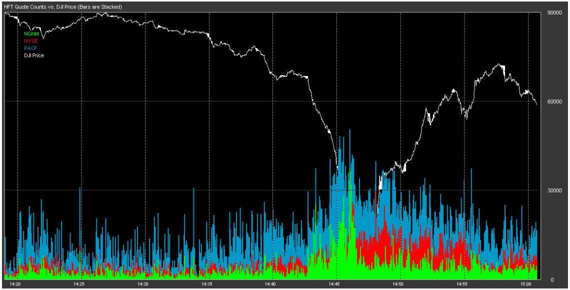
Time-varying Limit Order Book Networks

Shi Chen
Wolfgang Karl Härdle
Chong Liang
Melanie Schienle

Ladislaus von Bortkiewicz Chair of Statistics
Humboldt–Universität zu Berlin
Chair of Econometrics and Statistics
Karlsruher Institut für Technologie
http://lvb.wiwi.hu-berlin.de
https://statistik.econ.kit.edu

High Frequency Trading



- Rapid-fire trading makes decisions in *m*s.
- 10 seconds of HFT in super slow motion.

HF Data Overview

- Take the best ask price on one single trading day as example,

	Microsoft	Pfizer	Citigroup
NumObs ($*10^3$)	584.55	427.51	472.90

Table 1: *NumObs* denotes the number of tick-by-tick observations.

- Size of data sample, including 9 stocks over 2 months: 23.4 gigabyte (GB)

	Microsoft	JP & Morgan	Citigroup
Size, GB	5.0	3.3	3.5

Data Source

- Reconstructed limit order book (LOB) with NASDAQ traded stocks.
- Source: lobsterdata.com
  - NASDAQ'S historical TotalView - ITCH files
  - Detailed event information

LOBSTER
Limit Order Book System – The Efficient Reconstructor

Nasdaq

LOB Data Structure

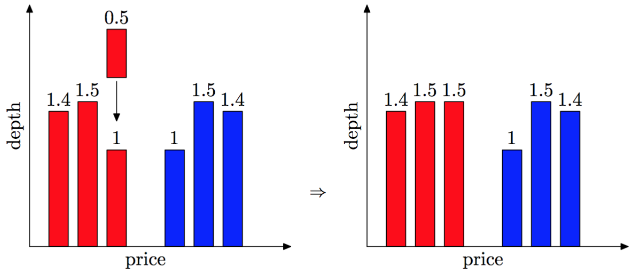
- Indicators for the type of event causing an update of the limit order book. e.g. 34713 sec = 9:38:55am; Direction−1: sell

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

- Evolution of the limit order book up to the requested number of levels.

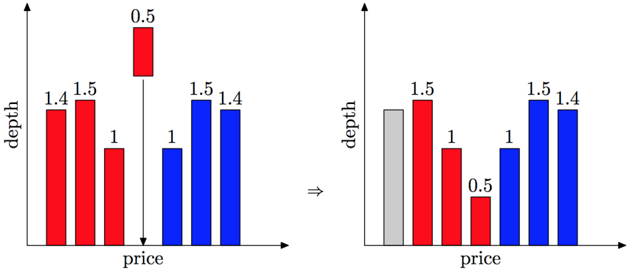
Ask Price 1	Ask Size 1	Bid Price 1	Bid Size 1	Ask Price 2	Ask Size 2	Bid Price 2	Bid Size 2	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1186600	9484	118500	8800	118700	22700	118400	14930	...
1186600	9384	118500	8800	118700	22700	118400	14930	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

LOB - Normal Limit Order



- arrival of a buy limit order with price 1000 and size 0.5 to be placed.

LOB - Aggressive Limit Order



- arrival of a buy limit order with price 1001 and size 0.5 to be posted inside of the current spread. ▸ more details: passive limit order

Research Interest

- How does the order flows interact with price dynamics, and further affect the market behavior?
- Are the impacts on price responding to incoming ask and bid market/limit orders widely symmetric? If not, how does the heterogeneous market impact caused by bid and ask order for various stocks affect the whole market?
- How to measure the impact of market/limit order quantitatively?

Challenges

1. Construct LOB network in the presence of microstructure noise (MN) and non-synchronous trading.
   - propose volume synchronization algorithm.
2. Extend current literature where the connectedness measures are often estimated by MA representation of VAR systems and restricted to Gaussian innovations
   - combine bootstrap-based generalized impulse response analysis with network construction.
3. Efficient programming to speed up numerical computation ($<$ 1 day).

LOB Network

Network structure with sparsity to find relevant intercon-nections

Individual stock network

- **stock $A \overset{impact}{\Longleftrightarrow}$ stock B**
- **with and without order flows**

LOB network

- **Asymmetric market sell/buy pressure**
- **Own-price/cross-price market impact**

- Focus on the dynamics of LOB networks and their evolution.

Programming Micro-Benchmarks

- Benchmark times relative to Cpp (smaller is better)

	Cpp gcc 4.8.5	Matlab R 2018a	Python 3.6.3	R 3.3.1
iteration-pi-sum	1.00	1.01	14.75	8.92
recursion-fibonacci	1.00	18.69	100.77	608.81
parse-integers	1.00	229.56	19.98	50.90
matrix-statistics	1.00	8.10	17.93	20.35
matrix-multiply	1.00	1.16	1.18	8.74
userfunc-mandelbrot	1.00	10.07	132.38	333.03

Outline

All QuantLets from www.quantlet.de

Microstructure Noise

- High Frequency (HF) data is contaminated by market frictions.
  - stale prices due to infrequent trading.
  - price discreteness, rounding
  - bid-ask bounds and misprints.
- Tick size is one cent for any stock over 1 USD.
- More liquid stocks have small tick sizes.

Basic Model

- MN contaminated Y_t (log price) with latent X_t,

$$Y_t = X_t + \varepsilon_t, \quad t \geq 0$$

  with $\mathrm{E}(\varepsilon_t | X) = 0$.

- Efficient log price X_t is semi-martingale, Delbaen and Schachermayer (1994),

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t \sigma_s dW_s$$

- $(a_s)_{s \geq 0}$ càdlàg drift process, $(\sigma_s)_{s \geq 0}$ càdlàg volatility process.

