Chen, Shi; Härdle, Wolfgang; Schienle, Melanie

Working Paper

High-dimensional statistical learning techniques for time-varying limit order book networks

IRTG 1792 Discussion Paper, No. 2021-015

Provided in Cooperation with:
Humboldt University Berlin, International Research Training Group 1792 "High Dimensional Nonstationary Time Series"

IRTG 1792 Discussion Paper 2021-015

High-dimensional Statistical Learning Techniques for Time-varying Limit Order Book Networks

Shi Chen *
Wolfgang Karl Härdle *2 *3 *4 *5 *6
Melanie Schienle *7

* Karlsruhe Institute of Technolog, Germany
*2 Humboldt-Universität zu Berlin, Germany
*3 Xiamen University, China
*4 Singapore Management University, Singapore
*5 Charles University, Czech Republic
*6 National Chiao Tung University, Taiwan
*7 Karlsruhe Institute of Technology, Germany

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

High-dimensional Statistical Learning Techniques for Time-varying Limit Order Book Networks*

Shi Chen[†], Wolfgang Karl Härdle[‡], Melanie Schienle[§]

Abstract

This paper provides statistical learning techniques for determining the full own-price market impact and the relevance and effect of cross-price and cross-asset spillover channels from intraday transactions data. The novel tools allow extracting comprehensive information contained in the limit order books (LOB) and quantify their impacts on the size and structure of price interdependencies across stocks. For correct empirical network determination of such dynamic liquidity price effects even in small portfolios, we require high-dimensional statistical learning methods with an integrated general bootstrap procedure. We document the importance of LOB liquidity network spillovers even for a small blue-chip NASDAQ portfolio.

JEL classification: C02, C13, C22, C45, G12

Keywords: limit order book, high-dimensional statistical learning, liquidity networks, high frequency dynamics, market impact, bootstrap, network

---

*Financial support of the European Unions Horizon 2020 research and innovation program "FIN- TECH: A Financial supervision and Technology compliance training programme" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA), the European Cooperation in Science & Technology COST Action grant CA19130 - Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry, the Deutsche Forschungsgemeinschaft's IRTG 1792 grant, the Yushan Scholar Program of Taiwan and the Czech Science Foundations grant no. 19-28231X / CAS: XDA 23020303 are greatly acknowledged.

†corresponding author; Karlsruhe Institute of Technology, ECON-Econometrics and Statistics, Blücherstr.17, 76185 Karlsruhe, Germany. Email: shi.chen@kit.edu.

‡Blockchain Research Center, Humboldt-Universität zu Berlin, Germany. Wang Yanan Institute for Studies in Economics, Xiamen University, China. Sim Kee Boon Institute for Financial Economics, Singapore Management University, Singapore. Faculty of Mathematics and Physics, Charles University, Czech Republic. National Chiao Tung University, Taiwan.

§Karlsruhe Institute of Technology, ECON-Econometrics and Statistics, Blücherstr.17, 76185 Karlsruhe, Germany.

1 Introduction

Today increasing attention is directed from understanding the high frequency market microstructure to measuring its effects on the overall market functioning (see e.g. O'Hara (2015), Obizhaeva and Wang (2013), Securities et al. (2010), Fleming et al. (2017)). We contribute to this literature by addressing three main research questions: 1. What is the full price market impact of a limit order when taking into account that time-varying channels and total amounts of interlinkages might exist among assets and that these spillovers might differ on different levels of the limit order book (LOB)? Does such a LOB network analysis lead to a substantial difference in price impact compared to cases ignoring cross-effects from the LOB? And is there a significant difference of ask and bid sides on the price impact which can be exploited in trading and hedging strategies? 2. Economically, if we interpret LOB spillovers as liquidity cross-effects among assets, is there additional linkage information in such networks contained beyond price and volatility spillovers? 3. For empirical answers to 1. and 2., how can we design adequate high-dimensional statistical procedures for correct estimation and testing of dynamic LOB networks with multiple assets and different LOB levels per asset? Standard time-series methods are intractable in these cases and pure machine learning techniques do not yield interpretable network links and cannot assess the significance of effects. Though the number of assets in our portfolio is not large, the network we construct not only includes the asset prices, but also the order book information at several levels of both bid and ask quotes per asset, and, consequently, the network becomes high dimensional and requires novel techniques for estimation and inference. We therefore develop novel statistical learning type procedures which are computationally efficient in order to handle the big data setting of high-frequency observations and large amounts of features even in small portfolios.

Generally, in a high-frequency trading (HFT) world, it is well documented that fundamental market concepts are affected by limit orders which contain additional information (see e.g. O'Hara (2015)). Empirical studies such as Evans and Lyons (2002), Obizhaeva and Wang (2013) and Hirschey (2020) have already shown that limit orders play an important role in determining price dynamics by investigating the relation between order flows and price dynamics. Though, strategic decisions inherent in LOB are naturally multi-asset based and thus lead to a closely interconnected complex market structure over time. If we focus on market making strategies of HFT (see Securities et al. (2010)[1]), they provide a similar service to that of traditional market makers by engaging in electronic liquidity provision (see e.g. Hendershott and Seasholes (2007), Hendershott et al. (2011), Menkveld (2013), Brogaard et al. (2014), Budish et al. (2015)). High frequency market making, however, places orders across different securities/markets while the classic approach generally focuses on profits from single asset bid-ask spreads. This causes securities/markets to become tightly interconnected on all levels of the order book. Thus for assessing the size and channels of the impact by an order placement, linkages between ask and bid quotes and several levels of depth across securities might play a crucial role. So far, such cross-asset effects have been neglected in the literature as their empirical quantification requires innovative high-dimensional statistical techniques.

While LOBs have been analyzed in a variety of ways (see e.g. Parlour (1998), Engle (2000), Roşu (2009), Handa et al. (2003), Bloomfield et al. (2005)), empirical evidence on the market impact of limit orders is rare, see Hautsch and Huang (2012), Eisler et al. (2012), Cont et al. (2014) and Fleming et al. (2017). Moreover, most existing literature focuses on the own-price market impact, i.e. the price impact caused by a stock's own order flow, at the level of individual stocks only and restricted to problems of moderate dimension. Though in practice, when there is a portfolio consisting of selected stocks from different sectors, each market participant may place market/limit orders on one side of the book across multiple stocks according to their trading strategies such as market making/statistical arbitrage, the local decisions and interactions between thousands of market participants will result in a high dimensional dynamic and connected system. Of particular interest is to measure the cross-price market impact, i.e., the impact of orders placed in stock A on the price changes of stock B. Consider, for example, stock A and B coming from the same sector and their prices are positive correlated, adding a buy limit order of stock A will therefore induce extra upwards pressure on stock B. We study these impacts in a per stock aggregated way in order to infer overall market consequences. Insights into the dynamics of the spillovers within an LOB system are vital for the pricing of assets, but requires practically valid dimension reduction techniques for already medium-sized portfolios and a valid generalized impulse response analysis to quantify the impacts. There already exist some theoretical models on spillovers across

[1]The Securities and Exchange Commission (SEC) divides HFT strategies into market-making strategies, and opportunistic trading such as arbitrage and directional (order anticipation and momentum ignition) trading.

assets via LOB-related liquidity channels such as Foucault et al. (2005). In particular, Cespa and Foucault (2014) describe a similar mechanism to ours and show cross-asset learning as a source of liquidity spillovers based on a two-period model with two assets. Our paper provides a novel way to empirically determine the liquidity spillovers within and across asset classes. Up to our knowledge, this paper is the first to focus empirically on the effect of trading activity on price dynamics in a time-varying pattern. We not only provide a comprehensive and thus robust measure for the market impact of order placements across different stocks, but also empirically examine the main spillover channels in a LOB portfolio as liquidity cross-effects.

In this paper, we provide three main contributions solving the three key research questions from above. First, this paper proposes a statistical learning methodology to fully quantify own-price and cross-price market impacts given potential time-varying LOB network interlinkages. The structure of this LOB network over time is completely derived from data - in particular without pre-specifying a topological form or a parametric distribution. In this sense, we offer a novel comprehensive data science approach to discern information (e.g., trading intensity) from trade data. As indicated in Easley et al. (2016), " Traders informed of good news profit by buying and traders informed of bad news profit by selling, thus creating a trade imbalance. Trade imbalance between buys and sells can also signal liquidity pressure in markets, leading to subsequent price movements." A variety of trade classification algorithms in the literature are devoted to discerning information from LOB data, such as the bulk volume technique of Easley et al. (2011) and tick rules. Our methodology departs from previous microstructure research in that we empirically quantify the liquidity pressure by estimating the market impacts of buys and sells placed in the book for each single stock. To achieve this, we model the prices and ask/bid market/limit orders for a portfolio in a large vector autoregressive (VAR) process, and derive the market impact as well as the correlation pattern among stock prices by a so-called generalized impulse response function (GI). For a shock to a certain variable in the large VAR system, the GI measures how fast its effects on other variables decay over time. We estimate the market impact as the dynamic impact from a current shock at any level of a limit order book on the future price of a stock (best bid/ask price or mid-price) in the system. Note that a comprehensive measure for the full market impact of each single stock has widespread consequences for investment and hedging strategies. In particular, in the context of optimal order execution for minimizing total transaction costs, a single-stock model as e.g. in Almgren and Chriss (2001) could be augmented by taking cross-stock interconnections into account.

Second, we are the first to examine potentially time-varying full system interdependencies across stocks through prices and the information contained in the market/limit orders. By this, we obtain a network of liquidity spillovers from high-frequency data over time. While there have been a number of recent studies extracting network linkages from financial data at daily or lower observation frequencies over time, such as e.g. Hautsch et al. (2014), Giglio et al. (2016), Acharya et al. (2017) in the systemic risk context, networks at the high-frequency scale come with additional challenges on the data level. Thus we tackle the non-synchronicity of trade occurrence in HFT by proposing the so-called size intensity as a measure for trading activity of ask/bid quotes placed "at market" or "at limit" over a moderate time interval. Furthermore we develop an algorithm to achieve volume synchronization which also alleviates microstructure noise arising from market frictions. It has been well documented that the financial market is connected to and shapes the transmission of risk across sectors (e.g. Billio et al. (2012), Elliott et al. (2014), Di Maggio et al. (2017)). Though, the time-varying risk transmission between components of different sectors and within the same sector has been rarely investigated, leaving an important source of risk to be further analyzed. We address this issue by evaluating the net spillover effects over a sample period which covers the Brexit announcement. In this way, we can identify spillover channels across stocks which we document to be important and significant in particular around unforeseen news events. The identification of these network linkages then allows to correctly quantify the full market impact of the LOB over time. This is key for investment and hedging decisions, while specific spillover channels across limit order books are also of independent interest for supervisors and regulators.

Third, we contribute to the empirical network estimation literature by proposing a bootstrap-based generalized impulse response analysis for a price impact network construction in a practically valid high dimensional setting. In order to track system interdependencies over time, we must model and estimate the time evolution of a large system containing not only prices of multiple stocks but also the market/limit orders on both sides of book for each single stock. This high-dimensional object enables us to distinguish market impacts on the granular level among each component but also on an aggregated sector level, allowing us to assess all cross-effects between prices and LOB.

However its high-dimensionality causes the model to become intractable with traditional statistical methods such as OLS. We therefore require a penalized learning technique for our time series which automatically shrinks irrelevant elements to zero and makes identification thus possible in a high-dimensional setting. Although such regularized estimators are standard for cross-sectional data, as e.g. the least absolute selection and shrinkage operator (Lasso) Tibshirani (1996), SCAD Fan and Li (2001), adaptive Lasso Zou (2006), elastic net Zou and Hastie (2005), Dantzig selector Candes and Tao (2007), it is only recently that such shrinkage techniques became available for time series data (e.g., see Negahban and Wainwright (2011), Kock and Callot (2015), Basu et al. (2015), Wu and Wu (2016), Belloni et al. (2012) and Belloni et al. (2013)). We contribute to this literature, by proposing a bootstrap technique which preserves the sparsity structure from estimation of a vector autoregressive model (VAR) in its impulse responses. Thus the relevant elements for the dynamics and for the market impact coincide. In this way, we improve on the classical Diebold and Yılmaz network approach (Diebold and Yılmaz (2014)) which builds on symmetric generalized forecast error variance decomposition (GFEVD) after fitting a VAR. In particular, our bootstrap technique produces directed effects under mild assumptions no longer requiring linearity in the time series or normally distributed error terms such as in Lanne and Nyberg (2016). Thus we are able to determine directed market impacts from a high dimensional network of dynamic linkages between ask and bid quotes and several levels of depth across multiple stocks. But the suggested technical procedure might also be of independent interest in other network scenarios.

In our empirical results, we show that nine NASDAQ blue chip stocks across three different industries are sufficient to determine a significant effect of the LOB in market impact. In particular, we use HFT data of NASDAQ's historical TotalView-ITCH of the largest companies in capitalization for the sectors finance, health care and technology from 06/2016 to 07/2016 covering the time around the Brexit vote. From the underlying time-varying granular LOB-network we construct three types of networks on different aggregation levels for distilling key points of the market evolution addressing our three main research questions in a tailored way: By investigating the relation between order flows and price dynamics, we quantify the effects of order placements and provide empirical evidence of own-price and cross-price market impacts. We document that overall, cross-asset effects on prices amount to 30 up to more than 45% on average and document that a substantial share of these occur in the quotes. In the 9 stock system, LOB-effects already comprise around 21% to 26% of all price impacts which a price-only system would ignore. Moreover, this share substantially impacts the role of different stocks within the system as net price impact transmitters or receivers which is key for investment and hedging strategies. We also document how these proportions vary over time in particular around the Brexit date. In particular, we find that the financial sector dominates the market in the sense that it has the strongest influence on the others. Generally, financial stocks are dominated by LOB and cross-asset price effects. For most Healthcare and technology stocks cross-asset and cross-sector effects are more important. We show that the impacts caused by ask and bid orders are substantial but also asymmetric, where the NASDAQ market is quite sensitive to the market sell pressure. Moreover, we compare the impacts caused by market order and limit orders, and discuss how the depth of the book at which ask/bid orders are submitted is driving the price. Moreover, we are able to statistically assess not only the size and structure of effects but also test for their significance level and difference thus identify meaningful transmission channels.

The rest of the paper is organized as follows. Section 2 introduces the NASDAQ LOB market and presents the respective data preprocessing technique to tackle the non-synchronicity and microstructure noise of the LOB data with a volume-synchronization algorithm and pre-averaging. In Section 3 we outline the methodology. In particular, we show how to efficiently identify the underlying dynamic high-dimensional VAR structure of the system and how to obtain valid estimates of market impact via a bootstrap procedure. Moreover we characterize different aggregation levels of networks employed. The empirical results of the impact of LOB networks are shown in Section 4. In Section 5 we employ the network results to derive the total price dynamics over time of an uncertainty shock. Section 6 concludes. Generally, all technical details can be found in the Appendix.

2 HF LOB-Data and Data Preprocessing

2.1 Data Description

We use high-frequency limit order book data from NASDAQ's historical TotalView-ITCH. Our sample portfolio covers three main economic sectors financials, technology and health care which together comprise around two thirds in capitalization of all NASDAQ traded firms.[2] For each sector, it consists of the three largest blue-chip stocks by market capitalization. Thus we work with market prices and ask/bid side quotes up to level 3 for nine highly liquid stocks. Table 1 presents an overview of the covered stocks for each industry sector. The sample period comprises two months from 1st, June 2016 to 31st, July of 2016 containing the date of the Brexit vote which we take as an event of a mostly unanticipated shock. The data comprises both, visible and hidden orders for all normal trading days running from 9:30h to 16:00h ET. To avoid erratic effects during the market opening and closure, we discard the first and the last 15 minutes of each trading day. The summary statistics for the considered quantities of all stocks are reported in Table 2.

Industry	Stock	Company	MktCap (billion $)
Technology	IBM	International Business Machines Corp.	171.72
	MSFT	Microsoft Corporation	499.35
	T	AT&T Inc.	257.53
Healthcare	JNJ	Johnson & Johnson	328.91
	PFE	Pfizer Inc.	206.69
	MRK	Merck & Co. Inc.	181.56
Finance	JPM	JP Morgan Chase & Co.	326.04
	WFC	Wells Fargo & Company	293.39
	C	Citigroup Inc.	168.06

Table 1: Overview of the stocks in the considered portfolio. MktCap denotes the market capitalization by Feb 25th, 2017.

We work with the LOBSTER reconstruction of the LOB based from the NASDAQ TotalView-ITCH Data.[3] In particular, the LOBSTER pre-processing steps match a trade with a corresponding LOB by a transparent algorithm yielding ready-to-use time series of LOBs with the structure as shown in Figure 1. The algorithm identifies buy and sell quotes but also detects sub-trades arising from

[2]The industry breakdown of the NASDAQ market is: technology 45.38%, Health Care 11.43% and Financials 8.42% (as of 23.02.2018.)

