Modelling and Forecasting Liquidity Supply Using Semiparametric Factor Dynamics

Wolfgang Karl Härdle
Nikolaus Hautsch
Andrija Mihoci

Ladislaus von Bortkiewicz Chair of Statistics
Chair of Econometrics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://case.hu-berlin.de

Snapshot of a Limit Order Book (LOB)

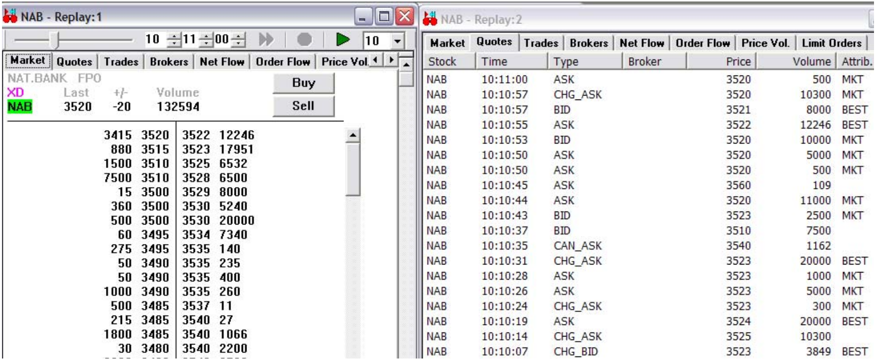


Figure 1: Snapshot of a LOB for National Australia Bank Ltd. (NAB)

LOB - Graphical Illustration

Figure 2: Limit order book for NAB on July 8, 2002

Objectives

- Modelling the LOB spatial and time structure using a dynamic factor model
  - Estimating and predicting factors and factor loadings
  - Understanding the dynamics of factor loadings
  - Impact of explanatory variables capturing the state of the market
- Forecasting demand and supply curves $\rightarrow$ liquidity supply
  - Extensive rolling window out-of-sample forecasting exercise
  - Forecasting evaluation against naive benchmark
  - Financial and economic applications

Statistical Challenges

- Require flexible framework for modelling and forecasting high-dimensional time-varying phenomenon
- Dimension reduction: extraction of common factors
- No obvious parametric model for factors
- Modelling philosophy: *smooth in space and parametric in time*
- Capturing dynamics by parametric multivariate TS model for factor loadings

Economic Implications

- LOB reflects liquidity supply on both sides of the market
- Information content: LOB reflects market's expectation
- Shape of order book curves drives instantaneous trading costs for given volumes
- Predicting transaction costs yields implications for splitting strategies: transaction costs vs. liquidity risks

Applications

Example: Trading Strategy

An investor decides to buy (sell) certain number of NAB shares (10,000 or 20,000) over the course of a trading day, starting from 10:30 until 15:55.

Which execution strategy should the investor follow:

(i) Splitting the buy (sell) order proportionally over the trading day (i.e. every 5 minutes)

(ii) Placing one buy (sell) order at a time where the predicted transaction costs using the DSFM approach are minimal?

Research Questions

- Does the DSFM successfully model liquidity supply?
- Do factor loadings and quote dynamics follow a vector error correction (VEC) specification?
- Does our method outperforms a naive forecasting benchmark and improves order execution strategies?

Outline

1. Motivation ✓
2. Limit Order Book Data
3. The Dynamic Semiparametric Factor Model (DSFM)
4. Modelling LOB Dynamics
5. Forecasting LOB Dynamics
6. Conclusions

The Data

- Limit order data from the Australian Stock Exchange (ASX)
  - Allows for complete reconstruction of the LOB at any time
  - Accounting for all LOB activities outside continuous trading

- Analyzing 4 stocks
  - Broken Hill Proprietary Ltd. (BHP)
  - National Australia Bank Ltd. (NAB)
  - MIM
  - Woolworths (WOW)

The Data

- Australian Stock Exchange (ASX)
  - Period covered: July 8 - August 16, 2002 (30 trading days)
  - Daily trading period: 10:15 - 15:55
  - LOB sampling frequency: 5 minutes

Descriptive Statistics

Orders	BHP	NAB	MIM	WOW
Limit orders				
(i) buy (bid side)	50012	28850	9551	13234
(ii) sell (ask side)	32053	25953	6474	11318
Market orders				
(i) buy	28030	16304	4115	7260
(ii) sell	16755	15142	2789	6464

