Technical Document
Finance Research Graz Data Services (FiRe Graz DS)

Corinna Blasch[1], Thomas Johann[2], Roland Mestel[*1] ,
Stefan Scharnowski[2,3], Erik Theissen[1,2], Christian Westheide[4]
and Lukas Zimmermann[2]

[1]University of Graz
[2]University of Mannheim
[3]Research Center SAFE (Goethe University Frankfurt)
[4]University of Vienna

This version: November 26, 2018

Abstract

In this document we describe how we construct the database "Finance Research Graz Data Services" (FiRe Graz DS). We collect quote (best bid and ask) and trade data from Wiener Börse AG. After applying several filters, we compute various market microstructure measures at a daily frequency.

1 Introduction

Market liquidity is an important indicator for a financial market's quality and efficiency. It affects, amongst others, trading profits, asset prices, real investment decisions, and corporate actions. While the market liquidity situation is well documented for large capital markets, such as the USA, Germany or the UK, studies on small developed markets are relatively rare. As regards Austria, the lack of scientific attentiveness for its capital market and especially its liquidity can be mostly attributed to a shortage of adequate data necessary for capturing the multidimensional and complex concept. We address this problem by providing the database "Finance Research Graz Data Services" (FiRe Graz DS). It contains numerous daily market-microstructure measures of liquidity and trading activity for stocks traded on the Vienna Stock Exchange (Wiener Börse).[1]

[*]Corresponding author: University of Graz, Institute of Banking and Finance, Universitaetsstrasse 15/F2, 8010 Graz, Austria. roland.mestel@uni-graz.at

[1]The database "FiRe Graz DS" is closely related to the Market Microstructure Database Xetra (MMDB-Xetra), which contains comparable liquidity and trading activity measures for the German stock market. In constructing the database "FiRe Graz DS" we apply the same data processing routines as used for creating MMDB-Xetra. For detailed information on MMDB-Xetra, see https://www.ifk-cfs.de/research/databases/market-microstructure-database-xetra.html.

In this document, we explain how we construct the database. Section 2 describes the sample, section 3 introduces the raw data that we then filter as described in section 4. In section 5, we list and explain all variables contained in the database. Section 6 provides definitions of some measures of liquidity. In an Appendix we report problems with the raw data.

2 Sample

The database "FiRe Graz DS" comprises numerous market-microstructure measures on a daily level for all equities and ETFs traded on the Vienna Stock Exchange via the electronic trading system Xetra. We exclusively focus on continuously traded instruments.[2] Data is available from November 5^{th}, 1999 to February 13^{th}, 2018. The final sample includes 1,139 different ISINs.

3 Data Source

We obtain daily txt- and rpt-files containing market data for all equities and ETFs traded on Xetra from Wiener Börse AG. We receive separate files for trades and quotes (best bid and ask), whereby both, trade data and quote data, are available from November 5^{th}, 1999 to February 13^{th}, 2018. The data fields in the raw data are:

Quote files

1. ISIN
2. DATE: format YYYYMMDD
3. TIMESTAMP: format YYYY-MM-DD HH:MM:SS.HSEC (HSEC subdivides each second into its hundredth; can take values from 0 to 99)
4. SEQUENCE: chronologically ascending number
5. CURRENT_TRADING_PHASE_FLAG: O = opening auction, C = continuous trading, A = intraday auction, F = closing auction, V = volatility interruption
6. TYPE: x1 = best bid, x2 = best ask, 4x = vola interruption, 6x = opening auction, 7x = continuous trading, 8x = intraday auction, 9x = closing auction
7. PRICE: best bid/best ask price
8. SIZE: best bid/best ask number of shares
9. MARKET_ORDER_INDICATOR: 1 = market order, 0 = no market order

Trade files

1. ISIN

[2]For detailed information on continuous trading, see https://www.wienerborse.at/en/trading/trading-information/market-models/.