[3]See https://lobsterdata.com1

	$NumObs$ $(*10^3)$	$AvgTrd$ $(*10^3)$	$AvgAP1$ (in \$)	$AvgBP1$ (in \$)	$AvgAS1$ (100 shrs)
IBM	118.25	5.82	153.07	153.04	1.92
MSFT	584.55	25.91	52.28	52.26	24.19
T	223.45	6.67	38.75	38.74	36.36
JNJ	172.77	8.17	113.99	113.98	4.11
PFE	427.51	12.49	34.83	34.82	41.96
MRK	188.84	5.82	56.70	56.68	7.43
JPM	414.35	11.49	65.48	65.46	9.47
WFC	275.29	10.91	50.90	50.89	18.02
C	472.90	12.19	46.82	46.81	14.19
	$AvgBS1$ (100 shrs)	$AvgAS2$ (100 shrs)	$AvgBS2$ (100 shrs)	$AvgAS3$ (100 shrs)	$AvgBS3$ (100 shrs)
IBM	2.17	1.95	2.26	2.09	2.26
MSFT	24.53	28.12	31.06	33.90	35.37
T	33.76	43.63	41.96	55.53	63.67
JNJ	3.62	5.86	4.44	7.74	4.90
PFE	42.29	48.07	48.09	50.94	55.68
MRK	7.36	14.34	11.30	24.20	13.87
JPM	9.45	13.10	11.82	17.41	15.09
WFC	17.01	20.68	17.72	23.58	19.05
C	12.97	18.58	16.48	22.23	19.60

Table 2: Summary statistics for all stocks in the sample. We report respective average values over the sample period. $NumObs$ denotes the average daily number of observations for each stock. $AvgTrd$ is the average daily number of trade executions. $AvgAP1$ ($AvgBP1$) is the average best ask (bid) price, $AvgAS1$ ($AvgBS1$) is the average volume of the market ask (bid) orders (1st level of LOB). $AvgAS2$ and $AvgBS2$ are the average volumes of best ask and bid limit orders (2nd level of LOB), and $AvgAS3$ and $AvgBS3$ are the volumes of orders placed at 3rd level of LOB.

the execution of a big market order against several (smaller) limit orders if they occur within less than one second after the previous trade and have the same initiation type. All such sub-trades are consolidated into a single trade. All orders and trades are time stamped up to the nanosecond and signed by a "Direction" and an "Event Type" ticker in the message file (see Figure 1). Each such quote is then associated with the corresponding trade and limit order book information. In particular, the k-th row in the "message" file (upper panel of Figure 1) describes the limit order event causing the change in the limit order book from line $k-1$ to line k in the "orderbook" file (lower panel of Figure 1).

For our study, we consider the mid price and the corresponding bid and ask sizes on the first three levels of LOB for each stock, i.e. we use the market order, best limit order and 2nd best limit order. As the LOBSTER reconstruction of the LOB mitigates large unilateral shifts in the LOB by

Time (sec)	Event Type	Order ID	Size	Price	Direction
⋮	⋮	⋮	⋮	⋮	⋮
34713.685155243	1	206833312	100	118600	-1
34714.133632201	3	206833312	100	118600	-1
⋮	⋮	⋮	⋮	⋮	⋮

"message" file

Ask Price 1	Ask Size 1	Bid Price 1	Bid Size 1	Ask Price 2	Ask Size 2	Bid Price 2	Bid Size 2	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1186600	9484	118500	8800	118700	22700	118400	14930	...
1186600	9384	118500	8800	118700	22700	118400	14930	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

"order book" file

Figure 1: Structure of the LOBSTER data. In the message file, the "Direction" ticker indicates whether a trade is initiated by a buyer or seller with '-1' for sell and '1' for a buy trade. The "Event Type" ticker marks the trading type: 1: Submission of a new order, 2: Cancellation (partial deletion of a order order), 3: Deletion (total deletion of a market/limit order), 4: Execution of a visible limit order, 5: Execution of a hidden limit order etc.

iceberg type trading effects, the focus on the mid-price appears justified for highly liquid stocks (see Table 2). As for liquid stocks, tick levels close to the best quotes are mostly filled, all limit prices can be determined with known constant distance from the best prices. Therefore we conclude that in our case, they do not contain any additional information but would only increase the dimension of the system.[4] Moreover, our own pre-studies and previous univariate studies abstracting from cross-effects between different assets (see e.g. Hautsch and Huang (2012)) have both shown that aggressive limit orders placed close to the best ask and bid have the highest market impact, with substantially lower price effects from increased distance from the spread. Accordingly, we focus only on the best three price levels in the book.

2.2 Volume Synchronization Algorithm

This section is devoted to the necessary data preparation steps addressing challenges of multivariate high-frequency data such as non-synchronicity and microstructure noise. In particular, we propose a

[4]For less liquid stocks, however, the limit prices might contain valuable information and should be included such as deeper levels of the book. It is possible to handle these larger scale situations with the presented methods but at the expense of precision of estimates with increasing dimensions in the system.

volume synchronization algorithm for our high-dimensional system of price and volumes at different levels across assets mitigating both problems.

For HFT price data at very high sampling frequencies, it is well documented that microstructure noise arising e.g. from market frictions contaminates the observed prices. While such noise-induced bias can of course be decreased with less frequent sampling at the cost of less precise estimates, generally an optimal sampling frequency attempts to balance the bias-variance tradeoff (see, e.g. Bandi and Russell (2006), Aït-Sahalia et al. (2005), Bandi and Russell (2008)). When working with the LOB, however, there are not only prices but also volumes at different levels of the LOB which must be aligned in time. Therefore we propose the size intensity $\tilde{S}_{t_j}$ as a trading activity measure,

$$\tilde{S}_{t_j} = S_{t_j}(t_{j+1} - t_j) \tag{1}$$

where t_j denotes the time stamp of jth LOB, S_{t_j} is the corresponding volume at t_j. While a sufficiently large time interval for prices helps to reduce microstructure noise, the time interval for trading volumes should be small enough to detect effects specific large orders in the book. In order to match price and volume frequencies extracting the maximum possible amount of information, we aggregate size intensities over the same moderate size time interval which is required for the pre-averaging of returns to eliminate microstructure noise.

Thus we suggest a pre-averaging approach in returns and volumes to achieve synchronization and elimination of microstructure noise. In the following we illustrate the key steps for preparing the raw HFT data. All technical details can be found in Appendix A. For ease of illustration, the volume synchronization algorithm can be divided into four steps,

Step 1: Set equally-spaced k time intervals starting at time T_0

$$T_0 + k\Delta T, \quad k = 0, 1, 2, \ldots, K$$

Step 2: Define the price $\tilde{P}$ and size $\tilde{S}$ at time $T_0 + k\Delta T$ as

$$\begin{aligned} \tilde{P}_{T_0+k\Delta T} &= P_{t_m},\ t_m = \max\{t_j;\ t_j \leq T_0 + k\Delta T\} \\ \tilde{S}_{T_0+k\Delta T} &= \sum_{T_0+(k-1)\Delta T \leq t_j \leq T_0+k\Delta T} S_{t_j}(t_{j+1} - t_j) \end{aligned}$$

where $\tilde{S}_{T_0+k\Delta T}$ are the size intensities defined in (1).

Step 3: Compute the changes of the log values to mitigate the impact of outliers and measure relative changes

$$\begin{aligned} \Delta p_{T_0+k\Delta T} &= \log \tilde{P}_{T_0+k\Delta T} - \log \tilde{P}_{T_0+(k-1)\Delta T} \\ \Delta s_{T_0+k\Delta T} &= \log \tilde{S}_{T_0+k\Delta T} - \log \tilde{S}_{T_0+(k-1)\Delta T} \end{aligned}$$

Step 4: Pre-average both $\Delta p_{T_0+k\Delta T}$ and $\Delta s_{T_0+k\Delta T}$ for removing microstructure noise

$$\Delta\tilde{p}_{T_0+k\Delta T} = \sum_{j=0}^{J} g_j \Delta p_{T_0+j\Delta T}, \quad \Delta\tilde{s}_{T_0+k\Delta T} = \sum_{j=0}^{J} g_j \Delta s_{T_0+j\Delta T}$$

where $g_j \geq 0$ and $\sum_{j=0}^{J} g_j = 1$, the details are in Appendix A.

Preparing data in this way alleviates microstructure noise, matches the price to the size in a moderate interval and solves the problem of non-synchronicity. We use volumes in log-scale to mitigate the effect of outliers as is standard in the empirical literature.

We divide the trading period into 1-minute intervals and pre-average both $\Delta\tilde{p}_t^{(n)}$, $\Delta\tilde{s}_t^{br(n)}$ and $\Delta\tilde{s}_t^{ar(n)}$ to reduce microstructure noise over 15-min, yielding 375 observations per day. With this we do not only mitigate microstructure noise but also ensure that the considered data is only mildly dependent over time which is crucial for the subsequent estimators of the large-dimensional system over time to yield consistent results. We explicitly checked that both final price and volume components are stationary due to the first differencing in the algorithm with respective univariate KPSS and ADF tests.[5]

[5] Results are omitted here for the sake of brevity but are available upon request.

3 Statistical Learning Techniques

Our goal is to extract the time-evolution of the LOB system across multiple assets over time in order to quantify the full market impact of a limit order. For each stock, we consider the mid-price on the first level and the corresponding bid and ask sizes on the first three levels jointly in the stock-specific variable $y_t^n \in \mathbb{R}^{7\times 1}$ defined as

$$y_t^{(n)\top} = [p^{(n)}, as_1^{(n)}, as_2^{(n)}, as_3^{(n)}, bs_1^{(n)}, bs_2^{(n)}, bs_3^{(n)}] \tag{2}$$

where $p^{(n)} = \Delta\tilde{p}_t^{(n)}$ is the pre-averaged price factor for stock n, $as_r^{(n)} = \Delta\tilde{s}_t^{ar(n)}$ stands for the corresponding rth level of ask size factor, whereas $bs_r^{(n)} = \Delta\tilde{s}_t^{br(n)}$ stands for the rth level of bid size factor for stock n. As our sample only consists of highly liquid stocks we do not include quote levels associated with $as_r^{(n)}$ and $bs_r^{(n)}$ in $y_t^{(n)}$. For such highly liquid stocks the spread is only narrow and the tick levels close to the best quotes are generally all filled such that limit prices occur on a fixed grid with constant distance to the corresponding best quotes. So in these cases, there is no gain in information from including the additional price data.[6] Though, for less liquid stocks, our model and the empirical strategy below can be readily extended to include these limit prices in $y_t^{(n)}$ as well.[7] This comes, however, at the expense of a much higher dimensionality of the empirical problem which then requires more observations and computational power to reach the same precision in estimation.

By stacking the LOB-vector $y_t^{(n)\top}$ for different N stocks together, we obtain the full system vector $Y_t \in \mathbb{R}^{7N\times 1}$ with LOB components as

$$Y_t^\top = [y_t^{(1)\top}, y_t^{(2)\top}, \ldots, y_t^{(N)\top}]; \tag{3}$$

By studying the time evolution of the large-dimensional Y_t in the following, we can characterize the market impact of a limit order and resulting network cross-effects over time.

[6] See also Hautsch and Huang (2011) who show that for liquid stocks trades absorbing more than one price level in the limit order book occur extremely rarely.

[7] The data input into the model can also be transformed to directly capture the dynamics relative spread changes or (ask-bid) depth imbalances.

3.1 A High-dimensional VAR as Statistical Learning Tools for Quotes and Depths

From our volume synchronization algorithm, all components in the LOB-vector $y_t^{(n)}$ are stationary and only weakly dependent. Therefore we can model the full system vector Y_t of quotes and depths of all considered assets from equation (3) with a VAR(p) model

$$Y_t \quad = \quad A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + u_t \tag{4}$$

where $Y_t \in \mathbb{R}^K$ with $K = 7N$ for N stocks in the portfolio for $t = 1, \ldots, T$. A_i are fixed but unknown $(K \times K)$ coefficient matrices[8] for $i = 1, \ldots, p$ and p is the unknown lag order allowing for influences of observations further in the past. By $u_t = (u_{1t}, u_{2t}, \ldots, u_{Kt})^\top \in \mathbb{R}^K$ we denote the *iid* unobserved innovations with variance Σ_u. Please note that we do not require a Gaussian distribution of u_t for the subsequent procedure which would be hard to justify for our LOB-vector Y_t, but we only need that each element in u_t has bounded $(4+\delta)$th moment with $\delta > 0$. This implies a non-trivial but crucial generalization of standard VAR estimation techniques in a set-up of large dimensions K as e.g. in Kock and Callot (2015).[9]

For estimating the unknown coefficients $A_1, \ldots, A_p$ we use the following matrix form of equation (4)

$$Y = AZ + U \tag{5}$$

with $Y = (Y_1, Y_2, \ldots, Y_T)$, $A = (A_1, A_2, \ldots, A_p)$, $Z = (Z_1, \ldots, Z_T)$, $U = (u_1, \ldots, u_T)$ and $Z_t = (Y_{t-1}, \ldots, \ldots, Y_{t-p})^\top$. We can re-write this in an OLS-vectorized version as follows

$$\begin{aligned} vec(Y) &= (Z^\top \otimes I_K) vec(A) + vec(U) && (6) \\ \mathbf{y} &= \mathbf{x}\beta + \mathbf{u} && (7) \end{aligned}$$

with $\mathbf{y} = vec(Y) \in \mathbb{R}^{KT \times 1}$, $\mathbf{u} = vec(U) \in \mathbb{R}^{KT \times 1}$, and $\beta = vec(A) \in \mathbb{R}^{K^2 p \times 1}$. Hence, the number

[8]Stationary of the VAR(p) in (4) ensured by standard conditions on A_i requiring the roots of $|I_K - \sum_{j=1}^{p} A_j z^j| = 0$ to lie outside unit circle.

[9]This is direct consequence of now available oracle results for dependent innovations from the statistical literature, see e.g. Wu and Wu (2016).

of parameters in β to be estimated is pK^2 from a total number KT of observations. Thus the ratio h of estimated elements relative to the number of observations is $h = \frac{Kp}{T} = \frac{7Np}{T}$ for N assets in the portfolio. In our empirical analysis later on, we re-estimate the VAR on a daily basis from 375 pre-averaged observations for 9 stocks in the portfolio in order to allow for potential time-variation in the VAR coefficients. Note that even for this moderate size portfolio we have $h > 0.5$ if only $p = 3$ lags enter the model, and $h > 1$ when $p \geq 6$ lags. This makes the model high-dimensional requiring techniques different from traditional OLS (see e.g. Lütkepohl (2005) for standard VAR estimation) in order to identify components in (7) from estimation. Note that for larger portfolios with higher N and more lags p even cases of $h > 1$ can easily occur, where there are more unknown parameters than observations.

A way to identify parameters in such cases is by imposing so-called sparsity, which implies that of the K^2p many elements in fact only a small number is effectively non-zero while the influence of the remaining parameters is negligible. This means that we just need to identify the active set from the data which contains all non-zero parameters, and then in this lower-dimensional set, identification and estimation are possible. For our LOB-system, sparsity appears as a plausible assumption as in fact we expect only a few components of the LOB to exert relevant effects at each point in time. In particular, we assume that the unknown components in (4) satisfy $\|\Sigma_u\|_2 < \infty$ and $\|(A_1, A_2, \ldots, A_p)\|_2 < \infty$.[10] Note that formally this technical Assumption implies a bound for column- and row-wise sums of elements of Σ_u and A in this large and fixed dimensional case.[11] If in each column or row the number of non-zero elements is small, i.e. when the matrices are sparse and only a limited amount of potential components of the full system LOB vector are relevant, these bounds are generally fulfilled.

For identification of the active set and estimation of the relevant variables in one step, we employ the following elastic net approach. We estimate $\hat{\beta}$ by minimizing the objective function $Q(\beta)$

$$Q(\beta) = \|\mathbf{y} - \mathbf{x}\beta\|_2^2 + \alpha_{1,T}\|\beta\|_1 + \alpha_{2,T}\|\beta\|_2^2 \tag{8}$$

[10] Note that norms here are matrix norms where $||A||_2$ denotes the maximum eigenvalue of $A^\top A$.

[11] Generally, in a statistical high-dimensional set-up, one considers also cases where the dimension $K(T)$ increases with sample size T. For our case here, the dimension is large but fix in T. Therefore the stated assumptions are sufficient for sparsity. Also a the usual $\beta_{\min}$-condition for the determination of the active set (see Kock and Callot (2015)) is trivially fulfilled for fixed K.

which is equivalent to minimizing $Q(A_1, A_2, \ldots, A_p)$ in components of A[12]

$$Q(A) = \sum_{t=1}^{T} \|Y_t - \sum_{j=1}^{p} A_j Y_{t-j}\|_2^2 + \alpha_{1,T} \sum_{j=1}^{p} \|vec(A_j)\|_1 + \alpha_{2,T} \sum_{j=1}^{p} \|vec(A_j)\|_2^2 \,. \tag{9}$$

Both versions of the criterion function employ two penalty terms which shrink the set of elements to the active set in the usual OLS squared loss function. While other penalized techniques such as e.g. the Lasso would also impose this necessary shrinkage, we prefer the elastic net to remedy the potentially strong correlation among regressors with the additional second penalty term in (8). Under sparsity, the estimator $\hat{\beta}$ of (8) is consistent in identifying the active set as those components which are estimated as different from zero. Moreover it contains mean squared error optimal estimates of all non-zero components for appropriate penalties $\alpha_{1,T}$ and $\alpha_{2,T}$.[13] Though, as estimates of relevant variables from penalized procedures are generally biased in finite samples, we re-estimate the model for the relevant variables only with OLS after model selection of the active set with the penalized approach (8) or (9) in the first step. This generally reduces shrinkage bias and ensures better model performance.

For deriving an estimate $\hat{\beta}$ from (8) or (9), a data-driven choice of tuning parameters $\alpha_{1,T}$ and $\alpha_{2,T}$ is crucial in order to control the imposed regularization. If regularization parameters are too large only a small amount of relevant variables will be admitted, missing out on potentially influential ones, i.e. throwing away useful information. In contrast, we lose identification and estimation precision if α is too small and the active set is too large. To balance sparsity and estimation accuracy, we choose moderately small tuning parameters using the Bayesian information criterion (BIC).[14] The model selection consistency results are available in Appendix C.

[12] Here, for vectors $v \in \mathbb{R}^d$ we use the usual norms $||v||_1 = \sum_{i=1}^{d} |v_i|$ and $||v||_2^2 = \sum_{i=1}^{d} |v_i|^2$.

[13] Estimates in the active set are also consistent for suitably vanishing penalties $\alpha_{1,T}$ and $\alpha_{2,T}$. Details are a direct consequence of results in Kock and Callot (2015).

[14] see e.g.Liang and Schienle (2018) for simulation evidence that BIC works well for penalty choices in time series settings.

