Charles University
Faculty of Mathematics and Physics

Selected Supplementary Exercises

NMST537 Time Series Analysis | NMEK432 Econometrics

November 2017 (ver. 1)

Radek Hendrych[1]

[1]Dept. of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Prague 8, Czech Republic. *E-mail*: hendrych@karlin.mff.cuni.cz

Contents

1 Time Series: Basic Concepts

1. Suppose that X_1, X_2, ... is a sequence of random variables with $\mathbb{E}(X_i) < \infty$ and $\mathbb{E}(X_i) = \mu$ for all $i \in \mathbb{N}$.

   (a) Show that the random variable $f(X_1, \ldots, X_n)$, $n \in \mathbb{N}$, that minimises
   $$\mathbb{E}\left[(X_{n+1} - f(X_1, \ldots, X_n))^2 \,|X_1, \ldots, X_n\right]$$
   is $f(X_1, \ldots, X_n) = \mathbb{E}\left(X_{n+1}\,|X_1, \ldots, X_n\right)$.

   (b) Deduce that the random variable $f(X_1, \ldots, X_n)$, $n \in \mathbb{N}$, that minimises
   $$\mathbb{E}\left[(X_{n+1} - f(X_1, \ldots, X_n))^2\right]$$
   is also $f(X_1, \ldots, X_n) = \mathbb{E}\left(X_{n+1}\,|X_1, \ldots, X_n\right)$.

   (c) If X_1, X_2, ... is i.i.d. with with $\mathbb{E}(X_i) < \infty$ and $\mathbb{E}(X_i) = \mu$ for all $i \in \mathbb{N}$, where μ is known, what is the minimum mean squared error predictor of X_{n+1} in terms of X_1, ..., X_n?

   (d) Under the conditions of point (c) show that the best linear unbiased estimator of μ in terms of X_1, ..., X_n is $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.

   (e) Under the conditions of point (c) show that $\bar{X}_n$ is the best linear predictor of X_{n+1} that is unbiased for μ.

   (f) If X_1, X_2, ... is i.i.d. with with $\mathbb{E}(X_i) < \infty$ and $\mathbb{E}(X_i) = \mu$ for all $i \in \mathbb{N}$, and if $S_0 = 0$, $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}$, what is the minimum mean squared error predictor of S_{n+1} is terms of S_1, ..., S_n?

2. Let $\{Z_t\}_{t\in\mathbb{Z}}$ be a sequence of independent normal random variables, each with mean 0 and finite variance σ^2, and let a, b and c be constants. Which, if any, of the following processes are (weakly) stationary? For each (weakly) stationary process specify the mean and autocovariance function.

   (a) $X_t = a + bZ_t + cZ_{t-2}$
   (b) $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$
   (c) $X_t = Z_t \cos(ct) + Z_{t-1} \sin(ct)$
   (d) $X_t = a + bZ_0$
   (e) $X_t = Z_0 \cos(ct)$
   (f) $X_t = Z_t Z_{t-1}$

3. If $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ are uncorrelated stationary sequences, show that $\{X_t + Y_t\}_{t\in\mathbb{Z}}$ is stationary with autocovariance function equal to the sum of the autocovariance functions of $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$.

4. Let $\{Z_t\}_{t\in\mathbb{Z}}$ be i.i.d. $\mathsf{N}(0, 1)$ noise and define
   $$X_t = \begin{cases} Z_t & \text{if } t \text{ is even,} \\ \frac{Z_{t-1}^2 - 1}{\sqrt{2}} & \text{if } t \text{ is odd.} \end{cases}$$

   (a) Show that $\{X_t\}_{t\in\mathbb{Z}}$ is $\mathsf{WN}(0, 1)$ but not i.i.d. noise.
   (b) Find $\mathbb{E}(X_{n+1}|X_1, \ldots, X_n)$ for n odd and n even and compare the results.

5. Let $\{x_1, \dots, x_n\}$ be observed values of a time series at times $1, \dots, n$, and let $\hat{\rho}(h)$ be the sample autocorrelation function at lag h.

