Financial Econometrics

Times Series Econometrics

Tim C.C. Hung 洪志清

March 21^{st}, 2022

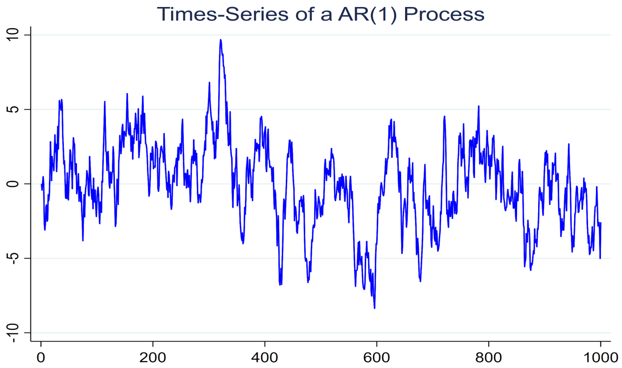
Data Type

- Different types of Data:
  - Cross-Sectional Data
    e.g. All monthly stock returns on TWSE in March
  - Times-Series Data
    e.g. Taiwan GDP from 1945 to present
  - Panel Data v.s. Pooled Cross-sectional Data
- Frequencies of Times-Series Data
  - Low frequency data: yearly, quarterly, monthly data
  - High frequency data: weekly, daily, intra-day data

Times-Series

- What's special about times-series data?
- There may be certain trends in times-series data.
  e.g. deterministic trend, seasonality, serial correlation, $\cdots$

- Example: an AR(1) process: $y_t = \rho y_{t-1} + \epsilon_t,\ \rho = 0.9, \epsilon_t \sim WN(0, \sigma^2)$



Times-Series Operators

- Lag operator: *L*

  $Ly_t = y_{t-1}$

- More about Lag operators:
  - $L^k y_t = y_{t-k}$
  - $L^k L^j y_t = L^k y_{t-j} = y_{t-j-k}$
  - $L^0 y_t = y_t$
  - $L^{-k} y_t = y_{t+k}$

- Difference operator: *D.* or Δ

  $D.y_t = \Delta y_t = y_t - y_{t-1} = y_t - Ly_t = (1-L)y_t$

  $\Delta^k y_t = y_t - y_{t-k} = (1-L^k)y_t$

Stationary

Definition (Weak Stationary)

A times series $\{\cdots, y_{t-3}, y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}, y_{t+3}, \cdots\}$ satisfy weak stationary **iff**

1. $E(y_t) = E(y_{t-k}) = \mu \quad \forall t, k$
2. $var(y_t) < \infty \quad \forall t$
3. $Cov(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma(k) \quad \forall t$

Definition (Strict Stationary)

For some k, $(t_1, t_2, \cdots, t_n)$, a times series $\{y_t\}$ satisfy strict stationary **iff**

$$(y_{t_1}, y_{t_2}, \cdots, y_{t_n})' \xrightarrow{d} (y_{t_1+k}, y_{t_2+k}, \cdots, y_{t_n+k})'$$

- Weak stationary is sometimes called covariance stationary.
- If Strict stationary $\&$ $E(y_t^2) < \infty$, $\rightarrow$ weak stationary.

White Noise

Definition (Stochastic Process)

A stochastic process $\{x_t\}$ is a sequence of random variables in time order.

Definition (White Noise)

A stochastic process $\{\epsilon_t\}$ is White Noise, or $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$, **iff**

1. $E(\epsilon_t) = 0 \quad \forall t$
2. $var(\epsilon_t) = \sigma_\epsilon^2 \quad \forall t$
3. $E(\epsilon_t \epsilon_{t-k}) = 0 \quad \forall t, k$

- You can clearly see that a white noise is stationary.
- A stronger condition is that $\epsilon_t, \epsilon_{t-k}$ are independent for $k \neq 0$
- If all these conditions hold and $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, then the process is Gaussian White Noise.

Autocovariance and Autocorrelation

National Taiwan University

Definition (Autocovariance)

The k^{th} autocovariance of a series $\{y_t\}$ is defined as:

$$\gamma(k) = cov(y_t, y_{t-k}) = E([y_t - E(y_t)][y_{t-k} - E(y_{t-k})])$$

- As you can tell, autocovariance is defined for weak stationary process.
- You can also see that $\gamma(0) = var(y_t)$

Definition (Autocorrelation)

The k^{th} autocorrelation of a covariance stationary process $\{y_t\}$ is defined as:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

Autocovariance and Autocorrelation

- As you can tell,

$$cov(y_t, y_{t-k}) = cov(y_{t-1}, y_{t-k-1}) == cov(y_{t-j}, y_{t-k-j}) \quad \forall j$$

autocovariance does not depend on t but only on the distance k.
i.e., autocovariance is defined for weak stationary process.

- You can also see that $\gamma(0) = var(y_t)$

- The correlation of y_t with y_{t+1} is a measure of how persistent a time series is, i.e., how strong is the tendency for a high observation today (this week, this month, this year, $\cdots$) to be followed by a high observation tomorrow (next week, next month, next year, $\cdots$).

