Time Series Analysis.

Model: • discrete time: $t = 1, 2, \cdots, T$

• continuous time: $t \in [0, 1]$. e.g. Stock Price.

Observe x_t, the realized value of some r.v. X_t

$\{x_t : t \in T_0\}$: data set. It is a realization of $\{X_t : t \in T_0\}$, which comes from $\{X_t : t \in T_1\}$, $T_0 \subset T_1$.

$\rightarrow$ Use Kolmogorov's theorem to guarantee that X_t exists.

• **Stationarity**: past $\approx$ future.

Def. The t.s. $\{X_t\}$ is strictly stationary if $\forall k \in \mathbb{N}$, $t_1\, t_2 \cdots t_k$, the joint dist. of $(X_{t_1}, X_{t_2}, \cdots, X_{t_k})$ and $(X_{t_1+h}, X_{t_2+h}, \cdots, X_{t_k+h})$ are the same.

intuition: shift time by $h \rightarrow$ statistical properties of the observed data are similar.

Problem: cannot test str. stat. (infinite set of $t_1 < t_2 < \cdots < t_k$ we have)

$k=1$: • X_1, X_2, $\cdots$, X_T has the same dist.

$k=2$: • $(X_1, X_2), (X_2, X_3), \cdots, (X_{T-1}, X_T)$ ......

• $(X_1, X_3), (X_2, X_4), \cdots, (X_{T-2}, X_T)$

$k=3$: • $(X_1, X_2, X_3), (X_2, X_3, X_4) \cdots$

Def Let $var(X_t) < \infty$, $\forall t$. The autocovariance $\gamma_X(\cdot, \cdot)$ of X_t is $\gamma_X(r, s) = cov(X_r, X_s) = \mathbb{E}[(X_r - \mathbb{E}[X_r])(X_s - \mathbb{E}[X_s])]$ $\forall r, s \in T_1$

$(\{X_t : t \in T_1\})$

Def The t.s. $\{X_t\}$ is weakly stationary (covariance stationary) if

(i) $\mathbb{E}[|X_t|^2] < \infty$, $\forall t$

(ii). $\mathbb{E}[X_t] = \mu$, $\forall t$

(iii). $\gamma_X(r, s) = \gamma_X(r+t, s+t)$, $\forall r, s, t$

Remark: If X_t is weakly stat. then $\gamma_X(r, s) = \gamma_X(r-s, 0)$

Let $\gamma_X(h) = \gamma_X(h, 0) = cov(X_{t+h}, X_t)$ $\forall t$, then $\gamma_X(\cdot)$ is the autocovariance of X_t.

f-n: function.

$\gamma(0) = cov(X_t, X_t) = Var(X_t)$

Def Autocorrelation f-n of weakly stat. t.s. $\{X_t\}$ is $\rho_X(h) = \frac{\gamma(h)}{\gamma(0)} = corr(X_{t+h}, X_t)$, $\forall t, h$

Other notations. $\gamma_h = \gamma_X(h)$

$\rho_h \equiv \gamma_X(h)/\gamma_X(0)$

• **Relationship btw. weak & strong stationary:**

1). $X_t \sim$ i.i.d Cauchy: $f(x; \mu, \gamma) = \frac{1}{\pi\gamma[1 + (\frac{x-\mu}{\gamma})^2]}$ heavy-tailed

$\Rightarrow \{X_t\}$ is str. stat.

$\{X_t\}$ is not weak stat since $\mathbb{E}[|X_t|^2] = \infty$

2) X_t str. stat + $\mathbb{E}[|X_1|^2] < \infty$

$\Rightarrow X_t$ weak stat.

$\mathbb{E}[|X_1^2|] < \infty \Rightarrow \mathbb{E}[|X_t|^2] < \infty$ $\forall t$

$\Downarrow$

$\mathbb{E}[X_1] < \infty \Rightarrow \mathbb{E}[X_t] < \infty$ $\forall t$

str. stat. with $k=2$: $\gamma_X(t+h, s+h) = \gamma_X(t, s)$

3) $X_t \sim$ independent.

$X_t \sim \begin{cases} N(1,1) & t\text{-even} \\ exp(1) & t\text{-odd} \end{cases}$

exponential: $f(x; \lambda) = \lambda e^{-\lambda x}$ $x \geq 0$

$\Rightarrow \mathbb{E}[X_t] = 1 \quad \forall t, \quad \mathbb{E}[|X_t|^2] < \infty, \ \forall t$

$$\gamma_X(h) = \begin{cases} 1 & \text{if } h=0 \\ 0 & \text{if } h \neq 0 \end{cases}$$

$\Rightarrow X_t$ is weakly stat. $\Rightarrow X_t$ is not str. stat. (X_1 is exponential, X_2 is normal)

e.g. 1. $Z_t \sim iid(0, \sigma^2)$

$X_t = Z_t + \theta Z_{t-1}, \ \theta \in \mathbb{R}$

Then. $\mathbb{E}[X_t] = 0$

$cov(Z_t + \theta Z_{t-1}, Z_t + \theta Z_{t-1}) = var(Z_t + \theta Z_{t-1}) = \sigma^2 + \theta^2\sigma^2$

$cov(X_{t+h}, X_t) = cov(Z_{t+h} + \theta Z_{t+h-1}, Z_t + \theta Z_{t-1})$

$cov(Z_{t+1} + \theta Z_t, Z_t + \theta Z_{t-1}) = cov(\theta Z_t, Z_t) = \theta \, var(Z_t) = \theta\sigma^2$

$$= \begin{cases} \sigma^2 + \theta^2\sigma^2 & \text{if } h=0 \\ \theta\sigma^2 & \text{if } |h|=1 \\ 0 & \text{if } |h|>1. \end{cases}$$

$\Rightarrow X_t$ is weakly stat. In fact also strictly stat.

e.g. 2. Y_t : weakly stat.

$$X_t = \begin{cases} Y_t & t\text{-even} \\ Y_t + 1 & t\text{-odd} \end{cases}$$

Then, $cov(X_{t+h}, X_t) = cov(Y_{t+h}, Y_t) = \gamma_Y(h)$

However, $\mathbb{E}[X_t] \neq$ constant. $\mathbb{E}[X_1] = \mathbb{E}[Y_1] + 1 \neq \mathbb{E}[X_2] = \mathbb{E}[Y_2]$

X_t is not weakly stat.

e.g. 3. (Random Walk)

$X_t \sim iid(0, \sigma^2)$

$S_t = x_1 + \cdots + x_t$

then. $\mathbb{E}[S_t] = 0, \ \forall t$

$cov(S_{t+h}, S_t) = cov(X_1 + X_2 + \cdots + X_{t+h}, X_1 + X_2 + \cdots + X_t)$

$= cov(X_1 + X_2 + \cdots + X_t, X_1 + X_2 + \cdots + X_t)$

$= \sigma^2 + \cdots + \sigma^2$

$= t\sigma^2 \quad \Rightarrow S_t$ is not weakly stat.

• **Proposition.** If $\gamma(\cdot)$ is the autocovariance f-n of weakly stat. X_t, $\rho(\cdot)$ is its autocorrelation f-n.

i) $\gamma(0) \geq 0, \quad \rho(0) = 1$

ii) $|\gamma(h)| \leq \gamma(0), \ |\rho(h)| \leq 1, \quad \forall h$

iii) $\gamma(h) = \gamma(-h), \ \rho(h) = \rho(-h)$

proof : i). $\gamma(0) = cov(X_t, X_t) = var(X_t) \geq 0$

$\rho(0) = \gamma(0)/\gamma(0) = 1$

iii) $\gamma(h) = cov(X_{t+h}, X_t) = cov(X_t, X_{t-h}) = cov(X_{t-h}, X_t) = \gamma(-h)$

$\rho(h) = \gamma(h)/\gamma(0) = \gamma(-h)/\gamma(0) = \rho(-h)$

ii) $|cov(X_{t+h}, X_t)| \leq \underbrace{\sqrt{var(X_{t+h})}}_{=\gamma(0)} \underbrace{\sqrt{var(X_t)}}_{=\gamma(0)}$ (Cauchy-Bunyakovsky-Schwartz ineq.)

proof: X, Y. r.v. $var(\lambda X + Y) \geq 0 \quad \forall \lambda \in \mathbb{R}$

$\lambda^2 var(X) + 2\lambda\, cov(X, Y) + var(Y) \geq 0 \quad \forall \lambda \in \mathbb{R}$.

$\therefore \ 4\, cov(X, Y)^2 - 4 var(X) var(Y) \leq 0$

$|cov(X, Y)| \leq \sqrt{var(X)}\sqrt{var(Y)}$.

$\cos(\alpha)\cos(\beta) = \frac{1}{2}\cos(\alpha+\beta) + \frac{1}{2}\cos(\alpha-\beta)$. $\quad \cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$.

• Ergodicity - LLN and CLT for time series • Long-run variance, HAC

Sample Setting:

$\{x_1, \cdots, x_T\}$ observed

$x_{T+1}, x_{T+2}, \cdots$ out of sample
$x_0, x_{-1}, \cdots$ pre-sample
} unobserved   estimate $\gamma_X(\cdot), \rho_X(\cdot)$

Def. The sample autocovar f-n of $\{x_1, \cdots, x_T\}$ is

$$\begin{cases} \hat{\gamma}(h) := \frac{1}{T}\sum_{j=1}^{T-h}(x_{j+h}-\bar{x})(x_j-\bar{x}) & , 0 \le h < T \\ \hat{\gamma}(h) := \hat{\gamma}(-h) & , -T < h < 0 \end{cases}$$

where $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$

The sample autocorr. is $\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$, $|h| < T$

When do sample moments converge to the true moments?

In cross-sectional, usually $X_t \sim$ iid, so LLN implies sample $\xrightarrow{p}$ true

e.g. $\varepsilon_t \sim iid\ N(0, \sigma^2)$

$\mu \sim N(0, \lambda^2)$, $\varepsilon_t \perp\!\!\!\perp \mu$

$Y_t = \mu + \varepsilon_t$, $t = 1, 2, 3, \cdots$

$\to \mathbb{E}[Y_t] = \mathbb{E}[\mu] + \mathbb{E}[\varepsilon_t] = 0$

$\frac{1}{T}\sum_{t=1}^{T} Y_t = \frac{1}{T}\sum_{t=1}^{T}(\mu + \varepsilon_t) = \mu + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t \xrightarrow{p} \mu \neq \mathbb{E}[Y_t]$

$\mathrm{cov}(Y_{t+h}, Y_t) = \begin{cases} \sigma^2 + \lambda^2 & , h = 0 \\ \lambda^2 & , h \neq 0 \end{cases} = \mathbb{E}[(\mu + \varepsilon_{t+h})(\mu + \varepsilon_t)]$

$\therefore Y_t$ is weakly stat. but $\frac{1}{T}\sum_{t=1}^{T} Y_t \not\xrightarrow{p} \mathbb{E}[Y_t]$

Def. A weakly stat. t.s. $\{X_t\}$ is ergotic for the mean if $\frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow{p} \mathbb{E}[X_t]$.

Intuition: 1) Many Universes

2) Observe one univ. with $\{y_1, \cdots, y_T\}$

3) Other universes have $\{y_1^{(i)}, y_2^{(i)}, \cdots, y_T^{(i)}\}$, $i = 1, 2, \cdots$

$\mathbb{E}[y_t] = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} y_t^{(i)}$

Ergodicity: temporal avg = spacial avg.

lack of memory across t. i.e. can use LLN

Stationarity: similarity across t

Lemma. If $\{X_t\}$ is weakly stationary and $\sum_{j=-\infty}^{\infty}|\gamma(j)| < \infty$, then $\frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow{p} \mathbb{E}[X_t]$, ergotic for the mean.

proof. WTS $\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{T} X_t - \mathbb{E}[X_t]| \ge \varepsilon) \to 0 \quad \forall \varepsilon$.

First, we show the Markov Inequality $\mathbb{P}(Y \ge c) \le \frac{\mathbb{E}[Y]}{c}$, $Y \ge 0$ r.v.

$\mathbb{E}[Y] = \int_0^{+\infty} y\,dF(y) \ge \int_c^{+\infty} y\,dF(y) \ge \int_c^{+\infty} c\,dF(y) = c\,\mathbb{P}(Y \ge c)$

Apply it. $\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{T} X_t - \mu|^2 \ge \varepsilon^2) \le \frac{1}{\varepsilon^2}\mathbb{E}[(\frac{1}{T}\sum_{t=1}^{T} X_t - \mu)^2]$

$= \frac{1}{\varepsilon^2}\mathrm{var}(\frac{1}{T}\sum_{t=1}^{T} X_t)$

$= \frac{1}{\varepsilon^2}\frac{1}{T^2}\sum_{i=1}^{T}\sum_{j=1}^{T}\mathrm{cov}(X_i, X_j)$

$= \frac{1}{\varepsilon^2}\frac{1}{T^2}(\gamma(0)T + \gamma(1)\underbrace{2(T-1)}_{(1\,2),(2\,1),\cdots} + \gamma(2)\underbrace{2(T-2)}_{(1\,3),(3\,1)\cdots} + \cdots + \gamma(T-1)\underbrace{\cdot 2}_{(1\,T),(T\,1)})$

$= \frac{1}{\varepsilon^2}\frac{1}{T}(\gamma(0) + 2(1-\frac{1}{T})\gamma(1) + 2(1-\frac{2}{T})\gamma(2) + \cdots + 2(1-\frac{T-1}{T})\gamma(T-1))$

$= \frac{1}{\varepsilon^2}\frac{1}{T}(\gamma(0) + 2\sum_{i=1}^{T-1}\gamma(i)\underbrace{(1-\frac{i}{T})}_{<1})$

$\le \frac{1}{\varepsilon^2}\frac{1}{T}\underbrace{\sum_{i=-\infty}^{\infty}|\gamma(i)|}_{<\infty}$ since $\gamma(h) = \gamma(-h)$

$\longrightarrow 0$ as $T \to \infty$

$\therefore \frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow{p} \mu$ as $T \to \infty$.

Corollary. Variance in CLT changes. with what?

If $X_t \sim iid(\mu, \sigma^2)$, then $\mathrm{var}(\frac{1}{T}\sum_{t=1}^{T} X_t) = \frac{1}{T^2}\cdot T\sigma^2 = \frac{\sigma^2}{T}$

Now, $\mathrm{var}(\frac{1}{T}\sum_{t=1}^{T} X_t) = \frac{1}{T}(\underbrace{\gamma(0)}_{\sigma^2} + \underbrace{2\sum_{i=1}^{T-1}\gamma(i)(1-\frac{i}{T})}_{\text{correlation across time}})$

$\gamma(0) + 2\sum_{i=1}^{T-1}\gamma(i)(1-\frac{i}{T}) \xrightarrow[T\to\infty]{} \sum_{i=-\infty}^{\infty}\gamma(i) = \gamma(0) + 2\sum_{i=1}^{\infty}\gamma(i)$ = long-run variance

? Why not $\frac{1}{T}$.

