Lectures on Unit Roots, Cointegration & Nonstationarity[‡]

by

Peter C. B. Phillips

Cowles Foundation, Yale University

April 1995

‡ Lecture notes for "Time Series Econometrics II" Econ 557b, Spring 1995

Review of Time Series I

Empirical features of econ. time series

serial dependence (temporal) – characterizing its form – exploring its properties $\gamma_k = E(X_t X_{t-k})$ $f(\lambda) = \frac{1}{2\pi}\sum \gamma_k e^{-ik\lambda}$

– parametric time series models $\gamma_k = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda$
AR, ARMA, ARMAX & vector analogues

– nonparametric
$E(y_t \mid y_{t-1}, y_{t-2} \cdots)$

joint dependence – structural modelling (SEM) $By_t + Cx_t = u_t$
– edifice of SE theory
– SEM's & VAR's $A(L)u_t = \varepsilon_t$ + ↑
– co-dependence / comovement / cointegration

nonstationarity – nature of trends < deterministic $\sum_{i=0}^{k} a_i t^i$, $\Delta^k X_t$ = const; stochastic $\Delta^k X_t = u_t$

– dependence between trends & stat. components

$$y_t = T_t + R_t$$
correlated

– { cotrending series, common trends } – sources of trend common = technology shock, demographics

– cointegration
– spurious reg ^{n}

volatility – conditional heterogeneity $\sigma_t^2 = E(u_t^2 \mid \mathcal{F}_{t-1})$ function of past history

– financial / stock prices

{ARCH, GARCH, Stochastic volatility}

outliers & non-Gaussianity
– er data
– stock prices $E|u_t|^r = \infty$ some r
– income / wealth distn's & changes over time

Methods of Inference

data + input (model + stoch. hyp.) ⟶ stat. instruments of estn, inference, diagnostic testing, model determination & forecasting

Classical: data X^n ⟶ statistic $\hat{\theta}_n$
dgp $pdf(X^n; \theta)$ ⟶ $pdf(\hat{\theta}_n)$ or asymp. equiv
$pdf(X_{n+1} \mid \theta, \mathcal{F}_n)$
$pdf(X_{n+1} \mid \hat{\theta}_n, \mathcal{F}_n)$ classical predictive density

Bayesian: jt den. $pdf(\theta, X^n)$ ⟶ $pdf(\theta \mid X^n)$ posterior density
$\pi(\theta)\, pdf(X^n \mid \theta)$
$pdf(X^n)$ { data density, dgp }
$pdf(X_{n+1} \mid \mathcal{F}_n)$ predictive density.

Model format

- data = signal + noise   useful framework for modelling a physical system over time

$$y_t = s_t + u_t$$

e.g. $s_t = a(L)y_t$ AR

$b(L)u_t$ MA

$a(L)y_t + b(L)u_t$ ARMA

$a(L)y_t + c(L)x_t + b(L)u_t$ ARMAX

- signal/noise ratio (SNR) $= \dfrac{var(s_t)}{var(u_t)}$

e.g. $y_t = \theta y_{t-1} + u_t$ $SNR = \theta^2 \dfrac{var(y_{t-1})}{var(u_t)}$

$|\theta| < 1$ $= \dfrac{\theta^2 \sigma^2/1-\theta^2}{\sigma^2}$

$= \theta^2/(1-\theta^2)$

$\theta = 1$ $SNR = \dfrac{t\sigma^2}{\sigma^2} = t \to \infty$

(signal is stronger when $\exists$ unit root)

$y_t = \theta u_{t-1} + u_t$ $SNR = \dfrac{\theta^2\sigma^2}{\sigma^2} = \theta^2$

$y_t = \theta + u_t$ $SNR = \dfrac{\theta^2 \frac{1}{n}\Sigma 1}{\sigma^2} = \dfrac{\theta^2}{\sigma^2}$

$y_t = \theta t + u_t$ $SNR = \dfrac{\theta^2 \frac{1}{n}\Sigma t^2}{\sigma^2} = \dfrac{\theta^2 \frac{n(n+1)}{2n}}{\sigma^2}$

sample variance of regressor $= \dfrac{\theta^2 n}{2\sigma^2} \to \infty$

signal stronger with linear trend

- affects rates of convergence

$y_t = \theta + u_t$ $\hat\theta = \bar y$, $\sqrt{n}(\hat\theta - \theta) = \dfrac{\Sigma u_t}{\sqrt{n}} \to N(0, \omega^2)$

$y_t = \theta t + u_t$ $\hat\theta - \theta = \dfrac{\Sigma t u_t}{\Sigma t^2}$, $n^{3/2}(\hat\theta - \theta) = \dfrac{\frac{1}{n^{3/2}}\Sigma t u_t}{\frac{1}{n^3}\Sigma t^2}$

$\to N(0, \omega^2 3)$

as $\frac{1}{n^3}\Sigma t^3 \to \frac{1}{3}$

- what if regressors in

$$y_t = x_t'\beta + u_t$$

have multiple signals? x_t < stationary / trends / stochastic trend / unit root   components

- need to rotate regressor space to get asymptotics + clarify signal

$$J = [J_1, J_2, J_3] \in O(k)$$

$$J_1' x_t = x_{1t} \quad \text{stationary}$$

$$J_2' x_t = x_{2t} \quad \text{has unit roots}$$

$$J_3' x_t = x_{3t} \quad \text{deterministic trends}$$

- reform model with rotated regressors

$$y_t = x_t' J J' \beta + u_t = x_{1t}'\beta_1 + x_{2t}'\beta_2 + x_{3t}'\beta_3 + u_t$$

$$J'X'XJ = \begin{pmatrix} X_1'X_1 & X_1'X_2 & X_1'X_3 \\ & X_2'X_2 & X_2'X_3 \\ & & X_3'X_3 \end{pmatrix}$$

$$\beta = J \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = J_1\beta_1 + J_2\beta_2 + J_3\beta_3$$

$$\hat\beta = J \begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \\ \hat\beta_3 \end{pmatrix}$$

$$\hat\beta - \beta = J_1(\hat\beta_1 - \beta_1) + J_2(\hat\beta_2 - \beta_2) + J_3(\hat\beta_3 - \beta_3)$$

$$\sqrt{n}(\hat\beta - \beta) = J_1 \sqrt{n}(\hat\beta_1 - \beta_1) + o_p(1) + o_p(1)$$

faster rates of convergence for unit roots & deterministic trends

$$\rightarrow J_1 N(0, \sigma^2 M_1)$$

$\in R(J_1)$ & is singular (confined to a k_1 dim subspace of $\mathbb{R}^k$

- our purpose to explore this.

Measure preserving (mp) maps & Ergodic theory

prob space $(\Omega, \mathcal{F}, P)$, $\{X_t\}_{-\infty}^{\infty}$ $\quad X_t : \Omega \to \mathbb{R}$

realizations $\quad x = (\cdots, x_{-1}, x_0, x_1, \cdots) \in \mathbb{R}_\infty$

coordinate representation

$$(\mathbb{R}_\infty, \mathcal{B}_\infty, P) \quad \{X_t\}_{-\infty}^{\infty} \qquad X_t(x) = x_t \in \mathbb{R}$$

picks off t'th element

$\mathcal{B}_\infty = \mathcal{B}(\mathbb{R}_\infty)$ = Borel σ field on $\mathbb{R}_\infty$

generated by cylinder sets like

$\overset{n}{\underset{-\infty}{\times}} \mathbb{R} \overset{m}{\underset{n+1}{\times}} B_i \overset{\infty}{\underset{m+1}{\times}} \mathbb{R}$

B_i = Borel set in $\mathbb{R}$ generated by

$(a,b]$ intervals $\quad P(a \le x \le b)$ [giving meaning to]

temporal displacements & shifts

$S : \Omega \to \Omega$ $\qquad x = (\cdots, x_{-1}, x_0, x_1, x_2, \cdots)$

$\qquad Sx = (\cdots, x_0, x_1, x_2, x_3, \cdots)$

$X : \Omega \to \mathbb{R}$ measurable function (i.e. r.v.)

$X^{-1}B \in \mathcal{F} \qquad B \in \mathbb{B}$

X takes mble sets (in $\mathcal{F}$) into mble sets in $\mathbb{R}$

$P(a \le x \le b) = P(X^{-1}[a,b])$

$X_1(x) = X(x) = x_1$

$X_2(x) = X(Sx) = x_2$

↑shifts points of space rather than r.v.

$U_S X(x) = X(Sx) = x_2 \qquad X_2 = U_S X$

$U_S^2 X(x) = X(S^2 x) = x_3 \qquad X_3 = U_S^2 X$

$\vdots$

$X_{n+1} = U_S^n X$

time series $\{X_n\}$ defined by r.v. X and map S

i.e. $X_n = U_S^{n-1} X$

$\{X_n\} \equiv$ strictly stationary time series if S m.p.

$P(E) = P(S^{-h}E) \ \forall h$ (preserves measure)

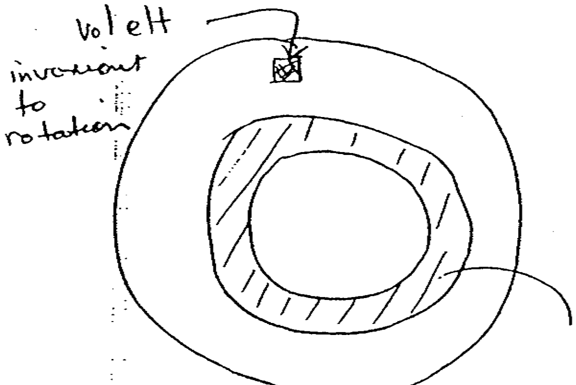
<u>Example</u> $(K, \mathcal{F}, P)$ $\qquad K = \{z \in \mathbb{C} : |z| \le 1\}$

P = normalized area $= m/\pi \qquad m$ = Lebesgue measure on $\mathbb{R}^2$

$S : K \to K$ rotation

$Sz = az$

$a = e^{i\theta}$



<u>Ergodicity</u>

- $S : \Omega \to \Omega$ is ergodic if $P(F) = 0, 1$ for all invariant events F

  (i.e. all F s.th $F = S^{-1}F$)

  i.e. all invariant events are ignorable or certain

- F in example above is invariant but $0 < P(F) < 1$ so S is not ergodic

- $X_n \; (= U_S^{n-1} X)$ is strictly stationary time series for S m.p.

  is ergodic if S is ergodic

<u>Example</u> (non ergodic, str. stat time series)

$X_t = U_t + Z \qquad \begin{matrix} U_t \equiv \text{iid uniform } [0,1] \\ Z \equiv N(0,1) \end{matrix} \Big\rangle$ indep

$F := \{ \ldots X_{-1} < 0, X_0 < 0, X_1 < 0 \ldots \}$ iff $Z < -1$

$P(F) = P(Z < -1) = \Phi(-1)$ & $0 < P(F) < 1$

so X_t is non ergodic. F is invariant under S but does not affect outcome of Z some events NOT experienced

- Ergodic theorem (SLLN)

$\{X_t\}$ str. stat & ergodic time series

$E|X_t| < \infty$

$$P\left[\lim_{n\to\infty} \frac{1}{n}\Sigma_1^n X_t = E(X_t)\right] = 1$$

If $\{X_t\}$ is str. stat but not ergodic

$$P\left[\lim_{n\to\infty} \frac{1}{n}\Sigma_1^n X_t = E(X_t \mid J)\right] = 1$$

where J = σ field of all invariant events

- Kolmogorov SLLN

$X_t \equiv$ iid , $E|X_t| < \infty$

$\frac{1}{n}\Sigma_1^n X_t \to E(X_t)$ a.s.

- M'ble functions of stat. time seri

$Y_t = Y_t(X)$, $X = \{X_t\}_{-\infty}^{\infty}$ str. stat & ergodic

then Y_t is str. stat & ergodic

$\frac{1}{n}\Sigma_1^n Y_t \to_{a.s} E(Y_t)$ if $E|Y_t| < \infty$

e.g. (i) $Y_t = X_t X_{t+k}$ is a mble function of X

$\frac{1}{n}\Sigma_1^n X_t X_{t+k} \to_{a.s} E(X_t X_{t+k}) = \gamma_k$ autocovari

(ii) $Y_t = \Sigma_{j=-\infty}^{\infty} a_j X_{t+j}$, $\Sigma_{-\infty}^{\infty}|a_j| < \infty$

Can show this series converges a.s. & hence is str. stat + ergodic

$\frac{1}{n}\Sigma_1^n Y_t \to_{a.s} E(Y_t) = \left(\Sigma_{j=-\infty}^{\infty} a_j\right) E(X_t)$.

(iii) $y_t = \theta y_{t-1} + u_t = \sum_0^\infty \theta^j u_{t-j}$ is str. stat & ergodic when $|\theta| < 1$

$$\frac{1}{n}\sum_1^n y_t \xrightarrow[a.s]{} E(y_t) = \left(\sum_0^\infty \theta^j\right) E(u_t) = 0$$

$$\hat{\theta} - \theta = \sum y_{t-1} u_t \Big/ \sum y_{t-1}^2 = \frac{1}{n}\sum_1^n y_{t-1} u_t \Big/ \frac{1}{n}\sum y_{t-1}^2$$

$$\xrightarrow[a.s]{} E(y_{t-1} u_t) \Big/ E(y_{t-1}^2) = 0$$

ie $\hat{\theta} \xrightarrow[a.s]{} \theta$ if u_t is $iid(0, \sigma^2)$

$$\frac{1}{n}\sum y_t^2 \xrightarrow[a.s]{} E(y_t^2) = \sigma^2 / 1 - \theta^2$$

(iv) non ergodic case: $X_t = u_t + Z$

$$\frac{1}{n}\sum_1^n X_t = \frac{1}{n}\sum_1^n u_t + Z \xrightarrow[a.s]{} E(u_t) + Z$$

$$= \tfrac{1}{2} + Z \quad (\text{r.v.})$$

$$= E(X_t | \mathcal{I})$$

invariant field spanned by Z

Weak Dependence & Mixing

Mixing: if $S: \Omega \to \Omega$ is m.p. on $(\Omega, \mathcal{F}, P)$ then S is mixing (mixes the points of Ω) if

$$P(F \cap S^{-n} G) \to P(F) P(G)$$
$$\forall F, G \in \mathcal{F}$$

Strong Mixing: $\alpha_m = \sup_j \alpha(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+m}^\infty) \searrow 0$

$$\alpha(F_1, F_2) = \sup_{A \in F_1, B \in F_2} |P(A \cap B) - P(A) P(B)|$$

diff. betw. jt. prob & product of marginals

φ-mixing $\quad \varphi_m \searrow 0 \quad |P(B|A) - P(B)|$

ψ mixing $\quad \psi_m \searrow 0 \quad \left| \frac{P(A \cap B)}{P(A)P(B)} - 1 \right|$

$$\psi_m \geq \varphi_m \geq \alpha_m$$

<u>functions of mixing sequences</u>

$$Y_t = h(X_{t-\ell}, X_{t-\ell+1}, \dots, X_{t+k}) \quad \ell, k \text{ fixed \& finite}$$

mixing (α, φ, ψ) and at same rate

<u>ARMA models</u> $\quad a(L)\, y_t = b(L)\, u_t$

α-mixing if stable + u_t has cts distribution (dens wrt Leb. meas.)

<u>SLLN's for mixing sequences</u> (McLeish, 1975 AP)

1. $E(X_t) = 0$
2. X_t α-mixing of size $-r/(r-1)$ $\quad (r>1)$

i.e. $\alpha_m = O\left(\frac{1}{m L_m}\right)^{r/r-1}$

L_m = slowly varying with $\sum \frac{1}{m L_m} < \infty$ (eg. $(\ln m)^{1+\varepsilon}$)

3. $\sup_t E|X_t|^{r+\delta} < \infty$ some $\delta > 0$

then

$$\frac{1}{n} \sum_1^n X_t \xrightarrow[a.s]{} 0$$

<u>trade offs</u> $\quad$ moment conditions vs weak dependence

$r \uparrow$ more moments, less outliers $\quad$ more dependence allowed

$r \searrow 1$ more outliers, less moments $\quad$ less dependence allowed (limit $r=1$ i.i.d)

Linear Process Theory & BN decomposition

(Phillips & Solo, 1992 AS ; B–N JME 1982)

linear process: $X_t = C(L)\,\varepsilon_t \qquad \varepsilon_t \equiv$ iid / inid / mds

$$C(L) = \sum_{j=0}^{\infty} c_j L^j$$

BN decomposition

$$C(L) = C(1) + \tilde{C}(L)(L-1)$$

$$\tilde{C}(L) = \sum_{j=0}^{\infty} \tilde{c}_j L^j \, , \quad \tilde{c}_j = \sum_{j+1}^{\infty} c_s \quad \text{tail sum}$$

Summability conditions

1/2-summ.: $\sum j^{1/2} |c_j| < \infty \implies \sum_0^{\infty} \tilde{c}_j^{\,2} < \infty$

1-summ.: $\sum j\, |c_j| < \infty \implies \sum |\tilde{c}_j| < \infty$

(e.g.
$$\sum_0^{\infty} |\tilde{c}_j| = \sum_0^{\infty} \Big| \sum_{j+1} c_s \Big| \le \sum_0^{\infty} \sum_{j+1} |c_s| \qquad s \ge j+1$$
$$= \sum_{s=1}^{\infty} |c_s| \sum_{j=0}^{s-1} 1$$
$$= \sum_{s=1}^{\infty} s\,|c_s|$$
$$= \sum_{s=0}^{\infty} s\,|c_s| \;)$$

SLLN's: means

(i) $X_t = C(L)\,\varepsilon_t = C(1)\,\varepsilon_t + \tilde{C}(L)(L-1)\,\varepsilon_t$

(ii) $\sup_t E|\varepsilon_t|^{1+\delta} < \infty$

$$= C(1)\,\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t$$

$$\tilde{\varepsilon}_t = \tilde{C}(L)\,\varepsilon_t = \sum_0^{\infty} \tilde{c}_j \varepsilon_{t-j}$$

$$\bar{X} = \frac{1}{n}\sum_1^n x_t = C(1)\,\underbrace{\frac{1}{n}\sum_1^n \varepsilon_t}_{\xrightarrow{a.s.}\,0} + \underbrace{\frac{1}{n}\left(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_n\right)}_{\xrightarrow{a.s.}\,0}$$

by SLLN for iid, inid, mds → MG cgce Theorem

by Borel Cantelli Lemma or MG cgce Theorem

SLLN's : variances $X_t = C(L)\varepsilon_t = \sum_0^\infty c_j \varepsilon_{t-j}$

$$X_t^2 = (C(L)\varepsilon_t)^2 = \sum_0^\infty c_j^2 \varepsilon_{t-j}^2 + 2\sum_{k>j} c_j c_k \varepsilon_{t-j}\varepsilon_{t-k}$$

$$= \sum_0^\infty c_j^2 \varepsilon_{t-j}^2 + 2\sum_{r=1}^\infty \sum_{j=0}^\infty c_j c_{j+r} \varepsilon_{t-j}\varepsilon_{t-j-r} \qquad k = j+r$$

$$= f_0(L)\varepsilon_t^2 + 2\sum_{r=1}^\infty f_r(L)\varepsilon_t \varepsilon_{t-r}$$

$$f_r(L) = \sum_{j=0}^\infty c_j c_{j+r} L^j$$

$$= f_0(1)\varepsilon_t^2 + 2\varepsilon_t \varepsilon_{t-1}^f \qquad \varepsilon_{t-1}^f = \sum_{r=1}^\infty f_r(1)\varepsilon_{t-r}$$

+ differenced terms that are negligible when averaged

$$\frac{1}{n}\sum_1^n X_t^2 = f_0(1)\underbrace{\frac{1}{n}\sum_1^n \varepsilon_t^2}_{\xrightarrow[a.s]{} \sigma_\varepsilon^2} + 2\underbrace{\frac{1}{n}\sum_1^n \varepsilon_t \varepsilon_{t-1}^f}_{\xrightarrow[a.s]{} 0} + \text{neglig}$$

$$\xrightarrow[a.s]{} f_0(1)\sigma_\varepsilon^2 = \sigma_\varepsilon^2 \sum_0^\infty c_j^2$$

Central Limit Theory (CLT)

CLT : means $X_t = C(L)\varepsilon_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t$

$$\frac{1}{\sqrt{n}}\sum_1^n X_t = C(1)\frac{1}{\sqrt{n}}\sum_1^n \varepsilon_t + \underbrace{\frac{1}{\sqrt{n}}(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_n)}_{\xrightarrow[p]{} 0}$$

$$\xrightarrow[d]{} C(1)\, N(0, \sigma_\varepsilon^2) \quad \text{as} \quad \mathrm{var}\left(\frac{\tilde{\varepsilon}_n}{\sqrt{n}}\right) = \frac{1}{n}\mathrm{var}(\tilde{\varepsilon}_n)$$

directly by Lindeberg Levy theorem $\qquad = \frac{1}{n}\sum_0^\infty \tilde{c}_j^2 \sigma_\varepsilon^2 \to 0$

$$\equiv N(0, \sigma_\varepsilon^2 C(1)^2)$$

$$\equiv N(0, 2\pi f_x(0))$$

where $f_x(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^\infty E(X_t X_{t-h}) e^{-ih\lambda}$ = spectrum of X_t

$= C(e^{i\lambda})\,(\sigma^2)\,C(e^{i\lambda})^*$ transfer function form

CLT : variances

$$\varepsilon_t^f = \sum_{r=1}^{\infty} f_r(L)\,\varepsilon_t$$

$$X_t^2 = f_0(1)\,\varepsilon_t^2 + 2\varepsilon_t\,\varepsilon_{t-1}^f + \text{small (from above)}$$

$$E(X_t^2) = \sigma_\varepsilon^2 f_0(1) = \gamma_0$$

$$\frac{1}{\sqrt{n}}\sum_1^n (X_t^2 - \gamma_0) = f_0(1)\frac{1}{\sqrt{n}}\sum_1^n(\varepsilon_t^2 - \sigma_\varepsilon^2) + \frac{2}{\sqrt{n}}\sum_1^n \varepsilon_t\,\varepsilon_{t-1}^f + o_p(1)$$

$$\xrightarrow{d} N(0, v_0)$$

$$v_0 = f_0(1)^2(\mu_4 - \sigma_\varepsilon^4) + 4\,\sigma_\varepsilon^2\,\sigma_f^2$$

$$\sigma_f^2 = \operatorname{var}(\varepsilon_{t-1}^f) = \sum_{r=1}^{\infty} f_r(1)^2\,\sigma_\varepsilon^2$$

$$\mu_4 = E(\varepsilon_t^4)$$

CLT : covariances

$$X_t X_{t+h} = \left(\sum c_j\,\varepsilon_{t-j}\right)\left(\sum c_k\,\varepsilon_{t+h-k}\right)$$

$$= \sum c_j c_{j+h}\,\varepsilon_{t-j}^2 + \sum_j \sum_{k \neq h+j} c_j c_k\,\varepsilon_{t-j}\,\varepsilon_{t+h-k}$$

$$= f_h(L)\,\varepsilon_t^2 + \sum_{r \neq 0} f_{h+r}(L)\,\varepsilon_t\,\varepsilon_{t-r}$$

SLLN $\quad \frac{1}{n}\sum X_t X_{t+h} \xrightarrow{a.s} f_h(1)\,\sigma^2 = \gamma_h$

CLT $\quad \frac{1}{\sqrt{n}}\sum (X_t X_{t+h} - \gamma_h) = f_h(1)\frac{1}{\sqrt{n}}\sum_1^n(\varepsilon_t^2 - \sigma_\varepsilon^2) + \sum_{r=1}^{\infty}\left(f_{h+r}(1) + f_{h-r}(1)\right)\frac{1}{\sqrt{n}}\sum_1^n \varepsilon_t\,\varepsilon_{t-r}$

$$\to N(0, v_h)$$

$$v_h = f_h(1)^2(\mu_4 - \sigma_\varepsilon^2) + \sum_{r=1}^{\infty}\left(f_{h+r}(1) + f_{h-r}(1)\right)^2 \sigma_\varepsilon^4$$

Hilbert space, Projection geometry & Wold decomposition

Hilbert space

$$\mathcal{H} = L_2(\Omega, \mathcal{F}, P) = \{X \mid \int X^2 dP < \infty\}$$

inner product $(X, Y) = \int XY dP = E(XY)$

Projection

P_M is projection on $\mathcal{H}$ if $P_M^2 = P_M$

orthogonal projection if P_M self adjoint

i.e. $(P_M X, Y) = (X, P_M Y)$

$\forall X, Y \in \mathcal{H}$.

Decomposition

$$M = \mathcal{R}(P_M) \subset \mathcal{H}, \quad \mathcal{H} = M \oplus M^\perp$$

$$X = \underbrace{P_M X}_{\in M} + \underbrace{(1 - P_M) X}_{\in M^\perp}$$

$\perp$

$$\begin{aligned}(P_M X, (1-P_M)X) &= \int P_M X (1-P_M) X dP \\ &= \int X P_M (1-P_M) X dP \\ &= 0\end{aligned}$$

as P_M self adjoint

Orthogonal decomposition minimises distance (MSE) of X from M. Let $Y \in M$

$$\begin{aligned}\|X-Y\|_2^2 = \|X - P_M X + P_M X - Y\|_2^2 &= \|X - P_M X\|_2^2 + \|P_M X - Y\|_2^2 \\ &\geq \|X - P_M X\|_2^2\end{aligned}$$

equality when $Y = P_M X$

Conditional Expectations as L_2 Projections

Conditional Expectation

$(\Omega, \mathcal{F}, P)$, $\mathcal{G} \subset \mathcal{F}$ (sub σ-field of $\mathcal{F}$)

$E(\cdot \mid \mathcal{G}) : L_1(\Omega, \mathcal{F}, P) \to L_1(\Omega, \mathcal{G}, P)$

defined by $X \to E(X \mid \mathcal{G})$

defining property $\int_G X \, dP = \int_G E(X \mid \mathcal{G}) \, dP$

$\forall G \in \mathcal{G}$

L_2 Projection

$E(\cdot \mid \mathcal{G}) : L_2(\Omega, \mathcal{F}, P) \to L_2(\Omega, \mathcal{G}, P)$

is given by orthogonal projection

$$E(X \mid \mathcal{G}) = P_{\mathcal{G}} X \quad \text{a.s. } (P)$$

Note:

$$E\left(E(X \mid \mathcal{G}) \mid \mathcal{G}\right) = E(X \mid \mathcal{G}) \quad \text{a.s.}$$

redundant conditioning

so $E(\cdot \mid \mathcal{G})$ is an idempotent operator

Prediction

given $(\Omega, \mathcal{F}, P)$, $X_t \in L_2(\Omega, \mathcal{F}, P)$

$$\mathcal{M}_n = \left\{ \Sigma_0^\infty c_j X_{n-j} : \Sigma c_j^2 < \infty \right\}$$

linear manifold of $L_2(\Omega, \mathcal{F}, P)$ spanned by $(X_t)_{-\infty}^{n}$

prediction problem:

approximate X_{n+1} using $\mathcal{M}_n$ 1-step prediction

X_{n+h} using $\mathcal{M}_n$ multi-step prediction

<u>Solution</u>

$$\hat{X}_{n+1} = P_{M_n} X_{n+1}$$

<u>prediction error</u>

$$\varepsilon_{n+1} = X_{n+1} - \hat{X}_{n+1} = (1 - P_{M_n}) X_{n+1}$$

$$\in M_n^{\perp}$$

minimises MSE prediction

i.e.

$$P_{M_n} X_{n+1} = \min_{Y \in M_n} \| X_{n+1} - Y \|^2$$

<u>prediction error variance</u>

$$\sigma^2 = \| X_{n+1} - P_{M_n} X_{n+1} \|^2 = \| \varepsilon_{n+1} \|^2$$

$$= \int \varepsilon_{n+1}^2 \, dP$$

$$= E(\varepsilon_{n+1}^2)$$

<u>Purely deterministic process</u>

if $\sigma^2 = 0$, X_n is said to be purely deterministic

i.e. $X_n = P_{M_{n-1}} X_n$

$= P_{M_{n-2}} X_n \ldots = P_{M_{-\infty}} X_n$

a.s.

i.e. X_n invariant to n

e.g. $X_t = \underbrace{U_t}_{\text{iid Unif}[0,1]} + \underbrace{Z}_{N(0,1)}$

indep

Z is purely deterministic process

$= P_{M_{-\infty}} X_t$

Wold decomposition

If $(X_n)_{-\infty}^{\infty} \in L_2(\Omega, \mathcal{F}, P)$ covariance stationary time series

Then

$$X_n = \Sigma_0^\infty c_j \, \varepsilon_{n-j} + v_n = u_n + v_n$$

with

(i) $\varepsilon_n \equiv WN(0, \sigma^2)$ orthogonal sequence

(ii) $E(\varepsilon_n v_m) = 0 \quad \forall n, m$

ie. $(\varepsilon_n) \perp (v_m)$

(iii) v_n purely deterministic

Notes

let $\varepsilon_n = X_n - P_{M_{n-1}} X_n \in M_{n-1}^{\perp}$

prediction error by constr^n

let $E_n = \{ \Sigma_0^\infty c_j \, \varepsilon_{n-j} : \Sigma c_j^2 < \infty \}$

$$X_n = P_{E_n} X_n + (1 - P_{E_n}) X_n$$

$$\underbrace{u_n \qquad\qquad v_n}_{\perp}$$

properties of v_n:

$$\left. \begin{array}{l} v_n \in M_{n-1} \oplus [\varepsilon_n] \\ v_n \perp [\varepsilon_n] \end{array} \right\} \Rightarrow v_n \in M_{n-1}$$

$$\cdots\cdots\cdots\cdots \Rightarrow v_n \in M_{-\infty}$$

$v_n = P_{M_{-\infty}} X_n$ invariant

ie. v_n purely deterministic – its prediction error from own past is zero.

