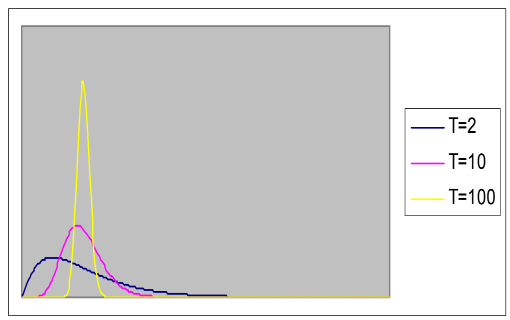
Asymptotic and small-sample distributions

A. White noise: review of asymptotic results

$$y_t = \mu + \varepsilon_t$$
$$E(\varepsilon_t) = 0$$
$$E(\varepsilon_t \varepsilon_s) = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\mu} = T^{-1} \sum_{t=1}^{T} y_t$$
$$E(\hat{\mu}) = \mu$$
$$E(\hat{\mu} - \mu)^2 = \sigma^2/T$$



Since

$$E(\hat{\mu}_T - \mu)^2 = \sigma^2/T$$

it follows that

$$\lim_{T\to\infty} E(\hat{\mu}_T - \mu)^2 = 0$$

or "$\hat{\mu}_T$ converges in mean square to μ"

Can also show that for any $\delta > 0$

$$\lim_{T\to\infty} \text{Prob}[|\hat{\mu}_T - \mu| > \delta] = 0$$

or "$\hat{\mu}_T$ converges in probability to μ"

Law of large numbers:

Under general conditions, the sample mean $(\hat{\mu}_T = T^{-1} \sum_{t=1}^{T} y_t)$ converges in probability to the population mean (μ)

To talk about a nondegenerate "asymptotic distribution," we must spread distribution out as T gets big

Consider:

$$\eta_T = \sqrt{T}(\hat{\mu}_T - \mu)$$

$$E(\eta_T) = \sqrt{T}[E(\hat{\mu}_T) - \mu]$$

$$= \sqrt{T}[\mu - \mu]$$

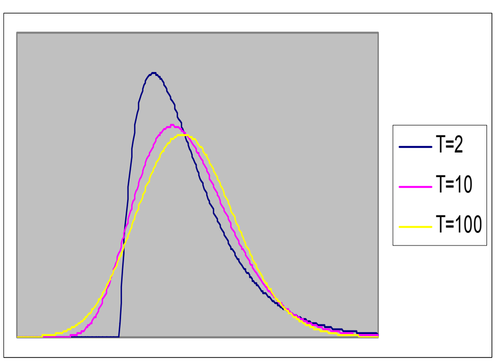
$$= 0 \quad \text{for all } T$$

$$\eta_T = \sqrt{T}(\hat{\mu}_T - \mu)$$

$$E(\eta_T^2) = TE(\hat{\mu}_T - \mu)^2$$

$$= T\sigma^2/T$$

$$= \sigma^2 \quad \text{for all } T$$



We say that a variable such as η_T "converges in distribution" to a random variable η if

$$\lim_{T\to\infty} \text{Prob}(\eta_T < x) = \text{Prob}(\eta < x)$$

for all continuity points x

Central Limit Theorem

Let $y_t = \mu + \varepsilon_t$

where ε_t is a martingale difference sequence $(E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_1) = 0)$ with $E(\varepsilon_t) = 0,\ E(\varepsilon_t^2 = \sigma^2),$ and $E(\varepsilon_t^4) < \infty$. Let

$$\hat{\mu}_T = T^{-1} \sum_{t=1}^{T} y_t$$

then $\sqrt{T}\,(\hat{\mu}_T - \mu)$ converges in distribution to a $N(0, \sigma^2)$ variable

Useful facts about convergence:

If $\mathbf{x}_T \xrightarrow{p} \mathbf{c}$ and $\mathbf{y}_T \xrightarrow{L} \mathbf{y}$, then

(a) $\mathbf{x}_T + \mathbf{y}_T \xrightarrow{L} \mathbf{c} + \mathbf{y}$

(b) $\mathbf{x}_T' \mathbf{y}_T \xrightarrow{L} \mathbf{c}' \mathbf{y}$

Note: any hypothesis test based on the asymptotic distribution is just an approximation.

