

Springer Series in Statistics

Jan G. De Gooijer

# Elements of Nonlinear Time Series Analysis and Forecasting

**Solutions Manual**

 Springer

This is the solutions manual for the book *Elements of Nonlinear Time Series and Forecasting*, published by Springer-Verlag, ISBN 978-3-319-43251-9 and ISBN 978-3-319-43252-6 (eBook). This release was created February, 2017; eventual future releases with corrections to errors will be published on my personal web page.

© 2017, by Jan G. De Gooijer

*This material may be reproduced for testing or instructional purposes by people using the main text.*

---

## Preface

This manual contains the solutions to the theoretical exercises, simulation exercises, and empirical exercises. The exercises have been gathered from many sources, some are even taken from (published) research papers. The theoretical exercises have been carefully chosen to reinforce concepts explained in the main text of the book or to develop and generalize them in significant ways. With few exceptions, almost all theoretical exercises can be solved with the knowledge of the book, and the background knowledge on linear time series methods, matrix calculus and Markov chain methods. It should be noted, however, that in some cases the mathematical proofs are lengthy and lack “elegance”. In my opinion, a long and drawn-out solution can often provide more insight than a concise proof. In applied courses, it may be preferable to skip much of the theoretical exercises entirely or cover them briefly.

The simulation exercises provide the reader with first-hand information on the behavior and performance of some of the theoretical results. Most simulations are performed using MATLAB and R computer code. For ease of access, I have included these codes in this manual. They are also available at the website of the book. I apologize for any omissions or errors in the codes. No endorsement or warranty, express or implied, should be drawn from the listing and/or description of the supplemental code. The number of replications is often fixed at 1,000, or 10,000. The sample size is taken in the range [100, 1,000] to allow for finite-sample statistical inference. Note, that in many cases the computations can be time demanding. The reader should keep in mind that round-off errors can occur – particularly in those computer codes involving many loops. Furthermore, to make the book as usable as possible, all simulation results are presented in the main text. This allows readers to focus directly on the problem at hand.

The empirical exercises are on concrete applications with a wide variety of data sets and topics. These exercises may help the reader to get a better understanding of specific nonlinear features in the underlying DGP. In all cases some background information is provided in the main text so that the general context of the data becomes clear. The solutions manual contains computer codes which should make the reproduction of some presented results possible and also facilitate further applications of the described methods to appropriate other data sets. The data and computer codes that are needed for the empirical exercises can be found at two locations: <http://www.jandegooijer.nl> and <http://extras.springer.com>.

As the intended audience of the book is rather wide, it is difficult to judge the level and background of the reader. Hence, I have not tried to provide a classification of the difficulty of the exercises. Most of the figures in the main text and this solutions manual are obtained using SigmaPlot.<sup>1</sup> In some cases, however, they are created by the computer software package EViews<sup>2</sup> and the R graphics facility. All figures are included in the set of supplementary material for this book. Of course, one can use any other graphics system for plotting data and computational results.

Although this solutions manual has been carefully checked, errors, typos, and inconsistencies in notation are bound to have crept into the text. Any readers spotting such errors, are kindly requested to contact me. Also, I welcome comments about the book and suggestions for improvement.

Amsterdam

*Jan G. De Gooijer*

---

<sup>1</sup>SigmaPlot is a registered trademark of Systat Software, Inc.

<sup>2</sup>EViews (Econometric Views) is a software package for Windows, used mainly for econometric time series analysis. It was developed by Quantitative Micro Software, now a part of IHS.

## Chapter 1

1.1 Suppose that  $X$  and  $Y$  are two random variables. The conditional mean of  $Y$  given  $X = x$  is a function of  $x$ , so that the mean of the conditional expectation is  $\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}[Y]$  (law of total expectation). Using this result, the mean of  $\{Y_t, t \in \mathbb{Z}\}$  is

$$\mathbb{E}[Y_t] = \mathbb{E}\left[\mathbb{E}[Y_t|(Y_{t-1}, Y_{t-2}, \dots)]\right] = \mathbb{E}\left[\sqrt{\alpha_0 + \alpha_1 Y_{t-1}^2} \varepsilon_t\right] = 0.$$

In the last step we used that  $\sigma_t = \sqrt{\alpha_0 + \alpha_1 Y_{t-1}^2}$  is a (nonlinear) function of  $(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$ , which is independent of  $\varepsilon_t$ . The variance of  $\{Y_t, t \in \mathbb{Z}\}$  is

$$\begin{aligned} \text{Var}(Y_t) &= \mathbb{E}[Y_t^2] = \mathbb{E}[\mathbb{E}[Y_t^2|(Y_{t-1}, Y_{t-2}, \dots)]] \\ &= \mathbb{E}[(\alpha_0 + \alpha_1 Y_{t-1}^2) \varepsilon_t^2] = \alpha_0 + \alpha_1 \mathbb{E}[Y_{t-1}^2], \end{aligned}$$

where we used that  $\mathbb{E}[Y_{t-1}^2 \varepsilon_t^2] = \mathbb{E}[Y_{t-1}^2] \mathbb{E}[\varepsilon_t^2]$  since  $\{Y_{t-1}\}$  is a (nonlinear) function of  $(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$ . We conclude that

$$\text{Var}(Y_t) = \alpha_0 + \alpha_1 \text{Var}(Y_{t-1}).$$

Iterative substitution of this result shows that  $\text{Var}(Y_t) = \alpha_0/(1 - \alpha_1)$  for  $0 < \alpha_1 < 1$ . In order to prove that  $\{Y_t, t \in \mathbb{Z}\}$  is a white noise (WN) process we compute the covariances  $\mathbb{E}[Y_t Y_{t-\ell}]$  for  $\ell \geq 1$ .

$$\begin{aligned} \mathbb{E}[Y_t Y_{t-\ell}] &= \mathbb{E}\left[\mathbb{E}[Y_t Y_{t-\ell} | (Y_{t-1}, Y_{t-2}, \dots)]\right] = \mathbb{E}\left[Y_{t-\ell} \mathbb{E}[Y_t | (Y_{t-1}, Y_{t-2}, \dots)]\right] \\ &= \mathbb{E}\left[Y_{t-\ell} \sqrt{\alpha_0 + \alpha_1 Y_{t-1}^2} \varepsilon_t\right] = \mathbb{E}\left[Y_{t-\ell} \sqrt{\alpha_0 + \alpha_1 Y_{t-1}^2}\right] \mathbb{E}[\varepsilon_t] = 0. \end{aligned}$$

So the process  $\{Y_t, t \in \mathbb{Z}\}$  is WN.

In order to prove that  $\{Y_t^2, t \in \mathbb{Z}\}$  follows an AR(1) process, we first show that the second moment of  $\{Y_t^2\}$  (that is, the fourth moment of  $\{Y_t, t \in \mathbb{Z}\}$ ) is constant. The fourth moment of  $\{Y_t, t \in \mathbb{Z}\}$  is

$$\begin{aligned} \mathbb{E}[Y_t^4] &= \mathbb{E}[\mathbb{E}[Y_t^4 | (Y_{t-1}, Y_{t-2}, \dots)]] = \mathbb{E}[\sigma_t^4 \mathbb{E}[\varepsilon_t^4]] = \mathbb{E}[\sigma_t^4] \mathbb{E}[\varepsilon_t^4] \\ &= 3\mathbb{E}[\alpha_0^2 + \alpha_1^2 Y_{t-1}^4 + 2\alpha_0 \alpha_1 Y_{t-1}^2] = 3\alpha_0^2 + 6\alpha_0 \alpha_1 \frac{\alpha_0}{1 - \alpha_1} + 3\alpha_1^2 \mathbb{E}[Y_{t-1}^4]. \end{aligned}$$

By iterative substitution it follows that the fourth moment is constant, provided that  $1 - 3\alpha_1^2 > 0$ , that is,  $3\alpha_1^2 < 1$ .

Now we prove that  $\{Y_t^2, t \in \mathbb{Z}\}$  follows an AR(1) process. There holds

$$Y_t^2 | (Y_{t-1}, Y_{t-2}, \dots) = \sigma_t^2 \varepsilon_t^2 = (\alpha_0 + \alpha_1 Y_{t-1}^2) \varepsilon_t^2,$$

which implies that

$$\mathbb{E}[Y_t^2 | (Y_{t-1}, Y_{t-2}, \dots)] = \alpha_0 + \alpha_1 Y_{t-1}^2 = \sigma_t^2.$$

Let  $\omega_t = Y_t^2 - \sigma_t^2$ , then we show as an auxiliary result that  $\{\omega_t, t \in \mathbb{Z}\}$  is a WN process. There holds

$$\begin{aligned} \mathbb{E}[\omega_t] &= \mathbb{E}[\mathbb{E}[\omega_t | (Y_{t-1}, Y_{t-2}, \dots)]] = \mathbb{E}[\mathbb{E}[Y_t^2 | (Y_{t-1}, Y_{t-2}, \dots)] - \sigma_t^2] \\ &= \mathbb{E}[\sigma_t^2 - \sigma_t^2] = 0 \\ \mathbb{E}[\omega_t^2] &= \mathbb{E}[\mathbb{E}[Y_t^4 | (Y_{t-1}, Y_{t-2}, \dots)] + \sigma_t^4 - 2\sigma_t^2 \mathbb{E}[Y_t^2 | (Y_{t-1}, Y_{t-2}, \dots)]] \\ &= \mathbb{E}[Y_t^4] - \mathbb{E}[\sigma_t^4] = \mathbb{E}[Y_t^4] - \mathbb{E}[\alpha_0^2 + \alpha_1^2 Y_{t-1}^4 + 2\alpha_0 \alpha_1 Y_{t-1}^2], \end{aligned}$$

which is constant because the second and fourth moments of  $\{Y_t, t \in \mathbb{Z}\}$  are constant. The autocorrelations of the process  $\{\omega_t\}$  are zero because, for  $\ell \geq 1$ ,  $\omega_{t-\ell} = Y_{t-\ell}^2 - \sigma_{t-\ell}^2 = Y_{t-\ell}^2 - \alpha_0 - \alpha_1 Y_{t-\ell-1}^2$  is a function of  $(Y_{t-1}, Y_{t-2}, \dots)$  so that

$$\begin{aligned} \mathbb{E}[\omega_t \omega_{t-\ell}] &= \mathbb{E}[\mathbb{E}[\omega_t \omega_{t-\ell} | (Y_{t-1}, Y_{t-2}, \dots)]] = \mathbb{E}[\omega_{t-\ell} \mathbb{E}[\omega_t | (Y_{t-1}, Y_{t-2}, \dots)]] \\ &= \mathbb{E}[\omega_{t-\ell} (\sigma_t^2 - \sigma_t^2)] = 0. \end{aligned}$$

The above results imply that  $Y_t^2 = \sigma_t^2 + \omega_t = \alpha_0 + \alpha_1 Y_{t-1}^2 + \omega_t$  where  $\{\omega_t, t \in \mathbb{Z}\}$  is a WN process. This shows that  $\{Y_t^2, t \in \mathbb{Z}\}$  is an AR(1) process.

If, on the other hand, we assume that the rewrite of  $\{Y_t^2, t \in \mathbb{Z}\}$  gives rise to an AR(1) process, the above proofs can be skipped. In that case, we can express the ARCH(1) process as

$$\begin{aligned} Y_t^2 &= \sigma_t^2 + Y_t^2 - \sigma_t^2 \\ &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \sigma_t^2 (\varepsilon_t^2 - 1) \\ &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \omega_t, \end{aligned}$$

where  $\omega_t = \sigma_t^2 (\varepsilon_t^2 - 1)$ . Thus, we can think of an ARCH(1) process as an AR(1) process for  $\{Y_t^2, t \in \mathbb{Z}\}$  driven by a new noise process  $\{\omega_t, t \in \mathbb{Z}\}$ . Let  $\mathcal{F}_{t-1}$  be the  $\sigma$ -field generated by  $Y_{t-1}, Y_{t-2}, \dots$ . It follows that  $\mathbb{E}(\omega_t | \mathcal{F}_{t-1}) = \mathbb{E}(\sigma_t^2 - \sigma_t^2) = 0$ . Nevertheless, the process  $\{\omega_t, t \in \mathbb{Z}\}$  *does not* have a constant conditional variance. In particular,

$$\mathbb{E}(\omega_t^2 | \mathcal{F}_{t-1}) = \mathbb{E}[\sigma_t^4 (\varepsilon_t^2 - 1)^2 | \mathcal{F}_{t-1}] = \sigma_t^4 \text{Var}(\varepsilon_t^2).$$

- 1.2** (a) The process is described by  $Y_t = \beta Y_{t-2} \varepsilon_{t-1} + \varepsilon_t$ , with  $\{\varepsilon_t\}$  an i.i.d. sequence such that  $\mathbb{E}(\varepsilon_t) = \mathbb{E}(\varepsilon_t^3) = 0$ ,  $\mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$ ,  $\mathbb{E}(\varepsilon_t^4) < \infty$ ,  $\beta^2 \mathbb{E}(\varepsilon_t^4) < 1$ ,  $\varepsilon_0 = 0$ , and  $Y_{-1} = Y_0 = 0$ . As  $\{Y_t, t \in \mathbb{Z}\}$  is a nonlinear function of  $(\varepsilon_t, \varepsilon_{t-1}, \dots)$  it follows that  $\varepsilon_{t-1}$  and  $Y_{t-2}$  are independent so that

$$\mathbb{E}(Y_t) = \beta \mathbb{E}(Y_{t-2} \varepsilon_{t-1}) + \mathbb{E}(\varepsilon_t) = \beta \mathbb{E}(Y_{t-2}) \mathbb{E}(\varepsilon_{t-1}) = 0.$$

For given  $\ell > 0$ , the covariance between  $Y_t$  and  $Y_{t-\ell}$  is therefore equal to

$$\begin{aligned} \mathbb{E}(Y_t Y_{t-\ell}) &= \mathbb{E}[(\beta Y_{t-2} \varepsilon_{t-1} + \varepsilon_t)(\beta Y_{t-\ell-2} \varepsilon_{t-\ell-1} + \varepsilon_{t-\ell})] \\ &= \mathbb{E}[\beta^2 \varepsilon_{t-1} \varepsilon_{t-\ell-1} Y_{t-2} Y_{t-\ell-2} + \beta \varepsilon_{t-1} \varepsilon_{t-\ell} Y_{t-2} \\ &\quad + \beta \varepsilon_t \varepsilon_{t-\ell-1} Y_{t-\ell-2} + \varepsilon_t \varepsilon_{t-\ell}] \\ &= \beta^2 \mathbb{E}(\varepsilon_{t-1}) \mathbb{E}(\varepsilon_{t-\ell-1} Y_{t-2} Y_{t-\ell-2}) + \beta \mathbb{E}(\varepsilon_{t-1}) \mathbb{E}(\varepsilon_{t-\ell} Y_{t-2}) \\ &\quad + \beta \mathbb{E}(\varepsilon_t) \mathbb{E}(\varepsilon_{t-\ell-1} Y_{t-\ell-2}) + \mathbb{E}(\varepsilon_t) \mathbb{E}(\varepsilon_{t-\ell}) \\ &= 0. \end{aligned}$$

So,  $\{Y_t, t \in \mathbb{Z}\}$  is an uncorrelated process.

To investigate whether  $\{Y_t, t \in \mathbb{Z}\}$  is a weak WN process we need to check whether it has a fixed variance (for  $t \geq 1$ ). There holds  $Y_1 = \beta Y_{-1} \varepsilon_0 + \varepsilon_1 = \varepsilon_1$  and  $Y_2 = \beta Y_0 \varepsilon_1 + \varepsilon_2 = \varepsilon_2$ , so that  $\text{Var}(Y_1) = \text{Var}(Y_2) = \sigma_\varepsilon^2$ . Further

$$\begin{aligned} \text{Var}(Y_3) &= \mathbb{E}(Y_3^2) = \mathbb{E}(\beta^2 Y_1^2 \varepsilon_2^2 + \varepsilon_3 \varepsilon_2 Y_1 + \varepsilon_3^2) \\ &= \beta^2 \sigma_\varepsilon^2 \mathbb{E}(Y_1^2) + \sigma_\varepsilon^2 = \beta^2 \sigma_\varepsilon^4 + \sigma_\varepsilon^2. \end{aligned}$$

This is different from  $\text{Var}(Y_2)$ , so  $\{Y_t, t \in \mathbb{Z}\}$  is *not* a WN process.

(b) The process  $\{Y_t^2, t \in \mathbb{Z}\}$  is equal to

$$Y_t^2 = \beta^2 Y_{t-2}^2 \varepsilon_{t-1}^2 + \varepsilon_t^2 + \varepsilon_t \varepsilon_{t-1} Y_{t-2}.$$

This process has expected values

$$\sigma_t^2 = \mathbb{E}(Y_t^2) = \beta^2 \sigma_\varepsilon^2 \sigma_{t-2}^2 + \sigma_\varepsilon^2 = \sigma_\varepsilon^2 (1 + \beta^2 \sigma_{t-2}^2).$$

From part (a) we have  $\sigma_1^2 = \sigma_2^2 = \sigma_\varepsilon^2$ , and for  $t \geq 3$  the variances  $\sigma_t^2$  can be computed recursively from the above formula. The covariance between  $Y_t^2$  and  $Y_{t-2}^2$  is given by

$$\begin{aligned} \text{Cov}(Y_t^2, Y_{t-2}^2) &= \mathbb{E}[(Y_t^2 - \sigma_t^2)(Y_{t-2}^2 - \sigma_{t-2}^2)] \\ &= \mathbb{E}(Y_t^2 Y_{t-2}^2) - \sigma_t^2 \sigma_{t-2}^2 \\ &= \mathbb{E}[(\beta^2 Y_{t-2}^2 \varepsilon_{t-1}^2 + \varepsilon_t^2 + Y_{t-2} \varepsilon_t \varepsilon_{t-1}) Y_{t-2}^2] - \sigma_\varepsilon^2 (1 + \beta^2 \sigma_{t-2}^2) \sigma_{t-2}^2 \\ &= \beta^2 \sigma_\varepsilon^2 \mathbb{E}(Y_{t-2}^4) + \sigma_\varepsilon^2 \sigma_{t-2}^2 - \sigma_\varepsilon^2 (\sigma_{t-2}^2 + \beta^2 \sigma_{t-2}^4) \\ &= \beta^2 \sigma_\varepsilon^2 (\mathbb{E}(Y_{t-2}^4) - \sigma_{t-2}^4). \end{aligned}$$

Let  $Z = Y_{t-2}^2$ , then  $Y_{t-2}^4 = Z^2$  and  $\sigma_{t-2}^4 = (\mathbb{E}(Z))^2$  so that the covariance between  $\{Y_t^2\}$  and  $\{Y_{t-2}^2\}$  is equal to  $\beta^2 \sigma_\varepsilon^2 \mathbb{E}[(Z - \mathbb{E}(Z))^2] = \beta^2 \sigma_\varepsilon^2 \text{Var}(Y_{t-2}^2)$ . Because  $Y_{t-2}^2$  is a random variable with positive variance, we conclude that  $\text{Cov}(Y_t^2, Y_{t-2}^2) > 0$ . This shows that  $\{Y_t^2, t \in \mathbb{Z}\}$  is a correlated process.

As an illustration, Figure 1.1(a) displays a generated subdiagonal time series (see Section 2.2 of the main text) of length  $T = 500$ , with  $\beta = 0.9$  and  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ . There are high amplitude oscillations, comparable to the extreme peaks in the plot of the magnetic field data in Figure 1.3 of the main text. We see that the variance of the process is time-dependent. The sample ACF of  $\{Y_t, t \in \mathbb{Z}\}$  in Figure 1.1(b) supports the result of part (a), with most peaks lying within the 95% large sample confidence interval  $\pm 1.96/\sqrt{500}$  (blue medium dashed lines). However, as we see from Figure 1.1(c), the sample ACF of  $\{Y_t^2, t \in \mathbb{Z}\}$  indicates that the squared series is not an uncorrelated process.

**1.3** Assume that  $\{Y_t, t \in \mathbb{Z}\}$  is fourth-order stationary. The second and fourth moments of an ARCH(1) process were computed in the solution of Exercise 1.1 with the results that

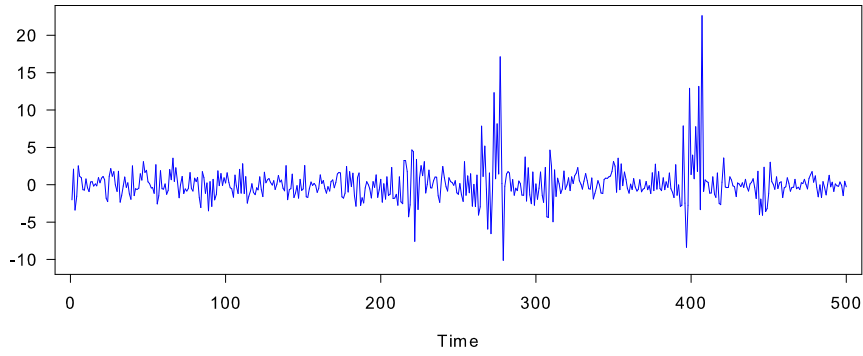
$$\begin{aligned} \mathbb{E}[Y_t^2] &= \frac{\alpha_0}{1 - \alpha_1}, \\ \mathbb{E}[Y_t^4] &= 3\alpha_0^2 + 6\alpha_0\alpha_1 \frac{\alpha_0}{1 - \alpha_1} + 3\alpha_1^2 \mathbb{E}[Y_{t-1}^4] = \frac{3\alpha_0^2(1 + \alpha_1)}{1 - \alpha_1} + 3\alpha_1^2 \mathbb{E}[Y_{t-1}^4]. \end{aligned}$$

By repetitive substitution of the lagged term  $\mathbb{E}[Y_{t-1}^4]$  it follows that

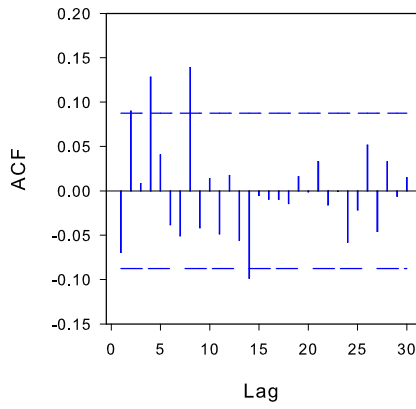
$$E[Y_t^4] = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)},$$

provided that  $1 - 3\alpha_1^2 > 0$ , that is,  $0 < \alpha_1 < \frac{1}{\sqrt{3}}$ . Thus,  $\gamma_Y(0) = \text{Var}(Y_t) = \alpha_0/(1 - \alpha_1)$ , and  $\gamma_Y(\ell) = 0$  for  $\ell \neq 0$ .

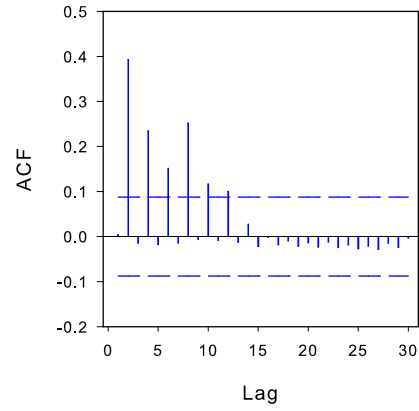
(a) A realization from a BL(0,0,2,1) model



(b) ACF series



(c) ACF squared series



**Figure 1.1:** (a) Realization of a subdiagonal BL(0, 0, 2, 1) process with  $\beta = 0.9$ ; (b) Sample ACF of  $\{Y_t\}_{t=1}^{500}$  with 95% asymptotic confidence limits (blue medium dashed lines); (c) Sample ACF of  $\{Y_t^2\}_{t=1}^{500}$ .

Now,

$$\begin{aligned} TCov(\hat{\gamma}_Y(h), \hat{\gamma}_Y(k)) &= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}(Y_t Y_{t+h} Y_s Y_{s+k}) - T \gamma_Y(h) \gamma_Y(k) \\ &= \sum_{-T < i < T} \left(1 - \frac{|i|}{T}\right) \sigma(i), \quad (0 \leq h \leq k < T), \end{aligned}$$

where

$$\sigma(i) = \text{Cov}(Y_1 Y_{1+h}, Y_{1+i} Y_{1+i+k}).$$

It is straightforward to see that

$$\begin{aligned} \sigma(0) &= \text{Var}(Y_1 Y_2) = \mathbb{E}(Y_1^2 Y_2^2) = \mathbb{E}[Y_1^2 (\alpha_0 + \alpha_1 Y_1^2) \varepsilon_2^2] \\ &= \alpha_0 \mathbb{E}(Y_1^2) + \alpha_1 \mathbb{E}(Y_1^4) \\ &= \frac{\alpha_0^2 (1 + 3\alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}. \end{aligned}$$

Since  $\sigma(i) = \text{Cov}(Y_1 Y_2, Y_{1+i} Y_{2+i}) = 0$  for  $i \geq 1$ , we obtain

$$T\gamma_Y^{-2}(0)\text{Var}(\hat{\gamma}_Y(1)) = \frac{(1 - \alpha_1)(1 + 3\alpha_1)}{1 - 3\alpha_1^2}.$$

Hence,  $\lim_{T \rightarrow \infty} T\gamma_Y^{-2}(0)\text{Var}(\hat{\gamma}_Y(1))$  increases as  $\alpha_1$  increases from 0 to  $1/\sqrt{3} \approx 0.5774$ . This result depends simultaneously on  $\alpha_1$  and  $\mathbb{E}(Y_1^4)$ .

**Remark:** As an additional example, consider the BL process  $Y_t = \beta Y_{t-2}\varepsilon_{t-1} + \varepsilon_t$ , where  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$  with  $\beta^2 \sigma_\varepsilon^2 < 1$ . In that case we have (cf. Exercise 4.2)

$$\begin{aligned} \sigma(0) &= \frac{\sigma_\varepsilon^4(1 + 2\beta^2\sigma_\varepsilon^2)}{(1 - \beta^2\sigma_\varepsilon^2)^2}, & \sigma(1) = \sigma(2) = 0, \\ \sigma(3) &= \frac{\beta^2\sigma_\varepsilon^6}{1 - \beta^2\sigma_\varepsilon^2}, & \sigma(i) = 0 \quad \text{for } i \geq 4. \end{aligned}$$

It is easy to see that  $\lim_{T \rightarrow \infty} T\gamma_Y^{-2}(0)\text{Var}(\hat{\gamma}_Y(1))$  is greater than 2 as  $|\beta| \geq 0.55$  with  $\gamma_Y(0) = \sigma_\varepsilon^2/(1 - \beta^2\sigma_\varepsilon^2)$ .

- 1.4** (a) Let  $f_{X,Y}(x, y)$ ,  $f_X(x)$ , and  $f_Y(y)$  denote the joint and marginal densities, respectively. Using the fact that  $\iint f_{X,Y}(x, y) dx dy = \int f_X(x) dx = \int f_Y(y) dy = 1$ , rewrite (1.18) as

$$\begin{aligned} I^{\text{KL}}(X, Y) &= \iint \left\{ (f_X(x)f_Y(y) - f_{X,Y}(x, y))/f_{X,Y}(x, y) \right. \\ &\quad \left. - \log \left( 1 + (f_X(x)f_Y(y) - f_{X,Y}(x, y))/f_{X,Y}(x, y) \right) \right\} f_{X,Y}(x, y) dx dy \\ &= \iint q(z) dx dy, \end{aligned}$$

where  $z = (f_X(x)f_Y(y) - f_{X,Y}(x, y))/f_{X,Y}(x, y)$  and  $q(z) = z - \log(1 + z)$  for  $z > -1$ . By examining the first derivative of  $q(\cdot)$ , it follows that  $q(\cdot)$  is non-negative and is equal to 0 if and only if  $z = 0$ . As  $f_{X,Y}(x, y)$  is non-negative and  $\iint f_{X,Y}(x, y) dx dy = 1$  then  $I^{\text{KL}}(X, Y)$  is non-negative and is 0 if and only if  $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ , which is the necessary and sufficient condition for  $X$  and  $Y$  to be independent.

The above results follow immediately from Theorem 3.1 of Kullback (1959, p. 14); see also Kullback et al. (1989). Or, since

$$\mathbb{E}\left(\frac{f_X(x)f_Y(y)}{f_{X,Y}(x, y)}\right) = \int f_{X,Y}(x, y) \frac{f_X(x)f_Y(y)}{f_{X,Y}(x, y)} dx dy = 0,$$

apply Jensen's inequality

$$\mathbb{E}\left(\log \frac{f_X(x)f_Y(y)}{f_{X,Y}(x, y)}\right) \leq 0,$$

with equality iff  $\mathbb{P}\left(\frac{f_X(x)f_Y(y)}{f_{X,Y}(x, y)} = 1\right) = 1$ .

- (b) Reformulate the problem as follows:  $f_{X,Y}(x, y) = f_{X,Y}(h(y), y) = g(y)$  for  $|x -$



$y| < \epsilon$  and 0 elsewhere ( $\epsilon > 0$ ). The marginal density for  $X$  is given by

$$\begin{aligned} f_X(x) &= \int f_{X,Y}(x,y)dy \\ &= \int_{x-\epsilon}^{x+\epsilon} g(y)dy \\ &= G(x+\epsilon) - G(x-\epsilon), \end{aligned}$$

where  $G(u) = \int_{-\infty}^u g(u)du$ . Note that

$$\lim_{\epsilon \rightarrow 0} (G(x+\epsilon) - G(x-\epsilon)) = \lim_{\epsilon \rightarrow 0} \int_{x-\epsilon}^{x+\epsilon} g(y)dy = 0.$$

Similarly, the marginal density for  $Y$  is given by

$$f_Y(y) = \int_{y-\epsilon}^{y+\epsilon} g(y)dx = 2\epsilon g(y).$$

Then

$$\lim_{\epsilon \rightarrow 0} \left( \frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)} \right) = \lim_{\epsilon \rightarrow 0} \left( \frac{1}{2\epsilon(G(x+\epsilon) - G(x-\epsilon))} \right) \rightarrow \infty,$$

which implies that  $I^{\text{KL}}(X, Y) \rightarrow \infty$  as  $\epsilon \rightarrow 0$ .

**1.5** (a) It follows from the independence result that

$$\mathbb{E}\left(\widehat{\mu}_{\nu,Y}^p \widehat{\mu}_{2,Y}^{-\nu p/2}\right) \mathbb{E}\left(\widehat{\mu}_{2,Y}^{\nu p/2}\right) = \mathbb{E}\left(\widehat{\mu}_{\nu,Y}^p\right), \quad (p = 1, 2, \dots),$$

so that the mean  $\mathbb{E}(\widehat{\tau}_Y)$  can be calculated from  $\mathbb{E}(\widehat{\mu}_{3,Y})$  and  $\mathbb{E}(\widehat{\mu}_{2,Y}^{3/2})$ . It is well known that for i.i.d. data, the mean of the third central moment is given by

$$\mathbb{E}(\widehat{\mu}_{3,Y}) = \frac{(n-1)(n-2)}{n^2} \mu_{3,Y}.$$

Since  $\mu_{3,Y} = 0$  for any symmetric distribution, we have  $E(\widehat{\tau}_Y) = 0$ . Further, the exact expression for the variance of  $\widehat{\tau}_Y$  is given by

$$\begin{aligned} \text{Var}(\widehat{\tau}_Y) &= \text{Var}\left(\frac{(n-2)}{\sqrt{n(n-1)}} \frac{k_3}{k_2^{3/2}}\right) \\ &= \frac{(n-2)^2}{n(n-1)} \frac{6n(n-1)}{(n-2)(n+1)(n+3)} = \frac{6(n-2)}{(n+1)(n+3)}. \end{aligned}$$

Similar as above the mean  $\mathbb{E}(\widehat{\kappa}_Y)$  can be calculated from  $\mathbb{E}(\widehat{\mu}_{2,Y}^2)$  and  $\mathbb{E}(\widehat{\mu}_{4,Y})$ . It is well known that  $\text{Var}(\bar{Y}) = \mu_{2,Y}/n$ . As given, and without loss of generality,  $\mu_Y = 0$ . Then we have

$$\mathbb{E}(\widehat{\mu}_{2,Y}) = \mathbb{E}\left(\frac{1}{n} \sum_i (Y_i - \bar{Y})^2\right) = \frac{1}{n} \mathbb{E}\left(\sum_i Y_i^2\right) - \mathbb{E}(\bar{Y}^2) = \frac{n-1}{n} \mu_{2,Y}.$$

We further have

$$\widehat{\mu}_{2,Y}^2 = \left(\frac{1}{n} \sum_i Y_i^2\right)^2 - 2\bar{Y}^2 \left(\frac{1}{n} \sum_i Y_i^2\right) + \bar{Y}^4.$$

Then, with  $\mu_Y = 0$ , we have

$$\begin{aligned}\frac{1}{n^2}\mathbb{E}\left(\sum_i Y_i^2\right)^2 &= \frac{\mu_{4,Y} + (n-1)\mu_{2,Y}^2}{n}, \\ \frac{1}{n^3}\mathbb{E}\left[\left(\sum_i Y_i\right)^2 \sum_i Y_i^2\right] &= \frac{\mu_{4,Y} + (n-1)\mu_{2,Y}^2}{n^2}, \\ \mathbb{E}(\bar{Y}^4) &= \frac{\mu_{4,Y} + 3(n-1)\mu_{2,Y}^2}{n^3}.\end{aligned}$$

Hence, on combining, we have

$$\mathbb{E}(\hat{\mu}_{2,Y}^2) = \mu_{2,Y}^2 + \frac{\mu_{4,Y} - 3\mu_{2,Y}^2}{n} - \frac{2\mu_{4,Y} - 5\mu_{2,Y}^2}{n^2} + \frac{\mu_{4,Y} - 3\mu_{2,Y}^2}{n^3}.$$

Rewriting  $\mathbb{E}(\hat{\mu}_{2,Y}^2)$ , and using similar calculations for  $\hat{\mu}_{4,Y}$ , we have

$$\begin{aligned}\mathbb{E}(\hat{\mu}_{2,Y}^2) &= \frac{n-1}{n^3}\mu_{2,Y}^2\left((n-1)\frac{\mu_{4,Y}}{\mu_{2,Y}^2} + (n^2 - 2n + 3)\right) \\ \mathbb{E}(\hat{\mu}_{4,Y}) &= \frac{n-1}{n^3}\mu_{2,Y}^2\left((n^3 - 3n + 3)\frac{\mu_{4,Y}}{\mu_{2,Y}^2} + 3(2n - 3)\right).\end{aligned}$$

Since  $\mu_{4,Y}/\mu_{2,Y}^2 = 3$  for a Gaussian distribution, we have the exact result

$$\mathbb{E}(\hat{\kappa}_Y) = \frac{\mathbb{E}(\hat{\mu}_{4,Y})}{\mathbb{E}(\hat{\mu}_{2,Y}^2)} = \frac{3n(n-1)}{n(n+1)}.$$

Clearly,  $\hat{\kappa}_Y$  is affected with a positive bias of order  $n^{-1}$ , while  $\hat{\tau}_Y$  is unbiased. The exact variance of  $\hat{\kappa}_Y$  follows directly from the expressions for  $k_2$ ,  $k_4$ , and  $\text{Var}(k_4 k_2^{-2})$ , i.e.,

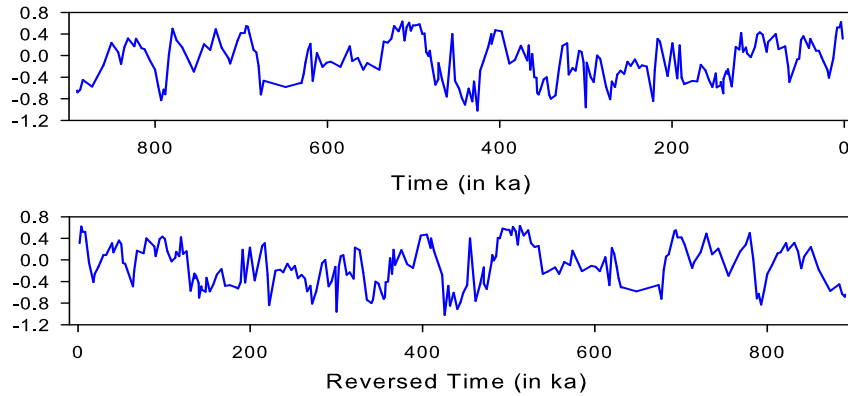
$$\begin{aligned}\text{Var}(\hat{\kappa}_Y) &= \text{Var}\left(\frac{(n-1)(n-2)(n-3)k_4}{n^2(n+1)\left(\frac{n-1}{n}k_2\right)^2} + \frac{3n^2(n-1)}{n^2(n+1)}\right) \\ &= \text{Var}\left(\frac{(n-2)(n-3)k_4}{(n-1)(n+1)k_2^2}\right) \\ &= \frac{(n-2)^2(n-3)^2}{(n-1)^2(n+1)^2} \frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)} \\ &= \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}.\end{aligned}$$

Note that for  $n \rightarrow \infty$  we have the asymptotic results  $\mathbb{E}(\hat{\tau}_Y) \rightarrow 0$ ,  $\text{Var}(\hat{\tau}_Y) \rightarrow 6$ ,  $\mathbb{E}(\hat{\kappa}_Y) \rightarrow 3$ , and  $\text{Var}(\hat{\kappa}_Y) \rightarrow 24$ ; confirming (1.5)

- (b) Clearly, using the results in part (a), we can define an alternative (denoted by the letter ‘‘a’’) JB test statistic as

$$\text{JB}^a = \frac{\hat{\tau}_Y^2}{\text{Var}(\hat{\tau}_Y)} + \frac{(\hat{\kappa}_Y - \mathbb{E}(\hat{\kappa}_Y))^2}{\text{Var}(\hat{\kappa}_Y)}.$$

Urzúa (1996) shows that the test statistic  $\text{JB}^a$  behaves much better for small- and medium-sized samples than the JB test statistic.



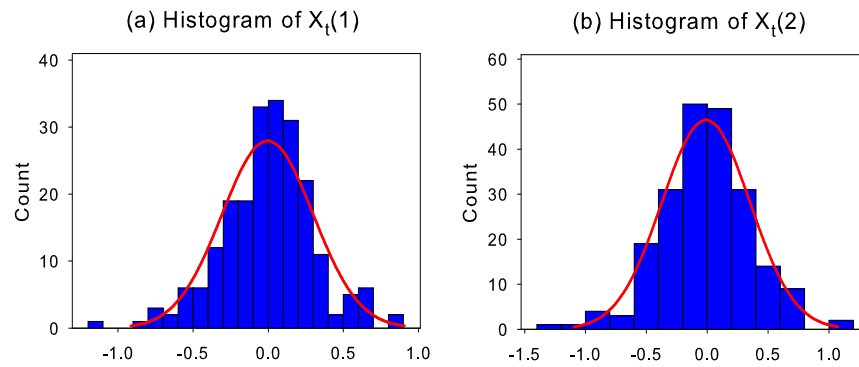
**Figure 1.2:** Climate change data set. Time plot of the  $\delta^{13}\text{C}$  series (top), and the reversed data plot (bottom).

- 1.6** With both time series the number of recurrences are homogeneously spread over the plot, which is typical of a stationary time series process. Both series display a quasi-periodic pattern, as indicated by the long diagonal lines with different distances parallel to and on the main diagonal. Comparing, however, the off-diagonal structure of the two series, we see that the evolution of states is quite different between the series. There is no evidence for drifts in the series, such as a time trend. A DGP with a slowly time-varying parameter, will result in a recurrence plot with white areas, or blanks, in the upper-left and lower-right corners.
- 1.7** The time series plot in Figure 1.9(a) of the main text shows a stationary series with some positive skewness. The directed scatter plot captures the positive skewness in the data through the sojourns in top right and top left corners. Therefore, the time series seems to be generated process by the stationary BL process (ii). In contrast, both plots in Figure 1.9(b) show a realization of  $Y_t = \varepsilon_t$ . This is a completely uncorrelated time series and so the directed scatter plot should tend towards circular noise. The time series wanders around a mean zero, with no real clustering of observations. The time series plot in Figure 1.9(c) shows that this particular realization has a transitional period with much larger variances than we see with a stationary process. The directed scatter plot can be interpreted as an attractor located at  $(Y_t, Y_{t-1}) = (0, 0)$ . The nonstationary variance can be viewed as this attractor being “switched” on and off.
- 1.8** (a) Both plots are displayed in the first panel of Figure 1.2. Clearly, the mirror image (bottom panel) of the  $\delta^{13}\text{C}$  series ( $Y_t$ ) is not similar to the original plot (top panel). Hence, the time series  $\{Y_t\}$  is *not* time-reversible. Note the large glacial-interglacial amplitudes.
- (b) The sample mean (sample standard deviation) of the series  $\{X_t(\ell) = Y_t - Y_{t-\ell}\}$  ( $\ell = 1, 2$ ) are given by  $-0.004$  (0.304) and  $-0.011$  (0.362). Figures 1.3(a) and (b) show histograms of  $\{X_t(\ell)\}$  ( $\ell = 1, 2$ ) with superimposed Gaussian distributions using sample means and standard deviations of the two series. The skewness (kurtosis) of  $X_t(1)$  and  $X_t(2)$  are, respectively,  $-0.213$  (4.003) and  $-0.164$  (3.747). If the process  $\{Y_t, t \in \mathbb{Z}\}$  is time-reversible (TR), the distri-

bution of the stochastic process  $\{X_t(\ell) = Y_t - Y_{t-\ell}\}$  is symmetric about zero; see Chapter 8 of the main text for more details. Clearly, both histograms indicate that this is indeed the case. Later, in Chapter 8, we will see that the null hypothesis of TR for the climate change series  $\delta^{13}\text{C}$  (denoted by  $Y_t$ ) cannot be rejected by a number of formal test statistics. The superimposed Gaussian distributions indicate that both time series  $X_t(\ell)$  are normally distributed.

- (c) Assuming *independence*, the JB test statistic has a  $p$ -value of 0.005 ( $\text{JB} = 10.634$ ) for the series  $\{X_t(1)\}_{t=1}^{215}$ , and a  $p$ -value of 0.052 for the series  $\{X_t(2)\}_{t=1}^{214}$  ( $\text{JB} = 5.932$ ). Hence, for  $\{X_t(1)\}$  normality is rejected at the conventional 5% significance level while the series  $\{X_t(2)\}$  is a “borderline” case. The  $p$ -values of the GJB test statistic *for weakly dependent data* confirm these observations, i.e.  $p$ -value GJB = 0.005 for the  $\{X_t(1)\}$  series (GJB = 10.544) and  $p$ -value GJB = 0.058 for the  $\{X_t(2)\}$  series (GJB = 5.710). These test results can be computed by means of the following MATLAB code.

```
function[JB, pvalueJB, GJB, pvalueGJB] = JB_GJBtest(data)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 1.8(c)
% File: JB_GJBtest.m
%
% Performs a Jarque-Bera (JB) (independent time series data)
% and a GJB (weakly dependent time series data) test for normality.
%
% INPUT:  data
% OUTPUT: JB, pvalueJB, GJB, pvalueGJB
%
% Reference:
% Jarque, C.M. and Bera, A.K. (1987).
%   A test for normality of observations and regression residuals.
%   International Statistical Review, 55(2), 163-172.
%   DOI: 10.2307/1403192.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
nr = length(data);
val = skewness(data)^2;
val = val + (1/4)*(kurtosis(data)-3)^2;
JB = (nr/6)* val;
pvalueJB = 1-chi2cdf(JB,2);
Z = data;
Z = Z(:)'-mean(Z);
y = filter(Z(nr:-1:1),1,Z);
acvf = y(nr:-1:1)/nr;
sum3 = 0;
sum4 = 0;
for i = 2:nr-1
    sum3 = sum3 + 2*acvf(i)^3;
    sum4 = sum4 + 2*acvf(i)^4;
end
F3 = acvf(1)^3+sum3;
F4 = acvf(1)^4+sum4;
```



**Figure 1.3:** Climate change data set. Histograms of the  $\{X_t(\ell) = Y_t - Y_{t-\ell}\}$  time series ( $\ell = 1, 2$ ) where  $\{Y_t\}_{t=1}^{216}$  denotes the  $\delta^{13}\text{C}$  time series.

**Table 1.1:** Climate change data set. Results of the Bai–Ng (2005) test statistics for the time series  $\{X_t(\ell) = Y_t - Y_{t-\ell}\}$  ( $\ell = 1, 2$ );  $p$ -values in parentheses. Test statistics are computed with a Parzen window, and  $\{Y_t\}_{t=1}^{216}$  denotes the  $\delta^{13}\text{C}$  time series.

Test statistic	$\hat{\pi}_{3,X}$ (Skewness)	$\hat{\pi}_{4,X}$ (Kurtosis)	$\hat{\pi}_{34,X}$ (Normality)
$X_t(1)$	-1.097 (0.864)	1.124 (0.131)	1.904 (0.386)
$X_t(2)$	-0.759 (0.776)	1.223 (0.111)	3.884 (0.143)

```

m2 = moment(data,2);
m3 = moment(data,3)^2/F3;
GJB = m3 + (1/4)*(moment(data,4)-3*m2*m2)^2/F4;
GJB = (nr/6)*GJB;
pvalueGJB = 1-chi2cdf(GJB,2);

```

**Remark:** The exercise can be extended by computing the three tests statistics proposed by Bai and Ng (2005) for weakly dependent data; see Section 1.3 of the main text. One way to obtain these results is by using the MATLAB codes given below. For the time series  $\{X_t(\ell)\}$  ( $\ell = 1, 2$ ) the test results, with  $p$ -values in parentheses, are summarized in Table 1.1. We see that in all cases the corresponding null hypotheses cannot be rejected at the 5% nominal significance level.

```

function [pi34,pvalue] = Normality_pi34(data)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 1.8(c)
% File: Normality_pi34.m
%
% Bai-Ng (JBES, 2005) Normality test statistics ``pi34''.
%
% INPUT: data
% OUTPUT: pi34, pvalue

```

```

%
% Reference:
% Bai, J. and Ng, S. (2005).
% Tests for skewness, kurtosis, and normality for time series
% data.
% Journal of Business & Economic Statistics, 23(1), 49-60.
% DOI: 10.1198/073500104000000271.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xx = data;
z = xx-mean(xx);
n = length(xx);
m2 = sum(z.*z)/(n-1);
m3 = sum(z.*z.*z)/(n-1);
m4 = sum(z.*z.*z.*z)/(n-1);
a = zeros(2,4);
a(1,1) = -3*m2;
a(2,2) = -6*m2;
a(1,3) = 1.0;
a(2,4) = 1.0;
% Start computing test statistic for normality:
vv = [z, (z.*z-m2), z.*z.*z, (z.*z.*z.*z-3*m2^2)];
[nsize,nreg] = size(vv);

% Find the LS approximation solution with the QR decomposition
for i = 1:nreg
    A = vv(2:nsize,i);
    y = vv(1:nsize-1,i);
    if issparse(A), R = qr(A);
        else R = triu(qr(A));
    end;
    be(i) = R\(R\'(A'*y));
    r = y-A*be(i);
    rrr = r'*r/nsize;
    se_b(i) = rrr;
end

beta = zeros(nreg,nreg);
bot = 0; top = 0;
for i = 1:nreg
    top = top+4*(be(i)^2)*se_b(i)^2/(1-be(i))^8;
    bot = bot+se_b(i)^2/((1-be(i))^4);
end
alpha = top/bot;
k = ceil(2.6614*(alpha*nsize)^(0.2));
if k > n/2; k = n/2;
end

% Variance-covariance matrix
vcv = vv'*vv/(n-1);

```

```

for i = 1:k
    x = i/k;
    if abs(x) >= 0 && abs(x) <= .5;    % Weights Parzen window
        w = 1-6*(abs(x))^2+6*abs(x)^3;
    else
        w = 2*(1-abs(x))^3;
    end;
    cov = vv(i+1:nsize, :)'*vv(1:nsize-i, :)/(n-1);
    vcv = vcv+w*cov;
    cov = vv(1:nsize-i, :)'*vv(i+1:nsize, :)/(n-1);
    vcv = vcv+w*cov;
end
d      = inv(eye(nreg)-beta');
omega = d*vcv*d';
se     = a*omega*a';
stat   = [m3, (m4-3*m2^2)]';
pi34   = stat'*inv(se)*stat*n; % end computing normality test
pvalue = 1-chi2cdf(pi34,2);

function [pi3,pvalue] = Skewness_pi3(data)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 1.8(c)
% File: Skewness_pi3.m
% Bai-Ng (JBES, 2005) Skewness test statistics "pi3".
%
% INPUT:  data
% OUTPUT: pi3, pvalue
%
% Reference:
% Bai, J. and Ng, S. (2005).
% Tests for skewness, kurtosis, and normality for time series
% data.
% Journal of Business & Economic Statistics, 23(1), 49-60.
% DOI: 10.1198/073500104000000271.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xx = data;
z   = xx-mean(xx);
n   = length(xx);
m2  = sum(z.*z)/(n-1);
m3  = sum(z.*z.*z)/(n-1);

% Start computing test statistic for skewness
a    = ones(2,1);
a(2,1) = -3*m2;
vv   = [z.*z.*z, z];
[nsize,nreg] = size(vv);

beta = zeros(nreg,nreg);
% Find the LS approximation solution with the QR decomposition

```

```

for i = 1:nreg
    A = vv(2:nsize,i);
    y = vv(1:nsize-1,i);
    if issparse(A),R = qr(A);
        else R = triu(qr(A));
    end
    b(i) = R\R\'(A'*y);
    r = y-A*b(i);
    rrr = r'*r/nsize;
    se_b(i) = rrr;
end

bot = 0; top = 0;
for i = 1:nreg
    top = top+4*(b(i)^2)*se_b(i)^2/(1-b(i))^8;
    bot = bot+se_b(i)^2/((1-b(i))^4);
end

alpha = top/bot;
k = ceil(2.6614*(alpha*nsize)^(0.2));
if k > n/2; k = n/2;
end

% Variance-covariance matrix
vcv = vv'*vv/(n-1);
for i = 1:k
    x = i/k;
    if abs(x) >= 0 && abs(x) <= .5; % Weights Parzen window
        w = 1-6*(abs(x))^2+6*abs(x)^3;
    else
        w = 2*(1-abs(x))^3;
    end
    cov = vv(i+1:nsize,:)*vv(1:nsize-i,:)/(n-1);
    vcv = vcv+w*cov;
    cov = vv(1:nsize-i,:)*vv(i+1:nsize,:)/(n-1);
    vcv = vcv+w*cov;
end
d = inv(eye(nreg)-beta');
omega = d*vcv*d';
se = sqrt(a'*omega*a)/sqrt(n);
pi3 = m3/se; % end computing test statistic for skewness
pvalue = 1-normcdf(pi3,0,1);

function[pi4,pvalue] = Kurtosis_pi4(data)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 1.8(c)
% File: Kurtosis_pi4.m
%
% Bai-Ng (JBES, 2005) Kurtosis test statistics ``pi4''.

```



```

%
% INPUT:  data
% OUTPUT: pi4, pvalue
%
% Reference:
% Bai, J. and Ng, S. (2005).
% Tests for skewness, kurtosis, and normality for time series
% data.
% Journal of Business & Economic Statistics, 23(1), 49-60.
% DOI: 10.1198/073500104000000271.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xx = data;
z = xx-mean(xx);
n = length(xx);
m2 = sum(z.*z)/(n-1);
m3 = sum(z.*z.*z)/(n-1);
m4 = sum(z.*z.*z.*z)/(n-1);
kappa = 3;
a = ones(3,1);
a(1,1) = 1;
a(2,1) = -4*m3;
a(3,1) = -2*m2*kappa;
vv = [(z.*z.*z.*z-m2^2),z,(z.*z-m2)];
[nsize,nreg] = size(vv);
beta = zeros(nreg,nreg);

% Find the LS approximation solution with the QR decomposition
for i = 1:nreg
    A = vv(2:nsize,i);
    y = vv(1:nsize-1,i);
    if issparse(A),R = qr(A);
        else R = triu(qr(A));
    end
    b(i) = R\'(R\'(A'*y));
    r = y-A*b(i);
    rrr = r'*r/nsize;
    se_b(i) = rrr;
end

bot = 0; top = 0;
for i = 1:nreg
    top = top+4*(b(i)^2)*se_b(i)^2/(1-b(i))^8;
    bot = bot+se_b(i)^2/((1-b(i))^4);
end

alpha = top/bot;
k = ceil(2.6614*(alpha*nsize)^(0.2));
if k > n/2; k = n/2;
end

```

```

% Variance-covariance matrix
vcv = vv'*vv/(n-1);
for i = 1:k
    x = i/k;
    if abs(x) >= 0 && abs(x) <= .5; % Weights Parzen window
        w = 1-6*(abs(x))^2+6*abs(x)^3;
    else
        w = 2*(1-abs(x))^3;
    end
    cov = vv(i+1:nsize,:)'*vv(1:nsize-i,:)/(n-1);
    vcv = vcv+w*cov;
    cov = vv(1:nsize-i,:)'*vv(i+1:nsize,:)/(n-1);
    vcv = vcv+w*cov;
end
d = inv(eye(nreg)-beta');
omega = d*vcv*d';
se = sqrt(a'*omega*a/(m2^4))/sqrt(n);
stat3 = m4/(m2^2);
pi4 = (stat3-kappa)/se; % end computing kurtosis test
pvalue = 1-normcdf(pi4,0,1);

```

## Chapter 2

2.1 Consider as an example the BL process in (2.18). It may be rewritten as

$$Y_t = 0.99Y_{t-1} + \omega_{t-2} - 0.5Y_{t-1}\omega_{t-2},$$

where  $\omega_t = \varepsilon_{t+1}$ . Thus, in terms of  $\{\omega_t\}$ , the model is redefined as a superdiagonal model. Similarly, consider the BL process in (2.17). Rewriting it as

$$Y_t = 0.99Y_{t-1} + \omega_{t-3} - 0.5Y_{t-2}\omega_{t-3},$$

where  $\omega_t = \varepsilon_{t+2}$ . Clearly, by advancing  $\varepsilon_t$ , we obtain a superdiagonal BL model. Of course, this operation is not physically realizable in real time.

The above result holds for general BL models. For a diagonal, taking  $L = 1$ , will result in a superdiagonal BL model of order  $(p, q + 1, P, Q + 1)$  with the MA coefficients  $\theta_j$  replaced by  $\theta_{j+1}$ , and the BL coefficients  $\psi_{jv}$  by  $\psi_{j,v+1}$  in (2.12). For the subdiagonal BL model, taking  $L = P$ , replacing  $\theta_j$  by  $\theta_{j+P}$  in (2.12), and  $\psi_{jv}$  by  $\psi_{j,v+P}$  yields a superdiagonal BL model of order  $(p, q + P, P, Q + P)$ .

2.2 Write

$$\varepsilon_t = Y_t - \sum_{i=1}^p \{\phi_i + \xi_i \exp(-\gamma Y_{t-1}^2)\} Y_{t-i} - \sum_{j=1}^q \{\theta_j + \tau_j \exp(-\gamma Y_{t-1}^2)\} \varepsilon_{t-j}.$$

Recursive substitution gives

$$\begin{aligned} \varepsilon_t = & Y_t + \sum_{k=1}^m \left\{ \sum_{i=1}^p \sum_{C_{r,i}^k \in \mathcal{F}_{1,i}(k)} \sum_{u \in C_{r,i}^k} (-1)^{r+1} \times \right. \\ & \prod_{s=1}^r A_i \left( t-1 - \sum_{l=1}^{s-1} j_l(u) \right) B_{j_i(u)} \left( t-1 - \sum_{l=1}^{s-1} j_l(u) \right) \\ & + \sum_{C_r^k \in \mathcal{F}_2(k)} \sum_{u \in C_r^k} (-1)^r \prod_{s=1}^r B_{j_s(u)} \left( t-1 - \sum_{l=1}^{s-1} j_l(u) \right) \left. \right\} Y_{t-k} \\ & + \left\{ \sum_{C_r^m \in \mathcal{F}_2(m)} \sum_{u \in C_r^m} (-1)^r \prod_{s=1}^r B_{j_s(u)} \left( t-1 - \sum_{l=1}^{s-1} j_l(u) \right) \right\} \varepsilon_{t-m}, \end{aligned}$$

where  $m \geq 1$  is an integer. Here

$$\begin{aligned} A_i(t-k) &= \phi_i + \xi_i \exp(-\gamma Y_{t-k}^2), \quad B_j(t-k) = \theta_j + \tau_j \exp(-\gamma Y_{t-k}^2), \\ C_{r,i}^k &= \left\{ (j_1, \dots, j_r) \in J^r : i + \sum_{s=1}^r j_s = k \right\}, \\ C_r^k &= \left\{ (j_1, \dots, j_r) \in J^r : \sum_{s=1}^r j_s = k \right\}, \\ \mathcal{F}_{1,i}(k) &= \{C_{r,i}^k : r \in \mathbb{Z}^+\}, \quad \mathcal{F}_2(k) = \{C_r^k : r \in \mathbb{Z}^+\}, \quad J = \{1, \dots, q\} \subset \mathbb{Z}^+. \end{aligned}$$

By a simple application of the strong law of large numbers and Jensen's inequality, it can be shown that the process is absolutely convergent almost surely, provided  $\max_{1 \leq j \leq q} (|\theta_j| + |\tau_j|) < 1$ . Moreover, the last term in curly brackets converges to zero in probability. This result can be used to prove that the process  $\{Y_t, t \in \mathbb{Z}\}$  is invertible; see also Section 3.5 of the main text.

- 2.3** (a) Consider first case (ii), with  $Y_t > 0$ , so the system starts in the unstable regime. Since  $Y_{t+1} = \phi_2 Y_t + \varepsilon_{t+1}$ , with  $\phi_2 \leq -1$ , it follows that  $Y_{t+1} > 0$  if  $\varepsilon_{t+1} > -\phi_2 Y_t > 0$ , which happens with probability  $p_{t+1} < (1/2)$ , because  $\{\varepsilon_t\}$  has a symmetric distribution around 0. Hence, the probability that the system stays in the nonstationary regime for  $T$  periods is, because  $\{\varepsilon_t\}$  is i.i.d.,  $p_{t+1} \times p_{t+2} \times \dots \times p_{t+T} < (1/2)^T$ , which clearly goes to zero as  $T \rightarrow \infty$ . Exactly the same applies to case (iii): if  $Y_t \leq 0$ , then  $Y_{t+1} \leq 0$  only if  $\varepsilon_{t+1} < 0 < \phi_1 Y_t$ , with probability  $p_{t+1} < (1/2)$ . So again, the probability that the system stays in the nonstationary regime for  $T$  periods is bounded by  $(1/2)^T$ .
- (b) This is most easily seen if we start at  $Y_1 = y \gg 0$ , so that  $\{\varepsilon_t\}$  is negligibly small relative to  $\phi_t Y_{t-1}$ , with  $\phi_t = \phi_1 I(Y_{t-1} \leq 0) + \phi_2 I(Y_{t-1} > 0)$ . As long as this is the case, the sign of  $\{Y_t\}$  will be equal to the sign of  $\phi_t Y_{t-1}$ , and because  $\phi_t < 0 \forall t$  this means that the process  $\{Y_t\}$  will alternate between positive and negative values, and hence  $\phi_t$  will alternate between  $\phi_1$  and  $\phi_2$ . This implies

$$\begin{aligned} Y_t &= \phi_t \phi_{t-1} Y_{t-2} + \varepsilon_t + \phi_t \varepsilon_{t-1} \\ &= \phi_1 \phi_2 Y_{t-2} + \varepsilon_t + \varepsilon_{t-1} \\ &= Y_{t-2} + \varepsilon_t + \phi_t \varepsilon_{t-1}, \end{aligned}$$

using the fact that  $\phi_1\phi_2 = 1$ . Thus,  $\{Y_1, Y_3, Y_5, \dots\}$  will behave as a (positive-valued) random walk, and similarly  $\{Y_2, Y_4, Y_6, \dots\}$  will follow a negative-valued random walk. So the process  $\{Y_t\}_{t \geq 0}$  will alternate between these two random walks, and hence is not stable.

- (c) Define  $\gamma = \max\{|\gamma_1|, |\gamma_\ell|\}$ , so that  $Y_{t-1} \leq -\gamma$  implies regime 1. Further  $Y_{t-1} > \gamma$  implies regime  $k$ , and the intermediate regimes occur within the band  $|Y_{t-1}| \leq \gamma$ . Whatever happens in those intermediate regimes, if they lead to unstable behavior then eventually  $\{Y_t\}$  will have to go outside the  $[-\gamma, \gamma]$  interval. If  $|\phi_1| < 1$  and  $|\phi_k| < 1$  (case (i)), then the process will mean-revert to zero, and hence will be pushed back inside the  $[-\gamma, \gamma]$  interval eventually. So  $\{Y_t\}$  alternates (possibly unstable) inside the band with (stable) periods outside the band, and hence will not explode or show trending behavior. If  $\phi_k \leq -1$  and  $\phi_1 \in (\frac{1}{\phi_\ell}, 1)$  (case (ii)), then after fluctuating in the middle regimes,  $\{Y_t\}$  will always reach the stationary regime  $Y_t \leq -\gamma$ , possibly after first visiting the explosive regime  $Y_t > \gamma$ , but it will never stay there long, analogously to part (a). So again, the process will be pushed back inside the band. The same occurs in case (iii) ( $\phi_1 \leq -1$  and  $\phi_k \in (\frac{1}{\phi_1}, 1)$ ).

**2.4** Assuming  $0 < \phi < 1$ , the proofs follow the same steps as in the paper by [Anděl et al. \(1984\)](#).

- (a) Let  $C = [2(1 - \phi^2)/\pi]^{1/2}$ . Given  $f(y)$ , the right-hand side (RHS) of the equation can be written as

$$\begin{aligned} \text{RHS} &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left\{-\frac{1}{2}(y + \phi x)^2\right\} [f(x) + f(-x)] dx \\ &= \frac{1}{\sqrt{2\pi}} C \int_0^\infty \exp\left\{-\frac{1}{2}(y + \phi x)^2\right\} \exp\left\{-\frac{1}{2}(1 - \phi^2)x^2\right\} dx, \end{aligned}$$

since  $\Phi(-\phi x) + \Phi(\phi x) = 1$ . Further, we have

$$\begin{aligned} \text{RHS} &= C \exp\left\{-\frac{1}{2}(1 - \phi^2)y^2\right\} \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left\{-\frac{1}{2}(x + \phi y)^2\right\} dx \\ &= C \exp\left\{-\frac{1}{2}(1 - \phi^2)y^2\right\} \Phi(-\phi y) = f(y). \end{aligned}$$

Obviously  $f(y) > 0$ , and

$$\int_{-\infty}^\infty f(y) dy = 2\sqrt{1 - \phi^2} \int_{-\infty}^\infty \varphi(\sqrt{1 - \phi^2}y) \Phi(-\phi y) dy$$

with  $\varphi(\cdot)$  the density of a standard normal distribution. Using  $\Phi(-\phi y) = 1 - \Phi(\phi y)$  we have

$$\begin{aligned} \mathcal{I} &\equiv \int_{-\infty}^\infty f(y) dy = 2\sqrt{1 - \phi^2} \left\{ \int_{-\infty}^\infty \varphi(\sqrt{1 - \phi^2}y) dy \right. \\ &\quad \left. - \int_{-\infty}^\infty \varphi(\sqrt{1 - \phi^2}y) \Phi(\phi y) dy \right\} \\ &= 2 \int_{-\infty}^\infty \varphi(u) du - \mathcal{I}, \end{aligned}$$

or  $2\mathcal{I} = 2$  as required. Therefore  $f(y)$  is a valid pdf.

**Remark 1:** Denote  $C(a, b)$  the Cauchy distribution with density

$$g(x) = \pi^{-1}b\{b^2 + (x - a)^2\}^{-1}, \quad b > 0, \quad -\infty < x < \infty.$$

Assume  $\{\varepsilon_t, t \in \mathbb{Z}\}$  is a strict WN process with Cauchy distribution  $C(0, 1)$ . Then it can be proved (Anděl and Bartoň, 1986) that the stationary distribution of the SETAR(2; 1, 1) process is given by

$$h(y) = \frac{2A}{\pi^2} \left\{ \frac{-[y \log A^{-2}(1 + y^2) + (A^2 - 1 + y^2) \arctan(y)]}{4A^2y^2 + (1 - A^2 + y^2)^2} + \frac{(1 + A)\pi}{(2A)[(1 + A)^2 + y^2]} \right\},$$

where  $A = \phi/(1 - \phi)$ .

**Remark 2:** Clearly, even for a very simple SETAR model the correspondence between the distribution of  $\{\varepsilon_t, t \in \mathbb{Z}\}$  and  $\{Y_t, t \in \mathbb{Z}\}$  is so complicated that only exceptionally we can expect to find explicit results. In fact, for many (non)linear time series processes the distribution function  $F(\cdot)$  of  $\{Y_t, t \in \mathbb{Z}\}$  is not specified completely and only the moments  $\mu_{k,Y} = \int y^k dF(y)$  ( $k = 1, \dots, n$ ) are known. Following Anděl (1987) and Anděl and Garrido (1988) we may, however, look for a distribution  $G(\cdot)$  which satisfies this condition. Write  $s_k = \mathbb{E}(\varepsilon_t^k)$ , with  $s_0 = 1$ . If  $n = 2r$  (an even number) and there are no other restrictions on  $G(\cdot)$ , then the question of the existence of  $G(\cdot)$  is answered by the following assertion (Krein and Nudelman, 1977, Chapter 5, Section 5).

*A sequence  $\{s_k\}_{k=0}^{2r}$  is a system of moments on  $(-\infty, \infty)$  if and only if the matrix  $\mathbf{A} = (a_{ij})$  ( $0 \leq i, j \leq r$ ) with  $a_{ij} = s_{i+j}$  is positive semidefinite.*

After calculating  $s_1, \dots, s_n$ , it is necessary to check whether they are moments. For a general even number of moments the following classical assertion (Krein and Nudelman, 1977, Chapter 5, Section 5, Ex. 1.4) can be used.

*Let  $s_1, \dots, s_{2r}$  be such numbers that the matrix  $\mathbf{A} = (s_{i+j})_{i,j=0}^r$  is positive semidefinite. Let  $\xi$  be an arbitrary real number. Then the polynomial*

$$Q(z) = \begin{vmatrix} s_0 & s_1 & \dots & s_{r-1} & 1 & 1 \\ s_1 & s_2 & \dots & s_r & \xi & z \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ s_{r+1} & s_{r+2} & \dots & s_{2r} & \xi^{r+1} & z^{r+1} \end{vmatrix}$$

*has  $r + 1$  different real roots  $z_1, \dots, z_{r+1}$ . (One of them is  $\xi$ ). The system of linear equations*

$$z_1^k p_1 + \dots + z_{r+1}^k p_{r+1} = s_k \quad (*)$$

*for  $k = 0, \dots, r$  has a unique solution  $p_1, \dots, p_{r+1}$ . This solution is non-negative and satisfies (\*) also for  $k = r + 1, \dots, 2r$ .*

Consider, as an example, the threshold model

$$Y_t = \begin{cases} \phi_1 Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} \leq c, \\ \phi_2 Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} > c. \end{cases}$$

Suppose we wish to get a stationary distribution function  $F(\cdot)$  of  $\{Y_t, t \in \mathbb{Z}\}$ . So we need to choose the distribution  $G(\cdot)$  of  $\{\varepsilon_t, t \in \mathbb{Z}\}$ .

Let  $p = \mathbb{P}(Y_t \leq c)$ ,  $q = \mathbb{P}(Y_t > c)$ ,  $w_k = \mathbb{E}(Y_t^k | Y_{t-1} \leq c)$ , and  $w_k^* = \mathbb{E}(Y_t^k | Y_{t-1} > c)$ . Then we have the following condition for the moments:

$$\mu_{k,Y} = \sum_{i=0}^k \binom{k}{i} (\phi_1^i w_i p + \phi_2^i w_i^* q) s_{k-i}. \quad (**)$$

For the special case that we wish  $Y_t \sim \mathcal{N}(0, 1)$  and  $c = 0$ , we have

$$p = q = 0.5, \quad \mu_{2r-1,Y} = 0, \quad \mu_{2r,Y} = 2^{-r} (2r)! / r!, \\ w_r^* = 2^{r/2} \pi^{-1/2} \Gamma\left(\frac{r+1}{2}\right), \quad w_r = (-1)^r w_r^* \quad (r = 1, 2, \dots).$$

Using (\*\*), the first four moments are given by

$$\begin{aligned} \mu_{1,Y} = 0 &= s_1 + (\phi_2 - \phi_1)(2\pi)^{-1/2}, \\ \mu_{2,Y} = 1 &= s_2 + (\phi_2 - \phi_1)(2/\pi)^{1/2} s_1 + (\phi_2^2 + \phi_1^2)/2, \\ \mu_{3,Y} = 0 &= s_3 + 3(\phi_2 - \phi_1)(2\pi)^{-1/2} s_2 + \frac{3}{2}(\phi_2^2 + \phi_1^2) s_1 + (\phi_2^3 - \phi_1^3)(2/\pi)^{1/2} \\ \mu_{4,Y} = 3 &= s_4 + 2(\phi_2 - \phi_1)(2/\pi)^{1/2} s_3 + 3(\phi_2^2 + \phi_1^2) s_2 \\ &\quad + 4(\phi_2^4 - \phi_1^4)(2/\pi)^{1/2} s_1 + 3(\phi_2^4 + \phi_1^4)/2. \end{aligned}$$

The solution is

$$\begin{aligned} s_1 &= (2\pi)^{-1/2}(\phi_1 - \phi_2), \\ s_2 &= 1 + \pi^{-1}(\phi_1 - \phi_2)^2 - (\phi_1^2 + \phi_2^2)/2, \\ s_3 &= (2\pi)^{-1/2}(\phi_1 - \phi_2)[3 + (3 - \pi)(\phi_1 - \phi_2)^2/\pi], \\ s_4 &= 3[1 + 2(\phi_1 - \phi_2)^2/\pi - (\phi_1^2 + \phi_2^2)] + 3\phi_1^2\phi_2^2 \\ &\quad + (\phi_1 - \phi_2)^2(\phi_1^2 + 4\phi_1\phi_2 + \phi_2^2)/\pi + 2(3 - \pi)(\phi_1 - \phi_2)^4/\pi^2. \end{aligned}$$

It is easy to extend the above procedure to first-order SETAR models with multiple regimes.

- (b) Making use of the hint, the first moment is readily obtained:

$$\begin{aligned} \mathbb{E}(Y_t) = \mu_Y &= \int_{-\infty}^{\infty} y f(y) dy = 2\sqrt{1 - \phi^2} \int_{-\infty}^{\infty} y \varphi(\sqrt{1 - \phi^2} y) \Phi(-\phi y) dy \\ &= -(2/\pi)^{1/2} \phi (1 - \phi^2)^{-1/2}. \end{aligned}$$

For the second moment we have

$$\begin{aligned} \mathbb{E}(Y_t^2) &= \int_{-\infty}^{\infty} y^2 f(y) dy = 2\sqrt{1 - \phi^2} \int_{-\infty}^{\infty} y^2 \varphi(\sqrt{1 - \phi^2} y) \Phi(-\phi y) dy \\ &= 1/(1 - \phi^2). \end{aligned}$$

Clearly, the expression for  $\text{Var}(Y_t)$  follows directly from the first two moments.

**Note:** The above formulas for  $f(y)$ ,  $\mathbb{E}(Y_t)$ , and  $\text{Var}(Y_t)$  remain valid for the case  $|\phi| < 1$  and the SETAR model is driven by a WN process  $\{\varepsilon_t\}$  with symmetrical

density. In fact, [Chan and Tong \(1986\)](#) and [Tong \(1990, p. 141\)](#) show that if  $g(\cdot)$  is a stationary density function of the AR(1) process  $\xi_t = \phi\xi_{t-1} + \varepsilon_t$ . Then the stationary density function  $f(y)$  of the SETAR process specified in Exercise 2.4 is given by

$$\begin{aligned} f(y) &= 2 \int_0^\infty g(x)f(y - \phi x)dx, \\ &= 2 \int_0^\infty \left(\frac{1 - \phi^2}{2\pi}\right)^{1/2} \exp\left\{-\frac{1}{2}(1 - \phi^2)x^2\right\} \times \\ &\quad \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y - \phi x)^2\right\} dx. \end{aligned}$$

Direct integration leads to the stationary marginal pdf  $f(y)$ , and we can see that  $f(y)$  and the expressions for the mean and variance are valid for  $|\phi| < 1$ .

**Remark 3:** For  $r = 3$  and  $r = 4$ , the central moments  $\mu_{r,Y} = \mathbb{E}[(Y_t - \mu_Y)^r]$  of the specified SETAR(2; 1, 1) process can be derived in the same way as above; see [Anděl et al. \(1984\)](#). In particular,

$$\mu_{3,Y} = \mu_3 - 3\mu_2\mu_1 + 2(\mu_1)^3, \quad \text{and} \quad \mu_{4,Y} = \mu_4 - 4\mu_3\mu_1 + 6\mu_2(\mu_1)^2 - 3(\mu_1)^4,$$

where (see, e.g., [Gradshteyn and Ryzhik \(1994, Sections 6.28–6.31\)](#))

$$\mu_1 \equiv \mu_Y \quad \text{and} \quad \mu_r \equiv \mathbb{E}(Y_t^r) = \int_{-\infty}^\infty y^r f(y)dy = \phi^{-r-1} [2(1 - \phi^2)/\pi]^{1/2} J_r,$$

with

$$J_n = \int_{-\infty}^\infty x^n \exp\left(-\frac{1}{2}kx^2\right) \Phi(-x) dx = \frac{n-1}{k} J_{n-2} - \frac{\mu_{n-1}}{k(k+1)^{n/2}} \quad (\text{Re } k > 0, n = 0, 1, \dots).$$

From the above expression for  $J_n$ , the following special cases follow directly

$$\begin{aligned} J_0 &= \left(\frac{\pi}{2k}\right)^{1/2}, \quad J_1 = \frac{-1}{k(k+1)^{1/2}}, \quad J_2 = k^{-1} J_0, \\ J_3 &= \left(\frac{2}{k} + \frac{1}{k+1}\right) J_1, \quad \text{and} \quad J_4 = \left(\frac{3}{k}\right) J_2. \end{aligned}$$

Moreover, after some algebra, it can be shown that the covariance at lag 1 is given by

$$\gamma_Y(1) = \mathbb{E}(Y_t Y_{t+1}) = \frac{\phi}{1 - \phi^2} + \frac{2\phi^2}{\pi(1 - \phi^2)^{1/2}} - \frac{2\phi}{\pi(1 - \phi^2)} \arctan\left(\sqrt{\frac{1 - \phi^2}{\phi^2}}\right).$$

Note, in [Anděl et al. \(1984, p. 96\)](#) the parameter  $\phi$  is missing in the numerator of the third term in the expression for  $\gamma_Y(1)$ . Given the above results, and assuming stationarity, the correlation coefficient at lag 1 is given by  $\rho_Y(1) = (\gamma_Y(1) - \mu_Y^2)/\text{Var}(Y_t)$ .

**2.5** (a) It is easy to see that

$$\begin{aligned} \mathbb{E}(\varepsilon_t \geq 0) &= 1/\sqrt{2\pi}, \quad \mathbb{E}(\varepsilon_t < 0) = -1/\sqrt{2\pi}, \\ \mathbb{E}(\varepsilon_i \varepsilon_j^+) &= \mathbb{E}(\varepsilon_i \varepsilon_j^-) = \begin{cases} 0 & \text{if } i \neq j, \\ 1/2 & \text{if } i = j, \end{cases} \\ \mathbb{E}(\varepsilon_i^+ \varepsilon_j^+) &= \mathbb{E}(\varepsilon_i^- \varepsilon_j^-) = \begin{cases} 1/2\pi & \text{if } i \neq j, \\ 1/2 & \text{if } i = j, \end{cases} \\ \mathbb{E}(\varepsilon_i^- \varepsilon_j^+) &= \begin{cases} -1/2\pi & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases} \end{aligned}$$

Then we have

$$\begin{aligned}\mu_Y &= \mathbb{E}(Y_t) = \mathbb{E}(\varepsilon_t + \theta^+ \varepsilon_{t-1}^+ + \theta^- \varepsilon_{t-1}^-) = \frac{\theta^+ - \theta^-}{\sqrt{2\pi}}, \\ \text{Var}(Y_t) &= 1 + \left( (\theta^+)^2 + (\theta^-)^2 \right) \frac{1}{2} - \mu_Y^2.\end{aligned}$$

(b) The marginal pdf of  $\{Y_t, t \in \mathbb{Z}\}$  is given by

$$\begin{aligned}f(y) &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left\{-\frac{1}{2}(y - \mu_+)^2\right\} \varphi(u) du \\ &\quad + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 \exp\left\{-\frac{1}{2}(y - \mu_-)^2\right\} \varphi(u) du.\end{aligned}$$

On substitution, we get

$$\begin{aligned}f(y) &= \frac{1}{\{1 + (\theta^+)^2\}^{1/2} \sqrt{2\pi}} \exp\left\{\frac{-y^2}{2\{1 + (\theta^+)^2\}}\right\} \Phi\left(\frac{\theta^+ y}{\{1 + (\theta^+)^2\}}\right) \\ &\quad + \frac{1}{\{1 + (\theta^-)^2\}^{1/2} \sqrt{2\pi}} \exp\left\{\frac{-y^2}{2\{1 + (\theta^-)^2\}}\right\} \Phi\left(\frac{-\theta^- y}{\{1 + (\theta^-)^2\}}\right).\end{aligned}$$

On specializing  $f(y)$  to the asMA(1) model with  $\theta^+ = -\theta^- \equiv \theta$ , we get

$$f(y) = \frac{2}{(1 + \theta^2)^{1/2} \sqrt{2\pi}} \exp\left\{\frac{-y^2}{2(1 + \theta^2)}\right\} \Phi\left(\frac{-\theta y}{(1 + \theta^2)^{1/2}}\right).$$

Setting  $\phi = \theta/(1 + \theta^2)^{1/2}$ , the above pdf is identical to the marginal density function of the SETAR(2; 1, 1) process

$$Y_t = \begin{cases} \phi Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} \leq 0, \\ -\phi Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} > 0, \end{cases}$$

see Exercise 2.4 of the main text. This establishes a duality between both models.