3.2 Estimating the Market Impact for a Large System

We quantify the market impact of a shock in limit orders as its implied expected influence on the ask and bid over time after its occurrence. In order to capture this effect, we use the generalized impulse response function GI (see Lanne and Nyberg (2016))

$$GI(l, \delta_{jt}, \mathcal{F}_{t-1}) \quad = \quad \mathsf{E}(Y_{t+l} \mid u_{jt} = \delta_{jt}, \mathcal{F}_{t-1}) - \mathsf{E}(Y_{t+l} \mid \mathcal{F}_{t-1}) \tag{10}$$

with the information set $\mathcal{F}_{t-1}$ consisting of all information entering (4) up to time point $t-1$ and a "shock" δ_{jt} in the j-th component of the system Y_t in (4). Generally, we use an impulse response type analysis in order to study the impact of a shock over time.

For a realistic empirical market impact analysis, we note that "shocks" in our LOB system might in practice well be endogenous, depending on the common history in all components and might also occur simultaneously in different components. Thus as a reference point, for the size of the "shock" we take the underlying dynamic VAR specification of the system up to t and measure typical simultaneous deviations across all j components at t. In particular, we focus on the case with

$$\delta_{jt} = \hat{u}^{\star}_{jt} \tag{11}$$

where $\hat{u}^{\star}_{jt}$ marks the j-th component in $\hat{u}^{\star}_{t}$ which is a random independent draw for time point t when sampling with replacement from the set of model-implied residuals $\{\hat{u}_t\}_{t=1}^{T}$ of (4) to get a "typical" deviation. Thus resampling from the distribution of those residuals, gives a realistic size of "shock" to be checked for the respective component given the dynamic specification. In this way, the considered shocks for the market impact of limit orders are endogenous and can occur simultaneously in several components leading to results which are fairly robust to the effective choice of underlying time series specification of the system. This is in contrast to a technical pseudo experimental set-up of a separate idiosyncratic unit shock from component j, i.e. taking $\delta_{jt} = 1$ with $\delta_{kt} = 0$ for $k \neq j$ which we use in Section 5 to derive the direction of price effects.

Moreover, in order to measure the propagation effect of a shock $\delta_t = (\delta_{1t,\ldots,\delta_{Kt}})$ at t on the system Y_{t+k} with $k > 0$, the basic object in (10) is the conditional expectation. However the penalized elastic net estimation of a high-dimensional VAR is nonlinear, thus the GI of (10) cannot be expressed as a closed form function over the forecast horizon l in contrast to standard impulse response functions for low-dimensional VARs. Therefore, we must use bootstrap resampling to compute GI for different l. We provide the detailed procedure in Appendix B.

For a fixed l or when taking a (weighted) average over different forecast horizons, studying the m-th component of $GI(\delta_{jt})$ shows the directed impact of a shock in j on m. Moving along all different component pairs we can use this to build a network of LOB-impact discussed in the following subsection.

3.3 Network Construction & Evaluation

We work with a standardized version of GI from (10) normed to $[0, 1]$ for easier interpretability of values across stocks. We thus define

$$\lambda_{ij,\mathcal{F}_{t-1}}(h) = \frac{\sum_{l=0}^{h} GI_i(l, \delta_{jt}, \mathcal{F}_{t-1})^2}{\sum_{j=1}^{K}\sum_{l=0}^{h} GI_i(l, \delta_{jt}, \mathcal{F}_{t-1})^2}, \quad i, j = 1, \ldots, K \tag{12}$$

where h is the maximal considered forecast horizon in the average and GI_i is the i-th component of GI. Obviously, $\lambda_{ij,\mathcal{F}_{t-1}}(h) \in [0, 1]$ measures the relative impact of a shock δ_{jt} of size (11) in the j-th component on element i of Y_t, relative to the total impact on i from shocks $\delta = (\delta_{1t}, \ldots, \delta_{Kt})$ according to (11) in all K components of Y_t at t after h periods.

The standardized GI links our estimated high-frequency market impact effects to the network literature on connectedness (see e.g. Diebold and Yılmaz (2014)). In particular, this literature generally characterizes systemic risk spillover effects as connectedness obtained from a generalized forecast error variance decomposition (GFEVD) of an underlying VAR (see Diebold and Yılmaz (2014) and Demirer et al. (2017)). It can be shown that directed spill-overs from GFEVD are scaled versions of our standardized generalized impulse response function $\lambda_{ij,\mathcal{F}_{t-1}}(h)$ (Lanne and Nyberg (2016)). In this sense, our bootstrap technique described below also opens up a way for this systemic risk literature to move from so far low-dimensional systems to estimation of large-dimensional directed networks.

For estimation of the standardized *GI* $\lambda_{ij,\mathcal{F}_{t-1}}$ and its respective confidence intervals in our high-dimensional set-up, we propose a bootstrap procedure. For the details of the computational steps of the resampling we refer to Appendix B. Note that with the bootstrap estimate $\lambda^b_{ij,\mathcal{F}_{t-1}}$ we preserve the sparsity imposed on the high-dimensional VAR estimation of (8) or (9) in order to focus on relevant network spill-over effects. This is a necessary key difference for high-dimensional systems to the standard estimation approach for impulse response functions in small fixed dimensions which relies on a moving average (MA) transformation of (9). In this way, however, it fails to keep the sparsity structure and hence encounters numerical accuracy or even feasibility problems in high dimensions. Moreover, the presented bootstrap technique also works for underlying nonlinear models and is not specific to a linear dynamic structure of the system as the MA-transformation. In particular, the numerical techniques for conditional mean forecast from nonlinear models for more than one period ahead are implemented in this paper for bootstrapping $\lambda_{ij,\mathcal{F}_{t-1}}(h)$ (see more details in Teräsvirta et al. (2010)).

For any forecast horizon h, the bootstrapped estimate $\lambda^b_{ij,\mathcal{F}_{t-1}}(h)$ of $\lambda_{ij,\mathcal{F}_{t-1}}(h)$ marks the directed impact of a shock in j on i in percentages and can be displayed as the directed effect of the node j on i in a network at time t. We collect all possible directed cross-effects in a corresponding Table 3 where the entries correspond to the network adjacency matrix. We refer to this table also as a connectedness table linking it to the network literature outlined above.

	x_1	x_2	$\dots$	x_K	From others
x_1	$\lambda^b_{11}(h)$	$\lambda^b_{12}(h)$	$\dots$	$\lambda^b_{1K}(h)$	$\sum_{j=1}^K \lambda^b_{1j}(h), j \neq 1$
x_2	$\lambda^b_{21}(h)$	$\lambda^b_{22}(h)$	$\dots$	$\lambda^b_{2K}(h)$	$\sum_{j=1}^K \lambda^b_{2j}(h), j \neq 2$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_K	$\lambda^b_{K1}(h)$	$\lambda^b_{K2}(h)$	$\dots$	$\lambda^b_{KK}(h)$	$\sum_{j=1}^K \lambda^b_{Kj}(h), j \neq K$
To others	$\sum_{i=1}^K \lambda^b_{i1}(h)$ $i \neq 1$	$\sum_{i=1}^K \lambda^b_{i2}(h)$ $i \neq 2$	$\dots$	$\sum_{i=1}^K \lambda^b_{iK}(h)$ $i \neq K$	$\frac{1}{K}\sum_{i=1,j=1}^K \lambda^b_{ij}(h)$ $i \neq j$

Table 3: Connectedness table of interest, estimated by bootstrap-based methods.

Accordingly, for each node i in the network we work with the following quantities. We denote the pairwise directional connectedness $C_{i\leftarrow j}$ from j to i by

$$C_{i\leftarrow j} = \lambda^b_{ij}(h) \tag{13}$$

Moreover, aggregating all effects of component i on other elements in the system, we call the total directional connectedness "to" $C_{i\leftarrow\bullet}$ (others to i) given by $C_{i\leftarrow\bullet} = C_{to,i} = \sum_{j=1}^{K} \lambda_{ij}^{b}(h), \quad$ for $i \neq j$. Note that by definition (12) we have $C_{i\leftarrow\bullet} = 1$. Analogously, the total directional connectedness "from" $C_{\bullet\leftarrow i}$ (from i to others) is defined as $C_{\bullet\leftarrow i} = C_{from,i} = \sum_{l=1}^{K} \lambda_{li}^{b}(h), \quad$ for $l \neq i$ and the corresponding net total directional connectedness C_i measures the direction and magnitude of the net spillover impacts of node i in the system as $C_i = C_{from,i} - C_{to,i} = C_{\bullet\leftarrow i} - C_{i\leftarrow\bullet} = C_{\bullet\leftarrow i} - 1$. Note that for any specified h, the estimated networks in Table 3 are time-varying as they are re-estimated in each rolling window where a new dynamic VAR specification is admitted (see above equation (7)). Thus for an overview over an investment period, we report the mean μ, median m and quantile $Q(\alpha)$ for levels $0 < \alpha < 1$ of all node impact measures aggregated over time for a specified h. [15] In the following we generally work with $h = 30$ which yields aggregated effects on a forecast horizon of 7.5 hours for observations at the 15-minute scale thus comprising around one business day.

3.4 Network Aggregation Types and Characteristics

We work with spillover networks as constructed in Table 3 on different aggregation levels. In our high-dimensional set-up, this allows to focus and structure results according to our research questions. We therefore distinguish three different types of aggregated networks where each network is a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of two core items: nodes (or vertices) $\mathcal{V}$ and edges $\mathcal{E}$. Nodes are the entities we are evaluating and edges are the connections between them.

Thus for the full LOB-network, the ordered nodes $\mathcal{V}$ are the elements of Y in (3), we write $\mathcal{V} = Y$, comprising prices and quotes of all assets $n = 1, \ldots, N$. The corresponding edges consist of all spillovers between these elements $\mathcal{E} = \{C_{i\leftarrow j} = C_{i\leftarrow j}(Y) | i, j = 1, \ldots, 7N\}$. Note that in particular, the full LOB-network contains 7 nodes per asset (1 price and 3 levels of ask-side and bid-side quotes) and spillover effects between assets can occur on all levels of the LOB. Thus, we study effects on prices of an aggregated ask and bid side where aggregation is either by levels of the LOB or by stocks per level of the LOB in a sparser network in order to distill the different channels of

[15] In particular, $\mu = \mu(C)$ and $Q(\alpha) = Q(C)(\alpha) = F^{-1}(\alpha) = \inf\{C : F(C) \geq \alpha\}$ where F is the empirical distribution of $C \in \{C_i, C_{i\leftarrow j}, C_{i\leftarrow\bullet}, C_{\bullet\leftarrow i}, deg_i, outdeg_i, indeg_i, Clos_i, Bet_i\}$ over time.

the LOB market impact.

Case 1: Own-price/Cross-price market impact For the own-price market impact, we study the effects of quotes on prices only within the LOB of each asset. Thus we work with the nodes of the full LOB-network $\mathcal{V} = Y$ but focus only on a selected subset of the edges in $\mathcal{E}$. Hence we have

$$\mathcal{E}_o = \left\{ C^o_{p^{(k+1)} \leftarrow as_r^{(k+1)}}, C^o_{p^{(k+1)} \leftarrow bs_r^{(k+1)}}, \text{ for } k = 0, \ldots, N-1, r = 1, \ldots 3 \right\} \tag{14}$$

Hence the network $\mathcal{G}_{own} = (\mathcal{V}, \mathcal{E}_o)$ displays the effects of different levels of quotes on the stock price within the LOB of each stock. In particular, $C^o_{p^{(k+1)} \leftarrow as_r^{(k+1)}}$ contains the effect of the ask-side level r on the own price of the stock $k+1$. For details how these edges and all other edge sets in the following are are calculated from $\mathcal{E}$ see Appendix D.

For the effects on prices across stocks, we reduce the complexity of the network by aggregating spillover effects of each ask and bid side per stock. For studying this cross-price market impact, we thus work with $\mathcal{G}_{cross} = (\mathcal{V}_1, \mathcal{E}_{c,1})$, where the ask-side (bid-side) of each stock is represented by a single node $as^{(n)}(bs^{(n)})$. Hence we define $\mathcal{V}_1 = \left((p^{(1)}, as^{(1)}, bs^{(1)}), \ldots, (p^{(N)}, as^{(N)}, bs^{(N)})\right)^\top$ and

$$\begin{aligned} \mathcal{E}_{c,1} &= \left\{ C^1_{p^{(k+1)} \leftarrow as^{(z+1)}}, C^1_{p^{(k+1)} \leftarrow bs^{(z+1)}}, \text{ for } k \neq z = 0, \ldots, N-1 \right\} \\ \mathcal{E}_{o,1} &= \left\{ C^1_{p^{(k+1)} \leftarrow as^{(k+1)}}, C^1_{p^{(k+1)} \leftarrow bs^{(k+1)}}, \text{ for } k = 0, \ldots, N-1 \right\} \end{aligned} \tag{15}$$

Note that $\mathcal{V}_1$ only consists of $3N$ nodes, where $C^1_{p^{(k+1)} \leftarrow as^{(z+1)}}$ marks the aggregated spillover effect of the ask-side of stock $z+1$ on the price of stock $k+1$ in the reduced network $\mathcal{V}_1$ with only 3 nodes per stock. For comparison, we also define own-price market effects in this aggregated network in (15) studying $(\mathcal{V}_1, \mathcal{E}_{o,1})$, $\mathcal{E}_{o,1}$ is the special case of $\mathcal{E}_{c,1}$ when $k = z$.

Case 2: Trade imbalance between bids and asks To measure potential asymmetry of bid and ask side effects on different stock prices, spillovers are aggregated across stocks in each quote level to determine the respective overall impact. We thus work with a network with graph $\mathcal{G}_s = (\mathcal{V}_2, \mathcal{E}_{s,2})$ where each level r of the ask (bid) side is represented by as_r (bs_r). It is $\mathcal{V}_2 =$

$\left((p^{(1)},\ldots,p^{(N)}),(as_1,as_2,as_3),(bs_1,bs_2,bs_3)\right)^\top$

$$\mathcal{E}_{s,2} = \left\{C^2_{p^{(k+1)}\leftarrow as_r}, C^2_{p^{(k+1)}\leftarrow bs_r},\ \text{for } k=0,\ldots,N-1,\ r=1,2,3\right\} \tag{16}$$

The number of nodes in $\mathcal{V}_2$ is only $N+6$ in this case, where $C^2_{p^{(k+1)}\leftarrow as_r}$ marks the aggregated spillover effects of the ask-side level r from all stocks on the price of stock $k+1$ in the reduced network $\mathcal{V}_2$. Note that in $\mathcal{V}_2$ we distinguish between different effects on the prices of different stocks. For an overview, however, we also look at asymmetric effects aggregated in the price effects across stocks. Hence we work with $\mathcal{G}_{sp} = (\mathcal{V}_3, \mathcal{E}_{s,3})$ where the price is denoted as p. We set $\mathcal{V}_3 = (p,(as_1,as_2,as_3),(bs_1,bs_2,bs_3))^\top$ and

$$\mathcal{E}_{s,3} = \left\{C^3_{p\leftarrow as_r}, C^3_{p\leftarrow bs_r},\ \text{for } r=1,2,3\right\} \tag{17}$$

In this network $\mathcal{G}_{sp}$, all effects are aggregated across stocks. The number of nodes is thus only 7 comprising aggregate price and ask and bid side effects on all three levels.

Case 3: Price-only effects and the fully LOB-aggregated network per stock For studying only price effects within the LOB-network, we reduce the number of nodes to prices only and select out $\mathcal{E}$ the appropriate edges between prices only. Thus we consider a network with graph $\mathcal{G}_p = (\mathcal{V}_4, \mathcal{E}_{p,4})$, where $\mathcal{V}_4 = Y_p = \left(p^{(1)},\ldots,p^{(N)}\right)$ and we denote the selected edges as $\mathcal{E}_{p,4} = \left\{C^{p,4}_{i\leftarrow j}\right\}$.

The price-only case is compared to the case where we aggregate all pairwise LOB-network spillover effects between two stocks. Thus we work with the network $\mathcal{G}_g = (\mathcal{V}_5, \mathcal{E}_{g,5})$ where each stock is represented by a node $g^{(n)}$. We have $\mathcal{V}_5 = (g^{(1)}, g^{(2)},\ldots,g^{(N)})^\top$ and we denote the aggregated edges as $\mathcal{E}_{g,5} = \left\{C^{g,5}_{i\leftarrow j}\right\}$.

4 Empirical Results

In this section, we focus on the key features of the full LOB-network over time. In particular, we determine the overall importance of cross-effects in the system by deriving own- and cross-price market impacts for our relatively small portfolio setup of 9 representative stocks (case 1). If derived effects already matter in this setting, their effect in larger portfolios and systems will scale up with the dimension. We also investigate if and where differences in the ask and bid-side impact occur and are statistically significant such that they could be employed for investment and hedging decisions (case 2). Moreover, we highlight the importance of limit order book information in quantifying the full market price effects (case 3). Thus we generally report summarized mean and quantile network impact measures over the considered time horizon of two months, but also study different subperiods around the Brexit vote separately. For the net effects, we also report the full time evolution in the last subsection.

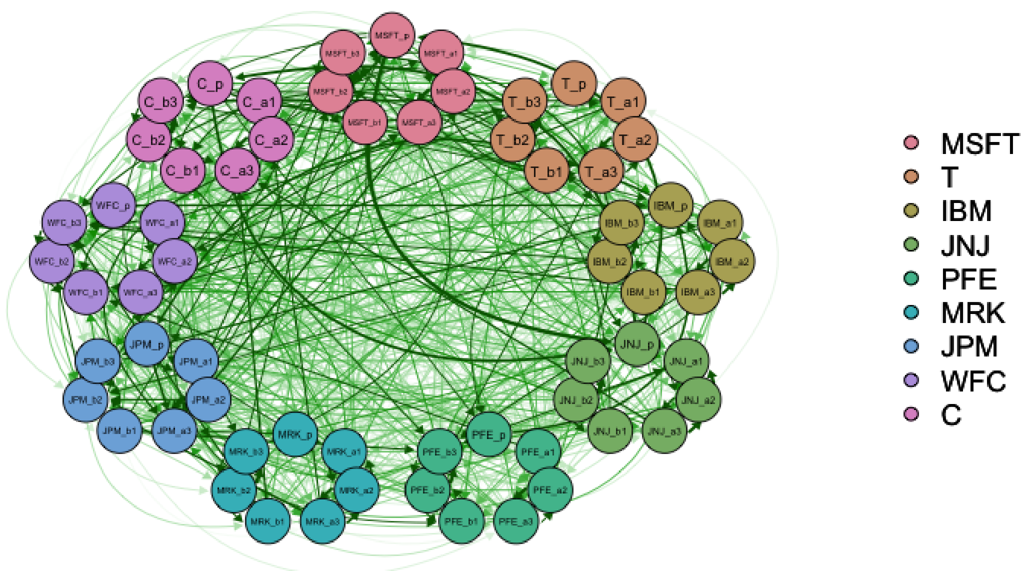

Figure 2: **LOB-network averaged over the whole sample period.** The figure visualizes the mean of the pairwise directional connectedness over the selected two month investment period. Specifically, the price factor and size factors that belong to the same stock appear in the same color. The network impacts are drawn as directed lines connecting two nodes, and quantified by the width of edges between two nodes.

