CHAPTER 1 Prerequisites

* $g: \Omega \to \mathbb{R}$, $f: \mathbb{R} \to \mathbb{R}$ measurable fns
$\Rightarrow f \circ g : \Omega \to \mathbb{R}$ is measurable.

* $F_X(x) = P_X((-\infty, x]) = P(X \leq x) = p$
$\Rightarrow x = F_X^{-1}(p)$ where $F_X : \mathbb{R} \to [0,1]$.

* $\epsilon_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else} \end{cases}$

* $E(X) = \int_{\mathbb{R}} x f_X(x)\, dx$ for continuous random variables

* $E(X) = \sum_{i=1}^{\infty} p_X(x_i) x_i$ for discrete random variables

* Chebyshev's inequality -
$\alpha > 0$, $g(\cdot) > 0$ on $\mathbb{R}^+$
$$P(|X| > \alpha) \leq g(\alpha)^{-1} E(g(|X|))$$

* $Var(X) = E((X - E[X])^2)$
* $Var(X) = E[X^2] - E[X]^2$
* $Cov(X,Y) = E((X - E[X])(Y - E[Y]))$
* $Cor(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)\, Var(Y)}} \in [-1,1]$

* Cauchy - Schwarz equality -
$$|E(XY)|^2 \leq E(|X||Y|)^2 \leq E(|X|^2)\, E(|Y|^2).$$

* $Cov(X,Y) = 0$
$\Rightarrow Cov(X,Y) = 0 \Rightarrow$ X and Y are uncorrelated or orthogonal.

* $X_1, \cdots, X_n$ - pairwise uncorrelated r.vs
$\alpha_1, \cdots, \alpha_n \in \mathbb{R}$
$$Var\left(\sum_{i=1}^{n} \alpha_i X_i\right) = \sum_{i=1}^{n} \alpha_i^2\, Var(X_i)$$

* $X_1, \cdots, X_n$ independent r.vs
$g_1, \cdots, g_n$ measurable fns such that
$E(g_1(X_1) \cdots g_n(X_n)) < +\infty$
$\Rightarrow E(g_1(X_1) \cdots g_n(X_n)) = E(g_1(X_1)) \cdots E(g_n(X_n))$

* Conditional prob of X given Y
$$P(X \in A \mid Y \in B) = \frac{P(X \in A, Y \in B)}{P(Y \in B)}$$

* Conditional density
$$f(x|y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)}, & \text{if } f_Y(y) \neq 0 \\ 0, & \text{else} \end{cases}$$

* Properties of the conditional expectation :-

(i) $a_1, a_2 \in \mathbb{R}$
X_1, X_2, Y r.vs
$$E(a_1 X_1 + a_2 X_2 | Y) = a_1 E(X_1|Y) + a_2 E(X_2|Y)$$

(ii) $E(E(X|Y)) = E(X)$

(iii) X independent of Y
$$E(X|Y) = E(X)$$

(iv) constant $a \in \mathbb{R}$
$$E(a|Y) = a$$

(v) $g: \mathbb{R} \to \mathbb{R}$ measurable fn
X, Y r.vs
$$E(g(Y)X|Y) = g(Y)E(X|Y)$$

(vi) $(X, Y_1, \cdots, Y_n)$ independent of Z
$$E(X|Y_1, \cdots, Y_n, Z) = E(X|Y_1, \cdots, Y_n)$$

* Bernoulli distribution
$\mathcal{X} = \{0,1\}$, $p \in [0,1]$
$P(X=1) = p$, $P(X=0) = 1-p$
$E(X) = p$, $Var(X) = p(1-p)$.

* Uniform distribution
$X \sim \mathcal{U}([a,b])$
$$f_X(x) = \begin{cases} (b-a)^{-1} & \text{if } x \in [a,b] \\ 0 & \text{else} \end{cases}$$

$E(X) = (a+b)/2$
$Var(X) = (b-a)^2/12$

* Normal distribution

$X \sim N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

* Student's t-distribution

$$f_X(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

* Central Limit Theorem

Let $(X_n, n \in \mathbb{N})$ be a sequence of i.i.d. r.v.s, each having mean μ and finite non-zero variance σ^2 and let $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. Then the distribution of the standardized sample mean tends to the standard normal distribution i.e., for all $x \in \mathbb{R}$

$$P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq x\right) \to \Phi(x).$$

* The cumulative distribution function of X is

$$F_X(x) = F_X(x_1, x_2, \cdots, x_n)$$
$$= P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n)$$

* The density function of X is

$$F_X(x) = f_X(x_1, x_2, \cdots, x_n)$$
$$= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(y_1, \cdots, y_n)\, dy_1 dy_2 \cdots dy_n$$

* Level = P(Reject H_0 | H_0 is true)

power = P(Reject H_0 | H_1 is true)

CHAPTER 2 Stationary time series and seasonality

log-returns P_t, $t=1,\cdots,n$

$r_t = \log(P_t) - \log(P_{t-1})$, $t=1,\cdots,n-1$

* Transformation logarithmic -
(i) to stabilize variance,
(ii) removes/reduces the skewness of the data.