   (a) If $x_t = a + bt$, where a and b are constants and $b \neq 0$, show that for each fixed $h \leq 1$ it holds: $\hat{\rho}(h) \to 1$ as $n \to \infty$.

   (b) If $x_t = c\cos(\omega t)$, where c and ω are constants ($c \neq 0$ and $\omega \in (-\pi, \pi]$), show that for each fixed h it holds: $\hat{\rho}(h) \to \cos(\omega h)$ as $n \to \infty$.

6. Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a stationary process with mean zero and let a and b be constants.

   (a) If $X_t = a + bt + s_t + Y_t$, where s_t is a seasonal component with period $d = 12$ (i.e. $s_t = s_{t+d}$ for all t), show that $\Delta\Delta_{12}X_t := (1-B)(1-B^{12})X_t$ (B denotes the lag operator) is stationary and express its autocovariance function in terms of that of $\{Y_t\}_{t\in\mathbb{Z}}$.

   (b) If $X_t = (a + bt)s_t + Y_t$, where s_t is a seasonal component with period $d = 12$ (i.e. $s_t = s_{t+d}$ for all t), show that $\Delta_{12}^2 X_t := (1-B^{12})^2 X_t$ (B denotes the lag operator) is stationary and express its autocovariance function in terms of that of $\{Y_t\}_{t\in\mathbb{Z}}$.

7. Let us consider the dataset ***DEATHS***, which contains the monthly accidental deaths in the USA during the period 1973-1978:

   (a) Display this time series, its histogram and its autocorrelation function. The presence of a strong seasonal component with period 12 is evident in the graph of the data and in the sample autocorrelation function.

   (b) Deseasonalise the data (e.g. by using some simple methods or dummy variables).

   (c) Estimate a suitable polynomial trend in the deseasonalised data.

   (d) Calculate estimated random errors and analyse their dependence structure.

2 Time Series: ARMA and ARIMA Models

1. Suppose that $X_1, X_2, \dots$ is a sequence of random variables with mean $\mu \in \mathbb{R}$ and autocorrelation function $\rho(\cdot)$. Show that the best predictor of X_{n+h} of the form $aX_n + b$ is obtained by choosing $a = \rho(h)$ and $b = \mu(1 - \rho(h))$.

2. Find the autocovariance function of the time series $X_t = Z_t + 0.3Z_{t-1} - 0.4Z_{t-2}$, where $\{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0,1)$.

3. Let $\{Y_t\}_{t\in\mathbb{Z}}$ be the AR(1) plus noise time series defined by

$$Y_t = X_t + W_t,$$

   where $\{W_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma_w^2)$ and $\{X_t\}_{t\in\mathbb{Z}}$ is the AR(1) process defined as

$$X_t = \phi X_{t-1} + Z_t,\ \{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma_z^2),\ \phi \in (-1, 1),$$

   and $\mathbb{E}(W_s Z_t) = 0$ for all s and t.

   (a) Show that $\{Y_t\}_{t\in\mathbb{Z}}$ is stationary and find its autocovariance function.

   (b) Show that the time series $U_t := Y_t - \psi Y_{t-1}$ is an MA(1) process.

   (c) Show that $\{Y_t\}_{t\in\mathbb{Z}}$ is an ARMA(1,1) process and express the three parameters of this model in terms of ϕ, σ_w^2 and σ_z^2.

4. Suppose that in a sample of size 100 from an AR(1) process $X_t = \phi X_{t-1} + Y_t$, $\{Y_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma^2)$, $\phi \in (-1, 1)$, with mean μ, $\phi = 0.6$ and $\sigma^2 = 1$ one obtains $\bar{x}_{100} = 0.271$. Construct an approximate 95% confidence interval for μ. Are the data compatible with the hypothesis that $\mu = 0$?

5. Let us consider the yearly dataset ***SUNSPOTS***, which contains the sunspot number in the years 1770-1869:

   (a) Display this time series and its sample autocorrelation function.