Definition (Sample Autocovariance)

$$\hat{\gamma}(k) = \widehat{cov(y_t, y_{t-k})} = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

where $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$

Definition (Sample Autocorrelation)

$$\hat{\rho}(k) = \frac{\widehat{cov(y_t, y_{t-k})}}{\widehat{var(y_t)}} = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

- You will soon see that the estimated $\hat{\rho}(1)$ coefficient happens to be the estimated coefficient of an AR(1) process.

Deterministic Trend

- One of the simplest times-series model is a model of deterministic trend:

$$y_t = \beta_0 + \beta_1 TIME_t + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$$

- $TIME_t$ could be a categorical variable that takes the value of t, i.e. $TIME_1 = 1, TIME_2 = 2, \cdots$

- Thus, if we have a time-series of $\{y_t\}_{t=1}^{T}$, the model can be rewritten as:

$$y_t = \beta_0 + \beta_1 t + \epsilon_t$$

- Deterministic trend models could also be of higher moments,

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \epsilon_t$$

Deterministic Trend

- Even with higher moments, we can still estimate the coefficients using OLS, where the explanatory (independent) variables are $t, t^2, t^3, \cdots$.

  Remarks: Why is that the case?

- After estimation, we can obtain $\hat{y}_t$

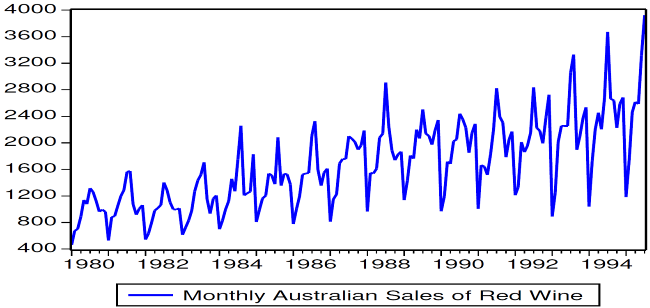
- The residual $\hat{\epsilon}_t = y_t - \hat{y}_t$
  is what we called a detrended data.

Seasonality

- Similar to deterministic trend, we may experience regular patterns within certain time window (cycle).

- For example, seasonality in GDP, Sales of product of a firm in different months, January effect in stock returns, Friday effect in attention (DellaVigna and Pollet (2009, *JF*)), etc.

Figure: Sales of wine in Australia (1980:1-1995:7) *from* 陳旭昇*(2013)*

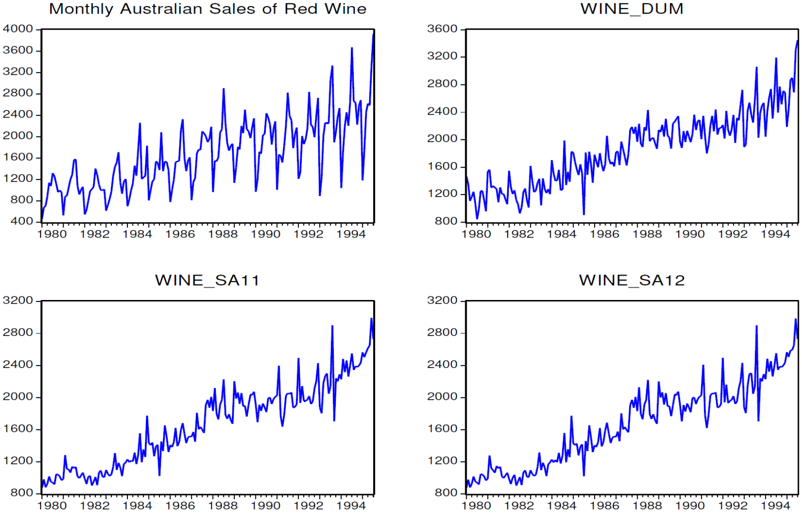


Seasonality

- Except the cases that our main focus is seasonality, we would want to exclude seasonality in the data.
- The most straightforward way is to use indicator (dummy) variables as regressors to exclude seasonality.

Season	D_1	D_2	D_3
Spring	1	0	0
Summer	0	1	0
Fall	0	0	1
Winter	0	0	0

Figure: Adjusted Sales of wine in Australia *from 陳旭昇(2013)*



AR(1) Process

Definition (AR(1) Model)

$$y_t = \alpha + \rho y_{t-1} + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$$

ρ is the first-order autoregressive coefficient.

