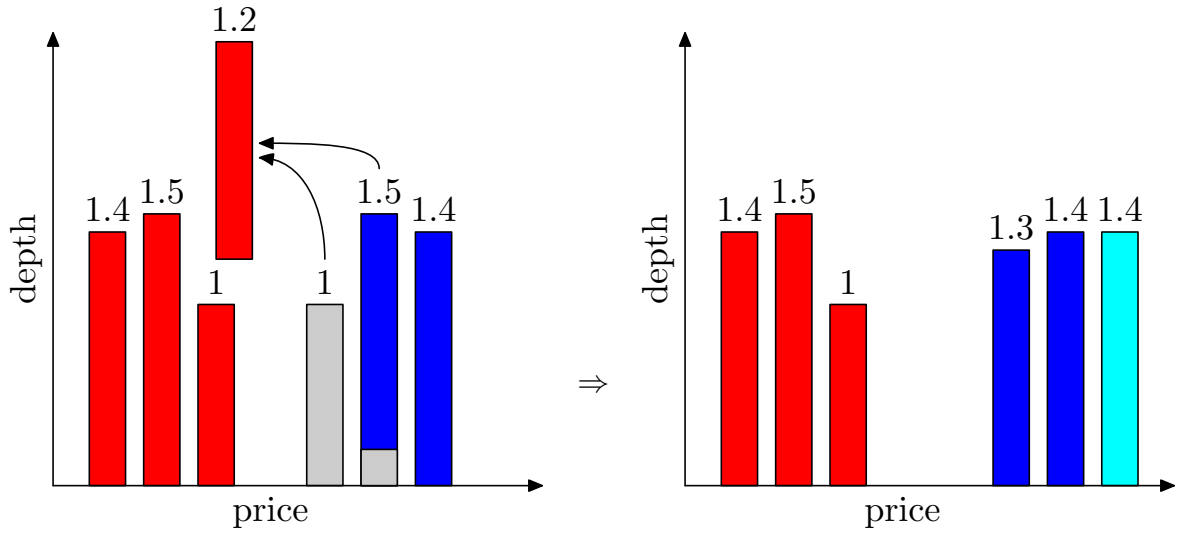


Figure 4 (Scenario 4 (aggressive market order)): An incoming bid (buy) market order with price 1003 and size 1.2 "walking up" the order book and simultaneously changing all depth levels on the bid side. The underlying shock vectors are $\delta_v = [0.26, -0.07, 0, 0, 0, 0]'$, $\delta_p = [0.001, 0]'$ and $\delta_x = [1, 0]'$.

3.3 Measuring the Market Impact

We quantify the market impact of incoming limit orders by the implied expected short-run and long-run shift of ask and bid quotes. This reaction is quantified by the impulse response function,

$$f(h; \delta_y, \delta_x) = \mathrm{E}[y_{t+h} | y_t + \delta_y, x_t + \delta_x, y_{t-1}, \cdots, x_{t-1}, \cdots] - \mathrm{E}[y_{t+h} | y_t, x_t, y_{t-1}, \cdots, x_{t-1}, \cdots], \tag{6}$$

where the shock on quotes and depths in the order book is denoted by $\delta_y := [\delta_p', \delta_v']'$ and h is the number of periods (measured in "order event" time).

Note that we do not have to orthogonalize the impulse since contemporaneous relationships between quotes and depths are captured by construction of the shock vector. Moreover, our data is based on the arrival time of orders avoiding time aggregation as another source of mutual dependence in high-frequency order book data.

Using impulse-response analysis to retrieve the market impact has two major advantages. First, in contrast to an analysis of estimated VEC coefficients which only reveals the immediate impact, it enables us to examine both long-run and short-run effects. Second, it allows us to straightforwardly quantify the joint effect induced by simultaneous changes of several variables given a certain state of other variables.

We consider two moving average (MA) representations of the cointegrated VARX

Scenario	limit order	shock vectors		
	(dir,price,size)	δ_v'	δ_p'	δ_x'
"normal limit order"	(Bid,1000, 0.5)	$[0, 0, 0, 0.4, 0, 0]$	$[0, 0]$	$[0, 0]$
"passive limit order"	(Bid,999, 0.5)	$[0, 0, 0, 0, 0.29, 0]$	$[0, 0]$	$[0, 0]$
"aggressive limit order"	(Bid,1001, 0.5)	$[0, 0, 0, -0.69, -0.4, 0.07]$	$[0, 0.001]$	$[0, 0]$
"normal market order"	(Bid,1002, 0.5)	$[-0.69, 0, 0, 0, 0, 0]$	$[0, 0]$	$[1, 0]$
"aggressive market order"	(Bid,1003, 1.2)	$[0.26, -0.07, 0, 0, 0, 0]$	$[0.001, 0]$	$[1, 0]$

Table 3: Shock vectors implied by the underlying five scenarios. Initial order book: $p_t^a = 1002$, $p_t^b = 1000$, second best ask price = 1003, second best bid price = 999, $V_t^{b,1} = 1$, $V_t^{b,2} = 1.5$, $V_t^{b,3} = V_t^{b,4} = 1.4$.

model. The first one is based on the reduced form given by equation (2). This representation allows us to compute the path of the response function over time. The second one is the Granger representation based on the VECM form in equation (1) which enables us to explicitly compute the permanent (long-run) response.

We start our discussion with the first MA representation. The companion VARX$(1, 1)$ form of the VARX(p, s) model in equation (2) is given by

$$Y_t = \boldsymbol{\mu} + \mathbf{A}Y_{t-1} + \mathbf{B}x_t + U_t, \tag{7}$$

where

$$\boldsymbol{\mu} := \underbrace{\begin{bmatrix} \mu \\ 0 \\ \vdots \\ 0 \\ -- \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{(Kp+2s)\times 1}, \; Y_t := \underbrace{\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \\ -- \\ x_t \\ \vdots \\ x_{t-s+1} \end{bmatrix}}_{(Kp+2s)\times 1}, \; U_t := \underbrace{\begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \\ -- \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{(Kp+2s)\times 1}, \; \mathbf{B} := \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ -- & -- \\ I_2 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}}_{(Kp+2s)\times 2}$$

and

$$\mathbf{A} := \underbrace{\left[\begin{array}{cccc|cccc} A_1 & \cdots & A_{p-1} & A_p & B_1 & \cdots & B_{s-1} & B_s \\ I_K & & 0 & 0 & 0 & \cdots & 0 & 0 \\ & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_K & 0 & 0 & \cdots & 0 & 0 \\ \hline & & & & 0 & \cdots & 0 & 0 \\ & & & & I_2 & \cdots & 0 & 0 \\ & & 0 & & & \ddots & \vdots & \vdots \\ & & & & 0 & \cdots & I_2 & 0 \end{array}\right]}_{(Kp+2s)\times(Kp+2s)}.$$

Successively substituting Y yields

$$Y_t = M_t + \sum_{i=0}^{t-1} \mathbf{A}^i \mathbf{B} x_{t-i} + \sum_{i=0}^{t-1} \mathbf{A}^i U_{t-i}, \tag{8}$$

where $M_t = A^t Y_0 + \sum_{i=0}^{t} A^i \boldsymbol{\mu}$ consists of terms of an initial value and a deterministic trend, which are irrelevant for the impulse-response analysis. Let $J := [I_K : 0 : \cdots : 0]$ be a $(K \times (Kp + 2s))$ selection matrix with $JY_t = y_t$. Pre-multiplying J on both sides of equation (8) gives

$$\begin{aligned} y_t &= JM_t + \sum_{i=0}^{t-1} J\mathbf{A}^i \mathbf{B} x_{t-i} + \sum_{i=0}^{t-1} J\mathbf{A}^i J' u_{t-i} \\ &= JM_t + \sum_{i=0}^{t-1} J\mathbf{A}^i [\mathbf{B} : J'] \begin{bmatrix} x_{t-i} \\ u_{t-i} \end{bmatrix}. \end{aligned} \tag{9}$$

Then, the linear impulse-response function according to equation (6) can be written as

$$f(h; \delta_y, \delta_x) = J\mathbf{A}^h [\mathbf{B} : J'] \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix}. \tag{10}$$

Given the consistent estimator specified in equation (5), the asymptotic distribution of the impulse-response function is obtained using the Delta method and is given by

$$\sqrt{T}(\hat{f} - f) \xrightarrow{d} \mathcal{N}(0, G_h \Sigma^*_{AB} G_h'), \tag{11}$$

where $G_h := \partial \operatorname{vec}(f) / \partial \operatorname{vec}(A_1, \cdots, A_p, B_1, \cdots, B_q)'$ and Σ^*_{AB} is the top-left $K(Kp + 2s) \times K(Kp + 2s)$ block of Σ_{AB}. As shown in the Appendix, G_h can be explicitly written as

$$G_h = \sum_{i=0}^{h-1} \left(\begin{bmatrix} \delta_x' & \delta_y' \end{bmatrix} \begin{bmatrix} \mathbf{B}' \\ J \end{bmatrix} (\mathbf{A}')^{h-1-i} \otimes J\mathbf{A}^i J' \right). \tag{12}$$

