Asymptotic Least Squares Theory

CHUNG-MING KUAN

Department of Finance & CRETA

October 19, 2010

Lecture Outline

Lecture Outline (cont'd)

6. $I(1)$ Variables

7. Autoregression of an $I(1)$ Variable
   - Asymptotic Properties of the OLS Estimator
   - Tests of Unit Root

8. Tests of Stationarity against $I(1)$

9. Regressions of $I(1)$ Variables
   - Spurious Regressions
   - Cointegration
   - Error Correction Model

When Regressors are Stochasitc

Given $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$, suppose that $\mathbf{X}$ is stochastic. Then, [A2](i) does not hold because $\mathbb{E}(\mathbf{y})$ can not be $\mathbf{X}\boldsymbol{\beta}_o$.

- It would be difficult to evaluate $\mathbb{E}(\hat{\boldsymbol{\beta}}_T)$ and $\mathrm{var}(\hat{\boldsymbol{\beta}}_T)$ because $\hat{\boldsymbol{\beta}}_T$ is a complex function of the elements of $\mathbf{y}$ and $\mathbf{X}$.
- Assume $\mathbb{E}(\mathbf{y} \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}_o$.
  - $\mathbb{E}(\hat{\boldsymbol{\beta}}_T) = \mathbb{E}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbb{E}(\mathbf{y} \mid \mathbf{X})\right] = \boldsymbol{\beta}_o$.
  - If $\mathrm{var}(\mathbf{y} \mid \mathbf{X}) = \sigma_o^2 \mathbf{I}_T$,

$$\mathrm{var}(\hat{\boldsymbol{\beta}}_T) = \mathbb{E}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathrm{var}(\mathbf{y} \mid \mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] = \sigma_o^2\,\mathbb{E}(\mathbf{X}'\mathbf{X})^{-1},$$

  which is not the same as $\sigma_o^2(\mathbf{X}'\mathbf{X})^{-1}$.
- $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is not normally distributed even when $\mathbf{y}$ is.

Q: Is the condition $\mathbb{E}(\mathbf{y} \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}_o$ realistic?

Suppose that $\mathbf{x}_t$ contains only one regressor y_{t-1}. Then, $\mathbb{E}(y_t \mid \mathbf{x}_1, \ldots, \mathbf{x}_T) = \mathbf{x}_t'\boldsymbol{\beta}_o$ implies

$$\mathbb{E}(y_t \mid y_1, \ldots, y_{T-1}) = \beta_o y_{t-1},$$

which is y_t with probability one. As such, the conditional variance of y_t,

$$\mathrm{var}(y_t \mid y_1, \ldots, y_{T-1}) = \mathbb{E}\{[y_t - \mathbb{E}(y_t \mid y_1, \ldots, y_{T-1})]^2 \mid y_1, \ldots, y_{T-1}\},$$

must be zero, rather than a positive constant σ_o^2.

Note: When $\mathbf{X}$ is stochastic, a different framework is needed to evaluate the properties of the OLs estimator.

Notations

- We observe $(y_t \ \mathbf{w}_t')'$, where $\mathbf{w}_t$ $(m \times 1)$ is the vector of all "exogenous" variables.
- $\mathcal{W}^t = \{\mathbf{w}_1, \ldots, \mathbf{w}_t\}$ and $\mathcal{Y}^t = \{y_1, \ldots, y_t\}$. Then, $\{\mathcal{Y}^{t-1}, \mathcal{W}^t\}$ generates a σ-algebra that is the information set up to time t.
- Regressors $\mathbf{x}_t$ $(k \times 1)$ are taken from the information set $\{\mathcal{Y}^{t-1}, \mathcal{W}^t\}$, and the resulting linear specification is

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t, \quad t = 1, 2, \ldots, T.$$

- The OLS estimator of this specification is

$$\hat{\boldsymbol{\beta}}_T = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{x}_t y_t\right).$$

Consistency

The OLS estimator $\hat{\boldsymbol{\beta}}_T$ is strongly (weakly) consistent for $\boldsymbol{\beta}^*$ if $\hat{\boldsymbol{\beta}}_T \xrightarrow{\text{a.s.}} \boldsymbol{\beta}^*$ ($\hat{\boldsymbol{\beta}}_T \xrightarrow{\mathbb{P}} \boldsymbol{\beta}^*$) as $T \to \infty$. That is, $\hat{\boldsymbol{\beta}}_T$ will be eventually close to $\boldsymbol{\beta}^*$ in a proper probabilistic sense when "enough" information becomes available.

[B1] (i) $\{\mathbf{x}_t \mathbf{x}_t'\}$ obeys a SLLN (WLLN) with the almost sure (prob.) limit

$$\mathbf{M}_{xx} := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\mathbf{x}_t \mathbf{x}_t'),$$

which is nonsingular.

[B1] (ii) $\{\mathbf{x}_t y_t\}$ obeys a SLLN (WLLN) with the almost sure (prob.) limit

$$\mathbf{m}_{xy} := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\mathbf{x}_t y_t).$$

[B2] There exists a $\boldsymbol{\beta}_o$ such that $y_t = \mathbf{x}_t' \boldsymbol{\beta}_o + \epsilon_t$ with $\mathbb{E}(\mathbf{x}_t \epsilon_t) = \mathbf{0}$ for all t.

By [B1] and Lemma 5.13, the OLS estimator of $\hat{\boldsymbol{\beta}}_T$ is

$$\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t y_t\right) \to \mathbf{M}_{xx}^{-1}\mathbf{m}_{xy} \quad \text{a.s. (in probability).}$$

When [B2] holds, $\mathbb{E}(\mathbf{x}_t\mathbf{y}_t) = \mathbb{E}(\mathbf{x}_t\mathbf{x}_t')\boldsymbol{\beta}_o$, so that $\mathbf{m}_{xy} = \mathbf{M}_{xx}\boldsymbol{\beta}_o$, and $\boldsymbol{\beta}^* = \boldsymbol{\beta}_o$.

Theorem 6.1

Consider the linear specification $y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t$.

(i) When [B1] holds, $\hat{\boldsymbol{\beta}}_T$ is strongly (weakly) consistent for $\boldsymbol{\beta}^* = \mathbf{M}_{xx}^{-1}\mathbf{m}_{xy}$.

(ii) When [B1] and [B2] hold, $\boldsymbol{\beta}_o = \mathbf{M}_{xx}^{-1}\mathbf{m}_{xy}$ so that $\hat{\boldsymbol{\beta}}_T$ is strongly (weakly) consistent for $\boldsymbol{\beta}_o$.

Remarks:

- Theorem 6.1 is about consistency (not unbiasedness), and what really matters is whether the data are governed by some SLLN (WLLN).
- Note that [B1] explicitly allows $\mathbf{x}_t$ to be a random vector which may contain some lagged dependent variables ($y_{t-j}, j \geq 1$) and other random variables in the information set. Also, the random data may exhibit dependence and heterogeneity, as long as such dependence and heterogeneity do not affect the LLN in [B1].
- Given [B2], $\mathbf{x}_t'\beta$ is the correct specification for the linear projection of y_t, and the OLS estimator converges to the parameter of interest $\boldsymbol{\beta}_o$.
- A sufficient condition for [B2] is that there exists $\boldsymbol{\beta}_o$ such that $\mathbb{E}(y_t \mid \mathcal{Y}^{t-1}, \mathcal{W}^t) = \mathbf{x}_t'\boldsymbol{\beta}_o$. (Why?)

Corollary 6.2

Suppose that $(y_t \ \mathbf{x}_t')'$ are independent random vectors with bounded $(2+\delta)$ th moment for any $\delta > 0$, such that $\mathbf{M}_{xx}$ and $\mathbf{m}_{xy}$ defined in $[\mathrm{B1}]$ exist. Then, the OLS estimator $\hat{\boldsymbol{\beta}}_T$ is strongly consistent for $\boldsymbol{\beta}^* = \mathbf{M}_{xx}^{-1}\mathbf{m}_{xy}$. If $[\mathrm{B2}]$ also holds, $\hat{\boldsymbol{\beta}}_T$ is strongly consistent for $\boldsymbol{\beta}_o$.

Proof: By the Cauchy-Schwartz inequality (Lemma 5.5), the i th element of $\mathbf{x}_t y_t$ is such that

$$\mathbb{E}\,|x_{ti} y_t|^{1+\delta} \leq \left[\mathbb{E}\,|x_{ti}|^{2(1+\delta)}\right]^{1/2} \left[\mathbb{E}\,|y_t|^{2(1+\delta)}\right]^{1/2} \leq \Delta,$$

for some $\Delta > 0$. Similarly, each element of $\mathbf{x}_t\mathbf{x}_t'$ also has bounded $(1+\delta)$ th moment. Then, $\{\mathbf{x}_t\mathbf{x}_t'\}$ and $\{\mathbf{x}_t\mathbf{y}_t\}$ obey Markov's SLLN by Lemma 5.26 with the respective almost sure limits $\mathbf{M}_{xx}$ and $\mathbf{m}_{xy}$.