CLT

Theorem: Let $X_t = \mu + \sum_{j=0}^{\infty} c_j\varepsilon_{t-j}$, $\varepsilon_t \sim indep(0, \sigma^2)$, $\sum_{j=0}^{\infty}|c_j| < \infty$.

Then: $\sqrt{T}(\frac{1}{T}\sum_{t=1}^{T} X_t - \mu) \xrightarrow{d} N(0, J)$, $J = \sum_{h=-\infty}^{\infty}\gamma_X(h)$

proof?

Remark: $J < \infty$ because:

$h \ge 0$: $\gamma(h) = \mathrm{cov}(X_{t+h}, X_t) = \mathrm{cov}(\mu + \sum_{j=0}^{\infty} c_j\varepsilon_{t+h-j}, \mu + \sum_{j=0}^{\infty} c_j\varepsilon_{t-j})$

$= \mathrm{cov}(\mu + \sum_{j'=-h}^{\infty} c_{j'+h}\varepsilon_{t-j'}, \mu + \sum_{j=0}^{\infty} c_j\varepsilon_{t-j})$

$= \sum_{j=0}^{\infty} c_{j+h}c_j\sigma^2$

$c_0\varepsilon_{t+h}$	$c_0\varepsilon_t$
$c_1\varepsilon_{t+h-1}$	$c_1\varepsilon_{t-1}$
$\vdots$	$\vdots$
$c_h\varepsilon_{t+h-h}$	$c_h\varepsilon_{t-h}$
$c_{h+1}\varepsilon_{t+h-(h+1)}$	$c_{h+1}\varepsilon_{t-h-1}$

$J = \sum_{h=-\infty}^{\infty}\gamma_X(h)$

by $\gamma(h) = \gamma(-h)$

$= \gamma_X(0) + 2\sum_{h=1}^{\infty}\gamma_X(h)$

$= \sum_{i=0}^{\infty} c_i^2\sigma^2 + 2\sum_{h=1}^{\infty}\sum_{j=0}^{\infty} c_{j+h}c_j\sigma^2$

$= \sigma^2(\sum_{i=0}^{\infty} c_i^2 + 2\sum_{h=1}^{\infty}\sum_{j=0}^{\infty} c_{j+h}c_j)$

$= \sigma^2(2\sum_{h=0}^{\infty}\sum_{j=0}^{\infty} c_{j+h}c_j - \sum_{i=0}^{\infty} c_i^2)$

$(c_0^2 + c_1^2 + \cdots) + 2(c_1c_0 + c_2c_1 + c_2c_0 + \cdots)$

$2(c_0^2 + c_1c_0 + c_1^2 + c_2c_1 + c_2c_0 + c_2^2 + \cdots)$

From $\sum_{j=0}^{\infty}|c_j| < \infty$:

$|\sum_{h=0}^{\infty}\sum_{j=0}^{\infty} c_{j+h}c_j| \le \sum_{h=0}^{\infty}|\sum_{j=0}^{\infty} c_{j+h}c_j| \le \sum_{h=0}^{\infty}\sum_{j=0}^{\infty}|c_{j+h}||c_j|$

$\le (\sum_{i=0}^{\infty}|c_i|)^2 < \infty$

$|\sum_{i=0}^{\infty} c_i^2| < \infty$

$\therefore J = \sigma^2(\sum_{i=0}^{\infty} c_i^2 + 2\sum_{h=1}^{\infty}\sum_{j=0}^{\infty} c_{j+h}c_j) < \infty$

How to estimate J?

$J = \sum_{i=-\infty}^{\infty} \gamma(i)$ $\quad \hat{J} \underset{=}{?=} \sum_{h=-(T-1)}^{T-1} \hat{\gamma}(h)$ where

$$\hat{\gamma}(h) = \frac{1}{T}\sum_{j=1}^{T-h}(x_{j+h}-\bar{x})(x_j-\bar{x}) \ , \ T > h \geq 0$$

Why not $T-h$?

$$\hat{\gamma}(h) = \hat{\gamma}(-h), \ h \in [-(T-1), 0)$$

If $T-h$, this step can't happen

$$\hat{J} = \hat{\gamma}(0) + 2\sum_{h=1}^{T-1}\hat{\gamma}(h)$$

$$= \frac{1}{T}\left\{\sum_{j=1}^{T}(x_j-\bar{x})^2 + 2\sum_{h=1}^{T-1}\sum_{j=1}^{T-h}(x_{j+h}-\bar{x})(x_j-\bar{x})\right\}$$

$$(a_1+a_2+\cdots+a_n)^2$$
$$= a_1^2+a_2^2+\cdots+a_n^2$$
$$+2(a_2a_1+a_3a_2+a_3a_1+\cdots a_Ta_1)$$

$h=1$: a_2a_1, a_3a_2, …, $a_{T-1}a_{T-2}$, a_Ta_{T-1}

$h=2$: a_3a_1, a_4a_2, …, $a_{T-1}a_{T-3}$, a_Ta_{T-2}

…… $h=T-1$: a_Ta_1

symmetric.

$$= \frac{1}{T}\left(\sum_{j=1}^{T}(x_t-\bar{x})\right)^2$$

$$= \frac{1}{T}\left(\sum_{t=1}^{T}x_t - T\bar{x}\right)^2$$

$$= 0$$

If μ is known, $\bar{x} \to \mu$

$$\hat{J} = \frac{1}{T}\left(\sum_{t=1}^{T}(x_t-\mu)\right)^2$$

$$= \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\mu)\right)^2 \xrightarrow{d} \text{normal}^2$$

lecture note: $\hat{\gamma}(T-1) = \frac{1}{T}x_1x_T$?

However, due to something like $\hat{\gamma}(T-1) = \frac{1}{T}(x_T-\bar{x})(x_1-\bar{x})$

using this estimation formula, it has bad precision. Small # of observations

$\to$ Truncate: $\frac{1}{T}\sum_{h=-S_T}^{S_T}\hat{\gamma}(h)$, $S_T \ll T$ $\to$ can get $\hat{J} < 0$

HAC: Heteroskedasticity and autocorrelation consistent estimator

$$\to \hat{J} = \sum_{h=-(T-1)}^{T-1} \underbrace{k}_{\text{kernel}}\left(\frac{h}{\underbrace{S_T}_{\text{bandwidth}}}\right)\hat{\gamma}(h)$$

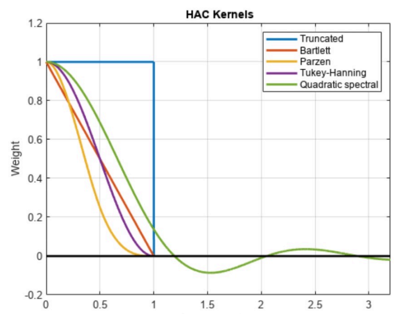
need $S_T \to \infty$ as $T \to \infty$.

Idea: assign larger weight to those that have smaller h, because they have more observation. h: distance.

e.g. $k(x) = \begin{cases} 1, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases}$ truncated kernel

If $x \in \mathbb{R}^n$ $n>1$, then can have $\hat{J}$ non-positive definite.

$k(x) = \begin{cases} 1-|x|, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases}$ Barklett kernel



QS $\quad k(x) = \frac{25}{12\pi^2x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$

$S_T \sim T^{1/5}$ $\quad$ QS achieves optimality.

- **White Noise**

Def. White noise is a sequence of r.v. $\{\varepsilon_t\}$ s.t. $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$, $\forall t \neq s$ $\quad \varepsilon_t \sim WN$.

If in addition, $\mathbb{E}[\varepsilon_t] = 0$ $\forall t$, $\text{Var}(\varepsilon_t) = \sigma^2$ $\forall t$,

$\varepsilon_t \sim WN(0, \sigma^2)$ $\quad$ Not a distribution!

- If $\varepsilon_t \sim iid(0, \sigma^2)$, then $\varepsilon_t \sim WN(0, \sigma^2)$

Indep $\Rightarrow$ $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$ $\quad \forall t \neq s$.

- $\varepsilon_t \sim iid(0, \sigma^2)$, $Z \perp\!\!\!\perp \varepsilon_t$, $\mathbb{E}[Z^2] < \infty$

$U_t := \varepsilon_t \cdot Z$

$\mathbb{E}[U_t] = \mathbb{E}[\varepsilon_t Z] = \mathbb{E}[\varepsilon_t]\mathbb{E}[Z] = 0$

$\mathbb{E}[U_t^2] = \mathbb{E}[\varepsilon_t^2 Z^2] = \mathbb{E}[\varepsilon_t^2]\mathbb{E}[Z^2] = \sigma^2\mathbb{E}[Z^2]$

$\forall t \neq s$: $\text{cov}(U_t, U_s) = \mathbb{E}[\varepsilon_t Z \varepsilon_s Z]$

$= \mathbb{E}[\varepsilon_t]\mathbb{E}[\varepsilon_s]\mathbb{E}[Z]^2 = 0$

$\Rightarrow U_t \sim WN(0, \sigma^2\mathbb{E}[Z^2])$

☆ U_t, U_s are dependent because of Z.

- **Moving Average**

$MA(1)$: $Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$

$MA(q)$: $Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$

$MA(\infty)$: $Y_t = \mu + \varepsilon_t + \sum_{j=1}^{\infty}\theta_j\varepsilon_{t-j} = \mu + \sum_{j=0}^{\infty}\theta_j\varepsilon_{t-j}$

$\theta_0 = 1$, $\sum_{j=0}^{\infty}\theta_j^2 < \infty$, $\varepsilon_t \sim WN(0, \sigma^2)$

$\Uparrow$ implies

Other Assumptions: $\sum_{j=0}^{\infty}|\theta_j| < \infty$, $\sum_{j=0}^{\infty}j|\theta_j| < \infty$

MA is always covariance stationary.

$\mathbb{E}[Y_t] = \mathbb{E}[\mu + \sum_{j=0}^{\infty}\theta_j\varepsilon_{t-j}] = \mu$.

$\theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots$

$\theta_0\varepsilon_{t+k} + \theta_1\varepsilon_{t+k-1} + \cdots + \theta_k\varepsilon_t + \theta_{k+1}\varepsilon_{t-1} + \cdots$

$k \geq 0$, $\gamma(k) = \text{cov}(Y_t, Y_{t+k})$

$= \text{cov}(\mu + \sum_{j=0}^{\infty}\theta_j\varepsilon_{t-j}, \mu + \sum_{j=0}^{\infty}\theta_j\varepsilon_{t+k-j})$

$= \text{cov}(\sum_{j=0}^{\infty}\theta_j\varepsilon_{t-j}, \sum_{j'=-k}^{\infty}\theta_{j'+k}\varepsilon_{t-j'})$ $\quad j' = j-k$

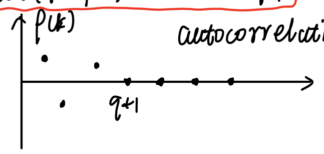
$= \sigma^2\sum_{j=0}^{\infty}\theta_j\theta_{j+k} < \infty$ does not depend on t.

since C-B-S ineq. $\left|\sum_{j=0}^{\infty}\theta_j\theta_{j+k}\right| \leq \sqrt{\sum_{j=0}^{\infty}\theta_j^2}\sqrt{\sum_{j=0}^{\infty}\theta_{j+k}^2}$ ($<\infty$ both)

For $MA(q)$: $\theta_{q+1} = \theta_{q+2} = \cdots = 0$, so,

$\gamma(k) = \text{cov}(Y_t, Y_{t+k}) = 0$ $\forall k > q$. ☆☆☆☆☆

Sharp cut-off: $\quad \rho(k)$ autocorrelation



For $MA(\infty)$: $\text{cov}(Y_t, Y_{t+k}) = \sigma^2\sum_{j=0}^{\infty}\theta_j\theta_{j+k} \to 0$ as $k \to \infty$

since $\left|\sum_{j=0}^{\infty}\theta_j\theta_{j+k}\right| \leq \sqrt{\sum_{j=0}^{\infty}\theta_j^2}\sqrt{\sum_{j=0}^{\infty}\theta_{j+k}^2} \to 0$ as $k \to \infty$

(more and more tail)

$$\sum_{j=0}^{\infty}\theta_{j+k}^2 = \sum_{j=k}^{\infty}\theta_j^2 \to 0 \text{ as } k \to \infty$$

$$\sum_{h=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}|\theta_{j+h}||\theta_j| = \left(\sum_{j=-\infty}^{\infty}|\theta_j|\right)^2$$

- **Why MA? Wold's Theorem**

Wold's Theorem:

Y_t : covariance stationary.
$\mathbb{E}[Y_t]=0$, then: $Y_t=\eta_t+\sum_{j=0}^{\infty}\theta_j\varepsilon_{t-j}$, $\theta_0=1$, $\sum_{j=0}^{\infty}\theta_j^2<\infty$

$\varepsilon_t\sim WN(0,\sigma^2)$, $\mathbb{E}[\varepsilon_t\eta_s]=0$, $\forall t,s$

Then: $\hat{\mathbb{E}}[\eta_{t+s}\mid Y_{t-1},Y_{t-2},\dots]\equiv\eta_{t+s}$ $\forall t,s$?

$\checkmark$ forecastable component

only linear combination of $\{Y_{t-i}\}$ are allowed. $\{\sum_{i=1}^{\infty}a_iY_{t-i}\}$

Intuition: $\{a_i^*\}=\underset{\{a_i\}}{\arg\min}\,\mathbb{E}[(Y_t-\sum_{i=1}^{\infty}a_iY_{t-i})^2]$ $\mathbb{E}[(Y_{t+s}-\sum_{i=1}^{\infty}a_iY_{t-i})^2]$

$\varepsilon_t:=Y_t-\underline{\sum_{i=1}^{\infty}a_i^*Y_{t-i}}$ Best linear forecast of Y_t based on past t.

$$Y_t=\varepsilon_t+\sum_{i=1}^{\infty}a_i^*Y_{t-i}$$
$$=\varepsilon_t+a_1^*Y_{t-1}+\sum_{i=2}^{\infty}a_i^*Y_{t-i}$$
$$=\varepsilon_t+a_1^*(\varepsilon_{t-1}+\sum_{i=1}^{\infty}a_i^*Y_{t-1-i})+\sum_{i=2}^{\infty}a_i^*Y_{t-i}$$
$$=\varepsilon_t+\underbrace{a_1^*}_{\theta_1}\varepsilon_{t-1}+\sum_{i=2}^{\infty}\tilde{a}_iY_{t-i}$$
$$=\varepsilon_t+a_1^*\varepsilon_{t-1}+\underbrace{\tilde{a}_2}_{\theta_2}(\varepsilon_{t-2}+\sum_{i=1}^{\infty}a_i^*Y_{t-2-i})+\sum_{i=3}^{\infty}\tilde{a}_iY_{t-i}$$
$$=\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad \sum_{j=0}^{\infty}\theta_j\varepsilon_{t-j}+$$

$\therefore\ \eta_t=Y_t-\sum_{j=0}^{\infty}\theta_j\varepsilon_{t-j}$, which is the limiting term in

In "nice" cases, $\eta_t=\mathbb{E}[Y_t]=0$ e.g. ergodicity.

e.g. Z: r.v. $\mathbb{E}[Z]=0$, $Y_t\equiv Z$ $\forall t$

$cov(Y_t,Y_s)=cov(Z,Z)=var(Z)\neq0$

For $MA(\infty)$, $cov(Y_t,Y_{t+k})\to0$ $k\to\infty$

but now $cov(Y_t,Y_{t+h})=cov(Z,Z)\not\to0$

$\Rightarrow Y_t$ cannot be represented as $MA(\infty)$ $\Rightarrow\eta$ is present

$\Rightarrow\eta_t=Z$. $\hat{\mathbb{E}}[\eta_{t+s}\mid Y_{t-1}\dots]\equiv Y_{t-1}\equiv Z$

e.g. $Z,\tilde{Z}$ – uncorrelated r.v.