Estimation Strategy

- Pre-averaging mitigates MN,
  - transaction or quote data are averaged over short time periods ranging from 30s to 5m.
  - the resulting average approximates the efficient price process much better than raw data.
- Portfolio selection using HF data
- Match price and size.

Pre-averaged Estimator

- Choose the "kernel" of $g(x) = x \wedge (1 - x)$, Podolskij et al. (2009), Christensen et al. (2010). More details of model setting
- Then pre-average returns with g,

$$\overline{Y}_i^n = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) \Delta_{i+j}^n Y$$

$$= -\sum_{j=0}^{k_n-1} \left\{ g\left(\frac{j+1}{k_n}\right) - g\left(\frac{j}{k_n}\right) \right\} Y_{i+j}^n$$

for $i = 0, \ldots, n - k_n + 1$.

Size Intensity

- In contrast to a moderate time interval for price to reduce the MN, the time interval for trading volumes should be small enough to capture the trading crowd.
- Define size intensity as $\tilde{S}_{t_j}$, t_j denotes the time stamp of jth LOB.

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

- The *size intensity* can be summed up over a given time interval and can be matched with returns.

Volume Synchronization Algorithm

1. Set equally-spaced time intervals $T_0 + k\Delta T$, $k = 0, 1, 2, \ldots, K$
2. Define the price and size at time $T_0 + k\Delta T$ as

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

3. The changes of the log values are

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

Volume Synchronization Algorithm - ctd

4. Pre-averaging both $\Delta p_{T_0+k\Delta T}$ and $\Delta s_{T_0+k\Delta T}$ by

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

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

where $g_j \geq 0$ and $\sum_{j=0}^{J} g_j = 1$.

hfhd_dataclean

Variables

- Prepare data in this way helps to:
  - alleviate microstructure noise
  - match the price to the size
  - solve non-synchronicity.
- Record the mid price on the first level, the ask and bid sizes on the first 3 levels,

$$y_t^{(n)\top} = [\Delta\tilde{p}_t^{(n)}, \Delta\tilde{s}_t^{a1(n)}, \Delta\tilde{s}_t^{a2(n)}, \Delta\tilde{s}_t^{a3(n)}, \Delta\tilde{s}_t^{b1(n)}, \Delta\tilde{s}_t^{b2(n)}, \Delta\tilde{s}_t^{b3(n)}]$$

This yields a vector,

$$Y_t^\top = [y_t^{(1)\top}, y_t^{(2)\top}, \dots, y_t^{(N)\top}]$$

that is stacked by $y_t^{(n)T}$ for different N stocks.

Penalized VAR Approach

- The VAR(p) model of Y_t,

$$
\begin{aligned}
Y_t &= \nu + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + u_t \qquad (1) \\
&= \nu + (A_1, A_2, \ldots, A_p) \left(Y_{t-1}^\top, Y_{t-2}^\top, \ldots, Y_{t-p}^\top \right)^\top + u_t
\end{aligned}
$$

- Assume (1) satisfies,
  1. The roots of $|I_K - \sum_{j=1}^{p} A_j z^j| = 0$ lie outside unit circle.
  2. u_t are i.i.d innovations;
     each element has bounded $(4+\delta)$th moment, $\delta > 0$.
  3. $\|\Sigma_u\|_2 < \infty$ and $\| (A_1, A_2, \ldots, A_p) \|_2 < \infty$.

Penalized VAR Approach

- For the multivariate case,

$$vec(Y) = (Z^\top \otimes I_K)vec(A) + vec(U) \tag{2}$$

- and

$$\begin{aligned} Y &= (Y_1, Y_2, \ldots, Y_T) \qquad & A = (A_1, A_2, \ldots, A_p) \\ Z_t &= (y_t, y_{t+1}, \ldots, y_{t-p+1})^\top \qquad & Z = (Z_0, Z_1, \ldots, Z_{T-1}) \end{aligned}$$

where A_i are $(K \times K)$, $K = 7N$.

- The compact form is equivalent to,

$$\mathbf{y} = (Z^\top \otimes I_K)\beta + \mathbf{u} = \mathbf{x}\beta + \mathbf{u}$$

Estimation Strategy

- Elastic net:

$$\arg\min_{\beta} \left(\|\mathbf{y} - \mathbf{x}\beta\|_2^2 + \alpha_{1,T}\|\beta\|_1 + \alpha_{2,T}\|\beta\|_2^2 \right)$$

equivalent to:

$$\arg\min_{A_1, A_2, \ldots, A_p} \sum_{t=1}^{T} \| Y_t - \sum_{j=1}^{p} A_j Y_{t-j} \|_2^2 + \alpha_{1,T} \sum_{j=1}^{p} \| vec(A_j) \|_1 + \alpha_{2,T} \sum_{j=1}^{p} \| vec(A_j) \|_2^2$$

  - $\|M\|_p$ depends on whether M is a vector or a matrix.
  - To avoid confusion, $vec(M)$ used to transform the object within $\|\|_p$ into a vector.
  - Reference: Basu et al. (2015); Liang and Schienle (2017).

Generalized Impulse Response Function

- Assume shocks hitting at j-th equation.
- Define the generalized impulse response GI, as shock δ_{jt} on j-th equation of y_t at horizon l,

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

- $\mathsf{E}(y_{t+1} | u_{jt} = \delta_{jt}, \mathcal{F}_{t-1})$ represents the expectation conditional on the history $\mathcal{F}_{t-1}$ and a fixed value of j-th shock on time t.

GI Analysis

$GI(l, \delta_{jt}, \mathcal{F}_{t-1})$

$\delta_{jt} : (\delta_{1t}, \delta_{2t}, \ldots, \delta_{Kt})^\top \sim e_j$
measure the persistent effect of a shock on the behavior of a series

$\delta_{jt} : (\delta_{1t}, \delta_{2t}, \ldots, \delta_{Kt})^\top \sim \hat{u}^\star_{jt} e_j$
use *GI* to construct connectedness measure

- *GI* with unit shock, Koop (1996). ▸ Measuring price direction
- *GI* with shock coming from system, Lanne and Nyberg (2016).
- Conditional mean forecast for more than one period ahead with bootstrap-based method. ▸ Bootstrap-based multistep forecast method

Fashionable Connectedness Measures

- Construct connectedness measures with generalized forecast error variance decomposition (GFEVD).
- DY-connectedness of Diebold and Yilmaz (2014): use GFEVD of Koop et al. (1996), Pesaran and Shin (1998)
  - The relative contributions to *h*-period impact of the shocks do not sum to unity.
  - GFEVD is constructed by orthogonalized impulse response function based on MA representation of VAR
- In this paper, we use the LN-type GFEVD,
  - LN-GFEVD of Lanne and Nyberg (2016) is a modification of GFEVD
  - The relative contributions to *h*-period impact of the shocks sum to unity, facilitates convenient interpretation.
  - The LN-GFEVD can be obtained in nonlinear models.