This approximation may be very good or very poor

How can we find the exact small-sample distribution of η_T?

(1) Derive analytically

e.g., if $y_t \sim \Gamma(\alpha, \beta)$

then $\sum_{t=1}^{T} y_t \sim \Gamma(T\alpha, \beta)$

and $T^{-1} \sum_{t=1}^{T} y_t \sim \Gamma(T\alpha, T\beta)$

(2) Parametric bootstrap

(a) Generate a sample of T observations from a $\Gamma(\alpha, \beta)$ distribution

(b) Calculate $\hat{\mu}_T = T^{-1} \sum_{t=1}^{T} y_t$ for this artificial sample

(c) Repeat steps (a) and (b) $10{,}000$ times

(d) Calculate fraction of these samples for which $\hat{\mu}_T$ exceeds value x of interest

(3) Nonparametric bootstrap

(a) Take the observed sample $y_1, y_2, \ldots, y_T$ as given numbers

(b) Define an artificial discrete-valued random variable X such that

Prob$(X = y_1) = 1/T$

Prob$(X = y_2) = 1/T$

$\vdots$

Prob$(X = y_T) = 1/T$

(c) Generate a sample of T independent values of the random variable X

(d) Calculate $v_T = T^{-1} \sum_{i=1}^{T} X_i$

(e) Repeat steps (c)-(d) $10,000$ times

(f) Calculate fraction of these samples for which v_T exceeds value x of interest

Asymptotic and small-sample distributions

A. White noise: review of asymptotic results

B. Stationary AR(1) process

Consider next an AR(1) process:

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

$\varepsilon_t \sim$ i.i.d. $(0, \sigma^2)$

$|\rho| < 1$

Then

$E(y_t) = 0$

Var$(y_t) = \sigma^2/(1-\rho^2)$

Estimate by OLS regression of y_t on y_{t-1}

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

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

$$\hat{\rho} = \rho + \frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

$$\sqrt{T}(\hat{\rho}_T - \rho) = \frac{\sqrt{T}\, T^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t}{T^{-1} \sum_{t=1}^{T} y_{t-1}^2}$$

Numerator of $\sqrt{T}(\hat{\rho}_T - \rho)$

$\sqrt{T}\, T^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t$

Notice $y_{t-1}\varepsilon_t$ is martingale difference

$E(y_{t-1}\varepsilon_t y_{t-s}\varepsilon_{t-s}) = 0$ for $s > 0$

with mean $E(y_{t-1}\varepsilon_t) = 0$ and variance

$E(y_{t-1}^2 \varepsilon_t^2) = \sigma^4/(1-\rho^2)$

So by central limit theorem,

$\sqrt{T}\, T^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{L} N(0, \sigma^4/(1-\rho^2))$

Denominator of $\sqrt{T}\,(\hat{\rho}_T - \rho)$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}^2$$

is the sample mean of y_{t-1}^2 which by law of large numbers converges in probability to its population mean,

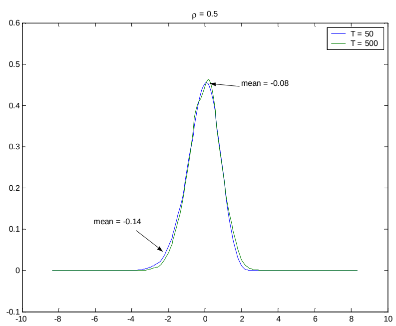
$$T^{-1}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{p} \sigma^2/(1-\rho^2)$$

