MA Advanced Econometrics: Applying Least Squares to Time Series

Karl Whelan

School of Economics, UCD

February 15, 2011

Part I

Time Series: Standard Asymptotic Results

OLS Estimates of $AR(n)$ Models Are Biased

- Consider the AR(1) model
$$y_t = \rho y_{t-1} + \epsilon_t \tag{1}$$
- The OLS estimator for a sample of size T is
$$\begin{aligned}\hat{\rho} &= \frac{\sum_{t=2}^{T} y_{t-1} y_t}{\sum_{t=2}^{T} y_{t-1}^2} &(2)\\ &= \rho + \frac{\sum_{t=2}^{T} y_{t-1}\epsilon_t}{\sum_{t=2}^{T} y_{t-1}^2} &(3)\\ &= \rho + \sum_{t=2}^{T}\left(\frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}\right)\epsilon_t &(4)\end{aligned}$$
- ϵ_t is independent of y_{t-1}, so $\mathbb{E}\left(y_{t-1}\epsilon_t\right) = 0$. However, ϵ_t is **not independent** of the sum $\sum_{t=2}^{T} y_{t-1}^2$. If ρ is positive, then a positive shock to ϵ_t raises current and future values of y_t, all of which are in the sum $\sum_{t=2}^{T} y_{t-1}^2$. This means there is a negative correlation between ϵ_t and $\frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$, so $\mathbb{E}\,\hat{\rho} < \rho$. (More on this later.)

Time Series: Serially Dependent Observations

- OLS estimates of AR models are biased. What about consistency? Do the estimates get closer to the correct value as samples get larger? The recipe for deriving asympototic properties of estimators has been to use a Law of Large Numbers and a Central Limit Theorem. Up to now, we have only discussed regressions using observations that are independently distributed and have used versions of the LLN and CLT for independent observations.
- However, observations from time series are not independent. For instance, for an $AR(n)$ process of the form

$$y_t = \alpha + \rho_1 y_{t-1} + \rho_1 y_{t-1} + \ldots + \rho_n y_{t-n} + \epsilon_t \tag{5}$$

  each observation depends on what happens in the past.
- LLNs and CLTs may not work for time series. For example, suppose y_t is highly autocorrelated, so that when the series has high values, it tends to stay high and when the series is low, it tends to stay low. This might mean that even if we have a lot of observations, we can't necessarily be sure that the sample average is a good estimator of the population average.
- It turns out there are some conditions under which WLLNs and CLTs hold for time series but these conditions sometimes don't hold.

Some Definitions

- We say that a time series of observations $\{y_t\}$ is **covariance (weakly) stationary** if $\mathbb{E}\, y_t = \mu$ for all t and $\text{Cov}\,(y_t, y_{t-k})$ is independent of t.
- As y_t moves up and down, then $\mathbb{E}\,(y_t|y_{t-1})$ will generally change. However, the stationarity here refers to the *ex ante* unconditional distribution, i.e. the distribution of outcomes that would have been expected before time has begun.
- We say that $\{y_t\}$ is **strictly stationary** if the unconditional joint distribution of $\{y_t, y_{t-1}, ..., y_{t-k}\}$ is independent of t for all k.
- For a weakly stationary series, let $\text{Cov}\,(y_t, y_{t-k}) = \gamma(k)$. We say that $\{y_t\}$ is **ergodic** if $\gamma(k) \rightarrow 0$ as $k \rightarrow \infty$.
- Let $\mathcal{F}_t = \{y_t, y_{t-1}, ..., y_{t-k}\}$. We say that e_t is a **martingale difference sequence** (MDS) if $\mathbb{E}\,(e_t|\,\mathcal{F}_{t-1}) = 0$.
- Armed with these definitions, we can state some theorems that allow us to make statements about the asympototic behaviour of least squares estimates of time series models.