* A stochastic process $X=(X_t, t\in T)$ is called iid noise with mean μ and variance σ^2 if the sequence of r.vs $(X_t, t\in T)$ is i.i.d with $E[X_t]=\mu$ and $Var[X_t]=\sigma^2 \ \forall t\in T$. $X \sim IID(\mu,\sigma^2)$

* Eg of iid noise is the binary process - flipping of a fair coin.
$P(X_1=1) = P(X_1=-1) = \frac{1}{2}$.
$E[X_1]=0 \quad Var[X_1]=1$.

* $X_n, n\in\mathbb{Z}$ i.i.d noise

$$\begin{cases} S_0 = 0 \\ S_n = \sum_{i=1}^{n} X_i = S_{n-1} + X_n, n\in\mathbb{N} \end{cases}$$

$(S_n, n\in\mathbb{N}_0)$ is a random walk.

* A time series X is said to be a Gaussian time series if all finite dimensional vectors are multivariate Gaussian distributed. A useful fact of the multivariate normal distribution is that any linear combination of the components of a multivariate normal random vector is also normal.
$[(X_t, t\in T)$ is Gaussian iff $\sum_{i=1}^{n} a_i X_{t_i}$ is Gaussian, $n\in\mathbb{N}$, $a\in\mathbb{R}^n$.$]$

* Let $X=(X_t, t\in T)$ be a stochastic process with $Var(X_t) < +\infty \ \forall t\in T$.
The mean fn $\mu_X: T\to\mathbb{R}$ of X is
$\mu_X(t) = E[X_t] \ \forall t\in T$.
The covariance fn $\gamma_X: T\times T\to\mathbb{R}$ is
$\forall r,s\in T \ \gamma_X: T\times T\to\mathbb{R}$ is $\gamma_X(r,s) = Cov(X_r,X_s) = E[(X_r-E[X_r])(X_s-E[X_s])]$

* Let $X=(X_t, t\in\mathbb{Z})$ be a time series with $Var(X_t)<\infty$ for all $t\in\mathbb{Z}$. The time series X is called (weakly) stationary if
(i) $\exists \mu\in\mathbb{R} \mid \mu_X(t) = \mu \ \forall t\in\mathbb{Z}$ and
(ii) $\gamma_X(r,s) = \gamma_X(r+h, s+h) \ \forall r,s,h\in\mathbb{Z}$.

* A time series $X=(X_t, t\in\mathbb{Z})$ (for which $Var(X_t)<\infty \ \forall t\in\mathbb{Z}$ is not necessarily true) is said to be strictly stationary if the r.vs $(X_1,\cdots,X_n)$ and $(X_{1+h},\cdots,X_{n+h})$ have the same joint distributions $\forall h\in\mathbb{Z}$ & $n\in\mathbb{N}$.

* A weakly stationary Gaussian time series is also strictly stationary since the normal distribution is completely determined by its mean & covariance.

* Let X be a stationary times series.
Auto covariance fn (ACVF) $\gamma_X: \mathbb{Z}\to\mathbb{R}$ of X is
$\gamma_X(h) = Cov(X_{t+h}, X_t)$ for $h\in\mathbb{Z}$ & $t\in\mathbb{Z}$.
Autocorrelation fn (ACF) $\rho_X: \mathbb{Z}\to[-1,1]$ of X is
$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$ for $h\in\mathbb{Z}$.

* γ_X is well-defined due to the stationarity of X and since the autocovariance fn is symmetric
i.e., $\gamma_X(h) = \gamma_X(-h)$, $h\in\mathbb{Z}$.

* A stochastic process $X=(X_t, t\in\mathbb{Z})$ is called a white noise with mean μ and variance σ^2 if it is a stationary process with $E(X_t)=\mu$, $t\in\mathbb{Z}$, and for $h\in\mathbb{Z}$

$$\gamma_X(h) = \begin{cases} \sigma^2 & \text{if } h=0 \\ 0 & \text{else} \end{cases} . \quad X \sim WN(\mu,\sigma^2).$$

* iid noise $\Rightarrow$ white noise.
* white noise $\Rightarrow$ iid noise if the random variables are independent and identically distributed.
* A Gaussian white noise is iid, since random variables that are uncorrelated and jointly normal are independent.
* Let $X=(X_t, t\in\mathbb{N})$ be a time series.
Sample mean: $\bar{X}_n = n^{-1}\sum_{t=1}^{n} X_t$
Sample autocovariance fn: $\hat{\gamma}(h) = n^{-1}\sum_{t=1}^{n-h}(X_{t+h}-\bar{X})(X_t-\bar{X})$
$h=0,\cdots,n-1$

* X - stationary time series with μ and γ_X.
$\lim_{n\to\infty} Var(\bar{X}_n) = \lim_{n\to\infty} E((\bar{X}_n-\mu)^2) = 0$
if $\sum_{|h|<\infty} |\gamma_X(|h|)| < +\infty$

* Reasonable $n \geq 50$, $h \leq n/4$.