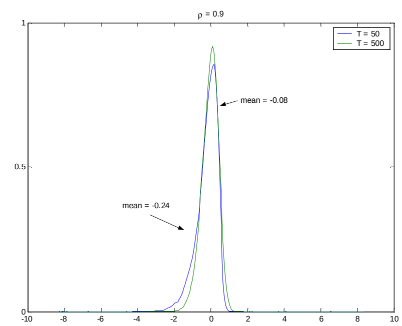
Conclusion:

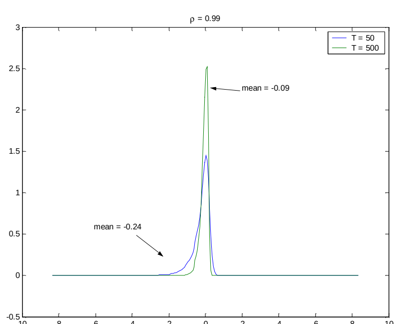
$$\sqrt{T}\,(\hat{\rho}_T - \rho) = \frac{\sqrt{T}\,T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{T^{-1}\sum_{t=1}^{T} y_{t-1}^2}$$

is asymptotically a $N(0, \sigma^4/(1-\rho^2))$ variable divided by the constant $\sigma^2/(1-\rho^2)$, or

$$\sqrt{T}\,(\hat{\rho}_T - \rho) \xrightarrow{L} N(0, 1-\rho^2)$$







Observations

$\hat{\rho}$ is downward-biased in small samples

This bias is more severe as ρ gets bigger

Normal approximation gets better as T gets bigger

Variance of $\widehat{\rho}$ around ρ gets smaller as ρ gets bigger

$$\sqrt{T}\,(\widehat{\rho}_T - \rho) \xrightarrow{L} N(0, 1-\rho^2)$$

Next consider the OLS t-statistic

$$\tau_\rho = \frac{\widehat{\rho} - \rho}{\widehat{\sigma}_{\widehat{\rho}}}$$

$$\widehat{\sigma}^2_{\widehat{\rho}} = \frac{\widehat{\sigma}^2}{\sum_{t=1}^{T} y_{t-1}^2}$$

$$\tau_\rho = \frac{\sqrt{T}\,(\widehat{\rho}_T - \rho)}{\sqrt{\widehat{\sigma}^2_T/(T^{-1}\sum y_{t-1}^2)}}$$

Numerator of τ_ρ

$$\sqrt{T}\,(\widehat{\rho}_T - \rho) \xrightarrow{L} N(0, 1-\rho^2)$$

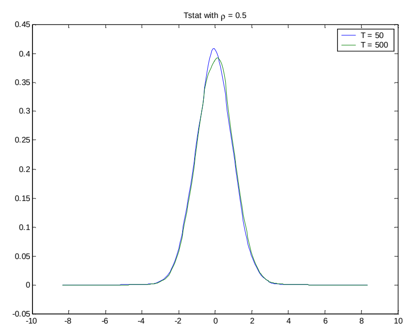
Denominator of τ_ρ

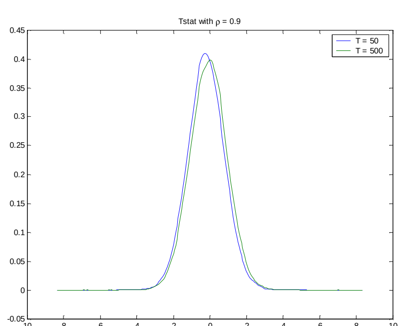
$$\sqrt{\widehat{\sigma}^2_T/(T^{-1}\sum y_{t-1}^2)} \xrightarrow{p} \sqrt{\sigma^2/[\sigma^2/(1-\rho^2)]}$$

