

# Financial Econometrics and Volatility Models

## Introduction to High Frequency Data

Eric Zivot

May 17, 2010

## Lecture Outline

- Introduction and Motivation
- High Frequency Data Sources
- Challenges to Statistical Modeling
- Market Microstructure Effects: Bid-Ask Bounce
- Nonsynchronous Trading
- Stylized Facts of High Frequency Data

## Reading

- APDVP, chapter 12.
- Yan, B. and Zivot, E. (2003). "Analysis of High-Frequency Financial Data with S-PLUS".

## Introduction and Motivation

What is high frequency data?

- Ten years ago is was daily data
  - Large data sets consisted of 1000s of stocks over 20-30 years (e.g. Center for Research in Security Prices (CRSP) data.
  - 5 - 10 million observations, 100MB compressed data

- Now it is intra-day tick-by-tick or transaction level data on prices, quotes, volume, order book
  - Large data sets consist of 1000s of stocks over 10-15 years (e.g. New York Stock Exchange (NYSE) TAQ data)
  - Trillions or Gazillions of observations, 100s of terabytes of compressed data

## Academic Research Topics

- Market microstructure theory
  - Price discovery and market quality
  - Modeling and estimating liquidity
  - Strategic behavior of market participants
  - Modeling real-time dynamics of trading process

- Event studies
- Estimation and testing of continuous-time models
- Volatility modeling, estimation and forecasting

## Finance Industry Applications

- Short-term trading
  - Pairs trading
  - Arbitrage strategies
  - Event analysis
- Transaction cost and price impact modeling
  - Order execution
  - Market making



- Derivatives pricing
  - Continuous-time models
  - Volatility estimation
  
- Risk Management

## Sources of Historical High Frequency Data

- Equity - NYSE TAQ
- FX - Olsen & Associates
- Options - Berkeley Options Database
- Fixed Income - GovPX

## Commercial Data Resuppliers and Sources of Real-time Data

- Wharton Data Services (WRDS)
- Bloomberg, Reuters
- OneMarketData OneTick Database ([www.onetick.com](http://www.onetick.com)): equities, fixed income, futures, options and FX.
- TICKDATA ([www.tickdata.com](http://www.tickdata.com)): equities, fixed income, futures, options and FX.
- Thompson-Reuters FastTick

## **NYSE Trades and Quotes (TAQ) Database**

Released by NYSE and provides intraday information for stocks traded on NYSE, NASDAQ-AMEX and SmallCap issues starting in 1993.

TAQ does not include transaction data that is reported outside of the Consolidated Tape hours of operation.

As of August 2000, those hours are 8:00am to 6:30pm EST.

As of March 4, 2004, the tape opens at 4:00am EST. Trading in NYSE-listed securities between 8:00am – 9:30am by other markets are also not in TAQ.

**Trade information:** All trades, time-stamped to the second, for all stocks traded on NYSE & regional affiliates, and the NASDAQ-AMEX

- Do not know trading parties
- Do not know if trade is buyer or seller initiated

**Quote information:** all best bid-ask quotes posted by specialists (NYSE, AMEX) and by market makers (NASDAQ) for all stocks

## **Olsen & Associates FOREX Databases**

Company founded by Richard Olsen

- Commercial providers of high quality intra-day foreign exchange data
- Research institute for the analysis of high frequency data

Sponsored three international conferences on the analysis of high frequency time series

- Made available three historical data sets - used in my empirical papers

## Olsen & Associates FOREX Databases

- Indicative (non-binding) dealer quotes on spot exchange rates for a wide assortment of currency pairs published over the Reuters network
- 24 hour market
- No transaction or volume information
- Bid/Ask quotes by dealer/institution
- Data are pre-filtered using proprietary data cleaning technology (magic Olsen filter)

## Challenges to Statistical Modeling

- Huge number of observations
- Dirty data
- Irregularly spaced observations
- Multiple observations with same time stamp
- Heavy-tailed return distributions