**2.6** (a) Recall the basic statistical theory about the relationship between moments and cumulants. In particular, the 3rd-order cumulant of a strictly stationary process  $\{Y_t, t \in \mathbb{Z}\}$  with mean  $\mu_Y$  is defined as

$$\begin{aligned}\gamma_Y(\ell_1, \ell_2) &= \mathbb{E}[(Y_t - \mu_Y)(Y_{t+\ell_1} - \mu_Y)(Y_{t+\ell_2} - \mu_Y)] \\ &= \mathbb{E}(Y_t Y_{t+\ell_1} Y_{t+\ell_2}) - \mu_Y [\mathbb{E}(Y_t Y_{t+\ell_1}) + \mathbb{E}(Y_t Y_{t+\ell_2}) + \mathbb{E}(Y_{t+\ell_1} Y_{t+\ell_2})] \\ &\quad + 2\mu_Y^3.\end{aligned}$$

The full structure of  $\gamma_Y(\ell_1, \ell_2)$  may be obtained by computing the following special cases:

$$\begin{aligned}\gamma_Y(0, 0) &= \mu_{Y,3} - 3\mu_{2,Y}\mu_Y + 2\mu_Y^3 \\ \gamma_Y(-\ell, -\ell) &= \mathbb{E}(Y_t Y_{t-\ell}^2) - 2\mu_Y \mathbb{E}(Y_t Y_{t-\ell}) - \mu_Y \mu_{2,Y} + 2\mu_Y^3 \\ \gamma_Y(0, -\ell) &= \mathbb{E}(Y_t^2 Y_{t-\ell}) - \mu_Y (2\mathbb{E}(Y_t Y_{t-\ell}) + \mu_{2,Y}) + 2\mu_Y^3 \\ \gamma_Y(-\ell, -\ell - s) &= \mathbb{E}(Y_t Y_{t-\ell} Y_{t-\ell-s}) \\ &\quad - \mu_Y (\mathbb{E}(Y_t Y_{t-\ell}) + \mathbb{E}(Y_t Y_{t-\ell-s}) + \mathbb{E}(Y_{t-\ell} Y_{t-\ell-s})) + 2\mu_Y^3,\end{aligned}$$



where  $\mu_{s,Y} = \mathbb{E}(Y_t^s)$  ( $s \geq 2$ ). All the 3rd-order cumulants can be obtained using the following symmetry relations (see, e.g., [Subba Rao and Gabr \(1984, p. 59\)](#))

$$\begin{aligned}\gamma_Y(0, \ell) &= \gamma_Y(\ell, 0) = \gamma_Y(-\ell, -\ell) \\ \gamma_Y(\ell, \ell) &= \gamma_Y(-\ell, 0) = \gamma_Y(0, -\ell) \\ \gamma_Y(s, s + \ell) &= \gamma_Y(s + \ell, s) = \gamma_Y(-s, \ell) = \gamma_Y(\ell, -s) \\ &= \gamma_Y(-s - \ell, -\ell) = \gamma_Y(-\ell, -\ell - s).\end{aligned}$$

Time-reversibility may be determined by verifying  $\gamma_Y(-\ell_1, -\ell_2) = \gamma_Y(\ell_1, \ell_2)$ . The NEAR(1) process  $\{Y_t, t \in \mathbb{Z}\}$  is distributed as an exponential process with marginal moments  $\mu_Y = 1$ ,  $\mu_{2,Y} = 2$ , and  $\mu_{3,Y} = 6$ . Moreover, it is easy to see that

$$\mathbb{E}(\varepsilon_t) = \mu_\varepsilon = p + (1-p)b, \quad \mathbb{E}(\varepsilon_t^2) = \mu_{2,\varepsilon} = 2[p + (1-p)b^2],$$

where  $b = (1-\alpha)\beta$ . Then it is straightforward to derive

$$\begin{aligned}\mathbb{E}(Y_t, Y_{t-\ell}) &= \begin{cases} \alpha\beta\mu_{2,Y} + \mu_\varepsilon & \ell = 1, \\ \alpha\beta\mathbb{E}(Y_t Y_{t-\ell+1}) + \mu_\varepsilon & \ell \geq 2, \end{cases} \\ \mathbb{E}(Y_t Y_{t-\ell}^2) &= \begin{cases} \alpha\beta\mu_{3,Y} + 2\mu_\varepsilon & \ell = 1, \\ \alpha\beta\mathbb{E}(Y_t Y_{t-\ell+1}^2) + 2\mu_\varepsilon & \ell \geq 2, \end{cases} \\ \mathbb{E}(Y_t^2 Y_{t-\ell}) &= \begin{cases} \alpha\beta^2\mu_{3,Y} + 2\alpha\beta\mu_{2,Y}\mu_\varepsilon + \mu_{2,\varepsilon} & \ell = 1, \\ \alpha\beta^2\mathbb{E}(Y_t^2 Y_{t-\ell+1}) + 2\alpha\beta\mu_\varepsilon\mathbb{E}(Y_t Y_{t-\ell+1}) + \mu_{2,\varepsilon} & \ell \geq 2. \end{cases}\end{aligned}$$

Defining the following constants

$$\xi_\beta(\ell) = \frac{1-\beta^\ell}{1-\beta}, \quad \xi_{\alpha\beta}(\ell) = \frac{1-(\alpha\beta)^\ell}{1-\alpha\beta}, \quad \xi_{\alpha\beta^2}(\ell) = \frac{1-(\alpha\beta^2)^\ell}{1-\alpha\beta^2},$$

we obtain the following closed form representation for the expectations

$$\begin{aligned}\mathbb{E}(Y_t Y_{t-\ell}) &= 2(\alpha\beta)^\ell + \mu_\varepsilon \xi_{\alpha\beta}(\ell) \\ \mathbb{E}(Y_t Y_{t-\ell}^2) &= 6(\alpha\beta)^\ell + 2\mu_\varepsilon \xi_{\alpha\beta}(\ell) \\ \mathbb{E}(Y_t^2 Y_{t-\ell}) &= 6(\alpha\beta^2)^\ell + 4(\alpha\beta)^\ell \mu_\varepsilon \xi_\beta(\ell) + \mu_{2,\varepsilon} \xi_{\alpha\beta^2}(\ell) \\ &\quad + 2\mu_\varepsilon^2 \alpha\beta \frac{1}{1-\alpha\beta} (\xi_{\alpha\beta^2}(\ell) - (\alpha\beta)^{\ell-1} \xi_\beta(\ell)).\end{aligned}$$

The cumulant structure is given by

$$\begin{aligned}\gamma_Y(0, 0) &= 2, \quad \gamma_Y(-\ell, -\ell) = 2(\alpha\beta)^\ell, \\ \gamma_Y(0, -\ell) &= 6(\alpha\beta)^\ell + 4(\alpha\beta)^\ell [\mu_\varepsilon \xi_\beta(\ell) - 1] \\ &\quad + 2\mu_\varepsilon [\mu_\varepsilon \alpha\beta \frac{1}{1-\alpha\beta} \{\xi_{\alpha\beta^2}(\ell) - (\alpha\beta)^{\ell-1} \xi_\beta(\ell)\} - \xi_{\alpha\beta}(\ell)] \\ &\quad + \mu_{2,\varepsilon} \xi_{\alpha\beta^2}(\ell).\end{aligned}$$

So,  $\gamma_Y(-\ell, -\ell) \neq \gamma_Y(0, -\ell) = \gamma_Y(\ell, \ell)$ , implying that the process is *not* time-reversible.

(b) Similar to (2.51), we have

$$\mathbb{E}(Y_t Y_{t-\ell}^2) = \frac{\mathbb{E}(Y_t) \mathbb{E}(Y_t^{\alpha^\ell+2})}{\mathbb{E}(Y_t^{\alpha^\ell})}, \quad \mathbb{E}(Y_t^2 Y_{t-\ell}) = \frac{\mathbb{E}(Y_t^2) \mathbb{E}(Y_t^{2\alpha^\ell+1})}{\mathbb{E}(Y_t^{2\alpha^\ell})}.$$

The 3rd-order cumulant structure is given by

$$\begin{aligned}\gamma_Y(0,0) &= 2, \quad \gamma_Y(-\ell, -\ell) = \frac{\mu_Y}{\mu_{\alpha^\ell, Y}} (\mu_{\alpha^\ell+2, Y} - 2\mu_{\alpha^\ell+1, Y}) \\ \gamma_Y(0, -\ell) &= \frac{\mu_{2, Y} \mu_{2\alpha^\ell+1, Y}}{\mu_{2\alpha^\ell, Y}} - 2 \frac{\mu_Y \mu_{\alpha^\ell+1, Y}}{\mu_{\alpha^\ell, Y}},\end{aligned}$$

where

$$\mu_{\alpha, Y} = \Gamma(1 + \alpha) \quad \alpha \in (0, 1), \quad \mu_Y = 1, \quad \text{and} \quad \mu_{2, Y} = 2.$$

Clearly,  $\gamma_Y(-\ell, -\ell) \neq \gamma_Y(0, -\ell) = \gamma_Y(\ell, \ell)$ , implying that the process is *not* time-reversible.

**2.7** Two-state Markov chain  $S_t \in \{1, 2\}$  with switching probabilities  $0 < w_1 < 1$  and  $0 < w_2 < 1$ .

(a) For the stationary probabilities we have to solve

$$\begin{aligned}\pi_1 &= (1 - w_1)\pi_1 + w_2\pi_2 \\ \pi_2 &= w_1\pi_1 + (1 - w_2)\pi_2 \\ \pi_1 + \pi_2 &= 1.\end{aligned}$$

The first two equations both lead to  $w_1\pi_1 = w_2\pi_2$ . Combining this with the third equation  $w_1\pi_1 = w_2(1 - \pi_1)$ . Hence  $\pi_1 = w_2/(w_1 + w_2)$ , which implies  $\pi_2 = 1 - \pi_1 = w_1/(w_1 + w_2)$ . This implies  $\mu = \mathbb{E}(S_t) = 1 \times \pi_1 + 2 \times \pi_2 = (2w_1 + w_2)/(w_1 + w_2)$ .

(b) If  $w_1 + w_2 = 1$ , define  $p = w_1 = 1 - w_2$ . We have

$$\begin{aligned}\mathbb{P}(S_t = 2 | S_{t-1} = 1) &= 1 - \mathbb{P}(S_t = 1 | S_{t-1} = 1) = w_1 = p, \\ \mathbb{P}(S_t = 2 | S_{t-1} = 2) &= 1 - \mathbb{P}(S_t = 1 | S_{t-1} = 2) = 1 - w_2 = p.\end{aligned}$$

Hence,  $\mathbb{P}(S_t = j | S_{t-i} = i)$  is independent of  $i$ , and so  $\{S_t - 1\}$  is an i.i.d. sequence with  $\mathbb{P}(S_{t-1} = 1) = 1 - \mathbb{P}(S_t - 1 = 0) = p$ , i.e., an i.i.d. Bernoulli sequence.

(c) We have  $\mathbb{E}(S_t | S_{t-1}, S_{t-2}, \dots) = \mathbb{E}(S_t | S_{t-1})$  because of the Markov property, and

$$\begin{aligned}\mathbb{E}(S_t | S_{t-1} = 1) &= (1 - w_1) + 2w_1 = 1 + w_1, \\ \mathbb{E}(S_t | S_{t-1} = 2) &= w_2 + 2(1 - w_2) = 2 - w_2.\end{aligned}$$

The proposed expression

$$\mathbb{E}(S_t | S_{t-1}) = \mu(1 - \phi) + \phi S_{t-1} = 2w_1 + w_2 + (1 - w_1 - w_2)S_{t-1}$$

gives  $\mathbb{E}(S_t | S_{t-1} = 1) = 1 + w_1$  and  $\mathbb{E}(S_t | S_{t-1} = 2) = 2 - w_2$ , which agrees with the conditional expectation given above, and hence confirms the autoregressive structure  $\mathbb{E}(S_t | S_{t-1}, S_{t-2}, \dots) = \mu(1 - \phi) + \phi S_{t-1}$ . Note that  $|\phi| < 1$ , so the model is stationary, because  $w_1 \in (0, 1)$  and  $w_2 \in (0, 1)$ .

**2.8** Skewness in the TGARCH(1,1) model  $r_t = Y_t = \sigma_t \varepsilon_t$ , with  $\sigma_t^2 = \alpha_0 + (\alpha_1 + \gamma_1 I(Y_{t-1} < 0))Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ .

(a)  $\mathbb{E}(Y_t^3) = \mathbb{E}(\varepsilon_t^3 \sigma_t^3) = \mathbb{E}(\varepsilon_t) \mathbb{E}(\sigma_t^3) = 0$ , because  $\{\varepsilon_t\}$  and  $\{\sigma_t\}$  are independent processes and  $\mathbb{E}(\varepsilon_t) = 0$ . Hence,  $\tau_Y = \mathbb{E}(Y_t^3) / \sigma_Y^2 = 0$ .

(b) For the two-period return  $r_t[2] = Y_t + Y_{t-1}$ , we find

$$\begin{aligned}\mathbb{E}(r_t[2]) &= \mathbb{E}((Y_t + Y_{t-1})^3) = \mathbb{E}(Y_t^3 + 3Y_t^2Y_{t-1} + 3Y_tY_{t-1}^2 + Y_{t-1}^3) \\ &= 3\mathbb{E}(Y_t^2Y_{t-1}),\end{aligned}$$

because  $\mathbb{E}(Y_t^3) = \mathbb{E}(Y_{t-1}^3) = 0$ , and  $\mathbb{E}(Y_tY_{t-1}^2) = \mathbb{E}(\varepsilon_t\sigma_tY_{t-1}^2) = \mathbb{E}(\varepsilon_t)\mathbb{E}(\sigma_tY_{t-1}^2) = 0$ . Next,

$$\begin{aligned}\mathbb{E}(Y_t^2Y_{t-1}) &= \mathbb{E}(\varepsilon_t^2)\mathbb{E}(\sigma_t^2Y_{t-1}) \\ &= \mathbb{E}((\alpha_0 + (\alpha_1 + \gamma_1I(Y_{t-1} < 0))Y_{t-1}^2 + \beta_1\sigma_{t-1}^2)Y_{t-1}) \\ &= \mathbb{E}(\gamma_1I(Y_{t-1} < 0)Y_{t-1}^3),\end{aligned}$$

because  $\mathbb{E}(Y_{t-1}) = 0$ ,  $\mathbb{E}(Y_{t-1}^3) = 0$ , and  $\mathbb{E}(\sigma_{t-1}^2Y_{t-1}) = \mathbb{E}(\sigma_{t-1}^3)\mathbb{E}(\varepsilon_{t-1}) = 0$ . Next,

$$\mathbb{E}(I(Y_{t-1} < 0)Y_{t-1}^3) = \mathbb{E}(I(\varepsilon_{t-1} < 0)\varepsilon_{t-1}^3\sigma_{t-1}^3) = \int_{-\infty}^0 \varepsilon^3 dF(\varepsilon)\mathbb{E}(\sigma_{t-1}^3) < 0,$$

where  $F(\varepsilon)$  is the distribution function of  $\{\varepsilon_t\}$ . It is not easy to find an explicit expression for  $\mathbb{E}(\sigma_{t-1}^3)$ , but we know it is positive. Combining terms gives

$$\mathbb{E}((r_t[2])^3) = 3\gamma_1 \int_{-\infty}^0 \varepsilon^3 dF(\varepsilon)\mathbb{E}(\sigma_{t-1}^3).$$

We also have  $\text{Var}(r_t[2]) = \mathbb{E}(r_t[2]^2) = \mathbb{E}(Y_t^2) + \mathbb{E}(Y_{t-1}^2) + 2\mathbb{E}(Y_tY_{t-1}) = 2\sigma_Y^2$ , where  $\sigma_Y^2 = \alpha_0/(1 - \alpha_1 - \frac{1}{2}\gamma_1 - \beta_1)$  with  $\mathbb{P}(Y_{t-1} < 0) = 1/2$ , such that

$$\tau(r_t[2]) = \frac{\mathbb{E}(r_t[2]^3)}{\text{Var}(r_t[2])^{3/2}} = 3\gamma_1 \frac{\int_{-\infty}^0 \varepsilon^3 dF(\varepsilon)\mathbb{E}(\sigma_{t-1}^3)}{2\sqrt{2}\sigma_Y^3},$$

so  $\tau(r_t[2]) < 0$  if  $\gamma_1 > 0$ .

**2.9** Using the R-nnet package, the results in Table 2.1 of the main text can be computed as follows.

```
#####
# R code: Exercise 2.9
# File: R-script-AR-NN.r
#
# Estimating and forecasting AR-NN(k;p,...,p) models
# with p = 7 and 8, and k = 0,1,...,5 (number of hidden units).
#
# Note: The results need not be exactly
#       as shown in the main text since they depend heavily on the
#       initial weights chosen by random in the R-function nnet.
#####
library(nnet)

# load the data into workspace
eeg <- read.table("c:\\...\\eeg.dat")
mod <- list()
# Last H = 80 observations are used for forecasting
```

```

hor <- 80
set.seed(1)

n <- length (eeg[,1])
nhor <- n-hor
n1 <- n-1-hor
n2 <- n-2-hor
n3 <- n-3-hor
n4 <- n-4-hor
n5 <- n-5-hor
n6 <- n-6-hor
n7 <- n-7-hor
n8 <- n-8-hor

for (p in (7:8)){
  pp <- p+1

  # Range of initial random weights [-rang1,rang1]
  rang1 <- 1/max(eeg)
  y <- eeg[pp:nhor,1]

  # Obtain the input variables eeg(t-1),...,eeg(t-p))
  if (p == 7){
    nr <- n7
    p1 <- p-1
    p2 <- p-2
    p3 <- p-3
    p4 <- p-4
    p5 <- p-5
    p6 <- p-6
    eeg.x <- cbind(eeg[p:n1,1],eeg[p1:n2,1],eeg[p2:n3,1],eeg[p3:n4,1],
                  eeg[p4:n5,1],eeg[p5:n6,1],eeg[p6:n7,1])
  }
  if (p == 8){
    nr <- n8
    p1 <- p-1
    p2 <- p-2
    p3 <- p-3
    p4 <- p-4
    p5 <- p-5
    p6 <- p-6
    p7 <- p-7
    eeg.x <- cbind(eeg[p:n1,1],eeg[p1:n2,1],eeg[p2:n3,1],eeg[p3:n4,1],
                  eeg[p4:n5,1],eeg[p5:n6,1],eeg[p6:n7,1],eeg[p7:n8,1])
  }

  for (k in (0:5)){
    # Build a p-k-1 network with skip layer connections and linear inputs
    eeg.nn <- nnet(eeg.x,y,size=k,linout=T,skip=T,maxit=10000,decay=1e-2,

```

```

                                reltol=1e-7, abstol=1e-7, rang=rang1)
npar <- p+1+k*(p+2)

# Sum of squared residuals
sse <- sum((y-predict(eeg.nn, eeg.x))^2)

# Estimate of the residual variance
resvar <- sse/nr

# AIC and SBIC
aic <- nr*log(sse) + 2*npar
bic <- nr*log(sse) + npar*log(nr)

# Setup the input variables in the forecasting subsample
if (p == 7){
eeg.p <- cbind(eeg[551:630,1], eeg[550:629,1], eeg[549:628,1],
              eeg[548:627,1], eeg[547:626,1], eeg[546:625,1], eeg[545:624,1])
}
if (p == 8){
eeg.p <- cbind(eeg[551:630,1], eeg[550:629,1], eeg[549:628,1],
              eeg[548:627,1], eeg[547:626,1], eeg[546:625,1], eeg[545:624,1],
              eeg[544:623,1])
}

# Compute the forecasts
yh <- predict(eeg.nn, eeg.p)
# The observed values the forecasting subsample
yo <- eeg[552:n,1]
# Compute and print the sum of squares of forecast errors
ssfe <- sum((yo-yh)^2)
rootmse <- sqrt(ssfe/80)
mae <- sum(abs(yo-yh))/80
# Print summary results
stats <- cbind(p,k,npar,resvar,aic,bic,rootmse,mae)
print(stats)
} # end loop for k
} # end loop for p

```

The results in Table 2.2 of the main text can be computed with the following set of commands.

```

> p=8
> k=4
> nr <- n8
> p1 <- p-1
> p2 <- p-2
> p3 <- p-3
> p4 <- p-4
> p5 <- p-5
> p6 <- p-6
> p7 <- p-7

```

```

> eeg.x <- cbind(eeg[p:n1,1],eeg[p1:n2,1],eeg[p2:n3,1],eeg[p3:n4,1],
+               eeg[p4:n5,1],eeg[p5:n6,1],eeg[p6:n7,1],eeg[p7:n8,1])
> eeg.nn <- nnet(eeg.x,y,size=k,linout=T,skip=T,maxit=10000,decay=1e-2,
+               reltol=1e-7,abstol=1e-7,rang=rang1)
> summary(eeg.nn)

```

**2.10** Using the R-tsDyn package the estimation results for the SETAR(2;2,2) model with delay  $d = 2$  can be obtained as follows.

(a)

```

#####
# R code: Exercise 2.10(a)
# File: SETAR_estimation.r
#
# SETAR(2;2,2) model estimation.
#####
library(tsDyn)
mod <- list()
data <- ts(USunemplmnt_logistic)
mod[["setar"]] <- setar(data,m=3,mL=2,mH=2,thDelay =2)
summary(mod[["setar"]])

```

The output is given by:

Non linear autoregressive model

SETAR model (2 regimes)

Coefficients:

Low regime:

phiL.1	phiL.2	const L
1.6062678	-0.6830262	-0.2215264

High regime:

phiH.1	phiH.2	const H
1.655925	-0.746213	-0.219406

Threshold:

-Variable:  $Z(t) = + (0) X(t) + (0) X(t-1) + (1) X(t-2)$

-Value: -2.537

Proportion of points in low regime: 86.35%

High regime: 13.65%

Residuals:

Min	1Q	Median	3Q	Max
-0.236202	-0.032783	-0.005116	0.028719	0.216038

Fit:

residuals variance = 0.003094, AIC = -1442, MAPE = 1.431%

Coefficient(s):

Estimate	Std. Error	t value	Pr(> t )
----------	------------	---------	----------

```

const L -0.221526    0.046361   -4.7782  3.043e-06 ***
phiL.1   1.606268    0.049435   32.4925 < 2.2e-16 ***
phiL.2  -0.683026    0.050753  -13.4579 < 2.2e-16 ***
const H -0.219406    0.169932   -1.2911  0.1978662
phiH.1   1.655925    0.174492    9.4900 < 2.2e-16 ***
phiH.2  -0.746213    0.192578   -3.8749  0.0001369 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1

```

Threshold

Variable:  $Z(t) = + (0) X(t) + (0) X(t-1) + (1) X(t-2)$

Value: -2.537

Intuitively, the SETAR(2; 2, 2) model says that the dynamics of unemployment act differently depending on the magnitude of the recent change in the unemployment rate. In the upper (first) regime, the negative intercept indicates a downward trend. The roots of the characteristic equation  $z^2 - 1.606z + 0.683026 = 0$  are complex ( $0.8031 \pm 0.1949i$ ), which indicate possible cyclical behavior in  $\{Y_t\}$ . The negative intercept in the lower (second) regime is not significantly different from zero. The roots of  $z^2 - 1.656z + 0.746 = 0$  are complex ( $0.8280 \pm 0.2464i$ ), which is another indication for cyclical behavior in the series. Note, however, that using the model option `selectSETAR(data, m= ..., thDelay=...)` with various combinations of the arguments `m` and `thDelay` does not result in a unique threshold value `th` on the basis of the minimum value of the pooled-AIC criterion.

- (b) The estimation results can be obtained, using the R-CSETAR set of functions available at the website of this book. For ease of reference, we reproduce these functions below.

```

#####
# R code: Exercise 2.10(b)
# File: CSETAR_estimation.r
#
# Set of R functions for estimating CSETAR models, programmed by
# K.S. Chan.
#
# Reference:
# Chan, K.S. and Tsay, R.S. (1998).
# Limiting properties of the least squares estimator of a
# continuous threshold autoregressive model.
# Biometrika, 85(2), 413-426.
# DOI: 10.1093/biomet/85.2.413.
#####
"create.x" <- function(x,start,p,d){
# Create the matrix [x(t-1),..x(t-p),x(t-d)];
# no x(t-d) between x(t-1) and x(t-p).
# This function is called by main
  n <- length(x)
  xy <- NULL
  for(i in (1:p))

```

```

        xy <- cbind(xy, x[(start - i):(n - i)])
        xy <- xy[, - d]
        cbind(xy, x[(start - d):(n - d)])
    }
    "derivative" <- function(x, r, is.constant.jump = F){
        m <- dim(x)[2]
        z <- x[, m]
        res <- cbind(x[,- m], pmin(z-r, 0), pmax(z-r,0))
        if(is.constant.jump) {
            I <- z <= r
            res <- cbind(x[,c(-1,- m)],I,I * z,(1 - I),(1 - I)*z)
        }
        res
    }
    "formh" <- function(x,theta,resi,covariate = NULL){
        # This function is a "work horse" of main
        theta1 <- rev(theta)
        r <- theta1[1]
        phid.minus <- theta1[3]
        phid.plus <- theta1[2]
        m <- dim(x)[2]
        z <- x[, m]
        I <- z <= r
        work <- cbind(1,x[,- m],pmin(z - r,0),pmax(z - r, 0),-(
            phid.minus * I+phid.plus*(1 - I)),covariate)
        apply(work, 2, mult, resi)
    }
    "is.missing" <- function(x){
        any(x == "NA")
    }
    "mult" <- function(a, b){
        a * b
    }
    "main" <- function(y,p,d,start=p+1,a=0.25,b=0.75,Print=T,series.name=
        "", is.grid.search=T,npts.grid=100,standard=F,covariate=NULL,
        covariate.name=NULL){
        #####
        # This function fits a continuous TAR model with
        # delay d and order p in both regimes.
        # It uses the method of grid search to minimize the
        # sum of squared one-step ahead error.
        # INPUT:
        # y      = time series
        # start  = starting value
        # p      = order
        # d      = delay
        # a      = beginning fraction of obs over which the search is done
        #          (default=.25)
        # b      = ending " " (default=.75)
    }

```



```

# npts.grid = no of grid points (default=100)
# standard = logical variable of whether or not to standardize
#             the data (default=F)
# covariate = matrix of covariates. (default=none)
# covariate.name = labels for the covariates
#
# OUTPUT:
# thd = estimated thd
# ls = the ls object of the regression fit with the threshold
#       fixed at thd
# coef = estimated coefficients=(phi0,phi1,...,phip,phid-,phid+,r)
# cov = var-cov matrix for coef
# sigma1 = noise std dev in the lower regime
# sigma2 = noise std dev in the upper regime
#####
AIC.linear <- NA
lab <- c("constant", paste("lag", 1:p, sep = ""))
lab <- lab[ -(d + 1)]
lab <- c(lab,paste(paste("lag",d,sep=""),c("-", "+"),sep=""),
         "threshold", covariate.name)
y1 <- y[y != "NA"]
if(standard)
  y <- (y - mean(y1))/var(y1)^0.5
  n <- length(y1)
  old.cand <- sort(y1[(start:n)-d])
  n.cand <- length(old.cand)
  lbound <- sum(old.cand == min(old.cand))
  ubound <- sum(old.cand == max(old.cand))
  s <- (a1<-max((2*p+1),lbound+p+1,round(n.cand*a))):(b1<-
        min(n.cand -(2*p+1),n.cand-ubound-(p+1),round(
          n.cand*b)))
  cand <- old.cand[s]
if(is.grid.search)
  cand <- old.cand[a1]+((old.cand[b1]-old.cand[a1])*(0:
    npts.grid))/npts.grid
  c1 <- cand[1]
  frac1 <- floor(sum(old.cand <= c1)/n.cand*100)
  c2 <- cand[length(cand)]
  frac2 <- ceiling(sum(old.cand <= c2)/n.cand*100)
  mse <- 0*cand
  x <- create.x(y,start,p,d)
  y2 <- y[ -(1:p)]
if(is.matrix(covariate))
  covariate <- covariate[ -(1:p),]
else covariate <- covariate[ -(1:p)]
i <- 1
for(r in cand) {
  xwork <- cbind(derivative(x,r),covariate)
  select <- !apply(cbind(xwork, y2),1,is.missing)

```

```

        mse[i] <- ls.diag(lsfit(xwork[select,],y2[select]))$std.dev
        i <- i + 1
    }
    thd <- cand[sortindex <- sort.list(mse)[1]]
    xwork <- cbind(derivative(x, thd), covariate)
    select <- !apply(cbind(xwork, y2), 1, is.missing)
    ls <- lsfit(xwork[select, ], y2[select])
    theta <- c(ls$coef, thd)
    n <- sum(select)
    n1 <- n - p
    if(is.matrix(covariate))
        covariate1 <- covariate[select, ]
    else covariate1 <- covariate[select]
    V <- formh(x[select, ],drop=F),theta,ls$residuals,covariate=
        covariate1)
    V <- t(V) %*% V/n
    U <- formh(x[select, ],drop=F),theta,1,covariate=covariate1)
    U <- t(U) %*% U/n
    Uinv <- solve(U)
    cov <- Uinv %*% V %*% Uinv/n
    coef <- c(ls$coef, thd)
    z <- x[select, dim(x)[2]]
    I <- z <= thd
    sigma1 <- sum(ls$residuals^2 * I)/(sum(I))
    sigma1 <- sigma1^0.5
    sigma2 <- sum(ls$residuals^2 * (1 - I))/sum(1 - I)
    sigma2 <- sigma2^0.5
    res <- ls$residuals
    std.res <- ((res * I)/sigma1) + ((res * (1 - I))/sigma2)
    fit <- y2[select] - res
    std.err <- diag(cov)^0.5
    coef <- c(ls$coef, thd)
    names(coef) <- lab
    dimnames(cov) <- list(lab, lab)
    names(std.err) <- lab
    coef.stder <- signif(rbind(coef, std.err), 4)
    dimnames(coef.stder) <- list(c("coef", "std. err."), lab)
    if(Print) {
        n <- length(y1)
    }
    if(series.name != "")
        cat(" \n\n The time series being analyzed is ",series.name,"\n")
        cat(" CLS estimation of \n continuous TAR model with p =\n",
            round(p, 3), " d = ", round(d, 3), "\n")
        cat("\n threshold searched from the ", signif(a1/n * 100,3),
            " percentile to the ", signif(b1/n * 100, 3),
            " percentile\n")
    if(is.grid.search)
        cat(" using a grid search of ", npts.grid + 1,
            " points evenly distributed\n over the range (" ,

```

```

        signif(range(cand), 4), "\n")
cat("\n The fitted model is: \n")
print(coef.stder)
cat("\n")
cat(" RMS for 1st (<= threshold) \n and 2nd regime resp. are = ",
signif(sigma1^2, 4), ",", signif(sigma2^2, 4), "\n")
cat("\n")
cat(" no. of observations in 1st regime (<= threshold)",
    " and 2nd regime\n resp. are = ", sum(I),
    ",", sum(1 - I), "\n")
cat(" The threshold estimate is approximately the ",signif(sum(
    old.cand <= thd)/n.cand * 100, 3),
    " percentile of all threshold data\n")
cat("\n (In the case that the threshold estimate occurs at a tie,",
    "\n the above percentile may appear to lie outside ",
    "\n the search range for the threshold.)\n\n")
    }
invisible(list(thd=thd,ls=ls,cand=cand,mse=mse,cov=cov,
    coef=coef,sigma1=sigma1,sigma2=sigma2,std.err=
    std.err,std.res=std.res,n1=sum(I),n2=sum(1 - I),lab
    = list(c("coef","std.err."),lab),fit=fit))
}

```

Next, the CLS estimates can be obtained by the following set of commands:

```

> USunemplmnt <- read.table("C:/.../USunemplmnt_logistic.dat",
    quote="\")
> data <- ts(USunemplmnt)
> main(data,2,2)

```

The output is given by:

```

CLS estimation of continuous TAR model with p = 2 d = 2
threshold searched from the 24.6 percentile to the 74.6
percentile using a grid search of 101 points evenly distributed
over the range ( -3.047 -2.639 )

```

The fitted model is:

	constant	lag1	lag2-	lag2+	threshold
coef	1.5750	1.60300	-0.67220	-0.60330	-2.6390
std. err.	0.2202	0.06676	0.07882	0.07446	0.1813

```

RMS for 1st (<= threshold)
and 2nd regime resp. are = 0.00364 , 0.001793

```

```

no. of observations in 1st regime (<= threshold) and 2nd regime
resp. are = 188 , 62

```

```

The threshold estimate is approximately the 75.2
percentile of all threshold data

```

(In the case that the threshold estimate occurs at a tie, the above percentile may appear to lie outside the search range for the threshold.)

Comparing with the SETAR model in part (a), we observe that the two models are similar, i.e. the lag-1 and lag-2 coefficients in part (a) are somewhat close to each other, which is indicative of using a CSETAR model. Also the threshold estimates are close to each other, and significantly different from zero at the 5% nominal level. Clearly, the CSETAR model needs fewer parameters than the SETAR model to describe the DGP underlying the series.

- (c) Using the R-MSwM package a two-state MS-AR model can be fitted to the data set USunemplmnt\_matrix.dat. First, however, we fit an AR(2) model to the data.

```
#####
# R code: Exercise 2.10(c)
# File: MS-AR.r
#####
# Optional: rename the time series variables in the 250 * 3 matrix
#           as follows:
renamed_data <- with(USunemplmnt_matrix, data.frame(y=V1,y_1=V2,
                                                    y_2=V3))
#-----
library(MSwM)
mod = lm(y~y_1+y_2,renamed_data) # A linear AR(2) model
summary(mod)
```

The output is given by:

Call:

```
lm(formula = y ~ y_1 + y_2, data = renamed_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.225616	-0.032498	-0.003171	0.030594	0.230100

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-0.14974	0.03531	-4.24	3.16e-05	***
y_1	1.60480	0.04805	33.40	< 2e-16	***
y_2	-0.65797	0.04826	-13.63	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05697 on 247 degrees of freedom

Multiple R-squared: 0.9634, Adjusted R-squared: 0.9631

F-statistic: 3254 on 2 and 247 DF, p-value: < 2.2e-16

The diagnostics indicate that the linear model fits the data well. Diagnostic plots, using the function `plot(mod)`, seem to confirm this observation. Next, we try to improve on these results by fitting a two-state MS-AR(2) model to the data. To indicate that all the parameters can be different in both periods, the

switching parameter (`sw`) is set to a vector with four components with value equal to `TRUE`; see, e.g., the paper by J.A. Sanchez-Espigares and A. Lopez-Moreno (2014) entitled “MSwM examples”, and available at: <http://cran.r-project.org/web/packages/MSwM/vignettes/examples.pdf>.

```
> mod.mswm=msmFit(mod,k=2,sw=c(T,T,T,T),control=list(parallel=F))
> summary(mod.mswm)
```

The output is given by:

Markov Switching Model

```
Call: msmFit(object = mod, k = 2, sw = c(T, T, T, T),
             control = list(parallel = F))
```

```
           AIC          BIC   logLik
-805.4394 -751.1819 408.7197
```

Coefficients:

Regime 1

-----

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)(S)	-0.0455	0.0330	-1.3788	0.168
y_1(S)	1.3010	0.0382	34.0576	< 2.2e-16 ***
y_2(S)	-0.3130	0.0396	-7.9040	2.665e-15 ***

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.02688384

Multiple R-squared: 0.989

Standardized Residuals:

Min	Q1	Med	Q3
-0.053144586	-0.012171286	0.000389544	0.011634994
Max			
0.057245018			

Regime 2

-----

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)(S)	-0.2196	0.0694	-3.1643	0.001555 **
y_1(S)	1.6580	0.0971	17.0752	< 2.2e-16 ***
y_2(S)	-0.7386	0.0999	-7.3934	1.432e-13 ***

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.0770452

Multiple R-squared: 0.9501

**Table 2.1:** Structure of the NeSETAR regimes for the Jökulsá Eystri river time series.

$i$	First-stage	Second-stage ( $j$ )		
2	$Q_{t-2} \leq 92 \text{ m}^3/\text{s}$	$\bar{T}_t \leq -2^\circ \text{C}$ (low temp.) (regime 1)	$-2^\circ \text{C} < \bar{T}_t \leq 1.8^\circ \text{C}$ (moderate temp.) (regime 2)	$\bar{T}_t > 1.8^\circ \text{C}$ (high temp.) (regime 3)
1	$Q_{t-2} > 92 \text{ m}^3/\text{s}$	$\forall \bar{T}_t$ (regime 4)		

Standardized Residuals:

	Min	Q1	Med	Q3
	-0.256000432	-0.011531917	-0.001652446	0.010458948
	Max			
	0.196921581			

Transition probabilities:

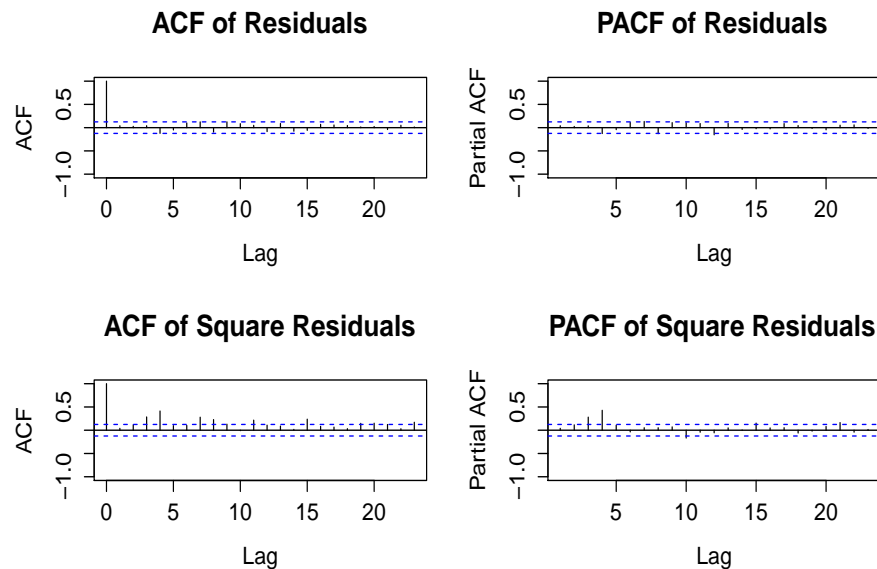
	Regime 1	Regime 2
Regime 1	0.93203735	0.08772958
Regime 2	0.06796265	0.91227042

The model implies that in the second regime (second state), the unemployment rate has a downward trend with an AR(2) polynomial possessing complex characteristic roots. This feature of the model is similar to the first regime of the SETAR model. In the first regime (first state) the characteristic roots are 0.9824 and 0.3186, which differs from the second regime of the SETAR model and the linear AR(2) model, both model-fits having complex characteristic roots. In both regimes the multiple R-squared has high values. The transition probabilities matrix has high values which indicates that it is difficult to change from one regime to the other.

Using the function `plotDiag(mod.mswm)`, we see (Figure 2.1) that the sample ACF and sample PACF of the residuals do not differ from a WN process at the 5% nominal significance level. This is not the case for the squared residuals. The functions `plotProb(mod.mswm,which=2)` and `plotProb(mod.mswm,which=1)` provide the two graphs in Figure 2.2. Both graphs show that the periods for each regime have been detected perfectly.

- 2.11** (a) Table 2.1 shows the structure of the fitted NeSETAR model. It is clear that  $k_1 = 2$ ,  $\ell_{1,2} = 1$ ,  $\ell_{2,2} = 3$ ,  $d_1 = 2$ ,  $d_2 = 0$ ,  $\hat{r}_1 = 92 \text{ m}^3/\text{s}$ ,  $\hat{r}_{2,1} = -2^\circ \text{C}$ , and  $\hat{r}_{2,2} = 1.8^\circ \text{C}$ .

The model reveals that temperature plays a dominant role in modulating the riverflow dynamics. When the temperature is below  $-2^\circ \text{C}$ , riverflow is not significantly affected by temperature or precipitation (regime 1). On the other hand, both current temperature, 3-days earlier temperature, and one-days earlier precipitation have effect on current-days riverflow when the temperature is higher than  $1.8^\circ \text{C}$  (regime 3). However, comparing regime 2 with regime 3, the lags and size of effect depend on whether the temperature is slightly above freezing or high. The reverse influence of temperature 3 days earlier in regime 3 on riverflow is understandable, because the effect of snow melting lasts only for a short

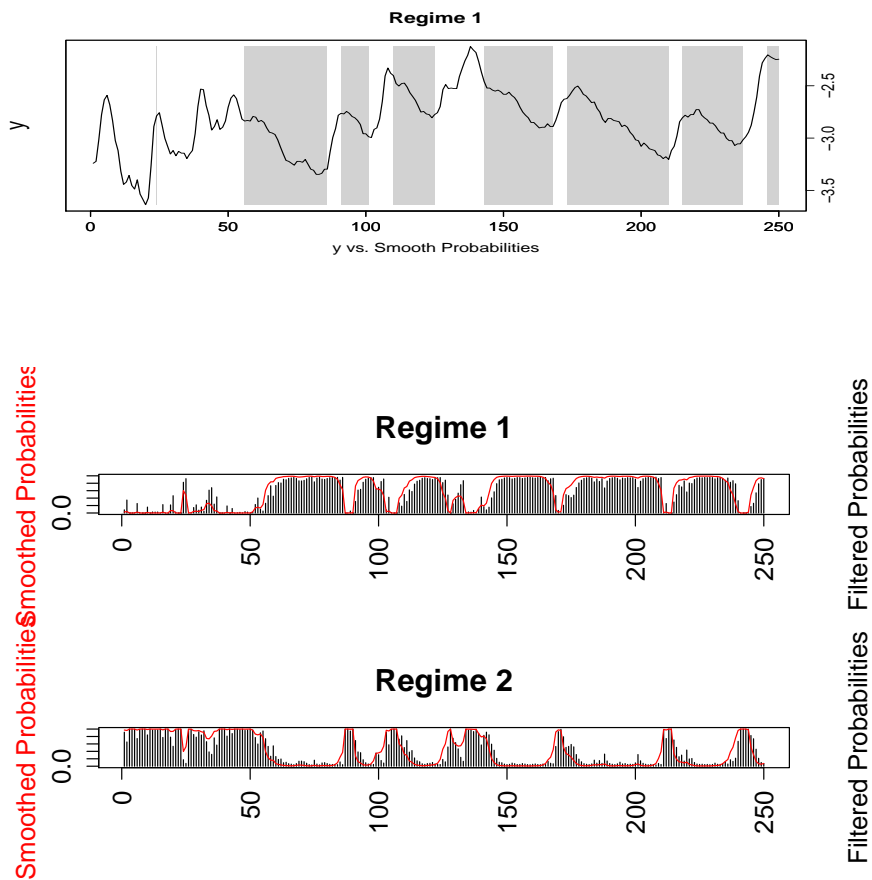


**Figure 2.1:** Jökulsá Eystri river time series. Sample ACF and sample PACF of the residuals of the fitted MS-AR model using `R-plotDiag(mod.mswm)`.

period. In regime 4,  $Q_t$  strongly depends on current, one- and three days ago temperature, and on the level of the previous riverflow  $Q_{t-1}$ . This appears to be reasonable. One possible reason for the heavy dependence on temperature is that the Jökulsá Eystri river basin has a glacier: as a result the effect of  $T_t$  goes beyond producing spring snowmelt.

- (b) To estimate the second stage threshold parameter, say  $\hat{r}_{2,1} = -2$ , we use the following R code.

```
#####
# R code: Exercise 2.11(b) first part
# File: Friedman's_supersmoother.r
#
# Friedman's supersmoother. This code is a modified version of
# S-Plus code provided by Tess Astatkie.
# INPUT: jokulsa.dat (1,086 * 32) matrix.
# OUTPUT: Plot of Friedman's supersmoother: Q on AvT.
#####
x      <- jokulsa
x[,33] <- (x[,24]+x[,25]+x[,26])/3 # Average temperature (AvT)
names(x) <- c("Q", "Q1", "Q2", "Q3", "Q4", "Q5", "Q6", "Q7", "Q8",
             "Q9", "Q10", "P", "P1", "P2", "P3", "P4", "P5", "P6", "P7",
             "P8", "P9", "P10", "Tm", "T1", "T2", "T3", "T4", "T5", "T6",
             "T7", "T8", "T9", "AvT")
x     <- x[x[,3] <= 92, ]
x1x  <- supsmu(x[,33],x[,1])$x   # Friedman's supersmoother
x1y  <- supsmu(x[,33],x[,1])$y
```



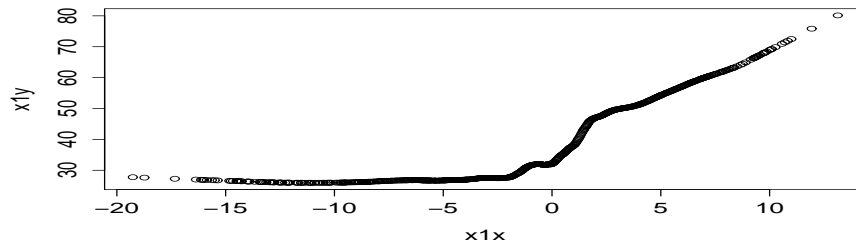
**Figure 2.2:** Jökulsá Eystrí river time series. Upper panel: response variable indicating which observations are associated to regime 1 using R function `plotProb(mod.mswm,which=1)`. Lower panel: filtered and smoothed probabilities for both regimes in the MSM-AR model using R function `plotProb(mod.mswm,which=1)`.

```
plot(x1x,x1y)
```

Figure 2.3 shows the supersmoother estimates of  $Q_t(x1y)$  on  $\bar{T}_t(x1x)$  conditional on  $Q_{t-2} \leq 92 \text{ m}^3/\text{s}$ . We see an impression of three linear pieces, an almost horizontal linear function until approximately  $\bar{T}_t = -2^\circ\text{C}$ , a strong positive linear function between  $-2^\circ\text{C} < \bar{T}_t \leq 2^\circ\text{C}$ , and a mildly positive linear function beyond  $\bar{T}_t = 2^\circ\text{C}$ .

Clearly, some fine-tuning of the threshold parameter estimate  $\hat{r}_{2,1}$  is needed. Bacon and Watts (1971) suggest that estimation of parameters in a two-regime threshold model can be done by reducing the problem to estimation of the parameters in a reparameterized piecewise linear regression with smooth transition at  $r$ . The two-regime model, with threshold  $r$  and  $\{u_i\}$  following an AR(1)





**Figure 2.3:** Jökulsá Eystrí river time series. Supersmoother of  $Q_t$  on  $\bar{T}_t$  ( $Q_{t-2} \leq 92$   $m^3/s$ ).

process, is given by

$$y_i = \begin{cases} \theta_0 + \theta_1 x_i + u_i & \text{if } x \leq r, \\ (\theta_0 + \theta_1 r) + \theta_2 (x - r) + u_i & \text{if } x > r, \end{cases} \quad (*)$$

where  $u_i = \phi u_{i-1} + \epsilon_i$  and with  $\{\epsilon_i\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ . The reparameterized model is given by

$$y_i = \phi y_{i-1} + \beta_0 (1 - \phi) + \beta_1 [(x_i - r) - \phi (x_{i-1} - r)] + \beta_2 \{[(x_i - r)^2 + [\delta/2]^2]^{1/2} - \phi [(x_{i-1} - r)^2 + [\delta/2]^2]^{1/2}\} + \epsilon_i, \quad (**)$$

where the parameters in (\*\*) are related to those in (\*) by

$$\beta_0 = \theta_0 + \theta_1 r, \quad \beta_1 = (\theta_0 + \theta_1)/2, \quad \beta_2 = (\theta_2 - \theta_1)/2,$$

with  $\delta$  a smoothing parameter, playing the same role as  $\gamma$  in an LSTAR model. Next, we obtain a nonlinear least squares estimate of  $r$  by using the following R code.

```
#####
# R code: Exercise 2.11(b) second part
# File: Piecewise_linear_regression.r
#
# INPUT:  x1x and x1y from first part.
# OUTPUT: Parameter estimates of reparameterized linear regression
#         model with smooth transition at "r".
#####
n1 <- length(x1x)
smooa <- data.frame(x1y[2:n1], x1y[1:(n1-1)], x1x[2:n1],
                   x1x[1:(n1-1)])
names(smooa) <- c("Yt", "Y1t", "Xt", "X1t")
thr2a.fit <- nls(Yt~phi*Y1t+B0*(1-phi)+B1*((Xt-Xo)-
  phi*(X1t-Xo))+ B2*(((Xt-Xo)^2+(delta^2)/4)^0.5-
  phi*((X1t-Xo)^2+(delta^2)/4)^0.5), smooa,
start = list(phi = 0.9, B0 = 25, B1 = 2, B2 = 2,
             Xo = -1.8, delta = 1), trace=T) # Rough parameter estimates
list(summary(thr2a.fit)$parameters)
```

The output is given by:

```
[[1]]
      Estimate Std. Error  t value    Pr(>|t|)
phi    0.9978826 0.001087632 917.481872 0.000000e+00
B0    30.4508672 2.129407502  14.300160 2.194523e-40
B1     1.6190465 0.027196238  59.532001 4.240483e-261
B2     1.8660249 0.025803978  72.315397 1.418707e-307
Xo    -1.9614243 0.066831527 -29.348788 4.715549e-120
delta  0.3289523 0.203380583   1.617422 1.062869e-01
```

- (c) After forming subset data sets, one may use a stepwise regression procedure to find a preliminary model. Next, if the best delays have been identified and they are physically meaningful, one may fit a final NeSETAR model to the data. Below is the relevant R code to reassemble the series in the original time order. It lists the actual flow values, the fitted values, the residuals, and the normalized residuals. Finally, the code provides plots of the sample ACF and sample PACF; see Figure 2.4.

```
#####
# R code: Exercise 2.11(c)
# File: NeSETAR_estimation.r
#
# INPUT:  jokulsa.dat (1,086 * 32) matrix.
# OUTPUT: Parameter estimates of NeSETAR model and plots of
#         sample ACF and sample PACF of the normalized residuals.
#####
x      <- jokulsa                # Reintroduce the original data set
x[,33] <- (x[,24]+x[,25]+x[,26])/3 # Average temperature (AvT)
x[,34] <- x[,12]^2                # Quadratic terms
x[,35] <- x[,13]^2
x[,36] <- x[,14]^2
x[,37] <- x[,15]^2
x[,38] <- x[,16]^2
x[,39] <- x[,17]^2
x[,40] <- x[,18]^2
x[,41] <- x[,19]^2
names(x) <- c("Q", "Q1", "Q2", "Q3", "Q4", "Q5", "Q6", "Q7", "Q8",
             "Q9", "Q10", "P", "P1", "P2", "P3", "P4", "P5", "P6", "P7",
             "P8", "P9", "P10", "Tm", "T1", "T2", "T3", "T4", "T5", "T6",
             "T7", "T8", "T9", "AvT", "Ps", "P1s", "P2s", "P3s", "P4s",
             "P5s", "P6s", "P7s")
x1  <- x[x[,3] <= 92 & x[,33] <=-2, ]
fit1 <- lm(Q ~ Q1, x1)
x2  <- x[x[,3] <= 92 & x[,33] > -2 & x[,33] <= 1.8, ]
fit2 <- lm(Q ~ Q1 + Q2 + P1 + Tm - 1, x2)
x3  <- x[x[,3] <= 92 & x[,33] > 1.8, ]
fit3 <- lm(Q ~ Q1 + Q2 + P1s + Tm + T3 -1, x3)
x4  <- x[x[,3] > 92, ]
fit4 <- lm(Q ~ Q1 + Tm + T1 + T3, x4)
```

```

res <- c(1:1085)
for(i in 1:length (resid(fit1))) {
  res[as.integer(names(resid(fit1)[i]))] <- resid(fit1)[i]}
for(i in 1:length (resid(fit2))) {
  res[as.integer(names(resid(fit2)[i]))] <- resid(fit2)[i]}
for(i in 1:length (resid(fit3))) {
  res[as.integer(names(resid(fit3)[i]))] <- resid(fit3)[i]}
for(i in 1:length (resid(fit4))) {
  res[as.integer(names(resid(fit4)[i]))] <- resid(fit4)[i]}
nres <- c(1:185)
for(i in 1:length (resid(fit1))) {
  nres[as.integer(names(resid(fit1)[i]))] <-
  resid(fit1)[i]/sqrt(1.6)}
for(i in 1:length (resid(fit2))) {
  nres[as.integer(names(resid(fit2)[i]))] <-
  resid(fit2)[i]/sqrt(20.3)}
for(i in 1:length (resid(fit3))) {
  nres[as.integer(names(resid(fit3)[i]))] <-
  resid(fit3)[i]/sqrt(49)}
for(i in 1:length (resid(fit4))) {
  nres[as.integer(names(resid(fit4)[i]))] <-
  resid(fit4)[i]/sqrt(146)}
Normresid.Jokulsa <- nres
Residuals.Jokulsa <- res
Fitted.Jokulsa <- jokulsa[,1] - Residuals.Jokulsa
Flow.Jokulsa <- jokulsa[,1]
Jokulsa.fit <- data.frame(Flow.Jokulsa, Fitted.Jokulsa,
                          Residuals.Jokulsa, Normresid.Jokulsa)
acf(Normresid.Jokulsa)
pacf(Normresid.Jokulsa)

```

The parameter estimates of the four-regime NeSETAR model are given by:

```

> fit1$coefficients
(Intercept)      Q1
  4.833466      0.817592

> fit2$coefficients
              Q1              Q2              P1              Tm
1.417020077 -0.399618495 -0.003110877  0.519265557

> fit3$coefficients
              Q1              Q2              P1s              Tm              T3
1.296205158 -0.315028732 -0.006475488  1.214949412 -0.914085532

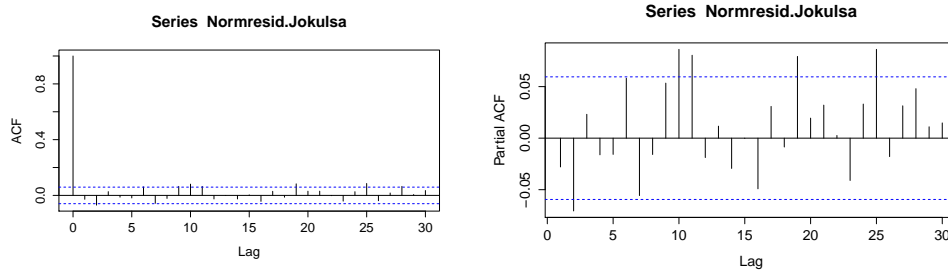
> fit4$coefficients
Call:
lm(formula = Q ~ Q1 + Tm + T1 + T3, data = x4)

```

Coefficients:

(Intercept)	Q1	Tm	T1	T3
49.2127	0.4485	3.4697	3.7486	-6.0845

The sample ACF does not suggest any serial correlations in the normalized residuals. However, for the sample PACF significant values can be seen at lags 2, 10, 11, 19, and 25. This suggests that some further improvements of the fitted NeSETAR model are required. In this context, we note that for reliable stepwise regression the number of observations (39) in the fourth regime is low. So, one may try to improve the identification of this particular regime.



**Figure 2.4:** Jökulsá Eystri river time series. Sample ACF and sample PACF of the normalized residuals together with asymptotic 95% confidence bands (blue medium dashed lines).

**2.12** (a) Let  $Y_{t-1} + \delta \leq 0$  and  $Y_{t-1} \leq 0$ . Expressing the model for  $Y_{t+1}$  gives

$$Y_{t+1} = (\phi_1 + \phi_2)((\phi_1 + \phi_2)Y_t + \varepsilon_t) + \varepsilon_{t+1}.$$

Then, choosing a shock of size  $\varepsilon_t = \delta$ , setting  $\varepsilon_{t+1} = 0$ , and conditioning on the information set  $\omega_{t-1}$ , gives the conditional expectations with and without the shock as

$$\begin{aligned} \mathbb{E}_{t-1}[Y_{t+1} | \varepsilon_t = \delta, \varepsilon_{t+1} = 0] &= (\phi_1 + \phi_2)((\phi_1 + \phi_2)Y_t + \delta) + 0, \\ \mathbb{E}_{t-1}[Y_{t+1} | \varepsilon_t = 0, \varepsilon_{t+1} = 0] &= (\phi_1 + \phi_2)((\phi_1 + \phi_2)Y_t + 0) + 0. \end{aligned}$$

So the traditional impulse (TI) response function is given by

$$\begin{aligned} \text{TI}(1, \delta, \omega_{t-1}) &= \mathbb{E}_{t-1}[Y_{t+1} | \varepsilon_t = \delta, \varepsilon_{t+1} = 0] - \mathbb{E}_{t-1}[Y_{t+1} | \varepsilon_t = 0, \varepsilon_{t+1} = 0] \\ &= (\phi_1 + \phi_2)\delta. \end{aligned}$$

Similarly, for the case  $Y_{t-1} + \delta > 0$  and  $Y_{t-1} > 0$  it follows that  $\text{TI}(1, \delta, \omega_{t-1}) = \phi_1\delta$ . Jointly with the remaining two cases, we have in summary

$$\text{TI} = \begin{cases} (\phi_1 + \phi_2)\delta & \text{if } Y_{t-1} + \delta \leq 0 \text{ and } Y_{t-1} \leq 0, \\ (\phi_1 + \phi_2)\delta + \phi_1((\phi_1 - \phi_2) - \phi_1)Y_{t-1} & \text{if } Y_{t-1} + \delta \leq 0 \text{ and } Y_{t-1} > 0, \\ \phi_1\delta + (\phi_1 + \phi_2)(\phi_1 - (\phi_1 + \phi_2))Y_{t-1} & \text{if } Y_{t-1} + \delta > 0 \text{ and } Y_{t-1} \leq 0, \\ \phi_1\delta & \text{if } Y_{t-1} + \delta > 0 \text{ and } Y_{t-1} > 0. \end{cases}$$

It is clear that the TI response function depends on the combined magnitude of the history  $Y_{t-1}$  and the shock  $\delta$ , relative to the value of the threshold at  $r = 0$ .

**Table 2.2:** Impulse response functions of a stationary SETAR(2; 1, 1) process and a stationary AR(1) process.

$H$	SETAR(2; 1, 1)		AR(1)	
	$\delta = 1$	$\delta = -1$	$\delta = 1$	$\delta = -1$
0	0.9982	-0.9998	1.0000	-1.0000
1	0.7397	-0.5536	0.4000	-0.4000
2	0.5595	-0.3661	0.1600	-0.1600
3	0.4343	-0.2654	0.0640	-0.0640
4	0.3438	-0.2022	0.0256	-0.0256
5	0.2757	-0.1586	0.0102	-0.0102
6	0.2234	-0.1266	0.0041	-0.0041
7	0.1824	-0.1023	0.0016	-0.0016
8	0.1497	-0.0833	0.0007	-0.0007
9	0.1233	-0.0683	0.0003	-0.0003
10	0.1020	-0.0562	0.0001	-0.0001

The TI response function is not symmetric in the size and sign of the shock. Positive and negative values of  $\delta$  yield impulse responses that are not opposite in sign, as for instance in the case of a linear time series model; see also part (b).

- (b) The bootstrap simulation results are summarized in Table 2.2. For comparison the analytic TI response function for a stationary AR(1) process with parameter  $\phi = (0.9 - 0.5) = 0.4$  is included. Reversing the shocks from  $\delta = 1$  to  $\delta = -1$  for the SETAR(2; 1, 1) process results in asymmetrical GIRFs, with the responses decaying at a faster rate for  $\delta = -1$ . In contrast, for the AR(1) process the absolute values of the impulse responses are the same across all horizons  $H$ . Moreover, for both  $\delta = 1$  and  $\delta = -1$ , the AR(1)-TI responses decay (geometrically) at a much faster rate than the responses for the SETAR process. The relevant MATLAB computer code is given below.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 2.12(b)
% File: nlm_girf.m
%
% Program to compute GI response functions (GIRF).
% Based on Koop et al. (1996), J. of Econometrics, 74(1), 119-147.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function nlm_girf( )
    clear all % clear variables from memory
    clc      % clear command window
    rng(123) % seeds the random number generator
    % Parameters
    delta = 1;          % Size of the shock
    phi1  = 0.9;        % AR(1), and first TAR parameter
    phi2  = -0.50;      % second TAR parameter
    phi   = phi1+ phi2;

```

```

B      = 1000;          % Number of Bootstrap replications
H      = 10;           % Maximum forecast horizon of impulses
% Simulate the data
e = randn(B,1);
y = zeros(B,1);
for i = 2:B
    y(i) = phi1*y(i-1) + phi2*y(i-1)*(y(i-1)<=0) + e(i);
end

% Total number of draws needed for a given initial history
% is B-1, and H+1 for the number of impulse horizons.
BH     = (B-1)*(H+1);
impulse = zeros(B-1,H+1);

% Loop through the data to change the
% initial history (condition)
for i=1:B-1
    % Bootstrap residuals
    ind    = fix(rand(BH,1)*(B-1) + 1 );
    e_boot = e(ind);
    e_boot = reshape(e_boot,H+1,B-1);
    ye0    = zeros(H+1,B-1);
    ye1    = zeros(H+1,B-1);

    % Loop through horizon of impulse responses
    for j = 2:B-1
        % Initial condition based on a bootstrap draw
        ye0(:,j) = model(e_boot(:,j),e_boot(1,j),phi1,phi2);
        % Initial condition based on history (i subscript)
        % plus shock delta
        ye1(:,j) = model(e_boot(:,j),e(i)+delta,phi1,phi2);
    end
    % Average over horizon given an initial condition
    impulse(i,:) = (mean(ye1,2) - mean(ye0,2))';
end

% Average each impulse across histories for SETAR process
impulse_girf = mean(impulse)';

% AR(1) Linear impulse responses
impulse_linear = delta*recserrar(zeros(H+1,1),1,phi);

format short
disp(' GI Response AR(1) Linear ')
disp([impulse_girf impulse_linear]);
end
%
%----- Functions -----
% Model used to compute GI Response Function.

```

```

%-----
function y = model(e,y0,phi1,phi2)
    y = y0 + zeros(length(e),1);
    for i=2:length(y)
        % SETAR model
        y(i) = phi1*y(i-1) + phi2*y(i-1)*(y(i-1)<=0) + e(i);
    end
end

function result = recserar(x,y0,a)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 2.12(b)
% PURPOSE: computes a vector of autoregressive recursive series.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% USAGE: result = recserar(x,y0,a)
% where: x = matrix of dimension (n,k)
%         y0 = matrix of dimension (p,k)
%         a = matrix of dimension (p,k)
%-----
% RETURNS:
%         results(1:n,1:k) = contains columns computed
%         recursively with result=y0 for rows 1:p, and
%         result(j,:)=result(j-1,:).*a(1:p,:) + x(j,:)
%         for rows p+1:n.
%
% written by:
% Kit Baum, Dept. of Economics, Boston College
% Chestnut Hill MA 02467 USA, baum@bc.edu
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if (nargin ~= 3)
    error('Wrong number of arguments to recserar');
end
[n1 k1] = size(x);
[p1 k2] = size(y0);
[p2 k3] = size(a);
if (k1 ~= k2)
    error('recserar x, y0 must have same number of columns');
end
if (k1 ~= k3)
    error('recserar x, a must have same number of columns');
end
if (p1 ~= p2)
    error('recserar y0, a must have same number of rows');
end
result = zeros(n1,k1);
for j=1:p1;
    result(j,:) = y0(j,:);
end
for j=(p1+1):n1;

```

```

result(j,:) = x(j,:);
for k=1:p1;
    result(j,:) = result(j,:)+a(k,:).*result(j-k,:);
end
end

```

## Chapter 3

**3.1** For  $m = 1, 2, \dots$ , we set

$$\begin{aligned}
 r_y &= \mathbb{E}_{\boldsymbol{\theta}}\{1 + Y_t^m | Y_{t-1} = y\} \\
 &= 1 + (\phi + \xi e^{-\gamma y^2})^m y^m + \sum_{k=1}^m \binom{m}{k} (\phi + \xi e^{-\gamma y^2})^{m-k} \mathbb{E}(\varepsilon_t^k) y^{m-k} \\
 &= (1 + y^m) \left\{ 1 - \frac{\ell(y; \boldsymbol{\theta})}{1 + y^m} \right\},
 \end{aligned}$$

where

$$\ell(y; \boldsymbol{\theta}) = [1 - (\phi + \xi e^{-\gamma y^2})^m] y^m - \sum_{k=1}^m \binom{m}{k} (\phi + \xi e^{-\gamma y^2})^{m-k} \mathbb{E}(\varepsilon_t^k) y^{m-k},$$

and  $\boldsymbol{\theta} = (\phi, \xi, \gamma)'$ . Note that

$$\lim_{|y| \rightarrow \infty} \frac{\ell(y; \boldsymbol{\theta})}{1 + y^m} = 1 - \phi^m.$$

If  $|\phi| < 1$ , then  $1 - \phi^m > 0$  and there exists  $k > 0$  and  $\delta > 0$ , such that  $\ell(y; \boldsymbol{\theta}) / (1 + y^m) \geq \delta \forall |y| > k$ . This implies  $\mathbb{E}_{\boldsymbol{\theta}}\{1 + Y_t^m | Y_{t-1} = y\} \leq (1 + y^m)(1 - \delta) \forall |y| > k$ . Since  $\sup_{|y| \leq k} |r_y| < \infty$ , it follows from [Feigin and Tweedie \(1985, Thm. 1\)](#)<sup>3</sup> that  $\{Y_t, t \in \mathbb{Z}\}$  is geometrically ergodic for each  $\boldsymbol{\theta} \in \Theta$  and  $\mathbb{E}|Y_t|^m < \infty, \forall m = 1, 2, \dots$ , where  $\Theta = \{\boldsymbol{\theta} : |\phi| < 1, \gamma > 0\}$ .

**3.2** The reconstruction errors can be written as

$$\begin{aligned}
 e_t &= \varepsilon_t - \widehat{\varepsilon}_t \\
 &= \psi(\varepsilon_{t-1})\varepsilon_{t-1} - \psi(\widehat{\varepsilon}_{t-1})\widehat{\varepsilon}_{t-1} \\
 &= \psi(\widehat{\varepsilon}_{t-1})(\varepsilon_{t-1} - \widehat{\varepsilon}_{t-1}) - (\psi(\widehat{\varepsilon}_{t-1}) - \psi(\varepsilon_{t-1}))\varepsilon_{t-1} \\
 &= \psi(\widehat{\varepsilon}_{t-1})e_{t-1} - (\psi(\widehat{\varepsilon}_{t-1}) - \psi(\varepsilon_{t-1}))\varepsilon_{t-1}.
 \end{aligned}$$

Hence,

$$|e_t| = |\psi(\widehat{\varepsilon}_{t-1})e_{t-1} + (\psi(\widehat{\varepsilon}_{t-1}) - \psi(\varepsilon_{t-1}))\varepsilon_{t-1}| \leq \gamma|e_{t-1}| + 2\gamma|\varepsilon_{t-1}|.$$

Recursive substitution gives

$$|e_t| \leq 2\gamma|\varepsilon_{t-1}| + \gamma\{2\gamma|\varepsilon_{t-2}| + \gamma(2\gamma|\varepsilon_{t-3}| + \gamma(\dots))\} \leq 2\gamma \sum_{j=1}^t |\varepsilon_{t-j}| \gamma^{j-1} + \gamma^t |e_0|.$$

<sup>3</sup>See also Section 3.4.2 of the main text.



Since  $\widehat{\varepsilon}_0 = 0$ ,  $e_0 = \varepsilon_0$ . So, we have

$$|e_t| \leq 2\gamma \sum_{j=1}^{t-1} |\varepsilon_{t-j}| \gamma^{j-1} + |\varepsilon_0| (2\gamma \gamma^{t-1} + \gamma^t) \leq 3\gamma \sum_{j=1}^t |\varepsilon_{t-j}| \gamma^{j-1}.$$

Using Jensen's inequality  $\mathbb{E}|\varepsilon_t|^n \mathbb{E}|\varepsilon_t|^m \leq \mathbb{E}|\varepsilon_t|^{n+m}$ , we get

$$\mathbb{E}|\varepsilon_{t-j_1}| \cdots \mathbb{E}|\varepsilon_{t-j_m}| \leq \mathbb{E}|\varepsilon_t|^m \leq c < \infty.$$

So, for  $m \geq 1$ , it follows that

$$\mathbb{E}|e_t|^m \leq 3^m \gamma^m c \left( \sum_{j=1}^t \gamma^{j-1} \right)^m \leq c^* (1 - \gamma)^{-m},$$

which shows that  $\{Y_t, t \in \mathbb{Z}\}$  is invertible in case the coefficients are unknown.

- 3.3** (a) With  $g(\varepsilon_{t-1}; \beta) = -\beta \varepsilon_{t-1}^2$ , the reconstruction errors satisfy the stochastic equation

$$\begin{aligned} e_t &= g(\varepsilon_{t-1}; \beta) - g(e_{t-1} + \varepsilon_{t-1}; \beta) = -\beta \varepsilon_{t-1}^2 + \beta (e_{t-1} + \varepsilon_{t-1})^2 \\ &= e_{t-1} (\beta e_{t-1} + 2\beta \varepsilon_{t-1}). \end{aligned}$$

Linearizing around  $e_t = 0$  yields the homogeneous stochastic equation  $e_t = 2\beta \varepsilon_{t-1} e_{t-1}$ . Its solution is given by

$$e_t = e_0 (2\beta)^t \prod_{s=1}^t \varepsilon_{s-1}.$$

In particular,

$$|e_t| = |e_0| \exp \left( t \sum_{s=1}^t \{ \log |2\beta| + \log |\varepsilon_{s-1}| \} / t \right).$$

Hence, if  $\log |2\beta| + \mathbb{E} \log |\varepsilon_t| < 0$  the model is locally invertible. Since  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ ,  $\mathbb{E} \log |\varepsilon_t| = -(C + \log 2)/2$  where  $C = 0.57721 \cdots$  is Euler's constant, it follows that a condition for local invertibility is  $|\beta| < -(1/2) \mathbb{E} \log |e_t| = 0.3176$ .

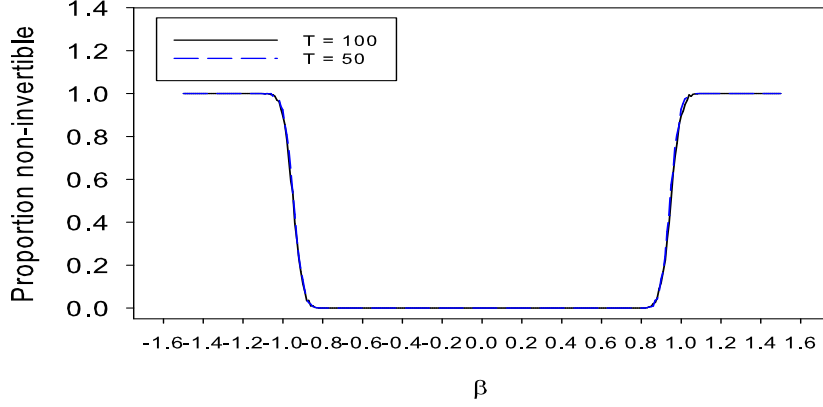
- (b) Using Algorithm 3.1 in the main text, Figure 3.1 gives a graphical representation of the proportion of non-invertible BL models, out of 1,000 MC replications, for  $T = 50$  (blue medium dashed line) and 100 (black solid line). The parameter  $\beta$  was taken in the range  $[-1.5, 1.5]$  with step size of 0.01. As can be seen there is not much difference between the two sample sizes. In both cases the model inverts for  $|\beta| < 0.85$ .

- 3.4** (a) Consider the more general  $\text{BL}(p, 0, Q, 1)$  model

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{u=1}^Q \psi_{1u} Y_{t-u} \varepsilon_{t-1} + \varepsilon_t.$$

Let  $s = \max(p, Q)$ , and  $\mathbf{Y}_t = (Y_t, \dots, Y_{t-s+1})'$ . The BL process in state space representation is given by

$$\mathbf{Y}_t = \mathbf{A} \mathbf{Y}_{t-1} + \mathbf{B} \mathbf{X}_{t-1} \varepsilon_{t-1} + \mathbf{C} \varepsilon_t,$$



**Figure 3.1:** Proportion of non-invertibility for the BL model  $Y_t = \varepsilon_t + \beta\varepsilon_{t-1}^2$  for  $T = 50$  and  $T = 100$ ; 1,000 MC replications.

where

$$\mathbf{A} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p & 0 & \cdots & 0 \\ & & & \mathbf{I}_{p-1} & & \mathbf{0}_{(p-1) \times (s-1)} & & \\ & & & & & & & \end{pmatrix}_{s \times s}$$

$$\mathbf{B} = \begin{pmatrix} \psi_{11} & \cdots & \psi_{1Q} & 0 & \cdots & 0 \\ & & & \mathbf{0}_{(s-1) \times s} & & \end{pmatrix}_{s \times s}, \quad \mathbf{C} = (1, 0, \dots, 0)'$$

Assume for simplicity of notation that  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ . Then, following Liu (1985), define

$$\boldsymbol{\mu}(t) = \mathbb{E}(\mathbf{Y}_t), \quad \mathbf{V}(t) = \mathbb{E}(\mathbf{Y}_t \mathbf{Y}_t'), \quad \mathbf{U}(t) = \mathbb{E}(\mathbf{Y}_t \mathbf{Y}_t' \varepsilon_t), \quad \mathbf{W}(t) = \mathbb{E}(\mathbf{Y}_t \mathbf{Y}_t' \varepsilon_t^2).$$

Since  $\{Y_t\}$  and  $\{\varepsilon_{t'}\}$  are independent for all  $t' > t$ , we have  $\mathbb{E}(\mathbf{Y}_t \varepsilon_t) = \mathbf{C}$ . From the state space representation, we have by recursive substitution

$$\boldsymbol{\mu}(t+1) = \mathbf{A}\boldsymbol{\mu}(t) + \mathbf{B}\mathbf{C} = \mathbf{A}^t \boldsymbol{\mu}(1) + \left( \sum_{j=0}^{t-1} \mathbf{A}^j \right) \mathbf{B}\mathbf{C}.$$

Then a sufficient condition for  $\lim_{t \rightarrow \infty} (\mathbf{A}^t \boldsymbol{\mu}(1) + (\sum_{j=0}^{t-1} \mathbf{A}^j) \mathbf{B}\mathbf{C}) < \infty$  is that  $\rho(\mathbf{A}) < 1$  (called *first-order stationarity*). Since in that case  $\lim_{t \rightarrow \infty} \boldsymbol{\mu}(t) = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{B}\mathbf{C} \doteq \boldsymbol{\mu}$ .

For *second-order stationarity*, we have

$$\mathbf{V}(t) = \mathbf{A}\mathbf{V}(t-1)\mathbf{A}' + \mathbf{A}\mathbf{U}(t-1)\mathbf{B}' + \mathbf{B}\mathbf{U}(t-1)\mathbf{A}' + \mathbf{B}\mathbf{W}(t-1)\mathbf{B}' + \mathbf{C}\mathbf{C}',$$

where

$$\begin{aligned} \mathbf{U}(t) &= \mathbf{A}\boldsymbol{\mu}(t-1)\mathbf{C}' + \mathbf{C}\boldsymbol{\mu}'(t-1)\mathbf{A}' + \mathbf{B}\mathbf{C}\mathbf{C}' + \mathbf{C}\mathbf{C}'\mathbf{B}', \\ \mathbf{W}(t) &= \mathbf{A}\mathbf{V}(t-1)\mathbf{A}' + \mathbf{A}\mathbf{U}(t-1)\mathbf{B}' + \mathbf{B}\mathbf{U}(t-1)\mathbf{A}' + \mathbf{B}\mathbf{W}(t-1)\mathbf{B}' + 2\mathbf{C}\mathbf{C}' \\ &= \mathbf{V}(t) + 2\mathbf{C}\mathbf{C}'. \end{aligned}$$

Suppose the condition for first-order stationarity is satisfied, then

$$\begin{aligned}\mathbf{U}(t) &\doteq \mathbf{U} = \mathbf{A}\boldsymbol{\mu}\mathbf{C}' + \mathbf{C}\boldsymbol{\mu}'\mathbf{A}' + \mathbf{B}\mathbf{C}\mathbf{C}' + \mathbf{C}\mathbf{C}'\mathbf{B}', \\ \mathbf{V}(t) &= \mathbf{A}\mathbf{V}(t-1)\mathbf{A}' + \mathbf{B}\mathbf{V}(t-1)\mathbf{B}'.\end{aligned}$$

Let  $\text{vec}\{\mathbf{V}(t)\} = (\mathbf{V}(t)_{.1}, (\mathbf{V}(t)_{.2}, \dots, (\mathbf{V}(t)_{.s})')'$ , where  $(\mathbf{V}(t))_{.i}$  is the  $i$ th column of  $\mathbf{V}(t)$  ( $i = 1, \dots, s$ ). Since  $\text{vec}\{\mathbf{DFG}\} = (\mathbf{G}' \otimes \mathbf{D})\text{vec}(\mathbf{F})$ , for any three squared matrices,  $\mathbf{D}$ ,  $\mathbf{F}$  and  $\mathbf{G}$  of the same order. Then we have

$$\text{vec}\{\mathbf{V}(t)\} = (\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B})\text{vec}\{\mathbf{V}(t-1)\}.$$

Now, a sufficient condition for second-order stationarity is that

$$\rho(\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B}) < 1.$$

For model (3.46) with  $s = 1$ , we have  $\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B} = \phi^2 + \psi^2$ . Hence, the condition for second-order stationarity reduces to  $\phi^2 + \psi^2 < 1$ .

It should be mentioned that in the case  $p = Q = 1$ , the above result was earlier obtained by [Pham and Tran \(1981\)](#). In particular, let  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ . Then, if  $|\phi| < 1$ , a necessary and sufficient condition for the existence of a strictly stationary process  $\{Y_t, t \in \mathbb{Z}\}$  satisfying the model (3.46) is  $\phi^2 + \psi^2\sigma_\varepsilon^2 < 1$ . Clearly, replacing  $\phi$ ,  $\psi$  and  $\sigma_\varepsilon^2$  by their estimates, we see that the fitted model (3.47) satisfies this condition.

- (b) From the results of [Liu \(1985\)](#), summarized in Table 3.3 of the main text, a sufficient condition for invertibility of the  $\text{BL}(p, 0, Q, 1)$  process is given by  $\mathbb{E} \log |\mathbf{C}'\mathbf{B}\mathbf{Y}_t| < 0$ . A more useful sufficient condition is  $\mathbb{E}(\mathbf{C}'\mathbf{B}\mathbf{Y}_t)^2 < 1$ . For the  $\text{BL}(1, 0, 1, 1)$  model (3.47) with  $p = Q = 1$ , the condition for invertibility becomes  $\psi^2 E(Y_t^2) < 1$ .

