Linear Time Series

Christian Francq

http://christian.francq140.free.fr/

CREST-ENSAE

Chapter 6: Asymptotic properties of the OLS estimator and of the unit root tests

Outline

1. Asymptotic properties of the OLS estimator
   - Martingale difference
   - Estimating the stationary AR(1)
   - ARMA Estimation
2. Unit root tests
3. Other asymptotic results

Martingale

In a fair game of pure chance between 2 players, the fortune of each player is a martingale.

Definition

Let $(Y_t)_{t\in\mathbb{N}}$ be a sequence of real random variables and $(\mathscr{F}_t)_{t\in\mathbb{N}}$ a sequence of sigma-fields. The sequence $(Y_t, \mathscr{F}_t)_{t\in\mathbb{N}}$ is a martingale if and only if

1. $\mathscr{F}_t \subset \mathscr{F}_{t+1}$;
2. Y_t is $\mathscr{F}_t$-measurable;
3. $E|Y_t| < \infty$;
4. $E(Y_{t+1}|\mathscr{F}_t) = Y_t$.

When $(Y_t)_{t\in\mathbb{N}}$ is said to be a martingale, we implicitly take $\mathscr{F}_t = \sigma(Y_u, u \leq t)$.

Martingale increments

In a fair game, the variation in a given player's fortune is a martingale difference.

Definition

Let $(\epsilon_t)_{t\in\mathbb{N}}$ be a sequence of real random variables and $(\mathscr{F}_t)_{t\in\mathbb{N}}$ a sequence of sigma-fields. The sequence $(\epsilon_t, \mathscr{F}_t)_{t\in\mathbb{N}}$ is a martingale difference iff

1. $\mathscr{F}_t \subset \mathscr{F}_{t+1}$;
2. ϵ_t est $\mathscr{F}_t$-measurable;
3. $E|\epsilon_t| < \infty$;
4. $E(\epsilon_{t+1}|\mathscr{F}_t) = 0.$

Examples of martingale differences

- If $(Y_t)_{t\in\mathbb{N}}$ is a martingale, then $\epsilon_t = Y_t - Y_{t-1}$ is a martingale difference.
- If $(\epsilon_t)_{t\in\mathbb{N}}$ is a martingale difference then $X_t = \sum_{i=0}^{t} \epsilon_i$ is a martingale.
- A semi-strong white noise is a martingale difference.
- If (X_t) is a causal AR(1) with semi-strong white noise

$$X_t = aX_{t-1} + \epsilon_t, \qquad |a| < 1,$$

then $\{\epsilon_t X_{t-1}, \sigma(\epsilon_u, u \leq t)\}$ is a martingale difference.

CLT for triangular martingale difference

Illustration

Lindeberg's CLT

Assume that, $\forall n > 0$, $(\epsilon_{nk}, \mathscr{F}_{nk})_{k\in\mathbb{N}}$ is a square integrable martingale difference. Let $\sigma^2_{nk} = E(\epsilon^2_{nk}|\mathscr{F}_{n(k-1)})$. If

$$\sum_{k=1}^{n} \sigma^2_{nk} \to \sigma^2_0 \text{ in probability as } n \to \infty,$$

where $\sigma_0 > 0$, and

$$\forall \varepsilon > 0, \quad \sum_{k=1}^{n} E\epsilon^2_{nk} 1_{\{|\epsilon_{nk}| \geq \varepsilon\}} \to 0 \text{ as } n \to \infty,$$

then $\sum_{k=1}^{n} \epsilon_{nk} \xrightarrow{\mathscr{L}} \mathcal{N}(0, \sigma^2_0)$

Encompasses the usual CLT

Application of the ergodic theorem and Lindeberg's CLT

If (X_t) is a causal AR(1)

$$X_t = aX_{t-1} + \epsilon_t, \qquad |a| < 1,$$

then the OLS of a

$$\hat{a} = \frac{\frac{1}{n}\sum_{t=2}^{n} X_t X_{t-1}}{\frac{1}{n}\sum_{t=2}^{n} X_{t-1}^2} \to \frac{\gamma(1)}{\gamma(0)} = a \quad \text{a.s.}$$

by the ergodic theorem, and

$$\sqrt{n}\{\hat{a} - a\} = \frac{n^{-1/2}\sum_{t=2}^{n} \epsilon_t X_{t-1}}{n^{-1}\sum_{t=2}^{n} X_{t-1}^2} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1 - a^2)$$

by Lindeberg's CLT.

1. Asymptotic properties of the OLS estimator
   - Martingale difference
   - Estimating the stationary AR(1)
   - ARMA Estimation
2. Unit root tests
3. Other asymptotic results

Assumptions on $\phi(B)X_t = \psi(B)\epsilon_t,\ (\epsilon_t) \sim BB(0,\sigma^2)$.

Let $\theta_0 = (\phi_1,\ldots,\phi_p,\psi_1,\ldots,\psi_q)'$ be the unknown parameter.