New Connectedness Measure

- The LN-GFEVD is defined by j-th shock hitting i-th variable at time t with h forecast horizon,

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

  - measuring the relative contribution of a shock δ_{jt} to the j-th equation in relation to the total impact of all K shocks on the i-th variable in y_t after h periods.

- Conditional mean forecast for more than one period ahead with bootstrap yields $\lambda_{ij}^{b}(h)$.
- The pairwise directional connectedness is $C_{i\leftarrow j} = \lambda_{ij}^{b}(h)$

Connectedness Table

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

Table 2: Connectedness table of interest, estimated by bootstrap.

- LN-GFEVD is economically interpretable since $\sum_{j=1}^K \lambda_{ij}^b(h) = 1$.
- The bootstrapped $C_{i \leftarrow j} = \lambda_{ij}^b(h)$ relies neither on the ordering of the variables nor on the distribution of the innovations.

Dataset

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

- Sample data from 01.06 to 29.07.2016, MktCap is the market capitalization by Feb 23rd, 2018.

Dataset -ctd

- Divide the trading period into 1-minute intervals, i.e. $\Delta T = 1$ min.
  - Calculate mid price on the first level, obtain $\Delta \tilde{p}_t^{(n)}$
  - Calculate corresponding bid and ask sizes (i.e. size intensity) on the first three levels, i.e. market order, best limit order and 2nd best limit order, i.e. $\Delta \tilde{s}_t^{a1(n)}, \Delta \tilde{s}_t^{a2(n)}, \Delta \tilde{s}_t^{a3(n)}, \Delta \tilde{s}_t^{b1(n)}, \Delta \tilde{s}_t^{b2(n)}, \Delta \tilde{s}_t^{b3(n)}$
- Then pre-average the above values over smooth-window of 15 minutes.
- Upon the estimates of the sparse HD VAR model, calculate the bootstrapped LN-GFEVD and corresponding connectedness at horizon $h = 30$ for every trading day.

Individual Stock Network

- A network is a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with,
  - nodes (or vertices) $\mathcal{V}$: entities we are evaluating
  - edges $\mathcal{E}$: connections between entities
- Cross-stock network $\mathcal{G}_p = (\mathcal{V}_p, \mathcal{E}_p)$ with only price factors $p^{(n)}$,

$$\mathcal{V}_p = p^{(n)}, \quad n = 1, \ldots, N \quad \text{and} \quad \mathcal{E}_p = C_{i \leftarrow j}, \quad i, j \in \mathcal{V}_p \quad (3)$$

Pairwise Stock Connectedness

- To evaluate the node importance, use
  - degree centrality $deg(\mathcal{V})$: interpreted as a form of popularity.
  - closeness centrality $Clos(\mathcal{V})$: distinguish influencers in the network.
  - betweenness centrality $Bet(\mathcal{V})$: help to decide which nodes act as "bridges" between nodes in a network

Pairwise Stock Connectedness - ctd

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$\mu_{outdeg}(\mathcal{V}_p)$	3.26	3.33	2.52	3.33	3.07	2.81	2.83	2.50	3.40
$\mu_{indeg}(\mathcal{V}_p)$	3.88	2.86	3.43	3.57	2.38	1.69	3.21	2.60	3.45
$\mu_{Clos}(\mathcal{V}_p)$	167.70	163.99	173.45	159.09	159.43	154.56	175.89	171.95	157.81
$\mu_{Bet}(\mathcal{V}_p)$	6.00	7.55	7.98	4.33	4.43	3.02	5.98	5.24	6.07

Table 3: Period: 06-07.2016. $\mu.$ is the mean.

- Citigroup, AT&T and Johnson&Johnson are central in the network, they are choice maker. Meanwhile JNJ is a choice receiver with high "in-degree".
- Conventional centrality measure works well for probing certain phenomena, it fails to capture the node's spreading potential, e.g. JNJ

Including Order Flows

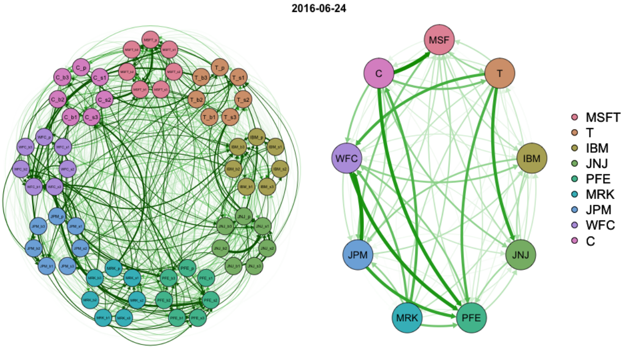


Figure 1: Left panel: the full sample network plot. Right panel: the aggregated network plot of nine stocks, on 24.06.2016. hfhd_cirnet

Including Order Flows - graph

- Aggregated individual stock network $\mathcal{G}_g = (\mathcal{V}_g, \mathcal{E}_g)$:

$$
\begin{aligned}
\mathcal{V}_g &= v_g^{(n)} && (4)\\
v_g^{(n)} &= p^{(n)} + \sum_r bs_r^{(n)} + \sum_r as_r^{(n)}, \quad n = 1, \dots, N && (5)\\
\mathcal{E}_g &= C_{i\leftarrow j} = \lambda_{ij}, \quad i, j \in \mathcal{V}_g && (6)
\end{aligned}
$$

- $as_r^{(n)}$ and $bs_r^{(n)}$ are the r-th level ask/bid size factors for stock n.
- By including size factors from LOB, we are able to investigate how the network is affected by the presence of liquidity effects.