$\mathbb{E}[Z]=\mathbb{E}[\tilde{Z}]=0$, $Var(Z)=Var(\tilde{Z})=\sigma^2$

$Y_t=Z\cos(\lambda t)+\tilde{Z}\sin(\lambda t)$ $\lambda\in\mathbb{R}$

$\mathbb{E}[Y_t]=0$ $Var(Y_t)=\mathbb{E}[(Z\cos(\lambda t)+\tilde{Z}\sin(\lambda t))^2]$
$$=\sigma^2(\cos^2(\lambda t)+\sin^2(\lambda t))$$
$$=\sigma^2$$

$$cov(Y_t,Y_{t+h})=\mathbb{E}[(Z\cos(\lambda t)+\tilde{Z}\sin(\lambda t))(Z\cos(\lambda(t+h))+\tilde{Z}\sin(\lambda(t+h)))]$$
$$=\sigma^2[\cos(\lambda t)\cos(\lambda(t+h))+\sin(\lambda t)\sin(\lambda(t+h))]$$
$$=\sigma^2\cos(\lambda h)$$

$\cos(\lambda(t+h))\not\to0$ as $h\to\infty$ $\Rightarrow Y_t$ is not $MA(\infty)$

$\eta_t:=Z\cos(\lambda t)+\tilde{Z}\sin(\lambda t)\equiv Y_t$

$\hat{\mathbb{E}}[\eta_{t+s}\mid Y_{t-1}\dots]\overset{?}{=}\eta_{t+s}$

$$\begin{cases}Y_{t-1}=Z\cos(\lambda(t-1))+\tilde{Z}\sin(\lambda(t-1))\\ Y_{t-2}=Z\cos(\lambda(t-2))+\tilde{Z}\sin(\lambda(t-2))\end{cases}$$

system of linear equation on Z & $\tilde{Z}$ $\Rightarrow$ can solve for Z & $\tilde{Z}$.

$\Rightarrow$ can calculate η_{t+s}

If $\mathbb{E}[Y_t]=\mu$, $X_t:=Y_t-\mu$, $\mathbb{E}[X_t]=\mathbb{E}[Y_t-\mu]=0$

$MA(\infty)$ impossible to estimate $\to$ truncate to $MA(q)$

Autoregression

$AR(1)$: $Y_t=\alpha+\beta Y_{t-1}+\varepsilon_t$

$AR(k)$: $Y_t=\alpha+\beta_1Y_{t-1}+\dots+\beta_kY_{t-k}+\varepsilon_t$, $\varepsilon_t\sim WN(0,\sigma^2)$

Lag of Y_t is Y_{t-1}. 2^{nd} lag of Y_t is Y_{t-2}

Lag Operator

Def. lag operators, denoted L, maps Y_t to Y_{t-1}. i.e.

$$LY_t=Y_{t-1}$$

- $L^2Y_t=L(LY_t)=L(Y_{t-1})=Y_{t-2}$ • $L^iY_t=Y_{t-i}$, $\forall i\in\mathbb{N}$
- $L(Y_t+X_t)=Z_{t-1}=Y_{t-1}+X_{t-1}$
- $L(\beta Y_t)=Z_{t-1}=\beta Y_{t-1}=\beta LY_t$
- lag polynomial: $\theta(L)=\theta_0+\theta_1L+\theta_2L^2+\dots+\theta_iL^i$

  $\theta(L)Y_t=\theta_0Y_t+\theta_1Y_{t-1}+\theta_2Y_{t-2}+\dots+\theta_iY_{t-i}$
- lag polynomials can be multiplied.

  $\theta(L)\varphi(L)=\varphi(L)\theta(L)$ (algebraic rules for numbers).

  e.g. $(1-\theta_1L)(1-\theta_2L)=1-\theta_1L-\theta_2L+\theta_1\theta_2L^2=(1-\theta_2L)(1-\theta_1L)$
- Some lag polynomial can be inverted.

Lemma: If $|\beta|<1$, then $(1-\beta L)$ is invertible and
$$(1-\beta L)^{-1}=\sum_{i=0}^{\infty}\beta^iL^i$$

Need. $(1-\beta L)^{-1}(1-\beta L)\equiv1$
$$(\sum_{i=0}^{\infty}\beta^iL^i)(1-\beta L)=\sum_{i=0}^{\infty}\beta^iL^i-\sum_{i=0}^{\infty}\beta^{i+1}L^{i+1}$$
$$=1+\sum_{i=1}^{\infty}\beta^iL^i-\sum_{i=1}^{\infty}\beta^iL^i=1$$

Need $|\beta|<1$ for $\sum_{i=0}^{\infty}\beta^iL^i$ to be well defined.

Intuition: $\sum_{i=0}^{\infty}\beta^i=\frac{1}{1-\beta}$ iff $|\beta|<1$. If $|\beta|>1$, $\sum_{i=0}^{\infty}\beta^i$ not finite.

To invert a higher-order lag polynomial: factorize.

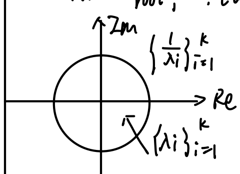
e.g. $(1-\theta_1L-\theta_2L^2)^{-1}=(1-\lambda_1L)^{-1}(1-\lambda_2L)^{-1}$ $\lambda_i=\frac{1}{root_i}$ can be complex.

If $|\lambda_1|<1$, $|\lambda_2|<1$, $(1-\lambda_iL)^{-1}=\sum_{n=0}^{\infty}(\lambda_i)^nL^n$, $i=1,2$.

$(1-\theta_1L-\theta_2L^2)^{-1}=(\sum_{n=0}^{\infty}\lambda_1^nL^n)(\sum_{n=0}^{\infty}\lambda_2^nL^n)$

Generally: $(1-\theta_1L-\dots-\theta_kL^k)^{-1}=(1-\lambda_1L)^{-1}(1-\lambda_2L)^{-1}\cdots(1-\lambda_kL)^{-1}$

$\lambda_i=\frac{1}{root_i}$: can be complex. Need $|\lambda_i|<1$ $\forall i=1,\dots,k$.



$|roots|>1$

$root=a+bi$

$\sqrt{a^2+b^2}<1$

Remark: $MA(q)$: $Y_t=\varepsilon_t+\theta_1\varepsilon_{t-1}+\dots+\theta_q\varepsilon_{t-q}$
$$=(1+\theta_1L+\dots+\theta_qL^q)\varepsilon_t$$

If invertible: $\varepsilon_t=(1+\theta_1L+\dots+\theta_qL^q)^{-1}Y_t$. Then also called $MA(q)$ invertible.

$AR(1)$: $Y_t=\alpha+\beta Y_{t-1}+\varepsilon_t \Leftrightarrow (1-\beta L)Y_t=\alpha+\varepsilon_t$

If $|\beta|<1$, then $Y_t=(1-\beta L)^{-1}(\alpha+\varepsilon_t)=\sum_{i=0}^{\infty}\beta^iL^i(\alpha+\varepsilon_t)$

$MA(\infty)$ represent. $=\sum_{i=0}^{\infty}\beta^i(\alpha+\varepsilon_{t-i})=\alpha\frac{1}{1-\beta}+\sum_{i=0}^{\infty}\beta^i\varepsilon_{t-i}$

Alternatively, via iterative back-substitution:

$$Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t$$
$$= \alpha + \beta(\alpha + \beta Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$
$$= \alpha(1+\beta) + \beta^2 Y_{t-2} + \beta\varepsilon_{t-1} + \varepsilon_t$$
$$= \alpha(1+\beta) + \beta^2(\alpha + \beta Y_{t-3} + \varepsilon_{t-2}) + \varepsilon_t + \beta\varepsilon_{t-1}$$
$$= \alpha(1+\beta+\beta^2) + \beta^3 Y_{t-3} + \varepsilon_t + \beta\varepsilon_{t-1} + \beta^2\varepsilon_{t-2}$$
$$= \alpha(1+\beta+\dots+\beta^{h-1}) + \beta^h Y_{t-h} + \varepsilon_t + \beta\varepsilon_{t-1} + \dots + \beta^{h-1}\varepsilon_{t-h+1}$$
$$\xrightarrow{h\to\infty} \alpha\frac{1}{1-\beta} + \sum_{i=0}^{\infty}\beta^i\varepsilon_{t-i}$$

With MA(∞) representation:

$$E[Y_t] = E\left[\frac{\alpha}{1-\beta} + \sum_{i=0}^{\infty}\beta^i\varepsilon_{t-i}\right] = \frac{\alpha}{1-\beta}$$

$h \geq 0$, $\quad cov(Y_t, Y_{t+h}) = cov\left(\sum_{i=0}^{\infty}\beta^i\varepsilon_{t-i}, \sum_{i=0}^{\infty}\beta^i\varepsilon_{t+h-i}\right)$

$$= \sigma^2\sum_{i=0}^{\infty}\beta^i\beta^{i+h} = \sigma^2\beta^h\frac{1}{1-\beta^2}$$

In particular: $var(Y_t) = \sigma^2\frac{1}{1-\beta^2}$. (*)

Notice: • if $|\beta|>1$, then $1-\beta^2<0$, (*) doesn't make sense.

• if $|\beta|=0$, then $\frac{1}{1-\beta^2}=\infty$, infinite variance.

∴ Same with invertible lag operator, $|\beta|<1$.

Another way to calculate moments:

$E[Y_t] = ?$ Assume Y_t weakly stat.

$$E[Y_t] = E[\alpha + \beta Y_{t-1} + \varepsilon_t] = \alpha + \beta E[Y_{t-1}] + E[\varepsilon_t] = \alpha + \beta E[Y_t] + E[\varepsilon_t]$$
$$= \alpha + \beta E[Y_t]$$
$$\Rightarrow E[Y_t] = \frac{\alpha}{1-\beta}.$$

$var(Y_t) = ?$ Assume Y_t is weakly stat.

$$Var(Y_t) = var(\alpha + \beta Y_{t-1} + \varepsilon_t) = \beta^2 var(Y_{t-1}) + var(\varepsilon_t) + 2\beta\, cov(Y_{t-1}, \varepsilon_t)$$
$$= \beta^2 var(Y_t) + \sigma^2 + 0$$
$$\Rightarrow var(Y_t) = \frac{\sigma^2}{1-\beta^2}$$

$h \geq 0$

$$\gamma(h) = cov(Y_{t+h}, Y_t) = cov(Y_t, Y_{t-h}) = cov(\alpha + \beta Y_{t-1} + \varepsilon_t, Y_{t-h})$$
$$= \beta\, cov(Y_{t-1}, Y_{t-h}) + cov(\varepsilon_t, Y_{t-h}) = \beta\gamma(h-1) + 0$$

∴ Yule-Walker Equation: $\gamma(h) = \beta\gamma(h-1) = \beta^2\gamma(h-2) = \beta^h\gamma(0) = \beta^h\frac{\sigma^2}{1-\beta^2}$

Autocorrelation: $\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \beta^h$ (geometrically decay).

• Impulse Response Function (IRF)

$IRF_h := \frac{\partial Y_{t+h}}{\partial \varepsilon_t}$ effect of ε_t on Y_{t+h}, assuming $\varepsilon_{t+1}\cdots\varepsilon_{t+h}$ do not change.

$$Y_{t+h} = \alpha + \beta Y_{t+h-1} + \varepsilon_{t+h} = \alpha + \beta(\alpha + \beta Y_{t+h-2} + \varepsilon_{t+h-1}) + \varepsilon_{t+h}$$
$$= \sum_{i=0}^{h}\alpha\beta^i + \sum_{i=0}^{h}\beta^i\varepsilon_{t+h-i} + \beta^{h+1}Y_{t-1}$$
$$= \beta^h\varepsilon_t + \sum_{i=0}^{h-1}\beta^i\varepsilon_{t+h-i} + \alpha\sum_{i=0}^{h}\beta^i + \beta^{h+1}Y_{t-1}$$

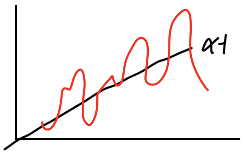
$\Rightarrow IRF_h = \beta^h \quad$ If $|\beta|<1$, then $\beta^h \to 0$ as $h \to 0$

no long-run effect.

• $\beta = 1$: unit-root random walk

$1-\beta L = 0 \Leftrightarrow L = \frac{1}{\beta}$ if $\beta=1$, $\to$ root $=1$.

$$Y_t = \alpha + Y_{t-1} + \varepsilon_t = 2\alpha + Y_{t-2} + \varepsilon_t + \varepsilon_{t-1}$$
$$= \alpha t + Y_0 + \sum_{i=1}^{t}\varepsilon_t$$



linear time trend if $\alpha \neq 0$

Let Y_0 be constant. Then,

$$E[Y_t] = E\left[\alpha t + Y_0 + \sum_{i=1}^{t}\varepsilon_t\right] = \alpha t + Y_0$$
$$var(Y_t) = var\left(\alpha t + Y_0 + \sum_{i=1}^{t}\varepsilon_t\right) = var\left(\sum_{i=1}^{t}\varepsilon_t\right) = \sigma^2 t \uparrow_t \infty.$$

Shocks ε_i do not disappear as $t \uparrow \infty$.

• Autoregression

AR(k): $Y_t = \alpha + \beta_1 Y_{t-1} + \dots + \beta_k Y_{t-k} + \varepsilon_t$, $\varepsilon_t \sim WN(0, \sigma^2)$

$(1-\beta_1 L - \dots - \beta_k L^k)Y_t = \alpha + \varepsilon_t$. $(1-\beta_1 L - \dots - \beta_k L^k)^{-1} = (1-\lambda_1 L)^{-1}\cdots(1-\lambda_k L)^{-1}$

need $|\lambda_i|<1 \ \forall i = 1\dots k$

then: Y_t is MA(∞). $Y_t = (1-\beta_1 L - \dots - \beta_k L^k)^{-1}(\alpha + \varepsilon_t)$

$\Rightarrow Y_t$ is covar. stat. MA(∞) is always cov. stat.