* Sample autocovariance f_n / k-D sample covariance matrix.

$$\hat{\Gamma}_k = \begin{pmatrix} \hat\gamma(0) & \hat\gamma(1) & \cdots & \hat\gamma(k-1) \\ \hat\gamma(1) & \hat\gamma(0) & \cdots & \hat\gamma(k-2) \\ \vdots & \vdots & & \vdots \\ \hat\gamma(k-1) & \hat\gamma(k-2) & \cdots & \hat\gamma(0) \end{pmatrix}$$

* $\hat\Gamma_k$ is nonnegative definite.

* Sample autocorrelation matrix
$\hat R_k = \hat\gamma(0)^{-1}\hat\Gamma_k$. (nonsingular if $\hat\gamma(0)>0$)

* If $\hat\rho(h)$ displays a periodic behaviour i.e, the peaks at lag h are similar to the peaks at lag $h-k$, then there is evidence of the presence of a seasonal component.

* Given a dataset, the first thing is to check if there is a temporal structure to it i.e., if it is something other than iid noise. If $Y=(Y_1,\dots,Y_n)$ is a sequence of iid r.vs with finite variance, then the sample autocorrelations $\hat\rho(h), h=1,2,3,\dots$ for suff large n approx iid and $N(0,n^{-1})$ distributed. Given a dataset $(y_1,\dots,y_n)$, an informal hypothesis test is to check if 95% of the computed values of $\hat\rho$ should fall below the bounds $\pm\frac{1.96}{\sqrt n}$ if the dataset is a realization of Y. Otherwise reject the null hypothesis of Y being iid noise.
Theoretically, we check if there is any temporal structure to the data by hypothesis testing.

* Portmanteau test / Box-Pierce test
$H_0: Y\sim IID(\mu,\sigma^2)$
$H_1: Y\nsim IID(\mu,\sigma^2)$
Test statistic $\lambda = n\sum_{i=1}^{h}\hat\rho(i)^2$.
Null hypothesis is rejected at $\alpha\in(0,1)$, if $\lambda > \chi^2_{1-\alpha,h}$.

* Ljung-Box test
$\lambda = n(n+2)\sum_{i=1}^{h}\frac{\hat\rho(i)^2}{n-i}$. $[\chi^2_{0.95,4} = 9.49]$

* Data $x=(x_t, t=1,\dots,205)$
Log returns $y=(y_t, t=1,\dots,204)$ $y_t=\log(y_{t+1})-\log(y_t)$
Absolute $|y|=(|y_t|, t=1,\dots,204)$

Good idea to apply a couple of transformations to the data when testing for iid, so that we do not fail to reject the null hypothesis by mistake.

* Goal of forecasting a stationary time series with μ and γ is to predict $(X_{n+h}, h>0)$ in terms of $(X_t, t=1,\dots,n)$. We find best predictors in the sense of minimum mean squared errors.

* X, Y r.vs. Y is approximation of X.
is $MSE(Y,X) = E[(Y-X)^2]$.

* Best predictor of X_t for $t\in\mathbb{Z}$ given X^n
$b_t(X^n) = E(X_t|X^n)$

* Best linear predictor of X_t is
$b_t(X^n) = a_0 + a_1X_{t_n} + a_2X_{t_{n-1}} + \dots + a_nX_{t_1}$

prediction equations
(i) $E(X_t - b_t^l(X^n)) = 0$,
(ii) $E(X_{t_j}(X_t - b_t^l(X^n))) = 0 \ \forall j=1,\dots,n$.

* $a_0 = \mu\left(1-\sum_{i=1}^{n}a_i\right)$
$\Gamma_n(a_1,\dots,a_n)' = (\gamma(t-t_n),\dots,\gamma(t-t_1))'$
$\Gamma_n = (\gamma(t_{n+1-j}-t_{n+1-i}))_{i,j=1}^{n}$
$MSE(b_t^l(X^n),X_t) = E((b_t^l(X^n)-X_t)^2)$
$=\gamma(0) - (a_1,\dots,a_n)(\gamma(t-t_n),\dots,\gamma(t-t_1))'$.

* AR(1) $X_t - \phi_1X_{t-1} = Z_t$, $Z\sim WN(0,\sigma^2)$, $|\phi_1|<1$.
Z_t is uncorrelated with X_{t-j} for $j>0$.
$\gamma_X(0) = \frac{\sigma^2}{1-\phi_1^2}$
$\gamma_X(h) = \frac{\sigma^2\phi_1^{|h|}}{1-\phi_1^2}$
$b_{n+1}^l(X^n) = \phi_1X^n$
$MSE(b_{n+1}^l(X^n), X_{n+1}) = \sigma^2$.