Time Series Versions of LLN and CLT

- **Ergodic Theorem**: If y_t is strictly stationary and ergodic and $\mathbb{E}|y_t| < \infty$ than as $T \to \infty$

$$\frac{1}{T}\sum_{i=1}^{T} y_i \xrightarrow{p} \mathbb{E}(y_t) \tag{6}$$

- **MDS Central Limit Theorem**: If u_t is a strictly stationary and ergodic MDS and $\mathbb{E}(u_t u_t') = \Omega < \infty$, then as $T \to \infty$

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{T} u_i \xrightarrow{d} N(0, \Omega) \tag{7}$$

- The following will be useful in applying these results
  - If y_t is strictly stationary and ergodic and $x_t = f(y_t, y_{t-1}.....)$ is a random variable, then x_t is also strictly stationary and ergodic.
  - For the AR(1) model $y_t = \alpha + \rho y_{t-1} + \epsilon_t$, the series is strictly stationary and ergodic if $|\rho| < 1$.
  - For the AR(k) model $y_t = \alpha + \rho_1 y_{t-1} + ... + \rho_k y_{t-k} + \epsilon_t$, the series is strictly stationary and ergodic if the roots of the polynomial $\rho_1 L + ... + \rho_k L^k$ are all less than one in absolute value.

Estimating an AR(k) Regression

- Consider estimating AR(k) model $y_t = \alpha + \rho_1 y_{t-1} + ... + \rho_k y_{t-k} + \epsilon_t$. Let

$$x_t = \begin{pmatrix} 1 & y_{t-1} & y_{t-2} & \dots & y_{t-k} \end{pmatrix}' \quad (8)$$

$$\beta = \begin{pmatrix} \alpha & \rho_1 & \rho_2 & \dots & \rho_k \end{pmatrix}' \quad (9)$$

- The vector x_t is strictly stationary and ergodic, which means that $x_t x_t'$ also is. Thus, we can use the ergodic theorem to show that

$$\frac{1}{T} \sum_{i=1}^{T} x_t x_t' \xrightarrow{p} \mathbb{E}\left(x_t x_t'\right) = Q \quad (10)$$

- We can also show that $x_t \epsilon_t$ is stationary and ergodic, so

$$\frac{1}{T} \sum_{i=1}^{T} x_t \epsilon_t \xrightarrow{p} \mathbb{E}\left(x_t \epsilon_t\right) = 0 \quad (11)$$

- This means OLS estimators, though biased, are consistent:

$$\hat{\beta} = \beta + \left(\frac{1}{T} \sum_{i=1}^{T} x_t x_t'\right)^{-1} \left(\frac{1}{T} \sum_{i=1}^{T} x_t \epsilon_t\right) \xrightarrow{p} Q^{-1}.0 = 0 \quad (12)$$

Asymptotic Distribution of OLS Estimator

- Let $u_t = x_t e_t$. This is a MDS because

$$\mathbb{E}\left(u_t \middle| \mathcal{F}_{t-1}\right) = \mathbb{E}\left(x_t e_t \middle| \mathcal{F}_{t-1}\right) = x_t\, \mathbb{E}\left(e_t \middle| \mathcal{F}_{t-1}\right) = 0 \tag{13}$$

- Applying the MDS version of the CLT

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{T} x_t e_t \xrightarrow{d} N\left(0, \Omega\right) \tag{14}$$

where

$$\Omega = \mathbb{E}\left(x_t x_t' e_t^2\right) \tag{15}$$

- This means that if y_t is an $AR(k)$ process that is strictly stationary and ergodic and $\mathbb{E}\, y_t^4 < \infty$ then

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, Q^{-1}\Omega Q^{-1}\right) \tag{16}$$

- The condition $\mathbb{E}\, y_t^4 < \infty$ is required for the covariance matrix to be finite.

