Definition of Prediction Task

Are financial data forecastable???

First, we have to define the prediction task:

Let $P_t$ be a random variable defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, which is also called stochastic basis, where $\Omega$ is space of outcomes, $\mathcal{F}$ is $\sigma$-algebra of the subsets of $\Omega$, and $\mathbb{P}$ is a probability measure on $\mathcal{F}$ and $\{\mathcal{F}_t\}$ is the usual filtration. A conditional probability $\mathbb{P}[P_{t+1} | \mathcal{F}_t]$ is conditional probability of the set $P_t$ being evaluated with the information available in the $\sigma$-algebra $\mathcal{F}$.

Now, let's assume economic agent's utility functions:

$$u(W_{t+h}) = g\left(P_{t+h}, \hat{\gamma}(P_{t+h})\right),$$

where agent's utility $u(.)$ depends on $P$ in time $t + h$, decision function $\gamma(.)$ and forecast $\hat{P}$ with forecasting horizon $h \geq 1$, $w$ as a reward variable.
Say that in time horizon 1, agent’s utility depends on the realization of $p_{t+1}$, and accuracy of it’s forecast, $\hat{p}_{t+1}$. Thus forecasting is defined as major factor of a decision rule.

Let $E\left[P_{t+h} \mid \mathcal{F}_t\right] = \hat{P}_{t+h} = h(X_t, \theta)$ be an expectation of $P_{t+h}$ conditional on the information set $\mathcal{F}_t$, where $\theta \in \Theta$ is unknown vector of parameters, where $\Theta \subseteq \mathbb{R}^k$ is compact and observable at time $t$, $X_t$ is an $\mathcal{F}_t$-measurable vector of variables.

$X_t$ can include any information - past, exogeneous, indicators, and also can miss some crucial information, which will have impact on forecast !!!!

Utility will be negatively correlated with forecast error:

$$e_{t+h} = p_{t+h} - \hat{p}_{t+h}$$

max $u(W_{t+h})$ implies optimal forecast $P_{t+h} = \arg\min_{\hat{\theta} \in \Theta} E[\mathcal{L}(P_{t+h}, X, \theta, \alpha) \mid \mathcal{F}_t]$, where $\alpha$ is degree of assymetry, $\mathcal{L}(.)$ is loss function, which might depend only on forecast errors, so it will take form of $\mathcal{L}(e_{t+h} \mid \mathcal{F}_t)$.

---

**Random Walk**

- **Random Walk Hypothesis**

  A time series $\{p_t\}$ is a random walk process if:

  $$p_t = \mu + p_{t-1} + e_t,$$

  where $p_t = \ln(P_t)$, thus $r_t = \mu + e_t$, where $\mu$ is drift, and we distinguish between random walk with or without drift. Random walk without drift is special case of AR(1) process, with coefficient 1, thus it is not stationary. Hence, random walk model is unit-root nonstationary time series.

  We distinguish between 3 forms of Random Walk:

  RW1: $e_t$ is independent and identically distributed $\sim$ iid, or $N \sim (0,\sigma^2)$, with conditional mean
and variance $P_0 + \mu t$ and $\sigma^2 t$

**RW2:** $e_t$ is independent, thus allows for heteroskedasticity, Test using filter rules, technical analysis

**RW3:** $e_t$ is uncorrelated, thus allows for dependence in higher moments Test using autocorrelations, variance ratios, long horizon regressions.

RW is sufficient but not necessary condition for EHM. RW is nonstationary, conditional mean and variance are linear in time for all RW forms.

- **Example of Random Walk with drift $\mu$ and $\epsilon \in N(0, \sigma)$:**

  Initial settings $\mu=0$ (RW without drift), and $N(0,1)$ - Gaussian White Noise

![Random Walk Example](QF_I_Lecture3.nb)
When Time is Continuous ...

Random Walk is discrete-time, let's make it continuous.

- **Simple Problem**

Suppose investor makes a deposit of P $ into a cash account that pays interest rate \( r \) 100% p.a., compounded continuously.