   (b) Fit AR(p) models with mean for $p = 1, 2, 3$ and compare them by the Akaike information criterion (select the best one according to this criterion).

   (c) Calculate estimated random errors for the model from (b) and display them.

   (d) Verify the model selected in (b) by using appropriate diagnostic tools.

   (e) Predict the next ten values of the sunspot series.

6. Determine which of the following ARMA processes are casual and which of them are invertible (in all cases $\{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, 1)$):

   (a) $X_t + 0.2X_{t-1} - 0.48X_{t-2} = Z_t$

   (b) $X_t + 1.9X_{t-1} + 0.88X_{t-2} = Z_t + 0.2Z_{t-1} + 0.7Z_{t-2}$

   (c) $X_t + 0.6X_{t-1} = Z_t + 1.2Z_{t-1}$

   (d) $X_t + 1.8X_{t-1} + 0.81X_{t-2} = Z_t$

   (e) $X_t + 1.6X_{t-1} = Z_t - 0.4Z_{t-1} + 0.04Z_{t-2}$

7. For those processes in Problem 5 that are casual, compute and graph their autocorrelation and partial autocorrelation functions (using software).

8. For those processes in Problem 5 that are casual, compute the first six coefficients $\psi_0, \ldots, \psi_5$ in the casual representation $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, $\psi_j \in \mathbb{R}$ for all $j \in \mathbb{N}_0$.

9. Compute the (partial) autocorrelation function of the AR(2) process

$$X_t = 0.8X_{t-2} + Z_t,\ \{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma^2),\ \sigma^2 \in (0, \infty).$$

10. Let $\{Y_t\}_{t\in\mathbb{Z}}$ be the ARMA plus noise time series defined by

$$Y_t = X_t + W_t,$$

where $\{W_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma_w^2)$ and $\{X_t\}_{t\in\mathbb{Z}}$ is the ARMA(p,q) process defined as

$$(1 - \phi_1 B - \cdots - \phi_p B^p)X_t = (1 + \theta_1 B + \cdots + \theta_q B^q)Z_t,\ \{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma_z^2),$$

$\phi_1, \ldots, \phi_p \in \mathbb{R}$, $\theta_1, \ldots, \theta_p \in \mathbb{R}$, $\phi_p \neq 0$, $\theta_q \neq 0$, and $\mathbb{E}(W_s Z_t) = 0$ for all s and t. Note that B denotes the lag operator.

   (a) Show that $\{Y_t\}_{t\in\mathbb{Z}}$ is stationary. Find its autocovariance function.

   (b) Show that the process $U_t := (1 - \phi_1 B - \cdots - \phi_p B^p)Y_t$ is an MA(r) process, where $r = \max(p, q)$. Show that $\{Y_t\}_{t\in\mathbb{Z}}$ is an ARMA(p,r) process.

11. Show that the following two MA(1) processes

$$\begin{aligned} X_t &= Z_t + \theta Z_{t-1},\ \{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma^2), \\ X_t &= \tilde{Z}_t + \frac{1}{\theta}\tilde{Z}_{t-1},\ \{\tilde{Z}_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma^2\theta^2), \end{aligned}$$

where $0 < |\theta| < 1$ and $\sigma^2 \in (0, \infty)$, have the same autocovariance function.

12. By matching the autocovariances and sample autocovariances at lags 0 and 1, fit a model of the form

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t,\ \{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma^2),\ \mu, \phi \in \mathbb{R},\ \sigma^2 \in (0, 1),$$

to the dataset ***STRIKES***. Use the fitted model to compute the best linear predictor of the number of strikes in 1981. Estimate the mean square error of this prediction and construct 95% prediction bounds for the number of strikes in 1981 assuming that $\{Z_t\}_{t\in\mathbb{Z}} \sim$ i.i.d $\mathsf{N}(0, \sigma^2)$. Verify this model.

