Econometrics II TA Session #9

Ryo Sakamoto

Jan 12, 2021

5 Time Series Analysis

- 5.1 Introduction
  - 1. Stationarity
  - 2. Ergodicity
  - 3. Auto-covariance Function
  - 4. Auto-correlation Function
  - 5. Sample mean
  - 6. Sample Auto-covariance
  - 7. Correlogram
  - 8. Lag Operator
  - 9. Likelihood Function
- 5.2 Autoregressive Model
  - 1. AR(p) Model
  - 2. Stationarity
  - 3. Partial Autocorrelation Coefficient

1. Stationarity

Let $y_1, \cdots, y_T$ be time series data.

(a) **Weak Stationarity:**

$$\mathbb{E}(y_t) = \mu, \qquad \forall t \in \{1, \cdots, T\},$$
$$\mathbb{E}[(y_t - \mu)(y_{t-\tau} - \mu)] = \gamma(\tau), \qquad \forall t \in \{1, \cdots, T\} \text{ and } \tau = 0, 1, 2, \cdots.$$

- Both of the first and second moments do not depend on time t.
- The second moment depends on time difference τ, not time itself.

1. Stationarity

(b) **Strong Stationarity:**

- Let $f(y_{t_1}, y_{t_2}, \cdots, y_{t_r})$ be the joint distribution of $y_{t_1}, y_{t_2}, \cdots, y_{t_r}$.

$$f(y_{t_1}, y_{t_2}, \cdots, y_{t_r}) = f(y_{t_1+\tau}, y_{t_2+\tau}, \cdots, y_{t_r+\tau}), \quad \forall (t_1, \cdots, t_r) \in \mathbb{R}^r \text{ and } \forall \tau \in \mathbb{N}.$$

- e.g.) The probability that the weather of 12/1, 2, 3. 4 is (sunny, cloudy, rainy, sunny) is equal to the probability that the weather of 12/4, 5, 6, 7 is (sunny, cloudy, rainy, sunny).
- All the moments are same for all τ.

2. Ergodicity

- As time difference between two data is large, the two data become independent.
- $y_1, \cdots, y_T$ is said to be **ergodic in mean** when $\overline{y}$ converges in probability to $\mathbb{E}(y_t)$, i.e.,

$$\frac{1}{T}\sum_{t=1}^{T} y_t \xrightarrow{\mathbb{P}} \mathbb{E}(y_t), \quad \text{as} \quad T \to \infty.$$

- Roughly speaking, it is a time series version of the Law of Large Numbers.

3. Auto-covariance Function

- The auto-covariance of y_t and $y_{t-\tau}$ is

$$\mathbb{E}[(y_t - \mu)(y_{t-\tau} - \mu)] = \gamma(\tau), \quad \tau = 0, 1, 2, \cdots .$$

- It has a property of symmetricity:

$$\gamma(\tau) = \gamma(-\tau),$$

where

$$\gamma(-\tau) = \mathbb{E}[(y_t - \mu)(y_{t+\tau} - \mu)].$$

4. Auto-correlation Function

- The auto-correlation is given by

$$\rho(\tau) = \frac{\mathbb{E}[(y_t - \mu)(y_{t-\tau} - \mu)]}{\sqrt{Var(y_t)}\sqrt{Var(y_{t-\tau})}} = \frac{\gamma(\tau)}{\gamma(0)}.$$

- Note that for all $t \in \{1, \cdots, T\}$,

$$\begin{aligned} \gamma(0) &= \mathbb{E}[(y_t - \mu)(y_t - \mu)] \\ &= \mathbb{E}\left[(y_t - \mu)^2\right] \\ &= Var(y_t). \end{aligned}$$

5. Sample Mean

- The sample mean is give by

$$\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t.$$

- Note that $\widehat{\mu}$ is a consistent estimator of y_t if the sample has a property of ergodicity.

6. Sample Auto-covariance

- The sample auto-covariance is given by

$$\widehat{\gamma}(\tau) = \frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{\mu})(y_{t-\tau} - \widehat{\mu}).$$

- This is the empirical counterpart of the auto-covarince function:

$$\gamma(\tau) = \mathbb{E}[(y_t - \mu)(y_{t-\tau} - \mu)].$$

7. Correlogram

- The correlogram (sample auto-correlation) is given by

$$\widehat{\rho}(\tau) = \frac{\widehat{\gamma}(\tau)}{\widehat{\gamma}(0)} = \frac{\frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{\mu})(y_{t-\tau} - \widehat{\mu})}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{\mu})(y_t - \widehat{\mu})}.$$

- This is the empirical counterpart of the auto-correlation function:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \frac{\mathbb{E}[(y_t - \mu)(y_{t-\tau} - \mu)]}{\sqrt{Var(y_t)}\sqrt{Var(y_{t-\tau})}}.$$

