$a<0 \Rightarrow \rho_\tau(X_1,X_2)=-1$

Let g and h be strictly increasing

Then $\rho_\tau(g(X_1),h(X_2)) = \rho_\tau(X_1,X_2)$ because

$$\underbrace{\text{sign}\{(X_1-X_1^*)(X_2-X_2^*)\}}_{\text{if}} = \underbrace{\text{sign}\{(g(X_1)-g(X_1^*))(h(X_2)-h(X_2^*))\}}_{\text{then}}$$

$\therefore \rho_\tau(X_1,X_2)$ depends only on the copula of (X_1,X_2)

Reason $\rho_\tau(X_1,X_2) = \rho_\tau(\underbrace{F_{X_1}(X_1), F_{X_2}(X_2)}_{\text{dist.}=\text{copula}})$

$\begin{pmatrix} X_{i,1} \\ X_{i,2} \end{pmatrix}$ $i=1,\dots,n$ is an iid sample

$$\hat{\rho}_\tau(X_1,X_2) = \underbrace{\binom{n}{2}^{-1}}_{\text{binomial}=\,{}_nC_2} \sum_{1\le i\le j\le n} \text{sign}\{(X_{i,1}-X_{j,1})(X_{i,2}-X_{j,2})\}$$

Spearman's rule

(corr. of transformed) ← $\rho_S(X_1,X_2) = \text{corr}(F_{X_1}(X_1), F_{X_2}(X_2))$ → depends only on copula not on marginals

Sample version

$\begin{pmatrix} X_{i,1} \\ X_{i,2} \end{pmatrix}$ $i=1,\dots n$

$\text{rank}(X_{i,1}) = \sum_{j=1}^{n} I\{X_{i,1} \le X_{j,1}\}$

① 7	18	②
② 3	15	①
③ 11	21	③

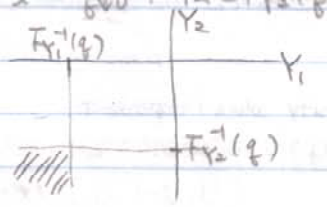
$\hat{\rho}_S(X_1,X_2)$ = sample correlation of $[\text{rank}(X_{i,1}), \text{rank}(X_{i,2})]$

$\Rightarrow (3,3)\ (5,5), (2,2) \cdots \Rightarrow$ ranks are indentical $\Rightarrow$ corr $=1$

Coefficient of lower tail dependence

$P(A|B) = \frac{P(A\cap B)}{P(B)}$ → $\lambda_l = \lim_{q\downarrow 0} P(Y_2 \le F_{Y_2}^{-1}(q) \mid Y_1 \le F_{Y_1}^{-1}(q)) = \lim_{q\downarrow 0} \frac{P(Y_2 \le F_{Y_2}^{-1}(q)\ \&\ Y_1 \le F_{Y_1}^{-1}(q))}{q} = \lim_{q\downarrow 0} \frac{P(F_{Y_2}(Y_2)\le q\ \&\ F_{Y_1}(Y_1)\le q)}{q}$

(asymmetric in Y_1 and Y_2)



$= \lim_{q\downarrow 0} \frac{C_{(Y_1,Y_2)}(q,q)}{q}$ ⇒ depends only on the copula

Gaussian meta-distribution : distribution whose copula is gaussian

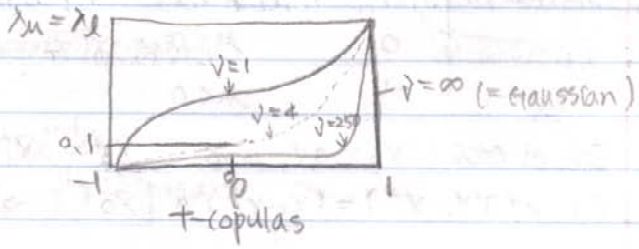
$$\lambda_l = \begin{cases} 0 & \text{if } \rho<1 \\ 1 & \text{if } \rho=1 \quad (\because P(F_{Y_2}(Y_2)\le q)=q \quad P(F_{Y_1}(Y_1)\le q)=q) \end{cases}$$

t-copula

2 parameters : ν, ρ

$\lambda_l = 2F_{t,\nu+1}\left[-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right]$

t CDF with $(\nu+1)$ df



t-copulas

$$\lim_{q\uparrow 1} C_{(Y_1,Y_2)}(q,q) = 1$$
$$\Downarrow$$

$$\lambda_u = \lim_{q\uparrow 1} P(Y_2 \geq F_{Y_2}^{-1}(q) \mid Y_1 \geq F_{Y_1}^{-1}(q)) = 2 - \lim_{q\uparrow 1} \frac{1 - C_{(Y_1,Y_2)}(q,q)}{1-q}$$

Gaussian, t-copulas, $\lambda_l = \lambda_u$

Calibrating Copulas

Data $Y_i = \begin{pmatrix} Y_{i,1} \\ \vdots \\ Y_{i,d} \end{pmatrix}$ iid

Model $C_Y(u_1, \dots, u_d; \theta_c)$ parameter vector

$f_{Y_1}(y_1; \theta_1), \dots, f_{Y_d}(y_d; \theta_d)$

$\theta = \begin{pmatrix} \theta_c \\ \theta_1 \\ \vdots \\ \theta_d \end{pmatrix}$

$f_Y(y_1, \dots, y_d; \theta) = C_Y\{F_{Y_1}(y_1;\theta_1), \dots, F_{Y_d}(y_d;\theta_d); \theta_c\} f_{Y_1}(y_1;\theta_1) \cdots f_{Y_d}(y_d;\theta_d)$

log-likelihood $= \sum_{i=1}^{n} \{ \underbrace{\log C_Y(F_{Y_1}(Y_{i,1},\theta_1), \dots, F_{Y_d}(Y_{i,d},\theta_d); \theta_c)}_{①} + \underbrace{\log f_{Y_1}(Y_{i,1};\theta_1) + \dots + \log f_{Y_d}(Y_{i,d};\theta_d)}_{②} \}$

MLE Max log-likelihood over θ.

Potential Problem : convergence to local max due to poor values.