Example AR(1) Regression

- Consider using OLS to estimate the coefficient ρ for the series

$$y_t = \rho y_{t-1} + \epsilon_t \tag{17}$$

- The previous results tell us that

$$\sqrt{T}\left(\hat{\rho} - \rho\right) \xrightarrow{d} N\left(0, \omega\right) \tag{18}$$

where

$$\omega = \frac{\mathbb{E}\left(y_{t-1}^2 \epsilon_t^2\right)}{\left(\mathbb{E}\, y_{t-1}^2\right)^2} \tag{19}$$

- Letting $\text{Var}\left(\epsilon_t\right) = \sigma_\epsilon^2$, we can calculate the asymptotic variance of y_t from

$$\text{Var}\left(y_t\right) = \rho^2 \text{Var}\left(y_{t-1}\right) + \sigma_\epsilon^2 \Rightarrow \text{Var}\left(y_t\right) \rightarrow \frac{\sigma_\epsilon^2}{1 - \rho^2} = \sigma_y^2 \tag{20}$$

- Thus

$$\omega = \frac{\mathbb{E}\left(y_{t-1}^2 \epsilon_t^2\right)}{\left(\mathbb{E}\, y_{t-1}\right)^2} = \frac{\sigma_y^2 \sigma_\epsilon^2}{\left(\sigma_y^2\right)^2} = \frac{\sigma_\epsilon^2}{\sigma_y^2} = 1 - \rho \Rightarrow \sqrt{T}\left(\hat{\rho} - \rho\right) \xrightarrow{d} N\left(0, 1 - \rho^2\right) \tag{21}$$

Part II

Unit Roots

What Happens When $\rho = 1$?

- Consider the process

$$y_t = y_{t-1} + \epsilon_t \tag{22}$$

where $\text{Var}(\epsilon_t) = \sigma^2$ for all t and which began with the observation y_0.

- We can apply repeated substitution to this series to get

$$y_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + ... + y_0 \tag{23}$$

- This implies that

$$\text{Var}(y_t) = \sigma^2 t \tag{24}$$

$$\text{Cov}(y_t, y_{t-k}) = \text{Cov}(\epsilon_t + \epsilon_{t-1} + ..., \epsilon_{t-k} + \epsilon_{t-k-1} + ...) \tag{25}$$

$$= \sigma^2 (t - k) \tag{26}$$

- This series is not covariance stationary (the covariances depend on t) and it's not ergodic (covariances with far past observations don't go to zero).
- If we consider a process of form $y_t = \alpha + y_{t-1} + \epsilon_t$, then it's not covariance stationary, not ergodic and $\mathbb{E}\, y_t \to \infty$ as $t \to \infty$.

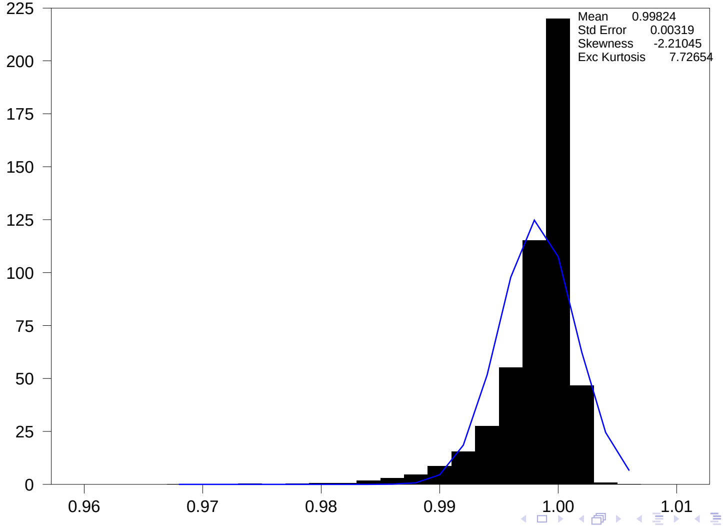
Asymptotics with Unit Root Processes?

