

Time Series - Practical Exercises

Questions 1 to 8 are based on the exercises at the end of chapter 2 of Enders (2010, 2004). Question 9 is from the exercises at the end of Chapter 3. Some of the results have been changed to correspond more closely to the conventions used in our practical work.

1. [Enders 2.6] Consider the second-order stochastic difference equation: $y_t = 1.5y_{t-1} - 0.5y_{t-2} + \varepsilon_t$.
 - (a) Find the characteristic roots of the homogeneous equation.
 - (b) Demonstrate that the roots of $1 - 1.5L + 0.5L^2$ are the reciprocals of your answer in part (a)
 - (c) Given initial conditions for y_0 and y_1 , find the solution for y_t , in terms of the current and past values of the $\{\varepsilon_t\}$ sequence.
 - (d) Find the forecast function for y_{T+s} , (i.e., find the solution for any value of y_{T+s} , given the values of y_T and y_{T-1}).
 - (e) Find $Ey_t, Ey_{t+1}, var(y_t), var(y_{t+1}),$ and $cov(y_{t+1}, y_t)$.

2. [Enders 2.8] The file entitled SIM_2.csv contains the simulated data sets used in chapter 2 of Enders (2010). The first series, denoted Y1, contains the 100 values of the simulated $AR(1)$ process used in Section 7 of Chapter 2. Use this series to perform the following tasks (Note: Due to differences in data handling and rounding, your answers need only approximate those presented here):
 - (a) Plot the sequence against time. Does the series appear to be stationary?
 - (b) Verify that the first 12 coefficients of the ACF and the PACF¹ are as follows:
ACF
1 : 0.739 0.584 0.471 0.389 0.344 0.335
7: 0.297 0.325 0.269 0.201 0.189 0.082
PACF
1: 0.739 0.083 0.030 0.026 0.060 0.089
7: -0.017 0.144 -0.100 -0.065 0.070 -0.204
Ljung-Box Q-Statistics: $Q(8) = 177.58, Q(16) = 197.84, Q(24) = 201.28$.
 - (c) Use the data to verify the results given below.

¹PACF results in Enders (2004) are derived using the Yule-Walker method. The default in EViews is not to use this option. Thus you may find considerable differences in the default values of the PACF calculated by Eviews and those given in Enders (2010).

	Model 1 $y_t = a_1 y_{t-1} + \varepsilon_t$	Model 2 $y_t = a_1 y_{t-1} + \varepsilon_t + \beta_{12} \varepsilon_{t-12}$
Degrees of Freedom	99	98
Sum of Squared residuals	85.21	85.17
Estimated a_1 (standard error)	0.7905 (0.0624)	0.7950 (0.0642)
Estimated β_{12} (standard error)		-0.0335 (0.1141)
AIC; SBC	AIC=2.707; SBC=2.733	AIC=2.737; SBC=2.779
Ljung-Box Q-statistics for the residuals (significance) level in parenthesis	Q(8)=6.14 (0.524) Q(16)=15.74 (0.399) Q(24)=21.02 (0.580)	Q(8)=6.19 (0.402) Q(16)=15.66 (0.335) Q(24)=20.88 (0.528)

(d) Estimate the series as an $AR(2)$ process without an intercept. You should obtain:

Included observations: 98				
Coefficient	Estimate	Standard Error	t -statistic	Probability
α_1	0.710	0.101	7.014	0.0000
α_2	0.105	0.101	1.036	0.3030

Ljung-Box Q-Statistics: Q(8) = 4.7910 Q(16) = 15.182 Q(24) = 19.700

(e) Estimate the series as an $ARMA(1, 1)$ process without an intercept. You should obtain:

Included observations: 99				
Coefficient	Estimate	Standard Error	t -statistic	Probability
α_1	0.846	0.069	12.248	0.0000
β_1	-0.147	0.126	-1.162	0.2481

Verify that the ACF and PACF of the residuals do not indicate any serial correlation.

3. [Enders 2.9] The second column in file SIM_csv contains the 100 values of the simulated $ARMA(I, 1)$ process used in Section 2.7 of Enders (2010). This series is entitled Y2. Use this series to perform the following tasks (Note.. Due to differences in data handling and rounding, your answers need only approximate those presented here): (To facilitate comparisons all models should be estimated over the same sample. The final model can of course be re-estimated over the entire sample).