Pseudo-MLE

① Estimate θ_1 uses $Y_{1,1}, \dots, Y_{n,1}$ ... Estimate θ_d using $Y_{1,d}, \dots, Y_{n,d}$

② Plug $\hat{\theta}_1, \dots, \hat{\theta}_d$ into part ① of the log-likelihood

③ Maximize part ① over θ_c to get $\hat{\theta}_c$

2 uses of the pseudo-MLE

a) use pseudo-MLE as final estimate

b) use pseudo-MLE as starting values of MLE (∵ pseudo MLE ≈ MLE)

Semiparametric pseudo-MLE

No parametric models are used for $f_{Y_1}, \dots, f_{Y_d}$

A parametric model is used for the copula

Replace $F_{Y_j}(y; \hat{\theta}_j)$ by $\hat{F}_{Y_j}(y) = \underbrace{\frac{1}{n+1} \sum_{i=1}^{n} I\{Y_{i,j} \leq y\}}_{\text{Empirical CDF (EDF)}}$

Usually, the EDF as $\frac{1}{n}$ not $\frac{1}{n+1}$

If $\frac{1}{n}$ is used, then $\hat{F}_{Y_j}(Y_{i,j}) = 1$ for one choice i. However $C_Y(u_1, \dots, u_d)$ often equals $+\infty$ if some $u_j = 1$

With semiparametric pseudo-MLE, we estimate θ_c only

Even pseudo-MLE can have many parameters

Ex: Gaussian copula

$\Omega = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1d} \\ & \ddots & & \vdots \\ & 0 & & \rho_{d-1,d} \\ & & & 1 \end{bmatrix}$ $\quad$ # parameter $= 1+2+3+\dots+(d-1) = \frac{(d-1)d}{2}$ can be big

For Gaussian and t-copulas, Ω can be estimated "directly" using rank correlation.

Ω is not the corr for Y but corr that defines gaussian copula

Suppose $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix}$ has meta-Gaussian distribution with copula $C^{Gauss}(u_1, \dots, u_d; \Omega)$

$\rho_\tau(Y_i, Y_j) = \frac{2}{\pi} \arcsin(\Omega_{ij})$ $\quad$ ($\arcsin = \sin^{-1}$)

$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ has a meta-Gaussian distribution

Copula of X is $C^{Gauss}(\cdot, \Omega)$

$\frac{6}{\pi}\sin^{-1}(\frac{x}{2})$ x if $|x| \le 1$

$\rho_\tau(X_i, X_j) = \frac{2}{\pi}\sin^{-1}(\Omega_{ij})$ pearson corr. $\quad$ (6.27)

$\Rightarrow \rho_S(X_i, X_j) = \frac{6}{\pi}\sin^{-1}(\frac{\Omega_{ij}}{2}) \approx \Omega_{ij}$ $\quad$ (6.28)

For a meta-t, (6.27) holds but not (6.28)

Gaussian: use spearman correlation matrix to estimate Ω.

This can be the final estimate or the starting value of MLE

t-copulas: Invert (6.27): $\hat{\Omega}_{ij} = \sin(\frac{\pi}{2}\rho_\tau(X_i, X_j))$

↑ plug in estimates

$\hat{\Omega}$ is symmetric but may not be positive definite.

1) check if all eigenvalues are positive - if 'yes' use $\hat{\Omega}$ as the estimate of Ω.

If there are negative eigen values, $\hat{\Omega} = \underbrace{O}_{\text{eigen vectors}} \underbrace{\text{diag}(\lambda_1, \dots, \lambda_d)}_{\text{eigen values}} O^T \Rightarrow \tilde{\Omega} = O\,\text{diag}(\max(\varepsilon, \lambda_i))_{\varepsilon > 0,\ i=1,\dots,d} O^T$

$\tilde{\Omega}$ is symmetric and positive definite

However, the diagonals of $\tilde{\Omega}$ are not all equal to 1

$\hat{\hat{\Omega}}_{ij} = \frac{\tilde{\Omega}_{ij}}{\sqrt{\tilde{\Omega}_{ii}\tilde{\Omega}_{jj}}}$

$\hat{\hat{\Omega}}$ is symmetric, positive definite, and a correlation matrix

Archimedean Copulas.

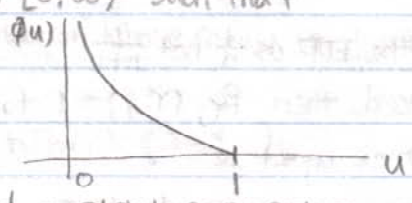
$C(u_1, \dots, u_d) = \phi^{-1}\{\phi(u_1) + \dots + \phi(u_d)\}$

Here, ϕ is a function from $[0,1]$ to $[0, \infty)$ such that

1) ϕ is decreasing

2) $\phi(0) = \infty$

3) $\phi(1) = 0$



Ex. $\phi(u) = -\log(u)$

If ϕ satisfies 1)-3), it is called a strict generator.

If $\phi(0)$ is finite, the ϕ is non-strict generator.

Ex. $\phi(u) = -\log u$

Generate the independence copula: $\phi^{-1}(t) = e^{-t}$

$C(u_1, \dots, u_d) = \phi^{-1}\{\phi(u_1) + \dots + \phi(u_d)\} = \exp\{\log u_1 + \dots + \log u_d\} = u_1 \cdots u_d$.

Let's check that $C(u_1, \cdots, u_d)$ is a CDF on $[0,1]^d$

① $C(u_1, \ldots, u_d)$ is increasing in each u_i ✓

② $C(0, \cdots, 0) = 0$ ✓

③ $\lim_{u_j \to 1 \forall j} C(u_1, \ldots, u_d) = \phi^{-1}(0) = 1$

Is it a copula?

$C(1, \ldots, 1, u_j, 1, \ldots, 1) = \phi^{-1}(\phi(u_j)) = u_j$ ✓

Frank Copula $C^{Fr}(u) = -\log\left[\frac{e^{-\theta u}-1}{e^{-\theta}-1}\right]$ $\Rightarrow$ strict : $u=0 \Rightarrow C^{Fr}(u) = \infty$

$(\phi^{Fr})^{-1}(y) = \frac{-\log[e^{-y}\{e^{-\theta}-1\}+1]}{\theta}$

$C^{Fr}(u_1, u_2) = -\frac{1}{\theta}\log\left[1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right]$

case: $\theta = 0$. as $\theta \to 0$, $\phi^{Fr}(u) \to -\log(u)$ independence copula

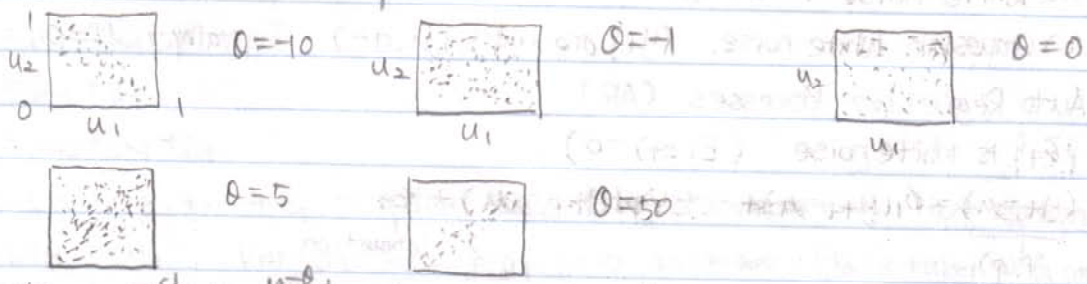
$\theta = -\infty$ Counter-monotonicity $(d=2)$

$\theta < 0$ negative correlation

$\theta = 0$ indep.

