Introduction to Algorithmic Trading Strategies
Lecture 6

Technical Analysis: Linear Trading Rules

Haksun Li
haksun.li@numericalmethod.com
www.numericalmethod.com

Outline

- Moving average crossover
- The generalized linear trading rule
- P&Ls for different returns generating processes
- Time series modeling

References

- Emmanual Acar, Stephen Satchell. Chapters 4, 5 & 6, Advanced Trading Rules, Second Edition. Butterworth-Heinemann; 2nd edition. June 19, 2002.

Assumptions of Technical Analysis

- History repeats itself.
- Patterns exist.

Does MA Make Money?

- Brock, Lakonishok and LeBaron (1992) find that a subclass of the moving-average rule does produce statistically significant average returns in US equities.
- Levich and Thomas (1993) find that a subclass of the moving-average rule does produce statistically significant average returns in FX.

Moving Average Crossover

- Two moving averages: slow (n) and fast (m).
- Monitor the crossovers.
- $B_t = \left(\frac{1}{m}\sum_{j=0}^{m-1} P_{t-j}\right) - \left(\frac{1}{n}\sum_{j=0}^{n-1} P_{t-j}\right), n > m$
- Long when $B_t \geq 0$.
- Short when $B_t < 0$.

How to Choose n and m?

- It is an art, not a science (so far).
- They should be related to the length of market cycles.
- Different assets have different m and n.
- Popular choices:
  - (150, 1)
  - (200, 1)

AMA(n , 1)

- $B_t \geq 0 \text{ iff } P_t \geq \left(\frac{1}{n}\sum_{j=0}^{n-1} P_{t-j}\right)$
- $B_t < 0 \text{ iff } P_t < \left(\frac{1}{n}\sum_{j=0}^{n-1} P_{t-j}\right)$

GMA(n , 1)

- $B_t \geq 0$ iff $P_t \geq \left(\prod_{j=0}^{n-1} P_{t-j}\right)^{\frac{1}{n}}$
  - $R_t \geq -\sum_{j=1}^{n-2} \frac{n-(j+1)}{n-1} R_{t-j}$ (by taking log)
- $B_t < 0$ iff $P_t < \left(\prod_{j=0}^{n-1} P_{t-j}\right)^{\frac{1}{n}}$
  - $R_t < -\sum_{j=1}^{n-2} \frac{n-(j+1)}{n-1} R_{t-j}$ (by taking log)

What is n?

- $n = 2$
- $n = \infty$

Acar Framework

- Acar (1993): to investigate the probability distribution of realized returns from a trading rule, we need
  - the explicit specification of the trading rule
  - the underlying stochastic process for asset returns
  - the particular return concept involved

Empirical Properties of Financial Time Series

- Asymmetry
- Fat tails

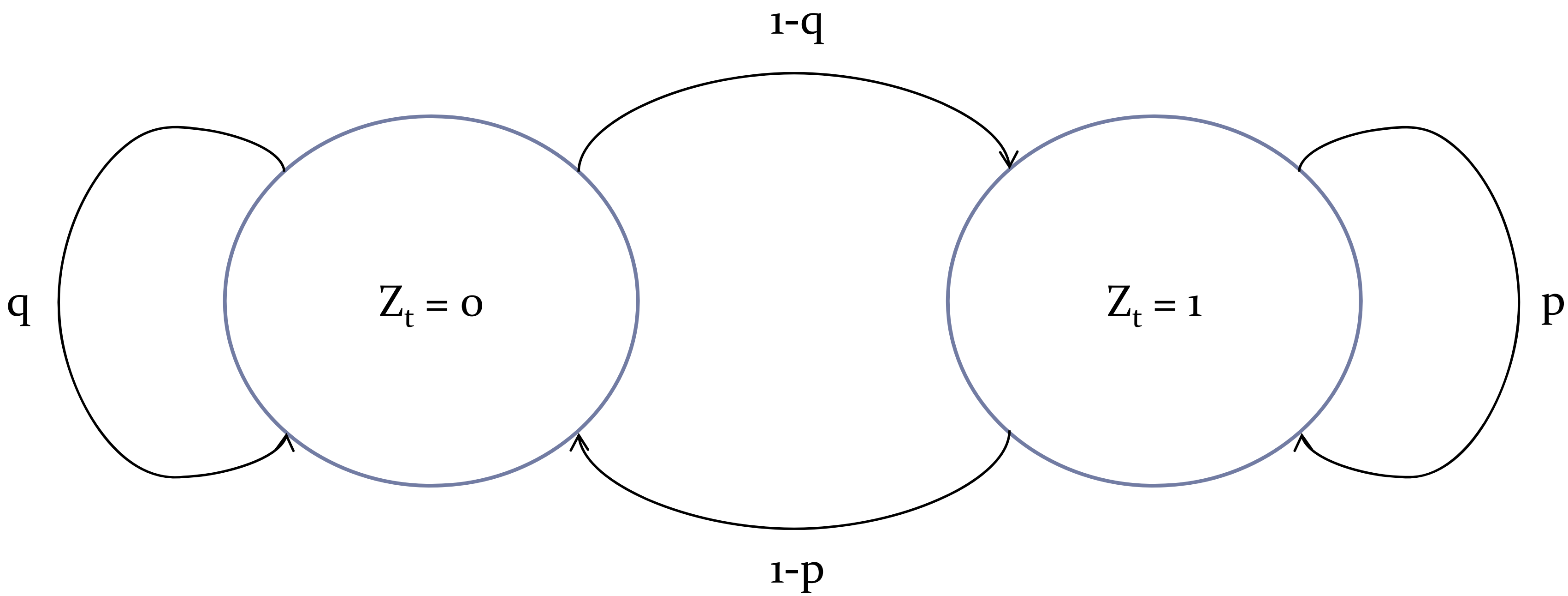
Knight-Satchell-Tran Intuition

- Stock returns staying going up (down) depends on
  - the realizations of positive (negative) shocks
  - the persistence of these shocks
- Shocks are modeled by gamma processes.
- Persistence is modeled by a Markov switching process.