In order to compute the long-run effect, we apply Granger's Representation Theorem to model (1) yielding

$$y_t = C \sum_{i=1}^{t} \left(u_i + \sum_{j=1}^{s} B_j x_{i-j} + \mu \right) + C_1(L) \left(u_t + \sum_{j=1}^{s} B_j x_{t-j} + \mu \right) + V, \tag{13}$$

where

$$C = \beta_\perp \left(\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right)^{-1} \alpha'_\perp \, . \tag{14}$$

Here, L is the lag operator and the power series $C_1(z)$ is convergent for $|z| < 1 + \xi$ for some $\xi > 0$. V depends on initial values, such that $\beta' V = 0$. The Granger representation decomposes the cointegrated process into a random walk term (C term), a stationary process (C_1 term) and a deterministic term (V). Because of the convergence of the series $C_1(z)$, the response implied by this sub-process will be zero in the long run. Moreover, since the deterministic term V is irrelevant for the impulse response, the permanent response of the system is determined by the first term in equation (13). Note that the shock (δ_y, δ_x) causes this term changing by $C(\delta_y + \sum_{j=1}^{s} B_j \delta_x)$. Thus, we can express the permanent response as

$$\bar{f}(\delta_x, \delta_y) := \lim_{h \to \infty} f(h; \delta_y, \delta_x) = C \left[\sum_{j=1}^{s} B_j : I_K \right] \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix}. \tag{15}$$

Note that given α and β, $\alpha_\perp$ and $\beta_\perp$ are not uniquely identified. However, the right hand side of equation (14) is invariant with respect to the choice of these bases. Therefore, $\bar{f}(\delta_x, \delta_y)$ is unique given the parameters and the shock vector in model (1). In practice, estimated responses and their covariances are obtained by replacing the unknown parameters in equation (10), (11) and (15) by their estimates.

4 Estimation Results

The underlying order book data contains bid and ask quotes as well as five levels of depth. Preliminary analyzes show that the depths on the fourth and fifth levels do not have significant effects on bid and ask quotes. Therefore, in our empirical study, we only use market depths up to the third level. In order to make the analysis tractable, we reduce the computational burden induced by the high number of observations by separately estimating the model for each of the 43 trading days. This strategy allows us also to address possible structural changes, e.g., due to stock specific news announcements or overnight effects. The market impact is then computed as the monthly average of

individual (daily) impulse response functions. To account for a structural break on September 1, 2008, due to the change of the tick size for some stocks in our sample, we treat the two months August and September separately.

For sake of brevity we refrain from presenting all individual results for the 30 analyzed stocks in this paper. We rather illustrate the analyzed effects for the stock Fortis (FOR in Table 1) in August 2008. Fortis is one of the most actively traded stocks and is representative for a major part of the market. The results for the remaining stocks and the remaining periods are provided in a web appendix on `http://amor.cms.hu-berlin.de/~huangrui/project/impact_of_orders/`. As one can see in the web appendix and discussed in more detail in Section 5.5, the effects are qualitatively remarkably similar across the market though the magnitudes of market impacts differ in dependence of underlying stock-specific characteristics.

The following estimation results are based on a VARX$(15, 15)$ specification which is selected based on residual diagnostics and information criteria. Testing for serial correlation using the Ljung-Box test according to Ljung and Box (1978) reveals almost no remaining serial correlation in the residuals for all regressions based on a 1% level using ten lags. The corresponding statistics are also recorded in the web appendix.

4.1 Statistical Properties of Market Depth

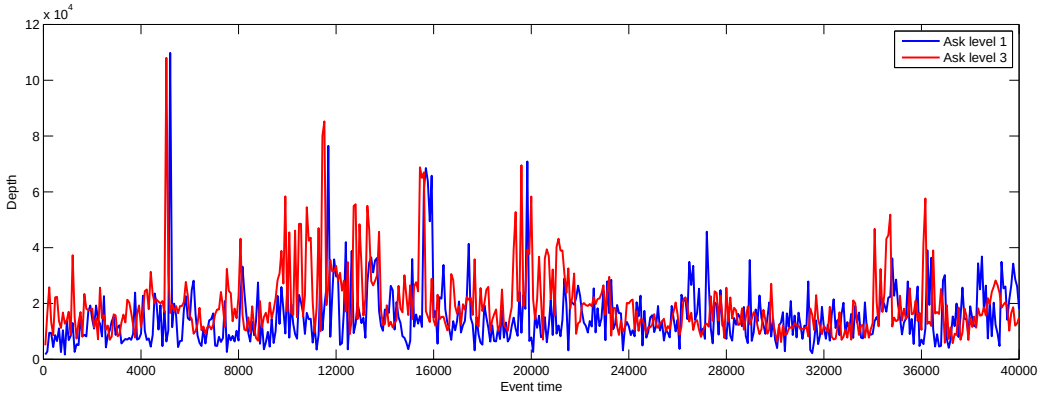


Figure 5: Time series of market depth in the order book. Trading of Fortis, Euronext, Amsterdam, August 1st, 2008.

Figure 5 provides time series plots of depths on the best ask and third best ask level of the order book for a single (though representative) trading day for Fortis. A general finding is that the depth behind the market is typically greater than that at the market. Furthermore, there is evidence for co-movements between the individual

depth levels, partially because of the "shift" effect induced by aggressive orders, e.g., market orders who completely absorb the best price levels.

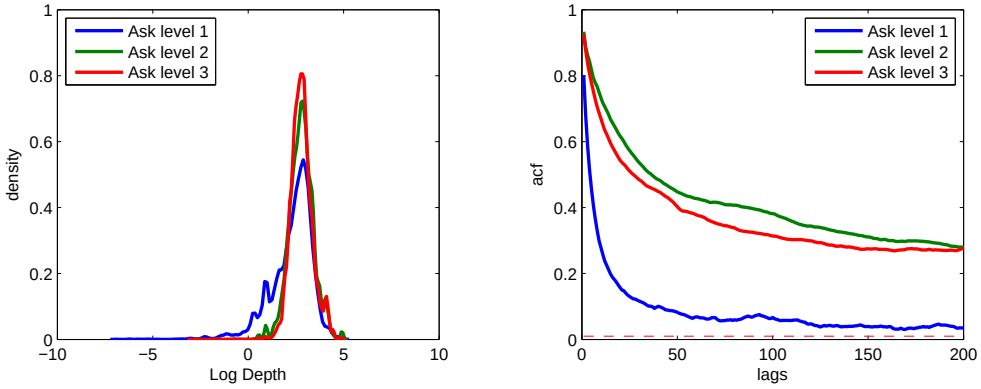


Figure 6: Left: Kernel density estimates of market depths. **Right:** Autocorrelation functions of market depths. Trading of Fortis, Euronext, Amsterdam.

Figure 6 depicts the unconditional distributions and autocorrelation functions of log market depth. We observe that the distributions of depths behind the market are similar, though they are quite different from those at the market. The same pattern is also observed for the autocorrelation functions. These empirical peculiarities are obviously due to the fact that there is more order activity at the market than behind the market. Consequently, market depth is more frequently changed at the best level inducing a lower persistence than at higher levels. This might also explain why the unconditional distribution of depth is more dispersed than that of depth behind the market.

4.2 Estimated Cointegration Relationships

For sake of brevity, we refrain from showing the individual estimates of $\mathbf{A}$ and $\mathbf{B}$. Ultimately, the effects induced by $\mathbf{A}$ and $\mathbf{B}$ are revealed by the impulse response analysis shown below. Nevertheless, it is interesting to highlight the estimated cointegration relationships. According to Johansen's trace statistics we identify seven cointegration relationships. Table 4 shows the estimated cointegrating vectors for a representative trading day. They are ordered according to their corresponding eigenvalues reflecting their contributions to the likelihood function. Figure 7 depicts the time series of the corresponding cointegration relationships. It turns out that the estimated cointegration relationships are quite different from the simple difference between ask and bid quotes yielding the bid-ask spread and shown in Figure 7. Compared to the spread

Variable	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
p^a	-0.9982	1.0000	-1.0000	-0.9999	-1.0000	1.0000	0.9457
p^b	1.0000	-0.9864	0.9978	1.0000	0.9837	-0.6954	-1.0000
$v^{a,1}$	-0.0205	-0.1328	0.0398	0.0285	0.0692	-0.0976	-0.0746
$v^{a,2}$	0.0078	0.0396	-0.0344	-0.0664	0.1399	-0.6558	-0.3732
$v^{a,3}$	-0.0073	-0.0102	0.0267	0.0143	-0.2263	-0.6543	-0.3146
$v^{b,1}$	-0.0081	0.1334	0.0339	0.0635	0.0392	0.0863	-0.0652
$v^{b,2}$	0.0002	-0.0462	-0.0556	0.1328	-0.0207	0.8649	-0.2855
$v^{b,3}$	0.0000	0.0288	0.0367	-0.1859	-0.0558	0.9881	-0.2033

Table 4: Representative estimates of the cointegrating vectors. The vectors are sorted according to their corresponding eigenvalues. Trading of Fortis at Euronext, Amsterdam.

which reflects a very discrete behavior, the cointegration relationships are much more "smooth". We also tested whether the estimated cointegration relationships are indeed different from the bid-ask spread, i.e., $R'\beta = 0$ with $R = [1, 1, 0, \ldots, 0]'$. The corresponding likelihood ratio test as described in Section 3.1 rejects this hypothesis at 1% significance level for all regressions (except one).