13. Find the Yule-Walker estimates of ϕ_1, ϕ_2 and σ^2 in the AR(2) model

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t,\ \{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma^2),\ \phi_1, \phi_2 \in \mathbb{R},\ \sigma^2 \in (0, \infty).$$

14. Consider the AR(2) model

$$X_t = \phi X_{t-1} + \phi^2 X_{t-2} + Z_t,\ \{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma^2),\ \phi \in \mathbb{R},\ \sigma^2 \in (0, \infty).$$

   (a) For what values of ϕ is this a casual process?
   (b) The following sample moments are computed after observing $X_1, \ldots, X_{200}$: $\hat{\gamma}(0) = 6.060$ and $\hat{\rho}(1) = 0.687$, where $\hat{\gamma}(\cdot)$ denotes the estimated autocovariance function and $\hat{\rho}(\cdot)$ denotes the estimated autocorrelation function. Find estimates of ϕ and σ^2 by solving the Yule-Walker equations. If one finds more than one solution, the casual one is preferred.

15. Given two observations x_1, x_2 ($|x_1| \neq |x_2|$) from the casual AR(1) process satisfying

$$X_t = \phi X_{t-1} + Z_t,\ \{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma^2),\ \phi \in (-1, 1),\ \sigma^2 \in (0, \infty),$$

find the (quasi) maximum likelihood estimates of ϕ and σ^2.

16. Suppose that $\{X_t\}_{t\in\mathbb{Z}}$ is an ARIMA(p, d, q) process satisfying the difference equation

$$\phi(B)(1 - B)^d X_t = \theta(B) Z_t,\ \{Z_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(0, \sigma^2),\ \sigma^2 \in (0, \infty),$$

where $\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ and $\theta(B) = (1 + \theta_1 B + \cdots + \theta_q B^q)$ (B denotes the lag operator), $\phi_1, \ldots, \phi_p \in \mathbb{R}$, $\theta_1, \ldots, \theta_p \in \mathbb{R}$, $\phi_p \neq 0$, $\theta_q \neq 0$, $d \in \mathbb{N}$. Show that these difference equations are also satisfied by the process $W_t = X_t + A_0 + A_1 t + \cdots + A_{d-1} t^{d-1}$, where $A_0, \ldots, A_{d-1}$ are arbitrary random variables.

17. Apply the augmented Dickey-Fuller and KPSS test to the levels of Lake Huron data (consider the yearly dataset ***LAKE***, which contains the average level in feet of Lake Huron in the years 1875-1972). Interpret the outputs. Identify, estimate and verify the model suitable for this dataset.

18. Consider the dataset ***AIRPASS***, which contains the number of international airline passengers (in thousands) for each month from January 1949 through December 1960, with the last twelve values deleted. Find an ARIMA model for the logarithms of the given data. Estimate and verify the model. Construct and display 12-step ahead forecasts and the corresponding 95% prediction bounds.

19. Consider the dataset ***AIRPASS***, which contains the number of international airline passengers (in thousands) for each month from January 1949 through December 1960, with the last twelve values deleted. Decompose the logged series into the trend, seasonal and residual component. Find an appropriate ARMA model for the residual component. Estimate and verify the model. Construct and display 12-step ahead forecasts and the corresponding 95% prediction bounds. Compare the twelve forecast errors found from this approach with those found in Problem 18.

20. Consider the dataset ***TUNDRA*** . It contains the average maximum temperature over the month of February for the years 1895-1993 in an area of the USA whose vegetation is characterised as tundra.

    (a) Fit a straight line to the data using OLS. Is the slope of the line significantly different from zero?
    (b) Find an appropriate ARMA model to the residuals from the OLS fit in (a).
    (c) Calculate the MLE estimates of the intercept and the slope of the line and the ARMA parameters in (a). Is the slope of the line significantly different form zero?
    (d) Use the model to forecast the average maximum temperature for the years 1994 to 2004.