Knight-Satchell-Tran Process

- $R_t = \mu_l + Z_t \varepsilon_t - (1 - Z_t)\delta_t$
  - μ_l: long term mean of returns, e.g., 0
  - ε_t, δ_t: positive and negative shocks, non-negative, i.i.d
- $f_\varepsilon(x) = \dfrac{{\lambda_1}^{\alpha_1} x^{\alpha_1 - 1}}{\Gamma(\alpha_1)} e^{-\lambda_1 x}$
- $f_\delta(x) = \dfrac{{\lambda_2}^{\alpha_2} x^{\alpha_2 - 1}}{\Gamma(\alpha_2)} e^{-\lambda_2 x}$

Knight-Satchell-Tran Z_t



Stationary State

- $\Pi = \frac{1-q}{2-p-q}$
- $R_t = \mu_l + \varepsilon_t \geq \mu_l$, with probability Π
- $R_t = \mu_l - \delta_t < \mu_l$, with probability $1 - \Pi$

GMA(2, 1)

- Assume the long term mean is 0, $\mu_l = 0$.
- $(B_t \geq 0) \equiv (R_t \geq 0) \equiv (Z_t = 1)$
- $(B_t < 0) \equiv (R_t < 0) \equiv (Z_t = 0)$

Naïve MA Trading Rule

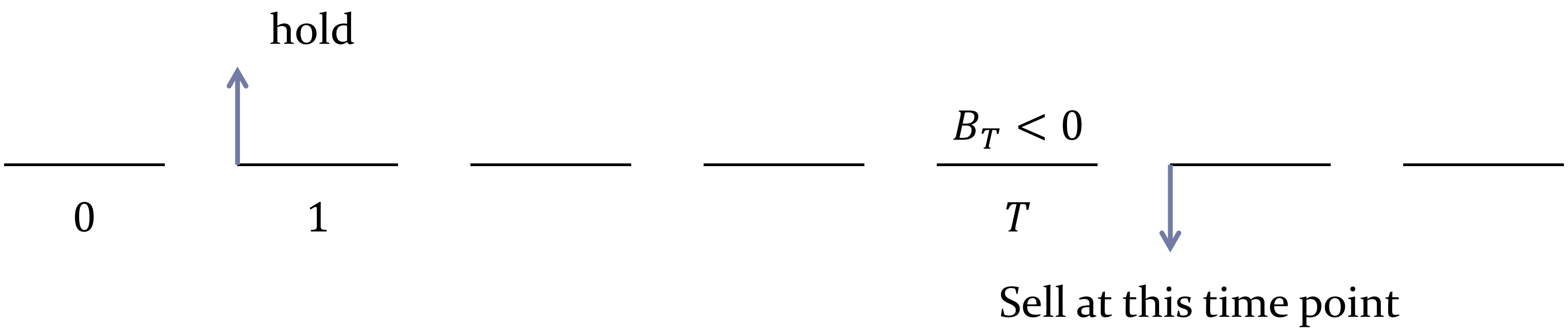
- Buy when the asset return in the present period is positive.
- Sell when the asset return in the present period is negative.

Naïve MA Conditions

- The expected value of the positive shocks to asset return >> the expected value of negative shocks.
- The positive shocks persistency >> that of negative shocks.

T Period Returns

- $RR_T = \sum_{t=1}^{T} R_t \times I_{\{B_{t-1} \geq 0\}}$



Holding Time Distribution

- $P(N = T)$
- $= P(B_T < 0, B_{T-1} \geq 0, \dots, B_1 \geq 0, B_0 \geq 0)$
- $= P(Z_T = 0, Z_{T-1} = 1, \dots, Z_1 = 1, Z_0 = 1)$
- $= P(Z_T = 0, Z_{T-1} = 1, \dots, Z_1 = 1 | Z_0 = 1) P(Z_0 = 1)$
- $= \begin{cases} \Pi p^{T-1}(1-p), & \mathrm{T} \geq 1 \\ 1 - \Pi, & \mathrm{T}=0 \end{cases}$

Conditional Returns Distribution (1)

- $\Phi_{RR_T|N=T}(s) = \mathrm{E}\left[e^{\left\{i\left[\sum_{t=1}^{T} R_t \times I_{\{B_{t-1} \geq 0\}}\right]s\right\}} | N = T\right]$
- $= \mathrm{E}\left[e^{\left\{i\left[\sum_{t=1}^{T} R_t \times I_{\{B_{t-1} \geq 0\}}\right]s\right\}} | B_T < 0, B_{T-1} \geq 0, \ldots, B_0 \geq 0\right]$
- $= \mathrm{E}\left[e^{\left\{i\left[\sum_{t=1}^{T} R_t\right]s\right\}} | Z_T = 0, Z_{T-1} = 1, \ldots, Z_1 = 1\right]$
- $= \mathrm{E}\left[e^{\{i[\varepsilon_1 + \cdots + \varepsilon_{T-1} - \delta_T]s\}}\right]$
- $= \begin{cases} {\Phi_\varepsilon}^{T-1}(s)\Phi_\delta(-s), & \mathrm{T} \geq 1 \\ \Phi_\delta(-s), & \mathrm{T} = 0 \end{cases}$

Unconditional Returns Distribution (2)