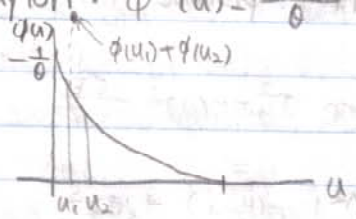
$\theta > 0$ positive corr.

$\theta = +\infty$ co-monotonicity



Clayton : $\phi^{Cl}(u) = \frac{u^{-\theta}-1}{\theta}$, $\theta \geq -1$ (strict if $\theta \geq 1$)

$\theta < 0$. $\frac{u^{-\theta}-1}{\theta}$



redefinition $C^{Cl}(u_1 u_2) = \phi^{-1}\{\phi(u_1)+\phi(u_2)\} \triangleq 0$ if $\phi(u_1)+\phi(u_2) \geq \phi(0)$

$\theta = -0.7$

no data

$\theta = -0.98$

$\theta = -1 \Rightarrow$ counter mono.

$\theta = 5$

Time Series Analysis

$y_1, \ldots, y_n$ is a sequence of data. y_t, where t is the time index

Stationarity : y_t is stationary if $(y_1, \ldots, y_m) \overset{D}{=} (y_{1+k}, \ldots, y_{m+k})$ $\forall m, \forall k$

In particular. $E(y_t)=\mu$, $Var(y_t)=\sigma^2=r_0$, $Cov(y_t, y_{t+h})=r(|h|)$ ← time interval, $r(h)=r(-h)$

r does not depend on t and is called the covariance function of $\{y_t\}$

$Corr(y_t, y_{t+h})=\frac{Cov(y_t, y_{t+h})}{r(0)}=\frac{r(h)}{r(0)}=\rho(h)$

$\rho(h)$ is called the auto-correlation function.

$r(h)$ is called the auto-covariance function.

To estimate $\hat{r}(h)=\frac{1}{n}\sum_{j=1}^{n-h}(y_{j+h}-\bar{y})(y_j-\bar{y})$

$\widehat{ACF}=\hat{\rho}(h)=\frac{\hat{r}(h)}{r(0)}$ $h=0,1,2\cdots$ ⇒ We want parametric models for the analysis

White Noise

(a) Weak white noise. $y_1, y_2, \cdots$ is weak white noise if:

1) $E(y_t)=\mu$
2) $Var(y_t)=\sigma^2$

} same as a stationary process.

3) $Cov(y_t, y_{t+h})=0$ if $h\neq 0$ → The process is uncorrelated

Ex. Garch processes

(b) iid White Noise: $y_1, y_2, \cdots$ are iid, mean is a constant, variance and all other moments are a constant. Also is <u>not</u> correlated. variance $<\infty$.

iid white noise + $Var(y_t)<\infty$ ⇒ weak white noise

(c) Gaussian White noise: $\{y_t\}$ are iid $N(\mu, \sigma^2)$ Typically, $\mu=0$

Auto Regressive Processes (AR)

$\{\varepsilon_t\}$ is white noise $(E(\varepsilon_t)=0)$.

$$\underbrace{(y_t-\mu)}_{AR(p)}=\phi_1(y_{t-1}-\mu)+\cdots+\phi_p(y_{t-p}-\mu)+\underbrace{\varepsilon_t}_{\text{innovation}}$$

$\phi_1, \cdots, \phi_p$ are parameters. p is the order of the AR process

(a) AR(1): $y_t-\mu=\phi(y_{t-1}-\mu)+\varepsilon_t$

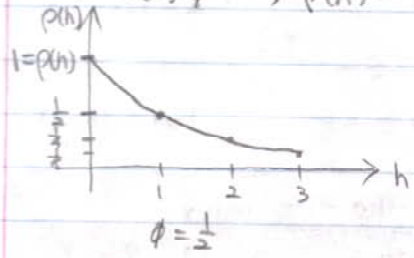
Fact. Stationary process ⇒ $|\phi|<1$, Therefore $\sigma_y^2=r(0)=\frac{\sigma_\varepsilon^2}{1-\phi^2}$

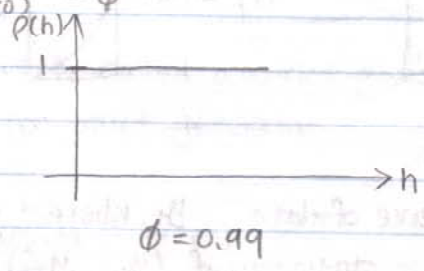
Proof:

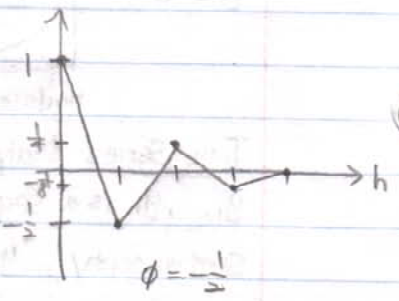
$r(0)=Var(y_t)=Var[\phi(y_{t-1}-\mu)]+Var[\varepsilon_t]=\phi^2\underbrace{Var(y_{t-1})}_{r(0)}+\sigma_\varepsilon^2$

$\therefore r(0)=\phi^2 r(0)+\sigma_\varepsilon^2$

$r(h)=r(0)\phi^h \Rightarrow \rho(h)=\frac{r(h)}{r(0)}=\phi^h$







• $\{y_t\}$ is only stationary if

1) $|\phi| < 1$

2) $E(y_1) = \mu$

3) $Var(y_1) = r(0) = \frac{\sigma_\varepsilon^2}{1-\phi^2}$

] if it starts with other mean and variance, the process y_t is not stationary, but converges to a stationary process

$(y_t - \mu) = \phi(y_{t-1} - \mu) + \varepsilon_t$

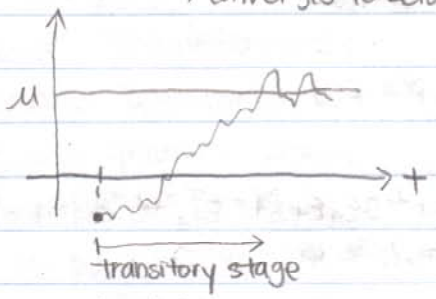
$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t$

$E(y_2) = \mu + \phi(E(y_1) - \mu)$

$E(y_3) = \mu + \phi(E(y_2) - \mu) = \mu + \phi^2(E(y_1) - \mu)$

$E(y_k) = \mu + \underbrace{\phi^{k-1} E(E(y_1) - \mu)}$

↳ converges to zero, geometrically fast as $k \to \infty$



mean reverting process, it is a characteristic of stationary process.