Example: Given the specification: $y_t = \alpha y_{t-1} + e_t$, suppose that $\{y_t^2\}$ and $\{y_t y_{t-1}\}$ obey a SLLN (WLLN). Then, the OLS estimator of α is such that

$$\hat{\alpha}_T \to \frac{\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(y_t y_{t-1})}{\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(y_{t-1}^2)} \quad \text{a.s. (in probability).}$$

When $\{y_t\}$ indeed follows a stationary AR(1) process:

$$y_t = \alpha_o y_{t-1} + u_t, \quad |\alpha_o| < 1,$$

where u_t are i.i.d. with mean zero and variance σ_u^2, we have $\mathbb{E}(y_t) = 0$, $\text{var}(y_t) = \sigma_u^2/(1-\alpha_o^2)$ and $\text{cov}(y_t, y_{t-1}) = \alpha_o \,\text{var}(y_t)$. We have

$$\hat{\alpha}_T \to \frac{\text{cov}(y_t, y_{t-1})}{\text{var}(y_t)} = \alpha_o, \quad \text{a.s. (in probability).}$$

When $\mathbf{x}_t'\boldsymbol{\beta}_o$ is not the linear projection, i.e., $\mathbb{E}(\mathbf{x}_t\epsilon_t) \neq \mathbf{0}$,

$$\mathbb{E}(\mathbf{x}_t y_t) = \mathbb{E}(\mathbf{x}_t\mathbf{x}_t')\boldsymbol{\beta}_o + \mathbb{E}(\mathbf{x}_t\epsilon_t).$$

Then, $\mathbf{m}_{xy} = \mathbf{M}_{xx}\boldsymbol{\beta}_o + \mathbf{m}_{x\epsilon}$, where

$$\mathbf{m}_{x\epsilon} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\mathbf{x}_t\epsilon_t).$$

The limit of the OLS estimator now reads

$$\boldsymbol{\beta}^* = \mathbf{M}_{xx}^{-1}\mathbf{m}_{xy} = \boldsymbol{\beta}_o + \mathbf{M}_{xx}^{-1}\mathbf{m}_{x\epsilon}.$$

Example: Given the specification: $y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t$, suppose

$$\mathbb{E}(y_t \mid \mathcal{Y}^{t-1}, \mathcal{W}^t) = \mathbf{x}_t'\boldsymbol{\beta}_o + \mathbf{z}_t'\boldsymbol{\gamma}_o,$$

where $\mathbf{z}_t$ are in the information set but distinct from $\mathbf{x}_t$. Writing

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_o + \mathbf{z}_t'\boldsymbol{\gamma}_o + \epsilon_t = \mathbf{x}_t'\boldsymbol{\beta}_o + u_t,$$

we have $\mathbb{E}(\mathbf{x}_t u_t) = \mathbb{E}(\mathbf{x}_t\mathbf{z}_t')\boldsymbol{\gamma}_o \neq \mathbf{0}$. It follows that

$$\hat{\boldsymbol{\beta}}_T \to \mathbf{M}_{xx}^{-1}\mathbf{m}_{xy} = \boldsymbol{\beta}_o + \mathbf{M}_{xx}^{-1}\mathbf{M}_{xz}\boldsymbol{\gamma}_o,$$

with $\mathbf{M}_{xz} := \lim_T \sum_{t=1}^{T} \mathbb{E}(\mathbf{x}_t\mathbf{z}_t')/T$. The limit can not be $\boldsymbol{\beta}_o$ unless $\mathbf{x}_t$ is orthogonal to $\mathbf{z}_t$, i.e., $\mathbb{E}(\mathbf{x}_t\mathbf{z}_t') = \mathbf{0}$.

Example: Given $y_t = \alpha y_{t-1} + e_t$, suppose that

$$y_t = \alpha_o y_{t-1} + \epsilon_t, \quad |\alpha_o| < 1,$$

where $\epsilon_t = u_t - \pi_o u_{t-1}$ with $|\pi_o| < 1$, and $\{u_t\}$ is a white noise with mean zero and variance σ_u^2. Here, $\{y_t\}$ is a weakly stationary ARMA(1,1) process. We know $\hat{\alpha}_T$ converges to $\operatorname{cov}(y_t, y_{t-1}) / \operatorname{var}(y_{t-1})$ almost surely (in probability). Note, however, that $\epsilon_{t-1} = u_{t-1} - \pi_o u_{t-2}$ and

$$\mathbb{E}(y_{t-1}\epsilon_t) = \mathbb{E}[y_{t-1}(u_t - \pi_o u_{t-1})] = -\pi_o \sigma_u^2.$$

The limit of $\hat{\alpha}_T$ is then

$$\frac{\operatorname{cov}(y_t, y_{t-1})}{\operatorname{var}(y_{t-1})} = \frac{\alpha_o \operatorname{var}(y_{t-1}) + \operatorname{cov}(\epsilon_t, y_{t-1})}{\operatorname{var}(y_{t-1})} = \alpha_o - \frac{\pi_o \sigma_u^2}{\operatorname{var}(y_{t-1})}.$$

The OLS estimator is inconsistent for α_o unless $\pi_o = 0$.

Remark: Given the specification: $y_t = \alpha y_{t-1} + \mathbf{x}_t' \boldsymbol{\beta} + e_t$, suppose that

$$y_t = \alpha_o y_{t-1} + \mathbf{x}_t' \boldsymbol{\beta}_o + \epsilon_t,$$

such that ϵ_t are serially correlated (e.g., AR(1) or MA(1)). The OLS estimator is inconsistent because $\alpha_o y_{t-1} + \mathbf{x}_t' \boldsymbol{\beta}_o$ is not the linear projection, a consequence of the joint presence of a lagged dependent variable and serially correlated disturbances.

Asymptotic Normality

By asymptotic normality of $\hat{\boldsymbol{\beta}}_T$ we mean:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) \xrightarrow{D} \mathcal{N}(\mathbf{0},\ \mathbf{D}_o),$$

where $\mathbf{D}_o$ is a p.d. matrix. We may also write

$$\mathbf{D}_o^{-1/2}\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) \xrightarrow{D} \mathcal{N}(\mathbf{0},\ \mathbf{I}_k).$$

Given the specification $y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t$ and [B2], define

$$\mathbf{V}_T := \operatorname{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t\epsilon_t\right).$$

[B3] $\{\mathbf{V}_o^{-1/2}\mathbf{x}_t\epsilon_t\}$ obeys a CLT, where $\mathbf{V}_o = \lim_{T\to\infty}\mathbf{V}_T$ is p.d.

- The normalized OLS estimator is

$$
\begin{aligned}
\sqrt{T}(\hat{\beta}_T - \beta_o) &= \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{x}_t \epsilon_t \right) \\
&= \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \mathbf{V}_o^{1/2} \left[\mathbf{V}_o^{-1/2} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{x}_t \epsilon_t \right) \right] \\
&\xrightarrow{D} \mathbf{M}_{xx}^{-1} \mathbf{V}_o^{1/2} \mathcal{N}(\mathbf{0}, \mathbf{I}_k).
\end{aligned}
$$

Theorem 6.6

Given $y_t = \mathbf{x}_t'\beta + e_t$, suppose that [B1](i), [B2] and [B3] hold. Then,

$$
\sqrt{T}(\hat{\beta}_T - \beta_o) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{D}_o), \quad \mathbf{D}_o = \mathbf{M}_{xx}^{-1} \mathbf{V}_o \mathbf{M}_{xx}^{-1}.
$$

Corrollary 6.7

Given $y_t = \mathbf{x}_t'\beta + e_t$, suppose that $(y_t\ \mathbf{x}_t')'$ are independent random vectors with bounded $(4+\delta)$th moment for any $\delta > 0$ and that [B2] holds. If $\mathbf{M}_{xx}$ defined in [B1] and $\mathbf{V}_o$ defined in [B3] exist,

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) \xrightarrow{D} \mathcal{N}(\mathbf{0},\, \mathbf{D}_o), \quad \mathbf{D}_o = \mathbf{M}_{xx}^{-1}\mathbf{V}_o\mathbf{M}_{xx}^{-1}.$$

Proof: Let $z_t = \boldsymbol{\lambda}'\mathbf{x}_t\epsilon_t$, where $\boldsymbol{\lambda}$ is such that $\boldsymbol{\lambda}'\boldsymbol{\lambda} = 1$. If $\{z_t\}$ obeys a CLT, then $\{\mathbf{x}_t\epsilon_t\}$ obeys a multivariate CLT by the Cramér-Wold device.
Clearly, z_t are independent r.v. with mean zero and $\text{var}(z_t) = \boldsymbol{\lambda}'[\text{var}(\mathbf{x}_t\epsilon_t)]\boldsymbol{\lambda}$. By data independence,

$$\mathbf{V}_T = \text{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t\epsilon_t\right) = \frac{1}{T}\sum_{t=1}^{T}\text{var}(\mathbf{x}_t\epsilon_t).$$

Proof (Cont'd):

The average of $\text{var}(z_t)$ is then

$$\frac{1}{T}\sum_{t=1}^{T}\text{var}(z_t) = \boldsymbol{\lambda}'\mathbf{V}_T\boldsymbol{\lambda} \rightarrow \boldsymbol{\lambda}\mathbf{V}_o\boldsymbol{\lambda}.$$

By the Cauchy-Schwartz inequality,

$$\mathbb{E}\,|x_{ti}y_t|^{2+\delta} \leq \left[\mathbb{E}\,|x_{ti}|^{2(2+\delta)}\right]^{1/2}\left[\mathbb{E}\,|y_t|^{2(2+\delta)}\right]^{1/2} \leq \Delta,$$

for some $\Delta > 0$. Similarly, $x_{ti}x_{tj}$ have bounded $(2+\delta)$ th moment. It follows that $x_{ti}\epsilon_t$ and z_t also have bounded $(2+\delta)$ th moment by Minkowski's inequality. Then by Liapunov's CLT,

$$\frac{1}{\sqrt{T(\boldsymbol{\lambda}'\mathbf{V}_o\boldsymbol{\lambda})}}\sum_{t=1}^{T} z_t \xrightarrow{D} \mathcal{N}(0,\,1).$$

Example: Consider $y_t = \alpha y_{t-1} + e_t$. Case 1: $y_t = \alpha_o y_{t-1} + u_t$ with $|\alpha_o| < 1$, where u_t are i.i.d. with mean zero and variance σ_u^2. Note

$$\operatorname{var}(y_{t-1}u_t) = \mathbb{E}(y_{t-1}^2)\,\mathbb{E}(u_t^2) = \sigma_u^4/(1-\alpha_o^2),$$

and $\operatorname{cov}(y_{t-1}u_t, y_{t-1-j}u_{t-j}) = 0$ for all $j > 0$. A CLT ensures:

$$\frac{\sqrt{1-\alpha_o^2}}{\sigma_u^2\sqrt{T}} \sum_{t=1}^{T} y_{t-1}u_t \xrightarrow{D} \mathcal{N}(0,\,1).$$

As $\sum_{t=1}^{T} y_{t-1}^2/T$ converges to $\sigma_u^2/(1-\alpha_o^2)$, we have

$$\frac{\sqrt{1-\alpha_o^2}}{\sigma_u^2}\,\frac{\sigma_u^2}{1-\alpha_o^2}\,\sqrt{T}(\hat{\alpha}_T - \alpha_o) = \frac{1}{\sqrt{1-\alpha_o^2}}\sqrt{T}(\hat{\alpha}_T - \alpha_o) \xrightarrow{D} \mathcal{N}(0,\,1),$$

or equivalently, $\sqrt{T}(\hat{\alpha}_T - \alpha_o) \xrightarrow{D} \mathcal{N}(0,\,1-\alpha_o^2)$.