**HOW DOES THE ACCOUNT EVOLVE AS FUNCTION OF TIME T ???**

\[ y'(t) = r \cdot y(t) \]

or equally \( \frac{dy}{dt} = f(t, y); \ y(0) = y_0 \)

amount of change in the account balance, or rate of change can be expressed by ODE - Ordinary Differential Equation

initial condition is \( y(0) = P \)

SOLUTION to this simple problem is:

\[ y[t] \rightarrow e^{rt} \cdot P \]
Brownian Motion

When we consider ODE from previous example, and add a random effect, we get SDE - Stochastic Differential Equation (we use SDE's to model stock price)

* Stochastic, from the Greek "stochos" or "aim, guess", means of, relating to, or characterized by conjecture and randomness. A stochastic process is one whose behavior is non-deterministic in that a state does not fully determine its next state.

\[ d p(t) = a(t, p(t)) \, dt + \sigma(t, p(t)) \, dB(t), \]

where \( dB(t) \) is "differential" of Brownian motion \( B(t) \), and \( \frac{\sigma(t, p(t))}{p(t)} \) is volatility, \( a(t, p(t)) \) is deterministic trend, or drift, and \( dB(t) \) is iid with zero mean and variance \( dt: dB(t) \sim N(0, \sqrt{dt}) \)

Brownian motion is thus continuous time random walk, which is most important stochastic process.
- **Application to Stock Price**

the dynamics of the price process can be expressed as:

\[ \frac{\Delta P}{P} = \mu \Delta t + \sigma \Delta W, \]

where \( \mu \Delta t \) is called deterministic component of price (trend, drift) and \( \sigma \Delta W \) random component, or volatility (W is Wiener process).

If \( \Delta t \to 0 \), we get a continuous-time case:

\[ \frac{dP}{P} = \mu dt + \sigma dW. \]

If we apply Ito's lemma to this SDE, we will be able to derive corestones of derivative pricing, like Black-Scholes.

(see next lectures, now let's get back to RW)

- **Example of Brownian Motion with drift \( \mu \) and \( \epsilon \in N(0, \sigma) \):**
Fundamental properties of the Brownian Motion

\[ E_{s \to B(t)} B(t) = B(s) \]
\[ E_{s \to B(t)} (B(t) - B(s))^2 = t - s, \text{ for } t \geq s. \]
Tests of Random Walks 1 and 2

- Sequences and Reversals (RW1)

\[ p_t = \mu + p_{t-1} + e_t, \quad e_t \sim N(0, \sigma^2) \]

Let's denote \( I_t = \begin{cases} 
1 & \text{if } r_t = p_t - p_{t-1} > 0 \\
0 & \text{if } r_t = p_t - p_{t-1} \leq 0
\end{cases} \).

Cowles and Jones (1937) test compares frequency of sequences (consecutive returns of same sign) and reversals (opposite). Cowles-Jones ratio is defined as:

\[
\hat{CJ} = \frac{N_1}{N_0} = \frac{\sum_{t=1}^{n} (I_t I_{t+1} + (1-I_t)(1-I_{t+1}))}{n-N_0} = \frac{\hat{\pi}_s}{1-\pi_s} \rightarrow \frac{\pi_s}{1-\pi_s} = CJ = 1,
\]

where \( \pi_s \) is probability of a sequence, and \( e_t \) is assumed to be symmetric, thus \( P[r_t > 0] = P[r_t \leq 0] = \frac{1}{2} \).

But most stocks exceeded the 1, because most stocks has some drift. So we can reconsider the case:

\[ I_t = \begin{cases} 
1 & \text{with probability } \pi \\
0 & \text{with probability } 1 - \pi
\end{cases} \]

and \( \pi = P[r_t > 0] = \phi\left(\frac{\mu}{\sigma}\right) \). Thus for drift \( \mu > 0, \pi > \frac{1}{2} \) and vice versa, and Cowles-JOnes ratio can be defined as:

\[
\hat{CJ} = \frac{\pi^2 + (1-\pi)^2}{2\pi(1-\pi)} \geq 1.
\]

it can be easily shown, that \( \hat{CJ} \sim N\left(\frac{\pi_s}{1-\pi_s}, \frac{\pi_s(1-\pi_s) + 2(\pi^3 + (1-\pi)^3 - \pi^2)}{n(1-\pi_s)^3}\right) \)
Example

Cowles-Jones Statistics

- Drift: 0.086
- Volatility: 0.311

\[ CJ = 1.0305 \]

Runs (RW1)

Another common test for RW1 is **runs test**, in which the number of sequences of consecutive positive and negative returns (runs) is tabulated, and compared against sampling distribution.

Let's consider \( I_t \in \{1, 0, 0, 1, 1, 1, 0, 1, 0, 0\} \). We have 3 runs of 1 with lengths (1,3,1) and 3 runs of 0 with length(2,1,2).

Idea is similar, and it can be shown (i.e. in CLM chapter 2), that the distribution of the number of runs converges to a normal distribution asymptotically when:

\[
z_i \sim N\left(0, \pi_i(1 - \pi_i) - 3 \pi_i^2 (1 - \pi_i)^2 \mu \right)
\]

Tests for RW 2

RW2 may be tested using filter rules (buy if price moves by + x %, short(sell) if price moves by - x %) or Technical Analysis (we will discuss in detail in one of further Lectures)
Rests of RW 3

- **Autocorrelation**

Another group of tests is based on serial autocorrelation. Under the RW hypothesis, $e_t$'s are uncorrelated.