If Y_t is covariance stat, then $E[Y_t] = \mu$, $\forall t$.

$$\mu = E[Y_t] = E[\alpha + \beta_1 Y_{t-1} + \dots + \beta_k Y_{t-k} + \varepsilon_t]$$
$$= \alpha + (\beta_1 + \dots + \beta_k)\mu + 0$$
$$\Rightarrow \mu = \frac{\alpha}{1-(\beta_1+\dots+\beta_k)}$$

$\alpha = \mu - \mu(\beta_1 + \dots + \beta_k)$ plug in:

$$\Rightarrow Y_t - \mu = (Y_{t-1}-\mu)\beta_1 + \dots + (Y_{t-k}-\mu)\beta_k + \varepsilon_t$$

"deviations from the mean" form.

Another useful way to work with AR(k):

$$\underbrace{\begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-k+1} \end{pmatrix}}_{k\times 1} = \underbrace{\begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_{k-1} & \beta_k \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}}_{\substack{k\times k \\ =F}} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-k} \end{pmatrix} + \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$\xi_t = (Y_t, Y_{t-1}, \dots, Y_{t-k+1})'$, $\bar{\alpha} = (\alpha, 0, \dots, 0)'$, $\omega_t = (\varepsilon_t, 0, \dots, 0)'$

$\Rightarrow \xi_t = \bar{\alpha} + F\xi_{t-1} + \omega_t \quad$ AR(1).

Roots $1-\beta_1 L - \dots - \beta_k L^k = 0$ are $|\cdot|>1$ is equivalent to eigenvalues of F are $|\cdot|<1$ $\Rightarrow$ cov. stat. ☆☆☆

e.g. $k=2$: $F = \begin{pmatrix} \beta_1 & \beta_2 \\ 1 & 0 \end{pmatrix}$ $\det(F-\lambda I) = 0 \Leftrightarrow \begin{vmatrix} \beta_1-\lambda & \beta_2 \\ 1 & -\lambda \end{vmatrix} = 0$

$\lambda(\lambda-\beta_1) - \beta_2 = 0 \quad \lambda^2(1 - \frac{1}{\lambda}\beta_1 - \frac{1}{\lambda^2}\beta_2) = 0 \quad$ Solutions: $L^* = \frac{1}{\lambda^*}$

General k,

$$\det(F-\lambda I) = \begin{vmatrix} \beta_1-\lambda & \beta_2 & \beta_3 & \cdots & \beta_{k-1} & \beta_k \\ 1 & -\lambda & 0 & \cdots & 0 & 0 \\ 0 & 1 & -\lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\lambda \end{vmatrix}$$

$$= (\beta_1 - 1)(-1)^{k-1}\lambda^{k-1} - \beta_2(-1)^{k-2}\lambda^{k-2} + \beta_3(-1)^{k-3}\lambda^{k-3} - \beta_4(-1)^{k-4}\lambda^{k-4}$$
$$+ \dots + (-1)^{k+1}\beta_k(-1)^{k-k}\lambda^{k-k}$$
$$= (-1)^k\lambda^{k-1}(\lambda-\beta_1) + (-1)^{2+1}\beta_2(-1)^{k-2}\lambda^{k-2} + (-1)^{3+1}\beta_3(-1)^{k-3}\lambda^{k-3} + (-1)^{4+1}\beta_4(-1)^{k-4}\lambda^{k-4}$$
$$+ \dots + (-1)^{k+1}\beta_k(-1)^0\lambda^0$$
$$= (-1)^k\lambda^{k-1}(\lambda-\beta_1) - (-1)^k\beta_2\lambda^{k-2} - (-1)^k\beta_3\lambda^{k-3} - \dots - (-1)^k\beta_k\lambda^{k-k}$$
$$= (-1)^k\lambda^k\left(1 - \frac{1}{\lambda}\beta_1 - \frac{1}{\lambda^2}\beta_2 - \dots - \frac{1}{\lambda^k}\beta_k\right)$$

$1-\frac{1}{\lambda}\beta_1-\frac{1}{\lambda^2}\beta_2-\cdots-\frac{1}{\lambda^k}\beta_k=0 \qquad L^*=1/\lambda^* \qquad |\lambda^*|<1 \Leftrightarrow |L^*|>1$

$1-\beta_1 L-\beta_2 L^2-\cdots-\beta_k L^k=0$

In deviation from the mean form.

$$\begin{pmatrix} Y_t-\mu \\ Y_{t-1}-\mu \\ \vdots \\ Y_{t-k+1}-\mu \end{pmatrix} = \underbrace{\begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_{k-1} & \beta_k \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}}_{=F} \begin{pmatrix} Y_{t-1}-\mu \\ Y_{t-2}-\mu \\ \vdots \\ Y_{t-k}-\mu \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

☆ Deviation from the mean

$\eta_t := (Y_t-\mu, Y_{t-1}-\mu, \dots, Y_{t-k+1}-\mu)'$ $k\times 1$

$\mathbb{E}[\eta_t]=(0,\cdots,0)'$, $w_t=(\varepsilon_t,0,\cdots,0)'$.

$\eta_t = F\eta_{t-1}+w_t$ AR(1) without intercept.

$w_t \sim WN\left(\vec{0}, \underbrace{\begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}}_{Q}\right)$

$$\Sigma_0 := var(\eta_t) = \mathbb{E}[\eta_t\eta_t'] = \mathbb{E}\begin{pmatrix} (Y_t-\mu)^2 & \cdots & (Y_{t-k+1}-\mu)(Y_t-\mu) \\ (Y_t-\mu)(Y_{t-1}-\mu) & \cdots & (Y_{t-k+1}-\mu)(Y_{t-1}-\mu) \\ \vdots & & \vdots \\ (Y_t-\mu)(Y_{t-k+1}-\mu) & \cdots & (Y_{t-k+1}-\mu)^2 \end{pmatrix}$$

Matrix of first $k-1$ autocov $\gamma(1)\dots\gamma(k-1)$ and var $\gamma(0)$.

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

$F cov(\eta_{t-1}, w)F'$?

Or: $\Sigma_0 = Var(\eta_t) = Var(F\eta_{t-1}+w_t)$

$= F Var(\eta_{t-1})F' + Var(w_t) + F\underline{\mathbb{E}[\eta_{t-1}w_t']} + \underline{\mathbb{E}[w_t\eta_{t-1}']}F'$

$= F\Sigma_0 F' + Q$, where $Q = var(w_t)$ $\quad =0$

since η_{t-1} only contains $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots$

$w_t=(\varepsilon_t, 0, \cdots, 0)'$

$\boxed{\Sigma_0 = F\Sigma_0 F' + Q}$ system of lin. eqns on $\gamma(0)\ \gamma(1) \cdots \gamma(k-1)$

k unknowns $(\gamma(0), \gamma(1), \cdots, \gamma(k-1))$, $\underline{k \text{ eqns}}$. Σ_0 solvable. Why? Because:

$$F = \begin{pmatrix} \overset{A}{\beta_1 \ \beta_2 \cdots} & \overset{B}{\beta_k} \\ I_{k-1} & 0_{k-1\times 1} \end{pmatrix} \qquad A: 1\times k-1 \quad B: 1\times 1$$

$$\Sigma_0 = \begin{pmatrix} \overset{D}{\Sigma_0(1:k-1, 1:k-1)} & E \\ E' & G \end{pmatrix} \qquad D: k-1\times k-1 \quad E: k-1\times 1 \quad F: 1\times k-1 \quad G: 1\times 1$$

since Σ_0 symmetric

$$F\Sigma_0F' = \begin{pmatrix} A & B \\ I_{k-1} & 0 \end{pmatrix}\begin{pmatrix} D & E \\ E' & G \end{pmatrix}\begin{pmatrix} A' & I_{k-1} \\ B' & 0 \end{pmatrix}$$

all the dimensions agree!

$$= \begin{pmatrix} (AD+BE')A' + (AE'+BG)B' & AD+BE' \\ DA'+EB' & D \end{pmatrix}$$

B & D are symmetric, so $AD+BE' = (DA'+E'B)'$

$$= \begin{pmatrix} \cdot & \cdots \\ \vdots & \Sigma_0(1:k-1, 1:k-1) \end{pmatrix} = \Sigma_0 - Q$$

$\therefore \Sigma_0(1:k-1, 1:k-1) = \Sigma_0(2:k, 2:k)$

$cov(Y_{t-i}, Y_{t-j}) = cov(Y_{t-1-i}, Y_{t-1-j})$

$\therefore \Sigma_0 = F\Sigma_0F' + Q$

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(k-1) \\ \gamma(1) & & & \\ \vdots & & \Sigma_0(1:k-1, 1:k-1) & \\ \gamma(k-1) & & & \end{pmatrix} = \begin{pmatrix} (AD+BE')A' + (AE'+BG)B' + \sigma^2 & AD+BE' \\ DA'+EB' & \Sigma_0(1:k-1, 1:k-1) \end{pmatrix}$$

$\therefore$ Only 1st row and column gives non-trivial equation. And they are the same. The rest: $\Sigma_0(1:k-1, 1:k-1) = \Sigma_0(1:k-1, 1:k-1)$. $\underline{k\text{-eqns}}$.

$\Sigma_j := cov(\eta_t, \eta_{t-j}) \quad \forall j \geq 1 \qquad k\times k$ matrix

$$\Sigma_j = \begin{pmatrix} \mathbb{E}[Y_tY_{t-j}] & \mathbb{E}[Y_tY_{t-j-1}] & \cdots \\ \mathbb{E}[Y_{t-1}Y_{t-j}] & \mathbb{E}[Y_{t-1}Y_{t-j-1}] & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \quad \text{ignoring } \mu$$

$$= \begin{pmatrix} \gamma(j) & \gamma(j+1) & \cdots & \gamma(j+k-1) \\ \gamma(j-1) & \gamma(j) & \cdots & \gamma(j+k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(j-k+1) & \gamma(j-k+2) & \cdots & \gamma(j) \end{pmatrix} \quad k\times k \text{ matrix, not symmetric}$$

$\Sigma_j = \mathbb{E}[\eta_t\eta_{t-j}'] = \mathbb{E}[(F\eta_{t-1}+w_t)\eta_{t-j}']$

$= F\mathbb{E}[\eta_{t-1}\eta_{t-j}'] + \underbrace{\mathbb{E}[w_t\eta_{t-j}']}_{=0}$ $\quad w_t=(\varepsilon_t, 0, \cdots, 0)$

$= F\Sigma_{j-1}$ $\quad \eta_{t-j}$ only has $\varepsilon_{t-j}, \varepsilon_{t-j-1}, \cdots$, $t>t-j$

$\underline{\Sigma_j = F\Sigma_{j-1} = F^2\Sigma_{j-2} = \cdots = F^j\Sigma_0}$. ☆☆ AR(1): Geometric.

1) solve for Σ_0, easier. 2). get all Σ_j.

- **Autoregressive Moving Average**

ARMA(k, q) $Y_t = \alpha + \beta_1 Y_{t-1} + \cdots + \beta_k Y_{t-k} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$

$\underbrace{(1-\beta_1 L - \cdots - \beta_k L^k)}_{\beta(L)} Y_t = \alpha + \underbrace{(1+\theta_1 L + \cdots + \theta_q L^q)}_{\theta(L)}\varepsilon_t$

$\beta(L)Y_t = \alpha + \theta(L)\varepsilon_t$

ARMA representation is not unique!

- $Y_t = \beta Y_{t-1} + \varepsilon_t = \sum_{i=0}^{\infty}\beta^i\varepsilon_{t-i}$ if $|\beta|<1$
- If $\beta(L)Y_t = \theta(L)\varepsilon_t$

then $\alpha(L)\beta(L)Y_t = \alpha(L)\theta(L)\varepsilon_t$

e.g. $Y_t = \varepsilon_t \Rightarrow (1-\beta L)Y_t = (1-\beta L)\varepsilon_t$

$Y_t = \beta Y_{t-1} + \varepsilon_t - \beta\varepsilon_{t-1}$.

- **Introduction to Forecast**

Y: random var. $\hat{Y}$: point forecast of Y (non-random)

$e := Y - \hat{Y}$ forecast error.

- Loss function: $L(Y, \hat{Y})$ Special case: $L(Y-\hat{Y}) \equiv L(e)$

e.g. $L(e)=e^2$ quadratic loss $\quad L(e)=|e|$ absolute loss

$L(e) = \begin{cases} ae^2, & e \geq 0 \\ be^2, & e<0 \end{cases}$ $\quad a, b>0$ asymmetric

e.g. Monopolist sells a product at price p.

Demand is linear $Q = 2a - p$

$\max_p \pi(p) \to p^* = a \qquad \pi(p^*) = a^2$

a is unobserved $\to$ forecast $\hat{a} \to$ set $p = \hat{a}$

$\pi(\hat{a}) = \hat{a}(2a-\hat{a})$

profit loss $= \pi(a) - \pi(\hat{a}) = a^2 - \hat{a}(2a-\hat{a})$

$= (\hat{a}-a)^2$ quadratic loss.

- Risk $R(\hat{Y}) = \mathbb{E}[L(Y-\hat{Y})]$ risk of a forecast

general case: $R(\hat{Y}) = \mathbb{E}[L(Y;\hat{Y})]$

Def. The optimal (best) point forecast solves $\underset{\hat{Y}}{\arg\min}\, R(\hat{Y})$

- $L(e) = e^2$

$R(\hat{Y}) = \mathbb{E}[(Y-\hat{Y})^2] = \mathbb{E}[Y^2 - 2\hat{Y}Y + \hat{Y}^2]$
$= \mathbb{E}[Y^2] - 2\hat{Y}\mathbb{E}[Y] + \hat{Y}^2$

min $\to \hat{Y}^* = \mathbb{E}[Y]$ ∴ optimal forecast under quadratic loss is mean.

- $L(e) = |e|$

$R(\bar{Y}) = \mathbb{E}[|Y-\bar{Y}|] = \int_{-\infty}^{\bar{Y}} (\bar{Y}-y) f_Y(y)dy + \int_{\bar{Y}}^{\infty} (y-\bar{Y}) f_Y(y)dy$

Since $\frac{\partial}{\partial x}\int_{-\infty}^{x} g(y,x)\,dy = g(x,x) + \int_{-\infty}^{x} \frac{\partial}{\partial x} g(y,x)dy$

$\frac{\partial}{\partial \bar{Y}}\mathbb{E}[|Y-\bar{Y}|] = \int_{-\infty}^{\bar{Y}} f_Y(y)dy - \int_{\bar{Y}}^{\infty} f_Y(y)dy = F(\bar{Y}) - (1-F(\bar{Y}))$

F.O.C: $0 = F(\bar{Y}) - (1-F(\bar{Y})) \Rightarrow F(\bar{Y}) = \frac{1}{2}$ $\bar{Y}^* = \text{median}(Y)$

∴ optimal forecast under absolute loss is the median.