Example (cont'd): When $\{y_t\}$ is a random walk:

$$y_t = y_{t-1} + u_t.$$

We already know $\text{var}(T^{-1/2} \sum_{t=1}^{T} y_{t-1} u_t)$ diverges with T and hence $\{y_{t-1} u_t\}$ does not obey a CLT. Thus, there is no guarantee that normalized $\hat{\alpha}_T$ is asymptotically normally distributed.

Theorem 6.9

Given $y_t = \mathbf{x}_t'\beta + e_t$, suppose that $[\mathrm{B1}](\mathrm{i})$, $[\mathrm{B2}]$ and $[\mathrm{B3}]$ hold. Then,

$$\widehat{\mathbf{D}}_T^{-1/2}\sqrt{T}(\hat{\beta}_T - \beta_o) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{I}_k),$$

where $\widehat{\mathbf{D}}_T = (\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t'/T)^{-1}\widehat{\mathbf{V}}_T(\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t'/T)^{-1}$, with $\widehat{\mathbf{V}}_T \xrightarrow{\mathbb{P}} \mathbf{V}_o$.

Remarks:

1. Theorem 6.6 may hold for weakly dependent and heterogeneously distributed data, as long as these data obey proper LLN and CLT.
2. Normalizing the OLS estimator with an inconsistent estimator of $\mathbf{D}_o^{-1/2}$ destroys asymptotic normality.

Consistent Estimation of Covariance Matrix

- Consistent estimation of $\mathbf{D}_o$ amounts to consistent estimation of $\mathbf{V}_o$.
- Write $\mathbf{V}_o = \lim_{T\to\infty} \mathbf{V}_T = \lim_{T\to\infty} \sum_{j=-T+1}^{T-1} \mathbf{\Gamma}_T(j)$, with

$$\mathbf{\Gamma}_T(j) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} \mathbb{E}(\mathbf{x}_t\epsilon_t\epsilon_{t-j}\mathbf{x}_{t-j}'), & j = 0,1,2,\ldots, \\ \frac{1}{T}\sum_{t=-j+1}^{T} \mathbb{E}(\mathbf{x}_{t+j}\epsilon_{t+j}\epsilon_t\mathbf{x}_t'), & j = -1,-2,\ldots. \end{cases}$$

- When $\{\mathbf{x}_t\epsilon_t\}$ is weakly stationary, $\mathbb{E}(\mathbf{x}_t\epsilon_t\epsilon_{t-j}\mathbf{x}_{t-j}')$ depends only on the time difference $|j|$ but not on t. Thus,

$$\mathbf{\Gamma}_T(j) = \mathbf{\Gamma}_T(-j) = \mathbb{E}(\mathbf{x}_t\epsilon_t\epsilon_{t-j}\mathbf{x}_{t-j}'), \quad j = 0,1,2,\ldots,$$

and $\mathbf{V}_o = \mathbf{\Gamma}(0) + \lim_{T\to\infty} 2\sum_{j=1}^{T-1}\mathbf{\Gamma}(j)$.

Eicker-White Estimator

Case 1: When $\{\mathbf{x}_t \epsilon_t\}$ has no serial correlations,

$$\mathbf{V}_o = \lim_{T\to\infty} \mathbf{\Gamma}_T(0) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\epsilon_t^2 \mathbf{x}_t \mathbf{x}_t').$$

- A heteroskedasticity-consistent estimator of $\mathbf{V}_o$ is

$$\widehat{\mathbf{V}}_T = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_t^2 \mathbf{x}_t \mathbf{x}_t',$$

  which permits conditional heteroskedasticity of unknown form.

- The Eicker-White estimator of $\mathbf{D}_o$ is:

$$\widehat{\mathbf{D}}_T = \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \hat{e}_t^2 \mathbf{x}_t \mathbf{x}_t' \right) \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1}.$$

- The Eicker-White estimator is "robust" when heteroskedasticity is present and of an unknown form.
- If ϵ_t are also conditionally homoskedastic: $\mathbb{E}\big(\epsilon_t^2 \mid \mathcal{Y}^{t-1}, \mathcal{W}^t\big) = \sigma_o^2$,

$$\mathbf{V}_o = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\big[\mathbb{E}\big(\epsilon_t^2 \mid \mathcal{Y}^{t-1}, \mathcal{W}^t\big)\mathbf{x}_t\mathbf{x}_t'\big] = \sigma_o^2\,\mathbf{M}_{xx}.$$

Then, $\mathbf{D}_o$ is $\mathbf{M}_{xx}^{-1}\mathbf{V}_o\mathbf{M}_{xx}^{-1} = \sigma_o^2\,\mathbf{M}_{xx}^{-1}$, and it can be consistently estimated by

$$\widehat{\mathbf{D}}_T = \hat{\sigma}_T^2 \left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1},$$

as in the classical model.

Case 2: When $\{\mathbf{x}_t \epsilon_t\}$ exhibits serial correlations such that

$$\mathbf{V}_T^{\dagger} = \sum_{j=-\ell(T)}^{\ell(T)} \mathbf{\Gamma}_T(j) \rightarrow \mathbf{V}_o,$$

where $\ell(T)$ diverges with T, we may try to estimate $\mathbf{V}_T^{\dagger}$.

- A difficulty: The sample counterpart $\sum_{j=-\ell(T)}^{\ell(T)} \widehat{\mathbf{\Gamma}}_T(j)$, which is based on the sample counterpart of $\mathbf{\Gamma}_T(j)$, may not be p.s.d.
- A heteroskedasticity and autocorrelation-consistent (HAC) estimator that is guaranteed to be p.s.d. has the following form:

$$\widehat{\mathbf{V}}_T^{\kappa} = \sum_{j=-T+1}^{T-1} \kappa\left(\frac{j}{\ell(T)}\right) \widehat{\mathbf{\Gamma}}_T(j), \qquad (1)$$

where κ is a kernel function and $\ell(T)$ is its bandwidth.

- The estimator of $\mathbf{D}_o$ due to Newey and West (1987),

$$\widehat{\mathbf{D}}_T^{\kappa} = \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \widehat{\mathbf{V}}_T^{\kappa} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1},$$

  is robust to both conditional heteroskedasticity of ϵ_t and serial correlations of $\mathbf{x}_t \epsilon_t$.

- The Eicker-White and Newey-West estimators do not rely on any parametric model of cond. heteroskedasticity and serial correlations.

- κ satisfies: $|\kappa(x)| \leq 1$, $\kappa(0) = 1$, $\kappa(x) = \kappa(-x)$ for all $x \in \mathbb{R}$, $\int |\kappa(x)| \, \mathrm{d}x < \infty$, κ is continuous at 0 and at all but a finite number of other points in $\mathbb{R}$, and

$$\int_{-\infty}^{\infty} \kappa(x) e^{-ix\omega} \, \mathrm{d}x \geq 0, \quad \forall \omega \in \mathbb{R}.$$

Some Commonly Used Kernel Functions

1. Bartlett kernel (Newey and West, 1987): $\kappa(x) = 1 - |x|$ for $|x| \leq 1$, and $\kappa(x) = 0$ otherwise.

2. Parzen kernel (Gallant, 1987):

$$\kappa(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & |x| \leq 1/2, \\ 2(1 - |x|)^3, & 1/2 \leq |x| \leq 1, \\ 0, & \text{otherwise}; \end{cases}$$

3. Quadratic spectral kernel (Andrews, 1991):

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

4. Daniel kernel (Ng and Perron, 1996): $\kappa(x) = \dfrac{\sin(\pi x)}{\pi x}$.

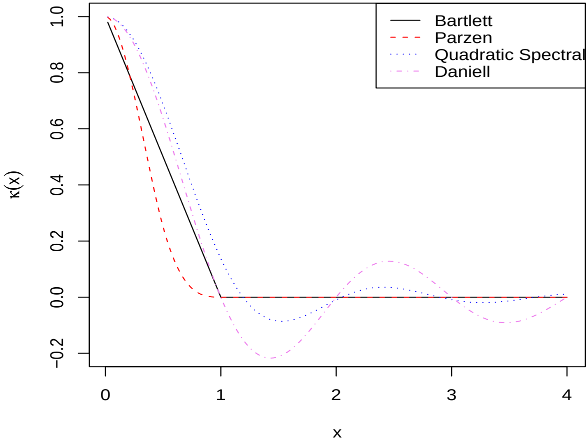

Figure: The Bartlett, Parzen, quandratic spectral and Daniel kernels.

Remarks:

- Bandwidth $\ell(T)$: It can be of order $o(T^{1/2})$, Andrews (1991). (What does this imply?)
- The Bartlett and Parzen kernels have the bounded support $[-1, 1]$, but the quadratic spectral and Daniel kernels have unbounded support.
- Andrews (1991): The quadratic spectral kernel is to be preferred in HAC estimation.
  - Rate of convergence: $O(T^{-1/3})$ for the Bartlett kernel, and $O(T^{-2/5})$ for the Parzen and quadratic spectral.
  - The quadratic spectral kernel is more efficient asymptotically than the Parzen kernel, and the Bartlett kernel is the least efficient.
- The optimal choice of $\ell(T)$ is an important issue in practice.