21. Suppose that the daily log return of a security, $\{r_t\}_{t\in\mathbb{Z}}$, follows the model

$$r_t = 0.01 + 0.2r_{t-2} + \varepsilon_t,$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a Gaussian white noise series with mean zero and variance 0.02. What are the mean and variance of the return series r_t? Compute the lag -1 and lag -2 autocorrelations of r_t. Assume that $r_{100} = -0.01$ and $r_{99} = 0.02$. Compute the 1-step and 2-step ahead forecasts of the return series at the forecast origin $t = 100$. What are the associated standard deviations of the forecast errors?

22. Consider the monthly log returns of CRSP equal-weighted index from January 1962 to December 2016. One may obtain the data from the CRSP website directly.[2]

    (a) Build an AR model for the series and check the fitted model.
    (b) Build an MA model for the series and check the fitted model.
    (c) Compute 1-step and 2-step ahead forecasts of the AR and MA models built in the previous two questions.
    (d) Compare the fitted AR and MA models.

23. Let $\{X_1, \ldots, X_{142}\}$ denote the data in the file ***WINE*** (it represents monthly sales in kilolitres of red wine by Australian winemakers from January 1980 through October 1991) and let $\{Y_1, \ldots, Y_{142}\}$ denote their natural logarithms. Denote by m the sample mean of the differenced series $\Delta_{12}Y_t = (1 - B^{12})Y_t$ (B denotes the lag operator).

    (a) Fit the MA(12) model for $\Delta_{12}Y_t - m$ and verify it.
    (b) Use the model in (a) to compute forecasts $X_{131}, \ldots, X_{142}$.
    (c) Compute the mean squared error for the 12 forecasts obtained in (b).
    (d) Repeat steps (b) and (c) for the corresponding forecasts obtained by applying the seasonal Holt-Winters method (with period 12) to the logged data $\{Y_1, \ldots, Y_{142}\}$.
    (e) Repeat steps (b) and (c) for the corresponding forecasts obtained by applying the non-seasonal Holt-Winters method to the logged data $\{Y_1, \ldots, Y_{142}\}$.
    (f) Compare the mean squared errors obtained by these three methods.

[2] `http://www.crsp.com/products/software-access-tools`, last access 13th November 2017.

3 Time Series: Multivariate ARMA Models

1. Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a stationary process and define the bivariate process $X_{t1} = Y_t$, $X_{t2} = Y_{t-d}$, where $d \neq 0$. Show that $\{(X_{t1}, X_{t2})^\top\}_{t\in\mathbb{Z}}$ is stationary and express its cross-correlation function in terms of the autocorrelation function of $\{Y_t\}_{t\in\mathbb{Z}}$. If the autocorrelation function of $\{Y_t\}_{t\in\mathbb{Z}}$, $\rho_Y(h)$, fulfills that $\rho_Y(h) \to 0$ as $h \to \infty$, show that there exist a lag k for which $\rho_{12}(k) > \rho_{12}(0)$ ($\rho_{12}(\cdot)$ denotes the cross-correlation function).

2. Determine the covariance matrix function of the VARMA(1,1) process satisfying

$$\boldsymbol{X}_t - \boldsymbol{\Phi}\boldsymbol{X}_{t-1} = \boldsymbol{Z}_t + \boldsymbol{\Theta}\boldsymbol{Z}_{t-1},\ \{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}} \sim \mathsf{WN}(\boldsymbol{0}, \boldsymbol{I}_2),$$

   where $\boldsymbol{I}_2$ is the 2×2 identity matrix and $\boldsymbol{\Phi} = \boldsymbol{\Theta} = \begin{pmatrix} 0.5 & 0.5 \\ 0.0 & 0.5 \end{pmatrix}$.

3. Consider the dataset ***STOCK7***. It contains the daily returns on seven different stock market indices from 27th April 1998 through 9th April 1999. Fit a multivariate autoregression model to the trivariate series consisting of the returns on the Dow Jones Industrials, All Ordinaries and Nikkei indices. Check the model for goodness of fit, verify it and interpret the results.

