Financial Econometrics and Volatility Models
Introduction to High Frequency Time Series

Eric Zivot
Updated: May 1, 2009

TAQ Data in ASCII Form

MSFT: 5/1/97 – 5/15/97 (2 weeks)

- 98,724 trades; 20,656 quotes
- Extracted from TAQ DVD to ASCII file

```
cond |ex |symbol |corr |g127 |price |siz |tdate |tseq |ttim |
T   |T   |MSFT   |0   |0   |121.125   |1500   |01MAY1997   |0   |28862   |
T   |T   |MSFT   |0   |0   |121.5625   |500   |01MAY1997   |0   |28944   |
T   |T   |MSFT   |0   |0   |121.5625   |1000   |01MAY1997   |0   |29000   |
T   |T   |MSFT   |0   |0   |121.5625   |1200   |01MAY1997   |0   |29002   |
T   |T   |MSFT   |0   |0   |121.625   |1000   |01MAY1997   |0   |31095   |
```

TAQ Data in S-PLUS

Representation as `timeSeries` object in S-PLUS

```
> msftt.ts[1:5,]
          Positions Cond  Ex Symbol Corr G127  Price    Size Seq
5/1/1997 8:01:02 T     T  MSFT    0    0    121.1250 1500 0
5/1/1997 8:02:24 T     T  MSFT    0    0    121.5625  500 0
5/1/1997 8:03:20 T     T  MSFT    0    0    121.5625 1000 0
5/1/1997 8:03:22 T     T  MSFT    0    0    121.5625 1200 0
5/1/1997 8:38:15 T     T  MSFT    0    0    121.6250 1000 0
```

Dates are in timeDate object

Data is in a data frame

Olsen Data in S-PLUS

USD/EUR spot rate quotes: 3/11/2001-3/17/2001 (2 weeks)

+ 126,988 quotes

```
> eurusd.ts[1:5,]
           Positions     Bid    Ask Institution
3/11/2001 22:01:35 0.9326 0.9330 ONEC
3/11/2001 22:01:37 0.9326 0.9331 AREX
3/11/2001 22:09:34 0.9326 0.9331 NWHK
3/11/2001 22:09:36 0.9327 0.9332 AREX
3/11/2001 22:11:08 0.9322 0.9327 NWHK
```

Aligning Time Series

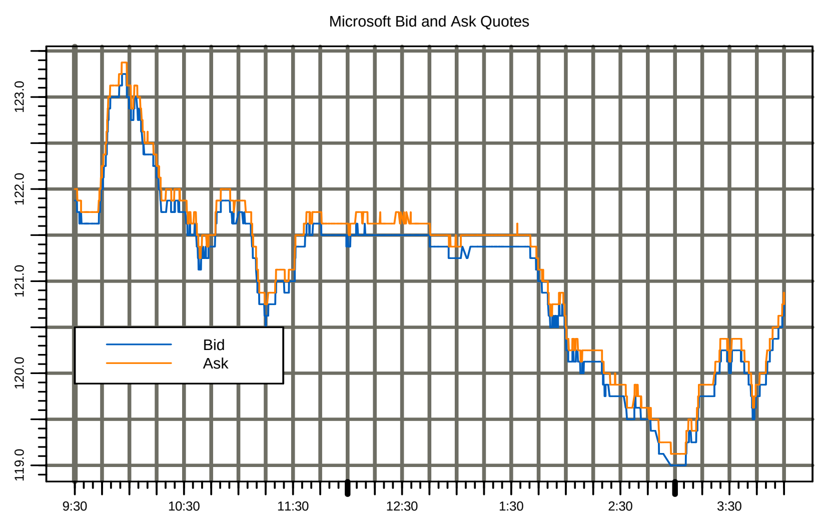
```
> msftt.ts[1:5,"Price"]
        Positions    Price
 5/1/1997 9:30:02 122.000
 5/1/1997 9:30:06 122.125
 5/1/1997 9:30:09 122.000
 5/1/1997 9:30:10 122.000
 5/1/1997 9:30:14 122.125
> msftq.ts[1:5,"Bid"]
        Positions      Bid
 5/1/1997 9:30:01 122.000
 5/1/1997 9:30:06 122.000
 5/1/1997 9:30:13 122.000
 5/1/1997 9:30:14 121.875
 5/1/1997 9:30:17 121.875
```

```
> align.ts =
   align(msftq.ts[,"Bid"],
+ pos = positions(msftt.ts),
+ how = "nearest")

> align.ts[1:5]
        Positions      Bid
 5/1/1997 9:30:02 122.000
 5/1/1997 9:30:06 122.000
 5/1/1997 9:30:09 122.000
 5/1/1997 9:30:10 122.000
 5/1/1997 9:30:14 121.875
```

Other align options: `drop`, `before`, `after`, `interep`

Plotting high frequency data



How Much Data Can You Analyze in S-PLUS?