[Pham and Tran \(1981\)](#) arrive at the above result by first writing (3.47) in an equivalent Markovian representation

$$Y_t = Z_{t-1} + \varepsilon_t, \quad Z_t = \{\phi + \psi\varepsilon_t\}Z_{t-1} + \{\phi + \theta\varepsilon_t\}\varepsilon_t,$$

where  $\{\varepsilon_t\}$  are i.i.d. random variables with  $\mathbb{E}(\varepsilon_t) = 0$ ,  $\mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$ , and  $\mathbb{E}(\varepsilon_t^4) < \infty$ . Then, assuming stationarity, it is easy to see that

$$\mathbb{E}(Z_t) = \frac{\psi\sigma_\varepsilon^2}{1-\phi},$$

$$\text{Var}(Z_t) = (\phi^2 + \psi^2\sigma_\varepsilon^2)\text{Var}(Z_t) + \text{Var}\left\{\psi\varepsilon_t^2 + \left(\phi + \frac{\psi^2\sigma_\varepsilon^2}{1-\phi}\right)\varepsilon_t\right\},$$

where the last term in  $\text{Var}(Z_t)$  is positive, given the assumption that  $\varepsilon_t^2$  is almost surely not a linear function of  $\varepsilon_t$ . Therefore  $(1 - \phi^2 - \psi^2\sigma_\varepsilon^2)\text{Var}(Z_t) > 0$ , which implies that  $1 - \phi^2 - \psi^2\sigma_\varepsilon^2 > 0$ . From the above results, and some reordering of terms, it follows that

$$\text{Var}(Z_t) = \frac{\sigma_\varepsilon^2}{1 - \phi^2 - \lambda^2} \left\{ v\lambda^2 + \left( \frac{\lambda^2}{1-\phi} + \phi \right)^2 + 2c\lambda \left( \frac{\lambda^2}{1-\phi} + \phi \right) \right\},$$

where  $\lambda = \psi\sigma_\varepsilon$ ,  $v = \text{Var}(\varepsilon_t^2)/\sigma_\varepsilon^4$ , and  $c = \mathbb{E}(\varepsilon_t^3)/\sigma_\varepsilon^3$ . Then

$$\text{Var}(Y_t^2) = \frac{\sigma_\varepsilon^2}{1 - \phi^2 - \lambda^2} \left\{ 1 + v\lambda^2 + \frac{4\phi\lambda^2}{1-\phi} + 2c\lambda \left( \frac{\lambda^2}{1-\phi} + \phi \right) \right\}.$$

Thus, the invertibility condition becomes

$$\left(v + \frac{4\phi}{1-\phi}\right)\lambda^4 + 2\lambda^2 + 2c\lambda^3\left(\frac{\lambda^2}{1-\phi} + \phi\right) - (1-\phi^2) < 0.$$

In the Gaussian case, the above result reduces to (3.48).

(c) Substituting the parameter estimates into condition (3.48) gives

$$\begin{aligned} 2(1 + 0.5421)(0.0284)^4 + 2(1 - 0.5421)(0.0284)^2 - (1 - 0.5421)^2(1 + 0.5421) \\ = -0.3234 < 0. \end{aligned}$$

So, the invertibility condition is satisfied.

**3.5** Let  $\beta_{i,t} = \theta_i + \sum_{j=0}^P \psi_{ij}Y_{t-j}$  ( $i = 1, \dots, \bar{Q}$ ), where  $\bar{Q} = \max\{q, Q\}$ . Then the BL model becomes

$$Y_t = \phi_0 + \sum_{j=1}^p \phi_j Y_{t-j} + \sum_{j=1}^{\bar{Q}} \beta_{j,t-j} \varepsilon_{t-j} + \theta_0 \varepsilon_t.$$

Define

$$Z_{i,t} = \phi_0 + \sum_{j=i}^p \phi_j Y_{t+i-j} + \sum_{j=i}^{\bar{Q}} \beta_{j,t+i-j} \varepsilon_{t+i-j},$$

and observe that  $Y_t = Z_{1,t} + \theta_0 \varepsilon_t$ . Moreover, it follows that

$$\begin{aligned} Z_{i,t} &= Z_{i+1,t-1} + \phi_i Y_t + \beta_{i,t} \varepsilon_t \quad (i = 1, \dots, \bar{Q} - 1), \\ Z_{\bar{Q},t} &= \phi_0 + \sum_{j=\bar{Q}}^p \phi_j Y_{t+\bar{Q}-j} + \beta_{\bar{Q},t} \varepsilon_t. \end{aligned}$$

So, by replacing  $Y_{t+1-i}$  by  $Z_{\bar{Q}+i,t}$  ( $i = 1, \dots, \bar{P} = \max\{p - \bar{Q}, P\}$ ) in the first equation we obtain the representation

$$\begin{aligned} Z_{i,t} &= (\phi_i + \psi_{i,0} \varepsilon_t) Z_{1,t-1} + Z_{i+1,t-1} + \sum_{j=1}^{\bar{P}} \psi_{ij} \varepsilon_t Z_{\bar{Q}+j,t-1} \\ &\quad + \{(\phi_i \theta_0 + \theta_i) \varepsilon_t + \theta_0 \psi_{i,0} \varepsilon_t^2\} \quad (i = 1, \dots, \bar{Q} - 1). \end{aligned}$$

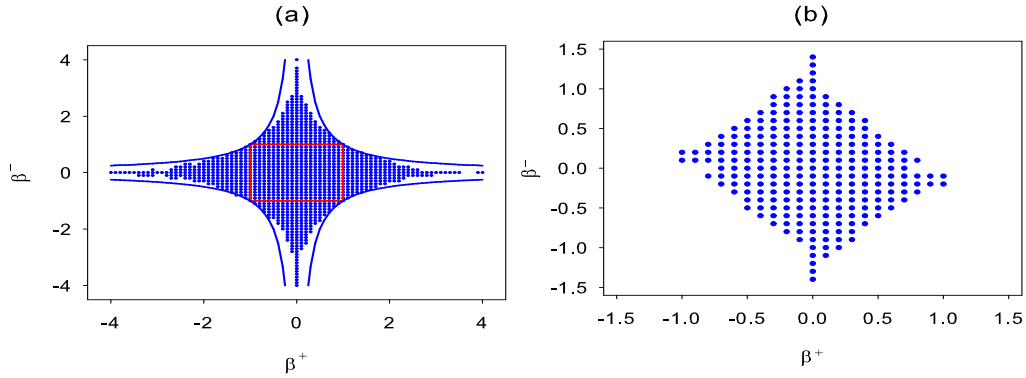
For the second equation, we have

$$\begin{aligned} Z_{\bar{Q},t} &= (\phi_{\bar{Q}} + \psi_{\bar{Q},0} \varepsilon_t) Z_{1,t-1} + \sum_{j=1}^{\bar{P}} (\phi_{\bar{Q}+j} + \psi_{\bar{Q},j} \varepsilon_t) Z_{\bar{Q}+j,t-1} \\ &\quad + \{(\phi_{\bar{Q}} \theta_0 + \theta_{\bar{Q}}) \varepsilon_t + \theta_0 \psi_{\bar{Q},0} \varepsilon_t^2\}, \end{aligned}$$

while

$$Z_{\bar{Q}+1,t} = Z_{1,t-1} + \theta_0 \varepsilon_t, \quad Z_{\bar{Q}+i,t} = Z_{\bar{Q}+i-1,t-1}.$$

From these results, it can be seen that the general form of  $\mathbf{A}_t$  and  $\mathbf{B}_t$  has the structure  $\mathbf{A}_t = \mathbf{A}_0 + \mathbf{A}_1 \varepsilon_t$  and  $\mathbf{B}_t = \mathbf{B}_0 + \mathbf{B}_1 \varepsilon_t + \mathbf{B}_2 \varepsilon_t^2$ .



**Figure 3.2:** (a) Contour plot of the invertibility region for the asMA(1) model for  $T = 100$  (blue dots) and the function  $|\beta^-| = 1/|\beta^+|$  (blue solid lines), and Wecker's invertibility region  $|\beta^+| < 1$  and  $|\beta^-| < 1$  (red solid lines); (b) Invertibility region based on the criterion  $|\mathbb{E}_k| \rightarrow 0$  as  $t \rightarrow \infty$  following from the model specification in Exercise 3.6 of the main text.

- 3.6** (a) Using Algorithm 3.1 of the main text, Figure 3.2(a) shows a graphical representation of the empirical invertibility region (blue dots) for a time series of size  $T = 100$ , based on 1,000 MC replications.
- (b) Figure 3.2(a) also includes the analytical invertibility region (red solid lines) derived by Wecker (1981). It is interesting to see that the empirical region indicates invertibility over a much wider range of parameter values than the one based on the conditions  $|\beta^+| < 1$  and  $|\beta^-| < 1$ . In fact, a simple approximation to the empirical invertibility region is given by the necessary condition  $|\beta^+||\beta^-| < 1$ . The blue solid lines in Figure 3.2(a) represent this latter condition graphically. Obviously, this invertibility region is much closer to the empirically found region than the region enclosed by the red solid lines.
- 3.7** (a) Rewrite the asMA(1) model in the form

$$\begin{aligned}
 Y_1 &= \varepsilon_1 \\
 (1 - \beta^+ B)Y_2 &= \varepsilon_2 \quad \text{if } \varepsilon_1 > 0, \\
 (1 - \beta^- B)Y_2 &= \varepsilon_2 \quad \text{if } \varepsilon_1 < 0, \\
 (1 - \beta^+ B + \beta^{+2} B^2)Y_3 &= \varepsilon_3 \quad \text{if } \varepsilon_1 > 0 \text{ and } \varepsilon_2 > 0, \\
 (1 - \beta^- B + \beta^{-2} B^2)Y_3 &= \varepsilon_3 \quad \text{if } \varepsilon_1 < 0 \text{ and } \varepsilon_2 < 0, \\
 (1 - \beta^- B + \beta^+ \beta^- B^2)Y_3 &= \varepsilon_3 \quad \text{if } \varepsilon_1 > 0 \text{ and } \varepsilon_2 < 0, \\
 (1 - \beta^+ B + \beta^+ \beta^- B^2)Y_3 &= \varepsilon_3 \quad \text{if } \varepsilon_1 < 0 \text{ and } \varepsilon_2 > 0, \\
 &\vdots
 \end{aligned}$$

with  $B$  the backward shift operator. Clearly, at any time point  $t$  the asMA(1) model has a “dual” NLAR representation with  $2^{t-1}$  different linear AR polynomials. When  $\{\varepsilon_t\}$  is a WN sequence with zero mean and symmetric marginal distribution, each polynomial arises with equal probability. Let  $\beta(k) = \beta^+ I(\varepsilon_k \geq 0) + \beta^- I(\varepsilon_k < 0)$ . Then, for arbitrary  $t$ , the resulting model has the

form

$$Y_t = \beta(t-1)Y_{t-1} - \beta(t-1)\beta(t-2)Y_{t-2} + \cdots - \beta(t-1)\cdots\beta(1)Y_1 + \varepsilon_t.$$

(b) As every combination  $\beta(t-1)\cdots\beta(t-k) = (\beta^+)^j(\beta^-)^{(k-j)}$  ( $j = 0, \dots, k$ ) arises with binomial probability  $\mathbb{P}(j) = [k!/(j!(k-j)!)](\frac{1}{2})^k$  the probability for each possible parameter combination in the autoregressive representation can be calculated. The density function  $\mathbb{P}(j)$  is symmetric. The expected value of the parameter combinations is given by  $\mathbb{E}_k = \sum_{j=0}^k \mathbb{P}(j)(\beta^+)^j(\beta^-)^{k-j}$ . Hence, another way to define invertibility for the asMA(1) model is to require that  $|\mathbb{E}_k| \rightarrow 0$  as  $k \rightarrow \infty$ . Figure 3.2(b) shows a graphical representation of the invertibility contour for  $k = 50$ .

(c) Assume  $|\beta_i| < |\beta(\varepsilon_{t-1})|$  ( $i = 1, 2$ ). The reconstruction errors are

$$|e_t| \leq |\beta(\varepsilon_{t-1})\varepsilon_{t-1} - \beta(\widehat{\varepsilon}_{t-1})\widehat{\varepsilon}_{t-1}|.$$

If  $\varepsilon_{t-1}, \widehat{\varepsilon}_{t-1} \geq 0$ , i.e. both values lie on one side of the threshold, we have  $\beta(\varepsilon_{t-1}) = \beta(\widehat{\varepsilon}_{t-1})$ . Hence,

$$|e_t| \leq |\beta(\varepsilon_{t-1})||\varepsilon_{t-1} - \widehat{\varepsilon}_{t-1}| = |\beta(\varepsilon_{t-1})||e_{t-1}|.$$

If, on the other hand,  $\varepsilon_{t-1} \geq 0$  and  $\widehat{\varepsilon}_{t-1} < 0$ , we have

$$\begin{aligned} |e_t| &\leq |\beta(\varepsilon_{t-1})||\varepsilon_{t-1}| + |-\beta(\widehat{\varepsilon}_{t-1})||\widehat{\varepsilon}_{t-1}| \\ &\leq |\beta(\varepsilon_{t-1})|(|\varepsilon_{t-1}| + |\widehat{\varepsilon}_{t-1}|) = |\beta(\varepsilon_{t-1})|(\varepsilon_{t-1} - \widehat{\varepsilon}_{t-1}) = |\beta(\varepsilon_{t-1})||e_{t-1}|. \end{aligned}$$

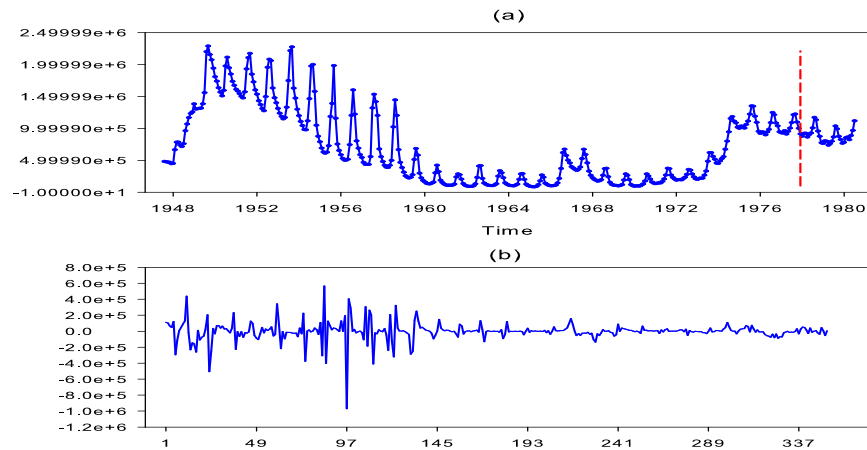
Finally, if  $\varepsilon_{t-1} < 0$  and  $\widehat{\varepsilon}_{t-1} \geq 0$ , one may proceed exactly as above. Thus, in summary, we have  $|e_t| \leq |\beta(\varepsilon_{t-1})||e_{t-1}|$ .

Now, let  $|b_i| < |\beta(\varepsilon_{t-1})|$  ( $i = 1, 2$ ), and assuming that the root  $\lambda_1$ , say, of the first-order equation  $f(z) = z - |\beta(\varepsilon_{t-1})| = 0$  is smaller than unity. Then  $\{Y_t, t \in \mathbb{Z}\}$  is invertible in the sense that  $\mathbb{E}|e_t| \rightarrow 0$  as  $t \rightarrow \infty$ . This latter result can be seen as follows. Let  $\alpha_i = |\beta(\varepsilon_{t-1})|\alpha_{i-1}$  ( $i = 1, 2, \dots$ ) with  $\alpha_0 = \mathbb{E}|e_0|$ . By repeated substitution, and assuming the model is correctly specified, it follows that  $\mathbb{E}|e_t| \leq \alpha_t \forall t \geq 1$ . Now, for all  $t < t_0$ , and using  $|e_t| \leq |\beta(\varepsilon_{t-1})||e_{t-1}|$ , we have

$$\mathbb{E}|e_{t_0}| \leq \mathbb{E}(|\beta(\varepsilon_{t-1})||e_{t_0-1}|) = |\beta(\varepsilon_{t-1})|\mathbb{E}|e_{t_0-1}| \leq |\beta(\varepsilon_{t-1})|\alpha_{t_0-1} = \alpha_{t_0}.$$

Hence, the sequence  $\{\mathbb{E}|e_t|\}$  weakly majorizes (or dominates) from above the sequence  $\{\alpha_t\}$ . The homogeneous solution of the first-order difference equation  $\alpha_i - |\beta(\varepsilon_{t-1})|\alpha_i = 0$  is given by  $\alpha_t = A_1\lambda_1^t$ , where  $A_1$  is a constant which follows from solving  $\alpha_0 = \mathbb{E}|e_0|$ . Because  $|\lambda_1| < 1$ , we have  $|\alpha_t| \leq |A_1||\lambda_1|^t$ , and consequently  $\alpha_t \rightarrow 0$  as  $t \rightarrow \infty$ , which completes the proof. Hence, it is clear that the invertibility region has the form  $|\beta_i| < 1$ , ( $i = 1, 2$ ), and  $|\beta_1 + \beta_2| < 1$ . However, since  $|\beta_i| < 1$ , the last inequality can occur only if  $|\beta_1| + |\beta_2| < 1$ . Note that the resulting invertibility region is identical to the one in Figure 3.2(b).

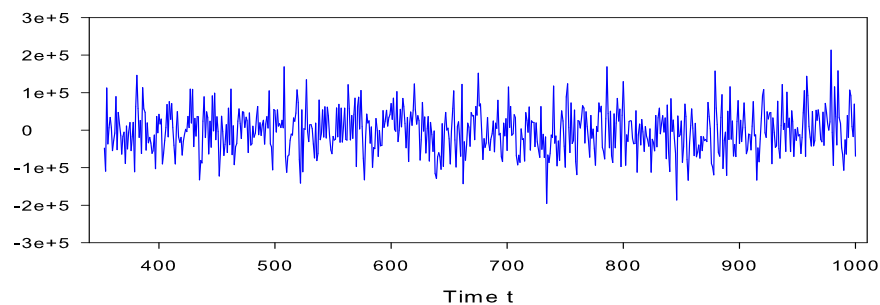
**3.8** Figure 3.3(a) shows a plot of the time series  $\{X_t\}_{t=1}^{389}$ . Note the seasonal pattern in the data with a period of 12 months. Moreover, we see a steep upward trend for the early time period 1948 – 1950, and a slow downward trend for the period 1950 – 1964. Hence, it makes sense to analyze the series  $Y_t = (1-B)(1-B^{12})X_t$ ; see Figure 3.3(b).



**Figure 3.3:** (a) Monthly West German unemployment time series  $(X_t)$  covering the time period 1948–1980 (389 observations). The red medium dashed vertical line denotes the start of the out-of-sample forecasting period; (b) The differenced time series  $\{Y_t = (1 - B)(1 - B^{12})X_t\}_{t=1}^{365}$ .

Note that the parameters of the fitted BL model are very small. The characteristic roots of the polynomial  $\phi(z) = 1 - 0.0874z + 0.1261z^2 - 0.0426z^9 - 0.2556z^{11} + 0.5067z^{12}$  are lying outside the unit circle (the reader may want to check this statement), indicating that the process is stationary. Now, using the invertibility check in Algorithm 3.1, we observe that the fitted subset BL model is indeed invertible. In particular, in Figure 3.4 we see that the series  $\{\hat{e}_t\}_{t=353}^{1,000}$  remains stable for all lead periods up to time  $t = N = 1,000$ .

To ease reproduction of the above results the MATLAB code is given below. The code can be quite easily modified to assess invertibility of other nonlinear models.



**Figure 3.4:** Plot of  $\{e_t\}_{t=353}^{1,000}$  versus time  $t$  for the subset BL model fitted by Subba Rao and Gabr (1984) to the differenced monthly West German unemployment time series.

```

function[res,hN]=Empinv(data,p,pAR,pAR1,f,f1,ip,ip1,iq1,cnst,s2,N)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 3.8 (see also Algorithm 3.1)
% File: Empinv.m
%
% Checking empirical invertibility of a mixed (subset) BL model
%       with a linear AR part.
% INPUT:
% data   = time series
% p      = maximum AR order
% pAR    = number AR lags;
% pAR1   = number of mixed lag AR terms
% f      = (pAR * 1) parameter vector for pure AR part
%       (Note: input assumes all lags are on RHS)
% f1     = (pAR1 * 1) parameter vector for mixed BL part
% ip     = (pAR * 1) vector with lags for pure AR part
% ip1    = (pAR1 * 1) vector with lags for  $Y_{\{t-ip1\}}*e_{\{t-iq1\}}$ 
% iq1    = (qBL1 * 1) vector with lags for  $Y_{\{t-ip1\}}*e_{\{t-iq1\}}$ 
% cnst   = constant
% s2     = residual variance
% N      = number of "after sample" observations
%
% OUTPUT:
% res    = "after sample" residuals
% hN     = estimate of  $E[e^{\{2\}}_{\{t\}}]$ 
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
y      = data;
T      = length(data);
qBL1 = pAR1;

% Obtain "within sample" residuals
e = zeros(p,1);
for it=(p+1):T
    v1 = 0;
    v2 = 0;
    for j=1:pAR
        v1 = v1+f(j)*y(it-ip(j)); % Linear AR part
    end
    for j=1:qBL1
        v2 = v2+f1(j)*y(it-ip1(j))*e(it-iq1(j));
    end
    e(it) = y(it)-v1-v2-cnst;
end

% Generate "after sample" observations  $Y_{\{T+1\}}, \dots, Y_{\{N\}}$ 
% using past observations of  $Y(t)$ , past values of  $e(it)$ ,
% and values of  $\epsilon\text{-tilde}(t)$ 
rng('default') % The same random numbers are produced as if you
                % restarted MATLAB

```



```

u3      = normrnd(0,1,N,1);

for it=T+1:N
    e(it) = sqrt(s2)*u3(it); % series epsilon-tilde_{t}
    v1    = 0;
    v2    = 0;
    for j=1:pAR
        v1 = v1+f(j)*y(it-ip(j));
    end
    for j=1:qBL1
        v2 = v2+f1(j)*y(it-ip1(j))*e(it-iq1(j));
    end
    y(it) = e(it)+v1+v2+cnst;
end

% Obtain "after sample" residuals and hN=sum(e^2)/(it-T).
% For small values of the ratio (ht/it), the model is assumed to be
% invertible
hN = zeros(N,1);
ht = 0;
for it=T+1:N
    v1 = 0;
    v2 = 0;
    for j=1:pAR
        v1 = v1+f(j)*y(it-ip(j));
    end
    for j=1:qBL1
        v2 = v2+f1(j)*y(it-ip1(j))*e(it-iq1(j));
    end
    res(it) = y(it)-v1-v2-cnst; % (Y-tilde)-(fitted "after sample" values)
    ht      = ht+(res(it)-u3(it))*(res(it)-u3(it));
    hN(it)  = ht/(it-N);
end

```

## Chapter 4

4.1 We begin by noting that from the definition of  $f_Y(\cdot, \cdot)$ , we have

$$f_Y(\omega_1, \omega_2) = f_Y(\omega_2, \omega_1), \quad f_Y(-\omega_1, -\omega_2) = f_Y^*(\omega_1, \omega_2),$$

where  $*$  denote complex conjugate. Furthermore, since  $\exp(2\pi i) = 1$ , we have for any two integers  $k, \ell$ ,

$$f_Y(\omega_1 + k, \omega_2) = f_Y(\omega_1, \omega_2 + k) = f_Y(\omega_1 + k, \omega_2 + \ell) = f_Y(\omega_1, \omega_2).$$

From these facts, all values of  $f_Y(\cdot, \cdot)$  can be obtained for  $0 \leq \omega_1 \leq \omega_2 \leq 1$ . We next show that in fact we only need  $f_Y(\cdot, \cdot)$  for  $0 \leq \omega_1 \leq \omega_2 \leq \frac{1}{2}$ . If  $\omega_1$  and  $\omega_2$  are both in  $[\frac{1}{2}, 1]$ , we use

$$f_Y(\omega_1, \omega_2) = f_Y^*(-\omega_1, -\omega_2) = f_Y^*(1 - \omega_1, 1 - \omega_2),$$

since  $1 - \omega_1$  and  $1 - \omega_2$  are both in  $[0, \frac{1}{2}]$ . If  $\omega_1 \in [0, \frac{1}{2}]$  while  $\omega_2 \in [\frac{1}{2}, 1]$ , we use

$$f_Y(\omega_1, \omega_2) = f_Y(\omega_1, -\omega_1 - \omega_2) = f_Y(-\omega_1 - \omega_2, \omega_2), \quad (*)$$

which follows from the symmetry relations (4.2). For example,

$$\begin{aligned} f_Y(\omega_1, -\omega_1 - \omega_2) &= \sum_{u, v=-\infty}^{\infty} \gamma_Y(u, v) \exp\{-2\pi i u \omega_1 - 2\pi i v(-\omega_1 - \omega_2)\} \\ &= \sum_{u, v=-\infty}^{\infty} \gamma_Y(u - v, -v) \exp\{-2\pi i(u - v)\omega_1 \\ &\quad - 2\pi i(u - v)\omega_1 - 2\pi i(-v)\omega_2\} \\ &= f_Y(\omega_1, \omega_2). \end{aligned}$$

The symmetry relations for  $\gamma_Y(\cdot, \cdot)$  follow from writing the order of the three terms in its defining expectation in three different orders, which of course doesn't change the value of the expectation.

If  $\omega_1 + \omega_2 \in [\frac{1}{2}, 1]$ , then  $1 - \omega_1 - \omega_2 \in [0, \frac{1}{2}]$ , and we can get  $f_Y(\omega_1, \omega_2)$  from  $f_Y(\omega_1, 1 - \omega_1 - \omega_2)$  as in the first equality in (\*) while if  $\omega_1 + \omega_2 \in [1, 1\frac{1}{2}]$ , then  $2 - \omega_1 - \omega_2 \in [\frac{1}{2}, 1]$  and we can get  $f_Y(\omega_1, \omega_2)$  using the second equality in (\*) by

$$\begin{aligned} f_Y(\omega_1, \omega_2) &= f_Y(2 - \omega_1 - \omega_2, \omega_2) = f_Y^*(1 - (2 - \omega_1 - \omega_2), 1 - \omega_2) \\ &= f_Y^*(\omega_1 + \omega_2 - 1, 1 - \omega_2), \end{aligned}$$

and  $\omega_1 + \omega_2 - 1$  and  $1 - \omega_2$  are both in  $[0, \frac{1}{2}]$ . Finally, the fact the  $f_Y(\omega_1, \omega_2) = f_Y(\omega_1, 1 - \omega_1 - \omega_2)$  gives that we only need  $f_Y(\omega_1, \omega_2)$  for  $\omega_2 = 1 - \omega_1 - \omega_2$ , i.e., for  $\omega_2 = (1 - \omega_1)/2$ .

- 4.2** (a) The process is given by  $Y_t = \varepsilon_t + \beta Y_{t-2} \varepsilon_{t-1}$  with  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$  which is identical to the BL process considered in Exercise 1.2. We already showed that  $\mathbb{E}(Y_t Y_{t-k}) = 0$  for  $k > 0$  and that  $\text{Var}(Y_t) = \gamma_Y(0) = \sigma_\varepsilon^2 / (1 - \beta^2 \sigma_\varepsilon^2)$ . Now

$$\begin{aligned} \mathbb{E}(Y_t Y_{t-k} Y_{t-\ell}) &= \mathbb{E}[(\varepsilon_t + \beta Y_{t-2} \varepsilon_{t-1})(\varepsilon_{t-k} + \beta Y_{t-k-2} \varepsilon_{t-k-1}) Y_{t-\ell}] \\ &= \mathbb{E}[(\varepsilon_t \varepsilon_{t-k} + \beta Y_{t-2} \varepsilon_{t-1} \varepsilon_{t-k} + \beta Y_{t-k-2} \varepsilon_{t-k-1} \varepsilon_{t-k-1} \\ &\quad + \beta^2 Y_{t-2} Y_{t-k-2} \varepsilon_{t-1} \varepsilon_{t-k-1}) Y_{t-\ell}] \\ &= \begin{cases} \beta \sigma_\varepsilon^2 \mathbb{E}(Y_{t-2}^2) = \beta \sigma_\varepsilon^4 / (1 - \beta^2 \sigma_\varepsilon^2) & \text{if } k = 1, \ell = 2, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Evaluating  $\mathbb{E}(Y_t^2 Y_{t-1}^2)$  gives, because of stationarity,

$$\begin{aligned} \mathbb{E}(Y_t^2 Y_{t-1}^2) &= \mathbb{E}(\varepsilon_t^2 Y_{t-1}^2 + 2\beta \varepsilon_t \varepsilon_{t-1} Y_{t-1}^2 Y_{t-2} + \beta^2 Y_{t-1}^2 \varepsilon_{t-1}^2 Y_{t-2}^2) \\ &= \sigma_\varepsilon^2 \mathbb{E}(Y_t^2) + \beta^2 \mathbb{E}\{(\varepsilon_{t-1}^2 + 2\beta \varepsilon_{t-1} Y_{t-3} \varepsilon_{t-2} \\ &\quad + \beta^2 Y_{t-3}^2 \varepsilon_{t-2}^2) \varepsilon_{t-1}^2 Y_{t-2}^2\} \\ &= \sigma_\varepsilon^2 \mathbb{E}(Y_t^2) + 3\beta^2 \sigma_\varepsilon^4 \mathbb{E}(Y_t^2) + \beta^4 \sigma_\varepsilon^2 \mathbb{E}(Y_{t-2}^2 Y_{t-3}^2 \varepsilon_{t-2}^2). \end{aligned}$$

Now, consider the third term in the above expression. After expanding  $Y_{t-2}^2$

first and then expanding  $Y_{t-3}^2$  we successively have, because of stationarity,

$$\begin{aligned}\mathbb{E}(Y_{t-2}^2 Y_{t-3}^2 \varepsilon_{t-2}^2) &= \mathbb{E}[(\beta^2 Y_{t-4}^2 \varepsilon_{t-4}^2 + \varepsilon_{t-2}^2 + 2\beta \varepsilon_{t-2} \varepsilon_{t-4} Y_{t-4}) Y_{t-3}^2 \varepsilon_{t-2}^2] \\ &= \beta^2 \mathbb{E}(Y_{t-4}^2 Y_{t-3}^2 \varepsilon_{t-4}^2 \varepsilon_{t-2}^2) + 3\sigma_\varepsilon^4 \mathbb{E}(Y_t^2) \\ &= \beta^2 \mathbb{E}[Y_{t-4}^2 (\beta^2 Y_{t-5}^2 \varepsilon_{t-5}^2 + \varepsilon_{t-3}^2 + 2\beta \varepsilon_{t-3} \varepsilon_{t-5} Y_{t-5}) \\ &\quad \times \varepsilon_{t-4}^2 \varepsilon_{t-2}^2] + 3\sigma_\varepsilon^4 \mathbb{E}(Y_t^2) \\ &= \beta^2 [\beta^2 \sigma_\varepsilon^4 \mathbb{E}(Y_{t-4}^2 Y_{t-5}^2 \varepsilon_{t-4}^2) + 3\sigma_\varepsilon^6 \mathbb{E}(Y_t^2)] + 3\sigma_\varepsilon^4 \mathbb{E}(Y_t^2).\end{aligned}$$

So,

$$(1 - \beta^4 \sigma_\varepsilon^4) \mathbb{E}(Y_{t-2}^2 Y_{t-3}^2 \varepsilon_{t-2}^2) = (3\beta^2 \sigma_\varepsilon^6 + 3\sigma_\varepsilon^4) \mathbb{E}(Y_t^2).$$

Substituting the above result in  $\mathbb{E}(Y_t^2 Y_{t-1}^2)$  gives

$$\begin{aligned}\mathbb{E}(Y_t^2 Y_{t-1}^2) &= (1 + 3\beta^2 \sigma_\varepsilon^2) \frac{(1 - \beta^2 \sigma_\varepsilon^2) \sigma_\varepsilon^2}{(1 - \beta^2 \sigma_\varepsilon^2)} \mathbb{E}(Y_t^2) \\ &\quad + 3\beta^4 \frac{(\beta^2 \sigma_\varepsilon^6 + \sigma_\varepsilon^4) \sigma_\varepsilon^2}{(1 + \beta^2 \sigma_\varepsilon^2)(1 - \beta^2 \sigma_\varepsilon^2)} \mathbb{E}(Y_t^2) \\ &= (1 + 2\beta^2 \sigma_\varepsilon^2 - 3\beta^4 \sigma_\varepsilon^4) \mathbb{E}(Y_t^2) + 3\beta^4 \sigma_\varepsilon^4 \mathbb{E}(Y_t^2) \\ &= (1 + 2\beta^2 \sigma_\varepsilon^2) \mathbb{E}(Y_t^2) \\ &= (1 + 2\beta^2 \sigma_\varepsilon^2) \frac{\sigma_\varepsilon^4}{(1 - \beta^2 \sigma_\varepsilon^2)^2}.\end{aligned}$$

(b) Because of stationarity, the one-step ahead MSFE is given by

$$\begin{aligned}\text{MSFE}(1) &= \mathbb{E}[(Y_{t+1} - Y_{t+1|t}^Q)^2] = \mathbb{E}[(Y_{t+1} - c_{1,2} Y_t Y_{t-1})^2] \\ &= \mathbb{E}(Y_t^2) - 2c_{1,2} \mathbb{E}(Y_{t+1} Y_t Y_{t-1}) + c_{1,2}^2 \mathbb{E}(Y_t^2 Y_{t-1}^2).\end{aligned}$$

Minimizing with respect to  $c_{1,2}$  gives

$$c_{1,2} = \frac{\mathbb{E}(Y_{t+1} Y_t Y_{t-1})}{\mathbb{E}(Y_t^2 Y_{t-1}^2)} = \frac{\beta \sigma_\varepsilon^4}{1 - \beta^2 \sigma_\varepsilon^2} \frac{(1 - \beta^2 \sigma_\varepsilon^2)^2}{\sigma_\varepsilon^4 (1 + 2\beta^2 \sigma_\varepsilon^2)} = \beta \frac{1 - \beta^2 \sigma_\varepsilon^2}{1 + 2\beta^2 \sigma_\varepsilon^2}.$$

(c) The variance of the error of the best quadratic predictor is

$$\begin{aligned}\mathbb{E}(Y_t - Y_{t+1|t}^Q)^2 &= \sigma_Y^2 - \frac{\beta^2 (1 - \beta^2 \sigma_\varepsilon^2)^2 \sigma_\varepsilon^4 (1 + 2\beta^2 \sigma_\varepsilon^2)}{(1 + 2\beta^2 \sigma_\varepsilon^2)^2 (1 - \beta^2 \sigma_\varepsilon^2)^2} \\ &= \sigma_Y^2 - \frac{\beta^2 \sigma_\varepsilon^4}{1 + 2\beta^2 \sigma_\varepsilon^2} \\ &= \sigma_Y^2 \left[ 1 - \frac{\sigma_\varepsilon^2}{\sigma_Y^2} \left( \frac{\beta^2 \sigma_\varepsilon^2}{1 + 2\beta^2 \sigma_\varepsilon^2} \right) \right].\end{aligned}$$

Let  $x = \beta^2 \sigma_\varepsilon^2$ , then the first derivative of  $f(x) = \sigma_Y^2 [1 - (\sigma_\varepsilon^2 / \sigma_Y^2) \{x / (1 + 2x)\}]$  is  $f'(x) = (\sigma_\varepsilon^2 / \sigma_Y^2) \{3 / (1 + 2x)^2\}$ . Thus, the maximum reduction relative to  $\sigma_Y^2$  is achieved when  $(1 + 2x) = \sqrt{3}$ , or  $\beta^2 \sigma_\varepsilon^2 = (\sqrt{3} - 1) / 2$ ; i.e. when the process is normalized the maximum reduction is  $1 - (\sqrt{3} - 1) / 2 = 1 - 0.866 = 0.134$ .

**4.3** (a) The process is given by  $Y_t = \varepsilon_t(1 + \varepsilon_{t-1}) + (\eta_t^2 - 1)$  with  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ ,  $\{\eta_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ , and both error processes are independent of each other. To derive the

bispectrum we first need to obtain expressions for the 3rd-order cumulants. In particular,

$$\begin{aligned}\gamma_Y(0, 0) &= \mathbb{E}(Y_t^3) = \mathbb{E}[\{(\varepsilon_t + \varepsilon_t \varepsilon_{t-1}) + (\eta_t^2 - 1)\}^3] \\ &= \mathbb{E}[(\varepsilon_t + \varepsilon_t \varepsilon_{t-1})^3 + 3(\varepsilon_t + \varepsilon_t \varepsilon_{t-1})^2(\eta_t^2 - 1) \\ &\quad + 3(\varepsilon_t + \varepsilon_t \varepsilon_{t-1})(\eta_t^2 - 1)^2 + (\eta_t^2 - 1)^3] \\ &= 0 + 0 + 0 + (15 - 3 \times 3 + 3 - 1) = 8,\end{aligned}$$

using the fact that for a unit normal distribution, symmetrical about mean zero,  $\mathbb{E}(Y_t^r) = 0$  if  $r$  is odd, and  $\mathbb{E}(Y_t^r) = (r-1)(r-3)\cdots 3 \cdot 1$  if  $r$  is even. Next,

$$\begin{aligned}\gamma_Y(0, -1) &= \mathbb{E}(Y_t^2 Y_{t-1}) = \mathbb{E}[\{(\varepsilon_t + \varepsilon_t \varepsilon_{t-1})^2 + (\eta_t^2 - 1)\}^2 \\ &\quad + 2(\varepsilon_t + \varepsilon_t \varepsilon_{t-1})(\eta_t^2 - 1)\{(\varepsilon_{t-1} + \varepsilon_{t-1} \varepsilon_{t-2}) + (\eta_{t-1}^2 - 1)\}] \\ &= \mathbb{E}[(\varepsilon_t^2 + 2\varepsilon_t^2 \varepsilon_{t-1} + \varepsilon_t^2 \varepsilon_{t-1}^2)(\varepsilon_{t-1} + \varepsilon_{t-1} \varepsilon_{t-2}) \\ &\quad + (\eta_t^4 - 2\eta_t^2 + 1)(\varepsilon_{t-1} + \varepsilon_{t-1} \varepsilon_{t-2}) \\ &\quad + 2(\varepsilon_t + \varepsilon_t \varepsilon_{t-1})(\varepsilon_{t-1} + \varepsilon_{t-1} \varepsilon_{t-2})(\eta_t^2 - 1) \\ &\quad + (\varepsilon_t^2 + 2\varepsilon_t \varepsilon_{t-1} + \varepsilon_t^2 \varepsilon_{t-1}^2)(\eta_{t-1}^2 - 1) \\ &\quad + (\eta_t^4 - 2\eta_t^2 + 1)(\eta_{t-1}^2 - 1) \\ &\quad + 2(\varepsilon_t + \varepsilon_t \varepsilon_{t-1})(\eta_t^2 - 1)(\eta_{t-1}^2 - 1)] \\ &= 2 + 0 + 0 + (1 + 1 - 1 - 1) + (3 - 2 + 1 - 3 + 2 - 1) + 0 \\ &= 2.\end{aligned}$$

Clearly,  $\gamma_Y(0, -1) = \gamma_Y(-1, 0) = \gamma_Y(1, 1)$ . Moreover, it can be shown that  $\gamma_Y(u, v) = 0 \forall |u|, |v| > 1$ . Hence, the bispectrum becomes

$$\begin{aligned}f_Y(\omega_1, \omega_2) &= \sum_{u, v=-\infty}^{\infty} \gamma_Y(u, v) \exp\{-2\pi i(\omega_1 u + \omega_2 v)\} \\ &= \gamma_Y(1, 1) \exp\{-i(\omega_1 + \omega_2)\} + \gamma_Y(-1, 0) \exp(i\omega_1) \\ &\quad + \gamma_Y(0, -1) \exp(i\omega_2) \\ &= 2[\exp\{-i(\omega_1 + \omega_2)\} + \exp(i\omega_1) + \exp(i\omega_2)] + 8.\end{aligned}$$

- (b) Using the bispectrum  $f_Y(\omega_1, \omega_2)$  and setting  $\alpha = \beta = 1/4$  and  $\gamma = 0$ , the left-hand side of (\*) is equal to

$$\begin{aligned}&f_Y(\alpha, \beta) f_Y(\gamma, 0) f_Y(-\alpha + \gamma, -\beta - \gamma) \\ &= [2\{\exp(-i\pi) + 2\exp(i\pi/2)\} + 8] \times [2\{\exp(0) + 2\exp(0)\} + 8] \\ &\quad \times [2\{\exp(i\pi) + 2\exp(-i\pi/2)\} + 8] = (6 + 4i) \times (14) \times (6 - 4i) = 728.\end{aligned}$$

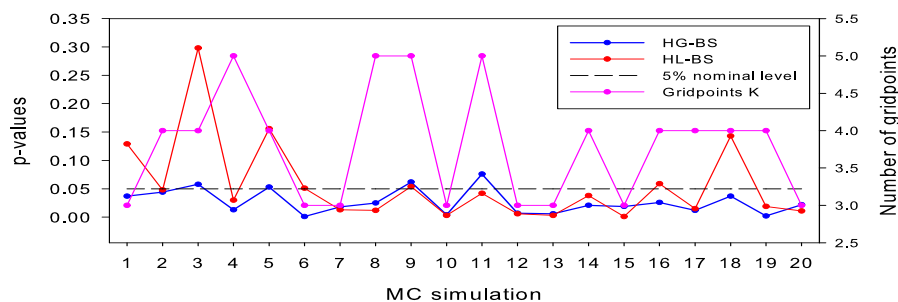
Similarly, the right-hand side of (\*) is equal to

$$\begin{aligned}&f_Y(\beta, \alpha) f_Y(0, -\alpha - \beta) f_Y(-\alpha + \gamma, -\gamma) = \\ &[2\{\exp(-i\pi/2) + \exp(i\pi/2) + \exp(0)\} + 8] \times [2\{\exp(i\pi) + \exp(0) \\ &\quad + \exp(-i\pi)\} + 8] \times [2\{\exp(i\pi/2) + \exp(-i\pi/2) + \exp(0)\} + 8] \\ &= 10 \times 6 \times 10 = 600.\end{aligned}$$

Therefore, it can be easily shown that the bispectrum (\*) cannot be represented in the triple-product form (4.10).

- 4.4 (a) The file Exercise44.r contains R functions for computing the bootstrap-based Gaussianity and linearity test statistics summarized in Algorithm 4.4. Based on the functions in this file, the following R code can be used for the MC simulations.

```
#####
# R code: Exercise 4.4(a)
# File: Main_4-4(a).r
#
# Main file for computing bootstrapped Gaussianity and linearity
# test statistics. Subfunctions are in the file "Exercise44.r".
#
# INPUT:
# data    = time series (here, USunemplmnt_first_dif.dat)
# repnum  = number of MC replications (default = 20)
# bootnum = number of BS replications (default = 1,000)
# M = M_{b} = bispectral bandwidth (default = 4)
# p = Bootstrap AR order drawn from the discrete uniform
#     distribution on {4,5,...,15}
# rows = number of gridpoints K, drawn from the discrete uniform
#     distribution on {3,4,5}
# c = const*M_{b}, with const a U[1.5,3] random deviate
#
# OUTPUT:
# Hinich's Gaussianity test statistic (1st column)
# Hinich's linearity test statistic (2nd column)
# BS version of Hinich's Gaussianity test, HG-BS (3rd column)
# BS version of Hinich's linearity test, HL-BS (4th column)
#####
UNEMP <- ts(data)
repnum <- 20
bootnum <- 1000
M.b <- 4
set.seed(100) # set position in pseudo-random sequence
UNEMPresults <- array(dim=c(repnum,4))
parameters <- array(dim=c(repnum,4))
for(i in 1:repnum){
  rows <- round(runif(1,2.5,5.5))
  const <- runif(1,1.5,3)
  M.s <- const*M.b
  ar.order <- round(runif(1,3.5,15.5))
  ht <- Hinich.test(UNEMP,rows,M.b,M.s)
  bt <- boot.test(bootnum,UNEMP,ar.order,rows,M.b,M.s)
  UNEMPresults[i,] <- c(ht$p_G,ht$p_L,bt$boot.p_G,bt$boot.p_L)
  parameters[i,] <- c(rows,M.b,M.s,ar.order)
  print(i)
}
reject <- double(4) # Rejection of null hypothesis
for(j in 1:4){
  sumtt <- 0
```



**Figure 4.1:** U.S. unemployment rate data set.  $P$ -values of the HG–BS test statistic (blue solid line), and the HL–BS test statistic (red solid line). Number of selected gridpoints  $K$  (placed on the right side of the graph) versus the number of MC simulations.

```

for(i in 1:repnum){
  if(UNEMResults[i,j]<0.05){#Nominal significance level=5%
    sumtt <- sumtt +1
  }
  reject[j] <- sumtt/repnum
}

```

Figure 4.1 shows  $p$ -values of the bootstrapped version of Hinich’s Gaussianity (HG–BS) test statistic, and  $p$ -values of the bootstrapped version of Hinich’s linearity test statistic (HL–BS). Both statistics are computed for 20 MC simulations each consisting of 1,000 BS replicates. Hinich’s Gaussianity test statistic *without bootstrapping* always has a  $p$ -value less than 0.05. The bootstrapped version of the Hinich test statistic rejects Gaussianity in 80% of the 20 randomly selected user-defined parameters (blue solid line). Similarly, Hinich’s linearity test statistic rejects the null hypothesis in 100% of the selected parameters, while the  $p$ -values of the bootstrapped version of this test statistic (red solid line) indicate that the first differences of the quarterly U.S. unemployment series follows a nonlinear time series process in 65% of the selected parameters. The averaged  $p$ -values are 0.0272 (HG–BS) and 0.0566 (HL–BS).

Of course, the outcomes of the test statistics depend on the simulated sample, and on the user-specified parameters in the simulation. Moreover, the test results will differ between different MC simulations. Note, however, that Spearman’s rank correlation coefficient is positive for the pair of variables ( $p$ -value HL–BS,  $K$ ) and ( $p$ -value HL–BS,  $p$ ) with  $p$ -values of 0.0179 and 0.0339, respectively. On the other hand, there are no significant relationships, at the 5% nominal level, between other combinations of  $p$ -values of the bootstrapped and non-bootstrapped test statistics, and the randomly selected parameters  $K$  and  $p$  respectively.

- (b) The file `goodnessfit.m` contains the MATLAB code for computing the goodness-of-fit test statistics  $AD_m^G$ ,  $CvM_m^G$ ,  $AD_m^L$ , and  $CvM_m^L$ ; see Algorithm 4.3. For ease of replication, the code is given below.

```

function [CvM,AD,CvML,ADL] = goodnessfit(x,cparm,nfft)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 4.4(b) (also used for computing Table 4.2)
% File: goodnessfit.m
%
% Goodness-of-fit tests statistics CvM and AD for testing
% Gaussianity and Linearity based on the paper by
% Jahan, N. and Harvill, J.L. (2008).
%
% INPUT:
% x      = input time series: column vector
% cparm = resolution (smoothing) parameter c;
%         0.5<c<1.0 [default=0.51], increasing c reduces the
%         variance, but increases the bias.
% nfft  = FFT length [default = 128]
%
% OUTPUT:
% CvM and AD test statistics for the Gaussianity hypothesis
% CvML and ADL test statistics for the linearity hypothesis
%
% References:
% Hinich, M.J. (1982). Testing for Gaussianity and linearity of
% stationary time series. Journal of Time Series Analysis, 3(3),
% 169-176.
% DOI: 10.1111/j.1467-9892.1982.tb00339.x.
% Jahan, N. and Harvill, J.L. (2008). Bispectral-based
% goodness-of-fit tests of Gaussianity and linearity of
% stationary time series. Communications in Statistics:
% Theory and Methods, 37(20), 3216-3227.
% DOI: 10.1080/03610920802133319.
% Stephens, M.A. (1986). Tests based on EDF statistics.
% In: D'Agostino, R.B. and Stephens, M.A. (Eds.):
% Goodness-of-Fit Techniques. Marcel Dekker, New York.
% ----- parameter checks: -----
if (min(size(x)) ~= 1)
    error('x: should be a vector')
end
nsamp = length(x);
if (exist('nfft') ~= 1)
    nfft = 128;
end

nrecs = floor(nsamp/nfft);
nrecs = max(nrecs,1);
ksamp = min(nfft,nsamp);
if (nfft > nsamp)
    disp(' glstat results are unreliable if zero-padding is severe: ')
    disp(' fft length= ',int2str(nfft),' data length=',int2str(nsamp))
end

```

```

if (exist ('cparm') ~= 1)
    cparm = 0.51;
end
if (cparm >= 1.0)
    error('cparm cannot be greater than or equal to 1.')
end
if (cparm <= 0.5 && nrecs == 1)
    error('cparm: for single segments: allowed range is (0.5,1.0)')
end

M = fix(nfft^cparm); M = M + 1 - rem(M,2);
Msmuth = M;
% ----- Estimate raw bispectrum -----
mrow = fix(nfft/3)+1; ncol = fix(nfft/2)+1;
F = zeros(mrow,ncol);
S = zeros(nfft,1);

mask = hankel(1:mrow,mrow:mrow+ncol-1);
for k=1:nrecs
    y = x(1:ksamp);
    xf = fft(y-mean(y), nfft);
    xc = conj(xf);
    S = S + xf .* xc; % Power spectrum
    F = F + xf(1:mrow) * xf(1:ncol).' .* ...
        reshape (xc(mask), mrow, ncol) ;
end
F = F/(nfft*nrecs); % `Raw' bispectrum
S = S/(nfft*nrecs); % `Raw' power spectrum

% ----- Zero out area outside principal domain -----
F(1:mrow,1:mrow) = triu(F(1:mrow,1:mrow));
Q = ones(mrow, ncol);
Q(1:mrow,1:mrow) = triu(Q(1:mrow,1:mrow)) + eye(mrow);

% The 2f1 + f2 = 1 line:
r = (rem(nfft,3) == 2);
for k=mrow+1:ncol-r
    j = k-mrow;
    Q(mrow-2*j+2:mrow, k) = zeros(2*j-1,1);
    F(:, k) = F(:,k).* Q(:,k);
    Q(mrow-2*j+1, k) = 2; % Factor 2 on boundary
end

F = F(2:mrow,2:ncol); % In principal domain, no dc terms
Q = Q(2:mrow,2:ncol);
mrow = mrow-1;
ncol = ncol-1;
% ----- Smooth the estimated spectrum and bispectrum -----

```



```

% Msmuth * Msmuth box-car smoother
% only every Msmuth term in the smoothed output is retained
% "independent" estimates

M = Msmuth;
m1 = rem(mrow,M); m2=rem(ncol,M);
m1 = - m1 + M * (m1 ~= 0);
m2 = - m2 + M * (m2 ~= 0);
F = [F,zeros(mrow,m2);zeros(m1,ncol+m2)];
Q = [Q,zeros(mrow,m2);zeros(m1,ncol+m2)];
k = size(F)/Msmuth;
k1 = k(1); k2 = k(2);

% Apply the boxcar smoother
% can replace the next ten lines or so with
% B = kron(eye(k1),ones(1,Msmuth))*F*kron(eye(k2),ones(Msmuth,1))
% /Msmuth^2;
% Q = kron(eye(k1),ones(1,Msmuth))*Q*kron(eye(k2),ones(Msmuth,1));

ind = 1:Msmuth;
B = zeros(k1,k2); Q1 = B;
for i=1:k1
    for j=1:k2
        t = F((i-1)*Msmuth+ind,(j-1)*Msmuth+ind );
        B(i,j) = mean(t(:));
        t = Q((i-1)*Msmuth+ind,(j-1)*Msmuth+ind );
        Q1(i,j) = sum(t(:));
    end
end
Q = Q1;

% -----
% At this point B corresponds to B in eqn.(2.6) of Hinich's
% paper and Q corresponds to the definition following eqn. (2.8)
% -----
M = Msmuth;
S = conv(S, ones(M,1))/M;
S = S(M+1:M:M+nfft-1);
S1 = S(1:k1) * S(1:k2)' .* hankel(S(2:k1+1),S(k1+1:k1+k2));
S1 = ones(k1,k2) ./ S1;
ind = find(Q > 0);
Q = Q(ind);
Bic = S1(ind) .* abs(B(ind)).^2; % Squared bicoherence
scale = 2 * (Msmuth^4) / nfft;
Xstat = scale * Bic ./ Q; %  $2|X_{m,n}|^2$ , eqn.(2.9)

% -----
% ----- Gaussianity EDF tests with fully -----
% specified distribution CvM and AD
% -----

Xsort = sort(Xstat);

```

```

q          = 0.5*erfc(-Xsort/sqrt(2)); % Normal CDF
Mlength = length(Xstat);           % Length M

ind = 0;
for i = 1:Mlength
    if(q(i)<1 && q(i)>0)           % Remove zeros and ones
        ind      = ind+1;
        qnew(ind) = q(i);
    end
end

mAD = length(qnew); % Effective length AD test (Not used)
m    = Mlength;     % Length CvM test

s1 = 0;
for i = 1:m
    s1 = s1+(q(i)-(2*i-1)/2/m)^2;
end
CvM1 = (s1+1/12/m)/m;

% Modification for a fully specified distribution
CvM = (CvM1-0.4/m+0.6/m^2)*(1+1/m);

if(CvM >= 0.00 && CvM < 0.0275);
    pvCvMG = 1- exp(-13.953 + 775.5*CvM - 12542.61*CvM^2);
elseif(CvM >= 0.0275 && CvM < 0.051);
    pvCvMG = 1- exp(-5.903 +179.546*CvM -1515.29*CvM^2);
elseif(CvM < 0.092 && CvM > 0.051);
    pvCvMG = exp(0.886 - 31.62*CvM + 10.897*CvM^2);
% in Table 4.9 of Stephens (1986); upper tail
elseif(CvM > 0.092);
    pvCvMG = exp(1.111 - 34.242*CvM + 12.832*CvM^2);
end

fx = normcdf(Xsort,mean(Xsort),std(Xsort));
i  = 1:m;
S  = sum((((2*i)-1)/m)*(log(fx)+log(1-fx(m+1-i))));
AD = -m-S;

% Correction factor for small sample sizes: case normal
ADa = AD*(1 + 0.75/m + 2.25/m^2);
if (ADa >= 0.00 && ADa < 0.200);
    pvADG = 1 - exp(-13.436 + 101.14*ADa - 223.73*ADa^2);
elseif (ADa >= 0.200 && ADa < 0.340);
    pvADG = 1 - exp(-8.318 + 42.796*ADa - 59.938*ADa^2);
elseif (ADa >= 0.340 && ADa < 0.600);
    pvADG = exp(0.9177 - 4.279*ADa - 1.38*ADa^2);
else (ADa >= 0.600 && ADa <= 13);
    pvADG = exp(1.2937 - 5.709*ADa + 0.0186*ADa^2);

```

```

end

disp(['CvM Gaussianity test= ', num2str(CvM), ', ...
p-value CvM = ', num2str(pvCvMG), ', AD Gaussianity test= ', ...
num2str(AD), ', p-value AD = ', num2str(pvADG)])
% -----
% ----- Linear EDF tests CvM and AD -----
% -----

df      = 2;
xmean  = mean(Xstat);
if(xmean>df)
    lam  = xmean-df;
else
    lam  = 0;
end
h = 1-(2*(df+lam)*(df+3*lam)/(3*(df+2*lam)^2)); % Exponent
Ytrans1 = (Xstat/(df+lam)); % Transformed variables
for i = 1:Mlength
    Ytrans(i) = (Ytrans1(i)^h);
end

muY1 = 1+h*(h-1)*(df+2*lam)/(df+lam)^2;
muY = muY1-h*(h-1)*(2-h)*(1-3*h)*((df+2*lam)^2)/(2*(df+lam)^4);

sigma2Y1 = h*h*(2*(df+2*lam))/((df+lam)^2);
sigma2Y = sigma2Y1*(1-(1-h)*(1-3*h)*(df+2*lam))/((df+lam)^2);

for i = 1:Mlength
    Ystand(i) = (Ytrans(i)-muY)/sqrt(sigma2Y);
end
Ysort = sort(Ystand);

qlin = 0.5*erfc(-Ysort/sqrt(2)); % Normal CDF
mlin = length(Ysort); % Effective length M

i = (1:mlin);
CvML = (1/12/mlin+sum((qlin-(2*i-1)/2/mlin).^2))/mlin;

% Modification of CvM for unknown mean and variance
% as given in Table 4.7 of Stephens (1986)
CvML = CvML*(1+0.5/mlin);

% Asymptotic p-values for modified CvM as given in Table 4.9
% of Stephens (1986)
if(CvML >= 0.00 && CvML < 0.0275);
    pvCvML = 1- exp(-13.953 + 775.5*CvML - 12542.61*CvML^2);
elseif(CvML >= 0.0275 && CvML < 0.051);
    pvCvML = 1- exp(-5.903 +179.546*CvML -1515.29*CvML^2);
elseif(CvML < 0.092 && CvML >0.051);

```

**Table 4.1:** *P-values of the Gaussianity and linearity test statistics for the first-differences of the quarterly U.S. unemployment rate.*

Gaussianity (G)			Linearity (L)					
GOF Tests <sup>(1)</sup>		Btstrp <sup>(2)</sup>	GOF Tests <sup>(1)</sup>		Btstrp <sup>(2)</sup>		MSFE <sup>(3)</sup>	
$AD_m^G$	$CvM_m^G$	$T^G$	$AD_m^L$	$CvM_m^L$	$Z_{IQR}^L$	$Z_{IDR}^L$	$Z_{80\%}^L$	$G_7^{T,K}$
$3.3 \times 10^{-7}$	0.028	0.013	0.106	0.826	0.043	0.008	0.054	0.872

```

pvCvML = exp(0.886 - 31.62*CvML + 10.897*CvML^2);
% in Table 4.9 of Stephens (1986); upper tail
elseif(CvML > 0.092);
    pvCvML = exp(1.111 - 34.242*CvML + 12.832*CvML^2);
end

fxs = normcdf(Ysort,mean(Ysort),std(Ysort))';
i = (1:m1in);
S = sum((((2*i)-1)/m1in)*(log(fxs)+log(1-fxs(m1in+1-i)))));
ADL = -m1in-S;

% Modification of AD test for unknown mean and variance
ADL = ADL*(1+0.75/m1in+2.25/(m1in^2));
if(ADL >= 0.00 && ADL < 0.200);
    pvADL = 1 - exp(-13.436 + 101.14*ADL - 223.73*ADL^2);
elseif (ADL >= 0.200 && ADL < 0.340);
    pvADL = 1 - exp(-8.318 + 42.796*ADL - 59.938*ADL^2);
elseif(ADL < 0.600 && ADL >= 0.340)
    pvADL = exp(0.9177 - 4.279*ADL - 1.38*ADL^2);
elseif(ADL >= 0.600 && ADL<= 13)
    pvADL=exp(1.2937 - 5.709*ADL + 0.0186*ADL^2);
end

disp(['CvM Linearity test = ',num2str(CvML),', p-value CvM = ',
...,num2str(pvCvML),', AD Linearity test = ', num2str(ADL), ',
..., p-value AD = ', num2str(pvADL)])

```

Using the MATLAB code for computing the test statistics  $AD_m^G$ ,  $CvM_m^G$ ,  $AD_m^L$ , and  $CvM_m^L$ , the nonlinear toolkit package for computing the test statistics  $T^G$ ,  $Z_{IQR}^L$ ,  $Z_{IDR}^L$ , and  $Z_{80\%}^L$ , and the MATLAB code for computing the Terdik–Máth test statistic  $G_{7,T}^{(K)}$ , one can obtain the  $p$ -values given in Table 4.1. Comparing the  $p$ -values of the bootstrapped test statistic  $T^G$ ,  $Z_{IQR}^L$ , and  $Z_{IDR}^L$  with the  $p$ -values 0.0272 (HG–BS) and 0.0566 (HL–BS) given in part (a), we see that they largely agree. Note, however, that the values of  $M$  are different. Indeed, replacing  $M \equiv M_b = 4$  by  $M_b = 23$  in the MC simulations, and using the same set of parameters as in part (a), the bootstrapped form of Hinich’s Gaussianity test statistic resulted in no rejections of the Gaussianity hypothesis, and the bootstrapped form of Hinich’s linearity test statistic rejected the null hypothesis for only 5% of the selected parameters. So, the test statistics are sensitive to

the choice of  $M_b$ .

**Remark:** A more formal comparison of the above linearity/Gaussianity test statistics can be based on the MATLAB function `simulation_tests.m`.

- 4.5 (a) To generate the time series in parts (a) and (b), add the following R functions to the set of functions used in Exercise 4.4. Note, these files are already included in the file `Exercise45.r`.

```
dataWN<-function(n){
  m <- 1000
  b <- 0
  evars <- rnorm(n+m,0,1)
  Data <- rep(b/(1),n+m) # starts with 0
  for(j in 2:(n+m)){Data[j]<- evars[j]}
  data <- Data[-(1:m)]
}

dataWNchi <- function(n){
  m <- 1000
  b <- 0
  evars <- rnorm(n+m,0,1)^2
  Data <- rep(b/(1),n+m) # starts with 0
  for(j in 2:(n+m)){Data[j]<- evars[j]}
  data <- Data[-(1:m)]
}

databil<-function(n){
  b <- 0.4 # Parameter value BL process
  m <- 1000
  evars <- rnorm(n+m,0,1)
  Data <- rep(b/(1),n+m) # starts with 0
  for(j in 2:(n+m)){Data[j]<-b*Data[j-1]*evars[j-1]+evars[j]}
  data <- Data[-(1:m)]
}
```

Next, the  $p$ -values of the test statistics can be generated by the following R code.

```
repnum <- 100
DATAresults <- array(,dim=c(repnum,4))
parameters <- array(,dim=c(repnum,4))
for(i in 1:repnum){
# MAKE APPROPRIATE CHANGES:databil = BL ; dataWN = WN
#                               dataWNchi = Chi-squared(1 df)
  data <- databil(250)
  bootnum <- 200
  rows <- 8
  M.b <- 4
  M.s <- 8
  ar.order <- 15
  ht <- Hinich.test(data,rows,M.b,M.s)
  bt <- boot.test(bootnum,data,ar.order,rows,M.b,M.s)
```

**Table 4.2:** Empirical rejection rates based on 100 MC simulations;  $T = 250$ ,  $B = 200$ .

DGP	$T_{\text{IQR}}^G$	$Z_{\text{IQR}}^L$	Btstrp	
			$T_{\text{IQR}}^G$	$Z_{\text{IQR}}^L$
i.i.d. $\mathcal{N}(0,1)$	0.22	0.11	0.05	0.03
i.i.d. $\chi_1^2$	1.00	0.55	1.00	0.06
Bilinear	1.00	1.00	1.00	0.98

```

DATAresults[i,] <- c(ht$p_G,ht$p_L,bt$boot.p_G,bt$boot.p_L)
# MAKE APPROPRIATE CHANGES:
print("Bilinear, T=250, rep = 100, Mb=4")
print(i)
}

```

The  $p$ -values are stored in the files DATAresults\_normal.dat, DATAresults\_chi.dat, and DATAresults\_bilinear. The empirical rejection rates for parts (a) and (b) can be computed using the following R code.

```

reject<- double(4)
for(j in 1:4){
  sumtt<- 0
  for(i in 1:repnum){
    if(DATAresults[i,j]<0.05){
      sumtt <- sumtt +1
    }
  }
  reject[j] <- sumtt/repnum
}

```

The results are summarized in Table 4.2. Ideally, the empirical rejection rate of the Gaussianity and linearity test statistics should be approximately equal to the nominal significance level  $\alpha = 0.05$ . Clearly, the bootstrapped test statistics do a much better job than the non-bootstrapped test statistics which are too high. For the  $\chi_1^2$  series, we see high empirical rejection rates of the Gaussianity tests in the bootstrap and non-bootstrap case, and low rejection rates of the linearity test in the bootstrap case. However, the rejection rates for  $Z_{\text{IQR}}^L$  in the non-bootstrap case are far too high. We noticed that for large values of  $T$  the empirical sizes of the test statistics improve, but they are still far from satisfactory. If plenty of computing time is available one may explore this option. Also, other types of nonlinear processes may be investigated.

- (b) The BL process (4.19) with  $\beta = 0.4$  is neither Gaussian nor linear. Hence, the null hypotheses of both Gaussianity and linearity are false. Observe, in all cases that the empirical rejection rates are high, as expected.

## Chapter 5

5.1 The ASTMA( $k$ ) model is given by

$$Y_t = \varepsilon_t + \theta_k \varepsilon_{t-k} + \psi_k \varepsilon_{t-k} G(\gamma \varepsilon_{t-k}), \quad \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2),$$

where the transition function  $G(\cdot)$  satisfies conditions (i) and (ii) in Section 5.1 of the main text. Assume, for simplicity, that  $\theta_k = 0$ . The LM-type test statistic for testing linearity against an ASTMA model now simplifies, as  $Y_t = \varepsilon_t$  under  $\mathbb{H}_0$ . It consists of testing  $\mathbb{H}_0^*: \beta = 0$  in the auxiliary regression model

$$Y_t = \beta Y_{t-k}^2 + \eta_t.$$

The LS estimator of  $\beta$  is  $\hat{\beta} = \sum_{t=k+1}^T (Y_t Y_{t-k}^2) / (\sum_{t=k+1}^T Y_{t-k}^4)$ , so that the  $t$ -statistic is given by

$$t_\beta = \frac{\hat{\beta}}{(\widehat{\text{Var}}(\hat{\beta}))^{1/2}} = \frac{\sum_{t=k+1}^T Y_t Y_{t-k}^2}{(\sum_{t=k+1}^T Y_{t-k}^4)^{1/2}}.$$

Replacing the numerator of  $t_\beta$  by its large sample approximation (under  $\mathbb{H}_0$ ) yields the WJ test statistic. Note that, with finite samples,  $t_\beta$  does *not* follow a Student  $t$  distribution with  $T - 1$  degrees of freedom under  $\mathbb{H}_0$ ; it is called a “pseudo- $t$ ” test statistic. The  $t_\beta$  statistic will, however, be asymptotically distributed as  $\mathcal{N}(0, 1)$  under quite weak conditions.

**5.2** It is convenient to rewrite the asAR( $p$ ) model in a more compact (matrix) form using the matrices

$$\mathbf{A}_\phi = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \phi_1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & & & \\ \phi_p & & \ddots & \ddots & & \\ \vdots & \ddots & & \ddots & 1 & 0 \\ 0 & \cdots & \phi_p & \cdots & \phi_1 & 1 \end{pmatrix}, \quad \mathbf{A}_\alpha = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & & & \\ \alpha_p & & \ddots & \ddots & & \\ \vdots & \ddots & & \ddots & 0 & 0 \\ 0 & \cdots & \alpha_p & \cdots & \alpha_1 & 0 \end{pmatrix}.$$

Let  $\mathbf{A}_{\phi,\alpha} = \mathbf{A}_\phi + \mathbf{A}_\alpha \mathbf{I}_{\varepsilon,T}$ , where  $\mathbf{I}_{\varepsilon,T}$  is a  $T \times T$  diagonal matrix defined as  $\text{diag}\{I(\varepsilon_1 \geq 0), \dots, I(\varepsilon_T \geq 0)\}$ . Then the asAR( $p$ ) model can be written as  $\mathbf{A}_{\phi,\alpha} \mathbf{y} = \boldsymbol{\varepsilon}$ , or  $\mathbf{y} = \mathbf{A}_{\phi,\alpha}^{-1} \boldsymbol{\varepsilon}$ . The derivative of the matrix  $\mathbf{A}_{\phi,\alpha}$  with respect to  $\theta_i$  ( $i = 1, 2, \dots, 2p$ ) is given by

$$\frac{\partial \mathbf{A}_{\phi,\alpha}}{\partial \theta_i} = \begin{cases} \mathbf{B}^i & (i = 1, \dots, p), \\ \mathbf{B}^i \mathbf{I}_{\varepsilon,T} + \mathbf{A}_\alpha \mathbf{D}_{\varepsilon,T} \frac{\partial \boldsymbol{\varepsilon}}{\partial \theta_i} & (i = p+1, \dots, 2p), \end{cases}$$

where  $\mathbf{B}$  is a  $T \times T$  matrix with ones on the first lower diagonal and all other elements equal to zero, and

$$\mathbf{D}_{\varepsilon,T} = \text{diag}\left(\frac{\partial I(\varepsilon_1 \geq 0)}{\partial \varepsilon_1}, \dots, \frac{\partial I(\varepsilon_T \geq 0)}{\partial \varepsilon_T}\right).$$

The concentrated log-likelihood function has the form

$$L_T(\boldsymbol{\theta}) = -\frac{1}{2} \log \left( \frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{T} \right).$$

Let  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}', \mathbf{0}')'$  be the vector of restricted estimates of  $\boldsymbol{\theta}$  with the associated vector of residuals  $\hat{\boldsymbol{\varepsilon}} = \mathbf{A}_{\hat{\boldsymbol{\alpha}}} \mathbf{y}$  and the vector of fitted values  $\hat{\mathbf{y}}$ . Then the components of the score vector are given by

$$\frac{\partial L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\mathbb{H}_0} = -\frac{1}{\sigma_\varepsilon^2} \left( \frac{\partial \mathbf{A}_{\phi,\alpha}^{-1}}{\partial \theta_i} \mathbf{y} \right)' \boldsymbol{\varepsilon} \Big|_{\mathbb{H}_0} = -\frac{1}{\hat{\sigma}_\varepsilon^2} \begin{cases} (\mathbf{A}_{\hat{\boldsymbol{\alpha}}}^{-1} \hat{\boldsymbol{\varepsilon}}_{-i})' \hat{\boldsymbol{\varepsilon}} & (i = 1, \dots, p), \\ (\mathbf{A}_{\hat{\boldsymbol{\alpha}}}^{-1} \hat{\boldsymbol{\varepsilon}}_{-i+p}^+)' \hat{\boldsymbol{\varepsilon}} & (i = p+1, \dots, 2p). \end{cases}$$

The outer-product estimate of the Fisher information matrix under  $\mathbb{H}_0$  is given by the  $p \times p$  matrix  $\widehat{\Sigma}_{22} - \widehat{\Sigma}_{21}\widehat{\Sigma}_{11}^{-1}\widehat{\Sigma}_{12}$  with elements

$$\widehat{\Sigma}_{11,ij} = \frac{\widehat{\mathbf{y}}_{-i}'\widehat{\Sigma}_{\varepsilon}^2\widehat{\mathbf{y}}_{-j}}{(\widehat{\varepsilon}'\widehat{\varepsilon})^2}, \quad \widehat{\Sigma}_{12,ij} = \frac{(\widehat{\mathbf{y}}_{-i}^+)' \widehat{\Sigma}_{\varepsilon}^2(\widehat{\mathbf{y}}_{-j})}{(\widehat{\varepsilon}'\widehat{\varepsilon})^2}, \quad \widehat{\Sigma}_{22,ij} = \frac{(\widehat{\mathbf{y}}_{-i}^+)' \widehat{\Sigma}_{\varepsilon}^2(\widehat{\mathbf{y}}_{-j}^+)}{(\widehat{\varepsilon}'\widehat{\varepsilon})^2} \quad (1 \leq i, j \leq p),$$

where  $\widehat{\Sigma}_{12} = \widehat{\Sigma}_{21}'$  and  $\widehat{\Sigma}_{\varepsilon}^2 = \text{diag}(\widehat{\varepsilon}_1^2, \dots, \widehat{\varepsilon}_T^2)$ . Given these results, the LM-type test statistic follows directly. Under  $\mathbb{H}_0$  the test statistic has a  $\chi_p^2$  distribution, as  $T \rightarrow \infty$ .

- 5.3** (a) An LM-type test statistic can be derived using the LM testing framework discussed in the preamble of Section 5.1 of the main text. Similar as before, the idea is to linearize the model using the first-order approximations

$$f_i(\boldsymbol{\alpha}'\mathbf{Y}_t) = g_{1,i}\boldsymbol{\alpha}'\mathbf{Y}_t \quad (i = 1, \dots, p) \quad \text{and} \quad g_j(\boldsymbol{\beta}'\mathbf{Y}_t) = h_{1,j}\boldsymbol{\beta}'\mathbf{Y}_t \quad (j = 1, \dots, q),$$

where  $g_{1,i} = (df_i(z)/dz)|_{z=0}$ , and  $h_{1,j} = (dg_j(z)/dz)|_{z=0}$ . Inserting these approximations in the model yields the auxiliary linear regression model

$$Y_t = \sum_{i=1}^p a_i Y_{t-i} + \sum_{j=1}^q b_j W_{j,t} + \sum_{i=1}^p \mathbf{c}'_i \mathbf{U}_{i,t} + \sum_{j=1}^q \mathbf{d}'_j \mathbf{V}_{j,t} + \eta_t,$$

where  $\mathbf{U}_{i,t} = Y_{t-i}\mathbf{Y}_t$ ,  $\mathbf{V}_{j,t} = W_{j,t}\mathbf{Y}_t$ ,  $\mathbf{c}_i = \phi_i g_{1,i}\boldsymbol{\alpha}_i$ , and  $\mathbf{d}_j = \theta_j h_{1,j}\boldsymbol{\beta}_j$ . Assume that the regressors in the above model are linearly independent, so that the linearity hypothesis holds if and only if  $\mathbf{c}_i = \mathbf{0}$  and  $\mathbf{d}_j = \mathbf{0}$ . It is convenient to write the above model in matrix form, i.e.

$$Y_t = \mathbf{x}'_{1,t}\boldsymbol{\tau}_1 + \mathbf{x}'_{2,t}\boldsymbol{\tau}_2 + \eta_t,$$

where

$$\mathbf{x}_{1,t} = (Y_{t-1}, \dots, Y_{t-p}, W_{1,t}, \dots, W_{q,t})', \quad \mathbf{x}_{2,t} = (\mathbf{U}'_{1,t}, \dots, \mathbf{U}'_{p,t}, \mathbf{V}'_{1,t}, \dots, \mathbf{V}'_{q,t})', \\ \boldsymbol{\tau}_1 = (a_1, \dots, a_p, b_1, \dots, b_q)' \quad \text{and} \quad \boldsymbol{\tau}_2 = (\mathbf{c}'_1, \dots, \mathbf{c}'_p, \mathbf{d}'_1, \dots, \mathbf{d}'_q)'.$$

Under  $\mathbb{H}_0$  we have

$$\frac{\partial \varepsilon_t}{\partial \boldsymbol{\alpha}_i} = -\phi_i g_{1,i} \mathbf{U}_{i,t} \quad \text{and} \quad \frac{\partial \varepsilon_t}{\partial \boldsymbol{\beta}_j} = -\theta_j h_{1,j} \mathbf{V}_{j,t}.$$

Then an explicit expression for the LM-type test statistic is given by

$$\text{LM}_T(\boldsymbol{\phi}, \boldsymbol{\theta}) = T \frac{\text{SEE}_0 - \text{SSE}_1(\boldsymbol{\phi}, \boldsymbol{\theta})}{\text{SSE}_0}.$$

Here  $\text{SSE}_0 = \sum_{t=1}^T \widehat{\varepsilon}_t^2$  is the residual sum of squares from the regression of  $Y_t$  on  $\mathbf{x}_{1,t}$ , using LS.

Furthermore,  $\text{SSE}_1(\boldsymbol{\phi}, \boldsymbol{\theta})$  is the residual sum of squares from an auxiliary least squares regression of  $\widehat{\varepsilon}_t$  on  $-\mathbf{x}_{1,t}$ ,  $-\phi_i g_{1,i} \mathbf{U}_{i,t}$  ( $i = 1, \dots, p$ ) and  $-\theta_j h_{1,j} \mathbf{V}_{j,t}$  ( $j = 1, \dots, q$ ). By assumption,  $\phi_i g_{1,i}$  and  $\theta_j h_{1,j}$  are all non-zero so that  $\text{SSE}_1(\boldsymbol{\phi}, \boldsymbol{\theta})$  is identical to the residual sum of squares  $\text{SSE}_1 = \sum_{t=1}^T \widehat{\eta}_t^2$  with  $\widehat{\eta}_t$  the residuals from the regression of  $\widehat{\varepsilon}_t$  on  $\mathbf{x}_{1,t}$  and  $\mathbf{x}_{2,t}$ . Thus, the LM-type test statistic is given by  $\text{LM}_T = T(\text{SEE}_0 - \text{SSE}_1)/\text{SSE}_0$ .

- (b) In this case the nonlinear model is given by

$$Y_t = \sum_{i=1}^p (a_i + \phi_i f_i(\boldsymbol{\alpha}'\mathbf{Y}_t)) Y_{t-i} + \sum_{j=1}^q (b_j + \theta_j g_j(\boldsymbol{\beta}'\mathbf{Y}_t)) W_{j,t} + \varepsilon_t.$$



Similar to part (a), we first linearize the above model to obtain a first-order approximation. Then, under  $\mathbb{H}_0^*$ , we have

$$\frac{\partial \varepsilon_t}{\partial \boldsymbol{\alpha}} = -(\phi_1 g_{1,1} \mathbf{U}_{1,t} + \cdots + \phi_p g_{1,p} \mathbf{U}_{p,t}), \quad \frac{\partial \varepsilon_t}{\partial \boldsymbol{\beta}} = -(\theta_1 h_{1,1} \mathbf{V}_{1,t} + \cdots + \theta_q h_{1,q} \mathbf{V}_{q,t}).$$

Denote these partial derivatives by  $\mathbf{U}_t(\boldsymbol{\phi})$  and  $\mathbf{V}_t(\boldsymbol{\theta})$  respectively. Then the LM-type test statistic is given by

$$\text{LM}_T^*(\boldsymbol{\phi}, \boldsymbol{\theta}) = T \frac{\text{SEE}_0 - \text{SSE}_1^*(\boldsymbol{\phi}, \boldsymbol{\theta})}{\text{SSE}_0}.$$

Here,  $\text{SSE}_1^*(\boldsymbol{\phi}, \boldsymbol{\theta})$  is the residual sum of squares in the least squares regression of  $\widehat{\varepsilon}_t$  on  $-\mathbf{x}_{1,t}$ ,  $\mathbf{U}_t(\boldsymbol{\phi})$  and  $\mathbf{V}_t(\boldsymbol{\theta})$ , and  $\text{SSE}_0$  is given in part (a). Since  $\mathbf{U}_t(\boldsymbol{\phi})$  and  $\mathbf{V}_t(\boldsymbol{\theta})$  are linear combinations of  $U_{i,t}$  ( $i = 1, \dots, p$ ) and  $V_{j,t}$  ( $j = 1, \dots, q$ ) respectively, it follows that  $\text{SSE}_1 = \inf_{\boldsymbol{\phi}, \boldsymbol{\theta}} \text{SSE}_1^*(\boldsymbol{\phi}, \boldsymbol{\theta})$ . Hence, an appropriate LM-type test statistic is given by  $\sup_{\boldsymbol{\phi}, \boldsymbol{\theta}} \text{LM}_T^*(\boldsymbol{\phi}, \boldsymbol{\theta})$ .

- 5.4 (a) The simulation results in Tables 5.3 and 5.4 of the main text were obtained using modified versions of two of the GAUSS-routines to accompany the paper by Pitarakis (2006). For computation of the  $F_T^{(1,2)}$  test statistic in Table 5.3, use can be made of the GAUSS, R, or MATLAB routines written by Bruce Hansen which are publicly available at <http://www.ssc.wisc.edu/~bhansen/>.

There are substantial differences in empirical power between the case where the test is implemented on a correctly specified model order, and the case where the order  $p$  is estimated by either the AIC or the BIC method. The worst performance is displayed when the lag length is estimated via the BIC method. This is both under  $T = 200$  and  $T = 500$ , and both DGPs. Some improvement can be noticed as  $T$  increases, but still the empirical power results are poor. Under DGP (i) and  $T = 200$ , AIC (BIC) selects in about 38% (24%) of the cases the true order  $p_0 = 2$ . For DGP (ii) and  $T = 200$  these percentages are, respectively 30% (AIC) and 15% (BIC). In general, however, BIC has a high frequency of underfitting the true order.