2. DATE: format YYYYMMDD
3. TIMESTAMP: format YYYY-MM-DD HH:MM:SS.HSEC (HSEC subdivides each second into its hundredth; can take values from 0 to 99)
4. TRADE_ID: unique identifier for a trade
5. SEQUENCE: chronologically ascending number
6. CURRENT_TRADING_PHASE_FLAG: O = opening auction, C = continuous trading, A = intraday auction, F = closing auction, V = volatility interruption
7. TYPE: x1 = best bid, x2 = best ask, 4x = vola interruption, 6x = opening auction, 7x = continuous trading, 8x = intraday auction, 9x = closing auction, 30 = correction of total volume and total turnover after change (A, L)
8. PRICE: price of a transaction
9. UNITS: number of shares traded
10. TOTAL_VOLUME: cumulative trading volume (in units)
11. TOTAL_TURNOVER: cumulative trading value (in EUR)
12. FLAG for change/deletion of a trade: A = change, L = deletion
13. CORRECTION_TICK_ID: TRADE_ID of the original trade in case of change or deletion

4 Data Filtering

Each day we append quote and trade data for a single asset by the respective ISIN using the statistical software Stata (version 15.1). We then apply several filters to the intraday data to remove extreme outliers and data errors. Before filtering we have 1.191 bn firm-level intraday observations. The filters remove approximately 70mn observations (5.84% of the sample), leaving us with 1.122 bn firm-level intraday observations.

Within the raw data, some few data is missing. We report these data problems in the Appendix.

The filters are applied in the following order:

1. Drop ISINs that are not traded continuously.
2. Drop data lacking information on whether it was a bid quote, an ask quote, or a trade.
3. Drop observations "outside" the trading hours.[3]
4. Drop trades if trade price or volume are not strictly positive.

[3]Trading hours for continuous trading: 9:20 - 17:30 before January 2^{nd}, 2009 and 9:00 - 17:30 afterwards.

5. Drop trades if the trade price is six standard deviations of daily trade prices below (above) the minimum (maximum) of the opening price, mean daily price or closing price.

6. Eliminate price spikes: If the transaction price of a trade is 20 percent plus 1 EUR above the price of the previous trade and the following trade, the trade is excluded. If the transaction price of a trade is 20 percent minus 1 below the price of the previous and the following trade, the trade is excluded.

7. Drop quotes if the ask or bid price is zero or negative.

8. Drop quotes if the ask price is less or equal to the bid price.

9. Drop quotes if the depth is not strictly positive.

10. Drop quotes if the quote midpoint is 50 EUR or less and the quoted half-spread is greater than 2.50 EUR; drop quotes if the quote midpoint is greater than 50 EUR and the quoted half-spread is greater than 5 percent.

11. Drop quotes if the bid price of a quote is less than the minimum of the first bid, the average bid, or the last bid price of the stock-day minus six standard deviations of the bid price. Drop quotes if the ask price of a quote is more than the maximum of the first ask, the average ask, or the last ask price of the stock-day plus six standard deviations of the ask price.

5 Variables

We construct the following variables for each day:

Identifiers and time variable

1. **isin**: ISIN as in XETRA

2. **date**: Date (DD.MM.YYYY)

Auctions

We differentiate between four kinds of auctions:

- Opening auction (o)
- Intraday auction (a)
- Closing auction (f)
- Volatility interruption (v)

In what follows, $\circledast$ is a placeholder for either of the auction types.

3. **auction_price_**$\circledast$: The price of a $\circledast$-auction. In case of multiple volatility auctions, their mean price is given.

4. **auction_unit_⊛**: The volume of a ⊛-auction measured as the number of shares. In case of multiple volatility auctions, their total volume is given.

Opening and closing trades

5. **opening_trade_price:** Price of the first trade (non-auction trades)
6. **opening_trade_volume:** Volume of the first trade (non-auction-trades)
7. **closing_trade_price:** Price of the last trade (non-auction trades)
8. **closing_trade_volume:** Volume of the last trade (non-auction trades)

Maximum and minimum trade prices

9. **max_price:** Maximum trade price
10. **min_price:** Minimum trade price

Volatility measure

11. **midpoint_vola_5:** *Midpoint Volatility*[4] (5 minute intervals)

Number, volume and value of trades (continuous trading)

12. **final_buys**: Total number of buyer-initiated trades[5]
13. **final_sells**: Total number of seller-initiated trades
14. **final_trades**: Total number of trades
15. **sum_buy_vol**: Total volume of buyer-initiated trades
16. **sum_buy_val**: Total value (price*volume) of buyer-initiated trades
17. **sum_sell_vol**: Total volume of seller-initiated trades
18. **sum_sell_val**: Total value (price*volume) of seller-initiated trades

Order imbalance

19. **oib**: *Order Imbalance*
20. **oib_val**: *Order Imbalance* in terms of value (price*volume)
21. **oib_vol**: *Order Imbalance* in terms of volume

The following variables are aggregated as time-weighted and equally-weighted means over all quotes during continuous trading. The variable names contain the suffixes ***time*** and ***equal***, indicated by ⋇ below.

