Time Series:

information ratio = $\dfrac{\text{ptf return} - \text{benchmark return}}{STD_{\text{ptf v.s. bench}}}$

return $R_i = \dfrac{S_{i+1}-S_i}{S_i} = \mu dt + \sigma dX_t$.

time series: discrete time, continuous state.

Modeling a dynamical system.

$y(t+1) = \underbrace{\phi_1 y(t) + \psi_1 u(t)}_{\text{information}} + \underbrace{w(t)}_{\text{noise}}$.

$\gamma_x(r,s) = \mathbb{E}(X_r X_s) - \mathbb{E}(X_r)\mathbb{E}(X_s)$
$= cov(X_r, X_s)$.

$\rho_x(h) = \dfrac{\gamma_x(h)}{\gamma_x(0)} = corr(X_t, X_{t+h}) \le 1$

Stationary process: 1. $\mathbb{E}(X_t^2) < \infty$, 2. $\mathbb{E}(X_t) = m$ 3. $\gamma_x(r,s) = \gamma_x(r-s)$.
(weakly).

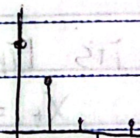
white noise: $\mathbb{E}(X_t) = 0$, $\mathbb{E}(X_t^2) = \sigma^2$ $\gamma_x(r,s) = \begin{cases} \sigma^2, & r=s \\ 0, & r \ne s \end{cases}$ $WN(0,\sigma^2)$.

random walk, not stationary, $var(X_t) \uparrow$ as more walks.

Moving Average: weighted sum of the most recent values of X_t.

MA(1): $Y_t = X_t + \theta X_{t-1}$, $X_t \in WN(0,\sigma^2)$, stationary. ✓

MA(q) = $Y_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) Z_t$.



Autoregressive $var(X_t) = \dfrac{\sigma^2}{1-\phi^2} \rightarrow \gamma_x(h) = \mathbb{E}((\phi X_{t-1} + Z_t) X_{t-h}) = \phi \gamma_x(h-1) = \dfrac{\phi^h \sigma^2}{1-\phi^2}$

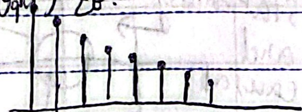
AR(1). $X_t = \phi X_{t-1} + Z_t$. $Z_t \in WN(0,\sigma^2)$.

$|\phi| < 1$, stationary.
$|\phi| = 1$, random walk.
$|\phi| > 1$, increasing volatile.

AR(p) $(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = Z_t$.

ARMA(p,q) $(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) Z_t$.

$\Rightarrow X_t = \psi(B) Z_t$.



If $X_t - \mu$ is ARMA(p,q), then X_t is ARMA(p,q) with mean μ.

$X_t = \underset{\text{trend}}{m_t} + \underset{\text{seasonal component}}{S_t} + \underset{\text{stationary time series.}}{Y_t}$

1. extracting deterministic components: detrending, difference.
2. model selection, parameter estimation
3. diagnostic test: residuals → WN, etc.

① $\nabla^n X_t = (1-B)^n X_t$. each differencing removes one order trend.

② $\nabla_d X_t = X_t - X_{t-d}$ ⚠ too much differencing magnified error when forecasting

$= (1-B^d)X_t \Rightarrow$ for seasonal component $S_t = S_{t+d}$, d known.

Wold's Decomposition:

any stationary process can be represented by infinite order MA model

$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, where $\sum_{j=0}^{\infty} (\psi_j)^2 < \infty$, MA($\infty$)

$E(X_t) = 0$. $E(X_t^2) = \sigma^2 \sum_{j=0}^{\infty} (\psi_j)^2 < \infty$. $E(X_t X_{t-h}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$.

EX.1 For AR(1). $X_t = \phi X_{t-1} + Z_t \Rightarrow (1-\phi B)X_t = Z_t \Rightarrow X_t = \frac{1}{1-\phi B} Z_t$

$\frac{1}{1-\phi B} = (1 + \phi B + \phi^2 B^2 + \dots +)$

$\frac{1}{1-x} = (1 + x + x^2 + \dots)$

$\Rightarrow X_t = (1 + \phi B + \phi^2 B^2 + \dots) Z_t$. for $|x| < 1$

E.X.2 For MA(1), $X_t = Z_t + \theta Z_{t-1} = (1+\theta B) Z_t$. $\frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots$

$\xRightarrow{|\theta|<1}$ $Z_t = (1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \dots) X_t$ AR(∞).