Wald Test

Null hypothesis: $\mathbf{R}\boldsymbol{\beta}_o = \mathbf{r}$

- Want to check if $\mathbf{R}\hat{\boldsymbol{\beta}}_T$ is sufficiently "close" to $\mathbf{r}$.
- By Theorem 6.6, $(\mathbf{R}\mathbf{D}_o\mathbf{R}')^{-1/2}\sqrt{T}\mathbf{R}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$, where $\mathbf{D}_o = \mathbf{R}\mathbf{M}_{xx}^{-1}\mathbf{V}_o\mathbf{M}_{xx}^{-1}\mathbf{R}'$.
- Given a consistent estimator for $\mathbf{D}_o$:

$$\widehat{\mathbf{D}}_T = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\widehat{\mathbf{V}}_T\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1},$$

with $\widehat{\mathbf{V}}_T$ be a consistent estimator of $\mathbf{V}_o$, we have

$$(\mathbf{R}\widehat{\mathbf{D}}_T\mathbf{R}')^{-1/2}\sqrt{T}\mathbf{R}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{I}_q).$$

The Wald test statistic is

$$\mathcal{W}_T = T(\mathbf{R}\hat{\boldsymbol{\beta}}_T - \mathbf{r})'(\mathbf{R}\widehat{\mathbf{D}}_T\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}_T - \mathbf{r}).$$

Theorem 6.10

Given $y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t$, suppose that [B1](i), [B2] and [B3] hold. Then under the null, $\mathcal{W}_T \xrightarrow{D} \chi^2(q)$, where q is the number of hypotheses.

- Data are not required to be serially uncorrelated, homoskedastic, or normally distributed.
- The limiting χ^2 distribution of the Wald test is only an approximation to the exact distribution.

Example: Given the specification $y_t = \mathbf{x}_{1,t}'\mathbf{b}_1 + \mathbf{x}_{2,t}'\mathbf{b}_2 + e_t$, where $\mathbf{x}_{1,t}$ is $(k-s)\times 1$ and $\mathbf{x}_{2,t}$ is $s \times 1$.

- Hypothesis: $\mathbf{R}\boldsymbol{\beta}_o = \mathbf{0}$, where $\mathbf{R} = [\mathbf{0}_{s\times(k-s)} \ \mathbf{I}_s]$.
- The Wald test statistic is

$$\mathcal{W}_T = T\hat{\boldsymbol{\beta}}_T'\mathbf{R}'\left(\mathbf{R}\widehat{\mathbf{D}}_T\mathbf{R}'\right)^{-1}\mathbf{R}\hat{\boldsymbol{\beta}}_T \xrightarrow{D} \chi^2(s),$$

  where $\widehat{\mathbf{D}}_T = (\mathbf{X}'\mathbf{X}/T)^{-1}\hat{\mathbf{V}}_T(\mathbf{X}'\mathbf{X}/T)^{-1}$. The exact form of $\mathcal{W}_T$ depends on $\widehat{\mathbf{D}}_T$.

- When $\widehat{\mathbf{V}}_T = \hat{\sigma}_T^2(\mathbf{X}'\mathbf{X}/T)$ is consistent for $\mathbf{V}_o$, $\widehat{\mathbf{D}}_T = \hat{\sigma}_T^2(\mathbf{X}'\mathbf{X}/T)^{-1}$ is consistent for $\mathbf{D}_o$, and the Wald statistic becomes

$$\mathcal{W}_T = T\hat{\boldsymbol{\beta}}_T'\mathbf{R}'\left[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\boldsymbol{\beta}}_T/\hat{\sigma}_T^2,$$

  which is s times the standard F statistic.

Lagrange Multiplier (LM) Test

- Given the constraint $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, the Lagrangian is

$$\frac{1}{T}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + (\mathbf{R}\boldsymbol{\beta} - \mathbf{r})'\boldsymbol{\lambda},$$

  where $\boldsymbol{\lambda}$ is the $q \times 1$ vector of Lagrange multipliers. The solutions are:

$$\ddot{\boldsymbol{\lambda}}_T = 2\big[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\big]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}_T - \mathbf{r}),$$

$$\ddot{\boldsymbol{\beta}}_T = \hat{\boldsymbol{\beta}}_T - (\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\ddot{\boldsymbol{\lambda}}_T/2.$$

- The LM test checks if $\ddot{\boldsymbol{\lambda}}_T$ (the "shadow price" of the constraint) is sufficiently "close" to zero.

By the asymptotic normality of $\sqrt{T}(\mathbf{R}\hat{\beta}_T - \mathbf{r})$,

$$\mathbf{\Lambda}_o^{-1/2}\sqrt{T}\ddot{\boldsymbol{\lambda}}_T \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{I}_q),$$

where $\mathbf{\Lambda}_o = 4(\mathbf{R}\mathbf{M}_{xx}^{-1}\mathbf{R}')^{-1}(\mathbf{R}\mathbf{D}_o\mathbf{R}')(\mathbf{R}\mathbf{M}_{xx}^{-1}\mathbf{R}')^{-1}$. Let $\ddot{\mathbf{V}}_T$ be a consistent estimator of $\mathbf{V}_o$ based on the constrained estimation result. Then,

$$\ddot{\mathbf{\Lambda}}_T = 4\left[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\right]^{-1}\left[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\ddot{\mathbf{V}}_T(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\right]$$
$$\left[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\right]^{-1},$$

and $\ddot{\mathbf{\Lambda}}_T^{-1/2}\sqrt{T}\ddot{\boldsymbol{\lambda}}_T \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$. The LM statistic is

$$\mathcal{LM}_T = T\ddot{\boldsymbol{\lambda}}_T'\ddot{\mathbf{\Lambda}}_T^{-1}\ddot{\boldsymbol{\lambda}}_T.$$

Theorem 6.12

Given $y_t = \mathbf{x}_t'\beta + e_t$, suppose that [B1](i), [B2] and [B3] hold. Then under the null, $\mathcal{LM}_T \xrightarrow{D} \chi^2(q)$, where q is the number of hypotheses.

Writing $\mathbf{R}\hat{\beta}_T - \mathbf{r} = \mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{X}\ddot{\beta}_T)/T = \mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{X}'\ddot{\mathbf{e}}/T$, $\ddot{\boldsymbol{\lambda}}_T = 2\big[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\big]^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{X}'\ddot{\mathbf{e}}/T$. The LM test is then

$$\mathcal{LM}_T = T\ddot{\mathbf{e}}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\big[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\ddot{\mathbf{V}}_T(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\big]^{-1} \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\ddot{\mathbf{e}}.$$

That is, the LM test requires only constrained estimation.

Note: Under the null, $\mathcal{W}_T - \mathcal{LM}_T \xrightarrow{\mathbb{P}} 0$; if $\mathbf{V}_o$ is known, the Wald and LM tests would be algebraically equivalent. (why?)

Example: Testing whether one would like to add additional s regressors to the specification: $y_t = \mathbf{x}'_{1,t}\mathbf{b}_1 + e_t$.

- The unconstrained specification is

$$y_t = \mathbf{x}'_{1,t}\mathbf{b}_1 + \mathbf{x}'_{2,t}\mathbf{b}_2 + e_t,$$

  and the null hypothesis is $\mathbf{R}\beta_o = \mathbf{0}$ with $\mathbf{R} = [\mathbf{0}_{s\times(k-s)}\ \mathbf{I}_s]$.

- The constrained estimator is $\ddot{\beta}_T = (\ddot{\mathbf{b}}'_{1,T}\ \mathbf{0}')'$, with $\ddot{\mathbf{b}}_{1,T} = (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y}$.

- Letting $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$ and $\ddot{\mathbf{e}} = \mathbf{y} - \mathbf{X}_1\ddot{\mathbf{b}}_{1,T}$, suppose that $\ddot{\mathbf{V}}_T = \ddot{\sigma}^2_T(\mathbf{X}'\mathbf{X}/T)$ is consistent for $\mathbf{V}_o$ under the null, where $\ddot{\sigma}^2_T = \sum_{t=1}^T \ddot{e}^2_t/(T-k+s)$. Then, the LM test is

$$\mathcal{LM}_T = T\ddot{\mathbf{e}}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\left[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\right]^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\ddot{\mathbf{e}}/\ddot{\sigma}^2_T.$$

Using the formula for the inverse of a partitioned matrix,

$$\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' = [\mathbf{X}_2'(\mathbf{I}-\mathbf{P}_1)\mathbf{X}_2]^{-1},$$

$$\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = [\mathbf{X}_2'(\mathbf{I}-\mathbf{P}_1)\mathbf{X}_2]^{-1}\mathbf{X}_2'(\mathbf{I}-\mathbf{P}_1).$$

As $\mathbf{X}_1'\ddot{\mathbf{e}} = \mathbf{0}$ and $(\mathbf{I}-\mathbf{P}_1)\ddot{\mathbf{e}} = \ddot{\mathbf{e}}$, the LM statistic is

$$\begin{aligned}\mathcal{LM}_T &= \ddot{\mathbf{e}}'(\mathbf{I}-\mathbf{P}_1)\mathbf{X}_2[\mathbf{X}_2'(\mathbf{I}-\mathbf{P}_1)\mathbf{X}_2]^{-1}\mathbf{X}_2'(\mathbf{I}-\mathbf{P}_1)\ddot{\mathbf{e}}/\ddot{\sigma}_T^2 \\ &= \ddot{\mathbf{e}}'\mathbf{X}_2[\mathbf{X}_2'(\mathbf{I}-\mathbf{P}_1)\mathbf{X}_2]^{-1}\mathbf{X}_2'\ddot{\mathbf{e}}/\ddot{\sigma}_T^2 \\ &= \ddot{\mathbf{e}}'\mathbf{X}_2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{X}_2'\ddot{\mathbf{e}}/\ddot{\sigma}_T^2.\end{aligned}$$

Note $\ddot{\mathbf{e}}'\mathbf{X}_2\mathbf{R} = [\mathbf{0}_{1\times(k-s)}\ \ddot{\mathbf{e}}'\mathbf{X}_2] = \ddot{\mathbf{e}}'\mathbf{X}$. A simple version of the LM test is

$$\mathcal{LM}_T = \frac{\ddot{\mathbf{e}}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\ddot{\mathbf{e}}}{\ddot{\mathbf{e}}'\ddot{\mathbf{e}}/(T-k+s)} = (T-k+s)R^2,$$

where R^2 is the non-centered R^2 of the auxiliary regression of $\ddot{\mathbf{e}}$ on $\mathbf{X}$.