On 32 bit operating systems theoretical limit is 4GB of addressable memory

On Windows, practical limit is closer to 1.5GB

S-PLUS memory requirements

- # of bytes required for data = r*c*8*4.5
- r = rows, c = columns, 8 = bytes for numeric data, 4.5 = avg # of data copies for modeling functions
- Ex: Data set with 98,672 rows and 507 columns requires about 1.8 GB memory

S-PLUS and R Code for Analysis of High Frequency Data

Yan, B. and E. Zivot (2004). “Analysis of High-Frequency Data with S-PLUS”, Working Paper, Department of Economics, University of Washington

Library is incorporated into S+FinMetrics 3.0

R package RTAQ by Jonathan Cornelissen and Kris Boudt

R package Realized by Scott Payseur (former UW PhD student)

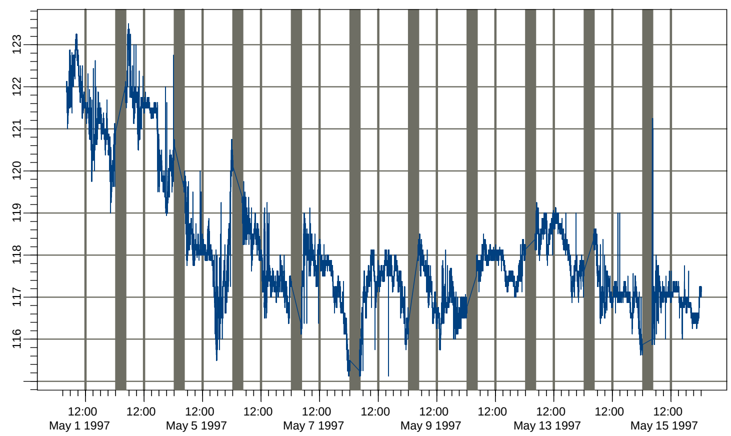
Overview of S-PLUS Library HF

- Access data from TAQ and Olsen FxFx databases
- Perform data cleaning and graphical diagnostics
- Define exchange and market time
- Construct market variables
  - Price change, B/A spread, duration, trade direction, realized volatility
- Enhancements to S-PLUS functions `align` and `aggregateSeries` to better handle HF financial data
- Construction of realized variance measures
- Nonparametric estimation of intra-day periodicities

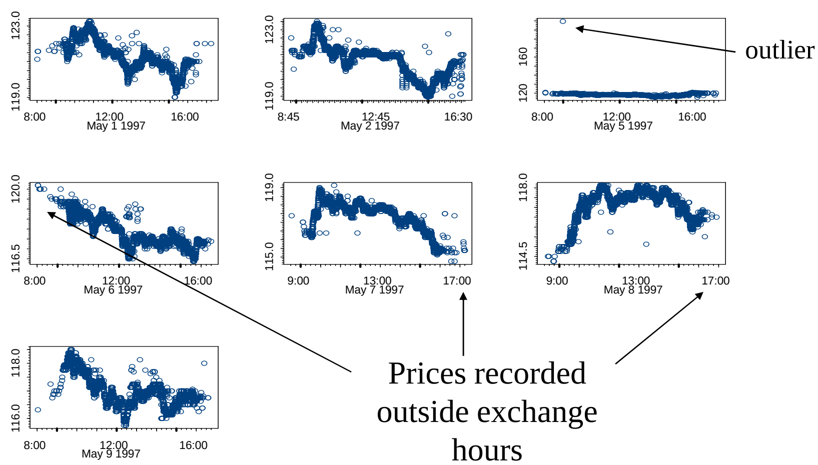
Data Cleaning and Graphics

- Common Data Errors
  - Mis-ordered time-stamps
  - Data recording errors
  - Missing or partial data
  - Time stamps outside of trading hours
- Graphical Diagnositics are Essential!!!
  - Must be careful because large amount of HF data may overwhelm plotting functions
  - HF function **`plotByDays()`**