• $\phi = 1$ not stationary

$y_t = \mu + (y_{t-1} - \mu) + \varepsilon_t$

$y_t = y_{t-1} + \varepsilon_t$

$= y_{t-2} + \varepsilon_{t-1} + \varepsilon_t$

$\vdots$

$= y_0 + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$ → random walk (not mean reverting, not stationary)

$E(y_t | y_0) = y_0$, $Var(y_t | y_0) = t \cdot \sigma_\varepsilon^2 \to \infty$ as $t \to \infty$. (y_0 is taken as a constant)

Now, $y_t - y_{t-1} = \varepsilon_t$ → stationary. The changes are stationary and that apply to financial data

(b) AR(p) : $(y_t - \mu) = \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t$

B = backwards operator, Lag opperator

$By_t = y_{t-1}$, $B\mu = \mu$

$(1 - B\phi_1 - B^2\phi_2 - \cdots - B^p\phi_p)(y_t - \mu) = \varepsilon_t$ → polynomial representation

$(1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p)$ → tells us if the process is stationary it has p roots $r_1 \cdots r_p$.

$(1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p) = (1 - \frac{x}{r_1})(1 - \frac{x}{r_2}) \cdots (1 - \frac{x}{r_p})$

⇒ Stationary ⇔ $|\frac{1}{r_j}| < 1$ for $j = 1 \ldots p$

⇔ $|r_j| > 1$.

If $|r_j| = 1$ for some j, then y_t is unit-root nonstationary

Time series
→ stationary (same mean).
→ correlation depends on time length

Estimation

Conditional least-squares : $\hat{\mu}, \hat{\phi}_1, \dots, \hat{\phi}_p$ minimize

$$\sum_{t=p+1}^{n} \{(Y_t-\mu)-(\phi_1(Y_{t-1}-\mu)+\dots+\phi_p(Y_{t-p}-\mu))\}^2$$

This is the MLE assuming Gaussian White Noise and using conditional density of $(Y_{p+1},\dots,Y_n)$ given $(Y_1,\dots,Y_p)$

FullMLE maximizes the likelihood for $(Y_1,\dots,Y_n)$

Conditional LS estimator $\approx$ MLE

MA(q) process

$$Y_t=\mu+\varepsilon_t+\theta_1\varepsilon_{t-1}+\dots+\theta_q\varepsilon_{t-q}$$

not observed → regression is a bit difficult

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

$var(Y_t)$
$=Var(\mu+\varepsilon_t+\theta\varepsilon_{t-1}) \rightarrow \sigma_Y^2=\sigma_\varepsilon^2+\theta^2\sigma_\varepsilon^2=\sigma_\varepsilon^2(1+\theta^2)$

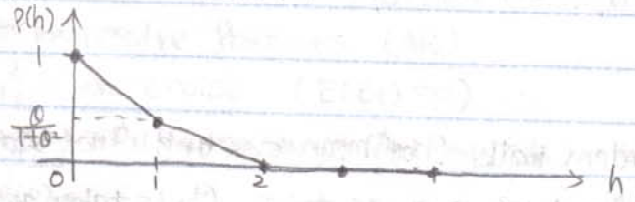
$r(0)=\sigma_\varepsilon^2(1+\theta^2)$

$r(1)=E\{(\varepsilon_t+\theta\varepsilon_{t-1})(\varepsilon_{t-1}+\theta\varepsilon_{t-2})\}=E\{\cancel{\varepsilon_t\varepsilon_{t-1}}+\theta\cancel{\varepsilon_t\varepsilon_{t-2}}+\theta\varepsilon_{t-1}^2+\theta^2\cancel{\varepsilon_{t-1}\varepsilon_{t-2}}\}=\theta\sigma_\varepsilon^2$

$\rho(1)=\frac{\theta\sigma_\varepsilon^2}{(1+\theta^2)\sigma_\varepsilon^2}=\frac{\theta}{1+\theta^2}$ (−) sign in lecture notes !! → incorrect)

$r(2)=E\{(\varepsilon_t+\theta\varepsilon_{t-1})(\varepsilon_{t-2}+\theta\varepsilon_{t-3})\}=0$

$r(h)=0$ for all $h>1$



MA(2) process

$r(0)=E(\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2})^2=\sigma_\varepsilon^2(1+\theta_1^2+\theta_2^2)$

$r(1)=E[(\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2})(\varepsilon_{t-1}+\theta_1\varepsilon_{t-2}+\theta_2\varepsilon_{t-3})]=\sigma_\varepsilon^2(\theta_1+\theta_1\theta_2)$

$r(2)=E[(\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2})(\varepsilon_{t-2}+\theta_1\varepsilon_{t-3}+\theta_2\varepsilon_{t-4})]=\sigma_\varepsilon^2\theta_2$

$r(h)=0 \quad h>2$

MA(q)

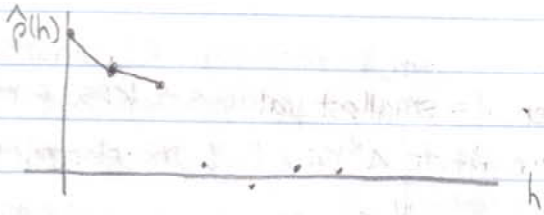
$1=\theta_0 \quad \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4 \quad \cdots \quad \theta_q$

$h=3 \quad 0 \quad 0 \quad 0 \quad 1 \quad \theta_1 \quad \cdots \quad \theta_{q-h}$

$r(h)=\theta_3+\theta_4\theta_1+\dots+\theta_q\theta_{q-3}$

General Formula

$$r(h)=\begin{cases}\sigma_\varepsilon^2\left(\sum_{j=0}^{q-h}\theta_j\theta_{j+h}\right) & h\le q\\ 0 & h>q\end{cases}$$



$\Rightarrow$ use MA(2)

ARMA(p, q)

$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$

ARMA(1,1)

In R: auto.arima – choose p and q that minimize either AIC or BIC. (see lab).