(a) Plot the sequence against time. Does the series appear to be stationary? Verify that the first 12 coefficients of the ACF and the PACF are as follows:

ACF

1: -0.834 0.597 -0.440 0.350 -0.319 0.332

7: -0.337 0.317 -0.276 0.179 -0.084 0.038

PACF

1: -0.834 -0.328 -0.194 -0.015 -0.140 0.849

7: 0.000 0.010 0.017 -0.199 -0.084 -0.021

(b) Verify the results in the table below.

	Estimates	Q-Statistics	AIC/SBC
Model 1 AR(1)	a_1 :-0.835(.053)	Q(8)=26.19(.000) Q(24)=41.63(.010)	AIC=3.316 BIC=3.342
Model 2 ARMA(1,1)	a_1 :-0.690(.077) b_1 :-0.681(.079)	Q(8)=3.81(.703) Q(24)=14.18(.895)	AIC=3.051 BIC=3.103
Model 3 AR(2)	a_1 :-1.168(.094) a_2 :-0.378(.093)	Q(8)=11.74(.068) Q(24)=23.28(.386))	AIC=3.179 BIC=3.232

(c) Estimate the process using a pure MA(2) model. You should obtain:

Included observations: 100				
Coefficient	Estimate	Standard Error	<i>t</i> -statistic	Probability
β_1	-1.27	0.086	-14.73	0.0000
β_2	0.542	0.087	6.22	0.000
AIC	3.240			
BIC	3.292			

Verify that the Ljung-Box Q-Statistics are:

$$Q(8) = 28.836. \text{ Significance level } 0.000$$

$$Q(16) = 33.895 \text{ Significance level } 0.002$$

$$Q(24) = 36.425. \text{ Significance level } 0.027$$

(d) On the basis of this analysis which model would you choose?

4. [Enders 2.10] The third column in file SIM_2.csv contains the 100 values of the simulated AR(2) process used in Section 2.7. This series is entitled Y3. Use this series to perform the following tasks (Note: Due to differences in data handling and rounding, your answers need only approximate those presented here):

(a) Plot the sequence against time. Verify the ACF and the PACF coefficients reported in Section 2.7. Compare the sample ACF and PACF to the those of a theoretical AR(2) process.

(b) Estimate the series as an AR(*I*) process. You should find that the estimated AR(*I*) coefficient and the *t*-statistic are

$$y_t = \underset{(5.24)}{0.467} y_{t-1} + \varepsilon_t$$

Why is an AR(*I*) model inadequate?

(c) Could an ARMA(1, 1) process generate the type of sample ACF and PACF found in part a? Estimate the series as an ARMA(1, 1) process. You should obtain:

Included observations: 99				
Coefficient	Estimate	Standard Error	<i>t</i> -statistic	Probability
α_1	0.183	0.159	1.15	0.2523
β_1	0.510	0.140	3.64	0.0004

Use the Ljung-Box Q-Statistics to show that the ARMA(1,1) model is inadequate.

(d) Estimate the series as an AR(2) process to verify the results reported in the text.

5. [Enders p70-72² and Exercise 2.13] As you read more of the time-series literature, you will find that different authors and different software packages report the AIC and the SBC (BIC) in

²These pages contain a good account of the relative appropriateness of the AIC and BIC. Where the true data generating process is an ARMA(p,q), with finite p and q, BIC consistently estimates p and q in large samples. Where the true values of p and/or q may be infinite then BIC gives a better approximation to the true generating process.

various ways. The examples in Enders use

$$AIC = T \log SSR + 2n \quad (1)$$

$$BIC = T \log SSR + n \log T \quad (2)$$

where SSR = sum of squared residuals. EViews, in general, uses the following

$$AIC = -2 \log L/T + \frac{2n}{T} \quad (3)$$

$$SBC = -2 \log L/T + \frac{n \log T}{T} \quad (4)$$

$$(5)$$

where L = maximised value of the likelihood function. Other sources use

$$AIC = \frac{\exp(2n/T) \cdot SSR}{T} \quad (6)$$

$$BIC = \frac{T^{n/T} SSR}{T} \quad (7)$$

This is only an example of the different forms available. If you find this confusing you are correct. The problem is that some of the formulae are more convenient in certain circumstances. If you estimate one model with one piece of software and extract its AIC and then estimate a second model with a second piece of software and extract the AIC for the second model it is likely that the two estimates of AIC **are not compatible**. You will need to do a little more work. I have seen cases where an econometric package used different versions of AIC for different estimation procedures but this was generally obvious.

