We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Modelling Limit Order Book Volume Covariance Structures

Andrija Mihoci

Additional information is available at the end of the chapter

Abstract

Limit order volume data have been here analysed using key multivariate techniques: principal components, factor and discriminant analysis. The focus lies on understanding of the covariance structure of posted quantities of the asset to be potentially sold or bought at the market. Employing the methods to data of 20 blue chip companies traded at the NASDAQ stock market in June 2016, one observes that two principal components account for approximately 85–95% of order book variation. The most important factor related to order book data variation has furthermore been the demand side (variability). The order book data variation, moreover, successfully classifies stock price movements. Potential applications include improving order execution strategies, designing trading algorithms and understanding price formation.

Keywords: limit order book, multivariate techniques, principal components analysis, factor analysis, discriminant analysis

1. Introduction

The limit order book (LOB) trading mechanism became the dominant way to trade assets on financial markets. Since the limit order book represents liquidity supply of assets on a market, it essentially reflects the demand for as well as the supply of assets above the equilibrium price-volume point. Its variation is affecting the liquidity and price dynamics of an asset, and thus, the goal of this study is to conduct a comprehensive multivariate analysis of the limit order book (variation) data.

Here we model the covariance structures of order book data of several assets by employing key multivariate methods. Theodore W. Anderson synthesized various subareas of the subject

and has influenced the direction of recent and current research in theoretical multivariate analysis [1]. The principal components, factor and discriminant analysis remain quite popular dimension-reduction and classification techniques that are applied in many research fields.

Multivariate techniques are, for example, recently used in financial econometrics of limit order book markets. The principal component analysis is performed in the studies about commonalities in liquidity (measures), see, for example [2, 3], or while analysing price impact data [4]. The dynamics of liquidity supply curves is captured by the so-called dynamic semiparametric factor model in [5], whereas [6] characterize traders' behaviour using discriminant analysis.

Our focus lies on understanding of the variability of posted quantities of the asset, to be potentially sold or bought at the market. The volume (variation) at every order book level is analysed as a random variable, and thus we do not suppress the order book information through, for example, liquidity measures or reward functions. In this chapter, we consider the (full) structure of the covariance matrices. Potential applications thus include improving order execution strategies, understanding price formation and liquidity commonalities, designing trading algorithms.
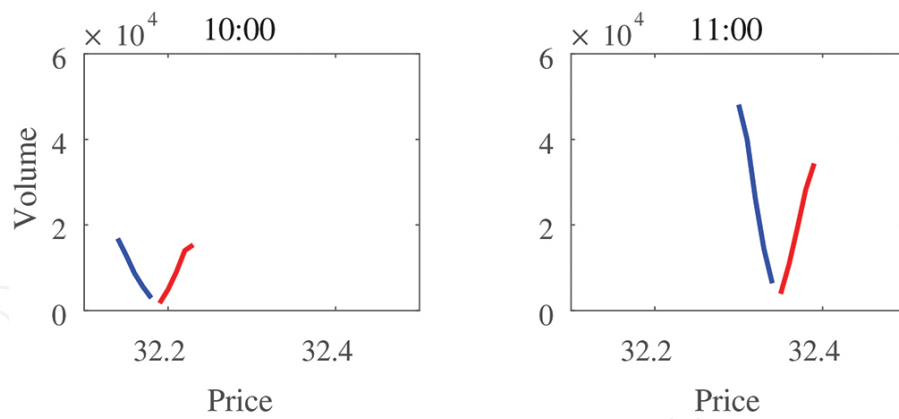
This study is organised as follows: after the limit order book data have been described in Section 2, the statistical methods are presented in Section 3. Empirical results are provided in Section 4, and Section 5 concludes.

2. Limit order book data

The limit order book of an asset lists the volume of pending buying or selling orders at given prices for the asset under consideration and here we analyse its variance-covariance structure. At a fixed time point, the order book essentially represents a snapshot of the asset's demand and supply curves above the market equilibrium quantity level. The volume to be potentially bought forms the asset's demand (bid) side, whereas the volume to be potentially sold depicts the asset's supply (ask) side. To be more precise, the order book bid and ask curves represent liquidity supply, thus quantities above the equilibrium volume level, as orders below the equilibrium (would) have been traded at the market.