Including Order Flows - ctd

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$\mu_{outdeg}(\mathcal{V}_g)$	128.83	147.02	132.71	129.76	127.95	125.48	120.31	113.50	123.00
$\mu_{indeg}(\mathcal{V}_g)$	136.29	122.50	121.24	118.31	121.88	121.00	136.79	133.98	136.60
$\mu_{Clos}(\mathcal{V}_g)$	0.13	0.13	0.12	0.13	0.13	0.13	0.13	0.13	0.13
$\mu_{Bet}(\mathcal{V}_g)$	4.17	3.26	2.36	2.07	2.79	2.19	3.12	3.57	3.69

Table 4: Period: 06-07.2016. $\mu.$ is the mean.

- Produces different results compared to those obtained for pairwise stock network in Table 2.
- The impacts caused by less important nodes may be neglected by conventional centrality measures, this will potentially cause inaccuracy and thus result in the poor performance.

Including Order Flows

- Conventional centrality measures are rarely accurate when the majority of nodes are not highly influential in the network.
- Use the quantiles to measure the time-varying net spillover effects for each stock,

$$
\begin{aligned}
C_i &= C_{\bullet\leftarrow i} - 2r - 1 = \sum_j C_{j\leftarrow i} - 2r - 1, \quad i,j \in \mathcal{V}_g \\
Q_{C_i}(\alpha) &= F^{-1}(\alpha) = \inf\{C_i : F(C_i) \geq \alpha\} \quad (7) \\
\alpha &= 5\%, 15\%, 50\%, 85\%, 95\% \quad (8)
\end{aligned}
$$

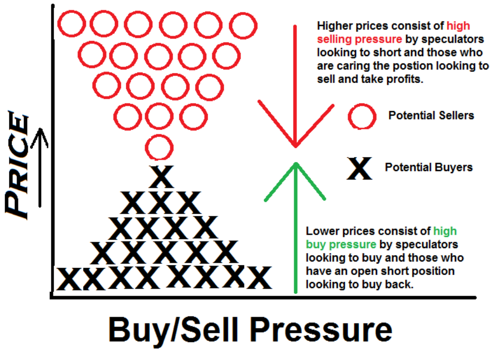
- Quantiles of net connectedness performs better than conventional centrality measures.

Including Order Flows

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$Q_{C_i}(0.05)$	-3.19	-3.51	-3.72	-3.79	-2.84	-3.38	-2.57	-2.80	-3.35
$Q_{C_i}(0.15)$	-2.15	-3.01	-3.24	-2.99	-2.16	-2.62	-1.97	-2.30	-2.61
$Q_{C_i}(0.50)$	0.17	-0.70	-0.91	-0.50	0.14	-0.71	0.97	0.34	0.35
$Q_{C_i}(0.85)$	2.01	1.47	1.31	2.27	1.78	2.47	3.68	3.21	3.04
$Q_{C_i}(0.95)$	3.84	2.54	2.99	2.70	3.81	3.28	5.11	4.63	4.74

- JP Morgan is most influential in the network, while IBM and AT&T are main risk receivers in the aggregated system.
- Financial companies are dominant stocks driving the networks
- The aggregated individual stock network is a better measure of how central a stock is within the network since it takes into consideration the trading volumes.

Market Interaction



Asymmetric Market Sell/Buy Pressure

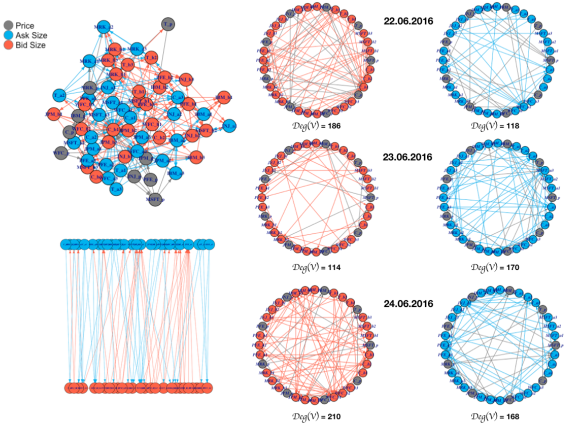


Figure 2: LOB networks from 22.06.2016-24.06.2016. 23.06: Brexit

Asymmetric Market Sell/Buy Pressure - formula

- The LOB network during Brexit announcement: The impacts on returns respond to ask and bid limit orders are not symmetric.
- Construct a graph $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ to study the asymmetric impact from aggregated size factors to price factors,

$$\mathcal{V}_s = \left(p^{(n)}, \sum_n bs_r^{(n)}, \sum_n as_r^{(n)}\right) \quad n = 1, \ldots, N \tag{9}$$

$$\mathcal{E}_s = C_{i \leftarrow j} \quad i \in \{p^{(n)}\}, \quad j \in \left\{\sum_n bs_r^{(n)}, \sum_n as_r^{(n)}\right\} \tag{10}$$

Asymmetric Market Sell/Buy Pressure - ctd

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$\mu C_{p(n) \leftarrow \sum as\mathbf{1}}$	0.44	0.43	0.45	0.34	0.45	0.68	0.86	0.53	0.83
$\mu C_{p(n) \leftarrow \sum as\mathbf{2}}$	0.53	0.41	0.59	0.48	0.60	0.65	0.33	0.50	0.64
$\mu C_{p(n) \leftarrow \sum as\mathbf{3}}$	0.90	0.69	0.54	0.33	0.58	0.89	1.03	0.54	0.39
$\mu C_{p(n) \leftarrow \sum bs\mathbf{1}}$	0.12	0.44	0.47	0.40	0.32	0.41	0.46	0.61	0.40
$\mu C_{p(n) \leftarrow \sum bs\mathbf{2}}$	0.50	0.35	0.38	0.20	0.37	0.22	0.48	0.20	0.28
$\mu C_{p(n) \leftarrow \sum bs\mathbf{3}}$	0.63	0.37	0.24	0.39	0.51	0.38	0.43	0.73	0.47

Table 5: Summary of the aggregated impacts from size factors to the stock price factor from 06.2016-07.2017.

- The higher are the values, the stronger are the stocks affected by trading activities over time.
- Financial stocks: price patterns are highly related to market trading activity.