- Recursively, y_t can be rewritten as:

$$\begin{aligned}
y_t &= \alpha + \rho y_{t-1} + \epsilon_t \\
&= \alpha + \rho(\alpha + \rho y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= \alpha(1+\rho) + \rho^2 y_{t-2} + \epsilon_t + \rho\epsilon_{t-1} \\
&= \alpha(1 + \rho + \rho^2 + \cdots) + (\epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + \cdots) + \lim_{k\to\infty} \rho^k y_{t-k} \\
&= \alpha\sum_{j=0}^{\infty}\rho^j + \sum_{j=0}^{\infty}\rho^j\epsilon_{t-j} + \lim_{k\to\infty}\rho^k y_{t-k}
\end{aligned}$$

AR(1) Process

National Taiwan University

- When we have $|\rho| < 1$:

$$
\begin{aligned}
y_t &= \alpha \sum_{j=0}^{\infty} \rho^j + \sum_{j=0}^{\infty} \rho^j \epsilon_{t-j} + \lim_{k\to\infty} \rho^k y_{t-k} \\
&= \underbrace{\frac{\alpha}{1-\rho}}_{\mu} + \sum_{j=0}^{\infty} \rho^j \epsilon_{t-j} + 0 \\
&= \mu + \sum_{j=0}^{\infty} \rho^j \epsilon_{t-j}
\end{aligned}
$$

AR(1) Process

- The Expectation:

$$E(y_t) = E(y_{t-k}) = \mu$$

- The Variance:

$$\begin{aligned}
\gamma(0) &= var(y_t) \\
&= var(\sum_{j=0}^{\infty} \rho^j \epsilon_{t-j}) \\
&= var(\epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + \cdots) \\
&= \sigma_\epsilon^2(1 + \rho^2 + \rho^4 + \cdots) \\
&= \frac{\sigma_\epsilon^2}{1-\rho^2} < \infty \quad (\text{for some } |\rho| < 1)
\end{aligned}$$

AR(1) Process

- The Autocovariance:

$$
\begin{aligned}
\gamma(j) &= cov(y_t, y_{t-j}) \\
&= E[(y_t - \mu)(y_{t-j} - \mu)] \\
&= E[(\epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2}\cdots)(\epsilon_{t-j} + \rho\epsilon_{t-1-j} + \rho^2\epsilon_{t-2-j}\cdots)] \\
&= \rho^j E[\epsilon_{t-j}\epsilon_{t-j}] + \rho^{j+2} E[\epsilon_{t-j-1}\epsilon_{t-j-1}] + \cdots \\
&= \sigma_\epsilon^2 \rho^j [1 + \rho^2 + \rho^4 + \cdots] \\
&= \frac{\sigma_\epsilon^2 \rho^j}{1-\rho^2} < \infty \quad \text{(for some } |\rho| < 1)
\end{aligned}
$$

- The Autocorrelation:

$$
\rho(j) = \frac{\sigma_\epsilon^2 \rho^j}{1-\rho^2} \Big/ \frac{\sigma_\epsilon^2}{1-\rho^2} = \rho^j
$$

AR(1) Process

National Taiwan University

- As we can see, when $|\rho| < 1$, an AR(1) process has an invariant mean, finite variance, and autocovariance independent to t.

Corollary (Stationary AR(1) Model)

For an AR(1) process: $y_t = \alpha + \rho y_{t-1} + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$

The stationary condition is: $|\rho| < 1$

AR(1) Process

National Taiwan University

- To stay simple, we can detrend y_t to exclude the intercept!
- For some $y_t = \alpha + \rho y_{t-1} + \epsilon_t$
- Let's minus μ on both sides:

$$y_t - \mu = \alpha - \mu + \rho y_{t-1} + \epsilon_t$$

- Let's add and minus a $\rho\mu$ term on the RHS:

$$\begin{aligned} x_t \equiv y_t - \mu &= \alpha - \mu + \rho\mu + \rho(y_{t-1} - \mu) + \epsilon_t \\ &= \alpha - (1-\rho)\mu + \rho(y_{t-1} - \mu) + \epsilon_t \\ &= \rho(y_{t-1} - \mu) + \epsilon_t \\ &= \rho x_{t-1} + \epsilon_t \end{aligned}$$

- x_t is an AR(1) process with no intercept and 0 mean.

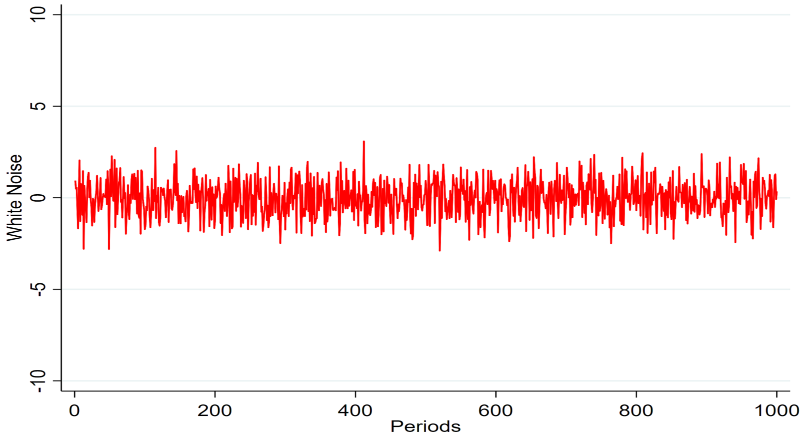
AR(1) Process

- The variance-covariance matrix for x_t is:

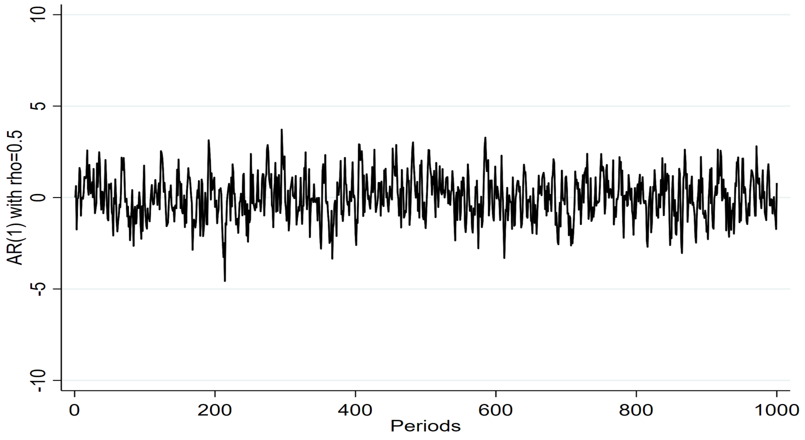
$$\frac{\sigma_\epsilon^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix} \equiv \sigma_\epsilon^2 \Omega$$

AR(1) Simulations: White Noise

國立臺灣大學 National Taiwan University

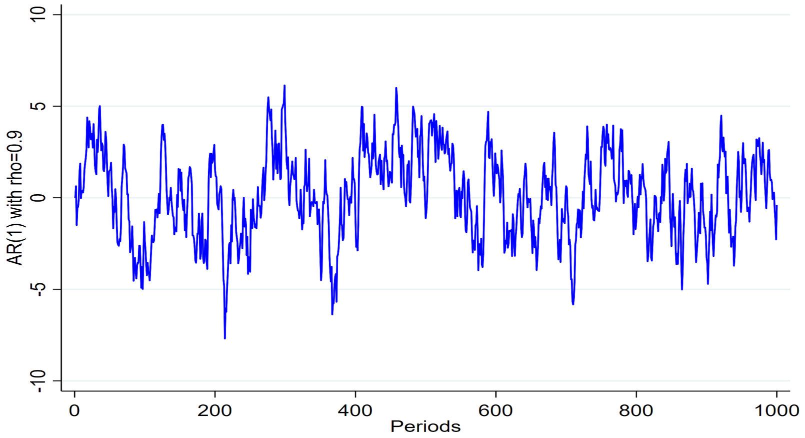


AR(1) Simulations: $\rho = 0.5$



AR(1) Simulations: $\rho = 0.9$

National Taiwan University



AR(p) Process

National Taiwan University

Definition (AR(p) Model)

$$y_t = \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_p y_{t-p} + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$$

the model is called the p^{th}-order autoregressive model.

- Using the lag operator, we can rewrite y_t as:

$$\rho(L) y_t = \alpha + \epsilon_t,$$

  where

$$\rho(L) = 1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p$$

AR(p) Process

Definition (AR(p) Stationary Condition)

For a time series y_t

$$y_t = \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_p y_{t-p} + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$$

The stationary condition is:

$$\rho(z) = 1 - \rho_1 z - \rho_2 z^2 - \cdots - \rho_p z^p = 0$$

the norms of all the roots, $||z_i||_{i=1}^p$, for the above equation > 1.

Alternatively, the roots all fall outside the unit circle.

- Example 1: AR(1), $y_t = \beta y_{t-1} + u_t$
  - The solution to: $1 - \beta z = 0$ is $z = \frac{1}{\beta}$
  - The AR(1) stationary condition would be: $|\frac{1}{\beta}| > 1$, or $|\beta| < 1$

- Example 2: AR(2), $y_t = 0.5 + 0.3y_{t-1} + 0.4y_{t-2} + e_t$
  - The solution to: $1 - 0.3z - 0.4z^2 = 0$ is $z = \frac{5}{4}$ or -2
  - The AR(2) stationary condition would be: $|z_1| = |\frac{5}{4}| > 1$ and $|z_2| = |-2| > 1$

- Example 3: AR(2), $y_t = 0.5 + 0.3y_{t-1} - 0.4y_{t-2} + e_t$
  - The solution to: $1 - 0.3z + 0.4z^2 = 0$ is $z = \frac{3}{8} \pm \frac{\sqrt{151}}{8}i$
  - The AR(2) stationary condition would be: $||z_1|| = ||z_2|| = \sqrt{(\frac{3}{8})^2 + (\frac{\sqrt{151}}{8})^2} = \sqrt{2.5} > 1$

AR(p) Process

- If the norm of any root of $z < 1$, the AR(p) process is explosive.
- If the norm of any root of $z = 1$, the AR(p) process has an unit root.

Definition (Sufficient Condition for AR(p) Stationary)

$$\sum_{i=1}^{p} |\rho_i| < 1$$

If alternatively,

$$\sum_{i=1}^{p} \rho_i = 1$$

then the series has at least one unit root.