Interpreting the estimated cointegrating vectors we can derive several interesting implications. The first five cointegration relationships are mostly linear combinations of spreads and depths. Specially, the first one is quite similar to the pure bid-ask spread since the coefficients for the depth variables are comparably small. The second cointegration relationship seems to involve the balance of market depth since the coefficients of $v^{a,1}$ and $v^{b,1}$ are similar in magnitude and opposite in sign. The most interesting relationships are implied by the last two cointegrating vectors in which the coefficients associated with the quotes are quite different and relatively large. This indicates that depth has a significant impact on the long-term relationship between quotes. Intuitively, the connection between ask and bid quotes becomes weaker (and thus deviates from the spread) if the depth is less balanced between both sides of the market. Hence, depth has a significant impact on quote dynamics and should be explicitly taken into account in a model for quotes. These findings support the idea of a cointegration model for *both* quotes and depth.

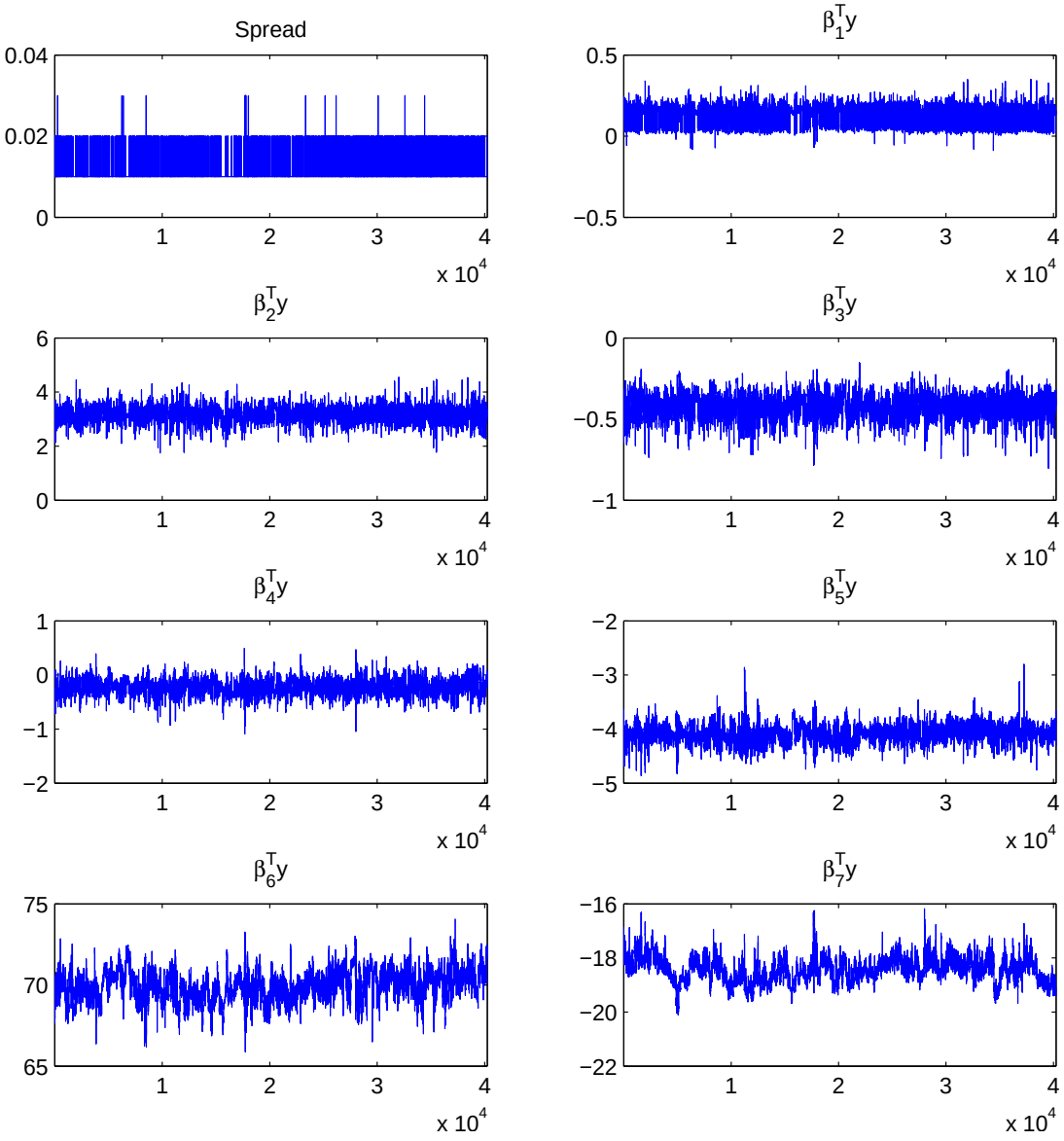


Figure 7: Time series of estimated cointegration relationships. The corresponding cointegrating vectors are documented in Table 4. Trading of Fortis at Euronext, Amsterdam, August 1st, 2008.

5 Estimated Market Impact

5.1 Limit Orders Placed At or Behind the Market

We start by considering the impact of an incoming at-the-market limit order as described in Scenario 1 in Section 3.2. Figure 8 shows the impulse responses induced by ask and bid limit orders with a size equal to half of the depths on their corresponding

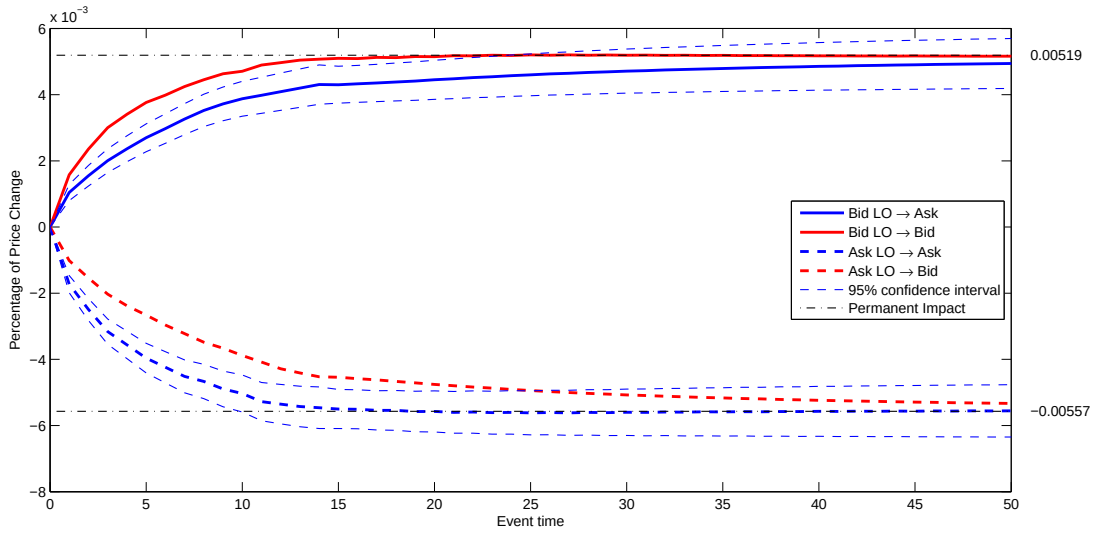


Figure 8: Percentage changes of ask and bid quotes induced by incoming bid/ask limit orders placed at the market (level one) with a size equal to the half of the depth on the first level. The marked number on the vertical axes indicates the magnitude of the permanent impact. The blue dotted lines indicate the corresponding 95%-confidence intervals. Trading of Fortis at Euronext, Amsterdam. LO: limit order.

best levels.[6] The impulse response function starts at zero since such a limit order does not directly change best ask and bid quotes. As expected, both ask and bid quotes tend to increase (decrease) significantly after the arrival of a bid (ask) limit order. Induced by the cointegration setting, the quotes naturally converge to a (new) permanent level at which the information content of the incoming limit order is completely incorporated. The confidence intervals reflect that the shift is statistically highly significant.

We observe that quotes adjust relatively quickly reaching the new level after approximately 20 lags. Recall that time is measured in terms of limit order book activities. Hence, the adjustment speed measured in physical time ultimately depends on the underlying frequency of order activities and differs across the market. However, the fact that the speed of stock-specific quote adjustments (in terms of a "limit order clock") is widely stable across the market, indicates that such a business time scale is appropriate for market-wide comparisons across stocks.