If we have extra info: $\mathbb{E}[\cdot] \to \mathbb{E}[\cdot|X]$ $\underset{\hat{Y}}{\arg\min}\,\mathbb{E}[L(Y-\hat{Y})|X]$

Different loss function → different choice of point optimal forecast.

Forecasting Time Series

AR(1): $Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t$, $|\beta|<1$, $\varepsilon_t \sim WN(0,\sigma^2)$

Assume: $\mathbb{E}[\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\cdots] = 0$ ε_t is unforecastable

$\mathbb{E}[\varepsilon_t|\cdots] = \mathbb{E}[\varepsilon_t] = 0$

$\mathbb{E}[Y_{t+1}|Y_t, Y_{t-1},\cdots] = \mathbb{E}[\alpha + \beta Y_t + \varepsilon_{t+1}|Y_t,\cdots]$
$= \alpha + \beta Y_t + \mathbb{E}[\varepsilon_{t+1}|Y_t,\cdots]$ same information.
$= \alpha + \beta Y_t + \mathbb{E}[\varepsilon_{t+1}|\varepsilon_t,\cdots]$ since Y_t cov. stat, $Y_t = (1-\beta L)^{-1}(\alpha+\varepsilon_t)$
$= \alpha + \beta Y_t$

$\mathbb{E}[Y_{T+h}|Y_T,\cdots] = \mathbb{E}[\alpha+\beta Y_{T+h-1}+\varepsilon_{T+h}|Y_T,\cdots]$
$= \alpha(1+\beta+\cdots+\beta^{h-1}) + \beta^h Y_T + \mathbb{E}[\varepsilon_{T+h}+\beta\varepsilon_{T+h-1}+\cdots + \beta^{h-1}\varepsilon_{t+1}|Y_T\cdots]$
$= \alpha\frac{1-\beta^h}{1-\beta} + \beta^h Y_T$

- forecast error: $Y_{T+h} - \mathbb{E}[Y_{T+h}|Y_T]$
$= Y_{T+h} - \mathbb{E}[\alpha+\beta Y_{T+h-1}+\varepsilon_{t+h}|Y_T]$
$= \varepsilon_{T+h} + \beta\varepsilon_{T+h-1} + \cdots + \beta^{h+1}\varepsilon_{T+1}$

- AR(k): $Y_t = \alpha + \beta_1 Y_{t-1} + \cdots + \beta_k Y_{T-k} + \varepsilon_t$. stat. $\varepsilon_t \sim WN(0,\sigma^2)$

$\mathbb{E}[\varepsilon_t|\varepsilon_{t-1},\cdots] = 0$

$\mathbb{E}[Y_{T+1}|Y_T, Y_{T-1},\cdots] = \alpha + \beta_1 Y_T + \beta_2 Y_{T-1} + \cdots + \beta_k Y_{T-k+1} + \mathbb{E}[\varepsilon_{T+1}|Y_T\cdots]$
$= \alpha + \beta_1 Y_T + \beta_2 Y_{T-1} + \cdots + \beta_k Y_{T-k+1}$.

$\mathbb{E}[Y_{T+2}|Y_T, Y_{T-1},\cdots] = \alpha + \beta_1\mathbb{E}[Y_{T+1}|Y_T\cdots] + \beta_2 Y_T + \cdots + \beta_k Y_{T-k+2}$
$= \alpha + \beta_1(\alpha+\beta_1 Y_T + \cdots + \beta_k Y_{T-k+1}) + \beta_2 Y_T + \cdots + \beta_k Y_{T-k+2}$.

For generalized horizon h: either iterate h times (very messy)
or work with vector AR(1) $\eta_t = F\eta_{t-1} + w_t$.

MA(1): $Y_t = \alpha + \varepsilon_t + \theta\varepsilon_{t-1}$ $\varepsilon_t \sim WN(0,\sigma^2)$ $\mathbb{E}[\varepsilon_t|\varepsilon_{t-1}\cdots] = 0$

$|\theta|<1$: guarantees $\varepsilon_t = (1+\theta L)^{-1}(-\alpha+Y_t) = \sum_{j=0}^{\infty}(-\theta)^j(Y_{t-j}-\alpha)$

$\mathbb{E}[Y_{T+h}|Y_T\cdots] = \alpha + \mathbb{E}[\varepsilon_{T+h}+\theta\varepsilon_{T+h-1}|Y_t,\cdots]$
$= \alpha + \mathbb{E}[\varepsilon_{T+h}+\theta\varepsilon_{T+h-1}|\varepsilon_T,\cdots]$
$= \begin{cases}\alpha+\theta\varepsilon_T & \text{if } h=1\\ \alpha & \text{if } h>1\end{cases}$
$= \begin{cases}\alpha+\theta\sum_{i=0}^{\infty}(-\theta)^i(Y_{T-i}-\alpha), & h=1\\ \alpha, & h>1\end{cases}$

MA(q): $Y_t = \alpha + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$. $\varepsilon_t \sim WN(0,\sigma^2)$

$\mathbb{E}[\varepsilon_t|\varepsilon_{t-1},\cdots] = 0$ + invertibility:
$(1+\theta_1 L+\cdots+\theta_q L^q)^{-1}$ exists.

$\mathbb{E}[Y_{T+h}|Y_T,\cdots] = \alpha + \mathbb{E}[\varepsilon_{T+h}+\theta_1\varepsilon_{T+h-1}+\cdots+\theta_q\varepsilon_{T+h-q}|\varepsilon_T,\cdots]$
$= \begin{cases}\alpha+\theta_h\varepsilon_T+\cdots+\theta_q\varepsilon_{T+h-q}, & h\le q\\ \alpha, & h>q\end{cases}$

e.g. h=1: $= \alpha + \theta_1\varepsilon_T + \cdots + \theta_q\varepsilon_{T+1-q}$
h=2: $= \alpha + \theta_2\varepsilon_T + \cdots + \theta_q\varepsilon_{T+2-q}$

• Martingale Difference Sequence (MDS)

$\mathbb{E}[\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\cdots] = 0$

[Generally, $\mathbb{E}[\varepsilon_t|\mathcal{F}_{t-1}] = 0$, $\mathcal{F}_{t-1}$: filteration

MDS ⇒ WN, since:

1) $\mathbb{E}[\varepsilon_t] = \mathbb{E}[\mathbb{E}[\varepsilon_t|\varepsilon_{t-1},\cdots]] = \mathbb{E}[0] = 0$
2) $\gamma(h) = \mathbb{E}[\varepsilon_t\varepsilon_{t+h}] = \mathbb{E}\{\mathbb{E}[\varepsilon_t\varepsilon_{t+h}|\varepsilon_{t+h-1},\varepsilon_{t+h-2},\cdots]\}$
$h>0$
$= \mathbb{E}\{\varepsilon_t\,\mathbb{E}[\varepsilon_{t+h}|\varepsilon_{t+h-1},\varepsilon_{t+h-2},\cdots]\}$
$= \mathbb{E}[\varepsilon_t\cdot 0] = 0$

∴ $\gamma(h) = 0$, $\forall h\neq 0$.

• Model Selection

(I) Testing

H_0: AR(k-1) v.s. H_1: AR(k) i.e. $\beta_k \overset{?}{=} 0$ in $Y_t = \alpha+\beta_1 Y_{t-1}+\cdots+\beta_k Y_{t-k}+\varepsilon_t$

Reject → AR(k). Not Reject → AR(k-1)

→ H_0: AR(k-2) v.s. H_1: AR(k-1) i.e. $\beta_{k-1} \overset{?}{=} 0$ in $Y_t = \alpha+\beta_1 Y_{t-1}+\cdots+\beta_{k-1}Y_{t-k+1}+\varepsilon_t$

Reject → AR(k-1). Not Reject → AR(k-2)

→ Repeat. Get AR(k^*)

- Multiple Testing: hard to control size.
- Inconsistent: e.g. true AR(1). Test AR(1) vs AR(2)

$H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$ on 5% level. → In 5% cases we reject and choose AR(2).

(II) Information Criteria

Idea: evaluate the fit of AR(k) and penalize for larger k.

1) Estimate AR(k), get residuals

$\hat{e}_t(k) := Y_t - \hat{\alpha}(k) - \hat{\beta}_1(k)Y_{t-1} - \cdots - \hat{\beta}_k(k)Y_{t-k}$

$\hat{\sigma}^2(k) := \frac{1}{T}\sum_t \hat{e}_t^2(k)$ sum of squared residuals.

2) Choose

$\hat{k} = \underset{k\le\bar{k}}{\arg\min}\Big(\underbrace{\ln\hat{\sigma}^2(k)}_{-2\text{ log-likelihood}} + \overbrace{\underbrace{(1+k)}_{\#\text{ parameters in AR(k)}}g(T)}^{\text{penalty}}\Big)$

- $\bar{k}<\infty$: upper limit on (known) order of AR
- $g(T) = \begin{cases}\frac{2}{T} & \text{AIC (Akaike Information Criteria)}\\ \frac{\ln T}{T} & \text{BIC (Bayes / Schwarz Inf. Cri.)}\end{cases}$

$\frac{2}{T} < \frac{\ln T}{T}$ ⇒ AIC chooses (weakly) larger model than BIC

BIC: Bayes Rule. Goal: find true model. (Unreliable if true model is outside the considered class).

AIC: min forecast error. Goal: find the best forecast model.

- Def. $\hat{k}$ is (weakly) consistent if $\mathbb{P}(\hat{k}=k_0) \xrightarrow{T\to\infty} 1$. k_0: true order.
- BIC is consistent. AIC: $\mathbb{P}(\hat{k} \ge k_0) \to 1$ as $T\to\infty$. can choose larger model.
- Lemma. If $g(T) \xrightarrow[T\to\infty]{} 0$, then $\mathbb{P}(\hat{k}\ge k_0) \xrightarrow[T\to\infty]{} 1$.

Pf. $S_j := \ln \hat\sigma^2(j) + (1+j)g(T)$

$\mathbb{P}(\hat{k}\ge k_0) = \mathbb{P}(\min\{S_{k_0}, S_{k_0+1}, \dots, S_{\bar{k}}\} < \min\{S_0, S_1, \dots, S_{k_0-1}\})$?

$\ge \mathbb{P}(S_{k_0} < \min\{S_0, S_1, \dots, S_{k_0-1}\})$

$= \mathbb{P}(\ln\hat\sigma^2(k_0) + (1+k_0)g(T) < \ln\hat\sigma^2(j) + (1+j)g(T), \forall j = 0, \dots, k_0-1)$

$= 1 - \mathbb{P}(\ln\hat\sigma^2(k_0) + (1+k_0)g(T) \ge \ln\hat\sigma^2(j) + (1+j)g(T)$, for some $j \in \{0, \dots, k_0-1\})$

$\ge 1 - \sum_{j=0}^{k_0-1} \mathbb{P}(\ln\hat\sigma^2(k_0) + (1+k_0)g(T) \ge \ln\hat\sigma^2(j) + (1+j)g(T))$

$= 1 - \sum_{j=0}^{k_0-1} \mathbb{P}\left(\ln\frac{\hat\sigma^2(j)}{\hat\sigma^2(k_0)} - (k_0-j)g(T) \le 0\right)$

$\hat\sigma^2(j) = \frac{1}{T}\sum_t (Y_t - \hat\alpha(j) - \hat\beta_1(j)Y_{t-1} - \dots - \hat\beta_j(j)Y_{t-j})^2$

$\hat\sigma^2(j) \xrightarrow[T\to\infty]{} \mathbb{E}[(Y_t - a^j - b_1^j Y_{t-1} - \dots - b_j^j Y_{t-j})^2]$

where $\{a^j, b_1^j \dots b_j^j\} = \underset{\{a, b_1, \dots, b_j\}}{\arg\min} \mathbb{E}[(Y_t - a - b_1 Y_{t-1} - \dots - b_j Y_{t-j})^2]$

True model $k_0 > j$ since $j \in \{0, \dots, k_0-1\}$, then:

$\underset{\{a, b_1, \dots, b_j\}}{\arg\min} \mathbb{E}[(Y_t - a - b_1 Y_{t-1} - \dots - b_{k_0} Y_{t-k_0})^2]$ $\Rightarrow a = \alpha, b_1 = \beta_1, \dots b_{k_0} = \beta_{k_0}$

$\to \mathbb{E}[\varepsilon_t^2] = \sigma^2$

$\hat\sigma^2(k_0) \xrightarrow[T\to\infty]{} \sigma^2 < \sigma^2(j)$ (since True model has k_0 lags, we cannot fit as good as it with only j lags)

$\therefore \ln\frac{\hat\sigma^2(j)}{\hat\sigma^2(k_0)} \to \ln\frac{\sigma^2(j)}{\sigma^2} > 0$, $(k_0-j)g(T) \to 0$ as $T\to\infty$.

$\therefore \mathbb{P}\left(\ln\frac{\hat\sigma^2(j)}{\hat\sigma^2(k_0)} - (k_0-j)g(T) \le 0\right) \to 0$ as $T \to 0$.

$\therefore \mathbb{P}(\hat{k} \ge k_0) \to 1$ as $T\to\infty$ if $g(T) \to 0$ as $T\to\infty$.

- Lemma: If $Tg(T) \xrightarrow{T\to\infty} \infty$, then $\mathbb{P}(\hat{k} \le k_0) \to 1$ as $T\to\infty$.

Similar proof ideas with some modifications.

Verification: AIC: $g(T) = \frac{2}{T} \to 0$, $Tg(T) = 2 \not\to \infty$, so $\mathbb{P}(\hat{k} \ge k_0) \to 1$

BIC: $g(T) = \frac{\ln T}{T} \to 0$, $Tg(T) = \ln T \to \infty$, so $\mathbb{P}(\hat{k} = k_0) \to 1$

consistent

- Implementation: AR(1) $\hat{e}_1 = Y_1 - \hat\alpha - \hat\beta \underline{Y_0}$ unobservable

then $\hat{e}_1$ cannot calculate.

AR(2) $\hat{e}_2 = Y_2 - \hat\alpha - \hat\beta_1 Y_1 - \hat\beta_2 \underline{Y_0}$ unobservable

$\to$ start with $t = \bar{k}+1$ T = restricted sample

(so all models have the same set of residuals).

• Vector Autoregression (VAR)

$\underset{(n\times1)}{Y_t} = \underset{(n\times1)}{\alpha} + \underset{(n\times n)}{\varphi_1}\underset{(n\times1)}{Y_{t-1}} + \dots + \underset{(n\times n)}{\varphi_k}\underset{(n\times1)}{Y_{t-k}} + \underset{(n\times1)}{\varepsilon_t}$ VAR(k)

$Y_t = \begin{pmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{nt} \end{pmatrix}$ purposes: ① Forecast econ time series.