- (b) Comparing the empirical correct decision frequencies based on the BIC method with the empirical power of the  $F_T^{(1,2)}$  test statistic in Table 5.3 of the main text using estimated lag lengths, we note substantial improvements in power for the penalty-based model selection approach under both sample sizes. Still, for DGP (ii) these gains are far below the empirical power of the  $F_T^{(1,2)}$  test statistic using the true model specification. The differences in performance of the AIC and BIC methods may also be seen in the final selection of the SETAR model. For instance, for DGP (ii) AIC (BIC) selects the “true” SETAR(2; 2, 2) model, with  $d = 2$ , in about 35.6% (10.7%) of the 1,000 replications. So, for these models and sample sizes, the use of the AIC-based specification strategy is recommended. Note that by setting  $p_{\max} = 6$ , gives a portfolio of  $p_{\max}(p_{\max} + 1)/2 + (p_{\max} + 1) = 28$  models (21 nonlinear and 7 linear, including white noise) to select from.

## Chapter 6

- 6.1 The proposed estimator of  $\tau$  is defined as  $\tilde{\tau} = \widehat{\gamma}_Y(1, 2)/\sigma_\varepsilon^2 \text{Var}(Y_t)$ . The condition  $\tau^4 \sigma_\varepsilon^4 < 1/3$  implies the condition for stationarity, causality and ergodicity of the process  $\{Y_t, t \in \mathbb{Z}\}$ , i.e.  $\tau^2 \sigma_\varepsilon^2 < 1$ . By ergodicity, we have  $\lim_{T \rightarrow \infty} \widehat{\gamma}_Y(1, 2) = \gamma_Y(1, 2) =$

$\tau\sigma_{\varepsilon}^2\sigma_Y^2$ . Hence, as  $T \rightarrow \infty$ ,  $\tilde{\tau} \rightarrow \tau$  a.s. Moreover, as an immediate result of [Grahm \(1995, Thm. 3.1\)](#), the law of iterated logarithm (LIL) holds with  $\mathcal{O}(S_T)$  order of convergence, where  $S_T = \{T/\log \log T\}^{-1/2}$ .

Now  $|\tilde{\tau} - \hat{\tau}| \leq |\tilde{\tau} - \tau| + |\hat{\tau} - \tau|$ . We know that  $|\hat{\tau} - \tau| \rightarrow 0$  a.s. In this case a.s. convergence is equivalent to  $L_2$  convergence in the space of continuous functions defined on the interval  $[0, 1]$ . Hence, it is sufficient to prove that, as  $T \rightarrow \infty$ ,

$$\int (|\tilde{\tau} - \tau| + |\hat{\tau} - \tau|)^2 df_Y(y) \rightarrow 0,$$

where  $f_Y(\cdot)$  is the pdf of  $\{Y_t, t \in \mathbb{Z}\}$ . Since  $\int (\hat{\tau} - \tau)^2 df_Y(y) \rightarrow 0$  and  $\int (\tilde{\tau} - \tau)^2 df_Y(y) \rightarrow 0$  when  $T \rightarrow \infty$ , and applying the Cauchy–Schwarz inequality for  $\int (|\tilde{\tau} - \tau| + |\hat{\tau} - \tau|)^2 df_Y(y)$ , the result holds. The order of convergence follows directly from the above results and [Theorem 3.1 of Grahm \(1995\)](#).

**Remark:** For ease of reference, we present [Theorem 3.1](#) below. To this end, it is appropriate to represent the BL( $p, q, k, r$ ) process (6.27), with  $\mathbb{E}(Y_t) = 0$ , by the state space representation

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \sum_{i=1}^k \varepsilon_{t-i} \mathbf{B}_i \mathbf{Y}_{t-1} + \mathbf{C}\varepsilon_t, \quad Y_t = \mathbf{D}\mathbf{Y}_t, \quad (*)$$

where  $\mathbf{Y}_t = (Y_t, Y_{t-1}, \dots, Y_{t-h+1})'$  with  $h = p \vee r$ ,  $\mathbf{D} = (1, 0, \dots, 0)' \in \mathbb{R}^h$ ,  $\varepsilon_t = (\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q})' \in \mathbb{R}^{q+1}$ , and the matrices

$$\underbrace{\mathbf{A}}_{h \times h} = \begin{pmatrix} \phi_1 & \cdots & \phi_p & 0 & \cdots & 0 \\ 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & 0 & \cdots \end{pmatrix}, \quad \underbrace{\mathbf{B}_i}_{h \times h} = \begin{pmatrix} \tau_{i1} & \cdots & \tau_{ir} & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 & \cdots & 0 \end{pmatrix}, \quad \underbrace{\mathbf{C}}_{h \times (q+1)} = \begin{pmatrix} 1 & \psi_1 & \cdots & \psi_q \\ 0 & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

After this preparation, we can state the theorem as follows.

Let  $\{Y_t, t \in \mathbb{Z}\}$  be the BL process defined by (\*). If for some  $\nu \in \mathbb{N}$  the corresponding sequence  $\{\varepsilon_t, t \in \mathbb{Z}\}$  is a  $2\nu k$ th order symmetric innovation sequence, and moreover  $\rho(\mathbf{\Gamma}_{2\nu}) < 1$ , where  $\mathbf{\Gamma}_{2\nu}$  is a matrix uniquely determined by the parameters of the BL process (see, e.g., the solution to [Exercise 3.4\(a\)](#)), then the following assertions hold.

- (i) There exists a strictly stationary, causal and ergodic solution  $\{Y_t, t \in \mathbb{Z}\}$  of (\*) with  $\mathbb{E}(Y_t^{2\nu}) < \infty$ . The unique strictly stationary solution is the first component of the vector process  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  with

$$\mathbf{Y}_t = \sum_{n=1}^{\infty} \left\{ \prod_{i=1}^n \mathbf{\Psi}(t+1-i) \right\} \mathbf{C}\varepsilon_{t-n} + \mathbf{C}\varepsilon_t,$$

where  $\mathbf{\Psi}(t) = \mathbf{A} + \sum_{i=1}^k \varepsilon_{t-i} \mathbf{B}_i$  and the infinite sum can be interpreted either as an almost sure or  $\mathcal{L}^{2\nu}$  limit.

- (ii) To give the result concerning the asymptotics of the empirical mixed moments of the process  $\{Y_t, t \in \mathbb{Z}\}$ , we need the following definitions:  
Let  $\mathcal{K} = \{0, k_1, \dots, k_i : k_1 \leq \dots \leq k_i, 0 \leq i \leq \nu - 1, k_j \in \mathbb{N} \text{ for } 0 \leq j \leq i\}$

denote the set of indices for identifying the several mixed moments of the process up to the order  $\nu$ . For  $\mathbf{k}_i = (0, k_1, \dots, k_i) \in \mathcal{K}$  we further define  $\mu_k = \mathbb{E}(Y_0, Y_{k_1}, \dots, Y_{k_i})$  and  $\eta_{k,t} = Y_t Y_{t+k_1} \cdots Y_{t+k_i} - \mu_k$ . Moreover, for  $u \in [0, 1]$  we define the process of partial sums  $S_{T,k}(u) = \sum_{t=1}^{[Tu]} \eta_{k,t}$ , and a continuous version  $\xi_{T,k}(u) = S_{T,k}(u) + (Tu - [Tu])\eta_{k,[Tu]+1}$ . Then for all  $k \in \mathbb{N}$  and for all  $\gamma = (\gamma_1, \dots, \gamma_k)' \in \mathbb{R}^k$  and  $\mathbf{k}_i \in \mathcal{K}$  ( $1 \leq i \leq k$ ), we have

$$\left\{ T^{-1/2} \sum_{i=1}^k \gamma_i \xi_{T,k}(u) \right\} \xrightarrow[C[0,1]]{D} \{c_\gamma, B(u)\}, \quad u \in [0, 1],$$

where  $\{B(u), u \in [0, 1]\}$  denotes the standard Brownian motion, and

$$c_\gamma = \lim_{T \rightarrow \infty} \mathbb{E} \left[ \left\{ \sum_{i=1}^k \gamma_i \xi_{T,k}(1) \right\}^2 \right] / T$$

exists with  $0 \leq c < \infty$  and the LIL, i.e.  $\{T \log \log T\}^{-1/2} \sum_{i=1}^k \gamma_i \xi_{T,k}$  is relatively compact in  $(C[0, 1], \|\cdot\|_\infty)$ .

- 6.2** (a) Taking expectations with the diagonal BL process, we have  $\mathbb{E}(Y_t) = \tau \mathbb{E}(Y_{t-1} \varepsilon_{t-1})$ . Multiplying both sides of the BL process with  $\varepsilon_t$  and taking expectations, gives  $\mathbb{E}(Y_t \varepsilon_t) = \sigma_\varepsilon^2$ . Hence,  $\mathbb{E}(Y_t) = \tau \sigma_\varepsilon^2$ . To get the variance of  $\{Y_t, t \in \mathbb{Z}\}$ , squaring both sides of the BL process and taking expectations gives

$$\mathbb{E}(Y_t^2) = \tau^2 \mathbb{E}(Y_{t-1}^2 \varepsilon_{t-1}^2) + \sigma_\varepsilon^2.$$

Multiplying the expression for  $Y_t^2$  with  $\varepsilon_t^2$ , and taking expectations gives

$$\mathbb{E}(Y_t^2 \varepsilon_t^2) = \lambda^2 \mathbb{E}(Y_{t-1}^2 \varepsilon_{t-1}^2) + 3\sigma_\varepsilon^4,$$

where  $\lambda = \tau \sigma_\varepsilon$ . The solution of this difference equation converges if  $\lambda^2 < 1$ , which follows from repeated substitution in which case we have

$$\mathbb{E}(Y_t^2 \varepsilon_t^2) = 3\sigma_\varepsilon^4 / (1 - \lambda^2).$$

So  $\mathbb{E}(Y_t^2) = \sigma_\varepsilon^2 (2\lambda^2 + 1) / (1 - \lambda^2)$ , and combined with  $\mathbb{E}(Y_t)$  gives the desired expression for  $\gamma_Y(0)$ . For  $\gamma_Y(1)$ , first consider

$$\begin{aligned} \mathbb{E}(Y_t Y_{t-1}) &= \tau \mathbb{E}(Y_{t-1}^2 \varepsilon_{t-1}) = 2\tau^2 \mathbb{E}(\varepsilon_{t-1}^2) \mathbb{E}(Y_{t-2} \varepsilon_{t-2}) \\ &= 2\tau^2 \sigma_\varepsilon^4 = 2\mathbb{E}^2(Y_t) \end{aligned}$$

so that  $\gamma_Y(1) = \mathbb{E}^2(Y_t) = \sigma_\varepsilon^2 \lambda^2$ . For the computation of lag  $\ell$  autocovariances, we have

$$\begin{aligned} \mathbb{E}(Y_t Y_{t-\ell}) &= \tau \mathbb{E}(Y_{t-1} \varepsilon_{t-1} \varepsilon_{t-\ell}), \quad (\ell > 1), \\ &= \tau \mathbb{E}((\tau Y_{t-2} \varepsilon_{t-2} + \varepsilon_{t-1}) \varepsilon_{t-1} Y_{t-\ell}) \\ &= \tau \mathbb{E}(\varepsilon_{t-1}^2 Y_{t-\ell}) = \tau^2 \sigma_\varepsilon^4. \end{aligned}$$

So  $\gamma_Y(\ell) = 0$  for  $\ell > 1$ . Hence,

$$\rho_Y(\ell) = \begin{cases} 1, & \ell = 0, \\ \lambda^2 (1 - \lambda^2) / (1 + \lambda^2 + \lambda^4), & \ell = 1, \\ 0, & |\ell| \geq 2. \end{cases}$$

- (b) The autocovariance structure of the diagonal BL process is very similar to that of an MA(1) process  $Y_t = \varepsilon_t + \tau\varepsilon_{t-1}$  ( $|\tau| < 1$ ,  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_{\varepsilon, \text{MA}}^2)$ ) in the sense that the autocovariances are all zero for lags  $|\ell| \geq 2$ . Hence, an identification problem arises.

By equating the first order autocorrelation  $\rho_Y(1)$  of the diagonal BL process with the corresponding expression of the MA(1) process, and considering the invertibility condition  $|\tau| < 1$ , it follows that

$$\tau = \frac{(1 + \lambda^2 + \lambda^4) - \sqrt{1 + 2\lambda^2 - \lambda^4 + 10\lambda^6 - 3\lambda^8}}{2\lambda^2(1 - \lambda^2)}.$$

Similarly, by equating the corresponding expressions for  $\gamma_Y(0)$ , it follows that

$$\sigma_{\varepsilon, \text{MA}}^2 = \frac{\sigma_{\varepsilon}^2(1 + \tau^2)(1 - \lambda^2)}{1 + \lambda^2 + \lambda^4}.$$

This last expression can be used to define the penalty, say  $(\sigma_{\varepsilon, \text{MA}}^2 - \sigma_{\varepsilon}^2)/\sigma_{\varepsilon}^2$ , from misclassifying a diagonal BL(0, 0, 1, 1) process as an MA(1) process.

**Remark:** To distinguish between the diagonal BL and the MA(1) model one may, for instance, consider the ACF of the process  $\{Y_t^2, t \in \mathbb{Z}\}$ . For an MA(1) the ACF of  $\{Y_t^2, t \in \mathbb{Z}\}$  is identified by the ACF as an MA(1) process. This can be seen as follows

$$\mathbb{E}(Y_t^2 Y_{t-\ell}^2) = \tau^2 \mathbb{E}(\tau^2 Y_{t-2}^2 Y_{t-\ell}^2 \varepsilon_{t-1}^2 \varepsilon_{t-2}^2) + 3\tau^2 \sigma_{\varepsilon}^4 \mathbb{E}(Y_t^2) + \sigma_{\varepsilon}^2 \mathbb{E}(Y_t^2) \quad (*)$$

by substituting twice for  $Y_{t-1}^2$ . For  $\ell > 1$ , the first term on the right-hand side can be written as  $\tau^4 \sigma_{\varepsilon}^2 \mathbb{E}(Y_{t-2}^2 Y_{t-\ell}^2 \varepsilon_{t-2}^2)$ . If we rewrite (\*) with  $t-1$  replacing  $t$ , we obtain

$$\mathbb{E}(Y_{t-1}^2 Y_{t-\ell}^2) = \tau^2 \mathbb{E}(Y_{t-2}^2 Y_{t-\ell}^2 \varepsilon_{t-2}^2) + \sigma_{\varepsilon}^2 \mathbb{E}(Y_{t-\ell}^2).$$

Using the stationarity assumption, we have

$$\mathbb{E}(Y_{t-1}^2 Y_{t-\ell}^2) = \tau^2 \sigma_{\varepsilon}^2 \mathbb{E}(Y_t^2 Y_{t-\ell+1}^2) + (2\tau^2 \sigma_{\varepsilon}^4 + \sigma_{\varepsilon}^2) \mathbb{E}(Y_t^2).$$

Hence,

$$\begin{aligned} \text{Cov}(Y_t^2, Y_{t-\ell}^2) &= \tau^2 \sigma_{\varepsilon}^2 \text{Cov}(Y_t^2, Y_{t-\ell+1}^2) + (2\tau^2 \sigma_{\varepsilon}^4 + \sigma_{\varepsilon}^2) \mathbb{E}(Y_t^2) \\ &\quad + (\tau^2 \sigma_{\varepsilon}^2 - 1) \mathbb{E}^2(Y_t^2) \\ &= \tau^2 \sigma_{\varepsilon}^2 \text{Cov}(Y_t^2, Y_{t-\ell+1}^2), \quad (\ell > 1). \end{aligned}$$

So, if we denote the ACF of  $Y_t^2$  by  $\rho_Y^{(2)}(\ell)$ , we have  $\rho_Y^{(2)}(\ell) = \tau^2 \sigma_{\varepsilon}^2 \rho_Y^{(2)}(\ell-1)$ ,  $\ell > 1$ . For  $\rho^{(2)}(1)$ , we consider

$$\mathbb{E}(Y_t^2 Y_{t-1}^2) = \tau^2 \mathbb{E}(Y_{t-1}^4 \varepsilon_{t-1}^2) + \sigma_{\varepsilon}^2 \mathbb{E}(Y_t^2).$$

Writing  $Y_{t-1}^4$  in full and multiplying the resulting equation with  $\varepsilon_{t-1}^2$  gives

$$\mathbb{E}(Y_t^2 Y_{t-1}^2) = \tau^6 \sigma_{\varepsilon}^2 \mathbb{E}(Y_{t-2}^4 \varepsilon_{t-2}^4) + 18\tau^4 \sigma_{\varepsilon}^4 \mathbb{E}(Y_{t-2}^2 \varepsilon_{t-2}^2) + 15\tau^2 \sigma_{\varepsilon}^6 + \sigma_{\varepsilon}^2 \mathbb{E}(Y_t^2).$$

Finally, computing  $\mathbb{E}(Y_{t-2}^4 \varepsilon_{t-2}^4)$  and some algebra gives

$$\rho^{(2)}(1) = \frac{3\lambda^2(2 + 3\lambda^2 + 5\lambda^4 + \lambda^6 - 8\lambda^8)}{1 + 4\lambda^2 + 40\lambda^4 + 18\lambda^6 - 54\lambda^8},$$

provided  $\lambda^2 < 1/3$ . Clearly, using the ACF, the process  $\{Y_t^2, t \in \mathbb{Z}\}$  is identified as an ARMA(1, 1) process when the true DGP of  $\{Y_t, t \in \mathbb{Z}\}$  is the diagonal BL(0, 0, 1, 1) process. Hence, the ACF of the process  $\{Y_t^2, t \in \mathbb{Z}\}$  can be used to distinguish between the MA(1) process and the diagonal BL(0, 0, 1, 1) process. To distinguish between the diagonal BL(0, 0, 1, 1) process and the MA(1) process, one has to consider the third-order moment since it is not zero for the BL process. In fact, the existence of the third and fourth moments is also required to derive the asymptotic distribution of the sample ACVF of the process. Some more algebra gives

$$\mathbb{E}(Y_t - \mu_Y)^3 = 2\lambda^3 \sigma_\varepsilon^3 \frac{2\tau^3(4 + 5\lambda^2)}{1 - \lambda^2},$$

provided  $\lambda^2 < 1$ . Note that the above expression is *not* in agreement with the third central moment given by Granger and Andersen (1978a, p. 52). That is, the factor  $\sigma_\varepsilon^3$  is missing from equation (6.25) in their monograph. Moreover,

$$\mathbb{E}(Y_t - \mu_Y)^4 = \frac{\sigma_\varepsilon^4(3 + 9\lambda^2 + 57\lambda^4 + 93\lambda^6 + 117\lambda^8 + 135\lambda^{10})}{(1 - \lambda^2)(1 - 3\lambda^4)},$$

provided  $3\lambda^4 < 1$ . This expression differs a factor 3 with the fourth central moment (6.26) given by Granger and Andersen (1978a, p. 53). This last result may explain the high values of the coefficient of kurtosis in column 5 of Table 1 of Granger and Andersen (1978a); see also De Gooijer and Heuts (1987).

- (c) From part (a), we know that  $\mathbb{E}(Y_t) = \tau\sigma_\varepsilon^2$ . We also see that  $\gamma_Y(0)$  is a function of  $\tau$  and  $\sigma_\varepsilon^2$ . If we substitute  $\gamma_Y(1)/\sigma_\varepsilon^4$  for  $\tau^2$  in  $\gamma_Y(0)$ , then  $\gamma_Y(0)$  becomes a quadratic equation in  $\sigma_\varepsilon^2$ . Solving for  $\sigma_\varepsilon^2$ , we then have

$$\sigma_\varepsilon^2 = 2^{-1}[\{\gamma_Y(0) - \gamma_Y(1)\} \pm \{\gamma_Y^2(0) - 6\gamma_Y(0)\gamma_Y(1) - 3\gamma_Y^2(1)\}^{1/2}],$$

with the condition that  $\rho_Y(1) = \gamma_Y(1)/\gamma_Y(0) < 0.1547$ . Substituting for  $\gamma_Y(0)$  in the above expression, we can show that

$$2\lambda^4 + 2\lambda^2 - 1 \leq 0,$$

depending on whether the negative or positive solution of  $\sigma_\varepsilon^2$  is chosen. For the positive solution we have  $2\lambda^4 + 2\lambda^2 - 1 < 0$ , which is a sufficient condition for invertibility (Pham and Tran, 1981). The latter condition reduces to

$$|\lambda| < \{(\sqrt{3} - 1)/2\}^{1/2} \equiv \lambda_0 = 0.605.$$

Thus, we can find a region in the parameter space where the process is invertible.

- (d) The following CLT will be used below.

*Let  $\{Z_t, t \in \mathbb{Z}\}$  be a strictly stationary sequence of  $m$ -dependent random variables with  $\mathbb{E}(Z_t) = 0$ , and autocovariance function  $\text{Cov}(Z_t, Z_{t+\ell}) \neq 0$ , then  $1/\sqrt{T} \sum_{t=1}^T Z_t \xrightarrow{D} \mathcal{N}(0, \sum_{\ell=-m}^m \text{Cov}(Z_t, Z_{t+\ell}))$ .*

First, however, we focus on the mean and autocovariance of  $\{U_{t,m}\}$  and  $\{W_{t,m}\}$ . It is easy to see that  $\mathbb{E}(U_{t+\ell,m}) = \sigma_\varepsilon \lambda \forall \ell$ , and the lag  $\ell$  autocovariance of  $\{U_{t,m}\}$  is given by

$$\mathbb{E}(U_{t,m}U_{t+\ell,m}) = \begin{cases} \sigma_\varepsilon^2 + 3\sigma_\varepsilon^2 \sum_{j=1}^m \lambda^{2j}, & \ell = 0, \\ 2\sigma_\varepsilon^2 \lambda^2, & |\ell| = 1, \\ \sigma_\varepsilon^2 \lambda^2, & |\ell| \geq 2. \end{cases} \quad (**)$$

Similarly, we have  $\mathbb{E}(W_{t+\ell,m}) = 0 \forall \ell$ . Moreover, for a fixed  $m$ , the lag  $\ell$  autocovariance of  $\{W_{t,m}\}$  is given by

$$\mathbb{E}(W_{t,m}W_{t+\ell,m}) = \begin{cases} 3\sigma_\varepsilon^2 \sum_{j=m+1}^{\infty} \lambda^{2j}, & \ell = 0, \\ 0, & |\ell| \geq 1. \end{cases} \quad (***)$$

Next, for a fixed  $m$ , we define the process  $X_t = U_{t,m} - \mu_U$ , where  $\mu_m = \mathbb{E}(U_{t,m})$ . Then  $\{X_t, t \in \mathbb{Z}\}$  is a sequence of stationary  $m$ -dependent random variables. Further, we have  $\mathbb{E}(X_t) = 0$ , and from (\*\*)

$$\mathbb{E}(X_t, X_{t+\ell}) = \text{Cov}(U_{t,m}, U_{t+\ell,m}) = \begin{cases} \sigma_\varepsilon^2(1 - \lambda^2 + 3 \sum_{j=1}^m \lambda^{2j}), & \ell = 0, \\ \sigma_\varepsilon^2 \lambda^2, & |\ell| = 1, \\ 0, & |\ell| \geq 2. \end{cases}$$

Moreover,

$$T^{-1/2} \sum_{t=1}^T X_t = T^{1/2}(\bar{U}_T - \mu_U)$$

and

$$\sum_{\ell=-m}^m \text{Cov}(X_t, X_{t-\ell}) = \sigma_\varepsilon^2 \left( 1 + \lambda^2 + 3 \sum_{j=1}^m \lambda^{2j} \right).$$

Therefore, the process  $\{X_t, t \in \mathbb{Z}\}$  satisfies the conditions of the CLT. The final part of the proof follows directly from the above results.

- (e) Using  $Q_{m,T}$  and  $R_{m,T}$  defined by the hint, we have  $\sqrt{T}(\bar{Y}_T - \mu_Y) = Q_{m,T} + R_{m,T}$ . For a fixed  $m$ , we have from (\*\*\*) that

$$\begin{aligned} \text{Var}(R_{m,T}) &= T^{-1} \sum_{t,s=1}^T \text{Cov}(W_{t,m}, W_{s,m}) \\ &= T^{-1} \sum_{\ell=-(T-1)}^{T-1} (T - |\ell|) \text{Cov}(W_{t,m}, W_{t+\ell,m}) \\ &= 3\sigma_\varepsilon^2 \sum_{j=m+1}^{\infty} \lambda^{2j}. \end{aligned}$$

Therefore,  $\lim_{m \rightarrow \infty} \text{Var}(R_{m,T}) = 0$ . Hence, by Chebychev's inequality, we have  $R_{m,T} \xrightarrow{P} 0$  uniformly in  $T$ . It then follows that the asymptotic distribution of  $\sqrt{T}(\bar{Y}_T - \mu_Y)$  is the same as that of  $Q_{m,T}$ , which has been shown in part (c) to converge to a normal distribution as  $T \rightarrow \infty$ . Using the autocovariances given in part (a), the variance of  $\sqrt{T}(\bar{Y}_T - \mu_Y)$  converges to  $\gamma_Y(0) + 2\gamma_Y(1)$ , as  $m \rightarrow \infty$ , which is identical to  $\sigma_\varepsilon^2(1 + 3\lambda^2 - \lambda^4)/(1 - \lambda^2)$ .

Thus, let  $\{Y_t, t \in \mathbb{Z}\}$  be a stochastic process defined by  $Y_t = \tau Y_{t-1} \varepsilon_{t-1} + \varepsilon_t$  with zero-mean Gaussian white noise. Assume that the model is invertible, i.e. the condition  $|\lambda| < 0.605$  holds. Furthermore, without loss of generality, let  $\sigma_\varepsilon^2 = 1$ . Then, in view of the above results, the asymptotic distribution of  $\hat{\tau}$  is given by

$$\sqrt{T}(\hat{\tau} - \tau) \xrightarrow{D} \mathcal{N}(0, (1 + 3\tau^2 - \tau^4)/(1 - \tau^2)),$$

which completes the proof. In closing, we refer to Exercise 6.5 of the main text where we consider another moment estimator of  $\tau$  when  $\sigma_\varepsilon^2$  is unknown.

- 6.3 (a) From the log-quasi conditional likelihood, the first order derivative with respect to  $\boldsymbol{\theta}$  becomes

$$G(\mathbf{y}, \boldsymbol{\theta}) = \frac{\partial Q_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2} \sum_t \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \left\{ \frac{1}{\sigma_t^2} - \frac{(Y_t - \mu_t)^2}{\sigma_t^4} \right\} + \sum_t \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{(Y_t - \mu_t)}{\sigma_t^2}.$$

Hence,

$$g(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \mathbb{E}[G(\mathbf{y}, \boldsymbol{\theta}) | \tilde{\boldsymbol{\theta}}] = -\frac{1}{2} \sum_t \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \left\{ \frac{1}{\sigma_t^2} - \frac{\tilde{\sigma}_t^2 + (\tilde{\mu}_t - \mu_t)^2}{\sigma_t^4} \right\} + \sum_t \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \frac{(Y_t - \mu_t)}{\sigma_t^2},$$

where  $\tilde{\mu}_t = \mu(Y_{t-1}, \dots, Y_{t-p}, \tilde{\boldsymbol{\theta}})$  and  $\tilde{\sigma}_t^2 = \sigma^2(Y_{t-1}, \dots, Y_{t-p}, \tilde{\boldsymbol{\theta}})$ . Thus,

$$\begin{aligned} \frac{\partial g(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta})}{\partial \tilde{\boldsymbol{\theta}}} &= -\frac{1}{2} \sum_t \left\{ -\frac{1}{\sigma_t^4} \left( \frac{\partial \tilde{\sigma}_t^2}{\partial \tilde{\boldsymbol{\theta}}} \right) - \frac{2(\tilde{\mu}_t - \mu_t)}{\sigma_t^4} \left( \frac{\partial \tilde{\mu}_t}{\partial \tilde{\boldsymbol{\theta}}} \right) \right\} \left( \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \right)' \\ &\quad + \sum_t \frac{1}{\sigma_t^2} \left( \frac{\partial \tilde{\mu}_t}{\partial \tilde{\boldsymbol{\theta}}} \right) \left( \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \right)'. \end{aligned}$$

The Fisher information matrix is given by

$$\mathcal{I}(\boldsymbol{\theta}) \equiv \frac{\partial g(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta})}{\partial \tilde{\boldsymbol{\theta}}} \Big|_{\tilde{\boldsymbol{\theta}}=\boldsymbol{\theta}} = \frac{1}{2} \sum_t \frac{1}{\sigma_t^4} \left( \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \right)' + \sum_t \frac{1}{\sigma_t^2} \left( \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \right)'.$$

Substituting the expressions for  $G(\mathbf{y}, \boldsymbol{\theta})$  and  $\partial g(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta})/\partial \tilde{\boldsymbol{\theta}}$  in the approximate expression of  $G(\mathbf{y}, \boldsymbol{\theta})$  it follows that

$$\begin{aligned} \sum_t \frac{1}{2} \frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \left\{ \left( \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \right)' (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \sigma_t^2 - (Y_t - \mu_t)^2 \right\} \\ + \sum_t \frac{1}{\sigma_t^2} \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \left\{ \left( \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} \right)' (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) - (Y_t - \mu_t) \right\} = 0. \end{aligned}$$

Rearranging terms gives expression (6.44) of the main text.

- (b) Recall  $\mathbf{Z}_t = \begin{pmatrix} \partial \sigma_t^2 / \partial \boldsymbol{\theta} \\ \partial \mu_t / \partial \boldsymbol{\theta} \end{pmatrix}$ . The partial derivative of  $\mu_t$  with respect to  $\boldsymbol{\theta}$  are given by

$$\frac{\partial \mu_t}{\partial \phi_0^{(i)}} = I_t^{(i)} \quad (i = 1, 2), \quad \frac{\partial \mu_t}{\partial \phi_j^{(i)}} = I_t^{(i)} Y_{t-j} \quad (i = 1, 2; j = 1, \dots, p_i),$$

and  $\partial \mu_t / \partial \alpha_j^{(i)} = 0$  ( $i = 1, 2; j = 0, 1, \dots, q_i$ ). The partial derivative of  $\sigma_t^2$  with respect to  $\boldsymbol{\theta}$  are given by

$$\begin{aligned} \frac{\partial \sigma_t^2}{\partial \phi_0^{(u)}} &= -2 \sum_{i=1}^2 I_t^{(i)} \left( \sum_{j=1}^{q_i} \alpha_j^{(i)} \varepsilon_{t-j} I_{t-j}^{(u)} \right) \quad (u = 1, 2), \\ \frac{\partial \sigma_t^2}{\partial \phi_v^{(u)}} &= -2 \sum_{i=1}^2 I_t^{(i)} \left( \sum_{j=1}^{q_i} \alpha_j^{(i)} \varepsilon_{t-j} Y_{t-j-v} I_{t-j}^{(u)} \right) \quad (u = 1, 2; v = 1, \dots, p_i), \\ \frac{\partial \sigma_t^2}{\partial \alpha_0^{(i)}} &= I_t^{(i)} \quad (i = 1, 2), \quad \frac{\partial \sigma_t^2}{\partial \alpha_j^{(i)}} = \varepsilon_{t-j}^2 I_t^{(i)} \quad (i = 1, 2; j = 1, \dots, q_i). \end{aligned}$$

Hence, stacking up by  $t$ , an explicit expression for  $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_T)'$  follows directly from  $\mathbf{Z}_t$ .

- 6.4 (a) Clearly,  $\mathbf{z}_t = (\nabla_{\boldsymbol{\theta}} h_t) h_t^{-1} = \mathbf{0}'$  and hence  $\mathbb{E}[\mathbf{V}(\varepsilon_{t-\ell}) \mathbf{z}_t'] = \mathbf{0}$  and  $\boldsymbol{\Omega} = \boldsymbol{\Upsilon} = \mathbb{E}[\mathbf{w}_t \mathbf{w}_t']$ . Consequently,

$$\boldsymbol{\Lambda}(\ell) = \mathbb{E}[\nabla \mathbf{U}(\varepsilon_t)] \otimes \mathbb{E}[\mathbf{V}(\varepsilon_{t-\ell}) \mathbf{w}_t'] \quad \text{and} \quad \boldsymbol{\Delta}(\ell) = \mathbb{E}[\mathbf{U}(\varepsilon_t) \varepsilon_t] \otimes \mathbb{E}[\mathbf{V}(\varepsilon_{t-\ell}) \mathbf{w}_t'].$$

Observe that  $\boldsymbol{\Lambda}(\ell) = \mathbf{0}$  if the condition  $\mathbb{E}[\nabla \mathbf{U}(\varepsilon_t)] = \mathbf{0}$  holds. In that case the asymptotic null distribution of  $\sqrt{T-\ell} \hat{\boldsymbol{\rho}}(\ell)$  will be free from estimation uncertainty. Now when  $P = Q = 1$ , and  $(u_1(\varepsilon_t), v_1(\varepsilon_{t-\ell})) = (\varepsilon_t, \varepsilon_{t-\ell})$ , we have  $\mathbb{E}[\nabla \mathbf{U}(\varepsilon_t)] = \mathbb{E}[\mathbf{U}(\varepsilon_t) \varepsilon_t] = (1, \dots, 1)'$ . This implies  $\boldsymbol{\Lambda}(\ell) = \boldsymbol{\Delta}(\ell) = \mathbb{E}[\mathbf{V}(\varepsilon_{t-\ell}) \mathbf{w}_t']$ , and  $\mathbf{A}(\ell, \ell') = -\boldsymbol{\Lambda}(\ell) \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda}'(\ell')$ . So, after some simple rewriting, we obtain the  $(i, j)$ th element of the asymptotic variance-covariance matrix  $\boldsymbol{\Sigma}(\ell)$  as specified in part (a).

- (b) Assume, we have a residual time series  $\{\hat{\varepsilon}_t\}_{t=1}^T$ . Let  $\hat{\boldsymbol{\rho}} = (\hat{\rho}_\varepsilon(1), \dots, \hat{\rho}_\varepsilon(M))'$  with  $\hat{\rho}_\varepsilon(\ell) = \sum_{t=\ell+1}^T (\hat{\varepsilon}_t - \bar{\hat{\varepsilon}})(\hat{\varepsilon}_{t-\ell} - \bar{\hat{\varepsilon}}) / \sum_{t=1}^T (\hat{\varepsilon}_t - \bar{\hat{\varepsilon}})^2$  and  $\bar{\hat{\varepsilon}} = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t$ . Then, given the results in part (a) and Section 6.3 of the main text, the large sample distribution of  $\hat{\boldsymbol{\rho}} \sqrt{T}$  is normal with mean  $\mathbf{0}$  and variance-covariance matrix  $\boldsymbol{\Sigma} = \mathbf{I}_M - \sigma_\varepsilon^{-2} \mathbf{m}' \mathbf{V}^{-1} \mathbf{m}$ . A diagnostic test statistic for residual serial correlation is given by

$$\mathcal{Q}_T(M) = T \hat{\boldsymbol{\rho}}' \hat{\boldsymbol{\Sigma}}^- \hat{\boldsymbol{\rho}},$$

where  $\hat{\boldsymbol{\Sigma}}^-$  is a consistent estimate of  $\boldsymbol{\Sigma}^-$ , with  $\boldsymbol{\Sigma}^-$  a generalized inverse of  $\boldsymbol{\Sigma}$ . Under the null hypothesis of no residual autocorrelation,  $\mathcal{Q}_T(M)$  is asymptotically  $\chi_M^2$  distributed, as  $T \rightarrow \infty$ . Of course, in practice  $\mathbf{V}$  and  $\mathbf{m}$  need to be replaced by consistent estimates. Then, under the null hypothesis of no serial correlation,  $\mathcal{Q}_T(M)$  is asymptotically distributed as a  $\chi_r^2$  random variable with  $r = \text{rank}(\mathbf{V})$ .

Depending on the type of nonlinear process, some algebra is required to derive explicit expressions of  $\mathbf{V}$  and  $\mathbf{m} = \{(m_{i,j})\}$  ( $i, j = 1, 2$ ). As an illustration, consider a strictly stationary time series  $\{Y_t, t \in \mathbb{Z}\}$  generated by the SETAR(2; 1, 1) model

$$Y_t = \phi_1 Y_{t-1}^- + \phi_2 Y_{t-1}^+ + \varepsilon_t,$$

where  $Y_{t-1}^- = Y_{t-1} I(Y_{t-1} \leq 0)$  and  $Y_{t-1}^+ = Y_{t-1} I(Y_{t-1} > 0)$ , and  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ . Then we have

$$\mathbf{V} = \begin{pmatrix} \mathbb{E}(Y_{t-1}^-)^2 & 0 \\ 0 & \mathbb{E}(Y_{t-1}^+)^2 \end{pmatrix},$$

$$m_{i,j} = \sigma_\varepsilon^{-2} \mathbb{E}(\varepsilon_t Y_{t+i-1}^-) \frac{\mathbb{E}(\varepsilon_t (Y_{t+j-1}^-)^2)}{\mathbb{E}(Y_{t-1}^-)^2} + \sigma_\varepsilon^{-2} \mathbb{E}(\varepsilon_t Y_{t+i-1}^+) \frac{\mathbb{E}(\varepsilon_t (Y_{t+j-1}^+)^2)}{\mathbb{E}(Y_{t-1}^+)^2}.$$

Consistent estimators of  $\mathbb{E}(\varepsilon_t Y_{t+i-1}^\pm)$  and  $\mathbb{E}^2(Y_{t-1}^\pm)$  are  $T^{-1} \sum_{t=1}^T (Y_t - \phi_1 Y_{t-1}^- - \phi_2 Y_{t-1}^+) Y_{t+i-1}^\pm$  and  $T^{-1} \sum_{t=1}^T (Y_{t-1}^\pm)^2$ , respectively.

- 6.5 Table 6.1 contains various simulation results. Observe, that when estimates of  $\tau$  are near the boundary  $\tau < (1/105)^{1/8} = 0.5589$  the number of effective MC replications



decreases considerably. Overall the CLS estimator  $\hat{\tau}$  performs better than the estimator  $\tilde{\tau}$ , both in terms of higher coverage probabilities and shorter 95% confidence intervals (CIs). Note that for  $\tau = \pm 0.6$  the expression for  $\text{Var}(\tilde{\tau})$  results in larger CIs than the CIs obtained for  $\tau = \pm 0.4$ . This may be due to the fact that the formula for  $\text{Var}(\tilde{\tau})$  is based on the following approximation

$$\frac{1}{T} \text{Var}\left(\sum_t Y_t\right) = \text{Var}(Y_t) + 2 \sum_{i=1}^4 \text{Cov}(Y_{t+i}, Y_t) + o(\lambda^2),$$

assuming that  $\text{Cov}(Y_{t+\ell}, Y_t) = o(\lambda^2)$  for  $\ell > 4$ . So the approximation is based on only a few autocovariance terms.

The estimation results can be computed by using a converted version of the MATLAB function BL\_CLS.m given in the file Example\_6-5.zip; see also below.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 6.5
% File: exercise_6-5.zip
%
% BL estimation with estimators beta-widehat and beta-tilde.
%
% BL model: Y_{t} = beta * eps_{t-1} * Y_{t-2} + eps_{t},
%           with eps_{t} ~ N(0,1) distributed.
%
% OUTPUT:
% CovCLS   = Coverage probability of beta-widehat estimator
% CovT     = Coverage probability of beta-tilde estimator
% LengthCLS = Average length 95% CI: beta-widehat estimator
% LengthT   = Average length 95% CI: beta-tilde estimator
%
% Reference:
% Giordano, F. and Vitale, C. (2003).
% CLS asymptotic variance for a particular relevant bilinear time
% series model.
% Statistical Methods & Applications, 12(2), 169-185.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
n   = 500;
m   = 100;
beta = 0.5;
iter = 1000;
sum1 = 0;
sum2 = 0;
VY   = 1/(1-beta*beta);    % Var[Y_{t}]

for it=1:iter
    eps = normrnd(0,1,n+m,1);
    y   = zeros(n+m,1);
    for i=3:n+m
        y(i) = beta*eps(i-1)*y(i-2)+eps(i);
    end
end

```

```

end

data=y(m+1:m+n);

% Start: CLS estimation
y2 = data.*data;
y2lag = y2(3:n);
yy = y2(1:n-2);
X = [ones(size(y2lag)) yy];
a = X\y2lag; % a(1) is estimated LS variance of eps_{t}

num12 = 0;
den = 0;
num113 = 0;
num014 = 0;
for i=1:n-4
    num12 = num12 + data(i+2)*data(i+1)*data(i);
    den = den + data(i)*data(i);
end
c0 = den/(n-4);

bestgr = (num12/(a(1)*den));
cond1 = bestgr^8;
if (cond1 < (1/105))
    sum1 = sum1+1;
    bgr(sum1,1) = bestgr;
    biasgr(sum1,1) = bestgr-beta;
    varestgr(sum1,1) = (1+22*bestgr^2+9*bestgr^2-6*bestgr^2*VY)/(VY);
end

besttil = num12/(n*VY);
cond2 = besttil^8;
if (cond2 < (1/105))
    sum2 = sum2+1;
    btil(sum2,1) = besttil;
    biastil(sum2,1) = besttil-beta;
    termtil = (1-3*besttil^4)*(1-15*besttil^6);
    vartilde(sum2,1) = (1-besttil^2)*(183*besttil^6+...
        42*besttil^4+14*besttil^2+1)/termtil;
end

end

meanbgr = mean(bgr);
stdbestgr = mean(sqrt(varestgr))/sqrt(n);

meanbtil = mean(btil);
stdbesttil = mean(sqrt(vartilde))/sqrt(n);

% Percentage of times beta is in the 95% CI

```

```

sum3 = 0;
for i=1:sum1
    lb1(i) = bgr(i,1) - 1.96* sqrt(varestgr(i,1))/sqrt(n);
    ub1(i) = bgr(i,1) + 1.96* sqrt(varestgr(i,1))/sqrt(n);
    lengthgr(i) = ub1(i)-lb1(i);
    if (beta>=lb1(i) && beta<=ub1(i))
        sum3 = sum3+1;
    end
end
CovCLS = (sum3/sum1)*100;
LengthCLS = mean(lengthgr);

sum4 = 0;
for i=1:sum2
    lb2(i) = btil(i,1) - 1.96* sqrt(vartilde(i,1))/sqrt(n);
    ub2(i) = btil(i,1) + 1.96* sqrt(vartilde(i,1))/sqrt(n);
    lengthtil(i)=ub2(i)-lb2(i);
    if (beta>=lb2(i) && beta<=ub2(i))
        sum4 = sum4+1;
    end
end
CovT = (sum4/sum2)*100;
LengthT = mean(lengthtil);

```

**Table 6.1:** Some summary measures for the estimators  $\hat{\tau}$  (CLS) and  $\tilde{\tau}$  of Exercise 6.5;  $T = 1,000$ , 1,000 MC replications, and a nominal 95% coverage probability.

$\tau$	$\hat{\tau}$ (CLS estimator)			$\tilde{\tau}$		
	Effective number of replicates	Coverage (in %)	Average length	Effective number of replicates	Coverage (in %)	Average length
-0.6	539	64.75	0.25	450	62.89	0.41
-0.4	982	95.62	0.25	957	87.27	0.28
-0.1	1,000	97.20	0.14	1,000	95.50	0.13
0.1	1,000	96.00	0.14	1,000	94.20	0.13
0.4	978	95.09	0.25	959	87.73	0.28
0.6	538	60.59	0.25	486	65.61	0.40

**Remark:** No closed form expression for the asymptotic variance of the CLS of  $\tau$  in model (6.35) exists when the errors are not normally distributed. Bootstrapping methods may be used for this purpose. However, bootstrapping requires sampling from the estimated errors, and refitting the BL model, which can be computationally demanding. Another approach is to use the profile empirical likelihood method. Using this method and adopting model (6.35) with  $\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$  and  $t(8)$ , Feng et al. (2013) present empirical coverage probabilities for four proposed confidence intervals and the bootstrap confidence interval. Provided there is sufficient computing time,

**Table 6.2:** Empirical coverage probabilities for the confidence intervals  $I_\alpha^{(1)}$ ,  $I_\alpha^{(2)}$ ,  $I_\alpha^{(3)}$  proposed by Feng et al. (2013), and the bootstrap interval  $I_\alpha^*$  with levels  $\alpha = 0.9$  and  $0.95$ , 1,000 random samples.

$(T, \beta, \varepsilon_t)$	$I_{0.9}^*$	$I_{0.9}^{(1)}$	$I_{0.9}^{(2)}$	$I_{0.9}^{(3)}$	$I_{0.95}^*$	$I_{0.95}^{(1)}$	$I_{0.95}^{(2)}$	$I_{0.95}^{(3)}$
(100, 0.1, $\mathcal{N}$ )	0.949	0.878	0.871	0.861	0.990	0.935	0.929	0.930
(100, 0.2, $\mathcal{N}$ )	0.912	0.872	0.874	0.864	0.971	0.934	0.931	0.929
(100, 0.3, $\mathcal{N}$ )	0.851	0.880	0.875	0.869	0.919	0.930	0.931	0.926
(100, 0.4, $\mathcal{N}$ )	0.830	0.879	0.884	0.869	0.830	0.931	0.931	0.923

the reader is invited to replicate their MC simulation results for the case  $n = 100$  and  $\tau = 0.1, \dots, 0.4$ , using the functions in the file `exercise_6-5-remark.r`. Using these functions, Table 6.2 reports some empirical coverage probabilities. Compare these results with those given by Feng et al. (2013, Table 1, first 4 top rows).

**6.6** (a)+(b) (**First part**)

Note that if  $\hat{\rho}_Y(1) \geq 0.1547$  (see, e.g., the solution to Exercise 6.2(c)) no real solution exists for  $\hat{\tau}^*$ , and hence the method of moments fails to yield an estimator for  $\tau$ . The third column of Table 6.3 shows the moment estimates  $\hat{\tau}^*$  averaged over 1,000 MC replications. Columns 4 and 5 show the bootstrap moment estimator and its corresponding standard deviation, both based on  $B = 1,000$  bootstrap (BS) replicates. These latter results were obtained as follows.

- (i) Compute the mean-deleted residuals  $\{\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \bar{\hat{\varepsilon}}\}_{t=1}^T$ , where  $\bar{\hat{\varepsilon}} = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t$  and

$$\hat{\varepsilon}_t = Y_t - \hat{\tau}^* Y_{t-1} \hat{\varepsilon}_{t-1}, \quad (\hat{\varepsilon}_0 = 0).$$

- (ii) Draw (with replacement) bootstrap pseudo residuals  $\{\varepsilon_t^*\}_{t=1}^T$  from the EDF of the centered residuals  $\tilde{\varepsilon}_t$  and generate the bootstrap sample  $\{Y_t^*\}_{t=1}^T$  as  $Y_t^* = \varepsilon_t^* + \hat{\tau}^* Y_{t-1}^* \varepsilon_{t-1}^*$ .
- (iii) Repeat steps (i) – (ii),  $B$  times. For each series  $\{Y_t^*\}$  obtain the bootstrap estimates of the moment estimator.
- (iv) Let  $\{\hat{\tau}^{*(b)}\}_{b=1}^B$  denote the set of bootstrap moment estimates. Calculate their sample mean and sample standard deviation as

$$\bar{\hat{\tau}}_{\text{BS}}^* = \frac{1}{B} \sum_{b=1}^B \hat{\tau}^{*(b)} \quad \text{and} \quad \hat{\sigma}_{\hat{\tau}_{\text{BS}}^*} = \left( \frac{1}{B-1} \sum_{b=1}^B (\hat{\tau}^{*(b)} - \bar{\hat{\tau}}_{\text{BS}}^*)^2 \right)^{1/2}.$$

The CLS estimate  $\hat{\tau}$  of  $\tau$  follows from minimizing  $\sum_{t=1}^T \varepsilon_t^2(\tau)$ , where  $\varepsilon_t(\tau) = Y_t + \sum_{j=1}^{\infty} (-1)^j \tau^j Y_{t-j} \prod_{k=1}^j Y_{t-k}$ . The estimating equation is given by  $S_T(\tau) = \sum_{t=1}^T e_t(\tau) e_t'(\tau) = 0$ , where  $e_t'(\tau) = \sum_{j=1}^t (-1)^j j \tau^{j-1} Y_{t-j} \prod_{k=1}^j Y_{t-k}$ . Column 6 shows the CLS estimates averaged over 1,000 MC replications. For comparison purposes, we have executed the same bootstrap procedure as above for the CLS estimates. The simulation results, denoted by  $\bar{\hat{\tau}}_{\text{BS}}$  and  $\hat{\sigma}_{\hat{\tau}_{\text{BS}}}$ , are summarized in columns 7 – 8 of Table 6.3. It is clear that the bootstrap CLS estimate  $\bar{\hat{\tau}}_{\text{BS}}$  is close to its corresponding estimate  $\hat{\tau}$  for all values of  $|\tau|$  and both sample sizes.

**Table 6.3:** Simulation and bootstrap results for the moment estimator  $\hat{\tau}^*$  of Exercise 6.6, and for the CLS estimator  $\hat{\tau}$ ; 1,000 MC replications, and  $B = 1,000$  bootstrap replicates.

$T$	$\tau$	Moment estimator			CLS estimator		
		$\hat{\tau}^*$	$\hat{\tau}_{BS}^*$	$\hat{\sigma}_{\hat{\tau}_{BS}^*}$	$\hat{\tau}$	$\hat{\tau}_{BS}$	$\hat{\sigma}_{\hat{\tau}_{BS}}$
500	-0.4	-0.3737	-0.2847	0.1924	-0.3975	-0.3951	0.0227
	-0.2	-0.2060	-0.2188	0.0734	-0.2002	-0.1992	0.0280
	0.2	0.2042	0.2045	0.0648	0.1995	0.2003	0.0279
	0.4	0.3735	0.3034	0.1552	0.3958	0.3950	0.0228
1,000	-0.4	-0.3866	-0.3329	0.1459	-0.3976	-0.3972	0.0144
	-0.2	-0.2027	-0.2054	0.0449	-0.1996	-0.1997	0.0184
	0.2	0.2048	0.2087	0.0472	0.2001	0.2001	0.0182
	0.4	0.3917	0.3242	0.1626	0.3986	0.3982	0.0134

Note that there is less agreement between the bootstrap moment estimates  $\hat{\tau}_{BS}^*$  and  $\hat{\tau}^*$  for  $\tau = \pm 0.4$ .

**Remark:** If  $\sigma_\varepsilon^2$  is known, and assuming  $|\lambda| = |\tau\sigma_\varepsilon| < 0.605$ , the moment estimator has the same asymptotic distribution as given in Exercise 6.2(e). When, however,  $\sigma_\varepsilon^2$  is *unknown* and assuming  $|\lambda| < 0.605$ , Kim et al. (1990) show that the moment estimator  $\hat{\tau}^*$  is such that, as  $T \rightarrow \infty$ ,

$$\sqrt{T}(\hat{\tau}^* - \tau) \xrightarrow{D} \mathcal{N}(0, \beta' \Sigma^* \beta),$$

where

$$\beta = (2\sigma_\varepsilon^2)^{-1} \begin{pmatrix} 2 \\ -\tau\{1 + (1 - 2\lambda^2 + 4\lambda^4)(1 - 4\lambda^2 + 8\lambda^6 + 4\lambda^8)^{-1/2}\} \\ \tau\{1 + 3(1 + 2\lambda^2)(1 - 4\lambda^2 + 8\lambda^6 + 4\lambda^8)^{-1/2}\} \end{pmatrix},$$

and  $\Sigma^*$  is a  $3 \times 3$  symmetric matrix whose elements are

$$\begin{aligned} \sigma_{11}^* &= \sigma_\varepsilon^2(1 + 3\lambda^2 - \lambda^4)/(1 - \lambda^2), \\ \sigma_{12}^* &= 4\sigma_\varepsilon^3\lambda(1 + 4\lambda^2 + \lambda^4)/(1 - \lambda^2), \\ \sigma_{13}^* &= 2\sigma_\varepsilon^3\lambda(2 + 7\lambda^2 + 3\lambda^4 + 6\lambda^6)/(1 - \lambda^2), \\ \sigma_{22}^* &= 2\sigma_\varepsilon^4(1 + 15\lambda^2 + 18\lambda^4 + 44\lambda^6 + 42\lambda^8 - 102\lambda^{10})/(1 - \lambda^2)^3(1 - 3\lambda^4), \\ \sigma_{23}^* &= \sigma_\varepsilon^4\lambda^2(30 + 106\lambda^2 + 239\lambda^4 + 159\lambda^6 - 132\lambda^8)/(1 - \lambda^2)^2, \\ \sigma_{33}^* &= \sigma_\varepsilon^4(1 + 23\lambda^2 + 80\lambda^4 + 76\lambda^6 - 84\lambda^8 - 120\lambda^{10} - 216\lambda^{12})/(1 - \lambda^2)(1 - 3\lambda^4). \end{aligned}$$

### Second part

Table 6.4 shows the asymptotic standard deviation of the moment estimator  $\hat{\tau}^*$  calculated on the basis of the above results, with  $\sigma_\varepsilon^2 = 1$ . We see that for  $\tau = \pm 0.2$  and  $T = 1,000$  the results are almost identical to the bootstrap estimates of the standard deviation of the moment estimator given in Table 6.3, column five. However, for  $\tau = \pm 0.4$  the results in these two tables do not match very well. Definitely, more observations are needed before the bootstrap method is asymptotically valid for the moment estimator.

The following two MATLAB codes can be used to replicate the results in Tables 6.3 and 6.4.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 6.6, first part
% File: exercise_6-6.m (contains first and second part)
%
% MC and BS results for the moment estimator and the LS estimator.
%
% BL model:  $Y_{\{t\}} = \beta * \text{eps}_{\{t-1\}} * Y_{\{t-2\}} + \text{eps}_{\{t\}}$ ,
%           with  $\text{eps}_{\{t\}} \sim N(0,1)$  distributed.
%
% MM = Method of moments estimator,
%        $2\text{mean}(Y)/[(g_0-g_1) + (g_0^{\{2\}}-6g_0g_1-3g_1^{\{2\}})^{\{1/2\}}]$ 
% LS = Least squares estimator
% BS = Bootstrap estimator
%
% totestm = [LS, MM, Bootst MM, Bootstr LS;
%           std.LS (no BS),std. MM (no BS),std. Bootst MM,std. Bootstr LS]
% table = Results for MM and LS estimators, as shown in Table
%         6.2 of the Solutions Manual.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
s = RandStream('mcg16807','Seed',0); % Set the seed
RandStream.setGlobalStream(s)

x(1) = 0;
y = zeros(1000);
nboot = 1000;
sumMM = 0;
nrep = 1000;
n = 1100;
b = 0.4;

% True percentage points of Normal distribution
p = [0.01 0.025 0.05 0.1 0.203 0.305 0.4 0.5 ...
     0.6 0.695 0.797 0.9 0.95 0.975 0.99];
z = norminv(p,0,1);

t = [-2.2 -2 -1.8 -1.6 -1.4 -1.2 -1 -0.8 -0.6 -0.4 -0.2 0 ...
     0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% MM estimator %%%%%%%%%
for jrep = 1:nrep % Start replications
    r = normrnd(0,1,1,n+1);
    for i = 2:n % Generation of X's
        x(i) = b*x(i-1)*r(i-1)+r(i);
    end
    y = x(101:n); % 100 observations for warming-up
    ny = length(y);
    avgy = mean(y); % Calculation of moments
    vary = var(y)*(ny-1)/ny;

```

```

sum = 0;
for i = 1:ny-1
    sum = sum+(y(i)-avgy)*(y(i+1)-avgy); % Lag 1 sample ACVF
end
covy = sum/ny;

ry = covy/varvy;
if (ry < 0.1547)
    sumMM = sumMM+1;
bhead = (2*avgy)/((varvy-covy)+sqrt(varvy^2-6*varvy*covy-3*covy^2));
bMM(sumMM) = bhead; % The MM estimate
end

%%%%%%%%%%%% LS %%%%%%%%%%%%%%
% Initialization
nnLS = 0;
n1 = ny-1;
bLS = b;
b0 = 0;

%
% Calculation of LS estimate

dbc = b0-bLS;
for uu = 1:100
    if abs(dbc) > 0.00001
        b0 = bLS;
        nnLS = nnLS+1;

%% Generation of e's
e(1) = y(1);
for j = 2:ny
    e(j) = y(j)-b0*y(j-1)*e(j-1);
end

%% Generation of w's
w(1) = 0;
for j = 2:ny
    w(j) = e(j-1)*y(j-1)-b0*w(j-1)*y(j-1);
end

%% Computation of delta bLS
suma = 0;
sumb = 0;
for i = 1:ny
    suma = suma+e(i)*w(i);
    sumb = sumb+w(i)^2;
end
dbc = suma/sumb;

%% The improved LS estimate
bLS = b0+dbc;
end

```

```

    end % end uu=1:100
        bestLS(jrep) = bLS;
end % end replications j=1:nrep NO BOOTSTRAPPING

betaMM = mean(bMM); % Overall mean MM estimator
StdbetaMM = sqrt(var(bMM)); % Overall standard dev
betaLS = mean(bestLS);
StdbetaLS = sqrt(var(bestLS)); % Overall std dev LS

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Bootstrapping MM estimator %%%%%%%%%%
res(1) = y(1);
resLS(1) = y(1);
for j = 2:ny
    res(j) = y(j)-betaMM*y(j-1)*res(j-1);
    resLS(j) = y(j)-betaLS*y(j-1)*resLS(j-1);
end
ares = mean(res);
aresLS = mean(resLS);

for i = 1:ny
    ecen(i) = res(i)-ares;
    ecenLS(i) = resLS(i)-aresLS;
end
id = ceil(rand(ny,nboot)*ny); % Generate a matrix of integers
                                % that range from 1 to ny to use
                                % as indices into the centered residuals

bstres = ecen(id); % Use this index to 'sample' from the
                    % centered residuals

bstresLS = ecenLS(id);

xst = zeros(ny,nboot);
xstLS = zeros(ny,nboot);

for j = 1:nboot
    xst(1,j) = bstres(1,j);
    xstLS(1,j) = bstresLS(1,j);
    for i = 2:ny
        xst(i,j) = bstres(i,j)+betaMM*xst(i-1,j)*bstres(i-1,j);
        xstLS(i,j) = bstresLS(i,j)+betaLS*xstLS(i-1,j)*bstresLS(i-1,j);
    end
end

axst = mean(xst);
vxst = var(xst)*(ny-1)/ny;

for j = 1:nboot
    xx = (xst(:,j)-axst(j));
    xx0 = xx(1:ny-1);

```



```

    xx1      = xx(2:ny);
    cxst(j) = xx0'*xx1/ny;
    rxst(j) = cxst(j)/vxst(j);    % Lag 1 autocorrelation
end

nb = 0;
for j = 1:nboot
    if rxst(j) < 0.1547    % MM method fails if rxst > 0.1547
                        % (no real solution)
        nb = nb+1;        % Bootstrap sample size
        bhst(j) = 2*axst(j)/((vxst(j)-cxst(j))+sqrt(vxst(j)^2 ...
                        -6*vxst(j)*cxst(j)-3*cxst(j)^2));
    end
end

% Computation of BS sample mean and variance of the MM estimator
betaBSMM = mean(bhst);
StdBSMM  = sqrt(var(bhst)*(nb-1)/nb);

%%%%%%%%%%%%%% Bootstrapping LS estimator %%%%%%%%%%%%%%%
%% Computation of BS LS estimate

for jj = 1:nboot
% Initialization
    nn      = 0;
    n1      = ny-1;
    bBLS    = b+0.1;
    bOB     = 0;

% Calculation of LS estimate
    for uu = 1:100
        if abs(bOB-bBLS)> 0.00001
            bOB = bBLS;
            nn  = nn+1;
%% Generation of e's
            e(1,jj) = xstLS(1,jj);
            for j = 2:ny
                e(j,jj) = xstLS(j,jj)-bOB*xstLS(j-1,jj)*e(j-1,jj);
            end
%% Generation of w's
            w(1,jj) = 0;
            for j = 2:ny
                w(j,jj) = e(j-1,jj)*xstLS(j-1,jj)-bOB*w(j-1,jj)*xstLS(j-1,jj);
            end
%% Computation of delta bLS
            suma = 0;
            sumb = 0;
            for i = 1:ny
                suma = suma+e(i,jj)*w(i,jj);

```

```

        sumb = sumb+w(i,jj)^2;
    end
    dbc = suma/sumb;
%% The improved LS estimate
    bBLS = b0B+dbc;
end
end % end of loop uu=1:1000
bestBBLS(jj) = bBLS;
end % end of loop jj=1:nboot

% Computation of BS LS estimate sample mean and std. dev.
betaBSLS = mean(bestBBLS);
StdBSLS = sqrt(var(bestBBLS)*(nb-1)/nb);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Bootstrap MM percentages %%%%%%%%%%%
nc = zeros(15,1);
nh = zeros(23,1);

for k = 1:nboot % betaMM = No BS
    zb(k) = (bhst(k)-betaMM)/StdBSMM; % bhst = BS results
    for i = 1:15
        if zb(k) <= z(i)
            nc(i) = nc(i)+1;
        end
    end
    for j = 1:23
        if zb(k) <= t(j)
            nh(j) = nh(j)+1;
        end
    end
end
for i = 1:15
    bcdfMM(i) = nc(i)/1000; % Approximate CDF
end
for j = 1:23
    bfrqMM(j) = nh(j)/1000; % Relative frequency
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Bootstrap LS percentages %%%%%%%%%%%
nc = zeros(15,1);
nh = zeros(23,1);

for k = 1:nboot % betaMM = No BS
    zb(k) = (bestBBLS(k)-betaLS)/StdBSLS; % bhstLS = BS results LS
    for i = 1:15
        if zb(k) <= z(i)
            nc(i) = nc(i)+1;
        end
    end
end
end

```

```

        for j = 1:23
            if zb(k) <= t(j)
                nh(j) = nh(j)+1;
            end
        end
    end
end
for i = 1:15
    bcdfLS(i) = nc(i)/1000;    % Approximate CDF
end
for j = 1:23
    bfrqLS(j) = nh(j)/1000;    % Relative frequency
end

table = [betaMM,betaBSMM,StdBSMM,betaLS,betaBSLS,StdBSLS];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 6.6, second part
% File: exercise_6-6.m (contains first and second part)
%
% Computation of asymptotic standard deviation (sdMM) of the
% MM estimator tau^{*} based on Kim et al. (1990).
%
% Reference:
% Kim, W.K., Billard, L., and Basawa, I.V. (1990).
% Estimation for the first-order diagonal BL time series model.
% Journal of Time Series Analysis, 11(3), 215-229.
% DOI: 10.1111/j.1467-9892.1990.tb00053.x.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
n    = 1000;    % sample size
se   = 1;
tau  = 0.4;    % parameter
lam  = tau*se;
s11  = (se^2)*(1+3*lam^2-lam^4)/(1-lam^2);
s12  = 4*(se^3)*lam*(1+4*lam^2+lam^4)/(1-lam^2);
s13  = 2*(se^3)*lam*(2+7*lam^2+3*lam^4+6*lam^6)/(1-lam^2);
s22  = 2*(se^4)*(1+15*lam^2+18*lam^4+44*lam^6+42*lam^8- ...
    102*lam^(10))/(((1-lam^2)^3)*(1-3*lam^4));
s23  = (se^4)*(lam^2)*(30+106*lam^2+239*lam^4+159*lam^6- ...
    132*lam^8)/(1-lam^2)^2;
s33  = (se^4)*(1+23*lam^2+80*lam^4+76*lam^6-84*lam^8- ...
    120*lam^(10)-216*lam^(12))/((1-lam^2)*(1-3*lam^4));
SS   = [s11 s12 s13; s12 s22 s23; s13 s23 s33];

b1   = se^(-2);
b2   = -tau*(2*se^2)^(-1)*(1+(1-2*lam^2+4*lam^4)* ...
    (1-4*lam^2+8*lam^6+4*lam^8)^(-1/2));
b3   = tau*(2*se^2)^(-1)*(1+3*(1+2*lam^2)* ...
    (1-4*lam^2+8*lam^6+4*lam^8)^(-1/2));
beta = [b1 b2 b3];

```

**Table 6.4:** Asymptotic standard deviation of the moment estimator  $\hat{\tau}^*$  for the case  $\tau = \pm 0.1, \dots, \pm 0.4$ , and  $T = 500$  and 1,000.

$T$	$\tau$			
	$\pm 0.1$	$\pm 0.2$	$\pm 0.3$	$\pm 0.4$
500	0.0488	0.0626	0.0927	0.1675
1,000	0.0345	0.0442	0.0655	0.1184

```
sdMM = sqrt(beta * SS * beta' / n);
```

**6.7** For  $\tau = 0.6$  and  $\sigma_\varepsilon^2 = 1$  the condition for asymptotic normality of the CLS estimator  $\hat{\tau}$  of  $\tau$  is violated ( $\lambda^8 = \tau^8 \sigma_\varepsilon^8 > 1/105$ ). In this case, no results on the asymptotic distribution or the rate of convergence of  $\hat{\tau}$  are available, apart from the fact that  $\hat{\tau}$  is still strongly consistent, because for proving consistency only  $\tau^4 \sigma_\varepsilon^4 < 1/3$  is required. Figure 6.1 shows the histograms of  $\sqrt{T}(\hat{\tau} - \tau)$  and  $\sqrt{T}(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)$  for  $T = 250$  and  $T = 1,000$ .

- (a) From Figures 6.1(a) and (c) and the corresponding summary statistics, we observe that for both values of  $T$  the estimator  $\hat{\tau}$  is far from normally distributed. In fact, as  $T$  increases the non-normal feature of the simulated data set increases. However, from both the mean and the median values we see that the estimator is still consistent. Note the negative mean and median of the distribution of  $\sqrt{T}(\hat{\tau} - \tau)$ . This was observed earlier in Example 6.5 of the main text.
- (b) Figures 6.1(b) and (d) and the corresponding summary statistics indicate that for both values of  $T$  the estimator  $\hat{\sigma}_\varepsilon^2$  is not normally distributed. However, as can be seen from the values of the skewness, kurtosis, and the Jarque–Bera (JB) normality test, the non-normal features are less prominent than those reported in Figures 6.1(a) and (c). The histograms and the summary statistics indicate the consistency of  $\hat{\sigma}_\varepsilon^2$ , but the estimator overestimates the parameter  $\sigma_\varepsilon^2$ .

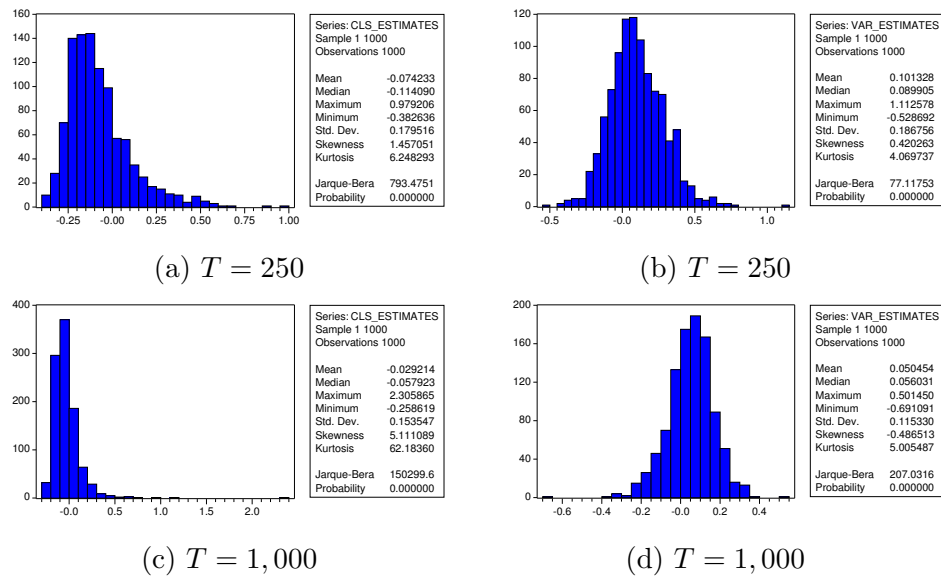
**Remark:** Recently, Tan and Wang (2016) propose the LASSO (least absolute shrinkage and selection) method in combination with the GCV method, the LARS (least angle regression) method, and the stepwise regression method for the selection and estimation of diagonal and subdiagonal BL models. The paper includes an analysis (and data) of the real annual ice condition at Bohai, China, for the time period 1966–1993.

**6.8** (a) The LSTAR(2; 1, 1) model is given by

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + (\xi_0 - \xi_1 Y_{t-1}) / (1 + \exp(-\gamma(Y_{t-1} - c))) + \varepsilon_t, \quad \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),$$

where  $\phi_0 = 1$ ,  $\phi_1 = 0.9$ ,  $\xi_0 = 3$ ,  $\xi_1 = 1.7$ ,  $\gamma = 10$ , and  $c = 5$ . As  $Y_{t-1} \rightarrow -\infty$ , the behavior of  $\{Y_t, t \in \mathbb{Z}\}$  is governed by the AR(1) process  $1 + 0.9Y_{t-1} + \varepsilon_t$ , and as  $Y_{t-1} \rightarrow +\infty$  the behavior of  $\{Y_t, t \in \mathbb{Z}\}$  is governed by  $4 - 0.8Y_{t-1} + \varepsilon_t$ . Thus, in both regimes the process is stationary (local stationarity).

The LSTAR parameters estimates can be obtained by means of the following R code.



**Figure 6.1:** Histograms of  $\sqrt{T}(\hat{\tau} - \tau)$  ((a) and (c)), and histograms of  $\sqrt{T}(\hat{\sigma}_{\varepsilon}^2 - \sigma_{\varepsilon}^2)$  ((b) and (d)) for  $\tau = 0.6$ ,  $T = 250$  and 1,000; 1,000 MC replications.

```
#####
# R code: Exercise 6.8
# File: Exercise_6-8.r
#####
library(tsDyn)
set.seed(1234)
nsim <- 100 # 100 Monte Carlo replications
yy <- matrix(nrow=200,ncol=nsim) # initialize
y <- matrix(nrow=300,ncol=1)
f <- matrix(nrow=300,ncol=1)
ynew <- matrix(nrow=200,ncol=1)
for(it in 1:nsim){
  y[1] <- 0 # starting condition
  f[1] <- 0
  u <- rnorm(300)
  for(i in 2:300){
    f[i] <- 1/(1+exp(-10*(y[i-1]-5)))
    y[i] <- (1+0.9*y[i-1])+(3-1.7*y[i-1])*f[i]+u[i]
  }
  ynew <- y[101:300] # effective sample size is T=200
  yy[,it] <- cbind(ynew)
}
coeff <- array(1,c(6,1,nsim)) # 6 parameters
for(j in 1:nsim) {
  mod.lstar <- lstar(yy[,j],m=2,d=1,mL=1,mH=1,mTh=c(0,1),
    starting.control=list(gammaInt=c(0,20),
```

**Table 6.5:** Parameter estimates of an LSTAR(2; 1, 1) model computed with the R-tsDyn package;  $T = 1,000$ , 100 MC replications.

Statistic	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\gamma}$	$\hat{c}$
Mean	1.39	0.57	0.62	-0.40	10.54	1.63
Standard deviation	0.64	0.41	1.20	0.78	9.62	1.14

```

                                nGamma=100))
# Restricting the initial value of gamma to the interval
# (0,20), and setting the number of smoothing values
# (gamma) in the grid at 100
  coeff[, ,j] <- mod.lstar$coefficients
}
tot <- cbind(coeff[1, ,],coeff[2, ,],coeff[3, ,],coeff[4, ,],
            coeff[5, ,],coeff[6, ,])
meantotal <- c(6) # initialize
stdtotal <- c(6)
for(j in 1:6){
  meantotal[j] <- mean(coeff[j, ,]) # mean
  stdtotal[j] <- sd(coeff[j, ,]) # standard deviation
}

```

- (b) Figures 6.2(a) – (f) show histograms and descriptive statistics of the LSTAR parameter estimates for  $T = 200$ . In all cases the Jarque–Bera (JB) test statistic rejects normality. Thus, one has to be wary about standard errors and Student  $t$ -statistics of fitted LSTAR models. The reason is that the nonlinear LS estimates do not satisfy the normality assumption. The overall averages of the point estimates of all parameters are not very close to their true values. The estimated values of  $\phi_0$ ,  $\phi_1$ ,  $\xi_1$  and  $\gamma$ , but not  $\xi_0$  and  $c$ , are contained within a confidence interval of  $\pm 2$  the standard deviations of the actual values. Nevertheless, it is clear that it is crucial to provide the optimization routine with very good initial values of  $\gamma$  and  $c$ . For  $T \rightarrow \infty$ , the estimators will convergence to the true values. Consider, for instance, the LSTAR model with  $T = 1,000$  and 100 MC replications. In that case we obtain the (improved) estimation results given in Table 6.5.

- (c) **Optional:** The S-Plus code consists of lines 2–18 of the R code in part (a) plus the following additional lines.

```

gam    = array(1,c(1,1,nsim)) # 1 parameter
coeff  = array(1,c(4,1,nsim)) # 4 parameters
thresh = array(1,c(1,1,nsim)) # 1 parameter
for(j in 1:nsim) {
  mod.lstar <- STAR(yy[,j],p=1,d=1)
  gam[, ,j] <- mod.lstar$gamma
  coeff[, ,j] <- mod.lstar$coef
  thresh[, ,j] <- mod.lstar$threshold
}

```

**Table 6.6:** Parameter estimates of an LSTAR(2;1,1) model computed with the S-Plus FinMetrics package;  $T = 200$ , and 100 MC replications.

Statistic	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\gamma}$	$\hat{c}$
Mean	1.04	0.83	4.14	-0.76	1486.19	2.77
Median	1.03	0.86	4.69	-0.90	6.45	2.94
Maximum	1.52	1.21	7.52	1.36	22026.47	3.54
Minimum	0.50	-0.35	-5.70	-1.62	1.82	1.49
Standard deviation	0.18	0.20	2.56	0.56	4876.92	0.49
Skewness	0.10	-2.62	-1.43	1.29	3.85	-0.68
Kurtosis	3.31	15.93	5.40	5.05	16.82	2.48

```

mean(cbind(thresh[1,,]))
mean(cbind(gam[1,,]))
sd(cbind(thresh[1,,]))
sd(cbind(gam[1,,]))
meancoeff <- c(4)      # initialize
sdcoeff   <- c(4)
for(j in 1:4){        # 4 parameters
  meancoeff[j] <- mean(coeff[j,,])
  sdcoeff[j]   <- sd(coeff[j,,])
}
tot = cbind(coeff[1,,],coeff[2,,],coeff[3,,],coeff[4,,],
            gam[1,,],thresh[1,,])

```

Using S-Plus FinMetric version 2.0.4 (2006), and no constraints on the search for the optimal value of  $\gamma$ , the estimation results are given in Table 6.6.

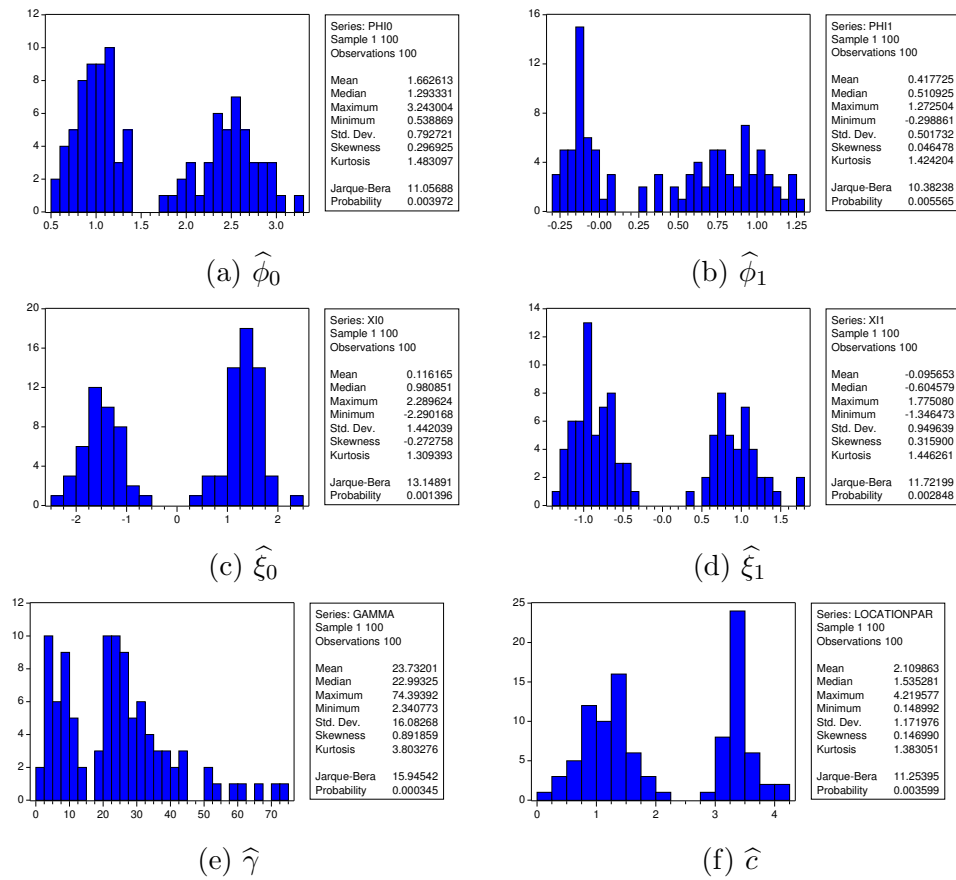
We observe that all parameters are very close to their true values, except for  $\gamma$ . As compared to the descriptive statistics corresponding to Figures 6.2(a) – (f), it is evident that there are marked differences. Clearly, it pays off to restrict the search for the optimal value of  $\gamma$  in some way, an option which is not available in the current version of the STAR command. Despite these findings, and in the absence of a benchmark nonlinear data set, it is not possible to say which computer code is “best”.

- 6.9** (a) The Lin–Mudholkar test statistic, denoted by  $Z_2$ , is given by expression (1.7) of the main text. If a time series  $\{Y_i\}_{i=1}^n$  consists of i.i.d. normal variables, the statistic  $Z = (n/3)^{1/2}Z_2$  is asymptotically normally distributed with mean 0 and unit variance. Below is the MATLAB code for computing  $Z$ . Its value is  $Z = 0.3083$  ( $p$ -value = 0.3789). Hence, the null hypothesis of normality cannot be rejected for the SETAR residuals.

```

function[Z,P] = LINMUD(X)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 6.9(a)
% File: Exercise_6-9.m
%
% Purpose: Provide the Lin-Mudholkar statistic to test the

```



**Figure 6.2:** Histograms of the LSTAR parameter estimates for  $T = 200$ ; 100 MC replications.

```

%          assumption that the input N*1 data vector X is normally
%          distributed.
% Input:   X = data
% Output:  Z-statistic
%          P-value
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
N = length(X);
for I=1:N
    T      = X(I);
    X(I)   = 0;
    BX(I)  = mean(X);
    R(I)   = dot(X,X);
    V(I)   = R(I)-(BX(I)*BX(I)/(N-1))*N*N;
    V(I)   = V(I)/N;
    ILS = 1;
    if V(I) < 0
        ILS = -1;
    end
end

```



```

end
V(I) = ILS*V(I);
if V(I) == 0
    V(I) = 0;
else
    V(I) = V(I)^(0.333333);
    V(I) = ILS*V(I); % V = cube root of
%           (Nth of (SS of X)-(N-1)th of (SS of X))
end
Y(I) = V(I);
X(I) = T;
end
Rcorr = corrcoef(X,Y);
R      = Rcorr(1,2);
Z2     = 0.5*log((1+R)/(1-R));
Z      = Z2/sqrt(3/N); % Z = asymptotically N(0,1) distributed
P      = 1-normcdf(Z,0,1);

```

- (b) The Doornik–Hansen test statistic is based on the Bowman and Shenton (1975) test statistic. Actually, this test statistic is identical to the Jarque–Bera (JB) test statistic for i.i.d. data (see Section 1.3.1 of the main text), although its construction is not the same. Let  $Z_1$  and  $Z_2$  denote the transformed skewness and kurtosis, where the transformation creates a test statistic much closer to the standard normal. The transformation for the skewness is as in D’Agostino (1970), and the kurtosis is transformed from a gamma to a chi-square distribution using the Wilson–Hilferty cubed root transformation, which then is translated into standard normal. The test statistic is given by  $E = Z_1^2 + Z_2^2$ . Under the null hypothesis of normality it is distributed asymptotically as a  $\chi_2^2$  distribution; see Doornik and Hansen (2008, Appendix) for details about the transformation of the third and fourth moments.