Generally, we calculate the bootstrapped normalized market impact $\lambda^b_{ij,\mathcal{F}_{t-1}}(h)$ and the resulting full LOB-network spillover matrix C (see Table 3) at horizon $h = 30$ comprising about one business day for every trading day from the respective estimates of the sparse high-dimensional VAR model (9). Thus, for each trading day, we obtain a corresponding full LOB impact network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The averaged value of $\mathcal{G}$ over the selected two month investment period is depicted in Figure 2. However, it is not easy to decipher useful information based on all pairwise connectedness. We take that as a start, but in order to address the key points above, we work with tailored aggregated versions.

4.1 Market Impact: Overall Size and Spillover Channels

In this subsection, we investigate the mean size and distribution of the LOB market price impact over the two months sample 06/01/2016 to 07/31/2016. Thus by determining time average spillover effects and their interquartile range to assess associated risk and deviations, we study key structural features of the market as a base for LOB-based investment strategies.

4.1.1 The Importance of Cross-asset and Cross-price Market Impacts

With our approach for each stock, we can quantify the share of own-price market impact originating from different sides of the stock's own LOB with values in $\mathcal{E}_{o,1}$ of (15) versus the cross-price effects of $\mathcal{E}_{c,1}$ from the level-aggregated network $\mathcal{G}_{cross} = (\mathcal{V}_1, \mathcal{E}_1)$ for each trading day. Separating the quote effects by stock allows for detailed insights into spillovers between stocks originating from different sides of the book. In this way, we detect liquidity cross-effects among stocks as captured by the LOB.

Stressing the importance of cross-asset versus own-asset effects aggregating all own-price LOB components, we also calculate the cross-asset impact as the aggregated cross-price effects of prices and quotes from all other assets on each stock. In general, all different cross-asset price effects among the considered 9 stocks account for at least 30 % of total effects on prices and can amount to up to 45% in the case of the impact of Citigroup's stock price. This follows from Tables 4, 5, and 6, where row-wise all impacts from both sides of the LOB and the prices sum to 1 by definition (12) such that any element in a specific row in each of the three tables denotes the percentage of received overall price impact of the respective company in this row of the first column from the specific source displayed in the table. Among these cross-asset effects, generally quote effects from the LOB dominate price only effects comprising 20 to 26 percentage points of the above shares. For all stocks, ask-side impacts exceed bid-side ones within the LOB-effects by up to 60% (JPM), see Tables 4 and 5) indicating market asymmetries.

On a more detailed level, Table 4 provides the time average of the ask side effects which we complement and contrast with the corresponding bid-side effects in Table 5. Generally we group all stocks in industry blocks and find substantial cross-price effects on the off-diagonal of each table between stocks and sectors which are not symmetric as expected. With a detailed analysis of $\mathcal{E}_1$ we can compare the effects of trades on stock prices. The cross-price market impacts $\mathcal{E}_{c,1}$ are reported as off-diagonal elements, while own-price effects $\mathcal{E}_{o,1}$ are displayed on the diagonal for the ask-side in Table 4 and the bid-side in Table 5 .

m \ l	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C	$\sum$
MSFT	3.38	0.68	1.82	0.65	1.76	0.73	**5.46**	0.54	3.66	18.68
T	3.08	1.47	1.00	0.62	1.86	2.58	1.08	1.29	2.38	15.37
IBM	1.52	0.38	3.06	1.33	0.91	1.57	1.41	**3.58**	2.05	15.80
JNJ	1.69	0.62	1.04	0.45	1.47	1.05	1.37	0.31	**3.49**	11.47
PFE	1.07	0.96	0.44	0.13	2.06	**4.83**	1.86	2.89	2.12	16.37
MRK	3.18	1.17	0.43	0.83	2.44	**5.27**	**4.15**	2.25	2.57	22.28
JPM	2.09	1.10	1.81	0.72	2.68	1.13	**9.22**	1.34	2.16	22.24
WFC	1.22	2.38	1.70	0.55	1.93	1.22	0.79	**5.37**	0.60	15.77
C	2.55	1.11	2.37	0.84	2.57	1.33	**4.51**	1.23	2.01	18.51

Table 4: **Ask-side**: Each element of the inner table contains the aggregated market impact $\mu(C^1_{p^{(m)}\leftarrow as^{(l)}})$ of the ask trades of stock l on the price of stock m as defined in (15) displayed as mean μ over the entire sample period 06/2016-07/2016. The last column aggregates all ask-side effects on the price of stock m as $\mu(\sum_{l=1}^{N} C^1_{p^{(m)}\leftarrow as^{(l)}})$. All numbers are in percentages. Own-price effects are on the diagonal, while cross-price effects are off-diagonal. Bold numbers highlight the 9 largest relative effects.

On both sides of the book separately, cross-price effects exceed own price effects at least by a factor of 3 and up to 10 on the bid side and up to 22 on the ask side. Individually between firms, cross-effects from financial companies are the largest within the system both, to companies in the tech and the health sector and from both sides of the LOB. Often these impacts from financial firms even exceed the own-price impact of specific firms, see e.g. the spillover of JP Morgan (JPM) on Merck (MRK) and Microsoft (MSFT) on the ask side where the latter is larger than all other own-price and cross-price effects from the ask quotes but also JPM on IBM and Wells Fargo (WFC) on Johnson & Johnson (JNJ) on the bid-side. Within-sector price effects from the LOB are most relevant for the financial companies where own-price effects dominate in particular on the ask-side. This is in contrast to price-only effects where impacts from the LOB are ignored. These are

m \ l	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C	$\sum$
MSFT	**4.02**	2.26	0.62	0.53	0.59	1.67	0.41	0.89	1.61	12.58
T	1.36	0.84	0.22	1.04	1.41	**3.67**	0.92	1.10	1.03	11.60
IBM	0.79	1.29	0.66	0.13	0.58	0.97	**3.47**	1.85	1.15	10.89
JNJ	0.63	0.85	0.30	1.30	0.86	0.99	0.50	1.90	**2.59**	9.93
PFE	2.12	0.36	1.10	0.19	1.13	0.37	1.43	**4.08**	1.23	12.00
MRK	0.72	0.49	0.25	0.25	1.35	**3.37**	1.59	0.84	1.27	10.12
JPM	1.66	0.47	1.25	0.97	1.39	0.59	**3.32**	1.87	2.16	13.68
WFC	1.99	1.29	0.30	0.73	1.37	0.83	1.75	**5.50**	1.67	15.45
C	1.02	1.80	0.42	1.24	0.84	1.08	1.41	1.12	**2.54**	11.48

Table 5: **Bid-side**: Each element of the table contains the aggregated market impact $\mu(C^1_{p^{(m)}\leftarrow bs^{(l)}})$ of the bid trades of stock l on the price of stock m as defined in (15) displayed as mean μ over the entire sample period 06/2016-07/2016. The last column aggregates all bid-side effects on the price of stock m as $\mu(\sum_{l=1}^{N} C^1_{p^{(m)}\leftarrow bs^{(l)}})$. All numbers are in percentages. Own-price effects are on the diagonal, while cross-price effects are off-diagonal. Bold numbers highlight the 9 largest relative effects.

depicted in $\mathcal{G}_p = (\mathcal{V}_4, \mathcal{E}_{p,4})$ with estimated average impact values in Table 6 as defined in Case 3. Here own-price effects on the diagonal largely dominate the picture and cross-price effects appear negligible. Moreover there is no visible dominance of the financial sector in cross-asset spillovers.

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C	$\sum$
MSFT	47.96	1.99	2.72	8.55	2.18	0.02	2.73	1.35	1.25	68.74
T	0.75	64.56	0.40	2.37	1.02	1.49	0.90	0.86	0.68	73.03
IBM	0.23	2.00	59.22	3.46	2.82	0.14	5.07	0.21	0.16	73.31
JNJ	6.24	1.36	2.84	57.33	1.32	0.16	0.70	2.23	6.41	78.60
PFE	0.31	0.94	5.01	1.27	61.76	0.04	1.72	0.13	0.44	71.63
MRK	2.16	1.92	1.78	2.68	0.84	54.34	2.85	0.47	0.55	67.60
JPM	1.32	1.03	2.43	1.36	2.52	0.25	53.94	1.11	0.11	64.07
WFC	0.25	1.28	4.26	0.71	1.56	0.31	1.57	58.76	0.08	68.78
C	4.15	2.47	2.72	6.30	2.78	0.40	0.87	0.77	49.54	70.00
$\sum$	63.38	77.56	81.39	84.03	76.79	57.16	70.37	65.89	59.22	635.78

Table 6: **Price-only effects**: Each element of the table measures the pairwise connectedness $\mu_{\mathcal{E}_{p,4}}$ of the price of stock l on the price of stock m as defined in Case 3 displayed as mean μ over the entire sample period 06/2016-07/2016. All numbers are in percentages.

If consider the structure of the quote effects from different levels of the book, we find that all levels contribute and there is no dominant first level effect. For ease of exposition, we report the level structure of own-price effects on the diagonal of Tables 4 and 5, but the picture of the cross-price effects is similar.[16] Thus we work with the own-price parts of $(\mathcal{V}_2, \mathcal{E}_2)$ and obtain the estimated average granular structure of the own-price market impacts in Table 7 where we mark the maximum effect per stock in bold. Respective maxima appear on all levels often with substantial effects on the third level. As limit orders are stored in the LOB and executed according to price priority, large trading quantities may cause a price drop or rise. In particular, if there is an arrival of a market order that is sufficiently large to match all of the best bids, then the limit order will be updated with a lower best bid price. It follows that the depth of the book at which limit orders are submitted is driving the price.

quote level $r\backslash$ stock m	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
ask level 1	0.47	**0.90**	1.28	0.06	0.31	**2.95**	3.58	**2.47**	0.26
ask level 2	0.34	0.31	0.47	0.13	0.07	1.42	0.38	2.17	**1.43**
ask level 3	2.57	0.26	**1.30**	**0.25**	**1.69**	0.90	**5.26**	0.74	0.32
ask all	3.38	1.47	3.05	0.44	2.07	5.27	9.22	5.38	2.01
bid level 1	0.09	**0.59**	0.14	0.05	**0.60**	1.14	1.12	**2.47**	0.70
bid level 2	1.35	0.18	0.29	0.05	0.43	0.08	0.83	0.89	0.42
bid level 3	**2.58**	0.07	**0.23**	**1.20**	0.11	**2.16**	**1.37**	2.14	**1.42**
bid all	4.02	0.84	0.66	1.30	1.14	3.38	3.32	5.50	2.54

Table 7: **Details of own-price effects in different quotes:** Each element in the white cells of the table contains the aggregated own-price market impact $\mu(C^o_{p^{(m)} \leftarrow as_r^{(m)}})$ for stock m by ask quote level r in the upper part and $\mu(C^o_{p^{(m)} \leftarrow bs_r^{(m)}})$ for stock m by bid quote level r as defined in (14) displayed as mean μ over the entire sample period 06/2016-07/2016. The grey cells aggregate the effects on each side per stock. A bold number highlights the largest relative effect in each column, i.e. per stock.

4.1.2 Trade Imbalance between Bid and Ask Side

For larger orders, the depth of the LOB impacts the price with aggregated level effects, and asymmetries between the two sides of the book reflect market instabilities and investment opportunities. In particular, Tables 4 and 5 suggest a strong asymmetry in the LOB with generally larger ask-side effects than bid-side ones.

In this section, we show that these effects are systematic by formally testing for asymmetry. For ease of exposition, we just report results for aggregated ask/bid size and price factors over all

[16]These results are available on request.

stocks as in $\mathcal{G}_{sp} = (\mathcal{V}_3, \mathcal{E}_{s,3})$ (see (17)). In this network $\mathcal{G}_{sp}$, the price and each of the three ask and bid levels correspond to one node comprising respective cumulative effects across all stocks quote effects on all three levels. Table 8 reports the distribution of aggregated different quote levels on the price in quantiles and detects overall asymmetry in the mean and even more strongly in the median for each quote level despite the slightly skewed distribution. Note that this aggregated overall result of stronger impacts on prices caused by market sell pressure is consistent with stock specific results (See Table 10 below and the corresponding median version in Table 12). We then

quote level	$Q_{\mathcal{E}_{s,3}}(0.25)$	$Q_{\mathcal{E}_{s,3}}(0.50)$	$Q_{\mathcal{E}_{s,3}}(0.75)$	$\mu_{\mathcal{E}_{s,3}}$
as_1	0.12	0.27	**0.67**	**0.50**
as_2	**0.19**	**0.30**	0.55	0.47
as_3	**0.17**	**0.35**	**0.83**	**0.59**
bs_1	0.09	0.18	0.39	0.36
bs_2	0.08	0.16	0.41	0.30
bs_3	0.13	0.29	0.63	0.42

Table 8: **Trade imbalance between bids and asks**: The table displays aggregated effects $\mathcal{E}_{s,3} = C^{s,3}_{i\leftarrow j}$ of different quote levels on prices by adding up the impacts across all stocks. $C^{s,3}_{p\leftarrow as_r}$ or $C^{s,3}_{p\leftarrow bs_r}$ correspond to the aggregated impacts of asks and bids on prices respectively. The columns contain different quantile levels $Q_{\mathcal{E}_{s,3}}(\alpha)$ and the mean $\mu_{\mathcal{E}_{s,3}}$ of these elements over the entire sample period 06/2016-07/2016. The largest two values are marked in bold.

test for a significant overall asymmetry of bid and ask trades in $\mathcal{E}_{s,3}$ and confirm the significance of the effect. More precisely, we take $\mu_1 = \frac{1}{3}\sum_{r=1}^{3}\mu(C^3_{p\leftarrow as_r})$ as the mean of the overall impact from selling orders and $\mu_2 = \frac{1}{3}\sum_{r=1}^{3}\mu_{C^3_{p\leftarrow bs_r}}$ as the mean of the overall impact from buying orders over the sample period. Hence we test $H_0 : \mu_1 - \mu_2 = 0$ against the alternative $H_a : \mu_1 - \mu_2 > 0$.[17] Formally, we employ a pooled t-test which rejects the H_0 below the 1%-level (p-value 0.003144). This result is also confirmed by a Welsh-test where we obtain the p-value 0.003168. This can also be executed on a more granular per stock level with similar results.

[17] Note that in our data generally deviations of H_0 only occur from larger overall ask side then bid side effects. Thus this one sided alternative is empirically valid and produces better power of the test than a more general two-sided alternative.

4.1.3 Systematic Differences in Topologies of Directed Price Impact Networks: Price-only versus the Fully LOB-aggregated Effects

To understand how the impact of a shock originating in one stock can be transmitted and amplified to the other stocks, almost all existing literature focus on network linkages stock prices only. In the following we make use of the information contained in the LOB to examine the impact of full system interdependencies across stocks by constructing the fully LOB-aggregated network per stock. We find systematic differences in the structure of the classic but only partial price-only network $\mathcal{G}_p$ to the complete LOB-aggregated network $\mathcal{G}_g$, which affect the construction of potential investment or hedging strategies.

As price-only effects account for up to 50-65% of all impacts according to Table 6, the individual stock network $\mathcal{G}_p = (\mathcal{V}_4, \mathcal{E}_{p,4})$ in Case 3 with price factors only might contain the basic structure of price-impact interconnection channels. In the corresponding panel A of Table 9, however, the reported total directional connectedness "from" C_{from} and "to" C_{to} suggest that the stock price of Johnson&Johnson has the strongest influence on the other stock prices and is also a large scale receiver of cross-effects, as suggested by its high connectedness values of both inbound and outbound effects. The role of the financial sector companies only appears minor both as transmitters or receivers displaying even negative net effects. Though if we compare the price-only impact network to the aggregated nine stock network $\mathcal{G}_g = (\mathcal{V}_5, \mathcal{E}_{g,5})$ where we cumulate all price and LOB volume effects for each single stock as in Case 3, the structure fundamentally changes leading to substantial differences in the roles of net transmitters and receivers in the network. By including the size factors in $\mathcal{G}_g$, we are able to investigate how the network is affected by the presence of LOB liquidity effects which not only produce relevant cross-asset price impact effects (Tables 4, 5) but also change the overall network topology systematically. In panel B of Table 9 we observe higher "from"-connectedness values for all financial stocks, suggesting that the finance sector has the strongest influence on the other stocks in the portfolio. Meanwhile, the "to"-connectedness of all financial stocks is the lowest leading to the largest positive net influence in the system. Generally, technology stocks receive the highest inbound effects from the system. On the stock level, we find that JP Morgan is the stock with the highest "net"-connectedness to others exhibiting the highest

Panel A: Price-only network effects

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$\mu_{C_{from}}(\mathcal{E}_{p,4})$	0.15	0.13	0.22	**0.27**	0.15	0.03	0.16	0.07	0.10
$\mu_{C_{to}}(\mathcal{E}_{p,4})$	**0.21**	0.08	0.14	**0.21**	0.10	0.13	0.10	0.10	0.20
$\mu_{C_{net}}(\mathcal{E}_{p,4})$	0.05	0.05	**0.08**	0.05	0.05	-0.10	0.06	-0.03	-0.11

Panel B: Effects from the fully aggregated LOB

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$\mu_{C_{from}}(\mathcal{E}_{g,5})$	3.09	2.30	2.10	2.23	2.67	2.34	**3.47**	3.12	3.17
$\mu_{C_{to}}(\mathcal{E}_{g,5})$	2.80	**3.05**	2.98	2.83	2.65	2.69	2.47	2.51	2.53
$\mu_{C_{net}}(\mathcal{E}_{g,5})$	0.17	-0.70	-0.91	-0.50	0.14	-0.71	**0.97**	0.34	0.35

Table 9: **Identification of transmitters and receivers of price impacts**: In Panel A, based on the price-only network $\mathcal{G}_p = (\mathcal{V}_4, \mathcal{E}_{p,4})$, we compute the "to"-connectedness C_{to} as weighted number of effects that the stock n directs to (prices of) other stocks, the "from"-connectedness C_{from} as weighted number of outbound links to the price factor $p^{(n)}$ of stock n, and "net"-connectedness C_{net} as the difference $C_{from} - C_{to}$. $\mu_{\cdot}$ is the mean over the given sample period 06/2016-07/2016 and the maximum value per measure is marked in bold. In panel B, we report the respective measures for the fully LOB-aggregated network $\mathcal{G}_g = (\mathcal{V}_5, \mathcal{E}_{g,5})$.

influence on the considered system, while the technology company like IBM is a main impact receiver in the aggregated system. Overall, financial companies appear as dominant stocks driving the price-impact network over time. This key role is only detectable from the aggregated individual stock network which considers the information contained in the LOB.