A1: X is strictly stationary and ergodic.*

A2: The polynomials $\phi(z) = \phi_{\theta_0}(z)$ and $\psi(z) = \psi_{\theta_0}(z)$ have their roots outside the unit disk and have no common root.

A3: $p+q>0$ and $\phi_p \neq 0$ or $\psi_q \neq 0$ (by convention $\phi_0 = \psi_0 = 1$).

A4: $\sigma^2 > 0$.

A5: $\theta_0 \in \Theta$ where Θ is a compact subset of the parameter space

$$\left\{\theta = (\theta_1,\ldots,\theta_{p+q}) \in \mathbb{R}^{p+q} :\ \phi_\theta(z)\psi_\theta(z) \neq 0 \quad \forall |z| \leq 1\right\}.$$

*The assumption is usually replaced by the stronger assumption that (ϵ_t) is a strong white noise

Least squares estimator

For all $\theta \in \Theta$, let

$$\epsilon_t(\theta) = \psi_\theta^{-1}(B)\phi_\theta(B)X_t = X_t + \sum_{i=1}^{\infty} c_i(\theta)X_{t-i}.$$

From observations $X_1, X_2, \ldots, X_n$, one can approximate $\epsilon_t(\theta)$, for $0 < t \leq n$, by $e_t(\theta)$ defined recursively by

$$e_t(\theta) = X_t - \sum_{i=1}^{p} \theta_i X_{t-i} + \sum_{i=1}^{q} \theta_{p+i} e_{t-i}(\theta)$$

where $e_0(\theta) = e_{-1}(\theta) = \ldots = e_{-q+1}(\theta) = X_0 = X_{-1} = \ldots = X_{-p+1} = 0$. It is said that $\hat{\theta}_n$ is an ordinary least squares (OLS) estimator if, almost surely,

$$Q_n(\hat{\theta}_n) = \min_{\theta \in \Theta} Q_n(\theta), \qquad Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} e_t^2(\theta).$$

Strong consistency of the OLS estimator

Link with the MLE

If (ϵ_t) were iid $\mathcal{N}(0,\sigma^2)$, the distribution of X_t given $\{X_u, u < t\}$ would be $\mathcal{N}\left\{EL(X_t \mid X_u, u < t), \sigma^2\right\}$, that is with density

$$f_\theta(X_t \mid X_u, u < t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\epsilon_t^2(\theta)}{2\sigma^2}}.$$

The OLS estimator of θ is thus equivalent to the Gaussian quasi-maximum likelihood.

Consistency

Under the assumptions **A1-A5**, $\hat{\theta}_n \longrightarrow \theta_0$ with probability 1 when $n \to \infty$.

Sketch of the proof

Asymptotic distribution of the OLS estimator

A6 : θ_0 belongs to the interior of Θ.

A7 : (ϵ_t) is a strong white noise.

A7 can be replaced by less restrictive assumptions, allowing for conditional heteroscedastic errors (see this paper).

Asymptotic normality

Under **A2-A7**, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \overset{d}{\rightsquigarrow} \mathcal{N}\left\{0, \sigma^2 J^{-1}\right\}, \quad J = E\frac{\partial \epsilon_t(\theta_0)}{\partial \theta}\frac{\partial \epsilon_t(\theta_0)}{\partial \theta'}.$$

Sketch of proof

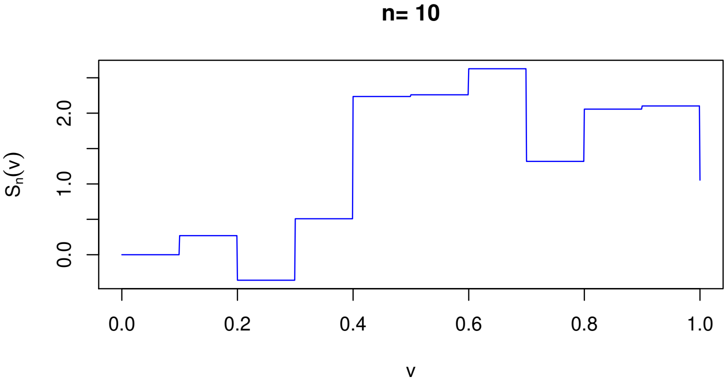
References

- Fuller (1976) and Dickey and Fuller (1979, *JASA*) are the first to obtain the (asymptotic) distribution of the OLS estimator of an AR with unit root, and provide statistical tables for unit root tests.
- Phillips (1987 *Econometrica*, 1988 *Biometrika*) expresses the limiting distribution as functionals of the Brownian motion.