AR(p) Process: Choice of p

Definition (Akaike Information Criterion (AIC))

$$AIC(p) = log(\frac{UV(p)}{T}) + (p+1)\frac{2}{T}$$

where $UV(p)$ is the unexplained variance (sum of squared residuals),

$$UV(p) = \sum_t \hat{\epsilon_t}^2,$$

$$\hat{\epsilon_t}^2 = y_t - \hat{y}_t = y_t - (\hat{\alpha} + \hat{\rho}_1 y_{t-1} + \cdots + \hat{\rho}_p y_{t-p})$$

Definition (Bayesian Information Criterion (BIC))

$$BIC(p) = log(\frac{UV(p)}{T}) + (p+1)\frac{logT}{T}$$

- Pick the p with the lowest *IC*s!

MA(1) and MA(q) Process

National Taiwan University

Definition (Moving Average Models)

If a stochastic process $\{y_t\}$ is a weighted average of current and past shocks, we call it a moving average model.

MA(1), first-order moving average model:

$$y_t = \epsilon_t + \theta \epsilon_{t-1},$$

$$\epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$$

MA(q), q^{th}-order moving average model:

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q},$$

MA(q) Process: Stationary

National Taiwan University

- Now, let $q \to \infty$: $y_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j}$
- For a MA model to be well defined, we need its partial sum (i.e. $\sum_{j=0}^{n} \theta_j \epsilon_{t-j}$) to converge in mean square to some random variable.

Definition (MA(∞) Stationary Condition)

For $y_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j}, \quad \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$

The stationary condition is:

$$\sum_{j=0}^{\infty} |\theta_j| < \infty$$

This also indicates that $\{\theta_j\}$ is absolutely summable.

- Quick Takeaway: for finite q, MA(q) is definitely stationary.

MA(q) Process: Invertibility

National Taiwan University

- MA(q) can be rewritten as: $y_t = \theta(L)\epsilon_t$
- If the roots of $\theta(z) = 0$ all have norms larger than 1 (all fall outside the unit circle), MA(q) can be further rewritten as:

$$\frac{1}{\theta(L)} y_t = \epsilon_t$$

- This property is called the invertiblility of MA(q) series.

National Taiwan University

Definition (Absolute Summable Inverses of Lag Polynomials)

For some

$$\beta(L) = 1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_p L^p$$

If the roots for $\beta(z) = 0$ all fall outside the unit circle, then

$$\beta(L)^{-1} = \phi(L) = \phi_0 + \phi_1 L + \phi_2 L + \cdots$$

where

$$\sum_{j=0}^{\infty} |\phi_j| < \infty$$

MA(q) Process: Invertibility

國立臺灣大學 National Taiwan University

- For example: a MA(1) process: $x_t = \epsilon_t + \theta\epsilon_{t-1}, \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$

$$\begin{aligned} x_t &= \epsilon_t + \theta\epsilon_{t-1} \\ &= \epsilon_t + \theta(x_{t-1} - \theta\epsilon_{t-2}) \\ &= \epsilon_t + \theta x_{t-1} - \theta^2(x_{t-2} - \theta\epsilon_{t-3}) \\ &= \epsilon_t + \theta x_{t-1} - \theta^2 x_{t-2} + \theta^3 x_{t-3} - \theta^4 x_{t-4} + \cdots \end{aligned}$$

- Thus, we can rewrite ϵ_t as a series of $\{x_t\}$

$$\epsilon_t = x_t - \theta x_{t-1} + \theta^2 x_{t-2} - \theta^3 x_{t-3} + \theta^4 x_{t-4} - \cdots$$

- Before we move on, assume for stationarity, can you derive the mean, variance, autocovariance, and autocorrelation of a AR(1) model and a MA(1) model *directly*?

ARMA model

National Taiwan University

- Now, let's combine AR models with MA models.
- Let's consider a ARMA(p,q) model:

$$y_t = \rho_1 y_{t-1} + \cdots + \rho_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

- Let

$$\rho(L) = 1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p$$
$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$$

- The ARMA(p,q) model can be rewritten as:

$$\rho(L) y_t = \theta(L) \epsilon_t$$

Definition (ARMA(p,q) Stationary Condition)

Given that

$$\rho(L)y_t = \theta(L)\epsilon_t$$

If the root of $\rho(z) = 0$ all fall outside the unit circle, y_t is stationary.

ARMA model

National Taiwan University

- If y_t is stationary, we can rewrite ARMA(p,q) as MA(∞):

$$
\begin{aligned}
y_t &= \underbrace{\rho(L)^{-1}\theta(L)}_{\psi(L)}\epsilon_t \\
&= [\frac{1+\theta_1 L+\theta_2 L^2+\cdots+\theta_q L^q}{1-\rho_1 L-\rho_2 L^2-\cdots-\rho_p L^p}]\epsilon_t \\
&= \psi(L)\epsilon_t \\
&= \epsilon_t+\psi_1\epsilon_{t-1}+\psi_2\epsilon_{t-2}+\cdots \\
&= MA(\infty)
\end{aligned}
$$

Remark: $\psi_0 = 1$,

$$\sum_{j=0}^{\infty}|\psi_j| < \infty$$

Definition (ARMA(p,q) Model Invertibility)

Given that

$$\rho(L)y_t = \theta(L)\epsilon_t$$

If the root of $\theta(z) = 0$ all fall outside the unit circle, y_t is invertible.