- $\Phi_{RR_T}(s) = \sum_{T=0}^{\infty} \mathrm{E}\left[e^{\left\{i\left[\sum_{t=1}^{T} R_t \times I_{\{B_{t-1} \geq 0\}}\right]s\right\}} \middle| N = T\right] P(N = T)$
- $= \sum_{T=1}^{\infty} \Pi p^{T-1}(1-p){\Phi_{\varepsilon}}^{T-1}(s)\Phi_{\delta}(-s) + (1-\Pi)\Phi_{\delta}(-s)$
- $= (1-\Pi)\Phi_{\delta}(-s) + \Pi(1-p)\frac{\Phi_{\delta}(-s)}{1-p\Phi_{\varepsilon}(s)}$

Long-Only Returns Distribution

- $\Phi_{RR_T}(s|R_0 \geq 0) = \frac{(1-p)\Phi_\delta(-s)}{1-p\Phi_\varepsilon(s)}$
- Proof: make $P(Z_0 = 1) = \Pi = 1$

I.I.D Returns Distribution

- $\Phi_{RR_T}(s) = \frac{q\Phi_\delta(-s)[1+p-p\Phi_\varepsilon(s)]}{1-p\Phi_\varepsilon(s)}$
- Proof:
  - $p + q = 1$
  - make $\Pi = \frac{1-q}{2-p-q} = 1 - q = p$

Expected Returns

- $\mathrm{E}(RR_T) = -i\Phi_{RR_T}{}'(0)$
- $= \frac{1}{1-p}\{\Pi p\mu_\varepsilon - (1-p)\mu_\delta\}$
- When is the expected return positive?
  - $\mu_\varepsilon \geq \frac{1-p}{\Pi p}\mu_\delta$, shock impact
  - $\mu_\varepsilon \gg \mu_\delta$, shock impact
  - $\Pi p \geq 1-p$, if $\mu_\varepsilon \approx \mu_\delta$, persistence

GMA(∞,1) Rule

- $P_t \geq \left(\prod_{j=0}^{n-1} P_{t-j}\right)^{\frac{1}{n}}$
- $\ln P_t \geq \frac{1}{n}\sum_{j=0}^{n-1} \ln P_{t-j}$
- $\ln P_t \geq \mu_1$

GMA(∞,1) Returns Process

- $\ln P_t = \mu_l + Z_t \varepsilon_t - (1 - Z_t)\delta_t$
- $R_t = \ln P_t - \ln P_{t-1}$
- $= Z_t \varepsilon_t - Z_{t-1}\varepsilon_{t-1} - (1 - Z_t)\delta_t + (1 - Z_{t-1})\delta_{t-1}$

Returns As a MA(1) Process

- $\mathrm{E}(R_r) = 0$
- $\mathrm{Var}(R_r) = 2[\Pi({\sigma_\varepsilon}^2 + {\mu_\varepsilon}^2) + (1 - \Pi)({\sigma_\delta}^2 + {\mu_\delta}^2)]$
- $\mathrm{E}\left(R_{t-i}R_{t-j}\right) = \begin{cases} -[\Pi({\sigma_\varepsilon}^2 + {\mu_\varepsilon}^2) + (1 - \Pi)({\sigma_\delta}^2 + {\mu_\delta}^2)] \\ 0 \end{cases}$

GMA(∞,1) Expected Returns

- $\Phi_{RR_T}(s) =$
  $(1-\Pi)q[\Phi_\delta(s) + \Phi_\delta(-s)] +$
  $[1-p(1-\Pi)][\Phi_\varepsilon(s) + \Phi_\varepsilon(-s)]$
- $\mathrm{E}(RR_T) = -[1-p(1-\Pi)][\mu_\varepsilon + \mu_\delta]$

MA Using the Whole History

- An investor will always expect to lose money using GMA(∞,1)!
- An investor loses the least amount of money when the return process is a random walk.

Optimal MA Parameters

- So, what are the optimal n and m?

Linear Technical Indicators

- As we shall see, a number of linear technical indicators, including the Moving Average Crossover, are really the “same” *generalized* indicator using different parameters.

The Generalized Linear Trading Rule

- A linear predictor of weighted lagged returns
  - $F_t = \delta + \sum_{j=0}^{t} d_j X_{t-j}$
- The trading rule
  - Long: $B_t = 1$, iff, $F_t > 0$
  - Short: $B_t = -1$, iff, $F_t < 0$
- (Unrealized) rule returns
  - $R_t = B_{t-1} X_t$
    - $R_t = -X_t$ if $B_{t-1} = -1$
    - $R_t = +X_t$ if $B_{t-1} = +1$

Buy And Hold

- $B_t = 1$

Predictor Properties

- Linear
- Autoregressive
- Gaussian, assuming X_t is Gaussian
- If the underlying returns process is linear, F_t yields the best forecasts in the mean squared error sense.

Returns Variance

- $\mathrm{Var}(R_t) = \mathrm{E}\left({R_t}^2\right) - (\mathrm{E}(R_t))^2$
- $= \mathrm{E}\left({B_{t-1}}^2 {X_t}^2\right) - (\mathrm{E}(R_t))^2$
- $= \mathrm{E}\left({X_t}^2\right) - (\mathrm{E}(R_t))^2$
- $= \sigma^2 + \mu^2 - (\mathrm{E}(R_t))^2$

Maximization Objective

- Variance of returns is inversely proportional to expected returns.
- The more profitable the trading rule is, the less risky this will be if risk is measured by volatility of the portfolio.
- Maximizing returns will also maximize returns per unit of risk.