MSFT Trade Price: 5/1/97 – 5/15/97



Trade Price by Day



Creating Market Variables

- Price/Quote Changes
  - +Price impact analysis
  - +Price Discovery
- Durations – time between events
  - +Many types of duration
    - Transaction, quote, price, volume
  - +Liquidity modeling
- Spreads (Bid/Ask)
  - +Market maker behavior
- Trade Direction – Buy/Sell Indicators
  - +Demand modeling
- Volatility Measures
  - +Derivatives pricing, Value-at-Risk

Complications

- Must separate overnight from intra-day changes
- Restrict data to exchange hours (Equity) or business week (FX)
- Need to deal with holidays, daylight savings times (DST), market closures
- Remove intraday seasonalities (diurnal effects) prior to modeling

Compute Price Changes

```
> msftt.ts = ExchangeHoursOnly(ts = msftt.ts,
+             exch.hours = c("9:30", "16:00"),
+             start.include = T, close.include = T)

> pcTicks.msft = PriceChgInInterv(msftt.ts[, "Price"],
+                 ticksize = 1/8,
+                 interv.type = "daily",
+                 bound.hours = c("9:30", "16:00"))

> pcTicks.msft[1:3]
        Positions Price
 5/1/1997 9:30:06  1
 5/1/1997 9:30:09 -1
 5/1/1997 9:30:10  0
```


Compute Duration Between Trades

```
> duration.msftt = DurationInInterv(x = msftt.ts,
+                 units = "seconds",
+                 interv.type = "daily",
+                 bound.hours = c("9:30", "16:00"))
> duration.msftt[1:5, ]
Positions Duration.in.seconds
5/1/1997 9:30:06 4
5/1/1997 9:30:09 3
5/1/1997 9:30:10 1
5/1/1997 9:30:14 4
5/1/1997 9:30:14 0
```

Compute Bid/Ask Spread

```
> spread.msft = getSpread(ask = msftq.ts[, "Ask"],
+                         bid = msftq.ts[, "Bid"],
+                         ticksize = 1/8)
> spread.msft[1:5, ]
Positions Spread
5/1/1997 9:30:14 1
5/1/1997 9:30:17 2
5/1/1997 9:30:17 1
5/1/1997 9:30:21 1
5/1/1997 9:30:57 1
```


Trade Direction – Buy or Sell Indicator

- TAQ Consolidated Tape does not indicate if transaction is "buyer" or "seller" initiated
- Use Lee-Ready rule to infer trade direction
  - Trade is "buy" if price > mid-quote lagged 5 seconds
  - Trade is "sell" if price < mid-quote lagged 5 seconds
  - Trade is "indeterminate" if price = mid-quote lagged 5 seconds
- Requires merge of Trade and Quote data

Compute Trade Direction

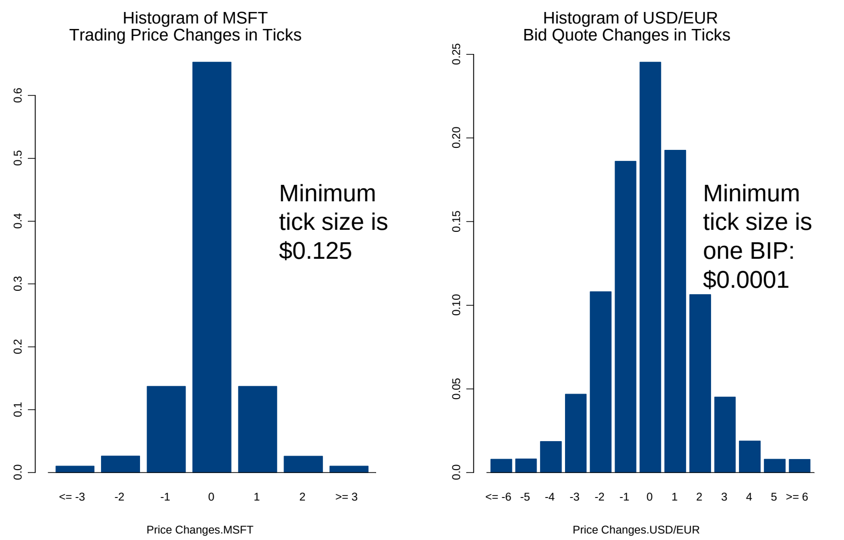
```
> mq.msft = getMidQuote(ask = msftq.ts[,"Ask"],
+                       bid = msftq.ts[, "Bid"])
> trade.direc.msft =
+ tradeDirec(trade = msftt.ts[, "Price"],
+            mq = mq.msft,
+            timeLag = "5s")

> trade.direc.msft[1:5,]
        Positions BuySellDirec
 5/1/1997 9:30:02 0
 5/1/1997 9:30:06 1
 5/1/1997 9:30:09 0
 5/1/1997 9:30:10 0
 5/1/1997 9:30:14 1
```


