**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

<table>
<thead>
<tr>
<th>cond</th>
<th>ex</th>
<th>symbol</th>
<th>corr</th>
<th>g127</th>
<th>price</th>
<th>siz</th>
<th>tdate</th>
<th>tseq</th>
<th>ttim</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>MSFT</td>
<td>0</td>
<td>0</td>
<td>121.125</td>
<td>1500</td>
<td>01MAY1997</td>
<td>0</td>
<td>28862</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>MSFT</td>
<td>0</td>
<td>0</td>
<td>121.5625</td>
<td>500</td>
<td>01MAY1997</td>
<td>0</td>
<td>29344</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>MSFT</td>
<td>0</td>
<td>0</td>
<td>121.5625</td>
<td>1000</td>
<td>01MAY1997</td>
<td>0</td>
<td>29000</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>MSFT</td>
<td>0</td>
<td>0</td>
<td>121.5625</td>
<td>1200</td>
<td>01MAY1997</td>
<td>0</td>
<td>29002</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>MSFT</td>
<td>0</td>
<td>0</td>
<td>121.625</td>
<td>1000</td>
<td>01MAY1997</td>
<td>0</td>
<td>31095</td>
</tr>
</tbody>
</table>
### TAQ Data in S-PLUS

**Representation as `timeSeries` object in S-PLUS**

```r
> msftt.ts[1:5,]

<table>
<thead>
<tr>
<th>Positions Cond</th>
<th>Ex</th>
<th>Symbol</th>
<th>Corr</th>
<th>GI27</th>
<th>Price</th>
<th>Size</th>
<th>Seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/1/1997 8:01:02 T T MSFT 0 0 121.1250 1500 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/1/1997 8:02:24 T T MSFT 0 0 121.5625 500 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/1/1997 8:03:20 T T MSFT 0 0 121.5625 1000 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/1/1997 8:03:22 T T MSFT 0 0 121.5625 1200 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/1/1997 8:38:15 T T MSFT 0 0 121.6250 1000 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Dates are in `timeDate` object

Data is in a data frame

### Olsen Data in S-PLUS


+ 126,988 quotes

```r
> eurusd.ts[1:5,]

<table>
<thead>
<tr>
<th>Positions</th>
<th>Bid</th>
<th>Ask</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/11/2001 22:01:35</td>
<td>0.9326</td>
<td>0.9330</td>
<td>ONEC</td>
</tr>
<tr>
<td>3/11/2001 22:01:37</td>
<td>0.9326</td>
<td>0.9331</td>
<td>AREX</td>
</tr>
<tr>
<td>3/11/2001 22:09:34</td>
<td>0.9326</td>
<td>0.9331</td>
<td>NWHK</td>
</tr>
<tr>
<td>3/11/2001 22:09:36</td>
<td>0.9327</td>
<td>0.9332</td>
<td>AREX</td>
</tr>
<tr>
<td>3/11/2001 22:11:08</td>
<td>0.9322</td>
<td>0.9327</td>
<td>NWHK</td>
</tr>
</tbody>
</table>
```
Aligning Time Series

> msftt.ts[1:5, "Price"]

<table>
<thead>
<tr>
<th>Positions</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/1/1997 9:30:02</td>
<td>122.000</td>
</tr>
<tr>
<td>5/1/1997 9:30:06</td>
<td>122.125</td>
</tr>
<tr>
<td>5/1/1997 9:30:09</td>
<td>122.000</td>
</tr>
<tr>
<td>5/1/1997 9:30:10</td>
<td>122.000</td>
</tr>
<tr>
<td>5/1/1997 9:30:14</td>
<td>122.125</td>
</tr>
</tbody>
</table>

> msftq.ts[1:5, "Bid"]

<table>
<thead>
<tr>
<th>Positions</th>
<th>Bid</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/1/1997 9:30:01</td>
<td>122.000</td>
</tr>
<tr>
<td>5/1/1997 9:30:06</td>
<td>122.000</td>
</tr>
<tr>
<td>5/1/1997 9:30:13</td>
<td>122.000</td>
</tr>
<tr>
<td>5/1/1997 9:30:14</td>
<td>121.875</td>
</tr>
<tr>
<td>5/1/1997 9:30:17</td>
<td>121.875</td>
</tr>
</tbody>
</table>

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

> align.ts[1:5]

<table>
<thead>
<tr>
<th>Positions</th>
<th>Bid</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/1/1997 9:30:02</td>
<td>122.000</td>
</tr>
<tr>
<td>5/1/1997 9:30:06</td>
<td>122.000</td>
</tr>
<tr>
<td>5/1/1997 9:30:09</td>
<td>122.000</td>
</tr>
<tr>
<td>5/1/1997 9:30:10</td>
<td>122.000</td>
</tr>
<tr>
<td>5/1/1997 9:30:14</td>
<td>121.875</td>
</tr>
</tbody>
</table>

Other align options: drop, before, after, interp

Plotting high frequency data

Microsoft Bid and Ask Quotes

![Graph showing high frequency data with blue and orange lines for bid and ask prices over time from 9:30 AM to 3:30 PM on May 2, 1997.]
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

\[ \text{# of bytes required for data} = r \times c \times 8 \times 4.5 \]

\( r \) = 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


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 Diagnostics are Essential!!!
  + Must be careful because large amount of HF data may overwhelm plotting functions
  + HF function plotByDays()
Trade Price by Day

Prices recorded outside exchange hours

outlier
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

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

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