2.1. NASDAQ market data and descriptive statistics

At the NASDAQ stock market, one of the world's largest securities exchange, the orders are posted nearly instantaneously and the limit orders are executed in the received order. To visualize a limit order book, consider the data of Intel Corp. (INTC) on 30 June 2016, obtained from the data provider LOBSTER (lobsterdata.com). The number of shares to be potentially bought or sold at different prices at 10:00 and 11:00 are depicted in **Figure 1**. For example, at 10:00 at prices 32.14 (fifth best bid price) and 32.18 (best bid price), there are 16,834 and 2927 stocks demanded, respectively. At the same time, the number of offered shares at prices 32.19 (best ask price) and 32.23 (fifth best ask price) similarly equals 1700 and 15,355, respectively. At 11:00, one furthermore observes that the order book shifted to the direction of higher prices. We attribute this movement to the (observed) increased demand pressure.

Figure 1. Observed limit order book for Intel Corp. (INTC) on 30 June 2016 at 10:00 and 11:00. The monotonically decreasing (increasing) functions represent the demand (supply) curves.

At the NASDAQ order book driven securities exchange, there are several event types that influence the bid and ask curves, namely submissions of new limit orders, cancellations, deletions and executions (lobsterdata.com). Our data set allows us thus to reconstruct all order book activities of a particular company over the course of a trading day. For a description of trading that is common to most limit order book markets, see, for example [7].

The order book volume at given price level represents here a p-dimensional random variable. Denoted by $S = \left(S_1,...,S_p\right)^\top$, $S_1 < ... < S_p$, the price and by $X = \left(X_1,...,X_p\right)^\top$, the associated volume vector. The limit order book of an asset is given by the pairs

$$\left\{S_1, X_1\right\},...,\left\{S_p, X_p\right\}. \tag{1}$$

The expected volume vector is denoted by $\mathsf{E}\left[X\right] = \mu$, and the object of our interest, the limit order book volume variance-covariance matrix by

$$\mathsf{Var}\left(X\right) = \Sigma = \mathsf{E}\left[\left(X - \mathsf{E}\left[X\right]\right)\left(X - \mathsf{E}\left[X\right]\right)^\top\right], \tag{2}$$

here Σ is a symmetric $p \times p$ matrix whose mean diagonal elements depict the variances of the pending volume at fixed price levels $1,...,p$.
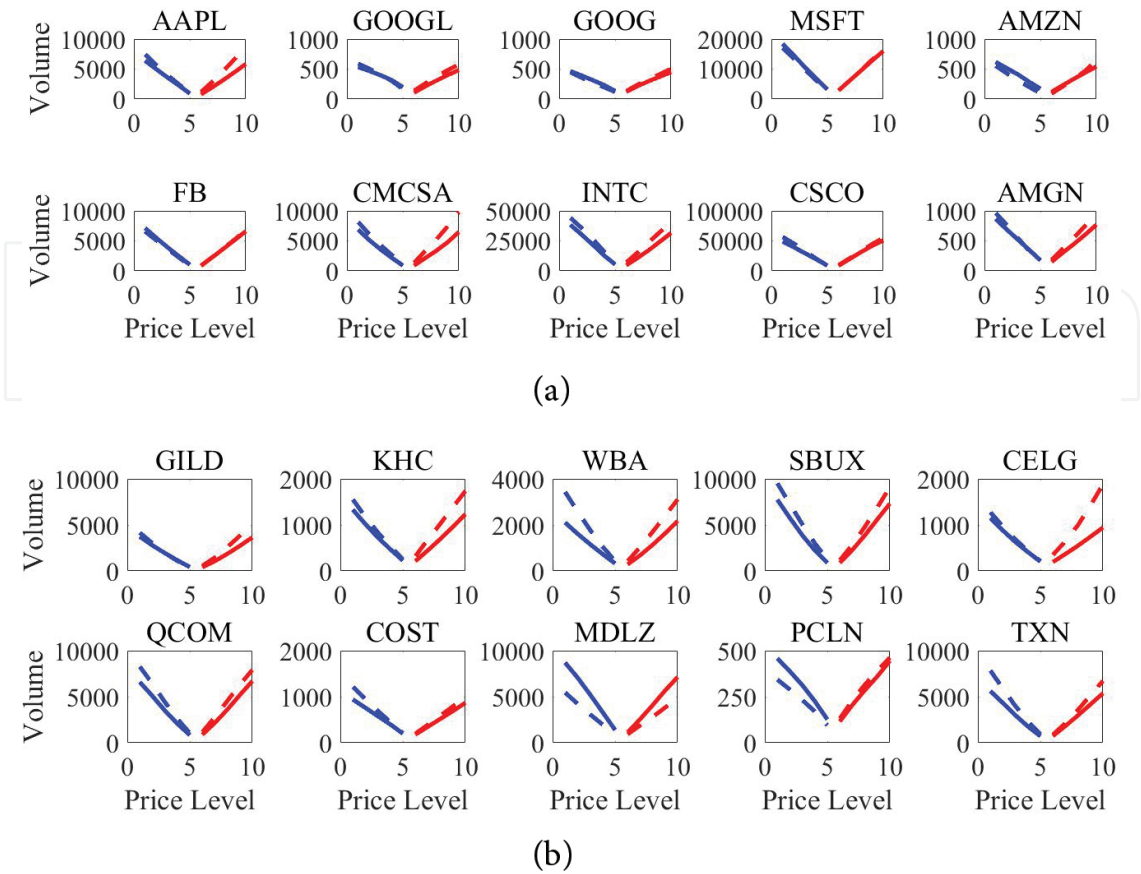
Limit order book data of the 20 largest stocks traded at the NASDAQ stock market have been collected for the purpose of our analysis. In modelling of the high-dimensional covariance structures of this object, we set $p = 10$. The volume at the demand side is thus represented by the variables $X_1,...,X_5$ and the variables $X_6,...,X_{10}$ form the supply regime. Since the "Brexit" referendum results had a significant influence on the stock market movements, we correspondingly focus on the order book activities on 27 June 2016 (S&P 500 at lowest level after the voting) and 30 June 2016 (upward movement of the S&P 500 series).