Partial sums

Let (η_t) be an iid $(0,1)$ sequence. The partial sums are defined by $S_0 = 0$, $S_k = \eta_1 + \ldots + \eta_k$, and the càdlàg function

$$S_n(v) = \frac{1}{\sqrt{n}} S_{[nv]},\ v \in [0,1] \quad \left(S_n(v) = \frac{S_k}{\sqrt{n}} \text{ for } v \in \left[\frac{k}{n}, \frac{k+1}{n} \right[\right).$$



Partial sums

Let (η_t) be an iid $(0,1)$ sequence. The partial sums are defined by $S_0 = 0,\ S_k = \eta_1 + \ldots + \eta_k$. Let the càdlàg function

$$S_n(v) = \frac{1}{\sqrt{n}} S_{[nv]},\ v \in [0,1] \quad \left(S_n(v) = \frac{S_k}{\sqrt{n}} \text{ for } v \in \left[\frac{k}{n}, \frac{k+1}{n} \right) \right).$$

Note that

- $S_n(0) = 0$;
- the increments $S_n(v_1) - S_n(v_0), \ldots, S_n(v_k) - S_n(v_{k-1})$ are independent for all k and all $0 \le v_0 \le v_1 \le \cdots \le v_k \le 1$;
- $S_n(v) = \frac{\sqrt{v}}{\sqrt{nv}} S_{[nv]} \xrightarrow{\mathscr{L}} \mathcal{N}(0,v)$ as $n \to \infty$ and $v \in [0,1]$ fixed;
- $S_n(u) - S_n(v) = \frac{\sqrt{u-v}}{\sqrt{nu-nv}} \sum_{i=[nv]+1}^{[nu]} \eta_i \xrightarrow{\mathscr{L}} \mathcal{N}(0, u-v)$ for $u > v$.

Brownian motion on $[0,1]$

A standard Brownian motion (or Wiener's process) is a process $\{W(v), 0 \leq v \leq 1\}$ satisfying:

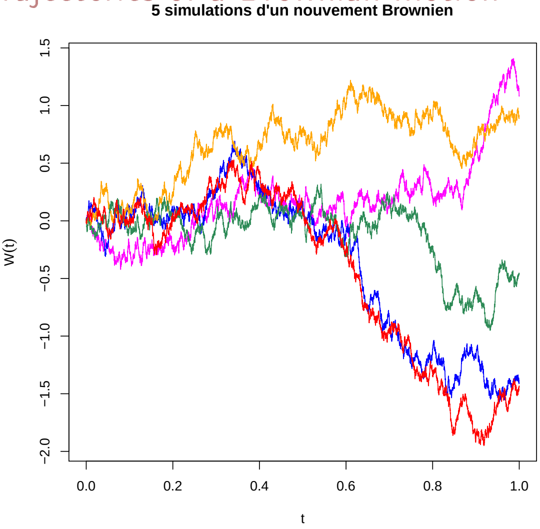
- $W(0) = 0$
- independent increments
- continuous trajectories
- $W(u) - W(v) \sim \mathcal{N}(0, |u-v|)$.

This is a Gaussian process with mean zero and covariance function

$$\text{Cov}\{W(u), W(v)\} = \text{Cov}\{W(u), W(u) + W(v) - W(u)\} = u$$

when $v \geq u$.

Simulated trajectories of a Brownian motion



Donsker's functional CLT

In the space $\mathscr{D}([0,1])$ of the càdlàg functions with the Skorokhod distance[†] we have the weak convergence

$$S_n(\cdot) \Longrightarrow W(\cdot),$$

with the previous notations.
The weak convergence means that for all functional $f : \mathscr{D}([0,1]) \to \mathbb{R}$ which is continuous in the Skorokhod distance sense, we have

$$f(S_n(\cdot)) \xrightarrow{\mathscr{L}} f(W(\cdot)).$$

[†]Intuitively, two càdlàg functions are close if, after a small deformation of the abscissa of one function, they are uniformly close on $[0,1]$ (for a precise definition, see Billingsley, P. (1999) *Convergence of Probability Measures.* Wiley)

Application of Donsker and continuous mapping theorems

Assume that $X_0 = 0$, $X_t = X_{t-1} + \epsilon_t$, $\epsilon_t = \sigma\eta_t$ where (η_t) is an iid (0,1) sequence and $\sigma > 0$. Setting $S_k = \eta_1 + \ldots + \eta_k$ and noting that $S_n(v) = S_k/\sqrt{n}$ when $v \in [\frac{k}{n}, \frac{k+1}{n}[$, we have

$$\int_0^1 S_n(v)dv = \sum_{k=0}^{n-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} S_n(v)dv = \sum_{k=0}^{n-1} n^{-3/2} S_k = \sigma^{-1} n^{-3/2} \sum_{k=1}^{n} X_{k-1},$$

which shows that

$$n^{-3/2} \sum_{t=1}^{n} X_{t-1} \xrightarrow{\mathscr{L}} \sigma \int_0^1 W(v)dv \sim \mathcal{N}(0, \sigma^2/3).$$

For the last result, we use an invariance principle: we can assume $\eta_t \sim \mathcal{N}(0, \sigma^2)$ since the limit does not depend on the distribution of η_t.