An interesting fact is that bid quotes tend to increase more quickly than ask quotes after the arrival of a bid limit order. A reverse effect is observed after the arrival of an ask limit order. This asymmetry introduces a one-sided and temporary decrease of

[6] In all figures illustrating impulse responses, the legend "A → B" is interpreted to reflect "the impact on B induced by A".

the bid-ask spread. We explain this phenomenon by the fact that impatient traders observing an incoming limit order on the same side of the market tend to post more aggressively to liquidate their positions or increase the execution probabilities thereof. As a result, they place limit orders inside the spread shifting bid quotes upward. Moreover, the higher liquidity supply on the bid side generates a (delayed) liquidity demand on the ask side shifting ask quotes upward as well. We thus refer this phenomenon to be a liquidity-motivated effect.

Our findings can be interpreted in terms of pure market mechanisms. The market equilibrium is perturbed by a limit order in two ways. On one hand, the limit order indicates an investor's willingness to buy or sell and thus increases supply or demand of the underlying asset. The market price changes in order to incorporate this temporary imbalance of supply and demand. One the other hand, an incoming limit order increases the supply of liquidity in the market. Narrowing of the spread reduces transaction costs and causes a re-balancing of supply and demand of liquidity.

The significant long-term effect induced by an incoming limit order indicates that it contains private information on the value of assets. This finding is in contrast to the common assumption in theoretical literature that informed traders only take liquidity but do not provide it. On the other hand, it is supported by the experiment by Bloomfield, O'Hara, and Saar (2004) showing that informed traders use order strategies involving both market orders and limit orders to optimally capitalize their informational advantage.

Given the setting of the book we observe that a limit order increasing first level depth by 50% shifts quotes by 0.5-0.6 basis points. Though this is generally rather small, it is economically significant if the tick size is small. Obviously, these magnitudes ultimately depend on the (relative) order size as well as on underlying stock specific characteristics. The impact of the latter will be discussed in more detail in Section 5.5.

In order to explore the role of the order's position in the book, Figure 9 depicts the bid prices' reactions induced by incoming bid limit orders placed at the market (level one) and behind the market (level two and three).[7] We observe a negative correlation between the magnitude of price reactions and the orders' distance from the spread. The at-the-market limit order induces significantly faster market reactions than the behind-the-market limit order. Nonetheless, the long-term impact of level one and level two limit orders is only approximately 20% smaller. Hence, it turns out that behind-the-

[7]The sizes of the orders are assumed to be the same. Nonetheless, the ultimate magnitudes of shocks are different since we assume that the initial order book equals to the monthly average in which the depth on level two and three are approximately 1.5 times of that on level one.

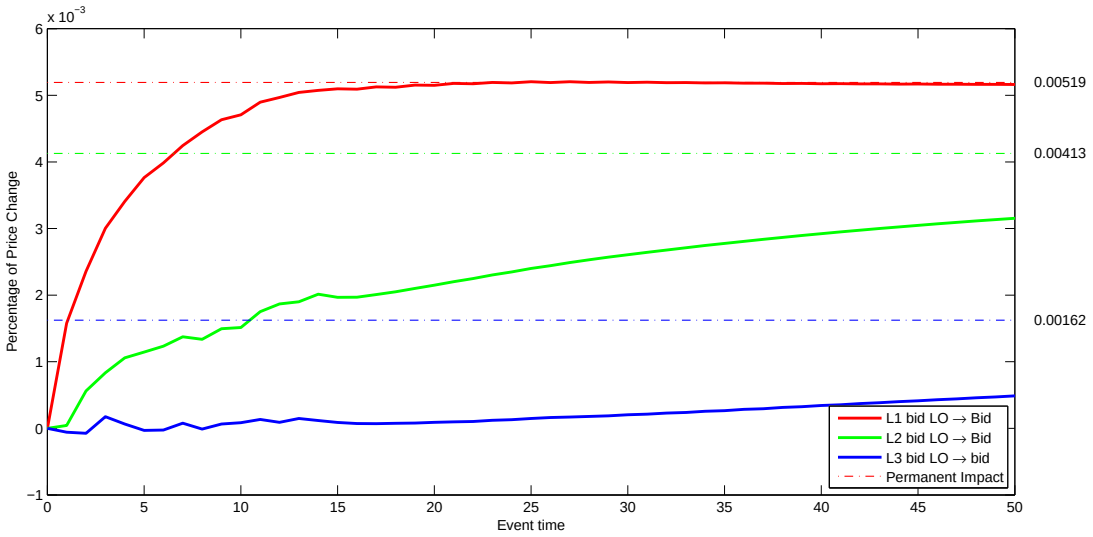


Figure 9: Percentage changes of bid quotes induced by incoming bid limit orders placed at the market (level one) and behind the market (level two and three). The order size equals to half of that at the best bid. The initial order book equals to the corresponding monthly average shown in Table 1. The marked number on the vertical axes indicates the magnitude of the permanent impact. Trading of Fortis at Euronext, Amsterdam. L1: level one. L2: level two. L3: level three. LO: limit order.

market orders can significantly shift the market though the quote adjustment is slower.[8] This result holds for level two orders and (to a weaker extent) for level three orders. However, for orders posted deeper in the book virtually no market impacts can be identified.

Eom, Lee, and Park (2009) find evidence that traders could have made extra profits using microstructure-based manipulations on the Korean Exchange (KRX) during a period between 2001 and 2002. In this period, KRX disclosed the total quantity on each side of the order book without fully disclosing the prices at which these orders have been placed. The manipulation strategy resulted in placing huge numbers of behind-the-market limit orders on the opposite side of the market inducing price moves in the favorite direction without having these orders executed. Our finding shows that this kind of manipulation is indeed possible. However, whether it is economically profitable in Euronext Amsterdam ultimately depends on (relative) order sizes. In order to move prices in her favorite direction, the trader has to submit rather big limit orders close to the market. Then, she obviously faces the risk that these orders are likely to be picked

[8] In order to improve the graphical illustrations, we refrain from showing the corresponding confidence intervals. They are quite similar to those shown in Figure 8.

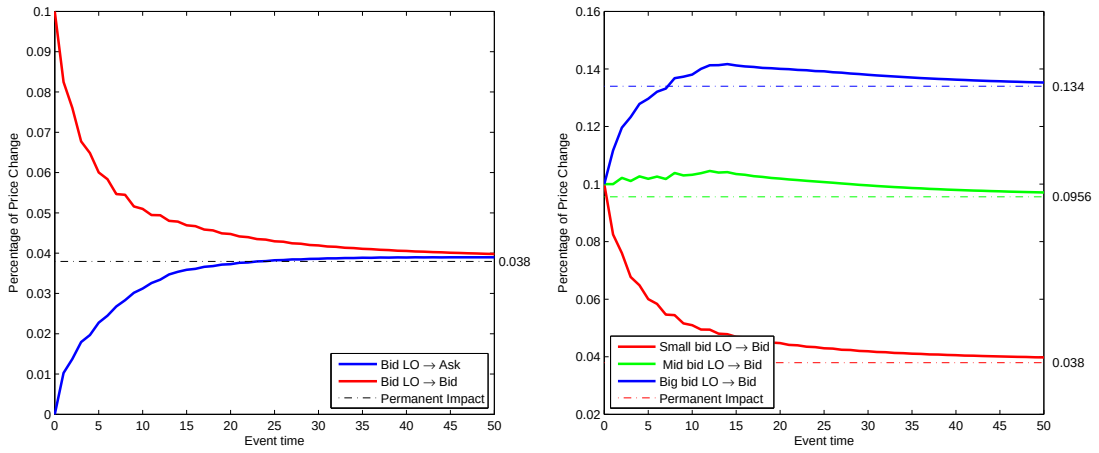


Figure 10: Left: Percentage changes of bid and ask quotes induced by incoming bid limit orders placed inside of the spread with a size equal to half of the depth at the best bid. **Right:** Percentage changes of the bid quote induced by incoming bid limit orders placed inside of the spread with different sizes. The bid limit orders increase the bid price by 0.1%. The initial order book equals to the corresponding monthly average shown in Table 1. Small size: 50% of the depth at the best bid. Mid size: 10 times of the depth at the best bid. Big size: 50 times of the depth at the best bid. Trading of Fortis at Euronext, Amsterdam. LO: limit order.

up.