		2016-06-27	2016-06-30	% Decr.
Apple Inc.	AAPL	1,805,688	1,124,082	37.7
Alphabet Inc.	GOOGL	236,436	178,569	24.5
Alphabet Inc.	GOOG	202,449	152,442	24.7
Microsoft Corporation	MSFT	1,778,587	777,538	56.3
Amazon.com, Inc.	AMZN	212,951	245,500	−15.3
Facebook, Inc.	FB	863,979	521,138	39.7
Comcast Corporation	CMCSA	694,958	367,544	47.1
Intel Corporation	INTC	1,260,947	603,475	52.1
Cisco Systems, Inc.	CSCO	870,147	477,008	45.2
Amgen Inc.	AMGN	171,111	135,631	20.7
Gilead Sciences, Inc.	GILD	635,498	443,749	30.2
The Kraft Heinz Company	KHC	133,353	166,864	−25.1
Walgreens Boots Alliance, Inc.	WBA	278,448	336,216	−20.7
Starbucks Corporation	SBUX	804,742	410,650	49.0
Celgene Corporation	CELG	338,872	304,187	10.2
QUALCOMM Incorporated	QCOM	709,635	419,285	40.9
Costco Wholesale Corporation	COST	150,007	141,545	5.6
Mondelez International, Inc.	MDLZ	414,248	699,600	−68.9
The Priceline Group Inc.	PCLN	85,459	59,628	30.2
Texas Instruments Incorporated	TXN	798,510	475,992	40.4

Table 1. Number of limit order book observations on 27 and 30 June 2016 and the change (decrease) in % for the largest 20 stocks at NASDAQ.

The number of daily order book changes varies considerably across the investigated stocks, that is, between 59,628 and 1,805,688, see **Table 1**. After the referendum results, there have been many order book changes present, as compared to the trading activities on 30 June 2016. For almost all stocks, the number of changes then decreased quite substantially.

The majority of the companies had on 30 June 2016 interestingly more stocks (on average) listed at the given price levels of the order book compared to that on 27 June 2016, see **Figure 2**. For convenience, denote the observed $n \times p$ volume data matrix by x. Here the expected value of X is estimated by

Figure 2. Estimated average volume of the order book data for selected stocks on 27 June 2016 (solid) and 30 June 2016 (dashed).