Table 1: Number of orders from July 8 to August 16, 2002

Notation and Data Preprocessing

- Seasonally adjusted bid/ask side volume
$$Y^b_{t,j} = \widetilde{Y}^b_{t,j}/s^b_t \in \mathbb{R}^{101} \text{ and } Y^a_{t,j} = \widetilde{Y}^a_{t,j}/s^a_t \in \mathbb{R}^{101}$$
- Best bid/ask price: $\widetilde{S}^b_{t,101}$, $\widetilde{S}^a_{t,1}$
- Relative price deviations from best bid/ask quotes: $S^b_{t,j}$, $S^a_{t,j}$
- Capturing intraday seasonality using FFF approximation

$$s_t = \delta \cdot \bar{t} + \sum_{m=1}^{M} \left\{ \delta_{c,m} \cos\left(\bar{t} \cdot 2\pi m\right) + \delta_{s,m} \sin\left(\bar{t} \cdot 2\pi m\right) \right\},$$

where $\bar{t} \in (0, 1]$.

Intraday Seasonalities in Liquidity Supply

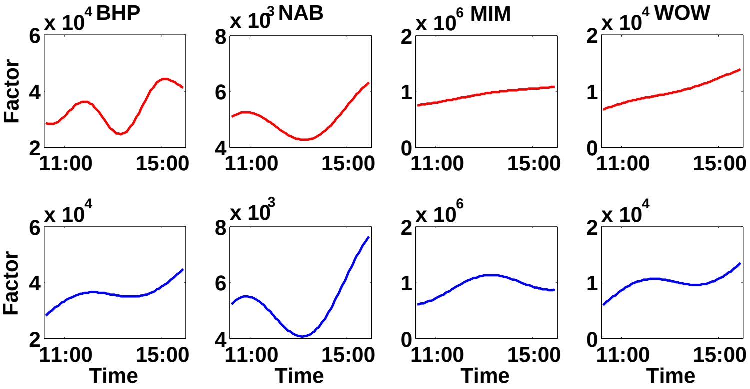


Figure 3: Seasonal factors for quantities at $\widetilde{S}^b_{t,101}$ (red) and $\widetilde{S}^a_{t,1}$ (blue)

The Dynamic Semiparametric Factor Model

- Orthogonal L-factor model of an observable J-dimensional random vector - Park et al. (2009), Fengler et al. (2007)

$$Y_{t,j} = m_{0,j} + Z_{t,1} m_{1,j} + \cdots + Z_{t,L} m_{L,j} + \varepsilon_{t,j}$$

$m(\cdot) = (m_0, m_1, \ldots, m_L)^\top$ - tuple of functions
$m_l : \mathbb{R}^d \to \mathbb{R}$ - time-invariant factors
$Z_t = (1_T, Z_{t,1}, \ldots, Z_{t,L})^\top$ - factor loadings

- Including explanatory variables $X_{t,j}$

$$Y_{t,j} = \sum_{l=0}^{L} Z_{t,l} m_l (X_{t,j}) + \varepsilon_{t,j} = Z_t^\top m(X_{t,j}) + \varepsilon_{t,j}$$

Estimation

- Efficient nonparametric method

$$Z_t^\top m(X) = \sum_{l=0}^{L} Z_{t,l} m_l(X) = \sum_{l=0}^{L} Z_{t,l} \sum_{k=1}^{K} a_{l,k} \psi_k(X) = Z_t^\top A \psi(X)$$

$\psi(\cdot) = (\psi_1, \ldots, \psi_K)^\top$ - basis functions (tensor B-spline basis)
$A = (a_{l,k}) \in \mathbb{R}^{(L+1)\times K}$ - coefficient matrix

$$\left(\widehat{Z}_t, \widehat{A}\right) = \arg\min_{Z_t, A} \sum_{t=1}^{T} \sum_{j=1}^{J} \{Y_{t,j} - Z_t^\top A \psi(X_{t,j})\}^2$$

- Minimization by Newton-Raphson algorithm

Implementation

Selection of L and K

- Explained variance

$$EV(L) = 1 - \frac{\sum_{t=1}^{T} \sum_{j=1}^{J} \{Y_{t,j} - \sum_{l=0}^{L} \widehat{Z}_{t,l} \widehat{m}_l (X_{t,j})\}^2}{\sum_{t=1}^{T} \sum_{j=1}^{J} \{Y_{t,j} - \bar{Y}\}^2}$$

Statistical Inference

- Difference between $\widehat{Z}_t$ and Z_t can be asymptotically neglected
- TS models can be used for modelling of $\widehat{Z}_t$