5.2 Limit Orders Placed Inside Of the Spread

Limit orders placed inside of the bid-ask spread perturb the order book dynamics in a more complex way. Apart from providing liquidity to the order book, they directly improve the best quotes. This quote adjustment induces a reduction of the spread, establishes a new best quote level and correspondingly shifts all depth levels on the corresponding side of the book upward (or downward, respectively). The system seeks the new equilibrium on a path recovering from the immediate quote change and simultaneously re-balancing liquidity. Given our setting, we assume that a bid limit order inside of the spread induces an automatic 0.1% increase of the best bid quote. However, as shown in the left plot of Figure 10, the long-run price impact is just 0.04%. Hence, the immediate quote movement is reverted back by approximately 60%. This is induced either by sell trades picking up the posted volume or by cancellations on the bid side. Similarly, liquidity demand on the ask side shifts the ask quote upward by 0.04%. Hence, overall we observe an asymmetric re-balancing of quotes and a corresponding re-widening of the spread.

The right plot of Figure 10 compares the effects of incoming bid limit orders of

different sizes but with same limit price posted inside of the bid-ask spread and thus improving bid quotes again by 0.1%. Interestingly, we observe quite different impulse response patterns in dependence of the order size. In case of a comparably small order, the posted volume is obviously quickly picked up, shifting the bid quote back. Hence, similar to the effect shown in the left plot of Figure 10, the automatic quote improvement is reverted back by more than 60%. In contrast, large volumes overbidding the prevailing quote cause a long-term upward movement of the bid quote. Relative to the initial shift of the bid price we observe a further approximately 35% price increase. Hence, extraordinary large orders are not likely to be picked up and rather induce strong buy pressure moving the market upwards. For smaller (though still comparably large) orders, adverse selection and signaling effects seem to counterbalance each other. As a consequence, the bid quote is hardly changed and the long run effect is close to the immediate price improvement. Note that in this particular example, the monthly average spread is approximately 0.14% implying that the hypothetical limit order improving the bid quote by 0.1% is indeed very aggressive. Consequently, it is very likely to be picked up by market orders and thus its size must be quite huge to effectively shift the market. As shown in the web appendix, for less aggressive limit orders placed inside of the bid-ask spread, smaller order sizes are sufficient to induce signaling effects and to ultimately "scare" the market.

5.3 Market Impact of Trades

Figure 11 shows the market impacts induced by incoming bid (buy) and ask (sell) market orders. We assume that the trade sizes correspond to 50% of the prevailing depth. Consequently, these market orders do not "walk up" (or down, respectively) the book and thus best ask and bid quotes are unaffected. Hence, the quote adjustments shown in Figure 11 are subsequent quote responses to trade arrivals. Both the bid and ask quotes increase (decrease) sharply after the arrival of a buy (sell) market order. Hence, the arrival of a buy (sell) market order induces aggressive posting on the bid (ask) side resulting in further buy (sell) market orders and bid (ask) limit orders posted inside of the spread. Similar to the findings for limit orders, we find evidence for asymmetric adjustments of the two sides of the market. It turns out that bid (buy) market orders shift the ask quote more quickly and strongly than the bid quote. The reverse is true for ask (sell) market orders. This result indicates that trades temporarily increase the spreads which is in contrast to the effects induced by limit orders. Engle and Patton (2004) report similar findings by analyzing quote data from the NYSE. They show that trades have a positive impact on spreads, but do not identify

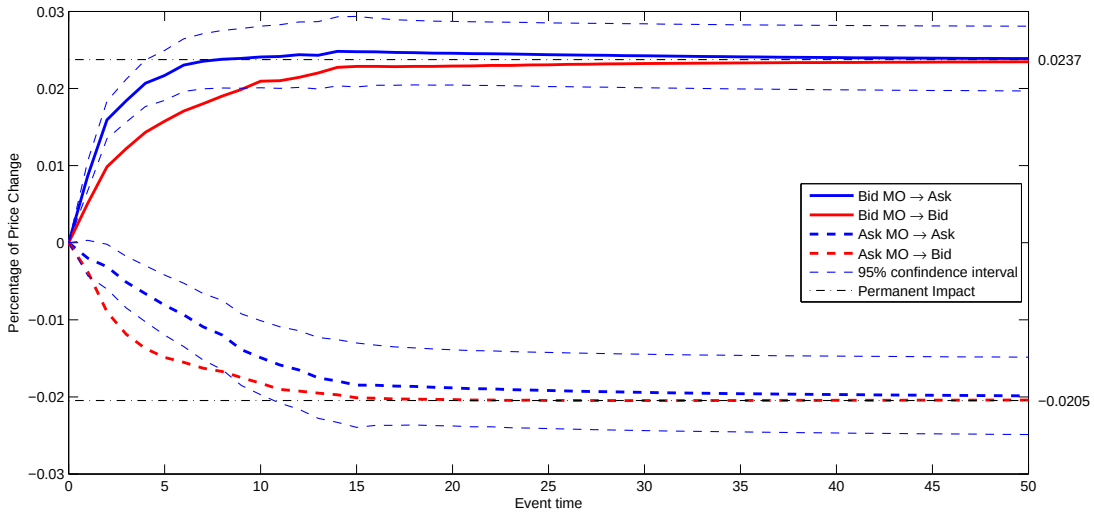


Figure 11: Percentage changes of ask and bid quotes induced by incoming bid/ask (buy/sell) market orders with a size equal to half of the depth on their corresponding first levels. The marked number on the vertical axes indicates the magnitude of the permanent impact. Trading of Fortis at Euronext, Amsterdam. MO: Market order.

whether this impact is permanent or only transitory. Using impulse-response analysis based on a structural VEC model, Escribano and Pascual (2006) also find that spreads (permanently) widen after the arrival of trades. Note that these effects contradict implications of asymmetric information based market microstructure models, such as Glosten and Milgrom (1985) and Easley and O'Hara (1992), where trades should resolve the uncertainty regarding existing information and should result in declining spreads.

The left plot of Figure 12 depicts the quote reactions induced by an aggressive market order "walking up" the book (Scenario 4 in Section 3.2). It absorbs the best ask level and shifts the best quote to the originally second best level which is assumed to be 10 basis points higher than the previous best ask. Similarly to the effects induced by aggressive limit orders we observe that the initial shift of the best ask is reverted back by approximately 35% inducing a long-run ask increase of 6.4 basis points. Simultaneously, aggressive posting on the bid side shifts bid quotes upward. Hence, the initially widened spread reverts back in an asymmetric way causing more quote movements on the bid side than on the ask side. The responses mirror the corresponding effects induced by aggressive bid limit orders (cf. Figure 10), where the spread is initially narrowed and then asymmetrically re-widened causing also more movements on the bid side than on the ask side.

The right plot of Figure 12 compares the market impacts on the ask quote induced

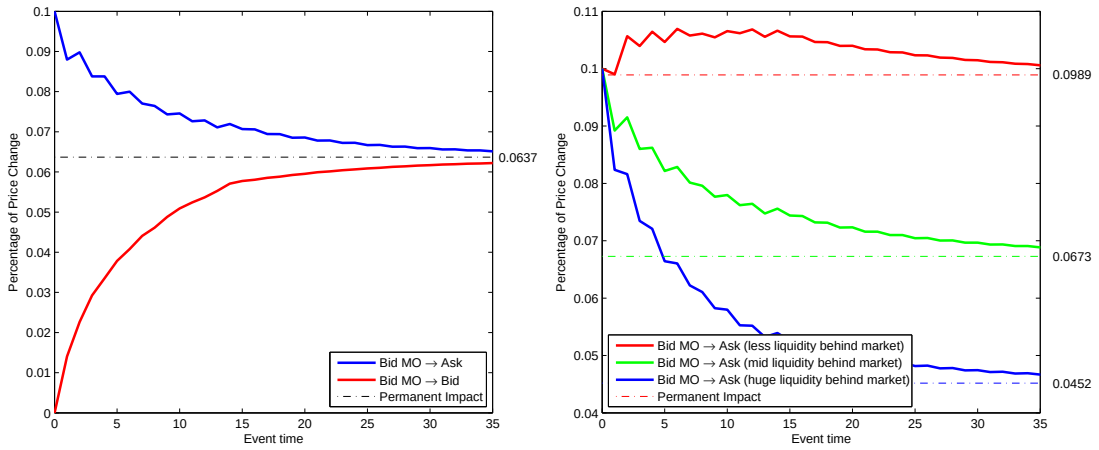


Figure 12: Left: Percentage changes of bid and ask quotes induced by an aggressive bid (buy) market order with a size exceeding the depth at the best ask by 20%. The second best ask price is assumed to be 0.1% higher than the best ask, where the depths behind the market are 1.5 times of the depth at the market. **Right:** Percentage changes of the ask quote induced by an aggressive bid (buy) market order with a size equal to the depth at the best ask when there is different depth at the second best level. Case 1: the depth at the second best ask level is 10% of that at the best ask; Case 2: the depth at the second best ask level equals to that at the best ask level; Case 3: the depth at the second best ask level is 500% of that at the best ask. The marked number on the vertical axes indicate the magnitude of the permanent impact. Trading of Fortis at Euronext, Amsterdam. MO: Market order.

by a buy market order in situations of different depth behind the market. It is assumed that the order just absorbs the first ask level and thus induces an instantaneous ask price increase by 10 basis points. In line with the results discussed above, in all three scenarios the initially shifted ask quote is reverted back. However, it turns out that the magnitude of this quote reversion critically depends on the prevailing depth behind the market. In fact, the existence of a huge level two depth reverts the ask quote back by approximately 55%. We explain this fact by a strong sell pressure induced by huge sell volume queued on the ask side. Conversely, in case of only small prevailing depth behind the market, the existing sell pressure is obviously weaker causing the incoming buy order to (upward) shift the market more strongly. In the extreme case of a very thin market, we even observe a temporary additional quote increase.