Quoted spreads

[4]Italic terms are defined in section 6.
[5]We use the algorithm of Lee and Ready (1991) to classify trades into buys and sells.

22. **quoted_⁎:** *Quoted Spread*
23. **rel_quoted_⁎:** *Relative Quoted Spread*

Quoted depth

24. **bid_depth_⁎:** *Depth* on the bid side
25. **ask_depth_⁎:** *Depth* on the ask side
26. **bid_depth_euro_⁎:** *Depth* on the bid side in terms of value (price*volume)
27. **ask_depth_euro_⁎:** *Depth* on the ask side in terms of value (price*volume)

The following variables are aggregated as volume-weighted and equally-weighted means over all trades during continuous trading. The variable names contain the suffixes ***value*** and ***equal***, indicated by ⊠ below.

Trade prices

28. **trade_price_⊠:** Trade price

Effective spreads and price impact

29. **effective_⊠:** *Effective Spread*
30. **rel_effective_⊠:** *Relative Effective Spread*
31. **pi_5_⊠:** *Price Impact* (5 minute intervals)

6 Definitions

We use the index i for a firm $i \in I$ (total number of firms), $t \in T$ for a trading day, τ for time at trading day t, and the index $j \in J$ for the number of the trade or the corresponding quote immediately before the trade, where J is the total number of trades for a stock i on day t.

- **Quoted spread**:

$$quoted_{it\tau} = ask_{it\tau} - bid_{it\tau} \tag{1}$$

- **Relative quoted spread**:

$$rel\_quoted_{it\tau} = \frac{ask_{it\tau} - bid_{it\tau}}{\frac{1}{2} * (ask_{it\tau} + bid_{it\tau})} \tag{2}$$

- **Midpoint**:

$$M_{itj} = \frac{bid_{itj} + ask_{itj}}{2}. \tag{3}$$

- **Effective spread**:

$$effective_{itj} = 2 * |P_{itj} - M_{itj}| \tag{4}$$

- **Relative effective spread**:

$$rel\_effective_{itj} = \frac{2 * |P_{itj} - M_{itj}|}{M_{itj}} = \frac{effective_{itj}}{M_{itj}} \tag{5}$$

- **Midpoint volatility**:

  Let M_{itf}^{5min} be the last midpoint of a five minute interval f and F the total number of daily five minute intervals without missing midpoint.

$$midpoint\_vola_{it} = \frac{1}{F-1} \sum_{f=2}^{F} \left(\frac{M_{itf}^{5min} - M_{it(f-1)}^{5min}}{M_{it(f-1)}^{5min}} \right)^2 \tag{6}$$

- **Price impact**:

  Let M_{itj}^{5min} be the midpoint five minutes after M_{itj} and let

$$q_{itj} = \begin{cases} 1 & \text{if trade } itj \text{ is a buy} \\ -1 & \text{if trade } itj \text{ is a sell} \end{cases}$$

  Then the price impact is defined as

$$PriceImpact_{itj} = \frac{\left(M_{itj}^{5min} - M_{itj} \right) * q_{itj}}{M_{itj}} \tag{7}$$

- **Order imbalance**:

  Order imbalance is the difference between buyer and seller initiated transactions relative to all transactions, computed using either the number of transactions, the number of shares traded in these transactions, or the value of these transactions:

$$OIB_{it} = \frac{final\_buys_{it} - final\_sells_{it}}{final\_buys_{it} + final\_sells_{it}} \tag{8}$$

$$OIB_{it}^{vol} = \frac{sum\_buy\_vol_{it} - sum\_sell\_vol_{it}}{sum\_buy\_vol_{it} + sum\_sell\_vol_{it}} \tag{9}$$

$$OIB_{it}^{val} = \frac{sum\_buy\_val_{it} - sum\_sell\_val_{it}}{sum\_buy\_val_{it} + sum\_sell\_val_{it}} \tag{10}$$