* MA(1) $X_t = Z_t + \theta_1Z_{t-1}$, $Z\sim WN(0,\sigma^2)$
$\gamma_X(0) = (1+\theta_1^2)\sigma^2$
$\gamma_X(1) = \gamma_X(-1) = \theta_1\sigma^2$
$\gamma_X(h) = 0 \ \forall |h|>1$.
$m_t = \sum_{j=0}^{k} a_j t^j$ $MSE(b_{n+1}^l(X^n), X_{n+1}) = (1+\theta_1^2)\sigma^2 - a_1\theta_1\sigma^2$

* Stochastic process X can be split into $X_t = m_t + s_t + Y_t$. This is called the classical decomposition model. $m:\mathbb{Z}\to\mathbb{R}$ is a slowly changing f_n called the trend component, $s:\mathbb{Z}\to\mathbb{R}$ is a f_n of period d referred to as the seasonal component i.e., $s_{t+d}=s_t$ and $\sum_{j=1}^{d}s_j=0$ and $Y=(Y_t, t\in\mathbb{Z})$ is a stationary time series with mean zero.

* Backward shift operator $BX_t = X_{t-1}$.
lag-d differencing operator $\nabla_d X_t = (1-B^d)X_t = X_t - X_{t-d}$

CHAPTER 3 Linear time series models

A time series X is called an autoregressive process of order p or AR(p) process if X is stationary and if $\forall t \in \mathbb{Z}$,
$$X_t - \sum_{j=1}^{p} \phi_j X_{t-j} = Z_t, \text{ where } Z \sim WN(0,\sigma^2)$$

* A time series X is called a moving average process of order q or MA(q) process if X is stationary and if $\forall t \in \mathbb{Z}$, $X_t = Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j}$, where $Z \sim WN(0,\sigma^2)$.

* A time series X is an ARMA(p,q) process if X is stationary and if $\forall t \in \mathbb{Z}$,
$$X_t - \sum_{j=1}^{p} \phi_j X_{t-j} = Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j}, \quad (*)$$
where $Z \sim WN(0,\sigma^2)$ and the polynomials $\left(1 - \sum_{j=1}^{p} \phi_j z^j\right)$ and $\left(1 + \sum_{j=1}^{q} \theta_j z^j\right)$ have no common zeros.

* $X \to$ ARMA(p,q) with mean μ
  $X - \mu \to$ ARMA(p,q)

* ARMA(p,q) can be rewritten as
$$\phi(B) X_t = \theta(B) Z_t$$

* A stationary solution X of eq(*) exists and is unique if and only if $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j \neq 0 \; \forall z \in \mathbb{C}$ with $|z| = 1$.

* An ARMA(p,q) process X is causal or a causal fn of Z if $\exists$ a real-valued sequence $(\psi_j, j \in \mathbb{N}_0) \mid \sum_{j=0}^{\infty} |\psi_j| < +\infty$ and $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \; \forall t \in \mathbb{Z}$. $\to$ MA(∞)

* An ARMA(p,q) process X is causal iff $1 - \sum_{j=1}^{p} \phi_j z^j \neq 0 \neq 0 \; \forall z \in \mathbb{C}$ with $|z| \leq 1$.

* A causal ARMA(p,q) process has a unique stationary solution.

* Since X is causal, $\psi(z)\phi(z) = \theta(z)$
  or, $(\psi_0 + \psi_1 z + \dots)(1 - \phi_1 z - \dots - \phi_p z^p) = (1 + \theta_1 z + \dots + \theta_q z^q)$

* An ARMA(p,q) process X is invertible if $\exists$ a real-valued sequence $(\pi_j, j \in \mathbb{N}_0) \mid \sum_{j=0}^{\infty} |\pi_j| < +\infty$ and $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \; \forall t \in \mathbb{Z}$. $\to$ AR(∞)

* An ARMA(p,q) process is invertible iff $1 + \sum_{j=1}^{q} \theta_j z^j \neq 0 \; \forall z \in \mathbb{C}$ with $|z| \leq 1$.

* For ARMA(p,q) process,
$$\gamma(h) = E(X_{t+h} X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}$$

* For ARMA(p,q) process,
$$\gamma(k) - \sum_{j=1}^{p} \phi_j \gamma(k-j) = \sigma^2 \sum_{j=0}^{q-k} \theta_{k+j} \psi_j, \quad 0 \leq k \leq q$$
$$\gamma(k) - \sum_{j=1}^{p} \phi_j \gamma(k-j) = 0, \quad k > q$$
where $\psi_j = 0$ for $j < 0$, $\theta_0 = 1$, $\theta_j = 0$ for $j \notin \{0, \dots, q\}$.

* For ARMA(p,q) process,
  $\alpha(0) = 1$
  $\alpha(h) = \phi_{hh}, \; h \geq 1$
  $=$ last component of ϕ_h
$$\phi_h = \left((\gamma(i-j))_{i,j=1}^{h}\right)^{-1} (\gamma(1), \gamma(2), \dots, \gamma(h))'$$

* To choose an appropriate AR(p) model is to look at the sample PACF $\hat{\alpha}$. If $\hat{\alpha}(h)$ is significantly different from zero, $h = 1, \dots, p$, for $h > p$, an AR(p) model is a good choice for the data. By C.L.T, around 95% of the sample PACF values beyond lag p should fall within the bounds $\pm \frac{1.96}{\sqrt{n}}$ which is justified by the fact that the sample PACF values at lags greater than p are approximately independent $N(0, n^{-1})$ distributed r.v.s.