A practical problem faced by many market participants is the fundamental choice between posting a market order or a limit order. A direct comparison of the market impacts induced by these two types of orders is shown in Figure 13. In both cases, the posted order does not *directly* change the best quote. We observe that the resulting

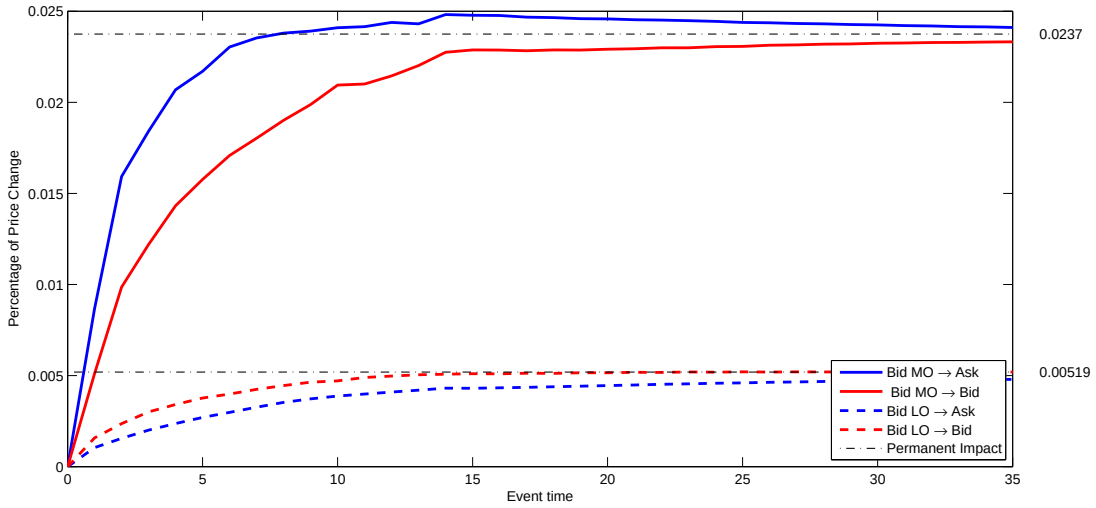


Figure 13: Percentage changes of ask and bid quotes induced by a bid (buy) market order and a bid limit order of similar size placed at the market. The order size is half of the depth at the best bid. The depths at the best bid and the best ask in the order book are assumed to be equal. LO: limit order; MO: market order.

long-run effect of trades is significantly greater than that of an equal-size limit orders. Actually, the price shift induced by a market order is approximately four times larger than that of a comparable limit order. Moreover, market orders also cause quicker market reactions. Finally, inferring from the "gap" between ask and bid curves, it is shown that market orders change the spread more dramatically than limit orders. Hence, the willingness to cross the bid-ask spread is obviously a stronger signal for private information than that induced by a comparable limit order.

Note that the comparison holds for "normal" order types placed on the best quote, but not necessarily for more aggressive orders. As discussed above, the long-term effects of aggressive limit orders and market orders critically depend on their (relative) size and the current state of the book. Therefore, an ultimate comparison of market impacts induced by both types of orders under comparable conditions is rather difficult. Nevertheless, our results show that limit orders do have a significant long-term effect and can significantly "scare" the market.

5.4 Robustness of Results

Selecting the appropriate lag order in VARX models is cumbersome in practice when a substantial cross-section of stocks is analyzed over a comparably long period. In order to analyze the sensitivity of our results regarding the choice of the lag order in the

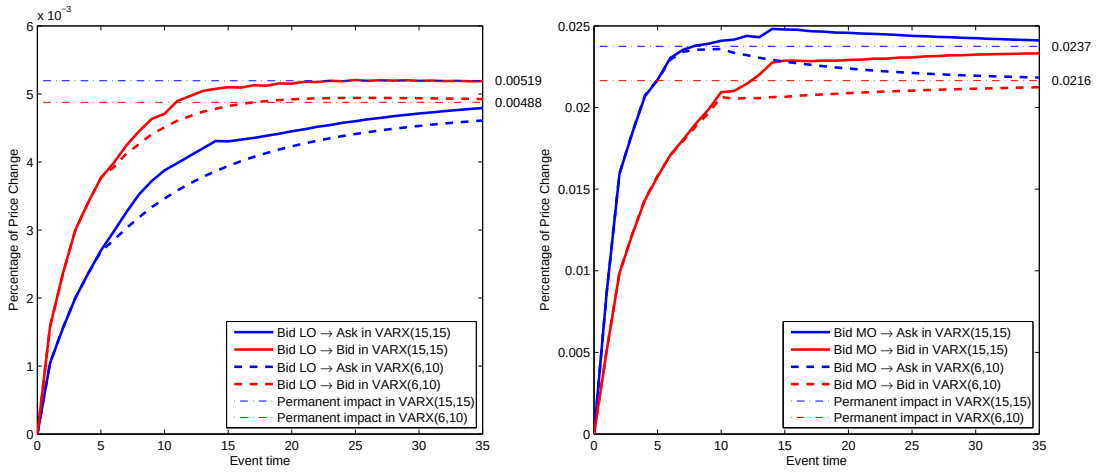


Figure 14: Robustness of results. Market impacts of a bid limit order estimated by a VARX$(15, 15)$ and a VARX$(6, 10)$ specification. Trading of Fortis, Euronext Amsterdam.

VARX model, Figure 14 compares the market impacts of a bid limit order and that of a normal buy market order predicted by a VARX$(15, 15)$ model with those induced by a VARX$(6, 10)$ specification using trading of Fortis in August, 2008. It turns out that despite a misspecification of the lag length and remaining serial correlation in the residuals, the impulse response estimates of a VARX$(6, 10)$ are quite close to that of a VARX$(15, 15)$. This is in line with results reported by Jorda (2005) using a VAR(2) to estimate impulse-response functions of an underlying VAR(12) model.

5.5 Cross-Sectional Evidence

The complete empirical analysis has been conducted for 29 other stocks traded at Euronext Amsterdam using a VARX$(15, 15)$ specification. The corresponding results are shown in the appendix on the companion web site at `http://amor.cms.hu-berlin.de/~huangrui/project/impact_of_orders/`. It turns out that the results reported in the previous sections are qualitatively stable and representative for a wide cross-section of stocks. Nevertheless, we observe that the *magnitudes* of market impacts vary across the market and seem to be driven by underlying liquidity characteristics. To gain insights into these relationships, we run a simple cross-sectional regression of absolute average market impacts on the average stock-specific trading frequency, trading volume as well as the minimum tick size. I.e.,

$$M_i = \gamma_0 + \gamma_1 N_i + \gamma_2 S_i + \gamma_3 V_i + \varepsilon_i, \tag{16}$$

where M_i denotes the absolute permanent impact of stock i induced by a bid/ask limit order, N_i is the average number of trades per day, S_i represents the normalized tick size, and V_i denotes the normalized transaction volume per day. Particularly,

$$S = \frac{\text{tick size} \times 100}{\text{the average of closing prices}}, \quad V = \frac{\text{adjusted trading volume per day}}{\text{number of outstanding shares}} \times 100.$$

The scenarios we consider below are similar to those studied in Section 3.2. The initial order book for each stock equals its monthly average.

Scenario "normal limit order" and "normal market order" : These scenarios are identical to that in Section 3.2.

Scenario "aggressive limit order" : An incoming order of a size which is half to the depth at the corresponding best price is posted inside of the spread and improving the corresponding quote by one tick.

Scenario "aggressive market order" : An incoming market order with a size equal to the depth at the corresponding best price and thus absorbing the first level in the book.

Scenario	γ_0	γ_1	γ_2	γ_3	R^2
"normal limit order"	0.0033	−0.0013	0.0419	–	0.67
	(13.25)	(−6.65)	(21.84)		
	0.0026	−0.0015	0.040	0.0012	0.72
	(11.28)	(−9.75)	(21.17)	(8.2)	
"aggressive limit order"	0.005	−0.0017	0.095	–	0.77
	(9.42)	(−5.26)	(14.97)		
	0.0041	−0.002	0.0933	0.0015	0.79
	(9.63)	(−6.65)	(14.23)	(7.19)	
"normal market order"	0.034	−0.014	0.14	−−	0.53
	(23.97)	(−9.45)	(4.71)		
	0.027	−0.0168	0.1257	0.013	0.65
	(16.73)	(−17.94)	(4.04)	(7.67)	
"aggressive market order"	0.0463	−0.0181	0.4488	−−	0.66
	(21.89)	(−8.40)	(11.15)		
	0.0358	−0.0214	0.4242	0.0178	0.74
	(14.76)	(−16.07)	(10.57)	(7.24)	

Table 5: Parameter estimates based on equation (16). The numbers in brackets denote heteroskedasticity robust $t-$statistics according to White (1980).