We graphically display the structure for the fully LOB-aggregated price impact network $\mathcal{G}_g$ for the first trading day after the Brexit announcement as an example. In accordance with the discussion in section 3.3, we depict the estimated full sample directional connectedness Table 3 in left panel of Figure 3. The size of the directed connectedness between two nodes is marked by the thickness of the arrow connecting them. The price factor and size factors that belong to the same company appear in the same color, the width of edges between two nodes represents the connectedness. The full sample network plot illustrates that the LOB factors of all stocks are largely interconnected with other stocks' prices, thus being very informative about the total aggregated directional connectedness $\mathcal{G}_g$. However, it is not easy to decipher all pairwise connectedness effects from the granular factor network on the left of Figure 3. Thus on the right panel, all connectedness effects are aggregated per stock. Hence for each stock there is only a single node in the network and links between nodes represent the overall "from"' and "to" impacts on the system, i.e., aggregating the directed connectedness of both price and size factors for each stock. The respective links of Citigroup and JP Morgan and Wells Fargo reveal that they are the stocks that generated the highest “to”-connectedness, whereas the other six stocks mainly act as price impact receivers as in the time means Table 9, panel B.

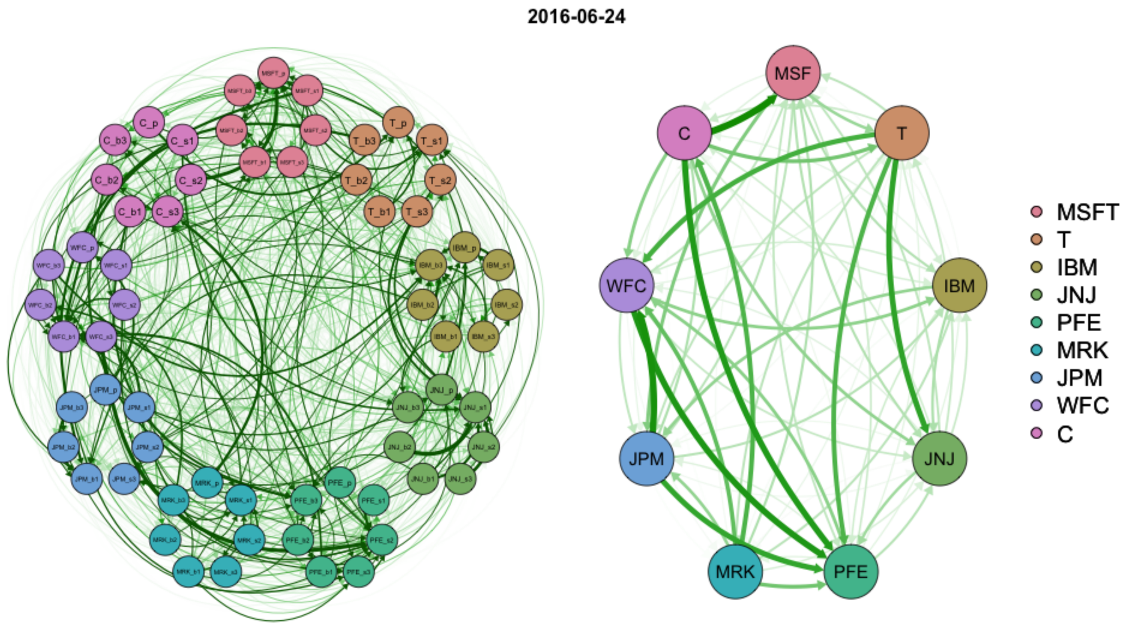

Figure 3: **Left panel: the LOB-network plot** : We visualize the full LOB-network matrix as defined in Table 3, on 24.06.2016. For each stock n, it contains 7 nodes $(p^{(n)}, as_1^{(n)}, as_2^{(n)}, as_3^{(n)}, bs_1^{(n)}, bs_2^{(n)}, bs_3^{(n)})$ and we place them in the clockwise order within each small ring using the same color. Hence we have in total 63 nodes for the selected nine stock portfolio. The relationships between any two nodes are associated with the sign and magnitude of the corresponding connectedness measure.
Right panel: Fully LOB-aggregated network per stock: We plot the aggregated network $\mathcal{G}_g = (\mathcal{V}_5, \mathcal{E}_{g,5})$ on 24.06.2016. For stock n, we merge the nodes $(p^{(n)}, as_1^{(n)}, as_2^{(n)}, as_3^{(n)}, bs_1^{(n)}, bs_2^{(n)}, bs_3^{(n)})$ into another node defined as $g^{(n)}$ in $\mathcal{V}_5$. The resulting aggregated network contains 9 nodes.

4.2 LOB Market Impact over Time

Macroeconomic announcements and news are often accompanied by price moves, while the impacts on how prices respond to ask and bid limit orders are not symmetric. With the United Kingdom European Union membership referendum at 23.06.2016 voting for leave (Brexit) during our sample period, we investigate if and how this largely unexpected event affects the market. Empirically it has been documented that within 24 hours of the Brexit vote, the British pound plunged by almost 10 per cent to a three-decade low, share prices sank and gilt yields slid. Trading volume declined during the event, but rebounded to record highs after. We examine first in subperiods around the Brexit vote, if and how the event affected price impacts yielding potential changes in the cross-asset or cross-price LOB structure in the system. Then we take a glance at the network spillover structure directly at the days before, at and after the Brexit voting date to study immediate short-term effects.

4.2.1 Effects in Subperiods

We explore potential time-variation in the cross-asset and cross-price market impacts in different subperiods around the Brexit vote. In particular, we separately study the subperiods period I: 01.06.2016-21.06.2016 before the Brexit vote, period II: 22.06.2016-11.07.2016 during which the Brexit decision was announced and period III: 12.07.2016-31.07.2016 for the medium range impact of the vote. Note that all three subperiods are approximately of equal length and partition the original full sample of the previous subsection.

Figure 4 provides an overview of the results by comparing different sources of price impact shares over the three periods. Overall, we see that there is some time-variation in the decomposition of price impact proportions but the relatively small size and the shape of these variations are such that the full sample time averages of the last subsection remain informative. Over time, the LOB quote effects on prices (all blue parts together) are always well above 20% amounting to up to 50% with the largest effects in the Brexit vote period 2. Its largest part consistently originates from cross-asset effects (dark blue) which exceeds the own-price LOB effects mostly by a factor larger than 2. Over all periods, all cross-asset effects (from prices and LOB together) marked in light blue and green are above 30% up to more than 50% where the major part comes from the LOB (light blue). Generally the largest single share of the price impact is the own-price effect from the stock's own past price (complement of the color bars to 1), which is smallest during the second period with the Brexit announcement and larger but similar in size in the pre- and post-Brexit subperiods.

For each subperiod, we also look deeper into the LOB structure of price impacts examining the

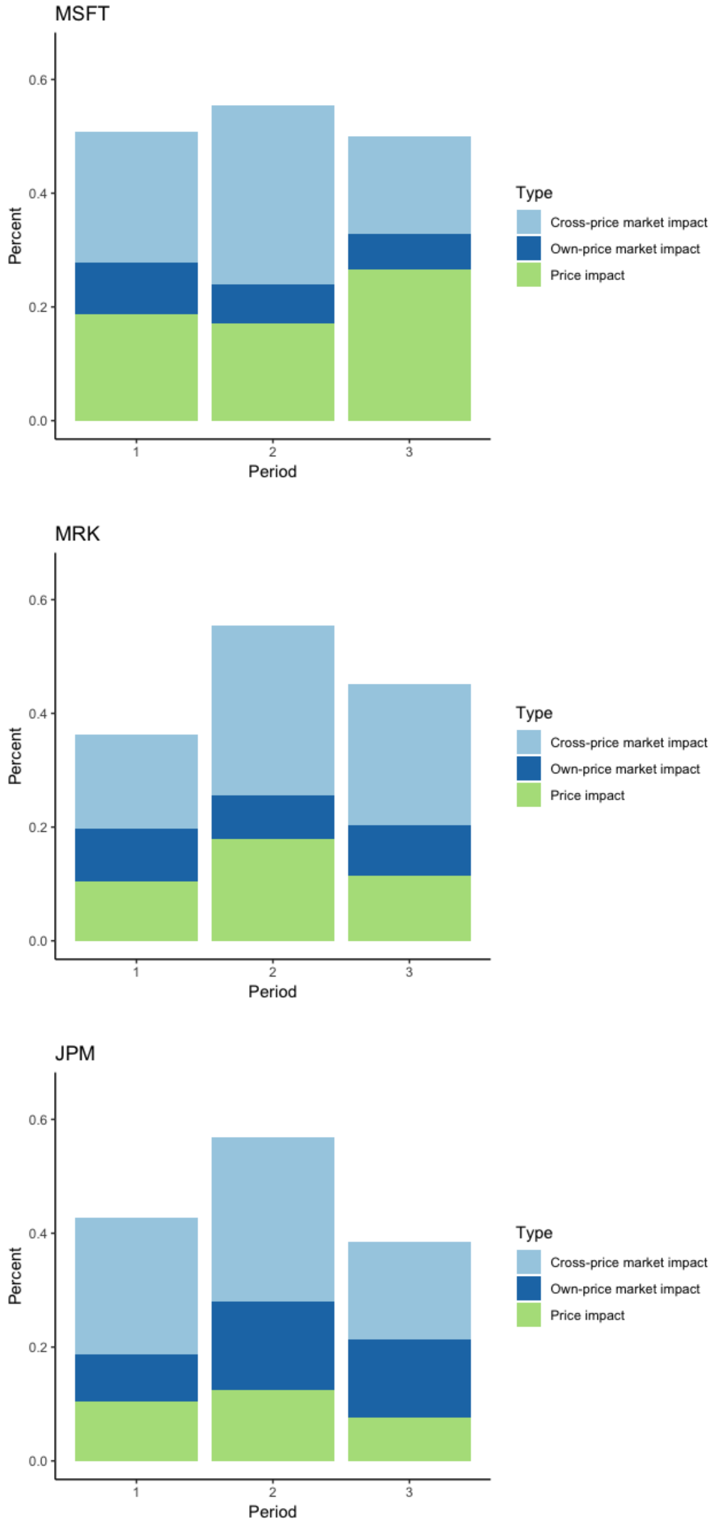

Figure 4: **Price impacts in subperiods**: For three exemplary stocks from each sector, this figure graphically depicts the composition of received price impacts from different sources as percentage aggregates over time in the three subperiods 1: 01.06.2016-21.06.2016; 2: 22.06.2016-11.07.2016 including the Brexit vote, and 3: 12.07.2016-31.07.2016. The green bars denote the cross-asset relative impact share received from the other stock prices. In light blue we mark the percent of cross-price market impact from the LOB. The dark navy blue bar displays the percentage of own-price market impact from the LOB on the specific stock price. In each period, the remaining share to the 100% is the own-price impact from the (lagged) stock's price. The results for the other six stocks are shown in Appendix, see Figure 8 and 9.

asymmetry in bid and ask side effects and the time-variation of effects across different quote levels in the comprehensive network graph $\mathcal{G}_s = (\mathcal{V}_2, \mathcal{E}_{s,2})$ in (16). Thus Table 10 provides the estimated average effects in $\mathcal{E}_{s,2}$, i.e. aggregated bid and ask effects across stocks in each quote level but also detailed results for each quote level for all stocks over the three periods. On all quote levels, we see time-variation of effects which for some stocks is more pronounced in the deeper levels of the book (e.g. T, ask side; IBM bid side). Overall a relatively large proportion of effects originates from the deeper levels of the book. In particular the relative peak of cross-asset LOB effects in period II in Figure 4 stems for all three stocks from different levels in the book.

In general, we observe that over all sub-periods the ask-side effects dominate the bid side effects cumulated over all levels, but also mostly separately per level and even mostly per cell in Table 10 on each level and in each period. As aggregated price impacts of ask quotes capture the selling pressure, larger values of ask trades indicate that selling trades have a stronger price impact than buying trades. These patterns are confirmed by estimated time medians of links in $\mathcal{E}_{s,2}$ as reported in Table 12, Appendix E.

4.2.2 Immediate Effects around the Brexit Vote

Figure 5 graphically displays the LOB networks $\mathcal{G}_s = (\mathcal{V}_2, \mathcal{E}_{s,2})$ at the day before, at and after the Brexit vote consisting of price factors (in grey), ask size factors (in blue) and bid size factors (in red) at different levels, thus 7 nodes per stock (one grey plus 3 red and 3 blue). Non-zero connectedness $C_{i \leftarrow j}$ between any two nodes is marked by an arrow in the color of the source factor j.

When investigating not only price impacts, but all spillover effects, the relative importance of price nodes at the boundary of the upper left Figure 5 is small relative to LOB factors which are more central to the system on 22.06.2016, the day before the Brexit vote. This plot is representative for many days before the Brexit. The importance of within-LOB effects is displayed in the lower part of the figure. For the three days around the Brexit vote, we separate price and ask-side effects from price and bid-side effects in different network parts. The graphs indicate that the Brexit announcement, induced substantial changes in the network structure with different effects on ask and bid sides. The price&bid size factor network is less connected at the day of the vote (23.06.2016)

quote level	period	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
	I	0.40	0.45	0.53	0.60	0.41	0.34	0.52	0.73	1.18
as_1	II	0.34	0.42	0.53	0.16	0.53	0.53	0.91	0.49	0.26
	III	0.57	0.43	0.27	0.24	0.41	1.17	1.13	0.39	1.05
	full	0.44	0.43	0.45	0.34	0.45	0.68	**0.86**	0.53	**0.83**
	I	0.86	0.56	1.25	0.82	0.50	0.66	0.42	0.27	0.78
as_2	II	0.56	0.27	0.16	0.45	1.24	0.86	0.44	0.29	0.78
	III	0.16	0.40	0.36	0.17	0.06	0.43	0.14	0.95	0.34
	full	0.53	0.41	0.59	0.48	0.60	**0.65**	0.33	0.50	**0.64**
	I	0.77	0.93	0.30	0.25	0.63	0.73	1.49	0.43	0.12
as_3	II	1.80	0.64	0.46	0.35	0.80	1.04	1.23	0.69	0.66
	III	0.14	0.50	0.86	0.38	0.31	0.92	0.38	0.49	0.38
	full	**0.90**	0.69	0.54	0.33	0.58	0.89	**1.03**	0.54	0.39
	I	0.07	0.33	0.46	0.33	0.25	0.03	0.19	0.16	0.18
bs_1	II	0.18	0.71	0.49	0.73	0.25	0.71	0.68	1.21	0.45
	III	0.12	0.29	0.46	0.15	0.45	0.49	0.50	0.46	0.57
	full	0.12	0.44	0.47	0.40	0.32	0.41	0.46	**0.61**	0.40
	I	0.19	0.32	0.47	0.16	0.43	0.07	0.34	0.17	0.11
bs_2	II	0.23	0.36	0.17	0.11	0.35	0.34	0.55	0.28	0.43
	III	1.10	0.37	0.50	0.33	0.33	0.26	0.57	0.15	0.29
	full	**0.50**	0.35	0.38	0.20	0.37	0.22	**0.48**	0.20	0.28
	I	0.94	0.29	0.17	0.29	0.17	0.75	0.27	0.92	0.32
bs_3	II	0.72	0.40	0.38	0.39	1.08	0.30	0.65	0.64	0.61
	III	0.24	0.40	0.17	0.49	0.29	0.10	0.37	0.64	0.47
	full	0.63	0.37	0.24	0.39	0.51	0.38	0.43	**0.73**	0.47

Table 10: **Quote level effects**: Each element of the upper half of the table contains the aggregated market impact $\mu(C^{s,2}_{p^{(m)}\leftarrow as_r})$ of the ask side level r across all stocks on the price of stock m as defined in (16) displayed as mean μ over different parts of the full sample period 01.06/2016-31.07/2016. The three subperiods are of approximately equal length with period I: 01.06.2016-21.06.2016; period II: 22.06.2016-11.07.2016 during which the Brexit decision was announced; period III: 12.07.2016-31.07.2016. In the lower half of the table corresponding bid-side effects $\mu(C^{s,2}_{p^{(m)}\leftarrow bs_r})$ are reported. All numbers are multiplied by 10. Bold (grey) numbers highlight the largest (smallest) relative effects for each quote level over the full sample period.

while the level of interconnections from this side only increases on the next day. This is different for the price&ask size factor network which starts of at a lower level of interconnectedness which directly increases at the day of the announcement exceeding the current level of interconnectedness from the bid and price part but only then. Thus at the day of the vote in moment of political uncertainty, the buying pressure is much weaker and selling pressure slightly stronger.