$$\widehat{\mu} = n^{-1}\mathcal{X}^{\top}1_n \tag{3}$$

with a $n \times 1$ vector of ones denoted by 1_n. The average posted quantities moreover exhibit a symmetric pattern while comparing the estimated volume at the bid and ask sides.
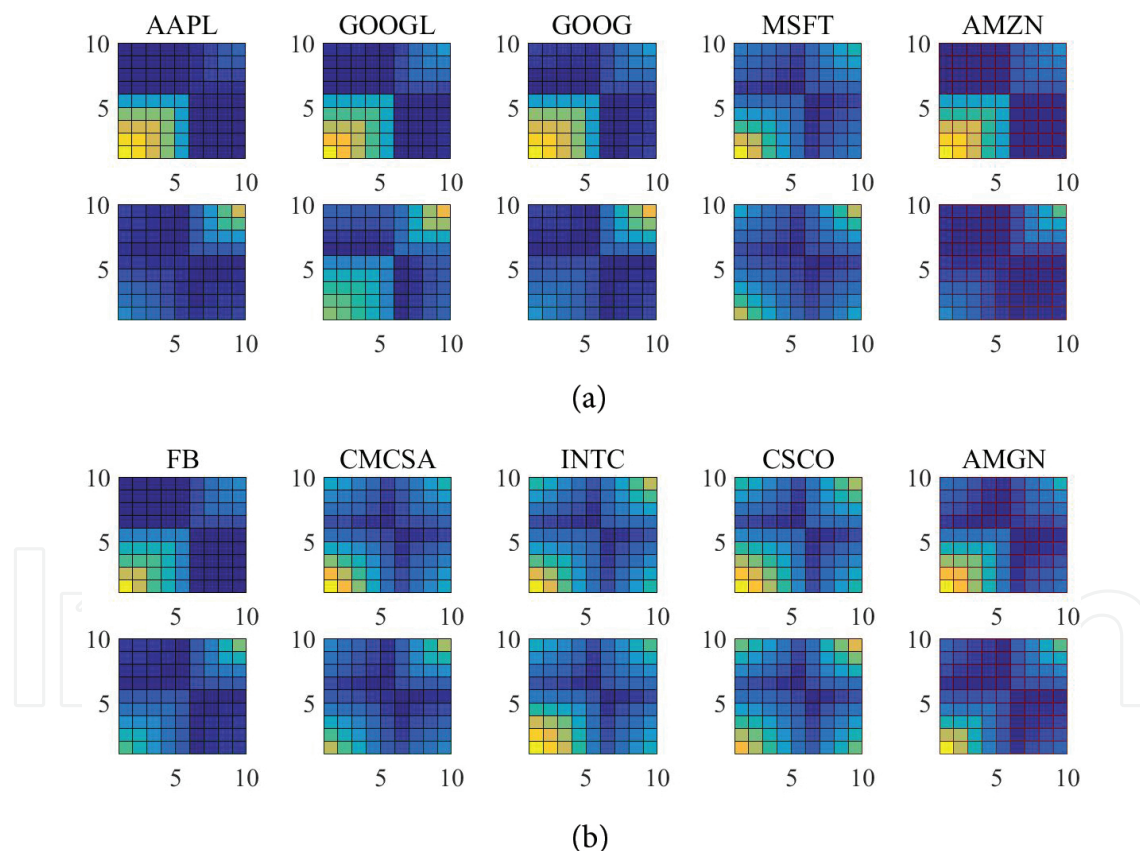
2.2. Covariance structure estimation

The results above indicate that the order book change count as well as the estimated average volume vector changed (substantially) on 30 June as compared to the market situation on the 27 June 2016. Having estimated the mean vector, we are ready to focus on the (potential) changes in the variance-covariance matrices, that is, covariance structures of the order book data. The covariance matrix of the order book volume is estimated by

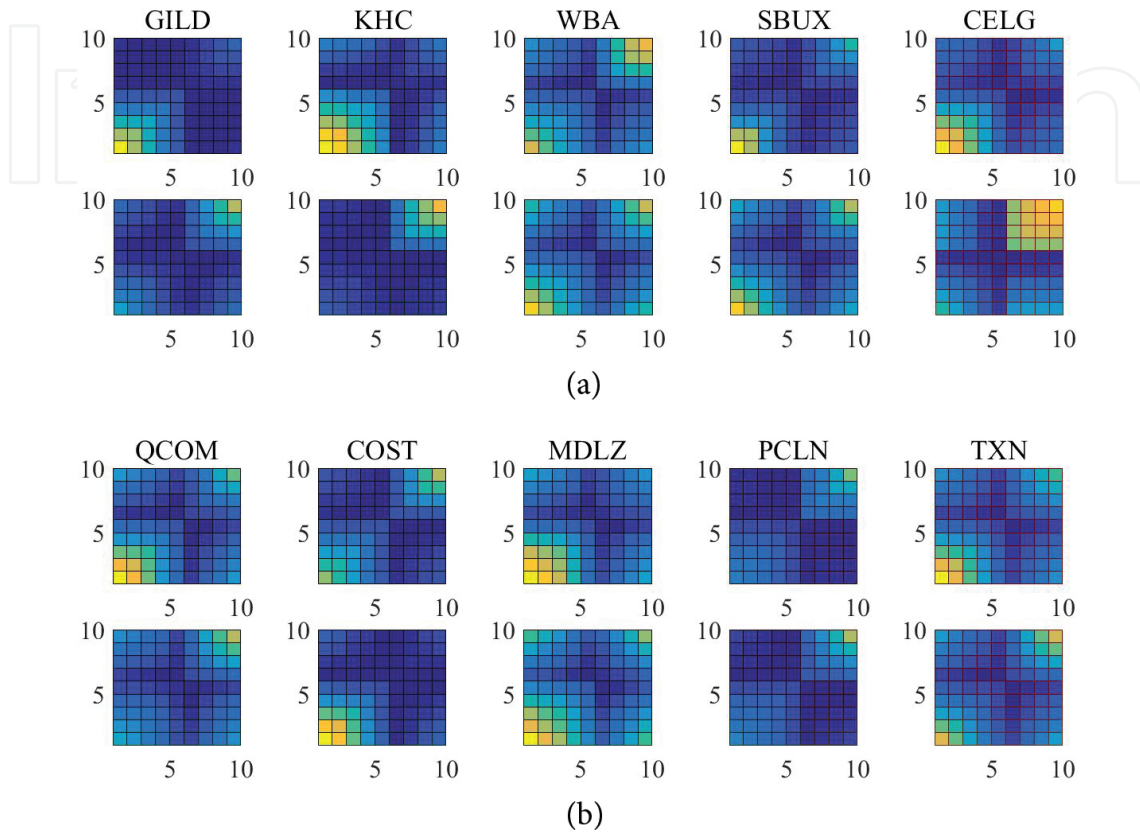$$\widehat{\Sigma} = n^{-1}\mathcal{X}^{\top}\mathcal{H}\mathcal{X} \tag{4}$$