For every scenario, we consider average market impacts of both bid and ask orders for 30 stocks estimated over two months resulting in 120 observations for each

regression. Table 5 reports the corresponding estimation results for two versions of the model: one with included trading volume and one without.

The high R^2 values, ranging between 65% and 79%, show that most of the cross-sectional variation of market impact can be indeed explained by the three explanatory variables. It turns out that the trading volume (though its parameter is significant) does not provide much explanatory power. This result indicates that the trading frequency rather than the trade size drives the strength of market responses to limit order arrivals. Furthermore, we observe that the trading frequency has a negative influence on the market impact of limit orders. Hence, in case of a slower trading, a single order obviously conveys more information.

The tick size is positively related to the magnitude of permanent impacts in all scenarios. For aggressive limit orders, this finding is not surprising as the implied price improvement is (relatively) higher for stocks trading on larger tick sizes. Since in these cases, also the spreads between best and second best quotes are higher, the immediate price shift by the arrival of an aggressive market order is larger as well. In the scenarios "normal limit order" and "normal market order", a higher tick size and thus an increase of the price discreteness makes it more likely that investors are forced to under-react or over-react in response to incoming information inducing higher deviations between quoted prices and the "true" underlying efficient price. Our findings show that in these situations, investors rather tend to over-react after the arrival of a limit order.

6 Conclusions

In this paper, we quantify the market impact of incoming limit orders in a limit order book market. Best bid and ask quotes as well as three levels of depth on both sides of the market are modelled based on a cointegrated VAR system. Incoming limit orders are represented in terms of shocks to the system. Limit order characteristics as well as the corresponding state of the book are captured by the specific design of the shock vector. This allows us to distinguish between limit orders of different aggressiveness (reflected by their distance to the market) and different sizes as well as between different states of the book. The market impacts on ask and bid prices are quantified by the estimated impulse response function using appropriate statistical inference.

Employing this modelling framework we analyze the limit order book processes of 30 stocks traded on Euronext Amsterdam over two months in 2008. The model is estimated using the highest possible frequency accounting for all order book changes during continuous trading. Parameter estimates and diagnostics indicate that the proposed

model captures the high-frequency order book dynamics quite well.

Based on the empirical analysis we can summarize the following findings: First, we find clear evidence for cointegration relationships between ask and bid quotes and corresponding depths. While some cointegration relationships are similar to the bid-ask spread, others show that depth has a distinct effect on quote dynamics and on the connection between ask and bid quotes. Second, limit orders do have significant long-term effects on quotes. This is even true for limit orders placed behind the market though these effects decline with the limit order's distance to the market. While incoming limit orders temporarily decrease the spread, market orders induce a temporary widening. Third, the speed of spread convergence as well as the direction of price movements after the arrival of aggressive limit orders undercutting (or overbidding, respectively) best ask and bid prices depends strongly on the incoming limit order's size. While small orders seem to face adverse selection risks and are likely to be picked up quickly, for larger orders information signaling effects seem to dominate pushing the market in the opposite direction. Fourth, the decrease (increase) of spreads after the arrival of an aggressive limit (market) order is reverted back asymmetrically inducing more quote movements on the side where the order has been placed. Fifth, the long-run market impact of aggressive market orders walking up (or down) the book is the lower the larger the queued depth behind the market. Sixth, the effects are qualitatively remarkably stable over the cross-section of the market. Variations in the magnitudes of market impacts are well explained by the underlying stock-specific trading frequency and minimum tick size.

Our empirical results also show that the proposed framework is useful and appropriate to capture order book dynamics on high frequencies. By modelling quotes *and* several levels of depth the model implicitly captures also the multivariate dynamics of mid-quotes, returns, spreads, spread changes as well as depth imbalances. In this sense, the suggested high-frequency cointegrated VAR model can serve as a workhorse for various applications in this area.

References

Bloomfield, Robert, Maureen O'Hara, and Gideon Saar, 2004, The "make or take" decision in an electronic market: evidence on the evolution of liquidity, *Journal of Financial Econometrics* 75, 165–199.

Dufour, Alfonso, and Robert F. Engle, 2000, Time and the impact of a trade, *Journal of Finance* 55, 2467–2498.

Easley, David, and Maureen O'Hara, 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 577–605.

Engle, R.F., and A.J. Patton, 2004, Impact of trades in an error-correction model of quote prices, *Journal of Financial Markets* 7, 1–25.

Eom, K.S., E. Lee, and K.S. Park, 2009, Microstructure-based manipulation: Strategic behavior and performance of spoofing traders, Discussion paper University of California at Berkeley.

Escribano, Alvaro, and Roberto Pascual, 2006, Asymmetries in bid and ask responses to inonovation in the trading process, *Empirical Economics* 30, 913–946.

Glosten, L.R, and P.R. Milgrom, 1985, Bid, ask and transaction prices in a market-maker market with heterogeneously informed trades, *Journal of Financial Economics* 14, 71–100.

Hansen, P. R., and A. Lunde, 2006, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24, 127–161.

Härdle, Wolfgang K., Nikolaus Hautsch, and Andrija Mihoci, 2009, Modelling and forecasting liquidity supply using semiparametric factor dynamics, Discussion Paper 2009/18 Collaborative Research Center 649 "Economic Risk", Humboldt-Universität zu Berlin.

Hasbrouck, J., 1991, Measuring the information content of stock trades, *The Journal of Finance* 46, 179–207.

Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* 59, 1551–80.

——— , and K. Juselius, 1990, Maximum likelihood estimation and inference on cointegration - with applications to the demand for money, *Oxford Bulletin of Economics and Statistics* 52, 169–210.

Jorda, O., 2005, Estimation and inference of impulse response by local projections, *American Economic Review* 95, 161–182.

Juselius, Katarina, 2007, *The Cointegrated VAR Model: Methodology and Applications* (Oxford University Press).

Lee, C., and M. Ready, 1991, Inferring trade direction from intraday data, *Jounal of Finance* 46, 733–746.

Ljung, G.M., and G.E.P. Box, 1978, On a measure of lack of fit in time series models, *Biometrika* 65, 297–303.

Lo, I., and S.G. Sapp, 2006, A structural error-correction model of best prices and depths in the foreign exchange limit order market, Workingpaper Bank of Canada.

Lütkepohl, Helmut, 2005, *New Introduction to Multiple Time Series Analysis* (Springer).

Lütkepohl, H., and H.E. Reimers, 1992, Impulse response analysis of cointegrated systems, *Jouranl of Economic Dynamics and Control* 16, 53–78.

Potters, M., and J.P. Bouchaud, 2003, More statistical properties of order books and price impact, *Physica A* 324, 133–140.

White, H., 1980, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica* 48, 817–838.

A Adaptive time window matching algorithm

In our database, transaction data and order book data are recorded in separate files stemming from different recording systems. As a result, the time stamps in the two data sets have different time distances to exchange time. In accordance with the institutional settings of Euronext, we design an adaptive time window matching algorithm which contains three main steps.

Step 1 Exact matching. The algorithm picks up a time stamp of a trade and opens a specified time window, e.g. $[-10, 10]$ seconds around this time stamp. Then, a procedure picks every order book record in this time window and performs the following analysis: If (i) the trade price equals to the best bid (ask) price and the difference of the best bid (ask) size between this order book record and the previous one equals to the trade size or (ii) the trade price equals to the previous best bid (ask) price, the best bid (ask) size equals to the trade size and the best bid (ask) price decreases (increases), it matches this order book record with the corresponding trade and records the delay time between the trade and the order book. If no match is achieved for all order book records in the time window, the trade remains to be unmatched.

Step 2 Inexactly matching. The algorithm picks up an unmatched trade record's time stamp and opens a time window of size which is twice the average delay time computed in Step 1. If (i) the trade price equals to the best bid (ask) price and the best bid (ask) size is less than the previous one or (ii) the best bid (ask) price decreases (increases), it matches the trade with the current order book. If no match is achieved for all order book records in the time window, the trade remains to be unmatched.

Step 3 Round time matching. The algorithm picks up an unmatched trade and matches it with an order book record that is closed to the trade's time stamp plus the average delay time.