Linear Process:

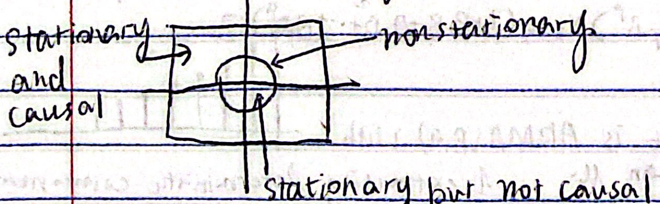
→ $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, $Z_t \in WN(0, \sigma^2)$.

it's linear causal if

$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ (X_t only depend on previous value of Z_t).

Stationarity of ARMA(p,q). $\phi(B) X_t = \theta(B) Z_t$

⇒ a stationary solution exists iff no roots of $\phi(B)$ lies ON unit circle



if causal, $\psi(B) = \frac{\theta(B)}{\phi(B)} = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots$

Thm. $X_t = ARMA(p_1, q_1)$. $Y_t = ARMA(p_2, q_2)$. $Z_t = ARMA((p_1+p_2), \max(p_1+q_2, p_2+q_1))$

$Z_t = X_t + Y_t$

X_t is ARIMA(p, d, q) if $Y_t \equiv \nabla^d X_t = (1-B)^d X_t$ is ARMA(p,q)

↳ non-stationary. ↘ integer — stationary causal.

ARFIMA: d can take non-integer values

Parameter Calibration.

MLE: Gaussian Process:

$$\begin{bmatrix} X_{t_1} \\ X_{t_2} \\ \vdots \\ X_{t_n} \end{bmatrix} \approx MVN(\mu, \Sigma), \quad f(X) = (2\pi)^{-\frac{n}{2}} (\det \Sigma)^{-\frac{1}{2}} e^{[-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)]}$$

$$L(X, \theta) = (2\pi)^{-\frac{n}{2}} (\det \Sigma(\theta))^{-\frac{1}{2}} e^{[-\frac{1}{2} X^T \Sigma^{-1}(\theta) X]}$$

① $X_i \approx N(\mu_i, \Sigma_{ii})$. ② $X_i \perp\!\!\!\perp X_j$ if $\Sigma_{ij} = 0$ ③ $AX \approx MVN(A\mu, A^T \Sigma A)$

$MLE(\hat{\theta}) = \underset{\theta}{\arg\max}\, L(X, \theta)$.

$\log L(X; \psi) = -\frac{n}{2} \log(2\pi) + \frac{1}{2} \log |\Gamma_n|^{-1} - \frac{1}{2} (X-\mu)^T \Gamma_n^{-1} (X-\mu)$.

difficult to implement.

easier: $f(x_2, x_1; \psi) = f(x_1; \psi) f(x_2 | x_1; \psi)$. $\quad X_t = \phi_0 + \phi X_{t-1} + Z_t$.

$f(x_n, x_{n-1}, \cdots, x_1; \psi) = f(x_1) \prod_{i=2}^{n} f(x_i | x_{i-1})$.

$$\log L(X, \psi) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi^2}\right) - \frac{\left(x_1 - \frac{\phi_0}{1-\phi}\right)^2}{2\frac{\sigma^2}{1-\phi^2}} - \frac{n-1}{2}\log(2\pi) - \frac{n-1}{2}\log(\sigma^2) - \sum_{i=2}^{n} \frac{(x_i - \phi_0 - \phi x_{i-1})^2}{2\sigma^2}$$

$$f(x_n | x_{n-1}; \psi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x_n - \phi_0 - \phi x_{n-1})^2}{2\sigma^2}\right).$$

Yule-Walker: for AR(p). $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = Z_t$.

multiply both side by X_{t-j}.

$E(X_t X_{t-j} - \phi_1 X_{t-1} X_{t-j} - \cdots - \phi_p X_{t-p} X_{t-j}) = E(Z_t X_{t-j})$. $\quad \left\{ \begin{array}{l} (p+1) \text{ equations} \\ (p+1) \text{ unknowns.} \end{array} \right.$

$$E(Z_t X_{t-j}) = \sum_{i=0}^{\infty} \psi_i E(Z_t Z_{t-j-i}) = \begin{cases} \psi_0 \sigma^2, & j = 0 \\ 0, & j = 1, 2, \cdots p \end{cases}$$

$$\begin{bmatrix} \gamma_x(0) & \gamma_x(1) & \cdots & \gamma_x(p-1) \\ \gamma_x(1) & \gamma_x(0) & & \vdots \\ \vdots & \vdots & \ddots & \\ \gamma_x(p-1) & \gamma_x(p-2) & \cdots & \gamma_x(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \gamma_x(1) \\ \gamma_x(2) \\ \vdots \\ \gamma_x(p) \end{bmatrix} + \gamma_x(0) - \phi_1 \gamma_x(1) - \cdots - \phi_p \gamma_x(p) = \sigma^2$$