Problem: Many time series are stationary. (In reality, not stationary but log returns are stationary)

Solution: Take differences.

Backwards operator: $BY_t = Y_{t-1}$, $B\mu = \mu$ (μ: constant)

Difference operator: $\Delta = 1 - B$.

$\Delta Y_t = (1-B)Y_t = Y_t - Y_{t-1}$

$\Delta^2 = (1-B)^2 = 1 - 2B + B^2 \Rightarrow \Delta^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$

1) Check if $\{Y_t\}$ is stationary
   - graphical
   - unit-root tests

   If stationary, STOP.

2) If not stat. check if ΔY_t is stationary.
   YES $\rightarrow$ STOP
   NO $\rightarrow$ check if $\Delta^2 Y_t$ is stat.

d = minimum # times Y_t must be differenced to get a stat. series.

ARIMA (p, d, q)

Y_t is ARIMA(p, d, q) if $\Delta^d Y_t$ is ARMA(p, q)

Unit root tests

?? $\rightarrow$ ADF = Augmented Didsey-Fuller

H_0: unit root nonstationary

H_1: stationary

$\therefore$ small p-value $\Rightarrow$ stationary

KPSS test

H_0: stationary

H_1: unit-root nonstationary

∴ large p-value ⇒ stationary

auto.arima : If d is unspecified, then d = smallest value s.t. KPSS p-value > α (0.05)

After d is selected, ARMA models are fit to $\Delta^d Y_t$. p, q are selected to minimize AIC or BIC

Residuals : $\hat{\varepsilon}_t$ are estimates of ε_t

Model fits ⇒ $\{\hat{\varepsilon}_t\}$ are "essentially" white Noise

Model does not fit ⇒ $\{\hat{\varepsilon}\}$ have autocorrelation

Residuals — AR(1)

$(Y_t - \mu) = \phi(Y_{t-1} - \mu) + \varepsilon_t$

$\varepsilon_t = (Y_t - \mu) - \phi(Y_{t-1} - \mu)$

$\hat{\varepsilon}_t = (Y_t - \hat{\mu}) + \hat{\phi}(Y_{t-1} - \hat{\mu})$

$AR(p) = y_t - \mu = \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t$

$\phi(B)(y_t - \mu) = \varepsilon_t$

$\varepsilon_t = (y_t - \mu) - \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu)$

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

$y_t - \mu = \theta(B)\varepsilon_t$

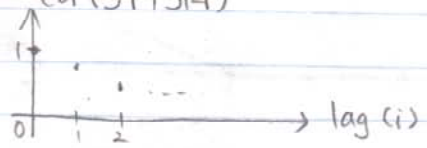
$ARMA(p,q)$: $\phi(B)(y_t - \mu) = \theta(B)\varepsilon_t$

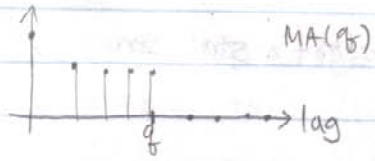
$AR(1)$: $y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t$

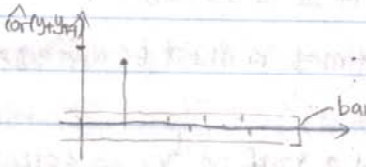
$y_t - \mu = \phi B(y_t - \mu) + \varepsilon_t \Rightarrow (1 - \phi B)(y_t - \mu) = \varepsilon_t$ $\quad (B\mu = \mu,\ By_t = y_{t-1})$

$MA(1)$: $y_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1} = (1 + \theta B)\varepsilon_t$ $\quad (\because \theta\varepsilon_{t-1} = \theta B\varepsilon_t)$

ACF : $Cor(y_t, y_{t-i})$







⇒ $\widehat{Corr}$ drops suddenly

→ consider MA

nonstationary → stationary (ARMA)

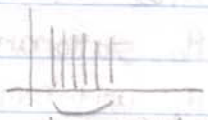
Differencing $y_t - y_{t-1}$, $y_{t-1} - y_{t-2} \cdots$

→ If this still nonstat. another differencing $(y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \cdots$

keep differencing until get stationary

$y_t - y_{t-1} = y_t - By_{t-1} = \underbrace{(1-B)}_{=\Delta}y_t$

$ARIMA(p,d,q)$: $\phi(B)\Delta^d y_t = \theta(B)\varepsilon_t$ → nonstationary b/c of Δ^d



decay very slowly

Conditional LS $\min \Sigma(y_t - \phi y_{t-1})$.

AR(p) → estimate $\phi_1 \cdots \phi_p$

MA(q) → $\theta_1 \cdots \theta_q$

ARMA(p,q)

→ $AIC = -2\log lik + 2p$ ($p = p$ for AR $= q$ for MA $= p+q$ for ARMA)

using AIC → find which one is the best fit

① Estimate parameters then choose model from AIC

② obtain residuals $\hat{\varepsilon}_1 \cdots \hat{\varepsilon}_n$

③ Assess model – ACF based on $\hat{\varepsilon}_1 \cdots \hat{\varepsilon}_n$

Forecast $y_{n+1} = u + \phi_1(y_n - u) + \varepsilon_{n+1}$

current time index n : $\hat{y}_{n+1} = \hat{u} + \hat{\phi}_1(y_n - \hat{u})$ (assume $\varepsilon_{n+1} = 0$)

$y_{n+1} = u + \phi_1(y_n - u) + \phi_2(y_{n-1} - u) + \varepsilon_{n+1}$

$\hat{y}_{n+1} = \hat{u} + \hat{\phi}_1(y_n - \hat{u}) + \hat{\phi}_2(y_{n-1} - u)$

$\hat{y}_{n+2} = \hat{u} + \hat{\phi}_1(\underbrace{y_{n+1}}_{\text{we don't know this value → use } \hat{y}_{n+1}} - \hat{u}) + \hat{\phi}_2(y_n - u)$

$y_{n+1} - \hat{u} = \varepsilon_{n+1} + \hat{\theta}\varepsilon_n = 0 + \hat{\theta}\hat{\varepsilon}_n$

