Statistics B3: Time Series Analysis
Solutions to Supplementary Time Series Analysis exam
2016-05-25

Department of Statistics, Uppsala University

Spring 2016

Task 1

A)

A stochastic process is a sequence of random variables.

B)

In order to be *covariance stationary*, a stochastic process need to fulfill the following

- $E[Y_t] = \mu, \forall\, t$, i.e. the mean is constant over time.
- $V(Y_t) = \gamma_0 < \infty \,\forall\, t$, i.e. the variance is finite and constant over time.
- $Cov(Y_{t+j}, Y_t) = Cov(Y_j, Y_0) \,\forall\, t, j$, i.e. the covariance is independent of *t* and only a function of the lag length *j*.

The third condition actually implies the second condition.

C)

The process

$$Y_t = \alpha_0 + \alpha_1 t + \phi_1 Y_{t-1} + e_t$$

is trend-stationary for $\alpha_1 \neq 0$, $|\phi_1| < 1$ and any (finite) value of α_0. The transformed process

$$Y_t^* = Y_t - \alpha_1 t = \alpha_0 + \phi_1 Y_{t-1} + e_t$$

is covariance stationary with the specified parameters.

D)

The process is previous subtask is difference-stationary for $\alpha_1 = 0$, $\phi_1 \geq 1$ and any (finite) value of α_0. The transformed process

$$\nabla^d Y_t = (1 - B)^d Y_t$$

is covariance stationary with the specified parameters such that $Y_t \sim I(d)$.

Task 2

A)

- **Identification**: Identify the DGP, purpose being to find a DGP that is suitable for modelling the data, this can be done by looking at correlograms.
- **Estimation**: Estimate the parameters of the model, this can be done by OLS or maximum likelihood estimation.
- **Diagnostics**: Investigate to what extent the model captures the systematic variation in the data. The residuals should be white noise, i.e. uncorrelated. This can be done by looking at correlograms and performing statistical tests on the estimated parameters and residuals.
- **Forecasting**: Use the model to forecast conditional mean of the process. Forecasts can be done based on conditional means.

B)

$$Y_t = \phi Y_{t-1} + e_t$$
$$e_t \sim N(0, \sigma^2)$$

$$E[Y_t] = E[\phi Y_{t-1} + e_t] = \phi E[Y_{t-1}] + E[e_t] = \phi E[Y_{t-1}]$$

assume stationarity, $E[Y_t] = E[Y_{t-1}]$, i.e. $|\phi| < 1$

$$E[Y_t](1 - \phi_1) = 0$$
$$E[Y_t] = \frac{0}{1 - \phi} = 0.$$

C)

$$\gamma_0 = V(Y_t) = V(\phi Y_{t-1} + e_t)$$
$$= \phi^2 V(Y_{t-1}) + V(e_t) + 2\phi Cov(Y_{t-1}, e_t)$$
$$= \phi^2 V(Y_{t-1}) + \sigma^2 + 0,$$

assume stationarity, $V(Y_t) = \gamma_0 \ \forall \ t$, i.e. $|\phi| < 1$

$$\gamma_0 = \phi^2 \gamma_0 + \sigma^2$$
$$\gamma_0 - \phi^2 \gamma_0 = \sigma^2$$
$$\gamma_0(1 - \phi^2) = \sigma^2$$
$$\gamma_0 = \frac{\sigma^2}{1 - \phi^2}.$$

Task 3

A)

By recursion,

$$E[X_t] = E[X_{t-1} + e_t]$$
$$E[X_t] = E[(X_{t-2} + e_{t-1}) + e_t]$$
$$E[X_t] = E[((X_{t-3} + e_{t-2}) + e_{t-1}) + e_t]$$
$$E[X_t] = E[(((X_{t-4} + e_{t-3}) + e_{t-2}) + e_{t-1}) + e_t]$$
$$\vdots$$
$$E[X_t] = E[X_0 + e_1 + e_2 + ... + e_{t-1} + e_t],$$

$$E[X_t] = E[X_0] + E[e_1] + E[e_2] + ... + E[e_{t-1}] + E[e_t] = 0$$

B)