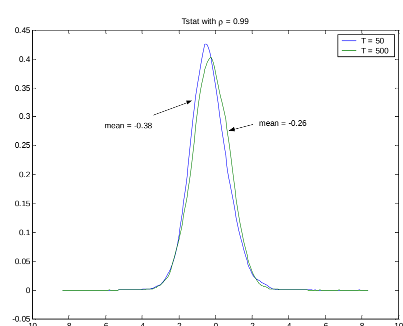
$$= \sqrt{1-\rho^2}$$

Conclusion

$$\tau_\rho \xrightarrow{L} N(0,1)$$







Observations

τ_ρ is negatively skewed relative to the $N(0,1)$ in small samples

This skew is more severe as ρ gets bigger

Normal approximation gets better as T gets bigger

Summary of asymptotic results:

$$\sqrt{T}(\hat{\rho}_T - \rho) \xrightarrow{L} N(0, 1-\rho^2)$$

$$\tau_\rho \xrightarrow{L} N(0,1)$$

Conjecture: when $\rho = 1$,

$$\sqrt{T}(\hat{\rho}_T - \rho) \xrightarrow{p} 0$$

$\tau_\rho \xrightarrow{L}$ something nonstandard

Asymptotic and small-sample distributions

A. White noise: review of asymptotic results
B. Stationary AR(1) process
C. Nonstationary AR(1) process

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

true $\rho_0 = 1$

$$\hat{\rho} = \frac{\sum_{t=1}^T y_{t-1} y_t}{\sum_{t=1}^T y_{t-1}^2}$$

$$= \rho_0 + \frac{\sum_{t=1}^T y_{t-1}\varepsilon_t}{\sum_{t=1}^T y_{t-1}^2}$$

$$\sqrt{T}(\hat{\rho}_T - \rho_0) \xrightarrow{p} 0$$

$$T(\hat{\rho}_T - \rho_0) = \frac{T^{-1}\sum_{t=1}^T y_{t-1}\varepsilon_t}{T^{-2}\sum_{t=1}^T y_{t-1}^2}$$

Numerator of $T(\hat{\rho}_T - \rho_0) = T^{-1}\sum_{t=1}^T y_{t-1}\varepsilon_t$

Notice under $H_0 : \rho_0 = 1$

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

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

$$y_{t-1}\varepsilon_t = (1/2)(y_t^2 - y_{t-1}^2 - \varepsilon_t^2)$$

$$T^{-1}\sum_{t=1}^T y_{t-1}\varepsilon_t = T^{-1}(1/2)\left(y_T^2 - y_0^2 - \sum_{t=1}^T \varepsilon_t^2\right)$$

Numerator of $T(\hat{\rho}_T - \rho_0)$

$$= T^{-1}(1/2)\left(y_T^2 - y_0^2 - \sum_{t=1}^T \varepsilon_t^2\right)$$

Suppose $y_0 = 0$ (asymptotically same)

$$T^{-1}\sum_{t=1}^T \varepsilon_t^2 \xrightarrow{p} \sigma^2$$

by law of large numbers

$$T^{-1}y_T^2 = T^{-1}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_T)^2$$
$$= \left(\sqrt{T}\,T^{-1}\sum_{t=1}^{T}\varepsilon_t\right)^2$$

which by central limit theorem converges to the square of a $N(0,\sigma^2)$ $= \sigma^2\chi^2(1)$

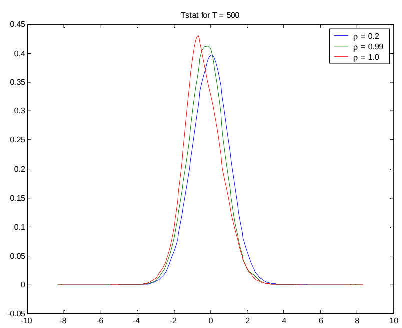
Conclusion: numerator of $T(\hat{\rho}_T - \rho_0)$

$$= T^{-1}(1/2)\left(y_T^2 - y_0^2 - \sum_{t=1}^{T}\varepsilon_t^2\right)$$

converges in distribution to

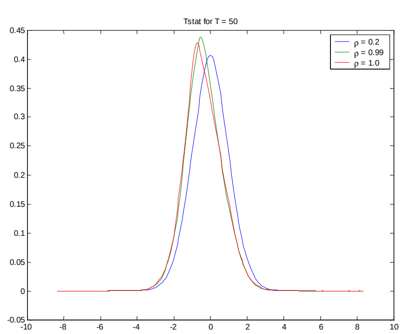
$\sigma^2(1/2)[\chi^2(1) - 1]$

Denominator of $T(\hat{\rho}_T - \rho_0)$ also has a nonstandard asymptotic distribution when $\rho_0 = 1$, and t-statistic τ_ρ also has nonstandard distribution