$\hat{Y}_{t+1} = Y_t + \Delta\hat{Y}_t$

band → prediction interval

Construct Prediction Interval

① $\varepsilon \sim N(0, \sigma^2)$

CI $\frac{\hat{y} - y}{sd(\hat{y})} \sim N(0,1)$ $\to 0$

PI $\frac{\hat{y} - y}{sd(\text{forecast error})}$ → $\sqrt{0 + \sigma^2}$ (0 is from ε_{n+1})

$y_{n+1} - \hat{y}_{n+1} = [u + \phi(y_n - u) + \varepsilon_{n+1}] - \{\hat{u} + \hat{\phi}(y_n - \hat{u})\} = u - \hat{u} + (\phi - \hat{\phi})y_n - (\phi u - \hat{\phi}\hat{u}) + \varepsilon_{n+1}$

$y_{n+1} - \hat{y}_{n+1} \approx \varepsilon_{n+1}$ $\to \sigma^2$

$y_{n+2} - \hat{y}_{n+2} = \phi\varepsilon_{n+1} + \varepsilon_{n+2}$ $\to (\phi^2 + 1)\sigma^2$

$y_{n+k} - \hat{y}_{n+k} = \phi^{k-1}\varepsilon_{n+1} + \phi^{k-2}\varepsilon_{n+2} + \cdots + \phi\varepsilon_{n+k-1} + \varepsilon_{n+k}$ → $[(\phi^{k-1})^2 + (\phi^{k-2})^2 + \cdots + \phi^2]\sigma^2$

⇒ Band is getting wider

• Forecast : $\hat{Y}_{n+1} \cdots$

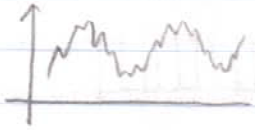
• Forecast limit ① $Y_1 \cdots Y_n$ → $\hat{Y}_{n+1} + \hat{\varepsilon}_{n+1}$, $\hat{Y}_{n+2} + \hat{\varepsilon}_{n+2}$ ...

get this by resampling

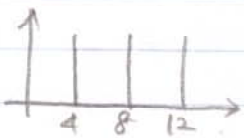
② $Y_1 \cdots Y_n$ → $\hat{Y}_{n+1} + \hat{\varepsilon}^*_{n+1}$ $\hat{Y}_{n+2} + \hat{\varepsilon}^*_{n+2}$

③ B.

Seasonal effects



→ ACF



find a model for each of these

1	2	3	4
Y_1	Y_2	Y_3	Y_4
Y_5	Y_6	Y_7	Y_8

$Y_t - Y_{t-s}$

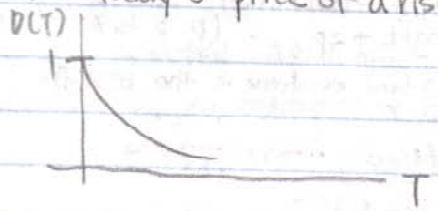
→ get rid of seasonal effect

$\Delta_s = (1 - B^s)Y_t$

→ fit ARMA or ARIMA

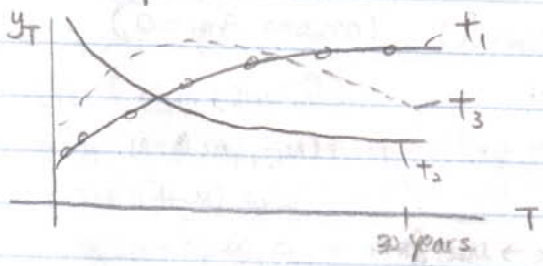
Ex. yields

$D(T)$ = today's price of a risk-free bond that pays \$1 at T



y_T solves $D(T) = \exp[-y_T \cdot T]$

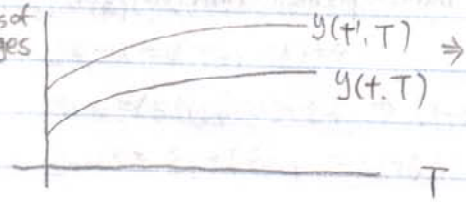
$y_T = \frac{-\log[D(T)]}{T}$



$y(t, T)$ (t: calendar time, T: maturity)

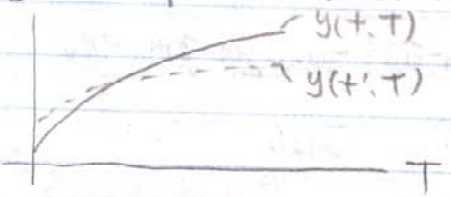
$T_1, \ldots, T_M$

Types of changes

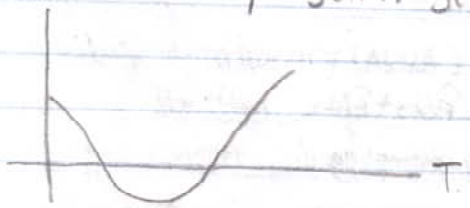


$\Rightarrow$ parallel shift : $y(t, T) - y(t', T)$ = constant in T.

Change-in-slope $y(t, T) - y(t', T) = b_0 + b_1 T$



Change in convexity $y(t, T) - y(t', T)$ is convex or concave.



PCA :

$\underline{Y}_i = \begin{pmatrix} Y_{i,1} \\ \vdots \\ Y_{i,t} \end{pmatrix}$ $\Rightarrow$ yield curves on day i ($i \to t$, $T_1, \ldots, T_d$)

$E(\underline{Y}_i) = 0$

$\mathrm{Cov}(\underline{Y}_i) = \Sigma$

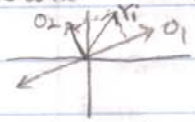
Ω = cov or corr matrix

$\Omega = O\,\mathrm{diag}(\underbrace{\lambda_1, \dots, \lambda_d}_{\text{eigenvalues}})O^T$ where $I = OO^T = O^TO$, $O = (\underbrace{o_1, \dots, o_d}_{\text{eigen vectors}})$, $\lambda_1 > \lambda_2 > \dots > \lambda_d > 0$

o_1 is the solution to max $\alpha^T\Omega\alpha = \mathrm{Var}(\alpha^T Y)$ s.t. $\|\alpha\| = 1$

Moreover, $\lambda_1 = o_1^T\Omega o_1$, o_1 = direction of max. variability

o_1 is called the 1st principal axis



$d=2$

o_2 is the solution to max $\alpha^T\Omega\alpha$ s.t. $\|\alpha\| = 1$ and $\alpha^T o_1 = 0$

$\lambda_2 = o_2^T\Omega o_2$

o_k is the direction of max variation among directons orthogonal to $o_1, \dots, o_{k-1}$

$\lambda_k = o_k^T\Omega o_k$, $o_1, \dots, o_d$ = principal axis

$\{o_k^T Y_1, \dots, o_k^T Y_n\}$ = kth principal component

$\underline{Y} = \begin{pmatrix} Y_1^T \\ \vdots \\ Y_n^T \end{pmatrix}$ $n \times d$ matrix of data

$\underline{S} = (\underline{Y}\underline{O}) = \begin{pmatrix} o_1^T Y_1 & \cdots & o_d^T Y_1 \\ \vdots & & \vdots \\ o_1^T Y_n & \cdots & o_d^T Y_n \end{pmatrix}$ $n \times d$ matrix of principal components

delete from lecture notes

$o_d^T Y_n = Y_n^T o_d$

$\underline{S} = \underline{Y}\underline{O} = \begin{pmatrix} Y_1^T \\ \vdots \\ Y_n^T \end{pmatrix}(o_1 \cdots o_d)$

Principal components are uncorrelated.