where $\mathcal{H} = I_n - n^{-1}1_n1_n^\top$, with identity matrix I_n, denotes the centring matrix, and 1_n represents a $n \times 1$ vector of ones [8]. The empirical results are displayed in **Figures 3** and **4**, for the mega-cap and large-cap stocks, respectively. Since the analysed order book volume vector is a 10-dimensional object, $X = (X_1, \ldots, X_{10})^\top$, the axes at every graphical display represent the index of the random variable(s) under consideration. In total, there are 100 estimated covariance values displayed at each graph, that is, all values of the 10×10 matrix $\widehat{\Sigma}$. For example, the upper left square of every graph denotes the estimated covariance between X_1 and X_1 (which equals the estimated variance of X_1); the lower left square represents the estimated covariance between X_1 and X_{10}, etc. The MATLAB function 'pcolor' has been used for generating **Figures 3** and **4**. The matrix values are used to define the vertex colours by scaling the values to map to the full range of the 'colourmap', see the MATLAB documentation for more details. Note that a darker (blue) colour shows a larger value of the estimated covariance between the random variables and vice versa.

(a)

(b)

Figure 3. Estimated covariance structure of order book data: mega-cap stocks on 27 June 2016 (upper panel for each stock) and 30 June 2016 (lower panel for each stock).

Our empirical results indicate several interesting findings. One observes a relatively stronger variation in the individual volume variables than the covariance levels across all stocks. We aim identifying the linear combination that is responsible for the largest proportion of the data

variation. There are furthermore relatively larger covariance levels between the bid and ask sides on 30 June 2016 in comparison with the levels on 27 June 2016, indicating a stronger impact of one market side on order book variation immediately after the referendum results. Our analysis aims particularly to select the most important factor associated with this variation.

Figure 4. Estimated covariance structures of order book data: large-cap stocks on 27 June 2016 (upper panel for each stock) and 30 June 2016 (lower panel for each stock).

3. Statistical modelling

3.1. Modelling framework

Recall, we model the limit order book volume as a p-dimensional random vector X and denote its expected value by μ, a $p \times 1$ vector, and the covariance matrix by $\mathrm{Var}(X) = \Sigma$, a $p \times p$ matrix. After observing n realizations of X, that is, after obtaining the $n \times p$ order book volume matrix $\mathcal{X}$, the parameters μ and Σ are estimated by expressions (3) and (4), respectively.

Among multivariate techniques that deal with dimension reduction of high-dimensional random vectors, in volume covariance structure modelling we focus on the principal components, factor and discriminant analysis. Multivariate techniques deal with simultaneous relationship among variables and differ from univariate and bivariate analysis in that they

direct attention away from the analysis of the mean and variance of single variable or from the pairwise relationship between two variables, to the analysis of the covariances and correlations among three or more variables [9].

3.2. Principal components analysis

Principal component analysis focuses on standardised principal components of a high-dimensional random variable. It has been first introduced by Karl Pearson for nonstochastic variables and by Harold Hotelling for random vectors [10]. The low dimensional representation enables us to study the correlation between the principal components and the original data; here our goal is to find the standardized linear combination of the order book volume vector X that is associated with the largest order book variation. The technique is based on a very useful theorem [11], the spectral decomposition theorem. General results about eigenvalues and eigenvectors for square matrices and those for symmetrical matrices are provided in [12].