```r
> pcTicks.msft[1:3]

<table>
<thead>
<tr>
<th>Positions</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/1/1997</td>
<td>9:30:06</td>
</tr>
<tr>
<td>5/1/1997</td>
<td>9:30:09</td>
</tr>
<tr>
<td>5/1/1997</td>
<td>9:30:10</td>
</tr>
</tbody>
</table>
```

Compute Duration Between Trades

```r
> duration.msftt = DurationInInterv(x = msftt.ts,
+     units = "seconds",
+     interv.type = "daily",
+     bound.hours = c("9:30", "16:00"))
```

```r
> duration.msftt[1:5,]

<table>
<thead>
<tr>
<th>Positions</th>
<th>Duration.in.seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/1/1997</td>
<td>9:30:06: 4</td>
</tr>
<tr>
<td>5/1/1997</td>
<td>9:30:09: 3</td>
</tr>
<tr>
<td>5/1/1997</td>
<td>9:30:10: 1</td>
</tr>
<tr>
<td>5/1/1997</td>
<td>9:30:14: 4</td>
</tr>
<tr>
<td>5/1/1997</td>
<td>9:30:14: 0</td>
</tr>
</tbody>
</table>
```
Compute Bid/Ask Spread

```r
> 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

```r
> 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,]

<table>
<thead>
<tr>
<th>Positions</th>
<th>BuySellDir</th>
<th>Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5/1/1997</td>
<td>9:30:02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5/1/1997</td>
<td>9:30:06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5/1/1997</td>
<td>9:30:09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5/1/1997</td>
<td>9:30:10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5/1/1997</td>
<td>9:30:14</td>
</tr>
</tbody>
</table>
```

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

Histogram of MSFT Trading Price Changes in Ticks

Minimum tick size is $0.125

Histogram of USD/EUR Bid Quote Changes in Ticks

Minimum tick size is one BIP: $0.0001

Serial Correlation and Bid-Ask Bounce

• Result: Bid-Ask spread introduces negative lag-1 serial correlation in an asset return

• Intuition comes from Roll’s (1984) model

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

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

\[ S = P_{\text{Ask}} - P_{\text{Bid}} \]

\[ I_t = \begin{cases} 1 \text{ with probability } 0.5 \\ -1 \text{ with probability } 0.5 \end{cases} \]
### Descriptive Analysis: Price Change

<table>
<thead>
<tr>
<th>MSFT Price Changes</th>
<th>i-th Trade</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i-1)th Trade</td>
<td>“+”</td>
</tr>
<tr>
<td>“+”</td>
<td>787</td>
</tr>
<tr>
<td>“0”</td>
<td>8449</td>
</tr>
<tr>
<td>“-”</td>
<td>7630</td>
</tr>
</tbody>
</table>

### Descriptive Analysis: Spread

- Histogram of MSFT Spread in Ticks
- Histogram of USD/EUR Spread in Ticks

Copyright ©2007
Eric Zivot. All Rights Reserved.
• 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

ACF of Number of Trades in 5-min Intervals:
MSFT (lags up to 3 days)
Diurnal Effects in Duration: MSFT Transactions

ACF of 5-min Mean Durations:
MSFT (lags up to 3 days)

5-min Mean Durations: MSFT
(Averaging across 11 trading days)

Intraday Trading Sessions for 24 Hour FX Market

<table>
<thead>
<tr>
<th></th>
<th>Asian</th>
<th>European</th>
<th>American</th>
<th>Post-American</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours in GMT</td>
<td>22:00 -</td>
<td>06:00 -</td>
<td>12:00 -</td>
<td>18:00 -</td>
</tr>
<tr>
<td></td>
<td>06:00</td>
<td>12:00</td>
<td>18:00</td>
<td>22:00</td>
</tr>
</tbody>
</table>
Diurnal Effects in Quote Activity: USD/EUR

ACF of Number of Quotes in 5-min Intervals:
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)