Descriptive Analysis of High Frequency Data

- Price changes of transaction prices and quotes are discrete valued variables, only taking values in multiples of tick sizes.
- There is tendency for price reversal, or bid-ask bounce in transaction price changes.
- Typically during active trading periods, several trades or quotes may appear to occur at the “same” time and share the same time stamp. Consequently, there may be a significant fraction of transactions with zero durations.
- Prices are often recorded at regular intervals (e.g. every 5 minutes) but not all assets trade at the same time or with the same frequency. This may cause cross correlation between returns, serial correlation in portfolio returns and negative serial correlation in individual returns.

Descriptive Analysis: Price Change



- Result: Bid-Ask spread introduces negative lag-1 serial correlation in an asset return
- Intuition comes from Roll's (1984) model

$$P_t = P_t^* + I_t \cdot \frac{S}{2}$$

$$P_t^* = \text{ constant fundamental value independent of } S$$

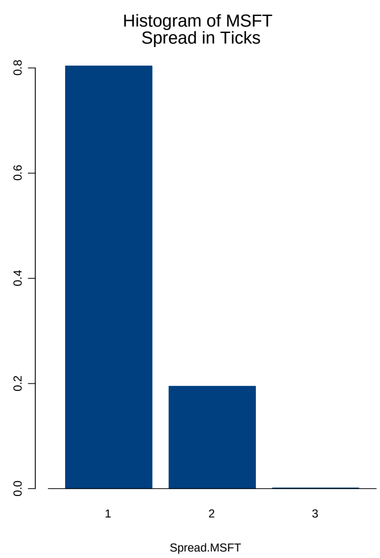
$$S = P_{Ask} - P_{Bid}$$

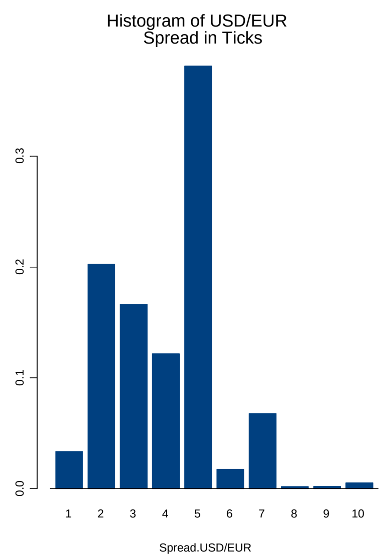
$$I_t = \begin{cases} 1 & \text{with probability } 0.5 \\ -1 & \text{with probability } 0.5 \end{cases}$$

Descriptive Analysis: Price Change

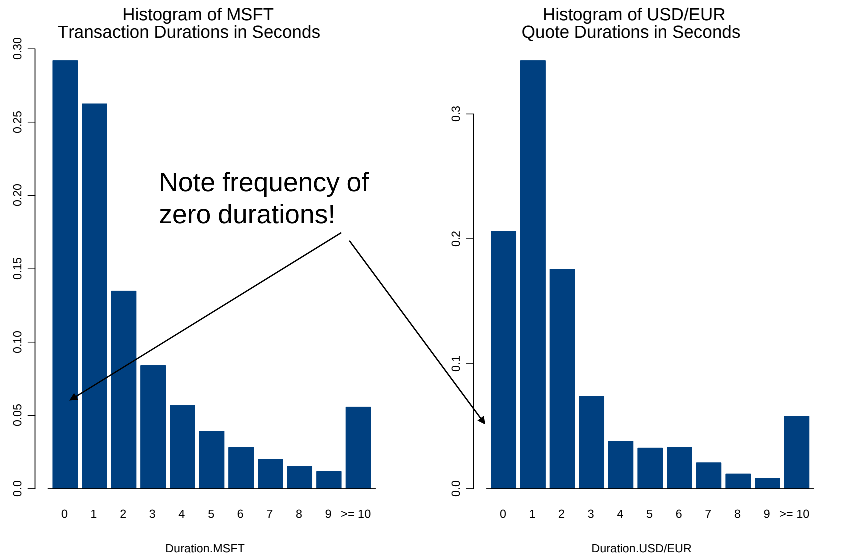
MSFT Price Changes	*i*th Trade		
(*i*-1)th Trade	"+"	"0"	"-"
"+"	787	8058	8020
"0"	8449	46869	8077
"-"	7630	8468	757