Optimal linear predictors

$$X_{n+1} = \sum_{j=0}^{\infty} c_j \varepsilon_{n+1-j} + v_{n+1}$$

1-step

$$\hat{X}_{n+1} = P_{M_n} X_{n+1} = \sum_{j=1}^{\infty} c_j \varepsilon_{n+1-j} + v_{n+1}$$

$v_{n+1} \in M_{-\infty} \subset M_n \quad \forall n$

h-step

$$\hat{X}_{n+h,h} = P_{M_n} X_{n+h} = \sum_{j=h}^{\infty} c_j \varepsilon_{n+h-j} + v_{n+h}$$

prediction error:

$$\begin{aligned} \varepsilon_{n+h,h} &= X_{n+h} - \hat{X}_{n+h,h} \\ &= \sum_{j=0}^{h-1} c_j \varepsilon_{n+h-j} \end{aligned}$$

note serial dependence for $h > 1$

prediction error variance

$$\sigma_h^2 = E(\varepsilon_{n+h,h}^2) = \left(\sum_{j=0}^{h-1} c_j^2\right) \sigma^2$$

Martingales (MG's)

MG: $Y_n \in L_1(\Omega, \mathcal{F}, P)$ $\qquad \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \cdots$ filtration

$$E(Y_{n+1} \mid \mathcal{F}_n) = Y_n$$

MDS (martingale difference sequence)

$$u_n = Y_n - E(Y_n \mid \mathcal{F}_{n-1}) \; ; \quad E(u_n \mid \mathcal{F}_{n-1}) = 0$$

non linear innovations as $u_n \perp \mathcal{F}_{n-1}$

$$E(Z_n u_n) = 0 \qquad \text{for all } Z_n \ \mathcal{F}_{n-1} \text{ mble r.v.'s}$$

Martingale Convergence Theorem (MGCT)

$Y_n \equiv MG$, $\sup_n E|Y_n| < \infty$ then

$$Y_n \to_{a.s} Y \qquad \text{some } Y \in L_1(\Omega, \mathcal{F}, P)$$

or $Y_n \equiv MG$, $\sup_n E(Y_n^2) < \infty$ then

$$Y_n \to Y \qquad \text{some } Y \in L_2(\Omega, \mathcal{F}, P)$$

Example

$$Y_n = \Sigma_1^n \varepsilon_t / t \qquad \varepsilon_t \equiv mds(0, \sigma_t^2)$$

$$\sup_t \sigma_t^2 < \infty$$

then

$$E(Y_n^2) = \Sigma_1^n \sigma_t^2 / t^2$$

$$< \sup_t \sigma_t^2 \left(\Sigma_1^\infty \frac{1}{t^2}\right) < \infty$$

so

$$Y_n \to_{a.s} Y = \Sigma_1^\infty \frac{\varepsilon_t}{t} \quad \text{cgt series}$$

Now apply Kronecker lemma

$$\frac{1}{n} \Sigma_1^n \varepsilon_t = \frac{1}{n} \Sigma_1^n t \left(\frac{\varepsilon_t}{t}\right) \to_{a.s.} 0$$

gives SLLN for mds

Maximal inequality for MG's

$Y_n \equiv MG$, $p \geq 1$, $\lambda > 0$ then

$$\lambda^p \, P\left(\max_{i \leq n} |Y_i| > \lambda\right) < E|Y_n|^p$$

e.g. for $Y_n = S_n = \sum_1^n u_t$ $u_t \equiv iid(0, \sigma^2)$, $p = 2$

Kolmogorov's inequality

$$P\left(\max_{k \leq n} |S_k| \geq \lambda\right) \leq \frac{var(S_n)}{\lambda^2}$$

Tchebycheff's inequality

$$P\left(|S_n| \geq \lambda\right) \leq \frac{var(S_n)}{\lambda^2}$$

Kronecker Lemma (converts cgt series → seq that cgt to zero)

x_n seq. of real nos s.th. $\sum_1^\infty x_n$ cgt

$b_n \nearrow \infty$ (eg. $b_n = n$)

$$\frac{1}{b_n} \sum_1^n b_i x_i \to 0$$

Toeplitz Lemma (weighted average of seq has same limit)

$a_n \to a$, $w_{ni} \geq 0$ weights with

$\sum_{i=1}^n w_{ni} = 1$, $w_{ni} \downarrow 0$ as $n \to \infty$ $\forall i$

$$\sum_{i=1}^n w_{ni} a_i \to a$$

Example

$S_n = \sum_1^n a_k$, $\sigma_n = \frac{1}{n} \sum_{k=1}^n S_k$ average of partial sums

then $\sigma_n = \frac{1}{n} \sum_{k=1}^n (n - (k-1)) a_k$

a_1
$a_1 + a_2$
$a_1 + a_2 + a_3$
$\vdots$
$n a_1 \; (n-1) a_2 \; (n-2) a_3$

$$= \sum_{k=1}^n \left(1 - \frac{k-1}{n}\right) a_k \quad \text{Cesaro sum}$$

$$\to \sum_{k=1}^\infty a_k \quad \text{if cgt}$$

Spectral Theory & Discrete Fourier Transforms (dft's)

Spectrum

X_t stationary with $\gamma_h = E(X_t X_{t+h})$ autocovariance

and $\sum_0^\infty |\gamma_h| < \infty$ summability condition

$$f_x(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_h e^{-i\lambda h}$$

cgt uniform & $f_x(\lambda)$ cts

dft

$$w(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{i\lambda t}$$

$$\lambda_s = \frac{2\pi s}{n} \quad \text{fundamental frequencies}$$

1:1 transformation between $(X_t)_1^n$ & $(w(\lambda_s))_0^{n-1}$

$$\begin{bmatrix} w(\lambda_0) \\ \vdots \\ w(\lambda_{n-1}) \end{bmatrix} = U \begin{pmatrix} X_1 \\ \\ X_n \end{pmatrix} \qquad U = \left[\left(\frac{e^{2\pi jk/n}}{\sqrt{2\pi n}} \right)_{jk} \right]_{n\times n}$$

$$UU^* = \frac{1}{2\pi} I \quad , \quad (2\pi)^{1/2} U = \text{unitary matrix}$$

Periodogram

$$I_x(\lambda) = w_x(\lambda) w_x(\lambda)^* = |w_x(\lambda)|^2$$

$$E(I_x(\lambda)) \to f_x(\lambda) \qquad \text{asymptotically unbiased}$$

$$\text{var}(I_x(\lambda)) \to \begin{cases} f_x(\lambda)^2 & \lambda \neq 0, \pi \\ 2f_x(\lambda)^2 & \lambda = 0, \pi \end{cases}$$

so $I_x(\lambda)$ is inconsistent estimator of spectrum

Spectral estimator

$$\hat{f}_x(\omega) = \frac{1}{m} \sum_{\lambda_s \in B} I_x(\lambda_s) \qquad B = \left\{ \omega - \frac{\pi}{2M} < \lambda_s < \omega + \frac{\pi}{2M} \right\}$$

$n = 2mM$

$M, m \to \infty$ as $n \to \infty$

band of width $\frac{\pi}{M}$

$= m \; (2\pi/n)$

(# in band) increment in λ_s

$\hat{f}_x(\omega) \xrightarrow[p]{} f_x(\omega)$ consistent

$m^{1/2}\left(\hat{f}_x(\omega) - f_x(\omega)\right) \to_d N(0, V(\omega))$ asymp. normal

$V(\omega) = \text{const. } f(\omega)^2$

tradeoff (i) bigger is band B — higher is m
smaller the variance
larger the bias

(ii) smaller is band B — smaller is m
lower is bias
larger is variance

optimal choice of bandwidth M?
- depends on kernel
- minimizes asymptotic MSE criterion
- relies on "plug in" value of $f(\omega)$

Asymptotic theory of dft

$X_t = C(L)\,\varepsilon_t$ use generalized BN decomposition

$$C(L) = C(e^{i\lambda}) + \tilde{C}_\lambda(L)\left(Le^{-i\lambda} - 1\right)$$

$$\tilde{C}_\lambda(L) = \sum_{s=0}^{\infty} \tilde{c}_{s\lambda} L^s$$

$$\tilde{c}_{s\lambda} = e^{-i\lambda s} \sum_{s+1}^{\infty} c_k e^{i\lambda k}$$

dft

$$w_x(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_1^n X_t e^{i\lambda t}$$

$$= C(e^{i\lambda})\, w_\varepsilon(\lambda) + \frac{1}{\sqrt{2\pi n}}\left(\tilde{\varepsilon}_{0\lambda} - e^{i\lambda n}\tilde{\varepsilon}_{n\lambda}\right)$$

$$\tilde{\varepsilon}_{t\lambda} = \tilde{C}_\lambda(L)\,\varepsilon_t$$

$$= C(e^{i\lambda})\, w_\varepsilon(\lambda) + o_p(1)$$

$$w_\varepsilon(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_1^n \varepsilon_t e^{i\lambda t}$$

ε_t : iid $(0, \sigma_\varepsilon^2)$

asymptotics for $w_\varepsilon(\lambda)$

Suppose $\lambda_s = \frac{2\pi s}{n} \in B$ so $\lambda_s \to \omega$ as $n \to \infty$

$$w_\varepsilon(\lambda_s) \to_d N_c\left(0, \frac{\sigma^2}{2\pi}\right), \text{ as } E\left(w_\varepsilon(\lambda) w_\varepsilon(\lambda)^*\right) = E\left(I_\varepsilon(\lambda)\right) \to f_\varepsilon(\lambda) = \frac{\sigma^2}{2\pi}$$

Hence

$$w_X(\lambda_s) = C(e^{i\lambda_s})\, w_\varepsilon(\lambda_s) + o_p(1)$$
$$\to_d N_c\left(0, \frac{\sigma^2}{2\pi} C(e^{i\omega}) C(e^{i\omega})^*\right)$$
$$\equiv N_c(0, f_X(\omega))$$

Spectra for parametric models

AR $\quad a(L) X_t = \varepsilon_t \quad f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|a(e^{i\lambda})|^2}$

ARMA $\quad a(L) X_t = b(L) \varepsilon_t \quad f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{|b(e^{i\lambda})|^2}{|a(e^{i\lambda})|^2}$

Linear process

$X_t = C(L) \varepsilon_t \quad f_X(\lambda) = \frac{\sigma^2}{2\pi} C(e^{i\lambda}) C(e^{i\lambda})^*$

AR(1) $\quad X_t = \theta X_{t-1} + \varepsilon_t$

$$f_X(\lambda) = \frac{\sigma^2/2\pi}{|1-\theta e^{i\lambda}|^2} = \frac{\sigma^2/2\pi}{1+\theta^2 - 2\theta\cos\lambda}$$

at $\theta = 1$

$$f_X(\lambda) = \frac{\sigma^2/2\pi}{|1-e^{i\lambda}|^2} \sim \frac{\sigma^2}{2\pi\lambda^2} = O\left(\frac{1}{\lambda^2}\right)$$

as $\lambda \to 0$

i.e.

$f_X(\lambda)$ has discontinuity at $\lambda = 0$ and is not integrable over $(-\pi, \pi)$

Probability & Random Elements on Function Spaces

We are mainly concerned with two function spaces (we want to give these spaces a structure that makes them as close as possible to $(\mathbb{R}, d_e)$ & this is achieved by using metric for closeness that makes them separable + complete).

$C[0,1]$ = space of continuous functions on $[0,1]$ interval, endowed with the uniform metric

$$d_u(f,g) = \sup_t |f(t)-g(t)| \quad , f, g \in C[0,1]$$

which makes $C[0,1]$ a complete metric space (Banach space)

<u>completeness</u>: a metric space (M, d) is complete if it contains all its limit points (limits of all Cauchy sequences)

$(C[0,1], d)$ is complete because, if $\{f_n(t)\}$ is a Cauchy sequence in $C[0,1]$, then $f_n(t)$ converges on $\mathbb{R}$ for a given t, say to $f(t)$. But because of the uniform metric $f_n(t) \to f(t)$ uniformly and hence $f(t)$ is continuous. Thus $f(t) \in C[0,1]$.

- Not all metric spaces are complete. e.g. (Q, d_e) where Q – rationals, d_e = Euclidean metric. Then $x_n = 1 + 1/1! + \ldots 1/n!$ is a Cauchy sequence as $|x_n - x_{n+1}| = 1/(n+1)! \to 0$ But $x_n \to e \notin Q$.

- When we come to discuss convergence in distⁿ in metric space completeness is important. We don't want the prob mass escaping from the space as $n \to \infty$

<u>separability</u>: a space is separable if it contains a countable dense subset

a space is <u>not</u> separable if it contains a noncountable discrete subset

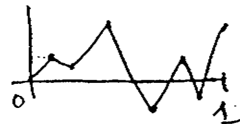
↑

(points [illegible])

e.g: $\mathbb{R}$ is separable because $\mathbb{Q}$ is a countable dense subset

$C[0,1]$ is separable because the rational polygonal functions

$$f(t) = \frac{p}{q} + \frac{m}{n}\left(t - \frac{i-1}{k}\right), \quad \frac{i-1}{k} \le t < \frac{i}{k}$$

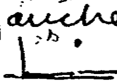


$p, q, m, n \in \mathbb{Z}$

are dense. This is because $\mathbb{Q}$ is dense in $\mathbb{R}$ and so the family $f(t)$ becomes dense in $C[0,1]$.

- separability is important because if it does not hold then not all the Borel sets of the space are measurable
  - e.g. there are noncountable subsets of $\mathbb{R}$ that are not Lebesgue measurable. These can be used to construct noncountable discrete subsets in function spaces that are not measurable as we see below
- also separability ensures that weak cgce on product spaces iff weak cgce on component spaces (Billingsley, p 21)

$D[0,1]$ = space of real valued functions with left limits and right continuous (CADLAG — fns continue à droit, limites à gauche)

excludes  isolated pt function as not right cts

not separable under d_u. Consider

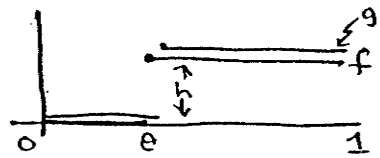
$$f_\theta(t) = \begin{cases} 0 & t < \theta \\ 1 & t \ge \theta \end{cases} \qquad \theta \in [0,1)$$

- The set of functions $\{f_\theta(t),\ \theta \in [0,1]\}$ is uncountable. But $d_u(f_\theta, f_{\theta'}) = 1 \ \forall \theta \ne \theta'$. So the set is also discrete – hence $(D[0,1], d_u)$ is not separable
- This means we can construct spheres $S(f_\theta, \tfrac{1}{2})$ around each point $f_\theta \in D[0,1]$. Take points $\theta \in H$ an uncountable non mble set of $\mathbb{R}$. Then although the spheres $S(f_\theta, \tfrac{1}{2})$

are in the Borel σ-field of $D[0,1]$ = open sets of $D[0,1]$ – we cannot attach a measure to them that corresponds to the Lebesgue measure on $\mathcal{H}$ (note that $m[0,1]=1$ and is a proper probability measure = uniform)

i.e. $\mu[(S(f_\theta, 1/2), \theta \in \mathcal{H})]$, not exist as $m(\mathcal{H})$ doesn't exist.
$= m(\mathcal{H})$

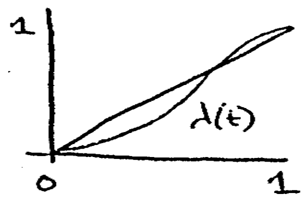
<u>uniform metric</u> $d_u(f,g) = \sup_t |f(t) - g(t)|$



here f & g are "close" by usual common sense standards. However
$d_u(f,g) = h$ (bdd above zero)

<u>Skorohod metric</u> allows time deformations $t \to \lambda(t)$ so that functions like f, g are close if the discontinuities are close in "magnitude" & "timing".

$$d_s(f,g) = \inf_{\lambda \in \Lambda} \left\{ \sup_t |f(t) - g(\lambda(t))| + \sup_t |t - \lambda(t)| \right\}$$

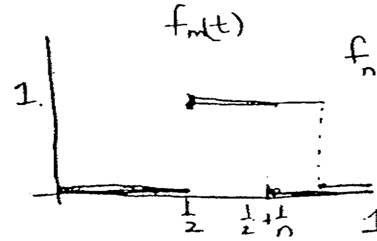


$\Lambda = \{ \lambda(t)$ strictly increasing function $[0,1] \to [0,1]$ continuous $\}$

$(D[0,1], d_s)$ is separable. eg. $f(t) = \frac{m}{n}$, $\frac{i-1}{k} \le t < \frac{i}{k}$ is dense under d_s (ie. rational valued jump functions)

but <u>not</u> complete eg. $f_n(t) = 1_{[\frac{1}{2}, \frac{1}{2}+\frac{1}{n})} \to \begin{matrix} 1 & t = 1/2 \\ 0 & t \ne 1/2 \end{matrix}$

$\notin D[0,1]$



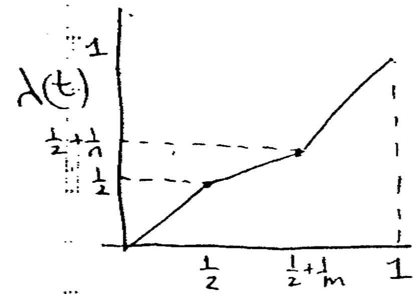
$f_n(t)$ has limit function $f(t) = 1, t=1$; $0, t \neq 1/2$

that is not in $D[0,1]$

$(m < n$ say$)$

$$d_S(f_m, f_n) = \begin{cases} 1 & \text{(= uniform distance between } f_m \text{ \& } f_n \text{, if } \lambda(t) \text{ not the best time deform}^{n}) \\ \left|\frac{1}{n} - \frac{1}{m}\right| \end{cases}$$

$$\lambda(t) = \begin{cases} t & t \leq 1/2 \\ \frac{m}{n}(t-\frac{1}{2}) + \frac{1}{2}, & t > 1/2,\ t \leq \frac{1}{2}+\frac{1}{m} \\ a\left(t - (\frac{1}{2}+\frac{1}{m})\right) + (\frac{1}{2}+\frac{1}{n}) & t \geq \frac{1}{2}+\frac{1}{n} \end{cases}$$



$$a = \frac{\frac{1}{2}-\frac{1}{n}}{\frac{1}{2}-\frac{1}{m}} = \frac{nm-2m}{nm-2n}$$

- idea behind $d_S(\ )$ metric is that we want $\lambda(t) \sim t$, i.e. only small time deformations allowable. This means we can get a situation like the above where
$$d_S(f_m, f_n) = \left|\frac{1}{n} - \frac{1}{m}\right| \to 0 \quad \text{as } n, m \to \infty$$
but the limit function $f(\cdot) \notin D[0,1]$. So space is not complete.

- An "equivalent" metric, devised by Billingsley (1968) under which the space is complete is:

$$d_B(f,g) = \inf_{\lambda \in \Lambda} \left\{ \sup_t |f(t) - g(\lambda(t))| + \sup_{s \neq t} \left| \log \frac{\lambda(t)-\lambda(s)}{t-s} \right| \right\}$$

approx slope of λ

This requires slope of $\lambda(\ )$ to be close to unity (i.e. $\log(\text{slope}) \sim 0$) rather than $\lambda(t) \sim t$. For the above example we get

$$d_B(f_n, f_m) = \min\left\{1, \left|\log \frac{m}{n}\right|\right\} \not\to 0$$

which doesn't converge to zero. So $f_n(t)$ is not cgt in $(D[0,1], d_B)$. The space is complete.

$(D[0,1], d_B)$ is complete and separable (Billingsley d3)
and d_B is equivalent to d_S
(i.e. generates same topology of open sets)

strictly speaking

for $\varepsilon > 0$ $\exists\, \delta > 0$ s.th $\begin{cases} d_B(f,g) < \delta & \Rightarrow \quad d_S(f,g) < \varepsilon \\ d_S(f,g) < \delta & \Rightarrow \quad d_B(f,g) < \varepsilon \end{cases}$

so structure of the two spaces $(D[0,1], d_S)$ and $(D[0,1], d_B)$ is the same - all the change of metric does is to relabel axes and points - but while you may get cgce in one space you may not in the other (as in the above example).

Remark

- As indicated above $(D[0,1], d_u)$ is a complete metric space (all cgt seq's under d_u are in $D[0,1]$) but is not separable.

(*) • $d_B(f,g) \le d_u(f,g)$ as upper bound occurs when $\lambda(t) = t$, & we minimise over $\lambda \in \Lambda$

- Hence uniform cgce in $D[0,1]$ implies d_B / d_S cgce in $D[0,1]$

- A Skorohod (S-) open set is necessarily U-open so

$$\mathcal{D} \subset \mathcal{U}$$

Skorohod-Topology (coarser) Uniform Topology (finer)
- $d_u(f,g) > d_S(f,g)$. helps us to differentiate f & g
- more open sets.

i.e. $S_{d_S}(f,\varepsilon) = \{g \mid d_S(g,f) < \varepsilon\}$ is open sphere (S-open) in $D[0,1]$

e.g. $f_n \xrightarrow[u]{} f \Rightarrow f_n \xrightarrow[S]{} f$ from (*) above

Thus $\sim [f_n \xrightarrow[u]{} f] \Leftarrow \sim [f_n \xrightarrow[S]{} f]$ Billingsley p 150.
i.e. U-open set $\Leftarrow$ S-open set

Examples of function space random elements

(1) Partial sum process : this is a very natural element in $D[0,1]$

$$X_n(r) = \frac{1}{\sqrt{n}} \sum_1^{[nr]} u_j \qquad [nr] = \text{integer part of } nr$$

u_j stationary, $I(0)$

$$= \frac{1}{\sqrt{n}} \sum_1^{k-1} u_j \qquad \text{for} \quad \frac{k-1}{n} \le r < \frac{k}{n}$$

$$= \frac{1}{\sqrt{n}} S_{[nr]} \qquad r \in [0,1]$$

with $S_0 = 0$

$X_n(r) \in D[0,1]$ for all n. In effect, $X_n(r)$ measures a scaled ($1/\sqrt{n}$) partial sum of the errors u_j up to a certain fraction (r) of the total sample (n). Graphically, the partial sum process looks like

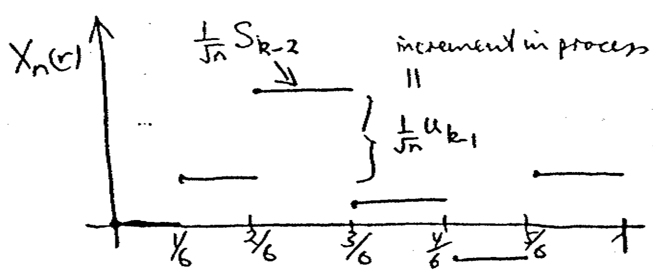


Fig 1.

Note. $X_n(1) = \lim_{r \to 1-} X_n(r)$ to avoid isolated point at $r = 1$

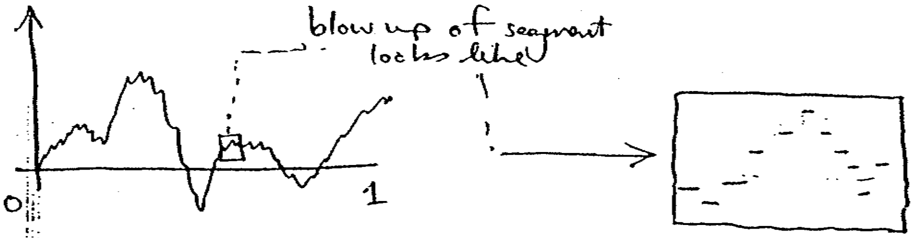
Here $n=6$ and we split the $[0,1]$ interval into segments of length $1/n$

$X_n(r)$ is a constant on each segment

$$\lim_{r \to \frac{k-1}{n}} X_n(r) = \frac{1}{\sqrt{n}} S_{k-1} \quad \text{(right continuity)}$$

signified by "dot" on line segments in Figure

As $n \nearrow \infty$ the segments move closer & become shorter (with smaller jumps $\frac{1}{\sqrt{n}} u_{k-1}$, at least when $E(u_t^2) < \infty$) & we end up with a continuous curve (in appearance) but when the segments are blown up they look like Fig 1



(ii) Continuous version of partial sum process

$$\bar{X}_n(r) = \frac{1}{\sqrt{n}} S_{[nr]} + \frac{nr-[nr]}{\sqrt{n}} u_{[nr]+1} \in C[0,1]$$

$$\frac{k-1}{n} \le r < k/n$$

$$\lim_{r \to \frac{k-1}{n}} \bar{X}_n(r) = \frac{1}{\sqrt{n}} S_{k-1} \quad \text{(right continuity)}$$
$$= \bar{X}_n(k-1/n)$$

$$\lim_{r \to \frac{k}{n}} \bar{X}_n(r) = \frac{1}{\sqrt{n}} S_{k-1} + \frac{1}{\sqrt{n}} u_{k-1+1} = \frac{1}{\sqrt{n}} S_k$$

(left continuity)

$$= \bar{X}_n(k/n)$$

Here the jumps in $\frac{1}{\sqrt{n}} S_{[nr]}$ are eliminated by line segments that connect the partial sums at each k/n $(k=0,\dots,n)$

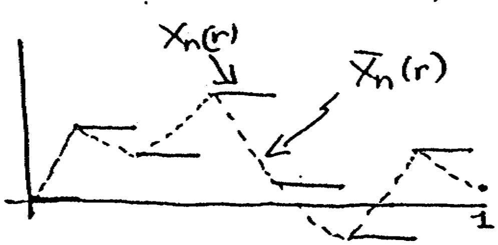
Note $\quad 0 \le nr - [nr] < 1 \quad$ for $\frac{k-1}{n} \le r < \frac{k}{n}$

i.e. $\quad nr-(k-1) < k-(k-1) = 1$

& $\quad nr-(k-1) \ge 0$

so
$$\frac{nr-[nr]}{\sqrt{n}} = O\left(\frac{1}{\sqrt{n}}\right) \quad \text{uniformly in } r \in \left[\frac{k-1}{n}, \frac{k}{n}\right).$$

and the asymptotic behaviour of $\bar{X}_n(r)$ is the same as $X_n(r)$.



Note. we can include $u_n/\sqrt{n}$ in def^{n} so that $\lim_{r \to 1-} \bar{X}_n(r) = \bar{X}_n(1)$; $\in D[0,1]$

(iii) Empirical cdf

Suppose $(X_t)_1^n$ is stationary sequence and

$F(x) = P(X_t \le x) = \text{cdf}(X_t)$. Then

$$F_n(x) = \frac{\# X_t \le x}{n} = \frac{1}{n}\sum_{t=1}^{n} 1(X_t \le x)$$

proportion of $X_t \le x$,

= empirical cdf

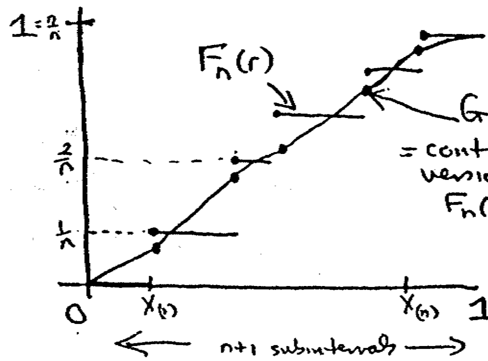
Note.

$$F_n(x) = \frac{1}{n}\sum_{t=1}^{n} 1(X_t \le x) \xrightarrow[a.s]{} E(1(X_t \le x)|J)$$
$$= E(1(X_t \le x)) \quad \text{if } X_t \text{ ergodic}$$
$$= \int 1(X \le x)\, dP$$
$$= P(X \le x)$$
$$= F(x)$$

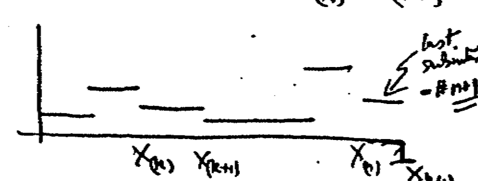
Empirical process

$$Y_n(r) = \sqrt{n}\left(F_n(r) - F(r)\right)$$

Uniform Case $F(r) = r$ on $[0,1]$



put mass $\frac{1}{n+1}$ on each subinterval $X_{(k)}$ $X_{(k+1)}$



$$G_n(x) = \frac{1}{n+1}\,\frac{1}{X_{(k+1)} - X_{(k)}}\int_{X_{(k)}}^{x} dt + G_n(X_k)$$

$X_{(k)} \le x < X_{(k+1)}$

$$\text{mass} = \frac{1}{n+1} = \left(\frac{1}{n+1}\right)\frac{1}{X_{(k+1)} - X_{(k)}}\int_{X_{(k)}}^{X_{(k+1)}} dt$$

$G_n(x) = \frac{1}{n+1}\,\frac{1}{X_1}\int_0^x dt$ over $0 < x \le X_1$

An important property of $D[0,1]$

Thm (Billingsley, p110)

$\forall x \in D[0,1]$, $\forall \varepsilon > 0$ $\exists$ partition $\{0 = t_0, \dots, t_k = 1\}$ of $[0,1]$ s.th.

$$\sup_{s,t \in [t_{i-1}, t_i)} |x(t) - x(s)| < \varepsilon$$

Remarks.

(1) The theorem implies that there are at most finitely many points at which the jump $|x(t) - x(t-)|$ exceeds a given +ve number (ε)

So $x(t) \in D[0,1]$ has at most a countable number of discontinuities

(2) Since there are only a finite # of pts for which $|x(t) - x(t-)|$ exceeds ε, let M_ε be the maximum discontinuity in these jumps. Then

$$\sup_t |x(t)| \leq \sup_{t \in [t_{i-1}, t_i)} |x(t)| + kM_\varepsilon < \infty$$

so $x(t)$ is bounded above.

(3) We also deduce that $x(t) \in D[0,1]$ can be uniformly (to $\varepsilon > 0$) approximated by simple functions that are constant over subinterval. Hence, $x(t)$ is a Borel measurable function.

Probability in Metric Spaces

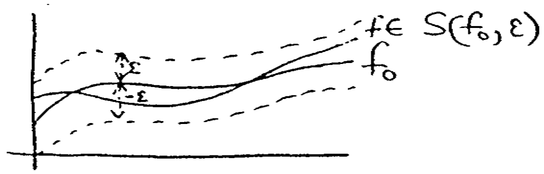
[ref's Parthasarathy (1967) Prob. measures on Metric Spaces
Billingsley (1968) Convergence of Measures]

Let (D, d) be a metric space that is separable & complete. We can build a probability space for D like the space $(\mathbb{R}, \mathcal{B}, P)$ for the real line $\mathbb{R}$ where $\mathcal{B}$ is the Borel σ-field.

Use the metric d to construct open sets like

$$S(f_0, \varepsilon) = \{ f \mid d(f, f_0) < \varepsilon \} = \text{open sphere around } f_0$$

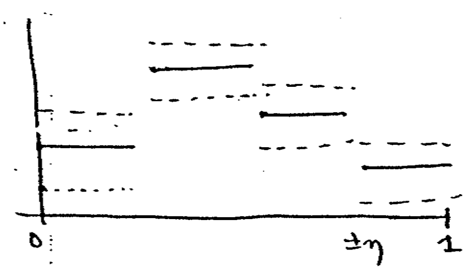
eg. (i) $D = C[0,1]$, $d = d_u$



$$d(f, f_0) = \sup_t |f(t) - f_0(t)| < \varepsilon$$

(ii) $D = D[0,1]$, $d = d_B = \inf_{\lambda \in \Lambda} \left\{ \sup |f(t) - g(\lambda(t))| + \|\lambda\| \right\}$

$$\|\lambda\| = \sup_{t \neq s} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|$$



The probability triple $(D, \mathcal{D}, P)$

(1) let $\mathcal{D}$ = Borel σ-field of D
= smallest σ-algebra of subsets of D. That contains the open sets

As usual we have

(i) $\{D, \phi\} \in \mathcal{D}$

(ii) $A \in \mathcal{D} \Rightarrow A^c \in \mathcal{D}$

(iii) $A_1, A_2, \ldots \in \mathcal{D} \Rightarrow \bigcap_1^\infty A_i \in \mathcal{D}, \ \bigcup_1^\infty A_i \in \mathcal{D}$

(2) Let P be a countably additive non negative set function on $\mathcal{D}$ with the properties

$$P(D) = 1 \quad , \quad P\left(\bigcup_1^\infty A_i\right) = \sum_1^\infty P(A_i) \,, \quad A_i \cap A_j = \phi$$
(disjoint A)

Then

$$(D, \mathcal{D}, P) = \text{probability space}.$$

Weak Convergence in $(D, \mathcal{D}, P)$

When $(D, \mathcal{D}, P) = (\mathbb{R}, \mathcal{B}, P)$, a probability space on the real line $\mathbb{R}$, we usually define weak convergence by:

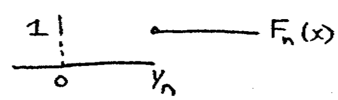
(a) cdf cgce: $F_n(x) \to F(x)$ at <u>all points of cty</u> of F (*)

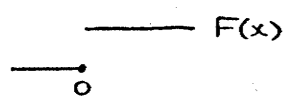
$F_n(x) = P(X_n \le x)$ cdf completely defines P on $\mathbb{R}$ as the intervals $(-\infty, x]$ generate $\mathcal{B}$.

we say $X_n \xrightarrow[d]{} X$

Requirement (*): this avoids the disappearance of mass

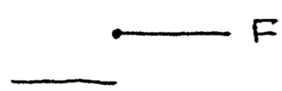
e.g. $F_n(x) = \varepsilon(x - \frac{1}{n}) = 0 \quad x < 1/n$
$\qquad\qquad\qquad\qquad\quad 1 \quad x \ge 1/n$



$\to$ 

as $F_n(0) = \varepsilon(-\frac{1}{n}) = 0 \to 0$

So $x = 0$ is not a pt of cty of F and we need to transfer mass in the limit function to pt $x = 0$ i.e. redefine $F(x)$ as.



[as $F(x) = P(X \le x)$ is cts <u>on right</u>]

(b) cf cgce: $cf_n(s) \to cf(s)$ pointwise and $cf(s)$ cts at $s = 0$

Remark $cf(s) = E(e^{isX}) = \int_{-\infty}^{\infty} e^{isx}\, dF(x)$, so ch. func $cf(s)$ is cts everywhere

So if limit function is to be a ch. func it must be cts. It is enough (sufficient) to require that the limit $cf(s)$ be cont. at $s=0$

Examples (i) $f_n(x) = \begin{cases} \frac{1}{2n} & x \in [-n, n] \\ 0 & \text{elsewhere} \end{cases}$ (pdf)

$$cf_n(s) = \frac{1}{2n}\int_{-n}^{n} e^{isx}\,dx = \frac{1}{2n}\,\frac{e^{ins} - e^{-ins}}{is}$$

$$= \frac{\sin(ns)}{ns}$$

$$\longrightarrow \begin{cases} 1 & s = 0 \\ 0 & s \neq 0 \end{cases}$$

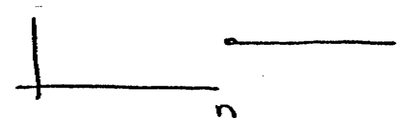
(not cts)

$f_n(x) \to 0$ everywhere (not a pdf)

$F_n(x) = \frac{1}{2n}\int_{-n}^{x} dt = \frac{x+n}{2n} \to \frac{1}{2} \quad \forall x$

(not a cdf)

(ii) $F_n(x) = \varepsilon(x-n)$



$cf_n(s) = \int e^{isx}\, dF_n(x) = e^{ins}$ not cgt

$F_n(x) \to 0 \quad \forall x$ not a proper cdf.

In the general probability space $(\mathcal{D}, \mathscr{D}, P)$ we cannot use pdf, cdf or even cf cgce (there is some scope for working with characteristic functional rather than ch. fnc's, however). It is most convenient to work directly with the sequence of probability measures $\{P_n\}$

Weak Convergence of $\{P_n\}$

Let $\{P_n\}$ be a seq. of prob. measures on $(\mathcal{D}, \mathscr{D})$

Then

$$P_n \Rightarrow P \qquad (P_n \text{ converges weakly to } P \text{ on } (\mathcal{D}, \mathscr{D}))$$

if

$$P_n(A) \to P(A) \quad \text{for all events } A \in \mathscr{D} \text{ s.th } P(\partial A) = 0$$

i.e. no atom of mass on boundary of A

where

$$\partial A = \text{boundary of } A = \bar{A} \cap \overline{A^c}$$

$$\bar{A} = \text{closure of } A, \quad \overline{A^c} = \text{closure of complement of } A$$

P– continuity sets

Sets $A \in \mathscr{D}$ for which $P(\partial A) = 0$ are called "P"– continuity sets.

Example 1 $\mathcal{D} = \mathbb{R}$, $\mathscr{D} = \mathcal{B}$

$$P_n \Rightarrow P \quad \text{iff} \quad F_n(x) \to F(x) \quad \text{at all cty points of } F$$

i.e. $P(X=x) = 0$

here

$$F(x) = P(X \le x), \quad \partial A = x \quad \text{with } A = (-\infty, x]$$

so

$$P(\partial A) = P(X = x) = 0$$

Remark The sets $\{y \mid y \le x\} = (-\infty, x]$ are a convergence determining class in $(\mathbb{R}, \mathcal{B}, P)$ i.e. cgce for these sets (intervals) implies cgce $\forall \mathcal{B}$.

Def$^{\underline{n}}$: $\mathcal{V}$ is a <u>cgce determining class</u> of $\mathcal{D}$ if convergence
$$P_n(A) \to P(A) \quad \forall \text{ P-cty sets } A \in \mathcal{V}$$
ensures weak cgce of P_n to P i.e.
$$P_n(A) \to P(A) \quad \forall \text{ P-cty sets } A \in \mathcal{D}.$$

Def$^{\underline{n}}$: $\mathcal{V}$ is a <u>determining class</u> of $\mathcal{D}$ if measures P & Q on $(D, \mathcal{D})$ are identical on $\mathcal{D}$ whenever they are identical on $\mathcal{V}$

The intervals $(-\infty, x]$ are determining and convergence determining in $(\mathbb{R}, \mathcal{B})$.