Expected Returns

- $\mathrm{E}(R_t) = \mathrm{E}(B_{t-1}X_t)$
- $= \mathrm{E}(B_{t-1}(\mu + \sigma N))$
- $= \sigma\, \mathrm{E}(B_{t-1}N) + \mu\, \mathrm{E}(B_{t-1})$
- $\mathrm{E}(B_{t-1}) = 1 \times \mathrm{P}(F_{t-1} > 0) + -1 \times \mathrm{P}(F_{t-1} < 0)$
- $= \mathrm{P}(F_{t-1} > 0) - \mathrm{P}(F_{t-1} < 0)$
- $= 1 - 2 \times \mathrm{P}(F_{t-1} < 0)$
- $= 1 - 2 \times \Phi\left(-\frac{\mu_F}{\sigma_F}\right)$

Truncated Bivariate Moments

- Johnston and Kotz, 1972, p.116
- $\mathrm{E}(B_{t-1}N) = \iint_{F_t>0} N - \iint_{F_t<0} N$
- $= \sqrt{\frac{2}{\pi}}\rho e^{-\frac{\mu_F{}^2}{2\sigma_F{}^2}}$
- Correlation:
  - $\rho = \mathrm{Corr}(X_t, F_{t-1})$

Expected Returns As a Weighted Sum

- $\mathrm{E}(R_t) = \sigma\,\mathrm{E}(B_{t-1}N) + \mu\,\mathrm{E}(B_{t-1})$
- $= \sigma\sqrt{\frac{2}{\pi}}\rho e^{-\frac{{\mu_F}^2}{2{\sigma_F}^2}} + \mu\left(1 - 2 \times \Phi\left(-\frac{\mu_F}{\sigma_F}\right)\right)$

a term for volatility

a term for drift

Praetz model, 1976

- Returns as a random walk with drift.
- $\mathrm{E}(R_t) = \mu(1 - 2f)$, f the frequency of short positions
- $\mathrm{Var}(R_t) = \sigma^2$

Comparison with Praetz model

- Random walk implies $\rho = \mathrm{Corr}(X_t, F_{t-1}) = 0$.
- $\mathrm{E}(R_t) = \mu\left(1 - 2 \times \Phi\left(-\frac{\mu_F}{\sigma_F}\right)\right)$

the probability of being short

- $\mathrm{Var}(R_t) = \sigma^2 + \mu^2 - \left\{\mu\left(1 - 2 \times \Phi\left(-\frac{\mu_F}{\sigma_F}\right)\right)\right\}^2$
- $= \sigma^2 + 4\mu^2\,\Phi\left(-\frac{\mu_F}{\sigma_F}\right)\left(1 - \Phi\left(-\frac{\mu_F}{\sigma_F}\right)\right)$

increased variance

Biased Forecast

- A biased (Gaussian) forecast may be suboptimal.
- Assume underlying mean $\mu = 0$.
- Assume forecast mean $\mu_F \neq 0$.
- $\mathrm{E}(R_t) = \sigma\sqrt{\frac{2}{\pi}}\rho e^{-\frac{{\mu_F}^2}{2{\sigma_F}^2}} \leq \sigma\sqrt{\frac{2}{\pi}}\rho$

Maximizing Returns

- Maximizing the correlation between forecast and one-ahead return.
- First order condition:
  - $\frac{\mu_F}{\sigma_F} = \frac{\mu}{\sigma\rho}$

First Order Condition

- Let x $= \frac{\mu_F}{\sigma_F}$
- $E(R_t) = \sigma\sqrt{\frac{2}{\pi}}\rho e^{-\frac{x^2}{2}} + \mu(1 - 2 \times \Phi(-x))$
- $\frac{d\ E(R_t)}{dx} = 0$
  - $\sigma\cancel{\sqrt{\frac{2}{\pi}}}\rho(-x)e^{\cancel{-\frac{x}{2}}} + \mu\cancel{\sqrt{\frac{2}{\pi}}}e^{\cancel{-\frac{x^2}{2}}} = 0$
  - $x = \frac{\mu_F}{\sigma_F} = \frac{\mu}{\sigma\rho}$

Fitting vs. Prediction

- If X_t process is Gaussian, no linear trading rule obtained from a finite history of X_t can generate expected returns over and above F_t.
- Minimizing mean squared error $\neq$ maximizing P&L.
- In general, the relationship between MSE and P&L is highly non-linear (Acar 1993).

Technical Analysis

- Use a finite set of historical prices.
- Aim to maximize profit rather than to minimize mean squared error.
- Claim to be able to capture complex non-linearity.
- Certain rules are ill-defined.

Technical Linear Indicators

- For any technical indicator that generates signals from a finite linear combination of past prices
  - Sell: $B_t = -1$ iff $\sum_{j=0}^{m-1} a_j P_{t-j} < 0$
- There exists an (**almost**) equivalent AR rule.
  - Sell: $\widetilde{B_t} = -1$ iff $\delta + \sum_{j=0}^{m-2} d_j X_{t-j} < 0$
  - $X_t = \ln \frac{P_t}{P_{t-1}}$
  - $\delta = \sum_{j=0}^{m-1} a_j$, $d_j = -\sum_{i=j}^{m-2} a_i$

Conversion Assumption

- $1 - \frac{P_{t-j}}{P_t} \approx \ln \frac{P_t}{P_{t-j}}$
- Monte Carlo simulation:
  - 97% accurate
  - 3% error.