8. Lag Operator

- The lag operator L^τ is used for taking a lag of τ periods:

$$L^\tau y_t = y_{t-\tau}, \quad \tau = 1, 2, \cdots .$$

9. Likelihood Function - Innovation Form

- Using the Bayes' rule, we have the joint density of $y_1, \cdots, y_T$ as follows

$$
\begin{aligned}
f(y_1, \cdots, y_T) &= f(y_T | y_{T-1}, \cdots, y_1) f(y_{T-1}, \cdots, y_1) \\
&= f(y_T | y_{T-1}, \cdots, y_1) f(y_{T-1} | y_{T-2}, \cdots, y_1) f(y_{T-2}, \cdots, y_1) \\
&\vdots \\
&= f(y_T | y_{T-1}, \cdots, y_1) f(y_{T-1} | y_{T-2}, \cdots, y_1) \cdots f(y_2 | y_1) f(y_1) \\
&= f(y_1) \prod_{t=2}^{T} f(y_t | y_{t-1}, \cdots, y_1).
\end{aligned}
$$

9. Likelihood Function - Innovation Form

- Therefore, the log-likelihood function is

$$\log f(y_1, \cdots, y_T) = \log f(y_1) + \sum_{t=2}^{T} \log f(y_t | y_{t-1}, \cdots, y_1).$$

- Under the normality assumption, $f(y_t | y_{t-1}, \cdots, 1)$ is given by the normal distribution:

$$y_t | y_{t-1}, \cdots, y_1 \ \sim \ \mathcal{N}_{\mathbb{R}}\Big(\mathbb{E}[y_t | y_{t-1}, \cdots, y_1], \ Var(y_t | y_{t-1}, \cdots, y_1)\Big).$$

1. AR(p) Model

- The AR(p) model is given by

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

- This expression can be rearranged as

$$\begin{aligned} y_t - \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} &= \epsilon_t \\ y_t - \phi_1 L y_t + \phi_2 L^2 y_t + \cdots + \phi_p L^p y_t &= \epsilon_t \\ (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) y_t &= \epsilon_t \\ \phi(L) y_t &= \epsilon_t. \end{aligned}$$

2. Stationarity

- Suppose that all the p solutions of x from $\phi(x) = 0$ are real numbers.
- Then, if the p solutions are greater than one, y_t is stationary.

- Suppose that the p solutions include imaginary numbers.
- Then, if the p solutions are outside unit circle, y_t is stationary.

3. Partial Autocorrelation Coefficient $\phi_{k,k}$

- The partial autocorrelation between y_t and y_{t-k}, denoted by $\phi_{k,k}$, is a measure of strength of the relationship between y_t and y_{t-k} after removing the influence of $y_{t-1}, \cdots, y_{t-k+1}$.
- In case of $k = 1$, there is no intermediate period between y_t and y_{t-1}.

  $\Rightarrow$ The partial autocorrelation coefficient is equivalent to the autocorrelation.

$$\phi_{1,1} = \rho(1) = \frac{\gamma(1)}{\gamma(0)}$$

3. Partial Autocorrelation Coefficient $\phi_{k,k}$

- In case of $k = 2$, to obtain the partial autocorrelation between y_t and y_{t-2} we need to remove the influence of y_{t-1}.

$$\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{2,1} \\ \phi_{2,2} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix}$$

- This matrix form corresponds to the following system of equations:

$$\phi_{2,1} + \rho(1)\phi_{2,2} = \rho(1),$$
$$\rho(1)\phi_{2,1} + \phi_{2,2} = \rho(2).$$

3. Partial Autocorrelation Coefficient $\phi_{k,k}$

- Solving the system, we have

$$\phi_{2,1} = \frac{\rho(1)[\rho(2) - 1]}{\rho(1)^2 - 1},$$
$$\phi_{2,2} = \frac{\rho(1)^2 - \rho(2)}{\rho(1)^2 - 1}.$$

3. Partial Autocorrelation Coefficient $\phi_{k,k}$

- Again, in case of $k = 2$,

$$\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{2,1} \\ \phi_{2,2} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix}$$

$$\Longleftrightarrow \begin{pmatrix} Corr(y_t, y_t) & Corr(y_t, y_{t-1}) \\ Corr(y_{t-1}, y_t) & Corr(y_{t-1}, y_{t-1}) \end{pmatrix} \begin{pmatrix} \phi_{2,1} \\ \phi_{2,2} \end{pmatrix} = \begin{pmatrix} Corr(y_t, y_{t-1}) \\ Corr(y_t . y_{t-2}) \end{pmatrix}$$

- Looking at this expression, we see that the partial autocorrelation coefficient $\phi_{2,2}$ is a measure of strength of the relationship between y_t and y_{t-2} after removing the inlfuence of y_{t-1}.