* For an AR(p) model,
$$\gamma(k) - \sum_{j=1}^{p} \phi_j \gamma(k-j) = \begin{cases} 0, & k \in \{1, \dots, p\} \\ \sigma^2, & k = 0 \end{cases}$$
  $\to$ Yule-Walker equations

* Yule-Walker estimation
$$(\hat{\phi}_1, \dots, \hat{\phi}_p) = \hat{R}_p^{-1} (\hat{\rho}(1), \dots, \hat{\rho}(p))'$$
$$\hat{\sigma}^2 = \hat{\gamma}(0)\left(1 - (\hat{\rho}(1), \dots, \hat{\rho}(p)) \hat{R}_p^{-1} (\hat{\rho}(1), \dots, \hat{\rho}(p))'\right)$$
where $\hat{R}_p = (\hat{\rho}(i-j))_{i,j=1}^{p}$.

* Maximum likelihood estimators of σ^2, ϕ, θ are determined from $\hat{\sigma}^2 = n^{-1} S(\hat{\phi}, \hat{\theta})$ and $(\hat{\phi}, \hat{\theta}) = \arg\min_{(\phi,\theta)} l(\phi, \theta)$
  Here $S(\phi, \theta) = \sum_{j=1}^{n} r_{j-1}^{-1} (X_j - \hat{X}_j)^2$,
  where $\hat{X}_j$ and r_{j-1}^{-1} can be computed using the parameters ϕ, θ.

l is the function given by

$$l(\phi,\theta) = \ln\left(n^{-1} S(\phi,\theta)\right) + n^{-1} \sum_{j=1}^{n} \ln v_{j-1}.$$

The minimization of l is done numerically.

* For MA(q) process,

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q} \theta_j \theta_{j+|h|} & |h| \leq q, \\ 0 & |h| > q. \end{cases}$$

* An autocorrelation function with "q peaks and then zero" along with a slowly decaying partial autocorrelation function indicates a MA(q) model.

* A partial autocorrelation function with "p peaks and then zero" along with a slowly decaying autocorrelation function indicates an AR(p) model.

* AICC – Akaike's information criterion biased-corrected

* BIC – Bayesian information criterion

* AICC criterion – Choose p, q, ϕ_p and θ_q to minimize

$$-2 \ln L(\phi_p, \theta_p, S(\phi_p, \theta_q)/n) + 2n \frac{p+q+1}{n-p-q-2}$$

where $\phi_p = (\phi_1, \cdots, \phi_p)$ and $\theta_q = (\theta_1, \cdots, \theta_q)$.

* BIC criterion – Choose p and q to minimize

$$(n-p-q) \ln\left(n\hat{\sigma}^2/(n-p-q)\right) + n\left(1 + \ln\sqrt{2\pi}\right) + (p+q) \ln\left(\left(\sum_{t=1}^{n} X_t^2 - n\hat{\sigma}^2\right) / (p+q)\right),$$

where $\hat{\sigma}^2$ denotes the maximum likelihood estimate of the white noise variance.

* Model building for ARMA processes

(1) Remove trend and seasonality until the data can be modeled as a stationary ~~time~~ series.

(2) Identify the order of the ARMA model for the time series by either looking at the ACF / PACF or by fitting successively higher order ARMA(p,q) to the data and choose a no. of candidate models with small AICC and/or BIC values.

(3) Estimate the final candidate using maximum likelihood.

(4) Compute the residuals $\hat{R}_t$ for different models and check that they are consistent with the specific distribution and temporal covariance structure for Z_t. The final model is chosen to be the one with residuals $\hat{R}_t$ most like Z_t. Another alternative is to reserve some test data at the start of the process and then compute forecasts for the test data using the candidate models. Then one chooses the final model as the one with the minimum forecast error.

* A SARIMA model without any seasonality is called an ARIMA model.

* Let X be a stochastic process and d a nonnegative integer. Then X is an ARIMA (p,d,q) process if the process Y defined by $Y_t = (1-B^d)^d X_t$ is a causal ARMA (p,q) process.

* We assume that we are given a Gaussian ARMA(p,q) process ($Z \sim \text{IID}\, \mathcal{N}(0,\sigma^2)$). Then for any fixed values ϕ, θ, σ^2, since the innovations $X_1 - \hat{X}_1, \cdots, X_n - \hat{X}_n$ are uncorrelated and Gaussian, they are independent, where $\hat{X}_1 = 0$, $\hat{X}_j = b_j^l((X_1, \cdots X_{j-1}))$, $j \geq 2$.

[CHAP]TER 4 ARCH and GARCH processes

Why we use these models?

① Since Z is IID, so is X which means that for two time points s and t, X_t and X_s should be independent.

② Volatility clustering i.e., periods of low and high variance tend to cluster together.