Likelihood Ratio (LR) Test

- The OLS estimator $\hat{\boldsymbol{\beta}}_T$ is also the MLE $\tilde{\boldsymbol{\beta}}_T$ that maximizes

$$L_T(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{T}\sum_{t=1}^{T}\frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{2\sigma^2}.$$

With $\hat{e}_t = y_t - \mathbf{x}_t'\tilde{\boldsymbol{\beta}}_T$, the unconstrained MLE of σ^2 is

$$\tilde{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_t^2.$$

- Given $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, let $\ddot{\boldsymbol{\beta}}_T$ denote the constrained MLE of $\boldsymbol{\beta}$. Then $\ddot{e}_t = y_t - \mathbf{x}_t'\ddot{\boldsymbol{\beta}}_T$, and the constrained MLE of σ^2 is

$$\ddot{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T}\ddot{e}_t^2.$$

For $H_0 : \mathbf{R}\boldsymbol{\beta}_o = \mathbf{r}$, the LR test compares the constrained and unconstrained L_T:

$$\mathcal{LR}_T = -2T\big(L_T(\ddot{\beta}_T, \ddot{\sigma}_T^2) - L_T(\tilde{\beta}_T, \tilde{\sigma}_T^2)\big) = T \log\left(\frac{\ddot{\sigma}_T^2}{\tilde{\sigma}_T^2}\right).$$

The null would be rejected if $\mathcal{LR}_T$ is far from zero.

Theorem 6.15

Given $y_t = \mathbf{x}_t'\boldsymbol{\beta} + e_t$, suppose that [B1](i), [B2] and [B3] hold and that $\tilde{\sigma}_T^2(\mathbf{X}'\mathbf{X}/T)$ is consistent for $\mathbf{V}_o$. Then under the null hypothesis,

$$\mathcal{LR}_T \xrightarrow{D} \chi^2(q),$$

where q is the number of hypotheses.

Noting $\ddot{\mathbf{e}} = \mathbf{X}(\tilde{\boldsymbol{\beta}}_T - \ddot{\boldsymbol{\beta}}_T) + \hat{\mathbf{e}}$ and $\mathbf{X}'\hat{\mathbf{e}} = \mathbf{0}$, we have

$$\ddot{\sigma}_T^2 = \tilde{\sigma}_T^2 + (\tilde{\boldsymbol{\beta}}_T - \ddot{\boldsymbol{\beta}}_T)'(\mathbf{X}'\mathbf{X}/T)(\tilde{\boldsymbol{\beta}}_T - \ddot{\boldsymbol{\beta}}_T).$$

We have seen

$$\ddot{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_T = -(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\left[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}'\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}_T - \mathbf{r}).$$

It follows that

$$\ddot{\sigma}_T^2 = \tilde{\sigma}_T^2 + (\mathbf{R}\tilde{\boldsymbol{\beta}}_T - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}']^{-1}(\mathbf{R}\tilde{\boldsymbol{\beta}}_T - \mathbf{r}),$$

and that

$$\mathcal{LR}_T = T\log\big(1 + \underbrace{(\mathbf{R}\tilde{\boldsymbol{\beta}}_T - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X}/T)^{-1}\mathbf{R}']^{-1}(\mathbf{R}\tilde{\boldsymbol{\beta}}_T - \mathbf{r})/\tilde{\sigma}_T^2}_{=:\, a_T}\big).$$

Owing to consistency of $\hat{\beta}_T$, $a_T \to 0$. The mean value expansion of $\log(1 + a_T)$ about $a_T = 0$ yields

$$\log(1 + a_T) \approx (1 + a_T^{\dagger})^{-1} a_T,$$

where $a_T^{\dagger}$ lies between a_T and 0 and converges to zero. Then,

$$\mathcal{LR}_T = T(1 + a_T^{\dagger})^{-1} a_T = T a_T + o_{\mathbb{P}}(1),$$

where $T a_T$ is the Wald statistic with $\widehat{\mathbf{V}}_T = \tilde{\sigma}_T^2(\mathbf{X}'\mathbf{X}/T)$. When this $\widehat{\mathbf{V}}_T$ is consistent for $\mathbf{V}_o$, $\mathcal{LR}_T$ has a limiting $\chi^2(q)$ distribution.

Note: The applicability of the LR test here is limited because it can not be made robust to conditional heteroskedasticity and serial correlation. (Why?)

Remarks:

- When the Wald test involves $\widehat{\mathbf{V}}_T = \tilde{\sigma}_T^2(\mathbf{X}'\mathbf{X}/T)$ and the LM test uses $\ddot{\mathbf{V}}_T = \ddot{\sigma}_T^2(\mathbf{X}'\mathbf{X}/T)$, it can be shown that

$$\mathcal{W}_T \geq \mathcal{LR}_T \geq \mathcal{LM}_T.$$

  Hence, conflicting inferences in finite samples may arise when the critical values are between two statistics.

- When $\widehat{\mathbf{V}}_T = \tilde{\sigma}_T^2(\mathbf{X}'\mathbf{X}/T)$ and $\ddot{\mathbf{V}}_T = \ddot{\sigma}_T^2(\mathbf{X}'\mathbf{X}/T)$ are all consistent for $\mathbf{V}_o$, the Wald, LM, and LR tests are asymptotically equivalent.

Power of Tests

Consider the alternative hypothesis: $\mathbf{R}\boldsymbol{\beta}_o = \mathbf{r} + \boldsymbol{\delta}$, where $\boldsymbol{\delta} \neq \mathbf{0}$.

- Under the alternative,

$$\sqrt{T}(\mathbf{R}\hat{\boldsymbol{\beta}}_T - \mathbf{r}) = \sqrt{T}\mathbf{R}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_o) + \sqrt{T}\boldsymbol{\delta},$$

  where the first term on the RHS converges and the second term diverges.

- We have $\mathbb{P}(\mathcal{W}_T > c) \to 1$ for any critical value c, because

$$\frac{1}{T}\mathcal{W}_T \xrightarrow{\mathbb{P}} \boldsymbol{\delta}'(\mathbf{R}\mathbf{D}_o\mathbf{R}')^{-1}\boldsymbol{\delta}.$$

  The Wald test is therefore a consistent test.

Instrumental Variable Estimator

- OLS inconsistency:
  1. A model omits relevant regressors.
  2. A model includes lagged dependent variables as regressors and serially correlated errors.
  3. A model involves regressors that are measured with errors.
  4. The dependent variable and regressors are jointly determined at the same time (simultaneity problem).
  5. The dependent variable is determined by some unobservable factors which are correlated with regressors (selectivity problem).
- To obtain consistency, let $\mathbf{z}_t$ $(k \times 1)$ be variables taken from $(\mathcal{Y}^{t-1}, \mathcal{W}^t)$ such that $\mathbb{E}(\mathbf{z}_t \epsilon_t) = \mathbf{0}$ and $\mathbf{z}_t$ are correlated with $\mathbf{x}_t$ in the sense that $\mathbb{E}(\mathbf{z}_t \mathbf{x}_t')$ is not singular.

- The sample counterpart of $\mathbb{E}(\mathbf{z}_t \epsilon_t) = \mathbb{E}[\mathbf{z}_t(y_t - \mathbf{x}_t' \boldsymbol{\beta}_o)] = \mathbf{0}$ is

$$\frac{1}{T} \sum_{t=1}^{T} [\mathbf{z}_t(y_t - \mathbf{x}_t' \boldsymbol{\beta})] = \mathbf{0},$$

which is a system of k equations with k unknowns.

- The solution is the instrumental variable (IV) estimator:

$$\hat{\boldsymbol{\beta}}_{T,\mathrm{IV}} = \left(\sum_{t=1}^{T} \mathbf{z}_t \mathbf{x}_t' \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{z}_t y_t \right) \xrightarrow{\mathbb{P}} \mathbf{M}_{zx}^{-1} \mathbf{m}_{zy} = \boldsymbol{\beta}_o,$$

under suitable LLN.

  - This is also a method of moment estimator, because it solves the sample counterpart of the moment conditions: $\mathbb{E}[\mathbf{z}_t(y_t - \mathbf{x}_t' \boldsymbol{\beta}_o)] = \mathbf{0}$.
  - This method breaks down when more than k instruments are available.

- Assume CLT: $T^{-1/2}\sum_{t=1}^{T}\mathbf{z}_t\epsilon_t \xrightarrow{D} \mathcal{N}(\mathbf{0},\, \mathbf{V}_o)$ with

$$\mathbf{V}_o = \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{z}_t\epsilon_t\right).$$

- The normalized IV estimator has asymptotic normality:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{T,\mathrm{IV}}-\boldsymbol{\beta}_o) = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{z}_t\epsilon_t\right) \xrightarrow{D} \mathcal{N}(\mathbf{0},\, \mathbf{D}_o),$$

  where $\mathbf{D}_o = \mathbf{M}_{zx}^{-1}\mathbf{V}_o\mathbf{M}_{zx}^{-1}$.

- Then, $\widehat{\mathbf{V}}_T^{-1/2}\sqrt{T}(\hat{\boldsymbol{\beta}}_{T,\mathrm{IV}}-\boldsymbol{\beta}_o) \xrightarrow{D} \mathcal{N}(\mathbf{0},\mathbf{I}_k)$, where $\widehat{\mathbf{V}}_T$ is a consistent estimator for $\mathbf{V}_o$.