Asymmetric Market Sell/Buy Pressure - ctd

	$Q_C(0.25)$	$Q_C(0.50)$	$Q_C(0.75)$	μ_C
$C_{\sum_N p \leftarrow \sum_N as_1^{(n)}}$	0.12	0.27	0.67	0.50
$C_{\sum_N p \leftarrow \sum_N as_2^{(n)}}$	0.19	0.30	0.55	0.47
$C_{\sum_N p \leftarrow \sum_N as_3^{(n)}}$	0.17	0.35	0.83	0.59
$C_{\sum_N p \leftarrow \sum_N bs_1^{(n)}}$	0.09	0.18	0.39	0.36
$C_{\sum_N p \leftarrow \sum_N bs_2^{(n)}}$	0.08	0.16	0.41	0.30
$C_{\sum_N p \leftarrow \sum_N bs_3^{(n)}}$	0.13	0.29	0.63	0.42

Table 6: Summary of the impacts from aggregated size factors to the aggregated price factor from 06.2016-07.2017.

- Results are consistent with Table 5. We can observe stronger impacts on prices caused by market sell pressure.

Asymmetric Market Sell/Buy Pressure - test

- let μ_1 be the mean of the overall impacts from selling orders over the sample period ($T = 42$), and μ_2 the corresponding mean of the overall impacts from buying orders,

$$
\begin{aligned}
\mu_1 &= \frac{1}{3T}\left(C_{t,\sum_N p \leftarrow \sum_N as_{\mathbf{1}}^{(n)}} + C_{t,\sum_N p \leftarrow \sum_N as_{\mathbf{2}}^{(n)}} + C_{t,\sum_N p \leftarrow \sum_N as_{\mathbf{3}}^{(n)}}\right) \\
\mu_2 &= \frac{1}{3T}\left(C_{t,\sum_N p \leftarrow \sum_N bs_{\mathbf{1}}^{(n)}} + C_{t,\sum_N p \leftarrow \sum_N bs_{\mathbf{2}}^{(n)}} + C_{t,\sum_N p \leftarrow \sum_N bs_{\mathbf{3}}^{(n)}}\right)
\end{aligned}
$$

- The hypothesis of interest can be expressed as,

$$
\begin{aligned}
H_0 &: \mu_1 - \mu_2 = 0 \\
H_a &: \mu_1 - \mu_2 > 0
\end{aligned}
$$

Asymmetric Market Sell/Buy Pressure - test

	t-statistics	p-value
Pooled t-test	2.7557	0.003144
Welsh t-test	2.7557	0.003168

Table 7: Comparison of two hypothesis tests, when assuming/not assuming equal standard deviation.

- Both the pooled t-test and the Welsh t-test give roughly the same results.
- Reject the null hypothesis, indicating that there is strong evidence of a significant larger impact from selling orders in the market.

Own-price Market Impact

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$\mu_{C_{p^{(n)} \leftarrow as\mathbf{1}^{(n)}}}$	0.47	0.90	1.28	0.06	0.31	2.95	3.58	2.47	0.26
$\mu_{C_{p^{(n)} \leftarrow as\mathbf{2}^{(n)}}}$	0.34	0.31	0.47	0.13	0.07	1.42	0.38	2.17	1.43
$\mu_{C_{p^{(n)} \leftarrow as\mathbf{3}^{(n)}}}$	2.57	0.26	1.30	0.25	1.69	0.90	5.26	0.74	0.32
$\sum \mu_{C_{p^{(n)} \leftarrow as^{(n)}}}$	3.38	1.47	3.05	0.44	2.07	5.27	9.22	5.38	2.01
$\mu_{C_{p^{(n)} \leftarrow bs\mathbf{1}^{(n)}}}$	0.09	0.59	0.14	0.05	0.60	1.14	1.12	2.47	0.70
$\mu_{C_{p^{(n)} \leftarrow bs\mathbf{2}^{(n)}}}$	1.35	0.18	0.29	0.05	0.43	0.08	0.83	0.89	0.42
$\mu_{C_{p^{(n)} \leftarrow bs\mathbf{3}^{(n)}}}$	2.58	0.07	0.23	1.20	0.11	2.16	1.37	2.14	1.42
$\sum \mu_{C_{p^{(n)} \leftarrow bs^{(n)}}}$	4.02	0.84	0.66	1.30	1.14	3.38	3.32	5.50	2.54

Table 8: The mean of own-price market impacts caused by market orders $\{as1^{(n)}, bs1^{(n)}\}$, limit orders $\{as2^{(n)}, bs2^{(n)}\}$ and $\{as3^{(n)}, bs3^{(n)}\}$ for each stock n. All numbers are multiplied by 100.

Cross-price Market Impact

- We measure the cross-price market impacts by adding up the impacts from all ask/bid orders for each stock.
- Denote the graph as $\mathcal{G}_{cross} = (\mathcal{V}_c, \mathcal{E}_c)$, with cross-price market impacts from the aggregated size factors to the price factor given by,

$$
\begin{aligned}
\mathcal{V}_c &= \left(p^{(m)}, \sum_r bs_r^{(n)}, \sum_r as_r^{(n)}\right) \\
\mathcal{E}_c &= C_{i\leftarrow j} \quad i \in \{p^{(m)}\}, \quad j \in \left\{\sum_r bs_r^{(n)}, \sum_r as_r^{(n)}\right\} \\
m, n &\in \{1, \dots, N\} \quad r = 1, 2, 3 \quad m \neq n
\end{aligned}
$$