- We have derived that the OLS to estimator applied to the $AR(1)$ process $y_t = \rho y_{t-1} + \epsilon_t$ has the property that $\sqrt{T}\left(\hat{\rho} - \rho\right) \xrightarrow{d} N\left(0, 1 - \rho^2\right)$.
- You might be tempted to think that we can use the logic underlying this argument to prove that variance of this asymptotic distribution goes to zero when $\rho = 1$, i.e. that the distribution collapses on the true value ρ. It turns out that this is indeed the case. However, you cannot use any of the previous arguments to prove this.
- Indeed, none of the previous arguments proving asymptotic normality hold because the assumptions underlying the Ergodic Theorem or the MDS version of the CLT do not hold in this case. The y_t series
  - Is not covariance stationary (variances increase as t gets larger).
  - Is not ergodic (the covariance with long-past observations does not go to zero).
  - Does not tend towards a finite variance
- Similarly, the previous arguments do not apply to any $AR(k)$ series in which one is a root of the polynomial $1 - \rho_1 L - .... - \rho_k L^k$. Series of this sort are referred to as **unit root processes**.

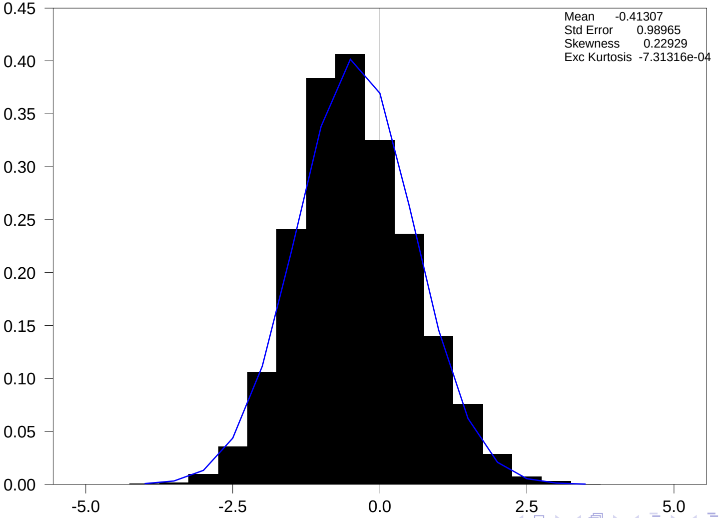
Non-Normal Distributions When $\rho = 1$ With No Drift

- For the process for which $y_t = y_{t-1} + \epsilon_t$ (i.e. a unit root with no intercept or drift) then the asymptotic distribution of the OLS estimator depends upon the regression specificiation:
  1. If no intercept is included and y_t is regressed on y_{t-1}, the distribution of $\hat{\rho}$ is non-Normal and skewed with most of the estimates below one (see next page) and the distribution of the t statistic testing $H_0 : \rho = 1$ is nonstandard.
  2. If an intercept is included in the regression, then the skewness and downward bias are far more serious (see the page after next). The critical value for rejecting the null of $\rho = 1$ at the 5% level changes from -1.95 with no intercept to -2.86 in large samples.
  3. The usual test procedures for these cases involve using the critical values derived by Dickey and Fuller (1976). While an analytical asympototic distribution exists, people usually use values from finite-sample distributions obtained by Monte Carlo (computer simulation) methods.

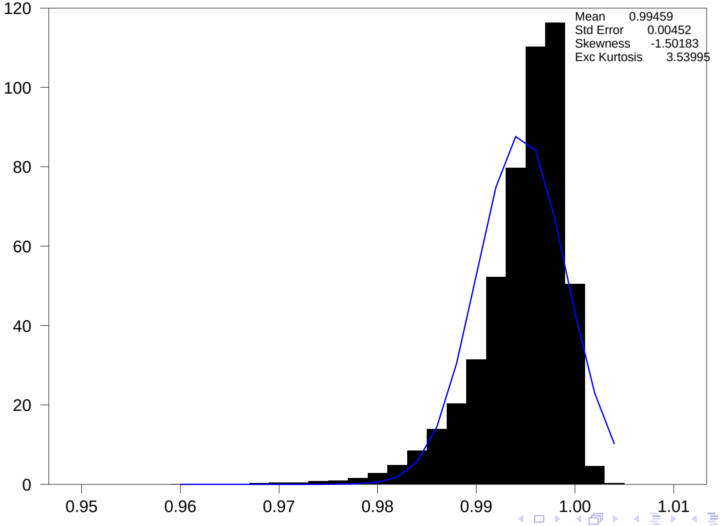
Distribution of $\hat{\rho}$ Under Unit Root (No Drift): No Constant in Regression ($T = 1000$)