Probability that $\tau_\rho < -1.96 = 0.05$

Probability that $N(0,1) < -1.96 = 0.025$



Asymptotic and small-sample distributions

A. White noise: review of asymptotic results
B. Stationary AR(1) process
C. Nonstationary AR(1) process
D. Stationary AR(p) process

$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$

$\varepsilon_t \sim$ i.i.d. $(0, \sigma^2)$

$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$

implies $|z| > 1$

e.g., $p = 1$ requires $|\phi_1| < 1$

$\mathbf{x}_t = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p})'$

$y_t = \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t$

OLS regression of y_t on $\mathbf{x}_t$

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

$$= \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \left(\sum_{t=1}^T \mathbf{x}_t (\mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t)\right)$$

$$= \boldsymbol{\beta} + \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \left(\sum_{t=1}^T \mathbf{x}_t \varepsilon_t\right)$$

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) = \left(T^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \left(T^{-1/2} \sum_{t=1}^T \mathbf{x}_t \varepsilon_t\right)$$

First term of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) =$

$$\left(T^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'\right)^{-1}$$

Each element is sample mean of stationary random variable. By law of large numbers, converges to population mean

$$\left(T^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \xrightarrow{p} \left[E(\mathbf{x}_t \mathbf{x}_t')\right]^{-1} = \boldsymbol{\Gamma}_0^{-1}$$

Second term of $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) =$

$$T^{-1/2} \sum_{t=1}^T \mathbf{x}_t \varepsilon_t$$

Each element is $\sqrt{T}$ times sample mean of martingale difference sequence.
By central limit theorem, this converges to a Normal variable with mean zero and variance $\sigma^2 \boldsymbol{\Gamma}_0$

Conclusion:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) = \left(T^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \left(T^{-1/2} \sum_{t=1}^T \mathbf{x}_t \varepsilon_t\right)$$

$\xrightarrow{L} \boldsymbol{\Gamma}_0^{-1} \mathbf{z}$ where $\mathbf{z} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Gamma}_0)$

or $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \xrightarrow{L} N(\mathbf{0}, \sigma^2 \boldsymbol{\Gamma}_0^{-1})$

where $\boldsymbol{\Gamma}_0 = \text{plim } T^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'$

Practical implication:

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \approx N\left(\mathbf{0}, \sigma^2 \left[\text{plim } T^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'\right]^{-1}\right)$$

$$\hat{\boldsymbol{\beta}} \approx N\left(\boldsymbol{\beta}, \sigma^2 \left[\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'\right]^{-1}\right)$$

which is the usual exact small-sample result for the classical regression model

Usual F test of hypothesis $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$

$$F = m^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \left[s^2 \mathbf{R} \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \mathbf{R}' \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$$

Under H_0,

$$F_T = m^{-1}[\mathbf{R}\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})]' \left[s^2 \mathbf{R} \left(T^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \mathbf{R}' \right]^{-1} [\mathbf{R}\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})]$$

But

$$[\mathbf{R}\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})] \xrightarrow{L} \mathbf{q} \sim N(\mathbf{0}, \sigma^2 \mathbf{R}\boldsymbol{\Gamma}_0^{-1}\mathbf{R}')$$

so

$$F_T \xrightarrow{L} m^{-1}\mathbf{q}' \left[\sigma^2 \mathbf{R}\boldsymbol{\Gamma}_0^{-1}\mathbf{R}' \right]^{-1} \mathbf{q} \sim m^{-1}\chi^2(\text{rank } \mathbf{R})$$

Conclusion: Usual t or F statistics of any hypothesis test for stationary AR(p) are asymptotically valid