

Modelling and Forecasting Liquidity Supply Using Semiparametric Factor Dynamics

Wolfgang Karl Härdle

Nikolaus Hautsch

Andrija Mihoci

Ladislaus von Bortkiewicz Chair of
Statistics

CASE - Center for Applied Statistics
and Economics

Humboldt-Universität zu Berlin

<http://ise.wiwi.hu-berlin.de>



Snapshot of a Limit Order Book (LOB)

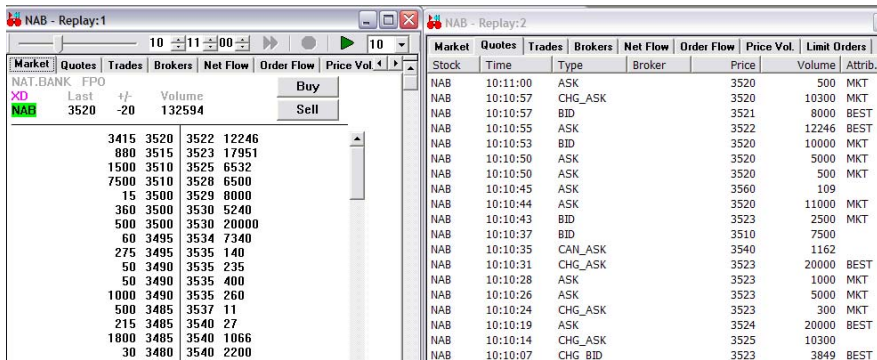


Figure 1: Snapshot of a LOB for National Australia Bank Ltd. (NAB)



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 factor dynamics
 - ▶ Impact of explanatory variables capturing the state of the market
- Forecasting demand and supply curves → liquidity supply
 - ▶ Extensive rolling window out-of-sample forecasting exercise
 - ▶ Forecasting evaluation against naive benchmark



Statistical Challenges

- ▣ Require flexible framework for modelling and forecasting high-dimensional time-varying phenomenon
- ▣ Dimension reduction: extraction of common factors driving the order book
- ▣ No obvious parametric model for factors: nonparametric approach
- ▣ 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 drive instantaneous trading costs for given volumes
- Predicting transaction costs yield implications for splitting strategies: transaction costs vs. liquidity risks



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)
- 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 ask/bid side volume at time t and level j :
 $Y_{t,j}^a = \tilde{Y}_{t,j}^a / s_t^a \in \mathbb{R}^{101}$ and $Y_{t,j}^b = \tilde{Y}_{t,j}^b / s_t^b \in \mathbb{R}^{101}$
- Seasonally adjusted ask side volume at time t and level j :
- Best ask/bid price at t : $\tilde{S}_{t,101}^b, \tilde{S}_{t,1}^a$
- Relative price deviations from best bid/ask quotes: $S_{t,j}^b$ and $S_{t,j}^a$
- Capturing intraday seasonality using FFF approximation:

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

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



Intraday Seasonalities in Liquidity Supply

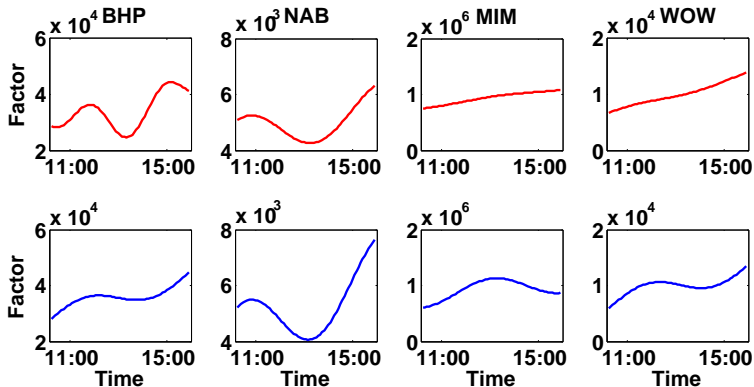


Figure 3: Seasonal factors for quantities at $\tilde{S}_{t,101}^b$ (red) and $\tilde{S}_{t,1}^a$ (blue)