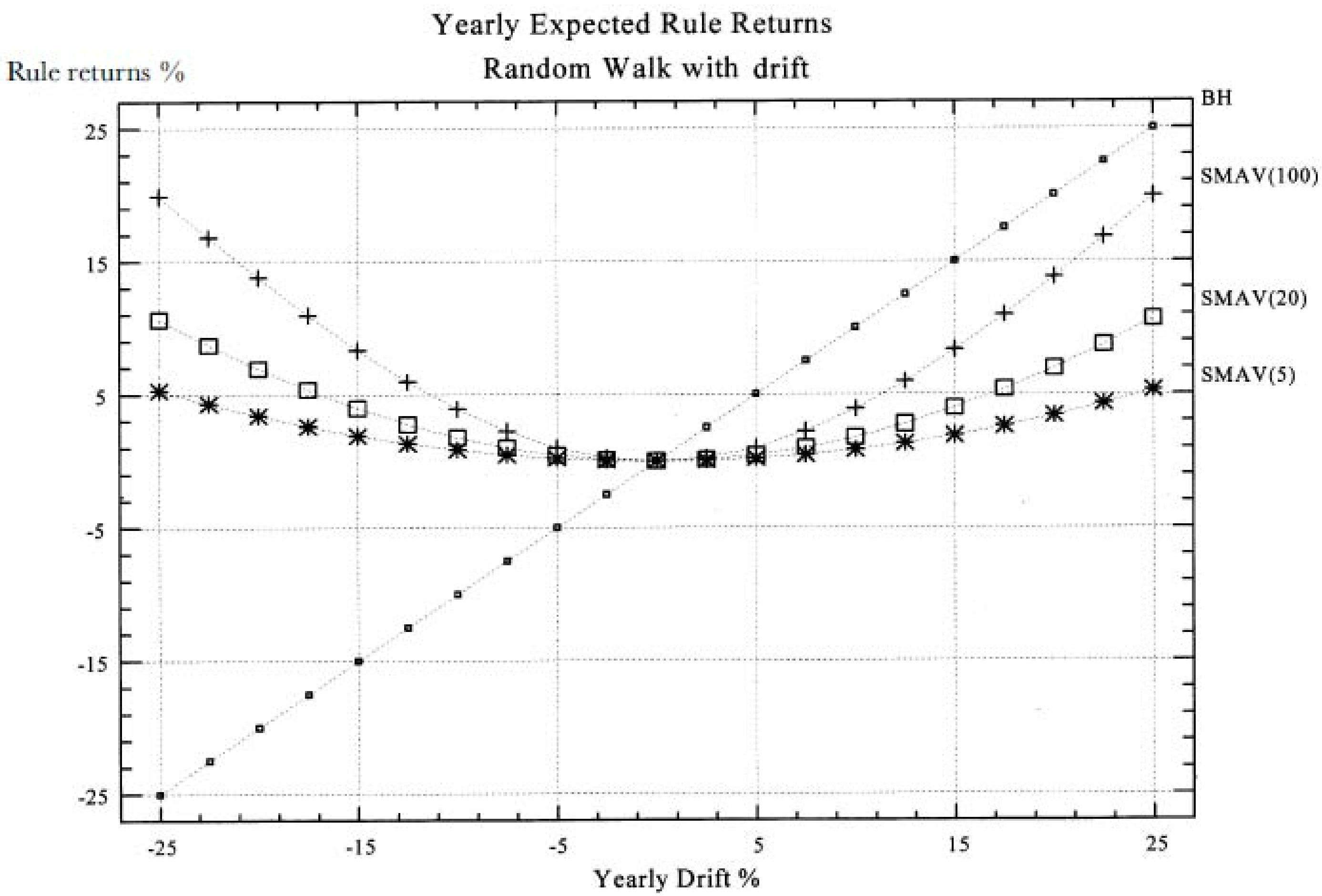
Example Linear Technical Indicators

- Simple order
- Simple MA
- Weighted MA
- Exponential MA
- Momentum
- Double orders
- Double MA