<u>Example 2</u> $D = \mathbb{R}_\infty$, $\mathcal{D} = \mathcal{B}_\infty$

The coordinate representation for a time series $\{X_n\}$ with trajectories
$$x = (x_1, x_2, \cdots) \in \mathbb{R}_\infty$$
The Borel field $\mathcal{B}_\infty$ is generated by product cylinders of the form
$$(*) \qquad \left(\mathop{\times}_{1}^{r} \mathbb{R}\right)\left(\mathop{\times}_{r+1}^{s} B_i\right)\left(\mathop{\times}_{s+1}^{\infty} \mathbb{R}\right) \qquad B_i \in \mathcal{B}$$
We can think of these sets another way, in terms of projections

<u>Projections</u>
$$\pi_k(x) = (x_1, \cdots, x_k) \in \mathbb{R}^k$$
is a finite dimensional (fidi) projection
$$\pi_k : \mathbb{R}_\infty \to \mathbb{R}^k$$

Pre-image of a projection

Let $H \in \mathcal{B}_k = \mathcal{B}(\mathbb{R}^k)$ be a Borel set in $\mathbb{R}^k$

Then

$$\pi_k^{-1} H = \text{preimage of } H \text{ under } \pi_k$$
$$= \text{cylinder set of form } (*) \text{ on p.65}$$

Finite dimensional sets

$$\{\pi_k^{-1} H, \; H \in \mathcal{B}_k\} \quad \forall k$$

$\equiv$ determining class + cgce determining class in $(\mathbb{R}_\infty, \mathcal{B}_\infty)$

i.e. $P_n(A) \to P(A)$ on $\{\pi_k^{-1} H, H \in \mathcal{B}_k, \forall k\}$

implies

$$P_n \Rightarrow P \quad \text{on } (\mathbb{R}_\infty, \mathcal{B}_\infty)$$

i.e. fidis are det. & cgce det. class on $(\mathbb{R}_\infty, \mathcal{B}_\infty)$

(proof: Billingsley p 19 & theorem 2.2 p 14-15)

Example 3 $\quad D = C[0,1], \; \mathcal{D} = \mathcal{C}$

$C = C[0,1]$ with uniform metric $d_u(f,g) = \sup_t |f(t) - g(t)|$

$\mathcal{C}$ - Borel σ-field of $C[0,1]$ = σ field generated by subsets of $C[0,1]$ that are open wrt d_u = σ-field generated by open spheres $S_u(f,\varepsilon) = \{g \in C \mid d_u(f,g) < \varepsilon\}$

Projections on C we define π_k-projections by

$$\pi_{t_1 \dots t_k}(x) = (x(t_1), \dots, x(t_k)) \in \mathbb{R}^k$$

so

$$\pi_{t_1 \dots t_k} : C[0,1] \to \mathbb{R}^k \quad \text{projection mapping}$$

$$\pi_{t_1 \dots t_k}^{-1} H = \text{finite dimensional set}$$
$$= \text{preimage of Borel set } H \in \mathcal{B}_k$$
$$\in \mathcal{C}$$

Remarks

(i) Note that the closed sphere

$$\underline{S}(f, \varepsilon) = \{ g \mid d_u(f,g) \le \varepsilon \}$$

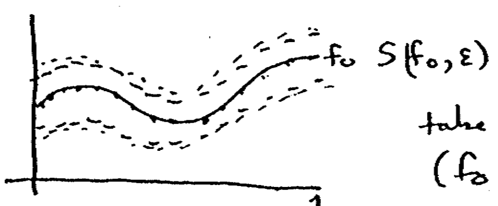
is the limit of the finite dimensional sets

$$(*) \quad \{ g \mid |f(i/k) - g(i/k)| \le \varepsilon, \ i = 1, \dots, k \}$$

(ii) $C[0,1]$ is separable, so each open set in $\mathcal{C}$ is a countable union of open spheres (& hence of closed spheres like $\bigcup_1^\infty \underline{S}(f, \varepsilon - \frac{1}{n}) = S(f, \varepsilon)$ = open sphere. Thus finite dimensional sets like $(*)$ generate $\mathcal{C}$. Hence

$$\boxed{\text{finite dimensional sets} = \text{determining class}}$$

ie. $P = Q$ on fidis, then $P = Q$ on $\mathcal{C}$



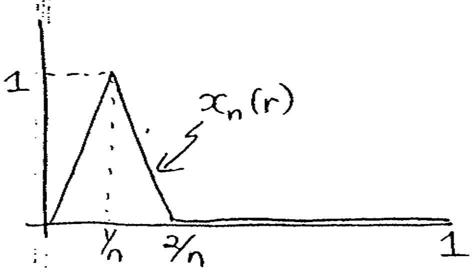
take limit & pts of comparison f & g $(f_0 \pm \varepsilon)$ are rationals = dense in $[0,1]$. By continuity, get the open sphere $S(f_0, \varepsilon)$.

Fidi's on $C[0,1]$ are NOT ~~cgce~~ determining

We illustrate with the following example from Billingsley (p.20)

P defined by $P(x = x_0) = 1$ $\quad x_0(r) = 0 \ \forall r$ zero function

P_n defined by $P_n(x = x_n) = 1$ where



$\equiv$ tent function $\in C[0,1]$

$$x_n(r) = \begin{cases} nr & 0 \le r \le 1/n \\ 2-nr & 1/n \le r \le 2/n \\ 0 & 2/n \le r \le 1 \end{cases}$$

Now $x_n(r) \longrightarrow x_0(r)$ pointwise $\forall r$

but NOT uniformly since

$$\sup_r |x_n(r) - x_0(r)| = d_u(x_n, x_0) = 1, \ \forall n$$

Differences between fidi's & ∞-dim sets (spheres)

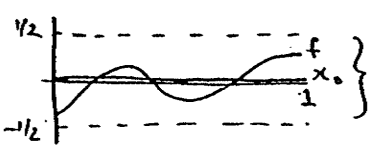
(i) $P_n(A) \to P(A)$ $\forall$ finite dim sets above

eq.^{n} if $A = \pi^{-1}_{1,\dots,k} H$ simply select N s.th $\frac{2}{n} \le t_j \ \forall j = 1,\dots,k, \ \forall n > N$

Then

$P_n(A) = P(A)$ $\quad \forall n > N$ because $x_n(r) = x_0(r)$ on this set of $\{t_j\}$

(ii) Now let $A = S(x_0, 1/2)$ = open sphere around x_0 (zero function)



$S(x_0, 1/2)$

Note $P(\partial A) = 0$ since $P(A) = 1$. so A is P-cty set

Note $x_n \notin S(x_0, 1/2) \ \forall n$

Thus $P_n(A) = 0 \ \forall n$ $\not\to 1 = P(A)$ $\Big\} \Rightarrow$ mass escapes from P_n as we take limit $n \to \infty$

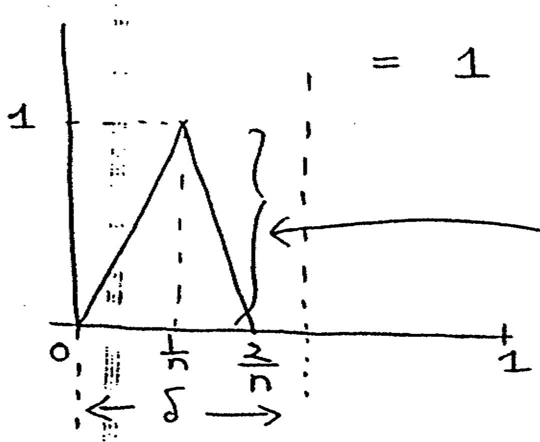
Remark 0

- Need to exclude functions (i.e. attach negligible probability to them in $C[0,1]$ as $n\to\infty$) whose fluctuations are too great, like the example above, if we are to get weak convergence of measures in $C[0,1]$.
- We can measure fluctuations using the modulus of continuity

$$w(x,\delta) = \sup_{|s-t|<\delta} |x(s)-x(t)|$$

- In the above case we have

$$w(X_n,\delta) = \sup_{|s-t|<\delta} |X_n(s)-X_n(t)|$$



$$= 1 \qquad \forall n > n_*$$

where $n_* \gg 2/\delta$

maximum fluctuation over $[0,\delta]$ is 1, i.e.

$$X_n(\tfrac{1}{n}) - X_n(\tfrac{2}{n}+\varepsilon) = 1$$

Implications

- $\forall \varepsilon>0, \forall \delta>0$ we can find n_* s.th $\forall n>n_*$

$$P(w(X_n,\delta)>\varepsilon) = 1$$

- i.e. it is NOT true that $\forall \varepsilon,\eta>0\ \exists \delta>0$ s.th for some n_0

$$P(w(X_n,\delta)>\varepsilon) < \eta \qquad \forall n>n_0$$

Remark 1 In above example if we set

$A = S(x_0, 1+\delta)$ for some $\delta > 0$, then

$$P_n(A) = 1 \quad \forall n$$
$$\to 1 = P(A) \quad \text{as } n \to \infty$$

So, in this case, there is no loss of mass. For the sphere $A = S(x_0, 1/2)$, mass appears from nowhere as $n \to \infty$, yet A is a P-cty set. Hence, failure of weak convergence

i.e. $x_n \not\Rightarrow x_0$ in $(C[0,1], \mathcal{C}, d_u)$

Remark 2

This example shows that fidi cgce is not enough to establish that $P_n \Rightarrow P$. We need something more than fidi cgce. What we need are:

(i) fidi cgce

\+ (ii) tightness of $\{P_n\}$

Def$^{\underline{n}}$ A family $\mathcal{P}$ of probability measures P on $(D, \mathcal{D})$ is tight if $\forall \varepsilon > 0$ $\exists$ compact set K s.th

$$P(K) > 1-\varepsilon \quad \forall P \in \mathcal{P}$$

i.e. is there a compact set in the space that contains almost all of the mass ~~for~~ all the measures in the family $\mathcal{P}$.

Note 1 Can apply this idea to a single probability measure, i.e. P is tight if $\exists K$ compact s.th $P(K) > 1-\varepsilon$

Clearly, P is tight if it has compact support (i.e. $\exists A \in \mathcal{D}$ s.th $P(A) = 1$ and A is compact)

P is also tight if it has σ-compact support (i.e. if $\exists A = \bigcup_i^\infty A_i$ with A_i compact s.th $P(A) = 1$)
For then $\exists$ compact set $A(n) = \bigcup_i^n A_i$ s.th given $\varepsilon > 0$ $P(A(n)) > 1-\varepsilon$, as $A(n) \to A$ and $P(A) = 1$)

Note 2: We have the following useful results

Th$^{\underline{\underline{m}}}$ (Billingsley p.10)

If D is a separable & complete space then every probability measure on $(D, \mathcal{D})$ is tight

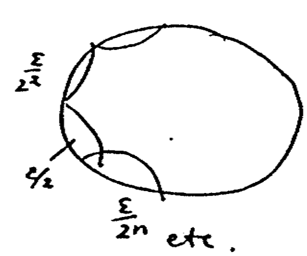
Pf Since D is separable $\exists$, for each n, a sequence $A_{n1}, A_{n2}, ..$ of open $1/n$ spheres that cover D
We can choose i_n s.th $P\left(\bigcup_{i \le i_n} A_{ni}\right) > 1 - \frac{\varepsilon}{2^n}$
Consider the set

$$A = \bigcap_{n \ge 1} \bigcup_{i \le i_n} A_{ni} \quad \text{which is} \subset \bigcup_{i \le i_N} A_{Ni}$$

for some N & is a finite $\varepsilon = \frac{1}{N}$ net for A

Thus, A is totally bounded (by def'n)

$$P(A) > 1 - \underbrace{\sum_{n=1}^{\infty} \frac{\varepsilon}{2^n}}_{\substack{\text{sum of measures} \\ \text{of pieces left out of} \\ \bigcup_{i \le i_n} A_{ni}}} = 1-\varepsilon$$



Since A is totally bounded it has compact closure $\bar{A}$. Thus $P(\bar{A}) > 1-\varepsilon$. So P is tight.
(Billingsley, p. 217)

Remarks

(i) $\mathbb{R}$ (and $\mathbb{R}^k$) are σ-compact spaces. So all probability measures on $\mathbb{R}$ & $\mathbb{R}^k$ are tight

(ii) $(C[0,1], \mathcal{C}$, $(D[0,1], \mathcal{D}, d_B)$ are separable & complete spaces. So all probability measures on these spaces are tight.

Examples

Ex 1. $P_n(X = n) = 1$, cdf $F_n(x) = 0 \quad x < n$.
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 1 \quad x \geq n$

n

(*) $\quad F_n(x) \to F(x) = 0 \quad \forall x$

- P_n is not tight because $\exists$ no compact K (here compact = closed & bdd as $\mathbb{R}$ is Euclidean space) for which given $\varepsilon > 0$

$$P_n(K) > 1 - \varepsilon \quad \forall n$$

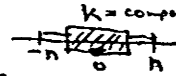
[take any compact K and choose n_0 such that $n_0 > \sup_{x \in K} |x|$. Then $P_n(K) = 0 \quad \forall n > n_0$]

- Note that although we get cgce of $F_n \to F$ at all points of continuity of F, we have lost all of the mass of P_n in the limit function F, which is not a cdf.

Ex 2. P_n = prob. measure of uniform dist^{n} on $[-n, n]$
pdf: $f_n(x) = 1/2n$ cdf $F_n(x) = \int_{-n}^{x} \frac{1}{2n} ds = \frac{1}{2n}(x+n)$

$$F_n(x) \to 1/2 \quad \forall x$$

K = compact set

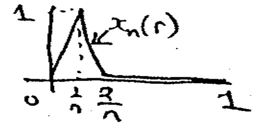


always non negligible mass outside for large enough n

Again limit of $F_n(x)$ is not a cdf and mass has escaped — here it has been smudged out as $n \to \infty$.

Ex 3 $(C[0,1], \mathcal{B})$: $P_n(x = x_n) = 1$

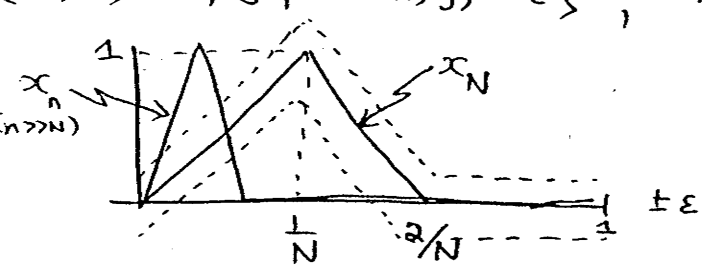
where x_n = tent function on p 68



- there is no compact set K of $C[0,1]$ that contains the sequence $\{x_n\}_{n \geq 2}$
- compactness in $C[0,1]$ requires that every open cover of K have a finite subcover

Consider the ε-net for $\{x_n\}$ given by the spheres

$$E = \{ S(x_n, \varepsilon) = \{ g \mid d_u(x_n, g) < \varepsilon \}, \quad n = 2, 3, \dots \}$$



Suppose there were a finite subcover

$$E_N = \{ S(x_n, \varepsilon), \ n \leq N \}$$

Clearly $x_n \notin E_N$ for $n \gg N$

- Note also that $x_n \to x_0$ = zero function pointwise but not uniformly in $C[0,1]$.

Thus although $x_n \in C[0,1]$ $\forall n$ the sequence $\{x_n\}_{n \geq 2}$ does NOT have a limit point in $C[0,1]$ under d_u. But every compact subset of $C[0,1]$ has the property that it is complete i.e. has all its limit points. Thus no compact set K of $C[0,1]$ can contain the infinite sequence $\{x_n\}_{n \geq 2}$

i.e. There is NO compact K s.th $P_n(K) > 1 - \varepsilon$, $\varepsilon > 0$.

L7

Functional Central Limit Theory

idea: Our object is to characterise the distribution as $n \to \infty$ of random elements that live in function spaces like $C[0,1]$, $D[0,1]$

e.g. the partial sum process $X_n(r) = \frac{1}{\sqrt{n}} \sum_1^{[nr]} u_j$

empirical process $Y_n(r) = \sqrt{n}\left(F_n(r) - F(r)\right)$

Since $X_n(r) \in D[0,1]$, the limit theory we obtain is described as a functional limit theory (here on $D[0,1]$) — or central limit theory on function spaces

assumptions & methods

- Note that $X_n(r)$ is just a function space version of the usual (Euclidean space) random element

$$\frac{1}{\sqrt{n}} S_n = \frac{1}{\sqrt{n}} \sum_1^n u_t$$

Indeed, when $r=1$ we have $X_n(1) = \frac{1}{\sqrt{n}} S_n$
In general, $X_n(r)$ just gives the standardised partial sum of u_t up to a certain fraction (r) of the overall sample

- Working with fractions of the sample & element like $X_n(r)$ turns out to be critical in unit root limit theory, where the whole trajectory of the process (rather than its end point) is important due to persistence in the shocks.
- We can expect many of the ideas / approaches & methods of CLT theory to carry over to FCLT theory for elements like $X_n(r)$

eg.

(i). need to control outlier occurrences through moment conditions

(ii) need to control temporal dependence

methods

(i) blocking / mixing / functions of mixing seq's

(ii) linear process decompositions

<u>terminology</u> We sometimes encounter the terminology "invariance principle" (IP) rather than FCLT. The reason goes back to the early literature (e.g. Erdos & Kac, Donsker) which looked at certain
1946, 1947 1951
specific functionals like

$$(*) \qquad \sup_r X_n(r) = \max_{i \le n} \tfrac{1}{\sqrt{n}} S_i$$

[Note: the importance of functionals like this in statistical tests of structural breaks – like the $\max_{[t_1,t_2]}(\text{Chow})$ test over subinterval $[t_1,t_2]$ or corresponding fractions $t_1/n, t_2/n$ of sample]

Originally, the limit theory of functionals like (*) of the process $X_n(r)$ were found under normality conditions on the underlying sequence (u_j). If one can establish, under certain conditions, the invariance of this limit result to the normality ass^n then we have an <u>invariance principle</u> or IP

<u>Present procedure</u>

- Find a limit law for $X_n(r)$ on a function space like $D[0,1]$ or $C[0,1]$ via an FCLT
- Then use continuous mapping theorems to map the limit law of $X_n(r)$ into the limit law for the functional like $\sup_r X_n(r)$

<u>Notation</u> Weak cgce of prob law P_n to P is denoted $P_n \Rightarrow P$. Similarly if $X_n(r) \equiv_d P_n$ (distributed according to P_n) we write $X_n(r) \Rightarrow X(r)$ or $X_n(r) \Rightarrow_d X(r)$ [illegible]

Partial sums of iid sequences

We establish an FCLT for partial sums of iid sequences & then use the Phillips-Solo BN decomposition approach to generalize this to linear processes.

Theorem (Donsker's theorem for partial sums)

Let $\{u_j\} \equiv iid(0,\sigma^2)$, convenient also to assume $E(u_j^4)<\infty$,

Then
$$X_n(r) = \frac{1}{\sqrt{n}\sigma}\sum_1^{[nr]} u_j \Longrightarrow W(r) \equiv BM(1)$$
$\equiv$ standard Brownian motion on $C[0,1]$

Recall $W(r)$ is completely defined by its properties

(i) $W(0)=0$ (Starts at origin)
(ii) $W(r) \equiv N(0,r)$ (Gaussian marginals)
(iii) $W(s)$ indep of $W(r)-W(s)$ $\quad 0\le s<r\le 1$ (indep. increm)
(iv) $W(r)$ has cts sample paths

The fact that there exists a random function on $C[0,1]$ with these properties is classical – the existence of Wiener measure (c.f. Billingsley p 62, th^{m} 9.1). It is demonstrated by showing that if the u_j component are normal and if $\overline{X}_n(r) \in C[0,1]$ is the cts version of $X_n(r)$ – c.f. p. 58 above – and P_n is the prob. measure of $\overline{X}_n(r)$ on $C[0,1]$ then $\{P_n\}$ is tight, the fidi's cge to those with properties that correspond to those of a Wiener process ((i) – (iv) above). Thus $P_n \Rightarrow P$, a prob measure on $C[0,1]$, and P is Wiener measure.

Proof It is convenient to work with $\overline{X}_n(r)$. Then all calculations are in $C[0,1]$.
We need to show: (a) fidi cgce (b) tightness for $\overline{X}_n(r)$. Start with:

(a) Consider the one dimensional dist^{n} $\overline{X}_n(r_1)$, given some $0<r_1\le 1$. We have $\overline{X}_n(r_1) = X_n(r_1) + \frac{nr_1-[nr_1]}{\sqrt{n}}u_{[nr_1]+1}$
so $|\overline{X}_n(r_1) - X_n(r_1)| \le \frac{1}{\sqrt{n}}u_{[nr_1]+1} \to_p 0$. Hence we can work with $X_n(r_1)$

Note

$$X_n(r_1) = \frac{1}{\sqrt{n}\,\sigma}\sum_1^{[nr_1]} u_j = \sqrt{\frac{n_1}{n}}\ \frac{1}{\sqrt{n_1}\,\sigma}\sum_1^{n_1} u_j \qquad n_1 = [nr_1]$$

Now $n_1/n = [nr_1]/n \to r_1$ as $n \to \infty$ and by the Lindeberg Levy CLT

$$\frac{1}{\sqrt{n_1}\,\sigma}\sum_1^{n_1} u_j \Rightarrow N(0,1).$$

Hence

$$\bar{X}_n(r_1) \sim X_n(r_1) \Rightarrow r_1^{1/2} N(0,1) \equiv N(0, r_1) \equiv W(r_1)$$

Next consider two-dim. fidi's eg. $(X_n(r_1), X_n(r_2))$ for $0 < r_1 < r_2 \le 1$. Equivalently we consider the vector

(*) $\left(X_n(r_1)\ ,\ X_n(r_2) - X_n(r_1)\right)$

(as an arbitrary lc of $(X_n(r_1), X_n(r_2))$ can always be written as a lc of (*)). Note (*) is

$$\left(\frac{1}{\sqrt{n}\,\sigma}\sum_1^{[nr_1]} u_j\ ,\ \frac{1}{\sqrt{n}\,\sigma}\sum_{[nr_1]+1}^{[nr_2]} u_j\right) = \left(\sqrt{\frac{n_1}{n}}\frac{1}{\sqrt{n_1}\,\sigma}\sum_1^{n_1} u_j\ ,\ \sqrt{\frac{n_2-n_1}{n}}\frac{1}{\sqrt{n_2-n_1}}\frac{1}{\sigma}\sum_{n_1+1}^{n_2} u_j\right)$$

$n_1 = [nr_1]$, $n_2 = [nr_2]$

$$\Rightarrow \Big(\underbrace{r_1^{1/2} N(0,1)\ ,\ (r_2 - r_1)^{1/2} N(0,1)}_{\text{indep}}\Big)$$

$$\equiv \left(N(0, r_1)\ ,\ N(0, r_2 - r_1)\right)$$

$$\equiv \left(W(r_1)\ ,\ W(r_2) - W(r_1)\right)$$

Higher order fidi's cge in the same way, just as req'd.

(b) To prove tightness note first that

$$\bar{X}_n(0) = 0 \quad \forall n \quad , \text{ so } \quad \bar{X}_n(0) \text{ is tight}$$

[let P_n be prob. measure of $\bar{X}_n(0)$ and set $K = \{\text{zero fnc}\}$, which is compact (a singleton). $P_n(K) = 1$. So $\bar{X}_n(0)$ is tight]

To show that $\bar{X}_n(r)$ is tight, we control the possible escape of probability mass by ensuring that the partial sums (S_i) which determine $X_n(r)$ up to a scale factor do not fluctuate too much. Specifically, from the lemma below $\{X_n(r)\}$ is tight if $\forall \varepsilon > 0, \exists \lambda > 1$ and n_0 s.th

$$P\left(\max_{i \le n} |S_{k+i} - S_k| > \lambda\sqrt{n}\right) < \frac{\varepsilon}{\lambda^2}, \quad \forall n \ge n_0, \ \forall k$$

Since the u_j in $S_k = \sum_1^k u_j$ are iid (it would be enough if they were just stationary) we can confine our attention to the $k=0$ case and need only prove

$$(77) \quad P\left(\max_{i \le n} |S_i| > \lambda\sqrt{n}\right) < \frac{\varepsilon}{\lambda^2} \qquad \forall n \ge n_0$$

Now by the maximal inequality for MG's we have:

$$P\left(\max_{i \le n} |S_i| > \lambda\sqrt{n}\right) < \frac{E|S_n|^p}{(\lambda\sqrt{n})^p} \qquad \text{some } p \ge 1$$

Set $p=4$ and assume that u_j has finite fourth moment $E u_j^4 < \infty$. Now

$$E(S_n^4) = E(S_n^2)^2 = E\left(\sum u_j^2 + 2\sum_{i<j} u_i u_j\right)^2$$
$$= E\Big\{\sum u_j^4 + 2\sum_{i<j} u_i^2 u_j^2$$
$$+ 2\left(\sum u_j^2\right)\left(2\sum_{i<j} u_i u_j\right)$$
$$+ 4\Big[\sum_{i<j} u_i^2 u_j^2 + 2\sum_{\substack{i<j \\ i\ne k \text{ or } j\ne \ell}} \sum_{k<\ell} u_i u_j u_k u_\ell\Big]$$

$$= n E(u^4) + 2(E(u^2))^2 \frac{n(n-1)}{2} + 0 + 4(E(u^2))^2 \frac{n(n-1)}{2}$$

$$E(S_n/\sqrt{n})^4 \longrightarrow 3\sigma^4 \equiv \sigma^4 E(Z^4) \quad \text{with } Z \equiv N(0,1)$$

Hence

$$P\left(\max_{i\le n}|S_i| > \lambda\sqrt{n}\right) < \frac{E|S_n/\sqrt{n}|^4}{\lambda^4} \to \frac{3\sigma^4}{\lambda^4}$$

(as $n\to\infty$)

$$< \frac{\varepsilon}{\lambda^2} \quad \text{for } \lambda>1 \text{ large enough}$$

Thus (ㅌ) holds for some $\lambda > 1$ and all $n > n_0$.

It follows that the partial sums S_i do not fluctuate too much and therefore $\{\bar{X}_n(r)\}$ is tight as required. Hence

$$\bar{X}_n(r) \Rightarrow W(r) \equiv BM(1)$$

on $(C, \mathcal{C})$.

Remarks

(i) The MG inequality used in the proof requires only that u_j = mds (not necessary for u_j to be iid).

(ii) We need $S_n/\sqrt{n} \to_d N(0,\sigma^2)$ as a fidi CLT and $|S_n/\sqrt{n}|^p$ to be uniformly integrable for some $p>2$. Then

$$E|S_n/\sqrt{n}|^p \to \sigma^p E|Z|^p \quad \text{for } Z \equiv N(0,1)$$

This will be so if

$$\sup_n E|S_n/\sqrt{n}|^{p+\delta} < \infty \qquad \text{some } \delta>0$$

(iii) In proving tightness of $\bar{X}_n(r)$ we use the following:

Lemma A Let $\bar{X}_n(r)$ be a seq. of random function on $C[0,1]$. $\{\bar{X}_n(r)\}$ is tight iff

(a) $\{\bar{X}_n(0)\}$ is tight (i.e. on the real line $\mathbb{R}$)

and

(b) $\forall \varepsilon, \eta > 0$, $\exists \delta$ $(0<\delta<1)$ and $\exists n_0$

s.th.

$$(*) \quad \frac{1}{\delta} P\left(\sup_{t<s<t+\delta} |\bar{X}_n(s) - \bar{X}_n(t)| \geq \varepsilon \right) \leq \eta$$

for $n > n_0$.

Proof (Billingsley pp 56–58).

Application to Partial sum processes

Let $\bar{X}_n(r) = \frac{1}{\sigma\sqrt{n}} S_{[nr]} + \frac{nr-[nr]}{\sqrt{n}\sigma} u_{[nr]+1} \in C[0,1]$

(i) $\bar{X}_n(0) = 0$, which is tight $\forall n$

(ii) note that

$$(**) \quad \sup_{t<s<t+\delta/2} |X_n(s) - X_n(t)| \leq 2 \max_{0<i<j-k} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k|$$

here $[nt] = k$, $[ns] = k+i$

as

and $\frac{k}{n} \leq t < \frac{k+1}{n}$, $\frac{j-1}{n} \leq t+\frac{\delta}{2} < \frac{j}{n}$

$$X_n(s) - X_n(t) = \frac{1}{\sigma\sqrt{n}}\left(S_{[ns]} - S_{[nt]}\right) + \frac{ns-[ns]}{\sqrt{n}\sigma} u_{[ns]+1} - \frac{nt-[nt]}{\sqrt{n}\sigma} u_{[nt]+1}$$

so $\displaystyle \sup_{t<s<t+\delta/2} = \max_{k<k+i<j} = \max_{0<i<j-k}$

and

$$\sup_{t<s<t+\frac{\delta}{2}} |X_n(s) - X_n(t)| < \max_{0<i<j-k} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k| + \max \frac{1}{\sqrt{n}\sigma} |u_{k+i}| + \max \frac{1}{\sqrt{n}\sigma} |u_{k+1}|$$

$$\leq 2 \max_{0<i<j-k} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k|$$

Thus we have

Lemma B (Billingsley Th^m 8.4, p 59)

$\{\bar{X}_n(r)\}$, the partial sum process, is tight if $\forall \varepsilon > 0$, $\exists \lambda > 1$ and n_0 s.th $\forall n \geq n_0$ we have

$$P\left(\max_{i \leq n} |S_{k+i} - S_k| > \lambda \sigma \sqrt{n}\right) < \frac{\varepsilon}{\lambda^2}, \forall k$$

Proof From Lemma A, $\bar{X}_n(r)$ is tight if (*) p.79 holds. In view of (**) on p.79 this is equivalent to

$$(*) \quad \frac{1}{\delta} P\left(\max_{0<i\leq j-k} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k| \geq \varepsilon\right) \leq \eta$$

Let $n > 4/\delta$ and since $\frac{j-1}{n} < t + \frac{\delta}{2} < \frac{j}{n}$ we have

$$j-1 < tn + \frac{\delta n}{2}$$

i.e.

$$j-1 < k+1 + \frac{\delta n}{2}$$

or

$$j-k < 2 + \frac{\delta n}{2}$$

$$\Rightarrow \quad j-k < \delta n$$

when $n \geq 4/\delta$

And so

$$\max_{0<i\leq[n\delta]} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k| \geq \varepsilon \qquad [B]$$

$$\Uparrow$$

whenever

$$\max_{0<i\leq j-k} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k| \geq \varepsilon \qquad [A]$$

Hence

$$(**) \quad \frac{1}{\delta} P\left(\max_{0<i\leq[n\delta]} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k| \geq \varepsilon\right) \leq \eta \Rightarrow (*)$$

i.e. $P[A] < P[B] < \eta$

Next let $m = [n\delta]$, then $(**)$ is

$$P\left(\max_{0<i\leq m} |S_{k+i} - S_k| \geq \varepsilon\sigma\sqrt{\frac{m}{\delta}}\right) \leq \eta\delta$$

and putting $\lambda = \frac{\varepsilon}{\sqrt{\delta}}$, which will be large if δ is small, we get

$$P\left(\max_{0<i\leq m} |S_{k+i} - S_k| > \lambda\sigma\sqrt{m}\right) \leq \frac{\eta\varepsilon^2}{\lambda^2}$$

giving the required result

Remark on Lemma A

- Let $w(X_n, \delta) = \sup_{|s-t|<\delta} |X_n(s) - X_n(t)|$ for $0 < \delta \leq 1$

be the "modulus of continuity" of $X_n(t)$ in $C[0,1]$, which controls fluctuations in X_n.

- Then the key requirement $(*)$ in Lemma A is that

$\forall\ \varepsilon, \eta > 0 \quad \exists\ \delta$ with $0 < \delta < 1$ s.th.

$$P\left(w(X_n, \delta) \geq \varepsilon\right) \leq \eta \quad \forall n > n_0 \text{ for some } n_0$$

i.e. modulus of continuity is "small" $(<\varepsilon)$ with probability arbitrarily $(\eta>0)$

Time Series Extensions of FCLT

- We now proceed to extend the FCLT so that it applies in a time series context. This was originally done in Billingsley's (1968) book using mixing processes & fns of mixing processes (NED seq's); and a general theory developed by McLeish (1974, 77, AP) and Herndorff (1984, 85, AP AS) in the prob. literature has proved useful in econometrics. The MG approx^{n} approach is also possible and is given in Hall & Heyde (1980).

- We shall use the linear process BN decomposition approach of Phillips-Solo (1992), i.e. suppose

(i) $X_t = C(L)\varepsilon_t \qquad C(L) = \sum_0^\infty c_j L^j$

$$\sum_0^\infty |j|^{1/2}|c_j| < \infty$$

(ii) $\varepsilon_t = iid(0, \sigma^2)$

Theorem (Phillips & Solo)

$$\frac{1}{\sqrt{n}}\sum_1^{[nr]} X_t \Rightarrow B(r) \equiv BM(\omega^2) \qquad \omega^2 = 2\pi f_{xx}(0)$$

Proof

Under (i) & (ii) we have the BN decomposition

$$X_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t \qquad \tilde{\varepsilon}_t = \tilde{C}(L)\varepsilon_t$$

with $\tilde{C}(L) = \sum_0^\infty \tilde{c}_j L^j$, $\tilde{c}_j = \sum_{j+1}^\infty c_s$

so that

$$(*) \quad \frac{1}{\sqrt{n}}\sum_1^{[nr]} X_t = C(1)\frac{1}{\sqrt{n}}\sum_1^{[nr]}\varepsilon_t + \frac{1}{\sqrt{n}}\left(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{[nr]}\right)$$

Now the first member on RHS of (*) satisfies the FCLT for iid sequences, i.e.

$$\frac{1}{\sqrt{n}\sigma}\Sigma_1^{[nr]}\varepsilon_t \Rightarrow W(r) = BM(1)$$

so that

$$C(1)\frac{1}{\sqrt{n}}\Sigma_1^{[nr]}\varepsilon_t \Rightarrow \sigma C(1)W(r) = B(r) \equiv BM(\omega^2)$$

with $\omega^2 = \sigma^2 C(1)^2 = 2\pi f_{xx}(0)$, where $f_{xx}(\lambda)$ is the spectrum of X_t.