The standardized linear combination of a p-dimensional variable $X = \left(X_1,...,X_p\right)^\top$ that maximizes the order book variation uses the first eigenvector associated to the first (largest) eigenvalue of the spectral decomposition $\Sigma = \Gamma \Lambda \Gamma^\top$, $\Lambda = \text{diag}\left(\lambda_1,...,\lambda_p\right)$, $\lambda_1 \leq ... \leq \lambda_p$ being the $p \times p$ diagonal matrix of eigenvalues and Γ the $p \times p$ matrix of associated eigenvectors. The second largest variance proportion is explained by the linear combination using the second eigenvector, etc. The principal components are given by $Y = \Gamma^\top(X - \mu)$.

In modelling order book data, we estimate the principal components by

$$\top = \left(\mathcal{X} - 1_n \widehat{\mu}^\top\right) \widehat{\Gamma} \tag{5}$$

with $\widehat{\Gamma}$ the estimated matrix of eigenvectors from the spectral decomposition of $\widehat{\Sigma} = \widehat{\Gamma} \widehat{\Lambda} \widehat{\Gamma}^\top$, and the estimated $p \times p$ dimensional diagonal matrix of eigenvalues $\widehat{\Lambda}$. For illustrative purpose, it often suffices to consider only the first two principal components, that is, the first two columns of the $n \times p$ matrix $\mathcal{Y}$.

3.3. Factor analysis

In factor analysis the random vector X is modelled as a linear combination of few common factors. The concept of latent factors seems to have been suggested by Francis Galton, the formulation and early development of factor analysis have their genesis in psychology and are generally attributed to Charles Edward Spearman [10]. Factor analysis aims to discover independent variables that describe the variation of a high-dimensional random variable with high explanatory power [13]. Formally, we consider a k-factor model

$$X = QF + U + \mu \tag{6}$$

where F and U denote the k and p dimensional common and specific factors, respectively [8]. The $p \times k$ matrix of factor loadings is denoted by Q. It is furthermore assumed that $E[F] = 0_k$, $\mathsf{Var}(F) = I_k$, $\mathsf{E}[U] = 0_p$, $\mathsf{Var}(U) = \Omega = \mathsf{diag}(\omega_1, \ldots, \omega_p)$, $\mathsf{Cov}(F, U) = 0_{k \times p}$.

The associated factor loadings represent the combinations which reflect the common variance part and the remaining variation is quantified through the covariance matrix of the specific factors. In practice, we are consequently interested in estimating the matrix of common factor loadings Q and the covariance matrix of the specific factor Ω. Here we utilise the maximum likelihood method: while assuming that the volume is multivariate normally distributed [8], the estimates are given by maximising the log-likelihood function, namely

$$\left(\widehat{Q}, \widehat{\Omega} \right) = \arg \max_{Q, \Omega} \left[-\frac{n}{2} \log \left\{ \left| 2\pi \left(QQ^\top + \Omega \right) \right| \right\} + \frac{n}{2} \, \mathsf{tr} \left\{ \left(QQ^\top + \Omega \right)^{-1} \widehat{\Sigma} \right\} \right] \tag{7}$$

where n denotes the sample size and $\widehat{\Sigma}$ the estimated covariance matrix, see Eq. (4).

3.4. Discriminant analysis

In discriminant analysis, multivariate data observations are classified into two or more known groups. A modern treatment of discriminant analysis and a brief history of discriminant analysis is included in [10]. In the analysis of group differences [13], for example, state two questions: (i) does there exists a significant difference between the groups (variation) and (ii) which variables are responsible in this aspect? In practice, a discriminant rule is used to classify existing and new observations and the number of correctly classified observations reflects the quality of the approach. Here we are interested in the classification accuracy: to which extend a price change can be expected (or not) at each order book entry based exclusively on observed volume data.

The linear Fisher's discriminant rule is based on a linear combination of data, say $\mathcal{X}a$, with a denoting a $p \times 1$ vector, and the idea is to find a that achieves a good separation [8, 14]. When the method is applied to two groups, one assumes that the data matrix $\mathcal{X}$ is split into two groups, say $\mathcal{X}_1$ and $\mathcal{X}_2$. Denote the sample sizes of these matrices by n_1 and n_2, the estimated mean vectors by μ_1 and μ_2, the estimated covariance matrices by $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$, and the centering matrices by $\mathcal{H}_1$ and $\mathcal{H}_2$. The linear combination that maximizes the ratio of the between-group-sum of squares to the within-group-sum of squares is given by

$$\hat{a} = \mathcal{W}^{-1} \left(\hat{\mu}_1 - \hat{\mu}_2 \right) \tag{8}$$

where the $p \times p$ matrix $\mathcal{W}$ is related to the between-group-sum of squares as follows: $a^\top \mathcal{X}_1^\top \mathcal{H}_1 \mathcal{X}_1 a + a^\top \mathcal{X}_2^\top \mathcal{H}_2 \mathcal{X}_2 a = a^\top \mathcal{W} a$.