Descriptive Analysis: Spread





Descriptive Analysis: Duration

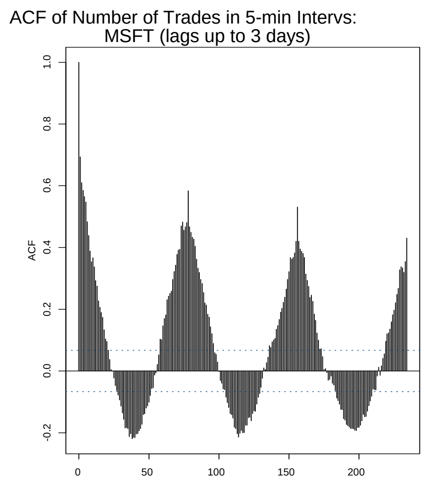


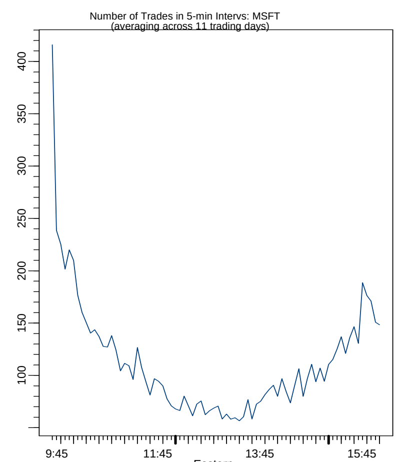
- Intraday calendar patterns (diurnal effects) have been found in the volatility of asset prices, transaction volumes,tick frequency, duration between ticks, and bid/ask spreads
- Equity activity variables, except duration, follow a reserve J-shaped pattern over trading hours. Duration follows an inverted U shape
- FX trading activities also follow an intra-day calendar pattern with three peaks corresponding to the business hours of three geographical trading centers (i.e. Asian, European, and American).

Nonparametric Estimation of Diurnal Effects

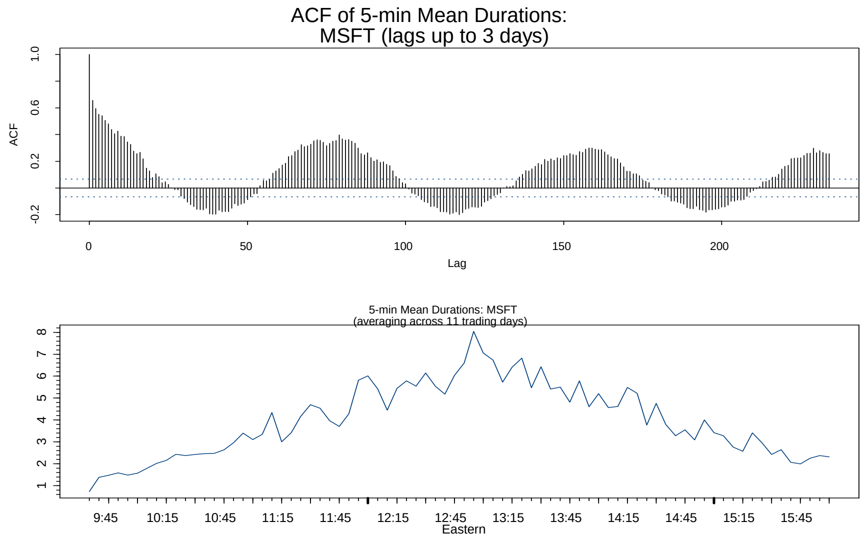
- Deterministic diurnal effects can be estimated by smoothing or averaging the variable in question across trading days.
- For example, the average duration between 9:00 and 9:35 for all of the observed trading days can be averaged to estimate the seasonal component of duration between 9:00 and 9:35. This can be done for all intraday time intervals.
- Alternatively one can use splines or trigonometric polynomials to capture diurnal effects.
- Statistical modeling should be done using seasonally adjusted variables.