- Long memory behavior
- Strong intra-day and intra-week periodicities
- Variables move in discrete increments
- Data for multiple assets seldom occur at the same time

## Market Microstructure Effects: Bid-Ask Bounce

- *Liquidity* is the ability to buy or sell significant quantities of a security quickly, anonymously, and with relatively little price impact.
- To provide liquidity, *market makers* are granted monopoly rights by an exchange to buy stocks from the public at a *bid price*,  $P_b$ , and to sell stocks to the public at an *ask price*,  $P_a$ .
- The bid-ask spread is the difference  $P_a - P_b$
- Economic sources for the bid-ask spread: (1) *order processing* costs; (2) *inventory* costs; (3) *adverse-selection* costs

## The Bid-Ask Spread and Negative Intra-day Return Serial Correlation

Roll's (1984) Model

$$\begin{aligned}P_t &= P_t^* + I_t \cdot \frac{S}{2} \\P_t^* &= \text{fundamental value independent of } S \\S &= P_a - P_b \\I_t &= \begin{cases} 1 & \text{with probability 0.5} \\ -1 & \text{with probability 0.5} \end{cases} \text{ and } iid\end{aligned}$$

Note

$$\begin{aligned}E[I_t] &= 0, \text{ var}(I_t) = E[I_t^2] = 1 \\E[\Delta I_t] &= 0, \text{ var}(\Delta I_t) = 2 \\cov(\Delta I_t, \Delta I_{t-1}) &= E[\Delta I_t \cdot \Delta I_{t-1}] \\&= -E[I_{t-1}^2] = -1\end{aligned}$$

Assume  $P_t^*$  is constant. Then

$$\begin{aligned}\Delta P_t &= \Delta I_t \cdot \frac{S}{2} \\ E[\Delta P_t] &= 0 \\ \text{var}(\Delta P_t) &= \frac{S^2}{4} \text{var}(\Delta I_t) = \frac{S^2}{2} \\ \text{cov}(\Delta P_t, \Delta P_{t-1}) &= \frac{S^2}{4} E[\Delta I_t \cdot \Delta I_{t-1}] = -\frac{S^2}{4} \\ \text{cov}(\Delta P_t, \Delta P_{t-j}) &= 0, \quad j > 1\end{aligned}$$

Therefore,

$$\begin{aligned}\text{corr}(\Delta P_t, \Delta P_{t-1}) &= \frac{-S^2/4}{S^2/2} = -0.5 \\ \text{corr}(\Delta P_t, \Delta P_{t-j}) &= 0\end{aligned}$$

Assume that  $P_t^*$  follows a random walk

$$\begin{aligned}\Delta P_t^* &= \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2) \\ \text{cov}(\varepsilon_t, I_t) &= 0\end{aligned}$$

Then

$$\begin{aligned}\Delta P_t &= \varepsilon_t + \Delta I_t \cdot \frac{S}{2} \\ \text{var}(\Delta P_t) &= \sigma^2 + \frac{S^2}{2} \\ \text{cov}(\Delta P_t, \Delta P_{t-1}) &= -\frac{S^2}{4}\end{aligned}$$

and so

$$\text{corr}(\Delta P_t, \Delta P_{t-1}) = \frac{-S^2/4}{\sigma^2 + S^2/2} \leq 0$$

## Nonsynchronous Trading

*Issue:* Prices are often recorded at regular intervals (e.g. daily closing price) but not all assets trade at the same time. Some implications are

- Lag 1 cross correlation between stock returns
- Lag 1 serial correlation in portfolio returns
- Possible negative serial correlation in individual stock return

## Intuition for First Two Points

Consider stocks A and B such that

$$r_{it} = \mu_i + \beta_i f_t + \varepsilon_{it}$$

$f_t$  = common market news component

$\varepsilon_{it}$  = firm specific component

$\varepsilon_{At}$  and  $\varepsilon_{Bt}$  are independent

A trades more frequently than B

## Implication

- Common news will generally influence  $r_{At}$  prior  $r_{Bt} \Rightarrow r_{At}$  will appear to lead  $r_{Bt}$