Scanned with CamScanner

using sample autocovariance

$$\begin{bmatrix}\hat\phi_1\\ \hat\phi_2\\ \vdots\\ \hat\phi_p\end{bmatrix} = \begin{bmatrix}\hat\gamma_x(0) & \hat\gamma_x(1) & \cdots & \hat\gamma_x(p-1)\\ \hat\gamma_x(1) & \ddots & & \vdots\\ \vdots & & & \\ \hat\gamma_x(p-1) & \cdots & & \hat\gamma_x(0)\end{bmatrix}^{-1} \begin{bmatrix}\hat\gamma_x(1)\\ \hat\gamma_x(2)\\ \vdots\\ \hat\gamma_x(p)\end{bmatrix}$$

$$\hat\gamma_x(0) - \hat\phi_1\hat\gamma_x(1) - \hat\phi_2\hat\gamma_x(2) - \cdots - \hat\phi_p\hat\gamma_x(p) = \hat\sigma^2.$$

NOT suitable for MA(q).

Conclusion. MLE less bias, less STD.
MLE more broad use.
MLE more computationally intensive.

Prediction:

$$P_n X_{n+h} = a_0 + a_1 X_n + a_2 X_{n-1} + \cdots + a_n X_1$$

$$S(a_0, a_1, \ldots, a_n) = \mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right], \quad \text{MSPE minimize.}$$

$$\frac{\partial S}{\partial a_0} = 0 \Rightarrow a_0 = \mu\left(1 - \sum_{j=1}^{n} a_j\right).$$

$$\gamma_x(h+i-1) = \sum_{j=1}^{n} \gamma_x(j-i)\, a_j$$

$$\underbrace{\begin{bmatrix}\gamma_x(0) & \gamma_x(1) & \cdots & \gamma_x(n-1)\\ \gamma_x(1) & \gamma_x(0) & & \gamma_x(n-2)\\ \vdots & & \ddots & \vdots\\ \gamma_x(n-1) & \cdots & & \gamma_x(0)\end{bmatrix}}_{\Gamma_n} \underbrace{\begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_n\end{bmatrix}}_{\alpha_n} = \underbrace{\begin{bmatrix}\gamma_x(h)\\ \gamma_x(h+1)\\ \vdots\\ \gamma_x(h+n-1)\end{bmatrix}}_{\gamma_n(h)}$$

$\alpha_n = \Gamma_n^{-1}\gamma_n(h)$. best linear predictor.

$$\mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right] = \gamma_x(0) - \alpha_n^T \gamma_n(h).$$

Scanned with CamScanner

State Space and Filtering

$X_t = AX_{t-1} + bZ_t$ ← state. dynamic update : prediction step.

$Y_t = CX_t + dV_t$ ← observation . update step. $Z_t, V_t = IIDN(0,1)$.

estimation of $X_t \mid Y_{t-1}, Y_{t-2}, \dots, Y_1$ ⇒ prediction problem.

estimation of $X_t \mid Y_t, Y_{t-1} \dots$ ⇒ filtering problem.

estimation of $X_t \mid Y_n, Y_{n-1} \dots Y_1$ ⇒ smoothing problem.
$(n > t)$.

probalistic state space model.

$X_t \approx p(X_t \mid X_{t-1})$

$Y_t \approx p(Y_t \mid X_t)$

linear state space + gaussian process → Kalmen Filter.

Filter has 2 steps : 1. predict at time t-1 2. measure Y_t and update

$Bel(X_{t-1}) = P(X_{t-1} \mid Y_{t-1})$ → $\boxed{\text{predict}}$ → $P(X_t \mid Y_{1:t-1})$ Time t-1

$Bel(X_t) = P(X_t \mid Y_{1:t})$ ← $\boxed{\text{Update}}$ ← ~~Pt~~ ← Y_t Time t

$P(X_t \mid X_{t-1})$ Markov

$$P(X_t \mid Y_{1:t-1}) = \int P(X_t \mid X_{t-1}, Y_{1:t-1}) P(X_{t-1} \mid Y_{1:t-1}) dX_{t-1} \quad \text{Prediction}$$

$P(Y_t \mid X_t)$

$$P(X_t \mid Y_t, Y_{1:t-1}) = \frac{P(Y_t \mid X_t, Y_{1:t-1})}{P(Y_t \mid Y_{1:t-1})} P(X_t \mid Y_{1:t-1}) \quad \text{Update.}$$

If $Bel(X_t) = N(0, \sigma^2)$, it becomes Kalman Filter.