② Design and evaluate econ. models

③ Evaluate the consequences of alternative policy actions.

$\mathbb{E}[\varepsilon_t] = \bar{0}$ & $\mathbb{E}[\varepsilon_t \varepsilon_s'] = \begin{cases} \bar{0}, & t \ne s \quad n\times n \text{ matrix of 0s} \\ \underline{0}, & t = s \quad \text{symmetric, p.d. matrix.} \end{cases}$

$(\Omega = \underline{0})$

$\iff \varepsilon_t \sim WN(0, \underline{0})$

- Can also write eq-ns for each coordinate of Y_t:

$Y_{it} = \alpha_i + \varphi_1^{i1} Y_{1t-1} + \dots + \varphi_1^{in} Y_{nt-1}$
$+ \varphi_2^{i1} Y_{1t-2} + \dots + \varphi_2^{in} Y_{nt-2}$
$+ \dots$
$+ \varphi_k^{i1} Y_{1t-k} + \dots + \varphi_k^{in} Y_{nt-k} + \varepsilon_{it}$

k lags of each coordinate

$\varphi_p^{ij} = \varphi_p[i, j]$

In lag Operator Notation:

$(I - \varphi_1 L - \dots - \varphi_k L^k) Y_t = \alpha + \varepsilon_t$

$= \varphi(L) Y_t = \alpha + \varepsilon_t$ $\varphi(L)$: lag operator n×n matrix

$\varphi(L)[i, j] = \delta_{ij} - \varphi_1^{ij} L - \dots - \varphi_k^{ij} L^k$

$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \ne j. \end{cases}$

Find $\mathbb{E}[Y_t]$: If Y_t is covar. stat, then

$\mu := \mathbb{E}[Y_t] = \mathbb{E}[\alpha + \varphi_1 Y_{t-1} + \dots + \varphi_k Y_{t-k} + \varepsilon_t]$

$= \alpha + \varphi_1 \mu + \dots + \varphi_k \mu + 0$

$\alpha = (I - \varphi_1 - \dots - \varphi_k)\mu$

plug back α: $Y_t - \mu = \varphi_1(Y_{t-1} - \mu) + \dots + \varphi_k(Y_{t-k} - \mu) + \varepsilon_t$.

"deviations from the mean" form.

$\underset{nk\times1}{\xi_t} := \begin{pmatrix} Y_t - \mu \\ Y_{t-1} - \mu \\ \vdots \\ Y_{t-k+1} - \mu \end{pmatrix} \begin{matrix} n\times1 \\ n\times1 \\ \vdots \\ n\times1 \end{matrix}$ $\underline{\xi_t = F \xi_{t-1} + V_t}$ $F = \begin{array}{r} n\times nk \to \\ \\ \end{array} \left(\begin{array}{cccc:c} \underset{n\times n}{\varphi_1} & \underset{n\times n}{\varphi_2} & \cdots & & \underset{n\times n}{\varphi_k} \\ \hdashline & \underbrace{I_{n(k-1)}}_{n(k-1)\times n(k-1)} & & & \underbrace{O_{n(k-1)\times n}}_{n(k-1)\times n} \end{array}\right)$

$V_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ $nk\times1$ $\mathbb{E}[V_t] = 0$

$\mathbb{E}[V_t V_s'] = 0_{nk\times nk}$

$t \ne s$

$\mathbb{E}[V_t V_t'] = \begin{array}{c} \left(\begin{array}{cccc} \underline{0}_{n\times n} & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 0 \end{array}\right) \end{array} \begin{array}{l} n\times nk \\ \\ n(k-1)\times nk \\ \\ \end{array}$

Need all the eigenvalues to be $|\cdot| < 1$ for stationarity.

(i.e. need $F^l \not\to \infty$ as $l \to \infty$)

$|F - \lambda I| = \begin{vmatrix} \varphi_1 - \lambda I_n & \varphi_2 & \cdots & \varphi_{k-1} & \varphi_k \\ I_n & -\lambda I_n & \cdots & O_{n\times n} & O_{n\times n} \\ O_{n\times n} & I_n & \cdots & O_{n\times n} & O_{n\times n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O_{n\times n} & O_{n\times n} & & I_n & -\lambda I_n \end{vmatrix}$

add last n-dim column $\cdot \frac{1}{\lambda}$ to the previous n-dim column

$= \begin{vmatrix} \varphi_1 - \lambda I_n & \varphi_2 & \cdots & \varphi_{k-1} + \frac{1}{\lambda}\varphi_k & \varphi_k \\ I_n & -\lambda I_n & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & & -\lambda I_n & 0 \\ 0 & 0 & & 0 & -\lambda I_n \end{vmatrix}$

keep on doing this

↓

turn it to a up-right triangle matrix.

$$= \begin{vmatrix} \varphi_1 - \lambda I_n + \frac{1}{\lambda}\varphi_2 + \frac{1}{\lambda^2}\varphi_3 + \cdots + \frac{1}{\lambda^{k-1}}\varphi_k & \varphi_2 + \varphi_1\varphi_3 & \cdots & \varphi_k \\ 0 & -\lambda I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\lambda I_n \end{vmatrix}$$

$$= (-\lambda)^{n(k-1)} \left| \varphi_1 - \lambda I_n + \frac{1}{\lambda}\varphi_2 + \frac{1}{\lambda^2}\varphi_3 + \cdots + \frac{1}{\lambda^{k-1}}\varphi_k \right|$$

$$= (-\lambda)^{n(k-1)} (-\lambda)^n \left| I_n - \frac{1}{\lambda}\varphi_1 - \frac{1}{\lambda^2}\varphi_2 - \cdots - \frac{1}{\lambda^k}\varphi_k \right| \quad \text{(pull out } -\lambda \text{ from an } n\times n \ |\cdots|)$$

$= 0 \leftarrow \det(F - \lambda I) = 0$

Eigenvalues of F solves $|I_n - \frac{1}{\lambda}\varphi_1 - \cdots - \frac{1}{\lambda^k}\varphi_k| = 0$

or $|I_n - z\varphi_1 - \cdots - z^k\varphi_k| = 0$, $z = \frac{1}{\lambda}$.

Need: $|\lambda| < 1 \iff |z| > 1$.

- **Vector MA(∞) Representation:**

$\xi_t = F\xi_{t-1} + V_t$

$(I - FL)\xi_t = V_t$

$$\xi_t = (I - FL)^{-1} V_t = \sum_{j=0}^{\infty} F^j V_{t-j}, \quad F^j = F \cdot F \cdots F$$

Another Way to get VMA(∞): iterated back-substitution.

$$\xi_t = F\xi_{t-1} + V_t = F(F\xi_{t-2} + V_{t-1}) + V_t$$

$$= F^2\xi_{t-2} + V_t + FV_{t-1} = \cdots = \underbrace{F^l \xi_{t-l}}_{\to 0 \text{ as } l \to \infty} + V_t + FV_{t-1} + \cdots + F^{l-1}V_{l-1}$$

$$= \sum_{j=0}^{\infty} F^j V_{t-j}$$

Back to Y_t. $\underset{n\times nk}{S} := (I_n, O_{n\times n(k-1)})$

$S\xi_t = Y_t - \mu$

$SV_t = \varepsilon_t$

$$S'\varepsilon_t = \begin{pmatrix} I_n \\ O_{n(k-1)\times n} \end{pmatrix} \varepsilon_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} = V_t$$

$$Y_t - \mu = S\xi_t = S\sum_{j=0}^{\infty} F^j V_{t-j} = S\sum_{j=0}^{\infty} F^j S'\varepsilon_{t-j} = \sum_{j=0}^{\infty} SF^jS'\varepsilon_{t-j}$$

$$= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{where } \psi_j = SF^jS'$$

$\psi = SF^jS' = n\times n$ corner of F^j (upper-left block)

$\psi_0 = SS'$.

- VMA(∞) is not unique.

Let $H = n\times n$ nonsingular matrix. $u_t = H\varepsilon_t$

Then: $\mathbb{E}[u_t] = H\mathbb{E}[\varepsilon_t] = 0$.

$$\mathbb{E}[u_t u_s'] = H\mathbb{E}[\varepsilon_t\varepsilon_s']H' = \begin{cases} 0, & t \neq s \\ H\Omega H', & t = s \end{cases} \Rightarrow u_t \sim WN(0, H\Omega H')$$

$$\therefore Y_t - \mu = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} = \sum_{j=0}^{\infty}\psi_j H^{-1}\underbrace{H\varepsilon_{t-j}}_{u_{t-j}} = \sum_{j=0}^{\infty}\underbrace{\psi_j H^{-1}}_{\tilde{\psi}_j} u_{t-j} = \underbrace{\sum_{j=0}^{\infty}\tilde{\psi}_j u_{t-j}}_{\text{Another VMA}(\infty)\text{ representation.}} \quad (\bigstar)$$

Cholesky Decomposition:

$\Omega = LL'$, L = lower triangle matrix

$\underset{\sim}{H} = L^{-1}$, $u_t = H\varepsilon_t$, $Var(u_t) = H\Omega H' = L^{-1}(LL')L^{-1'} = I_n$.

- **Impulse Response Functions**

$Y_t = \alpha + \varphi_1 Y_{t-1} + \cdots + \varphi_k Y_{t-k} + \varepsilon_t$. $\varepsilon_t \sim WN(0, \Omega)$

Estimate: OLS (eqn by eqn)

Y_{it} on constants, $Y_{1t-1}, \dots, Y_{1t-k}, Y_{2t-1}, \dots, Y_{2t-k}, \cdots Y_{nt-1}, \dots, Y_{nt-k}$.

If $\varepsilon_t \sim N(0, \Omega)$, then MLE $\iff$ OLS

$\hat{\Omega} = \frac{1}{T}\sum_t \hat{e}_t\hat{e}_t'$

OLS: consistent, asymptotically normal (assuming stationarity)

But, hard to interprete $n\times n$ matrices of coefficients.

$\Rightarrow$ Impulse Response Functions.

IRFs: $\left\{\frac{\partial Y_{it+s}}{\partial \varepsilon_{jt}}\right\}_{s=0,1,\dots}$ n^2 IRFs. $i \in \{1,\dots,n\}$, $j \in \{1,\dots,n\}$

- VMA(∞) representation: $Y_t = \mu + \sum_{l=0}^{\infty}\psi_l\varepsilon_{t-l}$

$$Y_{t+s} = \mu + \sum_{l=0}^{\infty}\psi_l\varepsilon_{t+s-l} \quad \therefore \frac{\partial Y_{it+s}}{\partial \varepsilon_{jt}} = \psi_s(i,j) \quad (*).$$

Problem: 1) VMA(∞) not unique

2) ε_{jt} and $\varepsilon_{j't}$ can be correlated: if ε_{jt} changes, then $\varepsilon_{j't}$ can also change. But in (*), we do not allow for such change.

Solution: Orthogonalize shocks.

Let $u_t = H\varepsilon_t$ s.t. $Var(u_t) = H\Omega H' = I_n$

e.g. Cholesky decomposition: $H = L^{-1}$ where $LL' = \Omega$.

$\to$ then, u_{it} and u_{jt} are uncorrelated.

$$(\bigstar) \Rightarrow \frac{\partial Y_{it+s}}{\partial u_{jt}} = \sum_{j'=1}^{n}\psi_s(i,j')H^{-1}(j',j)$$

since $Y_t = \mu + \sum_{l=0}^{\infty}\psi_l\varepsilon_{t-l} = \mu + \sum_{l=0}^{\infty}\psi_l H^{-1}u_{t-l}$.

Work with $\left\{\frac{\partial Y_{it+s}}{\partial u_{jt}}\right\}$ (n^2 plots).

How to estimate IRFs:

1) Estimate VAR by OLS

$$\to \begin{cases} \hat{\varphi}_l, \ l = 1, 2, \dots, k \\ \hat{\Omega} \end{cases}$$

2) Invert to get VMA(∞) $\to \tilde{\psi}_s$, $s = 0, 1, 2, \dots$

3) Find H s.t. $H\hat{\Omega}H' = I_n$. e.g. Cholesky decomposition $H = L^{-1}$, $LL' = \hat{\Omega}$

4) $\widehat{IRF}_s = \hat{\psi}_s L$

Standard Errors:

1). Delta Method: $\hat{\varphi}_l$, $\hat{\Omega}$ – asymptotically normal.

$$Y_t = \alpha + \varphi_1 Y_{t-1} + \cdots + \varphi_k Y_{t-k} + \varepsilon_t$$

$$= \alpha + \varphi_1(\alpha + \varphi_1 Y_{t-2} + \cdots + \varphi_k Y_{t-k-1} + \varepsilon_{t-1}) + \varphi_2 Y_{t-2} + \cdots + \varphi_k Y_{t-k} + \varepsilon_t$$

$$= \varepsilon_t + \varphi_1\varepsilon_{t-1} + \alpha + \varphi_1\alpha + (\varphi_1^2 + \varphi_2)Y_{t-2} + \cdots + \varphi_k Y_{t-k} + \varphi_1\varphi_k Y_{t-k-1}$$

$$= \varepsilon_t + \varphi_1\varepsilon_{t-1} + (\varphi_1^2 + \varphi_2)\varepsilon_{t-2} + \cdots$$

$$\Rightarrow \begin{cases} \psi_0 = I_n \\ \psi_1 = \varphi_1 \\ \psi_2 = \varphi_1^2 + \varphi_2 \end{cases}$$

- L is a function of $\hat{\Omega}$.

$\to$ use delta-method again.

Many non-linear Transformation:

→ imprecise estimation of $\hat{\varphi}_i, \hat{\Omega}$ are exacerbated.

→ Bootstrap.

II). Bootstrap:

(1). Estimate VAR by OLS → $\hat{\alpha}, \hat{\varphi}_1, \dots, \hat{\varphi}_k, \hat{\Omega}$

→ residuals. $\hat{\varepsilon}_t := Y_t - \hat{\alpha} - \hat{\varphi}_1 Y_{t-1} - \dots - \hat{\varphi}_k Y_{t-k}$, $t = k+1, \dots, T$

(2). Invert to get VMA(∞), IRFs.

(3). For $b = 1, \dots, B$:

(a) Form $Y_t^b = \hat{\alpha} + \hat{\varphi}_1 Y_{t-1}^b + \dots + \hat{\varphi}_k Y_{t-k}^b + \varepsilon_t^b$.

ε_t^b is randomly sampled (with replacement) from $\{\hat{\varepsilon}_\tau\}_{\tau=k+1}^{T}$

(b) Estimate 3(a) by OLS → $\hat{\alpha}^b, \hat{\varphi}_1^b, \dots, \hat{\varphi}_k^b, \hat{\Omega}^b$.