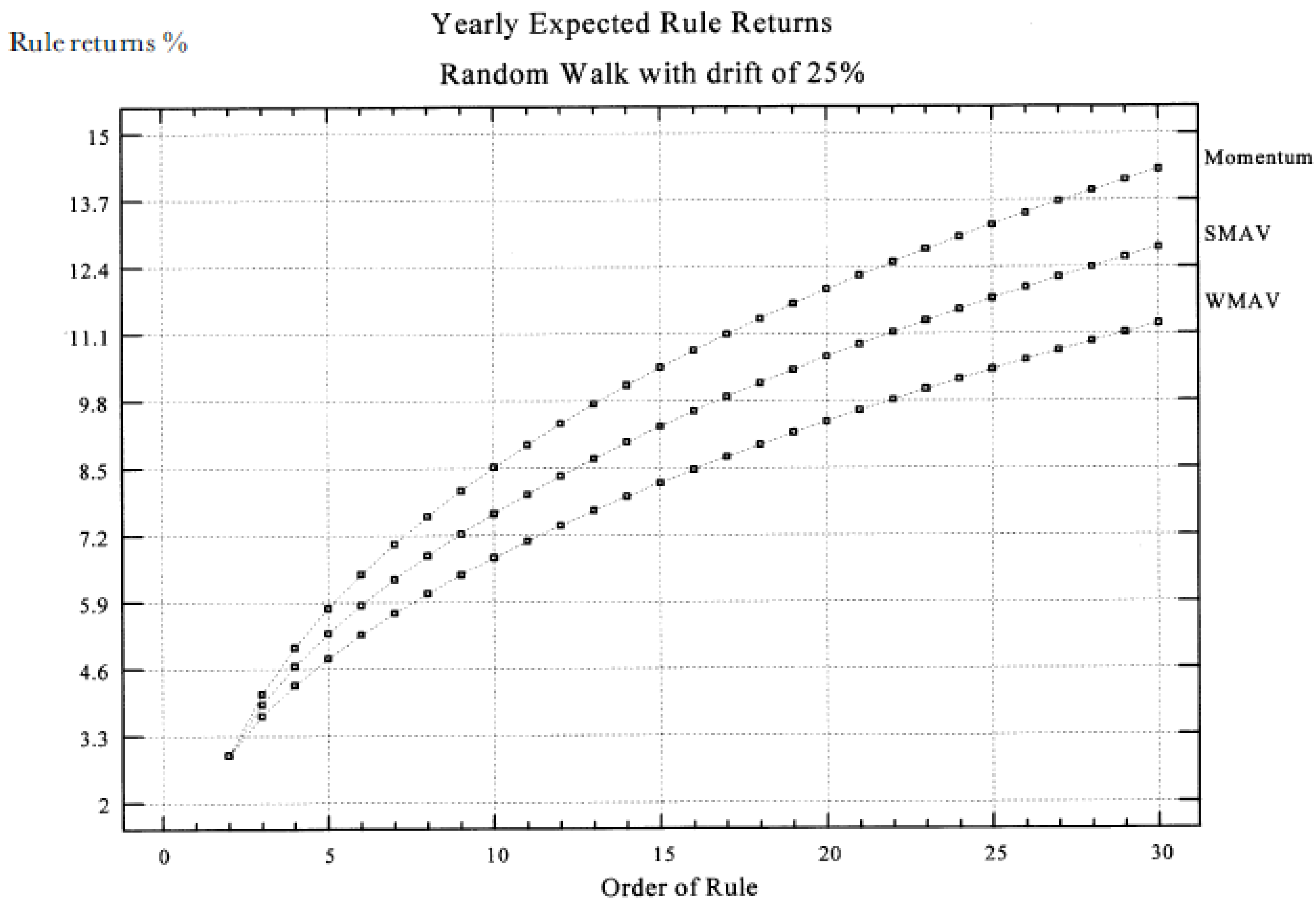
Returns: Random Walk With Drift

- $X_t = \mu + \varepsilon_t$
  - The bigger the order, the better.
  - Momentum > SMAV > WMAV
- How to estimate the *future* drift?
  - Crystal ball?
  - Delphic oracle?

Results


Yearly Expected Rule Returns
Random Walk with drift
Rule returns %
25
15
5
-5
-15
-25
-25
-15
-5
5
15
25
Yearly Drift %
BH
SMAV(100)
SMAV(20)
SMAV(5)

Results

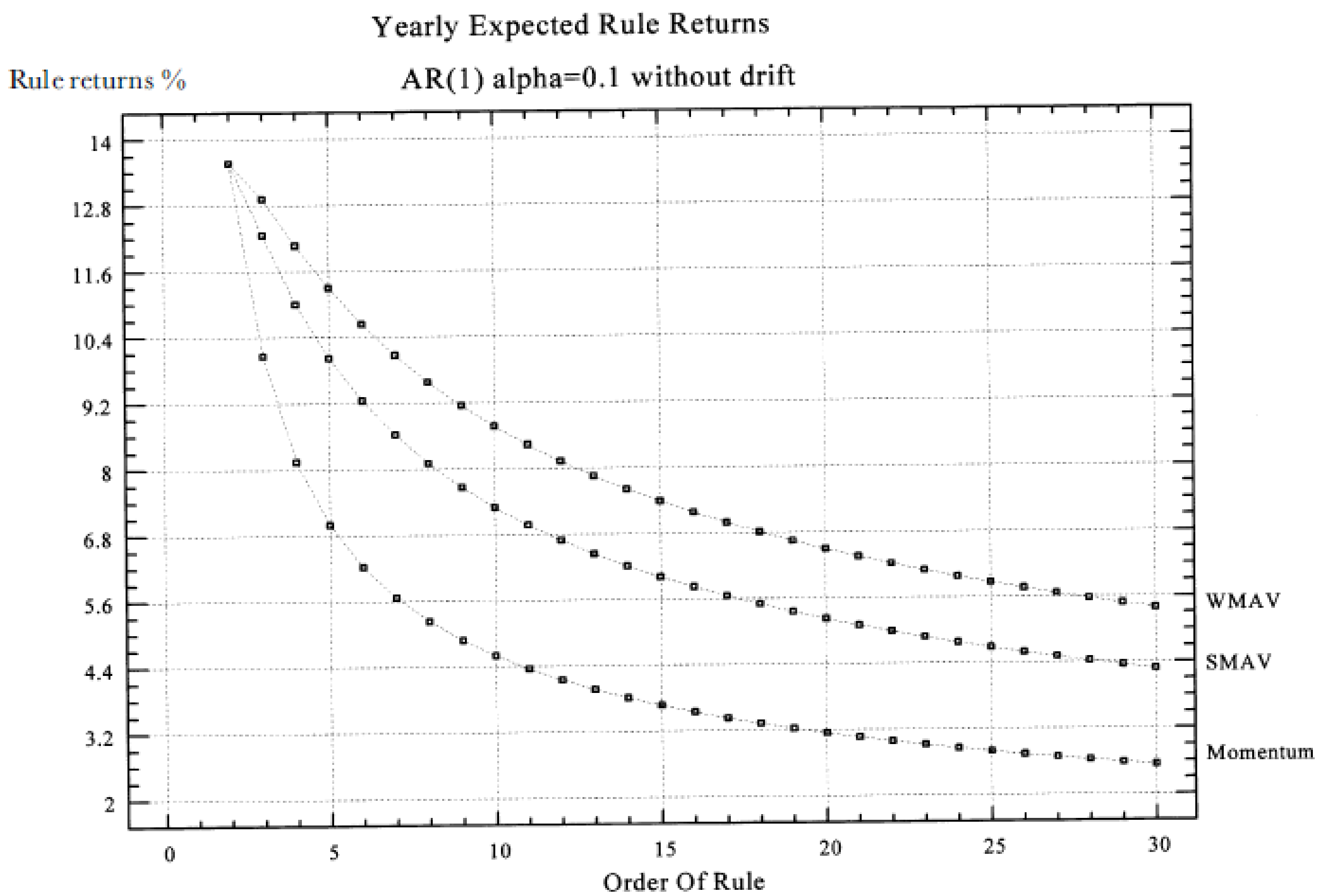


Yearly Expected Rule Returns
Random Walk with drift of 25%
Rule returns %
15
13.7
12.4
11.1
9.8
8.5
7.2
5.9
4.6
3.3
2
0
5
10
15
20
25
30
Order of Rule
Momentum
SMAV
WMAV