$H_0$: autocorrelation at various lags is equal to zero
$H_1$: autocorrelation at various lags is not equal to zero

The correlation coefficient between two random variables $X$ and $Y$ is defined as:

$$
\rho_{x,y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sqrt{E(X-\mu_x)^2 \cdot E(Y-\mu_y)^2}},
$$

where $\mu_x, \mu_y$ are the mean of $X$ and $Y$ resp., and it is assumed that variances exist. This coefficient measures the strength of linear dependence between $X$ and $Y$, while $-1 \leq \rho_{x,y} \leq 1$, and $\rho_{x,y} = \rho_{y,x}$. Thus two random variables are uncorrelated (no dependencies) if $\rho_{x,y} = 0$

Lag – 1 sample autocorrelation of series $x_t$ is defined as:

$$
\rho_l = \frac{\sum_{t=l+1}^{T} (x_t - \bar{x})(x_{t-l} - \bar{x})}{\sqrt{\sum_{t=1}^{T} (x_t - \bar{x})^2}}, \quad 0 \leq l < T - 1
$$
Example of ACF’s and PACF’s of random White Noise series

- Portmanteau Test

In finance, we often need to test jointly that several autocorrelations of $r_i$ are zero. For this, we can use Ljung and Box Q statistics: $Q(m) = T(T+2) \sum_{i=1}^{m} \frac{\hat{\rho}_i^2}{T-i}$. We test null hypothesis $H_0 : \rho_1 = \ldots = \rho_m = 0$ against alternative, $H_1 : \rho_i \neq 0$.
Stationarity

Definition
A time series \( \{r_t\} \) is said to be **strictly stationary** if the joint distribution of \( \{r_{t_1}, ..., r_{t_k}\} \) is identical to \( \{r_{t_1+\tau}, ..., r_{t_k+\tau}\} \) for all \( \tau \), where \( k \) is arbitrary positive integer and \( (t_1, ..., t_k) \) are positive integers. A time series \( \{r_t\} \) is said to be **weakly stationary** if both mean of \( r_t \) and covariance between \( r_t \) and \( r_{t-l} \) are time-invariant - \( E(r_t) = \mu \) and \( \text{Cov}(r_t, r_{t-l}) = \gamma_l \).

In other words, stationarity requires distribution of time series to be constant under time shift, weak stationarity, which is assumed more often requires only fluctuation with constant variation around constant level.

- **Why we need to care about stationarity?**
  - non-stationary time series influence its behavior and properties (persistence of shocks will be infinite)
  - spurious regression - regression of 2 variables trending over time could have high \( R^2 \) even if they are unrelated
  - it can be proved, that for non-stationary series, the standard assumptions, and thus testing is not valid.

- **Stochastic non-stationarity**
  Let's consider following process:
  \[
y_t = \mu + \phi \ y_{t-1} + u_t
  \]
  with \( \phi > 1 \) the system is explosive, shocks are infinite and become more influential in time
  \( \theta = 1 \) shocks persist in the system and never die
  \( \theta < 1 \) shocks gradually die away
- Detrending stochastic non-stationary series

Simplest way is to difference the data

\[ \Delta y_t = y_t - y_{t-1} \]

Using returns instead of prices in financial time series ensures this property

Unit Root Testing

Dickey-Fuller test

\[ H_0 : \phi = 1 \text{ in } y_t = \phi y_{t-1} + u_t, \text{ against one-sided alternative } H_A : \phi < 1 \]

\[ H_0 : \text{series contains a unit root} \]
\[ H_A : \text{series is stationary} \]

alternatively, we can use regression \[ \Delta y_t = \psi y_{t-1} + u_t, \] where we will test if \( \psi = 0 \) (\( \phi - 1 = \psi \))
Augmented Dickey Fuller Test (ADF)

if we have unmodeled autocorrelation in the dependent variable of the regression, we need to "augment" the test using p lags of the dependent variable:

\[ \Delta y_t = \psi y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + u_t \]

---

**Example on real world data**

Open Eviews workbook

---

**Homework #2**

Deadline: Monday 22.10.2007, 5 pm

*Homework may be returned in class, or sent via email to barunik@utia.cas.cz*

:] Exercise 1 [:

Choose 2 Stocks (if you do not know which ones, use the previous HW ;) ), and compute Auto-correlation Functions, and Ljung Box-Q statistics up to lag 30. Decide if the statistics are significant on any significance levels, discuss the implications * Please include your computations program with your results.