Modelling Liquidity Supply

- DSFM approaches
  - "Separated" approach - demand and supply separately, i.e. $Y^b_{t,j} \in \mathbb{R}^{101}$ and $Y^a_{t,j} \in \mathbb{R}^{101}$
  - "Combined" approach - whole LOB, $\left(-Y^b_{t,j}\,,\ Y^a_{t,j}\right) \in \mathbb{R}^{202}$

- Explanatory variables, $X_{t,j}$
  - Relative price levels, $S^b_{t,j}$ and $S^a_{t,j}$
  - Deseasonalized lagged 5 min buy/sell volume, Q^b_t and Q^s_t
  - Lagged 5 min log return and realized volatility, r_t and $V_t = r^2_t$

LOB Based on Relative Price Levels - Explained Variance

Approach	BHP	NAB	MIM	WOW
Bid side				
(i) Separated	0.964	0.965	0.996	0.975
(ii) Combined	0.921	0.936	0.975	0.914
Ask side				
(i) Separated	0.941	0.948	0.953	0.959
(ii) Combined	0.930	0.912	0.951	0.948

Table 2: EV of the estimated LOB data from July 8 to August 16, 2002

LOB and Relative Price Levels

Figure 4: True (solid) and estimated (dashed) LOB using the separated approach with two factors (EV$\approx$ 95%) on July 8, 2002

Estimated LOB Factors

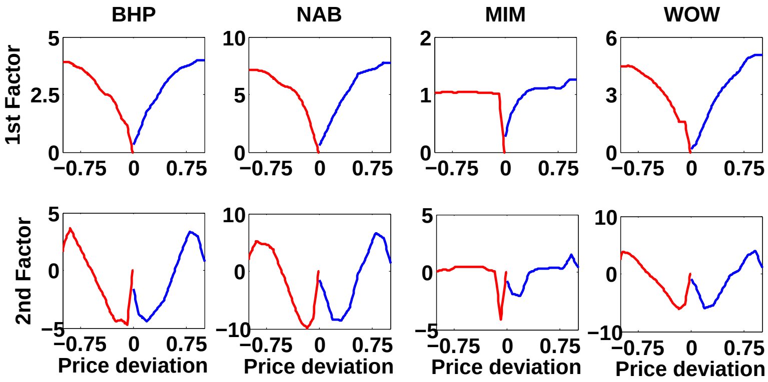


Figure 5: Estimated factors vs. relative price levels

Estimated Factor Loadings

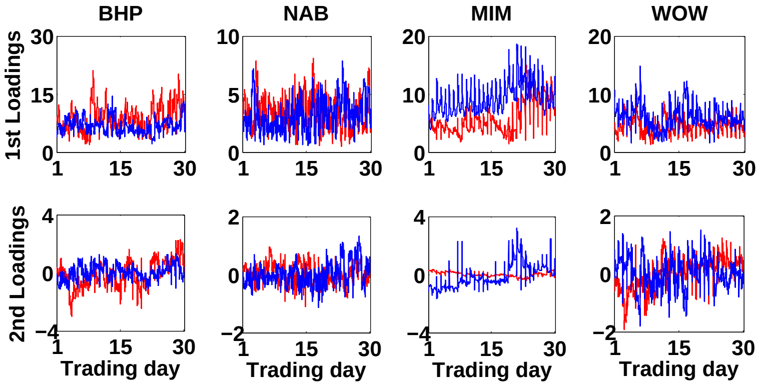


Figure 6: Estimated factor loadings vs. relative price levels

Vector Error Correction (VEC) Specification

- Engle and Patton (2004), Hautsch and Huang (2011)

$$z_t = \left(\widehat{Z}^b_{1,t}, \widehat{Z}^b_{2,t}, \widehat{Z}^a_{1,t}, \widehat{Z}^a_{2,t}, \Delta \log \widetilde{S}^b_{t,101}, \Delta \log \widetilde{S}^a_{t,1} \right)^\top$$

$\Delta \log \widetilde{S}^b_{t,101}$ and $\Delta \log \widetilde{S}^a_{t,1}$ - best bid and ask price return

$$z_t = c + \Gamma_1 z_{t-1} + \ldots + \Gamma_q z_{t-q} + \gamma \left(\log \widetilde{S}^b_{t-1,101} - \log \widetilde{S}^a_{t-1,1} \right) + \varepsilon_t$$

- Findings: BHP and WOW ($q = 3$), NAB ($q = 2$), MIM ($q = 4$)
  - Strong own-process dynamics, weak cross-dependencies
  - Quote changes are short-run predictable (up to 10-15 minutes)