③ X_t should be Gaussian but one observes "fat tails" (i.e, large +ve and -ve jumps more frequently than one would expect from a normal distribution. This means that the kurtosis $E(X_t^4)/E(X_t^2)^2$ is greater than 3, which is the value under the Gaussian IID model.

* A stochastic process is a random variance model if $X_t = \sigma_t Z_t \ \forall t \in \mathbb{Z}$ and where $Z = (Z_t, t \in \mathbb{Z})$ is IID$(0,1)$ and $\sigma = (\sigma_t, t \in \mathbb{Z})$ is an unspecified stochastic process called the volatility. If X_t can be written as a deterministic fn of $(Z_s, s \leq t) \ \forall t \in \mathbb{Z}$, then X is said to be causal. (independent of future)

* The main idea behind ARCH and GARCH processes is to incorporate the possibility of volatility clustering. One possibility would be to let σ_t be a function of time, but then we could end up with the variance of X_t being non constant, so that we would have a non-stationary model which would be hard to do estimate from data. Instead we allow for non-constant conditional variance

$$Var(X_t|X_{t-1}, X_{t-2}, \dots) = E((X_t - E(X_t))^2 | X_{t-1}, X_{t-2}, \dots) \neq \text{constant} \ \forall t \in \mathbb{Z}.$$

This is called conditional heteroscedasticity. This is accomplished by letting σ_t^2 be an ARMA like process in $X_t = \sigma_t Z_t$.

* A stochastic process $X = (X_t, t \in \mathbb{Z})$ is called an ARCH(p) process if it is stationary and if it satisfies the ARCH equations

$$X_t = \sigma_t Z_t, \text{ where } Z \sim IID(0,1),$$
$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2,$$
$$\alpha_0 > 0, \ \alpha_j \geq 0 \text{ for } j = 1, \dots, p.$$

ARCH - autoregressive conditional heteroscedasticity

* Since X is a causal ARCH(p) process, σ_t^2 can be written as a deterministic fn of $(Z_s, s \leq t-1)$ and is therefore independent of the random variables $(Z_s, s \geq t)$.

* $E(X_t^2) = \dfrac{\alpha_0}{1-\alpha(1)}$ where $1-\alpha(1) > 0$.

* If $E(X_t^4) < \infty$, it can be shown that $\{X_t^2, t \in \mathbb{Z}\}$ is an AR process, a fact that can be useful in the identification of ARCH processes.

* A stochastic process $X = (X_t, t \in \mathbb{Z})$ is called a GARCH(p,q) process if it is a stationary solution to the GARCH equations $X_t = \sigma_t Z_t$, where $Z \sim IID(0,1)$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2$$

with $\alpha_0 > 0$, $\alpha_j \geq 0$ for $j = 1, \dots, p$, $\beta_i \geq 0$ for $i = 1, \dots, q$.

* Existence of a GARCH(1,1) process - If $\alpha_1 + \beta_1 < 1$, there exists a stationary solution $X = (X_t, t \in \mathbb{Z})$ to the GARCH(1,1) equations that is given by the equation $X_t = \sigma_t Z_t$ where $Z \sim IID(0,1)$ and

$$\sigma_t^2 = \alpha_0 \left(1 + \sum_{i=1}^{\infty} (\alpha_1 Z_{t-1}^2 + \beta_1)(\alpha_1 Z_{t-2}^2 + \beta_1) \cdots (\alpha_1 Z_{t-i}^2 + \beta_1)\right)$$

It is unique, strictly stationary and causal. Conversely if $\alpha_1 + \beta_1 \geq 1$, then there is no non-zero stationary soln to the GARCH(1,1) equations for which σ_t can be written as a deterministic fn of $(Z_s, s < t) \ \forall t \in \mathbb{Z}$.

* For causal GARCH(p,q) process, since σ_t only depends on $(Z_r, r < t)$, it is independent of Z_t and we get for $s < t$

$$Cov(X_t, X_s) = E(Z_t Z_s \sigma_t \sigma_s) = E(Z_t) E(\sigma_t \sigma_s Z_s) = 0.$$

* For GARCH(1,1) model, the kurtosis is

$$\frac{E(X_t^4)}{E(X_t^2)^2} = \frac{\mu_4 (1 - (\alpha_1 + \beta_1)^2)}{1 - \beta_1^2 - 2\alpha_1\beta_1 - \mu_4 \alpha_1^2}.$$

* Existence of a GARCH(p,q) process -

If $\alpha(1)+\beta(1) = \sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 1$, there exists a unique weakly and strictly stationary causal soln $X=(X_t, t\in\mathbb{Z})$ to the GARCH(p,q) eqn s and a real-valued sequence $(\psi_j)_{j=0}^{\infty}$ such that $\sum_{j=0}^{\infty}|\psi_j|<\infty$ and σ^2 is given by

$$\sigma_t^2 = \psi_0 + \sum_{j=1}^{\infty}\psi_j X_{t-j}^2 .$$

Conversely, if $\alpha(1)+\beta(1) = \sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j \not< 1$, then no stationary and causal solution to the GARCH(p,q) eqn s for which σ_t can be written as a deterministic fn of $(Z_s, s<t)$ $\forall t\in\mathbb{Z}$ exists.