$O^TO = I$

$\begin{pmatrix} o_1^T \\ \vdots \\ o_d^T \end{pmatrix}(o_1 \cdots o_d) = I$

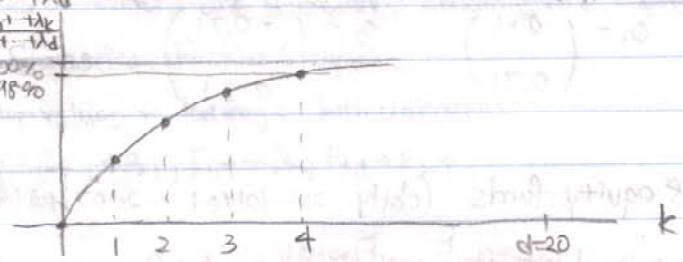
$\Rightarrow o_j^T o_k = 1$ $j = k$

$= 0$ $j \neq k$

Proof: $\mathrm{cov}(o_j^T Y_i, o_k^T Y_i) = E\{(o_j^T Y_i)(Y_i^T o_k)\} = \overset{1\times d}{o_j^T}\, \overbrace{E\{Y_i Y_i^T\}}^{d\times d}\, \overset{d\times 1}{o_k} = o_j^T \Omega o_k$ $\because E[Y_i] = 0$

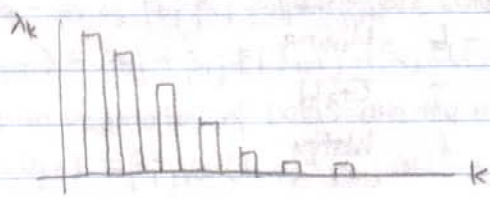
$= o_j^T O\,\mathrm{diag}(\lambda_1, \dots, \lambda_d)O^T o_k = e_j^T \mathrm{diag}(\lambda_1, \dots, \lambda_d)e_k$ if $j \neq k$

$\frac{\lambda_k}{\lambda_1 + \dots + \lambda_d}$ = fraction of total variation in the k-th direction



$\Rightarrow$ 4 principal components "explain" 95% of the variation

(can ignore 16 ...)



$\Rightarrow$ scree plot

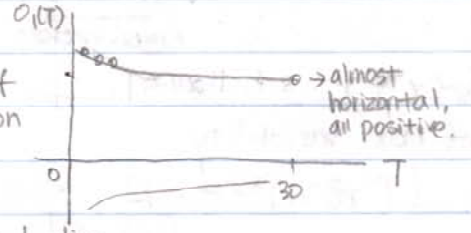
Treasury bonds

$T = 1, 3, 6$ months and $1, 2, 3, 5, 7, 10, 20, 30$ years

$\rightarrow d = 11$

PC 1

62% of variation



$\Rightarrow$ "Roughly" parallel shifts – all yields move in same direction

proportion = 62%

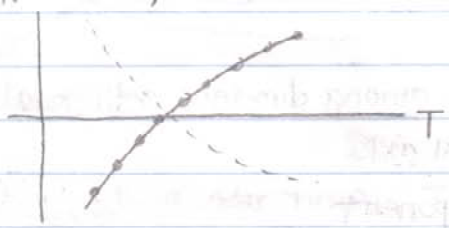
cumul. prop. = 62%

O_1 is d-dim

$O_1 = (O_{1,1} \cdots O_{1,11}) \Rightarrow S_{i,1}$

$Y_i = (Y_{i,1} \cdots Y_{i,11})$

PC2

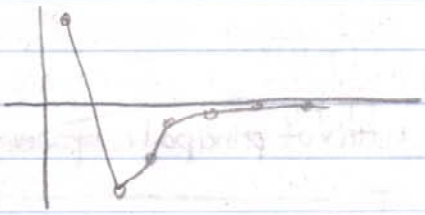


"Roughly" a change in slope

proportion = 25%

cumul. prop = 88%

PC3



Change in convexity

proportion = 7%

cumul. prop. = 95%

Issues

* use yields or changes in yields?

ok if yields are stationary, use if yields are nonstationary.

* corr or cov. matrix

Ex. $Cov = \begin{pmatrix} 10^6 & 900 \\ 900 & 1 \end{pmatrix}$ $O_1 = \begin{pmatrix} 1.0 \\ 0.0009 \end{pmatrix}$ $O_2 = \begin{pmatrix} -0.0009 \\ 1 \end{pmatrix}$

$\lambda_1 = 10^6$ $\lambda_2 = 0.19$ $\rightarrow$ λ_2 is much smaller than λ_1, so we may ignore λ_2 (most variations are from λ_1)

$\Rightarrow \frac{\lambda_2}{\lambda_1}$

$\Sigma = diag(\lambda_1 \lambda_2 \lambda_3 00 \cdots 0)^T$

only takes 3 principal components

$Corr = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$ $O_1 = \begin{pmatrix} 0.71 \\ 0.71 \end{pmatrix}$ $O_2 = \begin{pmatrix} -0.71 \\ 0.71 \end{pmatrix}$

$\lambda_1 = 1.9$ $\lambda_2 = 0.1$

PCA of equity funds – 8 equity funds (daily 2002/01/01 – 2007/05/31)