Returns: AR(1)

- $X_t = \alpha X_{t-1} + \varepsilon_t$
  - Auto-correlation is required to be profitable.
  - The smaller the order, the better. (quicker response)

Results



Yearly Expected Rule Returns
AR(1) alpha=0.1 without drift
Rule returns %
14
12.8
11.6
10.4
9.2
8
6.8
5.6
4.4
3.2
2
0
5
10
15
20
25
30
Order Of Rule
WMAV
SMAV
Momentum

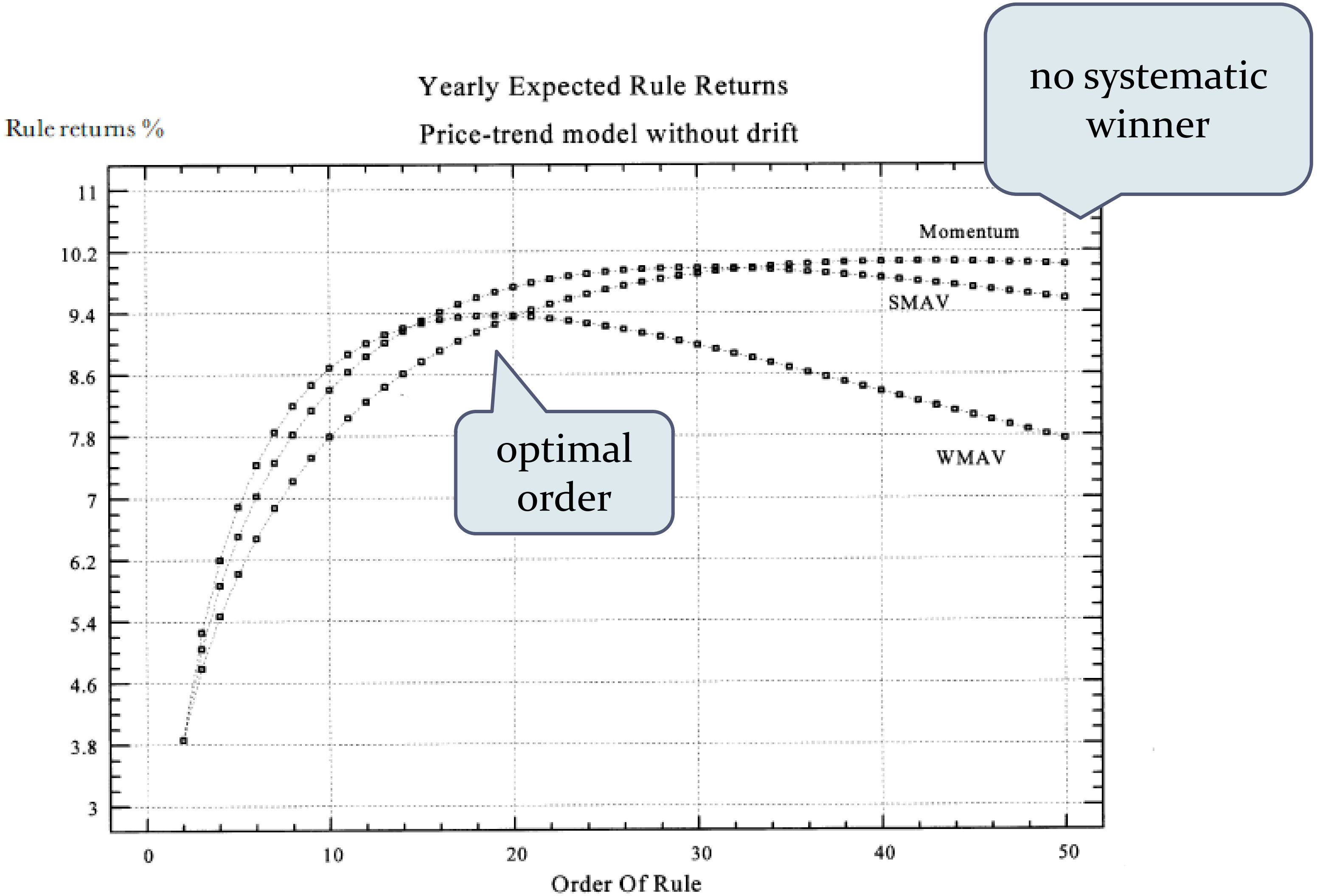
ARMA(1, 1)

AR

MA

- $(X_t - \mu) - p(X_{t-1} - \mu) = \varepsilon_t - q\varepsilon_{t-1}$
- Prices tend to move in one direction (trend) for a period of time and then change in a random and unpredictable fashion.
  - Mean duration of trends: $m_d = \frac{1}{(1-p)}$
- Information has impacts on the returns in different days (lags).
  - Returns correlation: $\rho_h = Ap^h$

