قسم الاحصاء وبحوث العمليات

المقرر: 336 إحص: تحليل السلاسل الزمنية

Tutorial set #2 solution

Question 1:

1- What is the difference between strict and weak stationarity? When we can say that weak stationarity leads to strict stationarity?

Strict stationarity:

If the set $(t_1, t_2, \dots, t_m)$ is subset of the time units, where $m = 1,2, \dots$ and if $k = \pm 1, \pm 2, \dots,$ then the stochastic process $\{Y_t\}$ is strictly stationary if the joint probability distribution function of the variables $\{Y_{t_1}, Y_{t_2}, \dots, Y_{t_m}\}$ is the same as the joint probability distribution function of the variables $\{Y_{t_1+k}, Y_{t_2+k}, \dots, Y_{t_m+k}\}$ for any set of time points $(t_1, t_2, \dots, t_m)$ and for any time lag k.

Weak stationarity:

the stochastic process $\{Y_t\}$ is weakly stationary if the second order moments exist, and satisfy:

a) The mean of the process μ_t is constant and do not depend on time t, that is:

$$\mu_t = E(Y_t) = \mu \quad ; t = 0, \pm 1, \pm 2, \dots$$

b) The variance of the process σ_t^2 is constant and do not depend on time t, that is:

$$\sigma_t^2 = V(Y_t) = \gamma(0) \quad ; t = 0, \pm 1, \pm 2, \dots$$

c) The covariance between any two variables depends only on the time lag between them. That is:

$$Cov(Y_t, Y_{t-k}) = \gamma(k)$$

And since the normal distribution is completely defined through its second order moments, so if the joint distribution of the variables $\{Y_{t_1}, Y_{t_2}, \dots, Y_{t_m}\}$ is the multivariate normal distribution then the weak stationarity leads to the strict stationarity.

2- An analyst has a time series data representing number of daily car accidents in a major road at Riyadh city. He applied the techniques of regression analysis to analyze the data set, by considering the dependent variable y_t as the number of daily car accidents, and the independent variable the time indices t=1,2,3,4,… representing days. So he applied the following simple linear regression model, $y_t = \beta_0 + \beta_1 t + \varepsilon_t$ comment on what he have done, do you think his analysis is always valid, discuss.

Solution:

Since the data set he has is a time series, then one would expect that there exists a serial correlation between the observations. Thus, I think what the analyst has done might be risky without checking that there is no serial correlation in the data. The results he would obtain – if there were actually correlation exist- would not be accurate, and the estimates of the coefficients β_i although unbiased but they do not have the minimum variance property, also the estimates of the standard errors of these coefficients underestimate the true standard errors of the estimated regression coefficient, thus the hypothesis tests and the confidence intervals for these coefficients that uses the t and F distributions are not applicable.

Question 2:

1- Assume the model:

$$Y_t = 1 + \varepsilon_t + \varepsilon_{t-1}$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with mean zero, and variance σ_ε^2. Find the autocorrelation function for the process $\{Y_t\}$, plot it and comment on the graph.

Solution:

$E(y_t) = 1$

$\gamma_k = Cov(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)]$

$\gamma_k = Cov(y_t, y_{t-k}) = E[(y_t - 1)(y_{t-k} - 1)]$

$= E(y_t y_{t-k}) - E(y_t) - E(y_{t-1}) + 1$

$= E(y_t y_{t-k}) - 1$

$= E[(1 + \varepsilon_t + \varepsilon_{t-1})(1 + \varepsilon_{t-k} + \varepsilon_{t-k-1})] - 1$

$$= E(1+\varepsilon_t+\varepsilon_{t-1}+\varepsilon_{t-k}+\varepsilon_t\varepsilon_{t-k}+\varepsilon_{t-1}\varepsilon_{t-k}+\varepsilon_{t-k-1}+\varepsilon_t\varepsilon_{t-k-1}+\varepsilon_{t-1}\varepsilon_{t-k-1})-1$$

for **k=0:**

$$\boldsymbol{\gamma_0} = E(1+\varepsilon_t+\varepsilon_{t-1}+\varepsilon_t+\varepsilon_t^2+\varepsilon_{t-1}\varepsilon_t+\varepsilon_{t-1}+\varepsilon_t\varepsilon_{t-1}+\varepsilon_{t-1}^2)-1$$
$$= 1+0+0+0+\sigma_\varepsilon^2+0+0+0+\sigma_\varepsilon^2-1 = \mathbf{2\sigma_\varepsilon^2}$$

for **k=1:**

$$\gamma_1 = 1+0+0+0+0+\sigma_\varepsilon^2+0+0+0-1 = \sigma_\varepsilon^2$$

for **k≥ 2:**

$$\gamma_k = 1+0+0+0+0+0+0+0+0-1 = 0$$

Thus the autocorrelation function for this process has the form:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} 1, & k=0 \\ 0.5, & k=1 \\ 0, & k\geq 2 \end{cases}$$

2- Find the autocorrelation function for the process, $Y_t = 1+\varepsilon_t - \varepsilon_{t-1}$, plot it and compare it with the ACF in part (1).