(c). Invert to get VMA(∞), IRFs.

(4). For each i, j, s. sort $\widehat{IRF}^b(i,j,s)$ in ascending order.

$\widehat{IRF}^{(1)}(i,j,s) \le \cdots \le \widehat{IRF}^{(B)}(i,j,s)$.

(5). Form a confidence interval.

$[\widehat{IRF}^{([B\alpha/2])}(i,j,s), \widehat{IRF}^{([B(1-\alpha/2)])}(i,j,s)]$ (Efron's Interval)

$\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$

Structural VAR

$B_0 Y_t = c + B_1 Y_{t-1} + \dots + B_k Y_{t-k} + V_t$, $V_t \sim WN(0, D)$

$B_0 = \begin{pmatrix} 1 & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & 1 & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & 1 \end{pmatrix}$ 1s on the main diagonal. B_0 describes contemporaneous equilibrium relationships.

e.g. supply & demand.

$S_t = c_1 + a_{12} d_t + \dots + V_{1t}$

$d_t = c_2 + a_{21} S_t + \dots + V_{2t}$

parameters: $\underset{c}{n} + \underset{B_1, \dots, B_k}{kn^2} + \underset{B_0}{n(n-1)} + \underset{D}{\frac{1}{2} n(n+1)}$

However, in VAR: $Y_t = \alpha + \varphi_1 Y_{t-1} + \dots + \varphi_k Y_{t-k} + \varepsilon_t$

parameters: $\underset{\alpha}{n} + \underset{\varphi_1, \dots, \varphi_k}{kn^2} + \underset{\Omega}{\frac{1}{2} n(n+1)}$

⇒ SVAR has more parameter than VAR.

⇒ Cannot, in general, identify SVAR from VAR.

Solution: impose extra restriction on SVAR.

1). B_0 is lower triangular

$B_0 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & 0 \end{pmatrix}$ $\frac{1}{2} n(n-1)$

Idea: order by the speed of reaction. e.g.

Y_{1t} = gdp (slowest to react). Y_{2t} = inflation (modest to react)

Y_{3t} = interest rate (fastest to react)

2) $E V_t V_t' = D$ is diagonal.

Idea: correlation is already captured by B_0, via simultaneous equations).

parameters: $n + kn^2 + \frac{1}{2} n(n+1)$ same in VAR and SVAR with restrictions 1), 2).

$\Omega = LL' = AGA'$ ← LDL decomp., $L = AG^{1/2}$

↑ Cholesky decomposition.

G-diag, A-lower triag with 1 on the main diag.

→ $D = (\mathrm{diag}\{L\})^2$, $B_0^{-1} = L(\mathrm{diag}(L))^{-1}$

Brownian Motion

Let $u_t \sim iid(0, \sigma^2)$, $t = 1, \dots, T$

$S(\tau) = \frac{1}{\sqrt{T}} \sum_{i=1}^{[T\tau]} u_i$ $\tau \in [0,1]$

$[x] = \max\{n \in \mathbb{Z} \text{ s.t. } n \le x\}$ e.g. $T = 101$, $\tau = \frac{1}{2}$, $[\frac{T}{2}] = 50$, $S(\frac{1}{2}) = \frac{1}{\sqrt{101}} \sum_{i=1}^{50} u_i$

$[T\tau] \in [T\tau - 1, T\tau]$ $\frac{[T\tau]}{T} \in [\tau - \frac{1}{T}, \tau] \to \tau$ as $T \to \infty$

$S(\tau) = \underbrace{\frac{\sqrt{[T\tau]}}{\sqrt{T}}}_{\to \sqrt{\tau}} \cdot \underbrace{\frac{1}{\sqrt{[T\tau]}} \sum_{i=1}^{[T\tau]} u_i}_{\xrightarrow{d} N(0, \sigma^2)}$ as $T \to \infty$

$S(\tau) \xrightarrow{d} \sqrt{\tau} N(0, \sigma^2) = N(0, \tau\sigma^2)$

common errors

• $\tau_1 \le \tau_2$

$\begin{pmatrix} S(\tau_1) \\ S(\tau_2) \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{[T\tau_1]}}{\sqrt{T}} \cdot \frac{1}{\sqrt{[T\tau_1]}} \sum_{i=1}^{[T\tau_1]} u_i \\ \frac{\sqrt{[T\tau_2]}}{\sqrt{T}} \cdot \frac{1}{\sqrt{[T\tau_2]}} \sum_{i=1}^{[T\tau_2]} u_i \end{pmatrix} \xrightarrow[T \to \infty]{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1 \sigma^2 & \tau_1 \sigma^2 \\ \tau_1 \sigma^2 & \tau_2 \sigma^2 \end{pmatrix} \right)$

Generalize:

$\begin{pmatrix} S(\tau_1) \\ S(\tau_2) \\ \vdots \\ S(\tau_k) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 \tau_1 & \sigma^2 \min\{\tau_1, \tau_2\} & \cdots & \vdots \\ \sigma^2 \min\{\tau_1, \tau_2\} & \sigma^2 \tau_2 & & \vdots \\ \vdots & & \ddots & \\ \cdots & \cdots & & \sigma^2 \tau_k \end{pmatrix} \right)$

However: our goal is to find something similar for all $\tau \in [0,1]$ (infinite, uncountable set of points τ)

→ Functional CLT

• Def. The Brownian motion or Wiener process is a stochastic process $W(\cdot)$ s.t.:

1) $W(0) = 0$ 2) For $0 \le t_1 \le t_2 \le \dots \le t_k$, increments $W(t_2) - W(t_1)$, $W(t_3) - W(t_2)$, …, $W(t_k) - W(t_{k-1})$ are independent.

3) Increments are Gaussian, $W(t) - W(s) \sim N(0, t-s)$ for $t > s$

4) $W(t)$ is almost surely continuous (i.e. W has continuous paths) $t \ge 0$

• Properties of $W(t)$:

1) $W(t) = W(t) - W(0) \sim N(0, t-0) = N(0, t)$

$E[W(t)] = 0$, $E[W^2(t)] = var(W(t)) = t$

density $f_{W(t)}(x) = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t})$

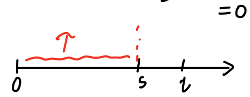
2) $cov(W(t), W(s)) = \min\{t, s\}$.

proof: Let $t \ge s$. then:

$cov(W(t), W(s)) = E[W(t)W(s)] = E[(W(t) - W(s) + W(s))W(s)]$

$= E[(W(t) - W(s))W(s)] + E[W^2(s)] = \underbrace{E[W(t) - W(s)]}_{=0} E[W(s)] + s = s$

3). Let $t \ge s$.



$E[W(t) \mid W(\tau), \tau \le s] = E[W(t) - W(s) + W(s) \mid W(\tau), \tau \le s]$

$= E[W(t) - W(s)] + W(s) = W(s)$. $W(t)$ is a martingale

• Wiener Process as a limit of random walk.

Let $X_1, X_2, \dots, \sim iid\, N(0,1)$. $\gamma\in[0,1]$

$S_T(\gamma) = \frac{1}{\sqrt{T}}\sum_{i=1}^{\lfloor T\gamma\rfloor} X_i$ where $\lfloor x\rfloor = \max\{z\in\mathbb{Z}\ s.t.\ n\le x\}$

• $S_T(\gamma) \overset{d}{\sim} N(0, \frac{\lfloor T\gamma\rfloor}{T}) \approx N(0,\gamma)$ as $\frac{\lfloor T\gamma\rfloor}{T}\in[\gamma-\frac1T, \gamma]$

• $S_T(\gamma)-S_T(s) = \frac{1}{\sqrt T}\sum_{i=\lfloor Ts\rfloor+1}^{\lfloor T\gamma\rfloor} X_i \overset{d}{\sim} N(0, \frac{\lfloor T\gamma\rfloor - \lfloor Ts\rfloor}{T}) \approx N(0,\gamma-s)$

• Increments $0\le t_1\le t_2\le\cdots\le t_k$: $S_T(t_2)-S_T(t_1)$, $S_T(t_3)-S_T(t_2)$, ……, $S_T(t_k)-S_T(t_{k-1})$ are independent. They do not share common X_i's.

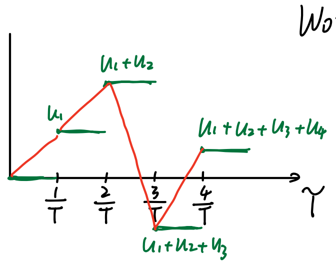
As $T\to\infty$, $S_T(\cdot)$ approaches $W(\cdot)$.

Theorem: (Functional CLT / Donsker's theorem)

Let $u_t\sim iid(0,\sigma^2)$. Then: $\frac{1}{\sqrt T}\sum_{i=1}^{\lfloor T\gamma\rfloor} u_i \xrightarrow[T\to\infty]{d} \sigma W(\gamma)$, $\gamma\in[0,1]$

$\xrightarrow{d}$: For r.v. we say $Y_n\xrightarrow{d}Y$ as $n\to\infty$ if $\forall$ cont. & bounded f-n $f:\mathbb R\to\mathbb R$, s.t. $E[f(Y_n)]\to E[f(Y)]$ as $n\to\infty$

Work with continuous functions!



$\sum_{t=1}^{\lfloor T\gamma\rfloor} u_t$: discontinuous

$\sum_{t=1}^{\lfloor T\gamma\rfloor} u_t + \frac{u_{\lfloor T\gamma\rfloor+1}}{\frac{t}{T}-\frac{t-1}{T}}(\gamma - \frac{\lfloor T\gamma\rfloor}{T})$ continuous.

$=\sum_{t=1}^{\lfloor T\gamma\rfloor} u_t + u_{\lfloor T\gamma\rfloor+1}(T\gamma-\lfloor T\gamma\rfloor)$

• Def. Functional $F: C[0,1]\to\mathbb R$

Sup-norm $d_\infty(g_1,g_2)=\sup_{\gamma\in[0,1]}|g_1(\gamma)-g_2(\gamma)|$

Benefit of d_∞: If $g_n\to g$ in d_∞ and $g_n\in C[0,1]$, then $g\in C[0,1]$.

• Continuity: F is continuous if $\forall\varepsilon>0, \exists\delta>0$, s.t. if $g_1,g_2\in C[0,1]$ and $d_\infty(g_1,g_2)<\delta$, then $|\underbrace{F(g_1)}_{\mathbb R}-\underbrace{F(g_2)}_{\mathbb R}|<\varepsilon$

• Boundedness: F is bounded if $\exists M>0$, s.t. $\forall g\in C[0,1]$, $|F(g)|\le M\|g\|_\infty \equiv M\sup_{\gamma\in[0,1]}|g(\gamma)|$.

e.g. 1) $F(g)=g(x^*)$, x^* = fixed 2) $F(g)=\int_A^B g(x)dx$, A, B fixed

• Unit Root Asymptotics

$\begin{cases} Y_t=\beta Y_{t-1}+\varepsilon_t, \ |\beta|<1 \ \text{stat. AR(1)} \quad \text{v.s.} \\ Y_t=Y_{t-1}+\varepsilon_t, \ (\text{i.e. } \beta=1) \ \text{unit root.}\end{cases}$ $\varepsilon_t\sim iid(0,\sigma^2)$

They have very different asymptotic behavior.

• $|\beta|<1$: $Y_t=\beta Y_{t-1}+\varepsilon_t = \sum_{i=0}^\infty \beta^i\varepsilon_{t-i}$ MA(∞) representation.

1). $\frac{1}{\sqrt T} Y_{\lfloor T\gamma\rfloor} = \frac{1}{\sqrt T}\sum_{i=0}^\infty \beta^i \varepsilon_{\lfloor T\gamma\rfloor-i} \xrightarrow{p} 0$ as $T\to\infty$

• $E[\frac{1}{\sqrt T}\sum_{i=0}^\infty\beta^i\varepsilon_{\lfloor T\gamma\rfloor-i}]=0$

• $Var(\frac{1}{\sqrt T}\underbrace{\sum_{i=1}^\infty \beta^i\varepsilon_{\lfloor T\gamma\rfloor-i}}_{=y_t,\ t=\lfloor T\gamma\rfloor}) = \frac1T var(y_t) = \frac1T\frac{\sigma^2}{1-\beta^2}\to 0$ as $T\to\infty$

2). $\bar Y := \frac1T\sum_{t=1}^T Y_t \xrightarrow{p} E[Y_t]=0$

3). OLS Estimate: $\hat\beta = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = \beta + \frac{\sum_t \varepsilon_t y_{t-1}}{\sum_t y_{t-1}^2}$

$\therefore \sqrt T(\hat\beta-\beta) = \dfrac{\frac{1}{\sqrt T}\sum_t \varepsilon_t y_{t-1} \xrightarrow{d} N(0,\frac{\sigma^4}{1-\beta^2})}{\frac1T\sum_t y_{t-1}^2 \xrightarrow{p} E[Y_{t-1}^2]=\frac{\sigma^2}{1-\beta^2}}$

Since: $E[\varepsilon_t y_{t-1}]=0$

$E[\varepsilon_t^2 y_{t-1}^2] = E[\varepsilon_t^2]E[y_{t-1}^2]=\sigma^2\cdot\frac{\sigma^2}{1-\beta^2}$ since independent.

$t>s$: $E[\varepsilon_t y_{t-1}\varepsilon_s y_{s-1}] = E[\varepsilon_t]E[y_{t-1}\varepsilon_s y_{s-1}]=0$.

Therefore, $\sqrt T(\hat\beta-\beta)\xrightarrow{d} N(0,1-\beta^2)$

4) t-stat: $(\hat\beta-\beta)/sd(\hat\beta)\xrightarrow{d}N(0,1)$

• $\beta=1$: $Y_t=Y_{t-1}+\varepsilon_t = Y_0+\sum_{s=1}^t\varepsilon_s$ Results are the same if Y_0 = const. or $Y_0=\bar o(\sqrt T)$ i.e. $\frac{Y_0}{\sqrt T}\xrightarrow[T\to\infty]{}0$.