3. Partial Autocorrelation Coefficient $\phi_{k,k}$

- Generally, we can express

$$\begin{pmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \rho(k-1) \\ \rho(1) & 1 & \cdots & \rho(k-3) & \rho(k-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho(k-2) & \rho(k-3) & \cdots & 1 & \rho(1) \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \vdots \\ \phi_{k,k-1} \\ \phi_{k,k} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k-1) \\ \rho(k) \end{pmatrix}$$

- Using Cramer's rule, we obtain $\phi_{k,k}$:

$$\phi_{k,k} = \left(\frac{\begin{vmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \color{red}{\rho(1)} \\ \rho(1) & 1 & \cdots & \rho(k-3) & \color{red}{\rho(2)} \\ \vdots & \vdots & \ddots & \vdots & \color{red}{\vdots} \\ \rho(k-2) & \rho(k-3) & \cdots & 1 & \color{red}{\rho(k-1)} \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & \color{red}{\rho(k)} \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \rho(k-1) \\ \rho(1) & 1 & \cdots & \rho(k-3) & \rho(k-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho(k-2) & \rho(k-3) & \cdots & 1 & \rho(1) \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & 1 \end{vmatrix}} \right)$$

- The main difference between cross sectional and time series analysis is whether random sampling is possible or not.
  - In cross sectional analysis, we can randomly choose a sample.
    $\Rightarrow$ i's observation is independent of j's one.
    $\Rightarrow$ We can rely on the LLN, which requires the independence of observations each other.

  - In time series analysis, we can not take observations randomly.
    $\Rightarrow$ The onservation in 2021 is dependent on that in 2020.
    $\Rightarrow$ We can not rely on the LLN, so another rule is necessary to consider the asymptotic properties of the estimators.
    $\Rightarrow$ Ergodicity !

- The chapter 20 in Greene provides the definition of Ergodicity as follows:

Fefinition 20.3 Ergodicity

A storong stationary time-series process, $\{z_t\}_{t=-\infty}^{t=\infty}$, is ergodic if for any two bounded functions that map vectors in the a and b dimensional real vector spaces to real scalars, $f : \mathbb{R}^a \to \mathbb{R}$ and $g : \mathbb{R}^b \to \mathbb{R}$,

$$\lim_{k\to\infty} |\mathbb{E}[f(z_t, z_{t+1}, \cdots, z_{t+a-1})g(z_{t+k}, z_{t+k+1}, \cdots, z_{t+k+b-1})]|$$
$$=|\mathbb{E}[f(z_t, z_{t+1}, \cdots, z_{t+a-1})]| \times |\mathbb{E}[g(z_{t+k}, z_{t+k+1}, \cdots, z_{t+k+b-1})]|$$

- Remember that Two random variables X and Y are independent iff $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$.

- The definition states essentially that if events re separated far enough in time, then they are asymptotically independent.
- Then, we have the following:

Theorem 20.1 The Ergodic Theorem

If $\{z_t\}_{t=-\infty}^{t=\infty}$ is a time-series process that is strongly stationary and ergodic and $\mathbb{E}[|z_t|]$ is a finite constant, and if $\bar{z}_T = (1/T)\sum_{t=1}^{T} z_t$, then $\bar{z}_T \xrightarrow{a.s.} \mu$, where $\mu = \mathbb{E}[z_t]$. Note that the convergence is almost surely in probability or in mean square.

- What we have in the ergodic theorem is, for sums of dependent observations, a counterpart to the LLN which requires sums of independent observations.

- Theorem 20.1 is extended as follows.

Theorem 20.2 Ergodicity of Functions

If $\{z_t\}_{t=-\infty}^{t=\infty}$ is a time-series process that is strongly stationary and ergodic and if $y_t = f(\{z_t\})$ is a measurable function in the probability space that defines z_t, then y_t is also stationary and ergodic. Let $\{\mathbf{z_t}\}_{t=-\infty}^{t=\infty}$ define a $K \times 1$ vector valued stochastic process - each element of the vector is an ergodic and stationary series, and the characteristics of ergodicity and stationarity apply to the joint distribution of the elements of $\{\mathbf{z_t}\}_{t=-\infty}^{t=\infty}$. Then, the ergodic theorem applies to functions of $\{\mathbf{z_t}\}_{t=-\infty}^{t=\infty}$.

- Theorem 20.2 produces the results we need to characterize the least squares (and other) estimators.

- Consider the following model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y}, \boldsymbol{\epsilon} \in \mathbb{R}^T$, $\boldsymbol{X} \in \mathcal{M}_{T\times K}(\mathbb{R})$ and $\boldsymbol{\beta} \in \mathbb{R}^K$.

- Then the OLS estimator is given by

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Y}) \\ &= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{\epsilon}) \\ &= f(\boldsymbol{X}, \boldsymbol{\epsilon}),\end{aligned}$$

which implies that the estimator is a function of random vectors $\boldsymbol{X}$ and $\boldsymbol{\epsilon}$.