Solution:

$$E(y_t) = 1$$
$$\gamma_k = Cov(y_t, y_{t-k}) = E[(y_t-1)(y_{t-k}-1)]$$
$$= E(y_t y_{t-k}) - 1$$
$$= E[(1+\varepsilon_t-\varepsilon_{t-1})(1+\varepsilon_{t-k}-\varepsilon_{t-k-1})]-1$$
$$= E(1+\varepsilon_t-\varepsilon_{t-1}+\varepsilon_{t-k}+\varepsilon_t\varepsilon_{t-k}-\varepsilon_{t-1}\varepsilon_{t-k}-\varepsilon_{t-k-1}-\varepsilon_t\varepsilon_{t-k-1}+\varepsilon_{t-1}\varepsilon_{t-k-1})-1$$

for k=0:

$$\gamma_0 = 1+0-0+0+\sigma_\varepsilon^2-0-0-0+\sigma_\varepsilon^2-1 = \mathbf{2\sigma_\varepsilon^2}$$

for k=1:

$$\gamma_1 = 1+0-0+0+0-\sigma_\varepsilon^2-0-0+0-1 = -\sigma_\varepsilon^2$$

for k≥ 2:

$$\gamma_k = 1+0-0+0+0-0-0-0+0-1 = 0$$

Thus, the autocorrelation function for this process has the form:

$$\rho_k = \frac{\gamma(k)}{\gamma(0)} = \begin{Bmatrix} 1, & k = 0 \\ -0.5, & k = 1 \\ 0, & k \geq 2 \end{Bmatrix}$$

We notice from the autocorrelation functions in (1) and (2) that in process in (1) observations that are one time lag apart are positively correlated with $\rho_1 = 0.5$, and that observations more than one time lag apart are not correlated. While the process in part (2) observations that are one time lag apart are negatively correlated with $\rho_1 = -0.5$, and that observations more than one time lag apart are not correlated.

Question 3:

If the series $\{Y_t\}$ ca n be expressed in the form:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$$

where $\{\varepsilon_t\}$ as defined as in Q.2 $[\boldsymbol{\varepsilon_{t\sim} i.i.d(0, \sigma_\varepsilon^2)}]$

1- Find the expectation, the variance and the autocorrelation function of the series.

$E(y_t) = E(\beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t) = \beta_0 + \beta_1 t + \beta_2 t^2$

$V(y_t) = V(\beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t) = \sigma_\varepsilon^2$

Because $f(t) = \beta_0 + \beta_1 t + \beta_2 t^2$ is a deterministic function and not random

$\gamma_{s,t} = Cov[(\beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t), (\beta_0 + \beta_1 s + \beta_2 s^2 + \varepsilon_s)] = 0, \quad s \neq t$

Thus, the autocorrelation function for this process has the form:

$$\rho_k = \frac{\gamma(k)}{\gamma(0)} = \begin{Bmatrix} 1, & k = 0 \\ 0, & k > 0 \end{Bmatrix}$$

2- Does this series fulfill the stationarity conditions? Discuss.

Since the mean function is a function of time (t), then the process is **not stationary**, although the variance function is not a function of time and that the ACF depends only on time lag.

Question 4:

The following data represent the total profit (in million riyals) for a company:

Year	1430	1431	1432	1433	1434	1435	1436	1437
Profit y_t	3	2	2	4	5	6.1	4.4	5.5

1- Calculate the coefficients of the Sample autocorrelation function (SACF) $\boldsymbol{r_k}$, and plot it.
2- Calculate the standard errors for these estimates.

solution:

$$r_{\mathbf{k}} = \hat{\rho}(\mathbf{k}) = \frac{\sum_{t=1}^{\mathbf{n-k}} (y_t - \bar{y})(y_{t+\mathbf{k}} - \bar{y})}{\sum_{t=1}^{\mathbf{n}} (y_t - \bar{y})^2}$$