Using the function `normality.test1` in the `R-normwhn.test` package the value of the Doornik–Hansen test statistic is 19.018 ( $p$ -value = 0.000). Hence, the SETAR residuals do not have a normal distribution. Clearly, this observation differs from the one in part (a) based on the Lin–Mudholkar test statistic. Some ideas about the “source” of non-normality, can be obtained from the following components:

<i>sk</i>	skewness statistic
<i>k</i>	kurtosis statistic
<i>rtb1</i>	skewness of standardized variable
<i>b2</i>	kurtosis of standardized variables
<i>z1</i>	skewness of transformed variable
<i>z2</i>	kurtosis of transformed variable
<i>pvalsk</i>	$p$ -value under null of no skewness
<i>pskneg</i>	$p$ -value under null of no negative skewness
<i>pskpos</i>	$p$ -value under null of no positive skewness
<i>pvalk</i>	$p$ -value under null of no kurtosis
<i>pkneg</i>	$p$ -value under null of no negative kurtosis
<i>pkpos</i>	$p$ -value under null of no positive kurtosis
<i>Ep</i>	value of the normality test statistic
<i>dof</i>	degrees of freedom
<i>Sig.Ep</i>	significance of normality test statistic

In fact, the null hypothesis “ $\mathbb{H}_0$ : data do not have kurtosis” is rejected at the 5% nominal significance level ( $p$ -value = 0.000). Also, the null hypothesis “ $\mathbb{H}_0$ : data do not have positive kurtosis” is rejected ( $p$ -value = 0.000). The above results are further supported by results obtained with the function `normality.test2` which allows for variable(s) being weakly dependent rather than independent. The associated value of the normality test statistic is 18.734 ( $p$ -value = 0.000).

**Note:** MATLAB Central offers two functions for computing the Doornik–Hansen omnibus multivariate (univariate) normality test statistic: `DorHanomunortest.m` and `MVomnibusnew.m`. Both functions produce the same test results as given above.

- (c) To calculate the bootstrap distribution of the two normality test statistics, using CLS residuals obtained from a time series  $\{Y_t\}_{t=1}^T$  fitted to a SETAR(2;  $p_1, p_2$ ) model, we briefly describe the main steps of the simulation algorithm. First, consider the homoskedastic bootstrap, which treats the errors as i.i.d. draws with the same variance across both regimes. The algorithm can be summarized as follows.
- (i) Generate a random sample  $\{\hat{\varepsilon}_t\}_{t=1}^T$  by sampling with replacement from the CLS residuals obtained from a SETAR(2;  $p_1, p_2$ ) model-fit with  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ .
  - (ii) Using fixed initial conditions  $(Y_0, \dots, Y_{-p+1})$  where  $p = p_1 \vee p_2$ , recursively generate a set of pseudo-observations  $\{Y_t^*\}_{t=1}^T$  given the SETAR(2;  $p_1, p_2$ ) model with parameters estimated previously by CLS. Below the initial conditions are set equal to zero.
  - (iii) Given the time series  $\{Y_t^*\}_{t=1}^T$ , calculate the value of the normality test statistic using the residuals from a new CLS–SETAR estimation.
  - (iv) Repeat steps (i) – (iii)  $B$  times (here  $B = 1,000$ ). Next, using the BS replicates, compute the empirical size of the normality test statistic under study.

In addition, it is also desirable to check if there is enough data to estimate the AR parameters in each regime. Preferably, a minimum of 20 observations per regime is required. If this is not the case, repeat the simulation step.

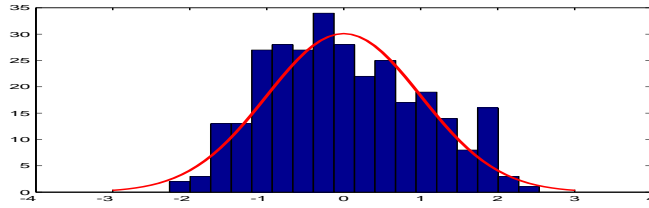
For the heteroskedastic case, step (i) is based on CLS residuals obtained from a SETAR model with regime specific variances, i.e.  $\{\varepsilon_t^{(i)}\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_i^2)$  ( $i = 1, 2$ ). Then, when simulating the EDF of  $\{Y_t^*, t \in \mathbb{Z}_1^T\}$ , use regime specific bootstrap residuals.

Table 6.7 gives the empirical sizes (rejection rates  $\times 100$ ) of the Lin–Mudholkar normality test statistic  $Z$ , and the Doornik–Hansen omnibus test statistic  $E$ . We see that for Gaussian WN errors the empirical sizes are close to the nominal sizes for both sample sizes  $T$ . For SETAR residuals, however, the rejection frequencies of the  $Z$  and  $E$  test statistics are significantly higher than the nominal levels. Therefore, both tests tend to reject the null hypothesis of the SETAR residuals far too often. Increasing the sample size  $T$  from 100 to 300 does not improve the size of both test statistics. One possible remedy of this size distortion problem is to use bootstrapped critical values rather than asymptotic critical values.

A common factor of both test statistics is that they rely on the departure of symmetry of possible alternatives to the normal distribution function. Figure

**Table 6.7:** Empirical sizes of the Lin–Mudholkar normality test statistic  $Z$ , and the Doornik–Hansen omnibus normality test statistic  $E$ , based on  $B = 10,000$  BS replicates.

Statistic	$T$	i.i.d. $\mathcal{N}(0, 1)$			SETAR-residuals					
		errors			Case (i)			Case (ii)		
		1%	2.5%	5%	1%	2.5%	5%	1%	2.5%	5%
$Z$	100	0.70	2.19	4.86	3.26	6.19	9.59	3.26	6.33	10.36
	300	1.05	2.58	5.20	4.85	8.24	11.82	4.47	7.86	11.87
$E$	100	1.62	3.02	5.37	5.35	9.54	14.76	5.54	9.91	14.84
	300	1.31	2.99	5.38	7.71	12.96	18.43	7.44	12.33	18.33



**Figure 6.3:** Histogram of a typical SETAR residual time series (case (ii)) with a normal distribution fit imposed;  $T = 300$ .

6.3 shows a histogram of a typical SETAR residual series (case (ii)) of size  $T = 300$  with a normal distribution function fit imposed. The  $p$ -values of the  $Z$  and  $E$  test statistics are 0.980 and 0.0014 respectively. So, the two normality test statistics lead to different conclusions at the 5% nominal significance level. In general, for asymmetric error distributions the use of Pearson residuals in diagnostic checking may lead to erroneous inference. Finally, it should be said that it is quite unrealistic to assume that the threshold value is known in practice.

## Chapter 7

7.1 (a) We find, for  $t \neq s$  such that  $Y_t$  is independent of  $Y_s$ ,

$$\begin{aligned}
 C_{1,Y}(h) &= \mathbb{P}(|Y_t - Y_s| < h) = \mathbb{E}(I(|Y_t - Y_s| \leq h)) \\
 &= \mathbb{E}[\mathbb{E}(I(|Y_t - Y_s| \leq h)|Y_s)] \\
 &= \mathbb{E}[\mathbb{P}(|Y_t - Y_s| \leq h|Y_s)] = \int_{-\infty}^{\infty} \mathbb{P}(|Y_t - y| \leq h) dF(y) \\
 &= \int_{-\infty}^{\infty} \mathbb{P}(y - h \leq Y_t \leq y + h) dF(y) \\
 &= \int_{-\infty}^{\infty} [F(y + h) - F(y - h)] dF(y) \equiv C,
 \end{aligned}$$

where the last step requires  $F(\cdot)$  to be absolutely continuous.

(b) If  $|t - s| > 1$ , then  $(Y_t - Y_s)$  is independent of  $(Y_{t+1} - Y_{s+1})$ , because they do

not have  $Y_j$  in common. This implies

$$\mathbb{P}(|Y_t - Y_s| \leq h, |Y_{t+1} - Y_{s+1}| \leq h) = \mathbb{P}(|Y_t - Y_s| \leq h) \mathbb{P}(|Y_{t+1} - Y_{s+1}| \leq h) = C^2.$$

If  $s = t + 1$ , then conditioning on  $Y_{t+1}$  gives, analogously to part (a),

$$\begin{aligned} \mathbb{P}(|Y_t - Y_{t+1}| \leq h, |Y_{t+1} - Y_{t+2}| \leq h) &= \int_{-\infty}^{\infty} \mathbb{P}(|Y_t - y| \leq h, |Y_{t+2} - y| \leq h | Y_{t+1} = y) dF(y) \\ &= \int_{-\infty}^{\infty} \mathbb{P}(|Y_t - y| \leq h) \mathbb{P}(|Y_{t+2} - y| \leq h) dF(y) \\ &= \int_{-\infty}^{\infty} [F(y+h) - F(y-h)]^2 dF(y) \equiv N. \end{aligned}$$

The same applies if  $s = t - 1$ ; in that case we condition  $|Y_{t+1} - Y_t|$  and  $|Y_{t-1} - Y_t|$  on  $Y_t$ .

- (c) Of the  $(T-2)(T-1)/2$  terms in the expression for  $\widehat{C}_{2,Y}(h)$ , there are  $T-2$  terms with expected value  $N$  (corresponding to  $j = i-1$ ), whereas the remaining  $(T-3)(T-2)/2$  terms have expected value  $C^2$ . Therefore,

$$\lim_{T \rightarrow \infty} \mathbb{E}[\widehat{C}_{2,Y}(h)] = \lim_{T \rightarrow \infty} \frac{2N + (T-3)C^2}{T-1} = C^2 = \{C_{1,Y}(h)\}^2.$$

- 7.2** (a) Note that the conditional mean of the ARCH(1) process is zero, and the conditional variance  $\mathbb{E}[\varepsilon_t^2 | (\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)] = 1 + \theta \varepsilon_{t-1}^2$ . The unconditional mean of the process is zero, and the unconditional variance is equal unity. The log-likelihood is given by

$$\ell(\mathbf{y}; \theta) = \log f_Y(y_1) + \sum_{t=2}^T \log f_{Y_t | Y_{t-1}}(y_t | y_{t-1}),$$

where  $\mathbf{y} = (y_1, \dots, y_T)'$  are observations of the process  $\{Y_t, t \in \mathbb{Z}\}$ . Although the first term can be calculated by determining the marginal density  $f_Y(y)$  up to order  $\theta$ , the stationary density must satisfy

$$\begin{aligned} f_Y(y; \theta) &= \int f_Y(z; \theta) f_{Y_t | Y_{t-1}}(y | z) dz \\ &= \int f_Y(z; \theta) \frac{1}{\{2\pi(1 + \theta z^2)\}^{1/2}} e^{-y^2/2(1 + \theta z^2)} dz. \end{aligned}$$

A Taylor expansion of the above transition probability function around  $\theta = 0$  gives

$$\begin{aligned} f_Y(y; \theta) &= (2\pi)^{-1/2} e^{-y^2/2} \int f_Y(z; \theta) \left[ 1 + \frac{1}{2} \theta z^2 (y^2 - 1) \right] dz + \mathcal{O}(\theta^2) \\ &= (2\pi)^{-1/2} e^{-z^2/2} \left[ 1 + \frac{1}{2} \theta (y^2 - 1) \right] + \mathcal{O}(\theta^2), \end{aligned}$$

where in the first step we have used

$$\begin{aligned} \frac{\partial}{\partial \theta} \Big|_{\theta=0} \frac{1}{(1 + \theta z^2)^{1/2}} e^{-y^2/2(1 + \theta z^2)} &= \\ \left( -\frac{1}{2} e^{-y^2/2(\theta y^2 + 1)} \frac{z^2(-y^2 + \theta z^2 + 1)}{(\theta z^2 + 1)^{5/2}} \right) \Big|_{\theta=0} &= \frac{1}{2} z^2 (y^2 - 1) e^{-y^2/2}, \end{aligned}$$

and in the second step  $\mathbb{E}(Z^2) = 1$ . The copula density is given by

$$c(u_t, u_{t-1}; \theta) = \frac{f_{Y_t|Y_{t-1}}(F^{-1}(u_t; \theta)|F^{-1}(u_{t-1}; \theta); \theta)}{f_Y(F^{-1}(u_{t-1}; \theta); \theta)},$$

where  $F(\cdot)$  is the marginal CDF. Because the effect of  $\theta$  is of order  $\mathcal{O}(\theta^2)$ , we have

$$\begin{aligned} \left. \frac{\partial}{\partial \theta} \right|_{\theta=0} \log c(u_t, u_{t-1}; \theta) &= \left. \frac{\partial}{\partial \theta} \right|_{\theta=0} \log f_{Y_t|Y_{t-1}}(\Phi^{-1}(u_t)|\Phi^{-1}(u_{t-1}); \theta) \\ &\quad - \left. \frac{\partial}{\partial \theta} \right|_{\theta=0} \log f_Y(\Phi^{-1}(u_{t-1}); \theta) \\ &= \left. \frac{\partial}{\partial \theta} \right|_{\theta=0} \log f_{Y_t|Y_{t-1}}(\Phi^{-1}(u_t)|\Phi^{-1}(u_{t-1}); \theta) \\ &= \left. \frac{\partial}{\partial \theta} \right|_{\theta=0} \left( -\frac{1}{2} \log(1 + \theta(\Phi^{-1}(u_{t-1}))^2) - \frac{1}{2} \frac{(\Phi^{-1}(u_t))^2}{1 + \theta(\Phi^{-1}(u_{t-1}))^2} \right) \\ &= -\frac{1}{2} [(\Phi^{-1}(u_{t-1}))^2 - 1] + \frac{1}{2} (\Phi^{-1}(u_t))^2 [(\Phi^{-1}(u_{t-1}))^2 - 1] \\ &= \frac{1}{2} [(\Phi^{-1}(u_t))^2 - 1] [(\Phi^{-1}(u_{t-1}))^2 - 1]. \end{aligned}$$

Replacing  $u_t$  by the realizations  $\hat{u}_t$ , and summing over all available pairs leads to the semiparametric (SP) score-type test statistic

$$Q_{\text{ARCH}}^{\text{SP}} = \sum_{t=2}^T (\Phi^{-1}(\hat{u}_t))^2 (\Phi^{-1}(\hat{u}_{t-1}))^2,$$

where in both terms the value 1 has been removed since it does not affect the distribution of the ranks. Using [Ibragimov and Linnik \(1971, Thm. 19.2.1\)](#) it can be shown, that under the null hypothesis of serial independence, and as  $T \rightarrow \infty$ ,

$$\frac{1}{\sqrt{T-1}} Q_{\text{ARCH}}^{\text{SP}} \xrightarrow{D} \mathcal{N}(0, 1).$$

Critical values for this test statistic and/or  $p$ -values can be obtained by MC simulation.

- (b) Note that the DGP is  $Y_t|Y_{t-1} \sim \mathcal{N}(\theta \text{sign}(Y_{t-1}), 1)$ . Similar as in part (a), the stationary density of the sign AR(1) process must satisfy

$$\begin{aligned} f_Y(y; \theta) &= \int f_Y(z; \theta) f_{Y_t|Y_{t-1}}(y|z) dz \\ &= \int f_Y(z; \theta) \frac{1}{(2\pi)^{1/2}} e^{-(y-\theta \text{sign}(z))^2/2} dz. \end{aligned}$$

Ignoring effects of order  $\mathcal{O}(\theta^2)$  in  $f_Y(\cdot)$ , the copula density is given by

$$\begin{aligned} c(u_t, u_{t-1}; \theta) &= \left. \frac{\partial}{\partial \theta} \right|_{\theta=0} \log f_{Y_t|Y_{t-1}}(\Phi^{-1}(u_t)|\Phi^{-1}(u_{t-1}); \theta) \\ &= \left. \frac{\partial}{\partial \theta} \right|_{\theta=0} -\frac{1}{2} \left( \Phi^{-1}(u_t) - \theta \text{sign}(\Phi^{-1}(u_{t-1})) \right)^2 \\ &= \Phi^{-1}(u_t) \text{sign}(\Phi^{-1}(u_{t-1})). \end{aligned}$$

Hence, replacing  $u_t$  by the realizations  $\widehat{u}_t$ , and summing over all available pairs leads to the score-type test statistic

$$Q_{\text{sAR}}^{\text{SP}} = \sum_{t=2}^T \Phi^{-1}(u_t) \text{sign}(\Phi^{-1}(u_{t-1})).$$

Under the null hypothesis of serial independence, and as  $T \rightarrow \infty$ , it can be shown that  $(1/\sqrt{T-1})Q_{\text{sAR}}^{\text{SP}} \xrightarrow{D} \mathcal{N}(0, 1)$ .

**7.3** We only consider the case  $\ell = 1$ . The case  $\ell > 1$  can be proved in the same way. Let  $\sigma_Y^2 = \text{Var}(Y_t)$ , and  $\rho = \text{Corr}(Y_t, Y_{t-1})$ . With  $f_Y(y) = (1/\sqrt{2\pi}\sigma_Y) \exp\{-\frac{1}{2}(y/\sigma_Y)^2\}$  and

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_Y^2} \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x}{\sigma_Y}\right)^2 - 2\rho\left(\frac{xy}{\sigma_Y^2}\right) + \left(\frac{y}{\sigma_Y}\right)^2\right\}\right\}.$$

Then it is easy to show that

$$\int f_{X,Y}^2(x, y) dx dy - \int f_X^2(x) f_Y^2(y) dx dy = \frac{1}{4\pi\sigma_Y^2} \{(1-\rho^2)^{-1/2} - 1\} \geq 0. \quad (*)$$

Using the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \Delta^*(1) &= \int f_{X,Y}^2(x, y) dx dy - \int f_X(x) f_Y(y) f_{X,Y}(x, y) dx dy \\ &\geq \int f_{X,Y}^2(x, y) dx dy \\ &\quad - \left\{ \int f_{X,Y}^2(x, y) dx dy \right\}^{1/2} \left\{ \int f_X^2(x) f_Y^2(y) dx dy \right\}^{1/2}. \quad (**) \end{aligned}$$

From (\*) it follows that  $\Delta^*(1) \geq 0$ . From its definition we have that  $\Delta^*(1) = 0$  if  $Y_t$  and  $Y_{t-1}$  are independent. On the other hand, if  $\Delta^*(1) = 0$ , then from (\*\*),

$$\int f_{X,Y}^2(x, y) dx dy \leq \int f_X^2(x) f_Y^2(y) dx dy,$$

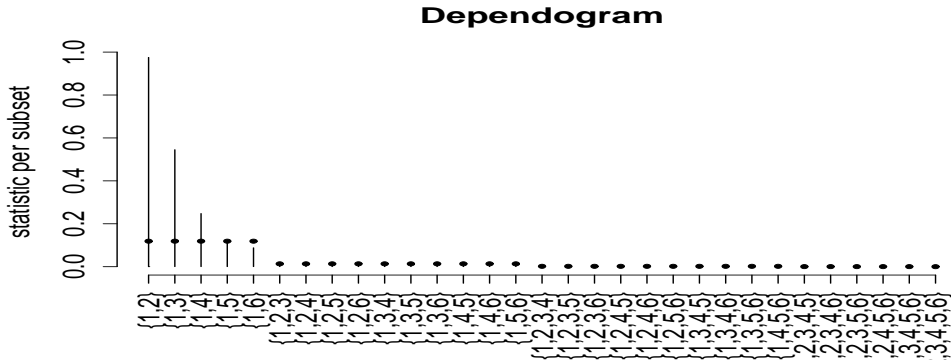
and it follows from (\*) that  $\rho = 0$ , i.e.,  $Y_t$  and  $Y_{t-1}$  must be independent.

**7.4** Assume that under the null hypothesis of serial independence the series  $\{e_t\}_{t=1}^T$  is randomly generated according to a density function  $f(\cdot)$ . Then, for sufficiently small  $h$ , we have

$$\begin{aligned} \frac{1}{h} \mathbb{P}(|e_t - e_s| < h) &= \frac{1}{h} \int_{-\infty}^{\infty} f(e_t) \int_{e_t-h}^{e_t+h} f(e_s) de_s de_t \\ &\approx 2 \int f(e_t)^2 de_t \equiv k. \end{aligned}$$

Moreover, for small  $h$ , it follows that

$$\begin{aligned} \mathbb{E}[C_{m,T}(e; h)] &= \binom{T}{2}^{-1} \sum_{t=1}^{T-1} \sum_{s=t+1}^T \mathbb{P}\left(\max_{i=0,1,\dots,m-1} |e_{t+i} - e_{s+i}| < h\right) \\ &\cong 2(hk)^m, \end{aligned}$$



**Figure 7.1:** Dependogram summarizing the results of the CvM test statistic of serial independence applied to the residuals of an RBF-AR(8) model fitted to the EEG recordings;  $\text{lag.max} = 5$ . A black dot denotes approximate bootstrapped critical values.

and

$$\lim_{T \rightarrow \infty} \lim_{h \rightarrow 0} \frac{\log \mathbb{E}[C_{m,T}(e; h)]}{\log h} = \lim_{h \rightarrow 0} \frac{m \log hk}{\log h} = m.$$

Besides

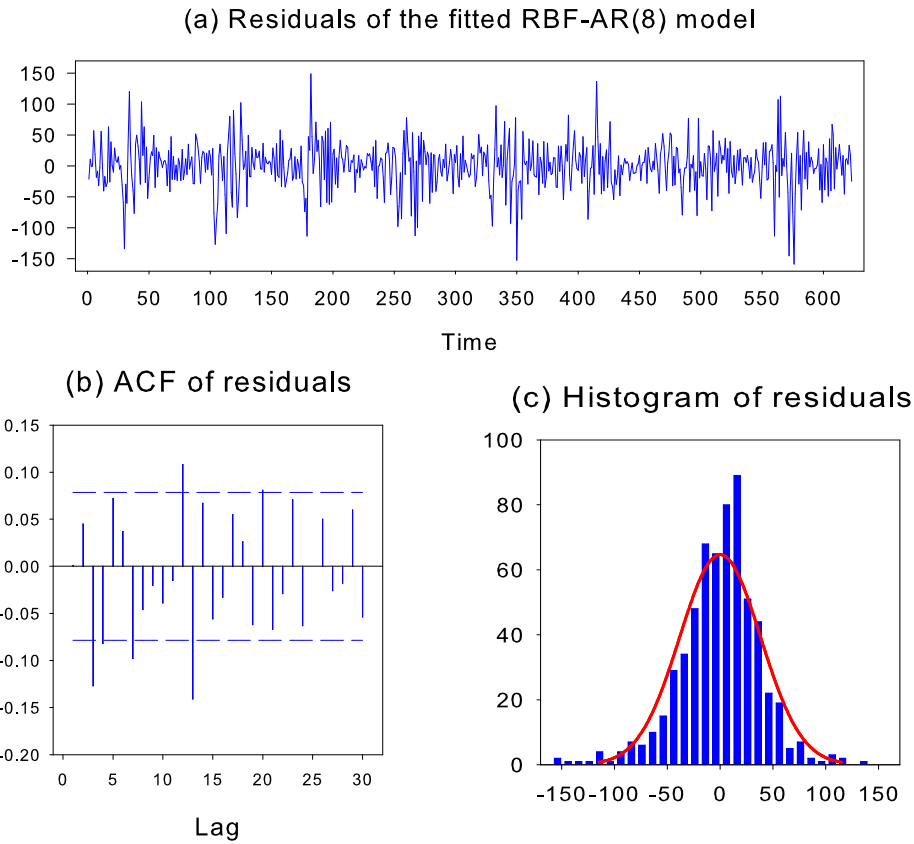
$$\lim_{h \rightarrow 0} \frac{\log \mathbb{P}(|e_t - e_s| < h)}{\log h} = 1.$$

By Jensen's inequality  $\mathbb{E}[\log C_{m,T}(e; h)] \leq \log(\mathbb{E}[C_{m,T}(e; h)])$ . So combining the above results, it follows that for large values of  $T$  and small values of  $h$  we have

$$\begin{aligned} \mathbb{E}[\log C_{m,T}(e; h)] &\leq \log(\mathbb{E}[C_{m,T}(e; h)]) \cong \log \mathbb{P}(|e_t - e_s| < h) \\ &\cong m \log k + m \log h, \end{aligned}$$

and where the equality holds in the linear case. Thus, in the linear regression  $\log C_{m,T}(e; h) = \alpha_m + \beta_m \log h + u$  the left-hand variable has a negative bias from  $m \log k + m \log h$ . Hence, small values of  $h$  imply small values of the bias. So, the estimated slope coefficients  $\hat{\beta}_m$  satisfy  $\mathbb{E}[\hat{\beta}_m] \leq m$ .

- 7.5** (a) Figure 7.2(a) shows a plot of the residuals. We see that at several places the residuals are particularly large, which indicates that the residual variance is nonhomogeneous (non-constant over time). This means that the model is not following the oscillatory behavior very well, compared with other parts of the series. Figure 7.2(b) shows the sample ACF of the residuals with medium dashed lines indicating 95% asymptotic confidence limits. Values exceeding these limits are at lags 3, 4, 7, 12, 13, and 20. The histogram of the residuals in Figure 7.2(c) is not extremely non-Gaussian, but it shows a slightly asymmetric shape around zero. Assuming the residuals are weakly dependent, the  $p$ -values of the skewness ( $\hat{\pi}_{3,e}$ ), kurtosis ( $\hat{\pi}_{4,e}$ ), and normality ( $\hat{\pi}_{34,e}$ ) test statistics (see Section 1.3.1) are 0.968, 0.000 and 0.001, respectively. Thus, normality is rejected at the 5% nominal significance level. The above observations suggest that the RBF-AR(8) model still has room for improvement, i.e. a more sophisticated model is needed for characterizing the local dynamics of the epilepsy oscillations.



**Figure 7.2: EEG recordings.** (a) Plot of the residuals of the fitted RBF-AR(8) model; (b) ACF of the residuals with approximate 95% confidence bands (blue medium dashed lines); and (c) histogram of the residuals;  $T = 623$ .

**Table 7.1:** Bootstrapped  $p$ -values of seven test statistics for high-dimensional serial independence applied to the residuals of the SETAR(2; 2, 2) model fitted to the log (base 10) of the Canadian lynx time series;  $T = 114$ ,  $B = 100$ . Blue-typed numbers indicate rejection of  $\mathbb{H}_0$  at the 5% nominal significance level.

$m$	BDS Rank-based BDS test statistics						
	$S_{m,T}$	$\tilde{I}_{m,T}^*$	$\tilde{M}_{m,T}^*$	$\tilde{T}_{m,T}^*$	$\tilde{I}_{m,T}$	$\tilde{M}_{m,T}$	$\tilde{T}_{m,T}$
2	0.24	0.02	0.68	0.02	0.65	0.68	0.08
4	0.32	0.00	0.66	0.00	0.80	0.66	0.00
6	0.42	0.00	0.52	0.00	0.62	0.52	0.00



(b) The computational aspects can be summarized as follows.

```
Ats <- ts(A,freq=1)           # A = residual time series, T = 623
## As we are going to consider lags up to 5, i.e., subsets of
## {1,...,6} whose cardinality is between 2 and 5 containing {1}.
d <- serialIndepTestSim(623,5) # lag.max = 5
```

```
## The next step consists in performing the three test statistics:
test <- serialIndepTest(Ats,d)
```

```
## Let us see the results:
> test
```

```
Global Cramer-von Mises statistic: 0.7139379 with
      p-value 0.0004995005
Combined p-values from the Mobius decomposition:
      0.0004995005 from Fisher's rule,
      0.0004995005 from Tippett's rule.
```

All three  $p$ -values indicate very strong evidence against the null hypothesis of serial independence. Unsurprisingly, the dependogram, which is tabulated below and summarized in Figure 7.1, indicates that independence is rejected for subsets  $\{1, 2\}$ ,  $\{1, 3\}$ , and  $\{1, 4\}$ . These test results are in accordance with those reported in part (a) based on the sample ACF of the residuals.

```
## Display the dependogram:
> dependogram(test,print=TRUE)
```

The `subset` statistics,  $p$ -values and critical values are:

	subset	statistic	pvalue	critvalue
1	{1,2}	0.9756086434	0.0004995005	1.188337e-01
2	{1,3}	0.5453707585	0.0004995005	1.188337e-01
3	{1,4}	0.2479490473	0.0004995005	1.188337e-01
4	{1,5}	0.1139513116	0.0034965035	1.188337e-01
5	{1,6}	0.0881524143	0.0064935065	1.188337e-01
6	{1,2,3}	0.0142100588	0.0004995005	1.327522e-02
7	{1,2,4}	0.0170914357	0.0014985015	1.327522e-02
8	{1,2,5}	0.0172097357	0.0004995005	1.327522e-02
9	{1,2,6}	0.0122954898	0.0024975025	1.327522e-02
10	{1,3,4}	0.0101665819	0.0104895105	1.327522e-02
11	{1,3,5}	0.0078805050	0.0524475524	1.327522e-02
12	{1,3,6}	0.0067533623	0.0924075924	1.327522e-02
13	{1,4,5}	0.0043661781	0.4420579421	1.327522e-02
14	{1,4,6}	0.0036824860	0.6798201798	1.327522e-02
15	{1,5,6}	0.0059334269	0.1643356643	1.327522e-02
16	{1,2,3,4}	0.0088659678	0.0004995005	1.749712e-03
17	{1,2,3,5}	0.0047901921	0.0004995005	1.749712e-03
18	{1,2,3,6}	0.0028819653	0.0004995005	1.749712e-03
19	{1,2,4,5}	0.0071092258	0.0004995005	1.749712e-03
20	{1,2,4,6}	0.0034064398	0.0004995005	1.749712e-03
21	{1,2,5,6}	0.0063859907	0.0004995005	1.749712e-03

22	{1,3,4,5}	0.0041963047	0.0004995005	1.749712e-03
23	{1,3,4,6}	0.0025498530	0.0004995005	1.749712e-03
24	{1,3,5,6}	0.0036447154	0.0004995005	1.749712e-03
25	{1,4,5,6}	0.0018868016	0.0004995005	1.749712e-03
26	{1,2,3,4,5}	0.0004205817	0.0004995005	2.267765e-04
27	{1,2,3,4,6}	0.0003616803	0.0004995005	2.267765e-04
28	{1,2,3,5,6}	0.0002997129	0.0004995005	2.267765e-04
29	{1,2,4,5,6}	0.0002659954	0.0004995005	2.267765e-04
30	{1,3,4,5,6}	0.0002228651	0.0034965035	2.267765e-04
31	{1,2,3,4,5,6}	0.0000927769	0.0004995005	3.348888e-05

The critical values are such that the simultaneous acceptance region has probability  $1 - 0.05$  under the null.

The individual rejection probability for any statistic obtained from the Mobius decomposition is  $1 - 0.9983467$  under the null.

**Remark:** A simple extension of this exercise is to compute the test statistics with jittered data to avoid possible ties.

**7.6** Tong's (1990) SETAR(2; 7, 2) model given in Table 7.5 of the main text is a refinement of the SETAR(2; 2, 2) model of this exercise. Originally the latter model was fitted to the yearly data covering the time period 1821 – 1919. Here, we consider the complete set of  $T = 114$  observations, and hence we test a time series of 112 residuals for serial independence.

(a)+(b) Bootstrapped  $p$ -values of the nine test statistics are presented in Table 7.1. We see that  $S_{m,T}$ ,  $\widetilde{M}_{m,T}^*$ ,  $\widetilde{I}_{m,T}$ , and  $\widetilde{M}_{m,T}$  fail to reject the null hypothesis of residual serial independence at the 5% nominal significance level, and for all values of  $m$ . Also, at  $m = 2$ , the  $p$ -values of the test statistic  $\widetilde{T}_{m,T}$  suggest that there is no evidence of serial dependence. However, at all values  $m$ , pronounced evidence of residual dependence can be noted from the  $p$ -values of  $\widetilde{I}_{m,T}^*$  and  $\widetilde{T}_{m,T}^*$ . Hence, these latter two test statistics, as well as the  $p$ -value of  $\widetilde{T}_{m,T}$  at  $m = 4$  and  $m = 6$ , alert us against uncritical acceptance of the SETAR(2; 2, 2) model.

**7.7** (a) For a MAR( $K; p_1, \dots, p_K$ ) model the conditional mean and conditional variance of  $\{Y_t, t \in \mathbb{Z}\}$  given  $\mathcal{F}_{t-1}$  are given by

$$\mathbb{E}(Y_t | \mathcal{F}_{t-1}) = \sum_{i=1}^K \pi_i (\phi_{i,0} + \phi_{i,1} Y_{t-1} + \dots + \phi_{i,p_i} Y_{t-p_i}) \equiv \sum_{i=1}^K \pi_i \mu_{i,t}$$

and

$$\text{Var}(Y_t | \mathcal{F}_{t-1}) = \sum_{i=1}^K \pi_i \sigma_i^2 + \sum_{i=1}^K \pi_i \mu_{i,t}^2 - \left( \sum_{i=1}^K \pi_i \mu_{i,t} \right)^2.$$

These expressions apply for any specification of the mixing weights (proportions)  $\pi_i$ , i.e. even when the weights are time dependent ( $\pi_{i,t}$ ). Hence, the Pearson residuals  $\widehat{\varepsilon}_t = (Y_t - \mathbb{E}(Y_t | \mathcal{F}_{t-1}, \widehat{\theta}_T)) / \{\text{Var}(Y_t | \mathcal{F}_{t-1}, \widehat{\theta}_T)\}^{1/2}$  can be computed straightforwardly from the fitted MAR model. Figure 7.3 shows the sample ACFs of  $\{\widehat{\varepsilon}_t\}_{t=1}^{112}$  and  $\{\widehat{\varepsilon}_t^2\}_{t=1}^{112}$ . The sample ACF of  $\{\widehat{\varepsilon}_t\}$  shows significant values at lags 10, 14, and 20, at the 5% nominal level. In contrast, the sample ACF



**Table 7.2:** Canadian lynx time series. Diagnostic test statistics for autocorrelation ( $A_{T,K_1}$ ) and heteroskedasticity ( $H_{T,K_2}$ );  $T = 114$ .  $P$ -values are given in parentheses.

$K_1 = K_2$	$A_{T,K_1}$	$H_{T,K_2}$
5	0.017 (1.000)	$4.281 \times 10^{-8}$ (1.000)
10	0.027 (1.000)	$1.490 \times 10^{-7}$ (1.000)
15	0.037 (1.000)	$4.595 \times 10^{-6}$ (1.000)
20	0.060 (1.000)	$2.460 \times 10^{-7}$ (1.000)
25	0.211 (1.000)	$3.095 \times 10^{-7}$ (1.000)
30	0.367 (1.000)	$6.453 \times 10^{-6}$ (1.000)

```

rng(1);

yy = data;
nr = length(yy);
ss1 = 0.0887; % As given by fitted MAR model, 1st line
ss2 = 0.2020;

term1(1) = yy(1)-0.7107;
term2(1) = yy(2)-0.9784;
term1(2) = yy(2)-0.7107-1.1022*yy(1);
term2(2) = yy(2)-0.9784-1.5279*yy(1);

for i=3:nr
    term1(i) = yy(i)-0.7107-1.1022*yy(i-1)+0.2835*yy(i-2);
    term2(i) = yy(i)-0.9784-1.5279*yy(i-1)+0.8871*yy(i-2);
end

cdftermyy = 0.3163.*normcdf(term1,0,1)+0.6837.*normcdf(term2,0,1);

fterm1 = normpdf(term1./ss1,0,1);
fterm1 = fterm1.*0.3163;
fterm2 = normpdf(term2./ss2,0,1);
fterm2 = fterm2.*0.6837;
ftot = fterm1 + fterm2;
f = log(ftot);
yyfr = [f';cdftermyy'];
r = yyfr;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Simulation part
T = 20000+2;
Tboots = 20000;
sigma1 = 0.0887;
sigma2 = 0.2020;

epsilon = normrnd(0,1,T,1);
eta = normrnd(0,1,T,1);

```

```

yn      = zeros(Tboots,1);
ar1     = zeros(Tboots,1);
ar2     = zeros(Tboots,1);

y(1)    = 0.3163*0.7107+0.6837*0.9784;    % mean of MAR process
yn(1)   = y(1);
xlim    = -norminv(0.3163,0,1);

boolean(1) = xlim-eta(1);

if(boolean(1)<0);
    yn(2) = 0.7107+1.1022*yn(1)+sigma1*epsilon(2);
else
    yn(2) = 0.9784+1.5279*yn(1)+sigma2*epsilon(2);
end

boolean(2) = xlim-eta(2);
if(boolean(2)<0);
    yn(3) = 0.7107+1.1022*yn(2)-0.2835*yn(1)+sigma1*epsilon(3);
else
    yn(3) = 0.9784+1.5279*yn(2)-0.8871*yn(1)+sigma2*epsilon(3);
end

for i=4:T
    boolean(i) = xlim-eta(i);
    if (boolean(i-1)<0);
        yn(i) = 0.7107+1.1022*yn(i-1)-0.2835*yn(i-2)+sigma1*epsilon(i);
    else
        yn(i) = 0.9784+1.5279*yn(i-1)-0.8871*yn(i-2)+sigma2*epsilon(i);
    end
end

apu2    = yn;
ynn     = yn(3:20002,1);
z0boots1 = apu2(3:(Tboots+2));    % two lags
z0boots2 = apu2(2:(Tboots+1));    % one lag
z0boots3 = apu2(1:Tboots);        % original
z0boots  = [z0boots1 z0boots2 z0boots3];
z0      = z0boots;

ar1 = 0.7107 + 1.1022*z0(:,1)-0.2835*z0(:,2);
ar2 = 0.9784 + 0.9784*z0(:,1)-0.8871*z0(:,2);

cdfterms = 0.3163*normcdf((z0boots3-ar1)./sigma1,0,1)+...
           0.6837*normcdf((z0boots3-ar2)./sigma2,0,1);

fterm1s = normpdf((z0boots3-ar1)./ss1,0,1);
fterm1s = fterm1s.*0.3163;
fterm2s = normpdf((z0boots3-ar2)./ss2,0,1);

```

```

fterm2s = fterm2s.*0.6837;
ftots   = fterm1s + fterm2s;
fsim    = log(ftots);
yysim   = [fsim;cdfterms];
apu3    = yyim;
nob     = 114;
rboots  = apu3((nob+1):2*nob,:);

```

```

apu4    = gradient(z0boots);
gboots  = apu4((nob+1):2*nob,:);
sboots  = apu4(1:nob,:);

```

The MATLAB code for computing the autocorrelation test statistic, given a time series of *univariate* quantile residuals, is given below.

```

function [T2] = atest(r,rboots,sboots,gboots,K1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 7.7(b)
% File: atest.m
%
% Converted from GAUSS code written by Leena Kalliovirta.
% Autocorrelation test statistic based on UNIVARIATE quantile
% residuals (QRs).
%
% INPUT:
% r      = vector of quantile residuals (T * 1).
% rboots = simulated vector of QRs using estimated model (Tboots * 1).
% sboots = simulated matrix of scores of QRs using estimated model
%         (Tboots * k) with k the number of columns of sboots.
% gboots = simulated matrix of derivatives of QRs using estimated
%         models (Tboots * k).
% K1     = number of lags used in the autocorrelation test.
%
% OUTPUT:
% T2     = scalar value of the test statistic, follows a Chi2(K1)
%         distribution under H0: no autocorrelation in QRs.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
T       = length(r);
Tboots  = length(rboots);
[~,k]   = size(sboots);

warning('off','all') % Disable all warnings.
%       Replace 'off' by 'on' to enable
% Warning: Matrix "cov" may be close to singular or badly scaled.
%       Results may be inaccurate.

cov     = inv((sboots'*sboots)/Tboots);
G2      = zeros(K1,k);
RR      = zeros(K1,1);
guboots = zeros(Tboots-K1,K1);

```

```

for t1=1:(Tboots-K1);
    help = ((rboots(t1+K1))*rboots(t1+K1-1));
    for i=2:K1
        help = [help;((rboots(t1+K1))*rboots(t1+K1-i))];
    end
    guboots(t1,:) = help;
end

Sigma2 = (horzcat(sboots(K1+1:Tboots,:),guboots))'...
        *(horzcat(sboots(K1+1:Tboots,:),guboots))/(Tboots-K1);

for i1=1:K1
    for j=1:k
        G2(i1,j) = (mean(rboots(1+i1:Tboots).*gboots(1:Tboots-i1,j)+...
            rboots(1:Tboots-i1).*gboots(1+i1:Tboots,j)))';
        RR(i1) = mean(r(1+i1:T).*r(1:T-i1));
    end
end

O2 = [G2*cov,eye(K1)]*Sigma2*[G2*cov,eye(K1)]';
T2 = (T-K1)*(RR')*inv(O2)*RR;    % Autocorrelation test statistic

```

The MATLAB code for the heteroskedasticity test statistic, given a series of *univariate* quantile residuals, is given below.

```

function [T3] = htest(r,rboots,sboots,gboots,K2)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 7.7(b)
% File: htest.m
%
% Converted from GAUSS code written by Leena Kalliovirta.
% Heteroskedasticity test statistic based on UNIVARIATE quantile
% residuals (QRs).
%
% INPUT:
% r      = the vector of quantile residuals (T * 1).
% rboots = simulated vector of QRs using estimated model (Tboots * 1).
% sboots = simulated matrix of scores of QRs using estimated model
%         (Tboots*k) with k the number of columns of sboots.
% gboots = simulated matrix of derivatives of QRs using estimated
%         models (Tboots * k).
% K2     = the number of lags used in the heteroskedasticity test.
%
% OUTPUT:
% T3     = scalar value of the test statistic, follows a Chi2(K2)
%         distribution under H0: no heteroskedasticity in QRs.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
T      = length(r);
Tboots = length(rboots);

```

```

[~,k] = size(sboots);

warning('off','all') % Disable all warnings.
% Replace 'off' by 'on' to enable
% Warning: Matrix "cov" may be close to singular or badly scaled.
% Results may be inaccurate.

cov = inv((sboots'*sboots)/Tboots);
G3 = zeros(K2,k);
RRR = zeros(K2,1);
gubts = zeros(Tboots-K2,K2);

for i2=1:K2
    for j=1:k
G3(i2,j) = 2*mean((rboots(K2+1-i2:Tboots-i2).^2-1).* ...
    rboots(K2+1:Tboots).*gboots(K2+1:Tboots,j)+...
    ((rboots(K2+1:Tboots).^2-1).* ...
    rboots(K2+1-i2:Tboots-i2)).* ...
    gboots(K2+1-i2:Tboots-i2,j))';
RRR(i2) = mean((r(K2+1:T).^2-1).*(r(K2+1-i2:T-i2).^2-1));
gubts(:,i2) = (rboots(K2+1:Tboots).^2-1).* ...
    (rboots(K2+1-i2:Tboots-i2).^2-1);
    end
end

Sigma3 = (horzcat(sboots(K2+1:Tboots,:),gubts))* ...
    (horzcat(sboots(K2+1:Tboots,:),gubts))/(Tboots-K2);
O3 = [G3*cov, eye(K2)]*Sigma3*[G3*cov, eye(K2)]';
T3 = T*(RRR')*inv(O3)*RRR; % Heteroskedasticity test statistic

```

**Remark 1:** Instead of the  $MAR(K; p_1, \dots, p_K)$  model with time invariant  $\pi_i$ 's, a more general model can be obtained by assuming that the mixing proportions are functions of a lagged observation  $Y_{t-d}$  ( $d \geq 1$ ); see, e.g., [Lanne and Saikkonen \(2003\)](#). That is,

$$f_{t-1}(Y_t) = \sum_{i=1}^K \frac{1}{\sigma_{i,t}} f_{\varepsilon} \left( \frac{Y_t - \phi_{i,0} - \phi_{i,1}Y_{t-1} - \dots - \phi_{i,p_i}Y_{t-p_i}}{\sigma_{i,t}} \right) \pi_{i,t-d}, \quad (*)$$

where

$$\pi_{i,t-d} = \mathbb{P}(r_{i-1} - Y_{t-d} \leq \eta_t < r_i - Y_{t-d}) = F_{\eta} \left( \frac{r_i - Y_{t-d}}{\sigma_{\eta}} \right) - F_{\eta} \left( \frac{r_{i-1} - Y_{t-d}}{\sigma_{\eta}} \right),$$

with  $F_{\eta}(\cdot)$  the CDF of  $\{\eta_t, t \in \mathbb{Z}\}$ . Assume that  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ ,  $\{\eta_s\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_{\eta}^2)$ , and the processes  $\{\varepsilon_t\}$  and  $\{\eta_s\}$  are independent of each other for all  $t$  and  $s$ . Furthermore, let

$$\pi_{i,t-d} = \begin{cases} 1 - \Phi((Y_{t-d} - r_1)/\sigma_{\eta}) & i = 1, \\ \Phi((Y_{t-d} - r_{i-1})/\sigma_{\eta}) - \Phi((Y_{t-d} - r_i)/\sigma_{\eta}) & i = 2, \dots, K-1, \\ \Phi((Y_{t-d} - r_{K-1})/\sigma_{\eta}) & i = K, \end{cases}$$



where  $r_0 < r_1 < \dots < r_{K-1} < r_K$  ( $r_0 = -\infty, r_K = \infty$ ). Then the resulting  $\text{MAR}(K; p_1, \dots, p_K, d)$  model can be viewed as an extension of a conventional SETAR model. In particular,

$$Y_t = \sum_{i=1}^K (\phi_{i,0} + \phi_{i,1}Y_{t-1} + \dots + \phi_{i,p_i}Y_{t-p_i} + \sigma_{i,t}\varepsilon_t) I(r_{i-1} + \eta_t \leq Y_{t-d} < r_i + \eta_t), \quad (**)$$

where  $I(r_{i-1} + \eta_t \leq Y_{t-d} < r_i + \eta_t) \equiv \frac{a}{b} I(Y_{t-d} < r_1 + \eta_t)$  for  $i = 1$ , and  $I(r_{i-1} + \eta_t \leq Y_{t-d} < r_i + \eta_t) \equiv I(Y_{t-d} \geq r_{K-1} + \eta_t)$  for  $i = K$ .

Observe that  $\{\eta_t\}$  is independent of  $\{Y_{t-j}\}$  ( $j \geq 1$ ). This implies that we can write  $f(Y_t, \eta_t | Y_{t-j}) = f_\eta(\eta_t) f(Y_t | Y_{t-j})$ . Next, we obtain  $f_{t-1}(Y_t)$  by integration using the indicator function  $I(\cdot)$  in  $(**)$  to divide the range of integration according to  $Y_{t-d}$ . Then we arrive at  $(*)$ . When  $\sigma_\eta^2 = 0$ ,  $(**)$  reduces to a SETAR model. Clearly, in the MAR model of  $(**)$  the threshold parameters are time dependent random variables as opposed to fixed thresholds in the SETAR model. An obvious extension of the MAR model  $(**)$  can be obtained by assuming that  $\sigma_{i,t}^2$  follows a GARCH process.

**Remark 2:** In addition to the model selection results shown in Section 7.5, one may use the R-autots package (Jang et al., 2011) to select the “best” (in the sense of minimizing the MFE) model for the  $\log_{10}$ -transformed Canadian lynx data. Using a tree-based model selection method, and 5-steps ahead predictions, the best model among the class of TAR models is an LSTAR model with two regimes. The best model among the class of combined ARMA–GARCH models is an ARMA(2, 1)–GARCH(1, 1) model. Overall, the LSTAR model is chosen as the final model with the lowest MFE.

7.8 (a) Firstly,

$$\mathbb{E}(\hat{f}_h(x)) = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \frac{1}{h} K \left( \frac{x - X_i}{h} \right) \right] = \int_{\mathbb{R}} K(z) f(x - hz) dz.$$

Observe that the last expression takes the form of a convolution of the kernel and the density function, i.e.  $\mathbb{E}(\hat{f}_h(x)) = (K_h * f)(x)$ . Expand  $f(x - hz)$  in a Taylor series about  $x$ , which is valid as  $h \rightarrow 0$ . For a  $\nu$ th-order kernel, we obtain

$$\begin{aligned} f(x - hz) &= f(x) - hz f^{(1)}(x) + \frac{1}{2} h^2 z^2 f^{(2)}(x) - \frac{1}{3} h^3 z^3 f^{(3)}(x) + \dots \\ &\quad + \frac{1}{\nu!} h^\nu z^\nu f^{(\nu)}(x) + o(h^\nu), \end{aligned}$$

which is uniformly in  $z$ .

Integrating term by term, using  $\int_{\mathbb{R}} K(z) dz = 1$  and the definition  $\mu_j(K) = \int_{\mathbb{R}} z^j K(z) dz$  with  $\mu_j(K) = 0$  ( $j = 1, \dots, \nu - 1$ ), we have

$$\mathbb{E}(\hat{f}_h(x)) = \int_{\mathbb{R}} K(z) f(x - hz) dz = f(x) + \frac{1}{\nu!} h^\nu \mu_\nu(K) f^{(\nu)}(x) + o(h^\nu).$$

So, for higher-order kernels the bias is proportional to  $h^\nu$ , which for  $\nu > 2$  is of lower order than  $h^2$  in the case of second-order kernels.

Similarly, since the kernel estimator is a linear estimator, and  $K(h^{-1}(x - X_i))$  is i.i.d., we have

$$\begin{aligned}\text{Var}(\widehat{f}_h(x)) &= \frac{1}{n} \{(K_h^2 * f)(x) - (K_h * f)^2(x)\} \\ &= \frac{1}{nh} \int_{\mathbb{R}} K^2(z) f(x - hz) dz - \frac{1}{n} \int_{\mathbb{R}} K(z) f(x - hz) dz \\ &= \frac{1}{nh} \int_{\mathbb{R}} K^2(z) \{f(x) + o(1)\} dz - \frac{1}{n} \int_{\mathbb{R}} K(z) \{f(x) + o(1)\} dz \\ &= \frac{1}{nh} R(K) f(x) + o\left(\frac{1}{nh}\right).\end{aligned}$$

(b) Using the results in part (a), we have the pointwise MSE

$$\begin{aligned}\text{MSE}(\widehat{f}_h(x)) &= \text{Bias}(\widehat{f}_h(x; h))^2 + \text{Var}(\widehat{f}_h(x; h)) \\ &\simeq \left(\frac{1}{\nu!} f^{(\nu)}(x) h^\nu \mu_\nu(K)\right)^2 + \frac{1}{nh} R(K) f(x) + o\left(\frac{1}{nh} + h^\nu\right) \\ &= \frac{\mu_\nu^2(K)}{(\nu!)^2} f^{(\nu)}(x)^2 h^{2\nu} + \frac{1}{nh} R(K) f(x) \\ &= \text{AMSE}(\widehat{f}_h(x)).\end{aligned}$$

Clearly, the first term in the third line increases in  $h$  while the second term decreases in  $nh$ : the bias-variance trade-off. Thus, as  $n \rightarrow \infty$ , we must have  $h \rightarrow 0$  and  $nh \rightarrow \infty$ . That is,  $h$  must decrease at a lower rate than  $n$ , a sufficient condition to establish pointwise consistency of the estimator  $\widehat{f}_h(\cdot)$ .

(c) Under the integrability assumptions on  $f(\cdot)$ , and using the results in part (b) the MISE is given by

$$\begin{aligned}\text{MISE}(\widehat{f}_h(x)) &= \int_{\mathbb{R}} (\widehat{f}_h(x) - f(x))^2 dx \\ &= \int_{\mathbb{R}} \left\{ \frac{1}{nh} R(K) f(x) + \frac{\mu_\nu^2(K)}{(\nu!)^2} f^{(\nu)}(x)^2 h^{2\nu} \right\} dx + o\left(\frac{1}{nh} + h^\nu\right) \\ &= \frac{1}{nh} R(K) + \frac{\mu_\nu^2(K)}{(\nu!)^2} h^{2\nu} R(f^{(\nu)}) + o\left(\frac{1}{nh} + h^\nu\right) \\ &= \text{AMISE}(\widehat{f}_h(x)) + o\left(\frac{1}{nh} + h^\nu\right),\end{aligned}$$

where AMISE is the asymptotic MISE, i.e. a large sample approximation to the MISE.

(d) Clearly,

$$\begin{aligned}\frac{\partial}{\partial h} \text{AMISE}(\widehat{f}_h(x)) &= \frac{\partial}{\partial h} \left( \frac{1}{nh} R(K) + \frac{\mu_\nu^2(K)}{(\nu!)^2} h^{2\nu} R(f^{(\nu)}) \right) \\ &= -\frac{R(K)}{nh^2} + 2\nu h^{2\nu-1} \frac{\mu_\nu^2(K)}{(\nu!)^2} R(f^{(\nu)}) = 0.\end{aligned}$$

Its solution leads to the required result, with corresponding AMISE

$$\inf_{h>0} \text{AMISE}(\widehat{f}_h(x)) = (1 + 2\nu) \left( \frac{\mu_\nu^2(K) R(K)^{2\nu} R(f^{(\nu)})}{(\nu!)^2 (2\nu)^{2\nu}} \right)^{1/(2\nu+1)} n^{-2\nu/(2\nu+1)}.$$

Notice that the first term of  $h_{\text{opt}}$  depends on the curvature of  $f(\cdot)$  which is unknown (but can be estimated). The second term can be calculated from results in Table 7.7 of Appendix 7A.

For second-order ( $\nu = 2$ ) kernels the optimal  $h$  is proportional to  $n^{-1/5}$ . For higher-order kernels the optimal rate is slower. Thus, since higher-order kernels have smaller bias, they can afford a larger bandwidth as opposed to second-order kernels. Equivalently, as  $n \rightarrow \infty$ , the best obtainable rate of convergence of the MISE of the kernel estimator is of order  $n^{-2\nu/(2\nu+1)}$ , which for low values of  $\nu$  is less efficient than the typical parametric rate of order  $n^{-1}$  to be expected in function estimation. Clearly, for  $\nu$  large, the convergence rate approaches  $n^{-1}$ .

- (e) From the solution in part (d) we have an expression for the optimal AMISE. So, using (A.5) in Appendix 7A, the efficiency of the kernel function  $K(\cdot)$  relative to the kernel function  $K^*(\cdot)$  is given by

$$\begin{aligned} \text{eff}(K) &= \left( \frac{\text{AMISE}(K)}{\text{AMISE}(K^*)} \right)^{(1+2\nu)/2\nu} \\ &= \left[ \left( \frac{\mu_\nu^2(K) R(K)^{2\nu} R(f^{(\nu)})}{(\nu!)^2 (2\nu)^{2\nu}} \right)^{1/(2\nu+1)} n^{-2\nu/(2\nu+1)} \right. \\ &\quad \left. / \left( \frac{\mu_\nu^2(K^*) R(K^*)^{2\nu} R(f^{(\nu)})}{(\nu!)^2 (2\nu)^{2\nu}} \right)^{1/(2\nu+1)} n^{-2\nu/(2\nu+1)} \right]^{(1+2\nu)/2\nu} \\ &= \left( \frac{\mu_\nu^2(K)}{\mu_\nu^2(K^*)} \right)^{1/2\nu} \frac{R(K)}{R(K^*)}. \end{aligned}$$

- (f) Let  $g(\cdot)$  be any density function taking values on  $\mathbb{R}$ . Set  $g_\sigma(x) = \sigma^{-1}g(x/\sigma)$ . Then  $g_\sigma^{(\nu)}(x) = \sigma^{-1-\nu}g^{(\nu)}(x/\sigma)$ . Given this set-up the roughness of the function  $g(\cdot)$  is given by

$$\begin{aligned} R(g_\sigma^{(\nu)})^{-1/(2\nu+1)} &= \left( \int_{\mathbb{R}} g_\sigma^{(\nu)}(x)^2 dx \right)^{-1/(2\nu+1)} \\ &= \left( \sigma^{-2-2\nu} \int_{\mathbb{R}} g^{(\nu)}(x/\sigma)^2 dx \right)^{-1/(2\nu+1)} \\ &= \left( \sigma^{-(2\nu+1)} \int_{\mathbb{R}} g^{(\nu)}(x)^2 dx \right)^{-1/(2\nu+1)} \\ &= \sigma R(g^{(\nu)})^{-1/(2\nu+1)}. \end{aligned}$$

Furthermore, with  $g_\sigma(\cdot) = \varphi(\cdot)$  the normal density, we have

$$\left( R(\varphi^{(\nu)}) \right)^{-1/(2\nu+1)} = 2 \left( \frac{\sqrt{\pi}\nu!}{(2\nu)!} \right)^{1/(2\nu+1)}.$$

Thus, on combining and replacing  $\sigma$  by  $\hat{\sigma}_X$ , the required result follows.

## Chapter 8

### 8.1 We have

$$\text{Var}\{\hat{\gamma}_Y^{(2,1)}(\ell)\} = \mathbb{E} \left[ \frac{1}{(T-\ell)^2} \sum_{t=\ell+1}^T \sum_{s=\ell+1}^T Y_t^2 Y_{t-\ell} Y_s^2 Y_{s-\ell} \right].$$

Since  $\mathbb{E}(Y_t^2 Y_{t-\ell} Y_s^2 Y_{s-\ell}) = \mu_{4,Y} \mu_{2,Y}$  for  $t = s$ , and  $\mathbb{E}(Y_t^2 Y_{t-\ell} Y_s^2 Y_{s-\ell}) = 0$  for  $t \neq s$ , it follows that  $\text{Var}\{\widehat{\gamma}_Y^{(2,1)}(\ell)\} = \mu_{4,Y} \mu_{2,Y} / (T - \ell)$ . An identical argument also shows that  $\text{Var}\{\widehat{\gamma}_Y^{(1,2)}(\ell)\} = \mu_{4,Y} \mu_{2,Y} / (T - \ell)$ . Next, since  $\mathbb{E}\{\widehat{\gamma}_Y^{(2,1)}(\ell)\} = \mathbb{E}\{\widehat{\gamma}_Y^{(1,2)}(\ell)\} = 0$ , we have

$$\text{Cov}\{\widehat{\gamma}_Y^{(2,1)}(\ell), \widehat{\gamma}_Y^{(1,2)}(\ell)\} = \mathbb{E}\left[\frac{1}{(T - \ell)^2} \sum_{t=\ell+1}^T \sum_{s=\ell+1}^T Y_t^2 Y_{t-\ell} Y_s Y_{s-\ell}^2\right].$$

Since  $\mathbb{E}(Y_t^2 Y_{t-\ell} Y_s Y_{s-\ell}^2) = \mu_{2,Y}^3$  for  $s = t - \ell$ ,  $\mathbb{E}(Y_t^2 Y_{t-\ell} Y_s Y_{s-\ell}^2) = 0$  for  $s \neq t - \ell$ , and since the condition  $s = t - \ell$  occurs  $T - 2\ell$  times in the calculation of  $\text{Cov}\{\widehat{\gamma}_Y^{(2,1)}(\ell), \widehat{\gamma}_Y^{(1,2)}(\ell)\}$ , we have

$$\text{Cov}\{\widehat{\gamma}_Y^{(2,1)}(\ell), \widehat{\gamma}_Y^{(1,2)}(\ell)\} = \mu_{3,Y}^3 \frac{T - 2\ell}{(T - \ell)^2}.$$

On substitution of the above results in

$$\text{Var}\{\widehat{\psi}_Y(\ell)\} = \text{Var}\{\widehat{\gamma}_Y^{(2,1)}(\ell)\} + \text{Var}\{\widehat{\gamma}_Y^{(1,2)}(\ell)\} - 2\text{Cov}\{\widehat{\gamma}_Y^{(2,1)}(\ell), \widehat{\gamma}_Y^{(1,2)}(\ell)\},$$

we obtain the required result (8.5) of the main text. If the underlying distribution is normal,  $\text{Var}\{\widehat{\psi}_Y(\ell)\} = 4\mu_Y^3$ . Recall, in Exercise 5.1 we noted that the standardized bivariate normal distribution is asymptotically normally distributed with mean zero and variance  $3(T - \ell)$ .

**8.2** Clearly,

$$\begin{aligned} \gamma_{\text{TR}}^{(i,j)}(\ell) &= \frac{1}{T_2 - T_1 - \ell} \int_{-T_2}^{-T_1 - \ell} f_{\text{TR}}^i(t) f_{\text{TR}}^j(t + \ell) dt \\ &= \frac{1}{T_2 - T_1 - \ell} \int_{-T_2}^{-T_1 - \ell} f^i(-t) f^j(-(t + \ell)) dt \\ &= -\frac{1}{T_2 - T_1 - \ell} \int_{T_2 - \ell}^{T_1} f^i(\xi + \ell) f^j(\xi) d\xi \\ &= \frac{1}{T_2 - T_1 - \ell} \int_{T_1}^{T_2 - \ell} f^j(\xi) f^i(\xi + \ell) d\xi = \gamma^{(j,i)}(\ell). \end{aligned}$$

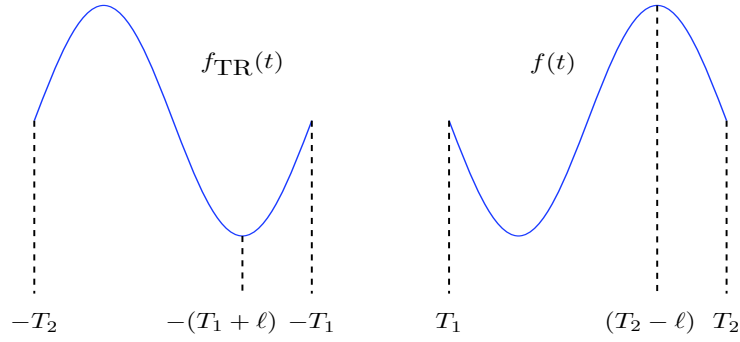
Now, assuming time-reversibility, we have

$$\begin{aligned} \gamma_{\text{TR}}^{(i,j)}(\ell) &= \frac{1}{T_2 - T_1 - \ell} \int_{-T_2}^{-T_1 - \ell} f_{\text{TR}}^i(t) f_{\text{TR}}^j(t + \ell) dt \\ &= \frac{1}{T_2 - T_1 - \ell} \int_{-T_2}^{-T_1 - \ell} f^i(-t) f^j(-(t + \ell)) dt \\ &= \frac{1}{T_2 - T_1 - \ell} \int_{-T_2}^{-T_1 - \ell} f^i(t + \xi) f^j(t + \xi + \ell) dt \\ &= \frac{1}{T_2 - T_1 - \ell} \int_{-T_2 + \xi}^{-T_1 - \xi - \ell} f^i(\eta) f^j(\eta + \ell) d\eta \\ &= \frac{1}{T_2 - T_1 - \ell} \int_{T_1}^{T_2 - \ell} f^i(\eta) f^j(\eta + \ell) d\eta = \gamma^{(i,j)}(\ell), \end{aligned}$$

where  $\eta = t + \xi$ . Inspection of Figure 8.1 shows that  $(T_2 - \xi) = -T_1$ , and  $-(T_1 + \xi) = T_2$ , which accounts for the changes in the limits of the last integration. An alternative way for expressing  $\gamma_{\text{TR}}^{(i,j)}(\ell)$  is  $\gamma^{(i,j)}(-\ell)$ , defined by

$$\begin{aligned}\gamma^{(i,j)}(-\ell) &= \frac{1}{T_2 - T_1 - \ell} \int_{T_1 + \ell}^{T_2} f^i(t) f^j(t - \ell) dt \\ &= -\frac{1}{T_2 - T_1 - \ell} \int_{-T_1 - \ell}^{-T_2} f^i(-\phi) f^j(-\phi - \ell) d\phi \\ &= \frac{1}{T_2 - T_1 - \ell} \int_{-T_2}^{-T_1 - \ell} f_{\text{TR}}^i(\phi) f_{\text{TR}}^j(\phi + \ell) d\phi = \gamma_{\text{TR}}^{(i,j)}(\ell),\end{aligned}$$

where  $\phi = -t$ .



**Figure 8.1:** Sketch of time-reversibility (TR); Exercise 8.2.

**8.3** (a) The numerator can be written as

$$\begin{aligned}\mathbb{E}[\{Y_t - Y_{t-\ell}\}^3] &= \mathbb{E}[Y_t^3 - 3Y_t^2 Y_{t-\ell} + 3Y_t Y_{t-\ell}^2 - Y_{t-\ell}^3] \\ &= 3\mathbb{E}(Y_t^2)^{3/2} \{\rho_Y^{(2,1)}(\ell) - \rho_Y^{(2,1)}(-\ell)\}.\end{aligned}$$

Similarly, the denominator can be written as

$$\mathbb{E}[\{Y_t - Y_{t-\ell}\}^2]^{3/2} = [2\mathbb{E}(Y_t^2)\{1 - \rho_Y^{(1,1)}(\ell)\}]^{3/2}.$$

On combining these results, we get the desired expression.

- (b) The derivative of  $\rho_Y^{(1,1)}(\ell)$  when  $\ell = 0$  is  $-\rho'_{11}(0) \approx -\rho_Y^{(1,1)}(\ell)/\ell$ . Further,  $\rho_Y^{(2,1)}(\ell) - \rho_Y^{(2,1)}(-\ell) \approx -2\ell\rho'_{21}(0)$ . Substituting these results in the expression in part (a) gives the approximation in part (b).
- (c) From part (b) we see that  $\mathbb{E}(X_t^3)/\mathbb{E}(X_t^2)^{3/2}$  is of order  $1/\ell^{1/2}$ . So, for small values of  $|\ell|$  time-irreversibility is most apparent, which seems also plausible on theoretical grounds. Explicit expressions of  $\rho_Y^{(2,1)}(\ell)$  for a particular nonlinear process are usually difficult to derive; see, e.g., Example 4.1 of the main text.

- 8.4 (a) The result that  $BY$  and  $(1-B)Y$  are independent  $\text{Gamma}(k\rho, \beta)$  and  $\text{Gamma}(k(1-\rho), \beta)$  random variables follows from [Patil and Seshadri \(1964\)](#). Then

$$\begin{aligned}\mathbb{E}[\exp\{-(v + Bu)Y\}] &= \mathbb{E}[\exp\{-(u + v)BY - u(1 - B)Y\}] \\ &= \mathbb{E}[\exp\{-(u + v)BY\}]\mathbb{E}[\exp\{-v(1 - B)Y\}],\end{aligned}$$

from which the result in part (a) follows.

- (b) The joint Laplace–Stieltjes transform is given by

$$\begin{aligned}\mathcal{L}_{Y_t, Y_{t-1}}(u, v) &= \mathbb{E}[\exp\{-uY_t - vY_{t-1}\}] \\ &= \mathbb{E}[\exp\{-uB_tY_{t-1} - uG_t - vY_{t-1}\}] \\ &= \mathbb{E}[\exp\{-uG_t\}]\mathbb{E}[\exp\{-(v + Bu)Y_{t-1}\}],\end{aligned}$$

because of stationarity, and where in the last step we used the assumed independence of  $G_t$  and  $Y_{t-1}$ . The Laplace–Stieltjes transform of  $G_t$  is  $\mathcal{L}(u) = \mathbb{E}[\exp\{-uG_t\}] = (\beta/(\beta + u))^k$ . Now, using the result in part (a), we have

$$\begin{aligned}\mathcal{L}_{Y_t, Y_{t-1}}(u, v) &= \left(\frac{\beta}{\beta + u}\right)^{k(1-\rho)} \left(\frac{\beta}{\beta + v + u}\right)^{k\rho} \\ &= \left(\frac{\beta}{\beta + u} \times \frac{\beta}{\beta + v}\right)^{k(1-\rho)} \left(\frac{\beta}{\beta + v + u}\right)^{k\rho}.\end{aligned}$$

- (c) Since  $\mathcal{L}_{Y_t, Y_{t-1}}(u, v)$  is symmetric in  $u$  and  $v$ , the joint distribution of  $Y_t$  and  $Y_{t-1}$  is symmetric. Also, since the joint distribution of any set of  $Y_t$ 's can be obtained from the result in part (b) and the marginal Gamma distribution, the BGAR(1) process is time-reversible, and therefore very limited in practice. Several models for broadening the structure of Gamma time series have been given in the literature; see, e.g., [Bakouch and Ristić \(2009\)](#), and the references therein.

### 8.5 Proofs:

- (a) (Sketch). The process is a second-order Markov chain. By introducing a bivariate state vector  $(Y_{t-1}, Y_t)'$  one arrives at a usual Markov process. The uniqueness of the invariant measure follows from the aperiodicity and irreducibility properties of the Markov chain.
- (b) (Sketch) The conditional distribution of  $Y_t$  given  $Y_{t-1} = u$  and  $Y_{t-2} = v$  differs from that of  $Y_{t-2}$  given  $Y_{t-1} = u$  and  $Y_t = v$ . Hence, there is asymmetry for  $m = 3$ , and by implication for all higher orders  $m$ .
- (c) (Sketch) It can be shown that the conditional distribution of  $Y_{t+\ell}$  given  $Y_t = u$  is uniform on the interval  $[0, 1)$  regardless the value of  $u$ . Since the marginal distribution of  $\{Y_t, t \in \mathbb{Z}\}$  is also  $U(0, 1)$ , the joint distribution of  $(Y_{t-\ell}, Y_t)$  is uniform on  $[0, 1]^2$ , and hence symmetric.

- 8.6 (a) Using the R-np package, the test results can be obtained as follows.

```
#####
# R code: Exercise 8.6
# File: exercise_8-6.r (Load R-np package)
#####
Ats <- ts(A,freq=1) # Replace "A" by the name of a data set
```

**Table 8.1:** Bootstrapped  $p$ -values of the  $\widehat{S}(\cdot)$  test statistics. Blue-typed numbers indicate rejection of the null hypothesis  $\mathbb{H}_0^{(\ell)}$  ( $\ell = 0, 1, 2, 3, 4$ ) of TR at the 5% nominal significance level.

Series	$\mathbb{H}_0^{(0)}$	$\mathbb{H}_0^{(\ell)}$			
		$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$
Unemployment rate <sup>(1)</sup>	0.000	0.020	0.040	0.101	0.000
EEG recordings	0.000	0.000	0.000	0.000	0.000
Magnetic field data	0.051	0.081	0.373	0.495	0.101
ENSO phenomenon	0.010	0.061	0.020	0.000	0.000
Climate change: $\delta^{13}\text{C}$	0.000	0.626	0.879	0.333	0.242
$\delta^{18}\text{O}$	0.020	0.848	0.061	0.000	0.000

<sup>(1)</sup> First differences of original data.

```
# Note: the test function converts time series data into
# numeric data. Alternatively, use: Anum <- as.matrix(A) to
# create a numeric vector.
```

```
npunitest(Ats, -Ats, boot.num=99)
# Note: by default the integral version of the statistic is
# computed (recommended). To speed up computations one may use:
# method=c("summation")
```

Table 8.1, column 2, shows  $p$ -values of the TR test statistic computed under  $\mathbb{H}_0^{(0)}$ . It is evident that, except for the magnetic field data, there is sufficient evidence to reject the null hypothesis of TR for almost all series.

- (b) The null hypothesis  $\mathbb{H}_0^{(1)}$  implies that  $f(Y_t - Y_{t-1})$  is symmetric when a series is TR. Letting  $X_t(1) = Y_t - Y_{t-1}$ , we have from the TR definition that  $f(x) = f(-x) \forall x$ , where  $f(x)$  denotes the pdf of the process  $\{X_t(1), t \in \mathbb{Z}\}$ . Table 8.1, column 3, shows  $p$ -values of the test statistic  $\widehat{S}(1)$  computed under  $\mathbb{H}_0^{(1)}$ . Interestingly, we find similar test results as under  $\mathbb{H}_0^{(0)}$  for the first three series. The conclusion based on the  $p$ -values for the last three series is, however, different.

**Remark:** If time permits, an obvious extension of question (b) is to compute the TR test statistic  $\widehat{S}(\ell)$  for lags  $\ell = 2, 3$ , and 4. The  $p$ -values of the corresponding test statistics are displayed in the last three columns of Table 8.1. Clearly, the test results for the series entitled Unemployment rate, EEG recordings, and ENSO phenomenon, indicate time-irreversibility for almost all lags  $\ell$ . On the other hand, the test statistic  $\widehat{S}(\ell)$  fails to reject the hypothesis of TR for the magnetic field data. The results for the two climate change series are somewhat “mixed” with rejections of the null hypothesis at certain lags. Moreover, comparing the  $p$ -values in Table 8.1 for lags  $\ell = 1, \dots, 4$  with the those reported in Table 8.1 of the main text under the heading “Nonparametric” for  $m = 2, \dots, 5$  (Note:  $\ell \equiv m - 1$ , here), we see that there is even less agreement between the results of the nonparametric test statistics  $S_{h,T}(m)$  and  $\widehat{S}(\ell)$ . Only for the magnetic field data and the  $\delta^{13}\text{O}$  time series, these two test statistics arrive at the

same conclusion, i.e. their  $p$ -values indicate that there is overwhelming evidence not to reject the null hypothesis of TR.

## Chapter 9

- 9.1 [(a)+(b)] Figure 9.1(a) shows plots of the simulated data (black circles), the true regression function (blue solid line), the NW local constant regression estimate with the Epanechnikov kernel (purple solid line), and the LL regression estimate with the Gaussian kernel (black solid line), both with  $h = 0.02$ . Clearly, the regression estimates are very rough and wiggly. In fact, the estimates  $\hat{\mu}_h^{\text{NW}}(\cdot)$  and  $\hat{\mu}_h^{\text{LL}}(\cdot)$  practically reproduce the data resulting in biased estimates of  $\mu(\cdot)$ . The size of the neighborhood around  $x$  is too small so that the kernel estimates are unable to closely resemble the true regression function. In other words, the two kernel regression estimates of  $\mu(\cdot)$  do not shift the chaff (noisy data) from the wheat (true regression function). For the NW estimator, it follows that  $\hat{\mu}_h^{\text{NW}}(x_t) \rightarrow \mathcal{K}_h(0)Y_t/\mathcal{K}_h(0) = Y_t$  as  $h \rightarrow 0$ .

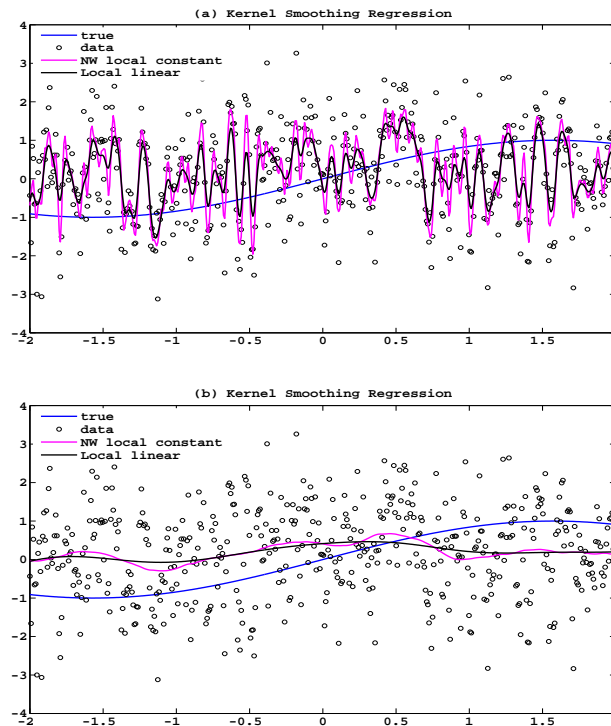
Using the MATLAB-`ksregress` function the classical CV criterion, the Shibata CV, Rice CV, full CV, final prediction CV, Akaike's CV, and GCV all give  $h_{\min} = 0.004$ . Consequently, the resulting regression estimates are even more noisy.

[(c)+(d)] It follows that the value of the bandwidth is given by  $h_{\text{rot}} = 1.1864 \times (4/1500)^{0.2} = 0.3626$ . Figure 9.1(b) shows plots of the data (black circles), the true regression function (blue solid line), the NW local constant regression estimate with the Epanechnikov kernel (purple solid line), and the LL regression estimate with the Gaussian kernel (black solid line) with this plug-in bandwidth. The regression estimates are fairly smooth and, in contrast with the results in Figure 9.1(a), resemble the true regression function more closely. In fact, as  $h$  becomes larger, the weights  $W_t(\cdot)$  will spread over larger neighborhoods of the time series values in the interval  $[-2, 2]$ . An extreme case is when  $h \rightarrow \infty$ , then  $\hat{\mu}_h^{\text{NW}}(x_t) \rightarrow \sum_{t=1}^T \mathcal{K}_h(0)Y_t / \sum_{t=1}^T \mathcal{K}_h(0) = T^{-1} \sum_{t=1}^T Y_t = \bar{Y}$ , suggesting that large bandwidths lead to an oversmoothed curve – the sample mean.