- **Depth bid/ask side**: *Depth* on the bid/ask side is defined as

$$d_{it\tau}^{bid} = bid\_quantity_{it\tau} \tag{11}$$

$$d_{it\tau}^{ask} = ask\_quantity_{it\tau} \tag{12}$$

- **Depth bid/ask side in terms of value**: *Depth* on the bid/ask side in Euro is calculated as

$$d_{it\tau}^{val,bid} = d_{it\tau}^{bid} * bid_{it\tau} \tag{13}$$

$$d_{it\tau}^{val,ask} = d_{it\tau}^{ask} * ask_{it\tau} \tag{14}$$

References

Lee, Charles M. C. and Mark J. Ready (1991). "Inferring trade direction from intraday data". In: *The Journal of Finance* 46.2, pp. 733–746.

Appendix

Tables 1 and 2 show problems with the raw data. In some cases, quote/trade data for one day are missing, in some cases they are incomplete.

Table 1: Data Problems Quote Data

Date	**Problem**
20.12.2000	Missing data before 12:30:01.440 o'clock.
03.07.2001	Raw data contain wrong time stamps from 15:50 o'clock.
30.08.2007	Missing data before 11:25:01.650 o'clock.
21.11.2011	Missing data before 09:02:02.270 o'clock.
21.09.2012	Missing data before 09:02:55.200 o'clock.
08.08.2016	Missing data.
14.09.2016	Missing data.
30.11.2016	Missing data.
03.11.2017	Raw data contain many observations with missing time stamp during final auction.
06.11.2017	Raw data contain many observations with missing time stamp during final auction.
07.11.2017	Raw data contain many observations with missing time stamp during final auction.
08.11.2017	Raw data contain many observations with missing time stamp during final auction.
09.11.2017	Raw data contain many observations with missing time stamp during final auction.
10.11.2017	Raw data contain many observations with missing time stamp during final auction.
13.11.2017	Raw data contain many observations with missing time stamp during final auction.
14.11.2017	Raw data contain many observations with missing time stamp during final auction.
15.11.2017	Raw data contain many observations with missing time stamp during final auction.
16.11.2017	Raw data contain many observations with missing time stamp during final auction.
17.11.2017	Raw data contain many observations with missing time stamp during final auction.
20.11.2017	Raw data contain many observations with missing time stamp during final auction.
21.11.2017	Raw data contain many observations with missing time stamp during final auction.
27.11.2017	Raw data contain many observations with missing time stamp during final auction.

28.11.2017	Raw data contain many observations with missing time stamp during final auction.
29.11.2017	Raw data contain many observations with missing time stamp during final auction.
30.11.2017	Raw data contain many observations with missing time stamp during final auction.
01.12.2017	Raw data contain many observations with missing time stamp during final auction.
29.12.2017	First observation at 05:35:26.200 o'clock, last observation at 17:07:46.753 o'clock (first observation with positive price and positive volume at 07:55:00.017 o'clock; last

Table 2: Data Problems Trade Data

Date	**Problem**
20.12.2000	Missing data before 12:36:13.210 o'clock.
28.12.2001	Missing data after 15:21:23.870 o'clock.
30.08.2007	Missing data before 11:30:09.320 o'clock.
30.09.2008	Missing data before 09:22:29.010 o'clock.
01.10.2008	Missing data before 09:20:03.000 o'clock.
10.10.2008	Missing data before 09:24:05.000 o'clock.
14.10.2008	Missing data before 09:22:28.000 o'clock.
17.02.2009	Missing data before 09:04:26.000 o'clock.
18.02.2009	Missing data before 09:15:26.040 o'clock.
19.02.2009	Missing data before 09:04:20.000 o'clock.
20.02.2009	Missing data before 09:08:10.010 o'clock.
23.02.2009	Missing data before 09:20:10.480 o'clock.
24.02.2009	Missing data before 09:25:34.040 o'clock.
25.02.2009	Missing data before 09:32:31.560 o'clock.
05.08.2011	Missing data before 09:02:18.000 o'clock.
10.08.2011	Missing data before 09:02:19.000 o'clock.
21.11.2011	Missing data before 09:02:18.510 o'clock.
29.12.2017	Missing data after 14:21:17.640 o'clock except for one observation at 18:07:46.750 o'clock.