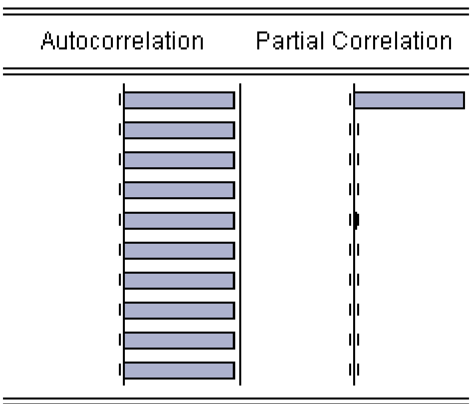
By recursion,

$$V[X_t] = V[X_{t-1} + e_t]$$
$$V[X_t] = V[(X_{t-2} + e_{t-1}) + e_t]$$
$$V[X_t] = V[((X_{t-3} + e_{t-2}) + e_{t-1}) + e_t]$$
$$V[X_t] = V[(((X_{t-4} + e_{t-3}) + e_{t-2}) + e_{t-1}) + e_t]$$
$$\vdots$$
$$V[X_t] = V[X_0 + e_1 + e_2 + ... + e_{t-1} + e_t],$$

$$V[X_t] = V[X_0] + V[e_1] + V[e_2] + ... + V[e_{t-1}] + V[e_t] + 2cov(e_1, e_2) + ... + 2cov(e_{t-1}, e_t) = t\sigma^2$$

C)

A correlogram of the realization of the process with 5000 observations:



D)

Differencing the process once gives

$$\nabla Y_t = (1-B)Y_t = Y_t - Y_{t-1} = e_t$$

Differenceing the process twice gives

$$\nabla^2 Y_t = (1-B)(Y_t - Y_{t-1}) = (1-B)e_t = e_t - e_{t-1}.$$

The process can be expressed as ARMA(0,1) where $\theta = 1$. The process is stationary but not invertible.

E)

The expected value is

$$E[\nabla^2 Y_t] = E[e_t - e_{t-1}] = E[e_t] - E[e_{t-1}] = 0.$$

F)

The variance is

$$\gamma_0 = V(\nabla^2 Y_t = V(e_t - e_{t-1})) = V(e_t) + V(e_{t-1}) - 2cov(e_t, e_{t-1}) = 2\sigma^2$$

G)

The first autocovariance is

$$\gamma_1 = Cov(\nabla^2 Y_t, \nabla^2 Y_{t-1})$$

$$= Cov(e_t - e_{t-1}, e_{t-1} - e_{t-2}) = Cov(e_t, e_{t-1}) - Cov(e_t, e_{t-2}) - Cov(e_{t-1}, e_{t-1}) + Cov(e_{t-1}, e_{t-2})$$
$$= -V(e_{t-1}) = -\sigma^2.$$

The first autocorrelation is

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-\sigma^2}{2\sigma^2} = -\frac{1}{2}.$$

Task 4

A)

1. Hypothesis

$$H_0 : \{Y_t\} \text{ has a unit root } \Leftrightarrow a = 0$$
$$H_1 : \{Y_t\} \text{ does not have a unit root } \Leftrightarrow a < 0$$

2. Significance level

$\alpha = 0.05$

3. Estimator(s)/Statistics

OLS estimator of a

5. Test statistic

$$ADF_{obs} = \frac{\hat{a} - 0}{\hat{\sigma}_{\hat{a}}}$$

6. Rejection rule and figure

The test statistic doesn't follow a standard distribution so we cannot draw a figure. Reject if p-value < 0.05

4. Assumptions

T is large

7. Calculations and results

P-value from Figure 5.2: $p = 0.0000 < 0.05 \Rightarrow H_0$ is not rejected.

8. Conclusion

We reject the null hypothesis that this process has a unit root at the 5% significance level.