The figures can be reproduced by using the following MATLAB code.

```
function [yhat,dydx,rlr,yinf] = smth(h,x,y,kern)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 9.1
% File: smth.m
%
% Purpose: Kernel smoothing function. Applies LL regression to
%          weighted data.
%
% INPUT:
% x      = abscissa values
% y      = ordinate values
% h      = smoothing parameter
% kern   = kernel smoother
%          kern = 0, uses Gaussian kernel
```





**Figure 9.1:** Data file: Yt-n500-sinus.dat. Kernel smoothing regression with bandwidths (a)  $h = 0.02$ , and (b)  $h = 0.3626$ .

```

%      kern = 1, uses 1/(1+x^2)
%      kern = 2, uses Epanechnikov kernel 0.75*(1-x^2) on [-1<x<1]
%
% OUTPUT:
% yhat = local linear smooth of data y, using abscissa x, bandwidth
%       h and kernel function kern
% dydx = local linear estimate of derivative of y
% rlry = -dydx/(yhat-yinf) where yinf is estimated roughly as that
%       portion of yhat for which dydx is smallish
% yinf (see above)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
zerotol = 1e-14;

difs    = meshgrid(x,x);
difs    = difs' - difs;
kernvals = g(difs/h,kern);
y       = meshgrid(y,y)';
s0      = sum(kernvals);
t0      = sum(kernvals.*y);
temp    = difs.*kernvals;

```

```

s1 = sum(temp);
t1 = sum(temp.*y);
temp = difs.*temp;
s2 = sum(temp);

% Add zerotol to avoid divide by zero on flat series
yhat = (s2.*t0 - s1.*t1)./((s0.*s2 - s1.*s1)+zerotol);

if nargout>1 % Only compute slope if required
    dydx = (s0.*t1 - s1.*t0)./((s0.*s2 - s1.*s1)+zerotol);
end
if nargout>2 % Only compute RLR and estimate yinf if rlr required
    sizex = max(size(x));
    q1 = floor(sizex/5);
    yinf = mean(yhat((4*q1):sizex));
    rlr = -1*dydx./(yhat-yinf);
end

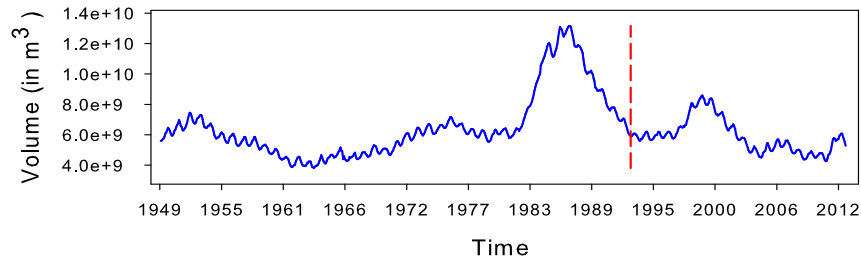
% Next, the subfunction required
% Create Kernel evaluating subfunction

function gx = g(x,kern)
if kern==0
    gx = exp(-(x.*x)/2) ; % Use x.*x not x.^2 for speed
elseif kern==1
    gx = 1./(1+(x.*x));
elseif kern==2
    gx = max(0.75*(1-x.*x),0);
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MAIN: Exercise 9.1
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
y = load('Yt-n500-sinus.dat');
x = linspace(-2,2,500);
h = 0.02;
for i=1:500
    z(i) = sin(x(i));
end
% Local linear regression
[yhat2,dydx,rlr,yinf] = smth(h,x,y,2); % Epanechnikov kernel
[yhat0,dydx,rlr,yinf] = smth(h,x,y,0); % Gauss kernel

plot(x,z);
hold on;
scatter(x,y,'black');
hold on;
plot(x,yhat2,'magenta');
hold on;

```



**Figure 9.2:** Great Salt Lake (GSL) monthly volume time series for the time period October 1949 – December 2012 (756 observations). The red medium dashed vertical line indicates the end of the time period under consideration.

```
plot(x,yhat0,'black')
title('(a) Kernel Smoothing Regression');
legend('true','data','NW local constant','Local linear','location','northwest');
```

- 9.2 (a) Below is a MATLAB function for univariate (unweighted)  $k$ -NN out-of-sample forecasting.

```
function[Ypred,actual,errors] = kNN(data,pset,p,method)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 9.2(a)+(b)
% File: kNN.m
%
% Converted from GAUSS routines (supplementary material) written by
% Ted Jaditz and Leigh A. Riddick.
%
% Reference:
% Jaditz, T. and Riddick, L.A. (2000).
% Time-Series Near-Neighbor Regression.
% Studies in Nonlinear Dynamics and Econometrics,
% 4(1), 35-44.
% DOI: 10.2202/1558-3708.1054.
%
% INPUT:
% data = Vector of T=size(data,1) time series Y_{t}
% pset = Size of the prediction set
%       P = {(Y_{t},X_{t}): N_{f}<t<= T}
% p     = 1,2. Dimension of X_{t}=(Y_{t},Y_{t+1},...,Y_{t+p-1})'
% method = type of "fitting" (window) set F:
%         1 = Fixed, i.e. F={(Y_{t},X_{t}): t<=N_{f}}
%         2 = Rolling, i.e. F_{t}={(Y_{i},X_{i}): t-N_{f}<i<t}
%         3 = Expanding, i.e. F_{t}={(Y_{i},X_{i}): i<t}
%
% OUTPUT:
% Ypred = One-step ahead forecasts for the value of k
```

```

%           that minimizes the CV estimate of MSE
% actual = Actual values
% error = Prediction error (= actual - Ypred)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if p == 1;
    x1 = data(p:size(data,1));
    X = x1;
elseif p == 2;
    x1 = data(p:size(data,1)-1);
    x2 = data(p-1:size(data,1)-2);
    X = [x1,x2];
end
y = data(p+1:size(data,1));
for i = 1:pset; % Start main loop, for each observation in F_{t}
    yi = y(size(y,1)-pset+i);
    xi = X(size(X,1)-pset+i,:);
    if method == 1;
        ypred = y(1:size(y,1)-pset);
        xpred = X(1:size(X,1)-pset,:);
    elseif method == 2;
        ypred = y(i:size(y,1)-pset+i-1);
        xpred = X(i:size(X,1)-pset+i-1,:);
    elseif method == 3;
        ypred = y(1:size(y,1)-pset+i-1);
        xpred = X(1:size(X,1)-pset+i-1,:);
    end
% Create distance table
    distance = zeros(size(xpred,1),1);
    for j = 1:size(xpred,1);
        distance(j) = max(abs(xpred(j,:)'-xi')); % supremum norm
    end
% Sort the data according to distance
    datasort = sortrows([distance,ypred,xpred]);
    sortedDistance = datasort(:,1);
    ypred = datasort(:,2);
    xpred = datasort(:,3:size(datasort,2));
% Loop through sorted data doing regressions
    XX = zeros(size(xpred,2),size(xpred,2));
    Xy = zeros(size(X,2),1);
    msekNN = zeros(size(xpred,1),1);
    for j = 1:size(xpred,1);
% Update matrices for this regression ...
        XX = XX+xpred(j,:)'*xpred(j,:);
        Xy = Xy+xpred(j,:)'*ypred(j);
% Solve for the parameter vector ...
        XXinv = pinv(XX); % Moore-Penrose pseudoinverse
        bij = XXinv*Xy;
        eij = yi-xi*bij;
        if j == 1 ;

```

```

        eij = yi-ypred(1);
    end
    msekNN(j) = msekNN(j)+eij*eij;
end % end of loop over neighbors, j=1:size(xpred,1)
end % end of loop over elements of the P set, i=1:pset
% Identify, rebuild the best near-neighbor
[C,whichOne] = min(msekNN/pset);
fprintf('Min. MSE occurs at %d where MSE = %d.\n',whichOne,C);
actual = y(size(y,1)-pset+1:size(y,1));
% Start rebuilding the k-NN forecasts
for i = 1:pset
    yiRe = y(size(y,1)-pset +i);
    xiRe = X(size(X,1)-pset+i,:);
    if method == 1;
        ypredRe = y(1:size(y,1)-pset);
        xpredRe = X(1:size(X,1)-pset,:);
    elseif method == 2;
        ypredRe = y(i:size(y,1)-pset+i-1);
        xpredRe = X(i:size(X,1)-pset+i-1,:);
    elseif method == 3;
        ypredRe = y(1:size(y,1)-pset+i-1);
        xpredRe = X(1:size(X,1)-pset+i-1,:);
    end
% Create distance table
distanceRe = zeros(size(xpredRe,1),1);
for j = 1:size(xpredRe,1);
    distanceRe(j) = max(abs(xpredRe(j,:)'-xiRe'));
end
% Sort the data according to distance
dataRe = sortrows([distanceRe,ypredRe,xpredRe]);
ypredRe = dataRe(:,2);
xpredRe = dataRe(:,3:size(dataRe,2));
% Build the data set to replicate the regression
XXRe = zeros(size(xpredRe,2),size(xpredRe,2));
XyRe = zeros(size(X,2),1);
for j = 1:whichOne;
    XXRe = XXRe+xpredRe(j,:)'*xpredRe(j,:);
    XyRe = XyRe+xpredRe(j,:)'*ypredRe(j);
end
XXinvRe = pinv(XXRe);
bijRe = XXinvRe*XyRe;
eijRe = yiRe-xiRe*bijRe;
error(i) = eijRe;
end % end of loop over elements of P set, i=1:pset
Ypred = actual-error';
end

```

**Remark:** The supremum norm may be replaced by, for instance, the Euclidean norm. [Jaditz and Riddick \(2000\)](#) discuss the use of a local weighting scheme

in the implementation of the  $k$ -NN regression estimation. Their MC simulation experiments, however, indicate that for a small number of observations the weighted matrices  $\mathbf{X}'_{(i)}\mathbf{X}_{(i)}$  can be much closer to singular (non-invertible) than the unweighted regression matrices.

For a uniform kernel, i.e.  $K_{k_n,i}(\mathbf{x}) = n/k_n$  if  $i \in N(\mathbf{x})$  and 0 otherwise, it is obvious that the specified  $k$ -NN regression estimator is a special case of the estimator  $\hat{\mu}^{k\text{-NN}}(\mathbf{x})$  defined by (9.33) in the main text. Clearly, other kernels or weight sequences can be used; see, e.g., Härdle (1990, Section 3.2).

Another extension of the above algorithm, is to use a discrete resampling (bootstrapping) kernel function which is based on the ordered set of nearest neighbor indices  $N(\mathbf{x})$ ; see, e.g., Lall and Sharma (1996). This will preserve the serial dependence structure of a time series process rather than making explicit assumptions about the DGP under study.

Figure 9.2 shows the GSL monthly volume time series for the time period October 1949 – December 2012 (756 observations). We see clear annual variation and a much longer period with longer periods associated with other dynamics. During the period 1983 – 1987 the lake underwent a modern era rise and threatened the Union Pacific Railroad, Interstate 80, the GSL minerals industry and the Salt Lake City airport. Lall and his co-workers (see, e.g., Abarbanel and Lall, 1996; Sangoyomi et al., 1996) present many studies discussing whether the dynamics are linear or nonlinear; deterministic or stochastic; the dimensionality of the series and its predictability.

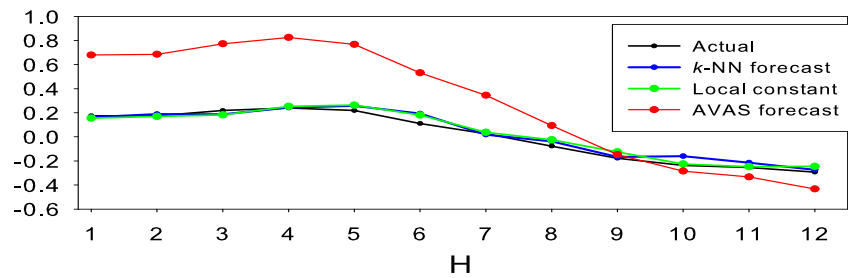
Our focus here is on nonparametric short-term forecasting. Let  $\{Y_t\}_{t=1}^{519}$  denote the time series under consideration, covering the time period October 1949 – December 1992. Some summary statistics are:

$$\begin{aligned}\bar{Y} &= 6.55 \times 10^9, & S_Y &= 2.18 \times 10^9, \\ \hat{\tau}_Y(\text{skewness}) &= 1.45, & \hat{\kappa}_Y(\text{kurtosis}) &= 4.52, \text{ and } \text{JB} = 232.60 \text{ (} p\text{-value} = 0.00\text{)}.\end{aligned}$$

Throughout the analysis, we focus on the standardized GSL time series. The first  $T = 507$  observations are used for parameter estimation and the last 12 observations (pset = 12) for out-of-sample forecasting.

- (b) • Using the  $k$ -NN regression algorithm of part (a) with an expanding window (method = 3) and  $p = 2$ , the minimum MSE value occurs at  $k_n^* = 4$  near neighbors with MSE of  $3.082 \times 10^{-5}$ . The plot of actual versus predicted values in Figure 9.3 confirms the excellent fit. Similar forecast results were obtained on the basis of the fixed and rolling window fitting sets.

As indicated earlier,  $k_n$  plays a role similar to the bandwidth  $h_n$  for kernel smoothers. In particular, the kernel estimator is based on specifying the radius for the local neighborhood and locating the points that fall in it, whereas the  $k$ -NN estimator specifies the number of points and determines the associated radius. If  $k_n = n$ , the  $k$ -NN smoother uses *all* of the observations, and  $\hat{\mu}^{k\text{-NN}}(\mathbf{x})$  then becomes the sample average of the series  $\{Y_t\}_{t=1}^n$  ( $n = T - p$ ). This may reduce the variance of the estimator. However, if  $\mu(\mathbf{x})$  is not actually constant, then  $\hat{\mu}^{k\text{-NN}}(\mathbf{x})$  will be biased for many values of  $\mathbf{x}$ . If, on the other hand,  $k_n = 1$  then the observations are reproduced at  $\mathbf{X}_{(i)}$ , i.e.  $\hat{\mu}^{k\text{-NN}}(\mathbf{x})$  is based on just one observation. In that case  $\mathbf{X}_{(i)}$  will



**Figure 9.3:** Forecasts results for the standardized **GSL** time series ( $T = 519$ ). Twelve one-step ahead forecasts, based on estimates obtained from an expanding training set starting at time index  $t = 505$ .

be close to  $\mathbf{x}$ , and the bias will be relatively small. Unfortunately, this will come at the cost of an increased variance of the  $k$ -NN estimator. So again, there is a bias-variance trade-off problem. In general,  $k_n = \sqrt{n}$  is usually a good choice for model order  $1 \leq p \leq 6$ , particularly for  $T \geq 100$ .

- Figure 9.3 reveals that the local constant regression (Nadaraya–Watson) estimator with a Gaussian kernel and a bandwidth obtained by CV (green solid line) has equally well forecasting performance as the  $k$ -NN regression estimator (blue solid line).

The following R code will generate the forecast results.

```
#####
# R code: Exercise 9.2(b)
# File: Local-constant-regression.r
#
# Local constant regression with a Gaussian product kernel
# and a single bandwidth obtained using cross-validation.
#
# INPUT:
# pset  = size of prediction set (here 12)
# p     = Markov order (here 2).
# OUTPUT:
# Ylc  = (pset * 1) vector of forecasts
#####
library(np)
zGSL  <- scale(GSL519) # standardized data
zGSLt <- ts(zGSL)
nr    <- length(zGSLt)
p     <- 2 # lag length (Markov order)
pset  <- 12
n     <- nr-pset-p
lagmat <- function(x,max.lag) embed(c(rep(NA,max.lag),x),max.lag+1)
xlag  <- lagmat(zGSLt,p)
Ylc   <- matrix(0,nrow=pset,ncol=1)
for (i in 1:pset){
```

```

Xactual <- window(cbind(xlag[3:nr,2],xlag[nc:nr,3]),end=n+i)
Yactual <- window(cbind(xlag[3:nr,1]),end=n+i)
Yc      <- c(Yactual)
bw      <- npregbw(xdat=Xactual,ydat=Yc)
fit.lc  <- npksum(txdat=Xactual,tydat=Yc,bws=bw$bw)$ksum/
         npksum(txdat=Xactual,bws=bw$bw)$ksum
Ylc[i,] <- fit.lc[n+i]
}

```

- For twelve one-step ahead forecasting, using an expanding data set and AVAS estimation, the following R code can be employed.

```

#####
# R code: Exercise 9.2(b)
# File: AVAS-estimation-forecasting.r
#
# INPUT:
# pset = size of prediction set (here 12)
# p    = Markov order (here 2)
#
# OUTPUT:
# Ypred = (pset * 1) vector of one-step ahead forecasts
#####
library(acepack)
zGSL <- scale(GSL519) # standardized data
zGSLt <- ts(zGSL)
nr    <- length(zGSLt)
p     <- 2 # lag length (Markov order)
pset  <- 12
n     <- nr-pset-p
lagmat <- function(x,max.lag)embed(c(rep(NA,max.lag),x),max.lag+1)
xlag  <- lagmat(zGSLt,p)

Ycmat1 <- matrix(0,nrow=pset,ncol=1)
Ypred  <- matrix(0,nrow=pset,ncol=1)
for (i in 1:pset){
  Xactual1 <- window(xlag[3:nr,2],end=n+i)
  Yactual1 <- window(xlag[3:nr,1],end=n+i)
  avasout1 <- avas(Xactual1,Yactual1)
  Ypred1    <- avasout1$ty[n+i]
  Ycmat1[i,] <- Ypred1 # 1-step ahead forecast, 1st lag

  Xactual2 <- window(xlag[3:nr,3],end=n+i)
  Yactual2 <- window(xlag[3:nr,1],end=n+i)
  avasout2 <- avas(Xactual2,Yactual2)
  Ypred2    <- avasout2$ty[n+i]
  Ypred[i,] <- Ypred2+Ypred1 # 1-step ahead forecast,
                             # 1st and 2nd lag
}

```

Figure 9.3 shows that the AVAS forecasting method (red solid line) performs



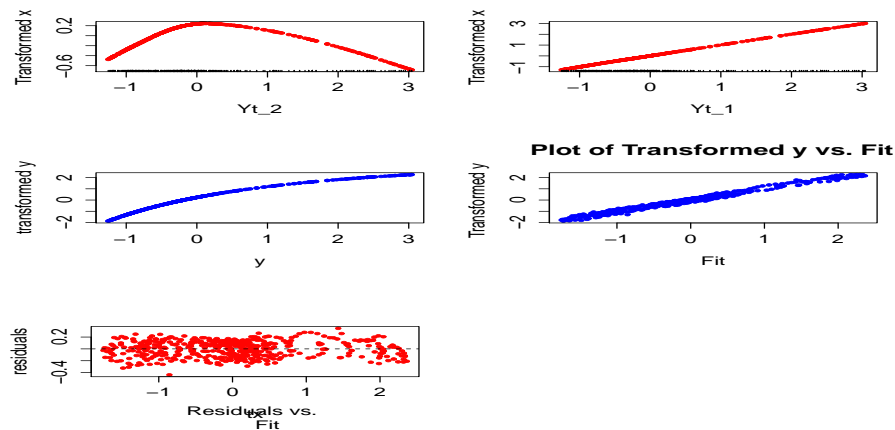


Figure 9.4: Plots of AVAS model fit for the standardized **GSL** time series;  $T = 519$ .

poorly as compared to the forecast performance of the  $k$ -NN method. For  $T = 519$ , Figure 9.4 displays details of the AVAS fit ( $R^2 = 0.983$ ) using 3 iterations. Clearly, AVAS indicates that  $Y_{t-2}$  is nonlinear (a quadratic function) while  $Y_{t-1}$  seems to be best transformed by a linear function. So, using these plots one can choose “parametric” forms of the transformations.

The following R code, written by Brant Deppa, can be used to reproduce the results in Figure 9.4.

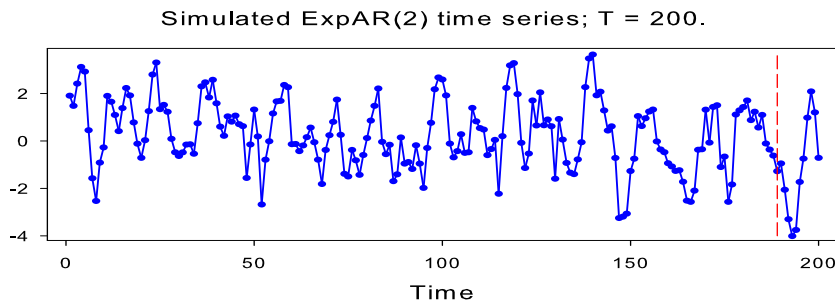
```
#####
# R code: Exercise 9.2(b)
# File: MAVASplot.r
#
# Plots of AVAS transformation results.
#
# INPUT:
# xmat = Matrix (e.g. Xactual) containing the independent
#       variables. Note: p > 1 independent variables
# y     = Vector (e.g. Yactual) with the response variable
# x     = List (e.g. avasout) containing 10 components
#####
mavasplot = function(xmat,y,x,nrow=2,ncol=2){
  par(mfrow=c(nrow,ncol),ask=T)
  for (i in 1:ncol(xmat)){
    plot(xmat[,i],x$tx[,i],xlab=dimnames(xmat)[[2]][i],
         ylab="Transformed x",pch=20,col="red")
    rug(xmat[,i])
  }
  plot(y,x$ty,xlab="y",ylab="transformed y",cex=0.7,
       pch=20,col="blue")
  fit <- rep(0,length(y))
}
```

```

for (k in 1:ncol(xmat)){
  fit <- fit+x$tx[,k]
}
plot(fit, x$ty,main="Plot of Transformed y vs. Fit",
     xlab="Fit",ylab="Transformed y ",pch=20,col="blue")
r <- x$ty-fit
plot(fit,r,xlab="tx",ylab="residuals",sub="Residuals vs.
Fit",pch=20,col="red")
abline(h=0,col="black",lty=2)
par(mfrow=c(1,1),ask=F)
invisible()
}

```

The MSFEs, i.e.  $\sum_{H=1}^{12} (Y_{t+H} - \hat{Y}_{t+1|t})^2 / 12$ , of the three forecasting methods are, respectively, 0.0016 ( $k$ -NN), 0.0014 (Local constant), and 0.1505 (AVAS). So, both  $k$ -NN regression and the local constant regression produce forecasts with the smallest MSFEs for this particular data set, and particular time period.



**Figure 9.5:** A particular realization of a simulated ExpAR(2) process;  $T = 200$ , ExpAR2.dat time series.

- 9.3** (a) Figure 9.5 shows a time series realization of the ExpAR(2) process  $Y_t = \{0.9 + 0.1 \exp(-Y_{t-1}^2)\}Y_{t-1} - \{0.2 + 0.1 \exp(-Y_{t-1}^2)\}Y_{t-2} + \varepsilon_t$ , where  $\{\varepsilon_t\} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ . The red medium dashed vertical line indicates the start of the out-of-sample forecast period. The plot suggests that the series is stationary with a constant mean and a constant variance. A sufficient condition for strict stationarity for this process is that the roots of the associated characteristic equation  $z^2 - c_1z - c_2 = 0$  are inside the unit circle, with  $c_1 = \max\{|0.9|, |0.9 + 0.1|\} = 1$ , and  $c_2 = \max\{|-0.2|, |-0.2 + 0.1|\} = 0.2$ . The solutions are  $z_1 = -0.7236$  and  $z_2 = -0.2764$ . Thus, the process  $\{Y_t, t \in \mathbb{Z}\}$  is stationary.
- (b) Using the function `regFit` in the R-fRegression package with the option `use="ppr"`, the output is as follows.

```

Title:
  Projection Pursuit Regression
Formula:
  y ~ x

```

```

Family:
  gaussian identity
Model Parameters:
-- Projection Direction Vectors --
      term 1      term 2
xV1  0.9390225 -0.99524382
xV2 -0.3438558  0.09741531
-- Coefficients of Ridge Terms --
      term 1      term 2
1.1040792  0.3844721
Residual Variance:
  0.7787314
Goodness of fit:
  2 terms
146.6212

```

Here, the roman type  $x$  denotes the two-period lagged data matrix of size  $(187 \times 2)$ , and  $y$  denotes the simulated time series of length  $T = (189 - 2) = 187$  (deleting the first two observations since  $p = 2$ ). Adopting the notation introduced in the main text, we have

$$\begin{aligned}\hat{\beta}_1 &= 1.104079, & \hat{\alpha}_1 &= (\hat{\alpha}_{1,1}, \hat{\alpha}_{1,2})' = (0.9390225, -0.3438558)' \\ \hat{\beta}_2 &= 0.3844721, & \hat{\alpha}_2 &= (\hat{\alpha}_{2,1}, \hat{\alpha}_{2,2})' = (-0.99524382, 0.09741531)',\end{aligned}$$

with the constraint  $\sum_{j=1}^2 \hat{\alpha}_{i,j}^2 = 1$ .

- (c) Using the function `regFit` in the `R-fRegression` package with the option `use="nnet"`, the output is as follows.

```

Title:
  Feedforward Neural Network Modeling
Formula:
  y ~ x
Family:
  gaussian identity
Model Parameters:
  a 2-2-1 network with 9 weights
  options were - linear output units
[1]  31.2494159 -22.9550428 -16.9108426 -1.6019634  0.2459704
     -0.1002664 -3.5628607 -0.4306908
[9]  23.2361972
Residual Variance:
  0.9665153

```

- (d) Table 9.1 contains the one-step ahead forecasts at times  $t = 190, \dots, 200$  for the PPR model, ANN model, and a linear regression model (benchmark) with two lagged predictors. Given these values, the out-of-sample one-step ahead mean square forecast errors (MSFEs) are given by 1.0653 (PPR), 1.2440 (ANN), and 1.3059 (linear model). So, the PPR model provides better forecasting results than the other two models for this particular time series realization. The out-of-sample one-step ahead values of the MAFE seems to support this tentative

conclusion with values 0.8698 (PPR), 0.9278 (ANN), and 0.9878 (linear model) respectively. Of course, these outcomes are based on a single replication. Further, we see that the residual variance of the PPR model (0.7787314) is considerably smaller than the residual variance of the ANN model (0.9665153). So, the PPR model has the best in-sample fit.

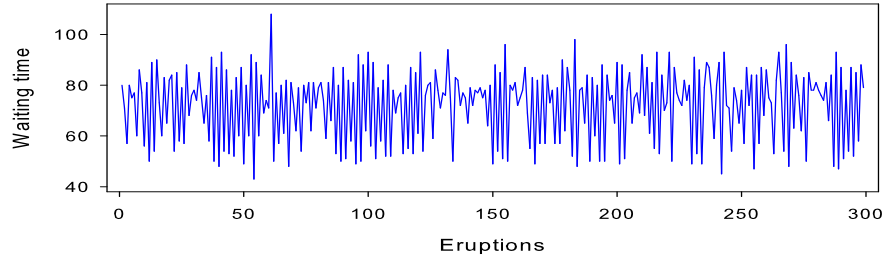
**Table 9.1:** One-step ahead forecasts at times  $t = 190, \dots, 200$ ; ExpAR2.dat time series.

	190	191	192	193	194	195	196	197	198	199	200
$Y_t$	-0.953	-2.062	-3.307	-4.017	-3.757	-1.736	-0.751	0.975	2.082	1.199	-0.719
PPR	-0.909	-1.239	-1.567	-2.255	-4.230	-2.370	-0.707	-0.149	1.096	1.878	0.540
ANN	-0.855	-0.451	-1.462	-2.062	-4.046	-2.506	-0.626	0.200	1.364	2.153	0.349
Linear	-0.863	-0.425	-1.503	-2.359	-2.762	-2.340	-0.617	-0.199	1.122	1.677	0.613

- 9.4** (a) Figure 9.6 shows a plot of the series. Figure 9.7 contains plots of the function  $f_2(p)$  ( $p = 1, \dots, 10$ ), defined in (9.10), for the conditional mean (black solid line), the conditional median (red medium dashed line), and the conditional mode (blue dotted line), based on the set of observations  $\{Y_t\}_{t=230}^{289}$ . As can be seen in all cases  $f_2(p)$  seems to stabilize, starting from a Markov coefficient  $p = 1$ . For the conditional mean and the conditional median, however, the two plotted lines are almost indistinguishable. Based on  $\arg \min_p f_2(p)$ , we have  $p = 1$  for the conditional mean, and  $p = 3$  for both the conditional median and the conditional mode. Note that the values of  $f_2(p)$  are much larger for the conditional mode than for the other two predictors, indicating that for this set of observations the values of the conditional mode are far off from the observed values. The minimum values of the functions  $f_1(p)$  and  $f_3(p)$  confirm the choice  $p = 3$  for the conditional median and the conditional mode, and the choice  $p = 1$  for the conditional mean.
- (b) The out-of-sample forecast results for each method are displayed in Table 9.2. Recall, the MSFE and RMAFE forecasting measures are respectively defined as

**Table 9.2:** Old Faithful geyser data set. Forecast results for the conditional mean, the conditional median, the conditional mode, and a naive forecasting rule; based on observations up to and including eruption number  $t = 289$ , and forecast horizon  $H = 1, \dots, 10$ .

	Eruption number ( $t$ )										MSFE	RMAFE
	290	291	292	293	294	295	296	297	298	299		
Waiting time ( $Y_t$ )	87	51	78	54	87	52	85	58	88	79		
Duration	1.57	4.41	1.56	4.25	2.08	4.05	2.04	4	4	2		
$Y_{289+H 289}^{\text{Mean}}$	85.36	61.97	81.01	65.05	77.59	67.77	77.88	66.96	76.27	70.18	93.80	13.72
$Y_{289+H 289}^{\text{Mdn}}$	85.18	58.30	82.31	63.32	80.16	72.27	80.16	69.05	78.01	72.27	91.04	13.07
$Y_{289+H 289}^{\text{Mode}}$	84.82	53.64	83.03	53.64	81.95	76.93	80.52	78.73	80.16	80.88	119.89	12.08
Naive forecast	55	80	55	80	55	80	55	80	80	55	690.20	37.46



**Figure 9.6:** Waiting time between the starts of successive eruptions of the Old Faithful geyser, Wyoming, August 1, 1985 – August 15, 1985;  $T = 299$ .

follows:

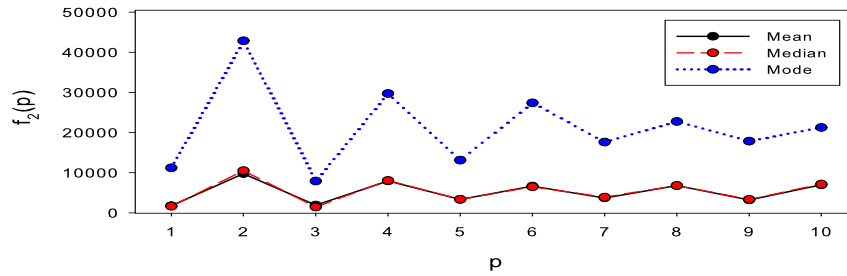
$$\text{MSFE} = \frac{1}{H_{\max}} \sum_{H=1}^{H_{\max}} (\widehat{Y}_{T+H|T}^{(\cdot)} - Y_{T+H})^2,$$

$$\text{RMAFE} = \frac{1}{H_{\max}} \sum_{H=1}^{H_{\max}} \frac{|\widehat{Y}_{T+H|T}^{(\cdot)} - Y_{T+H}|}{|Y_{T+H}|} \times 100,$$

where  $\widehat{Y}_{T+H|T}^{(\cdot)}$  denotes the conditional mean, the conditional median, or the conditional mode based on observations up till and including time  $t = T$ .

Clearly, the conditional mode performs the best as judged by the RMAFE. Of course, this result should be interpreted with care since it heavily depends on the last observed value. So, we cannot generalize our conclusion to apply to all starting values and forecast horizons. In fact, it is easy to verify that the conditional mode still outperforms the conditional mean and the conditional median when 10-steps ahead forecasts are based on observations up to and including eruption number  $t = 288$ .

- (c) Consider the following two methods for constructing a forecast interval (FI).
- i. *Conditional percentile interval* (see also Section 10.3.1 of the main text): Let  $\widehat{\xi}_q(\mathbf{x}) = \inf\{z : \widehat{F}(z|\mathbf{x}) \geq q\}$  be the nonparametric estimate of  $\xi_q(\mathbf{x})$  as defined by (9.6), where  $\widehat{F}(\cdot|\mathbf{x})$  is a nonparametric estimator of the conditional distribution function. From Gannoun (1990) it is known that  $\widehat{\xi}_q(\mathbf{x})$  converges uniformly and pointwise (on a compact set) to  $\xi_q(\mathbf{x})$ . So for each  $H$ , it is easy to compute a  $100\%(1 - \alpha)$  FI with  $\alpha \in [0, 1]$  (Note the change in notation, i.e.  $\alpha \equiv q$ ). This leads to the same interval for the three nonparametric forecasting methods.
  - ii. *Empirical method*: Assume that future forecast errors have the same distribution function as the distribution function of the past forecast errors, then FIs for future forecasts can be based on the EDF of the past forecast errors. In particular, for  $H \in \{1, \dots, H_{\max}\}$  and  $j \in \{T - r, \dots, T\}$ , with  $r$  a preselected large number, calculate the relative error  $\text{ER}_j(H) = |Y_{j+H}^{(\cdot)} - Y_{j+H}| / |Y_{j+H}|$ . Then, for each  $H$ , classify the  $r$   $\text{ER}_j(H)$ 's by increasing order and take the  $(1 - \alpha)$  one. This value is the  $(1 - \alpha)$ th empirical



**Figure 9.7:** Old Faithful geyser data set. The estimated function  $f_2(p)$  ( $p = 1, \dots, 10$ ) for the conditional mean, conditional median, and conditional mode: based on the time series  $\{Y_t\}_{t=230}^{289}$ .

conditional quantile, denoted by  $\tilde{\xi}_{(1-\alpha)}(H)$ . It is a natural estimator of  $\xi_{(1-\alpha)}(H)$  defined by

$$\mathbb{P}\left(\frac{|\hat{Y}_{j+H|j}^{(\cdot)} - Y_{j+H}|}{|Y_{j+H|j}^{(\cdot)}|} \leq \xi_{(1-\alpha)}(H)\right) = 1 - \alpha.$$

Obviously, for large values of  $r$  the estimates  $\tilde{\xi}_{(1-\alpha)}(H)$  become more reliable.

- (d) The results of the naive forecast rule are listed in the last row of Table 9.2. The MSFE and RMAFE results for this rule are dramatically larger than those based on the other three methods. Hence, the effectiveness of the naive forecasting rule is questionable.

The following MATLAB functions yield the forecasting results reported in rows 3 – 5 of Table 9.2.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 9.4
% File: cond_mean.m
% Coded by: Dawit Zerom
%
% Computing conditional MEAN forecasts.
%
% Reference:
% De Gooijer, J.G. and Zerom, D. (2000).
% Kernel based multi-step-ahead prediction of the U.S. short-term
% interest rate. Journal of Forecasting, 19(4), 335-353.
% DOI: 10.1002/1099-131x(200007)19:4%3C335::aid-for777%3E3.3.co;2-v
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
load    = waiting.dat;
datan   = waiting;
[nr,nc] = size(datan);
data1   = datan;
k       = 1;    % Maximum Markov order
mx      = 10;   % Maximum forecast horizon

```

```

x1      = data1(k:nr);
data    = x1;
z       = data;
n2      = 289;      % Start data set Geyser-waiting time
n3      = 289;      % End data set
df      = n3-n2+1;
n       = n2;
while n<=n3
    m = 1;
    while m<= mx
        n0 = n-m;
        y  = z(m+1:n,1);
        x0 = z(1:n0);
        sd = std(data1(1:n));
        hop = (sd)*((n)^(-1/(k+4))); % Optimal bandwidth
        sum1 = 0;
        sum2 = 0;
        for t1 = 1:n0
            vect1 = (z(n)-x0(t1))/hop;
            xx    = vect1*vect1';
            % Gaussian kernel
            sum1 = sum1+(exp(-0.5*xx))*((2*pi)^(-k/2));
            vect2 = (y(t1)*(exp(-0.5*xx))*((2*pi)^(-k/2)));
            sum2 = sum2+vect2;
        end
        if(sum1 == 0)
            sum1 = 1.0e-300;
        end
        ztmean(n,m) = sum2/sum1; % mth-step ahead forecast (until mx)
% made at time t=n=n2
        dif(n,m) = data1(n+m,1)- ztmean(n,m);
        m        = m+1;
    end
    n = n+1;
end
dif_mean    = dif(n2:n3,1:mx);          % Resize matrix
MSFE_mean  = sum(dif_mean*dif_mean')/mx;
RMAFE_mean = (100/mx)*sum(abs(dif_mean)'./abs(data1(n2+1:nr,1)));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 9.4
% File: cond_median.m
% Coded by: Dawit Zerom
%
% Computing conditional MEDIAN forecasts.
%
% Reference:
% De Gooijer, J.G. and Zerom, D. (2000).
% Kernel based multi-step-ahead prediction of the U.S.

```

```

% short-term interest rate.
% Journal of Forecasting, 19(4), 335-353.
% DOI: 10.1002/1099-131x(200007)19:4%3C335::aid-for777%3E3.3.co;2-v
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
load    = waiting.dat;
datan   = waiting;
[nr,nc] = size(datan);
data1   = datan;
k       = 1;    % max Markov order
mx      = 10;   % maximum forecast horizon
x1      = data1(k:nr);
data    = x1;
z       = data;
n2      = 289;  % Start data set Geyser-waiting time
n3      = 289;
df      = n3-n2+1;
n       = n2;
while n<=n3
    m = 1;
    while m<=mx;
        n0    = n-m;
        y     = z(m+1:n,1);
        x0    = z(1:n0);
        sd    = std(data1(1:n));
        hop   = (sd)*((n)^(-1/(k+4))); % optimal bandwidth
        a     = min(y)-3*hop;
        b     = max(y)+3*hop;
        step  = (b-a)/256;
        for l = 1:256;
            sum1 = 0;
            sum2 = 0;
            t1   = 1;
            while t1 <= n0;
                vect1 = (z(n)-x0(t1))/hop;
                xx    = vect1*vect1';
                sum1  = sum1+(exp(-0.5*xx))*((2*pi)^(-k/2));
                if y(t1) <= (a+l*step)
                    vect2 = (z(n)-x0(t1))/hop;
                    xx2   = vect2*vect2';
                    sum2  = sum2+(exp(-0.5*xx2))*((2*pi)^(-k/2));
                end
                t1 = t1+1;
            end
            if(sum1 ==0 )
                sum1 = 1.0e-300;
            end
            if (sum2/sum1) > 0.5 % Median
                break
            end
        end
    end
end

```



```

        end
        ztmedian(n,m) = a+l*step;
        dif(n,m)      = data1(n+m,1)-ztmedian(n,m);
        m             = m+1;
    end
    n = n+1;
end
dif_median = dif(n2:n3,1:mx);           % Resize matrix
MSFE_median = sum(dif_median*dif_median')/mx;
RMAFE_median = (100/mx)*sum(abs(dif_median)'./abs(data1(n2+1:nr,1)));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 9.4
% File: cond_mode.m
% Coded by: Dawit Zerom
%
% Computing conditional MODE forecasts.
%
% Reference:
% De Gooijer, J.G. and Zerom, D. (2000).
% Kernel based multi-step-ahead prediction of the U.S. short-term
% interest rate. Journal of Forecasting, 19(4), 335-353.
% DOI: 10.1002/1099-131x(200007)19:4<3C335::aid-for777>3E3.3.co;2-v
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
load    = waiting.dat;
datan   = waiting;
[nr,nc] = size(datan);
data1   = datan;
k       = 1;    % Maximum Markov order
mx      = 10;   % Maximum forecast horizon
x1      = data1(k:nr);
data    = x1;
z       = data;
n2      = 289;  % Start data set Geyser-waiting time
n3      = 289;
df      = n3-n2+1;
n       = n2;
while n <= n3;
    m = 1;
    while m <= mx;
        n0 = n-m;
        y  = z(m+1:n,1);
        x0 = z(1:n0);
        sd = std(data1(1:n));
        hop = (sd)*((n)^(-1/(k+4)));
        a  = min(y)-3*hop;
        b  = max(y)+3*hop;
        step = (b-a)/256;
        sum1 = 0;
    end
end

```

```

sum2 = 0;
i = 1;
while i <= n0;
    vect1 = (z(n)-x0(i))/hop;
    xx    = vect1*vect1';
    vect2 = (a-y(i))/hop;
    xx2   = vect2*vect2';
    sum1  = sum1+...
    (exp(-0.5*xx2))*((2*pi)^(-k/2))*(exp(-0.5*xx))*((2*pi)^(-k/2));
    sum2  = sum2+(exp(-0.5*xx))*((2*pi)^(-k/2));
    i     = i+1;
end
if(sum2==0)
    sum2 = 1.0e-300;
end
summax = (sum1/sum2);
ymax  = a;
l      = 1;
while l <= 256;
    y1   = a+l*step;
    sum1 = 0;
    sum2 = 0;
    i    = 1;
    while i <= n0;
        vect1 = (z(n)-x0(i))/hop;
        vect2 = (y1-y(i))/hop;
        xx    = vect1*vect1';
        xx2   = vect2*vect2';
        sum1  = sum1+...
        (exp(-0.5*xx))*((2*pi)^(-k/2))*(exp(-0.5*xx2))*((2*pi)^(-k/2));
        sum2  = sum2+(exp(-0.5*xx))*((2*pi)^(-k/2));
        i     = i+1;
    end
    if(sum2==0)
        sum2 = 1.0e-300;
    end
    sumy = (sum1/sum2);
    if(sumy > summax);
        summax = sumy;
        ymax  = y1;
    end
    l = l+1;
end
ztmode(n,m) = ymax;
dif(n,m)    = data1(n+m,1)-ztmode(n,m);
m           = m+1;
end
n = n+1;
end

```

**Table 9.3:** Comparison of MSFEs for  $H = 1, 10$  and  $20$  steps ahead forecasts made with glmboost, gamboost, MARS, and VAR (unrestricted) for the hourly river flow – rainfall data set; based on 15 forecast results. For each  $H$ , blue-typed numbers indicate the lowest MSFE.

$H$	glmboost	gamboost	MARS	VAR
1	99.543	149.834	286.340	120.518
10	1890.600	1830.708	1208.053	1463.838
20	190.495	159.390	75.380	193.327

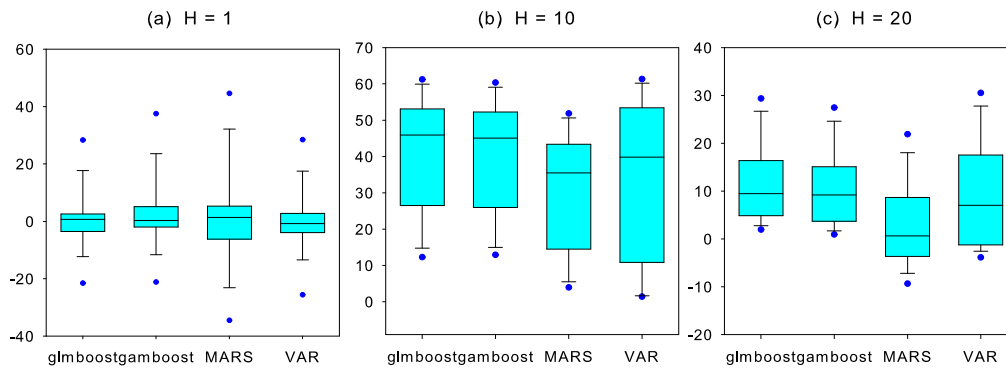
```
dif_mode = dif(n2:n3,1:mx); % Resize matrix
MSFE_mode = sum(dif_mode*dif_mode')/mx;
RMAFE_mode = (100/mx)*sum(abs(dif_mode)' ./abs(data1(n2+1:nr,1)));
```

- 9.5 (a) Table 9.3 shows the MSFEs of the four forecasting methods. Apparently, for short-term ( $H = 1$ ) forecasting, glmboost outperforms the other three forecasting methods. MARS performs comparably good for  $H = 10$  and  $H = 20$ . The forecasting performance of VAR (unrestricted) is about the same as glmboost and gamboost, but is not superior to MARS for 10 and 20 time periods ahead. Note, however, that the inclusion of one exogenous variable in the model means that boosting and MARS deal with 24 covariates (12 lags for each covariate).

To allow for a more closer comparison between the four forecasting methods, Figure 9.8 shows boxplots of the average squared forecast errors. For  $H = 1$ , MARS has the largest outliers at time points  $t = 373$  and  $t = 374$ . Also, glmboost and gamboost have outlying forecast errors at these particular time points. We see that MARS is superior to the other forecasting methods for  $H = 10$  and  $H = 20$ , both in terms of lowest mean and lowest median forecast errors.

- (b) For  $H = 1$ , the MSFEs of the two univariate boosting methods are larger than those given in Table 9.3 with values 165.285 (glmboost) and 221.157 (gamboost). For MARS the MSFE equals 190.586 which is much smaller than the MSFE value obtained with this forecasting method with lagged values of  $\{(Y_t, X_t)\}$ . For  $H = 10$  and  $H = 20$ , the three univariate forecasting perform equally well as their bivariate counterparts. So, the variable rainfall as an additional covariate does not necessarily improve the forecasting quality.

```
#####
# R code: Exercise 9.5
# File: boost.r
#####
rm(list = ls(all = T))
library(GAMBoost)
library(mboost)
library(vars)
library(mda)
library(zoo)
source("c:/.../boost_nlar//analysis/functions/helpers-forecasting.R")
```



**Figure 9.8:** Boxplots of the averaged squared forecast errors, based on 15 forecasts, for  $H=1, 10,$  and  $20$  of the river flow – rainfall data set.

```

source("c:/.../boost_nlar/./analysis/functions/helpers-general.R")
source("c:/.../boost_nlar/./analysis/functions/f-VAR.R")

Flowtot <- read.table("c:/.../flow.dat", quote="\"")
FLtot401 <- ts(Flowtot,freq=1)
Raintot <- read.table("c:/.../rain.dat", quote="\"")
RNtot401 <- ts(Raintot,freq=1)

##### Defaults #####
setup <- list(ychar = "FLtot401",
             xchar = c("FLtot401","RNtot401"),
             iniend = c(366,1),
             maxLag = 12,
             hors = c(1, 10, 20))

lsstp <- list(setup)

##### Parameter sets #####
parset.glmboost <- list(method = "glmboost",
                       mstop = 500,
                       dummy = FALSE)
parset.gamboost <- list(method = "gamboost",
                       df = 3.5,
                       knots = 20,
                       mstop = 500,
                       dummy = FALSE)
parset.mars <- list(method = "mars",
                   penalty = "log")
parset.VAR <- list(method = "VAR",
                  maxLag = setup$maxLag,
                  restrict = FALSE,
                  hor = NULL,

```

```

iter      = FALSE)

##### mboost with input variable #####
lsglmFLtot2 <- lapply(lsstp,function(z) dir.paradigm(z,parset.glmboost))
lsgamFLtot2 <- lapply(lsstp,function(z) dir.paradigm(z,parset.gamboost))
lsmarsFLtot2 <- lapply(lsstp,function(z) dir.paradigm(z,parset.mars))
lsvarFLtot2 <- lapply(lsstp,function(z) dir.paradigm(z,parset.VAR))

```

## Chapter 10

10.1 (a) Clearly,  $f_3(x) = \int_{a+\omega x^{1/2}}^{b+\omega x^{1/2}} f_2(y)dy$ , where simple calculations gives

$$f_2(x) = \int_{a+\omega x^{1/2}}^{b+\omega x^{1/2}} f_1(y)dy = \frac{a+b}{2} + \frac{2\omega}{3(b-a)} \left[ (b+\omega x^{1/2})^{3/2} - (a+\omega x^{1/2})^{3/2} \right].$$

with  $f_1(x) = (a+b)/2 + \omega x^{1/2}$ . Using Mathematica<sup>4</sup> (a symbolic mathematical computation program), we obtain

$$\begin{aligned} \int_{b+\omega x^{1/2}}^{b+\omega x^{1/2}} (a+wy^{1/2})^{3/2} dy &= \frac{4}{35w^2} \sqrt{a+w\sqrt{b+w\sqrt{x}}} \left( -2a^3 \right. \\ &\quad \left. + a^2w\sqrt{b+w\sqrt{x}} + 8aw^2(b+w\sqrt{x}) + 5w^3(b+w\sqrt{x})^{3/2} \right) \\ &\quad - \frac{4}{35w^2} \sqrt{a+w\sqrt{a+w\sqrt{x}}} \left( -2a^3 + a^2w(8w+\sqrt{a+w\sqrt{x}}) \right. \\ &\quad \left. + aw^3(5\sqrt{a+w\sqrt{x}}+8\sqrt{x}) + 5w^4\sqrt{a+w\sqrt{x}\sqrt{x}} \right), \end{aligned}$$

and

$$\begin{aligned} \int_{b+\omega x^{1/2}}^{b+\omega x^{1/2}} (b+wy^{1/2})^{3/2} dy &= -\frac{4}{35w^2} \sqrt{b+w\sqrt{a+w\sqrt{x}}} \left( -2b^3 \right. \\ &\quad \left. + b^2w\sqrt{a+w\sqrt{x}} + 8bw^2(a+w\sqrt{x}) + 5w^3(a+w\sqrt{x})^{3/2} \right) \\ &\quad + \frac{4}{35w^2} \sqrt{b+w\sqrt{b+w\sqrt{x}}} \left( -2b^3 + b^2w(8w+\sqrt{b+w\sqrt{x}}) \right. \\ &\quad \left. + bw^3(5\sqrt{b+w\sqrt{x}}+8\sqrt{x}) + 5w^4\sqrt{b+w\sqrt{x}\sqrt{x}} \right). \end{aligned}$$

Combining and rewriting the last expressions gives the required result, after substitution in  $f_3(x)$ .

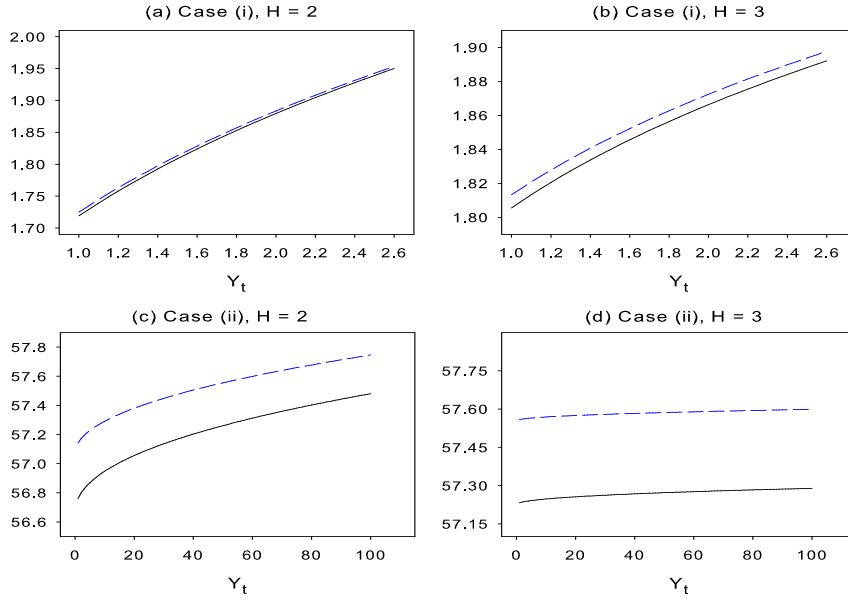
(b) With  $x - \omega x^{1/2} - a = 0$  and  $x - \omega^{1/2} - b = 0$ , we have

$$\alpha^2 = x_a^2 = (\omega + \sqrt{\omega^2 + 4a})/2, \quad \text{and} \quad \beta^2 = x_b^2 = (\omega + \sqrt{\omega^2 + 4b})/2.$$

Furthermore,

$$\begin{aligned} g_1(x) &= \omega x + (a+b)/2, \quad g_2(x) = \omega(\sqrt{\omega x + (a+b)/2}) + (a+b)/2, \quad \text{and} \\ g_3(x) &= \omega\sqrt{g_2(x)} + (a+b)/2. \end{aligned}$$

<sup>4</sup>Mathematica is a registered trademark of Wolfram, Corp.



**Figure 10.1:** Forecast functions  $g_H(x)$  (blue medium dashed lines) for the naive (SK) forecasting method and  $f_H(x)$  (black solid lines) for the exact LS forecast method;  $H = 2$  and 3.

Thus, in case (i) ( $\omega = 1, a = 0, b = 1$ )  $\alpha = 1$  and  $\beta = 2.618$ , and in case (ii) ( $\omega = 1, a = 0, b = 100$ )  $\alpha = 1$  and  $\beta = 110.512$ .

Figures 10.1(a) – (d) show the forecast functions  $g_H(x)$  (blue medium dashed lines) and  $f_H(x)$  (black solid lines) ( $H = 2$  and 3) for each case separately. In all cases, the differences between the naive method and the exact LS method is small. In Figure 10.1(d), we see that all forecasts with  $Y_t \in [0, 110.512]$  are very near to  $(a + b)/2 + \omega\sqrt{x} \approx 57$ . So, also in this case the forecasts obtained by both methods are practically the same.

- 10.2** (a) We consider separately the cases  $Y_t \leq r$ , and  $Y_t > r$  as this will influence the form of  $\mu(\cdot; \theta)$ . Moreover, in each case the range of integration will need to be decomposed into two sub-ranges depending on the value of  $Y_{t+1}$  as this in turn influences the form of  $\mu(Y_{t+1}; \theta)$ .

**Case 1:** Suppose  $Y_t \leq r$ , so that  $\mu(Y_t; \theta) = \phi_1 Y_t$  and relation (10.6) becomes

$$\begin{aligned}
 Y_{t+2|t} &= \int_{-\infty}^r \phi_1 Y_{t+1} g(Y_{t+1} - \phi_1 Y_t) dY_{t+1} + \int_r^{\infty} \phi_2 Y_{t+1} g(Y_{t+1} - \phi_1 Y_t) dY_{t+1} \\
 &= \int_{-\infty}^r \phi_1 \{\phi_1 Y_t + \varepsilon_t\} g(Y_{t+1} - \phi_1 Y_t) dY_{t+1} \\
 &\quad + \int_r^{\infty} \phi_2 \{\phi_1 Y_t + \varepsilon_t\} g(Y_{t+1} - \phi_1 Y_t) dY_{t+1}.
 \end{aligned}$$

Using the SETAR specification and changing the variable of integration we have

$$Y_{t+2|t} = \int_{-\infty}^{r-\phi_1 Y_t} \{\phi_1^2 Y_t + \phi_1 \varepsilon_{t+1}\} g(\varepsilon_{t+1}) d\varepsilon_{t+1} \\ + \int_{r-\phi_1 Y_t}^{\infty} \{\phi_1 \phi_2 Y_t + \phi_2 \varepsilon_{t+1}\} g(\varepsilon_{t+1}) d\varepsilon_{t+1}.$$

Noting that  $g(\varepsilon_t) = (1/\sigma_\varepsilon)\varphi(\varepsilon_t/\sigma_\varepsilon)$ , we get

$$Y_{t+2|t} = \phi_1^2 Y_t \Phi\left(\frac{r-\phi_1 Y_t}{\sigma_\varepsilon}\right) - \phi_1 \sigma_\varepsilon \varphi\left(\frac{r-\phi_1 Y_t}{\sigma_\varepsilon}\right) \\ + \phi_1 \phi_2 Y_t \left[1 - \Phi\left(\frac{r-\phi_1 Y_t}{\sigma_\varepsilon}\right)\right] + \phi_2 \sigma_\varepsilon \varphi\left(\frac{r-\phi_1 Y_t}{\sigma_\varepsilon}\right) \\ = \{\phi_1 p_{(1)} + \phi_2 (1-p_{(1)})\} Y_{t+1|t} + (\phi_2 - \phi_1) \sigma_{e,t+1} \varphi\left(\frac{r-Y_{t+1|1}}{\sigma_{e,t+1}}\right),$$

where

$$p_{(i)} = \Phi\left(\frac{r-\phi_i Y_t}{\sigma_{e,t+1}}\right) \quad (i = 1, 2).$$

**Case 2:** Suppose  $Y_t > r$ , so that  $\mu(Y_t; \boldsymbol{\theta}) = \phi_2 Y_t$ . Then, following similar arguments as in the first case, the optimal least squares two-step ahead prediction conditional on  $Y_t$  is given by  $Y_{t+2|t}$  in case 1, but now with  $Y_{t+1|t} = \phi_2 Y_t$ . Then combining both expressions gives the required result for the exact two-step ahead MMSFE forecast.

We observe from both cases that the two-step point forecast is a weighted average of the forecasts from regime 1,  $Y_{t+1|t} = \phi_1 Y_t$  and regime 2,  $Y_{t+1|t} = \phi_2 Y_t$  where the weights  $p_{(1)}$  and  $1-p_{(1)}$  are respectively the probabilities of the process being in the lower and higher regime at time  $t$ , assuming normality of the forecast error. We also note that if  $\phi_1 = \phi_2 = \phi$ , so the SETAR(2; 1, 1) model reduces to a linear AR(1) model, the nonlinearity of the prediction disappears and the result reduces to the usual expression of  $\phi^2 Y_t$ .

(b) Similar as in part (a) we have

$$\mathbb{E}(Y_{t+2}^2 | Y_t) = \int_{-\infty}^{\infty} \mathbb{E}(Y_{t+1}^2 | Y_t) g(Y_{t+1} - \lambda(Y_t)) dY_{t+1},$$

where  $\mathbb{E}[(\phi_i Y_{t+1} + \varepsilon_{t+1})^2 | Y_{t+1}] = \phi_i^2 Y_{t+1}^2 + \sigma_\varepsilon^2$  with  $i = 1$  if  $Y_{t+1} \leq r$  and  $i = 2$  if  $Y_{t+1} > r$ . Then we have

$$\mathbb{E}(Y_{t+2}^2 | Y_t) = \begin{cases} (1 + \phi_2^2) \sigma_\varepsilon^2 + \phi_1^2 \phi_2^2 Y_t^2 + (\phi_1^2 - \phi_2^2) (\sigma_\varepsilon^2 + \phi_1^2 Y_t^2) p_1 \\ - \sigma_\varepsilon (\phi_1^2 - \phi_2^2) (r + \phi_1 Y_t) \varphi\left(\frac{r-\phi_1 Y_t}{\sigma_\varepsilon}\right) & \text{if } Y_t \leq r, \\ (1 + \phi_2^2) \sigma_\varepsilon^2 + \phi_2^4 Y_t^2 + (\phi_1^2 - \phi_2^2) (\sigma_\varepsilon^2 + \phi_2^2 Y_t^2) p_2 \\ - \sigma_\varepsilon (\phi_1^2 - \phi_2^2) (r + \phi_2 Y_t) \varphi\left(\frac{r-\phi_2 Y_t}{\sigma_\varepsilon}\right) & \text{if } Y_t > r. \end{cases}$$

Combining the results from the two regimes gives the required expression of the second-order conditional moment of  $Y_{t+2}$  given  $Y_t$  with  $\sigma_{t+1|t}^2 = \sigma_\varepsilon^2$ . Note that setting  $\phi_1 = \phi_2 = \phi$  gives the variance  $\sigma_{t+2|t}^2 = \mathbb{E}(Y_{t+2}^2 | Y_t) - \mathbb{E}^2(Y_{t+2} | Y_t) = (1 + \phi^2) \sigma_\varepsilon^2$  which is the two-step ahead variance of the forecast error for a linear AR(1) model with parameters  $\phi$  and  $\sigma_\varepsilon^2$ .

- (c) As  $Y_t \rightarrow \pm\infty$  the fourth and fifth term in the expression for  $\sigma_{t+2|t}^2$  will tend to zero. The limiting values are determined from the first two terms and partly from the third term. That is, from the expression

$$\begin{aligned} & \sigma_\varepsilon^2 \Phi(z_{t+1|t}) + \sigma_\varepsilon^2 \Phi(-z_{t+1|t}) + \{\phi_1^2 \Phi(z_{t+1|t}) + \phi_2^2 \Phi(-z_{t+1|t})\} \sigma_\varepsilon^2 \\ & = \sigma_\varepsilon^2 \{1 + \phi_1^2\} \Phi(z_{t+1|t}) + \sigma_\varepsilon^2 (1 + \phi_2^2) - \sigma_\varepsilon^2 \{1 + \phi_2^2\} \Phi(z_{t+1|t}). \end{aligned}$$

Assume  $Y_t \leq r$  and  $\phi_1 > 0$ . Then it is unlikely ( $\Phi(\cdot) \rightarrow 0$ ) that the process will move out of the first regime  $(-\infty, r]$  and so the path taken is influenced solely by the first sub-model. Therefore, the two-step ahead forecast variance matches that of a linear AR(1) model with parameters  $\phi_1$  and  $\sigma_\varepsilon^2$ , i.e.  $\sigma_{t+2|t}^2 = \sigma_\varepsilon^2 \{1 + \phi_1^2\}$ . If, however,  $\phi_1 < 0$ ,  $\Phi(\cdot) \rightarrow 0$ , so that  $\sigma_{t+2|t}^2 = \sigma_\varepsilon^2 \{1 + \phi_2^2\}$ .

Now assume  $Y_t > r$  and  $\phi_2 > 0$ . In this case the same argument as above will apply if  $Y_t \rightarrow \infty$  (i.e.  $\Phi(\cdot) \rightarrow 0$ ), leading to  $\sigma_{t+2|t}^2 = \sigma_\varepsilon^2 \{1 + \phi_2^2\}$  which matches that of a linear AR(1) model with parameters  $\phi_2$  and  $\sigma_\varepsilon^2$ . If, however,  $\phi_2 < 0$  then it is most likely ( $\Phi(\cdot) \rightarrow 1$ ) that the process moves into the second regime  $[r, \infty)$  and the future path would now depend on the second sub-model, leading to the same variance as that of an AR(1) process with parameters  $\phi_1$  and  $\sigma_\varepsilon^2$ .

- 10.3** (a) We start off with a slightly more general model specification than in Example 10.1 of the main text, i.e. we assume that  $\{Y_t, t \in \mathbb{Z}\}$  follows the piecewise constant SETAR model:

$$Y_t = \alpha_i + \varepsilon_t, \quad \text{if } Y_{t-1} \in (r_{i-1}, r_i] = \mathbb{R}^{(i)} \quad (i = 1, 2, \dots, k).$$

Here,  $-\infty = r_0 < r_1 < \dots < r_k = \infty$  are the threshold values, and  $\{\varepsilon_t, t \in \mathbb{Z}\}$  is a sequence of zero mean i.i.d. random variables having an absolute continuous pdf  $g(\cdot)$ , and distribution function  $G(\cdot)$ , and finite variance  $\sigma_\varepsilon^2$ . Underlying the above model is a Markov chain  $M_t$  with states  $\{1, 2, \dots, k\}$ , where  $M_t = i$  if and only if  $Y_t \in \mathbb{R}^{(i)}$ . The transition probability matrix  $\mathbf{P}$  of the process  $\{M_t\}$  has elements

$$p_{ij} = \mathbb{P}(M_t = i | M_{t-1} = j) = G(r_i - \alpha_j) - G(r_{i-1} - \alpha_j).$$

Under fairly weak conditions on the parameters and the noise distribution, the process  $\{M_t\}$  will be ergodic and will possess a stationary distribution denoted by the row vector  $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_k)$ . This is the unique solution to the system  $\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P}$  with  $\sum_{i=1}^k \pi_i = 1$ . Let us assume that  $\mathbf{P}$  has distinct eigenvalues  $1 = \lambda_1, \lambda_2, \dots, \lambda_k$ , where  $|\lambda_i| < 1$  ( $i \geq 2$ ) under the assumptions on  $\{M_t\}$ . Let the corresponding right (left) eigenvectors be  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k$  ( $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_k$ ), normalized so that  $\mathbf{d}_i' \mathbf{e}_j = \delta_{ij}$ , Kronecker's delta.

Let  $f_{H,i}(y|\mathbf{y})$  denote the  $H$ -step predictive pdf of  $Y_{t+H}$  ( $H \geq 1$ ) given that  $Y_t \in \mathbb{R}^{(i)}$ . When  $H = 1$ , it is easy to check that  $f_{1,i}(y|\mathbf{y}) = g(y - \alpha_i)$ . The  $k \times 1$  vector  $\mathbf{f}_H(y|\mathbf{y})$  with  $i$ th element  $f_{H,i}(y|\mathbf{y})$  satisfies the equation

$$\mathbf{f}_H(y|\mathbf{y}) = \mathbf{P} \mathbf{f}_{H-1}(y|\mathbf{y}) = \mathbf{P}^{H-1} \mathbf{f}_1(y|\mathbf{y}) \quad (H \geq 2),$$

as can be easily seen on applying the law of total probability. The stationary marginal pdf is also easily derived to be

$$f(y) = \sum_{i=1}^k \pi_i g(y - \alpha_i).$$



Thus, the mean of  $\{Y_t, t \in \mathbb{Z}\}$  is given by  $\mu_Y = \sum_{i=1}^k \pi_i \alpha_i$  and its variance is given by  $\sigma_Y^2 = \sigma_\varepsilon^2 - \mu_Y^2 + \sum_{i=1}^k \pi_i \alpha_i^2$ .

Let  $\mu_{H,i}(y|\mathbf{y}) = \mathbb{E}(Y_{t+H}|Y_t \in \mathbb{R}^{(i)})$ , and  $\boldsymbol{\mu}_H(y|\mathbf{y})$  the  $k \times 1$  vector with  $i$ th element  $\mu_{H,i}(y|\mathbf{y})$ . Then the conditional mean is given by

$$\boldsymbol{\mu}_H(y|\mathbf{y}) = \mathbf{P}^{m-1} \boldsymbol{\mu}_1(y|\mathbf{y}) = \sum_{i=1}^k \lambda_i^{m-1} \mathbf{e}_i \mathbf{d}'_i \boldsymbol{\alpha},$$

where in the second step we used the spectral decomposition of  $\mathbf{P}$  and  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)'$ .

Now, consider the SETAR(2; 0, 0) process in Example 10.1 of the main text with  $\alpha_1 = -\alpha_2 \equiv \alpha$ , and  $\{\varepsilon_t\} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ . Thus, in the present case, the transition matrix  $\mathbf{P}$  reduces to a  $2 \times 1$  vector  $\mathbf{w}$  with elements  $(\Phi(r_i - \alpha) - \Phi(r_{i-1} + \alpha))$  ( $i = 1, 2$ ) where  $-\infty = r_0 < r_1 < r_2 = \infty$ . Let  $w_i^{(H)}$  ( $i = 1, 2$ ) denote the elements of the vector  $\mathbf{w}^{H-1}$ . Then it is easy to see that (10.9) and (10.10) follow from the above results.

(b) From (10.9) we have

$$\mu_{i,Y} = \mathbb{E}(Y_t^i) = \frac{1}{2} \left\{ \int_{-\infty}^{\infty} (y + \alpha)^i \varphi(y) dy + \int_{-\infty}^{\infty} (y - \alpha)^i \varphi(y) dy \right\}.$$

On expanding  $(y \pm \alpha)^i$  in a binomial form for each value of  $i$  ( $i = 1, 2, \dots$ ) and using the results  $\int_{-\infty}^{\infty} y^{2n} \varphi(y) dy = (2n - 1)!!$  and  $\int_{-\infty}^{\infty} y^{2n+1} \varphi(y) dy = 0$  ( $n = 0, 1, \dots$ ), the  $i$ th moment follows directly. Hence,  $\mathbb{E}(Y_t) = 0$ ,  $\text{Var}(Y_t) = \mathbb{E}(Y_t^2) = (1 + \alpha^2)$ . Moreover,  $\mathbb{E}(Y_t^3) = 0$ , and  $\mathbb{E}(Y_t^4) = 3 + \alpha^2 + \alpha^4$ .

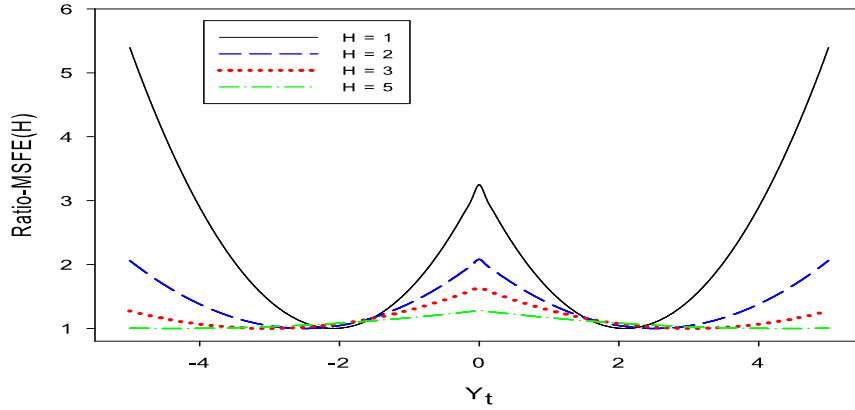
Denote the pdf of  $Y_{t-1}$  by  $f(y_{t-1})$ , and the joint pdf of  $Y_t$  and  $Y_{t-1}$  by  $g(y_t, y_{t-1}) = h(y_t|y_{t-1})f(y_{t-1})$ , with  $h(y_t|y_{t-1})$  the conditional pdf of  $Y_t$  given  $Y_{t-1} = y_{t-1}$ . Then the autocovariance at lag  $\ell = 1$  is given by

$$\begin{aligned} \gamma_Y(1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_t y_{t-1} g(y_t, y_{t-1}) dy_t dy_{t-1} \\ &= \int_{-\infty}^0 \int_{-\infty}^{\infty} y_t y_{t-1} g(y_t, y_{t-1}) dy_t dy_{t-1} + \int_0^{\infty} \int_{-\infty}^{\infty} y_t y_{t-1} g(y_t, y_{t-1}) dy_t dy_{t-1} \\ &= \int_{-\infty}^0 y_{t-1} f(y_{t-1}) \mathbb{E}(Y_t | Y_{t-1} \leq 0) dy_{t-1} \\ &\quad + \int_0^{\infty} y_{t-1} f(y_{t-1}) \mathbb{E}(Y_t | Y_{t-1} > 0) dy_{t-1} \\ &= \alpha \left\{ -\varphi(\alpha) + \frac{1}{2} \alpha \beta \right\} - \alpha \left\{ \varphi(\alpha) - \frac{1}{2} \alpha \beta \right\} = -2\alpha \varphi(\alpha) + \alpha^2 \beta, \end{aligned}$$

where  $\beta = 1 - 2\Phi(\alpha)$ . The above result can be easily generalized to autocovariances at lags  $\ell > 1$ .

(c) Note that

$$\begin{aligned} \text{MSFE}(Y_{t+H}^{\text{AR}}|t) &= \mathbb{E}[(Y_{t+H} - Y_{t+H}^{\text{AR}}|t)^2] \\ &= \text{MSFE}(Y_{t+H}^{\text{SETAR}}|t) + (Y_{t+H}^{\text{AR}}|t - Y_{t+H}^{\text{SETAR}}|t)^2, \end{aligned}$$



**Figure 10.2:** Ratio-MSFE( $H$ ) versus  $Y_t$  for the AR(1) process with parameter  $\phi = -0.7194$  and the SETAR(2; 0, 0) process with parameter  $a = 1.5$ .

and

$$\text{MSFE}(Y_{t+H}^{\text{SETAR}}) = \mathbb{E}[(Y_{t+H} - Y_{t+H|t}^{\text{SETAR}})^2] = 1 + \alpha^2 - (Y_{t+H|t}^{\text{SETAR}})^2.$$

Then, using  $Y_{t+H}^{\text{AR}} = \phi^H Y_t$ , we have the ratio of MSFEs

$$\begin{aligned} \text{Ratio-MSFE}(H) &\equiv \frac{\text{MSFE}(Y_{t+H}^{\text{AR}})}{\text{MSFE}(Y_{t+H}^{\text{SETAR}})} = 1 + \frac{(Y_{t+H|t}^{\text{AR}} - Y_{t+H|t}^{\text{SETAR}})^2}{\text{MSFE}(Y_{t+H}^{\text{SETAR}})} \\ &= 1 + \frac{\phi^H Y_t + \alpha\beta^{H-1}I(Y_t \leq 0) - \alpha\beta I(Y_t > 0)}{1 + \alpha^2(1 - \beta^{2H-2})}. \end{aligned}$$

- (d) From part (b) it follows that  $\rho_Y^{\text{SETAR}}(1) = -\alpha(2\varphi(\alpha) - \alpha\beta)/(1 + \alpha^2)$ . Equating this expression with  $\rho_Y^{\text{AR}}(1) = \phi$ , with  $\alpha = 1.5$ , it follows that  $\phi \approx -0.7194$ . Figure 10.2 shows plots of Ratio-MSFE( $H$ ) versus  $Y_t \in [-5, 5]$  for  $H = 1, 2$ , and 3. Clearly, the ratio of MSFEs is not necessarily a decreasing function in  $H$  at all forecast origins  $Y_t$ . Ratio-MSFE( $H$ )  $\rightarrow 1$  as  $H \rightarrow \infty$ , which implies that the multi-step ahead forecast of the AR(1) process is asymptotically as good as that of the true nonlinear SETAR(2; 0, 0) process.

- 10.4** (a) Using expression (10.80), we have for  $H = 1, 2, \dots$ ,

$$\begin{aligned} Y_{t+H}^{\text{LS}} &= \mathbb{E}(Y_{t+H} | \mathbf{Y}_t) = \alpha^H Y_t + \sum_{j=1}^H \alpha^{H-j} \mathbb{E}(\varepsilon_{t+j} | \mathbf{Y}_t) \\ &= \alpha^H Y_t + \mathbb{E}(\varepsilon_t)(1 - \alpha^H)/(1 - \alpha) = \alpha^H Y_t + \mu(1 - \alpha^H). \end{aligned}$$

Clearly,  $\mathbb{E}(e_{t+H}) = 0$ , i.e.  $Y_{t+H}^{\text{LS}}$  is an unbiased forecast of  $Y_{t+H}$ . Hence, the MSFE( $H$ ) is given by

$$\begin{aligned} \text{MSFE}(H) &= \text{Var}(Y_{t+H} | \mathbf{Y}_t) = \text{Var}(\alpha^{H-1}\varepsilon_{t+1} + \dots + \varepsilon_{t+H}) \\ &= \sum_{j=1}^H \alpha^{2(H-j)} \text{Var}(\varepsilon_{t+j}) = \mu^2(1 - \alpha^{2H}), \end{aligned}$$

where  $\mathbf{Y}_t = (Y_t, Y_{t-1}, \dots)'$ .

- (b) Let  $d = \mu(1 - \alpha)$ . Then, using expression (10.80), the MAFE of the one-step ahead LS forecast is given by

$$\begin{aligned} \text{MAFE}(1) &= \mathbb{E}(|Y_{t+1} - \alpha Y_t - d|) = \mathbb{E}(\varepsilon_{t+H} - d) \\ &= \alpha \mathbb{E}\{|-d|\} + (1 - \alpha) \mathbb{E}\{|E_{t+1} - d|\}. \end{aligned}$$

Now,

$$\begin{aligned} \mathbb{E}\{|E_{t+1} - d|\} &= - \int_0^d \frac{(y-d)}{\mu} e^{-y/\mu} dy + \int_d^\infty \frac{(y-d)}{\mu} e^{-y/\mu} dy \\ &= d - \mu + 2\mu e^{-d/\mu}. \end{aligned}$$

Substituting for  $d$ , the required results follows.

- 10.5** (a) The summation term in the expression for the sample ACF  $\hat{\gamma}_d^*(\ell)$  can be re-expressed as follows

$$\begin{aligned} \sum_{t=\ell+1}^P (d_t - \bar{d})(d_{t-\ell} - \bar{d}) &= \sum_{t=\ell+1}^P d_t d_{t-\ell} + (P - \ell) \bar{d}^2 - \bar{d} \sum_{t=\ell+1}^P d_t - \bar{d} \sum_{t=\ell+1}^P d_{t-\ell} \\ &= \sum_{\ell+1}^P d_t d_{t-\ell} + (P - \ell) \bar{d}^2 - \bar{d} \left( P \bar{d} - \sum_{t=1}^{\ell} d_t \right) \\ &\quad - \bar{d} \left( P \bar{d} - \sum_{t=P-\ell+1}^P d_t \right) \\ &= \sum_{t=\ell+1}^P d_t d_{t-\ell} - (P + \ell) \bar{d}^2 + \bar{d} \left( \sum_{t=1}^{\ell} d_t + \sum_{t=P-\ell+1}^P d_t \right). \end{aligned}$$

Taking the expectation of each term gives

$$\begin{aligned} \mathbb{E} \left[ \sum_{t=\ell+1}^P d_t d_{t-\ell} \right] &= (P - \ell) \gamma_d(\ell), \quad \mathbb{E}[(P + \ell) \bar{d}^2] = (P + \ell) \text{Var}(\bar{d}), \\ \mathbb{E} \left[ \bar{d} \left( \sum_{t=1}^{\ell} d_t + \sum_{t=P-\ell+1}^P d_t \right) \right] &= \mathbb{E} \left[ \frac{1}{P} \sum_{t=1}^P d_t \sum_{t=1}^{\ell} d_t + \frac{1}{P} \sum_{t=1}^P d_t \sum_{t=P-\ell+1}^P d_t \right] \\ &= \frac{2}{P} \left[ \sum_{i=1}^{\ell-1} (\ell - i) \gamma_d(i) + \sum_{i=0}^{P-\ell} \ell \gamma_d(i) + \sum_{i=1}^{\ell-1} (\ell - i) \gamma_d(P - \ell + i) \right]. \end{aligned}$$

Substituting the above results into the expected value of the sample ACF gives

$$\begin{aligned} \mathbb{E}(\hat{\gamma}_d^*(\ell)) &= (P - \ell)^{-1} \{ (P - \ell) \gamma_d(\ell) - (P + \ell) \text{Var}(\bar{d}) \\ &\quad + 2P^{-1} \left[ \sum_{i=1}^{\ell-1} (\ell - i) \gamma_d(i) + \sum_{i=0}^{P-\ell} \ell \gamma_d(i) + \sum_{i=1}^{\ell-1} (\ell - i) \gamma_d(P - \ell + i) \right] \} \\ &= \gamma_d(\ell) - (P - \ell)^{-1} (P + \ell) \text{Var}(\bar{d}) + \mathcal{O}(P^{-2}). \end{aligned}$$

Assuming that  $P$  is large relative to  $\ell$ , we have  $\mathbb{E}(\hat{\gamma}_d^*(\ell)) \approx \gamma_d(\ell) - \text{Var}(\bar{d})$ .

- (b) If the process  $\{d_t, t \in \mathbb{Z}\}$  is WN, it follows that  $\text{Var}(\bar{d}) = P^{-1}\gamma_d(0)$  and  $\gamma_d(\ell) = 0$  for  $\ell \neq 0$ . Hence, from the results in part (a), we have

$$\begin{aligned} \mathbb{E}(\hat{\gamma}_d^*(\ell)) &= \gamma_d(\ell) - (P - \ell)^{-1}(P + \ell)\text{Var}(\bar{d}) \\ &\quad + \frac{2}{P(P - \ell)} \left[ \sum_{i=1}^{\ell-1} (\ell - i)\gamma_d(i) + \sum_{i=0}^{P-\ell} \ell\gamma_d(i) + \sum_{i=1}^{\ell-1} (\ell - i)\gamma_d(P - \ell + i) \right] \\ &= \gamma_d(\ell) - (P - \ell)^{-1}(P + \ell)\text{Var}(\bar{d}) + 2P^{-1}(P - \ell)^{-1}\ell\gamma_d(0) \\ &= \gamma_d - (P - \ell)^{-1}((P + \ell) - 2\ell)\text{Var}(\bar{d}) = \gamma_d(\ell) - \text{Var}(\bar{d}) \\ &= \begin{cases} \gamma_d(0) - \text{Var}(\bar{d}) & \ell = 0, \\ -\text{Var}(\bar{d}) & \ell \neq 0. \end{cases} \end{aligned}$$

This time no approximation is necessary due to the nature of the WN process  $\{d_t, t \in \mathbb{Z}\}$ .

- 10.6** (a) The one-step ahead combined forecast error can be written as

$$\begin{aligned} e_{T+1|T}^C &= Y_{T+1|T} - Y_{T+1|T}^C \\ &= wY_{T+1|T} + (1 - w)Y_{T+1|T} - \{wY_{1,T+1|T} + (1 - w)Y_{2,T+1|T}\} \\ &= we_{1,T+1|T} + (1 - w)e_{2,T+1|T}. \end{aligned}$$

The variance of  $e_{T+1|T}^C$  is given by  $\text{Var}(e_{T+1|T}^C) = w^2\sigma_{1,e}^2 + (1 - w)^2\sigma_{2,e}^2 + 2w(1 - w)\sigma_{12}$ . This expression attains a minimum value if

$$\omega(\sigma_{1,e}^2 + \sigma_{2,e}^2 - 2\sigma_{12}) - \sigma_{2,e}^2 + \sigma_{12} = 0.$$

Solving this expression for  $w$ , we obtain the required result  $w^*$ .

- (b) Substituting  $w^*$  in  $\text{Var}(e_{T+1|T}^C)$  gives

$$\begin{aligned} \sigma_C^2(w^*) &= (w^*)^2(\sigma_{1,e}^2 + \sigma_{2,e}^2 - 2\sigma_{12}) + \sigma_{2,e}^2 - 2w^*(\sigma_{2,e}^2 - \sigma_{12}) \\ &= \sigma_{2,e}^2 + \frac{(\sigma_{2,e}^2 - \sigma_{12})^2}{\sigma_{1,e}^2 + \sigma_{2,e}^2 - 2\sigma_{12}} - 2\frac{(\sigma_{2,e}^2 - \sigma_{12})^2}{\sigma_{1,e}^2 + \sigma_{2,e}^2 - 2\sigma_{12}} \\ &= \sigma_{2,e}^2 - \frac{(\sigma_{2,e}^2 - \sigma_{12})^2}{\sigma_{1,e}^2 + \sigma_{2,e}^2 - 2\sigma_{12}}. \end{aligned}$$

The numerator of the second term is always positive. If  $\rho_{12} = 1$ , we have  $\sigma_{12} = \sigma_{1,e}\sigma_{2,e}$ . In this case the denominator reduces to  $(\sigma_{1,e} - \sigma_{2,e})^2$ , which is always positive. If  $\rho_{12} = -1$ , we have  $\sigma_{12} = -\sigma_{1,e}\sigma_{2,e}$ . Then the denominator reduces to  $(\sigma_{1,e} + \sigma_{2,e})^2$  which is always positive. In fact, for all values of  $\rho_{12} \in [-1, 1]$  the denominator of the second term is positive, including  $\rho_{12} = 0$ . Hence,  $\sigma_C^2(w^*) < \sigma_{2,e}^2$ . The equality  $\sigma_C^2(w^*) = \sigma_{2,e}^2$  holds if  $\sigma_{2,e}^2 = \sigma_{12}$ . This implies that  $w^* = 0$ , and hence  $Y_{T+1|T}^C \equiv Y_{2,T+1|T}$ .

- (c) Assume  $\sigma_{12} = 0$ , which implies that  $\rho_{12} = 0$ . Then  $w^* = \sigma_{2,e}^2/(\sigma_{1,e}^2 + \sigma_{2,e}^2)$ . If  $\sigma_{1,e}^2 \rightarrow 0$ , we see that  $w^* \rightarrow 1$ . In other words, almost all weight is given to the first forecast  $Y_{1,T+1|T}$ . Similarly, if  $\sigma_{2,e}^2 \rightarrow 0$ , we see that  $w^* \rightarrow 0$ , i.e. almost all weight is given to the second forecast  $Y_{2,T+1|T}$ .

Now, assume  $\rho_{12} = 1$ . So,  $w^* = \sigma_{2,e}/(\sigma_{2,e} - \sigma_{1,e})$ . If  $\sigma_{2,e} \rightarrow 0$ , we have  $w^* \rightarrow 0$ . In this case almost all weight is given to the second forecast. In contrast, if  $\sigma_{1,e} \rightarrow 0$ , we have  $w^* \rightarrow 1$ . A similar observation follows for the case  $\rho_{12} = -1$ .

- (d) Given the  $T - 1$  forecast errors, consistent estimates of  $\sigma_{1,e}^2 \equiv \sigma_{11}^2$ ,  $\sigma_{2,e}^2 \equiv \sigma_{22}^2$ , and  $\sigma_{12}$  can be obtained as  $\hat{\sigma}_{ij} = (1/(T - 1)) \sum_{t=1}^{T-1} e_{1,t+1|t} e_{2,t+1|t}$ . Substitution in  $w^*$  gives an estimate of the optimal weight.

**10.7** Note that  $\{Y_t, t \in \mathbb{Z}\}$  is a first order Markov process. It is easily verified that for all  $i \in \mathbb{Z}$ , the conditional expectation of a  $p$ th order Markov process  $\{Y_t, t \in \mathbb{Z}\}$  is given by

$$\mathbb{E}(g(Y_i)|Y_t, t \neq i) = \mathbb{E}(g(Y_i)|Y_t, t \neq i, |t - i| \leq p),$$

for all bounded measurable functions  $g(\cdot)$  on  $\mathbb{R}$ . Then, in view of this result, it is obvious that  $\mathbb{E}(Y_i|\mathbf{Y}^{-i}) = \mathbb{E}(Y_i|Y_{i-1}, Y_{i+1})$ . For ease of exposition, we consider the case  $Y_{i-1} \geq 0$ . Thus, we obtain

$$\begin{aligned} \widehat{Y}_i &= \mathbb{E}(Y_i|Y_{i-1}, Y_{i+1}) \\ &= \int_{-\infty}^{\infty} y_i f(y_i|Y_{i-1}, Y_{i+1}) dy_i \\ &= \int_{-\infty}^{\infty} y_i \frac{f(Y_{i+1}|y_i) f(y_i|Y_{i-1})}{f(Y_{i+1}|Y_{i-1})} dy_i \\ &= \frac{1}{f(Y_{i+1}|Y_{i-1})} \left( \int_0^{\infty} y_i \varphi_1(Y_{i+1} - \phi_1 y_i) \varphi_1(y_i - \phi_1 Y_{i-1}) dy_i \right. \\ &\quad \left. + \int_{-\infty}^0 y_i \varphi_2(Y_{i+1} - \phi_2 y_i) \varphi_1(y_i - \phi_1 Y_{i-1}) dy_i \right), \end{aligned}$$

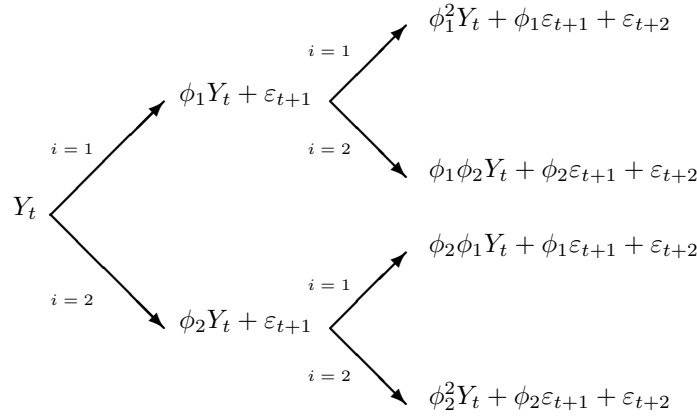
where  $\varphi_j(\cdot)$  is the density function of  $\mathcal{N}(0, \sigma_j^2)$  ( $j = 1, 2$ ), and  $f(Y_{i+1}|Y_{i-1})$  is the conditional density function of  $Y_{i+1}$  given  $Y_{i-1}$ , i.e.

$$\begin{aligned} f(Y_{i+1}|Y_{i-1}) &= \int_{-\infty}^{\infty} f(Y_{i+1}, y_i|Y_{i-1}) dy_i \\ &= \int_{-\infty}^{\infty} f(Y_{i+1}|y_i) f(y_i|Y_{i-1}) dy_i \\ &= \int_0^{\infty} \varphi_1(Y_{i+1} - \phi_1 y_i) \varphi_1(y_i - \phi_1 Y_{i-1}) dy_i \\ &\quad + \int_{-\infty}^0 \varphi_2(Y_{i+1} - \phi_2 y_i) \varphi_1(y_i - \phi_1 Y_{i-1}) dy_i. \end{aligned}$$

The integrals in the expression for  $\widehat{Y}_i$  follow by routine calculation. A similar result can be obtained for the case  $Y_{i-1} < 0$ . Finally, on combining, the required result follows.

**Remark:** Clearly, the conditional expectation can become a complicated expression for a simple NLAR model when there is only one missing observation. An alternative to the estimation of missing values, is to treat these values as parameters. Then, for instance, the likelihood function can be maximized with respect to the missing observations.

- 10.8** (a) Generally,  $\sigma_{t+2|t}^2$  attains its largest value as there is most uncertainty about the regime of  $Y_{t+1}$  in which the intermediate step lies. This is clearly seen in Figure 10.9(a) of the main text where two maxima occur in the variance function. On the other hand, in Figure 10.9(b) of the main text there is virtually no



**Figure 10.3:** Tree diagram of all possible paths from  $Y_t$  to  $Y_{t+2}$ .

uncertainty about the regime of  $Y_{t+1}$  whenever  $Y_t < -2$  and the uncertainty diminishes monotonically from  $Y_t \geq -2$ .

Starting at  $Y_t$ , the uncertainty about the future trajectories of the process depends on the parameter values of the model and the values of the process  $\{\varepsilon_t\}$ . The tree diagram shows all possible routes (paths) the SETAR process can take from  $Y_t$  to  $Y_{t+2}$ . Now, consider the SETAR process in Figure 10.9(a) with  $\phi_1 = 0.8$  and  $\phi_2 = -0.4$ . Assume that the series value at the forecast origin is  $Y_t = +7$ . Then the one-step ahead forecast  $Y_{t+1|t}$  is about  $\phi_2 Y_t = -0.4 \times 7 = -2.8$ . Hence, there is a good deal of uncertainty as to which path (sub-model) the process will take to forecast further. Now, assume  $\varepsilon_{t+1}$  takes a value less than 0.8. Then the value of the two-step ahead forecast  $Y_{t+2|t}$  is less than  $\phi_2 \phi_1 Y_t + \phi_1 \varepsilon_{t+1} = (-0.4 \times 0.8) \times 7 + 0.8 \times 0.8 = -1.6$ . If, however,  $\varepsilon_{t+1} > 0.8$  it is likely that the two-step ahead forecast will be above the threshold value. Thus, there is a substantial difference between the two possible forecasts and this leads to a large expected variance along with a large final two-step ahead forecast variance.

(b) The two-step ahead forecast can be written as

$$Y_{t+2|t} = \{\phi_1 \Phi(u) + \phi_2 (1 - \Phi(u))\} (r - u\sigma_\varepsilon) + (\phi_2 - \phi_1) \sigma_\varepsilon \psi(u).$$

Therefore the derivative is given by

$$\begin{aligned} \frac{dY_{t+2|t}}{du} &= -\phi_2 \sigma_\varepsilon + \{\phi_1 - \phi_2\} \varphi(u) (r - u\sigma_\varepsilon) - \{\phi_1 - \phi_2\} \Phi(u) \sigma_\varepsilon \\ &\quad + (\phi_2 - \phi_1) (-u\sigma_\varepsilon) \varphi(u) \\ &= -\phi_2 \sigma_\varepsilon + (\phi_1 - \phi_2) \{r\varphi(u) - \sigma_\varepsilon \Phi(u)\}. \end{aligned}$$

Similarly the two-step ahead forecast variance can be written as

$$\begin{aligned}\sigma_{t+2|t}^2 &= (1 + \phi_2^2)\sigma_\varepsilon^2 + \phi_2^2(r - u\sigma_\varepsilon)^2 + \{\phi_1^2 - \phi_2^2\}\Phi(u)\{(r - u\sigma_\varepsilon)^2 + \sigma_\varepsilon^2\} \\ &\quad + (\phi_2^2 - \phi_1^2)(2r - u\sigma_\varepsilon)\sigma_\varepsilon\varphi(u) - \widehat{Y}_{t+2|t}^2.\end{aligned}$$

Hence,

$$\begin{aligned}\frac{d\sigma_{t+2|t}^2}{du} &= -2\phi_2^2\sigma_\varepsilon(r - u\sigma_\varepsilon) + \{\phi_1^2 - \phi_2^2\}\varphi(u)\{(r - u\sigma_\varepsilon)^2 + \sigma_\varepsilon^2\} \\ &\quad - 2\sigma_\varepsilon(\phi_1^2 - \phi_2^2)\Phi(u)(r - u\sigma_\varepsilon) - \sigma_\varepsilon^2(\phi_2^2 - \phi_1^2)\varphi(u) \\ &\quad + (\phi_2^2 - \phi_1^2)(2r - u\sigma_\varepsilon)(-u\sigma_\varepsilon\varphi(u)) - 2Y_{t+2|t}\frac{dY_{t+2|t}}{du}.\end{aligned}$$

Solving  $d\sigma_{t+2|t}^2/du = 0$  numerically for values of the forecast origin in the range  $[-12, 12]$  gives as locations  $Y_t = -2.9$  and  $5.8$  respectively. Both extremes are in agreement with the locations of the maxima of the two-step ahead forecast variance function to be inferred from the plot in Figure 10.9(a).

**10.9** (a) With  $Z \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_Z^2)$ , and  $c$  and  $M$  constants, we have

$$\begin{aligned}\mathbb{E}(\exp(-c(Z + M)^2)) &= \frac{1}{\sqrt{2\pi\sigma_Z^2}} \int_{-\infty}^{\infty} \exp(-c(Z + M)^2) \exp\left(-\frac{1}{2\sigma_Z^2}Z^2\right) dZ \\ &= \frac{1}{\sqrt{2\pi\sigma_Z^2}} \int_{-\infty}^{\infty} \exp(-cZ^2 - 2cMZ - cM^2 - \frac{1}{2\sigma_Z^2}Z^2) dZ \\ &= \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp(-cM^2) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{A}{\sigma_Z^2}(Z^2 + 4\frac{Mc\sigma_Z^2}{A}Z)\right) dZ \\ &= \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left(-cM^2 + \frac{2c^2M^2\sigma_Z^2}{A}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{A}{\sigma_Z^2}\left(Z + \frac{2cM\sigma_Z^2}{A}\right)^2\right) dZ \\ &= \frac{1}{\sqrt{A}} \exp(-c_1M^2),\end{aligned}$$

where  $A = 1 + 2c\sigma_Z^2$  and  $c_1 = c/A$ . Let  $r(Z)$  be a function of  $Z$ . Then it is clear from the last integration that

$$\mathbb{E}\{r(Z) \exp(-c(Z + M)^2)\} = A^{-1/2} \exp(-c_1M^2) \mathbb{E}(R(V)),$$

where  $V \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(-2cM\sigma_Z^2, \sigma_Z^2/A)$ .

The result in (10.26) of the main text can be rewritten to obtain the expression for  $u_H$ . In particular, with  $U = V + M$ , it follows that

$$\begin{aligned}\mathbb{E}(U) &= \mathbb{E}(V + M) = M - \frac{2c}{A}\sigma_Z^2M \\ &= \frac{M}{A}(1 + 2c\sigma_Z^2 - 2c\sigma_Z^2) = \frac{M}{A}, \quad \text{since } A = 1 + 2c\sigma_Z^2.\end{aligned}$$

Hence, the required result is given by

$$\mathbb{E}\{r(Z + M) \exp(-c(Z + M)^2)\} = A^{-1/2} \exp(-c_1M^2) \mathbb{E}(R(U)),$$

where  $U \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(M/A, \sigma_Z^2/A)$ .

**Table 10.1:** Comparison of  $\text{MAFE}(H)$  and  $\text{MSFE}(H)$  for  $H = 5, 15,$  and 30-steps ahead forecasts made with NFE (normal forecast error), SK (skeleton), and LN (linearization); all numbers should be multiplied by  $10^3$ .

$H$	$\text{MAFE}(H)$			$\text{MSFE}(H)$		
	SK	NFE	LN	SK	NFE	LN
5	0.2756	0.2692	0.925	0.4919	0.4646	8.47
15	0.8416	0.8255	4.729	1.5288	1.4688	91.48
30	1.8162	1.7923	10.956	3.4993	3.3938	224.41

- (b) Recall from Example 6.1 of the main text that the specified ExpAR(1) model has two limiting points at  $\pm 0.2295$ , which are stable. So, we expect that values of the forecasts will oscillate between these two points. The results of the simulation experiment are summarized in Table 10.1. We also added results of the LN (linearization) forecasting method to the table. Clearly, the NFE method is best on both accuracy measures and across all forecast horizons. The SK method is the next best method, and the LN method is totally inadequate.

- 10.10** (a) The results in Table 10.3 of the main text can be calculated by means of the following R code, made available by Li Pan. Note, the computations can be time consuming.