- (a) Show that all three methods of calculating the AIC will necessarily select the same model. [Hint: The AIC will select model 1 over model 2 if the AIC from model 1 is smaller than that from model 2. Show that all three ways of calculating the AIC are monotonic transformations of each other.]
 - (b) Show that all three methods of calculating the SBC will necessarily select the same model.
 - (c) Select one of the three pairs above. Show that the AIC will never select a more parsimonious model than the SBC.
6. [Enders 2.11] The file QUARTERLY.XLS contains a number of series including the US index of industrial production (indprod), unemployment rate (urate) and producer price index for finished goods (finished). All the series run from 1960Q1 to 2008Q2.
- (a) Exercise with indprod.
 - i. Construct the growth rate of the series as $y_t = \log indprod_t - \log indprod_{t-1}$. Since the first few autocorrelations suggest an AR(1) process, estimate $y_t = 0.0035 + 0.543y_{t-1} + \varepsilon$ (the t -statistics are 3.50 and 9.02 respectively).
 - ii. Show that adding an MA term at lag 8 improves the fit and removes the serial correlation
 - (b) Exercise with urate
 - i. Examine the ACF of the series. It should be clear that the ACF decays slowly and that the PACF has a very negative spike at lag 2.

- ii. Temporarily ignore the necessity of differencing the series. Estimate *urate* as an AR(2) process including a constant. You should find that $a_1 + a_2 = 0.9980$
- iii. Form the difference as $y_t = \text{urate}_t - \text{urate}_{t-1}$. Show that the model $y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_4 \varepsilon_{t-4} + \beta_8 \varepsilon_{t-8} + \varepsilon_t$ exhibits no remaining serial correlation and has a better fit than $y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_4 y_{t-4} + \beta_8 \varepsilon_{t-8} + \varepsilon_t$ or $y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_4 y_{t-4} + \alpha_8 y_{t-8} + \varepsilon_t$.
- (c) Exercise with finished
- i. Construct the growth rate of the series as $y_t = \log \text{finished}_t - \log \text{finished}_{t-1}$
- ii. Show that an ARMA(1,1) model fits the data well but that the residual autocorrelation at lag 3 is a problem.
- iii. Show that an ARMA((1,3), 1) is not appropriate. Show that an ARMA(1, (1,3)) shows no remaining residual correlation.
- iv. Show that an AR(3) has the best fit and has no remaining serial correlation.
7. [Enders 2.12] The file QUARTERLY.XLS contains US interest rate data from 1960Q1 to 2008Q1. Form the interest rate spread (slope of the interest rate yield curve) by subtracting the T-bill rate from the 10-year rate.
- (a) Use the full sample period to obtain estimates of the AR(7) and an ARMA(1,1) model.³
- (b) Estimate the AR(7) and ARMA(1,1) over the period 1960Q1 to 1994Q3. Obtain the one-step-ahead forecast from each for the remainder of the sample. Let $[f_{1t}]$ denote the forecasts from the AR(7) model and $[f_{2t}]$ from the ARMA(1,1). You should find that the forecasts are such that $y_{1993Q3+t} = 0.0359 + 0.986f_{1t}$ and $y_{1993Q3+t} = -0.055 + 1.042f_{2t}$. Are the forecasts unbiased?
- (c) Use the Eviews forecast evaluation table to evaluate these forecasts
- (d) Construct the Diebold-Mariano test using the mean absolute error. Compare your results with those reported in section 10 of chapter 2 of Enders (2010).
- (e) Use the Granger-Newbold test to compare the AR(7) and ARMA(1,1) forecasts.
- (f) Construct the ACF and PACF of the first-difference of the spread. What type of model is suggested.
- (g) Estimate a model with 2 AR lags and MA lags at 3, 5 and 8. and show that it is appropriate.
8. [Enders 2.16] The file QUARTERLY.XLS contains the quarterly values of the Consumer Price Index (excluding food and fuel) that have not been seasonally adjusted (*CPINSA*). Form the inflation rate, as measured by the *CPI* as $\ln(\text{CPINSA}_t / \text{CPINSA}_{t-1})$
- (a) Plot the inflation rate. Does there appear to be a plausible seasonal pattern in the data? Seasonally difference the CPI using $\Delta_4 \ln \text{CPINSA}_t = \ln \text{CPINSA}_t - \ln \text{CPINSA}_{t-4}$. Find the ACF of this series and explain why it is not appropriate to apply Box-Jenkins methodology to this series
- (b) For convenience let $\pi_t = (1 - L)(1 - L^4) \ln \text{Cpinsa}_t$. Find an appropriate model for the π_t series. In particular consider the following four models for the series.
- | | |
|---|--|
| $\pi_t = a_0 + a_1 \pi_{t-1} + \varepsilon_t + \varepsilon_{t-4}$ | Model 1: AR(1) with Seasonal MA |
| $\pi_t = a_0 + (1 + a_1 L)(1 + a_4 L \pi_{t-1} + \varepsilon_t$ | Model 2: Multiplicative Autoregressive |
| $p i_t = a_0 + (1 + \beta_1 L)(1 + \beta_4 L) \varepsilon_t$ | Model 3: Multiplicative Moving Average |
| $\pi_t = a_0 + (a_1 \pi_{t-1} + \beta_1 \varepsilon_{t-1})(1 + \beta_4 \varepsilon_{t-4} + \varepsilon_t$ | Model 4: ARMA(1,1) with Seasonal MA |