1- We can easily find:

$$n = 8 \ ; \qquad \bar{y} = \frac{32}{8} = 4 \quad ; \quad \sum_{t=1}^{8} (y_t - 4)^2 = 16.82$$

also, we can find the pairs $(y_t - 4)$:

Year	1430	1431	1432	1433	1434	1435	1436	1437
$(y_t - 4)$	-1	-2	-2	0	1	2.1	0.4	1.5

according to the definition of SACF r_k then:

$$r_1 = \hat{\rho}(1) = \frac{\sum_{t=1}^{7} (y_t - 4)(y_{t+1} - 4)}{16.82}$$

$$r_1 = \frac{1}{16.82}[(-1)(-2) + (-2)(-2) + (-2)(0) + (0)(1) + (1)(2.1) + (2.1)(0.4) + (0.4(1.5)] = 0.5672$$

also,

$$r_2 = \hat{\rho}(2) = \frac{\sum_{t=1}^{6}(y_t - 4)(y_{t+2} - 4)}{16.82}$$

$$r_2 = \frac{1}{16.82}[(-1)(-2) + (-2)(0) + (-2)(1) + (0)(2.1) + (1)(0.4) + (2.1)(1.5)] = 0.2111$$

$$r_3 = \hat{\rho}(3) = \frac{\sum_{t=1}^{5}(y_t - 4)(y_{t+3} - 4)}{16.82}$$

$$r_3 = \frac{1}{16.82}[(-1)(0) + (-2)(1) + (-2)(2.1) + (0)(0.4) + (1)(1.5)] = -0.2794$$

$$r_4 = \hat{\rho}(4) = \frac{\sum_{t=1}^{4}(y_t - 4)(y_{t+4} - 4)}{16.82}$$

$$r_4 = \frac{1}{16.82}[(-1)(1) + (-2)(2.1) + (-2)(0.4) + (0)(1.5)] = -0.3567$$

$$r_5 = \hat{\rho}(5) = \frac{\sum_{t=1}^{3}(y_t - 4)(y_{t+5} - 4)}{16.82}$$

$$r_5 = \frac{1}{16.82}[(-1)(2.1) + (-2)(0.4) + (-2)(1.5)] = -0.3508$$

$$r_6 = \hat{\rho}(6) = \frac{\sum_{t=1}^{2}(y_t - 4)(y_{t+6} - 4)}{16.82}$$

$$r_6 = \frac{1}{16.82}[(-1)(0.4) + (-2)(1.5)] = -0.2021$$

$$r_7 = \hat{\rho}(7) = \frac{\sum_{t=1}^{1}(y_t - 4)(y_{t+7} - 4)}{16.82}$$

$$r_7 = \frac{1}{16.82}[(-1)(1.5)] = -0.0892$$

also, one can plot the SACF by two axes: x-axis having the lag times between the observations, and the y-axes: the corresponding SACF coefficients, the resulting figure is called the correlogram.

2- Calculate the standard errors for these estimates.

We can calculate the SE of r_k from Bartlet's equation:

$$SE(r_k) \cong \sqrt{\frac{1}{n}\left(1 + 2\sum_{j=1}^{q} r_j^2\right)} \quad , k > q$$

$$SE(r_1) \cong \sqrt{\frac{1}{8}(1)} = 0.3536$$

$$SE(r_2) \cong \sqrt{\frac{1}{8}(1 + 2\,r_1^2)} = 0.4532$$

$$SE(r_3) \cong \sqrt{\frac{1}{8}(1 + 2\,r_1^2 + 2\,r_2^2)} = 0.4654$$

$$SE(r_4) \cong \sqrt{\frac{1}{8}(1 + 2\,r_1^2 + 2\,r_2^2 + 2\,r_3^2)} = 0.4859$$

$$SE(r_5) \cong \sqrt{\frac{1}{8}(1 + 2\,r_1^2 + 2\,r_2^2 + 2\,r_3^2 + 2\,r_4^2)} = 0.5176$$

$$SE(r_6) \cong \sqrt{\frac{1}{8}(1 + 2\,r_1^2 + 2\,r_2^2 + 2\,r_3^2 + 2\,r_4^2 + 2\,r_5^2)} = 0.5465$$

$$SE(r_7) \cong \sqrt{\frac{1}{8}(1 + 2\,r_1^2 + 2\,r_2^2 + 2\,r_3^2 + 2\,r_4^2 + 2\,r_5^2 + 2\,r_6^2)} = 0.5532$$