$I(1)$ Variables

$\{y_t\}$ is said to be an $I(1)$ (integrated of order 1) process if $y_t = y_{t-1} + \epsilon_t$, with ϵ_t satisfying:

[C1] $\{\epsilon_t\}$ is a weakly stationary process with mean zero and variance σ_ϵ^2 and obeys an FCLT:

$$\frac{1}{\sigma_* \sqrt{T}} \sum_{t=1}^{[Tr]} \epsilon_t = \frac{1}{\sigma_* \sqrt{T}} y_{[Tr]} \Rightarrow w(r), \qquad 0 \leq r \leq 1,$$

where w is standard Wiener process, and σ_*^2 is the long-run variance of ϵ_t:

$$\sigma_*^2 = \lim_{T \to \infty} \operatorname{var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t \right).$$

- Partial sums of an $I(0)$ series (e.g., $\sum_{i=1}^{t} \epsilon_i$) form an $I(1)$ series, while taking first difference of an $I(1)$ series (e.g., $y_t - y_{t-1}$) yields an $I(0)$ series.
  - A random walk is $I(1)$ with i.i.d. ϵ_t and $\sigma_*^2 = \sigma_\epsilon^2$.
  - When $\epsilon_t = y_t - y_{t-1}$ is a stationary ARMA(p, q) process, y is an $I(1)$ process and known as an ARIMA$(p, 1, q)$ process.
- An $I(1)$ series y_t has mean zero and variance increasing linearly with t, and its autocovariances $\text{cov}(y_t, y_s)$ do not decrease when $|t - s|$ increases.
- Many macroeconomic and financial time series are (or behave like) $I(1)$ processes.

ARIMA vs. ARMA Processes

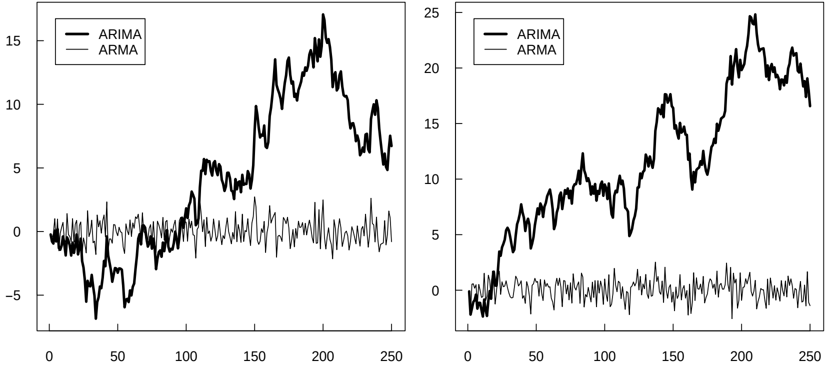

Figure: Sample paths of ARIMA and ARMA series.

$I(1)$ vs. Trend Stationarity

Trend stationary series: $y_t = a_o + b_o t + \epsilon_t$, where ϵ_t are $I(0)$.

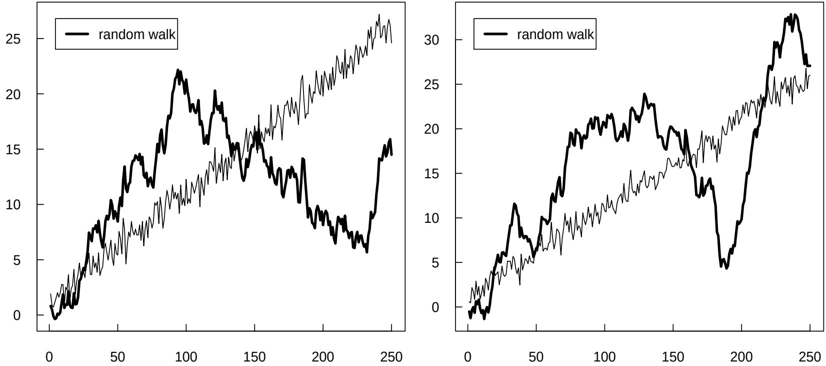

Figure: Sample paths of random walk and trend stationary series.

Autoregression of an $I(1)$ Variable

Suppose $\{y_t\}$ is a random walk such that $y_t = \alpha_o y_{t-1} + \epsilon_t$ with $\alpha_o = 1$ and ϵ_t i.i.d. random variables with mean zero and variance σ_ϵ^2.

- $\{y_t\}$ does not obey a LLN, and $\sum_{t=2}^T y_{t-1}\epsilon_t = O_{\mathbb{P}}(T)$ and $\sum_{t=2}^T y_{t-1}^2 = O_{\mathbb{P}}(T^2)$.
- Given the specification: $y_t = \alpha y_{t-1} + e_t$, the OLS estimator of α is:

$$\hat{\alpha}_T = \frac{\sum_{t=2}^T y_{t-1} y_t}{\sum_{t=2}^T y_{t-1}^2} = 1 + \frac{\sum_{t=2}^T y_{t-1}\epsilon_t}{\sum_{t=2}^T y_{t-1}^2} = 1 + O_{\mathbb{P}}(T^{-1}),$$

  which is T-consistent. This is also known as a super consistent estimator.

Lemma 7.1

Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Then,

(i) $T^{-3/2} \sum_{t=1}^{T} y_{t-1} \Rightarrow \sigma_* \int_0^1 w(r)\, \mathrm{d}r$;

(ii) $T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \Rightarrow \sigma_*^2 \int_0^1 w(r)^2\, \mathrm{d}r$;

(iii) $T^{-1} \sum_{t=1}^{T} y_{t-1}\epsilon_t \Rightarrow$
$$\frac{1}{2}[\sigma_*^2 w(1)^2 - \sigma_\epsilon^2] = \sigma_*^2 \int_0^1 w(r)\, \mathrm{d}w(r) + \frac{1}{2}(\sigma_*^2 - \sigma_\epsilon^2),$$

where w is the standard Wiener process.

Note: When y_t is a random walk, $\sigma_*^2 = \sigma_\epsilon^2$.

Theorem 7.2

Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Given the specification $y_t = \alpha y_{t-1} + e_t$, the normalized OLS estimator of α is:

$$T(\hat{\alpha}_T - 1) = \frac{\sum_{t=2}^{T} y_{t-1}\epsilon_t / T}{\sum_{t=2}^{T} y_{t-1}^2 / T^2} \Rightarrow \frac{\frac{1}{2}\left[w(1)^2 - \sigma_\epsilon^2/\sigma_*^2\right]}{\int_0^1 w(r)^2 \,\mathrm{d}r}.$$

where w is the standard Wiener process. When y_t is a random walk,

$$T(\hat{\alpha}_T - 1) \Rightarrow \frac{\frac{1}{2}\left[w(1)^2 - 1\right]}{\int_0^1 w(r)^2 \,\mathrm{d}r},$$

which does not depend on σ_ϵ^2 and σ_*^2 and is asymptotically pivotal.

Lemma 7.3

Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Then,

(i) $T^{-2} \sum_{t=1}^{T} (y_{t-1} - \bar{y}_{-1})^2 \Rightarrow \sigma_*^2 \int_0^1 w^*(r)^2 \, \mathrm{d}r;$

(ii) $T^{-1} \sum_{t=1}^{T} (y_{t-1} - \bar{y}_{-1}) \epsilon_t \Rightarrow \sigma_*^2 \int_0^1 w^*(r) \, \mathrm{d}w(r) + \frac{1}{2}(\sigma_*^2 - \sigma_\epsilon^2),$

where w is the standard Wiener process and $w^*(t) = w(t) - \int_0^1 w(r) \, \mathrm{d}r$.

Theorem 7.4

Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Given the specification $y_t = c + \alpha y_{t-1} + e_t$, the normalized OLS estimators of α and c are:

$$T(\hat{\alpha}_T - 1) \Rightarrow \frac{\int_0^1 w^*(r)\,\mathrm{d}w(r) + \frac{1}{2}(1 - \sigma_\epsilon^2/\sigma_*^2)}{\int_0^1 w^*(r)^2\,\mathrm{d}r} =: A,$$

$$\sqrt{T}\hat{c}_T \Rightarrow A\left(\sigma_* \int_0^1 w(r)\,\mathrm{d}r\right) + \sigma_* w(1).$$

In particular, when y_t is a random walk,

$$T(\hat{\alpha}_T - 1) \Rightarrow \frac{\int_0^1 w^*(r)\,\mathrm{d}w(r)}{\int_0^1 w^*(r)^2\,\mathrm{d}r}.$$

- The limiting results for autoregressions with an $I(1)$ variable are not invariant to model specification.
- All the results here are based on the data with DGP: $y_t = y_{t-1} + \epsilon_t$. intercept. These results would break down if the DGP is $y_t = c_o + y_{t-1} + \epsilon_t$ with a non-zero c_o; such series are said to be $I(1)$ with drift.
- $I(1)$ process with a drift:

$$y_t = c_o + y_{t-1} + \epsilon_t = c_o\, t + \sum_{i=1}^{t} \epsilon_i,$$

which contains a deterministic trend and an $I(1)$ series without drift.

Tests of Unit Root

1. Given the specification $y_t = \alpha y_{t-1} + e_t$, the unit root hypothesis is $\alpha_o = 1$, and a leading unit-root test is the t test:

$$\tau_0 = \frac{\left(\sum_{t=2}^{T} y_{t-1}^2\right)^{1/2}(\hat{\alpha}_T - 1)}{\hat{\sigma}_{T,1}},$$

where $\hat{\sigma}_{T,1}^2 = \sum_{t=2}^{T}(y_t - \hat{\alpha}_T y_{t-1})^2/(T-2)$.

2. Given the specification $y_t = c + \alpha y_{t-1} + e_t$, a unit-root test is

$$\tau_c = \frac{\left[\sum_{t=2}^{T}(y_{t-1} - \bar{y}_{-1})^2\right]^{1/2}(\hat{\alpha}_T - 1)}{\hat{\sigma}_{T,2}},$$

where $\hat{\sigma}_{T,2}^2 = \sum_{t=2}^{T}(y_t - \hat{c}_T - \hat{\alpha}_T y_{t-1})^2/(T-3)$.

Theorem 7.5

Let y_t be generated as a random walk. Then,

$$\tau_0 \Rightarrow \frac{\frac{1}{2}[w(1)^2 - 1]}{\left[\int_0^1 w(r)^2 \, \mathrm{d}r\right]^{1/2}},$$

$$\tau_c \Rightarrow \frac{\int_0^1 w^*(r) \, \mathrm{d}w(r)}{\left[\int_0^1 w^*(r)^2 \, \mathrm{d}r\right]^{1/2}}.$$

- For the specification with a time trend variable:

$$y_t = c + \alpha y_{t-1} + \beta\left(t - \frac{T}{2}\right) + e_t,$$

the t-statistic of $\alpha_o = 1$ is denoted as τ_t.