Cross-price Market Impact - Ask Side

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$\mu^{C}_{p(MSFT)\leftarrow\sum_r as_r^{(n)}}$	3.38	0.68	1.82	0.65	1.76	0.73	5.46	0.54	3.66
$\mu^{C}_{p(T)\leftarrow\sum_r as_r^{(n)}}$	3.08	1.47	1.00	0.62	1.86	2.58	1.08	1.29	2.38
$\mu^{C}_{p(IBM)\leftarrow\sum_r as_r^{(n)}}$	1.52	0.38	3.06	1.33	0.91	1.57	1.41	3.58	2.05
$\mu^{C}_{p(JNJ)\leftarrow\sum_r as_r^{(n)}}$	1.69	0.62	1.04	0.45	1.47	1.05	1.37	0.31	3.49
$\mu^{C}_{p(PFE)\leftarrow\sum_r as_r^{(n)}}$	1.07	0.96	0.44	0.13	2.06	4.83	1.86	2.89	2.12
$\mu^{C}_{p(MRK)\leftarrow\sum_r as_r^{(n)}}$	3.18	1.17	0.43	0.83	2.44	5.27	4.15	2.25	2.57
$\mu^{C}_{p(JPM)\leftarrow\sum_r as_r^{(n)}}$	2.09	1.10	1.81	0.72	2.68	1.13	9.22	1.34	2.16
$\mu^{C}_{p(WFC)\leftarrow\sum_r as_r^{(n)}}$	1.22	2.38	1.70	0.55	1.93	1.22	0.79	5.37	0.60
$\mu^{C}_{p(C)\leftarrow\sum_r as_r^{(n)}}$	2.55	1.11	2.37	0.84	2.57	1.33	4.51	1.23	2.01

Table 9: The mean of the market impacts caused by ask orders of stock m for each stock n. All numbers are multiplied by 100.

- The stock price can be affected not only by their own ask order flows, but also by the ask order flows of financial stocks

Cross-price Market Impact - Bid Side

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$\mu C_{p(MSFT) \leftarrow \sum_r bs_r^{(n)}}$	4.02	2.26	0.62	0.53	0.59	1.67	0.41	0.89	1.61
$\mu C_{p(T) \leftarrow \sum_r bs_r^{(n)}}$	1.36	0.84	0.22	1.04	1.41	3.67	0.92	1.10	1.03
$\mu C_{p(IBM) \leftarrow \sum_r bs_r^{(n)}}$	0.79	1.29	0.66	0.13	0.58	0.97	3.47	1.85	1.15
$\mu C_{p(JNJ) \leftarrow \sum_r bs_r^{(n)}}$	0.63	0.85	0.30	1.30	0.86	0.99	0.50	1.90	2.59
$\mu C_{p(PFE) \leftarrow \sum_r bs_r^{(n)}}$	2.12	0.36	1.10	0.19	1.13	0.37	1.43	4.08	1.23
$\mu C_{p(MRK) \leftarrow \sum_r bs_r^{(n)}}$	0.72	0.49	0.25	0.25	1.35	3.37	1.59	0.84	1.27
$\mu C_{p(JPM) \leftarrow \sum_r bs_r^{(n)}}$	1.66	0.47	1.25	0.97	1.39	0.59	3.32	1.87	2.16
$\mu C_{p(WFC) \leftarrow \sum_r bs_r^{(n)}}$	1.99	1.29	0.30	0.73	1.37	0.83	1.75	5.50	1.67
$\mu C_{p(C) \leftarrow \sum_r bs_r^{(n)}}$	1.02	1.80	0.42	1.24	0.84	1.08	1.41	1.12	2.54

Table 10: The mean of the market impacts caused by bid orders of stock m for each stock n. All numbers are multiplied by 100.

- Financial stocks have stronger cross-price market impacts compared with healthcare and technology stocks.

Price and Order Flows

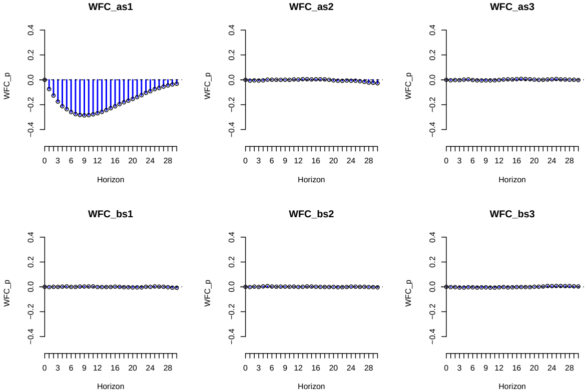
- Even though sufficiently large market order immediately affects the price direction, the bid/ask sizes alone do not give enough information on price direction.
- To measure the evolution of the market/limit order, recall

▸ Return to GI analysis

$$
\begin{aligned}
\delta_{jt} &: (\delta_{1t}, \delta_{2t}, \ldots, \delta_{Kt})^\top \sim e_j \\
GI(l, \delta_{jt}, \mathcal{F}_{t-1}) &= \mathsf{E}(y_{t+l} \mid u_{jt} = \delta_{jt}, \mathcal{F}_{t-1}) - \mathsf{E}(y_{t+l} \mid \mathcal{F}_{t-1})
\end{aligned}
$$

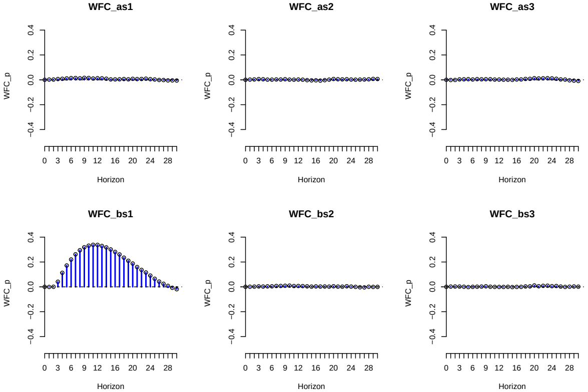
when the shock δ_{jt} is treated as one of the size factors $(\Delta\tilde{s}_t^{a1(n)}, \Delta\tilde{s}_t^{a2(n)}, \Delta\tilde{s}_t^{a3(n)}, \Delta\tilde{s}_t^{b1(n)}, \Delta\tilde{s}_t^{b2(n)}, \Delta\tilde{s}_t^{b3(n)})$ hitting the system. hfhd_marketimpact

Example - Wells Fargo



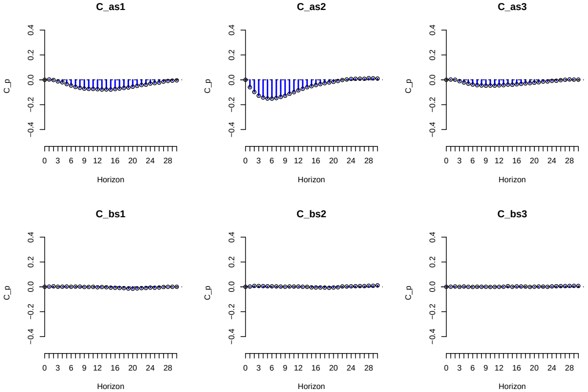
- Own-price market impact on 25.07.2016.
- The investors will start marking down their bid price when there is a wave of sell orders coming into the order book.