- If the root of $\theta(z) = 0$ all fall outside the unit circle,

$$\underbrace{\theta(L)^{-1}\rho(L)}_{b(L)} y_t = \epsilon_t$$

ARMA model

- We now have:

$$b(L)y_t = \epsilon_t$$

where

$$b(L) = 1 - b_1L - b_2L^2 - \cdots$$

- Therefore, ARMA(p,q) can be rewritten as a AR(∞) model:

$$y_t = \epsilon_t + b_1y_{t-1} + b_2y_{t-2} + \cdots = AR(\infty)$$

- Estimation: Sadly, we won't be able to introduce estimation of MA models and ARMA models without non-linear estimation methods. Please refer to 陳旭昇(2013):Chapter 4 for more details.

- Back to a normal AR(1) process.

$$y_t = \alpha + \rho y_{t-1} + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$$

- The stationary condition for y_t is $|\rho| < 1$
- We also know that if $|\rho| > 1$, y_t is an explosive series.
- What if, $\rho = 1$?

- When $\rho = 1$,

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

- It is called a random walk model with a drift term.

Unit Root: Random Walk

- Without a drift term,

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

- It is a simple random walk model. We also have:

$$E_t[y_t] = y_{t-1}$$

$\rightarrow$ the best predictor for the next period is the current value.

- Assume we begin from y_0. By iterations,
  - Without drift term:

$$y_t = \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1 + y_0$$

  - With a drift term:

$$y_t = \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1 + y_0 + \alpha t$$

- The αt term is the deterministic trend we have introduced earlier.
  If $\alpha > 0$, the times series gradually increase along with time.
  If $\alpha < 0$, the times series gradually decrease along with time.

- Other than the deterministic trend (固定趨勢), there is another type of trend called the the *stochastic trend (隨機趨勢)*.

Unit Root: Deterministic Trend

- Given

$$y_t = \alpha + \beta t + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} WN(0, \sigma_\epsilon^2)$$

- We can see that:

$$E[y_t] = \alpha + \beta t$$
$$E[y_{t+s}] = \alpha + \beta(t + s)$$

- $E[y_t] \neq E[y_{t+s}]$, suggesting that y_t is not stationary. But we could regress y_t on t and obtain the residual:

$$\hat{\epsilon}_t = y_t - \hat{y}_t = y_t - \hat{\alpha} - \hat{\beta} t$$

- For any time series like y_t for which we could restore stationarity after detrending the deterministic trend, we call it trend stationary (TS).

Unit Root: Stochastic Trend

- Stochastic trends, however, are more subtle, harder to detect, and could have more series impact to time-series econometrics.
- A stochastic trend is the prolong effect of the stochastic component. Imagine if the ϵ component does not die out and has permanent effect.
- For example, the AR(1) model with $\rho = 1$. $y_t = \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1$. Any huge ϵ_j could last till period t even if $t >> j$.

Definition (Unit Root and Stochastic Trend)

Given a AR(p) process:

$$\rho(L)y_t = \alpha + \epsilon_t$$

If the polynomial $\rho(z) = 1 - \rho_1 z - \rho_2 z^2 - \cdots - \rho_p z^p = 0$ has a root $= 1$.
This AR(p) process possesses a stochastic trend.
The AR(p) process is also said to have a unit root (單根).

- There are three types of serious problems caused by stochastic trends.

 1. The estimated autocorrelation coefficient has a small-sample downward bias.
 2. The t-statistic for the autocorrelation coefficient is not distributed normally.
 3. Spurious regression 虛假迴歸.

Unit Root: Small-Sample Bias

National Taiwan University

- Take AR(1) for example, assume the DGP is:

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

- Assume we don't know the true DGP and estimate the model as a AR(1) process, (*which suppose to be right!*)

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

- While the true $\beta_1 = 1$, $\hat{\beta}_1$ has a downward bias:

$$\text{Bias} = E[\hat{\beta}_1] - \beta_1 \approx -\frac{5.3}{T}$$

- The smaller T, the larger bias.

Unit Root: Non-Normal t

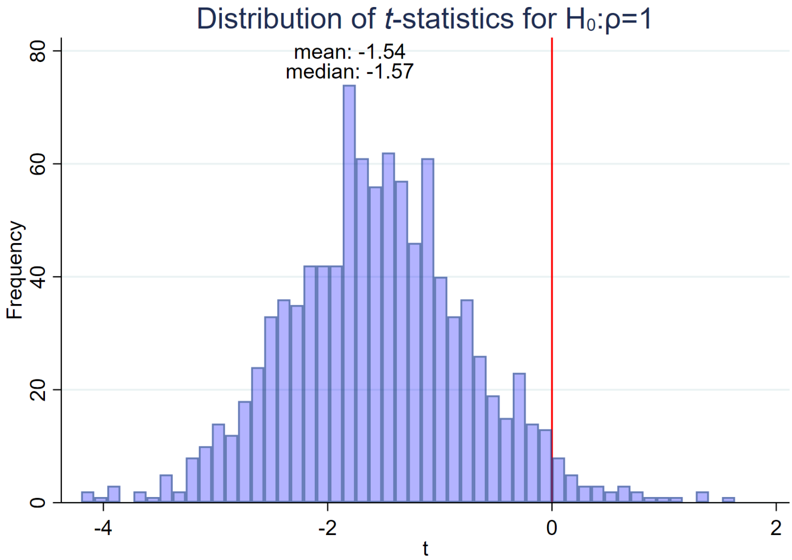
- Following the previous example.