³See section 10 of chapter 2 of Enders (2010) for a discussion of this series

9. [Enders 3.5] The file labelled ARCH.csv contains the 100 realisations of the simulated $\{y_t\}$ sequence used to create the Panel (d) of Figure 3.7.(on page 129 of Enders (2010)). Recall that this series was simulated as $y_t = 0.9y_{t-1} + \varepsilon_t$, where ε_t is the ARCH(1) error process $\varepsilon_t = \nu_t(1 + 0.8\varepsilon_{t-1})^{1/2}$. You should find the series has a mean of 0.263, a standard deviation of 4.894 with minimum and maximum values of -10.8 and 15.15, respectively.

- (a) Estimate the series using OLS and save the residuals. You should obtain

$$y_t = 0.944y_{t-1} + \varepsilon_t.$$

The t -statistic for $\hat{\alpha}_1$ is 26.51. Note that the estimated value of α_1 , differs from the theoretical value of 0.9. This is due to nothing more than sampling error; the simulated values of $\{y_t\}$ do not precisely conform to the theoretical distribution. However, can you provide an intuitive explanation of why positive serial correlation in the $\{\nu_t\}$ sequence might shift the estimate of α_1 upward in small samples?

- (b) Obtain the ACF and the PACF of the residuals. Use Ljung-Box Q-statistics to determine whether the residuals approximate white noise. You should find:

	1	2	3	4	5	6	7	8
ACF	0.149	0.004	-0.018	-0.012	0.068	0.003	-0.099	-0.151
PACF	0.149	-0.018	-0.016	-0.007	0.073	-0.019	-0.100	-0.123

$$Q(4) = 2.31, Q(8) = 6.39, Q(24) = 18.49.$$

- (c) Obtain the ACF and the PACF of the squared residuals. You should find:

	1	2	3	4	5	6	7	8
ACF	0.473	0.127	-0.057	-0.078	0.057	0.242	0.273	0.214
PACF	0.474	-0.125	-0.086	0.004	0.135	0.198	0.070	0.062

Based on the ACF and PACF of the residuals and the squared residuals, what can you conclude about the presence of ARCH errors?