4. Empirical results

An analysis of principal components often reveals relationships that were not previously suspected and thereby allows interpretations that would not ordinarily results [15]. Consider, for example, the proportion of order book variance explained by two principal components in **Table 2**. Two principal components are sufficient to describe the order book variation, since the explained proportions range between 0.81 and 0.96 (27 June 2016) and 0.78 and 0.97 (30 June 2016).

The limit order book variation of most companies is clearly stronger explained on 30 June 2016 as compared to the resulting explanatory power on 27 June 2016. The largest explained proportion increase is evident for smaller stocks, especially for SBUX, CELG, QCOM, COST and PCLN. Looking only at the descriptive results reported in **Table 1** one would conclude that the number of changes is apparently similar across all stocks. Now it is evident that the demand and supply curves of smaller stocks change relatively stronger during turbulent times (here during a downward price movement). We attribute this to the relatively lower liquidity of large-cap stocks as compared to the highly liquid mega-cap stocks.

	2016-06-27	2016-06-30		2016-06-27	2016-06-30
AAPL	0.87	0.87	FB	0.84	0.85
GOOGL	0.85	0.87	CMCSA	0.94	0.89
GOOG	0.87	0.87	INTC	0.93	0.95
MSFT	0.90	0.95	CSCO	0.96	0.97
AMZN	0.85	0.78	AMGN	0.88	0.85
	2016-06-27	**2016-06-30**		**2016-06-27**	**2016-06-30**
GILD	0.83	0.89	QCOM	0.90	0.95
KHC	0.93	0.88	COST	0.84	0.88
WBA	0.93	0.92	MDLZ	0.96	0.96
SBUX	0.83	0.91	PCLN	0.81	0.89
CELG	0.88	0.94	TXN	0.92	0.94

Table 2. Estimated proportion of explained order book volume variance by the first two principal components.

Factor analysis can be considered as an extension of principal component analysis, although both techniques can be viewed as attempts to approximate the covariance matrix; however, the approximation based on the factor analysis model is more elaborate [15]. In the sequel, we chose a $k = 1$ factor model since we are interested in selecting the driving factor of order book variation. The results are depicted in **Tables 3** and **4** for the mega-cap and large-cap companies, respectively, based on the estimated values of the factor loadings $\hat{Q}$.

	2016-06-27	2016-06-30		2016-06-27	2016-06-30
AAPL	Demand	Demand	FB	Supply	Demand
GOOGL	Demand	Demand	CMCSA	Demand	Demand
GOOG	Demand	Supply	INTC	Supply	Demand
MSFT	Supply	Supply	CSCO	Demand	Both
AMZN	Demand	Demand	AMGN	Supply	Supply

Table 3. Identified common factors based on the estimated factor loadings for investigated mega-cap and largest large-cap stocks.

	2016-06-27	2016-06-30		2016-06-27	2016-06-30
GILD	Supply	Demand	QCOM	Supply	Supply
KHC	Demand	Demand	COST	Supply	Supply
WBA	Supply	Supply	MDLZ	Supply	Demand
SBUX	Supply	Demand	PCLN	Demand	Demand
CELG	Demand	Supply	TXN	Supply	Supply

Table 4. Identified common factors based on the estimated factor loadings for investigated large-cap stocks.

Across all stocks, demand is selected as the most important factor on 30 June 2016. The price of the companies indeed reacted positively during this day. For most of the relatively illiquid large-cap stocks, interestingly, the same factor has been identified on both days. Its magnitude changed, as evident from the principal components analysis.

Discriminant analysis cannot usually provide an error-free method of assignment of data, because there may not be a clear distinction between the measured characteristics of the populations—that is, the groups may overlap [15]. We report the proportions of correctly classified price changes based only on volume data in **Tables 5** and **6** for the selected mega-cap and largest large-cap, and large-cap stocks, respectively.