Drivers of the Order Book Shape

Variable	BHP	NAB	MIM	WOW
Bid side				
Q_t^s	10.37	8.17	5.41	6.31
Q_t^b	10.42	8.41	4.37	6.29
r_t	21.93	23.09	39.47	175.40
V_t	95.74	87.12	258.37	-

Table 3: RMSE of the estimated LOB data from July 8, 2002 to August 16, 2002

Drivers of the Order Book Shape

Variable	BHP	NAB	MIM	WOW
Ask side				
Q_t^s	7.38	8.30	5.72	9.18
Q_t^b	7.30	8.42	7.22	8.88
r_t	18.00	22.13	45.54	236.08
V_t	78.62	63.63	192.87	-

Table 4: RMSE of the estimated LOB data from July 8, 2002 to August 16, 2002

Drivers of the Order Book Shape

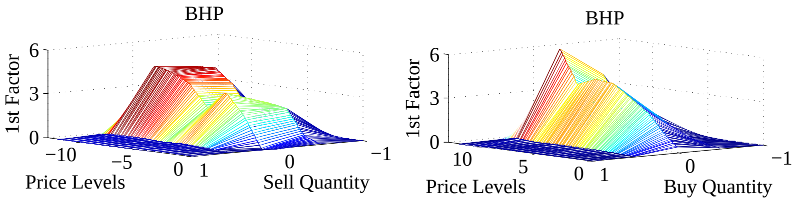


Figure 7: Estimated first factor of the bid (left) and the ask (right) side with respect to relative price levels and the past log traded sell (left) and buy (right) volume using the DSFM-Separated approach with two factors from 20020708 to 20020816

Impulse-Response Analysis

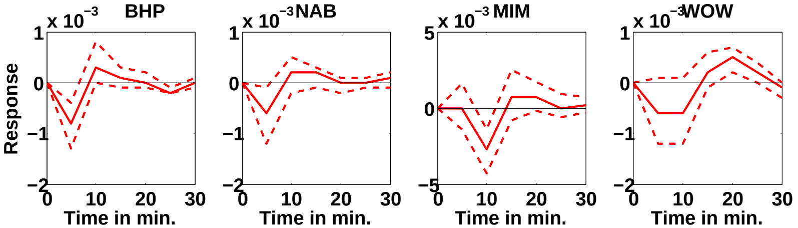


Figure 8: Responses of the best bid quote return to a one standard deviation shock in the estimated first bid factor loadings from July 8 to August 16, 2002. We employ the DSFM-separated approach with two factors. The response variable always enters the VEC specification in the first position. 95% confidence intervals are shown with dashed lines.

Impulse-Response Analysis

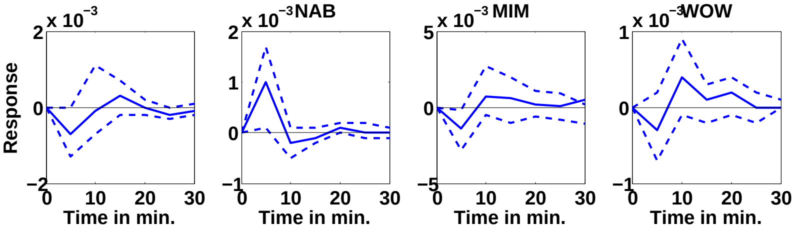


Figure 9: Responses of the best ask quote return to a one standard deviation shock in the estimated first ask factor loadings from July 8 to August 16, 2002. We employ the DSFM-separated approach with two factors. The response variable always enters the VEC specification in the first position. 95% confidence intervals are shown with dashed lines.

Forecasting Liquidity Supply

Setup

- 4 stocks, forecasting period: 20020722 - 20020816 (20 days)
- Forecasts for all 5 minute intervals until the end of a day

Strategies

- "DSFM-Separated" approach - estimated factors and factor loadings every 5 minutes (10 trading days)
- "Naive" approach - last observed LOB curve

LOB - Forecasting

Figure 10: True (solid) and forecasted LOB using the "DSFM-Separated" (dashed) and the "Naive" approach (black) on July 22, 2002

