Financial Data

Examples from Yahoo Finance

→ Daily fluctuations in price :
Open - High - Low - Close - Volume

SPY , GLD, ^GSPC = SPX , ^DJI

Spyder   gold   S&P 500   Dow Jones Ind.

Markets : NYSE (stocks)
NASDAQ (stocks)
CME (Commodities)
T-Bond Auctions, CBOT

ETFs
→ Exchange Traded Funds, trade like stocks

→ Daily data vs "tick" data
(Petabytes of tick data)
↳ transaction-by-transaction

Data availability improving with time,

esp. since 1984 when computer networks appeared

Price Scales – Global Markets

What are the reference units?

prices in \$, ¥, RMB, EUR, AUD etc

→ currency (transaction / translation) costs

Currencies change value because of:

– inflation/deflation

esp. since 1984 when computer networks appeared

Price Scales – Global Markets

What are the reference units?

prices in \$, ¥, RMB, EUR, AUD etc

→ currency (transaction / translation) costs

Currencies change value because of:

- inflation/deflation
- economic growth or recession
- fluctuations in global currency markets → leads to arbitrage such as "yen carry trade"
- trade imbalances
- capital flow changes
- increase/decrease in debt/GDP ratios

all these are random variables

Krugman pieces on borrowing in own currency

So stock prices can reflect changes in valuation of underlying currencies

- Define $Y(t)$ as price of a financial asset at time t.
- Common choices for financial random variables include:

(i) $Z(t) = Y(t+\Delta t) - Y(t)$

Problem here is that over time, all markets have increased in value due to inflation, so changes may have no real significance

(ii) Deflated or discounted prices

$$Z_D(t) = [Y(t+\Delta t) - Y(t)]\, D(t)$$

→ $D(t)$ = deflation or discounting factor

Problem here is that $D(t)$'s are difficult to compute, and are themselves random variables. Also, <u>no</u> unique choice of $D(t)$

Fractional change

(iii) Returns (or in %)

Best approach & most common

$$R(t) \equiv \frac{Y(t+\Delta t) - Y(t)}{Y(t)} = \frac{Z(t)}{Y(t)}$$

Has an advantage that inflation is automatically discounted, as long as $\boxed{\Delta t \to 0}$

An important quantity is the idea of Drawdowns

$$\text{drawdown } W(t) = -\min(R(t))$$

over a long time T.

(iv) another possible rand. var. is differences of natural logarithms

$$S(t) \equiv \ln Y(t+\Delta t) - \ln Y(t)$$

$$= \ln\left(\frac{Y(t+\Delta t)}{Y(t)}\right)$$

for small Δt, $S(t) \approx R(t)$

Note that this is a nonlinear transformation, and such transformations can lead to unwanted consequences

High Frequency Trading

HFT uses $R(t)$ with $\Delta t \to 0$ (μsec, for example). Speed of light is even important here. HFT guys take care to locate their servers as close as possible to the exchange servers where trades are paired up & made

Most commonly studied r.v.'s are

$S(t)$ and $R(t)$

Time Scales in Financial Data

What is appropriate choice of time?

- physical time
- trading (or market) time
- number of transactions ("natural time")

Indisputable choice is not possible

A <u>Physical time</u> - now, trading is global, 24/7. Futures are traded in London, Singapore, Hong Kong, Shanghai, Tokyo, etc.
→ difficult to know how to incorporate physical time into trading models in these circumstances

B. <u>Trading Time</u> (on a given exchange)
Elapsed time on a given exchange during open market hours. Most trades tend to take place on an exchange just after it opens, and just before it closes

Trading time is most common choice in models of trading dynamics

<u>Problems exist, however</u>:

(i) Information affecting dynamics of prices can be released during periods when markets are closed

(ii) In HFT, overnight price changes are treated as short-term price changes

(iii) market activity is often assumed to be uniform during market hours

⇑

Definitely a <u>bad</u> assumption - there is a daily cycle in trading operations, volume is highest at open & close (typically)

Foreign Exchange Markets - 3 different peaks in activity corresponding to open hours of US, Asia, Europe (London)

Results of analyses are sensitive to the time-period chosen, so one needs to be careful about definitions.

Stationary Processes & Time Correlations

Here we consider:

- definitions of stationary stoch. processes
- short & long range correlations
- empirical studies of financial data

If stoch variables are independent, "stationary" implies $x(t) \in iid$ //

Expectation or mean:

$$E\{x(t)\} \equiv \int_{-\infty}^{\infty} x\, f(x,t)\, dx \qquad n=1$$

where $f(x,t)$ gives prob of observing random value x at time t.

$$E\{x_1(t_1)\, x_2(t_2)\} \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2\, f(x_1,x_2;t_1,t_2)\, dx_1\, dx_2$$

→ Autocorrelation

$f(x_1,x_2;t_1,t_2)$ is joint prob density that x_1 is observed at t_1, and x_2 at t_2

Wikipedia

Autocorrelation:

$$R(t,t') = \frac{E[(x(t)-\mu_t)(x(t')-\mu_{t'})]}{\sigma_t \sigma_{t'}}$$

This is the definition common in statistics

If R depends only on $\tau = t-t'$ (statistically stationary) then

$$R = R(\tau) = \frac{E[(x(t)-\mu)(x(t+\tau)-\mu)]}{\sigma^2}$$

$$R(\tau) = R(-\tau) \quad \text{(even function)}$$

In signal processing, however, the term "autocorrelation" means something different:

$$R(\tau) = \int_{-\infty}^{\infty} x(t+\tau)\,\bar{x}(t)\,dt$$

$$= \int_{-\infty}^{\infty} x(t)\,\bar{x}(t-\tau)\,dt$$

$\bar{x}$ = complex conjugate