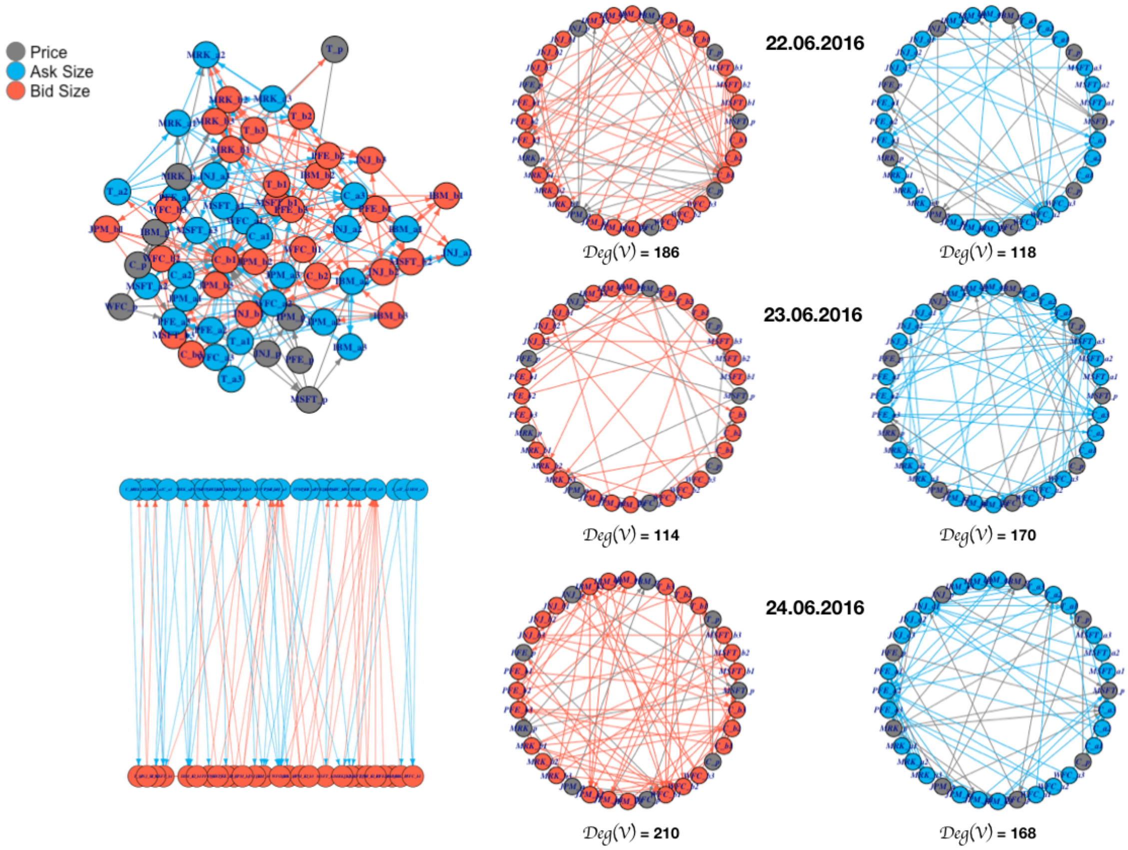

Figure 5: **LOB-network around the Brexit announcement 22.06.2016-24.06.2016 at different aggregation levels**. The upper left part of the figure depicts the full LOB-network on 22.06.2016. Generally, ask (bid) side nodes are marked in blue (red) and prices in grey. An arrow links two nodes i and j in the color of node j if $C_{i \leftarrow j}$ significantly differs from zero. The layout of the graph is data-driven allocating more connected nodes to the center using a version of the Fruchterman-Reingold-algorithm which minimizes the length of all paths. In the figure below for the same day, we display only interactions between quotes. On the right, we use a circular layout and separate price and ask side effects (on the right) from price and bid side effects (on the left) across three consecutive days from top to bottom. $Deg(\mathcal{V})$ counts the plain number of links in the network above.

5 Extension: Price Direction under an Exogenous Unit Shock

So far, we have analyzed the time-varying market impacts within and across asset classes studying induced "endogenous shocks". In the LOB market, however, huge sell/buy volumes queued on the ask/bid side could induce strong sell/buy pressure on the market which could therefore change the price rather as an idiosyncratic exogenous shock. This is consistent with our empirical findings that both ask and bid trades placed in the LOB may directly affect the price dynamics of one asset. Though, when a large buy or sell market order arrives, the market order will automatically execute which causes a temporary own-price market impact. Even though a large market order immediately affects the price direction, in this case, the size of bid/ask trading volumes alone does not provide enough information on the price direction. In market microstructure analysis, it is of interest to detect important transmission channels where trading activities on the LOB have significant influence on the asset price, which can be derived from the price impact of the arrival of large ask/bid market and limit orders.

We empirically address this with the high-dimensional tools proposed in Section 3.2 by measuring size and persistence of the effect on the asset price by a shock on the quote factors. In particular, for each quote level and each asset separately, we investigate the impact over time on the asset price in response to an idiosyncratic unit shock. This analysis helps to uncover the mechanism of price formation and could be exploited for arbitrage opportunities. For measuring the size of the impact of exogenous shocks in market/limit orders and for determining their temporary or robust impact shape over time, we resort to the generalized impulse response function GI defined componentwise in (10). Now, however, we analyze the case of a unit shock $\delta_t = (\delta_{1t}, \delta_{2t}, \ldots, \delta_{Kt})^\top \sim e_j$ on Y_t instead of (11), where e_j is a unit vector marking the idiosyncratic unit"shock" in component j. For comparability of impacts across factors and assets, we use exogenous unit shocks. We study the case where a unit shock can occur anywhere in the LOB, i.e. in $(as_1^{(n)}, as_2^{(n)}, as_3^{(n)}, bs_1^{(n)}, bs_2^{(n)}, bs_3^{(n)})$ for a stock n and focus in our results on own-price effects for ease of exposition. With this, we can not only detect the evolution of a unit shock from different quote levels over time but also test for a statistically significant impact. We set the forecast horizon of the GI to $l = 30$ minutes and use the bootstrap approach (see Appendix B) to obtain generalized impulse response curves for each trading day over two month investment period.

In the exemplary case of Wells Fargo on 25.07.2016, we observe a positive price impact from a shock in the bid market order and a negative effect from the first level on the bid-side as depicted in the left part of Figure 6. In particular the magnitudes of effects on the own price in response to bid and ask side shocks are strongly negatively correlated. Both impacts last for almost 10 minutes before the price shifts back, which yields sufficient time for high-frequency investors to make profit from arbitrage strategies. The presented results are representative for all other days, where unit shocks on the first ask and bid level induce significant price effects. As generally investors will start marking down (up) their bid (ask) price when there is a wave of sell (buy) orders coming into the order book, the empirical findings that the price factor tends to decrease (increase) significantly after the arrival of a large ask (bid) market order are in line with theory.

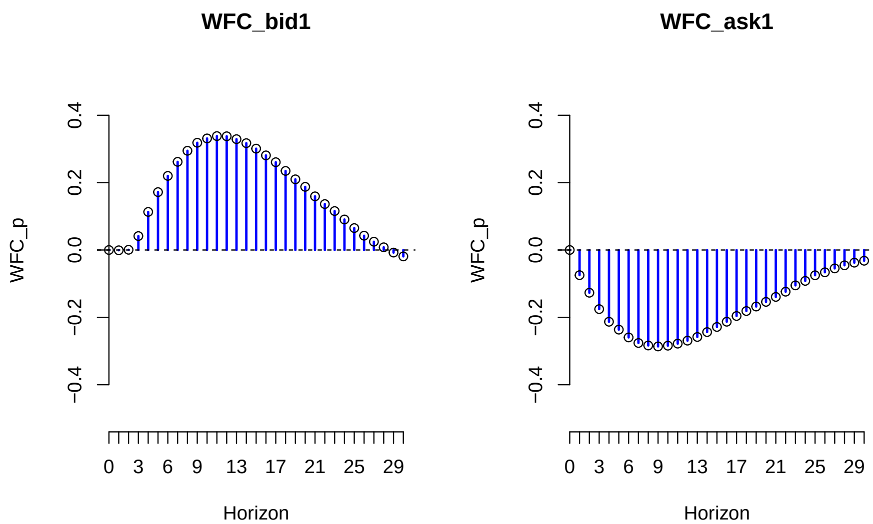

Figure 6: **Exogenous unit shock price impact for Wells Fargo**: This figures depicts the price impacts over time in response to a unit shock in the bid market order (left side) or in the ask market order (right side) of the same stock on 25.07.2016.

Figure 7 shows the shape of market impacts of orders posted deeper in the book for Citigroup (C). This suggests a positive pile-on of effects where a larger ask order may cumulate effects from different levels of the LOB. The estimated market impact lasts for almost 20 minutes, the price goes up after 10 minutes because the market investors may buy trades picking up the posted volume or by cancellations on the ask side.[18]

Table 11 summarizes days of significant price impacts from exogenous unit shocks in different levels of the LOB. For each trading day, we use $\ominus$ and $\oplus$ to mark a 5% significant negative and positive response of the price to the arrival of market/limit orders. In general, financial theory suggest that responses are positive to buying (bid trades) and negative to selling (ask trades). Our results suggest that overall on all levels, the group of financial stocks empirically shows this reaction, see e.g. Citigroup. This may be explained by the fact that the financial sector is leading the market as

[18]Complete results for all other stocks are available on request.

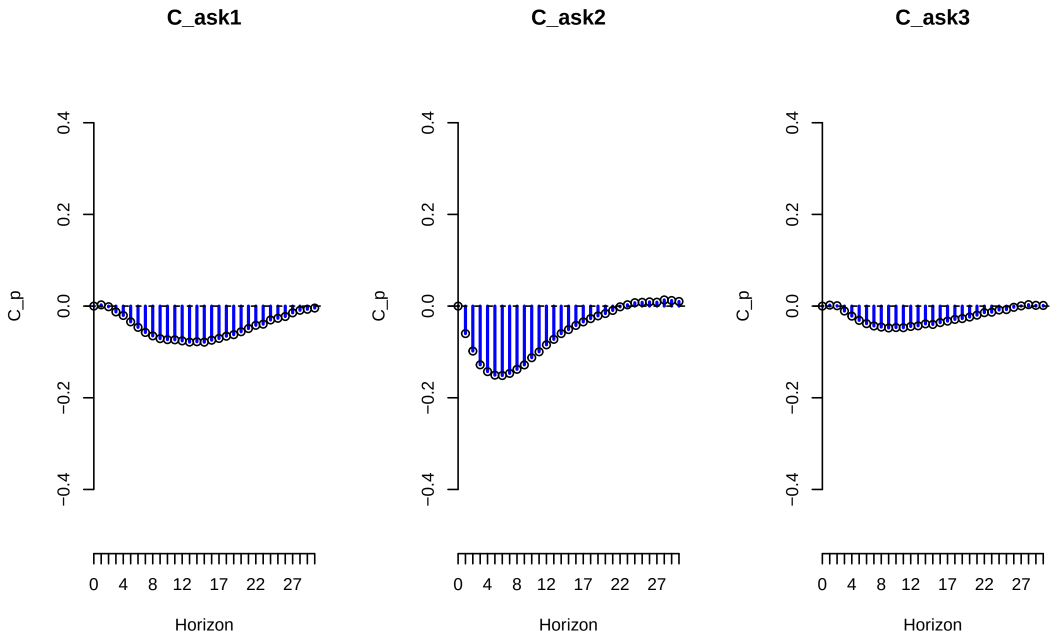

Figure 7: **Exogenous unit shock price impact for Citigroup**: This figures depicts the price impacts over time in response to a unit shock in different levels of the LOB of the same stock on 01.06.2016.

net price impact transmitters (see Table 9). This is also in line with other empirical findings which indicate that for financial stocks, the response in prices to changes in trading volumes is relatively stable and thus robust to statistical arbitrage, see e.g. Hautsch and Huang (2012). Interestingly, the healthcare and technology stocks sometimes show opposite results where their prices are positively linked to ask order flows in some cases. This might be as they are mostly net price-impact receiving stocks (see Table 9) with large shares of cross-asset effects. This result is consistent with our main findings in subsection 4.1.2 and 4.1.3. Overall, our findings could be key for algorithm traders for profitable strategies.

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$as1$		⊖ ⊖ ⊖ ⊖ ⊖		⊕ ⊕ ⊕		⊕⊖	⊖ ⊖ ⊖	⊖ ⊖ ⊖	⊖
$as2$	⊖ ⊖ ⊖	⊕	⊕⊕		⊕			⊖⊖	⊖ ⊖ ⊖ ⊖ ⊖⊖
$as3$	⊕⊕	⊕	⊕	⊕	⊕ ⊕ ⊕		⊖	⊕⊕	⊖
$bs1$	⊕ ⊖ ⊖	⊕ ⊖ ⊖⊖		⊕	⊕	⊕⊕	⊕ ⊕ ⊕⊖	⊕⊕	⊕
$bs2$	⊖	⊖			⊕⊖	⊕	⊕		⊕
$bs3$		⊖	⊖⊖	⊖⊖	⊖	⊖⊖	⊖		

Table 11: **Summary of own-price market impacts.** Symbols mark each trading day with a significant positive ⊕ and significant negative ⊖ market impact in response to an exogenous unit shock at the respective quote level.

6 Conclusion

In this paper, we suggest a statistical learning methodology in order to empirically quantify the general own- and cross-asset as well as the own- and cross-price impact of the LOB over time and identify relevant network spillover channels. Our empirical strategy focuses on real price effects from observable LOB movements which might happen simultaneously and cannot be treated in isolation while their occurrence and size often depend endogeneously on the state of the system in previous time periods. We find strong network interaction effects within the entire limit order book with a substantial impact on prices even in an only moderate dimensional portfolio. We document and quantify how these liquidity interlinkages from the LOB incorporate information beyond standard price channels and change price impact receiver or transmitter roles of stocks within the system. Moreover, we show substantial variations of effects over time by studying the period around the Brexit vote in different time granularity. For the case of a "large" volume shock in isolation, we also analyze the quasi-experimental scenario of an impact of an idiosyncratic unit shock in different components over time.

Technically, already small portfolios require statistical tools for large dimensional LOB systems for accurate quantification and identification of effects. Though, the analysis of such large dimensional LOB systems needs novel statistical techniques for data synchronization, estimation, forecasting and impulse response analysis with endogenous shocks in order to identify market impacts. These are of independent interest and also work for very general large-dimensional financial systems with complex interlinkages. In particular, we provide a bootstrap technology for generalized types of impulse responses as economically interpretable links, since standard tools as MA transformations are infeasible in our scenario and even more so for much larger portfolios. These also offer a link to the existing network connectedness literature which is prevalent for systemic risk analysis which has generally been technically restricted to smaller low-dimensional systems only potentially missing out on important cross-effects.

References

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies*, 30(1):2–47.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial studies*, 18(2):351–416.

Almgren, R. and Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40.

Bandi, F. M. and Russell, J. R. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics*, 79(3):655–692.

Bandi, F. M. and Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies*, 75(2):339–369.

Basu, S., Michailidis, G., et al. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Belloni, A., Chernozhukov, V., et al. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics*, 104(3):535–559.

Bloomfield, R., O'Hara, M., and Saar, G. (2005). The make or take decision in an electronic market: Evidence on the evolution of liquidity. *Journal of Financial Economics*, 75(1):165–199.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8):2267–2306.

Budish, E., Cramton, P., and Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621.

Candes, E. and Tao, T. (2007). The dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, pages 2313–2351.

Cespa, G. and Foucault, T. (2014). Illiquidity contagion and liquidity crashes. *The Review of Financial Studies*, 27(6):1615–1660.

Christensen, K., Kinnebrock, S., and Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159(1):116–133.

Cont, R., Kukanov, A., and Stoikov, S. (2014). The price impact of order book events. *Journal of financial econometrics*, 12(1):47–88.

Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische annalen*, 300(1):463–520.

Demirer, M., Diebold, F. X., Liu, L., and Yılmaz, K. (2017). Estimating global bank network connectedness. Technical report, National Bureau of Economic Research.

Di Maggio, M., Kermani, A., and Song, Z. (2017). The value of trading relations in turbulent times. *Journal of Financial Economics*, 124(2):266–284.

Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.

Easley, D., De Prado, M. L., and O'Hara, M. (2011). The microstructure of the flash crash: Flow toxicity, liquidity crashes and the probability of informed trading. *Journal of Portfolio Management*, 37(2):118–128.

Easley, D., de Prado, M. L., and O'Hara, M. (2016). Discerning information from trade data. *Journal of Financial Economics*, 120(2):269–285.

Eisler, Z., Bouchaud, J.-P., and Kockelkoren, J. (2012). The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance*, 12(9):1395–1419.

Elliott, M., Golub, B., and Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, 104(10):3115–53.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68(1):1–22.

Evans, M. D. and Lyons, R. K. (2002). Order flow and exchange rate dynamics. *Journal of political economy*, 110(1):170–180.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Fleming, M. J., Mizrach, B., and Nguyen, G. (2017). The microstructure of a us treasury ecn: The brokertec platform. *Journal of Financial Markets*.

Foucault, T., Kadan, O., and Kandel, E. (2005). Limit order book as a market for liquidity. *The review of financial studies*, 18(4):1171–1217.

Giglio, S., Kelly, B., and Pruitt, S. (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics*, 119(3):457–471.

Handa, P., Schwartz, R., and Tiwari, A. (2003). Quote setting and price formation in an order driven market. *Journal of financial markets*, 6(4):461–489.

Hautsch, N. and Huang, R. (2011). Limit order flow, market impact and optimal order sizes: evidence from nasdaq totalview-itch data. *Market Impact and Optimal Order Sizes: Evidence from NASDAQ TotalView-ITCH Data (August 22, 2011)*.

Hautsch, N. and Huang, R. (2012). The market impact of a limit order. *Journal of Economic Dynamics and Control*, 36(4):501–522.

Hautsch, N. and Podolskij, M. (2013). Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: theory, implementation, and empirical evidence. *Journal of Business & Economic Statistics*, 31(2):165–183.

Hautsch, N., Schaumburg, J., and Schienle, M. (2014). Financial network systemic risk contributions. *Review of Finance*, 19(2):685–738.