Root Mean Squared Prediction Errors

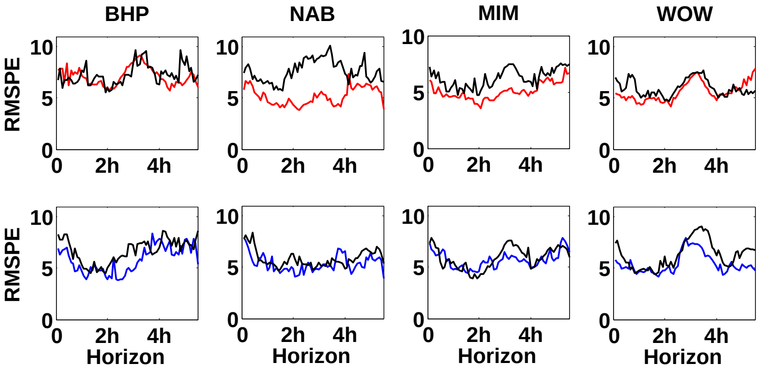


Figure 11: RMSPEs using DSFM (red and blue) and naive forecast (black) for all intervals during the day

Applications

Example: Trading Strategy

An investor decides to buy (sell) certain number of shares (5 or 10 times the average best bid/ask volume) over the course of a trading day, starting from 10:30 until 15:55.

Which execution strategy should the investor follow:

(i) Splitting the buy (sell) order proportionally over the trading day (i.e. every 5 minutes)

(ii) Placing one buy (sell) order at a time where the predicted transaction costs using the DSFM approach are minimal?

Trading Strategy

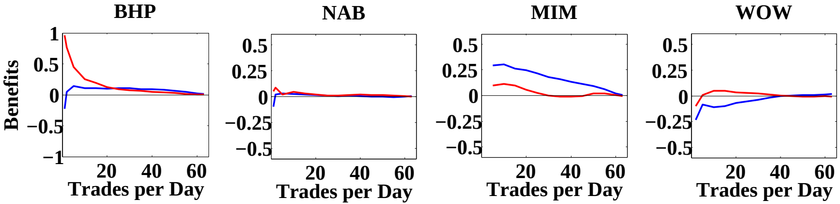


Figure 12: Average percentage DSFM gains by reduced transaction costs compared to an equal-splitting strategy when buying and selling shares. Daily volumes: BHP (175,000), NAB (25,000), MIM (1,860,000) and WOW (50,000). Period covered: 20020722-20020816.

Trading Strategy

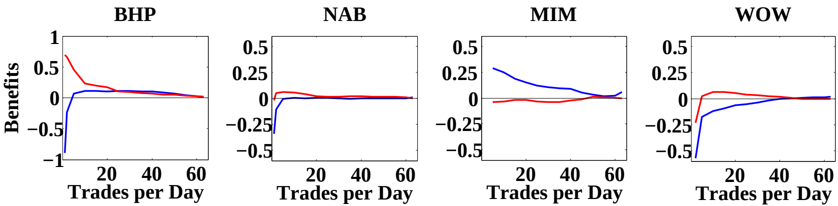


Figure 13: Average percentage DSFM gains by reduced transaction costs compared to an equal-splitting strategy when buying and selling shares. Daily volumes: BHP (350,000), NAB (50,000), MIM (4,650,000) and WOW (100,000). Period covered: 20020722-20020816.

Conclusions

(i) Modelling LOB Dynamics

- Two factors are sufficient to model LOB dynamics (slope, curvature) - explained variance approximately 95%
- Estimated factor loadings and quote dynamics follow a VEC specification
- Strong own-process dynamics, weak cross-dependencies
- The shape of the order book curves depends stronger on past trading volume than on past price movements or past volatility

Conclusions

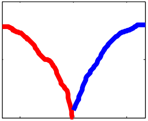
(ii) Forecasting LOB Dynamics

- The DSFM approach outperforms a naive benchmark and a proportional trading strategy
- Quote changes are short-run predictable (up to 10-15 minutes)
- Applications: improved order execution strategies
- Demand and supply curves are modelled and forecasted successfully

Modelling and Forecasting Liquidity Supply Using Semiparametric Factor Dynamics

Wolfgang Karl Härdle
Nikolaus Hautsch
Andrija Mihoci

Ladislaus von Bortkiewicz Chair of Statistics
Chair of Econometrics
C.A.S.E. – Center for Applied Statistics and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://case.hu-berlin.de



References

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

Fengler, M. R., Härdle, W. and Mammen, E.
A Dynamic Semiparametric Factor Model for Implied Volatility String Dynamics
Journal of Financial Econometrics **5**(2): 189-218, 2007

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

References

Park, B., Mammen, E., Härdle, W. and Borak, S.
Time Series Modelling With Semiparametric Factor Dynamics
Journal of the American Statistical Association **104**(485): 284-298, 2009