Diurnal Effects in Trading Activity: MSFT Stock





Diurnal Effects in Duration: MSFT Transactions

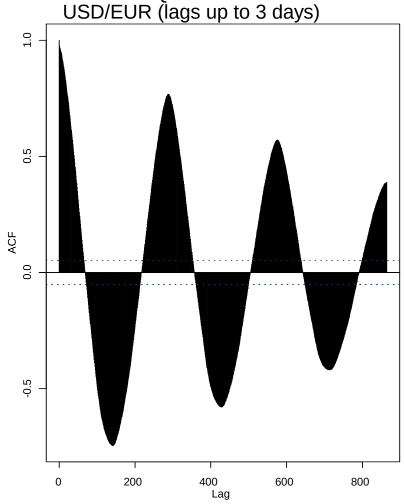


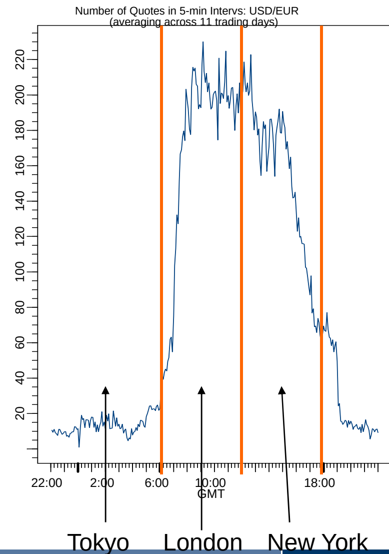
Intraday Trading Sessions for 24 Hour FX Market

	Asian	European	American	Post-American
Hours in GMT	22:00 - 06:00	06:00 - 12:00	12:00 - 18:00	18:00 - 22:00

Diurnal Effects in Quote Activity: USD/EUR

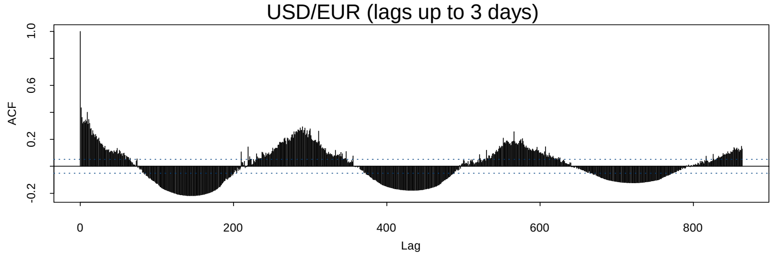
ACF of Number of Quotes in 5-min Intervs:
USD/EUR (lags up to 3 days)

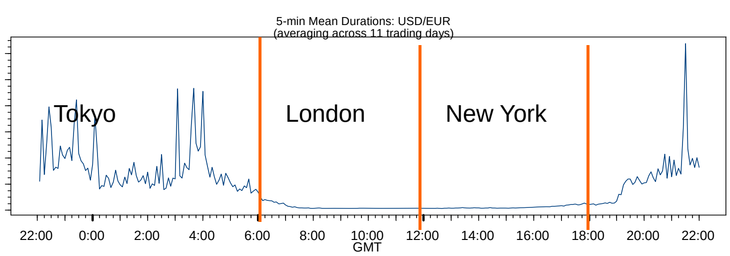




Diurnal Effects in Quote Duration:USD/EUR

ACF of 5-min Mean Durations:
USD/EUR (lags up to 3 days)





Texbook and Monograph References

Campbell, J., A. Lo, and C. MacKinlay. ***The Econometrics of Financial Markets***, Princeton University Press, 1997.

Tsay, R. ***Analysis of Financial Time Series***, John Wiley & Sons, 2002.

Gourerioux, C., J. Jasiak. ***Financial Econometrics***, Princeton University Press, 2001.

Dacarogna, M., M. Gencay, U.A. Muller, R. Olsen, O.V. Pictet. ***An Introduction to High Frequency Finance***, Academic Press, 2001.

Bauwens, L., P. Giot. ***Econometric Modeling of Stock Market Intraday Activity***. Kluwer, 2001.

Hasbrouck, J. ***Empirical Analysis of Market Micro-Structure***, Lecture notes, New York University, 2004.