Hendershott, T., Jones, C. M., and Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1):1–33.

Hendershott, T. and Seasholes, M. S. (2007). Market maker inventories and stock prices. *American Economic Review*, 97(2):210–214.

Hirschey, N. (2020). Do high-frequency traders anticipate buying and selling pressure? *Management Science*.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications*, 119(7):2249–2276.

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, 186(2):325–344.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, 74(1):119–147.

Lanne, M. and Nyberg, H. (2016). Generalized forecast error variance decomposition for linear and nonlinear multivariate models. *Oxford Bulletin of Economics and Statistics*, 78(4):595–603.

Liang, C. and Schienle, M. (2018). Determination of vector error correction models in high dimensions. *Journal of Econometrics*.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Menkveld, A. J. (2013). High frequency trading and the new market makers. *Journal of Financial Markets*, 16(4):712–740.

Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, pages 1069–1097.

Obizhaeva, A. A. and Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, 116(2):257–270.

Parlour, C. A. (1998). Price dynamics in limit order markets. *The Review of Financial Studies*, 11(4):789–816.

Podolskij, M., Vetter, M., et al. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3):634–658.

Roşu, I. (2009). A dynamic model of the limit order book. *The Review of Financial Studies*, 22(11):4601–4641.

Securities, Commission, E., et al. (2010). Concept release on equity market structure. *Federal Register*, 75(13):3594–3614.

Teräsvirta, T., Tjostheim, D., Granger, C. W., et al. (2010). Modelling nonlinear economic time series. *OUP Catalogue*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Wu, W. B. and Wu, Y. N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, 10(1):352–379.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

A Pre-averaging estimation

Suppose that we observe non-synchronous noisy data Y_t following,

$$Y_t = X_t + \varepsilon_t, \quad t \geq 0 \tag{18}$$

with efficient log price X_t is latent. The error term ε_t represents microstructure noise and is assumed to be independent and identically distributed with

$$\mathsf{E}(\varepsilon_t) = 0, \quad \mathsf{E}\left(\varepsilon_t^2\right) = \psi \tag{19}$$

The price process X_t follows a semi-martingale form, Delbaen and Schachermayer (1994),

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t \sigma_s dW_s \tag{20}$$

where $(a_s)_{s\geq 0}$ is a càdlàg drift process, $(\sigma_s)_{s\geq 0}$ is an adapted càdlàg volatility process, $(W_s)_{s\geq 0}$ is a Brownian motion. In addition, we assume X_t and ε_t are independent, i.e.

$$\mathsf{E}(\varepsilon_t \mid X) = 0 \tag{21}$$

If one can only observe Y_i^n at discrete times t, i indexes the time points with interval length Δ_n, the returns $\Delta_i^n Y$ is thus defined as,

$$Y_i^n = Y_{i\Delta_n}, \quad \Delta_i^n Y = Y_i^n - Y_{i-1}^n, \quad i = 1, \ldots, n \tag{22}$$

A pre-averaging is conducted to alleviate microstructure noise and solve non-synchronicity, we follow the notations originally used by Jacod et al. (2009). The basic idea is to construct smoothing functions to diminish the impact of the noise induced by ε_t. Specifically, there is a sequence of integers denoted as k_n which satisfies,

$$\exists \theta > 0, \quad k_n \sqrt{\Delta_n} = \theta + \mathcal{O}\left(\Delta_n^{\frac{1}{4}}\right) \tag{23}$$

and a continuous weight function $g : [0,1] \mapsto \mathbb{R}$. g is piecewise C^1 with a piecewise derivative $g^{'}$, $g(0) = g(1) = 0$, and $\int_0^1 g^2(s)ds > 0$. Furthermore, the following real-valued numbers and functions are associated with function g on $\mathbb{R}_+$,

$$
\begin{aligned}
\psi_1 &= \int_0^1 \{g^{'}(u)\}^2 du, \quad \psi_2 = \int_0^1 \{g(u)\}^2 du \\
\Phi_1(s) &= \int_s^1 g^{'}(u) g^{'}(u-s) du, \quad \Phi_2(s) = \int_s^1 g(u)g(u-s)du \\
\Phi_{ij} &= \int_0^1 \Phi_i(s)\Phi_j(s)du, \quad i,j = 1,2, \quad u \in [0,1]
\end{aligned}
\tag{24}
$$

Here we choose $g(x) = x \wedge (1-x)$, as in Podolskij et al. (2009), Christensen et al. (2010) and Hautsch and Podolskij (2013). Therefore we have

$$
\begin{aligned}
\psi_1 &= 1, \quad \psi_2 = \frac{1}{12}, \quad \Phi_{11} = \frac{1}{6} \\
\Phi_{12} &= \frac{1}{96}, \quad \Phi_{22} = \frac{151}{80640}
\end{aligned}
\tag{25}
$$

The pre-averaged returns $\overline{Y}_i^n$ associated with the weight function g are given as,

$$
\begin{aligned}
\overline{Y}_i^n &= \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) \Delta_{i+j}^n Y \\
&= -\sum_{j=0}^{k_n-1} \left\{ g\left(\frac{j+1}{k_n}\right) - g\left(\frac{j}{k_n}\right) \right\} Y_{i+j}^n, \quad i = 0, \ldots, n-k_n+1
\end{aligned}
\tag{26}
$$

The window size k_n defined in equation (23) is chosen of $\mathcal{O}\left(\sqrt{\frac{1}{\Delta_n}}\right)$, balance the noise $\overline{\varepsilon}_i^n = \mathcal{O}_p\left(\sqrt{\frac{1}{k_n}}\right)$ and the efficient price $\overline{X}_i^n = \mathcal{O}_p\left(\sqrt{k_n \Delta_n}\right)$.

B Bootstrap-based multistep forecast methods

Here we describe the computational steps to obtain the $\mathsf{E}(Y_{t+1}|u_{jt} = \delta_{jt}, \mathcal{F}_{t-1})$, GI, GFEVD via Bootstrap method, more details can be found in Koop et al. (1996), Lanne and Nyberg (2016), Teräsvirta et al. (2010).

1. Denote $\mathcal{F}_{t-1}$ as all the information prior to Y_t, and select a forecast horizon h.

2. Randomly sample N_B vectors of shocks $(\delta_{1t}, \delta_{2t}, \ldots, \delta_{Kt})^\top$ from the residuals of estimated model,

$$\delta_{jt} : (\delta_{1t}, \delta_{2t}, \ldots, \delta_{Kt})^\top \sim \hat{u}^\star_{jt} e_j \tag{27}$$

$$\hat{u}^\star_{jt} = Y_t - \left(\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_p\right) \left(Y_{t-1}^\top, Y_{t-2}^\top, \ldots, Y_{t-p}^\top\right)^\top = Y_t - g(Y_{t-1}) \tag{28}$$

3. Compute conditional multistep forecast $\mathsf{E}(y_{t+l}|\mathcal{F}_{t-1})$,

$$\begin{aligned} f_{t,0} &= g(Y_{t-1}) \\ f_{t,1} &= \mathsf{E}[Y_{t+1} \mid \mathcal{F}_{t-1}] = \mathsf{E}[g(f_{t,0} + \hat{v}^\star_t) \mid \mathcal{F}_{t-1}] \\ f_{t,2} &= \mathsf{E}[Y_{t+2} \mid \mathcal{F}_{t-1}] = \mathsf{E}[g(f_{t,1} + \hat{v}^\star_{t+1}) \mid \mathcal{F}_{t-1}] \\ &\cdots \end{aligned} \tag{29}$$

with $\hat{v}^\star_{t+l}, l = 1, \ldots, h$ are bootstrap samples from the set of residuals $\{\hat{v}_{t+l}\}_{t=1}^T$.

4. Repeat steps 3 for all N_B vectors of estimated innovations with bootstrap methods, iterating on the estimated model,

$$\begin{aligned} fb_{t,1} &= \frac{1}{N_B} \sum_{i=1}^{N_B} g(f_{t,0} + \hat{v}^{\star(i)}_t) \\ fb_{t,2} &= \frac{1}{N_B} \sum_{i=1}^{N_B} g(g(f_{t,0} + \hat{v}^{\star(i)}_t) + \hat{v}^{\star(i)}_{t+1}) \\ &\cdots \end{aligned} \tag{30}$$

5. By the same logic, we compute $\mathsf{E}(y_{t+l} \mid \delta_{jt} = u_{jt}, \mathcal{F}_{t-1})$ when the shock is given as $\delta_{jt} = \hat{u}^\star_{jt} e_j$,

$$\begin{aligned} \tilde{f}_{t,0} &= g(Y_{t-1}) \\ \tilde{f}_{t,1} &= \mathsf{E}[Y_{t+1} \mid \mathcal{F}_{t-1}] = \mathsf{E}[g(\tilde{f}_{t,0} + \delta_{jt}) \mid \mathcal{F}_{t-1}] \\ \tilde{f}_{t,2} &= \mathsf{E}[Y_{t+2} \mid \mathcal{F}_{t-1}] = \mathsf{E}[g(\tilde{f}_{t,1} + \hat{v}^\star_{t+1}) \mid \mathcal{F}_{t-1}] \\ &\cdots \end{aligned} \tag{31}$$

6. Repeat steps 5 for all N_B vectors of estimated innovations with bootstrap methods, iterating on the estimated model,

$$\begin{aligned} fb_{t,1} &= \frac{1}{N_B}\sum_{i=1}^{N_B} g(f_{t,0} + \hat{u}_{jt}^{\star(i)} e_j) \\ fb_{t,2} &= \frac{1}{N_B}\sum_{i=1}^{N_B} g(g(f_{t,0} + \hat{u}_{jt}^{\star(i)} e_j) + \hat{u}_{j,t+1}^{\star(i)} e_j) \\ &\dots \end{aligned} \tag{32}$$

7. Plug in the GI function

$$GI(l, \delta_{jt}, \mathcal{F}_{t-1}) = \mathsf{E}(y_{t+l} \mid u_{jt} = \delta_{jt}, \mathcal{F}_{t-1}) - \mathsf{E}(y_{t+l} \mid \mathcal{F}_{t-1}) \tag{33}$$

to obtain the relative contribution of a shock δ_{jt} to the i-th variable with horitzon h at time t,

$$\lambda_{ij,\mathcal{F}_{t-1}}(h) = \frac{\sum_{l=0}^{h} GI(l, \delta_{jt}, \mathcal{F}_{t-1})_i^2}{\sum_{j=1}^{K}\sum_{l=0}^{h} GI(l, \delta_{jt}, \mathcal{F}_{t-1})_i^2}, \quad i, j = 1, \dots, K \tag{34}$$

8. Repeat steps 2-6 for all histories.

9. Construct table 3 using averaged $\lambda_{ij,\mathcal{F}_{t-1}}(h)$ generated from step 7.

If there is a unit shock,

$$\delta_{jt} \quad : \quad (\delta_{1t}, \delta_{2t}, \dots, \delta_{Kt})^\top \sim e_j \tag{35}$$

then we can simiply replace $\hat{u}_{jt}^{\star} e_j$ of (27) with e_j of (35), and repeat the steps from 1 to 6 stated above, the generalized impulse response can be calculated based on (33), i.e.

$$GI(l, \delta_{jt}, \mathcal{F}_{t-1}) = \mathsf{E}(y_{t+l} \mid u_{jt} = \delta_{jt}, \mathcal{F}_{t-1}) - \mathsf{E}(y_{t+l} \mid \mathcal{F}_{t-1}) \tag{36}$$

We should note that if K is extremely large in empirical study, the denominator of equation (34) might be unnecessarily large due to accumulated noise caused by the large amount of irrelevant variables. Therefore one more step of prescreening is preferred to filter out less relevant variables.

C Model Selection Consistency

Proof of Theorem C.1

In this section we show the model selection consistency according to (9). The following theorem C.1 derives the statistical properties of the estimate using elastic net. Additionally, the second part of the theorem gives the asymptotic distribution of the estimate.

Theorem C.1. *Suppose that* $\alpha_{1,T}/\sqrt{T} \to 0$ *and* $\alpha_{2,T}/\sqrt{T} \to 0$*. Then the objective function* (9) *yields the following:*

1. $\lim_{T\to\infty} \mathbb{P}(\mathcal{A}_T^* = \mathcal{A}) = 1$;

   where $\mathcal{A}$ *is the set of indices for the non-zero elements of* $vec(A)$, $\mathcal{A}_T^*$ *is the set of indices for the non-zero elements of* $vec(\hat{A})$ *in* (9)

2. $\sqrt{T}vec(\hat{A}_T' - A') \to_d \mathcal{N}(0, \Sigma_u \otimes \Gamma^{-1})$

   where $\frac{1}{T}ZZ' \to_p \Gamma$.

The model selection is consistent, i.e. the true lags are selected with probability one.

Let $vec(\hat{A}) = vec(A) + vec(\frac{1}{\sqrt{T}}e_A)$, where e_A is an $K \times K$ matrix, define

$$\begin{aligned}
\Psi_T(e_A) &= \left\| vec(Y) - (Z' \otimes I_K)vec(A + \frac{1}{\sqrt{T}}e_A) \right\|_2^2 \\
&+ \alpha_1 \left\| vec(A + \frac{1}{\sqrt{T}}e_A) \right\|_1 \\
&+ \alpha_2 \left\| vec(A + \frac{1}{\sqrt{T}}e_A) \right\|_2^2
\end{aligned}$$

where $e_A = [e_{A,1}, \ldots, e_{A,P}]$. Each $e_{A,p}$, $p = 1, \ldots, P$ is a $K \times K$ matrix.

In order to find e_A so as to minimize $\Psi_T(e_A)$. This is equivalent to minimize

$$\begin{aligned}
\Psi_T(e_A) - \Psi_T(0) &= vec(\frac{1}{\sqrt{T}}e_A)'(ZZ' \otimes I_K)vec(\frac{1}{\sqrt{T}}e_A) \\
&- 2vec(U)'(Z' \otimes I_K)vec(\frac{1}{\sqrt{T}}e_A) \\
&+ \alpha_1 \sum_{p=1}^{P}\sum_{i,j=1}^{K}\big(|A_p(i,j) + \frac{1}{\sqrt{T}}e_{A,p}(i,j)| - |A_p(i,j)|\big) \\
&+ \alpha_2 \sum_{p=1}^{P}\sum_{i,j=1}^{K}\big(\frac{1}{T}e_{A,p}^2(i,j) + \frac{2}{\sqrt{T}}A_p(i,j)e_{A,p}^2(i,j)\big)
\end{aligned}$$

For the third term in the above equation, we have $\sqrt{T}\big(|A(i,j)+\frac{1}{\sqrt{T}}e_A(i,j)|-|A(i,j)|\big) \to sign(A(i,j))|e_A(i,j)|$ for $A(i,j) \neq 0$. By Slutsky's theorem, $\alpha_1\big(|A(i,j) + \frac{1}{\sqrt{T}}e_A(i,j)| - |A(i,j)|\big) \to_p 0$ when $\dfrac{\alpha_1}{\sqrt{T}} \to 0$.

For the last term, since $\dfrac{\alpha_2}{\sqrt{T}} \to 0$ we have $\alpha_2\big(\frac{1}{T}e_A^2(i,j) + \frac{2}{\sqrt{T}}A(i,j)e_A^2(i,j)\big) \to 0$.

In addition, by minimizing $\Psi_T(e_A) - \Psi_T(0)$, we have

$$\begin{aligned}
\sqrt{T}vec(\hat{A}_T - A) &= vec(\hat{e}_A) \\
&= \sqrt{T}(ZZ' \otimes I_K^{-1})(Z \otimes I_K)vec(U) \\
&= \sqrt{T}\left((ZZ')^{-1}Z \otimes I_K\right)vec(U)
\end{aligned}$$

Because $\frac{1}{T}ZZ' \to_p \Gamma$ and $\frac{1}{\sqrt{T}}vec(UZ') \to_d N(0, \Gamma \otimes \Sigma_u)$ with $\Sigma_u = \mathsf{E}(uu')$ hold, the asymptotic distribution of $\sqrt{T}vec(\hat{A}_T - A)$ is normal and the covariance matrix is

$$(\Gamma^{-1} \otimes I_K)(\Gamma \otimes \Sigma_u)(\Gamma^{-1} \otimes I_K)$$

Thus, $\sqrt{T}vec(\hat{A}_T' - A') \to_d \mathcal{N}(0, \Sigma_u \otimes \Gamma^{-1})$ $\square$

D Network Aggregation Details

Case 1: Own-price/Cross-price market impact We have

$$\begin{aligned}
\mathcal{V} &= Y \qquad (37)\\
\mathcal{E}_o &= \left\{ \begin{array}{c} C^o_{i\leftarrow j} = \begin{cases} C^o_{p^{(k+1)}\leftarrow as_r^{(k+1)}} = C_{7k+1\leftarrow 7k+1+r} & \text{for } i = j - r = 7k+1 \\ C^o_{p^{(k+1)}\leftarrow bs_r^{(k+1)}} = C_{7k+1\leftarrow 7k+4+r} & \text{for } i = j - 4 - r = 7k+1 \end{cases} \\ k = 0, \dots, N-1, r = 1, \dots 3 \end{array} \right\}
\end{aligned}$$

Hence the network $\mathcal{G}_{own} = (\mathcal{V}, \mathcal{E}_o)$ displays the effects of different levels of quotes on the stock price within the LOB of each stock.