```
#####
# R code: Exercise 10.10
# File: Table-10-3.r
#
# Ff = forward with fitted residuals; Fp = forward with predictive
# residuals; "Warming up = 100"; T = 100; B = 1,000; m = 500.
#
# Reference:
# Pan, L. and Politis, D.N. (2016).
# Bootstrap prediction intervals for linear, nonlinear and
# nonparametric autoregressions (with discussion).
# Journal of Statistical Planning and Inference, 177, 1-27.
# DOI: 10.1016/j.jspi.2014.10.003.
#####
f <- function(x){# SETAR(2;1,1) model with d=1, and N(0,1) errors
  if(x<=0) return(0.5*x)
  if(x>0) return(-0.4*x)
}
sim <- function(f,n){ # Simulation
  temp <- rep(0,(n+100))
  temp[1] <- 1
  e <- rnorm((n+100),0,1) # e <- rt(n+100,df=5) Student t(5)
  for (i in 1:((n-1)+100)){
    temp[i+1] <- f(temp[i])+e[i+1]
  }
  u <- temp[101:(100+n)]
}
```



```

}
bootstrap=function(u,fhat,residuals){
  bb = rep(0,(n+100))
  bb[1] = u[1]
  estar = sample(residuals,(n+101),replace=TRUE)
  for (i in 1:(n+99)){
    bb[i+1] <- fhat(bb[i])+estar[i+1]
  }
  b = bb[101:(n+100)]
  fv = fhat(u[n])+estar[(n+101)]#X_{n+1}^*
  list(b=b,fv=fv)
}
n = 100 # sample size
TAR.sim <- function(f,n){
  ##### Simulation and model estimation #####
  u <- sim(f,n)
  x <- u[1:(n-1)]
  y <- u[2:n]
  idx <- which(x>=0)
  if(length(idx)>=2 & length(idx)<=n-3){
    fit1 <- lm(y[idx]~x[idx])
    fit2 <- lm(y[-idx]~x[-idx])
    while(abs(fit1$coef[2])>=1 | abs(fit2$coef[2])>=1){
      u <- sim(f,n)
      x <- u[1:(n-1)]
      y <- u[2:n]
      idx <- which(x>=0)
      if(length(idx)>n-3){
        fit1 <- lm(y[idx]~x[idx])
        fit2 <- fit1
      }
      if(length(idx)<2){
        fit2 <- lm(y[-idx]~x[-idx])
        fit1 <- fit2
      }
      if(length(idx)>=2 & length(idx)<=n-3){
        fit1 <- lm(y[idx]~x[idx])
        fit2 <- lm(y[-idx]~x[-idx])
      }
    }
  }
  fres = c(fit1$resid,fit2$resid)
  fhat <- function(x){
    if(x>=0) y <- fit1$coef[2]*x+fit1$coef[1]
    if(x<0) y <- fit2$coef[2]*x+fit2$coef[1]
  }
}
else{
  fit <- lm(y~x)
  fres <- fit$resid
}

```

```

while(abs(fit$coef[2])>=1){
  u <- sim(f,n)
  x <- u[1:(n-1)]
  y <- u[2:n]
  fit <- lm(y~x)
}
fhat <- function(x) fit$coef[2] * x +fit$coef[1]
}
yhat <- fhat(u[n]) # Fitted value
fres <- fres-mean(fres) # Fitted residuals
sigmahat <- sd(fres) # Sigma
resid <- rep(0,(n-1)) # Predictive residuals
for (i in 1:(n-1)){
  xtemp <- x[-i]
  ytemp <- y[-i]
  idx.temp <- which(xtemp*x[i]>=0)
  if(length(idx.temp)>=2){
    xtemp <- xtemp[idx.temp]
    ytemp <- ytemp[idx.temp]
    tempfit <- lm(ytemp~xtemp)
  }
  else{
    tempfit <- lm(ytemp~xtemp)
  }
  resid[i] <- y[i]-predict(tempfit,data.frame(xtemp=x[i]))
}
pres <- resid-mean(resid)

##### Bootstrap #####
Ffstarhat <- rep(0,1000)
Ffstar <- rep(0,1000)
Fsf sd <- rep(0,1000)
for (i in 1:1000){
  result <- bootstrap(u,fhat,fres)
  ustar <- result$b
  xstar <- ustar[1:(n-1)]
  ystar <- ustar[2:n]
  idx.star <- which(xstar>=0)
  if(length(idx.star)<=n-3 & length(idx.star)>=2){
    fitstar1 <- lm(ystar[idx.star] ~ xstar[idx.star])
    fitstar2 <- lm(ystar[-idx.star] ~ xstar[-idx.star])
    fhatstar <- function(x){
      if(x>=0) y <- fitstar1$coef[2]*x+fitstar1$coef[1]
      if(x<0) y <- fitstar2$coef[2]*x+fitstar2$coef[1]
    }
    sigmastar <- sd(c(fitstar1$res,fitstar2$res))
  }
  else{
    fitstar <- lm(ystar~xstar)
  }
}

```

```

        fhatstar <- function(x) fitstar$coef[2]*x+fitstar$coef[1]
        sigmastar <- sd(fitstar$res)
    }
    Ffstarhat[i] <- fhatstar(u[n])
    Ffstar[i] <- result$fv
    Ffsfd[i] <- (Ffstar[i] -Ffstarhat[i])/sigmastar
}
Fpstarhat <- rep(0,1000)
Fpstar <- rep(0,1000)
Fpspd <- rep(0,1000)
for (i in 1:1000){
    result <- bootstrap(u,fhat,pres)
    ustar <- result$b
    xstar <- ustar[1:(n-1)]
    ystar <- ustar[2:n]
    idx.star <- which(xstar>=0)
    if(length(idx.star)<=n-3 & length(idx.star)>=2){
        fitstar1 <- lm(ystar[idx.star] ~ xstar[idx.star])
        fitstar2 <- lm(ystar[-idx.star] ~ xstar[-idx.star])
        fhatstar <- function(x){
            if(x>=0) y <- fitstar1$coef[2]*x+fitstar1$coef[1]
            if(x<0) y <- fitstar2$coef[2]*x+fitstar2$coef[1]
        }
        sigmastar <- sd(c(fitstar1$res,fitstar2$res))
    }
    else{
        fitstar <- lm(ystar~xstar)
        fhatstar <- function(x) fitstar$coef[2]*x+fitstar$coef[1]
        sigmastar <- sd(fitstar$res)
    }
    sigmastar <- sd(c(fitstar1$res,fitstar2$res))
    Fpstarhat[i] <- fhatstar(u[n])
    Fpstar[i] <- result$fv
    Fpspd[i] <- (Fpstar[i] -Fpstarhat[i])/sigmastar
}
##### Get the FI interval #####
# Ff (fitted residuals)
Ffdown <- yhat+quantile(Ffstar-Ffstarhat,.025)
Ffup <- yhat+quantile(Ffstar-Ffstarhat,.975)
Ffl <- Ffup-Ffdown
# Fp (predicted residuals)
Fpdown <- yhat+quantile(Fpstar-Fpstarhat,.025)
Fpup <- yhat+quantile(Fpstar-Fpstarhat,.975)
Fpl <- Fpup-Fpdown
# 1,000 future values
truevalue <- f(u[n]) + rnorm(1000)
Ff <- 0
Fp <- 0
for (i in 1:1000){

```

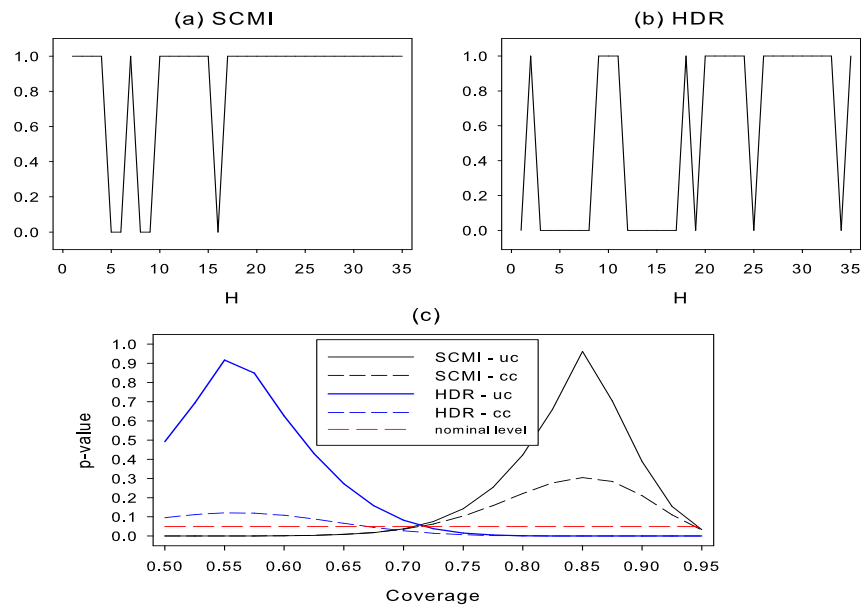
```

        if((truevalue[i]<Ffup) & truevalue[i]>Ffdown)
          Ff <- Ff+1
        if((truevalue[i]<Fpup) & truevalue[i]>Fpdown)
          Fp <- Fp+1
      }
      Ffc <- Ff/1000
      Fpc <- Fp/1000
      list(Ffc=Ffc,Fpc=Fpc,Ffl=Ffl,Fpl=Fpl)
    }
    Ffc <- rep(NA,500)
    Fpc <- rep(NA,500)
    Ffl <- rep(NA,500)
    Fpl <- rep(NA,500)
    for (i in 1:500){
      set.seed(5*i)
      results <- TAR.sim(f,n)
      Ffc[i] <- results$Ffc
      Fpc[i] <- results$Fpc
      Ffl[i] <- results$Ffl
      Fpl[i] <- results$Fpl
      if(i%%20==0) print(i)
    }
    print(mean(Ffc))
    print(mean(Fpc))
    print(mean(Ffl))
    print(mean(Fpl))
    print(sd(Ffl))
    print(sd(Fpl))

```

- (b) Bootstrap FI (BFI) has the lowest CVR for both distributions (a) and (b). In the case of the  $\mathcal{N}(0, 1)$  distribution, we see that the under-coverage of BFI with fitted residuals is partially corrected by the use of predictive residuals. For the Student  $t_5$  distribution the correction causes more coverage distortions than the standard normal distribution. We also see this effect in the values of  $\overline{FI}$ . The increase in variability of the interval length is the price one has to pay for using predictive residuals. This is a finite-sample effect since asymptotically the omission of a finite number of observations, through the use of the leave-one-out construction, does not make a real difference. In general, the BFI with fitted residuals performs reasonably well. This finding confirms the observations made in Example 10.6 of the main text.

- 10.11** (a) Figures 10.4(a) – (b) show the indicator (hit) function  $\{i_t^{(\alpha)}\}$  for SCMI and HDR. Figure 10.4(c) shows  $p$ -values of the  $LR_{uc}$  and  $LR_{cc}$  test statistics as a function of the coverage probability  $p$  with  $p$  ranging from 0.5 to 0.95. When  $p$  is approximately less than 0.71, we see that the uc (unconditional coverage) and the cc (conditional coverage) hypothesis both are rejected at the 5% nominal significance level for the SCMI-based FI. The reverse seems to be true for the HDR, rejecting the uc and cc hypotheses for coverage probabilities larger than approximately 0.7. In addition, the independence hypothesis is not rejected for SCMI with  $LR_{iid} = 2.37$  ( $p$ -value = 0.123). By contrast, the assumption of



**Figure 10.4:** Indicator (*hit*) function for SCMI and HDR in panels (a) and (b), respectively. Panel (c): empirical *p*-values of the LR test statistics for unconditional coverage (uc) and conditional coverage (cc).

independence is rejected for HDR ( $p$ -value = 0.04) at the 5% nominal significance level. Clearly, these test results reflect the fact that the SCMI-based FI is too wide at certain forecast horizons  $H$  (see Figure 10.6 in the main text), while the HDR-based FI is a bit too narrow.

- (b) For SCMI the observed value of the test statistic  $T_{N,[N/2]}$  with  $N = 30$  is  $-2.015$  ( $p$ -value = 0.998). Thus, we do not reject the independence (ind) hypothesis, which is in line with the conclusion based on the  $LR_{\text{ind}}$  test statistic in part (a). For HDR the observed value of the test statistic  $T_{N,[N/2]}$  with  $N = 19$  is  $1.214$  ( $p$ -value = 0.439). So, also in this case we do not reject the null hypothesis. Unfortunately, this result is not supported by the test result reported in part (a). Thus, one may tentatively conclude that both SCMI and HDR produce adequate independent FIs without clustering patterns.

The above test results can be obtained by means of the following R code.

```
#####
# R code: Exercise 10.11(b)
# File: Hit-function.r
# Coded by: P. Araujo Santos
#
# Reference:
# Araujo Santos, P. (2010).
# Interval forecasts evaluation: R programs for a
# new independence test. Notas e Comunicacoes CEAUL 17/2010.
# Available at: http://www.ceaul.fc.ul.pt/notas.html?ano=2010.
```

```

# Araujo Santos, P. and Fraga Alves, M.I. (2012).
# A new class of independence tests for interval forecasts
# evaluation.
# Computational Statistics & Data Analysis, 56(11), 3366-3380.
# DOI: 10.1016/j.csda.2010.10.002.
#####
# Hit sequence
hit <- c(1,1,1,1,0,0,1,0,0,1,1,1,1,1,1,0,1,1,1,1,1,1,1,
        1,1,1,1,1,1,1,1,1,1,1) # SCMI
# hit <- c(0,1,0,0,0,0,0,0,1,1,1,0,0,0,0,0,1,0,1,1,1,1,1,
#         0,1,1,1,1,1,1,1,1,0,1) # HDR
tt <- length(hit)
# Durations
no_hit_duration <- 0
j <- 1
zeros <- 0
for(i in 1:tt) {
  if (hit[i]<1){
    zeros <- zeros+1
  }
  else {
    no_hit_duration[j] <- zeros+1
    zeros <- 0
    j <- j+1
  }
}
no <- no_hit_duration
n <- length(no)

# T[0.5] Independence Test
no <- sort(no)
observed_T <- log(2)*(no[n]-1)/no[floor(0.5*n)]-log(n)
observed_T

# Simulation of the upper bound for the p-value
v <- 0
repl <- 250000 # Number of replications
v <- rep(0,times=repl)
for(i in 1:repl) {
  if((i/10000-floor(i/10000))==0){print(repl-i)}
  u <- runif(n)
  y <- -log(1-u)
  no_simul <- sort(y)
  v[i] <- log(2)*(no_simul[n]/no_simul[floor(n/2)])-log(n)
}
simulated_upper_bound_p_value <- length(v[v>=observed_T])/repl
observed_T
simulated_upper_bound_p_value

```

**10.12** For the LS predictor we have  $Y_{t+1|t}^{\text{LS}} = 0$ . The one-step ahead NL predictor is given by

$$Y_{t+1|t}^{\text{NL}} = \psi Y_{t-1} \hat{\varepsilon}_t,$$

where  $\hat{\varepsilon}_t = 0$ ,  $Y_t = 0$ , for  $t < 1$ . The usual procedure is to compute  $\{\hat{\varepsilon}_t\}_{t=1}^T$  recursively by using the relationship

$$\hat{\varepsilon}_t = Y_t - \psi Y_{t-2} \hat{\varepsilon}_{t-1}.$$

Since the model is completely known, the only relevance of the time series length  $T = 50$  is that the “estimated” errors need some time to converge. This seems to be the case for  $T = 50$ , but may not be so for  $T$  much smaller than 50. Figures 10.5(a) – (b) show the efficiency, as measured by the relative MSFE, for  $H = 1$  and 2.

(a)+(b) We see from Figure 10.5(a) that the best linear predictor is less efficient than the quadratic (Q) predictor (red solid line) for one-period ahead for sufficiently small values of  $\psi$  (approximately  $|\psi| < 0.47$ ). The disadvantage from using a linear predictor obviously decreases as the parameter  $\psi$  gets nearer to the boundary of the invertibility region since in that case the BL model will be more difficult to estimate. The theoretical efficiency loss is maximized for  $\psi = -0.38$  and  $\psi = 0.42$ . Also, we see that the one-step ahead nonlinear (NL) predictor is far more efficient (blue solid line) than the quadratic predictor for all values of  $\psi$ . Only for small values of  $|\psi|$  the quadratic predictor has a slightly smaller one-step ahead MSFE than the nonlinear predictor.

In forecasting two-steps ahead (Figure 10.5(b)), the quadratic predictor (based on one-step ahead forecasts) is solidly beaten by both the linear and nonlinear predictors for all values of  $|\psi|$ . Note that the efficiency becomes disastrously low for values of  $|\psi|$  near the boundary of the invertibility region.

**Remarks:**

- It can be shown (Terdik and Subba Rao, 1989, p. 284) that the MSE of the best  $H$ -step ahead linear predictor is always greater than the innovation variance of the process  $\{Y_t, t \in \mathbb{Z}\}$  satisfying the equation

$$Y_t = \phi_0 + \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{j=0}^R \sum_{v=0}^S \psi_{j,j+v} Y_{t-j-v} \varepsilon_{t-v},$$

where  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ .

- It is recommended to extend the MC simulation experiment to more elaborate subdiagonal and diagonal BL models. For instance, one may compare the forecasting performance of the nonlinear predictor and the quadratic predictor in case the DGP is given by

$$Y_t = \psi_1 Y_{t-2} \varepsilon_{t-1} + \psi_2 Y_{t-4} \varepsilon_{t-2} + \varepsilon_t, \quad \{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

Also, other forecast accuracy measures may be employed. We anticipate that for all BL models under study, the results will be disappointing for the quadratic predictor.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 10.12
% File: MSFE.m
%
% Purpose: Using MC simulation, the computer code computes
% the MSFEs of three predictors for forecasts obtained
% from a subdiagonal BL model, with
% NLP = Nonlinear predictor;
% QP = Quadratic predictor; and
% LP = Linear (least squares) predictor.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
nrep = 2000;
N = 100;

for b = 1:131
    beta = -0.65+(b-1)*0.01;
    index(b) = beta;
    for it = 1:nrep
        e = normrnd(0,1,154,1);
        y(1) = 0;
        y(2) = 0;

% Simulate time series
        for i = 3:154
            y(i) = beta*y(i-2)*e(i-1)+e(i); % BL process
        end
        w = y(103:154); % 52 data points, 100 observations for warming up
        w0 = 0;
        res(1) = w(1);
        res(2) = w(2)-beta*w0*res(1);
        res(3) = w(3)-beta*w(1)*res(2);
        for u = 4:52
            res(u) = w(u)-beta*w(u-2)*res(u-1);
        end

        wLtrue1(it) = w(51); % actual values
        wLtrue2(it) = w(52);

        NLP1(it) = beta*w(49)*res(50); % 1-step ahead NL
        QP1(it) = beta*w(49)*w(50); % 1-step ahead QP
        resNLP(51) = w(51)-NLP1(it);

        sum1 = 0;
        sum2 = 0;
        sum3 = 0;
        for u = 1:N
            ee = datasample(e,1);
            sum1 = sum1 + beta*w(50)*resNLP(51)+ee; % 2-step ahead NL:
            % 52 conditional on res(51) and w(50)

```



```

        sum2 = sum2 + beta*w(50)*QP1(it)+ee;    % 2-step ahead QP:
                                                % 52 conditional on QP(51) and w(50)
        sum3 = sum3 + ee;
    end
    NLP2(it) = sum1/N;
    QP2(it) = sum2/N;
    LP2(it) = sum3/N;
end % it=1:nrep

mseNLP1 = 0;
mseNLP2 = 0;
mseQP1 = 0;
mseQP2 = 0;
mseLP1 = 0;
mseLP2 = 0;

for it=1:nrep
    mseNLP1 = mseNLP1+(wLtrue1(it)-NLP1(it))^2;
    mseNLP2 = mseNLP2+(wLtrue2(it)-NLP2(it))^2;

    mseQP1 = mseQP1+(wLtrue1(it)-QP1(it))^2;
    mseQP2 = mseQP2+(wLtrue2(it)-QP2(it))^2;

    mseLP1 = mseLP1+wLtrue1(it)^2;    % 1-step ahead LS
                                        % forecasts are zero
    mseLP2 = mseLP2+(wLtrue2(it)-LP2(it))^2;
end

RMSE1(b) = mseNLP1/mseQP1;    % Nonlinear vs. QP
RMSE2(b) = mseNLP2/mseQP2;

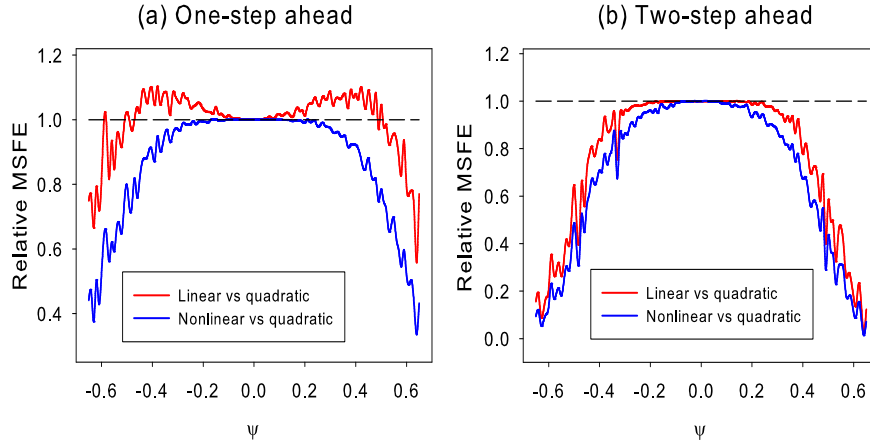
RMSLP1(b) = mseLP1/mseQP1;    % Linear vs. QP
RMSLP2(b) = mseLP2/mseQP2;

end % end beta loop

% Relative Mean Square Errors
RMSE_NLP_QP1 = [RMSE1;index]';
RMSE_NLP_QP2 = [RMSE2;index]';

RMSLP_LP_QP1 = [RMSLP1;index]';
RMSLP_LP_QP2 = [RMSLP2;index]';

```



**Figure 10.5:** One- and two-step ahead relative MSFEs for the BL model  $Y_t = \psi Y_{t-2} \varepsilon_{t-1} + \varepsilon_t$  with  $\{\varepsilon_t\} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ .

## Chapter 11

**11.1** Below we will make use of the properties (rules) for matrix norms given in Appendix 3.A of the main text for  $p = 2$ , and write  $\|\cdot\| = \|\cdot\|_2$ . From (11.14) the vector of innovations is given by

$$\widehat{\varepsilon}_t = \mathbf{Y}_t - \sum_{i=1}^p \Phi^i \mathbf{Y}_{t-i} - \left\{ \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u} \otimes \mathbf{I}_m] \right\} \widehat{\varepsilon}_{t-v}, \quad (v \in \{1, \dots, q\}; q \leq p),$$

so that the vector of reconstruction errors can be expressed as

$$\mathbf{e}_t = \varepsilon_t - \widehat{\varepsilon}_t = - \left\{ \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u} \otimes \mathbf{I}_m] \right\} \varepsilon_{t-v}, \quad (v \in \{1, \dots, q\}; q \leq p).$$

Given a vector of initial values  $\varepsilon_0$ , we get after  $s$  recursions

$$\varepsilon_t = (-1)^s \prod_{j=1}^s \left\{ \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u-(j-1)v} \otimes \mathbf{I}_m] \right\} \varepsilon_0.$$

Taking the norm, we get

$$\|\mathbf{e}_t\| \leq \prod_{j=1}^s \left\| \left\{ \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u-(j-1)v} \otimes \mathbf{I}_m] \right\} \right\| \|\varepsilon_0\|, \quad \text{by rule (A.9).}$$

Similarly, for the second moment of the process  $\{\mathbf{e}_t, t \in \mathbb{Z}\}$  we get

$$\begin{aligned} \|\mathbf{e}_t\|^2 &\leq \left\{ \prod_{j=1}^s \left\| \left\{ \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u-(j-1)v} \otimes \mathbf{I}_m] \right\} \right\| \|\mathbf{e}_0\| \right\}^2 \\ &\leq \left\{ \prod_{j=1}^s \left\| \left\{ \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u-(j-1)v} \otimes \mathbf{I}_m] \right\} \right\| \right\}^2 \|\mathbf{e}_0\|^2, \end{aligned}$$

where we used rule (A.10) in Appendix 3.A, and the following two properties: (i)  $\|\mathbf{A}'\| = \|\mathbf{A}\|$ , since  $\mathbf{A}$  and  $\mathbf{A}'$  have the same set of eigenvalues, and (ii)  $\|\mathbf{xx}'\| = \|\mathbf{x}\|^2$ , since  $\mathbf{xx}'$  has rank 1 and  $\mathbf{xx}' \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{x}'\mathbf{x} = \mathbf{x}\|\mathbf{x}\|^2$ ; thus the only non-zero eigenvalue of  $\mathbf{xx}'$  equals  $\|\mathbf{x}\|^2$ .

Now, we need to show that  $\|\mathbf{e}_t\|$  converges to zero in probability. Let

$$P_{s,t} = \prod_{j=1}^s \left\| \left\{ \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u-(j-1)v} \otimes \mathbf{I}_m] \right\} \right\|, \text{ and}$$

$$L_{j,t} = \log \left\| \left\{ \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u-(j-1)v} \otimes \mathbf{I}_m] \right\} \right\|.$$

Recall the following ergodic theorem (cf. Hannan, 1970; Hannan and Deistler, 2012, p. 203) for a general strictly stationary process.

(i) If  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  has finite first moments (i.e.  $\mathbb{E}\{|\mathbf{Y}_{j,t}|\} < \infty$ ), then

$$\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \mathbf{Y}_t = \mathbb{E}(\mathbf{Y}_t), \text{ a.s.}$$

(ii) If  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  has finite second moments (i.e.  $\mathbb{E}(Y_{j,t}^2) < \infty$ ), then (for  $s \geq 0$ )

$$\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \mathbf{Y}_{t+s} \mathbf{Y}'_t = \mathbb{E}(\mathbf{Y}_t \mathbf{Y}'_s), \text{ a.s.}$$

For proofs see, e.g., Rosanov (1967).

The above results extend to the following case. If  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  is ergodic and for any real-valued vector function  $f(\mathbf{y})$  with  $\mathbb{E}\{|f(\mathbf{Y}_{j,t})|\} < \infty$ , then with probability one,

$$\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T f(\mathbf{Y}_t) = \mathbb{E}\{f(\mathbf{Y}_t)\}.$$

Now, by the ergodic theorem, we get

$$\lim_{s \rightarrow \infty} s^{-1} \log P_{s,t} = \lim_{s \rightarrow \infty} s^{-1} \sum_{j=1}^s L_{s,t} = \mathbb{E} \left\{ \log \left\| \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u} \otimes \mathbf{I}_m] \right\| \right\} < 0,$$

which implies  $\lim_{s \rightarrow \infty} P_{s,t} = 0$  a.s. This follows from  $\log (P_{s,t})^{s^{-1}} \rightarrow \alpha < 0$ . So  $(P_{s,t})^{s^{-1}} \rightarrow \exp(\alpha) < 1$ . Choose any  $\epsilon > 0$ . If  $P_{s,t} > \epsilon$ , then  $\log (P_{s,t})^{s^{-1}} \rightarrow 1$ . Hence,  $P_{s,t} < \epsilon$  for  $s$  large enough.

The above result means that  $\lim_{t \rightarrow \infty} \|\mathbf{e}_t\| = 0$ , and by the Lebesgue dominated convergence theorem (cf. Rao (1973, p. 136)) we have  $\lim_{t \rightarrow \infty} \mathbb{E}\|\mathbf{e}_t\| = \mathbb{E}\{\lim_{t \rightarrow \infty} \|\mathbf{e}_t\|\} = 0$ . Next, using a similar argument gives  $\lim_{t \rightarrow \infty} \mathbb{E}\|\mathbf{e}_t\|^2 = \mathbb{E}\{\lim_{t \rightarrow \infty} \|\mathbf{e}_t\|^2\} = 0$ . Thus, the vector of reconstruction errors  $\mathbf{e}_t$  satisfies the invertibility condition; cf. condition (3.31) in the univariate case.

By using Jensen's inequality, we obtain

$$\exp \mathbb{E} \left\{ \log \left\| \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u} \otimes \mathbf{I}_m] \right\| \right\} \leq \mathbb{E} \left\{ \left\| \Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u} \otimes \mathbf{I}_m] \right\| \right\} < 1.$$

Clearly, this last condition is similar to condition (11.13) in the main text for the general multivariate BL model. In fact, this result also follows from similar arguments used by Granger and Andersen (1978a,b), i.e., the geometric mean of a set of positive numbers is less than or equal to the arithmetic mean.

Next, we consider a more useful condition for invertibility. In particular, using Jensen's inequality, the Cauchy–Schwarz inequality, rule (A.12) in Appendix 3.A, the property  $\|\mathbf{x} \otimes \mathbf{I}_m\| = \|\mathbf{x}\|$ , and the strict stationarity of  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ , we have

$$\begin{aligned} \log \exp \mathbb{E} \left\{ \log \|\Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u} \otimes \mathbf{I}_m]\| \right\} &\leq \log \mathbb{E} \|\Theta^v + \sum_{u=1}^p \Psi^{uv} [\mathbf{Y}_{t-u} \otimes \mathbf{I}_m]\| \\ &\leq \log \left\{ \|\Theta^v\| + \sum_{u=1}^p \|\Psi^{uv}\| \mathbb{E} \|\mathbf{Y}_t\| \right\} \\ &\leq \log \left\{ \|\Theta^v\| + \sum_{u=1}^p \|\Psi\| \sqrt{\mathbb{E} \|\mathbf{Y}_t\|^2} \right\}. \end{aligned}$$

Hence, the result in (11.16) follows.

**Remark:** In Exercise 6.2 we showed that it is difficult to distinguish a diagonal BL(0, 0, 1, 1) process from an MA(1) process using the ACF. Similarly, when two variables are generated by a bivariate diagonal BL process and examined with the ACF and CCF, there will be no indication that the DGP is nonlinear. For instance, suppose the processes  $\{X_t, t \in \mathbb{Z}\}$  and  $\{Y_t, t \in \mathbb{Z}\}$  are generated by

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} + \begin{pmatrix} 0 & \tau_1 \varepsilon_{1,t-1} \\ \tau_2 \varepsilon_{2,t-1} & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix}, \quad (*)$$

where  $\{\varepsilon_{i,t}\} \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_{i,\varepsilon}^2)$  ( $i = 1, 2$ ), and  $\varepsilon_{1,t}$  is independent of  $\varepsilon_{2,t}$ . For this model we have

$$Y_t = \varepsilon_{1,t} + \tau_1 \varepsilon_{1,t-1} \varepsilon_{2,t-1} + \tau_1 \tau_2 \varepsilon_{1,t-1} \varepsilon_{2,t-2} Y_{t-2},$$

so  $\mathbb{E}(Y_t) = 0$ . Next,

$$\sigma_Y^2 = \frac{\sigma_{1,\varepsilon}^2 + \tau_1^2 \sigma_{1,\varepsilon}^2 \sigma_{2,\varepsilon}^2}{1 - \tau_1^2 \tau_2^2 \sigma_{1,\varepsilon}^2 \sigma_{2,\varepsilon}^2} = \frac{\sigma_{1,\varepsilon}^2 (1 + \lambda_1^2)}{1 - \lambda^2},$$

where  $\lambda = \lambda_1 \lambda_2$ ,  $\lambda_1 = \tau_1 \sigma_{2,\varepsilon}$ , and  $\lambda_2 = \tau_2 \sigma_{1,\varepsilon}$ . Thus, for the processes  $\{X_t, t \in \mathbb{Z}\}$  and  $\{Y_t, t \in \mathbb{Z}\}$  to be *non-explosive* we need the condition  $\lambda^2 < 1$ .

First we consider the invertibility condition. Let  $e_{1,t} = \varepsilon_{1,t} - \widehat{\varepsilon}_{1,t}$  denote the reconstruction errors, where  $\widehat{\varepsilon}_{1,t}$  is generated from arbitrary values,  $\widehat{\varepsilon}_{1,0}$  and  $\widehat{\varepsilon}_{2,0}$ , and a sample of  $X_t$  and  $Y_t$ ,  $t \geq 0$ , using (\*). Subtracting  $Y_t = \widehat{\varepsilon}_{1,t} + \tau_1 \widehat{\varepsilon}_{1,t-1} X_{t-1}$  from the first equation in (\*) yields  $e_{1,t} = -\tau_1 X_{t-1} e_{1,t-1}$ . As this is like (8.34) in Granger and Andersen (1978a), we can use their result (8.42) that  $\tau_1^2 \mathbb{E}(X_t^2) < 1$  for invertibility. This gives the *invertibility condition*  $(\lambda_2(1 + \lambda_1^2)/(1 - \lambda)) < 1$  for  $\{X_t, t \in \mathbb{Z}\}$ , and likewise  $(\lambda_1^2(1 + \lambda_2^2)/(1 - \lambda^2)) < 1$  for  $\{Y_t, t \in \mathbb{Z}\}$ .

It is easy to see that the autocovariances of the two processes are zero, i.e.,  $\gamma_X(\ell) = 0 \forall \ell > 0$ , and  $\gamma_Y(\ell) = 0 \forall \ell > 0$ . The cross-covariances of the two processes are zero

except at zero lag. Then,

$$\begin{aligned}\mathbb{E}(X_t Y_t) &= \tau_1 \tau_2 \mathbb{E}(\varepsilon_{1,t} \varepsilon_{2,t} X_t Y_t) \\ &= \tau_1 \tau_2 (\mathbb{E}(\varepsilon_{1,t}^2 \varepsilon_{2,t}^2) + \tau_1 \mathbb{E}(\varepsilon_{1,t} \varepsilon_{2,t}^2 \varepsilon_{1,t-1} X_{t-1}) \\ &\quad + \mathbb{E}(\varepsilon_{1,t}^2 \varepsilon_{2,t} \varepsilon_{2,t-1} Y_{t-1}) + \tau_1 \tau_2 \mathbb{E}(\varepsilon_{1,t} \varepsilon_{2,t} + \varepsilon_{2,t-1} Y_{t-1} \varepsilon_{1,t-1} X_{t-1})) \\ &= \tau_1 \tau_2 (\sigma_{1,\varepsilon}^2 \sigma_{2,\varepsilon}^2 + 0 + 0 + 0) = \tau_1 \tau_2 \sigma_{1,\varepsilon}^2 \sigma_{2,\varepsilon}^2.\end{aligned}$$

If we define  $\rho_{X,Y}(\ell) = \mathbb{E}((X_t - \bar{X})(Y_{t-\ell} - \bar{Y}))/\sigma_X \sigma_Y$ , then the CCF of (\*) is given by

$$\rho_{X,Y}(\ell) = \begin{cases} \frac{\tau_1 \tau_2 \sigma_{1,\varepsilon} \sigma_{2,\varepsilon} (1 - \lambda^2)}{\sqrt{(1 + \lambda_1^2)(1 + \lambda_2^2)}}, & \ell = 0, \\ 0, & \ell \neq 0. \end{cases}$$

So, by examining the ACF and CCF of  $\{X_t, t \in \mathbb{Z}\}$  and  $\{Y_t, t \in \mathbb{Z}\}$  we might assume incorrectly that these processes were WN with instantaneous causality.

It is easily seen that using the correct model (\*), gives forecasting advantage over a WN model. However, the advantage extends only to one-step ahead forecasts, the two and longer step forecasts being identical. To see this, for model (\*) the  $H$ -step ahead forecast at time  $t$  is given by

$$\mathbb{E}(Y_{t+H}|t) = \begin{cases} \tau_1 \varepsilon_{1,t} X_t, & H = 1, \\ 0, & H > 1. \end{cases}$$

Define  $e_{1,t+H|t} = Y_{t+H} - \mathbb{E}(Y_{t+H}|t)$ , the  $H$ -step ahead forecast error. Then  $e_{1,t+1|t} = \varepsilon_{1,t+1}$  and  $e_{1,t+H|t} = Y_{t+H}$  if  $H > 1$ . For the WN model, the  $H$ -step ahead forecast error is given by  $e'_{1,t+1|t} = Y_{t+1}$  and  $e'_{1,t+H|t} = e'_{1,t+1|t}$  if  $H > 1$ . The loss from using the WN model for one-step ahead forecasts can be defined as

$$\text{loss} = \frac{\mathbb{E}((e'_{t+1|t})^2)}{\mathbb{E}((e_{t+1|t})^2)} = \frac{\sigma_Y^2}{\sigma_{1,\varepsilon}^2} = \frac{1 + \lambda_1^2}{1 - \lambda_1^2 \lambda_2^2}.$$

Since for invertibility we must have  $\tau_1^2 \sigma_X^2 < 1$  and  $\tau_2^2 \sigma_Y^2 < 1$ , then

$$1 \leq \text{loss} < \frac{1}{\sigma_{1,\varepsilon}^2 \tau_2^2} < \frac{1}{\lambda_2^2} \quad \text{for } \{Y_t, t \in \mathbb{Z}\}.$$

The third moments and the cross third moments can be used to show that the two processes are not instantaneously related WN. Another set of statistics that will distinguish the two processes from WN with instantaneous causality are the ACF and CCF of  $\{X_t^2, t \in \mathbb{Z}\}$  and  $\{Y_t^2, t \in \mathbb{Z}\}$ . Explicit expressions for these functions can be obtained routinely.

- 11.2** (a) Let  $U \sim \mathcal{N}_1(0, \sigma_{11})$  and  $\mathbf{V} \sim \mathcal{N}_{m-1}(\mathbf{0}, \mathbf{\Sigma}_{22})$ . Then it is well known from the theory of multivariate statistical analysis (see, e.g., [Johnson and Wichern, 2002](#)) that the conditional distribution of  $\mathbf{V}$  given that  $U = u$ , is normal and has

$$\text{Mean} = \mathbf{\Sigma}_{12} \sigma_{11}^{-1} u \quad \text{and} \quad \text{Covariance} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12} \sigma_{11}^{-1} \mathbf{\Sigma}'_{12}.$$

Denote the corresponding conditional density function by  $g_{m-1}(\cdot|u)$ .

(i) Since  $f_m(\mathbf{v}) = g_{m-1}(\mathbf{v}|u)f_1(u)$  with  $\mathbf{v} = (v_1, \dots, v_m)'$ , we have

$$\begin{aligned} \int_{A_i} x_1 f_m(\mathbf{x}) d\mathbf{x} &= \int_{A_i} u f_1(u) g_{m-1}(\mathbf{v}|u) du d\mathbf{v} \\ &= \int_{\mathbb{R}^{(i)}} u f_1(u) \left( \int_{\mathbb{R}^{m-1}} g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right) du = \mu_i = \sigma_{11} \frac{\mu_1}{\sigma_{11}}. \end{aligned}$$

In a similar way, for  $i > 1$ ,

$$\begin{aligned} \int_{A_i} x_j f_m(\mathbf{x}) d\mathbf{x} &= \int_{\mathbb{R}^{(i)}} f_1(u) \left( \int_{\mathbb{R}^{m-1}} v_j g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right) du \\ &= \int_{\mathbb{R}^{(i)}} f_1(u) \frac{\sigma_{j1}}{\sigma_{11}} u du = \frac{\sigma_{j1}}{\sigma_{11}} \mu_i. \end{aligned}$$

(ii) Clearly, for  $k = j = 1$  the term on the right-hand side becomes  $\sigma_i - \sigma_{11}\alpha_i + \sigma_{11}\alpha_i = \sigma_i$ . So similar to part (i), we have

$$\begin{aligned} \int_{A_i} x_1^2 f_m(\mathbf{x}) d\mathbf{x} &= \int_{A_i} u^2 f_1(u) g_{m-1}(\mathbf{v}|u) du d\mathbf{v} \\ &= \int_{\mathbb{R}^{(i)}} u^2 f_1(u) \left( \int_{\mathbb{R}^{m-1}} g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right) du = \sigma_i. \end{aligned}$$

For  $k = 1, j \neq 1$ , the term on the right-hand side becomes  $(\sigma_{j1}/\sigma_{11})\sigma_i$ . Again, this result follows in a similar way as above, i.e.

$$\begin{aligned} \int_{A_i} x_1 x_j f_m(\mathbf{x}) d\mathbf{x} &= \int_{A_i} u f_1(u) v_j g_{m-1}(\mathbf{v}|u) du d\mathbf{v} \\ &= \int_{\mathbb{R}^{(i)}} u f_1(u) \left( \int_{\mathbb{R}^{m-1}} v_j g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right) du = \int_{\mathbb{R}^{(i)}} u \frac{\sigma_{j1}}{\sigma_{11}} u du = \frac{\sigma_{j1}}{\sigma_{11}} \sigma_i. \end{aligned}$$

For  $k \neq 1, j \neq 1$ , we have

$$\begin{aligned} \int_{A_i} x_k x_j f_m(\mathbf{x}) d\mathbf{x} &= \int_{A_i} f_1(u) \left( \int_{\mathbb{R}^{m-1}} v_k v_j g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right) du \\ &= \int_{\mathbb{R}^{(i)}} f_1(u) \left\{ \int_{\mathbb{R}^{m-1}} \left( (v_k - \frac{\sigma_{k1}}{\sigma_{11}}u)(v_j - \frac{\sigma_{j1}}{\sigma_{11}}u) \right. \right. \\ &\quad \left. \left. + \frac{\sigma_{k1}}{\sigma_{11}}u v_j + \frac{\sigma_{j1}}{\sigma_{11}}u v_k - u^2 \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}^2} \right) g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right\} du \\ &= \int_{\mathbb{R}^{(i)}} f_1(u) \left( \sigma_{kj} - \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}} + \frac{\sigma_{k1}}{\sigma_{11}}u \frac{\sigma_{j1}}{\sigma_{11}}u + \frac{\sigma_{j1}}{\sigma_{11}}u \frac{\sigma_{k1}}{\sigma_{11}}u - u^2 \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}^2} \right) du \\ &= \left( \sigma_{kj} - \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}} \right) \alpha_i + \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}^2} \sigma_i = \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}} \left( \frac{\sigma_i}{\sigma_{11}} - \alpha_i \right) + \sigma_{kj} \alpha_i. \end{aligned}$$

(b) (i) Using result (i) in part (a), we have

$$\begin{aligned} \int_{A_i} \mathbf{r}' \mathbf{x} f_m(\mathbf{x}) d\mathbf{x} &= \int_{A_i} \sum_{j=1}^m r_j x_j f_m(\mathbf{x}) d\mathbf{x} \\ &= \sum_{j=1}^m r_j \int_{A_i} x_j f_m(\mathbf{x}) d\mathbf{x} = \frac{\mu_i}{\sigma_{11}} \mathbf{r}' \boldsymbol{\Sigma}_{12}^*, \end{aligned}$$

with  $\boldsymbol{\Sigma}_{12}^* = (\sigma_{11}, \boldsymbol{\Sigma}_{12})'$ .

(ii) Using result (ii) in part (a), we have

$$\begin{aligned}\int_{A_i} \mathbf{r}'\mathbf{xx}'\mathbf{s}f_m(\mathbf{x})d\mathbf{x} &= \int_{A_i} \left(\sum_{j=1}^m r_j x_j\right) \left(\sum_{k=1}^m s_k x_k\right) f_m(\mathbf{x})d\mathbf{x} \\ &= \sum_{j=1}^m \sum_{k=1}^m r_j s_k (\sigma_{kj}\sigma_{j1}\gamma_i + \sigma_{k1}\alpha_i) \\ &= \gamma_i \mathbf{r}'\boldsymbol{\Sigma}_{12}^* \boldsymbol{\Sigma}_{12}^{*\prime} \mathbf{s} + \alpha_i \mathbf{r}'\boldsymbol{\Sigma}_X \mathbf{s}.\end{aligned}$$

(c) To obtain the required moments, we introduce an  $m \times m$  orthogonal matrix  $\mathbf{Q}$  with first column given by the  $m$ -dimensional vector  $\mathbf{1}$ . Applying the nonsingular linear transformation  $\mathbf{Q}'\mathbf{x}$  to  $\mathbf{x}$  with Jacobian  $|\mathbf{Q}| \neq 0$ , we have

$$\mathbf{Q}'\mathbf{x} \in A_i \iff \mathbf{x} \in M_i, \quad (i = 1, 2).$$

Let  $\psi(\mathbf{x})$  be the density function of  $\mathcal{N}_m(\mathbf{0}, \boldsymbol{\Sigma}_X)$ , and  $\varphi(\mathbf{z})$  the density function of  $\mathcal{N}_m(\mathbf{0}, \mathbf{Q}'\boldsymbol{\Sigma}_X\mathbf{Q})$ . Then, given the above result, there exists a bounded continuous function  $g(\cdot)$  such that

$$\int_{M_i} g(\mathbf{x})\psi(\mathbf{x})d\mathbf{x} = \int_{\xi(M_i)} g(\xi^{-1}(\mathbf{x}))\psi(\xi^{-1}(\mathbf{x}))d\mathbf{x} = \int_{A_i} g(\mathbf{Qz})\varphi(\mathbf{z})d\mathbf{z}, \quad (i = 1, 2),$$

with  $\xi: \mathbf{X} \rightarrow \mathbf{Q}'\mathbf{X}$ . Hence, the first moment of  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  is given by

$$\mathbb{E}(\mathbf{Y}_t) = \mathbb{E}(\mathbf{X}_t) + \sum_{i=1}^2 \int_{M_i} \mathbf{B}_i \mathbf{x} \psi(\mathbf{x})d\mathbf{x} = \sum_{i=1}^2 \int_{A_i} \mathbf{B}_i \mathbf{Qz} \varphi(\mathbf{z})d\mathbf{z}.$$

Note, the covariance matrix  $\mathbf{Q}'\boldsymbol{\Sigma}_X\mathbf{Q}$  has as first column  $\mathbf{Q}'\boldsymbol{\Sigma}_X\mathbf{1}$ . Then, using result (i) in part (b), we have

$$\begin{aligned}\mathbb{E}(\mathbf{Y}_t) &= \sum_{i=1}^2 \frac{\mu_i}{\mathbf{1}'\boldsymbol{\Sigma}_X\mathbf{1}} \mathbf{B}_i \mathbf{Q} \mathbf{Q}' \boldsymbol{\Sigma}_X \mathbf{1} = \frac{1}{\mathbf{1}'\boldsymbol{\Sigma}_X\mathbf{1}} \sum_{i=1}^2 \mu_i \mathbf{B}_i \boldsymbol{\Sigma}_X \mathbf{1} \\ &= \frac{1}{\mathbf{1}'\boldsymbol{\Sigma}_X\mathbf{1}} \mu_2 (\mathbf{B}_2 - \mathbf{B}_1) \boldsymbol{\Sigma}_X \mathbf{1} = (2\pi \mathbf{1}'\boldsymbol{\Sigma}_X\mathbf{1})^{-1/2} \mathbf{B} \boldsymbol{\Sigma}_X \mathbf{1},\end{aligned}$$

since  $-\mu_1 = \mu_2 = (\mathbf{1}'\boldsymbol{\Sigma}_X\mathbf{1}/2\pi)^{1/2}$ .

The variance of  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  is given by

$$\text{Var}(\mathbf{Y}_t) = \mathbb{E}(\mathbf{X}_t \mathbf{X}_t') - \mathbb{E}(\mathbf{Y}_t) (\mathbb{E}(\mathbf{Y}_t))'.$$

Using result (ii) of part (b), the first term on the right-hand side can be written as

$$\begin{aligned}\mathbb{E}(\mathbf{X}_t \mathbf{X}_t') &= \boldsymbol{\Sigma}_X + \sum_{i=1}^2 \int_{M_i} \mathbf{B}_i \mathbf{xx}' \mathbf{B}_i' \psi(\mathbf{x})d\mathbf{x} = \boldsymbol{\Sigma}_X + \sum_{i=1}^2 \int_{A_i} \mathbf{B}_i \mathbf{Qzz}' \mathbf{Q}' \mathbf{B}_i' \varphi(\mathbf{z})d\mathbf{z} \\ &= \boldsymbol{\Sigma}_X + \sum_{i=1}^2 \gamma_i \mathbf{B}_i \boldsymbol{\Sigma}_X \mathbf{1} \mathbf{1}' \boldsymbol{\Sigma}_X' \mathbf{B}_i' + \sum_{i=1}^2 \alpha_i \mathbf{B}_i \boldsymbol{\Sigma}_X \mathbf{B}_i' \\ &= \boldsymbol{\Sigma}_X + \frac{1}{2} (\mathbf{B}_1 \boldsymbol{\Sigma}_X \mathbf{B}_1' + \mathbf{B}_2 \boldsymbol{\Sigma}_X \mathbf{B}_2'),\end{aligned}$$

since  $\alpha_1 = \alpha_2 = 1/2$ , and

$$\gamma_i = \frac{\sigma_i}{\sigma_{11}^2} - \frac{\alpha_i}{\sigma_{11}} = \frac{\mathbf{1}'\boldsymbol{\Sigma}_X\mathbf{1}}{2} \frac{1}{(\mathbf{1}'\boldsymbol{\Sigma}_X\mathbf{1})^2} - \frac{1/2}{\mathbf{1}'\boldsymbol{\Sigma}_X\mathbf{1}} = 0, \quad (i = 1, 2).$$

It is easy to see that the covariance matrix at lag 1 is given by

$$\begin{aligned} \text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-1}) &= \mathbb{E}\left(\sum_{i=1}^2 \mathbf{B}_i I(\mathbf{X}_{t-1} \in M_i) \mathbf{X}_{t-1} \mathbf{X}'_{t-1}\right) \\ &\quad - \mathbb{E}\left(\sum_{i=1}^2 \mathbf{B}_i I(\mathbf{X}_{t-1} \in M_i) \mathbf{X}_{t-1}\right) (\mathbb{E}(\mathbf{X}_{t-1}))' \\ &= \sum_{i=1}^2 \int_{M_i} \mathbf{B}_i \mathbf{x} \mathbf{x}' \psi(\mathbf{x}) d\mathbf{x} = \sum_{i=1}^2 \int_{A_i} \mathbf{B}_i \mathbf{Q} \mathbf{z} \mathbf{z}' \mathbf{Q}' \varphi(\mathbf{z}) d\mathbf{z} \\ &= \left(\sum_{i=1}^2 \gamma_i \mathbf{B}_i\right) \boldsymbol{\Sigma}_X \mathbf{1} \mathbf{1}' \boldsymbol{\Sigma}_X + \left(\sum_{i=1}^2 \alpha_i \mathbf{B}_i\right) \boldsymbol{\Sigma}_X \\ &= \frac{1}{2} (\mathbf{B}_1 + \mathbf{B}_2) \boldsymbol{\Sigma}_X. \end{aligned}$$

Clearly, when  $\boldsymbol{\Sigma}_X = \sigma_\varepsilon^2 \mathbf{I}_m$ , the above expressions reduce to, respectively,

$$\begin{aligned} \mathbb{E}(\mathbf{Y}_t) &= \frac{\sigma_\varepsilon}{\sqrt{2\pi}} \mathbf{B} \mathbf{1}, \quad \text{Var}(\mathbf{Y}_t) = \sigma_\varepsilon^2 \left( \mathbf{I}_m + \frac{1}{2} (\mathbf{B}_1 \mathbf{B}'_1 + \mathbf{B}_2 \mathbf{B}'_2) - \frac{1}{2\pi} \mathbf{B} \mathbf{1} \mathbf{1}' \mathbf{B}' \right), \\ \text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-1}) &= \frac{\sigma_\varepsilon^2}{2} (\mathbf{B}_1 + \mathbf{B}_2). \end{aligned}$$

Moreover, in the case  $m = 1$  (a univariate asMA(1) process) and  $\sigma_\varepsilon^2 = 1$ , the expressions for  $\mathbb{E}(\mathbf{Y}_t)$  and  $\text{Var}(\mathbf{Y}_t)$  reduce to the mean and variance given in Exercise 2.4(a) of the main text.

- 11.3** (a) Assume some standard regularity conditions are satisfied. Then, under  $\mathbb{H}_0: \boldsymbol{\Theta}_1 = \mathbf{0}$  in (11.69), the LM-type test statistic follows from the score matrix

$$\frac{\partial \log L_T(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\Theta}_1} = \sum_{t=1}^T \left\{ \mathbf{z}_t' X_t (\mathbf{Y}_t - \hat{\boldsymbol{\Theta}}_0' \mathbf{z}_t)' \hat{\boldsymbol{\Sigma}}_\varepsilon^{-1} \right\} = \mathbf{U}' (\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}}_1) \hat{\boldsymbol{\Sigma}}_\varepsilon^{-1},$$

where  $\hat{\boldsymbol{\theta}}$ ,  $\hat{\boldsymbol{\Theta}}_0 = \hat{\mathbf{B}}_1$  (an  $(mp + 1) \times m$  matrix) and  $\hat{\boldsymbol{\Sigma}}_\varepsilon$  (an  $m \times m$  matrix) are parameter estimates under the null hypothesis  $\mathbb{H}_0: \boldsymbol{\Theta}_1 = \mathbf{0}$ . This expression can be rewritten as

$$\begin{aligned} \mathbf{Q} &\equiv \mathbf{U}' (\mathbf{Y} - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}) \hat{\boldsymbol{\Sigma}}_\varepsilon^{-1} \\ &= \mathbf{U}' (\mathbf{I}_T - \mathbf{P}_X) (\mathbf{X} \mathbf{B}_1 + \mathbf{E}) \hat{\boldsymbol{\Sigma}}_\varepsilon^{-1} \\ &= \mathbf{U}' (\mathbf{I}_T - \mathbf{P}_X) \mathbf{E} \hat{\boldsymbol{\Sigma}}_\varepsilon^{-1}, \end{aligned}$$

where  $\mathbf{E} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_T)'$  is a  $T \times m$  matrix.

- (b) Under the null hypothesis,  $\hat{\boldsymbol{\Sigma}}_\varepsilon$  converges to  $\boldsymbol{\Sigma}_\varepsilon$  in probability. Further, because  $\text{vec}(\mathbf{E}') \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_T \otimes \boldsymbol{\Sigma}_\varepsilon)$ , it follows that the random matrix  $\mathbf{Q}$  converges in probability to a matrix normal distribution with mean zero and variance  $\mathbf{U}' (\mathbf{I}_T - \mathbf{P}_X) \mathbf{U} \otimes \boldsymbol{\Sigma}_\varepsilon^{-1}$  conditional on  $\mathbf{X}$  and  $\mathbf{U}$ . Let  $\mathbf{S} \equiv (\mathbf{U}' (\mathbf{I}_T - \mathbf{P}_X) \mathbf{U})^{-1/2} \mathbf{Q} \hat{\boldsymbol{\Sigma}}_\varepsilon^{1/2}$ .



Then, according to the above result,  $\mathbf{S}$  will asymptotically converge to a matrix normal distribution with mean zero and variance  $\mathbf{I}_T \otimes \mathbf{I}_T$ . So, we conclude that the chi-square version of the LM-type test statistic is given by

$$\begin{aligned} \text{LM}_{T,p}^{(1)}(m) &= \text{tr}\{\mathbf{S}'\mathbf{S}\} \\ &= \text{tr}\{\widehat{\Sigma}_\varepsilon^{1/2}\mathbf{Q}'[(\mathbf{U}'(\mathbf{I}_T - \mathbf{P}_X)\mathbf{U})^{-1/2}]'(\mathbf{U}'(\mathbf{I}_T - \mathbf{P}_X))^{-1/2}\mathbf{Q}\widehat{\Sigma}_\varepsilon^{1/2}\} \\ &= \text{tr}\{\widehat{\Sigma}_\varepsilon^{-1}(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}_1)'\mathbf{U}[\mathbf{U}'(\mathbf{I}_T - \mathbf{P}_X)\mathbf{U}]^{-1}\mathbf{U}'(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}_1)\}, \end{aligned}$$

which under the null hypothesis, and as  $T \rightarrow \infty$ , converges in probability to a  $\chi^2$  distribution with  $m(mp + 1)$  degrees of freedom.

- 11.4 (a) Because of independence, the joint distribution function  $F_{U^{(p)}}(u_1, u_2) = u_1 u_2$  and  $F_{U^{(p)}}(u_1, u_2) = F_{U_1}(u_1)F_{U_2}(u_2) = 1$ . Let  $Y_1 = U_1 U_2$ ,  $Y_2 = U_2$ . Since  $y_1 = u_1 u_2$  and  $y_2 = u_2$ , we have  $u_1 = y_1/y_2$  and  $u_2 = y_2$ . The determinant of the Jacobian for the inverse transformation is

$$J = \det\left(\frac{\partial(u_1, u_2)}{\partial(y_1, y_2)}\right) = \begin{vmatrix} \frac{1}{y_2} & -\frac{y_1}{y_2^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{y_2}.$$

Accordingly, the joint density function of  $(Y_1, Y_2)'$  is given by

$$f_{Y_1 Y_2} = f\left(\frac{y_1}{y_2}, y_2\right) \times \frac{1}{y_2} = \frac{1}{y_2}, \quad 0 < y_1 < y_2 < 1.$$

Since  $Y_1$  is the random variable of interest, integrating  $Y_2$  out of  $f_{Y_1 Y_2}$  over the permissible range gives

$$f_{Y_1}(y_1) = \int_{y_1}^1 y_2^{-1} dy_2 = [\log y_2]_{y_1}^1 = -\log y_1.$$

Hence, the distribution function is given by  $F_{Y_1}(y_1) = y_1 - y_1 \log(y_1)$  if  $0 < y_1 < 1$ .

- (b) Let  $Y_1 = U_1/U_2$ ,  $Y_2 = U_2$ . Since  $y_1 = u_1/u_2$  and  $y_2 = u_2$ , we have  $u_1 = y_1 y_2$  and  $u_2 = y_2$ . It is easy to see that in this case the determinant of the Jacobian for the inverse transformation is given by  $J = y_2$ . The joint density function of  $(Y_1, Y_2)'$  is then

$$f_{Y_1 Y_2} = f\left(\frac{y_1}{y_2}, y_2\right) \times y_2 = y_2, \quad 0 < y_1 < \infty, 0 < y_2 < 1.$$

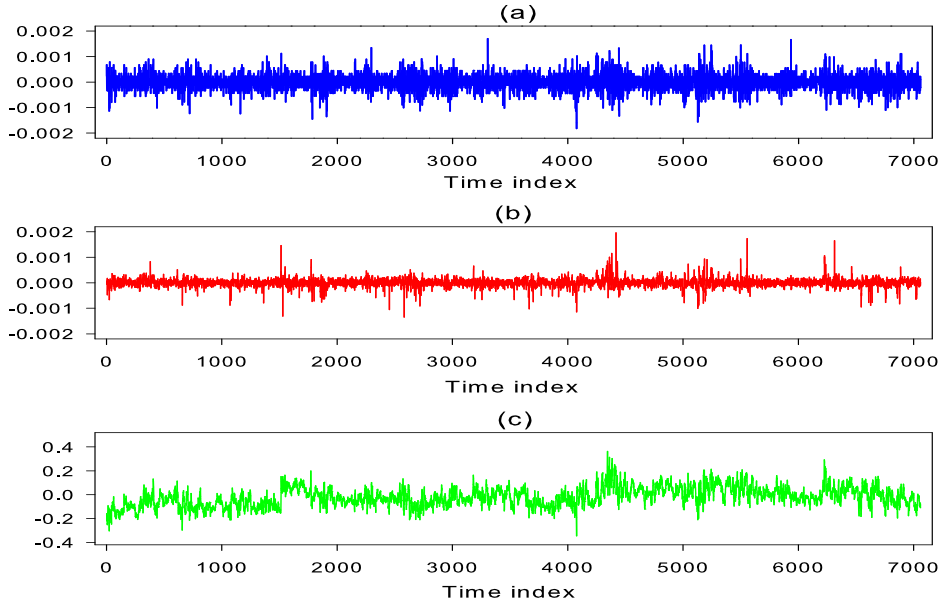
Since  $Y_1$  is the random variable of interest, integrating  $Y_2$  out of  $f_{Y_1, Y_2}$  over the permissible range gives

$$f_{Y_1}(y_1) = \begin{cases} \int_{y_1}^1 y_2 dy_2 = \frac{1}{2}, & \text{when } y_1 < 1, \\ \int_0^{1/y_1} y_2 dy_2 = \frac{1}{2y_1^2}, & \text{when } y_1 > 1. \end{cases}$$

From this result, we immediately obtain the distribution function given in Exercise 11.4(b).

- (c) Let  $Y_1 = (U_1 - \frac{1}{2})(U_2 - \frac{1}{2})$ ,  $Y_2 = (U_2 - \frac{1}{2})$ . Since  $y_1 = (u_1 - \frac{1}{2})(u_2 - \frac{1}{2})$  and  $y_2 = (u_2 - \frac{1}{2})$ , we have  $u_1 - \frac{1}{2} = y_1/y_2$  and  $u_2 - \frac{1}{2} = y_2$ . The determinant of the Jacobian for the inverse transformation is identical to the one given in part (a) above, i.e.  $J = 1/y_2$ . The joint density of  $(Y_1, Y_2)$  is then

$$f_{Y_1 Y_2} = f\left(\frac{y_1}{y_2} + \frac{1}{2}, y_2 + \frac{1}{2}\right) \times \left|\frac{1}{y_2}\right| = \left|\frac{1}{y_2}\right|, \quad -\frac{1}{2} < y_1 < y_2 < \frac{1}{2}.$$



**Figure 11.1:** Intraday transaction data set: Minute returns of S&P 500 index futures and cash prices and the associated threshold variable in May 1993: (a) first differences of  $\log(\text{future})$  ( $Y_{1,t}$ ), (b) first differences of  $\log(\text{price})$  ( $Y_{2,t}$ ), and (c)  $\{X_t = 100X_t^*\}$ ;  $T = 7,060$ .

Integrating  $Y_2$  out of  $f_{Y_1 Y_2}$  over the permissible range gives the density of  $Y_1$ , i.e.

$$f_{Y_1}(y_1) = \begin{cases} \int_{2y_1}^{1/2} \frac{1}{y_2} dy_2 & y_1 > 1/2, y_2 > 1/2, \\ \int_{-1/2}^{2y_1} -\frac{1}{y_2} dy_2 & y_1 > 1/2, y_2 < 1/2, \\ \int_{-1/2}^{-2y_1} \frac{1}{y_2} dy_2 & y_1 < 1/2, y_2 > 1/2, \\ \int_{-1/2}^{-2y_1} -\frac{1}{y_2} dy_2 & y_1 < 1/2, y_2 < 1/2, \end{cases}$$

$$= \begin{cases} \log(\frac{1}{2}) - \log(2y_1), & y_1 > 0, \\ \log(\frac{1}{2}) - \log(-2y_1), & y_1 < 0. \end{cases}$$

From which we obtain the distribution function as given by Exercise 11.4(c).