	2016-06-27	2016-06-30		2016-06-27	2016-06-30
AAPL	0.41	0.55	FB	0.42	0.41
GOOGL	0.50	0.59	CMCSA	0.58	0.51
GOOG	0.51	0.56	INTC	0.64	0.65
MSFT	0.61	0.64	CSCO	0.67	0.68
AMZN	0.54	0.56	AMGN	0.49	0.50

Table 5. Estimated proportion of correctly classified price changes based on volume data for investigated mega-cap and largest large-cap stocks.

	2016-06-27	2016-06-30		2016-06-27	2016-06-30
GILD	0.49	0.52	QCOM	0.50	0.64
KHC	0.47	0.50	COST	0.50	0.50
WBA	0.47	0.50	MDLZ	0.54	0.53
SBUX	0.45	0.55	PCLN	0.52	0.57
CELG	0.53	0.51	TXN	0.55	0.60

Table 6. Estimated proportion of correctly classified price changes based on volume data for investigated large-cap stocks.

The empirical findings suggest that limit order book volume data successfully classify price changes, especially on 30 June 2016, a day with relatively low number of order book entries. Here the first group contains entries with mid-quote price $(S_5 + S_6)/2$ change, and the second group entries without a change. Our results show that the classification rates changed positively quite significantly for extremely large and the smallest investigated stocks. The later ones, as discussed above, exhibit a relatively well understood covariance structure on 30 June 2016.

5. Conclusions

Limit order book data of 20 highly traded stocks at the NASDAQ market in June 2016 have been analysed. We select 2 days after the 'Britex' referendum, namely, 27 June (lowest S&P 500 level) and 30 June (recovery day). The variable of interest is the 10-dimensional order book volume data vector, that is, quantities pending at the five best levels of the demand side and at the five best supply side levels.

Two principal components account for approximately 85–95% of the order book data variation. The results of a one-factor model identify the demand (variation) as the most important factor explaining the order book covariance structure. The limit order book volume data variation is quite informative in predicting the price evolution (change or no change in the mid-quote) across all stocks and during the analysed trading activities. The mega-cap and the smallest investigated large-cap companies share almost the same classification performance. Finally, multivariate statistical techniques are successfully employed in covariance modelling of order book data.

Author details

Andrija Mihoci

Address all correspondence to: Andrija.Mihoci@b-tu.de

Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany

References

[1] R.J. Muirhead. *Aspects of Multivariate Statistics*. John Wiley & Sons, New York, 1982.

[2] J. Hasbrouck and D.J. Seppi. Common factors in prices, order flows, and liquidity. *Journal of Financial Economics*, 59:383–411, 2001.

[3] W. Bauer. Commonality in liquidity in pure order-driven markets. *National Centre of Competence in Research Working Paper*, 2004.

[4] H. Beltran-Lopez, P. Giot, and J. Grammig. Commonalities in the order book. *Financial Markets and Portfolio Management*, 23(3):209–242, 2009.

[5] W.K. Härdle, N. Hautsch, and A. Mihoci. Modelling and forecasting liquidity supply using semiparametric factor dynamics. *Journal of Empirical Finance*, 19(4):610–625, 2012.

[6] S. Yang, M. Paddrik, R. Hayes, A. Todd, A. Kirilenko, P. Beling, and W. Scherer. Behavior based learning in identifying high frequency trading strategies. In *2012 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr)*, pages 1–8, 2012.

[7] M.D. Gould, M.A. Porter, S. Williams, M. McDonald, D.J. Fenn, and S.D. Howison. Limit order books. *Quantitative Finance*, 13(11):1709–1742, 2013.

[8] W.K. Härdle and L. Simar. *Multivariate Statistical Analysis*. Springer, Berlin, 4 edition, 2015.

[9] W. R. Dillon and M. Goldstein. *Multivariate Analysis*. John Wiley & Sons, New York, 1984.

[10] N.C. Giri. *Multivariate Statistical Analysis*. Marcel Dekker, New York, 2 edition, 2004.

[11] B. Flury and H. Riedwyl. *Multivariate Statistics, A practical Approach*. Chapman and Hall, London, 1988.

[12] K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.

[13] K. Backhaus, B. Erichson, W. Plinke, and R. Weiber. *Multivariate Analysemethoden*. Springer, Berlin, Heidelberg, New York, 9 edition, 2000.

[14] L. Fahrmeir, W. Häußler, and G. Tutz. Diskriminanzanalyse. In L. Fahrmeir, A. Hamerle, and G. Tutz, editors, *Multivariate statistische Verfahren*, chapter 8, pages 357–436. Walter de Gryter & Co., Berlin, 2 edition, 1996.

[15] R.A. Johnson and D.W. Wichern. *Applied Multivariate Analysis*. Pearson Education, New Jersey, 6 edition, 2007.