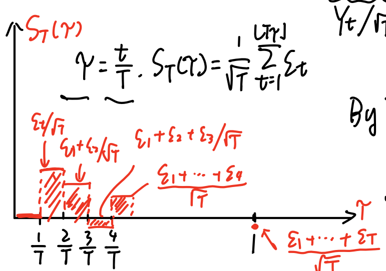
Assume $Y_0=0$, then $Y_t=\sum_{s=1}^t\varepsilon_s$

1) $\frac{1}{\sqrt T}Y_{\lfloor T\gamma\rfloor} = \frac{1}{\sqrt T}\sum_{s=1}^{\lfloor T\gamma\rfloor}\varepsilon_s \xrightarrow[T\to\infty]{d}\sigma W(\gamma)$ by FCLT

$t=\lfloor T\gamma\rfloor\in\{0,1,\dots,T\}$, $\gamma\in[0,1]$

2) $\frac{1}{\sqrt T}\bar Y := \frac{1}{\sqrt T}\frac1T\sum_{t=1}^T Y_t = \frac1T\sum_{t=1}^T\underbrace{\frac{1}{\sqrt T}\sum_{s=1}^t\varepsilon_s}_{Y_t/\sqrt T} = \frac1T\sum_{t=1}^T S_T(\frac tT)$

$=\int_0^1 S_T(\gamma)d\gamma + \frac1T S_T(1)$



By FCLT: $\int_0^1 S_T(\gamma)d\gamma \xrightarrow{d}\sigma\int_0^1 W(\gamma)d\gamma$

$\underbrace{\frac1T}_{0}\underbrace{S_T(1)}_{\sigma W(1) \text{ by FCLT}}\xrightarrow{T\to\infty}0$

Therefore $\frac{1}{\sqrt T}\bar Y\xrightarrow{d}\sigma\int_0^1 W(\gamma)d\gamma$. as $T\to\infty$

Similarly, $\frac{1}{T^{1+k/2}}\sum_{t=1}^T Y_t^k = \frac1T\sum_{t=1}^T(\frac{Y_t}{\sqrt T})^k = \underbrace{\frac1T}_{\text{width}}\underbrace{\sum_{t=1}^T\left(S_T(\frac tT)\right)^k}_{\text{weight}}$ area, i.e. integral.

$\xrightarrow{d}\sigma\int_0^1 W^k(\gamma)d\gamma$

3). $\hat\beta=\frac{\sum_t Y_tY_{t-1}}{\sum_t Y_{t-1}^2}=1+\frac{\sum_t\varepsilon_tY_{t-1}}{\sum_tY_{t-1}^2}$

$T(\hat\beta-1)=\dfrac{\frac1T\sum_t\varepsilon_tY_{t-1}}{\frac{1}{T^2}\sum_tY_{t-1}^2}$, $\frac{1}{T^2}\sum_tY_{t-1}^2 \xrightarrow{d}\sigma\int_0^1W^2(\gamma)d\gamma$ as above

$\frac1T\sum_t\varepsilon_tY_{t-1}=\frac1T\sum_{t=1}^T\varepsilon_t\sum_{s=1}^{t-1}\varepsilon_s=\frac1T\sum_{t>s}\varepsilon_t\varepsilon_s=(*)$

$Y_t^2=(\sum_{s=1}^T\varepsilon_s)^2=\sum_{s=1}^T\varepsilon_s^2+2\sum_{t>s}\varepsilon_t\varepsilon_s$

$\Rightarrow \sum_{t>s}\varepsilon_t\varepsilon_s=\frac12(Y_t^2-\sum_{s=1}^T\varepsilon_s^2)$

$(*)=\frac{1}{2T}(Y_t^2-\sum_{s=1}^T\varepsilon_s^2)=\frac12\left(\frac{Y_t}{\sqrt T}\right)^2-\frac12\underbrace{\frac1T\sum_{t=1}^T\varepsilon_s^2}_{\sigma^2 \text{ (LLN)}}$

$\left(\frac{Y_t}{\sqrt T}\right)^2=\left(\frac{1}{\sqrt T}\sum_{i=1}^T\varepsilon_i\right)^2 \xrightarrow{d} (\sigma W(1))^2$ by FCLT

$\beta=1$

$\xrightarrow{d}\frac{\sigma^2}{2}(W^2(1)-1)$

So $T(\hat\beta-1)\xrightarrow{d}\dfrac{\frac{\sigma^2}{2}(W^2(1)-1)}{\sigma^2\int_0^1W^2(\gamma)d\gamma}=\dfrac{W^2(1)-1}{2\int W^2(\gamma)d\gamma}$

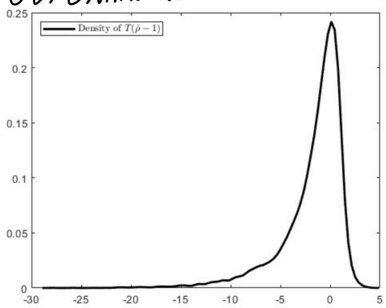
4) t-statistics $=\frac{\hat\beta-1}{\hat\sigma}\sqrt{\sum_{t=1}^TY_{t-1}^2}=\frac{T(\hat\beta-1)}{\hat\sigma}\sqrt{\frac{1}{T^2}\sum_{t=1}^TY_{t-1}^2}$

$\hat\sigma^2=\frac1T\sum_t(Y_t-\hat\beta Y_{t-1})^2\to\sigma^2$ in PS6 $\xrightarrow{d}\dfrac{W^2(1)-1}{\sigma 2\int W^2(\gamma)d\gamma}\sqrt{\sigma\int_0^1W^2(\gamma)d\gamma}$

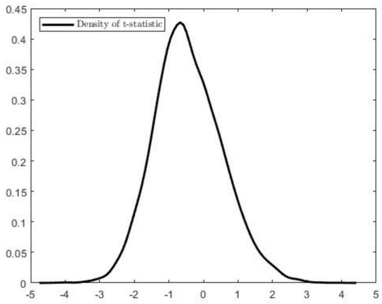
$=\dfrac{W^2(1)-1}{2\sqrt{\sigma\int_0^1W^2(\gamma)d\gamma}}$

Summary: very different asymptotic distribution and convergence rates ($\sqrt{T}$ vs T)

OLS estimation:



t-statistics



• Spurious Regression:

$Y_t = Y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid(0, \sigma^2)$, $X_t = X_{t-1} + u_t$, $u_t \sim iid(0, \lambda^2)$

$\varepsilon_t \perp\!\!\!\perp u_s$, $\forall t, s$.

Run a regression $Y_t = \beta X_t + V_t$, true $\beta = 0$.

However, $\hat{\beta} \gg 0$:

$$\hat{\beta} = \frac{\sum_t Y_t X_t}{\sum_t X_t^2} = \frac{\frac{1}{T^2}\sum_t Y_t X_t \xrightarrow{d} \sigma\lambda\int W_x(r)W_y(r)dr}{\frac{1}{T^2}\sum_t X_t^2 \xrightarrow{d} \lambda^2\int W_x^2(r)dr}$$

since $\frac{1}{T}\sum_t \underbrace{\frac{Y_t}{\sqrt{T}}}_{\sigma W_y(r)} \underbrace{\frac{X_t}{\sqrt{T}}}_{\lambda W_x(r)} \xrightarrow{d} \sigma\lambda\int_0^1 W_y(r)W_x(r)dr$

$$\xrightarrow[T\to\infty]{d} \frac{\sigma}{\lambda}\frac{\int_0^1 W_x(r)W_y(r)dr}{\int_0^1 W_x^2(r)dr}$$

non standard random variable

$t = \frac{\hat{\beta}}{\hat{\sigma}}\sqrt{\sum_t X_t^2} = \sqrt{T}\left(\begin{matrix}\text{random}\\ \text{variable}\end{matrix}\right) \to \infty$.

$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T (Y_t - \hat{\beta}X_t)^2 \quad \frac{1}{T}\sigma^2 \to$ random variable.

$H_0: \beta = 0$ is rejected as $|t| \nearrow \infty$

However, if $\begin{cases} Y_t = \beta X_t + \varepsilon_t, \ \beta \neq 0 \\ X_t = X_{t-1} + u_t \end{cases}$, then $T(\hat{\beta} - \beta) \xrightarrow{d} \left(\begin{matrix}\text{non-standard}\\ \text{random variable}\end{matrix}\right)$

i.e. $\hat{\beta} \xrightarrow[T\to\infty]{P} \beta$

Cointegration: X_t, Y_t are unit roots, but $X_t - \theta Y_t$ is stationary for some $\theta \in \mathbb{R}$

e.g. X_t, Y_t = % rates of maturities, $X_t - Y_t$ = spread

• Autoregressive Conditional Heteroskedasticity

Time-varying Volatility.

AR, MA: about $E[Y_t | \text{past}]$. e.g. in AR(1): $Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t$,

$E[Y_t | Y_{t-1}, \dots] = \alpha + \beta Y_{t-1}$

Nothing about $var(Y_t | \text{past})$. $E[\varepsilon_t^2 | \text{past}] = \sigma^2$, not depend on t.

Now: replace σ^2 with σ_t^2

$Y_t = X_t'\beta + u_t$. (X_t may include lags of Y_t. baseline: $X_t \equiv 1$, $Y_t = \beta + u_t$)

$\begin{cases} E_{t-1}[u_t] := E[u_t | u_{t-1}, \dots] = 0 \quad \text{(MDS)} \\ E_{t-1}[u_t^2] := E[u_t^2 | u_{t-1}, \dots] = \sigma_t^2, \quad u_t \sim WN(0, \sigma^2) \end{cases}$

☆ $\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2$, $\alpha_0 > 0$, $\alpha_1 \cdots \alpha_q \geq 0$ ARCH(q).

$E_{t-1}[u_t] = 0$, $E_{t-1}[u_t^2] = \sigma_t^2$,

So 1) $E[u_t] = E[E_{t-1}[u_t]] = 0$

2) $h > 0$: $E[u_t u_{t+h}] = E[E_{t+h-1}[u_t u_{t+h}]] = E[u_t \underline{E_{t+h-1}[u_{t+h}]}] = 0$

3) to let $E[u_t^2] = \sigma^2$, also need $\sum_{i=1}^q \alpha_i < 1$

ARCH(q) is weakly stationary iff $\sum_{i=1}^q \alpha_i < 1$.

If model is stat., then, $E[u_t^2] = \sigma^2 \quad \forall t$.

Thus, $\sigma^2 = E[u_t^2] = E[E_{t-1}[u_t^2]] = E[\sigma_t^2]$

$= E[\alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2] = \alpha_0 + \sum_{i=1}^q \alpha_i \underbrace{E[u_{t-i}^2]}_{=\sigma^2 \text{ by stat.}}$

$= \alpha_0 + \sigma^2 \sum_{i=1}^q \alpha_i$

$\Rightarrow \sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i} \qquad \alpha_0 = \sigma^2(1 - \sum_{i=1}^q \alpha_i)$

$\sum_{i=1}^q \alpha_i < 1$ guarantees that $\sigma^2 > 0$.

$\longrightarrow \sigma_t^2 - \sigma^2 = \sigma^2(1 - \sum_{i=1}^q \alpha_i) + \sum_{i=1}^q \alpha_i u_{t-i}^2 - \sigma^2$

$= \sum_{i=1}^q \alpha_i (u_{t-i}^2 - \sigma^2)$ deviation from the mean form.

Why Autoregressive conditional heteroskedasticity?

Let $V_t := u_t^2 - \sigma_t^2 = u_t^2 - E_{t-1}[u_t^2]$

$\longrightarrow E_{t-1}[V_t] := E[V_t | u_{t-1}, \dots] = E_{t-1}[u_t^2] - E_{t-1}[E_{t-1}[u_t^2]] \equiv 0$

$\longrightarrow \forall h > 0$: $E[V_t V_{t+h}] = E[E_{t+h-1}[V_t V_{t+h}]]$

$= E[V_t \underbrace{E_{t+h-1}[V_{t+h}]}_{=0}] = 0$

So $V_t \sim WN$ with 0 mean.

$\underset{u_t^2 - V_t}{\underset{\parallel}{\sigma_t^2}} = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2$

$\Rightarrow u_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + V_t$, AR(q) for u_t^2

• $\sum_{i=1}^q \alpha_i < 1$, $\alpha_i \geq 0 \ \forall i = 1, \dots, q$ implies stat. for this AR(q).

i.e. $E[u_t^2]$ = constant.

• The errors u_t are leptokurtic.

Kurtosis: $kurt(X) = \frac{E[(X - E[X])^4]}{(E[(X - E[X])^2])^2}$

If $X \sim N(\mu, \sigma^2)$, then $(E[(X - E[X])^2])^2 = (\sigma^2)^2 = \sigma^4$

$E[(X - E[X])^4] = 3\sigma^4$.

$\Rightarrow kurt(X) = \frac{3\sigma^4}{\sigma^4} = 3$

leptokurtic. $kurt > 3$: fatter tails than N

Platykurtic. $kurt < 3$: thinner tails than N

• Let $u_t | u_{t-1} \cdots \sim N(0, \sigma_t^2)$

Let $\varepsilon_t := \frac{u_t}{\sigma_t} \sim N(0, 1)$

$E_{t-1}[\varepsilon_t] = E_{t-1}\left[\frac{u_t}{\sigma_t}\right] = \frac{1}{\sigma_t} \overset{0}{\overset{\parallel}{E_{t-1}[u_t]}} = 0$

Known at $t-1$

$E_{t-1}[\varepsilon_t^2] = E_{t-1}\left[\frac{u_t^2}{\sigma_t^2}\right] = \frac{1}{\sigma_t^2} E_{t-1}[u_t^2] = \frac{\sigma_t^2}{\sigma_t^2} = 1$

$Kurt(u_t) = \frac{E[u_t^4]}{E[u_t^2]^2} = \frac{E[(\varepsilon_t \sigma_t)^4]}{(\sigma^2)^2}$

$= \frac{E[E_{t-1}[\sigma_t^4 \varepsilon_t^4]]}{\sigma^4} = \frac{E[\sigma_t^4 E_{t-1}[\varepsilon_t^4]]}{\sigma^4}$

$= \frac{3E[\sigma_t^4]}{\sigma^4}$ since $\varepsilon_t | u_{t-1}, \dots \sim N(0, 1)$

By Jensen's Ineq. $E[(\sigma_t^2)^2] \geq (E[\sigma_t^2])^2 = (\sigma^2)^2 = \sigma^4$

Strict unless σ_t^2 = const. i.e. if $\sigma_t^2 \neq$ const, $Kurt(u_t) > 3$.

Fatter than N tails → some moments may not exist.

Generalized Autoregressive conditional Heteroskedasticity.

GARCH

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad \text{GARCH}(p, q)$$

$$\begin{cases} \alpha_0 > 0, \quad \alpha_i \geq 0 \quad \forall\, i = 1, \cdots, q \\ \qquad\qquad \beta_j \geq 0 \quad \forall\, j = 1, \cdots, p \\ \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1 \end{cases}$$

$$\sigma^2 = \mathbb{E}[\sigma_t^2] = \alpha_0 + \sum_{i=1}^{q} \alpha_i \sigma^2 + \sum_{j=1}^{p} \beta_j \sigma^2$$

$$\Rightarrow \quad \sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}$$

$$V_t := u_t^2 - \mathbb{E}_{t-1}[u_t^2] = u_t^2 - \sigma_t^2$$

$$\sigma_t^2 = u_t^2 - V_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j (u_{t-j}^2 - V_{t-j})$$

$$u_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j u_{t-j}^2 + V_t - \sum_{j=1}^{p} \beta_j V_{t-j}$$

ARMA$(\max\{p, q\}, p)$ for u_t^2