$$t = \frac{\hat{\beta}_1 - \beta_1}{SE[\hat{\beta}_1]} \xrightarrow{d} N(0,1)$$

- However, when $\beta = 1$

$$t = \frac{\hat{\beta}_1 - \beta_1}{SE[\hat{\beta}_1]} = \frac{\hat{\beta}_1 - 1}{SE[\hat{\beta}_1]}$$

- The limit distribution is not standard normal!



Distribution of t-statistics for H0:ρ=1
mean: -1.54
median: -1.57
Frequency
0
20
40
60
80
-4
-2
0
2
t

Unit Root: Spurious Regression

National Taiwan University

- The idea of spurious regression is brought up by Granger and Newbold (1974).

- In general, for two independent and stationary x_t and y_t, if we run the following regression:

$$y_t = a_0 + a_1 x_t + u_t$$

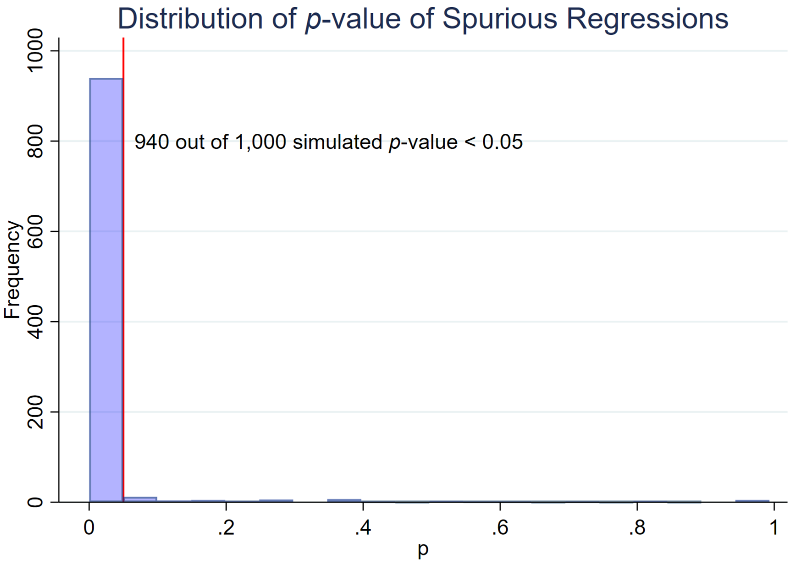
  We should expect for
  1. Insignificant $\hat{a}_1$
  2. Very low R^2

- However, if both x_t and y_t have stochastic trends, the probability of rejecting the null of $a_1 = 0$ is stunningly high! And we also have very high R^2s. This supposedly none-existing association is called a spurious regression.

- The following figures are the distributions of AR(1) coefficients and R^2 of 1,000 simulations.

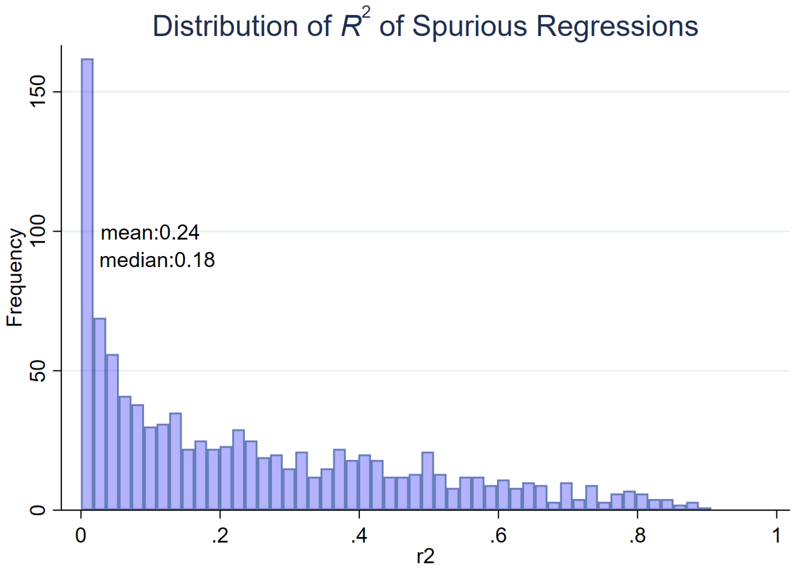
Unit Root: Bias and Non-Normal t

National Taiwan University



Unit Root: Bias and Non-Normal t

National Taiwan University



Unit Root Test

- The most common unit-root test is the Augmented Dickey-Fuller Test
- Consider a AR(k) model:

$$\rho(L)y_t = \mu + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$$

where $\rho(L) = 1 - \rho_1 L - \cdots - \rho_k L^k$

Definition (Dickey-Fuller Reparameterization)