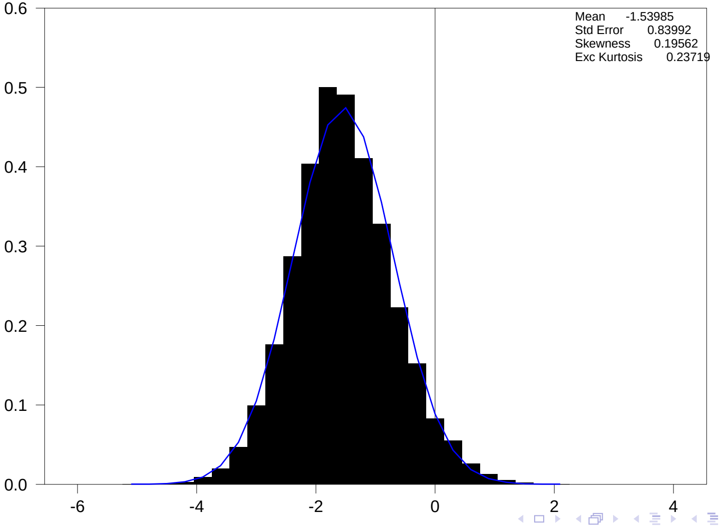
Distribution of t test of $H_0 : \rho = 1$ Under Unit Root (No Drift): No Constant in Regression ($T = 1000$)



Distribution of $\hat{\rho}$ Under Unit Root (No Drift): Constant in Regression ($T = 1000$)



Distribution of t test of $H_0 : \rho = 1$ Under Unit Root (No Drift): Constant in Regression ($T = 1000$)



Testing $\rho = 1$ for a Unit Root With Drift

- Most macroeconomic time series grow over time, so they are clearly not described by $y_t = \rho y_{t-1} + \epsilon_t$, which doesn't impart any trend to the series.
- So, for macroeconomic series, the question of whether the series has a unit root is usually phrased as whether the series has a deterministic time trend

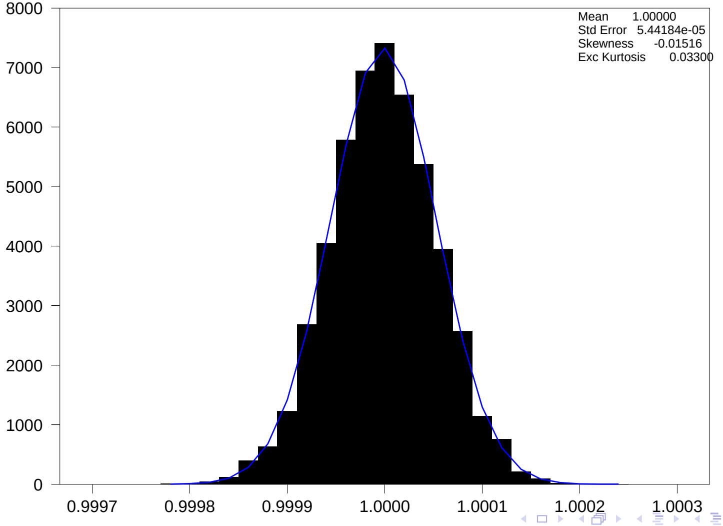
$$y_t = \alpha + \delta t + \rho y_{t-1} + \epsilon_t \tag{27}$$

where $0 < \rho < 1$ or whether the series has a stochastic trend, meaning the series is a unit root with drift