To prove the invariance principle (IP) for X_t we now need only show that the second member on RHS of (*) is negligible as $n \to \infty$. Since (*) is a function in $D[0,1]$ this requires that the <u>distance between the functions</u> $\xrightarrow{p} 0$ ie in Skorohod topology

$$d_B\left(\frac{1}{\sqrt{n}}\Sigma_1^{[nr]}X_t - C(1)\frac{1}{\sqrt{n}}\Sigma_1^{[nr]}\varepsilon_t,\ \underset{O(r)}{\text{zero function}}\right) \xrightarrow{p} 0$$

ie.

$$(*) \qquad d_B\left(\frac{1}{\sqrt{n}}(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{[nr]}),\ O(r)\right) \xrightarrow{p} 0$$

Note that

$$0 \le d_B(f,g) \le \sup_t |f(t) - g(t)|$$

so that (*) holds necessarily if

$$\sup_r \left|\frac{1}{\sqrt{n}}(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{[nr]})\right| \xrightarrow{p} 0$$

Now

$$\sup_r \left|\frac{1}{\sqrt{n}}(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{[nr]})\right| \le \frac{1}{\sqrt{n}}|\tilde{\varepsilon}_0| + \sup_r \left|\frac{\tilde{\varepsilon}_{[nr]}}{\sqrt{n}}\right|$$

Clearly, $\tilde{\varepsilon}_0/\sqrt{n} \xrightarrow{p} 0$ as $\mathrm{var}(\tilde{\varepsilon}_0) = \sigma^2\Sigma\tilde{c}_j^2/n \to 0$

and

$$\sup_r \frac{1}{\sqrt{n}}|\tilde{\varepsilon}_{[nr]}| = \max_{0\le k\le n} \frac{1}{\sqrt{n}}|\tilde{\varepsilon}_k|$$

Under (i) & (ii) $\tilde{\varepsilon}_k$ is stationary & in L_2 so that

$$P\left(\max_k \left|\frac{\tilde{\varepsilon}_k}{\sqrt{n}}\right| \geq \eta\right) = P\left(\max_k \frac{\tilde{\varepsilon}_k^2}{n} \geq \eta^2\right)$$

$$\leq \sum_{k=0}^{n} P\left(\frac{1}{n}\tilde{\varepsilon}_k^2 \geq \eta^2\right)$$

as if $\max_k \frac{\tilde{\varepsilon}_k}{n} > \eta^2$ then at least one of $\varepsilon_k/n > \eta^2$ $k = 0, .., n$

(A) $\Rightarrow$ (B) , $P[A] < P(B)$

$$= (n+1)\, P\left(\frac{1}{n}\tilde{\varepsilon}_0^2 \geq \eta^2\right)$$

by stationarity

$$\leq \frac{n+1}{n\eta^2} E\left[\tilde{\varepsilon}_0^2\, 1\left(|\tilde{\varepsilon}_0| > \sqrt{n}\,\eta\right)\right]$$

$\longrightarrow 0$ as $n \to \infty$

since $E(\tilde{\varepsilon}_0^2) < \infty$

$\longrightarrow 0$ as $n \to \infty$

Hence

$$\max_k \left|\frac{\tilde{\varepsilon}_k}{\sqrt{n}}\right| \xrightarrow{p} 0$$

and (*) holds, as required.

Remark(1)

(1) Since $\sup_r \left| \frac{1}{\sqrt{n}} \sum_1^{[nr]} X_t - C(1) \frac{1}{\sqrt{n}} \sum_1^{[nr]} \varepsilon_t \right| \xrightarrow{p} 0$ (Z)

we do not need to worry about the tightness of $\frac{1}{\sqrt{n}}\sum_1^{[nr]} X_t$. We already know that $\frac{1}{\sqrt{n}}\sum_1^{[nr]}\varepsilon_t$ is tight and this ensures that $\frac{1}{\sqrt{n}}\sum_1^{[nr]} X_t$ is tight because of (Z). This avoids the extra burden of having to show that $\frac{1}{\sqrt{n}}\sum_1^{[nr]} X_t$ is tight, which is what we would need to do if we used a (i) fidi cgce + (ii) tightness approach to establishing the FCLT for $\frac{1}{\sqrt{n}}\sum^{[nr]} X_t$

Remark (2) The above result can be generalized in a wide variety of ways – see Phillips & Solo (92) for details. For instance, if ε_t is an mds with dominating r.v. ε for which

(#) $$P(|\varepsilon_t| \geq x) \leq c\, P(|\varepsilon| \geq x)$$

and $$E(|\varepsilon|^{2+\delta}) < \infty \quad \text{for some } \delta > 0$$

then

$$\frac{1}{\sqrt{n}} \sum_1^{[nr]} X_t \Rightarrow \sigma_\varepsilon C(1) W(r)$$

where

$$\frac{1}{n} \sum_1^n E(\varepsilon_t^2 | \mathcal{F}_{t-1}) \rightarrow_{a.s} \sigma_\varepsilon^2$$

(ie. σ_ε^2 is limit of average conditional variance of ε_t)

Note A sequence satisfying (#) is said to be <u>strongly uniformly integrable</u> (sui)

Thus we can handle heterogeneous innovations with no difficulty in this theory by imposing a very mild 'sui' condition like (#). Series like X_t with i.n.i.d innovations ε_t can be handled in the same way.

Remark (3) In proving the Theorem we implicitly use the following result on a separable metric space (D, d)

Lemma If $X_n \Rightarrow X$ and $d(X_n, Y_n) \rightarrow_p 0$ then $Y_n \Rightarrow X$

(eg. Billingsley, p.25)

Note that on a <u>separable</u> metric space $d(X_n, Y_n)$ is a (measurable) r.v. This follows because the product space is separable and $d: D \times D \rightarrow \mathbb{R}$ is cts map (Billingsley p225)

The Continuous Mapping Theorem (cmt)

- Weak cgce like $X_n \Rightarrow X$ or $P_n \Rightarrow P$ is preserved under continuous mappings from the original metric space to another metric space. The most popular / common maps take a general metric space like $C[0,1]$, $D[0,1]$ into Euclidean space $(\mathbb{R}, \mathbb{R}^k)$ and correspond to functionals that take random functions into Euclidean space r.v.'s eg.

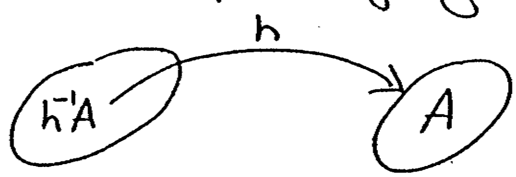
$$\int_0^1 X_n(r)\,dr \ , \quad \sup_{r\in[0,1]} X_n(r) \ , \quad \inf_{r\in[0,1]} X_n(r) \quad \text{etc}$$

- Let h be a mble map from $(D,d) \to (D',d')$
with Borel σ fields: $\mathcal{D}$ $\mathcal{D}'$

Then corresponding to every prob. measure P on $(D,\mathcal{D})$ h induces a prob measure Ph^{-1} on $(D',\mathcal{D}')$ as follows

$$P(h^{-1}(A)) = P(h^{-1}A) \quad \forall A \in \mathcal{D}'$$

here $h^{-1}A$ = preimage of $A \in \mathcal{D}'$ under h



$\in \mathcal{D}$ $\qquad\qquad$ $\in \mathcal{D}'$

$(D, \mathcal{D}, P)$ $\qquad\qquad$ $(D, \mathcal{D}', Ph^{-1})$

eg. $(\mathbb{R}, \mathcal{B}, Ph^{-1})$

The cmt theorem tells us that if h is cts

$$P_n \Rightarrow P \quad \text{implies} \quad P_n h^{-1} \Rightarrow Ph^{-1}$$

<u>Theorem</u> (cmt) Suppose

(i) $h : D \to D'$ is continuous

(ii) $P_n \Rightarrow P$ on $(D, \mathcal{D})$

then $P_n h^{-1} \Rightarrow P h^{-1}$ on $(D', \mathcal{D}')$

<u>Proof</u>

$$P_n h^{-1}(F) = P_n(h^{-1}F)$$

$$\to P(h^{-1}F) \quad \left\{ \begin{array}{l} \text{by weak cgce of } P_n \\ \text{for all } P\text{-cty sets} \\ \text{i.e. all sets for which} \\ P(\partial h^{-1}F) = 0 \end{array} \right.$$

$$= P h^{-1}(F) \quad \text{by def}^n \text{ of } Ph^{-1}$$

and $Ph^{-1}(\partial F) = P(h^{-1}\partial F)$
$= P(\partial h^{-1} F)$
$= 0$

so it holds

$\forall \; Ph^{-1}$ - cty sets F

<u>Note (1)</u>

$$h^{-1}\partial F = \partial h^{-1} F$$

since $x \in h^{-1}\partial F$ iff $h(x) \in \partial F$ iff $h(x) \in \overline{F}, \overline{F^c}$

iff $x \in h^{-1}\overline{F}, h^{-1}\overline{F^c}$

iff $x \in \overline{h^{-1}F}, \overline{h^{-1}F^c}$ as h cts & takes closed sets into closed sets

iff $x \in \partial h^{-1} F$

<u>Note (2)</u> The following so-called portmanteau theorem of weak cgce is useful

<u>Th$^{\underline{m}}$</u> (Billingsley p.11) Let P_n, P be measures on $(D, \mathcal{D})$

$P_n \Rightarrow P$ iff any of the following equivalent conditions hold:

(i) $P_n(F) \to P(F) \quad \forall \; P$-cty sets $F \in \mathcal{D}$

(ii) $\limsup_n P_n(F) \leq P(F) \quad \forall$ closed sets F

(iii) $\liminf_n P_n(F) \geq P(F) \quad \forall$ open sets F

(iv) $\int f \, dP_n \xrightarrow{n} \int f \, dP \quad \forall$ bdd uniformly cts real functions f

space of measures on $(D, \mathcal{D})$ (defines weak topology on

<u>Note (3)</u> Using (iv) we get the equivalent argument for weak cgce of $P_n h^{-1} \Rightarrow P h^{-1}$, viz

$$P_n \Rightarrow P \quad \text{implies} \quad \int f\, dP_n \to \int f\, dP \quad \forall \text{ bdd cts } f$$

$$\text{implies} \quad \int f(h(x))\, P_n(dx) \to \int f(h(x))\, d$$

$$\text{i.e.} \quad \int f\, P_n h^{-1}(dx) \to \int f\, P h^{-1}(dx)$$

$$\text{i.e.} \quad P_n h^{-1} \Rightarrow P h^{-1}$$

<u>Note (4)</u> The continuity assumption on h can be relaxed provided the set of discontinuities is negligible in the limit (P) measure. Thus:

<u>Theorem</u> (cmt) If $P_n \Rightarrow P$ and $P(D_h) = 0$, where D_h = set of discontinuities of h, then

$$P_n h^{-1} \Rightarrow P h^{-1}$$

<u>Proof</u>. Use criterion (ii) for weak cgce – then we need only show that for any closed set $F \in \mathcal{D}'$

$$\limsup_{n \to \infty} P_n h^{-1}(F) \le P h^{-1}(F)$$

We are given that $P_n \Rightarrow P$ so we do have

$$\limsup_{n \to \infty} P_n(h^{-1}F) \le \limsup_{n \to \infty} P_n\left(\overline{h^{-1}F}\right)$$

$$\le P\left(\overline{h^{-1}F}\right)$$

Now note ~~that~~ $\overline{h^{-1}F} \subseteq h^{-1}F \cup D_h$, i.e. a closure point of $h^{-1}F$ is either in $h^{-1}F$ (since F is closed) or in D_h, a point of discontinuity. So $P(\overline{h^{-1}F}) \le P(h^{-1}F) + P(D_h)$
$= P(h^{-1}F)$

Hence by criterion (ii) $P_n h^{-1} \Rightarrow P h^{-1}$

Examples of Continuous functionals

(i) $h : C[0,1] \to \mathbb{R}$ defined by $h(f) = \int_0^1 f(r)\,dr$

h is a cts functional because for any sequence $f_n \to f$ in $C[0,1]$ i.e. $d_u(f_n, f) \to 0$ we have

$$
\begin{aligned}
|h(f_n) - h(f)| &= \left|\int_0^1 (f_n - f)\,dr\right| \\
&\le \int_0^1 |f_n - f|\,dr \\
&\le \sup_r |f_n - f| \to 0
\end{aligned}
$$

as $d_u(f_n, f) = \sup_r |f_n - f| \to 0$

(ii) $h : D[0,1] \to \mathbb{R}$ defined by $h(f) = \int_0^1 f(r)\,dr$

(note all fncs in $D(0,1)$ are ible as they are cts. except for at most a countable set of pts

Now $f_n \to f$ in $(D[0,1], d_B)$ requires $d_B(f_n, f) \to 0$ and this implies that $\exists\, \lambda_n \in \Lambda$ s.th

$(*)\ \sup_r |f_n(\lambda_n(r)) - f(r)| \to 0$ and $\sup_r |\lambda_n(r) - r| \to 0\ (**)$

Now as above in (i)

$$
|h(f_n) - h(f)| \le \int_0^1 |f_n - f|\,dr \le \int_0^1 |f_n(r) - f_n(\lambda_n(r))|\,dr + \int_0^1 |f_n(\lambda_n(r)) - f(r)|\,dr
$$

we have:

(a) $\int_0^1 |f_n(\lambda_n(r)) - f(r)|\,dr \le \sup_r |f_n(\lambda_n(r)) - f(r)| \to 0$ by $(*)$

(b) $\int_0^1 |f_n(r) - f_n(\lambda_n(r))|\,dr = \sum_{i=0}^{k} \int_{t_{i-1}}^{t_i} |f_n(r) - f_n(\lambda_n(r))|\,dr$

$$
\le \sum_{i=0}^{k} \int_{t_{i-1}}^{t_i} \sup_r |f_n(r) - f_n(\lambda_n(r))|\,dr + N \sup_{r \in [0,1]} f_n(r) \sup_r |\lambda_n(r) - r|
$$

grid over which max dev $\le \varepsilon$ $\longrightarrow 0$

$< \varepsilon$ except for finite # of pts at ends of grid

finite # of pts where jump in $f_n \ge \varepsilon$ (see Th$^{\underline{m}}$ p. 60 above)

bdd

$\longrightarrow 0$ by $(**)$

First Application of FCLT & CMT

Sample mean of an I(1) process

$$\Delta X_t = u_t = C(L)\,\varepsilon_t$$

i.e.

$$X_t = X_{t-1} + u_t = \dots = \Sigma_1^t u_j + X_0$$

Let X_0 = finite variance r.v.

$$u_t = C(1)\,\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t \qquad \text{BN decomposition valid}$$

Theorem

$$\frac{1}{n^{3/2}} \Sigma_1^n X_t \Rightarrow \int_0^1 B(r)\,dr \qquad B(r) \equiv BM(\omega^2)$$

$$\omega^2 = 2\pi f_{uu}(0) = \sigma_\varepsilon^2\, C(1)^2$$

Proof

$$\Sigma_1^n X_t = \Sigma_{j=1}^n \left(S_{j-1} + u_j + X_0\right) \qquad S_k = \Sigma_1^k u_t$$

$$= \sqrt{n}\, \Sigma_{j=1}^n \left[\frac{1}{\sqrt{n}} S_{j-1}\right] + \Sigma_1^n u_j + n X_0$$

$$= n\sqrt{n} \sum_{j=1}^n \left[\int_{(j-1)/n}^{j/n} X_n(r)\,dr\right] + \Sigma_1^n u_j + n X_0$$

$$= n^{3/2} \int_0^1 X_n(r)\,dr + \Sigma_1^n u_j + n X_0$$

Hence

$$\frac{1}{n^{3/2}} \Sigma_1^n X_t = \int_0^1 X_n(r)\,dr + \frac{1}{n^{3/2}} \Sigma_1^n u_j + \frac{1}{\sqrt{n}} X_0$$

$$= \int_0^1 X_n(r)\,dr + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$\Rightarrow \int_0^1 B(r)\,dr$$

In effect (short-cut)

$$\frac{1}{n^{3/2}} \Sigma_1^n X_t = \frac{1}{n} \Sigma_1^n \frac{X_t}{\sqrt{n}} = \Sigma_1^n \frac{X_{[nr]}}{\sqrt{n}} \frac{1}{n}$$

$$\Rightarrow \int_0^1 B(r)\,dr$$

$$\left.\begin{array}{l} 1 \le t \le n \\ 0 \le r \le 1 \end{array}\right\} t = [nr]$$

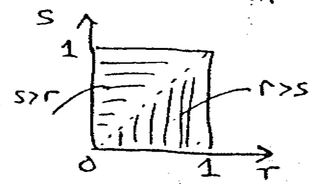
Limit Distribution $\int_0^1 B$

$$\int_0^1 B(r)\,dr \equiv N(0, v)$$

linear functional of Gaussian process are Gaussian

$$v = E\left(\int_0^1 B\right)^2$$

$$= 2 \int_0^1 \int_0^r E(B(r)\,B(s))\,ds\,dr$$



or $\displaystyle\int_0^1 \int_r^1 E(B(r)\,B(s))\,ds\,dr = \iint_{s>r} \cdots = \iint_{r>s} \cdots$

$$= \int_0^1 \int_r^1 \omega^2 \min(r,s)\,ds\,dr$$

$$= \int_0^1 \omega^2 r \int_r^1 ds\,dr = \omega^2 \int_0^1 (r - r^2)\,dr$$

$$= \omega^2 \left[\frac{r^2}{2} - \frac{r^3}{3}\right]_0^1$$

$$= \omega^2 \left(\frac{1}{2} - \frac{1}{3}\right)$$

$$= \omega^2 \frac{1}{6} \quad //$$

$$= 2\omega^2 \int_0^1 \int_0^r s\,ds\,dr$$

$$= 2\omega^2 \int_0^1 \frac{r^2}{2}\,dr$$

$$= 2\omega^2 \left.\frac{r^3}{6}\right|_0^1$$

$$= \frac{\omega^2}{3} \quad //$$

Note 1 $\quad E(B(r)\,B(s)) = \omega^2\, r \wedge s = \omega^2 \min(r,s)$

Note 2 In an entirely similar way to the theorem we get

$$\frac{1}{n^{3/2}} \sum_1^{[nr]} X_t \Rightarrow \int_0^r B(s)\,ds = \int_0^r B \quad \text{short-hand}$$

or

$$\left.\begin{aligned} \frac{1}{n^{3/2}} \sum_1^{[n\cdot]} X_t &\Rightarrow \int_0^{\cdot} B \\ X_{[n\cdot]} &\Rightarrow B(\cdot) \end{aligned}\right\} \text{another short-hand notation}$$

Sample variance of an I(1) process

Theorem

(a) $\frac{1}{n^2}\sum_1^n X_t^2 \Rightarrow \int_0^1 B(r)^2 dr$

(b) $\frac{1}{n^2}\sum_1^{[nr]} X_t^2 \Rightarrow \int_0^r B(s)^2 ds$

(c) $\frac{1}{n^2}\sum_1^n (X_t-\bar{X})^2 \Rightarrow \int_0^1 B(s)^2 ds - \left(\int_0^1 B(s)ds\right)^2$

Proof:

(a) $$\frac{1}{n^2}\sum_1^n X_t^2 = \frac{1}{n}\sum_1^n \left(\frac{1}{\sqrt{n}}S_{j-1} + \frac{1}{\sqrt{n}}u_j + \frac{1}{\sqrt{n}}X_0\right)^2$$

$$= \frac{1}{n}\sum_{j=1}^n \left\{\int_{(j-1)/n}^{j/n} X_n(r)^2 dr + \frac{2}{n^2}S_{j-1}u_j + \frac{2}{n^2}S_{j-1}X_0 + \frac{2}{n^2}u_j X_0 + \frac{1}{n^2}u_j^2 + \frac{1}{n^2}X_0^2\right\} dr$$

$$= \int_0^1 X_n(r)^2 dr + o_p(1)$$

$$\to_d \int_0^1 B(r)^2 dr$$

or short-cut method:

$$\frac{1}{n^2}\sum_1^n X_t^2 = \frac{1}{n}\sum_1^n\left(\frac{X_t}{\sqrt{n}}\right)^2 = \sum_1^n\left(\frac{X_{[nr]}}{\sqrt{n}} + o_p(1)\right)^2\frac{1}{n}$$

$$= \int_0^1 (X_n(r) + o_p(1))^2 dr$$

$$\to \int_0^1 B(r)^2 dr$$

(b) $$\frac{1}{n^2}\sum_1^{[nr]} X_t^2 = \frac{1}{n}\sum_1^{[nr]}\left(\frac{X_{[ns]}}{\sqrt{n}} + o_p(1)\right)\frac{1}{n}$$

$$= \int_0^r (X_n(s) + o_p(1))^2 ds$$

$$\to \int_0^r B(s)^2 ds$$

(c) $$\frac{1}{n^2}\sum_1^n (X_t-\bar{x})^2 = \frac{1}{n^2}\sum_1^n X_t^2 - \frac{1}{n^3}\left(\sum_1^n X_t\right)^2 = \frac{1}{n}\sum_1^n\left(\frac{1}{\sqrt{n}}X_t\right)^2 - \left(\frac{1}{n}\sum_1^n \frac{1}{\sqrt{n}}X_t\right)^2$$

$$\Rightarrow \int_0^1 B(r)^2 dr - \left(\int_0^1 B(r)dr\right)^2$$

Vector Brownian Motion, Product Spaces & the Multivariate FCLT

Much of our analysis in time series is with random vectors in $\mathbb{R}^k$. For I(1) vector time series this requires us to transform $\mathbb{R}^k$ random vectors into k-vector functions. Hence we work with the product spaces

- $C[0,1]^k = \overset{k}{\underset{1}{\times}} C[0,1]$ k Cartesian copies of $C[0,1]$

with typical element $x(t) = \begin{bmatrix} x_1(t) \\ x_k(t) \end{bmatrix} \in C[0,1]^k$

<u>metric</u>

$$d_u^k(f,g) = \max_{1 \le j \le k} d_u(f_j, g_j)$$

induces Borel σ-field $\mathcal{B}^k$

<u>space</u> $\left(C[0,1]^k, \mathcal{B}^k\right)$ separable & complete

<u>weak cgce</u>

$$P_n = P_n^1 \times \ldots \times P_n^k$$ product measure on $(C[0,1]^k, \mathcal{B}^k)$

$$P_n \Rightarrow P = P^1 \times \ldots \times P^k$$

iff $P_n^i \Rightarrow P^i$ on $(C[0,1], \mathcal{B})$

(because $C[0,1]^k$ is separable – Billingsley p21)

<u>tightness</u>

$\{P_n\}$ on $C[0,1]^k$ are tight iff all marginal measures ($P_n^1(A) = P_n(A \times C[0,1]^{k-1})$

Multivariate Brownian motion

(i) $W(t) = \begin{bmatrix} W_1(t) \\ W_k(t) \end{bmatrix} \in C[0,1]^k$

- $W_i(t) \equiv BM(1)$ on $C[0,1]$ $\forall i = 1, \dots, k$
- W_i & W_j independent $i \neq j$

$W(t) \equiv N(0, t I_k)$ fidi dist's

$W(t) = BM(I_k)$ notation

(ii) $B(t) = \begin{bmatrix} B_1(t) \\ B_k(t) \end{bmatrix} \in C[0,1]^k$

- $B_i(t) \equiv BM(\omega_{ii})$ on $C[0,1]$ $\forall i = 1, \dots, k$
- $E\{B_i(t) B_j(t)\} = \omega_{ij} t$ $\forall i, j$
- $E\{B(t) B(t)'\} = \Omega t$ $\Omega = (\omega_{ij})$

$B(t) \equiv BM(\Omega)$

Multivariate FCLT

(i) $\{u_j\} \equiv iid(0, \Omega)$ Ω $(k \times k)$

$$X_n(r) = \frac{1}{\sqrt{n}} \Omega^{-1/2} \sum_1^{[nr]} u_j \Rightarrow W(r) \equiv BM(I_k)$$

$$Y_n(r) = \frac{1}{\sqrt{n}} \sum_1^{[nr]} u_j \Rightarrow B(r) \equiv BM(\Omega)$$

Proof

(a) fidi's of $\lambda' Y_n(r) \underset{fd}{\Longrightarrow} \lambda' B(r)$; apply Cramer Wold device to get $Y_n(r)$ fidi

(b) tightness of $\{Y_n(r)\}$ follows from tightness of all marginals

$$= \frac{1}{\sqrt{n}} \sum_1^{[nr]} \lambda' u_i \Rightarrow BM(\lambda' 2\pi f_{uu}(0) \lambda)$$

(ii) $\begin{cases} \varepsilon_t \equiv iid(0,\Sigma), & \sum_0^\infty j^{1/2}\|C_j\| < \infty \\ u_t = C(L)\varepsilon_t = C(1)\,\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t \end{cases}$

$$X_n(r) = \frac{1}{\sqrt{n}} \sum_1^{[nr]} u_t \Rightarrow B(r) \equiv BM(\Omega)$$

$$\Omega = C(1)\,\Sigma\, C(1)' = 2\pi f_{uu}(0)$$

$$= lrvar(u_t)$$

<u>Proof</u> Entirely analogous to the scalar case considered on p. 82.

$D[0,1]^k = \overset{k}{\underset{1}{\times}} D[0,1]$ k Cartesian copies of $D[0,1]$

with typical element $x(t) = \begin{bmatrix} x_1(t) \\ \\ x_k(t) \end{bmatrix}$, $x_i \in D[0,1]$

<u>metric</u>

$$d_B^k(f,g) = \max_{1 \le j \le k} d_B(f_j, g_j)$$

induces Borel σ-field $\mathcal{D}^k$

<u>space</u> $(D[0,1]^k, \mathcal{D}^k)$ separable & complete

<u>Example</u> $X_n(r) = \frac{1}{\sqrt{n}} \sum_1^{[nr]} u_j$ as in (i) & (ii) above

$$\in D[0,1]^k$$

The FCLTe result above for $X_n(r)$ is an FCLT on $(D[0,1]^k, \mathcal{D}^k)$, but the limit process is $B(r) \in C[0,1]^k$. See Phillips–Durlauf, 1986, for detail of fidi's, fidi cgce & tightness in $D[0,1]^k$ (aided by separability of space under d_B^k metric)

Further Applications: Vector I(1) Analysis

[1] I(1) vector process: $\Delta X_t = u_t = C(L)\varepsilon_t$

$$\sum_0^\infty j^{1/2}\|C_j\| < \infty$$

$$\varepsilon_t \equiv iid(0, \Sigma_\varepsilon)$$

Theorem

(i) $\frac{1}{n^{3/2}} \sum_1^{[nr]} X_t \Rightarrow \int_0^r B(s)ds = \int_0^r B \qquad B \equiv BM(\Omega)$

$$\Omega = 2\pi f_{uu}(0) = C(1)\Sigma_\varepsilon C(1)'$$

(ii) $\frac{1}{n^2} \sum_1^{[nr]} X_t X_t' \Rightarrow \int_0^r BB'$

(iii) $\frac{1}{n^2} \sum_1^n (X_t - \bar{X})(X_t - \bar{X})' \Rightarrow \int_0^1 BB' - \left(\int_0^1 B\right)\left(\int_0^1 B'\right)$

Proof

Analogous to the scalar case, e.g.

$$\frac{1}{n^2}\sum_1^{[nr]} X_t X_t' = \frac{1}{n}\sum_1^{[nr]} \frac{X_t}{\sqrt{n}}\frac{X_t'}{\sqrt{n}}$$

$$= \sum_{j=1}^{[nr]} \int_{(j-1)/n}^{j/n} X_n(s) X_n(s)' ds + o_p(1)$$

$$X_n(s) = \frac{X_{[ns]}}{\sqrt{n}} \qquad \frac{j-1}{n} \le s < j/n$$

$$= \int_0^{[nr]/n} X_n(s) X_n(s)' ds + o_p(1)$$

$$\Rightarrow \int_0^r B(s) B(s)' ds$$

[2] I(2) process: $\Delta^2 X_t = u_t = C(L)\varepsilon_t$

i.e. $\Delta X_t = \sum_1^t u_s + d_0$

$$X_t = \sum_{j=1}^t \sum_1^j u_s + d_0 t + c_0$$

$$\frac{1}{n^{3/2}} X_t = \frac{1}{n} \Sigma_{j=1}^{t} \frac{1}{\sqrt{n}} \Sigma_1^j u_s + d_0 \frac{t}{n^{3/2}} + \frac{c_0}{n^{3/2}}$$

Let $t = [nr]$, then

$$\frac{1}{n^{3/2}} X_{[nr]} = \frac{1}{n} \Sigma_{j=1}^{[nr]} \frac{1}{\sqrt{n}} \Sigma_1^j u_s + o_p(1)$$

$$= \sum_{j=1}^{[nr]} \int_{(j-1)/n}^{j/n} U_n(p)\,dp + o_p(1)$$

$$U_n(p) = \frac{1}{\sqrt{n}} \Sigma_1^{[np]} u_s$$

$$\frac{j-1}{n} \le p < j/n$$

$$= \int_0^{[nr]/n} U_n(p)\,dp + o_p(1)$$

Thus

(iv) $\frac{1}{n^{3/2}} X_{[nr]} \Rightarrow \int_0^r B \qquad B = BM(\Omega)$

Similarly,

(v) $\frac{1}{n^4} \Sigma_1^{[nr]} X_t X_t' \Rightarrow \int_0^r \underline{B}\, \underline{B}'$

where $\underline{B}(s) = \int_0^s B(p)\,dp$

outline:

$$\frac{1}{n^4} \Sigma_1^{[nr]} X_t X_t' = \frac{1}{n} \Sigma_1^{[nr]} \left(\frac{1}{n^{3/2}} X_t\right)\left(\frac{1}{n^{3/2}} X_t'\right)$$

$$\rightarrow \int_0^r \underline{B}(s)\, \underline{B}(s)'\, ds$$

And

(vi) $\frac{1}{n^4} \Sigma_1^n (X_t - \bar{X})(X_t - \bar{X}) \Rightarrow \int_0^1 \underline{B}\, \underline{B}' - \left(\int_0^1 \underline{B}\right)\left(\int_0^1 \underline{B}'\right)$

Remark. • c_0, d_0 do not enter limit theory
• try $I(3), I(4)$ process limits

Now go back to an $I(1)$ process

$\Delta X_t = u_t = C(L)\varepsilon_t$ and consider interactions of trends and $I(1)$ processes, viz

(vii) $\frac{1}{n^{5/2}}\sum_1^n t X_t \Rightarrow \int_0^1 r B(r)dr$

<u>Proof</u> Same approach, viz

$$\frac{1}{n^{5/2}}\sum_1^n t X_t = \frac{1}{n}\sum_1^n \frac{t}{n}\frac{1}{\sqrt{n}}X_t$$

$$= \sum_1^n \int_{(j-1)/n}^{j/n} \frac{[nr]}{n} X_n(r)\,dr + o_p(1)$$

$$= \int_0^1 r X_n(r)\,dr + o_p(1)$$

$$\Rightarrow \int_0^1 r B(r)\,dr$$

Similarly if $X_t \equiv I(2)$ we have

(viii) $\frac{1}{n^{7/2}}\sum_1^n t X_t \Rightarrow \int_0^1 r \underline{B}(r)dr$

as

$$\frac{1}{n^{7/2}}\sum_1^n t X_t = \frac{1}{n}\sum_1^n \frac{t}{n}\frac{X_t}{n^{3/2}} \to \int_0^1 r \underline{B}$$

Hilbert Projections in $L_2[0,1]$

We work in the space of square ible functions on $[0,1]$, viz

$$L_2[0,1] = \{ f \mid \int f^2 < \infty \}$$

Define

$$1(r) \in L_2[0,1] \quad \text{by} \quad 1(r) = 1 \quad \forall r \in [0,1]$$

$$j(r) \in L_2[0,1] \quad \text{by} \quad j(r) = r^{j-1} \quad \forall r \in [0,1]$$

Recall that projections P in L_2 are defined by two properties:

(i) $P^2 = P$ <u>idempotent</u>

(ii) $(f, Pg) = (Pf, g)$ i.e. $P = P^*$ <u>self adjoint</u>

$$\forall f, g \in L_2$$

<u>Example</u>

$$P_1 f(r) = \left(\int_0^1 f \right) 1(r)$$

defines the operator P_1 by its action on the arbitrary element $f \in L_2[0,1]$

Note that:

(i) $P_1^2 f = P_1(P_1 f) = P_1\left(\int_0^1 f \; 1(r)\right) = \int_0^1 f \int 1 \; 1(r)$

$$= \int_0^1 f \; 1(r)$$

$$= P_1 f \quad \forall f$$

$$\boxed{\text{so } P_1^2 = P_1}$$

(ii) $(f, P_g) = \int_0^1 f \, Pg \, dr$

$$= \int_0^1 f(r) \left(\int_0^1 g(s)\right) 1(r) \, dr$$

$$= \left(\int_0^1 g(s) ds\right)\left(\int_0^1 f(r) dr\right)$$

$$= \int_0^1 P_1 f \; g(s) ds = (P_1 f, g)$$

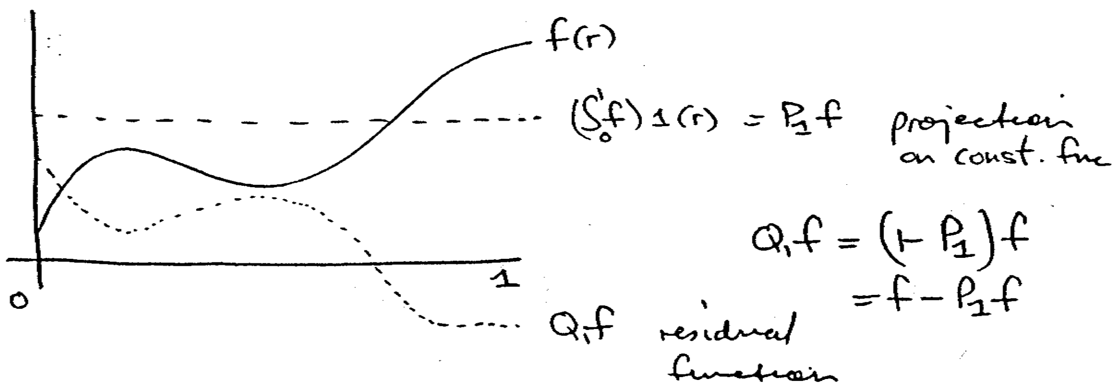
$$\boxed{\text{so } P_1^* = P_1}$$

Orthogonal decomposition of $f \in L_2[0,1]$

$$
\begin{aligned}
f_1 &= P_1 f + (1 - P_1) f \\
&= P_1 f + Q_1 f
\end{aligned}
$$

$P_1 f \perp Q_1 f$ as

$$
\begin{aligned}
(P_1 f, Q_1 f) &= \int_0^1 P_1 f \, Q_1 f \, dr = \int_0^1 P_1 f \, (f - P_1 f) \, dr \\
&= (P_1 f, f) - (P_1 f, P_1 f) \\
&= (P_1 f, f) - (P_1 f, f) \\
&= (P_1 f)^2 - (P_1 f)^2 \\
&= 0 .
\end{aligned}
$$



Application to $B(r) \in C[0,1]$

Note that $B(r) \in L_2[0,1]$ and Wiener measure induces a measure on the space $L_2[0,1]$ for which $P(f \in C[0,1]) = 1$, i.e.