Let $p = k - 1$, we can reparameterize $\rho(L)$ to get

$$\rho(L) = (1 - L) - \alpha_0 L - \alpha_1(L - L^2) - \cdots - \alpha_p(L^p - L^{p+1})$$

where

$$\alpha_0 = -1 + \sum_{j=1}^{k} \psi_j$$

$$\alpha_i = -\sum_{j=i+1}^{k} \psi_j, \quad for\ i = 1, 2, \cdots, p$$

Definition (Dickey-Fuller Reparameterization (Cont.))

We can rewrite the previous AR(k) process as:

$$\Delta y_t = \mu + \alpha_0 y_{t-1} + \alpha_1 \Delta y_{t-1} + \cdots + \alpha_p \Delta y_{t-p} + \epsilon_t$$

Unit Root Test

National Taiwan University

- Reparameterization Example: AR(3) model

$$y_t = \mu + \psi_1 y_{t-1} + \psi_2 y_{t-2} + \psi_3 y_{t-3} + \epsilon_t$$

- Now, let's reparameterize y_t:

$$\begin{aligned} y_t &= \mu + \psi_1 y_{t-1} - (\psi_2 + \psi_3)(y_{t-1} - y_{t-2}) - \psi_3(y_{t-2} - y_{t-3}) + \epsilon_t + (\psi_2 y_{t-1} + \psi_3 y_{t-1}) \\ &= \mu + (\psi_1 + \psi_2 + \psi_3) y_{t-1} - (\psi_2 + \psi_3)(y_{t-1} - y_{t-2}) - \psi_3(y_{t-2} - y_{t-3}) + \epsilon_t \\ &= \mu + (\psi_1 + \psi_2 + \psi_3) y_{t-1} - (\psi_2 + \psi_3)\Delta y_{t-1} - \psi_3 \Delta y_{t-2} + \epsilon_t \end{aligned}$$

- Let's subtract y_{t-1} on both RHS and LHS:

$$\Delta y_t = \mu + \underbrace{(-1 + \psi_1 + \psi_2 + \psi_3)}_{\alpha_0} y_{t-1} \underbrace{-(\psi_2 + \psi_3)}_{\alpha_1} \Delta y_{t-1} \underbrace{-\psi_3}_{\alpha_2} \Delta y_{t-2} + \epsilon_t$$

- A quick exercise together: AR(2): $y_t = 0.5 y_{t-1} + 0.5 y_{t-2} + \epsilon_t$

Unit Root Test

- Almost there!

- We know that if $\psi(z) = 0$ has a root that falls *on* the unit circle, i.e. $\psi(1) = 0$, y_t has an unit root.

$$\psi(1) = (1 - 1) - \alpha_0(1) - \alpha_1(1 - 1^2) - \cdots - \alpha_p(1^p - 1^{p+1}) = -\alpha_0$$

- Thus, the null hypotheses testing whether y_t has a unit root: $H_0 : \psi(1) = 0$, is equivalent to testing $H_0 : \alpha_0 = 0$!!!

Definition (Augmented Dickey-Fuller)

If the null hypotheses is that y_t has an unit root, and the alternative hypotheses is that y_t is stationary, consider the following regression:

$$\Delta y_t = \beta_0 + \delta y_{t-1} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + u_t$$

and we test: $H_0 : \delta = 0$ v.s. $H_1 : \delta < 0$

If the null hypotheses is that y_t has an unit root, and the alternative hypotheses is that y_t is *trend* stationary, consider the following regression:

$$\Delta y_t = \beta_0 + \alpha t + \delta y_{t-1} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + u_t$$

and we test: $H_0 : \delta = 0$ v.s. $H_1 : \delta < 0$

- The $\delta y_{t-1} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p}$ component is called the augmented part in ADF.

- We could choose the optimal p according to AIC or BIC.

$$ADF - t = \frac{\hat{\delta}}{SE(\hat{\delta})}$$

- Under the null: ADF-t is not standard normal. It has the following critical values:

ADF model	10%	5%	1%
with β_0	-2.57	-2.86	-3.43
with β_0 & deterministic trend	-3.12	-3.41	-3.96

Dealing with Unit Root

- Given the following unit root time series:

$$y_t = \alpha + y_{t-1} + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2)$$

- The most straightforward way to detrend the stochastic trend is to take first difference:

$$\Delta y_t = y_t - y_{t-1} = \alpha + \epsilon_t$$

- For time-series that retain stationary after taking first difference, we call it difference stationary.

- If y_t is stationary after first difference, we express it as: $y_t \sim I(1)$ (integrated of degree one).
- If y_t is already stationary, we express it as: $y_t \sim I(0)$.
- If y_t is stationary after d^{th} difference, we express it as: $y_t \sim I(d)$