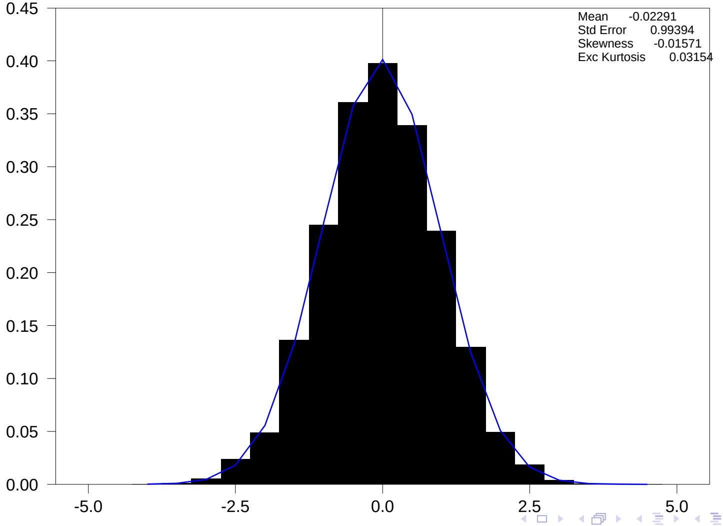
$$y_t = \delta + y_{t-1} + \epsilon_t \tag{28}$$

- If the true process is of the form (28), then the asymptotic distribution of the OLS estimator depends upon the regression specificiation:
  1. If we regress y_t on a constant and y_{t-1}, then $\hat{\rho}$ converges to a normal distribution and the usual t and F tests can be compared with their usual critical values. (See next two pages)
  2. If we estimate a "nesting" specification, regress y_t on a constant and y_{t-1} and a time trend, then we again get a skewed non-normal distribution. (See the following two pages).

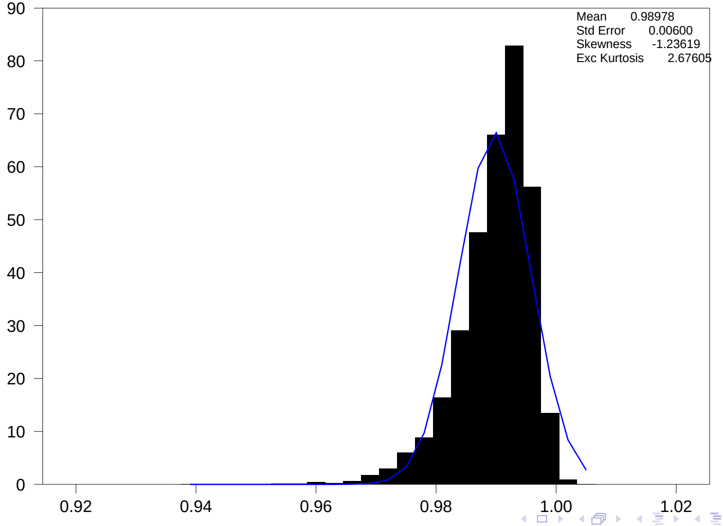
Distribution of $\hat{\rho}$ Under Unit Root With Drift: Constant But No Trend in Regression ($T = 1000$)