Dickey-Fuller distributions

Table: Some percentiles of the Dickey-Fuller distributions.

Test	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
τ_0	−2.58	−2.23	−1.95	−1.62	−0.51	0.89	1.28	1.62	2.01
τ_c	−3.42	−3.12	−2.86	−2.57	−1.57	−0.44	−0.08	0.23	0.60
τ_t	−3.96	−3.67	−3.41	−3.13	−2.18	−1.25	−0.94	−0.66	−0.32

- These distributions are not symmetric about zero and assume more negative values.
- τ_c assumes negatives values about 95% of times, and τ_t is virtually a non-positive random variable.

The Dickey-Fuller Distributions

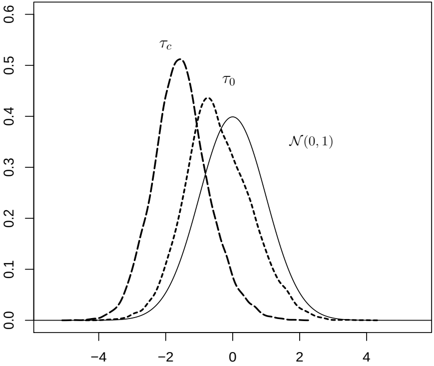

Figure: The distributions of the Dickey-Fuller τ_0 and τ_c tests vs. $\mathcal{N}(0,1)$.

Implementation

In practice, we estimate one of the following specifications:

1. $\Delta y_t = \theta y_{t-1} + e_t$.
2. $\Delta y_t = c + \theta y_{t-1} + e_t$.
3. $\Delta y_t = c + \theta y_{t-1} + \beta\big(t - \frac{T}{2}\big) + e_t$.

The unit-root hypothesis $\alpha_o = 1$ is now equivalent to $\theta_o = 0$.

- The weak limits of the normalized estimators $T\hat{\theta}_T$ are the same as the respective limits of $T(\hat{\alpha}_T - 1)$ under the null hypothesis.
- The unit-root tests are now computed as the t-ratios of these specifications.

Note: The Dickey-Fuller tests check only the random walk hypothesis and are invalid for testing general $I(1)$ processes.

Theorem 7.6

Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Then,

$$\tau_0 \Rightarrow \frac{\sigma_*}{\sigma_\epsilon} \left(\frac{\frac{1}{2}[w(1)^2 - \sigma_\epsilon^2/\sigma_*^2]}{\left[\int_0^1 w(r)^2 \, \mathrm{d}r\right]^{1/2}} \right),$$

$$\tau_c \Rightarrow \frac{\sigma_*}{\sigma_\epsilon} \left(\frac{\int_0^1 w^*(r) \, \mathrm{d}w(r) + \frac{1}{2}(1 - \sigma_\epsilon^2/\sigma_*^2)}{\left[\int_0^1 w^*(r)^2 \, \mathrm{d}r\right]^{1/2}} \right),$$

- Let $\hat{e}_t$ denote the OLS residuals and s^2_{Tn} a Newey-West type estimator of σ^2_* based on $\hat{e}_t$:

$$s^2_{Tn} = \frac{1}{T-1}\sum_{t=2}^{T}\hat{e}_t^2 + \frac{2}{T-1}\sum_{s=1}^{T-2}\kappa\left(\frac{s}{n}\right)\sum_{t=s+2}^{T}\hat{e}_t\hat{e}_{t-s},$$

with κ a kernel function and $n = n(T)$ its bandwidth.

- Phillips (1987) proposed the following modified τ_0 and τ_c statistics:

$$Z(\tau_0) = \frac{\hat{\sigma}_T}{s_{Tn}}\,\tau_0 - \frac{\frac{1}{2}(s^2_{Tn} - \hat{\sigma}^2_T)}{s_{Tn}\left(\sum_{t=2}^{T} y^2_{t-1}/T^2\right)^{1/2}},$$

$$Z(\tau_c) = \frac{\hat{\sigma}_T}{s_{Tn}}\,\tau_c - \frac{\frac{1}{2}(s^2_{T} - \hat{\sigma}^2_T)}{s_{Tn}\left[\sum_{t=2}^{T}(y_{t-1} - \bar{y}_{-1})^2\right]^{1/2}};$$

see also Phillips and Perron (1988).

The Phillips-Perron tests eliminate the nuisance parameters by suitable transformations of τ_0 and τ_c and have the same limits as those of the Dickey-Fuller tests.

Corollary 7.7.

Let $y_t = y_{t-1} + \epsilon_t$ be an $I(1)$ series with ϵ_t satisfying [C1]. Then,

$$Z(\tau_0) \Rightarrow \frac{\frac{1}{2}\left[w(1)^2 - 1\right]}{\left[\int_0^1 w(r)^2 \,\mathrm{d}r\right]^{1/2}},$$

$$Z(\tau_c) \Rightarrow \frac{\int_0^1 w^*(r) \,\mathrm{d}w(r)}{\left[\int_0^1 w^*(r)^2 \,\mathrm{d}r\right]^{1/2}}.$$

Augmented Dickey-Fuller (ADF) Tests

Said and Dickey (1984) suggest "filtering out" the correlations in a weakly stationary process by a linear AR model with a proper order. The "augmented" specifications are:

1. $\Delta y_t = \theta y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + e_t$.
2. $\Delta y_t = c + \theta y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + e_t$.
3. $\Delta y_t = c + \theta y_{t-1} + \beta\left(t - \frac{T}{2}\right) + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + e_t$.

Note: This approach avoids non-parametric kernel estimation of σ_*^2 but requires choosing a proper lag order k for the augmented specifications (say, by a model selection criteria, such as AIC or SIC).

KPSS Tests

$\{y_t\}$ is trend stationary if it fluctuates around a deterministic trend:

$$y_t = a_o + b_o\, t + \epsilon_t,$$

where ϵ_t satisfy [C1]. When $b_o = 0$, it is level stationary. Kwiatkowski, Phillips, Schmidt, and Shin (1992) proposed testing stationarity by

$$\eta_T = \frac{1}{T^2\, s_{Tn}^2} \sum_{t=1}^{T} \left(\sum_{i=1}^{t} \hat{e}_i \right)^2,$$

where s_{Tn}^2 is a Newey-West estimator of σ_*^2 based on $\hat{e}_t$.

- To test the null of trend stationarity, $\hat{e}_t = y_t - \hat{a}_T - \hat{b}_T\, t$.
- To test the null of level stationarity, $\hat{e}_t = y_t - \bar{y}$.

The partial sums of $\hat{e}_t = y_t - \bar{y}$ are such that

$$\sum_{t=1}^{[Tr]} \hat{e}_t = \sum_{t=1}^{[Tr]} (\epsilon_t - \bar{\epsilon}) = \sum_{t=1}^{[Tr]} \epsilon_t - \frac{[Tr]}{T} \sum_{t=1}^{T} \epsilon_t, \quad r \in (0, 1].$$

Then by a suitable FCLT,

$$\frac{1}{\sigma_* \sqrt{T}} \sum_{t=1}^{[Tr]} \hat{e}_t \Rightarrow w(r) - rw(1) = w^0(r).$$

Similarly, given $\hat{e}_t = y_t - \hat{a}_T - \hat{b}_T\, t$,

$$\frac{1}{\sigma_* \sqrt{T}} \sum_{t=1}^{[Tr]} \hat{e}_t \Rightarrow w(r) + (2r - 3r^2) w(1) - (6r - 6r^2) \int_0^1 w(s)\, \mathrm{d}s,$$

which is a "tide-down" process (it is zero at $r = 1$ with prob. one).

Theorem 7.8

Let $y_t = a_o + b_o\, t + \epsilon_t$ with ϵ_t satisfying [C1]. Then, η_T computed from $\hat{e}_t = y_t - \hat{a}_T - \hat{b}_T\, t$ is:

$$\eta_T \Rightarrow \int_0^1 f(r)^2 \, \mathrm{d}r,$$

where $f(r) = w(r) + (2r - 3r^2)w(1) - (6r - 6r^2)\int_0^1 w(s)\, \mathrm{d}s$.

Let $y_t = a_o + \epsilon_t$ with ϵ_t satisfying [C1]. Then, η_T computed from $\hat{e}_t = y_t - \bar{y}$ is:

$$\eta_T \Rightarrow \int_0^1 w^0(r)^2 \, \mathrm{d}r,$$

where w^0 is the Brownian bridge.

Table: Some percentiles of the distributions of the KPSS test.

Test	1%	2.5%	5%	10%
level stationarity	0.739	0.574	0.463	0.347
trend stationarity	0.216	0.176	0.146	0.119

- These tests have power against $I(1)$ series because η_T would diverge under $I(1)$ alternatives.
- KPSS tests also have power against other alternatives, such as stationarity with mean changes and trend stationarity with trend breaks. Thus, rejecting the null of stationarity does not imply that the series must be $I(1)$.

The KPSS Distributions

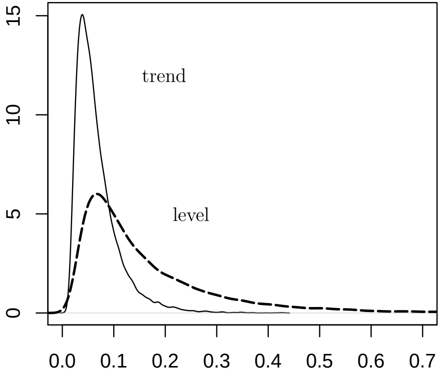

Figure: The distributions of the KPSS tests.

Spurious Regressions

- Granger and Newbold (1974): Regressing one random walk on the other typically yields a significant t-ratio. They refer to this result as spurious regression.
- Given the specification $y_t = \alpha + \beta x_t + e_t$, let $\hat{\alpha}_T$ and $\hat{\beta}_T$ denote the OLS estimators for α and β, respectively, and the corresponding t-ratios: $t_\alpha = \hat{\alpha}_T / s_\alpha$ and $t_\beta = \hat{\beta}_T / s_\beta$, where s_α and s_β are the OLS standard errors for $\hat{\alpha}_T$ and $\hat{\beta}_T$.
- $y_t = y_{t-1} + u_t$ and $x_t = x_{t-1} + v_t$, where $\{u_t\}$ and $\{v_t\}$ are mutually independent processes satisfying the following condition.