Example - Wells Fargo



- On 19.07.2016, the argument also holds for bid market order.
- Both impacts last for almost 10 min before the price shifts back, HF investors have enough reaction time to arbitrage.

Example - Citigroup



- The market impacts of orders posted deeper in the book.
- Positive pile-on effect where larger ask order may further perpetuating a price decrease, last for 20 min. (01.06.0216)

r_{size} Ratio

- For each trading day, we use $\ominus$ and $\oplus$ to represent the significant negative and positive response of price after the arrival of a market/limit order.
- Define a ratio denoted as r_{size} to measure the price direction of market impacts,

$$
\begin{aligned}
r_{size} &= \frac{|\operatorname{sgn}(GI_t)|}{42} \\
\operatorname{sgn}(GI_t) &= \begin{cases} -1 & -GI_t(h) > Q_{0.05}(GI_t(h)) \\ 0 & |GI_t(h)| \leq Q_{0.05}(GI_t(h)) \\ 1 & GI_t(h) > Q_{0.05}(GI_t(h)) \end{cases} \\
t &= 1 \ldots T, \quad h = 1, \ldots, 30
\end{aligned}
$$

Summary

	MSFT	T	IBM	JNJ	PFE	MRK	JPM	WFC	C
$as1$		$\ominus\ \ominus\ \ominus\ \ominus\ \ominus$		$\oplus\ \oplus\ \oplus$		$\oplus\ominus$	$\ominus\ \ominus\ \ominus$	$\ominus\ \ominus\ \ominus$	$\ominus$
$as2$	$\ominus\ \ominus\ \ominus$	$\oplus$	$\oplus\oplus$		$\oplus$			$\ominus\ominus$	$\ominus\ \ominus\ \ominus\ \ominus\ \ominus\ominus$
$as3$	$\oplus\oplus$	$\oplus$	$\oplus$	$\oplus$	$\oplus\ \oplus\ \oplus$		$\ominus$	$\oplus\oplus$	$\ominus$
$bs1$	$\oplus\ \ominus\ \ominus$	$\oplus\ \ominus\ \ominus\ominus$		$\oplus$	$\oplus$	$\oplus\oplus$	$\oplus\ \oplus\ \oplus\ominus$	$\oplus\oplus$	$\oplus$
$bs2$	$\ominus$	$\ominus$			$\oplus\ominus$	$\oplus$	$\oplus$		$\oplus$
$bs3$		$\ominus$	$\ominus\ominus$	$\ominus\ominus$	$\ominus$	$\ominus\ominus$	$\ominus$		
r_{size}	44%	46%	0%	14%	25%	57%	80%	78%	100%

- The group of financial stocks is of higher r_{size} values, finance sector is leading the market, the history information indicates that their response of price to trading volumes is stable and thus robust for statistical arbitrage.
- Healthcare and technology stocks are price-dominated, i.e., they have multiple risk sources except for their own trading activity. This is consistent with our main findings in LOB network analysis.

Conclusion

- LOB network is constructed in a MN context and non-synchronous trading.
- Network relies neither on the ordering of the variables nor on the distribution of the innovations, the resulting connectedness measures is economically interpretable.
- Main findings:
  - Network involving trading volumes is a better measure of the stock connectedness.
  - Significant market impact caused by the arrival of a large limit order is identified.
  - Order imbalance generally exists across stocks
  - Bootstrapped market impacts can be quantified.
  - The financial institutions are connected more closely compared with the firms come from other industry

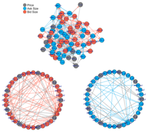
Time-varying Limit Order Book Networks

Shi Chen

Wolfgang Karl Härdle

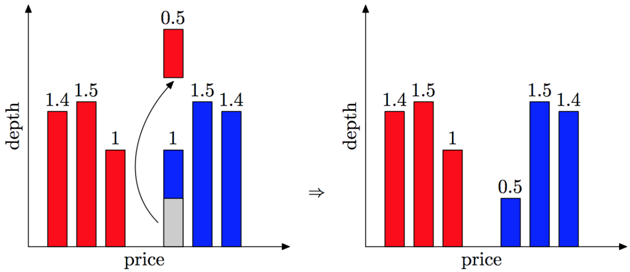
Chong Liang

Melanie Schienle



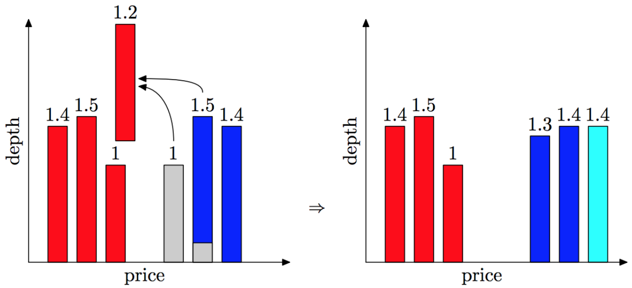
Ladislaus von Bortkiewicz Chair of Statistics
Humboldt–Universität zu Berlin
Chair of Econometrics and Statistics
Karlsruher Institut für Technologie
http://lvb.wiwi.hu-berlin.de

Normal Order Book



- An incoming bid (buy) market order with price 1002 and size 0.5 which results in a buy transaction.