4. The bivariate AR(4) model $\boldsymbol{x}_t - \boldsymbol{\Phi}_4\boldsymbol{x}_{t-4} = \boldsymbol{\phi}_0 + \boldsymbol{e}_t$ is a special seasonal model with periodicity 4, where $\{\boldsymbol{e}_t\}_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed normal random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Such a seasonal model may be useful in studying quarterly earnings of a company.

   (a) Assume that $\boldsymbol{x}_t$ is weakly stationary. Derive the mean vector and covariance matrix of $\boldsymbol{x}_t$.

   (b) Derive the necessary and sufficient condition of weak stationarity for $\boldsymbol{x}_t$.

   (c) Show that $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}_4\boldsymbol{\Gamma}_{\ell-4}$ for $\ell > 0$, where $\boldsymbol{\Gamma}_\ell$ is the lag ℓ autocovariance matrix of $\boldsymbol{x}_t$.

5. The bivariate MA(4) model $\boldsymbol{x}_t = \boldsymbol{e}_t - \boldsymbol{\Theta}_4\boldsymbol{e}_{t-4}$ is another seasonal model with periodicity 4, where $\{\boldsymbol{e}_t\}_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed normal random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Derive the covariance matrices $\boldsymbol{\Gamma}_\ell$ of $\boldsymbol{x}_t$ for $\ell = 0, \ldots, 5$.

4 Time Series: Nonlinear Models

1. Derive multi-step ahead forecasts for a GARCH(1,2) model at the forecast origin h.

2. Suppose that $r_1, \ldots, r_T$ are observations of the AR(1)–GARCH(1,1) returns

$$r_t = \mu + \phi r_{t-1} + e_t,\ e_t = \sigma_t\varepsilon_t,\ \sigma_t^2 = \omega + \alpha_1 e_{t-1}^2 + \beta_1\sigma_{t-1}^2,$$

   where ε_t is a standard Gaussian white noise series and μ, ϕ, ω, α_1 and β_1 are the real parameters. Derive the conditional log likelihood function of the data.

3. In the previous Problem 2, assume that ε_t follows a standardised Student-t distribution with ν degrees of freedom. Derive the conditional log likelihood function of the data.

4. Consider the daily prices of Intel stock from January 1981 to December 2016. One may obtain the data from the *Yahoo! Finance* website directly.[3] Transform the prices into log returns. Build a GARCH model for the transformed series and compute 1-step to 5-step ahead volatility forecasts.

[3] `https://finance.yahoo.com/quote/INTC/history?p=INTC`, last access 13th November 2017.

5. Consider the daily prices of General Motors (GM) stock from January 1980 to December 2016. One may obtain the data from the *Yahoo! Finance* website directly.[4]

   (a) Build a GARCH model with Gaussian innovations for the log returns of GM stock. Check the model and write down the fitted model.

   (b) Build a GARCH-M model with Gaussian innovations for the log returns of GM stock. What is the fitted model?

   (c) Build a GARCH model with Student-t distribution for the log returns of GM stock, including estimation of the degrees of freedom. Write down the fitted model. Let ν be the degrees of freedom of the Student-t distribution. Test the hypothesis $H_0 : \nu = 6$ versus $H_1 : \nu \neq 6$, using the 5% significance level.

   (d) Build an EGARCH model for the log returns of GM stock. What is the fitted model?

   (e) Obtain 1-step to 6-step ahead volatility forecasts for all the models obtained. Compare the forecasts.

6. Consider the daily prices of General Motors (GM) stock from January 1980 to December 2016. One may obtain the data from the *Yahoo! Finance* website directly.[5]

   (a) Build a GJR-GARCH model for the log return series using the standard Gaussian innovations. Write down the fitted model. Is the leverage effect significant at the 1% level?

   (b) Build a GJR-GARCH model for the log return series using the Student-t innovations. Write down the fitted model. Is the leverage effect significant at the 1% level?

   (c) Build a GJR-GARCH model for the log return series using the GED innovations. Write down the fitted model. Is the leverage effect significant at the 1% level?