* For GARCH(p,q) process,

$$E[X_t^2] = E[\sigma_t^2] = \frac{\alpha_0}{1-\alpha(1)-\beta(1)} .$$

Under the assumption that $E[\sigma_t^4]<\infty$, one can derive that $(X_t^2, t\in\mathbb{Z})$ is an ARMA$(\max\{p,q\}, q)$ process with generating polynomials $\phi(z)=1-\alpha(z)-\beta(z)$ and $\theta(z)=1-\beta(z)$ with mean $\alpha_0(1-\alpha(1)-\beta(1))^{-1}$

i.e., a process that can be represented by

$$X_t^2 - \sum_{i=1}^{\max\{p,q\}}(\alpha_i+\beta_i)X_{t-i}^2 = \alpha_0 + U_t - \sum_{j=1}^{q}\beta_j U_{t-j}$$

where $\alpha_i=0$ for $i>p$, $\beta_i=0$ for $i>q$ and $U_t = X_t^2-\sigma_t^2$ can be shown to be white noise.

* Conditional maximum likelihood estimation - The conditional maximum likelihood estimators $(\hat\alpha_0,\dots,\hat\alpha_p,\hat\beta_1,\dots,\hat\beta_q,\hat\theta_Z)$ are obtained as the values that maximize the conditional likelihood fn

$$L(\alpha_0,\dots,\alpha_p,\beta_1,\dots,\beta_q,\theta_Z) = \prod_{t=p+1}^{n}\frac{1}{\sigma_t}f_Z\left(\frac{x_t}{\sigma_t}\right)$$

where f_Z is the density of the white noise Z and θ_Z is any other parameter Z depends on. The likelihood fn is a fn of the parameters $\alpha_0,\dots,\alpha_p,\beta_1,\dots,\beta_q$ via σ_t, which is computed recursively, along with supposing that $\sigma_t=\sqrt{\hat\sigma^2}$, $t\le 0$
$X_t = 0$, $t\le 0$
where $\hat\sigma^2$ is the sample variance of $\{x_1,\dots,x_n\}$.

Equivalently, one obtains $(\hat\alpha_0,\dots,\hat\alpha_p,\hat\beta_1,$ as the values that minimize $-\ln L(\alpha_0,\dots,\alpha_p,$

* The residuals for this estimation are numbers $(x_t/\hat\sigma_t, t=p+1,p+2,\dots,n)$ where $\hat\sigma_t$ is the approx forecasted conditional variance given $(Z_s, s<t)$.

* $Z\sim \text{IIDN}(0,1)$

$$-\ln L(\alpha_0,\dots,\alpha_p,\beta_1,\dots,\beta_q) = \frac{1}{2}\sum_{t=p+1}^{n}\left(\ln 2\pi + \ln\sigma_t^2 + \frac{X_t^2}{\sigma_t^2}\right)$$

* Two-pass estimation of GARCH -

$$\left.\begin{aligned}\beta_i &= \hat\theta_i\\ \hat\alpha_i &= \hat\phi_i - \hat\theta_i\end{aligned}\right\}\ \text{Use M.L method to estimate parameters of ARMA } X^2 \text{ denoted by } \hat\phi_i \text{ and } \hat\theta_i .$$

* AICC criterion

$$-2\,\frac{n}{n-p}\ln L(\alpha_0,\dots,\alpha_p,\beta_1,\dots,\beta_q,\theta_Z) + 2n\,\frac{m+1}{n-m-2}$$

n = sample size, m = no. of non zero est. para.

* A time series X is said to be an ARMA-GARCH process if it is an ARMA(p,q) process driven by GARCH$(\tilde p,\tilde q)$ noise i.e., if it is stationary and

$$X_t - \sum_{j=1}^{p}\phi_j X_{t-j} = \sigma_t Z_t + \sum_{j=1}^{q}\theta_j\sigma_{t-j}Z_{t-j}$$

where $Z\sim \text{IID}(0,1)$, the polynomials $(1-\sum_{j=1}^{p}\phi_j z^j)$ and $(1+\sum_{j=1}^{q}\theta_j z^j)$ have no common zeros and $\sigma_t^2 = \alpha_0 + \sum_{j=1}^{\tilde p}\alpha_j\sigma_{t-j}^2 Z_{t-j}^2 + \sum_{i=1}^{\tilde q}\beta_i\sigma_{t-i}^2$ with $\alpha_0>0$, $\alpha_j\ge 0$ for $j=1,\dots,\tilde p$, $\beta_i\ge 0$ for $i=1,\dots,\tilde q$.

* $X \to$ ARMA - GARCH with mean μ
$\Rightarrow X-\mu \to$ ARMA - GARCH

* A stochastic process $X=(X_t, t\in\mathbb{Z})$ is called an EGARCH(p,q) process if it is stationary and satisfies the EGARCH eqn s $X_t = \sigma_t Z_t$, where $Z\sim \text{IID}(0,1)$ has a symmetric distribution i.e. Z_t and $-Z_t$ have the same distribution.