[C2] $\{u_t\}$ and $\{v_t\}$ are two weakly stationary processes with mean zero and respective variances σ_u^2 and σ_v^2 and obey an FCLT with:

$$\sigma_y^2 = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left(\sum_{t=1}^{T} u_t\right)^2, \quad \sigma_x^2 = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left(\sum_{t=1}^{T} v_t\right)^2.$$

We have the following results:

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} y_t \Rightarrow \sigma_y \int_0^1 w_y(r)\,\mathrm{d}r, \quad \frac{1}{T^2} \sum_{t=1}^{T} y_t^2 \Rightarrow \sigma_y^2 \int_0^1 w_y(r)^2\,\mathrm{d}r,$$

where w_y is a standard Wiener processes. Similarly,

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} x_t \Rightarrow \sigma_x \int_0^1 w_x(r)\,\mathrm{d}r, \quad \frac{1}{T^2} \sum_{t=1}^{T} x_t^2 \Rightarrow \sigma_x^2 \int_0^1 w_x(r)^2\,\mathrm{d}r.$$

We also have

$$\frac{1}{T^2}\sum_{t=1}^{T}(y_t-\bar{y})^2 \Rightarrow \sigma_y^2 \int_0^1 w_y(r)^2 \,\mathrm{d}r - \sigma_y^2 \left(\int_0^1 w_y(r)\,\mathrm{d}r\right)^2 =: \sigma_y^2 m_y,$$

$$\frac{1}{T^2}\sum_{t=1}^{T}(x_t-\bar{x})^2 \Rightarrow \sigma_x^2 \int_0^1 w_x(r)^2 \,\mathrm{d}r - \sigma_x^2 \left(\int_0^1 w_x(r)\,\mathrm{d}r\right)^2 =: \sigma_x^2 m_x,$$

where $w_y^*(t) = w_y(t) - \int_0^1 w_y(r)\,\mathrm{d}r$ and $w_x^*(t) = w_x(t) - \int_0^1 w_x(r)\,\mathrm{d}r$ are two mutually independent, "de-meaned" Wiener processes. Also,

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t=1}^{T}(y_t-\bar{y})(x_t-\bar{x}_t) \\
&\qquad \Rightarrow \sigma_y\sigma_x \left(\int_0^1 w_y(r)w_x(r)\,\mathrm{d}r - \int_0^1 w_y(r)\,\mathrm{d}r \int_0^1 w_x(r)\,\mathrm{d}r\right) \\
&\qquad =: \sigma_y\sigma_x m_{yx}.
\end{aligned}$$

Theorem 7.9

Let $y_t = y_{t-1} + u_t$ and $x_t = x_{t-1} + v_t$, where $\{u_t\}$ and $\{v_t\}$ are mutually independent and satisfy [C2]. Given the specification $y_t = \alpha + \beta x_t + e_t$,

(i) $\hat{\beta}_T \Rightarrow \dfrac{\sigma_y\, m_{yx}}{\sigma_x\, m_x}$,

(ii) $T^{-1/2}\hat{\alpha}_T \Rightarrow \sigma_y \left(\int_0^1 w_y(r)\,\mathrm{d}r - \dfrac{m_{yx}}{m_x} \int_0^1 w_x(r)\,\mathrm{d}r \right)$,

(iii) $T^{-1/2}\, t_\beta \Rightarrow \dfrac{m_{yx}}{(m_y m_x - m_{yx}^2)^{1/2}}$,

(iv) $T^{-1/2}\, t_\alpha \Rightarrow \dfrac{m_x \int_0^1 w_y(r)\,\mathrm{d}r - m_{yx} \int_0^1 w_x(r)\,\mathrm{d}r}{\left[(m_y m_x - m_{yx}^2) \int_0^1 w_x(r)^2\,\mathrm{d}r\right]^{1/2}}$,

w_x and w_y are two mutually independent, standard Wiener processes.

- While the true parameters should be $\alpha_o = \beta_o = 0$, $\hat{\beta}_T$ has a limiting distribution, and $\hat{\alpha}_T$ diverges at the rate $T^{1/2}$.
- Theorem 7.9 (iii) and (iv) indicate that t_α and t_β both diverge at the rate $T^{1/2}$ and are likely to reject the null of $\alpha_o = \beta_o = 0$ using the critical values from the standard normal distribution.
- Spurious trend: Nelson and Kang (1984) showed that, when $\{y_t\}$ is in fact a random walk, one may easily find significant time trend specification: $y_t = a + b\,t + e_t$.
- Phillips and Durlauf (1986) demonstrate that the F test (and hence the t-ratio) of $b_o = 0$ in the time trend specification above diverges at the rate T, which explains why an incorrect inference would result.

Cointegration

- Consider an equilibrium relation between y and x: $ay - bx = 0$. With real data (y_t, x_t), $z_t := ay_t - bx_t$ are equilibrium errors because they need not be zero all the time.
- y_t and x_t are both $I(1)$:
  - A linear combination of them, z_t, is, in general, an $I(1)$ series. Then, $\{z_t\}$ rarely crosses zero, and the equilibrium condition entails little empirical restriction on z_t.
  - When y_t and x_t involve the same random walk q_t such that $y_t = q_t + u_t$ and $x_t = cq_t + v_t$, where $\{u_t\}$ and $\{v_t\}$ are $I(0)$. Then,

$$z_t := cy_t - x_t = cu_t - v_t,$$

    which is a linear combination of $I(0)$ series and hence is also $I(0)$.

- Granger (1981), Granger and Weiss (1983), and Engle and Granger (1987): Let $\mathbf{y}_t$ be a d-dimensional vector $I(1)$ series. The elements of $\mathbf{y}_t$ are cointegrated if there exists a $d \times 1$ vector, $\boldsymbol{\alpha}$, such that $z_t = \boldsymbol{\alpha}'\mathbf{y}_t$ is $I(0)$. We say the elements of $\mathbf{y}_t$ are CI(1,1).
- The vector $\boldsymbol{\alpha}$ is a cointegrating vector. The space spanned by linearly independent cointegating vectors is the cointegrating space; the number of linearly independent cointegrating vectors is the cointegrating rank which is the dimension of the cointegrating space.
- If the cointegrating rank is r, we can put r linearly independent cointegrating vectors together and form the $d \times r$ matrix $\mathbf{A}$ such that $\mathbf{z}_t = \mathbf{A}'\mathbf{y}_t$ is a vector $I(0)$ series.
- The cointegrating rank is at most $d - 1$. (Why?)

Cointegrating Regression

- Cointegrating regression: $y_{1,t} = \boldsymbol{\alpha}'\mathbf{y}_{2,t} + \mathbf{z}_t$. Then, $(1\ \boldsymbol{\alpha}')'$ is the cointegrating vector and z_t are the regression (equilibrium) errors.
- When the elements of $\mathbf{y}_t$ are cointegrated, z_t is correlated with $\mathbf{y}_{2,t}$. Consistency of the OLS estimators do not matter asymptotically, but correlation would result in finite-sample bias and efficiency loss.
- Efficiency: Saikkonen (1991) proposed a modified co-integrating regression:

$$y_{1,t} = \boldsymbol{\alpha}'\mathbf{y}_{2,t} + \sum_{j=-k}^{k} \Delta\mathbf{y}'_{2,t-j}\mathbf{b}_j + e_t,$$

so that the OLS estimator of $\boldsymbol{\alpha}$ is asymptotically efficient.

Tests of Cointegration

- One can verify a cointegration relation by applying unit-root tests, such as the augmented Dickey-Fuller test and the Phillips-Perron test, to $\hat{z}_t$. The null hypothesis that a unit root is present is equivalent to the hypothesis of no cointegration.
- To implement a unit-root test on cointegration residuals $\hat{z}_T$, a difficulty is that $\hat{z}_T$ is not a raw series but a result of OLS fitting. Thus, even when z_t may be $I(1)$, the residuals $\hat{z}_t$ may not have much variation and hence behave like a stationary series.
- Engle and Granger (1987), Engle and Yoo (1987), and Davidson and MacKinnon (1993) simulated proper critical values for the unit-root tests on cointegrating residuals. Similar to the unit-root tests discussed earlier, these critical values are all "model dependent."

Table: Some percentiles of the distributions of the cointegration τ_c test.

d	1%	2.5%	5%	10%
2	-3.90	-3.59	-3.34	-3.04
3	-4.29	-4.00	-3.74	-3.45
4	-4.64	-4.35	-4.10	-3.81

- Drawbacks of cointegrating regressions:
    1. The choice of the dependent variable is somewhat arbitrary.
    2. This approach is more suitable for finding only one cointegrating relationship. One may estimate multiple cointegration relations by a vector regression.
- It is now typical to adopt the maximum likelihood approach of Johansen (1988) to estimate the cointegrating space directly.

Error Correction Model (ECM)

- When the elements of $\mathbf{y}_t$ are cointegrated with $\mathbf{A}'\mathbf{y}_t = \mathbf{z}_t$, then there exists an error correction model (ECM):

$$\Delta \mathbf{y}_t = \mathbf{B}\mathbf{z}_{t-1} + \mathbf{C}_1 \Delta \mathbf{y}_{t-1} + \cdots + \mathbf{C}_k \Delta \mathbf{y}_{t-k} + \nu_t.$$

- Cointegration characterizes the long-run equilibrium relations because it deals with the levels of $I(1)$ variables, and the ECM deals with the differences of variables and describes short-run dynamics.
- When cointegration exists, a vector AR model of $\Delta \mathbf{y}_t$ is misspecified because it omits $\mathbf{z}_{t-1}$, and the parameter estimates are inconsistent.
- We regress $\Delta \mathbf{y}_t$ on $\hat{\mathbf{z}}_{t-1}$ and lagged $\Delta \mathbf{y}_t$. Here, standard asymptotic theory applies because ECM involves only stationary variables when cointegration exists.