7. Suppose that the monthly log returns, in percentages, of a stock respect the following Markov switching model:

$$\begin{aligned} r_t &= 1.25 + e_t, \ e_t = \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \begin{cases} 0.10 e_{t-1}^2 + 0.93 \sigma_{t-1}^2 & \text{if } s_t = 1, \\ 4.24 + 0.10 e_{t-1}^2 + 0.78 \sigma_{t-1}^2 & \text{if } s_t = 2, \end{cases} \end{aligned}$$

   where the transition probabilities are

$$\mathbb{P}(s_t = 2 | s_{t-1} = 1) = 0.15, \ \mathbb{P}(s_t = 1 | s_{t-1} = 2) = 0.05.$$

   Suppose that $r_{100} = 6.0$, $\sigma_{100}^2 = 50.0$ and $s_{100} = 2$ with probability 1. What is the 1-step ahead volatility forecast at the forecast origin $t = 100$? Also, if the probability of $s_{100} = 2$ is reduced to 0.8, what is the 1-step ahead volatility forecast at the forecast origin $t = 100$?

5 Time Series: State-space Models

1. Consider the ARMA(1,1) model

$$y_t - 0.8 y_{t-1} = \varepsilon_t + 0.4 \varepsilon_{t-1}, \ \{\varepsilon_t\}_{t \in \mathbb{Z}} \sim \text{i.i.d. } \mathsf{N}(0, 0.49).$$

   Convert the model into a state-space form.

[4] `https://finance.yahoo.com/quote/GM/history?p=GM`, last access 13th November 2017.
[5] `https://finance.yahoo.com/quote/GM/history?p=GM`, last access 13th November 2017.

2. Consider the following AR(3) model

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \varphi_3 x_{t-3} + \varepsilon_t,\ \varepsilon_t \sim \text{i.i.d. } \mathsf{N}(0, \sigma_\varepsilon^2),\ \sigma_\varepsilon^2 \in (0, \infty),$$

and suppose that the observed data are

$$y_t = x_t + e_t,\ e_t \sim \text{i.i.d. } \mathsf{N}(0, \sigma_e^2),\ \sigma_e^2 \in (0, \infty),$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and $\{e_t\}_{t\in\mathbb{Z}}$ are independent and the initial values of x_j with $j \leq 0$ are independent of e_t and ε_t for $t > 0$.

(a) Convert the model into a state-space form.

(b) If $\mathbb{E}(e_t) = c$, which is not zero, what is the corresponding state-space form for the system?

6 Econometrics: Regression Models

1. Consider the least squares regression of $\boldsymbol{y}$ on K variables (with a constant) $\boldsymbol{X}$. Consider an alternative set of regressors $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{P}$, where $\boldsymbol{P}$ is a nonsingular matrix. Thus, each column of $\boldsymbol{Z}$ is a mixture of some of the columns of $\boldsymbol{X}$. Prove that the residual vectors in the regressions of $\boldsymbol{y}$ on $\boldsymbol{X}$ and $\boldsymbol{y}$ on $\boldsymbol{Z}$ are identical. What relevance does this have to the question of changing the fit of a regression by changing the units of measurement of the independent variables?

2. A data set consists of n observations on $\boldsymbol{X}_n$ and $\boldsymbol{y}_n$. The least squares estimator based on these n observations is $\boldsymbol{b}_n = (\boldsymbol{X}_n^\top \boldsymbol{X}_n)^{-1} \boldsymbol{X}_n \boldsymbol{y}_n$. Another observation, $\boldsymbol{x}_s$ and y_s, becomes available. Prove that the least squares estimator computed using this additional observation is

$$\boldsymbol{b}_{n,s} = \boldsymbol{b}_n + \frac{1}{1 + \boldsymbol{x}_s^\top (\boldsymbol{X}_n^\top \boldsymbol{X}_n)^{-1} \boldsymbol{x}_s} (\boldsymbol{X}_n^\top \boldsymbol{X}_n)^{-1} \boldsymbol{x}_s \left(y_s - \boldsymbol{x}_s^\top \boldsymbol{b}_n \right).$$

Note that the last term is e_s, the residual from the prediction of y_s using the coefficients based on $\boldsymbol{X}_n$ and $\boldsymbol{b}_n$. Conclude that the new data change the results of least squares only if the new observation on y cannot be perfectly predicted using the information already in hand.