Application of Donsker (continued)

By the same arguments

$$\int_0^1 S_n^2(v)dv = \sum_{k=0}^{n-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} \frac{S_k^2}{n} dv = \sum_{k=0}^{n-1} n^{-2} S_k^2 = \sigma^{-2} n^{-2} \sum_{k=1}^{n} X_{k-1}^2,$$

which shows that

$$n^{-2} \sum_{t=1}^{n} X_{t-1}^2 \xrightarrow{\mathscr{L}} \sigma^2 \int_0^1 W^2(v)dv.$$

Application of the LLN and CTL

We have

$$\begin{aligned} S_n^2 &= 2\sum_{1\le s<t\le n}\eta_t\eta_s+\sum_{t=1}^n\eta_t^2=2\sigma^{-1}\sum_{t=1}^n\eta_t X_{t-1}+\sum_{t=1}^n\eta_t^2 \\ &= 2\sigma^{-2}\sum_{t=1}^n\epsilon_t X_{t-1}+\sum_{t=1}^n\eta_t^2. \end{aligned}$$

Thus

$$\begin{aligned} n^{-1}\sum_{t=1}^n\epsilon_t X_{t-1} &= \frac{\sigma^2}{2}\left\{S_n^2(1)-n^{-1}\sum_{t=1}^n\eta_t^2\right\} \\ &\stackrel{\mathcal{L}}{\to} \frac{\sigma^2}{2}\left\{W^2(1)-1\right\}. \end{aligned}$$

Note that $W^2(1)\sim\chi_1^2$.

Asymptotic distribution of the statistic $n\hat{\pi}$ of DF

In view of the previous computations, we have

$$\begin{aligned} n\hat{\pi} &= n\left(\frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=2}^{n} X_{t-1}^2} - 1\right) = \frac{n^{-1}\sum_{t=1}^{n} \epsilon_t X_{t-1}}{n^{-2}\sum_{t=1}^{n} X_{t-1}^2} \\ &\xrightarrow{\mathcal{L}} \frac{\frac{1}{2}\left\{W^2(1) - 1\right\}}{\int_0^1 W^2(v)\,dv}. \end{aligned}$$

Strong mixing coefficients (Rossenblatt, 1956)

The strong mixing coefficients, $\alpha_u(k)$ $k \geq 1$, of a process $u = (u_t)$ are defined by

$$\alpha_u(k) = \sup_t \sup_{A \in \sigma(u_s,\, s \leq t),\, B \in \sigma(u_s,\, s \geq t+k)} |P(A \cap B) - P(A)P(B)|.$$

- When the u_t's are independent $\alpha_u(k) = 0 \quad \forall k \geq 1$.
- If $u_t = f_t(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-m+1})$ where the ϵ_t's are independent, then $\sigma(u_s,\, s \leq t) \subset \sigma(\epsilon_t, \epsilon_{t-1}, \ldots)$ and

$$\sigma(u_s,\, s \geq t+k) \subset \sigma(\epsilon_{t+k-m+1}, \epsilon_{t+k-m+2}, \ldots).$$

Thus $\alpha_u(k) = 0 \quad \forall k \geq m$.

Functional CLT for mixing processes

Let $u = (u_t)$ be a process such that

1. $Eu_t = 0$ for all t,
2. $\sum_{k=1}^{\infty} \{\alpha_u(k)\}^{\frac{\nu}{2+\nu}} < \infty$, for some $\nu > 0$,
3. $\sup_t E|u_t|^{2+\nu} < \infty$,
4. $\lim_{n\to\infty} \mathrm{Var}\{n^{-1/2} \sum_{t=1}^{n} u_t\} = \vartheta_u^2$ exists and $\vartheta_u^2 > 0$.

We say that ϑ_u^2 is the long run variance.

▸ MA example

Let the partial sums $S_k = u_1 + \ldots + u_k$ and the càdlàg function $S_n(v) = \frac{1}{\sqrt{n}\vartheta_u} S_{[nv]}$. Then, P and PP showed that

$$n^{-1} \sum_{t=1}^{n} u_t^2 \to \sigma_u^2 := \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E\left(u_t^2\right)$$

almost surely, and

$$S_n(\cdot) \Longrightarrow W(\cdot)$$

where $\{W(v), v \in [0,1]\}$ is a standard Brownian motion.

Behavior of the OLS estimator in Case 1

Assume $X_0 = 0$, $X_t = X_{t-1} + u_t$, where (u_t) satisfies the previous assumptions. We have

$$n(\hat{\rho}_n - 1) = \frac{n^{-1}\sum_{t=1}^{n} u_t X_{t-1}}{n^{-2}\sum_{t=1}^{n} X_{t-1}^2} \Rightarrow \frac{(1/2)\left\{\vartheta_u^2 W^2(1) - \sigma_u^2\right\}}{\vartheta_u^2 \int_0^1 W^2(v)\,dv}$$

$$= \frac{(1/2)\left\{W^2(1) - 1\right\}}{\int_0^1 W^2(v)\,dv} + \frac{(1/2)\left\{\vartheta_u^2 - \sigma_u^2\right\}}{\vartheta_u^2 \int_0^1 W^2(v)\,dv}$$