B Proof of Equation (12)

Recall three useful standard results (see, e.g., in the appendix in Lütkepohl (2005)) for any comfortable matrices A, B, C, D and vector a

$$\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B), \tag{B.1}$$

$$(A \otimes B)(C \otimes D) = AB \otimes CD, \tag{B.2}$$

$$\frac{\partial \operatorname{vec} A^h}{\partial a'} = \left[\sum_{i=0}^{h-1} (A')^{h-1-i} \otimes A^i\right] \frac{\partial \operatorname{vec}(A)}{\partial a'}. \tag{B.3}$$

Let Ψ denote $[B : J'] \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix}$ and $a := \operatorname{vec}(A_1, \cdots, A_p, B_1, \cdots, B_q)$. We first note that

$$a = \operatorname{vec}(J\mathbf{A})$$

and

$$\frac{\partial \operatorname{vec} \mathbf{A}}{\partial a'} = \frac{\partial \operatorname{vec}(J'J\mathbf{A})}{\partial a'} = \left(I_{Kp+2s} \otimes J'\right), \tag{B.4}$$

where we use (B.1) and the fact that $\partial \operatorname{vec}(J\mathbf{A})/\partial a' = I$. By further elaborating on (B.3), we have

$$\begin{aligned} G_h &= \frac{\partial \operatorname{vec}(J\mathbf{A}^h\Psi)}{\partial a'} = (\Psi' \otimes J)\frac{\partial \operatorname{vec}(\mathbf{A}^h)}{\partial a'} \\ &= (\Psi' \otimes J)\left[\sum_{i=0}^{h-1} (A')^{h-1-i} \otimes A^i\right] \frac{\partial \operatorname{vec}(A)}{\partial a'}. \end{aligned}$$

Equation (12) is found by inserting (B.4) and applying (B.2).

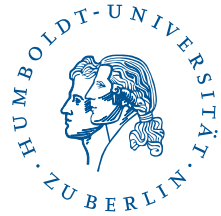
SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Implied Market Price of Weather Risk" by Wolfgang Härdle and Brenda López Cabrera, January 2009.
002 "On the Systemic Nature of Weather Risk" by Guenther Filler, Martin Odening, Ostap Okhrin and Wei Xu, January 2009.
003 "Localized Realized Volatility Modelling" by Ying Chen, Wolfgang Karl Härdle and Uta Pigorsch, January 2009.
004 "New recipes for estimating default intensities" by Alexander Baranovski, Carsten von Lieres and André Wilch, January 2009.
005 "Panel Cointegration Testing in the Presence of a Time Trend" by Bernd Droge and Deniz Dilan Karaman Örsal, January 2009.
006 "Regulatory Risk under Optimal Incentive Regulation" by Roland Strausz, January 2009.
007 "Combination of multivariate volatility forecasts" by Alessandra Amendola and Giuseppe Storti, January 2009.
008 "Mortality modeling: Lee-Carter and the macroeconomy" by Katja Hanewald, January 2009.
009 "Stochastic Population Forecast for Germany and its Consequence for the German Pension System" by Wolfgang Härdle and Alena Mysickova, February 2009.
010 "A Microeconomic Explanation of the EPK Paradox" by Wolfgang Härdle, Volker Krätschmer and Rouslan Moro, February 2009.
011 "Defending Against Speculative Attacks" by Tijmen Daniëls, Henk Jager and Franc Klaassen, February 2009.
012 "On the Existence of the Moments of the Asymptotic Trace Statistic" by Deniz Dilan Karaman Örsal and Bernd Droge, February 2009.
013 "CDO Pricing with Copulae" by Barbara Choros, Wolfgang Härdle and Ostap Okhrin, March 2009.
014 "Properties of Hierarchical Archimedean Copulas" by Ostap Okhrin, Yarema Okhrin and Wolfgang Schmid, March 2009.
015 "Stochastic Mortality, Macroeconomic Risks, and Life Insurer Solvency" by Katja Hanewald, Thomas Post and Helmut Gründl, March 2009.
016 "Men, Women, and the Ballot Woman Suffrage in the United States" by Sebastian Braun and Michael Kvasnicka, March 2009.
017 "The Importance of Two-Sided Heterogeneity for the Cyclicality of Labour Market Dynamics" by Ronald Bachmann and Peggy David, March 2009.
018 "Transparency through Financial Claims with Fingerprints – A Free Market Mechanism for Preventing Mortgage Securitization Induced Financial Crises" by Helmut Gründl and Thomas Post, March 2009.
019 "A Joint Analysis of the KOSPI 200 Option and ODAX Option Markets Dynamics" by Ji Cao, Wolfgang Härdle and Julius Mungo, March 2009.
020 "Putting Up a Good Fight: The Galí-Monacelli Model versus 'The Six Major Puzzles in International Macroeconomics'", by Stefan Ried, April 2009.
021 "Spectral estimation of the fractional order of a Lévy process" by Denis Belomestny, April 2009.
022 "Individual Welfare Gains from Deferred Life-Annuities under Stochastic Lee-Carter Mortality" by Thomas Post, April 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

023 "Pricing Bermudan options using regression: optimal rates of convergence for lower estimates" by Denis Belomestny, April 2009.
024 "Incorporating the Dynamics of Leverage into Default Prediction" by Gunter Löffler and Alina Maurer, April 2009.
025 "Measuring the effects of geographical distance on stock market correlation" by Stefanie Eckel, Gunter Löffler, Alina Maurer and Volker Schmidt, April 2009.
026 "Regression methods for stochastic control problems and their convergence analysis" by Denis Belomestny, Anastasia Kolodko and John Schoenmakers, May 2009.
027 "Unionisation Structures, Productivity, and Firm Performance" by Sebastian Braun, May 2009.
028 "Optimal Smoothing for a Computationally and Statistically Efficient Single Index Estimator" by Yingcun Xia, Wolfgang Härdle and Oliver Linton, May 2009.
029 "Controllability and Persistence of Money Market Rates along the Yield Curve: Evidence from the Euro Area" by Ulrike Busch and Dieter Nautz, May 2009.
030 "Non-constant Hazard Function and Inflation Dynamics" by Fang Yao, May 2009.
031 "De copulis non est disputandum - Copulae: An Overview" by Wolfgang Härdle and Ostap Okhrin, May 2009.
032 "Weather-based estimation of wildfire risk" by Joanne Ho and Martin Odening, June 2009.
033 "TFP Growth in Old and New Europe" by Michael C. Burda and Battista Severgnini, June 2009.
034 "How does entry regulation influence entry into self-employment and occupational mobility?" by Susanne Prantl and Alexandra Spitz-Oener, June 2009.
035 "Trade-Off Between Consumption Growth and Inequality: Theory and Evidence for Germany" by Runli Xie, June 2009.
036 "Inflation and Growth: New Evidence From a Dynamic Panel Threshold Analysis" by Stephanie Kremer, Alexander Bick and Dieter Nautz, July 2009.
037 "The Impact of the European Monetary Union on Inflation Persistence in the Euro Area" by Barbara Meller and Dieter Nautz, July 2009.
038 "CDO and HAC" by Barbara Choroś, Wolfgang Härdle and Ostap Okhrin, July 2009.
039 "Regulation and Investment in Network Industries: Evidence from European Telecoms" by Michał Grajek and Lars-Hendrik Röller, July 2009.
040 "The Political Economy of Regulatory Risk" by Roland Strausz, August 2009.
041 "Shape invariant modelling pricing kernels and risk aversion" by Maria Grith, Wolfgang Härdle and Juhyun Park, August 2009.
042 "The Cost of Tractability and the Calvo Pricing Assumption" by Fang Yao, September 2009.

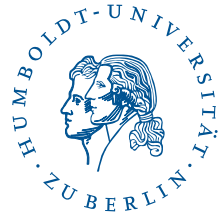
SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

043 "Evidence on Unemployment, Market Work and Household Production" by Michael C. Burda and Daniel S. Hamermesh, September 2009.
044 "Modelling and Forecasting Liquidity Supply Using Semiparametric Factor Dynamics" by Wolfgang Karl Härdle, Nikolaus Hautsch and Andrija Mihoci, September 2009.
045 "Quantifizierbarkeit von Risiken auf Finanzmärkten" by Wolfgang Karl Härdle and Christian Wolfgang Friedrich Kirchner, October 2009.
046 "Pricing of Asian temperature risk" by Fred Benth, Wolfgang Karl Härdle and Brenda López Cabrera, October 2009.
047 "MM-Stat – MultiMedia-Statistik: Statistische Datenanalyse – webbasiert, interaktiv und multimedial" by Sigbert Klinke, Dina Kuhlee, Christian Theel, Cornelia Wagner and Christian Westermeier, October 2009.
048 "Migration of the Highly Skilled: Can Europe catch up with the US?" by Lydia Mechtenberg and Roland Strausz, October 2009.
049 "A blocking and regularization approach to high dimensional realized covariance estimation" by Nikolaus Hautsch, Lada M. Kyj and Roel C.A. Oomen, October 2009.
050 "Generalized single-index models: The EFM approach" by Xia Cui, Wolfgang Karl Härdle and Lixing Zhu, October 2009.
051 "The Market Impact of a Limit Order" by Nikolaus Hautsch and Ruihong Huang, October 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".