3. For the classical normal regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with no constant term and K regressors, assuming that the true value of $\boldsymbol{\beta}$ is zero, what is the exact expected value of $F_{K,n-K} = (R^2/K)/[(1 - R^2)/(n - K)]$? Note that R^2 denotes the coefficient of determination.

4. Prove that $\mathbb{E}[\boldsymbol{b}^\top \boldsymbol{b}] = \boldsymbol{\beta}^\top \boldsymbol{\beta} + \sigma^2 \sum_{k=1}^{K} \frac{1}{\lambda_k}$, where $\boldsymbol{b}$ is the ordinary least squares estimator in the classical normal regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and λ_k is a characteristic root of $\boldsymbol{X}^\top \boldsymbol{X}$.

5. For the simple regression model $y_i = \mu + \varepsilon_i$, $\varepsilon_i \sim$ i.i.d. $\mathsf{N}(0, \sigma^2)$, prove that the sample mean is consistent and asymptotically normally distributed. Now consider the alternative estimator $\hat{\mu} = \sum_i w_i y_i$, $w_i = \frac{i}{n(n+1)/2}$. Note that $\sum_i w_i = 1$. Prove that this is a consistent estimator of μ and obtain its asymptotic variance.

6. Consider the dataset ***SHIPS*** containing the number of incidents of damage to a sample of ships, with the type of ship and the period when it was constructed. There are five types of ships and four different periods of construction. Use F tests and dummy variable regressions to test the hypothesis that there is no signficant "ship type effect" in the expected number of incidents. Now, use the same procedure to test whether there is a significant "period effect".

7. Does first differencing reduce autocorrelation? Consider the models $y_t = \boldsymbol{\beta}^\top \boldsymbol{x}_t + \varepsilon_t$, where $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ and $\varepsilon_t = u_t - \lambda u_{t-1}$. Compare the autocorrelation of ε_t in the original model with that of v_t in $y_t - y_{t-1} = \boldsymbol{\beta}^\top(\boldsymbol{x}_t - \boldsymbol{x}_{t-1}) + v_t$, where $v_t = \varepsilon_t - \varepsilon_{t-1}$.

7 Econometrics: Systems of Econometric Equations

1. Prove that in the following SUR system consisting of two regression equations:

$$\begin{aligned} \boldsymbol{y}_1 &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1, \\ \boldsymbol{y}_2 &= \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2, \end{aligned}$$

the SUR estimator is equivalent to equation-by-equation ordinary least squares if $\boldsymbol{X}_1 = \boldsymbol{X}_2$. Does your result hold if it is also known that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$?

8 Econometrics: Models for Discrete and Limited Responses

1. Suppose that a linear probability model is to be fit to a set of observations on a dependent variable y that takes values zero and one, and a single regressor x that varies continuously across observations. Obtain the exact expressions for the least squares slope in the regression in terms of the mean(s) and variance of x, and interpret the result.

2. Consider the dataset ***STRIKES_DUR*** containing the strike duration in days $[T]$ and unanticipated industrial production $[PROD]$ for a number of strikes in each of 9 years. Use the Poisson regression model to determine whether $PROD$ is a significant determinant of the number of strikes in a given year.

Note: If not specifically specified, use 5% significance level to draw conclusions in the exercises.

References

[1] P. J. Brockwell and R. A. Davis: *Introduction to Time Series and Forecasting.* Springer, Berlin 2002.

[2] W. H. Greene: *Econometric Analysis.* 5th edition, Prentice Hall, Upper Saddle River (NJ) 2003.

[3] R. S. Tsay: *Analysis of Financial Time Series.* Wiley, Hoboken (NJ) 2013.

Note: This material has been compiled employing the above mentioned sources [1]-[3].