▸ Proof

and, with $\hat{\sigma}_u^2 = \sum_{t=1}^{n}(X_t - \hat{\phi}_n X_{t-1})^2/(n-1)$,

$$n^2 \hat{\sigma}_{\hat{\rho}_n}^2 = \frac{\hat{\sigma}_u^2}{n^{-2}\sum_{t=1}^{n} X_{t-1}^2} \Rightarrow \frac{\sigma_u^2}{\vartheta_u^2 \int_0^1 W^2(v)\,dv}.$$

PP test in Case 1

Under the previous assumptions, if $\hat{\vartheta}_u^2$ is a weakly consistent estimator of ϑ_u^2, the PP test are based on

$$Z_\phi := n\left(\hat{\rho}_n - 1\right) - \frac{n^2 \hat{\sigma}^2_{\hat{\rho}_n}}{2\hat{\sigma}_u^2}\left(\hat{\vartheta}_u^2 - \hat{\sigma}_u^2\right) \xrightarrow{\mathcal{L}} \frac{(1/2)\left\{W^2(1) - 1\right\}}{\int_0^1 W^2(v)dv},$$

and

$$Z_t := \frac{\hat{\sigma}_u}{\hat{\vartheta}_u}\frac{\hat{\rho}_n - 1}{\hat{\sigma}_{\hat{\rho}_n}} - \frac{n\hat{\sigma}_{\hat{\rho}_n}}{2\hat{\sigma}_u\hat{\vartheta}_u}\left(\hat{\vartheta}_u^2 - \hat{\sigma}_u^2\right) \xrightarrow{\mathcal{L}} \frac{(1/2)\left\{W^2(1) - 1\right\}}{\left\{\int_0^1 W^2(v)dv\right\}^{1/2}}.$$

Similar results can be obtained in Cases 2 and 4.

Estimators of the long run variance

When (u_t) is stationary and $Eu_tu_{t+j} \to 0$ quickly enough as $j \to \infty$,

$$\vartheta_u^2 = \sigma_u^2 + 2\sum_{j=1}^{\infty} Eu_tu_{t+j}.$$

The empirical moment $n^{-1}\sum_{t=1}^{n-j}\hat{u}_t\hat{u}_{t+j}$ does not estimate Eu_tu_{t+j} well when j is large because it is based on only $n-j$ values of $\hat{u}_t$. Newey and West (1987, Econometrica) propose the so-called "heteroskedasticity and autocorrelation consistent" (HAC) estimator that weights the empirical moments:

$$\hat{\vartheta}_u^2 = \hat{\sigma}_u^2 + 2\sum_{j=1}^{\ell}\left\{1 - \frac{j}{\ell+1}\right\} n^{-1}\sum_{t=1}^{n-j}\hat{u}_t\hat{u}_{t+j}, \quad \hat{u}_t = X_t - \hat{\rho}_n X_{t-1},$$

where $\ell = \ell_n$ is a truncation parameter which tends to infinity slowly ($\ell \to \infty$ and $\ell = o(n^{1/4})$).

Estimators of the long run variance (continued)

More generally, a HAC estimator is of the form

$$\hat{\vartheta}_u^2 = \sum_{j=-(n-\ell)}^{n-\ell} \hat{\gamma}_u(j) K(j/\ell),$$

where ℓ is called the bandwidth and K is a density or a kernel (see Andrews, Econometrica, 1991).

Alternative estimator: since

$$\vartheta_u^2 = \sum_{h=-\infty}^{\infty} \gamma_u(h)$$

is 2π times the spectral density at 0 of (u_t), one can fit an AR(p_n) on $\hat{u}_t, t = 1, \ldots, n$ and estimate ϑ_u^2 by 2π the spectral density at 0 of this AR.

1. Asymptotic properties of the OLS estimator

2. Unit root tests

3. Other asymptotic results
   - KPSS for testing stationarity
   - Super-consistent OLS of a cointegration relationship
   - Spurious regression for non cointegrated I(1) variables

KPSS test (without drift)

Let the KPSS model

$$y_t = r_t + \epsilon_t, \quad r_t = r_{t-1} + u_t$$

for $t \geq 1$, r_0 constant, u_t iid $(0, \sigma_u^2)$ and ϵ_t iid $\mathcal{N}(0, \sigma_\epsilon^2)$.
The null hypothesis

$$H_0 : \sigma_u^2 = 0$$

of stationarity is rejected for large values of

$$LM = \frac{\sum_{k=1}^n S_k^2}{n^2 \hat{\sigma}_\epsilon^2}, \quad S_k = \sum_{i=1}^k e_i, \quad \hat{\sigma}_\epsilon^2 = n^{-1} \sum_{i=1}^n e_i^2,$$

where the e_i's are the residuals of the regression of y_t on 1.