Results

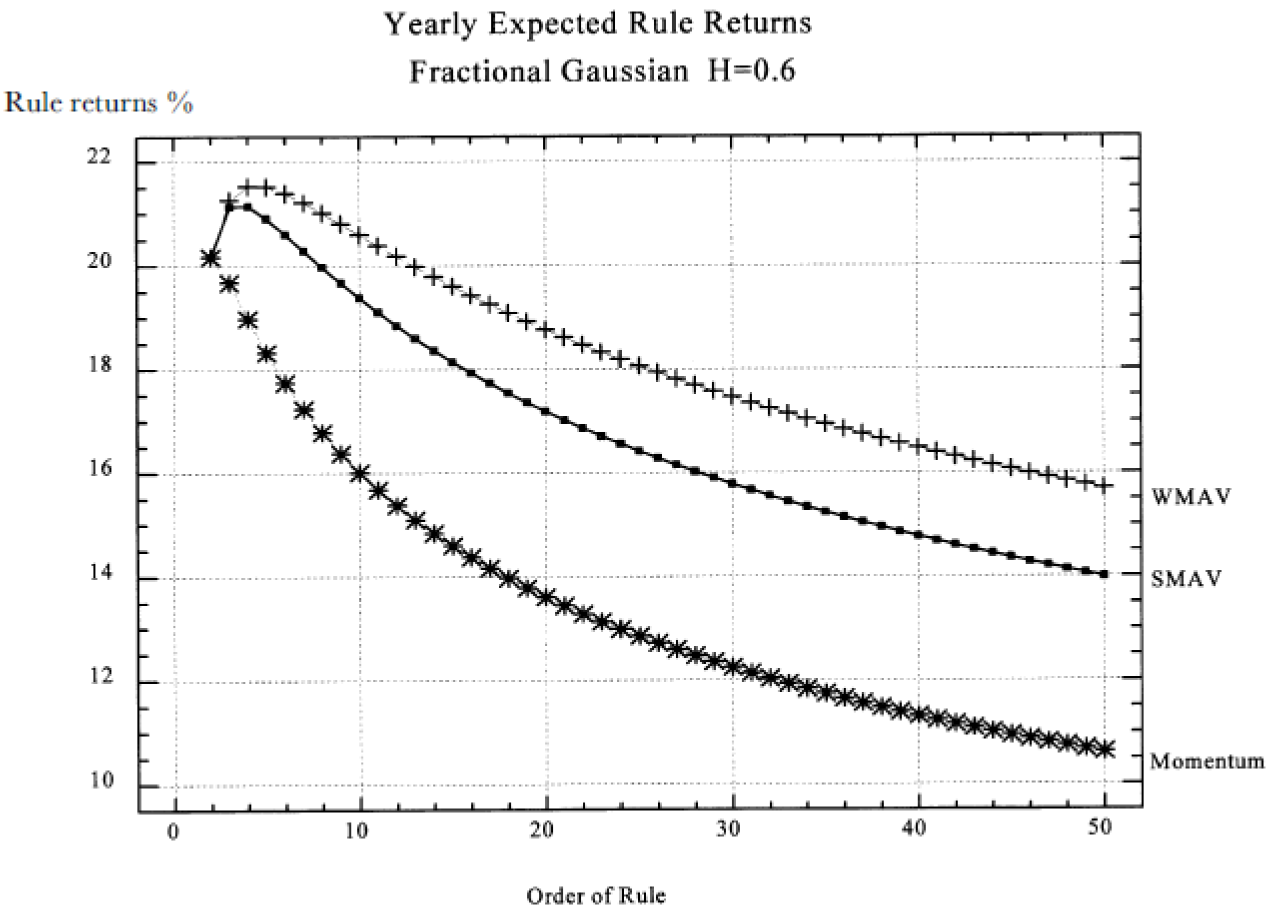


Yearly Expected Rule Returns
Price-trend model without drift
Rule returns %
11
10.2
9.4
8.6
7.8
7
6.2
5.4
4.6
3.8
3
0
10
20
30
40
50
Order Of Rule
Momentum
SMAV
WMAV
optimal order
no systematic winner

ARIMA(0, d, 0)

- $\nabla^d (X_t - \mu) = e_t$
- Irregular, erratic, aperiodic cycles.

Results



Yearly Expected Rule Returns
Fractional Gaussian H=0.6
Rule returns %
22
20
18
16
14
12
10
0
10
20
30
40
50
Order of Rule
WMAV
SMAV
Momentum

ARCH(p)

- $X_t = \mu + \left\{ \sqrt{\alpha_0 + \sum_{i=1}^{p} \alpha_i (X_{t-i} - \mu)^2} \right\} \varepsilon_t$
- $X_t - \mu$ are the residuals
- When $\mu = 0$, $\mathrm{E}(R_t) = 0$.

residual coefficients as a
function of lagged squared
residuals

AR(2) – GARCH(1,1)

AR(2)

GARCH(1,1)

- $X_t = a + b_1 X_{t-1} + b_2 X_{t-2} + \varepsilon_t$
- $\varepsilon_t = \sqrt{h_t} z_t$

innovations

- $h_t = \alpha_0 + \alpha_1 {\varepsilon_{t-1}}^2 + \beta h_{t-1}$

ARCH(1): lagged squared residuals

lagged variance

Results

- The presence of conditional heteroskedasticity will not drastically affect returns generated by linear rules.
- The presence of conditional heteroskedasticity, if unrelated to serial dependencies, may be neither a source of profits nor losses for linear rules.

Conclusions

- Trend following model requires positive (negative) autocorrelation to be profitable.
  - What do you do when there is zero autocorrelation?
- Trend following models are profitable when there are drifts.
  - How to estimate drifts?
- It seems quicker response rules tend to work better.
- Weights should be given to the more recent data.