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{j=1}^{p}\alpha_j g(Z_{t-j}) + \sum_{i=1}^{q}\beta_i\ln(\sigma_{t-i}^2)$$

where $g(x) = x + \lambda(|x| - E(|Z_t|))$ and $\alpha_0,\alpha_1,\dots,\alpha_p,\beta_1,\dots,\beta_q,\lambda\in\mathbb{R}$.

* EGARCH(p,q) - ① $E(Z_t|Z_t|) = 0$ $\forall t\in\mathbb{Z}$.
② $g(Z) = (g(Z_{t-1}), t\in\mathbb{Z}) \sim \text{WN}(0, 1+\lambda^2\,\text{Var}(|Z_t|))$
③ $(\ln(\sigma_t^2)/\alpha_{p'}, t\in\mathbb{Z})$ is an ARMA$(q, p-p')$ with mean $\mu = \alpha_0/(\alpha_{p'}(1-\beta(1)))$.
[p' is the first $j\in\mathbb{N}$ such that $\alpha_j\neq 0$.]

CHAPTER 5 Nonlinear models and nonparametric methods

A stochastic process $X = (x_t, t \in T)$ on some index set T is a Markov process if its conditional distribution function satisfies

$$P(X_h | X_s, s \leq t) = P(X_h | X_t) \text{ for } h > t.$$

If X is a discrete-time stochastic process, i.e., $T = \mathbb{N}$ or $\mathbb{Z}$, then the property becomes $P(X_h | X_t, X_{t-1}, \cdots) = P(X_h | X_t)$ for $h > t$ and the process is known as Markov chain.

* Example –

A time series $X = (X_t, t \in \mathbb{Z})$ follows a Markov switching autoregressive model (MSA) with two states if it satisfies

$$X_t = \begin{cases} c_1 + \sum_{i=1}^{p} \phi_{1i} X_{t-i} + Z_{1t} & \text{if } S_t = 1, \\ c_2 + \sum_{i=1}^{p} \phi_{2i} X_{t-i} + Z_{2t} & \text{if } S_t = 2 \end{cases}$$

where S assumes values in $\{1, 2\}$ and is a Markov chain with transition probabilities

$$P(S_t = 2 | S_{t-1} = 1) = \omega_1,$$
$$P(S_t = 1 | S_{t-1} = 2) = \omega_2$$

with $\omega_1, \omega_2 \in [0, 1]$. The time series $Z_1 = (Z_{1t}, t \in \mathbb{Z})$ and $Z_2 = (Z_{2t}, t \in \mathbb{Z})$ are IID$(0, \sigma^2)$ noises and independent of each other.

* Parametric bootstrap –

Given data $(x_1, x_2, \cdots, x_n)$, we want to forecast x_{n+h} for $h > 0$. We denote this forecast by $x_n(h)$. The parametric bootstrap computes forecasts of $x_{n+1}, \cdots, x_{n+h}$ sequentially as follows. For $i = 1, \cdots, h$ repeat:

(i) Generate a random sample of the driving noise at time $n+i$ according to the underlying model.

(ii) Compute $\tilde{x}_{n+i}$ using the generated sample, the model, the data, and the previous forecasts $x_n(1), \cdots, x_n(i-1)$.

(iii) Repeat the prev two steps K times to get K realizations $(\tilde{X}^{(k)}_{n+i}, k = 1, \cdots, K)$. A point forecast for X_{n+i} is then obtained via $X_n(i) = K^{-1} \sum_{k=1}^{K} \tilde{X}^{(k)}_{n+h}$.

* Directional measure.

Actual \ Predicted	Up	Down	
Up	m_{11}	m_{12}	m_{10}
Down	m_{21}	m_{22}	m_{20}
	m_{01}	m_{02}	m

$m = N - n = m_{01} + m_{02} = m_{10} + m_{20}$

Test statistic $\chi^2 = \sum_{i,j=1}^{2} \frac{(m_{ij} - m_{i0} m_{0j}/m)^2}{m_{i0} m_{0j}/m}$

∴ We say that our forecasts are better than random choice at level α if $\chi^2 \geq \chi^2_{1-\alpha}$ where $\chi^2_{1-\alpha}$ is the $1-\alpha$ percentile of the χ^2 distribution with 1 degree of freedom.

[For eg, the forecasts are better than random choice at the 5% level if $\chi^2 \geq \chi^2_{0.95}$]

$\alpha = 0.95$ d.o.f $= 1$

$\chi^2_{0.95,1} = 3.841$.

① Testing data

iid
seasonality
stationarity
(non)linearity
(G)ARCH effects

background of data
ACVF
stochastic variance
plot | visualize

② Model fitting

order selection (ARMA | GARCH)
parameter estimation (ARMA | GARCH)
nonparametric methods

③ forecasting

best (linear) predictors
ARMA forecasting
parametric bootstrap

one step ahead
iid (distr)

④ Error analysis

MSE
evaluation | prediction
sub sample

nonlinear method -
confidence intervals