KPSS test

We have $e_t = y_t - \overline{y}_n$ and $y_t = \epsilon_t + r_0$ under H_0. With the notation $S_n(v) = \sigma_\epsilon^{-1} n^{-1/2} \sum_{k=1}^{[nv]} \epsilon_k$, we have under the null

$$LM = \frac{\sum_{k=1}^n S_k^2}{n^2 \hat{\sigma}_\epsilon^2} = \int_0^1 \{S_n(v) - tS_n(1)\}^2 \, dv + o_P(1)$$
$$\xrightarrow{\mathscr{L}} \int_0^1 B^2(v) \, dv$$

where $B(v) = W(v) - tW(1)$ is called a Brownian bridge. ▸ Simulations
For the second equality, we note that

$$\frac{S_k}{\sqrt{n}\sigma_\epsilon} = \frac{\sum_{i=1}^k \epsilon_i - k\bar{\epsilon}_n}{\sqrt{n}\sigma_\epsilon} = S_n(v) - \frac{k}{n} S_n(1), \quad \frac{k}{n} \leq v < \frac{k+1}{n}.$$

Regression on cointegrated variables

Assume $Y_t = \beta X_t + u_t$ with $X_t = \sum_{i=1}^{t} \epsilon_i$ where (ϵ_t) and (u_t) are two independent strong white noises. Then $\hat{\beta} = \sum Y_t X_t / \sum X_t^2$ and for all $\delta > 0$

$$n^{1-\delta}(\hat{\beta} - \beta) = \frac{n^{-1-\delta} \sum u_t X_t}{n^{-2} \sum X_t^2} = o_P(1).$$

Indeed, with $S_n(v) = n^{-1/2} \sum_{i=1}^{[nv]} \epsilon_i$ and W a Brownian motion,

$$\frac{1}{n^2} \sum X_t^2 = \int_0^1 S_n^2(v) dvv \Rightarrow \sigma_\epsilon^2 \int_0^1 W^2(v) dv$$

and

$$\mathrm{Var}(n^{-1-\delta} \sum u_t X_t) = \sigma_u^2 \frac{1}{n^{2+\delta}} \sum_{t=1}^{n} t \sigma_\epsilon^2 = o(1).$$

The OLS estimator of the cointegration vector is thus super-consistent (*i.e.* at a rate larger than $\sqrt{n}$).

Regression on non cointegrated $I(1)$ variables

Spurious regression

Assume $X_t = \sum_{i=1}^{t} \epsilon_i$ and $Y_t = \sum_{i=1}^{t} e_i$ with (ϵ_t) and (e_t) two independent strong white noises. In the regression $Y_t = \hat{\beta} X_t + \hat{u}_t$ the vector $\hat{\beta}$ does not converge to 0 since, denoting $S_n^*(v) = n^{-1/2} \sum_{i=1}^{[nv]} e_i$ and W^* a Brownian motion independent of W,

$$n^{-2} \sum Y_t X_t = \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} S_n(v) S_n^*(v) dv \Rightarrow \sigma_u \sigma_e \int_0^1 W(v) W^*(v) dv$$

by using a multivariate extension of the Donsker CLT.

The end ☺ !

Sketch of proof for consistency ‡

It can be shown that, under the identifiability conditions $\mathbf{A2}-\mathbf{A4}$,

$$\epsilon_t(\theta) = \epsilon_t \;\; a.s. \quad \Rightarrow \quad \theta = \theta_0.$$

Under the conditions on the roots of the AR and MA polynomials, and the compactness of Θ, the initial values are asymptotically irrelevant :

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} |Q_n(\theta) - O_n(\theta)| = 0 \quad \text{where} \quad O_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \epsilon_t^2(\theta).$$

‡ For details, see this paper and the references therein

Minimization of the limit criterion

Since

$$\epsilon_t - \epsilon_t(\theta) = \sum_{i=1}^{\infty} \{c_i(\theta_0) - c_i(\theta)\} X_{t-i} \in \mathscr{H}_X(t-1) \perp \epsilon_t$$

the limit criterion $O_\infty(\theta) := E_{\theta_0}\epsilon_t^2(\theta)$ satisfies

$$\begin{aligned} O_\infty(\theta) &= E_{\theta_0}\{\epsilon_t(\theta) - \epsilon_t + \epsilon_t\}^2 \\ &= E_{\theta_0}\{\epsilon_t(\theta) - \epsilon_t\}^2 + E_{\theta_0}\epsilon_t^2 + 2\mathrm{Cov}\{\epsilon_t(\theta) - \epsilon_t, \epsilon_t\} \\ &= E_{\theta_0}\{\epsilon_t(\theta) - \epsilon_t\}^2 + \sigma^2 \geq \sigma^2, \end{aligned}$$

with equality iff $\epsilon_t(\theta) = \epsilon_t$ a.s., that is iff $\theta = \theta_0$. We thus have shown that limit criterion is minimized at θ_0:

$$\sigma^2 = O_\infty(\theta_0) < O_\infty(\theta), \quad \forall \theta \neq \theta_0.$$

This is **not suffiicent to conclude**
$(\lim_n \arg\min_\theta Q_n(\theta) = \arg\min_\theta \lim_n Q_n(\theta)?)$