- (d) Estimate the squared residuals as: $\varepsilon^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2$. You should verify that $\alpha_0 = 1.55$ (t -statistic=2.83) and $\alpha_1 = 0.474$ (t -statistic=5.28)

Show that the Lagrange multiplier ARCH(I) errors is: $TR^2 = 22.03$ with a significance level of 0.00000269.

- (e) Simultaneously estimate the $\{y_t\}$ sequence and the ARCH(I) error process using maximum likelihood estimation. You should find:

$$y_t = 0.886 y_{t-1} + \varepsilon_t \quad h_t = 1.16 + 0.663\varepsilon_{t-1}^2$$

(33.00) (4.37) (3.02)

10. The file ireland2.csv contains the total returns index for the ISEQ. (Total returns are calculated on the basis that dividends are reinvested).

- (a) Examine the series, ACF and PACF of the index in levels, first differences, logs and first differences of the logs. From an economic viewpoint which is the appropriate model for such a series. Does the data fit this viewpoint. Is there autocorrelation in the transformed series? Does this imply independence?

- (b) Examine the squared values of the first differences of the logs and comment on the ACF and PACF of these first differences. Is there autocorrelation in this series. Comment on its relevance.
- (c) Estimate an appropriate ARCH/GARCH model for the return series (First differences of logs)
11. The file `cpi.csv` contains data on the Irish CPI for the period 1957 Quarter 1 to 2006 Quarter 2. It also contains data on a variable `c` which has been proposed as an indicator of the level of the CPI. This indicator is exogenous, known in advance and can not be influenced by the authorities.
- (a) Load the data into EVIEWS and add a EVIEWS date variable to the data set.
- (b) Regress the log of the CPI on `c` and a constant and comment on the fit of the regression, t -statistics and the results of any specification tests that you consider appropriate. You should also examine the residuals from the regression. Using these results, is `c` a suitable predictor of the (log)CPI?
- (c) An enquiry into the origin of `c` shows that it is cumulative rainfall in Armagh for the period 1857Q1 to 1906Q2. Is the apparent relationship due to chance or is there a deeper explanation for the spurious nature of the regression.
- (d) List the kind of conclusions that can and can not be drawn from regression analysis with time series data
12. The data set `npext.csv` is an updated version of the Nelson-Plosser “unit root” data set and contains the fourteen U.S. economic time series set out below. All series are transformed by taking logarithms except for the bond yield. The sample period ends in 1988.
- year** Time index from 1860 until 1988.
- realgnp** Real GNP, [Billions of 1958 Dollars, [1909 - 1988]
- nomgnp** Nominal GNP, [Millions of Current Dollars], [1909 - 1988]
- gnpperca** Real Per Capita GNP, [1958 Dollars], [1909 - 1988]
- indprod** Industrial Production Index, [1967 = 100], [1860 - 1988]
- employmt** Total Employment, [Thousands], [1890 - 1988]
- unemploy** Total Unemployment Rate, [Percent], [1890 - 1988]
- gnpdefl** GNP Deflator, [1958 = 100], [1889 - 1988]
- cpi** Consumer Price Index, [1967 = 100], [1860 - 1988]
- wages** Nominal Wages (Average annual earnings per full-time employee in manufacturing), [current Dollars], [1900 - 1988]
- realwag** Real Wages, [Nominal wages/CPI], [1900 - 1988]
- M** Money Stock (M2), [Billions of Dollars, annual averages], [1889 - 1988]
- velocity** Velocity of Money, [1869 - 1988]
- interest** Bond Yield (Basic Yields of 30-year corporate bonds), [Percent per annum], [1900 - 1988]
- sp500** Stock Prices, [Index; 1941 - 43 = 100], [1871 - 1988]

For a selected series

- (a) Using graphs examine the series, its first difference, and whatever other transformations you consider necessary. Comment on what the graphs indicate.
- (b) Does economics (or common sense) tell you if the variable is a random walk, a random walk with drift or a random walk with drift and trend. Do economics and the graphs agree. Which ADF test are you going to use?
- (c) Choose a maximum lag order and then determine what is the appropriate lag length to use for an ADF or DF unit root test. (Use AIC, BIC and lag length reduction tests to determine the appropriate lag length).
- (d) Implement an ADF test for the appropriate lag length (or range of lag lengths). Interpret your results. How sensitive are your conclusions to the choice of lag length?
- (e) Implement an ERS test for a unit root. Compare the results of ADF and ERS tests