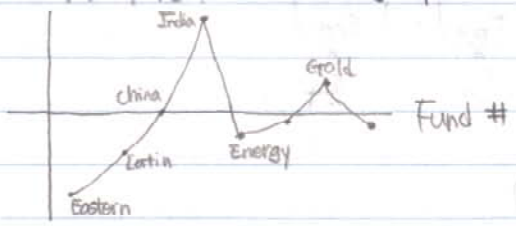
1 Eastern Europe
2 Latin
3 China
4 India

emerging market

5 Energy
6 Mining
7 Gold
8 Water

Need 6 PC's to explain 95% of variance

	Cum. Prop.	
PC1	46.7%	– leads + on all funds.
PC2	63.5%	– negative on gold and mining
PC3	75.1%	– large positive loading on India



CAPM

$R_{j,t}$ = return at time t on jth equity

$R_{M,t}$ = ——— " ——— market index

$\mu_{f,t}$ = risk-free rate

$$\underbrace{(R_{j,t}-\mu_{f,t})}_{\text{excess return}} = \underbrace{\beta_{0,j}}_{\substack{\text{Alpha} \\ H_0:\text{alpha}=0}} + \underbrace{\beta_{1,j}}_{\text{Beta-sensitivity of stock to the market}} \overbrace{(R_{M,t}-\mu_{f,t})}^{\text{market risk factor}} + \underbrace{\overbrace{\varepsilon_{j,t}}^{\text{unique risk factor}}}_{\text{assumed uncorrelated}}$$

$\beta_{0,j}$, $\beta_{1,j}$ doesn't depend on time

$\beta_{0,j}$ and $\beta_{1,j}$ can be estimated by time-series regression ($R_{j,t}-\mu_{f,t} \to Y$, $R_{M,t}-\mu_{f,t} \to X$)

Factor Models

I. Time Series Regression – factors are known – factor loadings are unknown

Ex. CAPM:

factor = excess return on market portfolio

loadings – betas

2 types of factors – macroeconomic ex. interest rate, ΔGDP

– fundamental ex. return on market index, ex. return on a hedge portfolio

II. Cross Sectional factor models

Values of factors are unknown. Loadings are known

III Statistical Factor Analysis

Factor values + loadings both unknown

$R_{j,t} = \beta_{0,j} + \beta_{1,j} F_{1,t} + \beta_{2,j} F_{2,t} + \varepsilon_{j,t}$

time

??

time-series β_{0j}, $\beta_{1,j}$, $\beta_{2,j}$ and $\{\varepsilon_{j,t}\}$ are estimated by regression one factor at a time

$E(R_{j,t}) = \beta_{0j} + \beta_{1j} E(F_{1,t}) + \beta_{2j} E(F_{2,t})$ Assume stationarity (R, F are stationary)

Plug-in estimates of betas and the means of the factors

$Var(R_{j,t}) = \beta_{1,j}^2 Var(F_{1,t}) + \beta_{2,j}^2 Var(F_{2,t}) + 2\beta_{1,j}\beta_{2,j} Cov(F_{1,t}, F_{2,t}) + \sigma_{\varepsilon,j}^2$

estimated by residual mean square

copula ~ Factor model

$j \neq j'$

$Cov(R_{j,t}, R_{j',t}) = \beta_{1,j}\beta_{1,j'} Var(F_{1,t}) + \beta_{2,j}\beta_{2,j'} Var(F_{2,t}) + (\beta_{1,j}\beta_{2,j'} + \beta_{1,j'}\beta_{2,j}) Cov(F_{1,t}, F_{2,t}) + 0$

Assuming $cor(\varepsilon_{j,t}, \varepsilon_{j',t}) = 0$

$$\underbrace{\begin{pmatrix} R_{1,t} \\ \vdots \\ R_{n,t} \end{pmatrix}}_{\underline{R}_t} = \underbrace{\begin{pmatrix} \beta_{0,1} \\ \vdots \\ \beta_{0,n} \end{pmatrix}}_{\underline{\beta}_0} + \underbrace{\begin{pmatrix} \beta_{1,1} & \cdots & \beta_{p,1} \\ \vdots & & \vdots \\ \beta_{1,n} & & \beta_{p,n} \end{pmatrix}}_{\underline{\beta}^T} \underbrace{\begin{pmatrix} F_{1,t} \\ \vdots \\ F_{p,t} \end{pmatrix}}_{\underline{F}_t} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{n,t} \end{pmatrix}}_{\underline{\varepsilon}_t}$$

$\underline{R}_t = \underline{\beta}_0 + \beta^T \underline{F}_t + \underline{\varepsilon}_t$

$Cov(\underline{\varepsilon}_t) = \Sigma_\varepsilon = diag(\sigma^2_{\varepsilon,1}, \ldots, \sigma^2_{\varepsilon,n})$

$Cov(\underline{F}_t) = \Sigma_F$

$$Cov(\underline{R}) = \underbrace{\Sigma_R}_{\frac{n(n+1)}{2}\text{ entries}} = \underbrace{\underline{\beta}^T}_{np\ \text{big}} \underbrace{\Sigma_F}_{p(p+1)} \underbrace{\underline{\beta}}_{np} + \underbrace{\Sigma_\varepsilon}_{n}$$

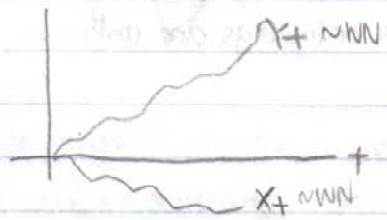
small

$\underline{\beta}$ – estimated by regression.

Σ_ε – estimated by the residual mean squares in regression

Σ_F – estimated by the sample covariance matrix

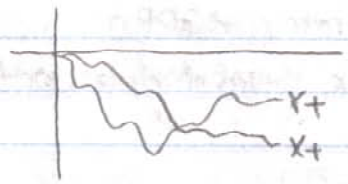
Danger when doing time series regressions: It is assumed that the variables are stationary.



$\Rightarrow Y_t = \beta_0 + \beta_1 Y_t + \varepsilon_t$ → nonstationary

$\hat{\beta}_1$ is negative with a small p-value

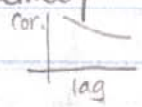
Y_t and X_t independent



$\Rightarrow \hat{\beta}_1 > 0$ and has small p-value

Spurious correlation or regression

Remedy – look at the ACF of the residuals – should see little correlation



residual $= \hat{\varepsilon}_t = Y_t - (\hat{\beta}_0 + \hat{\beta}_1 X_t)$

$Y_t = W_t + \varepsilon_t$, $X_t = \frac{1}{2} W_t + \varepsilon_t$ $\Rightarrow (Y_t - 2X_t) = (\varepsilon_t - 2\varepsilon_t)$ → White Noise ($\because \varepsilon_t, \varepsilon_t'$ uncorrelated)

random walk iid cointegrated white Noise