Uniform consistency of the criterion

Let $V_m(\theta^*)$ be the ball of center θ^* and radius $1/m$. Set $S_m(t) = \inf_{\theta \in V_m(\theta^*) \cap \Theta} \epsilon_t^2(\theta)$. The process $\{S_m(t)\}_t$ is stationary and ergodic. The ergodic theorem shows that, a.s.,

$$\inf_{\theta \in V_m(\theta^*) \cap \Theta} O_n(\theta) = \inf_{\theta \in V_m(\theta^*) \cap \Theta} \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\theta) \geq \frac{1}{n} \sum_{t=1}^{n} S_m(t) \to E_{\theta_0} S_m(t),$$

as $n \to \infty$. Since $\epsilon_t^2(\theta)$ is continuous at θ, $S_m(t)$ increases to $\epsilon_t^2(\theta^*)$ as $m \to +\infty$, and Beppo Levi entails

$$\lim_{m \to \infty} E_{\theta_0} S_m(t) = E_{\theta_0} \epsilon_t^2(\theta^*) = O_\infty(\theta^*).$$

We thus have

$$\liminf_{m \to \infty} \liminf_{n \to \infty} \inf_{\theta \in V_m(\theta^*) \cap \Theta} O_n(\theta) \geq O_\infty(\theta^*) > \sigma^2 \quad \forall \theta^* \in \Theta,\ \theta^* \neq \theta_0.$$

Uniform minimization of the criterion

We have seen that for all $\theta^* \in \Theta$, $\theta^* \neq \theta_0$, there exists a neighbourhood $V(\theta^*)$ of θ^* such that

$$\liminf_{n\to\infty} \inf_{\theta \in V(\theta^*)\cap\Theta} O_n(\theta) > \sigma^2, \quad a.s.$$

Since

$$\inf_{\theta\in\Theta} Q_n(\theta) \geq \inf_{\theta\in\Theta} O_n(\theta) - \sup_{\theta\in\Theta} |O_n(\theta) - Q_n(\theta)|$$

for all $\theta^* \in \Theta$, $\theta^* \neq \theta_0$, there exists a neighbourhood $V(\theta^*)$ of θ^* such that

$$\liminf_{n\to\infty} \inf_{\theta \in V(\theta^*)\cap\Theta} Q_n(\theta) > \sigma^2, \quad a.s.$$

Compactness arguments

Let $V(\theta_0)$ be an arbitrary neighbourhood of θ_0. The compact set Θ is covered by $V(\theta_0)$ and the union of the open sets $V(\theta^*)$, $\theta^* \in \Theta - V(\theta_0)$, where $V(\theta^*)$ satisfies the previous equation. By compactness Θ is covered by and a finite number of these open sets: there exist $\theta_1, \ldots, \theta_k$ such that $\bigcup_{i=0}^{k} V(\theta_i) \subset \Theta$ and, a.s.,

$$\inf_{\theta \in \Theta^*} Q_n(\theta) = \min_{i=0,1,\ldots,k} \inf_{\theta \in V(\theta_i) \cap \Theta} Q_n(\theta) = \inf_{\theta \in V(\theta_0) \cap \Theta} Q_n(\theta),$$

for n large enough. Thus, almost surely, $\hat{\theta}_n$ belongs to $V(\theta_0)$ for n large enough. Since $V(\theta_0)$ can be chosen arbitrarily small, the proof is complete.

Return

Proof of the asymptotic normality

In view of the consistency and **A6**, the derivative of the criterion vanishes at $\hat{\theta}_n$, at least for n large enough. Doing a Taylor expansion around θ_0 of each row of the derivative of the criterion, we obtain

$$0 = \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) + J_n^*\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$$

where the row i of the matrix J_n^* is of the form $\frac{\partial^2}{\partial\theta'\partial\theta_i}Q_n(\theta_i^*)$ with θ_i^* between $\hat{\theta}_n$ and θ_0. Showing that the initial values are asymptotically irrelevant, doing a Taylor expansion and applying the ergodic theorem, we show that

$$J_n^* = \frac{\partial^2}{\partial\theta\partial\theta'}O_n(\theta_0) + o_{p.s.}(1) \to J.$$

Proof of the asymptotic normality (continuation and end)

Setting $\boldsymbol{Z}_t = (-X_{t-1}, \cdots, -X_{t-p}, \epsilon_{t-1}, \cdots, \epsilon_{t-q})'$, we have

$$\frac{\partial \epsilon_t(\theta_0)}{\partial \theta} = \boldsymbol{Z}_t + \sum_{i=1}^{q} \psi_i \frac{\partial \epsilon_{t-i}(\theta_0)}{\partial \theta}.$$

If there exists $\boldsymbol{\lambda} \neq 0$ such that $\boldsymbol{\lambda}' \frac{\partial \epsilon_t(\theta_0)}{\partial \theta} = 0$ (a.s. and for all t), then $\boldsymbol{\lambda}' \boldsymbol{Z}_t = 0$ a.s. This entails that X_t follows an ARMA with smaller orders, which is impossible under the identifiability conditions. This shows that J is invertible, and we obtain

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -J^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t \frac{\partial}{\partial \theta} \epsilon_t(\theta_0) + o_P(1).$$

The Lindeberg CLT concludes.