The Dynamic Semiparametric Factor Model

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

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

$m(\cdot) = (m_0, m_1, \dots, m_L)^\top$ - tuple of functions

$m_l : \mathbb{R}^d \rightarrow \mathbb{R}$ - time-invariant factors

$Z_t = (1, Z_{t,1}, \dots, Z_{t,L})^\top$ - 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

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

$\psi(\cdot) = (\psi_1, \dots, \psi_K)^\top$ - basis functions (tensor B-spline basis)

$A = (a_{l,k}) \in \mathbb{R}^{(L+1) \times K}$ - coefficient matrix

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

- Minimization by Newton-Raphson algorithm



Implementation

Selection of L and K

- Explained variance:

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

Statistical Inference

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



Modelling Liquidity Supply

□ DSFM approaches:

- ▶ "Separated" approach - demand and supply separately, i.e.
 $Y_{t,j}^b \in \mathbb{R}^{101}$ and $Y_{t,j}^a \in \mathbb{R}^{101}$
- ▶ "Combined" approach - whole LOB, $(-Y_{t,j}^b, Y_{t,j}^a) \in \mathbb{R}^{202}$

□ Explanatory variables, $X_{t,j}$:

- ▶ Relative price levels, $S_{t,j}^b$ and $S_{t,j}^a$
- ▶ Deseasonalized lagged 5 min buy/sell volume, Q_t^b and Q_t^s
- ▶ Lagged 5 min log return r_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 - Estimation

Figure 4: True (solid) and estimated (dashed) LOB on July 8, 2002



Estimated LOB Factors

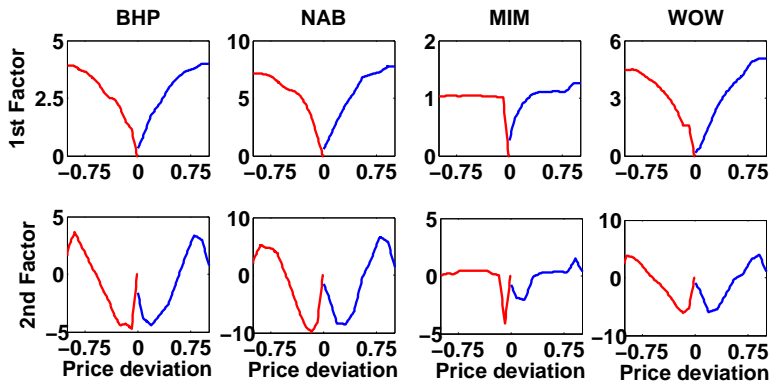


Figure 5: Estimated factors vs. relative price levels



Estimated Factor Loadings

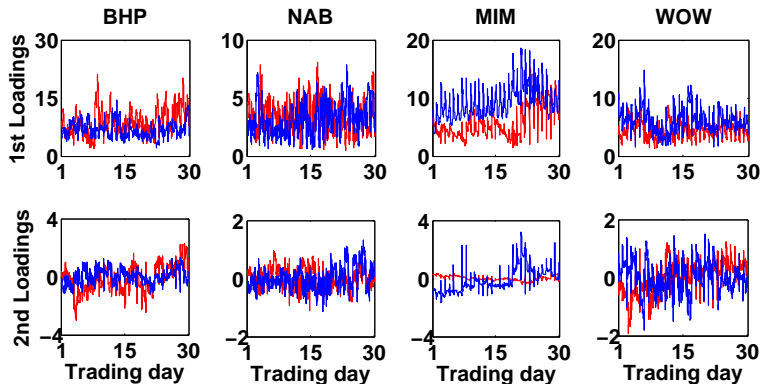


Figure 6: Estimated factor loadings vs. relative price levels



Properties of Estimated Factor Loadings

- Unit root hypothesis is rejected
- Stationarity not rejected
- No evidence for cointegration
- VAR(p) model for $\hat{Z}_t \in \mathbb{R}^4$:

$$\hat{Z}_t = c + B_1 \hat{Z}_{t-1} + \dots B_p \hat{Z}_{t-p} + \varepsilon_t$$

- Model selection based on BIC
- BHP and MIM - VAR(4), NAB - VAR(2), WOW - VAR(3)