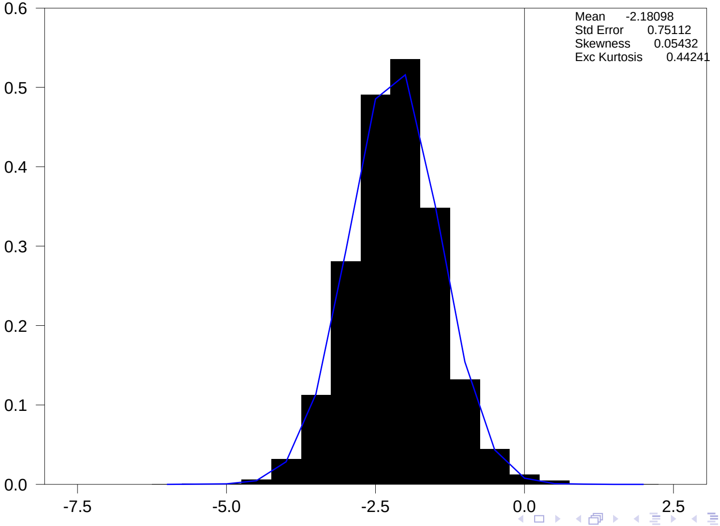
Distribution of t test of $H_0 : \rho = 1$ Under Unit Root With Drift: Constant But No Trend in Regression ($T = 1000$)



Distribution of $\hat{\rho}$ Under Unit Root With Drift: Constant and Trend in Regression ($T = 1000$)



Distribution of t test of $H_0 : \rho = 1$ Under Unit Root With Drift: Constant and Trend in Regression ($T = 1000$)



Computing Critical Values

- Note that you can easily use computer simulations to calculate critical values. In the case of testing for a unit root against a deterministic time trend, you could consult tables like those at the back of Hamilton's textbook to find out that the 5% critical value for rejecting the unit root is -3.41. Or you could do Monte Carlo simulation, save the results and calculate the fractiles. See below:

```
Statistics on Series TSTATS
Observations                   10000
Sample Mean                -2.180975      Variance               0.564183
Standard Error              0.751121      of Sample Mean         0.007511
t-Statistic (Mean=0) -290.362487      Signif Level           0.000000
Skewness                    0.054322      Signif Level (Sk=0) 0.026598
Kurtosis (excess)           0.442414      Signif Level (Ku=0) 0.000000
Jarque-Bera                86.472285      Signif Level (JB=0) 0.000000

Minimum                    -5.707974      Maximum                 1.524200
01-%ile                    -3.968187      99-%ile                -0.327494
05-%ile                    -3.416907      95-%ile                -0.938125
10-%ile                    -3.124171      90-%ile                -1.261622
25-%ile                    -2.663260      75-%ile                -1.705408
Median                     -2.176619
```

Unit Root Testing for $AR(k)$ Processes

- For the $AR(k)$ model $y_t = \alpha + \rho_1 y_{t-1} + ... + \rho_k y_{t-k} + \epsilon_t$, the series is strictly stationary and ergodic if the roots of the polynomial $1 - \rho_1 L - .... - \rho_k L^k$ are all less than one in absolute value.
- One is a root of this polyomial if $1 - \rho_1 - .... - \rho_k = 0 \Rightarrow \sum_{i=1}^{k} \rho_i = 1$.
- Note now that we can re-write an $AR(2)$ process as follows

$$\begin{aligned} y_t &= \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \epsilon_t && (29) \\ &= \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-1} - \rho_2 y_{t-1} + \rho_2 y_{t-2} + \epsilon_t && (30) \\ &= \alpha + (\rho_1 + \rho_2)\, y_{t-1} + \gamma \Delta y_{t-1} + \epsilon_t && (31) \end{aligned}$$

- So, $AR(k)$ series have a representation of the following form

$$y_t = \alpha + \gamma_1 \Delta y_{t-1} + ... + \gamma_{k-1} \Delta y_{t-k+1} + \left(\sum_{i=1}^{k} \rho_i \right) y_{t-1} + \epsilon_t \qquad (32)$$

- When testing for a unit root in an $AR(k)$ process, we can use the same Dickey-Fuller critical values for testing $\rho = 1$ in the augmented regression

$$y_t = \alpha + \gamma_1 \Delta y_{t-1} + ... + \gamma_{k-1} \Delta y_{t-k+1} + \rho y_{t-1} + \epsilon_t \qquad (33)$$