Return

Triangular sequence

Assume that, for all $n \geq 1$, $(\epsilon_{nk}, \mathscr{F}_{nk})_k$ satisfies the 4 conditions of the definition of a martingale difference, for $1 \leq k \leq n$

$$\begin{array}{llll}
\epsilon_{11} & & & \\
\epsilon_{21} & \epsilon_{22} & & \\
\epsilon_{31} & \epsilon_{32} & \epsilon_{33} & \\
\vdots & & & \\
\epsilon_{n1} & \epsilon_{n2} & \cdots & \epsilon_{nn} \\
\vdots & & &
\end{array}$$

We can extend the definition of ϵ_{nk} and $\mathscr{F}_{nk}$ for all $k \geq 0$ by setting $\epsilon_{n0} = 0$, $\mathscr{F}_{n0} = \{\emptyset, \Omega\}$ and $\epsilon_{nk} = 0$, $\mathscr{F}_{nk} = \mathscr{F}_{nn}$ for all $k > n$.

Return

The Lindeberg CLT encompasses the usual CLT

Let $Z_1, \cdots, Z_n$ be an iid sequence with finite variance,

$$\epsilon_{nk} = \frac{Z_k - EZ_k}{\sqrt{n}} \text{ and } \mathscr{F}_{nk} = \sigma(Z_1, \ldots, Z_k).$$

We have $\sigma_{nk}^2 = E\epsilon_{nk}^2 = n^{-1}\text{Var}(Z_0)$ and $\sigma_0^2 = \text{Var}(Z_0)$. Moreover

$$\begin{aligned} \sum_{k=1}^{n} E\epsilon_{nk}^2 1_{\{|\epsilon_{nk}| \geq \varepsilon\}} &= \sum_{k=1}^{n} n^{-1} \int_{\{|Z_k - EZ_k| \geq \sqrt{n}\varepsilon\}} |Z_k - EZ_k|^2 dP \\ &= \int_{\{|Z_1 - EZ_1| \geq \sqrt{n}\varepsilon\}} |Z_1 - EZ_1|^2 dP \to 0 \end{aligned}$$

because $\{|Z_1 - EZ_1| \geq \sqrt{n}\varepsilon\} \downarrow \emptyset$ et $\int_\Omega |Z_1 - EZ_1|^2 dP < \infty$. The Lindeberg condition is thus satisfied.

Return

MA example: $u_t = \sum_{i=0}^{q} \psi_i \epsilon_{t-i}$

where ϵ_t iid $(0, \sigma^2)$, $\sigma^2 > 0$, and $E|\epsilon_t|^{2+\nu} < \infty$ for $\nu > 0$.
Clearly, 1-3 hold true ($\|u_t\|_{2+\nu} \leq \|\epsilon_t\|_{2+\nu} \sum_{i=0}^{q} |\psi_i|$).
By stationarity of (u_t), we obtain

$$\begin{aligned}
\mathrm{Var}\left\{n^{-1/2}\sum_{t=1}^{n} u_t\right\} &= n^{-1}\sum_{h=-q}^{q}(n-|h|)\,\mathrm{Cov}\,(u_t, u_{t+h}) \\
&= \sum_{h=-q}^{q}(1-|h|/n)\sum_{i=\max\{0,-h\}}^{\min\{q,q-h\}} \psi_i \psi_{i+h}\sigma^2 \\
&\to \vartheta_u^2 = \sigma^2\left(\sum_{i=0}^{q}\psi_i\right)^2.
\end{aligned}$$

Condition 4 is satisfied if $\sum_{i=0}^{q} \psi_i \neq 0$ (i.e. (u_t) is not an MA(q) with unit root). Note that $\sigma_u^2 = E\left(u_t^2\right) = \sigma^2 \sum_{i=0}^{q} \psi_i^2$.

Return

Behaviour of (X_t) in Case 1

We have

$$\sum_{t=1}^{n} X_t^2 = \sum_{t=1}^{n} (X_{t-1} + u_t)^2 = \sum_{t=1}^{n} X_{t-1}^2 + 2\sum_{t=1}^{n} u_t X_{t-1} + \sum_{t=1}^{n} u_t^2.$$

Thus

$$\begin{aligned} n^{-1}\sum_{t=1}^{n} u_t X_{t-1} &= (1/2)\left\{n^{-1}X_n^2 - n^{-1}\sum_{t=1}^{n} u_t^2\right\} \\ &\Rightarrow (1/2)\left\{\vartheta_u^2 W^2(1) - \sigma_u^2\right\}. \end{aligned}$$

Behaviour of (X_t) (continued)

$$\int_0^1 S_n^2(v)dv = \sum_{k=0}^{n-1} \frac{1}{n^2\vartheta_u^2} S_k^2 = \frac{1}{n^2\vartheta_u^2}\sum_{k=1}^{n} X_{k-1}^2,$$

which shows that

$$\frac{1}{n^2}\sum_{t=1}^{n} X_{t-1}^2 \xrightarrow{\mathscr{L}} \vartheta_u^2 \int_0^1 W^2(v)dv.$$

Return

Simulations of a Brownian bridge

Return KPSS