Kalman Filter: $X_t = AX_{t-1} + V_t$. $V_t = IIDN(0, Q)$.

$Y_t = CX_t + Z_t$. $Z_t = IIDN(0, R)$.

Predict (time: t-1). ~~$X_{t|t-1}$~~ $\hat{X}_{t|t-1} = A\hat{X}_{t-1|t-1}$, $P_{t|t-1} = AP_{t-1|t-1}A^T + Q$

Measurement (time: t). Y_t is observed.

Update : $K_t = P_{t|t-1}C^T(CP_{t|t-1}C^T + R)^{-1}$. $P_{t|t} = (I - K_tC)P_{t|t-1}$.

$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t(Y_t - C\hat{X}_{t|t-1})$

Scanned with CamScanner

Volatility.

$r_t = \mu + \underbrace{\sigma_t \varepsilon_t}_{a_t}$, $\varepsilon_t \sim$ i.i.d. $N(0,1)$. $var(a_t) = \frac{\alpha_0}{1-\alpha_1}$ $kurt(a_t) = \frac{E(a_t^4)}{(E(a_t^2))^2} - 3\frac{E(a_t^4)}{(E(a_t^2))^2} > 3$

ARCH(1): $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2$, AR model on squared residuals

$V_t = a_t^2 - \sigma_t^2$. volatility changing.

$a_t^2 - V_t = \alpha_0 + \alpha_1 a_{t-1}^2 \leftarrow AR(1)$.

properties: returns uncorrelated; volatility clustering; excess kurtosis

GARCH(1).: $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$. ← substitute gives $AR(\infty)$. , $Var(a_t) = \frac{\alpha_0}{1-\alpha_1-\beta_1} = \sigma^2$

$a_t^2 - V_t = \alpha_0 + \alpha_1 a_{t-1} + \beta_1 (a_{t-1}^2 - V_{t-1}) \leftarrow ARMA(1,1)$. $(1 - \phi_1 B - \phi_2 B^2 - \ldots)X_t = (1 + \theta_1 B + \theta_2 B^2 + \ldots) Z_t$.

$(\alpha_1 + \beta_1)^K = 0.5$. up move has <u>same</u> impact on volatility than down move.

$L(\theta | x_1, x_2 \ldots x_N) = f(x_1, x_2, \ldots x_N | \theta) \overset{\text{if i.i.d}}{=\!=} \prod_{i=1}^{N} f(x_i | \theta)$.

$\ln L(\theta | x_1, x_2 - x_N) = \sum_{i=1}^{N} \ln f(x_i | \theta)$.

Ex. let $x_1, x_2, \ldots x_N$ be random sample i.i.d. $f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0. \\ 0, & \text{otherwise} \end{cases}$

$L(\lambda) = \lambda^N e^{-\lambda \Sigma x_i}$. $\ln(L(\lambda)) = N \ln \lambda - \lambda \sum_{i=1}^{N} x_i$

$\frac{\partial \ln(L(\lambda))}{\partial \lambda} = \frac{N}{\lambda} - \Sigma x_i = 0$. $\lambda = \frac{\Sigma x_i}{N} = \bar{x}$

GJR GARCH: $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \gamma_1 S_{t-1} a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, $S_{t-1} = \begin{cases} 1, & \text{if } a_{t-1} < 0 \\ 0, & \text{if } a_{t-1} \geq 0, \end{cases}$

time varying volatility ? stochastic.

$\ln(\sigma_t) = \alpha + \phi(\ln(\sigma_{t-1}) - \alpha) + \eta_t$ if $\eta_t \sim N(0, \sigma_\eta^2)$, then ~~log~~ $\ln(\sigma_t) \sim N$

↑ AR(1). ↑ rate of mean reversion. $E(\ln(\sigma_t)) = \alpha$, $Var(\ln(\sigma_t)) = \frac{\sigma_\eta^2}{1-\phi^2}$

Parameter Estimation: Generalized Methods of Moments.

Classical. $\hat{\theta} = f(m_1, m_2, \ldots m_K)$, $m_j = (\frac{1}{N}) \sum_{i=1}^{N} x_i^j$

GMM: $g(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^{T} f_t(r_t, \hat{\theta})$, $\min Q(\theta) = g(\theta)^T W g(\theta)$.

$W = S^{-1}$