For studying this cross-price market impact, we thus work with $\mathcal{G}_{cross} = (\mathcal{V}_1, \mathcal{E}_{c,1})$, where the ask-side (bid-side) of each stock is represented by a single node $as^{(n)}(bs^{(n)})$:

$$\begin{aligned}
\mathcal{V}_1 &= \Big((p^{(1)}, as^{(1)}, bs^{(1)}), \dots, (p^{(N)}, as^{(N)}, bs^{(N)})\Big)^\top \qquad (38)\\
\mathcal{E}_{c,1} &= \left\{ \begin{array}{c} C^1_{i\leftarrow j} = \begin{cases} C^1_{p^{(k+1)}\leftarrow as^{(z+1)}} = \sum_{r=1}^3 C_{7k+1\leftarrow 7z+1+r} & \text{for } i = 3k+1, j = 3z+2 \\ C^1_{p^{(k+1)}\leftarrow bs^{(z+1)}} = \sum_{r=1}^3 C_{7k+1\leftarrow 7z+4+r} & \text{for } i = 3k+1, j = 3z+3 \end{cases} \\ k \neq z = 0, \dots, N-1 \end{array} \right\} \\
\mathcal{E}_{o,1} &= \left\{ \begin{array}{c} C^1_{i\leftarrow j} = \begin{cases} C^1_{p^{(k+1)}\leftarrow as^{(k+1)}} = \sum_{r=1}^3 C_{7k+1\leftarrow 7k+1+r} & \text{for } i = j - 1 = 3k+1 \\ C^1_{p^{(k+1)}\leftarrow bs^{(k+1)}} = \sum_{r=1}^3 C_{7k+1\leftarrow 7k+4+r} & \text{for } i = j - 2 = 3k+1 \end{cases} \\ k = 0, \dots, N-1 \end{array} \right\}
\end{aligned}$$

Case 2: Trade imbalance between bids and asks To measure potential asymmetry of bid and ask side effects on different stock prices, spillovers are aggregated across stocks in each quote level to determine the respective overall impact. We thus work with a network with graph $\mathcal{G}_s = (\mathcal{V}_2, \mathcal{E}_{s,2})$

where each level r of the ask (bid) side is represented by as_r (bs_r)

$$
\begin{aligned}
\mathcal{V}_2 &= \Big((p^{(1)}, \ldots, p^{(N)}), (as_1, as_2, as_3), (bs_1, bs_2, bs_3)\Big)^\top \qquad (39) \\
\mathcal{E}_{s,2} &= \left\{ \begin{array}{c} C^{s,2}_{i\leftarrow j} = \begin{cases} C^2_{p^{(k+1)}\leftarrow as_r} = \sum_{l=0}^{N-1} C_{7k+1\leftarrow 7l+1+r} & \text{for } i = k+1, j = N+r \\ C^2_{p^{(k+1)}\leftarrow bs_r} = \sum_{l=0}^{N-1} C_{7k+1\leftarrow 7l+4+r} & \text{for } i = k+1, j = N+r+3 \end{cases} \\ k = 0, \ldots, N-1 \qquad r = 1,2,3 \end{array} \right\}
\end{aligned}
$$

For an overview, we work with $\mathcal{G}_{sp} = (\mathcal{V}_3, \mathcal{E}_{s,3})$ where the price is denoted as p

$$
\begin{aligned}
\mathcal{V}_3 &= (p, (as_1, as_2, as_3), (bs_1, bs_2, bs_3))^\top \qquad (40) \\
\mathcal{E}_{s,3} &= \left\{ \begin{array}{c} C^{s,3}_{i\leftarrow j} = \begin{cases} C^3_{p\leftarrow as_r} = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} C_{7k+1\leftarrow 7l+1+r} & \text{for } i = 1, j = 1+r \\ C^3_{p\leftarrow bs_r} = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} C_{7k+1\leftarrow 7l+4+r} & \text{for } i = 1, j = 4+r \end{cases} \\ r = 1,2,3 \end{array} \right\}
\end{aligned}
$$

In this network $\mathcal{G}_{sp}$, all effects are aggregated across stocks.

Case 3: Price-only effects and the fully LOB-aggregated network per stock For studying only price effects within the LOB-network, we consider a network with graph $\mathcal{G}_p = (\mathcal{V}_4, \mathcal{E}_{p,4})$

$$
\begin{aligned}
\mathcal{V}_4 &= Y_p = \left(p^{(1)}, \ldots, p^{(N)}\right)^\top \qquad (41) \\
\mathcal{E}_{p,4} &= \left\{C^{p,4}_{i\leftarrow j} = C_{7(i-1)+1\leftarrow 7(j-1)+1} \text{for } i \neq j = 1, \ldots, 7\right\}
\end{aligned}
$$

The price-only case is compared to the case where we aggregate all pairwise LOB-network spillover effects between two stocks. Thus we work with the network $\mathcal{G}_g = (\mathcal{V}_5, \mathcal{E}_{g,5})$ where each stock is represented by a node $g^{(n)}$.

$$
\begin{aligned}
\mathcal{V}_5 &= (g^{(1)}, g^{(2)}, \ldots, g^{(N)})^\top \qquad (42) \\
\mathcal{E}_{g,5} &= \left\{C^{g,5}_{i\leftarrow j} = \sum_{m=7i+1}^{7i+7}\sum_{l=7j+1}^{7j+7} C_{m\leftarrow l} \quad \text{for } i \neq j = 1, \ldots, N\right\}
\end{aligned}
$$

E Tables and figures

		MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$m(C^{s,2}_{p^{(n)}\leftarrow as_1})$	period I	1.29	0.69	0.75	1.76	0.31	0.98	0.35	0.70	1.89
	period II	0.34	1.79	0.74	0.33	0.26	0.88	4.42	2.10	0.55
	period III	0.27	0.64	0.81	0.09	1.11	1.56	0.68	1.45	1.84
	full period	0.39	0.96	0.75	0.42	0.48	1.25	0.97	0.99	**1.41**
$m(C^{s,2}_{p^{(n)}\leftarrow as_2})$	period I	0.99	0.50	4.38	0.53	0.42	1.33	0.50	0.82	0.85
	period II	3.77	1.82	0.21	1.81	3.26	4.15	2.60	0.52	0.91
	period III	0.37	0.31	0.69	0.13	0.54	0.88	0.32	0.75	0.74
	full period	0.96	**1.51**	0.80	0.83	0.75	1.41	0.52	0.69	0.79
$m(C^{s,2}_{p^{(n)}\leftarrow as_3})$	period I	0.97	0.42	1.91	0.24	0.37	0.36	6.20	0.26	0.33
	period II	9.82	2.20	0.39	1.11	2.43	2.11	1.26	2.10	2.68
	period III	0.73	1.53	1.31	2.07	0.34	0.75	1.26	0.95	0.96
	full period	1.33	1.24	0.85	0.81	0.44	0.69	**1.53**	0.53	0.65
$m(C^{s,2}_{p^{(n)}\leftarrow bs_1})$	period I	0.36	0.48	0.80	1.14	0.30	0.10	0.17	0.15	0.68
	period II	0.77	1.79	0.39	0.48	1.66	1.58	1.17	1.94	1.34
	period III	0.74	0.64	0.44	0.39	0.62	1.89	1.33	1.69	0.81
	full period	0.56	0.80	0.53	0.62	0.82	0.36	0.65	0.52	**0.90**
$m(C^{s,2}_{p^{(n)}\leftarrow bs_2})$	period I	1.30	0.37	1.94	0.26	0.57	0.10	0.19	0.78	0.62
	period II	0.34	1.75	0.30	0.16	0.62	0.74	0.94	0.61	1.00
	period III	1.35	0.77	0.14	0.89	1.09	0.41	0.64	0.46	0.65
	full period	0.78	**0.83**	0.64	0.30	0.77	0.32	0.33	0.54	0.81
$m(C^{s,2}_{p^{(n)}\leftarrow bs_3})$	period I	0.88	0.52	0.50	0.31	0.31	0.88	0.57	1.26	0.82
	period II	0.96	2.54	0.20	0.99	0.56	1.14	1.27	3.16	3.78
	period III	0.66	0.66	0.38	0.09	0.55	0.27	0.38	2.69	1.06
	full period	0.76	1.21	0.31	0.60	0.41	0.83	0.59	**1.82**	1.14

Table 12: **Quote level effects**: Each element of the upper half of the table contains the aggregated market impact $m(C^{s,2}_{p^{(l)}\leftarrow as_r})$ of the ask side level r across all stocks on the price of stock l as defined in (16) displayed as median m over different parts of the full sample period. In the lower half of the table corresponding bid-side effects $m(C^{s,2}_{p^{(l)}\leftarrow bs_r})$ are reported.All other specifications are as in Table 10.

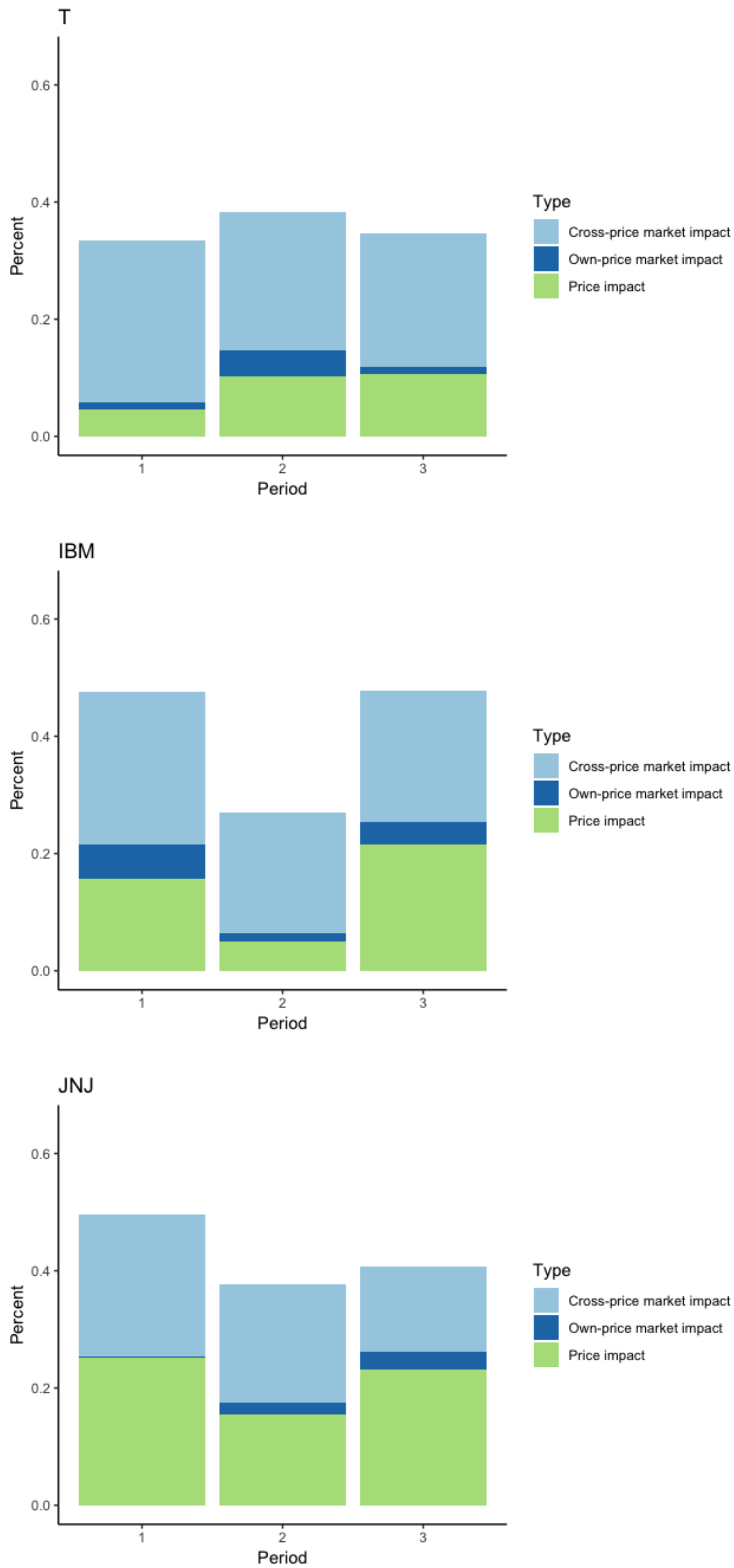

Figure 8: **Price impacts in subperiods**: This figure this graphically depicts the composition of received price impacts from different sources as percentage aggregates over time in the three subperiods for AT&T, IBM and Johnsons&Johnson. All colors and specifications are as in Figure 4.

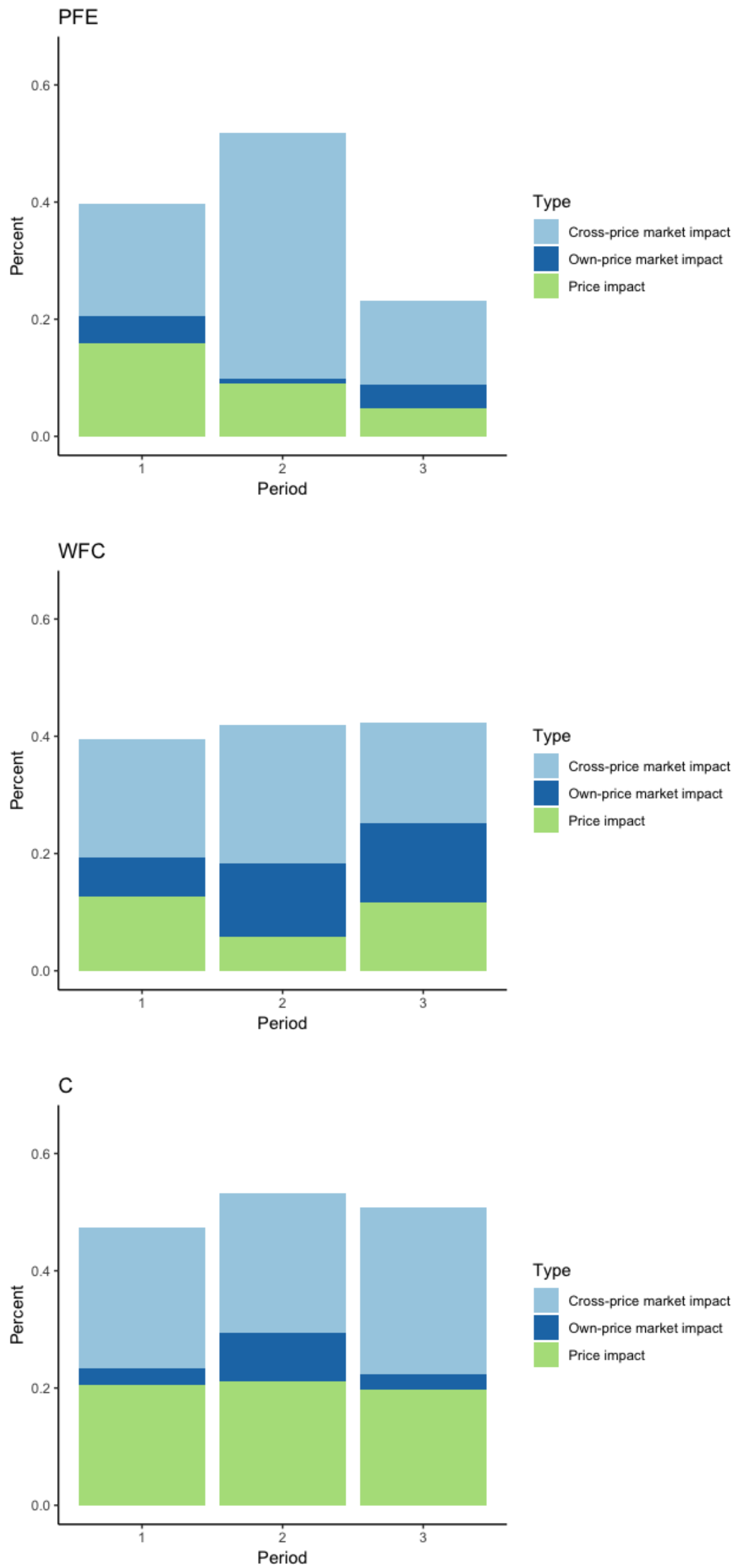

Figure 9: **Price impacts in subperiods**: This figure this graphically depicts the composition of received price impacts from different sources as percentage aggregates over time in the three subperiods for Pfizer, Wells Fargo and Citigroup. All colors and specifications are as in Figure 4.

IRTG 1792 Discussion Paper Series 2021

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Surrogate Models for Optimization of Dynamical Systems" by Kainat Khowaja, Mykhaylo Shcherbatyy, Wolfgang Karl Härdle, January 2021.

002 "FRM Financial Risk Meter for Emerging Markets" by Souhir Ben Amor, Michael Althof, Wolfgang Karl Härdle, February 2021.

003 "K-expectiles clustering" by Bingling Wang, Yingxing Li, Wolfgang Karl Härdle, March 2021.

004 "Understanding Smart Contracts: Hype or Hope?" by Elizaveta Zinovyev, Raphael C. G. Reule, Wolfgang Karl Härdle, March 2021.

005 "CATE Meets ML: Conditional Average Treatment Effect and Machine Learning" by Daniel Jacob, March 2021.

006 "Coins with benefits: on existence, pricing kernel and risk premium of cryptocurrencies" by Cathy Yi-Hsuan Chen, Dmitri Vinogradov, April 2021.

007 "Rodeo or Ascot: which hat to wear at the crypto race?" by Konstantin Häusler, Wolfgang Karl Härdle, April 2021.

008 "Financial Risk Meter based on Expectiles" by Rui Ren, Meng-Jou Lu, Yingxing Li, Wolfgang Karl Härdle, April 2021.

009 "Von den Mühen der Ebenen und der Berge in den Wissenschaften" by Annette Vogt, April 2021.

010 "A Data-driven Explainable Case-based Reasoning Approach for Financial Risk Detection" by Wei Li, Florentina Paraschiv, Georgios Sermpinis, July 2021.

011 "Valuing cryptocurrencies: Three easy pieces" by Michael C. Burda, July 2021.

012 "Correlation scenarios and correlation stress testing" by Natalie Packham, Fabian Woebbeking, July 2021.

013 "Penalized Weigted Competing Risks Models Based on Quantile Regression" by Erqian Li, Wolfgang Karl Härdle, Xiaowen Dai, Maozai Tian, July 2021.

014 "Indices on Cryptocurrencies: an Evaluation" by Konstantin Häusler, Hongyu Xia, August 2021.

015 "High-dimensional Statistical Learning Techniques for Time-varying Limit Order Book Networks" by Shi Chen, Wolfgang Karl Härdle, Melanie Schienle, August 2021.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.