Aggressive Order Book



- An incoming bid (buy) market order with price 1003 and size 1.2 "walking up" the order book.
- Passive limit order: arrival of a buy limit order with price 999 and size 0.5 to be posted behind the market. Back

Setting

- To construct the estimator, choose a sequence k_n of integers that satisfy,

$$k_n\sqrt{\Delta_n} = \theta + \mathcal{O}\left(\Delta_n^{\frac{1}{4}}\right), \text{for some } \theta > 0$$

  - TRTS (transaction time sampling): i indexes time points associated with Δ_n-th trade.
  - CTS (calendar time sampling): i indexes equal-spaced time intervals of length Δ_t with $n = \dfrac{t}{\Delta_n}$.
- Consider a continuous function $g : [0,1] \to \mathbb{R}$ which is piecewise continuously differentiable such that,
  - $g^{'}$ is a piecewise Lipschitz derivative
  - $g(0) = g(1) = 0$ and $\int_0^1 g^2(s)ds > 0$

Setting - ctd

- Choose the simple form of $g(x) = x \wedge (1-x)$, Podolskij et al. (2009), Christensen et al. (2010).

$$\psi_1 = 1, \quad \psi_2 = \frac{1}{12}, \quad \Phi_{11} = \frac{1}{6}, \quad \Phi_{12} = \frac{1}{96}, \quad \Phi_{22} = \frac{151}{80640}$$

- g is associated with the following real-valued numbers,

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

Back

Bootstrap-based multistep forecast method

The computation steps to yield bootstrapped GFEVD, Lanne and Nyberg (2016); Terasvirta et al. (2010),

1. $\mathcal{F}_{t-1}$ information prior to Y_t; forecast horizon h.
2. Randomly sample B vectors of shocks $(\delta_{1t}, \delta_{2t}, \dots, \delta_{Kt})^\top$ from the residuals,

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

$u^\star_{jt}$ coms from $\hat{u}_{jt}$,

$$\begin{aligned} \hat{u}_{jt} &= Y_t - \left(\hat{A}_1, \hat{A}_2, \dots, \hat{A}_p\right) \left(Y_{t-1}^\top, Y_{t-2}^\top, \dots, Y_{t-p}^\top\right)^\top \\ &= Y_t - g(Y_{t-1}) \end{aligned}$$

Bootstrap-based multistep forecast method

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

$$
\begin{aligned}
f_{t,0} &= g(Y_{t-1}) \\
f_{t,1} &= \mathsf{E}^{\star}[Y_{t+1}|\mathcal{F}_{t-1}] = \mathsf{E}^{\star}[g(f_{t,0} + u^{\star}_{t})|\mathcal{F}_{t-1}] \\
f_{t,2} &= \mathsf{E}^{\star}[Y_{t+2}|\mathcal{F}_{t-1}] = \mathsf{E}^{\star}[g(f_{t,1} + u^{\star}_{t+1})|\mathcal{F}_{t-1}] \\
&\ldots
\end{aligned}
$$

with $u^{\star}_{t+l}, l = 1, \ldots, h$ Bootstrap sample from residuals $\{\hat{u}_{t+l}\}_{t=1}^{T}$ over the sample period.

Bootstrap-based multistep forecast method

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

$$
\begin{aligned}
fb_{t,1} &= \frac{1}{B}\sum_{b=1}^{B} g(f_{t,0} + u_t^{\star(b)}) \\
fb_{t,2} &= \frac{1}{B}\sum_{b=1}^{B} g(g(f_{t,0} + u_t^{\star(b)}) + u_{t+1}^{\star(b)}) \\
&\dots
\end{aligned}
$$

Bootstrap-based multistep forecast method

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

$$
\begin{aligned}
f_{t,0} &= g(Y_{t-1}) \\
f_{t,1} &= \mathsf{E}^{\star}[Y_{t+1} \mid \mathcal{F}_{t-1}] = \mathsf{E}^{\star}[g(f_{t,0} + u^{\star}_{jt} e_j) \mid \mathcal{F}_{t-1}] \\
f_{t,2} &= \mathsf{E}^{\star}[Y_{t+2} \mid \mathcal{F}_{t-1}] = \mathsf{E}^{\star}[g(f_{t,1} + u^{\star}_{j,t+1} e_j) \mid \mathcal{F}_{t-1}] \\
&\cdots
\end{aligned}
$$

with $u^{\star}_{j,t+l}, l = 1, \ldots, h$ Bootstrap sample from residuals $\{\hat{u}_{j,t+l}\}_{t=1}^{T}$ over the sample period.

Bootstrap-based multistep forecast method

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

$$
\begin{aligned}
fb_{t,1} &= \frac{1}{B}\sum_{b=1}^{B} g(f_{t,0} + u_{jt}^{\star(b)} e_j) \\
fb_{t,2} &= \frac{1}{B}\sum_{b=1}^{B} g\{g(f_{t,0} + u_{jt}^{\star(b)} e_j) + u_{j,t+1}^{\star(b)} e_j\} \\
&\dots
\end{aligned}
$$

Bootstrap-based multistep forecast method

7. Plug in the GI function

$$\widehat{GI}(l, \delta_{jt}, \mathcal{F}_{t-1}) = \mathsf{E}^{\star}(y_{t+l}|u_{jt} = \delta_{jt}, \mathcal{F}_{t-1}) - \mathsf{E}^{\star}(y_{t+l}|\mathcal{F}_{t-1})$$

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

$$\hat{\lambda}_{ij,\mathcal{F}_{t-1}}(h) = \frac{\sum_{l=0}^{h} \widehat{GI}(l, \delta_{jt}, \mathcal{F}_{t-1})_i^2}{\sum_{j=1}^{K} \sum_{l=0}^{h} \widehat{GI}(l, \delta_{jt}, \mathcal{F}_{t-1})_i^2}, \quad i, j = 1, \ldots, K$$

8. Repeat steps 2-6 for all histories.
9. Construct connectedness table using averaged $\hat{\lambda}_{ij,\mathcal{F}_{t-1}}(h)$ generated from step 7. Return to LN-GFEVD

References

Belloni, A., Chernozhukov, V., et al.
Least squares after model selection in high dimensional sparse models
Bernoulli, 19(2):521177, 2013.

Bühlmann, Peter, and Sara Van De Geer.
Statistics for high-dimensional data: methods, theory and applications
Springer Science & Business Media, 2011.

References

Chen, Xiaohui, Mengyu Xu, and Wei Biao Wu
Covariance and precision matrix estimation for high-dimensional time series.
The Annals of Statistics 41.6 (2013): 2994-3021.

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

References

N Hautsch, M Podolskij
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.

Lanne, M. and Nyberg, H.
Generalized forecast error variance decomposition for linear and nonlinear multivariate models.
Oxford Bulletin of Economics and Statistics, 78(4):595173.