Note

$$
\begin{aligned}
\underline{B}(r) &= B(r) - \int_0^1 B \\
&= Q_1 B
\end{aligned}
$$

to constant fnc $1(r)$

$= \| \cdot \|$ . . . projection of B . . .

$$P_1 B = \left(\int_0^1 B\right) 1(r)$$ constant fn at level = mean of $B(r)$

In effect, $\underline{B}(r) = (1 - P_1)B$ = regression residual from the cts time OLS regression

$$\min_{\alpha} \int_0^1 (B(r) - \alpha)^2 dr$$

which leads to OLS estimator

$$\hat{\alpha} = \int_0^1 B(r)dr = \text{sample mean of } B(r)$$

so that we have the

<u>regression relation</u>

$$B(r) = \hat{\alpha} + \eta(r), \quad \text{with}$$

$$\eta(r) = B(r) - \int_0^1 B(r)dr = \underline{B}(r)$$

<u>Detrended Brownian Motion</u>

We can readily extend this idea to general polynomial functions $j(r) = r^j$ in $L_2[0,1]$.

Define the Hilbert projection

$$\underline{B}_p = QB = \text{projection of } B \text{ in } L_2[0,1]$$ onto the orthogonal complement of space spanned by $\{j(r) = r^j \,;\, j = 0, \dots, p\}$

Then

$$\underline{B}_p = \text{detrended BM}$$
$$= B(r) - \hat{\alpha}_0 - \hat{\alpha}_1 r - \dots - \hat{\alpha}_p r^p$$

where $\hat{\alpha}_i$ $(i=0,1,\dots,p)$ minimize the L_2 distance, i.e.

$$(*) \quad \min_{\alpha} \int_0^1 \left(B(r) - \alpha_0 - \alpha_1 r - \dots - \alpha_p r^p\right)^2 dr$$

$$= \min_{\alpha} \int_0^1 \left(B(r) - X(r)'\alpha\right)^2 dr \quad \text{say}$$

FOC are: $\int_0^1 d\alpha' X(r)\left[B(r) - X(r)'\alpha\right] dr = 0$

gives

$$(**) \quad \hat{\alpha} = \left[\int_0^1 X(r)X(r)'\right]^{-1}\left[\int_0^1 X(r)B(r)\right]$$

eg.

$p = 0 \qquad \hat{\alpha}_0 = \int_0^1 B(r)dr \qquad X(r) = 1(r)$

$$p = 1 \qquad \begin{bmatrix}\hat{\alpha}_0 \\ \hat{\alpha}_1\end{bmatrix} = \begin{bmatrix}1 & \int_0^1 s\,ds \\ \int_0^1 s\,ds & \int_0^1 s^2 ds\end{bmatrix}^{-1}\begin{bmatrix}\int_0^1 B(s)ds \\ \int_0^1 s\,B(s)ds\end{bmatrix}$$

$$= \begin{bmatrix}1 & 1/2 \\ 1/2 & 1/3\end{bmatrix}^{-1}\begin{bmatrix}\int_0^1 B(s)ds \\ \int_0^1 s\,B(s)ds\end{bmatrix}$$

$$= \begin{bmatrix}1/3 & -1/2 \\ -1/2 & 1\end{bmatrix}\begin{pmatrix}\int B \\ \int sB\end{pmatrix} 12$$

$$= \begin{bmatrix}1/3 \int B & -1/2 \int sB \\ -1/2 \int B & + \int sB\end{bmatrix} 12 = \begin{bmatrix}4\int B - 6\int sB \\ -6\int B + 12\int sB\end{bmatrix}$$

Here

$$QB = B(r) - \hat{\alpha}_0 - \hat{\alpha}_1 r$$

$$= B(r) - \left(4\int_0^1 B - 6\int_0^1 sB\right) - \left(12\int_0^1 sB - 6\int_0^1 B\right)r$$

$$= \text{detrended BM}$$

Form of PB, QB in General case

From (*) above we deduce that

$$PB = \hat{\alpha}' X(r) = \left(\int_0^1 B(s)X(s)'\right)\left(\int_0^1 X(s)X(s)'\right)^{-1} X(r)$$

= Hilbert projection of $B(r)$ onto space spanned by vector of functions $X(r)$.

c.f. $y'X(X'X)^{-1}X'$

$$QB = (1-P)B$$

$$= B(r) - \hat{\alpha}' X(r)$$

$$= B(r) - \left(\int_0^1 B(s)X(s)'\right)\left(\int_0^1 X(s)X(s)'\right)^{-1} X(r)$$

$$= B - \left(\int_0^1 B X'\right)\left(\int_0^1 X X'\right)^{-1} X$$

for short

c.f. $\hat{u}' = y' - y'X(X'X)^{-1}X'$

$= \underline{B}_X$, say. $\qquad = y'\left(I - X(X'X)^{-1}X'\right)$

Remark: The function space results of these projections are straightforward analogues of the Euclidean space projections & residuals in $\mathbb{R}^n$

Sample Moments for Filtered I(0) Series

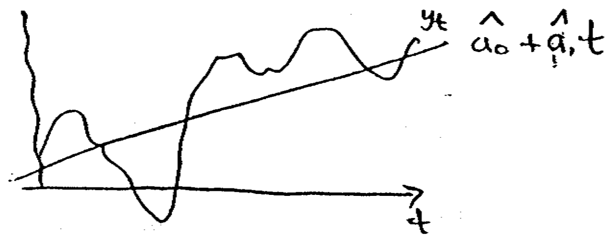
Let y_t be an I(0) series for which

$$\Delta y_t = u_t = C(L)\varepsilon_t \qquad \sum j^{1/2}\|C_j\| < \infty$$

Let $\underline{y}_t$ be the filtered series defined from the regression:

$$(1) \quad y_t = \hat{a}_0 + \hat{a}_1 t + \dots + \hat{a}_p t^p + \underline{y}_t$$

i.e. residual from a regression on trend (the detrended I(0) series)



Then the sample moments of $\underline{y}_t$ have the following limit behaviour:

$$\frac{1}{\sqrt{n}}\,\underline{y}_{[nr]} \Rightarrow \underline{B}_p(r)$$

$$\frac{1}{n^{3/2}}\sum_1^{[nr]} \underline{y}_t \Rightarrow \int_0^r \underline{B}_p$$

$$\frac{1}{n^2}\sum_1^{[nr]} \underline{y}_t\underline{y}_t' \Rightarrow \int_0^r \underline{B}_p\underline{B}_p'$$

Proof: Write (1) as

$$\underline{y}_t = y_t - \hat{A}x_t \qquad \hat{A} = [\hat{a}_0\ \hat{a}_1\ \ \hat{a}_p]$$

$n \times (p+1)$ matrix

$$\hat{A} = Y'X(X'X)^{-1} \qquad x_t = \begin{pmatrix} 1 \\ t \\ t^p \end{pmatrix}$$

Write

$$\frac{1}{\sqrt{n}}\underline{y}_{[nr]} = \frac{1}{\sqrt{n}} y_{[nr]} - \frac{1}{\sqrt{n}}\hat{A}\, x_{[nr]}$$

$$= \frac{1}{\sqrt{n}} y_{[nr]} - \frac{1}{\sqrt{n}}\hat{A} D_n D_n^{-1} x_{[nr]}$$

where $D_n = \operatorname{diag}(1, n, \dots, n^p)$

<u>Note:</u>

$$D_n^{-1} x_{[nr]} = \begin{bmatrix} 1 \\ [nr]/n \\ \vdots \\ [nr]^p/n^p \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ r \\ \vdots \\ r^p \end{bmatrix} = X(r), \text{ say.}$$

$$\frac{1}{\sqrt{n}}\hat{A} D_n = \frac{1}{\sqrt{n}} Y'X (X'X)^{-1} D_n$$

$$= \left(\frac{1}{\sqrt{n}} \Sigma_1^n y_t x_t' D_n^{-1}\right)\left(D_n^{-1} \Sigma_1^n x_t x_t' D_n^{-1}\right)^{-1}$$

$$= \left(\frac{1}{n}\Sigma_1^n \frac{y_t}{\sqrt{n}} x_t' D_n^{-1}\right) D_n^{-1} \frac{1}{n}\Sigma_1^n x_t x_t' D_n^{-1}\Big)^{-1}$$

Consider

$$\frac{1}{n}\Sigma_1^n \frac{y_t}{\sqrt{n}} x_t' D_n^{-1} = \frac{1}{n}\sum_1^n \frac{y_{[nr]}}{\sqrt{n}} x_{[nr]}' D_n^{-1}$$

$$= \Sigma_1^n \int_{(j-1)/n}^{j/n} \frac{y_{[nr]}}{\sqrt{n}} x_{[nr]}' D_n^{-1}\, dr$$

$$= \int_0^1 \frac{y_{[nr]}}{\sqrt{n}} x_{[nr]}' D_n^{-1}\, dr$$

$$\Rightarrow \int_0^1 B(r) X(r)'\, dr$$

and

$$\frac{1}{n}\Sigma_1^n D_n^{-1} x_t x_t' D_n^{-1} = \Sigma_1^n \int_{(j-1)/n}^{j/n} D_n^{-1} x_{[nr]} x_{[nr]}' D_n^{-1}\, dr$$

$$= \int_0^1 D_n^{-1} x_{[nr]} x_{[nr]}' D_n^{-1}\, dr$$

$$\rightarrow \int_0^1 X(r) X(r)'\, dr$$

Hence

$$\frac{1}{\sqrt{n}} \hat{A} D_n D_n^{-1} x_{[nr]} \Rightarrow \left(\int_0^1 B(r) X(r)'\right)\left(\int X(r) X(r)'\right)^{-1} X(r)$$

and

$$\frac{1}{\sqrt{n}} \underline{y}_{[nr]} = \frac{1}{\sqrt{n}} y_{[nr]} - \frac{1}{\sqrt{n}} \hat{A} D_n D_n^{-1} x_{[nr]}$$

$$\Rightarrow B(r) - \left(\int_0^1 B \cdot X'\right)\left(\int_0^1 X X'\right) X(r)$$

$$\equiv B_p(r)$$

detrended BM (degree p).

Asymptotic Theory of Spurious Regression

Phillips (1986), Durlauf & Phillips (1988)

- Empirical consequences of spurious reg^n important and were analyzed as early as Yule (1927)
- examples like { # ordained ministers in UK / alcoholism statistics
- Classic textbook warnings about serially correlated errors invalidating inference
- urge to difference data before regression
  (Granger & Newbold, 1974, dramatic Monte Carlo study of regression of two RW's that were independent

  results: heavily biased towards rejection of no relationship
  i.e. <u>acceptance</u> of <u>spurious relation</u>

  T=50, DW < R^2 often, t values reject $\beta=0$ 75% of time in

  $$y_t = \alpha + \beta x_t + u_t \qquad y_t,\ x_t \text{ RW's}$$

  suggest critical value of 11.2 instead of 1.96.

Prototypical Spurious Regression

$$\begin{aligned} \Delta y_t &= u_t = C_u(L)\,\varepsilon_{ut} \\ \Delta x_t &= v_t = C_v(L)\,\varepsilon_{vt} \end{aligned} \Big\rangle \text{ need not be independent}$$

i.e.

$$\Delta z_t = w_t = C(L)\,\varepsilon_t \qquad \text{lrvar}(w_t) = \Omega_w = C(1)\Sigma_\varepsilon C(1)' > 0$$

$$z_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix} \equiv I(1) \text{ processes (full rank)}$$

By FCLT

$$\frac{1}{\sqrt{n}} Z_{[nr]} = \begin{bmatrix} \frac{1}{\sqrt{n}} y_{[nr]} \\ \frac{1}{\sqrt{n}} x_{[nr]} \end{bmatrix} \Rightarrow \begin{bmatrix} B_y(r) \\ B_x(r) \end{bmatrix} = B_z(r)$$

$$\equiv BM(\Omega)$$

Regression

$$y_t = \hat{\alpha} + \hat{\beta}' x_t + \hat{u}_t = \hat{\gamma}' w_t + \hat{u}_t \text{, say}$$

Regression Asymptotics

(i)
$$\hat{\beta} = \left[\Sigma (x_t - \bar{x})(x_t - \bar{x})'\right]^{-1} \left[\Sigma (x_t - \bar{x})(y_t - \bar{y})\right]$$

$$= \left[\frac{1}{n^2} \Sigma_1^n (x_t - \bar{x})(x_t - \bar{x})'\right]^{-1} \left[\frac{1}{n^2} \Sigma (x_t - \bar{x})(y_t - \bar{y})\right]$$

$$\Rightarrow \left(\int_0^1 \underline{B}_x \underline{B}_x'\right)^{-1} \left(\int_0^1 \underline{B}_x \underline{B}_y\right) \equiv \xi$$ has r.v. limit not a constant!

as $\frac{1}{n^2} \Sigma_1^n (z_t - \bar{z})(z_t - \bar{z}) \Rightarrow \int_0^1 \underline{B}_z \underline{B}_z$

where $\underline{B}_z = B_z - \int_0^1 B_z$

(ii) $\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$, so $\hat{\alpha}$ diverges as

$$\frac{\hat{\alpha}}{\sqrt{n}} = \frac{1}{n^{3/2}} \Sigma_1^n y_t - \hat{\beta} \frac{1}{n^{3/2}} \Sigma_1^n x_t$$

$$\Rightarrow \int_0^1 B_y - \xi \int_0^1 B_x$$

(iii)
$$R^2 = 1 - \Sigma \hat{u}_t^2 / \Sigma (y_t - \bar{y})^2$$

$$= \hat{\beta}' \Sigma (x_t - \bar{x})(x_t - \bar{x})' \hat{\beta} / \Sigma (y_t - \bar{y})^2$$

$$\Rightarrow \xi' \int_0^1 \underline{B}_x \underline{B}_x' \xi' / \int \underline{B}_y^2$$

$$= \frac{\int_0^1 \underline{B}_y \underline{B}_x' \left(\int_0^1 \underline{B}_x \underline{B}_x\right)^{-1} \int_0^1 \underline{B}_x \underline{B}_y}{\int_0^1 \underline{B}_y^2}$$

This is just the R^2 in the cts time regression

(*) $\quad B_y(r) = \alpha_0 + \alpha_1 B_x(r) + \xi(r)$

(N) $\quad DW = \Sigma_2^n (\Delta \hat{u}_t)^2 / \Sigma_1^n \hat{u}_t^2 \to_p 0$

$= \frac{1}{n^2} \Sigma_2^n (\Delta \hat{u}_t)^2 / \frac{1}{n^2} \Sigma_1^n \hat{u}_t^2$

Now

$$\frac{1}{n^2} \Sigma_1^n \hat{u}_t^2 = \frac{1}{n^2} \Sigma (y_t - \bar{y})^2 - \hat{\beta} \frac{1}{n^2} \Sigma (x_t - \bar{x})(x_t - \bar{x})' \hat{\beta}$$

$$\Rightarrow \int_0^1 \underline{B}_y^2 - \int_0^1 \underline{B}_y \underline{B}_x' \left(\int \underline{B}_x \underline{B}_x'\right)^{-1} \int_0^1 \underline{B}_x \underline{B}_y$$

$= \int_0^1 \underline{B}_{y.x}^2 \qquad \underline{B}_{y.x} = \underline{B}_y - \int_0^1 B_y B_x \left(\int_0^1 B_x B_x'\right)^{-1} \underline{B}_x$ see p110

$\equiv$ RSS in cts time regression (*)

and

$$\frac{1}{n} \Sigma (\Delta \hat{u}_t)^2 = \frac{1}{n} \Sigma (\Delta y_t - \hat{\beta}' \Delta x_t)^2$$

$$= (1 \;\; -\hat{\beta}') \left(\frac{1}{n} \Sigma \Delta z_t \Delta z_t'\right) \begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix}$$

$$\Rightarrow (1 \;\; -\xi') \, \Sigma_w \begin{pmatrix} 1 \\ -\xi \end{pmatrix}$$

$$\Sigma_w = E(\Delta z_t \Delta z_t') = E(w_t w_t')$$

$$= \eta' \Sigma_w \eta \qquad \eta = \begin{pmatrix} 1 \\ -\xi \end{pmatrix} = \text{r.v.}$$

(v) $t_i = t(\hat{\beta}_i) = \hat{\beta}_i / s_{\hat{\beta}_i}$

$$= \hat{\beta}_i \Big/ \left\{ s^2 \left[\left(\Sigma (x_t - \bar{x})(x_t - \bar{x})' \right)^{-1} \right]_{ii} \right\}^{1/2}$$

$$s^2 = \tfrac{1}{n} \Sigma \hat{u}_t^2$$

From the above we have

$$\frac{s^2}{n} = \frac{1}{n^2} \Sigma \hat{u}_t^2 \Rightarrow \int_0^1 \underline{B}_y^2 - \int_0^1 \underline{B}_y \underline{B}_x' \left(\int \underline{B}_x \underline{B}_x' \right)^{-1} \int \underline{B}_x \underline{B}_y$$

$$= \int_0^1 \underline{B}_{y \cdot x}^2$$

where $\underline{B}_{y \cdot x}(r) = \underline{B}_y(r) - \int_0^1 \underline{B}_y \underline{B}_x' \left(\int \underline{B}_x \underline{B}_x' \right)^{-1} \underline{B}_x$

= projection residual of $\underline{B}_y(r)$ on space spanned by $\underline{B}_x(r)$

= RSS in cts time of regression of B_y on $\{1(r), \underline{B}_x(r)\}$

Hence

$$t_i = \frac{\hat{\beta}_i'}{\left\{ n \frac{s^2}{n} \left[n^2 \frac{1}{n^2} \Sigma (x_t - \bar{x})(x_t - \bar{x})' \right]_{ii}^{-1} \right\}^{1/2}}$$

$$= \frac{n^{1/2} \hat{\beta}_i'}{\left\{ \frac{s^2}{n} \left[\frac{1}{n^2} \Sigma (x_t - \bar{x})(x_t - \bar{x})' \right]_{ii}^{-1} \right\}^{1/2}}$$

diverges

as

$$\frac{1}{\sqrt{n}} t_i = \frac{\hat{\beta}_i}{\left\{ \frac{s^2}{n} \left[\frac{1}{n^2} \Sigma (x_t - \bar{x})(x_t - \bar{x})' \right]^{-1}_{ii} \right\}^{1/2}}$$

$$\Rightarrow \frac{\xi_i}{\left\{ \int_0^1 B_{y.x}^2 \left[\left(\int_0^1 B_x B_x' \right)^{-1} \right]_{ii} \right\}^{1/2}}$$

Hence

$$P(|t_i| > K) \to 1 \quad \text{as } n \to \infty$$

i.e. we will always reject the null hypothesis of no association

(vi) <u>Residual autocorrelations</u>

$$r_s^2 = \Sigma \hat{u}_t \hat{u}_{t-s} / \Sigma \hat{u}_t^2$$

$$= \tfrac{1}{n^2} \Sigma_1^n \hat{u}_t \hat{u}_{t-s} / \tfrac{1}{n^2} \Sigma \hat{u}_t^2$$

so

$$r_s^2 \xrightarrow[p]{} 1 \qquad \text{as } \tfrac{1}{n^2} \Sigma_1^n \hat{u}_t \hat{u}_{t-s}$$

$$= \tfrac{1}{n^2} \Sigma_1^n \hat{u}_{t-s}^2 + o_p(1)$$

$$\hat{u}_t = y_t - \hat{\alpha} - \hat{\beta} x_t$$
$$= y_t - \bar{y} - \hat{\beta}(x_t - \bar{x})$$
$$= y_{t-s} - \bar{y} - \hat{\beta}(x_{t-s} - \bar{x})$$
$$+ \underline{\Sigma_{t-s+1}^{t} (u_{yj} - \hat{\beta}\, u_{xj})}$$

$= I(0)$ - sum of s $I(0)$ components

(vii) Box Pierce statistic

$Q_k = n \sum_{s=1}^{k} r_s^2$

for stationary time series (i.i.d)

$\sqrt{n}\, r_s \to N(0,1)$

$n r_s^2 \to \chi_1^2$

$Q_k \to \chi_k^2$

diverges for $I(1)$ series

almost certainly find evidence of serial correlation

i.e. $P(Q_k > M) \to 1$

c.f. $DW \to_p 0$

almost certain evidence of serial correlation here.

Sample Moments of I(0) processes

$u_t = C(L)\varepsilon_t \,,\quad \varepsilon_t \equiv iid(0, \Sigma_\varepsilon)\,,\quad \left\{ \begin{array}{l} \sum j^{1/2}\|C_j\| < \infty \\ C(1) \text{ full rank} \end{array} \right.$

Limit theory: for partial sums has already been resolved by FCLT. But notice the following alternate routes

- $X_n(r) = \frac{1}{\sqrt{n}} X_{[nr]} = \frac{1}{\sqrt{n}} \sum_1^{[nr]} u_j \Rightarrow B(r) \equiv BM(\Omega) \quad \Omega = 2\pi f_{uu}(0)$
$= C(1)\Sigma_\varepsilon C(1)'$

- $X_n(r) = \frac{1}{\sqrt{n}} X_{[nr]} = \sum_1^{[nr]} \frac{u_j}{\sqrt{n}} = \sum_1^{[nr]} \int_{(j-1)/n}^{j/n} dX_n(s) \;:=$ Riemann integral

with $X_n(s) = \frac{1}{\sqrt{n}} \sum_1^{j-1} u_t \quad \frac{j-1}{n} \le s < j/n$

$= \int_0^{[nr]/n} dX_n(s)$ the Riemann integral is well defined as $X_n(s)$ is a step function with finite # of jumps

$\Rightarrow \int_0^r dB(s) \underset{\text{def'n}}{=} B(r)$

Here the Riemann integral is not well defined
Stieltjes

as BM $B(s)$ does not have the requisite property, i.e. it is NOT of bounded variation (see below).

Remark 1. We could define $\int_0^r dB(s) = B(r)$ directly on the basis of its properties (the above cgce & the FCLT) & intuitively $B(r)$ is just the sum of its increments prior to $t = r$.

<u>Remark 2</u> Let $f(r) \in C[0,1]$ be ctsly differentiable. Then we could <u>define</u> the integral as follows

$$(*) \quad \int_0^r f(s)\, dB(s) \underset{def}{=} \Big[B(s) f(s)\Big]_0^r - \int_0^r f'(s)\, B(s)\, ds$$

$$= B(r) f(r) - \int_0^r f'(s)\, B(s)\, ds$$

and this is a continuous mapping of $C[0,1] \to C[0,1]$. The definition $(*)$ is constructed by virtue of the analogy to "integration by parts", which is valid if f is cts and B is of bdd variation <u>or</u> if B is cts and f is of bdd variation & cts.

<u>Example 1</u>

$$\frac{1}{n^{3/2}} \sum_1^n t u_t = \sum_1^n \frac{t}{n} \frac{u_t}{\sqrt{n}} = \sum_1^n \int_{(j-1)/n}^{j/n} \frac{[ns]+1}{n}\, dX_n(s)$$

$$j - 1/n \le s < j/n$$

$$= \int_0^1 \frac{[ns]+1}{n}\, dX_n(s)$$

$$\Rightarrow \int_0^1 s\, dB(s)$$

$$\underset{def'n}{=} \Big[B_1(s)\, s\Big]_0^1 - \int_0^1 B(s)\, ds$$

$$= B(1) - \int_0^1 B(s)\, ds \qquad \text{demeaned BM}$$

<u>Note</u> Let $S_t = \sum_1^t u_j$ then

$$\Delta\left(\sum_1^n t S_t\right) = \sum_1^n \Delta t\, S_t + \sum_1^n t\, \Delta S_t$$

i.e. $\quad n S_n = \sum_1^n S_t + \sum_1^n t u_t$

$\left.\right\}$ i.e. $\frac{1}{n^{3/2}} \sum_1^n t u_t = \frac{1}{\sqrt{n}} S_n - \frac{1}{n^{3/2}} \sum_1^n S_t$

Example 2

$$\frac{1}{n^{a+1/2}}\sum_1^n t^a u_t = \sum_1^n \left(\frac{t}{n}\right)^a \frac{u_t}{\sqrt{n}} \sim \int_0^1 \left(\frac{[ns]}{n}\right)^a dX_n(s)$$

$$\Rightarrow \int_0^1 s^a \, dB(s)$$

Note

By direct calculation we have

$$\Delta\left(\sum_1^n t^a S_t\right) = \sum_1^n \Delta t^a S_t + \sum_1^n t^a \Delta S_t$$

$$= \sum_1^n \left(t^a - (t-1)^a\right) S_t + \sum_1^n t^a u_t$$

$$= \sum_1^n t^a \left[1 - \left(1-\frac{1}{t}\right)^a\right] S_t + \sum_1^n t^a u_t$$

$$= \sum_1^n t^a \left[1 - \left\{1 - \frac{a}{t} + O\left(\frac{1}{t^2}\right)\right\}\right] S_t + \sum_1^n t^a u_t$$

$$= a\sum_1^n \left[t^{a-1} S_t + o\left(t^{a-2} S_t\right)\right] + \sum_1^n t^a u_t$$

Hence

$$n^a S_n \sim a \sum_1^n t^{a-1} S_t + \sum_1^n t^a u_t$$

so that

$$\frac{1}{n^{a+1/2}} \sum_1^n t^a u_t = \frac{1}{\sqrt{n}} S_n - a\frac{1}{n}\sum_1^n \left(\frac{t}{n}\right)^{a-1} \frac{S_t}{\sqrt{n}}$$

$$\Rightarrow B(1) - a\int_0^1 s^{a-1} B(s)\, ds$$

Example 3 general case $g \in C^1$

$$\frac{1}{\sqrt{n}} \sum_1^n g\left(\frac{t}{n}\right) u_t \Rightarrow \int_0^1 g(s)\, dB(s)$$

$$(*) \qquad = g(1) B(1) - \int_0^1 g'(s) B(s)\, ds$$

Remark 1

This def'n of $\int_0^1 g\,dB$ for $g \in C^1$ using "hypothetical" integration by parts is one way of approaching the general idea of a stochastic integral

Remark 2

We can obtain the limit process (*) by direct calculation, as before. Note that

$$\Delta \sum_1^n h_t = \sum_1^n \Delta h_t \quad \text{by linearity of } \Delta \,\&\, \Sigma$$

and if $h_t = h_{1t} h_{2t}$ we have

$$\begin{aligned}
\Delta h_t &= h_{1t} h_{2t} - h_{1t-1} h_{2t-1} \\
&= (h_{1t} - h_{1t-1}) h_{2t} + h_{1t-1}(h_{2t} - h_{2t-1}) \\
&= \Delta h_{1t} h_{2t} + h_{1t-1} \Delta h_{2t}
\end{aligned}$$

Now set $h_t = g(t/n)$. Then

$$\Delta\left(\sum_1^n h_t S_t\right) = \sum_1^n \Delta(h_t S_t) = \sum_1^n \left[(\Delta h_t) S_t + h_{t-1}(\Delta S_t)\right]$$

i.e.

$$\underbrace{\Delta\left(\sum_1^n g\left(\tfrac{t}{n}\right) S_t\right)}_{= g(1) S_n} = \sum_1^n \left(\Delta g\left(\tfrac{t}{n}\right)\right) S_t + \sum_1^n g\left(\tfrac{t-1}{n}\right) u_t$$

Hence

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_1^n g\left(\tfrac{t-1}{n}\right) u_t &= g(1) \frac{S_n}{\sqrt{n}} - \sum_1^n \left(\Delta g\left(\tfrac{t}{n}\right)\right) \frac{S_t}{\sqrt{n}} \\
&= g(1) \frac{S_n}{\sqrt{n}} - \sum_1^n g'\left(\tfrac{t-1}{n}\right) \frac{1}{n} \frac{S_t}{\sqrt{n}} \\
&\Rightarrow g(1) B(1) - \int_0^1 g'(r) B(r)\,dr \\
&\underset{\text{def'n}}{\equiv} \int_0^1 g(r)\, dB(r)
\end{aligned}$$

Near-Integrated Processes & Roots local to Unity

I(1) process

$$\Delta X_t = u_t = C(L)\varepsilon_t \qquad \Sigma_j j^{1/2}\|C_j\| < \infty,$$

$C(1)$ full rank

$$\frac{1}{\sqrt{n}} X_{[nr]} \Rightarrow B(r) \equiv BM(\Omega) \qquad \Omega = C(1)\Sigma_\varepsilon C(1)'$$

Near- I(1) process

$$\Delta X_t = \tfrac{1}{n} D X_{t-1} + u_t$$

or

$$(*) \qquad X_t = A_n X_{t-1} + u_t \qquad A_n = I + \underbrace{\tfrac{1}{n} D}_{\to 0}$$

"local to unit" roots

Notes

(i) System (*) really defines a triangular array of the type $X_{t,n} = A_n X_{t-1,n} + u_t \quad t = 1, .., n$

i.e.

$\{\{X_{t,n}\}_{t=1}^{n}\}_{n=1}^{\infty} \equiv$ time series array

each row for fixed n is a time series with roots that are near unity

(ii) if $D = \mathrm{diag}(d_1, \ldots, d_k)$

$d_i > 0$ implies X_{it} is near explosive

$d_i < 0$ implies X_{it} is near stationary

(iii) it is often more convenient to replace (*) with the asymptotically equivalent system

$$X_t = A_n X_{t-1} + u_t \quad , \quad A_n = \exp\left(\tfrac{1}{n} D\right)$$
$$\sim I + \tfrac{1}{n} D$$

(iv) Near I(1) processes are very useful in developing more general asymptotics and in studying power functions of tests.

Theorem (Asymptotics for near I(1) processes)

Phillips (1987, 1988), Chan & Wei (1988)

(a) $\frac{1}{\sqrt{n}} X_{[nr]} \Rightarrow J_D(r) \equiv \int_0^r \exp\{(r-s)D\}\, dB(s)$

(b) $\frac{1}{n^{3/2}} \sum_1^n X_t \Rightarrow \int_0^1 J_D(r)dr$

(c) $\frac{1}{n^2} \sum_1^n X_t X_t' \Rightarrow \int_0^1 J_D J_D'$

where $B(s) \equiv BM(\Omega)$, $\Omega = 2\pi f_{uu}(0) = C(1)\Sigma_\varepsilon C(1)'$

and $J_D(r)$ is a vector diffusion process that satisfies the stochastic differential equation

$$d J_D(r) = D J_D(r) dr + dB(r), \qquad J_D(0) = 0$$

Proof To prove (a) we note that

$$X_t = \sum_{j=0}^{t-1} A_n^j u_{t-j} + A_n^t X_0$$

$$= \sum_{p=1}^{t} A_n^{t-p} u_p + A_n^t X_0 \qquad \begin{cases} t-j = p \\ t-p = j \end{cases}$$

i.e.

$$\frac{1}{\sqrt{n}} X_{[nr]} = \sum_{p=0}^{[nr]} A_n^{[nr]-p} \frac{1}{\sqrt{n}} u_p + A_n^{[nr]} \frac{X_0}{\sqrt{n}}, \quad A_n = \exp\{\tfrac{1}{n} D\}$$

$$= \sum_{p=0}^{[nr]} \exp\left\{ \frac{[nr]-p}{n} D \right\} \frac{u_p}{\sqrt{n}} + O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$= \sum_{j=0}^{[nr]} \int_{(j-1)/n}^{j/n} \exp\left\{ \left(\frac{[nr]}{n} - \frac{j}{n} \right) D \right\} dX_n(s) + O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$(j-1)/n \le s < j/n$$

$$\sim \int_0^r \exp\{(r-s)D\}\, dX_n(s) + O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$\Rightarrow \int_0^r \exp\{(r-s)D\}\, dB(s)$$

$$= J_D(r)$$

Similarly, we find for (b)

$$\frac{1}{n^{3/2}}\Sigma_1^n X_t = \frac{1}{n}\Sigma_1^n \frac{1}{\sqrt{n}} X_t = \Sigma_1^n \int_{(j-1)/n}^{j/n} \frac{1}{\sqrt{n}} X_{[ns]}\, ds$$

$$= \int_0^1 \frac{X_{[ns]}}{\sqrt{n}}\, ds \qquad \frac{j-1}{n} < s < j/n$$

$$\Rightarrow \int_0^1 J_D(s)\, ds$$

and for (c)

$$\frac{1}{n^2}\Sigma_1^n X_t X_t' = \frac{1}{n}\Sigma_1^n \frac{X_t}{\sqrt{n}}\frac{X_t'}{\sqrt{n}} \Rightarrow \int_0^1 J_D(r) J_D(r)' dr$$

<u>Remark</u>

$$J_D(r) = \int_0^r \exp\{(r-s)D\}\, dB(s) = \int_0^r g(r-s)\, dB(s)$$

Since $g \in C^1$ we can write this as

$$= g(0) B(r) - \int_0^r \frac{d}{ds} g(r-s)\, B(s)\, ds$$

$$(*) \qquad = B(r) + \int_0^r D\, g(r-s)\, B(s)\, ds$$

Now take differentials (wrt r) and note that

$$d \int_0^r D\, g(r-s) B(s) ds = D\, g(0)\, B(r)\, dr + \int_0^r D\, g'(r-s) B(s) ds\, dr$$

$$= \left\{ D\, B(r) + D^2 \int_0^r g(r-s) B(s) ds \right\} dr$$

$$= D\, J_D(r)\, dr \text{ from } (*)$$

Thus

$$d J_D(r) = dB(r) + D\, J_D(r)\, dr$$

i.e.

$$dJ_D(r) = D J_D(r) dr + dB(r) \qquad \text{diffusion equation driven by BM}$$

called (vector) Ornstein-Uhlenbeck process

Variation & Quadratic Variation

(1) k-variation of a function

Let $\Pi_m(t) = [0 = t_0 < t_1 < \ldots < t_m = t]$

be a partition of $[0,t]$ interval

$$V^k(f) = \lim_{m \to \infty} \sum_{j=0}^{m-1} |f(t_{j+1}) - f(t_j)|^k, \quad k > 0$$

is called the k-variation of f if the limit exists

(2) VF function (variation finite)

f is a VF function if

$$V_t(f) = \sup_{\Pi_m(t)} \sum_{j=0}^{m-1} |f(t_{j+1}) - f(t_j)| < \infty$$

or, we say f is of bounded variation

(3) Theorem (integration by parts)

If f is cts & g is a VF function then

$\int f\,dg$, $\int g\,df$ both exist as Riemann Stieltjes integral

and the integration by parts formula

$$\int_0^t f\,dg = [f(t)g(t) - f(0)g(0)] - \int_0^t g\,df$$

is valid

The bisection partition

(i) A useful partition of $[0,t]$ in practice is

$$\Pi_m(t) = \left[t_k = \frac{k}{2^n} \ , \ k = 0, 1, \dots, [2^n t] \right]$$

$$m = [2^n t]$$

Note that

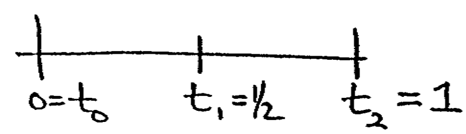
$$t_m = \frac{[2^n t]}{2^n} \longrightarrow t \quad \text{as} \quad n \to \infty$$

$$= t \quad \text{for } t \text{ integer}$$

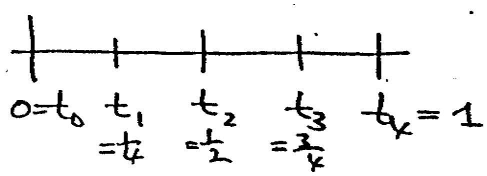
We often proceed as if t is an integer and then $[2^n t] = t$.