Estimation Results for VAR(p) Models

- VAR(p) specifications provide reasonable fits
- High own-dynamic persistence
- Cross-correlation dependency is more pronounced for illiquid stocks than for liquid ones
- Preferred VAR(p) specification used for forecasting



LOB and Economic Variables

Approach	BHP	NAB	MIM	WOW
Bid side				
(i) Traded buy quantity	18.45	159.52	22.01	15.16
(ii) Traded sell quantity	42.95	45.59	193.50	12.95
(iii) Log returns	17.27	12.49	1104.46	9.51
Ask side				
(i) Traded buy quantity	12.09	217.03	122.93	18.72
(ii) Traded sell quantity	37.00	19.97	61.58	17.32
(iii) Log returns	15.76	12.68	7750.16	16.40

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



Forecasting Setup

- Forecasting period: July 22 - August 16, 2002 (20 trading days)
- Rolling windows shifted over 5 minute grids
- Information set used at each 5min interval: last $T = 690$ demand and supply curves (10 trading days)
- LOB forecasts produced for all 5 minute intervals until the end of a trading day
- Forecasting approaches:
 - ▶ "DSFM-Separated" approach - estimated factors and factor loadings every 5 minutes
 - ▶ "Naive" approach - use last observed LOB curve



LOB - Forecasting

Figure 7: True (solid) and forecasted LOB using the "DSFM-Separated" (dashed) and the "Naive" approach (black) on July 22, 2002



Root Mean Squared Prediction Errors

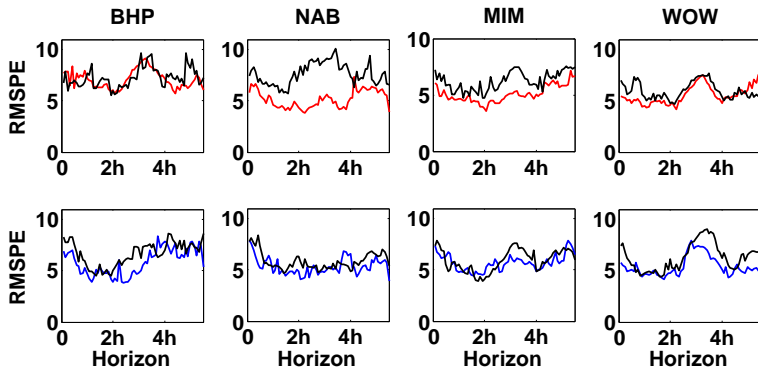


Figure 8: RMSPEs using DSFM (red and blue) and naive forecast (black) for all intervals during the day



Conclusions

- ▣ Two factors are sufficient to model LOB dynamics (slope, curvature)
- ▣ Both estimated factor loadings reject the unit root hypothesis
- ▣ Estimated factor loadings are not stationary processes
- ▣ Demand and supply curves are modelled and forecasted successfully
- ▣ Applications: analysed dynamics and predictability of equity value






Further Steps

- ▣ Studying market elasticities
- ▣ Linking factors and loadings to execution risks and execution probabilities
- ▣ Studying liquidity risks, GARCH, 'default' risks
- ▣ Financial applications



References

-  Fengler, M. R., Härdle, W. and Mammen, E. (2007)
A Dynamic Semiparametric Factor Model for Implied Volatility String Dynamics
Journal of Financial Econometrics 5(2): 189-218
-  Hall, A. D. and Hautsch, N. (2006)
Order aggressiveness and order book dynamics
Empirical Economics 30: 973-1005
-  Hall, A. D. and Hautsch, N. (2007)
Modelling the buy and sell intensity in a limit order book market
Journal of Financial Markets 10(3): 249-286



References



Hautsch, N. (2004)

Modelling Irregularly Spaced Financial Data - Theory and Practice of Dynamic Duration Models

Lecture Notes in Economics and Mathematical Systems,
Springer Verlag, Berlin



Park, B. and Mammen, E. and Härdle, W. and Borak, S.
(2009)

Time Series Modelling With Semiparametric Factor Dynamics
Journal of the American Statistical Association **104**(485):
284-298