13. The file `mypr3.csv` contains the following series

obs Year of observation,

m Logarithm of US M1,

p Logarithm of Prices (Price Deflator for US Net National Product)

y Logarithm of US real Net National Product

r Interest Rates

mr Logarithm of Real M1 ($mr = m - p$)

v Logarithm of Velocity ($v = y - m + p$)

Detailed definitions are in Appendix B of Stock and Watson (1993).⁴

- (a) Draw time series plots of
 1. real Net National Product and Real M1
 2. Log velocity and Interest Rates. Use separate axes for each variable.Do these graphs provide any indication of a possible long run money demand function
- (b) Complete Univariate Augmented Dickey-Fuller unit root tests for m , $m - p$, y , r , including 2 and 4 lags and report on your conclusions.
- (c) Complete DF-GLS unit root test choosing SIC as criterion for number of lags to be included. Compare these results with those derived from the standard ADF tests.
- (d) Test for co-integration of $m - p$, y and r using the Engle-Granger procedure. (There is no direct instruction in EViews to complete this test. You must do the appropriate OLS estimate, save the residuals and then calculate the appropriate ADF test for the residuals. The critical values given by EViews are for standard unit root tests and are not appropriate in this case. Correct Critical values are available in many textbooks. These critical values may depend on sample size and on number of variables (calculated in different ways in different tables). If you use the EViews default (SIC) to determine the number of lags you will use two lags and reject co-integration. Stock and Watson (1993) use one lag and can not reject co-integration. Replicate and comment.

⁴This paper examines various ways of estimating long-run relationships. For a summary of some of the analysis in this paper see Hayashi (2000).

- (e) Determine and appropriate lag-length for an appropriate VAR in the variables $m - p$, y and r . You should find that you 1 or 2 lags is appropriate depending on the information criteria used. Do specification tests for a VAR with 2 lags.
- (f) Do a Johansen test for co-integration in this VAR and determine an appropriate number of co-integrating vectors. Estimate the residuals in this VAR. Assuming one co-integrating relationship compare the results with the Engle-Granger results already obtained. Do these results provide evidence against a unit elasticity of real M1 with respect to income. Derive 95% confidence intervals for the semi-elasticity of real M1 with respect to interest rates.
- (g) Read the account of dynamic OLS (DOLS) in Stock and Watson (2007) or Hayashi (2000). This is a relatively simple way to estimate a long term relationship. Estimate a long run relationship between $m - p$, y and r using two lags and two leads of Δy and δr . You should get an equation of the form

$$m - p = 0.970y - 0.101r$$

$$(.046) \quad (.013)$$

Compare this result to those obtained using the Engle-Granger and Johansen methodology

14. Read Stock and Watson (2001).

- (a) Using a variety of lag determination criteria determine an appropriate order for the VAR in the paper. How do your results compare with those in the paper? What is your preferred lag length?
- (b) Replicate the descriptive statistics in table 1 in the paper. Repeat the analysis for your preferred lag length or if you prefer the Stock and Watson (2001) lag length your second choice and comment on the robustness of the analysis.
- (c) Replicate the impulse responses in figure 1. Compare these with those for the alternative model you analysed in the previous part.
- (d) Replicate the forecast evaluation in table 2 of the paper.
- (e) Replicate the impulse responses of monetary policy shocks reported in Figure 2.

References

- Enders, W. (2004). *Applied Econometric Time Series* (Second ed.). Wiley.
- Enders, W. (2010). *Applied Econometric Time Series* (3rd ed.). Wiley.
- Hayashi, F. (2000). *Econometrics*. Princeton University Press.
- Stock, J. H. and M. W. Watson (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61(4), 783-820.
- Stock, J. H. and M. W. Watson (2001). Vector autoregressions. *Journal of Economic Perspectives* 15(4), 101-115.
- Stock, J. H. and M. W. Watson (2007). *Introduction to Econometrics*. Pearson Addison Wesley.