(ii) the partition bisects itself as n increases. Thus, let $t = 1$ and consider

$n = 1$, $m = 2$



$n = 2$, $m = 4$



(iii) This bisection partition is especially useful in proving a.s. cvgce of the second variation, because $\sum_{n=1}^{\infty} \frac{1}{2^n} < \infty$ and we can use the Borel Cantelli lemma

Path Properties of Brownian Motion

Theorem $W = BM(1)$

(i) $V_t^1(W) = \infty$ a.s i.e. $W(t)$ is of unbounded variation

(ii) $V_t^2(W) = t$ a.s i.e. second variation is finite and $= t$ a.s

Remark We call the second variation the quadratic variation and use the notation

$$[W]_t = \int_0^t (dW)^2 = t \quad \text{a.s.}$$

called square bracket process and defined as

$$[W]_t = \lim_{n\to\infty} \sum_{k=1}^{2^n t} \left[W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right]^2 \quad \text{a.s.}$$

using the bisection partition

Proof We prove (ii) first. Note that

$$Y_k = W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right) \equiv N\left(0, \tfrac{1}{2^n}\right) \quad \text{iid } \forall k$$

Hence if

$$Q_n = \sum_{k=1}^{2^n t} \left[W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right]^2 = \sum_{k=1}^{2^n t} Y_k^2$$

we have

$$E(Q_n) = \sum_{k=1}^{2^n t} \frac{1}{2^n} = t, \quad \forall n \qquad \chi_1^2 \ \& \ \mathrm{var}(\chi_1^2) = 2$$

$$\mathrm{var}(Q_n) = \sum_{k=1}^{2^n t} \mathrm{var}(Y_k^2) = \sum_{k=1}^{2^n t} \frac{1}{2^{2n}} \mathrm{var}\left(\frac{Y_k^2}{1/2^n}\right) = 2\sum_{k=1}^{2^n t} \frac{1}{2^{2n}}$$

i.e.

$$\text{var}(Q_n) = \frac{2\,2^n t}{2^{2n}} = \frac{2t}{2^n} \longrightarrow 0 \quad \text{as } n\to\infty$$

Thus,

$$Q_n \underset{L_2}{\underset{P}{\longrightarrow}} t \quad \text{as} \quad n\to\infty$$

To prove a.s. cgce just use the Borel Contelli lemma

Lemma (Borel Contelli)

$$\sum_{n=1}^{\infty} P(E_n) < \infty \implies P(E_n \text{ i.o}) = 0$$

Note:

- $$E_n \text{ i.o} = \limsup_n E_n = \bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} E_n$$

  $\omega \in E_n$ infinitely often iff ω belongs to ∞'ly many of $\{E_n\}_1^\infty$

- proved in 1st course.

Let $E_n = [|Q_n - t| > \varepsilon]$, for some $\varepsilon > 0$. Then

$$\sum_{n=1}^{\infty} P(|Q_n - t| \geq \varepsilon) \leq \sum_{n=1}^{\infty} \text{var}(Q_n) = 2t \sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{2t}{1-1/2}$$

$$= 4t < \infty$$

Hence

$$P(|Q_n - t| > \varepsilon \text{ i.o}) = 0 \quad \forall \varepsilon > 0$$

which implies that

$$Q_n \underset{a.s}{\longrightarrow} t.$$

establishing part (ii) of the theorem

Lemma (Borel-Cantelli)

$$\sum_1^\infty P(E_n) < \infty \Rightarrow P(E_n \text{ i.o.}) = 0$$

Proof . E_n i.o. $= \limsup_n E_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E_n$

Consider $F_m = \bigcup_{n=m}^{\infty} E_n$. We have

$$F_m \supset F_{m+1} \supset F_{m+2} \supset \cdots$$ decreasing sequence

i.e. $F_m \searrow$ is decreasing and, in fact,

$$F_m \searrow F = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E_n$$

So by the monotone property

$$P(F) = \lim_{m\to\infty} P(F_m)$$

i.e.

$$P(E_n \text{ i.o.}) = P(\limsup_n E_n) = \lim_{m\to\infty} P(F_m)$$

But the condition that $\sum_{n=1}^{\infty} P(E_n) < \infty$ ensures that

$$\lim_{m\to\infty} \sum_{n=m}^{\infty} P(E_n) = 0$$

and

$$P(F_m) \le \sum_{n=m}^{\infty} P(E_n)$$ by def^n of F_m

Hence $P(F_m) \to 0$ and so

$$P(E_n \text{ i.o.}) = \lim P(F_m) = 0$$

Next we prove point (i). The first variation is:

$$V_t^1(W) = \lim_{n\to\infty} \sum_{k=1}^{2^n t} \left|W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right|$$

Since

$$Q_n = \sum_{k=1}^{2^n t} \left(W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right)^2$$

$$\leq \left[\max_{k\leq 2^n t} \left|W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right|\right] \underbrace{\sum_{k=1}^{2^n t} \left|W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right|}_{\mathbin{/\!/}}$$

$$= \max_{k\leq 2^n t} \left|W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right| V_{t,n}^1(W)$$

But $Q_n \to t$ a.s., shown above,

and

$$\max_{k\leq 2^n t} \left|W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right| \longrightarrow 0$$

by virtue of the continuity of $W(t)$

It follows that

$$V_{t,n}^1(W) \longrightarrow \infty \quad \text{a.s.}$$

i.e.

$$V_t^1(W) = \lim_{n\to\infty} \sum_{k=1}^{2^n t} \left|W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right|$$

$$= \lim_{n\to\infty} V_{t,n}^1(W)$$

$$= \infty$$

Direct proof that BW has unbounded variation

Let $v_m(W) = \sum_{j=1}^{m} |W(\frac{j}{m}) - W(\frac{j-1}{m})|$

Note that

$$W(\tfrac{j}{m}) - W(\tfrac{j-1}{m}) \equiv N(0, \tfrac{1}{m}) \equiv N_m$$
$$\equiv \text{iid } N_m$$

Hence

$$v_m(W) = \Sigma_{j=1}^{m} \text{ iid } |N_m|$$

so

$$E(v_m(W)) = m\, E|N_m|$$
$$= m \frac{1}{m^{1/2}} E|N|$$
$$= m^{1/2} a,$$

$$a = 2\int_0^\infty x \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \sqrt{\tfrac{2}{\pi}} \int_0^\infty x e^{-x^2/2} dx = \sqrt{\tfrac{2}{\pi}}\, \tfrac{1}{2}\int_0^\infty e^{-r/2} dr = \sqrt{\tfrac{2}{\pi}}\, e^{-r/2}\Big|_0^\infty = \sqrt{\tfrac{2}{\pi}}$$

$$b = 1 - \tfrac{2}{\pi}$$

$$\mathrm{var}(v_m(W)) = m\, \mathrm{var}|N_m|$$
$$= m \frac{1}{m} \mathrm{var}|N|$$
$$= \mathrm{var}|N|$$
$$= b, \text{ fixed say}$$

Then let $K_m = E(v_m(W)) = m^{1/2} a$

$$P\left(v_m(W) > \tfrac{1}{2}K_m\right) \geq P\left(|v_m(W) - K_m| < \tfrac{1}{2}K_m\right)$$

i.e.

$$-\tfrac{1}{2}K_m < v_n(W) - K_m < \tfrac{1}{2}K_m$$
$$\tfrac{1}{2}K_m < v_n(W) < \tfrac{3}{2}K_m$$

$$= 1 - P\left(|v_n(W) - K_m| > \tfrac{1}{2}K_m\right)$$
$$> 1 - \frac{E(v_n(W) - K_m)^2}{\frac{1}{2}K_m} \quad \text{by Tchebycheff.}$$
$$\to 1 \qquad \to 0$$

Remark

(a) Part (i) shows that the BM $W(t)$ has unbounded variation a.s.. Consequently, we cannot define integrals like

$$\int_0^t f(s)\,dB(s) \quad \text{as Riemann Stieltjes integral}$$

in general by a pathwise (i.e. given samp $\omega \in \Omega$, probability space) argument

(b) for continuously differentiable (possibly random) of bounded variation $f(s)$, we have earlier been able to define $\int_0 f(s)\,dB(s)$ using the form

$$\int_0^t f(s)\,dB(s) = f(t)B(t) - \int_0^t f'(s)\,B(s)\,ds$$

We now wish to extend this definition for random $f(s)$ with sample paths of unbounded variation;

(c) For example, we want to define

$$\int_0^1 B(s)\,dB(s)$$

By our earlier heuristics we would expect this to be the limit of sample covariances like

$$\frac{1}{n}\sum_1^n S_{t-1}u_t = \sum_1^n \frac{S_{t-1}}{\sqrt{n}}\frac{u_t}{\sqrt{n}} = \sum_1^n \int^{j/n} X_n(s)\,dX_n(s) = \int_0^1 X_n(s)\,dX_n(s)$$

More on the Quadratic Variation (q.v) of BM

From the theorem p.121 we have.

$$(*)\quad [W]_t \underset{\text{def}}{=} V_t^2(W) = t \quad \text{a.s.}$$

i.e.

$$\sum_{k=1}^{2^n t} \left[W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right]^2 \underset{n\to\infty}{\longrightarrow} t \quad \text{a.s}$$

Alternate Representation of q.v

Note that

$$\int_{(k-1)/2^n}^{k/2^n} dW \underset{\text{def}}{=} W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right),$$

so it seems reasonable to write the approximation

$$\int_{(k-1)/2^n}^{k/2^n} (dW)^2 \sim \left(W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right)^2$$

Then

$$\sum_{k=1}^{2^n t} \left[W\left(\tfrac{k}{2^n}\right) - W\left(\tfrac{k-1}{2^n}\right)\right]^2 \sim \sum_{k=1}^{2^n t} \int_{(k-1)/2^n}^{k/2^n} (dW)^2$$

$$= \int_0^t (dW)^2$$

Differential of the q.v process

In fact $[W]_t = \int_0^t (dW)^2 = q.v(W(t))$

and

$$d[W]_t = (dW)^2 = dt \quad \text{a.s}$$

in view of $(*)$.

Some further heuristics

$$dW = W(t+dt) - W(t) \equiv N(0, dt)$$

$$dt = \text{Lebesgue measure infinitesimal}$$

This implies

$$\frac{(dW)^2}{dt} \equiv \chi^2_1$$

so that

$$E\left[\frac{(dW)^2}{dt}\right] = 1 \quad , \quad \text{var}\left[\frac{(dW)^2}{dt}\right] = 2$$

ie

$$\text{var}\left((dW)^2\right) = 2(dt)^2 = 0$$

Because in conventional calculus when dt is infinitesimal $(dt)^2 = 0$

Thus

$$(dW)^2 = dt \qquad \text{a.s.}$$

(since $(dW)^2$ has zero variance it equals its mean with probability 1)

Stochastic Calculus

Second order quantities like $(dW)^2$ matter in stochastic calculus, where in conventional calculus second order infinitesimals are zero like $(dt)^2 = 0$.

The consequence is that in considering functions of stochastic processes like Brownian motion we need to carry 2nd

Theorem (Ito's formula)

$f \in C^2$, twice continuously differentiable

$W(t) \equiv BM(1)$, $B(t) \equiv BM(\omega^2)$

(a) $df(W(t)) = f'(W(t))\,dW(t) + \frac{1}{2} f''(W(t))\,dt$

(b) $df(B(t)) = f'(B(t))\,dB(t) + \frac{\omega^2}{2} f''(B(t))\,dt$

Proof

$$f(W+dW) = f(W) + f'(W)\,dW + \frac{1}{2} f''(W)(dW)^2 + o(dW)^2$$

i.e.

$$\begin{aligned} df &= f'(W)\,dW + \tfrac{1}{2} f''(W)\,dt + o(dt) \\ &= f'(W)\,dW + \tfrac{1}{2} f''(W)\,dt \end{aligned}$$

giving (a). (b) follows in the same way by using the fact that $B(t) = \omega W(t)$ and then

$$(dB)^2 = \omega^2 (dW)^2 = \omega^2 dt \quad \text{a.s.}$$

Example 1

$$\begin{aligned} d(B^2) &= 2B\,dB + \tfrac{1}{2}(2)\,\omega^2 dt \\ &= 2B\,dB + \omega^2 dt \end{aligned}$$

Hence

$$B\,dB = \tfrac{1}{2}\left[d(B^2) - \omega^2 dt\right]$$

$$\int_0^1 B\,dB = \tfrac{1}{2}\left[\int_0^1 d\,B^2 - \omega^2 \int_0^1 dt\right]$$

$= \frac{1}{2}[B^2(1)$ [illegible]

i.e.

$$\int_0^1 W\,dW = \frac{1}{2}\left[W(1)^2 - 1\right] \qquad (*)$$

Remark Note that if $W(t)$ were of bounded variation, integration by parts would yield

$$\int_0^1 W\,dW = W^2\Big|_0^1 - \int_0^1 W\,dW$$

i.e.

$$2\int_0^1 W\,dW = W(1)^2$$

or

$$\int_0^1 W\,dW = \frac{1}{2}W(1)^2 > 0$$

which is very different from $(*)$. The fact that $W(t) \equiv BM(1)$ is of unbounded variation matters a great deal. In effect, we have

$$(B + dB)^2 - B = 2B\,dB + (dB)^2$$

$$= 2B\,dB + \omega^2 dr \quad a.s$$

i.e.

$$d(B^2) = 2B\,dB + \omega^2 dr.$$

Example 2

$$d\left[1/B(r)\right] = -\frac{1}{B^2}dB + \frac{1}{2}\left(+\frac{2}{B^3}\right)(dB)^2$$

$$= \underbrace{-\frac{1}{B^2}dB}_{\text{martingale component}} + \underbrace{\frac{1}{B^3}dt\,\omega^2}_{\text{drift}}$$

Multivariate Extension of Ito formula

Let $B(r) \equiv BM(\Omega)$, $W(r) \equiv BM(I_k)$

We have the following q.v. processes

Lemma

$$d[B]_t = dB\,dB' = \Omega\,dt$$

$$d[W]_t = dW\,dW' = I_k\,dt$$

Proof

$$(dW_i)^2 = dt \qquad \forall i = 1, \dots, k \qquad \text{a.s.}$$

$$dW_i\,dW_j = 0 \quad \text{a.s.} \quad \forall i \neq j$$

We can confirm the latter by noting that

$$\left.\begin{array}{l} dW_i \equiv N(0, dt) \\ dW_j \equiv N(0, dt) \end{array}\right\} \text{indep } i \neq j$$

so

$$E(dW_i\,dW_j) = 0$$

$$E(dW_i\,dW_j)^2 = E(dW_i)^2 E(dW_j)^2 = (dt)^2$$
$$= 0 \text{ as}$$

Thus $\quad dW_i\,dW_j = 0 \quad$ a.s.

The result for $d[B]_t$ follows by noting that

$$B(t) = \Omega^{1/2} W(t)$$

$$dB = \Omega^{1/2} dW$$

$$dB\,dB' = \Omega^{1/2} dW\,dW' \Omega^{1/2} = \Omega\,dt$$

a.s.

<u>Theorem</u> (Vector Ito formula)

$$f \in C^2, \quad B \equiv BM(\Omega)$$

$$d\,f(B) = f'(B)\,dB + \tfrac{1}{2}\,\mathrm{tr}\left[f^{(2)}(B)\,\Omega\right]dt$$

<u>Proof</u>

$$f(B+dB) = f'(B)\,dB + \tfrac{1}{2}\,dB'\,f^{(2)}(B)\,dB + o(dB'dB)$$

$$= f'(B)\,dB + \tfrac{1}{2}\,\mathrm{tr}\left[f^{2}(B)\,\Omega\right]dt,$$

as required.

<u>Functions of t & $B(t)$</u>

<u>Theorem</u> (Extended Ito formula)

$$d\,f(t, B(t)) = f_t(t,B)\,dt + f_B(t,B)'\,dB + \tfrac{1}{2}\,\mathrm{tr}\left[f_{BB}(t,B)\,\Omega\right]dt$$

<u>Proof</u> Exactly as above

$$f(t+dt, B+dB) - f(t,B) = f_t\,dt + f_B'\,dB + \tfrac{1}{2}\,dB' f_{BB}\,dB + o(dt)$$

i.e.

$$d\,f(t,B) = f_t\,dt + f_B'\,dB + \tfrac{1}{2}\,\mathrm{tr}(f_{BB}\,\Omega)\,dt$$

General Cts Martingales & q.v. processes

$(\Omega, \mathcal{F}, P)$ = complete prob. space

$(\mathcal{F}_t)_{t \geq 0}$ = filtration $\quad \mathcal{F}_s \subset \mathcal{F}_t \quad \forall s<t$

right cts, so that $\mathcal{F}_t = \mathcal{F}_{t+} = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}$

M_t = cts squareable MG

i.e. (i) $E(M_t \mid \mathcal{F}_s) = M_s \quad \forall t > s$

(ii) $M_t \in L_2(P)$, i.e. $\int M_t^2 \, dP < \infty$

(iii) M_t has all sample paths cts

Examples

(1) $M_t = B(t) \equiv BM\,(\omega^2)$

(2) $M_t = B(t)^2 - t\omega^2$

$$
\begin{aligned}
E(M_t \mid \mathcal{F}_s) &= E\left(\left(B(s) + B(t) - B(s)\right)^2 \mid \mathcal{F}_s\right) - t\omega^2 \\
&= \left[B(s)^2 + \omega^2(t-s)\right] - t\omega^2 \\
&= B(s)^2 - s\omega^2 \\
&= M_s \quad \text{a.s.}
\end{aligned}
$$

note also that $B(t)^2 - t\omega^2 = 2\int_0^t B\,dB \equiv MG$

here $\int B\,dB$ is defined from the rule $dB^2 = 2B\,dB + \omega^2 dt$

(3) $M_t = \int_0^t B\,dB$

$$
\begin{aligned}
E(M_t \mid \mathcal{F}_s) &= \int_0^s B\,dB + E\left(\int_s^t B\,dB \mid \mathcal{F}_s\right) \\
&= M_s + E\left(\int_s^t B\,dB\right) \\
&= M_s + E\left\{\tfrac{1}{2}\left[B(r)^2 - \omega^2 r\right]\Big|_s^t\right\} \\
&= M_s
\end{aligned}
$$

Remark In both the above examples it is simplest to use stochastic calculus and show that

$$E(dM_t \mid \mathcal{F}_t) = 0$$

e.g.

(i)
$$M_t = W(t)^2 - t$$

$$dM_t = 2W(t)\,dW + \tfrac{2}{2}dt - dt$$

$$E(dM_t \mid \mathcal{F}_t) = E(2W(t)\,dW \mid \mathcal{F}_t)$$
$$= 2W(t)\,E(dW \mid \mathcal{F}_t)$$
$$= 0$$

(ii)
$$M_t = \int_0^t B\,dB$$

$$dM_t = B\,dB$$

$$E(dM_t \mid \mathcal{F}_t) = B\,E(dB \mid \mathcal{F}_t) = 0$$

In both cases $E(dB \mid \mathcal{F}_t) = 0$ because $dB(t) = B(t+dt) - B(t)$ is forward looking

Quadratic variation (q.v.) process

$$[M]_t = \operatorname*{plim}_{n \to \infty} \sum_{k=1}^{[2^n t]} \left(M\left(\tfrac{k}{2^n}\right) - M\left(\tfrac{k-1}{2^n}\right) \right)^2 \qquad (*)$$

$$= \int_0^t (dM)^2 \quad \text{in an alternative notation}$$

$[M]_t$ is a continuous increasing stochastic process

Example

$$M_t = \int_0^t B\,dB \,, \quad [M]_t = \omega^2 \int_0^t B^2(s)\,ds$$

proof note that $dM_{s} = B\,dB$

$$(dM_{s})^2 = B^2(s)(dB)^2 = \omega^2 B^2(s)\,ds$$

$$[M]_t = \int_0^t (dM_s)^2 = \omega^2 \int_0^t B(s)^2\,ds$$

Note. We can prove that the plim in the definition of $[M]_t$ exists and that the cgce is also in L_1 when $M_t \in L_2$. (eg. Ethier & Kurtz p67)

Stochastic Integration w.r.t M

- Our object is to define quantities like $\int_0^t X\,dM$ where M_t is cts L_2 MG and X_t is cts stochastic process.

- if X_t had sample paths that were of bounded variation we could define it using intg^n by parts, viz

$$\int_0^t X\,dM = X_t M_t - X_0 M_0 - \int_0^t M\,dX$$

(134

The steps involved in the definition of the general stochastic integral $\int_0^t X\,dM$ are as follows:

(i) Define $\int_0^t X\,dM$ for $X(t)$ a step function (simple function) of form

defines class of function $\mathcal{A}$

$$X(t) = \sum_{i=0}^{\infty} X(t_i)\, 1\{t_i \le t < t_{i+1}\}$$

$$0 \le t_0 < t_1 < \dots < t_n$$

and allow $t_n \to \infty$

so $X(t)$ is real bounded & right cts step fnc & $\mathcal{F}_t$-mble (assume each X_{t_i} is $\mathcal{F}_{t_i}$-mble)

as

$$\int_0^t X\,dM = \sum_{\substack{i \ge 0 \\ t_{i+1} \le t}} X(t_i)\left[M(t_{i+1}) - M(t_i)\right]$$

$$+ \; X(t_{m(t)})\left[M(t) - M(t_{m(t)})\right]$$

with $m(t) = \max\{i \ge 0 : t_i \le t\}$

(ii) Note that for $X \in \mathcal{A}$ and $M \in \mathcal{M}_2$ (= space of cts L_2 MG's) we have

$$\left[\int_0^t X\,dM\right]_t = \int_0^t X^2\, d[M]_t$$

Proof

$$d\left(\int_0^t X\,dM\right) = X_t\, dM_t$$

and $E(X_t dM_t | \mathcal{F}_t) = X_t E(dM_t | \mathcal{F}_t) = 0$ because M_t = mG. Also

$$\left[d\left(\int_0^t X dM\right)\right]^2 = (X_t dM)^2 = X_t^2 (dM)^2 = X_t^2 d[M]_t$$

so that by def^n of $[M]_t$

$$\left[\int_0^t X dM\right]_t = \int_0^t X^2 (dM)^2 = \int_0^t X^2 d[M]$$

$\underbrace{\qquad}_{\text{q.v. of } \int X dM}$

(iii) $$E\left(\int_0^t X dM\right)^2 = E\left\{\int_0^t X^2 d[M]_t\right\} = E\left[\int_0^t X dM\right]_t$$

__Proof__ $$E\left(\int_0^t X dM\right)^2 = E\left\{\sum_i \sum_j X(t_i) X(t_j) (M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j})\right\}$$

$$= E\left\{\sum_i \sum_j E\left(X(t_i) X(t_j)(M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j}) \Big| \mathcal{F}_{t_i}\right)\right\}$$

assume wlg $t_i < t_j$

$$= E\left\{\sum_i E\left(X(t_i)^2 (M_{t_{i+1}} - M_{t_i})^2 \Big| \mathcal{F}_{t_i}\right)\right\}$$

$$= E\left\{\sum_i X(t_i)^2 E\left[(M_{t_{i+1}} - M_{t_i})^2 | \mathcal{F}_{t_i}\right]\right\}$$

$$= E\left\{\sum_i X(t_i)^2 E\left[\left(\int_{t_i}^{t_{i+1}} dM\right)^2\right]\right\}$$

$$= E\left\{\sum_i X(t_i)^2 E\left(\int_{t_i}^{t_{i+1}} (dM)^2\right)\right\} \qquad \text{as } E(dM_t dM_s) = 0 \;\; s \neq t$$

$$= E\left\{\sum_i X(t_i)^2 E\left[\int_{t_i}^{t_{i+1}} (dM)^2 \Big| \mathcal{F}_{t_i}\right]\right\}$$

$$= E\left\{\sum_i X(t_i)^2 \int_{t_i}^{t_{i+1}} (dM)^2\right\}$$

$$= E\left\{\int_0^t X(s)^2 d[M]\right\}$$

(iv) $X \to \int X\,dM$ is a mapping: $\mathcal{A} \to \mathcal{M}_2$

space $(\mathcal{A}, \|\cdot\|_M) = \{ X \in \mathcal{A}$ with seminorm $E\{\int_0^t X^2(s)\, d[M]_s\} \}$

here "$d[M]_s$" = Stieltjes measure induced by the increasing process $[M]_t$

space $\mathcal{M}_2 = \{ M \equiv MG$ s.th $M(t)$ cts sample path & $E\, M(t)^2 = \int M^2 dP < \infty \}$

Remark The mapping $X \to \int X\,dM$ is an isometry in view of (iii) because

$\int_0^t X\,dM \equiv MG \in \mathcal{M}_2$ and

$$E\left(\int_0^t X\,dM\right)^2 = \int_0^t X^2\, d[M]_t$$

ie same distance (norm) as in $(\mathcal{A}, \|\,\|_M)$

(v) The above defines $\int_0^t X\,dM \ \forall X \in \mathcal{A}$.
To get a more general definition we extend $\mathcal{A}$ in the following way to $L_2[M] = \{ X(t)$ $\mathcal{F}_t$ mble

<u>Lemma</u> If $M \in \mathcal{M}_2$, $X \in L_2([M])$ then $\exists$ seq $X_n \in \mathcal{S}$ s.th

$$\lim_{n\to\infty} E\left\{\int_0^t (X_n - X)^2 \, d[M]\right\} = 0 \quad t \geq 0$$

<u>Proof</u> If X is bdd & cts then the stated result holds for $X_n(s) = X\left(\frac{[ns]}{n}\right)$

We can then extend this to $X \in L_2([M])$.

(c.f. Ethier & Kurtz p.282)

<u>Remark</u> In effect, this shows that the space $\mathcal{S}$ is dense in $L_2([M])$ – for all $X \in L_2([M])$ there is an $X_n \in \mathcal{S}$ close to it, and for this X_n we can define $\int X_n dM$ as above. We can then construct $\int X dM$ as the limit of $\int X_n dM$. Thus

(vi) For $M_t \in \mathcal{M}_2$ & $X \in L_2([M])$ $\exists$ process $\int_0^t X dM$ (a.s. unique) s.th. whenever

$$(*) \quad \sum_n \left[E\left\{\int_0^T (X_n - X)^2 \, d[M]\right\}\right]^{1/2} < \infty$$

we have

$$(**) \quad \sup_{0 \leq t \leq T} \left|\int_0^t X_n dM - \int_0^t X dM\right| \to 0 \quad T > 0$$

a.s., $L_2(P)$ as $n \to \infty$

i.e. $\int_0^t X_n \, dM \longrightarrow \int_0^t X \, dM$ a.s, & L_2

uniformly in t over $0 \le t \le T$, $\forall T > 0$

<u>Proof</u>

$$\int X_n \, dM \underset{a.s}{\longrightarrow} \int X \, dM$$

$$Z_n \underset{a.s}{\longrightarrow} Z \quad , \text{ say}$$

Need to show

$$(Z) \quad \sum_n P(|Z_n - Z| > \varepsilon) < \infty$$

then $\quad P(|Z_n - Z| > \varepsilon \text{ i.o}) = 0$

so $\quad Z_n \longrightarrow Z \quad$ a.s

Now (Z) is

$$\le \sum_n E|Z_n - Z| / \varepsilon^2$$

$$\le \sum_n \{E(Z_n - Z)^2\}^{1/2} / \varepsilon$$

by Cauchy Schwarz $|E(A1)| \le (E(A^2)E(1))^{1/2}$

$$= \sum_n E\left\{\int_0^t (X_n - X)^2 \, d[M]\right\}^{1/2} / \varepsilon$$

by the isometry between $\mathcal{M}_2$ & $L_2([M])$

$$< \infty \quad \text{by } (*)$$

The uniform cgce (in t) follows by a similar argument but uses an extra maximal inequality for MG's.

i.e.

$$\sup_{0\le t\le T} \left|\int_0^t X_n dM - \int_0^t X dM\right| \xrightarrow[a.s]{} 0$$

using Borel Cantelli lemma

Remark 1 In effect, (i) – (vi) define $\int_0^t X dM$ in terms of the (a.s unique) limit of simple integrals $\int_0^t X_n dM$ for $X_n \in \mathcal{S}$ and use the isometry

$$E\left(\int_0^t X dM\right)^2 = E\left(\int_0^t X^2 d[M]_t\right)$$

between the space $L_2([M])$ that are square ible wrt measure $[M]$ induced by M_t ($[M]_t$ is increasing q.v. process)

and the space of square ible MG processes $\int X dM$ $\mathcal{M}_2$

Remark 2 The MG & Doob inequalities for cts MG's $M_t \in \mathcal{M}_2$ are:

(a) $$P\left(\sup_{0\le t\le T} |M_t| > \lambda\right) < \frac{E|M_T|^p}{\lambda^p} \qquad p \ge 1$$

(MG maximal ineq.)

(b) $$E\left(\sup_{0\le t\le T} |M_t|^p\right) \le \frac{p}{p-1} E|M_T|^p \quad , \; p > 1$$

(Doob ineq.)

follow from discrete MG versions

Heuristics

(a) $X(t)$ $\mathcal{F}_t$ mble, previsible process

$M(t)$ $\mathcal{F}_t$ mble MG, $\in \mathcal{M}_2$

$$\underbrace{\int_0^t}_{\substack{\text{bet at}\\ \text{time } s}} \underbrace{X\, dM}_{\substack{\text{change in price}\\ s,\, s+ds}} \equiv \underbrace{MG}_{\substack{\text{accumulated wealth from}\\ \text{betting process}}} \in \mathcal{M}_2 \quad \text{cts square ible}$$

(b) in case $X \in \mathcal{S}$

$$= \sum_{\substack{i \geq 0 \\ t_{i+1} \leq t}} \underbrace{X(t_i)}_{\text{bet}} \underbrace{[M(t_{i+1}) - M(t_i)]}_{\text{change in price}} + X(t_{m(t)}) [M(t) - M(t_{m(t)})]$$

change in wealth over $[t_i, t_{i+1}]$

= total winnings at time t with sequence of bets at $t_0, t_1, \dots, t_{m(t)}$

= MG

Weak Convergence to Stochastic Integrals

We start with the sample covariance $\frac{1}{n}\Sigma_1^n S_{t-1} u_t$ where

$$S_t = \Sigma_1^t u_j \ , \quad S_0 = 0 \ , \quad u_t = C(L)\,\varepsilon_t$$

$$\varepsilon_t \equiv iid(0, \sigma_\varepsilon^2)$$

The simplest approach is to use "partial summation". Note

$$\Delta\left(\Sigma_1^n S_t^2\right) = \Sigma_1^n\left(\Delta S_t\, S_t + S_{t-1}\, \Delta S_t\right)$$

i.e.

$$S_n^2 = \Sigma_1^n\left(u_t S_t + S_{t-1} u_t\right)$$

$$S_n^2 = \Sigma_1^n u_t^2 + 2\,\Sigma_1^n S_{t-1} u_t$$

Thus

$$2\,\frac{1}{n}\Sigma_1^n S_{t-1} u_t = \frac{1}{n} S_n^2 - \frac{1}{n}\Sigma_1^n u_t^2$$

$$\Rightarrow B(1)^2 - \sigma^2$$

$$B(1) \equiv BM(\omega^2)$$

$$\omega^2 = \sigma_\varepsilon^2\, C(1)^2$$

$$= \sum_{h=-\infty}^{\infty} \gamma_h \qquad \gamma_h = E(u_t u_{t+h})$$

Hence

$$= \gamma_0 + 2\sum_{h=1}^{\infty}\gamma_h$$

$$\frac{1}{n}\Sigma_1^n S_{t-1} u_t \Rightarrow \frac{1}{2}\left(B(1)^2 - \gamma_0\right)$$

$$= \frac{1}{2}\left(B(1)^2 - \omega^2\right) + \frac{1}{2}\left(\omega^2 - \gamma_0\right)$$

$$= \int_0^1 B\,dB + \lambda$$

$$\lambda = \sum_{h=1}^{\infty}\gamma_h = \sum_{h=1}^{\infty} E(u_0 u_k)$$

Remark 1 In the special case where $u_t \equiv \text{iid}(0, \sigma^2)$ we have $\lambda = 0$ and

$$\frac{1}{n} \Sigma_1^n S_{t-1} u_t \Rightarrow \int_0^1 B\, dB$$

This is precisely what we would expect from the representation

$$\frac{1}{n} \Sigma_1^n S_{t-1} u_t = \Sigma_1^n \frac{S_{t-1}}{\sqrt{n}} \frac{u_t}{\sqrt{n}}$$

$$= \Sigma_1^n \int_{(j-1)/n}^{j/n} X_n(s)\, dX_n(s)$$

$$X_n(s) = \frac{S_{j-1}}{\sqrt{n}} \quad \frac{j-1}{n} \le s < j/n$$

$$= \int_0^1 X_n(s)\, dX_n(s)$$

We cannot use the CMT to deduce cgce to $\int_0^1 B\, dB$ as (*) is not a cts function of its argument for a large enough class i.e. it holds for $X_n(s)$ as Riemann Stieltjes integral but NOT for B=BM.