B)

1. Hypothesis

$$\begin{array}{ll} H_0: & \rho_1 = \rho_2 = ... = \rho_8 = 0 \\ H_1: & \text{At least one } \rho_j \neq 0, j = 1, 2, ..., 8 \end{array}$$

2. Significance level

$\alpha = 0.05$

3. Estimator(s)/Statistics

$\hat{\rho}_j, \;\; j = 1, 2, ..., 8$

4. Assumptions

T is large

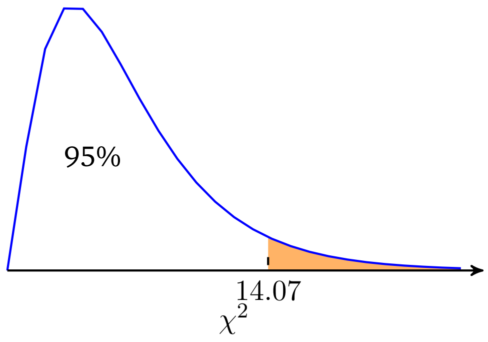
5. Test statistic

$Q_{LB} = T(T+2)\sum_{j=1}^{K} \dfrac{\hat{\rho}_j^2}{T-j} \sim \chi^2_{K-p-q-P-Q}$

$p:$	no. of AR-terms
$q:$	no. of MA-terms
$P:$	no. of Seasonal AR-terms
$Q:$	no. of Seasonal MA-terms

6. Rejection rule and figure

Reject if: $Q_{LB} > \chi^2_{7,\,0.05} = 14.07$



7. Calculations and results

From Figure 5: $Q_{LB} = 10.816 < \chi^2_{crit} = 14.07$

8. Conclusion

We cannot reject the null hypothesis that the eight first autocorrelations are simultaneously zero at the 5% significance level. Based on this test alone the AR(1) seems to be a good model.

C)

Test of the eighth autocorrelation

1. Hypothesis

$H_0: \ \rho_8 = 0 \quad H_1: \ \rho_8 \neq 0$

2. Significance level

$\alpha = 0.05$

3. Estimator

$\hat{\rho}_8$, *the sample autocorrelation for lag 8.*

4. Assumptions

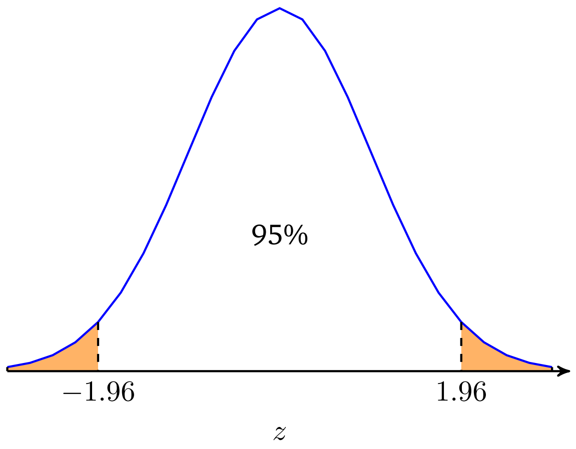
Large T

5. Test statistic

$z_{obs} = \frac{\hat{\rho}_8}{\sqrt{1/T}}$, under the null and large T : $z_{obs} \sim N(0,1)$

6. Rejection rule and figure

Reject if: $z_{obs} < -1.96$ or $z_{obs} > 1.96$



7. Calculations and results

$$z_{obs} = \frac{0.236}{\sqrt{1/98}} = 2.336$$

$$z_{obs} > 1.96 \Rightarrow \ H_0 \text{ rejected}$$

8. Conclusion

We reject the null hypothesis that the eighth autocorrelation is equal to zero at the 5 % significance level. Based on this test the AR(1) model does not capture all the systematic variation in the river flow data.

D)

The last model, $SARMA(1,0) \times (0,1)_8$ has the highest R^2 and the lowest information criteria. The model captures the spike at the eighth lag. These tools indicate that it is the best model. In addition, all the estimated parameters and the model as a whole are significant.

E)

The lag polynomials are:

$$\Phi(B) = 1$$
$$\phi(B) = (1 - \phi B)$$
$$\Theta(B) = (1 - \Theta B^8)$$
$$\theta(B) = 1.$$

F)

Using the lag polynomials from the previous subtask the model can be formulated as

$$\Phi(B)\phi(B)Y_t = \Theta(B)\theta(B)e_t$$
$$(1 - \phi B)Y_t = (1 - \Theta B^8)e_t$$
$$Y_t - \phi Y_{t-1} = e_t - \Theta e_{t-8}$$

and expressed with only Y_t on the left hand side the model is

$$Y_t = \phi Y_{t-1} - \Theta e_{t-8} + e_t.$$