In fact if $\int_0^1 B\, dB$ was defined in the same way, i.e. as Riemann Stieltjes integral, we would get the wrong answer as then we would have $\int_0^1 B\, dB = B^2 \Big|_0^1 - \int B\, dB$

i.e. $\int B\, dB = \frac{1}{2} B(1)^2$!

Remark 2 We tackle the general case by first considering iid innovations. Thus suppose $S_t = \Sigma_1^t u_j$, $u_t = C(L)\varepsilon_t$ and $\varepsilon_t \equiv iid(0, \sigma_\varepsilon^2)$. Consider

$$(*) \quad \frac{1}{n}\Sigma_1^n S_{t-1}\varepsilon_t = \Sigma_1^n \frac{S_{t-1}}{\sqrt{n}}\frac{\varepsilon_t}{\sqrt{n}}, \quad \varepsilon_t \perp S_{t-1}$$

If $\varepsilon_t \equiv N(0, \sigma_\varepsilon^2)$ Then $\varepsilon_t/\sqrt{n} \equiv N(0, \sigma_\varepsilon^2/n)$ and we could write

$$\frac{\varepsilon_t}{\sqrt{n}} \equiv \sigma_\varepsilon\left[W(s+\tfrac{1}{n}) - W(s)\right] \qquad \frac{t-1}{n} \le s < \frac{t}{n}$$

$$= B_\varepsilon(s+\tfrac{1}{n}) - B_\varepsilon(s), \qquad B_\varepsilon \equiv BM(\sigma_\varepsilon^2)$$

Then we could write $(*)$ as

$$\frac{1}{n}\Sigma_1^n S_{t-1}\varepsilon_t = \Sigma_1^n \frac{S_{t-1}}{\sqrt{n}}\left[B_\varepsilon(s+\tfrac{1}{n}) - B_\varepsilon(s)\right]$$

$$= \Sigma_1^n X_n(s_t)\left[B_\varepsilon(s_t+\tfrac{1}{n}) - B_\varepsilon(s_t)\right]$$

$$(**) \qquad = \int_0^1 X_n(s)\, dB_\varepsilon(s) \qquad s_t = \frac{t-1}{n}$$

Note $(**)$ is defined as a stochastic integral because $X_n(s)$ is a simple process in $\mathcal{A}$. But

$$X_n(s) \Rightarrow B_u(s) \equiv BM(\omega^2), \quad \omega^2 = \sigma_\varepsilon^2 C(1)^2$$

So we just now take a new probability space in which we have

(i) $\frac{\varepsilon_t}{\sqrt{n}} = B_\varepsilon(s+\tfrac{1}{n}) - B_\varepsilon(s)$

(rather than just distributionally equivalent)

(ii) $X_n(s) = \frac{1}{\sqrt{n}} S_{[ns]} \underset{a.s}{\longrightarrow} B_u(s)$

(a.s. cgce rather than just weak cgce)

[this is always possible by virtue of the Skorohod representation]

Then in this new space we have

$$\frac{1}{n}\sum_1^n S_{t-1}\varepsilon_t = \int_0^1 X_n(s)\, dB_\varepsilon(s)$$

$$\underset{as}{\longrightarrow} \int_0^1 B_u(s)\, dB_\varepsilon(s)$$

$$= \int_0^1 B_u\, dB_\varepsilon$$

(ie. the stochastic integral in general)

This means in the original space where we just have distributional equivalence we get cgce in law viz

$$\frac{1}{n}\sum_1^n S_{t-1}\varepsilon_t \Rightarrow \int_0^1 B_u\, dB_\varepsilon$$

General Case

$$u_t = C(L)\varepsilon_t \quad , \quad S_t = \Sigma_1^t u_j$$

$$\frac{1}{n}\Sigma_1^n S_{t-1} u_t \Rightarrow \int_0^1 B_u dB_u + \lambda$$

$$\lambda = \Sigma_{h=1}^\infty E(u_0 u_h)$$

Proof: Use BN decomposition

$$u_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t \quad , \quad \tilde{\varepsilon}_t = \tilde{C}(L)\varepsilon_t$$

Then

$$(**) \quad \frac{1}{n}\Sigma_1^n S_{t-1} u_t = \frac{1}{n}\Sigma_1^n S_{t-1}\varepsilon_t C(1) + \frac{1}{n}\Sigma_1^n S_{t-1}(-\Delta\tilde{\varepsilon}_t)$$

$$\xrightarrow[d]{} \int_0^1 B_u dB_\varepsilon C(1)$$

$$= \int_0^1 B_u dB_u$$

Consider

$$\Delta(S_t\tilde{\varepsilon}_t) = \underbrace{\Delta S_t}_{= u_t}\tilde{\varepsilon}_t + S_{t-1}\Delta\tilde{\varepsilon}_t$$

so

$$\underbrace{\Delta\left(\Sigma_1^n S_t\tilde{\varepsilon}_t\right)}_{= S_n\tilde{\varepsilon}_n} = \Sigma_1^n u_t\tilde{\varepsilon}_t + \Sigma_1^n S_{t-1}\Delta\tilde{\varepsilon}_t$$

i.e.

$$\frac{1}{n}\Sigma_1^n S_{t-1}\Delta\tilde{\varepsilon}_t = \frac{1}{n}S_n\tilde{\varepsilon}_n - \frac{1}{n}\Sigma_1^n u_t\tilde{\varepsilon}_t$$

$$\xrightarrow[p]{} 0 - E(u_0\tilde{\varepsilon}_0)$$

Note $E(u_t \tilde{\varepsilon}_t) = E\left[\left(\sum_{j=0}^{\infty} c_j \varepsilon_{t-j}\right)\left(\sum_{j'=0}^{\infty} \tilde{c}_{j'} \varepsilon_{t-j'}\right)\right]$

$$= \sum_{j=0}^{\infty} c_j \tilde{c}_j$$

$$= \sum_{j=0}^{\infty} c_j \sum_{s=j+1}^{\infty} c_s \quad (*)$$

$$= \sum_{s=1}^{\infty} c_s \sum_{j=0}^{s-1} c_j$$

Compare

$$\sum_{h=1}^{\infty} E(u_0 u_h) = \sum_{h=1}^{\infty} E\left[\left(\sum_{j=0}^{\infty} c_j \varepsilon_{-j}\right)\left(\sum_{j'=0} c_{j'} \varepsilon_{h-j'}\right)\right]$$

$$= \sum_{h=1}^{\infty} \sum_{j=0}^{\infty} c_j c_{h+j} \qquad \begin{matrix} -j = h-j' \\ j' = h+j \end{matrix}$$

$$= \sum_{j=0}^{\infty} c_j \sum_{h=1}^{\infty} c_{h+j}$$

$$= \sum_{j=0}^{\infty} c_j \sum_{s=j+1}^{\infty} c_s$$

$$= (*)$$

ie. going back to $(**)$ we have

$$\frac{1}{n} \sum_1^n S_{t-1} u_t \Rightarrow \int_0^1 B_u dB_u + \lambda$$

with

$$\lambda = E(u_0 \tilde{\varepsilon}_0)$$

$$= \sum_{h=1}^{\infty} E(u_0 u_h)$$

Matrix Case

$$S_t = \Sigma_1^t u_j \;, \qquad u_t = C(L)\varepsilon_t, \quad \varepsilon_t \equiv iid(0, \Sigma_\varepsilon)$$

(i) $$\frac{1}{n}\Sigma_1^n S_{t-1} u_t' \Rightarrow \int_0^1 B_u dB_u' + \Lambda$$

$$\Lambda = \sum_{h=1}^{\infty} E(u_0 u_h')$$

(ii) $$\frac{1}{n}\Sigma_1^n S_t u_t' \Rightarrow \int_0^1 B_u dB_u + \Delta$$

$$\Delta = \sum_{h=0}^{\infty} E(u_0 u_h') = \Sigma_u + \Lambda$$

Proof

(i) follows in exactly the same way as before

(ii) follows directly using the decomposition

$$\frac{1}{n}\Sigma_1^n S_t u_t' = \frac{1}{n}\Sigma_1^n S_{t-1} u_t' + \frac{1}{n}\Sigma_1^n u_t u_t'$$

Remark The above implies that if we have $S_t = \Sigma_1^t u_j$, $P_t = \Sigma_1^t v_j$ then

$$\frac{1}{n}\Sigma S_{t-1} v_t \Rightarrow \int_0^1 B_u dB_v + \lambda$$

$$\lambda = \Sigma_{h=1}^{\infty} E(u_0 v_h)$$

Unit Root Asymptotics : AR(1) regression

(1) $y_t = y_{t-1} + u_t \qquad u_t = C(L)\,\varepsilon_t\,, \ \varepsilon_t \equiv iid(0,\sigma_\varepsilon^2)$

$$\hat{\alpha} = \Sigma y_t y_{t-1} / \Sigma y_{t-1}^2$$

$$\hat{\alpha} - 1 = \Sigma y_{t-1} u_t / \Sigma y_{t-1}^2$$

$$n(\hat{\alpha}-1) = \tfrac{1}{n}\Sigma y_{t-1} u_t \Big/ \tfrac{1}{n^2}\Sigma y_{t-1}^2$$

Theorem

$$\tfrac{1}{n^2}\Sigma_1^n y_{t-1}^2 \Rightarrow \textstyle\int_0^1 B_u^2$$

$$\tfrac{1}{n}\Sigma_1^n y_{t-1} u_t \Rightarrow \textstyle\int_0^1 B_u\,dB_u + \lambda$$

$$\lambda = \Sigma_{h=1}^\infty E(u_0 u_h)$$

$$n(\hat{\alpha}-1) \Rightarrow \frac{\int_0^1 B_u\,dB_u + \lambda}{\int_0^1 B_u^2}$$

Proof directly from above using

$$y_t = S_t + y_0$$

Remark If $u_t \equiv iid(0,\sigma_u^2)$, $\lambda = 0$ &

$$n(\hat{\alpha}-1) \Rightarrow \frac{\int_0^1 B_u\,dB_u}{\int B_u^2} = \frac{\int W dW}{\int W^2}$$

$$= \frac{\tfrac{1}{2}(W(1)^2 - 1)}{\int_0 W^2} = \text{D/F dist}^{n}$$

free of nuisance parameter

(2) AR(1) + Tr(p) regression I

Model: $y_t = y_{t-1} + u_t$

fitted regression $\quad y_t = \hat{\alpha}\, y_{t-1} + \sum_{i=0}^{p} \hat{b}_i t^i + \hat{u}_t$

equivalent to: $\quad \underline{y}_t = \hat{\alpha}\, \underline{y}_{t-1} + \hat{u}_t$

$$\underline{y}_t = y_t - \sum_{i=0}^{p} \hat{\gamma}_i t^i = y_t - \hat{\gamma}' x_t$$

$$\hat{\alpha} = \Sigma_1^n \underline{y}_t\, \underline{y}_{t-1} \,/\, \Sigma_1^n \underline{y}_{t-1}^2$$

$$n(\hat{\alpha} - 1) = \tfrac{1}{n} \Sigma_1^n \underline{y}_{t-1} u_t \,/\, \tfrac{1}{n^2} \Sigma_1^n \underline{y}_{t-1}^2$$

Note

$$y_t = y_{t-1} + u_t \qquad y = y_{-1} + u$$

$$\underline{y} = Q_t\, y = Q_t\, y_{-1} + Q_t\, u$$

$$= \underline{y}_{-1} + Q_t\, u$$

$$\hat{\alpha} = \underline{y}' \underline{y}_{-1} \,/\, \underline{y}_{-1}' \underline{y}_{-1}$$

$$\hat{\alpha} - 1 = \underline{y}_{-1}' (\underline{y} - \underline{y}_{-1}) \,/\, \underline{y}_{-1}' \underline{y}_{-1}$$

$$= \underline{y}_{-1}' Q_t u \,/\, \underline{y}_{-1}' \underline{y}_{-1}$$

$$= \underline{y}_{-1}' u \,/\, \underline{y}_{-1}' \underline{y}_{-1}$$

as $Q_t^2 = Q_t$ = orthogonal projector

(149b)

Note

$$\frac{1}{n^2}\underline{y}_{-1}'\underline{y}_{-1} = \frac{1}{n^2}\sum_1^n \underline{y}_{t-1}^2 \Rightarrow \int_0^1 \underline{B}_p^2$$

$$B_p(r) = B(r) - \left(\int_0^1 B\underline{X}\right)\left(\int_0^1 \underline{X}\,\underline{X}'\right)^{-1}\underline{X}(r)$$

$$= \text{detrended BM}$$

$$\underline{X}(r)' = (1, r, \cdots, r^p)$$

$$\frac{1}{n}\underline{y}_{-1}'u = \frac{1}{n}y_{-1}'u - \hat{\gamma}'\frac{X'u}{n}$$

$$\Rightarrow \left(\int_0^1 B\,dB + \lambda\right) - \bar{\gamma}\int_0^1 \underline{X}\,dB$$

$$\bar{\gamma}' = \left(\int_0^1 BX\right)\left(\int_0^1 XX'\right)^{-1}$$

$$= \int_0^1 \underline{B}_p\,dB + \lambda$$

with

$$\underline{B}_p(r) = B(r) - \int_0^1 B\underline{X}\left(\int_0^1 \underline{X}\,\underline{X}'\right)^{-1}\underline{X}(r)$$

Note

$$\frac{1}{n}\hat{\gamma}'X'u = \frac{1}{n}y'X(X'X)^{-1}X'u$$

$$= \left(\frac{1}{n^{3/2}}\sum y_t x_t' D_n^{-1}\right)\left(\frac{1}{n}\sum D_n^{-1}x_t x_t' D_n^{-1}\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_t D_n^{-1} x_t u_t\right)$$

$$\Rightarrow \left(\int_0^1 B\underline{X}\right)\left(\int_0^1 \underline{X}\,\underline{X}'\right)^{-1}\left(\int_0^1 \underline{X}\,dB\right)$$

$$D_n^{-1}x_{[nr]} \to X(r) \qquad D_n = \mathrm{diag}(1, n, \ldots, n^p)$$

Thus

$$(*) \quad n(\hat{\alpha}-1) \Rightarrow \left[\int_0^1 \underline{B}_p \, dB + \lambda\right]\left[\int_0^1 \underline{B}_p^2\right]^{-1}$$

where

$$\underline{B}_p(r) = B(r) - \int_0^1 BX' \left(\int_0^1 XX'\right)^{-1} X(r)$$

$$X(r) = \begin{pmatrix} 1 \\ r \\ \vdots \\ r^p \end{pmatrix}$$

= detrended Brownian motion.

(3) t-ratios

$$t = (\hat{\alpha}-1)/s_{\hat{\alpha}} \quad , \quad s_{\hat{\alpha}}^2 = s^2 / \sum_1^n \underline{y}_{t-1}^2$$

$$s^2 = \frac{1}{n}\sum(\underline{y}_t - \hat{\alpha}\underline{y}_{t-1})^2$$

$$\rightarrow_p \sigma^2$$

$$t = \frac{n(\hat{\alpha}-1)}{n s_{\hat{\alpha}}} = \frac{n(\hat{\alpha}-1)}{\left(\frac{s^2}{\frac{1}{n^2}\sum \underline{y}_{t-1}^2}\right)^{1/2}}$$

$$\Rightarrow \frac{\int_0^1 \underline{B}_p \, dB + \lambda}{\left(\frac{\sigma^2}{\int_0^1 \underline{B}_p^2}\right)^{1/2}}$$

$$= \frac{\int_0^1 \underline{B}_p \, dB + \lambda}{\sigma / \left(\int_0^1 \underline{B}_p^2\right)^{1/2}}$$

if $u_t \equiv iid(0, \sigma_u^2)$ we get $\lambda = 0$

$B = BM(\sigma_u^2)$ and

$$t = \frac{n(\hat{\alpha}-1)}{n.s_{\hat{\alpha}}} \Rightarrow \frac{\int_0^1 \underline{B}_p\, dB}{\sigma_u \left(\int_0^1 \underline{B}_p^2\right)^{1/2}}$$

$$= \frac{\int_0^1 \underline{W}_p\, dW}{\left(\int_0^1 \underline{W}_p^2\right)^{1/2}}$$

nuisance parameter free

$W = BM(1)$

$\underline{W}_p =$ detrended $BM(1)$

(3) AR + Tr regression II

(*) Model: $y_t = \beta_0 + \beta_1 t + \ldots + \beta_{p-1} t^{p-1} + y_{t-1} + u_t$

implies
$$\begin{cases} y_t = \beta_0 t + \beta_1 t^2 + \ldots + \beta_{p-1} t^p + y_t^0 \\ y_t^0 = y_{t-1}^0 + u_t \end{cases}$$

Thus if (*) is true model note that

$$Q_t y = Q_t y^0$$ (using Q_t operator which is the same as before i.e. $= I - X(X'X)^{-1}X'$, $x_t' = (1, \ldots, t^p)$)

and hence when we regress on a p'th degree trend in (*) we get

$$\hat{\alpha} = \underline{y}'\underline{y}_{-1} / \underline{y}_{-1}'\underline{y}_{-1} = y'Q_t y_{-1} / y_{-1}' Q_t y_{-1}$$

$$= y^{0\prime} Q_t y_{-1}^0 / y_{-1}^{0\prime} Q_t y_{-1}^0$$

$$= \underline{y}^{0\prime} \underline{y}_{-1}^0 / \underline{y}^{0\prime} \underline{y}_{-1}^0$$

Hence for model (*) we get the same limit theory viz

$$n(\hat{\alpha}-1) \Rightarrow \frac{\int_0^1 \underline{B}_p dB + \lambda}{\int_0^1 \underline{B}_p^2}$$

$$t_\alpha = \frac{(\hat{\alpha}-1)}{s_{\hat{\alpha}}} \Rightarrow \frac{\int_0^1 \underline{B}_p dB + \lambda}{\sigma_u \left(\int_0^1 \underline{B}_p^2\right)^{1/2}}$$

Remarks

(i) These limit distributions form the basis of all our unit root tests & tests of stochastic trends vs deterministic trends.

(ii) Note that to eliminate the deterministic components we must detrend data using $x_t = (1, t, \ldots, t^p)$ i.e. full degree $= p$ even though in model (*) we have trend degree only of $p-1$. This is because, when there is a unit root, the data have a trend of degree p.

(iii) This means that in unit root tests we need to augment the trend degree in the regression in order to achieve the invariant statistics above (i.e. invariant to parameters of trend). This is important in what follows.

Unit Root Tests

Model:

$$(*) \quad y_t = \beta_0 + \beta_1 t + \dots + \beta_{p-1} t^{p-1} + \alpha y_{t-1} + u_t$$

H_0: test $\alpha = 1$: presence of unit root with maintained trend of degree $p-1$

H_1: $\alpha < 1$ (trend stationarity)

i.e. test unit root against trend (degree t^p) stationary

Under H_0 we have, as on p151

$$\begin{cases} y_t = \beta_0 t + \beta_1 t^2 + \dots + \beta_{p-1} t^p + y_t^0 \\ y_t^0 = y_{t-1}^0 + u_t \end{cases}$$

Test statistics || estimate α in model $(*)$ augmented with trend of degree p i.e. $y_t = \sum_{i=0}^{p} \beta_i t^i + \alpha y_{t-1} + u_t$ achieves invariance

A. $u_t \equiv iid(0, \sigma^2)$ DF framework

(i) coeff test: $n(\hat{\alpha}-1) \Rightarrow \dfrac{\int_0^1 \underline{B}_p dB}{\int \underline{B}_p^2} \equiv \dfrac{\int_0^1 \underline{W}_p dW}{\int_0^1 \underline{W}_p^2}$

(ii) t-ratio test: $t(\hat{\alpha}) \Rightarrow \dfrac{\int_0^1 B_p dB}{\sigma\left(\int_0^1 \underline{B}_p^2\right)^{1/2}} \equiv \dfrac{\int_0^1 \underline{W}_p dW}{\left(\int_0^1 \underline{W}_p^2\right)^{1/2}}$

nuisance parameter free functionals of BM

B $u_t = C(L)\varepsilon_t$, $\varepsilon_t \equiv iid(0, \sigma_\varepsilon^2)$ or heterogeneity allowed

(i) coeff. test

$$Z(\hat{\alpha}) = n(\hat{\alpha}-1) - \frac{\hat{\lambda}}{n^{-2}\sum_1^n \underline{y}_{t-1}^2}$$

(ii) t-ratio test

$$Z(t) = \frac{\hat{\sigma}}{\hat{\omega}} t(\hat{\alpha}) - \frac{\hat{\lambda}}{\hat{\omega}\left(n^{-2}\Sigma_1^n \underline{y}_{t-1}^2\right)^{1/2}}$$

Where

$\hat{\lambda}$ kernel based consistent estimator of $\lambda = \Sigma_{h=1}^{\infty} E(u_0 u_h)$

$\hat{\omega}^2$ kernel based consistent estimator of $\omega^2 = \Sigma_{h=-\infty}^{\infty} E(u_0 u_h)$
$= \text{lr var}(u_t)$

$\hat{\sigma}^2$ consistent estimator of $E(u_t^2)$

Theorem

$$Z(\hat{\alpha}) \Rightarrow \frac{\int_0^1 \underline{B}_p dB}{\int_0^1 \underline{B}_p^2} \equiv \frac{\int_0^1 \underline{W}_p dW}{\int_0^1 \underline{W}_p^2}$$

$$Z(t) \Rightarrow \frac{\int_0^1 \underline{B}_p dB}{\omega\left(\int_0^1 \underline{B}_p^2\right)^{1/2}} \equiv \frac{\int_0^1 \underline{W}_p dW}{\left(\int_0^1 \underline{W}_p^2\right)^{1/2}}$$

nuisance parameter free D/F limit distn's

Remark $Z(\hat{\alpha})$, $Z(t)$ tests are semi-parametric tests for a unit root. They allow for general error processes u_t & even allow for heterogeneity. Phillips (1987), Phillips & [illegible]

Proof

$$Z(\hat{\alpha}) = n(\hat{\alpha}-1) - \hat{\lambda} / \tfrac{1}{n^2}\Sigma_1^n \underline{y}_{t-1}^2$$

$$\Rightarrow \frac{\int_0^1 dB + \lambda}{\int_0^1 \underline{B}_p^2} - \frac{\lambda}{\int_0^1 \underline{B}_p^2} \equiv \frac{\int_0^1 \underline{B}_p \, dB}{\int_0^1 \underline{B}_p^2}$$

$$Z(t) = \frac{\hat{\sigma}}{\hat{\omega}} t(\hat{\alpha}) - \frac{\hat{\lambda}}{\hat{\omega}\left(n^{-2}\Sigma_1^n \underline{y}_{t-1}^2\right)^{1/2}}$$

$$\Rightarrow \frac{\sigma}{\omega} \frac{\int \underline{B}_p dB + \lambda}{\sigma\left(\int_0^1 \underline{B}_p^2\right)^{1/2}} - \frac{\lambda}{\omega\left(\int_0^1 \underline{B}_p^2\right)^{1/2}}$$

$$= \frac{\int \underline{B}_p \, dB}{\omega\left(\int_0^1 \underline{B}_p^2\right)^{1/2}} \equiv \frac{\int \underline{W}_p \, dW}{\left(\int_0^1 \underline{W}_p^2\right)^{1/2}}$$

Power Functions of these tests

Remark First note that these are one sided tests of $H_0: \alpha = 1$ against $H_1: \alpha < 1$ (i.e. difference stationarity vs trend stationarity)

Local Alternative

$$H_1 : \alpha = \exp\left(\tfrac{1}{n} c\right)$$

As we have seen earlier under local alternative we have

$$\tfrac{1}{\sqrt{n}} y_t \Rightarrow J_c(r) = \int_0^r \exp((r-s)c)\, dB(s)$$

$$t = [nr]$$

Similarly, we find

$$\frac{1}{\sqrt{n}}\, y_t \Rightarrow \underline{J_c}(r) = J_c(r) - \left(\int_0^1 J_c \underline{X}'\right)\left(\int_0^1 \underline{X}\underline{X}'\right)^{-1}\underline{X}(r)$$

$\underline{X}(r) = (1, r, \ldots, r^p)$

= detrended diffusion process

Then we get

<u>Theorem</u> Under H_1 local alternative to H_0 we have:

$$Z(\hat{\alpha}) \Rightarrow \frac{\int_0^1 \underline{J_c}\, dB}{\int_0^1 \underline{J_c}^2} + c$$

$$Z(t) \Rightarrow \frac{\int_0^1 \underline{J_c}\, dB + c\left(\int_0^1 \underline{J_c}^2\right)}{\omega\left(\int_0^1 \underline{J_c}^2\right)^{1/2}}$$

<u>Remarks</u> These limits are identical to those that apply for iid errors u_t (i.e. the D/F limit distributions under the alternative H_1). Thus, there is no loss of power asymptotically in making these semi-parametric corrections.

Note also: $J_c(r) = \int_0^r e^{(r-s)c} dB = \omega \int_0^r e^{(r-s)c} dW$

$=: J_c^W(r)$, say

so that

$$\underline{J_c}(r) = \omega\, \underline{J_c}^W(r)$$

$$\frac{\int_0^1 \underline{J_c}\, dB}{\int_0^1 \underline{J_c}^2} \equiv \frac{\omega^2 \int_0^1 \underline{J_c}^W dW}{\ldots} = \frac{\int_0^1 \underline{J_c}^W dW}{\ldots}$$

Proofs As before, but now with $\alpha = e^{\frac{1}{n}c}$, we set

(a) $n(\hat{\alpha}-\alpha) = \frac{1}{n}\underline{y}_{-1}'u / \frac{1}{n^2} y_{-1}' y_{-1}$

$$\Rightarrow \left(\int_0^1 \underline{J}_c dB + \lambda\right) / \left(\int_0^1 \underline{J}_c^2\right)$$

Hence

$$n(\hat{\alpha}-1) \Rightarrow c + \left(\int_0^1 J_c dB + \lambda\right) / \left(\int_0^1 \underline{J}_c^2\right)^{-1}$$

Then

$$Z(\hat{\alpha}) = n(\hat{\alpha}-1) - \hat{\lambda} / n^{-2} \Sigma_1^n \underline{y}_{t-1}^2$$

$$\Rightarrow c + \int_0^1 \underline{J}_c dB \Big/ \int_0^1 \underline{J}_c^2$$

(b) $t_\alpha = \dfrac{n(\hat{\alpha}-1)}{n S_{\hat{\alpha}}} \Rightarrow \dfrac{c + \left(\int_0^1 \underline{J}_c dB + \lambda\right) / \int_0^1 \underline{J}_c^2}{\sigma \left(\int_0^1 \underline{J}_c^2\right)^{-1/2}}$

$$Z(t) = \frac{\sigma}{\omega} t_\alpha - \frac{\hat{\lambda}}{\omega\left(n^{-2}\Sigma \underline{y}_{t-1}^2\right)^{+1/2}}$$

$$\Rightarrow \frac{c + \left(\int_0^1 \underline{J}_c dB + \lambda\right) / \int_0^1 \underline{J}_c^2}{\omega\left(\int_0^1 J_c^2\right)^{-1/2}} - \frac{\lambda}{\omega\left(\int_0^1 \underline{J}_c^2\right)^{1/2}}$$

$$= \frac{c\left(\int_0^1 \underline{J}_c^2\right) + \int_0^1 \underline{J}_c dB}{\omega\left(\int_0^1 \underline{J}_c^2\right)^{1/2}}$$

$$= \frac{c}{\omega}\left(\int_0^1 \underline{J}_c^2\right)^{1/2} + \frac{\int_0^1 \underline{J}_c dB}{\omega\left(\int_0^1 \underline{J}_c^2\right)^{1/2}}$$

Other Unit-Root Tests

1. The ADF test

model $y_t = y_{t-1} + u_t$ $\quad u_t = C(L)\varepsilon_t$

$\varepsilon_t \equiv iid(0, \sigma^2)$

assume $C(L)^{-1} = a(L)$ invertible
then model becomes in AR form

AR Representation

$$(1-L)\,a(L)\,y_t = \varepsilon_t$$

$$a(L) = \Sigma_0^p a_i L^i = 1 - \varphi(L), \text{ say}$$

i.e.

$$\Delta y_t = \Sigma_{i=1}^p \varphi_i \Delta y_{t-i} + \varepsilon_t$$

ADF test

$$\Delta y_t = a\, y_{t-1} + \Sigma_{i=1}^p \varphi_i \Delta y_{t-i} + \varepsilon_t$$

test $H_0: a = 0$ by t-ratio test

$$t = \hat{a} / s_{\hat{a}}$$

Theorem

(a) If $p \to \infty$, $p = o(n^{1/3})$ as $n \to \infty$, then ADF test is asymptotically valid i.e. has correct asymptotic size if $u_t = C(L)\varepsilon_t$ and has AR repⁿ $a(L)\,u_t = \varepsilon_t$ with $a(L)$ possibly ∞ dim.

(b) $$t \Rightarrow \frac{\int_0^1 W dW}{\left(\int_0^1 W^2\right)^{1/2}} \equiv \text{same lt distⁿ as } Z(t) \text{ test}$$

ADF test with trends

$$\Delta y_t = \beta_0 + \beta_1 t + \dots + \beta_p t^p + a\, y_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta y_{t-i} + \varepsilon_t$$

in this case we have the limit theory

$$t \Rightarrow \frac{\int_0^1 \underline{W}_p\, dW}{\left(\int_0^1 \underline{W}_p^2\right)^{1/2}} \equiv \text{same as } Z(t) \text{ test again}$$

Power function

under local alternatives $H_1 : a = e^{\frac{1}{n}c} - 1 \sim \frac{1}{n}c$

we have

$$t \Rightarrow \frac{\int_0^1 \underline{J}_c^W\, dW}{\left(\int_0^1 \underline{J}_c^{W\,2}\right)^{1/2}} \equiv \text{same power func as the } Z(t) \text{ test}$$

Remark

- The ADF & $Z(t)$ tests are asymptotically equivalent
- there is no coeff-based ADF test
- simulation studies have found

  ADF test less subject to size distortions especially when

  $u_t = \varepsilon_t - \theta \varepsilon_{t-1}, \quad \theta \sim 1$

  (Schwert, 1988, 1989; Phillips & Perron, 1988)

  ADF test generally has much less power than $Z(a)$, $Z(t)$

- Automated "optimal" bandwidth choices, make

Alternative Formulation

$$(*) \quad y_t = b(L)\, y_{t-1} + \varepsilon_t$$

Apply BN: $\quad b(L) = b(1) + (L-1)\,\tilde{b}(L)$

$$\tilde{b}(L) = \sum_{j=0}^{p-1} \tilde{b}_j L^j$$

$$\tilde{b}_j = \sum_{s=j+1}^{p} b_s$$

Then (*) is

$$y_t = b(1)\, y_{t-1} - \tilde{b}(L)\,\Delta y_{t-1} + \varepsilon_t$$

$$= b(1)\, y_{t-1} + \varphi(L)\,\Delta y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \underbrace{(b(1)-1)}_{=a}\, y_{t-1} + \varphi(L)\,\Delta y_{t-1} + \varepsilon_t$$

test $H: a = 0$ equivalent to $b(1) = 1$

if $b(1) = 1$

$$b(L) = 1 + (L-1)\,\tilde{b}(L)$$

and

$$1 - L\,b(L)$$

$$= (1-L) + (L-1)\,\tilde{b}(L)\,L$$

has root at unity.

Efficient Unit Root Tests

- When there is no trend / intercept in the model the D/F test is very close to being optimal – i.e. for local alternatives it comes close to having optimal power in relation to the best test computed for the known local alternative using the Neyman Pearson lemma (This is the so-called power envelope) and is based on the likelihood ratio

$$\frac{L_{H_0}(\alpha=1)}{L_{H_1}(\alpha=1+c/n,\ \text{given } c)}$$

- However, when there is a trend in the model the D/F & Z tests rely on trend removal by regression & it turns out that efficiency can be gained by "improving" the trend removal process.

- The reason is that the Grenander Rosenblatt result on efficiency of OLS regression trend removal no longer holds

Grenander-Rosenblatt Theorem

$$y_t = \beta' x_t + u_t \qquad u_t \equiv \mathrm{sd}(0, f_{uu}(\lambda))$$

$$x_t = \text{polynomial trend} = (1, t, \dots, t^p)$$

$\hat{\beta} = \text{OLS}$ is asymptotically efficient estimator of β if $f_{uu}(\lambda)$ is cts at $\lambda = 0$ i.e. $f_{uu}(0) < \infty$.

OLS detrending with near-unit roots

model ① $y_t = \beta' x_t + u_t$, $\quad u_t = \alpha u_{t-1} + e_t$

$$\alpha = \exp\left(\tfrac{1}{n} c\right)$$

$$x_t' = (t, t^2, \ldots, t^p)$$

note: cannot identify intercept in ① e.g. $\hat{\beta}_0 = \bar{y}$ diverges as $\frac{1}{n^{1/2}}\bar{y} = \frac{1}{n}\sum_1^n \frac{y_t}{\sqrt{n}} \to \int B$

$$D_n^{-1} x_t = \left(\tfrac{t}{n}, \left(\tfrac{t}{n}\right)^2, \ldots, \left(\tfrac{t}{n}\right)^p\right)', \quad D_n = \mathrm{diag}(n, n^2, \ldots, n^p)$$

OLS asymptotics (if $\alpha = \exp(\frac{1}{n}c)$)

$$n^{-1/2} D_n(\hat{\beta} - \beta) \Rightarrow \left(\int_0^1 X(r)X(r)'dr\right)^{-1}\left(\int_0^1 X(r)J_c(r)dr\right)$$

Proof

$$\hat{\beta} = (X'X)^{-1}X'y, \quad \hat{\beta} - \beta = (X'X)^{-1}X'u$$

$$n^{-1/2} D_n(\hat{\beta} - \beta) = \left[D_n^{-1} X'X D_n^{-1}\right]^{-1} D_n^{-1} \frac{X'u}{n^{1/2}}$$

$$= \left[\tfrac{1}{n} D_n^{-1} X'X D_n^{-1}\right]^{-1} \tfrac{1}{n} D_n^{-1} \frac{X'u}{\sqrt{n}}$$

Now

$$\tfrac{1}{n} D_n^{-1} X'X D_n = \tfrac{1}{n}\sum_1^n D_n^{-1} x_t x_t' D_n^{-1}$$

$$\to \int_0^1 X(r)X(r)'$$

$$X(r)' = (r, r^2, \ldots, r^p)$$

$$\tfrac{1}{n} D_n^{-1} \frac{X'u}{\sqrt{n}} = \tfrac{1}{n}\sum_1^n D_n^{-1} x_t \frac{u_t}{\sqrt{n}} \qquad \frac{u_{[nr]}}{\sqrt{n}} \Rightarrow J_c(r)$$

$$\Rightarrow \int_0^1 X(r) J_c(r) dr \qquad = \int_0^r e^{(r-s)c} dB_e$$

So

$$n^{-1/2} D_n(\hat{\beta} - \beta) \Rightarrow \left[\int_0^1 X(r)X(r)'\right]^{-1}\left[\int_0^1 X(r)J_c(r)\right]$$

Note

$$\int_0^1 X J_c \equiv N(0, V_c)$$

$$V_c = \omega_e^2 \int_0^1 \int_0^1 X(r)\, e^{(r+s)c}\, \frac{1}{2c}\left(1 - e^{-2c(r\wedge s)}\right) X(s)'\, dr\, ds$$

as:

$$V_c = \int_0^1 \int_0^1 X(r)\, E\left(J_c(r)\, J_c(s)\right) X(s)'\, ds$$

$$E\left[\int_0^r e^{(r-p)c}\, dB_e(p) \int_0^s e^{(s-q)c}\, dB_e(q)\right]$$

$$= e^{(r+s)c}\, \omega^2 \int_0^{r\wedge s} e^{-pc-pc}\, dp \qquad E\left(dB_e(p)\, dB_e(q)\right) = \omega_e^2 \begin{cases} dp & p = q \\ 0 & \text{elsewhere} \end{cases}$$

$$= e^{(r+s)c} \left[\frac{e^{-2pc}}{-2c}\right]_0^{r\wedge s} \omega^2$$

$$= e^{(r+s)c}\, \frac{1}{2c}\left[1 - e^{-2c(r\wedge s)}\right] \omega_e^2$$

<u>GLS detrending</u>

$$\tilde{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$$

where $\tilde{x}_t = x_t - \alpha x_{t-1} = (1-\alpha L)x_t$

$\tilde{y}_t = y_t - \alpha y_{t-1} = (1-\alpha L)y_t$

i.e. OLS on transformed model under local alternative $H_1 : \alpha = \exp(\frac{1}{n}c)$

$$(1-\alpha L)y_t = (1-\alpha L)x_t'\beta + (1-\alpha L)u_t$$

$$\tilde{y}_t = \tilde{x}_t \beta + e_t$$

GLS asymptotics

$$\tilde\beta = (\tilde X'\tilde X)^{-1}\tilde X'\tilde y \qquad \tilde\beta-\beta=(\tilde X'\tilde X)^{-1}\tilde X'e$$

Component asymptotics

$$\tilde x_t = (1-\alpha L)x_t = \left(1-L-\tfrac{c}{n}L\right)x_t \qquad \alpha = 1+\tfrac{c}{n}$$
$$= \left(\Delta - \tfrac{c}{n}L\right)x_t$$
$$= \Delta_c x_t \text{ , say}$$

take t^i element

$$\frac{1}{n^{i-1}}\Delta_c t^i = \frac{1}{n^{i-1}}\Delta t^i - \frac{c}{n}\frac{1}{n^{i-1}}(t-1)^i$$
$$= n\,\Delta\left(\frac{t}{n}\right)^i - c\left(\frac{t-1}{n}\right)^i$$
$$\sim i\left(\frac{t}{n}\right)^{i-1} - c\left(\frac{t}{n}\right)^i \text{ , as } \Delta\left(\frac{t}{n}\right)^i \sim i\left(\frac{t}{n}\right)^{i-1}\frac{1}{n}$$

let $t = [nr]$

$$\to i r^{i-1} - c r^i$$

Then, we have with $F_n = \mathrm{diag}(1, n, \ldots, n^{p-1})$

$$F_n^{-1}\tilde x_{[nr]} = \left[\left(\frac{1}{n^{i-1}}\Delta_c t^i\right)_i\right] \to \left[\left(i r^{i-1} - c r^i\right)\right]$$
$$= X_c(r) \text{ , say}$$

Then

$$\frac{1}{n}F_n^{-1}\tilde X'\tilde X F_n^{-1} = \frac{1}{n}\sum_1^n F_n^{-1}\tilde x_t\tilde x_t' F_n^{-1}$$
$$\to \int_0^1 X_c(r)X_c(r)'$$

Similarly, we get

$$\frac{1}{\sqrt{n}} F_n^{-1} \sum_1^n \tilde{x}_t e_t = \sum_1^n F_n^{-1} \tilde{x}_t \frac{e_t}{\sqrt{n}}$$

$$\Rightarrow \int_0^1 X_c(r)\, dB_e$$

$$e_t \equiv wd(0, f_{ee}(\lambda))$$

$$\omega_e^2 = 2\pi f_{ee}(0)$$

Thus, we have:

<u>GLS Asymptotics</u>

$$\sqrt{n}\, F_n (\tilde{\beta} - \beta) \Rightarrow \left(\int_0^1 X_c(r) X_c(r)'\right)^{-1} \left(\int_0^1 X_c(r)\, dB_e\right)$$

$$\equiv N\left(0,\ \omega_e^2 \left(\int_0^1 X_c(r) X_c(r)'\right)^{-1}\right)$$

<u>OLS / GLS Comparison</u>

Let $p = 1$ then we have $X(r) = r$
$D_n = n,\ F_n = 1$

OLS

$$n^{-1/2} D_n (\hat{\beta} - \beta) \Rightarrow \left(\int_0^1 X X'\right)^{-1} \left(\int_0^1 X J_c\right)$$

$$\equiv N\left(0,\ \left(\int_0^1 X X'\right)^{-1} \int_0^1\int_0^1 X(r)\, e^{(r+s)c} \frac{1}{2c}\left[1 - e^{-2c(r\wedge s)}\right] X(s)'\, dr\, ds \left(\int_0^1 X X'\right)^{-1}\right)$$

$$\equiv V_{OLS}$$

GLS

$$n^{1/2} F_n (\tilde{\beta} - \beta) \Rightarrow N\left(0,\ \omega_e^2 \left(\int_0^1 X_c X_c'\right)^{-1}\right)$$

$$= V_{GLS}$$

Now $\rho = 1$ implies $X(r) = r$ so

$$\int_0^1 X^2 = \int_0^1 r^2 = 1/3 \; ; \quad X_c(r) = 1 - cr$$

$$\sqrt{n}\,(\hat{\beta} - \beta) \Rightarrow 3 \int_0^1 r\, J_c(r)\, dr$$

$$\sqrt{n}\,(\tilde{\beta} - \beta) \Rightarrow \frac{\int_0^1 (1-cr)\, dB_e}{\int_0^1 (1-cr)^2}$$

Now calculate results for null hypothesis $H_0: \alpha = 1$ i.e. $c = 0$

$$\sqrt{n}\,(\hat{\beta} - \beta) \Rightarrow 3 \int_0^1 r\, B_e(r)\, dr \qquad J_c(r) = B_e \text{ when } c = 0$$

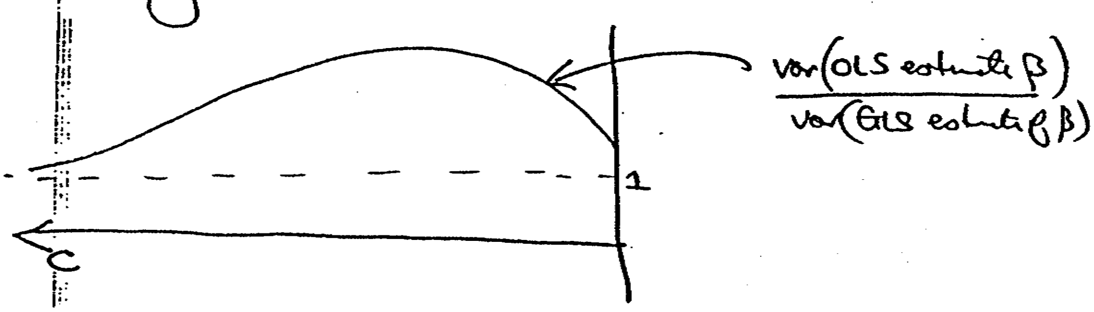
$$\equiv N\left(0, \frac{6\sigma^2}{5}\right)$$

$$\sqrt{n}\,(\tilde{\beta} - \beta) \Rightarrow \int_0^1 dB_e = B_e(1) \equiv N(0, \sigma^2)$$

i.e. OLS has $1/5 = 20\%$ more variability than GLS.

i.e. GLS detrending is more efficient for the estimation of the trend

Similar gains when $c < 0$ $c \neq 0$



(164b

$$\mathrm{var}\left(3\int_0^1 B_e(r)\,dr\right)$$

$$= 9\,E\left\{\int_0^1 r\,B_e(r)\int_0^1 s\,B_e(s)\,ds\right\}$$

$$= 2\times 9\left(\int_0^1 r\int_0^r s^2\,ds\,dr\right)\sigma_e^2$$

$$E\left(B_e(r)\,B_e(s)\right) = \sigma_e^2\, r\wedge s$$

$$= \sigma_e^2 s \quad \text{for } s<r$$

$$= 18\left(\int_0^1 r\,\frac{r^3}{3}\,dr\right)\sigma_e^2$$

$$= \frac{18}{3}\left(\int_0^1 r^4\,\sigma_e^2\right)$$

$$= 6\left(\frac{r^5}{5}\Big|_0^1\right)\sigma_e^2$$

$$= \frac{6}{5}\sigma_e^2$$

GLS-detrended Unit Root Tests

We can take advantage of the additional efficiency of the GLS trend estimator $\tilde{\beta}$ in the construction of unit root tests. Our model is

$$(*) \quad y_t = \beta' x_t + y^0_{t,c} \qquad y^0_{t,c} = \alpha\, y^0_{t-1,c} + e_t$$
$$\alpha = \exp\left(\tfrac{1}{n} c\right)$$

null $H_0: \alpha = 1$, $c = 0$

local alternative $H_1: c < 0$

<u>Proposal</u> (Elliot, Rothenberg & Stock)

- Estimate β in $(*)$ with a prespecified value of c (they use $c = -13.5$, where the point optimal test is tangent to the power envelope at power of 50%) and GLS detrending

  $\tilde{\beta} \to_p \beta \qquad \forall c$ finite but is more efficient than OLS especially around $c \sim -13.5$

  Then construct usual Z_α, Z_t, ADF or DF tests.

- Note "c" is NOT estimated – it is postulated and therefore does not influence the asymptotic theory. But there are differences in the asymptotics

- Note that $\underset{\sim}{y} = y - X\tilde{\beta} = y - X(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$ is not a projection residual directly – but is asymptotically equivalent to true GLS viz

True Model (null) $\qquad H_0: \alpha = 1, \; c = 0$

$$y_t = \beta' x_t + u_t \qquad \Delta u_t = e_t = C(L)\varepsilon_t$$

GLS detrending

$$\tilde{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$$

$\tilde{x}_t = x_t - \alpha x_{t-1}$

$\tilde{y}_t = y_t - \alpha y_{t-1}$

$\alpha = \exp\left(\frac{1}{n}\bar{c}\right)$

$\bar{c}$ fixed & given

GLS asymptotics under null

$$\tilde{y}_t = \beta' \tilde{x}_t + \tilde{u}_t$$

$$\begin{aligned}\tilde{u}_t &= u_t - \alpha u_{t-1} \\ &= \Delta_{\bar{c}} u_t \\ &= \Delta u_t - \frac{\bar{c}}{n} u_{t-1} \\ &= e_t - \frac{\bar{c}}{n} u_{t-1}\end{aligned}$$

$$\begin{aligned}\tilde{\beta} - \beta &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{u} \\ &= (\tilde{X}'\tilde{X})^{-1}\left(\tilde{X}'e - \frac{\bar{c}}{n}\tilde{X}'u\right)\end{aligned}$$

$$\sqrt{n}F_n(\tilde{\beta} - \beta) = \left(\frac{1}{n}F_n^{-1}\tilde{X}'\tilde{X}F_n^{-1}\right)^{-1}\left(\frac{1}{\sqrt{n}}F_n^{-1}\tilde{X}'e - \frac{\bar{c}}{n^{3/2}}F_n^{-1}\tilde{X}'u\right)$$

$$\Rightarrow \left(\int_0^1 X_{\bar{c}} X_{\bar{c}}'\right)^{-1}\left(\int_0^1 X_{\bar{c}}\, dB_e - \bar{c}\int_0^1 X_{\bar{c}} B_e\right)$$

Note

$\tilde{\beta} \to_p \beta$ so $\tilde{\beta}$ is consistent under the null (and under the alternative – where it is more efficient than OLS detrending)

D/F regression with detrended data by GLS

(1) We work with

$$\underset{\sim}{y} = y - X\tilde{\beta}$$

i.e.

$$\frac{1}{\sqrt{n}} \underset{\sim}{y}_{[nr]} = \frac{1}{\sqrt{n}} y_{[nr]} - \frac{1}{\sqrt{n}} x_{[nr]}'\left(\tilde{X}'\tilde{X}\right)\tilde{X}\tilde{y}$$

$$= \frac{1}{\sqrt{n}} y_{[nr]} - \frac{1}{\sqrt{n}} x_{[nr]}'\left[\beta + \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{u}\right]$$

$$= \frac{1}{\sqrt{n}} u_{[nr]} - \frac{1}{\sqrt{n}} x_{[nr]}'\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{u}$$

$$= \frac{1}{\sqrt{n}} u_{[nr]} - \frac{1}{\sqrt{n}} x_{[nr]}' F_n^{-1}\left(\frac{1}{n} F_n^{-1}\tilde{X}'\tilde{X}F_n^{-1}\right)^{-1}\frac{1}{n} F_n^{-1}\tilde{X}'\tilde{u}$$

$$= \frac{1}{\sqrt{n}} u_{[nr]} - \frac{1}{n} x_{[nr]}' F_n^{-1}\left(\frac{1}{n} F_n^{-1}\tilde{X}'\tilde{X}F_n^{-1}\right)^{-1}\frac{1}{\sqrt{n}} F_n^{-1}\tilde{X}'\tilde{u}$$

as $D_n^{-1} x_{[nr]} \to X(r)$ & $D_n = nF_n$

$$\Rightarrow B_e(r) - X(r)\left(\int_0^1 X_{\bar{c}} X_{\bar{c}}'\right)^{-1}\left(\int_0^1 X_{\bar{c}}\, dB_e - \bar{c}\int_0^1 X_{\bar{c}} B_e\right) + O_p\left(\tfrac{1}{n}\right)$$

$= \underset{\sim}{B}_e(r)$, say

(2) Next do DF on $\underset{\sim}{y}$ data, giving unit root test statistics with the GLS detrended data.

Thus

$$\tilde{\alpha} = \underset{\sim}{y}'\underset{\sim}{y}_{-1} / \underset{\sim}{y}_{-1}'\underset{\sim}{y}_{-1}$$

$$n(\tilde{\alpha}-1) = n^{-1}\underset{\sim}{y}_{-1}'\,\Delta \underset{\sim}{y} / \underset{\sim}{y}_{-1}'\underset{\sim}{y}_{-1}/n^2$$

Consider

$$\Delta \underset{\sim}{y} = \Delta y - \Delta X\tilde{\beta} = \Delta(y - X\tilde{\beta})$$
$$= \Delta\left[y - X\beta + X(\beta-\tilde{\beta})\right]$$
$$= \Delta u + \Delta X(\beta-\tilde{\beta})$$
$$= e + \Delta X(\beta-\tilde{\beta})$$

Then

$$n(\tilde{\alpha}-1) = \frac{n^{-1}\underset{\sim}{y}_{-1}'e + n^{-1}\underset{\sim}{y}_{-1}'\Delta X(\beta-\tilde{\beta})}{n^{-2}\underset{\sim}{y}_{-1}'\underset{\sim}{y}_{-1}}$$

Since

$$\frac{1}{\sqrt{n}}\underset{\sim}{y}_{[nr]} \Rightarrow \underset{\sim}{B}_e(r)$$

we have

$$\frac{1}{n^2}\underset{\sim}{y}_{-1}'\underset{\sim}{y}_{-1} \Rightarrow \int_0^1 \underset{\sim}{B}_e^2$$

$$n^{-1}\underset{\sim}{y}_{-1}'e \Rightarrow \int_0^1 \underset{\sim}{B}_e dB_e + \lambda$$

<u>Note.</u>

$$n^{-1}\underset{\sim}{y}_{-1}'\Delta X(\beta-\tilde{\beta}) = \frac{1}{n^{3/2}}\underset{\sim}{y}_{-1}'\Delta X F_n^{-1}\left[\sqrt{n}F_n(\beta-\tilde{\beta})\right]$$
$$= \left[\frac{1}{n}\sum_1^n\left(\underset{\sim}{y}_{t-1}\Delta x_t' F_n^{-1}\right)\right]\left[\sqrt{n}F_n(\beta-\tilde{\beta})\right]$$

Now $t = [nr]$

$$F_n^{-1}\Delta x_t = \left[\left(\frac{1}{n^{i-1}}\Delta t^i\right)_i\right] \longrightarrow \left[\left(i r^{i-1}\right)\right]$$

$$= X_0(r)$$

i.e. $c = 0$ case

So we have

$$\frac{1}{n}\sum_1^n \frac{\underset{\sim}{y}_{t-1}}{\sqrt{n}}\Delta x_t' F_n^{-1} \Longrightarrow \int_0^1 B_e X_0'$$

$$\sqrt{n}\,F_n(\beta - \tilde{\beta}) \Longrightarrow -\left(\int_0^1 X_{\bar{c}} X_{\bar{c}}'\right)^{-1}\left(\int_0^1 X_{\bar{c}}\, dB_e - \bar{c}\int_0^1 X_{\bar{c}} B_e\right)$$

i.e. $n^{-1}\underset{\sim}{y}_{-1}' X(\beta - \tilde{\beta})$

$$= \left(\frac{1}{n}\sum_1^n \frac{y_{t-1}}{\sqrt{n}}\Delta x_t' F_n^{-1}\right)\left(\sqrt{n}\,F_n(\beta - \tilde{\beta})\right)$$

$$\Longrightarrow -\left(\int_0^1 B_e X_0'\right)\left(\int_0^1 X_{\bar{c}} X_{\bar{c}}'\right)^{-1}\left(\int_0^1 X_{\bar{c}}\, dB_e - \bar{c}\int_0^1 X_{\bar{c}} B_e\right)$$

when $\bar{c} = 0$ this becomes

$$-\left(\int_0^1 B_e X_0'\right)\left(\int_0^1 X_0 X_0'\right)^{-1}\int_0^1 X_0\, dB_e$$

Hence

$$n(\hat{\alpha}-1) \Rightarrow \frac{\int_0^1 B_e dB_e + \lambda - \left(\int_0^1 B_e X_0'\right)\left(\int_0^1 X_{\bar{c}} X_{\bar{c}}'\right)^{-1}\left(\int_0^1 X_{\bar{c}} dB_e - \bar{c}\int_0^1 X_{\bar{c}} B_e\right)}{\int_0^1 \underset{\sim}{B_e}^2}$$

$$= \frac{\int_0^1 \underset{\sim}{B_e} dB_e + \lambda}{\int_0^1 \underset{\sim}{B_e}^2}$$

when $\bar{c} = 0$ this is

$$= \frac{\int_0^1 \underline{B_e}\, dB_e + \lambda}{\int_0^1 \underline{B_e}^2}$$

with $\underline{B_e} = B_e - \left(\int_0^1 B_e X_0'\right)\left(\int_0^1 X_0 X_0'\right)^{-1} X_0$

= detrended (OLS) BM B_e using trends $X_0 = \begin{pmatrix} 1 \\ r \\ r^{p-1} \end{pmatrix}$ i.e. only p'th-1 order

Conclusion

New GLS detrended unit root test has DIF limit distribution only when $\bar{c} = 0$ & then it involves p-1'th order detrending.

<u>Consider the linear-trend case</u>

$$X(r) = r \quad , X_{\bar c}(r) = 1-\bar c r$$

$$\underset{\sim}{B_e}(r) = B_e(r) - r\left(\int_0^1 (1-\bar c r)^2 dr\right)^{-1}\left\{\int_0^1 (1-\bar c r)\, dB_e\right.$$

$$\left. - \bar c \int_0^1 (1-\bar c r) B_e\right\}$$

$$= B_e(r) - r\left[\int_0^1 \left(1 - 2\bar c r + \bar c^2 r^2\right) dr\right]^{-1}\left\{\int_0^1 (1-\bar c r)\, dB_e\right.$$

$$\left. - \bar c \int_0^1 (1-\bar c r) B_e\right\}$$

$$= B_e(r) - r\left[1 - \cancel{2}\bar c \tfrac{1}{\cancel{2}} + \bar c^2 \tfrac{1}{3}\right]^{-1}\left\{\int_0^1 (1-\bar c r)\, dB_e\right.$$

$$\left. - \bar c \int_0^1 (1-\bar c r) B_e\right\}$$

$$= B_e(r) - 3r\left[3 - 3\bar c + \bar c^2\right]^{-1}\left\{\int_0^1 (1-\bar c r)\, dB_e\right.$$

$$\left. - \bar c \int_0^1 B_e (1-\bar c r)\right\}$$

$$= B_e(r) - \frac{r}{1-\bar c+\bar c^2/3}\left\{B_e(1) - \bar c \int_0^1 r\, dB_e\right.$$

$$\left. - \bar c \int_0^1 B_e + \bar c^2 \int_0^1 r B_e\right\}$$

$$\int_0^1 r\, dB_e = r B_e\Big|_0^1 - \int_0^1 B_e = B_e(1) - \int_0^1 B_e$$

$$= B_e(r) - \frac{r}{\quad} \cdots \left\{(1-\bar c) B_e(1) + \bar c^2 \int^1 r B_e\right\}$$

Testing Stationarity (KPSS, 1992 JOE)

Kwiatkowski, Phillips, Schmidt, Shin

Idea
- split a time series into a deterministic trend, a stochastic trend & a stationary component
- test if the stochastic trend is present

Structural Components Representation

$$y_t = \sum_0^p \beta_k t^k + r_t + u_t$$

(*) $$= x_t'\beta + r_t + u_t \text{, say}$$

$$r_t = r_{t-1} + v_t, \quad v_t \equiv iid(0, \sigma_v^2)$$

$$u_t = iid(0, f_{uu}(\lambda))$$

u_t, v_t independent

Likelihood function

- Assume Gaussianity: $u_t \equiv iid(0, \sigma_u^2)$, $v_t = iid(0, \sigma_v^2)$ indep
- $w_t = r_t + u_t$, say
- model is

(*

$$y_t = x_t'\beta + w_t$$

$$y = X\beta + w$$

$E(w) = 0$, $var(w) = var(r) + var(u)$

Set up the score from (*+) as

$$\tilde{\lambda} = \frac{\partial L(\tilde{\beta}, \tilde{\sigma}_u^2, \sigma_v^2 = 0)}{\partial \sigma_v^2}$$

$$= -\frac{1}{2\tilde{\sigma}_u^2} tr(A) + \frac{1}{2\tilde{\sigma}_u^4}(y - X\tilde{\beta})' A (y - X\tilde{\beta})$$

<u>Note</u> Here we use the restricted ML estimators $\tilde{\beta}$, $\tilde{\sigma}_u^2$ obtained under the null hypothesis that there is no unit root, i.e.

$$H_0: \sigma_v^2 = 0 \qquad r_t = 0 \quad a.s.$$

These restricted ML estimators are just the OLS estimators in the restricted model

$$y = X\beta + w, \quad w = u$$

i.e.

$$\tilde{\beta} = (X'X)^{-1}X'y$$

$$\tilde{\sigma}_u^2 = \frac{1}{n}(y - X\tilde{\beta})'(y - X\tilde{\beta}) = \frac{1}{n}\tilde{u}'\tilde{u}$$

<u>LM test of $\sigma_v^2 = 0$</u>

- This is based on the score $\tilde{\lambda}$
- idea is, as in all LM tests, how does likelihood func change as we move away from the null $\sigma_v^2 = 0$

- to compute the test we need to "studentize" the score $\tilde{\lambda}$
- First calculate

$$\tilde{\lambda} = -\frac{1}{2\tilde{\sigma}_u^2} tr(A) + \frac{1}{2\tilde{\sigma}_u^4} \tilde{u}' A \tilde{u}$$

- Note that $\tilde{u}$ estimates the error u and

$$var\left(\frac{u'Au}{2\sigma_u^2}\right) = \frac{1}{4\sigma_u^8} var\left[vec(A)'(u \otimes u)\right]$$

$$= \frac{1}{4\sigma_u^8} vec(A)' \, var(u \otimes u) \, vec(A)$$

$var(u^2) = 2\sigma_u^4$
under normality as
$\frac{u^2}{\sigma_u^2} \equiv \chi_1^2$

$$= \left(\frac{1}{4\sigma_u^8}\right) 2\sigma_u^4 \, tr(A^2)$$

$$= \frac{1}{2\sigma_u^4} tr(A^2)$$

LM statistic

$$LM = \frac{\tilde{\lambda}}{s_{\tilde{\lambda}}} = \frac{-\frac{1}{2\tilde{\sigma}_u^2} tr(A) + \frac{1}{2\tilde{\sigma}_u^4} \tilde{u}' A \tilde{u}}{\left(\frac{1}{2\tilde{\sigma}_u^4} tr(A^2)\right)^{1/2}}$$

$$= \frac{\tilde{u}' A \tilde{u}}{2^{1/2} \tilde{\sigma}^2 (tr(A^2))^{1/2}} - \frac{tr(A)}{2^{1/2} (tr(A^2))^{1/2}}$$

We may as well work with

$$LM = \tilde{u}'A\tilde{u} \,/\, \tilde{\sigma}_u^2$$

removing the scale coefficient & constant (asymptotically)

Alternative representation of LM statistic

$$\tilde{u}'A\tilde{u} = \tilde{u}'LL'\tilde{u} = \Sigma_1^n \tilde{S}_{t-1}^2 \quad \text{see below}$$

where

$$\tilde{S}_t = \Sigma_1^t \tilde{u}_j$$

so we get the representation

$$LM = \Sigma_1^n \tilde{S}_{t-1}^2 \,/\, \tilde{\sigma}_u^2$$

Note

$$L'\tilde{u} = \begin{pmatrix} 1 & \cdots & & 1 \\ 0 & 1 & \cdots & 1 \\ & & & \\ 0 & & & 1 \end{pmatrix} = \begin{bmatrix} \Sigma_1^n \tilde{u}_t \\ \Sigma_2^{n-1} \tilde{u}_t \\ \\ \tilde{u}_n \end{bmatrix}$$

since $\Sigma_1^n \tilde{u}_t = 0$

$\Sigma_i^n \tilde{u}_t = -\tilde{S}_{i-1}$

$$= \begin{bmatrix} 0 \\ -\tilde{u}_1 \\ -(\tilde{u}_1 + \tilde{u}_2) \\ \vdots \\ -\Sigma_1^{n-1} \tilde{u}_t \end{bmatrix}$$

$$= -\begin{bmatrix} \tilde{S}_0 \\ \tilde{S}_1 \\ \\ \tilde{S}_{n-1} \end{bmatrix}$$

so

$$\tilde{u}'LL'\tilde{u} = \Sigma_1^n \tilde{S}_{t-1}^2$$

Limit Theory under the null

$$\frac{1}{\sqrt{n}} S_{[nr]} = \frac{1}{\sqrt{n}} \sum_1^{[nr]} u_j \Rightarrow B(r) \equiv BM(\sigma_u^2)$$

whereas

$$\frac{1}{\sqrt{n}} \tilde{S}_{[nr]} = \frac{1}{\sqrt{n}} \sum_1^{[nr]} \tilde{u}_j$$

$$= \frac{1}{\sqrt{n}} \sum_1^{[nr]} u_j - \frac{1}{\sqrt{n}} \left(\sum_1^{[nr]} x_t' \right) (X'X)^{-1} X'u$$

$$= \underbrace{\left(\frac{1}{n} \sum_1^{[nr]} x_t' D_n^{-1}\right)}_{\Rightarrow \int X(s)} \underbrace{\left(\frac{1}{n} D_n^{-1} X'X D_n^{-1}\right)^{-1}}_{\Rightarrow \int X X'} \underbrace{\left(D_n^{-1} \frac{X'u}{\sqrt{n}}\right)}_{\int_0^1 X dB}$$

detrended errors

$$\tilde{u} = Q_X u = Q_X y$$

$$\Rightarrow B(r) - \left(\int_0^1 dB\, \underline{X}'\right) \left(\int_0^1 \underline{X}\, \underline{X}'\right)^{-1} \int_0^r \underline{X}(s) ds$$

$$= \int_0^r d\underset{\sim}{B}_X = \underset{\sim}{B}_X(r)$$

$$\underset{\sim}{B}_X(r) = B(r) - \left(\int_0^1 dB X'\right) \left(\int_0^1 \underline{X}\, \underline{X}'\right)^{-1} \int_0^r \underline{X}(s)$$

Remark

(i) When $X(r) = 1$ we get

$$\underset{\sim}{B}_0(r) = B(r) - r\int_0^1 dB = B(r) - rB(1)$$

$\equiv$ Brownian Bridge process

($\equiv$ tied down Brownian motion

= tied down to zero at $r=1$)

(ii) Note that in general

$$\underset{\sim}{B}_X(1) = B(1) - \left(\int_0^1 dB\, \underline{X}'\right)\left(\int_0^1 \underline{X}\,\underline{X}'\right)^{-1} \int_0^1 \underline{X}(r)$$

$$= B(1) - \left(\int_0^1 dB\, \underline{X}'\right)\left(\int_0^1 \underline{X}\,\underline{X}'\right)^{-1} \begin{pmatrix} 1 \\ 1/2 \\ \vdots \\ 1/(p+1) \end{pmatrix}$$

$$= B(1) - \int_0^1 dB\, \underline{X}' \begin{pmatrix} 1 \\ 0 \\ \vdots \end{pmatrix}$$

$B(1) - B(1)$

(iii) We call $\underset{\sim}{B}_X(r)$ a p'th-level Brownian Bridge (see Schmidt & Phillips 1992, Oxford Bulletin

Limit Theory for LM statistic

$$LM = \sum_1^n \tilde{S}_{t-1}^2 / \tilde{\sigma}_u^2$$

$$n^{-2} LM = \frac{1}{n^2} \sum_1^n \tilde{S}_{t-1}^2 / \tilde{\sigma}_u^2$$

$$= \frac{1}{n} \sum_1^n \left(\frac{1}{\sqrt{n}} \tilde{S}_{t-1} \right)^2 / \tilde{\sigma}_u^2$$

$$\Rightarrow \int_0^1 \underset{\sim}{B}_X^2 / \sigma_u^2$$

$$\equiv \int_0^1 \underset{\sim}{W}_X^2$$

where $\underset{\sim}{W}_X(r) = W(r) - \left(\int_0^1 dW \underline{X}'\right)\left(\int_0^1 \underline{X}\underline{X}'\right)^{-1} \underline{X}(r)$

= p-level standard Brownian Bridge

General Version of Test

- There is no reason why u_t has to be iid for the above limit theory to go through
- We therefore consider the case when

$$u_t = \text{wd}(0, f_{uu}(0))$$

$$\omega^2 = 2\pi f_{uu}(0) = \text{lrvar}(u_t)$$

LM test of $H: \sigma_v^2 = 0$

(same null hypothesis)

if $var(v_t) = 0$ $r_t = 0$ a.s.

<u>LM statistic</u>

$$LM = \Sigma_1^n \tilde{S}_{t-1}^2 / \tilde{\omega}_u^2$$

$\tilde{\omega}_u^2 =$ consistent estimator of $lrvar(u_t)$

$$n^{-2} LM = n^{-2} \Sigma_1^n \tilde{S}_{t-1}^2 / \tilde{\omega}_u^2$$

$$\Rightarrow \int_0^1 \underset{\sim}{W}_X^2 \quad \text{, as before}$$

<u>Form of the test</u>

- find CV (5%, 1% etc) of $\int_0^1 \underset{\sim}{W}_X^2$
- $H_0: \sigma_v^2 = 0$, $H_1: \sigma_v^2 > 0$
- Accept H_0 if

  $n^{-2} LM < CV$

  (i.e. LM not too big)
- reject H_0 if $n^{-2} LM > CV$

  (if $\sigma_v^2 > 0$ we expect $\tilde{S}_{t-1} = \Sigma_1^{t-1} \tilde{u}_s = O_p(n^{3/2})$

  and $\Sigma_1^n \tilde{S}_{t-1}^2 = O_p(n^4)$)

- The KPSS test is consistent

- Its power depends on $\tilde{\omega}_u^2$ and its bandwidth expansion rate (see KPSS paper for details)

$$n^{-2} LM = O_p(n/\ell) \quad \text{as } n \to \infty \text{ under } H_1$$

where ℓ is the lag truncation parameter in estimation of ω^2

Remark

- Using the general theory above we have a test for breaking trend stationarity.