- 11.5** (a) Figure 11.1 shows time series plots of the data. Here,  $Y_{1,t} = f_{t,\ell} - f_{t-1,\ell}$  and  $Y_{2,t} = s_t - s_{t-\ell}$ , where  $f_{t,\ell}$  is the log of the index future price at time  $t$  and maturity  $\ell$ , and  $s_t$  is the log of the security index price at time  $t$ . The variable  $X_t^*$  is assumed to be weakly stationary and have a continuous distribution. Its values follow from a version of the so-called cost-of-carry model (see, e.g., Tsay, 2010) which is given by

$$f_{t,\ell} - s_t = (r_{t,\ell} - q_{t,\ell})(\ell - t) + X_t^*,$$

where  $r_{t,\ell}$  is the risk-free premium,  $q_{t,\ell}$  is the dividend yield with respect to the cash price at time  $t$ , and  $(\ell - t)$  is the time to maturity of the futures contract.

The Tukey nonadditivity-type test statistic  $F_{T,p}^{(T)}(m)$ , and the  $F_{T,p}^{(O)}(m)$  test statistic both have zero  $p$ -values. Moreover, the  $p$ -values for the univariate version of

the  $F_{T,p}^{(O)}(m)$  test statistic (Algorithm 5.10) all indicate that the null hypothesis of linearity should be rejected. Also, the test statistic  $C_{T,p}(d, m)$  with  $d = 1$  has a zero  $p$  value. Hence, this test indicates threshold nonlinearity.

**Table 11.1:** CLS estimates of the bivariate VTAR model (11.99) for the intraday transaction data set;  $T = 7,060$ . Blue-typed numbers denote significant parameter values at the 5% nominal level.

	Lower regime ( $i = 1$ )		Upper regime ( $i = 2$ )	
	$Y_{1,t}$	$Y_{2,t}$	$Y_{1,t}$	$Y_{2,t}$
$\phi_0^{(i)}$	0.000	0.000	0.000	-0.000
$Y_{1,t-1}$	-0.061	0.083	-0.038	0.037
$Y_{1,t-2}$	0.000	0.148	0.095	0.111
$Y_{1,t-3}$	0.017	0.121	0.036	0.097
$Y_{1,t-4}$	0.005	0.082	-0.006	0.079
$Y_{1,t-5}$	0.003	0.082	0.001	0.048
$Y_{1,t-6}$	-0.016	0.056	-0.003	0.035
$Y_{1,t-7}$	0.010	0.048	-0.000	0.035
$Y_{1,t-8}$	0.002	0.021	-0.018	0.006
$Y_{2,t-1}$	-0.047	-0.011	-0.022	0.057
$Y_{2,t-2}$	-0.055	-0.083	-0.015	-0.015
$Y_{2,t-3}$	0.064	-0.021	-0.023	-0.007
$Y_{2,t-4}$	0.043	0.021	0.018	0.007
$Y_{2,t-5}$	0.079	0.028	-0.015	0.003
$Y_{2,t-6}$	0.014	0.047	0.058	0.003
$Y_{2,t-7}$	-0.037	0.016	-0.056	0.031
$Y_{2,t-8}$	-0.022	0.034	0.020	0.008
$X_{t-1}$	0.000	0.000	0.000	0.001

- (b) Using a grid search method with 300 points, AIC finds a threshold value  $\hat{r} = -0.024$ . The corresponding number of data points in the two regimes are 3,519 and 3,532, respectively. Table 11.1 contains details of the CLS parameter estimates. It is apparent that spot rates changes ( $Y_{2,t}$ ) are more time dependent than futures prices changes ( $Y_{1,t}$ ) with significant coefficients of  $X_{t-1}$  for  $Y_{2,t}$  in both regimes at the 5% nominal level, but not for  $Y_{1,t}$ . We also see that  $Y_{1,t}$  depends negatively on  $Y_{1,t-1}$ . This is in agreement with the so-called *bid-ask spread bounce*, i.e., a negative first-order serial correlation in the series of observed price changes and no serial correlation beyond lag 1. Finally, in lead-lag terms,  $Y_{1,t}$  leads  $Y_{2,t}$  in a stronger fashion than vice versa. This indicates that there is more market-wide information in the futures market than the cash market.
- (c) The values ( $p$ -values) of the  $LM_{T,p}^{(1)}(2)$  and  $F_{T,p}^{(1)}(2)$  test statistics are respectively 111.80 ( $0.089 \times 10^{-9}$ ) and 3.478 ( $0.121 \times 10^{-9}$ ). Thus, in both cases linearity is rejected in favor of LVSTAR nonlinearity. This further supports the results in part (a) that the series  $\{Y_{1,t}\}$  and  $\{Y_{2,t}\}$  are generated by a nonlinear DGP. In addition, we computed the LVSTAR-version of Wilks' lambda test statistic and Rao's modified  $F$  test statistic. The values ( $p$ -values) are respectively 111.845 ( $0.088 \times 10^{-9}$ ) and 3.505 ( $0.088 \times 10^{-9}$ ). All these test results were obtained

with the MATLAB function `scoretest.m` (available at the website of this book), which has been coded after the following R function.

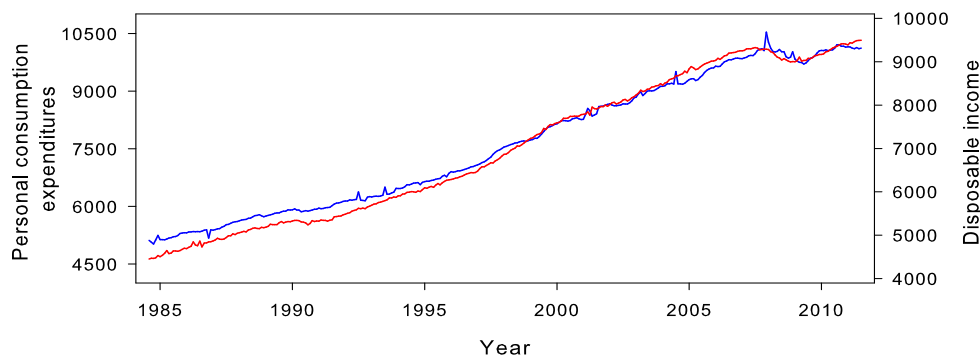
```
SCORETEST <- function(mE, mX, mZ, flag)
#####
# R code: Exercise 11.5(c)
# File: scoretest.r
# Coded by: Yukai Yang
#
# LM-type linearity test statistic versus LVSTAR nonlinearity
# INPUT: mE (matrix of residuals from VAR(p)-fit)
#        mX (matrix of regressors)
#        mZ (auxiliary regression matrix)
# OUTPUT: LM test statistic, first-order approximation
#         F test statistic (rescaled LM-type test statistic)
#         correcting a mistake in the degrees of freedom iDF2
#         Wilks' test statistic
#         Rao's F test statistic
#
# Reference:
# Yang, Y. (2012). Modelling Nonlinear Vector Economic Time Series,
#   PhD thesis, Aarhus University, Denmark. Available at:
#   http://pure.au.dk/portal/files/45638557/Yukai\_Yang\_PhD\_Thesis.pdf.
#####
{
  iT = dim(mE)[1]
  ip = dim(mE)[2]
  ix = dim(mX)[2]
  iz = dim(mZ)[2]
  mE = data.matrix(mE)
  mX = data.matrix(mX)
  mZ = data.matrix(mZ)

  iK = ix + iz
  iDF = iz * ip
  RSS0 = t(mE)%*%mE

# RSS0
  mXX = cbind(mX, mZ)
  mU = svd(mXX)$u
  mR = mE - mU%*%t(mU)%*%mE
  RSS1 = t(mR)%*%mR

# RSS1
  RO = svd(RSS0)$d
  R1 = svd(RSS1)$d
  LM = list()
  FT = list()
  WK = list()
}
```





**Figure 11.2:** Monthly U.S. personal disposable expenditures (blue solid line) and disposable income (red solid line) in millions of dollars for the time period January 1985 – December 2011 ( $T = 324$ ).

and early 2009 with the U.S. subprime mortgage crisis that coincided with the U.S. recession of December 2007 – June 2009. Both series have a steady upward stochastic trend, so it is reasonable to make them stationary by differencing the data. For the VAR( $p$ ) model order, AIC chooses  $p = 4$  and BIC indicates that  $p = 3$ . This latter choice is confirmed by the model selection criteria FPE and HQ (Hannan and Quinn, 1979). Hence, we set  $p = 3$  throughout the analysis.

Below are three R functions, written by Dao Li, to obtain the LSTAR–CNF model forecasts.

```
#####
# R code: Exercise 11.6 (LSTAR–CNF model forecasts)
# File: LSTAR–CNF–forecast.r
# Coded by: Dao Li
#
# Reference:
# Li, D. and He, C. (2013).
#   Forecasting with vector nonlinear time series models.
#   Working papers 2013:8, Dalarna University, Sweden.
#   Available at:
#   http://www.diva-portal.org/smash/get/diva2:606647/FULLTEXT02.pdf.
#####
library("MASS")
rm(list=ls())
myestimation <- function(data,p=3,dn=1,d=7,rank=1){
# Notes: p = 3 (order of LSTAR model);
#       d = 7 (delay of transition variable)
  n <- length(data[,1])
  T <- length(data[1,])
  y <- data
  tempx <- matrix(0,(n*p),(T-max(p,d)))
  for(i in 1:p){
```

```

    tempx[(1+n*(i-1)):(n+n*(i-1)),] <- y[, (max(p,d)-i+1):(T-i)]
  }
  tempy <- y[, (max(p,d)+1):T]
  x <- rbind(rep(1, (T-max(p,d))), tempx)
  st <- y[dn, max(p+1-d, 1):(T-d)]
  st1 <- round(quantile(st, 0.05), 2)
  st2 <- round(quantile(st, 0.95), 2)
# Grid search of (gamma, c)
  r <- seq(30, 60.4, 0.01)
  c <- seq(0.69, 0.75, 0.01)
  len1 <- length(r)
  len2 <- length(c)
  est1 <- matrix(0, len1, len2)
  for(k in 1:len1){ # search for gamma
    for(j in 1:len2){ # search for c
      #####
      # Estimation given one (gamma, c)
      z <- matrix(0, (n*p+1), (T-max(p,d)))
      for(i in 1:(n*p+1)){
        z[i,] <- x[i,]/(1+exp(-r[k]*(st-c[j])/sd(st)))
      }
      X <- rbind(x, z)
      resv <- tempy - tempy%*%t(X)%*%solve(X%*%t(X))%*%X
      Sigv <- resv%*%t(resv)/T
      Sigyz <- tempy%*%t(z)/T
      Sigyx <- tempy%*%t(x)/T
      Sigxx <- x%*%t(x)/T
      Sigxz <- x%*%t(z)/T
      Sigzz <- z%*%t(z)/T
      Sigyxx <- Sigyz - Sigyx%*%solve(Sigxx)%*%Sigxz
      Sigzxx <- Sigzz - t(Sigxz)%*%solve(Sigxx)%*%Sigxz

      tempres <- eigen(x=Sigv, symmetric=TRUE, only.values=FALSE)
      Sigv2 <- tempres$vector%*%diag(sqrt(tempres$values))%*%t(tempres$vector)
      matrix_eigen <- solve(Sigv2)%*%Sigyxx%*%solve(Sigzxx)%*%t(Sigyxx)%*%solve(Sigv2)
      result_eigen <- eigen(x=matrix_eigen, symmetric=TRUE, only.values=FALSE)
      eigen_value <- result_eigen$values
      eigen_vec <- result_eigen$vectors
      eigenvec <- matrix(0, n, rank)
      for(ind in 1:rank){
        eigenvec[,ind] <- eigen_vec[,ind] # each clmn is an eigenvector
        eigenvec[,ind] <- eigenvec[,ind]/sqrt(sum(eigenvec[,ind]^2))
      }
      A <- Sigv2%*%eigenvec # n times rank matrix
      B <- t(eigenvec)%*%solve(Sigv2)%*%Sigyxx%*%solve(Sigzxx)
      Gamma <- A%*%B
      Phi <- (Sigyx - A%*%B%*%t(Sigxz))%*%solve(Sigxx)

      reduced_v <- tempy - Phi%*%x - Gamma%*%z

```

```

normA      <- solve(Sigv2)%%reduced_v%%t(reduced_v)%%solve(Sigv2)/T
estl[k,j]  <- sum(diag(normA))
} # end search for c
} # end search for gamma
inrowcol   <- which(estl==min(estl),arr.ind=TRUE)
estr      <- r[inrowcol[1,1]]
estc      <- c[inrowcol[1,2]]
z         <- matrix(0,(n*p+1),(T-max(p,d)))
for(i in 1:(n*p+1)){
  z[i,]   <- x[i,]/(1+exp(-estr*(st-estc)/sd(st)))
}
X         <- rbind(x,z)
resv     <- tempy - tempy%%t(X)%%solve(X%%t(X),tol=1e-30)%%X
Sigv     <- resv%%t(resv)/T
Sigyz   <- tempy%%t(z)/T
Sigyx   <- tempy%%t(x)/T
Sigxx   <- x%%t(x)/T
Sigxz   <- x%%t(z)/T
Sigzz   <- z%%t(z)/T

Sigyzz  <- Sigyz - Sigyx%%solve(Sigxx)%%Sigxz
Sigzzx  <- Sigzz - t(Sigxz)%%solve(Sigxx)%%Sigxz

tempres  <- eigen(x=Sigv,symmetric=TRUE,only.values=FALSE)
Sigv2    <- tempres$vector%%diag(sqrt(tempres$values))%%t(tempres$vector)
matrix_eigen <- solve(Sigv2)%%Sigyzz%%solve(Sigzzx)%%t(Sigyzz)%%solve(Sigv2)
result_eigen <- eigen(x=matrix_eigen,symmetric=TRUE,only.values=FALSE)
eigen_value <- result_eigen$values
eigen_vec  <- result_eigen$vectors
eigenvec   <- matrix(0,n,rank)
for(ind in 1:rank){
  eigenvec[,ind] <- eigen_vec[,ind]/sqrt(sum(eigen_vec[,ind]^2))
}

A         <- Sigv2%%eigenvec #n times rank matrix
B         <- t(eigenvec)%%solve(Sigv2)%%Sigyzz%%solve(Sigzzx)
Gamma    <- A%%B
Phi      <- (Sigyx - A%%B%%t(Sigxz))%%solve(Sigxx)
reduced_v <- tempy - Phi%%x - Gamma%%z
return(list(gamma=estr,c=estc,estPhi=Phi,estA=A,estB=B,residual=reduced_v))
}

myoneT2forecast <- function(y,result,resB,h=12,j,Nb=1000)
# Notes: Nb = Number of bootstrap replicates;
#         h = Maximum forecast horizon
#         2 variables, delay transition variable = 7
# Do one-step ahead forecast up to h here
{
  oneT2newy <- matrix(0,2,h)

```



```

Tused      <- length(y[1,])
r          <- result$gamma
c          <- result$c
Phi        <- result$estPhi
A          <- result$estA
B          <- result$estB
vhatm      <- matrix(rowMeans(resB),2,1)
vhat       <- resB - matrix(1,1,Tused-7)%x%vhatm
Cfactor    <- t(chol(cov(t(vhat))))
vhat_ind   <- solve(Cfactor)%*%vhat
AB         <- A%*%B
sdst       <- sd(y[1,1:(Tused-7)])
##### h=1
if(h==1){
  tempG <- 1/(1+exp(-r*(y[1,(Tused-7+1)]-c)/sdst))
  oneT2newy[,h] <- Phi[,1]+(Phi[,2:3]+AB[,2:3]*tempG)%*%y[(Tused)]
                + (Phi[,4:5]+AB[,4:5]*tempG)%*%y[(Tused-1)]
                + (Phi[,6:7]+AB[,6:7]*tempG)%*%y[(Tused-2)]
}
##### h>1
if(h>1 && h<=j){
  temp1 <- sample(vhat_ind[1,],size=Nb*h,replace=TRUE)
  temp2 <- sample(vhat_ind[2,],size=Nb*h,replace=TRUE)
  vboots <- rbind(temp1,temp2)
  vboots <- Cfactor%*%vboots
  tempvhat1 <- matrix(vboots[1,]+rep(vhatm[1,],Nb*h),nrow=Nb,ncol=h)
  tempvhat2 <- matrix(vboots[2,]+rep(vhatm[2,],Nb*h),nrow=Nb,ncol=h)

  tempy1 <- matrix(y[(Tused)],2,Nb) # each y(t-1) at Tused~Tused-6
  tempy2 <- matrix(y[(Tused-1)],2,Nb)
  tempy3 <- matrix(y[(Tused-2)],2,Nb)
  tempy <- matrix(0,2,Nb)
  tempst <- rep(0,7)
  for(i in 1:h){
    if(i<=7){
      for(k in 1:Nb){
        tempvhat <- rbind(tempvhat1[k,i],tempvhat2[k,i])
        tempG <- 1/(1+exp(-r*(y[1,(Tused-7+i)]-c)/sdst))#st=y1(t-7)
        tempy[,k] <- Phi[,1]+(Phi[,2:3]+AB[,2:3]*tempG)%*%tempy1[,k]
                  + (Phi[,4:5]+AB[,4:5]*tempG)%*%tempy2[,k]
                  + (Phi[,6:7]+AB[,6:7]*tempG)%*%tempy3[,k]+tempvhat
      } # end loop for k
      tempy3 <- tempy2
      tempy2 <- tempy1
      tempy1 <- tempy
      tempst[i] <- mean(tempy[1,])
    } # end loop for i<7
  }
  if(i>7){
    for(k in 1:Nb){

```

```

tempvhat <- rbind(tempvhat1[k,i],tempvhat2[k,i])
tempG <- 1/(1+exp(-r*(tempst[i-7]-c)/sdst))#st=y1(t-7)
tempy[,k] <- Phi[,1]+(Phi[,2:3]+AB[,2:3]*tempG)%*tempy1[,k]
+ (Phi[,4:5]+AB[,4:5]*tempG)%*tempy2[,k]
+ (Phi[,6:7]+AB[,6:7]*tempG)%*tempy3[,k]+tempvhat
} # end loop for k
tempy3 <- tempy2
tempy2 <- tempy1
tempy1 <- tempy
} # end loop for i>7
oneT2newy[,i] <- matrix(c(mean(tempy[1,]),mean(tempy[2,])),2,1)
} # end loop for i in 1:h
} # end loop for h>1 and h<=j
if(h>1 && h>j){
temp1 <- sample(vhat_ind[1,],size=Nb*h,replace=TRUE)
temp2 <- sample(vhat_ind[2,],size=Nb*h,replace=TRUE)
vboots <- rbind(temp1,temp2)
vboots <- Cfactor%*vboots
tempvhat1 <- matrix(vboots[1,]+rep(vhatm[1,],Nb*h),nrow=Nb,ncol=h)
tempvhat2 <- matrix(vboots[2,]+rep(vhatm[2,],Nb*h),nrow=Nb,ncol=h)

tempy1 <- matrix(y[, (Tused)],2,Nb) # each y(t-1) at Tused~Tused-6
tempy2 <- matrix(y[, (Tused-1)],2,Nb)
tempy3 <- matrix(y[, (Tused-2)],2,Nb)
tempy <- matrix(0,2,Nb)
tempst <- rep(0,7)
for(i in 1:j){
if(i<=7){
for(k in 1:Nb) {
tempvhat <- rbind(tempvhat1[k,i],tempvhat2[k,i])
tempG <- 1/(1+exp(-r*(y[1,(Tused-7+i)]-c)/sdst))#st=y1(t-7)
tempy[,k] <- Phi[,1]+(Phi[,2:3]+AB[,2:3]*tempG)%*tempy1[,k]
+ (Phi[,4:5]+AB[,4:5]*tempG)%*tempy2[,k]
+ (Phi[,6:7]+AB[,6:7]*tempG)%*tempy3[,k]+tempvhat
} # end loop for k
tempy3 <- tempy2
tempy2 <- tempy1
tempy1 <- tempy
tempst[i] <- mean(tempy[1,])
} # end loop i<7
if(i>7) {
for(k in 1:Nb){
tempvhat <- rbind(tempvhat1[k,i],tempvhat2[k,i])
tempG <- 1/(1+exp(-r*(tempst[i-7]-c)/sdst)) # st=y1(t-7)
tempy[,k] <- Phi[,1]+(Phi[,2:3]+AB[,2:3]*tempG)%*tempy1[,k]
+ (Phi[,4:5]+AB[,4:5]*tempG)%*tempy2[,k]
+ (Phi[,6:7]+AB[,6:7]*tempG)%*tempy3[,k]+tempvhat
} # end loop for k
tempy3 <- tempy2

```

```

        tempy2 <- tempy1
        tempy1 <- tempy
    } # end loop for i>7
    oneT2newy[,i] <- matrix(c(mean(tempy[1,]),mean(tempy[2,])),2,1)
} # end loop for i in 1:j
} # end loop h>1 and h>j
return(oneT2newy)
}

myforecast <- function(data,h=12,H1,H2,PIN=1000){
T      <- length(data[1,])
newy   <- matrix(0,2,h*(H1-H2+1))
forePI <- matrix(0,PIN*2,h*(H1-H2+1))
for(j in H1:H2){
    print(j)
# Add one data once. Point forecast
y      <- data[,1:(T-j)]
Tused <- T-j
result <- myestimation(data=y)
res    <- result$residual
newy[,((H1-j)*h+1):((H1-j+1)*h)] <- myoneT2forecast(y,result,res,h,j)
# Prediction interval
r      <- result$gamma
c      <- result$c
Phi    <- result$estPhi
A      <- result$estA
B      <- result$estB
AB     <- A%*%B
sdst   <- sd(y[1,1:(Tused-7)])
for(PIk in 1:PIN){
    # For bootstrap data to estimate one set of parameter
    preInd <- sample(seq(1,(Tused-7),1),Tused-7,replace=TRUE)
    preres <- result$residual[,preInd]
    dataPI <- matrix(0,2,Tused)
    intInd <- sample(seq(1,(Tused-7),1),1,replace=TRUE)
    dataPI[,1:7] <- y[,intInd:(intInd+6)]
    for(PIt in 8:Tused) {
        tempG <- 1/(1+exp(-r*(dataPI[1,(PIt-7)]-c)/sdst))
dataPI[,PIt] <- Phi[,1]+(Phi[,2:3]+AB[,2:3]*tempG)%*%dataPI[, (PIt-1)]+(Phi[,4:5]
+AB[,4:5]*tempG)%*%y[, (PIt-2)]+(Phi[,6:7]+AB[,6:7]*tempG)%*%y[, (PIt-3)]+preres[,PIt-7]
    } # end loop for PIt
    newresult <- myestimation(dataPI)
    # For one set of errors
    preInd2 <- sample(seq(1,(Tused-7),1),Tused-7,replace=TRUE)
    #newres <- result$residual[,preInd2]
forePI[((PIk-1)*2+1):((PIk*2),((H1-j)*h+1):((H1-j+1)*h)]<-myoneT2forecast(y,
newresult,res,h,j)
    } # end loop for PIk
} # end loop for j in H1:H2
}

```

**Table 11.2:** *Monthly U.S. personal disposable expenditures. Forecast performance evaluation of an LSTAR(3)–CNF model versus a VAR(3) model. Blue-typed numbers denote significant parameter values at the 5% nominal level.*

Model	Maximum forecast horizon		
	$H = 1$	$H = 3$	$H = 6$
	<b>RMSFE</b>		
LVSTAR(3)–CNF	1.334	1.112	1.016
VAR(3)	1.000	0.991	1.009
	<b>Equal forecast accuracy: DM (MDM)</b>		
$Y_{1,t}$ and $Y_{2,t}$ from VAR	0.120 (0.119)	0.033 (0.032)	0.006 (0.006)
$Y_{1,t}$ from VAR	NaN (NaN)	<b>-2.254 (-2.222)</b>	-0.944 (-0.928)
$Y_{2,t}$ from VAR	<b>1.980 (1.964)</b>	1.660 (1.637)	0.537 (0.528)

```
return(list(pointf=newy, intervalf=forePI))
}
```

- (a) Within the R open source system, the add-on package `VARs` contains the function `predict` to produce forecasts from a VAR model. The `R-tsDyn` package offers the ready-to-use function `predict_rolling`. However, unlike the usual prediction methods, the function will not generate  $h = 1, \dots, h = H$ -step ahead forecasts, but only  $H$ -step ahead forecasts. The prediction errors used in parts (b) and (c) of this exercise are based on the function `varf.m` in the MATLAB Econometrics toolbox.
- (b) Table 11.2 contains the RMSFEs as well as the results of the DM and MDM test statistics to be considered in Exercise 11.6(c). In terms of RMSFEs, the VAR(3) model outperforms the LSTAR(3)–CNF model for all values of  $H$ . The difference becomes minimal as  $H$  increases. As an additional check the reader may also want to compute the GFESM measure, as defined by (11.96) in the main text. However, we expect that for this data set and time period, the overall conclusion will remain the same. That is, the linear VAR(3) is slightly better for short-term out-of-sample forecasting than the LVSTAR(3)–CNF model, indicating that the two time series  $\{Y_{1,t}\}$  and  $\{Y_{2,t}\}$  have no CNF.

Another extension is to apply the multivariate version of the DM test statistic proposed by Mariano and Preve (2012); see <http://www.mysmu.edu/staff/danielpreve> for some relevant MATLAB code.

- (c) The next two R functions can be used to compute the DM (MDM) results.

```
#####
# R code: Exercise 11.6(c)
# File: myDM-VmyDM.r
# Coded by: Dao Li
#
# Testing equal forecast accuracy.
#
# INPUT: h
```

```

# Benchmark (N=nonlinear, L=linear):
# d = (eN[1,]^2-eL[1,]^2), Y_{1,t} from VAR
# d = (eN[2,]^2-eL[2,]^2), Y_{2,t} from VAR
#####
myDM <- function (d, h, templag=NULL)
{
#d <- c(abs(e1))^power - c(abs(e2))^power
d.cov <- acf(d,na.action=na.omit,lag.max=templag,type="covariance",
plot=FALSE)$acf[, , 1]
d.var <- sum(c(d.cov[1], 2 * d.cov[-1]))/length(d)
STATISTIC <- mean(d, na.rm = TRUE)/sqrt(d.var)
return(STATISTIC) # normal
}
#####
# R code: Exercise 11.6(c)
# File: myDM-VmyDM.r
# Coded by: Dao Li
#
# Testing equal forecast accuracy in terms of "d" instead of
# "e1" and "e2".
#
# INPUT:
# h = forecast horizon; Benchmark (N = nonlinear, L = linear):
# d = cbind(eN[1,]^2-eL[1,]^2,eN[2,]^2-eL[2,]^2), Y_{i,t} from VAR
#####
VmyDM <- function (d,h,templag=NULL)
{
# d <- c(abs(e1))^power - c(abs(e2))^power
d.cov<-acf(d,na.action = na.omit,lag.max=templag,type="covariance",
plot=FALSE)$acf
n <- length(d[1,])
temp1 <- NULL
k <- 1
while(k <= n){
temp1 <- rbind(temp1,d.cov[1,1:n,k])
k <- k + 1
}
temp2 <- temp1
if(h >= 2){
for(j in 2:length(d.cov[,1,1]))
{
temp1 <- NULL
k <- 1
while(k <= n){
temp1 <- rbind(temp1,d.cov[j,1:n,k])
k <- k + 1
}
temp2 <- temp2+temp1+t(temp1)
}
}
}

```

}

Clearly, in terms of equal forecast accuracy both models produce the same overall results. However, for each of the component series  $\{Y_{1,t}\}$  and  $\{Y_{2,t}\}$  there is a slight preference for using the VAR(3) model for short-term, out-of-sample, forecasting as opposed to the LVSTAR(3)-CNF model. Hence, supporting the results in part (b).

**Remark:** As a further check, the reader may want to test the null hypothesis “*nonlinear forecasts encompass linear forecasts*” and, vice versa, the null hypothesis “*linear forecasts encompass nonlinear forecasts*”. The above two R functions can be easily modified for this purpose. For instance, assume that the null hypothesis is “*nonlinear forecasts encompass linear forecast at forecast horizon  $h = 1$* ”. Then the relevant function call is given by

```
VmyDM(d=cbind(eN[1,]^2-eL1[1,]*eN[1,],eN[2,]^2-eL1[2,]*eN[2,]),h=1).
```

Obviously, in the case of testing the null hypothesis “*linear forecasts encompass nonlinear forecasts*” the relevant function call is given by

```
VmyDM(d=cbind(eL1[1,]^2-eL1[1,]*eN[1,],eL1[2,]^2-eL1[2,]*eN[2,]),h=1).
```

So, in both cases, the test results follow from the correlation between  $e_{i,t}$  and  $e_{i,t} - e_{j,t}$  ( $i, j = 1, 2$ ). Alternatively, one may obtain test results based on the correlation between  $e_{i,t}$  and  $\hat{Y}_{j,t}$ .

The results for the MDM test statistic follow from:

```
modDM <- estDM*sqrt((n+1-2*hh+hh*(hh-1)/n)/n),
```

where `estDM` is the result based on `VmyDM`.

## Chapter 12

**12.1** By repeated application of the Kronecker product property, and using the notation introduced in Section 12.3 of the main text, it follows that

$$\begin{aligned} \mathbf{g}_Y(\omega_1) \otimes \cdots \otimes \mathbf{g}_Y(\omega_r) &= \{\mathbf{H}(\omega_1)\boldsymbol{\Sigma}_\varepsilon\mathbf{H}^*(\omega_1)\} \otimes \cdots \otimes \{\mathbf{H}(\omega_r)\boldsymbol{\Sigma}_\varepsilon\mathbf{H}^*(\omega_r)\} \\ &= \{\mathbf{H}(\omega_1) \otimes \cdots \otimes \mathbf{H}(\omega_r)\} \{\boldsymbol{\Sigma}_\varepsilon \otimes \cdots \otimes \boldsymbol{\Sigma}_\varepsilon\} \{\mathbf{H}^*(\omega_1) \otimes \cdots \otimes \mathbf{H}^*(\omega_r)\}, \quad (*) \end{aligned}$$

where  $\omega_r = -\sum_{j=1}^{r-1} \omega_j$  and  $(\omega_1, \dots, \omega_r) \in [0, 1]^r$ . Now, let  $\mathcal{H}_1 = \mathbf{H}(\omega_1) \otimes \cdots \otimes \mathbf{H}(\omega_r)$ ,  $\mathcal{H}_2 = \mathcal{H}_1^*$ , and let  $\boldsymbol{\Sigma}_\varepsilon^{[r]}$  denote the Kronecker products of  $r$  covariance matrices  $\boldsymbol{\Sigma}_\varepsilon$ . Then, from (\*), we can write

$$\{\mathbf{g}_Y(\omega_1) \otimes \cdots \otimes \mathbf{g}_Y(\omega_r)\}^{-1} = \mathcal{H}_2^{-1}(\boldsymbol{\Sigma}_\varepsilon^{[r]})^{-1}\mathcal{H}_1^{-1}.$$

Moreover, using the  $r$ th-order spectral density vector  $f_Y(\omega_1, \dots, \omega_{r-1}) = \mathbf{H}(\omega_1) \otimes \cdots \otimes \mathbf{H}(\omega_r)\mathbf{C}_r$ , we have

$$f_Y(\omega_1, \dots, \omega_{r-1}) = \mathcal{H}_1\mathbf{C}_r \quad \text{and} \quad f_Y^*(\omega_1, \dots, \omega_{r-1}) = \mathbf{C}_r'\mathcal{H}_2.$$

Hence, (12.26) follows immediately. Note that when  $m = 1$ , (12.26) reduces to

$$\frac{|f_Y(\omega_1, \dots, \omega_{r-1})|^2}{f_Y(\omega_1) \cdots f_Y(\omega_{r-1})f_Y(\omega_1 + \cdots + \omega_{r-1})} = \frac{c_r^2}{\sigma_\varepsilon^{2r}},$$

**Table 12.1:** Values of the sample ACF, the sample PACF, and the  $\widehat{R}(\ell)$  test statistic for the annual temperatures and tree ring width data set;  $T = 66$ . Blue-typed numbers denote significant values at the 5% nominal level.

Lag ( $\ell$ )	ACF		PACF		$\widehat{R}_{ij}(\ell)$	
1	0.017	0.211	0.017	0.211	0.257	0.280
	0.048	<b>0.355</b>	0.048	<b>0.355</b>	0.246	<b>0.332</b>
2	0.044	0.158	0.034	0.097	0.238	<b>0.330</b>
	0.137	0.146	0.120	0.010	0.238	0.242
3	-0.152	0.000	-0.165	-0.075	0.216	<b>0.305</b>
	0.040	<b>0.275</b>	-0.010	0.192	0.256	<b>0.323</b>
4	-0.144	0.182	-0.178	0.171	0.200	0.257
	-0.150	<b>0.260</b>	-0.152	0.097	0.266	<b>0.300</b>
5	0.065	0.065	0.130	0.016	0.279	0.212
	-0.005	0.220	0.044	0.134	0.291	0.281

where  $\text{Cum}(\varepsilon_t, \varepsilon_t) \equiv \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ , and  $c_r = \mathbb{E}(\varepsilon_t^r)$  is the  $r$ th-order cumulant of the process  $\{\varepsilon_t\}$ . With  $r = 3$ , the above result reduces to equation (4.12) in the main text.

**12.2** Table 12.1 contains various statistics for the tree ring data set: annual temperatures ( $Y_{1,t}$ ) and annual tree ring widths ( $Y_{2,t}$ ) for the time period 1907 – 1972 ( $T = 66$ ).

- The sample ACF detects only a linear relation between  $Y_{2,t}$  and  $Y_{2,t-1}$ , between  $Y_{2,t}$  and  $Y_{2,t-3}$ , and between  $Y_{2,t}$  and  $Y_{2,t-4}$  at the 5% nominal significance level. At the 10% nominal significance level there is also a relationship between  $Y_{1,t}$  and  $Y_{2,t-1}$ . The sample PACF cuts off after lag 1, suggesting a *linear* VAR(1) model for the data. Thus, the overall pattern of these statistics agrees with the observations made in Example 11.5 of the main text.
- In contrast to the results in part (a), the  $\widehat{R}(\ell)$  test statistic indicates a nonlinear relationship between  $Y_{2,t}$  and  $Y_{2,t-1}$ , between  $Y_{1,t}$  and  $Y_{2,t-2}$ , between  $Y_{1,t}$  and  $Y_{1,t-3}$ , between  $Y_{1,t}$  and  $Y_{2,t-3}$ , and between  $Y_{2,t}$  and  $Y_{t-4}$  at the 5% nominal significance level. One may, however, erroneously conclude from these results that there exists a causal nonlinear relationship from lagged values of  $Y_{2,t}$  to  $Y_{1,t}$ . In fact, both variables  $Y_{1,t}$  and  $Y_{2,t}$  are highly influenced by the ENSO phenomenon. Additionally, tree ring widths reflect changes in other climatic variables, such as precipitation.

To compute critical values of the  $\widehat{R}(\ell)$  test statistic for bivariate time series of size  $n$ , the relevant MATLAB codes are given below.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% MATLAB code: Exercise 12.2
% File: Rtest.m
%
% Purpose: To apply nonlinear vector time series lag
% identification by the R test statistic to a given data set.

```

```

%
% Converted from FORTRAN90 code written by Jane L. Harvill.
% INPUT: data = A BIVARIATE TIME SERIES of size (nr * nc=2)
% OUTPUT: rrr = values of R_{i,j}(\ell) test statistic
%           for \ell=1,...,nlag, i,j=1,...,nc, i.e.
%           two matrices of size (nlag * nc=2)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[nr, nc] = size(data);
in      = 30;    % 30-point Gaussian quadrature rule
nlag    = 5;    % Number of lags
x       = data;
rrr     = compr(nr,nlag,nc,x,in);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% SUBFUNCTIONS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [x,w]=lgwt(N,a,b)
% This script is for computing definite integrals using Legendre-Gauss
% Quadrature. Computes the Legendre-Gauss nodes and weights on an
% interval [a,b] with truncation order N.
% Suppose you have a continuous function f(x) which is defined on [a,b]
% which you can evaluate at any x in [a,b]. Simply evaluate it at all
% of the values contained in the x vector to obtain a vector f. Then
% compute the definite integral using sum(f.*w);
%
% Written by Greg von Winckel - 02/25/2004
N = N-1;
N1 = N+1; N2 = N+2;
xu = linspace(-1,1,N1)';
% Initial guess
y = cos((2*(0:N)+1)*pi/(2*N+2))+(0.27/N1)*sin(pi*xu*N/N2);
% Legendre-Gauss Vandermonde Matrix
L = zeros(N1,N2);
% Derivative of LGVM
Lp = zeros(N1,N2);
% Compute the zeros of the N+1 Legendre polynomial
% using the recursion relation and the Newton-Raphson method
y0 = 2;
% Iterate until new points are uniformly within epsilon of old points
while max(abs(y-y0))>eps
    L(:,1) = 1;
    Lp(:,1) = 0;
    L(:,2) = y;
    Lp(:,2) = 1;
    for k=2:N1
        L(:,k+1) = ((2*k-1)*y.*L(:,k)-(k-1)*L(:,k-1))/k;
    end
    Lp = (N2)*(L(:,N1)-y.*L(:,N2))./(1-y.^2);
    y0 = y;

```



```

        y = y0-L(:,N2)./Lp;
    end

    % Linear map from[-1,1] to [a,b]
    x = (a*(1-y)+b*(1+y))/2;
    % Compute the weights
    w = (b-a)./((1-y.^2).*Lp.^2)*(N2/N1)^2;
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    function f = kuniv(n,x,xt)      % Estimate of univariate density
    hx = 0.85*n^(-1/5);           % Bandwidth
    f = 0;
    for i=1:n
        f = f+tpdf((x-xt(i))/hx,4); % Univ. Student t density with df=4
    end
    f = f/(n*hx);
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    function r = compr(n,nlag,k,x,in)
    prod = x.*x;
    sum1 = sum(x(:,1:2))/n;
    sigx = sum(prod(:,1:2));
    for i=1:k
        sigx(i) = ((sigx(i)-n*sum1(i)^2)/(n-1))^(0.5);
    end
    for i=1:nlag
        for j=1:k
            for l=1:k
                for m=1:n-i
                    xt(m) = (x(m+i,j)- sum1(j))/sigx(j);
                    yt(m) = (x(m,l) - sum1(l))/sigx(l);
                end
                [rho,deltai] = hbivar2(in,n-i,xt,yt);
                r(i,j,l) = deltai;
                rt(i,j,l) = rho;
                if (deltai < 0)
                    deltai = 0;
                end
                r(i,j,l) = (1.0-exp(-2*deltai))^(0.5); % R(\ell) test statistic
            end
        end
    end
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    function [rho,h] = hbivar2(in,n,xt,yt)
    rho = sum(xt.*yt)/n;
    xlow1 = min(xt);
    xupp1 = max(xt);
    xlow2 = min(yt);
    xupp2 = max(yt);

    [qx,qwx] = lgwt(in,-1,1);

```

```

for i=1:in
    tempx = -qx(i)*(xupp1-xlow1+2)/2+(xupp1+xlow1)/2;
    tempy = -qx(i)*(xupp2-xlow2+2)/2+(xupp2+xlow2)/2;
    fx(i) = kuniv(n,tempx,xt);
    fy(i) = kuniv(n,tempy,yt);
end
h = 0;
for i=1:in
    for j=1:in
        tempx = -qx(i)*(xupp1-xlow1+2)/2+(xupp1+xlow1)/2;
        tempy = -qx(j)*(xupp2-xlow2+2)/2+(xupp2+xlow2)/2;
        f = kbivar(n,tempx,tempy,xt,yt,rho);
        h = h+(log(f)-log(fx(i))-log(fy(j)))*f*qwx(i)*qwx(j);
    end
end
h = h*(xupp1-xlow1+2)*(xupp2-xlow2+2)/4;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function f = kbivar(n,x,y,xt,yt,rho); % Estimate of bivariate density
% Product kernel of Student's t distributions with bandwidth:
hx = (0.85*(1-(rho^2))^(5/12))*(1+0.5*(rho^2))^(1/6)*n^(-1/6);
hy = hx;
f = 0;
for i=1:n
    f = f+tpdf((x-xt(i))/hx,4)*tpdf((y-yt(i))/hy,4);
end
f = f/(n*hx*hy);

```

### 12.3 (a)+(b)

Table 12.2 shows  $p$ -values of the modified nonlinear causality test statistic  $Q_{T,W}^*(h)$ , defined by (12.41) in the main text, for time periods P4 ( $T = 266$ ), P3 ( $T = 179$ ), and P2 ( $T = 301$ ). To allow for a comparison with the test results displayed in Figure 12.5 of the main text, we also included  $p$ -values for the time period P1 ( $T = 216$ ), at lag 1.

The test statistic  $Q_{T,W}^*(h)$  does not detect any causal relationship in the oldest time period P4. Also in time period P3, covering the late Pliocene, the test statistic suggests an absence of nonlinear causality among all variables, except for  $\delta^{18}\text{O} \rightarrow \delta^{13}\text{C}$ . Dust flux influences both  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$  in period P2 (early Pleistocene). Comparing time periods P2 and P1, it seems that the coupling between  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$  has become stronger over time. The role of the variables dust flux and insolation has, however, diminished. So evidence that these two variables are the driven force behind the other two variables is hardly available. This result is line with the conclusion of Diks and Mudelsee (2000). Essentially, the test results suggest a bi-directional causal nonlinear relationship between  $\delta^{13}\text{C}$ , an indicator of the strength of North Atlantic Deep Water, and  $\delta^{18}\text{O}$ , an indicator of the global ice volume.

For completeness the R code for computing the modified nonparametric causality test statistic  $Q_{T,W}^*(h)$  is given below. The HJ test statistic  $Q_{T,W}(h)$ , defined by (12.37) in the main text, can be downloaded from the website of this book; i.e. function `hj.r`.

```
#####
# R code: Exercise 12.3
# File: dp.r
# Coded by: Marcin Wolski
#
# Calculates the Diks-Panchenko (DP, 2006) p-values of the modified
# bivariate nonparametric causality test statistic.
#
# Reference:
# Diks, C. and Panchenko, V. (2006).
#   A new statistic and practical guidelines for nonparametric Granger
#   causality testing.
#   Journal of Economic Dynamics & Control, 30(9-10), 1647-1669.
#   DOI: 10.1016/j.jedc.2005.08.008.
#####
dp = function(x, y, lag=1, lagmax=1, epsilon=1.5) {
  N = length(x)
  IYij = IXVij = IYZij = IXYZij = 0
  disx = disy = disz = 0
  h = Cy = Cxy = Cyz = Cxyz = rep(0,N)
  T2 = 0
  n = N-lagmax

  # Calculating correlation integrals
  i = lagmax+1
  while(i<(N+1)) {
    Cy[i] = Cxy[i] = Cyz[i] = Cxyz[i] = 0
    j = lagmax+1
    while(j<(N+1)) {
      if(j!=i) {
        disx = disy = 0
        for(s in 1:lag) {
          disx = max(abs(x[i-s]-x[j-s]),disx)
        }
        for(s in 1:lagmax) {
          disy = max(abs(y[i-s]-y[j-s]),disy)
        }
        if(disy<=epsilon) {
          Cy[i] = Cy[i]+1
          if(disx<=epsilon) {
            Cxy[i] = Cxy[i]+1
          }
          disz = max(abs(y[i]-y[j]),disy)
          if(disz<=epsilon) {
            Cyz[i] = Cyz[i]+1
            if(disx<=epsilon) {
              Cxyz[i] = Cxyz[i]+1
            }
          }
        }
      }
    }
  }
}
```

```

    }
  }
  j = j+1
}

Cy[i] = Cy[i]/n
Cxy[i] = Cxy[i]/n
Cyz[i] = Cyz[i]/n
Cxyz[i] = Cxyz[i]/n
h[i] = h[i]+2*(Cxyz[i]*Cy[i]-Cxy[i]*Cyz[i])/6 # to check!
j = lagmax+1
while(j<(N+1)) {
  if(j!=i) {
    IYij = IXYij = IYZij = IXYZij = 0
    disx = disy = 0
    for(s in 1:(lag)) {
      disx = max(abs(x[i-s]-x[j-s]),disx)
    }
    for(s in 1:(lagmax)) {
      disy = max(abs(y[i-s]-y[j-s]),disy)
    }
    if(disy<=epsilon) {
      IYij = 1
      if(disx<=epsilon) {
        IXYij = 1
      }
      disz = max(abs(y[i]-y[j]),disy)
      if(disz<=epsilon) {
        IYZij = 1
        if(disx<=epsilon) {
          IXYZij = 1
        }
      }
    }
  }
  h[j]=h[j]+2*(Cy[i]*IXYZij-Cyz[i]*IXYij)/
  (6*n)+2*(Cxyz[i]*IYij-Cxy[i]*IYZij)/(6*n)
}
j = j+1
}
T2 = T2+Cxyz[i]*Cy[i]-Cyz[i]*Cxy[i]
i = +1
}
T2 = T2/n
i = lagmax+1
while(i<(N+1)){
  h[i] = h[i]-T2
i = i+1
}
# End of calculating correlation integrals

```

```

K      = floor(sqrt(sqrt(N)))
ohm    = rep(0,K)
ohm[1] = 1
for(k in 2:(K+1)) {
  ohm[k] = 2*(1-(k-1)/K)
}

# Autocovariance of h
cov = rep(0,K)
for(k in 1:K) {
  for(i in (lagmax+k):N) {
    cov[k] = cov[k]+h[i]*h[i-k+1]
  }
  cov[k] = cov[k]/(n-(k-1))
}

VT2 = 0
for(k in 1:K) {
  VT2 = VT2+9*ohm[k]*cov[k]
}

T2_TVAL = T2*sqrt(n)/sqrt(VT2)

# Calculating p-values
p_T2 = pnorm(-T2_TVAL)
return(p_T2)
}

```

**Remark 1:** In the derivation of the asymptotic distribution of the Diks–Panchenko (DP) bivariate causality test statistic the authors follow the approach of Powell and Stoker (1996). The main drawback is that i.i.d. data are assumed, which is obviously not the case in time series analysis. Another problem stems from the fact that the Powell–Stoker approach was originally developed in the context of 2nd order U-statistics. The DP test statistic involves U-statistics of the 3rd order, and the authors were forced to limit attention to testing only lag one dependence.

The papers by Yoshihara (1976) and Denker and Keller (1986) do not restrict the data to be i.i.d. On the other hand, their results are only applicable to statistics with kernels that are independent of the length  $T$  of the time series under study. Khashimov (1993, Thm. 2) gives a result for weakly dependent data and kernels that depends on  $T$  directly. This theorem, however, is concerned with 2nd order U-statistics. A sketch of the proof for kernels with three arguments is given by Opuchlik (2009) in an unpublished MPhil thesis. Moreover, the thesis contains MC simulation results showing that the DP test statistic is asymptotically normally distributed for  $0.25 \leq \beta \leq 0.5$ , with  $\beta$  the rate at which the bandwidth  $h$  shrinks (recall  $h = cT^{-\beta}$ , with  $c$  a scaling factor). This interval is much wider than the interval  $\beta \in (\frac{1}{4}, \frac{1}{3})$  derived by Diks and Panchenko.

**Remark 2:** The exercise can be extended by computing the sharpened version of the nonparametric Granger causality test statistic  $Q_{T,W}^s(h)$  (see (12.47) of the main text),

**Table 12.2: Extended climate change data set.** *P*-values obtained by the modified bivariate (pairwise) nonparametric causality test statistic  $Q_{T,W}^*(h)$ , with bandwidth fixed at  $h = 1.5$ , and embedding dimension (lag)  $\ell_X = \ell_Y = 1$ . The notation \*\* marks a *p*-value  $< 0.01$ , and \* marks a *p*-value lying between 0.01 and 0.05. Each entry provides information about the null hypothesis:  $Y_{i,t}$  (top row) does not cause  $Y_{j,t}$  ( $i \neq j; i, j = 1, \dots, 4$ ) (left column), i.e.  $Y_{i,t} \not\rightarrow Y_{j,t}$ .

	$\delta^{13}\text{C}$	$\delta^{18}\text{O}$	Dust flux	Insolation	$\delta^{13}\text{C}$	$\delta^{18}\text{O}$	Dust flux	Insolation
	<b>Period P1 (since 894 ka)</b>				<b>Period P2 (2470 – 987 ka)</b>			
$\delta^{13}\text{C}$		0.000**	0.044*	0.026*		0.072	0.002**	0.208
$\delta^{18}\text{O}$	0.004**		0.228	0.173	0.263		0.006**	0.034*
Dust flux	0.825	0.214		0.057	0.763	0.639		0.055
Insolation	0.780	0.807	0.455		0.349	0.312	0.628	
	<b>Period P3 (3585 – 2625 ka)</b>				<b>Period P4 (5000 – 3585 ka)</b>			
$\delta^{13}\text{C}$		0.024*	0.623	0.658		0.760	0.084	0.106
$\delta^{18}\text{O}$	0.196		0.363	0.567	0.398		0.398	0.755
Dust flux	0.485	0.168		0.816	0.184	0.599		0.067
Insolation	0.796	0.213	0.348		0.079	0.293	0.308	

proposed by Diks and Wolski (2016). The C source codes can be downloaded from the website <http://qed.econ.queensu.ca/jae/datasets/diks001/>.<sup>5</sup>

Table 12.3 shows *p*-values for the various relations. One may readily observe that there is a significant bi-directional nonlinear relationship between  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$ , for each of the two conditioning variables, i.e., insolation, and dust flux. This confirms earlier results presented in Example 12.6 of the main text. Also, after conditioning on dust flux, we see that insolation is still affecting  $\delta^{13}\text{C}$  in a nonlinear way.

Using the `dsdp_4vars.c` program, with conditioning on the variables insolation and dust flux jointly, we observe the following bi-directional causal nonlinear relationship  $\delta^{13}\text{C} \leftrightarrow \delta^{18}\text{O}$  (*p*-values in the range 1% – 5%), for both normal and uniform marginals.

**Remark 3:** The paper by Davidson et al. (2016) applies tests for Granger causality in the context of a linear VAR model using a different climate data set. The data can be downloaded from the website <http://people.exeter.ac.uk/jehd201/research.html#dps>.

- 12.4** (a) Table 12.4 shows values of  $|\widehat{\Sigma}_\varepsilon|$  for both the unrestricted and restricted (additive) PMARS-VARX models fitted to the data. Clearly, the unrestricted ARX model has the lowest value of  $|\widehat{\Sigma}_\varepsilon|$  at lag  $\ell = 15$ , with 57 terms, whereas the “best” additive ARX model is attained as lag  $\ell = 3$ , with 19 terms. Note that the

<sup>5</sup>Three files are available: `dp.c` (calculates the original DP test statistic with Gaussian kernel), `dsdp.c` (calculates the sharpened DP test statistic in a 3-variate setting), and `dsdp_4vars.c` (calculates the sharpened DP test statistic in a 4-variate setting). If the main program is `dp.c`, then the compiling command is: `GCC-O DP DP.C -LM`. The command should be in lower case letters and running in a Linux/Windows environment (e.g., within `Cygwin` for Windows users), while “dp” is the executive file.

**Table 12.3:** Extended climate change data set for time period P1.  $P$ -values obtained by the nonparametric Granger causality test statistic  $Q_{T,W}^s(h)$ , with bandwidth selection as suggested by Diks and Wolski (2016, Section 4), and embedding dimension (lag)  $\ell_X = \ell_Y = \ell_Q = 1$ . The data are standardized by transforming the marginals to either a normal (first line) or a uniform (second line) distribution. The notation \*\* marks a  $p$ -value  $< 0.01$ , and \* marks a  $p$ -value lying between 0.01 and 0.05. Each entry provides information about the null hypothesis:  $Y_{i,t}$  (top row) does not cause  $Y_{j,t}$  ( $i \neq j; i, j = 1, \dots, 3$ ) (left column), i.e.  $Y_{i,t} \not\rightarrow Y_{j,t}$ , with conditioning on a fourth variable.

		Conditioning on Insolation			Conditioning on Dust flux		
		$\delta^{13}\text{C}$	$\delta^{18}\text{O}$	Dust flux	$\delta^{13}\text{C}$	$\delta^{18}\text{O}$	Insolation
$\delta^{13}\text{C}$			0.015*	0.060	$\delta^{13}\text{C}$	0.004**	0.021*
			0.016*	0.083		0.010*	0.057
$\delta^{18}\text{O}$		0.002**		0.238	$\delta^{18}\text{O}$	0.002**	0.051
		0.036*		0.181		0.036*	0.066
Dust flux		0.466	0.181		Insolation	0.220	0.313
		0.401	0.460			0.237	0.401
		Conditioning on $\delta^{13}\text{C}$			Conditioning on $\delta^{18}\text{O}$		
		$\delta^{18}\text{O}$	Dust flux	Insolation	$\delta^{13}\text{C}$	Dust flux	Insolation
$\delta^{18}\text{O}$			0.224	0.057	$\delta^{13}\text{C}$	0.073	0.073
			0.216	0.096		0.097	0.127
Dust flux		0.169		0.330	Dust flux	0.489	0.326
		0.401		0.454		0.313	0.466
Insolation		0.312	0.104		Insolation	0.481	0.217
		0.423	0.051			0.550	0.131

**Table 12.4:** Determinant of the residual covariance matrices of the unrestricted and restricted (additive) PMARS-VARX models fitted to the Icelandic river flow data set.

PMARS model										
Lags										
	1	2	3	4	5	6	7	8	9	10
VARX	74.30	81.33	80.83	80.40	80.32	80.36	80.45	77.88	80.61	82.63
Additive VARX	132.26	112.72	102.15	102.57	102.51	100.29	100.40	100.74	100.92	102.98
Lags										
	11	12	13	14	15	16	17	18	19	20
VARX	79.97	82.91	83.05	83.20	44.12	84.29	74.81	75.17	61.10	61.20
Additive VARX	101.17	101.35	101.53	101.72	103.83	104.02	102.18	104.30	100.42	100.51

maximum number of lags (here, 15) of the unrestricted VARX model is the same as identified by AIC for the VTARX model in Table 11.4 of the main text.

- (b) The estimated residual covariance matrix for the “best” fitted PMARS–VARX model is given by

$$\widehat{\Sigma}_{\varepsilon, \text{VARX}} = \begin{pmatrix} 21.971 & 1.247 \\ 1.247 & 2.079 \end{pmatrix}, \quad R^2 = 0.95 (Q_{1,t}) \text{ and } 0.936 (Q_{2,t}),$$

where the time series  $\{Q_{1,t}\}$  denotes the daily flow of the Jökulsá Eystri river, and the time series  $\{Q_{2,t}\}$  is the daily flow of the Vatnsdalsá river. The pooled estimated residual covariance matrix follows from  $\widehat{\Sigma}_{\varepsilon}^{(1)}$  and  $\widehat{\Sigma}_{\varepsilon}^{(2)}$  in Table 11.4 of the main text, and is given by

$$\widehat{\Sigma}_{\varepsilon, \text{VTARX}} = \begin{pmatrix} 27.874 & 1.416 \\ 1.416 & 3.521 \end{pmatrix}, \quad |\widehat{\Sigma}_{\varepsilon, \text{VTARX}}| = 96.145.$$

So, in terms of residual variances, the unrestricted PMARS–VARX model is to be preferred over the VTARX model. Moreover, in terms of goodness-of-fit, there is no clear difference between the VTARX model and the best fitted additive PMARS–VARX model at lag  $\ell = 3$ , with  $R^2 = 0.926 (Q_{1,t})$  and  $R^2 = 0.891 (Q_{2,t})$ .

- (c) If we consider only terms with absolute coefficient value more than twice the estimated standard error, the unrestricted PMARS–VARX model so obtained is summarized in Table 12.5. The model reveals that temperature  $T_t$  plays a dominant role in modulating the dynamics of  $Q_{1,t}$ . For the variable  $Q_{2,t}$  we see three subsystems: low temperatures ( $\leq 0^\circ\text{C}$ ), moderate temperatures ( $0^\circ\text{C} < T_t \leq 1.5^\circ\text{C}$ ), and high temperatures ( $> 1.5^\circ\text{C}$ ). This is in good agreement with the subsystems obtained by Astatkie et al. (1997), albeit for the flow series  $\{Q_{1,t}\}$ ; see Exercise 2.11.

For the variable  $Q_{1,t-k}$  with  $k = 0, 1$ , and 2, eight threshold values are specified whereas 13 thresholds are detected for the variable  $Q_{2,t-k}$ . These threshold values may be used as proxies for the changes in soil moisture conditions in the basin of each river in the absence of measured indices. Thus, for instance, for the Jökulsá Eystri river a very wet soil condition is represented by  $Q_{1,t-k} \geq 134 \text{ m}^3/\text{s}$  ( $k = 1, 2$ ), and an extremely dry soil condition by the fact that  $Q_{1,t-1} < 30.2 \text{ m}^3/\text{s}$ .

It is interesting to see that precipitation at time  $t-1$  has only a minor effect on  $Q_{1,t-2}$ . In contrast, lagged values of the variable  $P_t$  have no effect on  $Q_{2,t}$ . This is different from the results presented in Exercise 2.11 and Table 11.4 of the main text. Finally, note from Table 12.5 that there seems to be a feedback relationship between lagged values of the two river flow series  $Q_{1,t}$  and  $Q_{2,t}$ . Sample cross-correlations computed between the series confirm this relationship. However, this result seems to be spurious and could be an indication that important variables like evapotranspiration and soil moisture indices are missing from the analysis; see also part (d) below.

- (d) Table 12.6 shows  $p$ -values for the modified bivariate nonparametric causality test statistic  $Q_{T,W}^*(h)$  with  $h = 1.5$  and lags  $\ell_{Q_1} = \ell_{Q_2} = 1, \dots, 8$ . There is no evidence to reject the null hypothesis ( $\mathbb{H}_0$ ): “ $Q_{1,t}$  does not Granger cause  $Q_{2,t}$ ”.



**Table 12.5:** *Estimated coefficients of the “best” fitted unrestricted PMARS–VARX model for the Icelandic river flow data set. Only terms with absolute coefficient value more than twice the estimated standard error are given.*

Terms	$Q_{1,t}$	$Q_{2,t}$	Terms	$Q_{1,t}$	$Q_{2,t}$
Constant	2.95	-1.62	$(Q_{1,t-1} - 30.2)_+$	-1.06	-0.22
$Q_{1,t-1}$	1.05	0.23	$(Q_{1,t-2} - 33.2)_+$	0.62	0.18
$Q_{2,t-1}$	-2.32	1.43	$(Q_{2,t-1} - 13)_+ \times Q_{1,t-2}$	0.05	–
$Q_{2,t-2}$	0.84	-0.42	$(Q_{1,t-2} - 134)_+ \times Q_{2,t-1}$	-0.46	–
$(Q_{1,t-1} - 98)_+$	-0.97	–	$(Q_{1,t-2} - 134)_+ \times Q_{2,t-2}$	0.27	0.10
$T_t$	-0.95	–	$(Q_{1,t-2} - 90)_+ \times T_t$	0.06	–
$Q_{1,t-1} \times T_t$	-0.13	–	$Q_{1,t-2} \times P_{t-1}$	-0.01	–
$Q_{2,t-1} \times T_t$	-0.10	-0.03	$(Q_{2,t-1} - 16.9)_+$	5.31	2.40
$Q_{2,t-1} \times (T_t - 0.3)_+$	–	0.18	$(Q_{2,t-1} - 10.8)_+$	11.50	–
$Q_{1,t-2} \times T_t$	-0.08	–	$(Q_{2,t-2} - 18.4)_+$	-4.20	-2.24
$Q_{1,t-1} \times Q_{2,t-1}$	0.16	0.01	$(Q_{2,t-1} - 11.9)_+$	-11.21	–
$Q_{1,t-1} \times Q_{2,t-2}$	-0.07	-0.02	$Q_{2,t-14}$	–	0.12
$(Q_{1,t-1} - 134)_+$	6.72	-0.56	$(Q_{2,t-14} - 32.9)_+$	–	0.65
$(Q_{1,t-2} - 134)_+$	2.35	-0.89	$Q_{2,t-1} \times Q_{2,t-14}$	0.02	-0.01
$(Q_{2,t-2} - 9.02)_+$	–	0.82	$(Q_{2,t-1} - 10.8)_+ \times T_t$	–	0.09
$(Q_{2,t-1} - 23)_+$	8.08	1.88	$(Q_{2,t-1} - 9.9)_+$	-3.88	–
$(Q_{2,t-2} - 9.9)_+$	-0.27	0.21	$(Q_{2,t-2} - 9.9)_+$	-2.36	0.67
$Q_{1,t-2} \times Q_{2,t-1}$	-0.09	-0.02	$(Q_{2,t-1} - 20.1)_+$	-4.75	-1.02
$Q_{1,t-2} \times Q_{2,t-2}$	0.03	0.01	$(Q_{2,t-1} - 25.7)_+$	-4.89	-1.84
$Q_{2,t-3} \times T_t$	0.07	0.02	$(Q_{1,t-1} - 50.9)_+$	-0.66	–
$Q_{2,t-1} \times (T_t - 1.5)_+$	–	-0.12	$(Q_{1,t-1} - 50.9)_+ \times T_t$	-0.05	–
$(Q_{2,t-1} - 23)_+ \times Q_{1,t-2}$	–	0.01	$(Q_{1,t-1} - 54.9)_+ \times T_t$	0.73	–
$(Q_{2,t-1} - 13)_+$	5.58	-0.15	$(Q_{2,t-1} - 25.7)_+ \times T_t$	-0.10	0.25
$Q_{1,t-1} \times (Q_{2,t-1} - 13)_+$	-0.11	–	$(Q_{2,t-1} - 25.7)_+ \times (T_t - 1.5)_+$	0.86	-0.94

**Table 12.6:** *Icelandic river flow data set.  $P$ -values of the modified bivariate (pairwise) nonparametric causality test statistic  $Q_{T,W}^*(h)$ , with bandwidth fixed at  $h = 1.5$  and lag lengths  $\ell_{Q_1} = \ell_{Q_2} = 1, \dots, 8$ .*

$\mathbb{H}_0$	Lags							
	1	2	3	4	5	6	7	8
$Q_{1,t} \rightsquigarrow Q_{2,t}$	0.874	0.953	0.991	0.996	0.996	0.996	0.997	0.995
$Q_{2,t} \rightsquigarrow Q_{1,t}$	0.009	0.012	0.012	0.013	0.013	0.014	0.020	0.026

On the other hand, there is sufficient evidence to reject the null hypothesis: “ $Q_{2,t}$  does not Granger cause  $Q_{1,t}$ ”. So, for all lag lengths, the test statistic detects a uni-directional nonlinear relationship  $Q_{2,t} \rightarrow Q_{1,t}$ .

**Remark:** As a further check of the adequacy of the fitted PMARS-VARX model, the multivariate quantile residuals may be tested for normality, autocorrelation, and heteroskedasticity using the MATLAB codes in the file `Matl-code-Quantile-residuals-multivariate.zip`.

# Computer Codes<sup>1</sup>

The computer codes below can be found at two locations:

<http://www.jandegooijer.nl> and <http://extras.springer.com> (ISBN 978-3-319-43251-9).

EXAMPLES	EXERCISES	MISCELLANEA
<b>Chapter 1</b>		
Example_1-5.zip (R code)	Exercise_1-8-M.zip (M code)	Resnick-vdBerg-test.txt (S-Plus)
Example_1-8.zip (F code & exe files)	Exercise_1-8-G.zip (G code)	
Example_1-9.zip (Data & M code)		
<b>Chapter 2</b>		
Example_2-8.zip (M code)	Exercise_2-9.zip (R code)	Gonzalo-Wolf-SETAR.zip (C code)
Example_2-9.zip (F code & exe file)	Exercise_2-10.zip (R code)	GRASP.zip (M code)
	Exercise_2-11.zip (R code)	SEASETAR.zip (F code)
	Exercise_2-12.zip (M code)	
<b>Chapter 3</b>		
Example_3-3.zip (R code)	Exercise_3-8.zip (M code)	
<b>Chapter 4</b>		
Example_4-4.zip (M code)	Exercise_4-4.zip (R code)	Hinich-tests.zip (F code)
Section_4-5.zip (M code)	Exercise_4-4b.zip (M code)	Nonlinear-toolbox.zip (exes file)
	Exercise_4-5.zip (R code)	Simulation-freq-domain- tests.zip (M code)
		Subba-Rao-program4.zip (F code)
<b>Chapter 5</b>		
Example_5-1.zip (M code)	Exercise_5-4.zip (G code)	AIC_new.f90 (F code)
Example_5-2.zip (F code & exe file)		BL-U-root-Charemza.zip (G code)
		LM-F6-test.g (G code)
		LR-Chan.zip (F code & exe file)
		ML-BL-U-root-Hristova.zip (G code)
		NLTS.f (F code)
		NLTS-S-Plus.zip (S-Plus)
		SETAR-two-regimes.zip (R code)

<sup>1</sup>File type: F = FORTRAN, G = GAUSS, and M = MATLAB.

EXAMPLES	EXERCISES	MISCELLANEA
<b>Chapter 6</b>		
Example_6-1.zip (M code)	Exercise_6-5.m (M code)	Amendola-Francq.zip (R code)
Example_6-2.zip (M & R codes)	Exercise_6-5-remark.r (R code)	Bagnato.zip (M code)
Example_6-3.zip (F code)	Exercise_6-6.m (M code)	Ling-Tong.zip (F code)
Example_6-4.zip (R code)	Exercise_6-8.r (R code)	STAR3-32bit.zip (F code)
Example_6-5.zip (M code)	Exercise_6-9.m (M code)	Strikholm-Teräsvirta.zip (G code)
Example_6-6.zip (Renamed exe and dll)		TARfit.src (G code)
Example_6-7.zip (M code)		
Example_6-8.zip (G code)		
<i>Section 6:4:</i>		<i>Threshold models:</i>
TARSO.zip (F code)		(Renamed exe and dll)
		DTGARCH-GA.zip
		MSETAR-GA.zip
		PLTAR-GA.zip
		SETAR-GA.zip
		Giovanis-GA.zip (M code)
		Multiple-regime-GA (M code)
<b>Chapter 7</b>		
<i>Example 7.6 and Section 7.5:</i>	Exercise_7-7.zip (M code)	
Rank-based-BDS.zip (C code)		
<i>Section 7.3.3:</i>		
Bagnato-et-al.zip (R code)		
Hong.zip (G code)		
Hong-Lee.zip (G code)		
Hong-White.zip (G code)		
Skaug-Tjostheim.zip (C code)		
<b>Chapter 8</b>		
<i>Section 8.2 and Application:</i>	Exercise_8-6.r (R code)	
Chen-TR-test (G code)		
Copula-based-TR-test (M code)		
Ramsey-Rothman.zip (F code & exe files)		
Rever2.zip (C code and Linux/Windows executable)		
<b>Chapter 9</b>		
Example_9-8.zip (R code)	Exercise_9-1.zip (M code)	Algorithm-93.ox (Ox code)
Example_9-9.zip (S-Plus)	Exercise_9-2a-b.zip (M code)	FCAR.zip (S-Plus)
Example_9-10.zip (SAS)	Exercise_9-2b.zip (R code)	FPE-additive.zip (G code)
	Exercise_9-4.zip (M code)	Mean_median.m (M code)
	Exercise_9-5.zip (R code)	
<b>Chapter 10</b>		
Example_10-2.zip (R code)	Exercise_10-10.zip (R code)	Pan-Politis.zip (R code)
Example_10-3.zip (F code)	Exercise_10-11b.zip (R code)	Regions.zip (G code)
Example_10-4.zip (M code)	Exercise_10-12.zip (M code)	
Example_10-6 (RATS & M codes)		
Example_10-7.zip (C code and executables)		
Example_10-8.zip (R code)		
Section10-1-1 (F code)		

EXAMPLES	EXERCISES	MISCELLANEA
<b>Chapter 11</b>		
Example_11-1.zip (M code)	Exercise_11-5c.zip (R code)	Appendix_11A.zip (M code)
Example_11-2.zip (M code)	Exercise_11-6.zip (R code)	MTARfor.f (F code)
Example_11-3.zip (M code)		Quantile-residuals.zip (M code)
Example_11-4.zip (M code)		Robust-VSETAR.zip (M code)
Example_11-5.zip (F code)		TEVCM.zip (F code)
Example_11-6.zip (F & M codes)		V-ARasMA-
Application.zip (F code & exe files)		asQGARCH.zip (RATS)
<b>Chapter 12</b>		
Example_12-1.zip (Figures)	Exercise_12-2.zip (M code)	Anoruo-noncausality.zip (M code)
Example_12-2.zip (M code)	Exercise_12-3.zip (C code)	Hiemstra-Jones.r (R code)
Example_12-3.zip (M & R codes)		POLYMARS.zip (S-Plus)
Example_12-4.zip (C & M codes)		VFCAR.zip (F & M codes)
Example_12-5.zip (F code & exe files)		
Example_12-6.zip (C & M codes and executables)		

## References\*

*Pages on which each reference is cited are given in square brackets.*

- Abarbanel, H.D.I. and Lall, U. (1996). Nonlinear dynamics of the Great Salt Lake: System identification and prediction. *Climate Dynamics*, 12(4), 287–297.  
DOI: [10.1007/bf00219502](https://doi.org/10.1007/bf00219502). □
- Anděl, J. (1987). On linear processes with given moments. *Journal of Time Series Analysis*, 8(4), 373–378. DOI: [10.1111/j.1467-9892.1987.tb00001.x](https://doi.org/10.1111/j.1467-9892.1987.tb00001.x). □
- Anděl, J., Netuka, I., and Zvára, K. (1984). On threshold autoregressive processes. *Kybernetika*, 20(2), 89–106. □
- Anděl, J. and Bartoň, T. (1986). A note on the threshold AR(1) model with Cauchy innovations. *Journal of Time Series Analysis*, 7(1), 1–5.  
DOI: [10.1111/j.1467-9892.1986.tb00481.x](https://doi.org/10.1111/j.1467-9892.1986.tb00481.x). □
- Anděl, J. and Garido, M. (1988). On stationary distributions of some time series models. *Transactions of the Tenth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*. Academia, Publishing House of the Czechoslovak Academy of Sciences, Prague, pp. 195–202. □
- Astatkie, T., Watts, D.G., and Watt, W.E. (1997). Nested threshold autoregressive (NeTAR) models. *International Journal of Forecasting*, 13(1), 105–116.  
DOI: [10.1016/s0169-2070\(96\)00716-9](https://doi.org/10.1016/s0169-2070(96)00716-9). □
- Bacon, D.W. and Watts, D.G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, 58(3), 525–534. DOI: [10.1093/biomet/58.3.525](https://doi.org/10.1093/biomet/58.3.525). □
- Bai, J. and Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*, 23(1), 49–60.  
DOI: [10.1198/073500104000000271](https://doi.org/10.1198/073500104000000271). □
- Bakouch, H.S. and Ristić, M.M. (2009). A Bivariate Beta-Gamma autoregressive process (BVBGAR(1)). *Communications in Statistics: Theory and Methods*, 38(7), 1113–1131.  
DOI: [10.1080/03610920802376330](https://doi.org/10.1080/03610920802376330). □
- Bowman, K.O. and Shenton, L.R. (1975). Omnibus contours for departures from normality based on  $\sqrt{b_1}$  and  $b_2$ . *Biometrika*, 62(2), 243–250. DOI: [10.1093/biomet/62.2.243](https://doi.org/10.1093/biomet/62.2.243). □

---

\*A DOI (Digital Object Identifier) number can be converted to a web address with the URL prefix <http://dx.doi.org/>. The URL will lead to the abstract of a paper or a book.

- Chan, K.S. and Tong, H. (1986). A note on certain integral equations associated with non-linear time series analysis. *Probability Theory and Related Fields*, 73(1), 153–159. DOI: [10.1007/bf01845999](https://doi.org/10.1007/bf01845999). □
- D'Agostino, R.B. (1970). Transformation to normality of the null distribution of  $g_1$ . *Biometrika*, 57(3), 679–681. DOI: [10.1093/biomet/57.3.679](https://doi.org/10.1093/biomet/57.3.679). □
- Davidson, J.E.H., Stephenson, D.B., and Turasie, A.A. (2016). Time series modelling of paleoclimate data. *Environmetrics*, 27(1), 55–65. DOI: [10.1002/env.2373](https://doi.org/10.1002/env.2373). □
- De Gooijer, J.G. and Heuts, R.M.J. (1987). Higher order moments of bilinear time series processes with symmetrically distributed errors. In T. Pukkila and S. Puntanen (Eds.), *Proceedings of the Second International Conference in Statistics*, Tampere, Finland, pp. 467–478. Available at: [https://www.researchgate.net/profile/Jan\\_G\\_De\\_Gooijer/publications?sorting=newest&page=6](https://www.researchgate.net/profile/Jan_G_De_Gooijer/publications?sorting=newest&page=6). □
- Denker, M. and Keller, G. (1986). Rigorous statistical procedures for data from dynamical systems. *Journal of Statistical Physics*, 44(1-2), 67–93. DOI: [10.1007/bf01010905](https://doi.org/10.1007/bf01010905). □
- Diks, C. and Mudelsee, M. (2000). Redundancies in the Earth's climatological time series. *Physics Letters, A* 275, 407–414. DOI: [10.1016/s0375-9601\(00\)00613-7](https://doi.org/10.1016/s0375-9601(00)00613-7). □
- Diks, C. and Wolski, M. (2016). Nonlinear Granger causality: Guidelines for multivariate analysis. *Journal of Applied Econometrics*, 31(7), 1333–1351. DOI: [10.1002/jae.2495](https://doi.org/10.1002/jae.2495). □
- Doornik, J.A. and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70, 927–939. DOI: [10.1111/j.1468-0084.2008.00537.x](https://doi.org/10.1111/j.1468-0084.2008.00537.x). □
- Feigin, P.D. and Tweedie, R.L. (1985). Random coefficient autoregressive processes: A Markov chain analysis of stationary and finiteness of moments. *Journal of Time Series Analysis*, 6(1), 1–14. DOI: [10.1111/j.1467-9892.1985.tb00394.x](https://doi.org/10.1111/j.1467-9892.1985.tb00394.x). □
- Feng, H., Peng, L., and Zhu, F. (2013). Interval estimation for a simple bilinear model. *Statistics & Probability Letters*, 83(10), 2152–2159. DOI: [10.1016/j.spl.2013.05.037](https://doi.org/10.1016/j.spl.2013.05.037). □
- Gannoun, A. (1990). Estimation non paramétrique de la médiane conditionnelle: médianogramme et méthode du noyau. *Publication de l'Institut de Statistique de l'Université de Paris*, XXXV, 11–22. □
- Gradshteyn, I.S. and Ryzhik, I.M. (1994). *Table of Integrals, Series, and Products* (5th edn.). Academic Press, New York. □
- Grahn, T. (1995). A conditional least squares approach to bilinear time series estimation. *Journal of Time Series Analysis*, 16(5), 509–529. DOI: [10.1111/j.1467-9892.1995.tb00251.x](https://doi.org/10.1111/j.1467-9892.1995.tb00251.x). □
- Granger, C.W.J. and Andersen, A.P. (1978a). *An Introduction to Bilinear Time Series Models*. Vandenhoeck & Ruprecht, Göttingen. □
- Granger, C.W.J. and Andersen, A.P. (1978b). On the invertibility of time series models. *Stochastic Processes and their Applications*, 8(1), 87–92. DOI: [10.1016/0304-4149\(78\)90069-8](https://doi.org/10.1016/0304-4149(78)90069-8). □
- Hannan, E.J. (1970). *Multiple Time Series*. Wiley, New York. DOI: [10.1002/9780470316429](https://doi.org/10.1002/9780470316429). □

- Hannan, E.J. and Deistler, M. (2012). *The Statistical Theory of Linear Systems*. Classics in Applied Mathematics (CL70), SIAM, Philadelphia (Originally published: Wiley, New York, 1988). DOI: [10.1137/1.9781611972191](https://doi.org/10.1137/1.9781611972191).  $\square$
- Hannan, E.J. and Quinn, B.G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, B 41(2), 190–195.  $\square$
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, UK. DOI: [10.1017/ccol0521382483](https://doi.org/10.1017/ccol0521382483).  $\square$
- Ibragimov, I.A. and Linnik, Yu.V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.  $\square$
- Jaditz, T. and Riddick, L.A. (2000). Time-series near-neighbor regression. *Studies in Non-linear Dynamics and Econometrics*, 4(1), 35–44. DOI: [10.2202/1558-3708.1054](https://doi.org/10.2202/1558-3708.1054).  $\square$
- Jang, D., Oh, H.-S., and Dim, D. (2011). Extending the scope of automatic time series model selection: The package autots for R. *Communications of the Korean Statistical Society*, 18(3), 319–331. DOI: [10.5351/ckss.2011.18.3.319](https://doi.org/10.5351/ckss.2011.18.3.319).  $\square$
- Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis* (5th edn.). Prentice Hall, Upper Saddle River, New Jersey.  $\square$
- Kalliovirta, L. (2012). Misspecification tests based on quantile residuals. *Econometrics Journal*, 15(2), 358–393. DOI: [10.1111/j.1368-423x.2011.00364.x](https://doi.org/10.1111/j.1368-423x.2011.00364.x).  $\square$
- Khashimov, Sh.A. (1993). The central limit theorem for generalized  $U$ -statistics for weakly dependent vectors. *Theory of Probability & Its Applications*, 38(3), 456–469. DOI: [10.1137/1138042](https://doi.org/10.1137/1138042).  $\square$
- Kim, W.K., Billard, L., and Basawa, I.V. (1990). Estimation for the first-order diagonal bilinear time series model. *Journal of Time Series Analysis*, 11(3), 215–229. DOI: [10.1111/j.1467-9892.1990.tb00053.x](https://doi.org/10.1111/j.1467-9892.1990.tb00053.x).  $\square$
- Krein, M.T. and Nudelman, A.A. (1977). *The Markov Moment Problem and Extremal Problems* (Translations of Mathematical Monographs, Vol. 50). Providence: American Mathematical Society.  $\square$
- Kullback, S. (1959). *Information Theory and Statistics*. Wiley, New York.  $\square$
- Kullback, S., Keegel, J., and Kullback, J. (1987). *Topics in Statistical Information Theory*. Springer-Verlag, New York. DOI: [10.1007/978-1-4615-8080-5](https://doi.org/10.1007/978-1-4615-8080-5).  $\square$
- Lall, U. and Sharma, A. (1996). A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, 32(3), 679–693. DOI: [10.1029/95wr02966](https://doi.org/10.1029/95wr02966).  $\square$
- Lanne, M. and Saikkonen, P. (2003). Modeling the U.S. short-term interest rate by mixture autoregressive processes. *Journal of Financial Econometrics*, 1(1), 96–125. DOI: [10.1093/jjfinec/nbg004](https://doi.org/10.1093/jjfinec/nbg004).  $\square$
- Liu, S.-I. (1985). Theory of bilinear time series models. *Communications in Statistics: Theory and Methods*, 4(10), 2549–2561. DOI: [10.1080/03610926.1985.10524941](https://doi.org/10.1080/03610926.1985.10524941).  $\square$
- Mariano, R.S. and Preve, D. (2012). Statistical tests for multiple forecast comparison. *Journal of Econometrics*, 169(1), 123–130. DOI: [10.1016/j.jeconom.2012.01.014](https://doi.org/10.1016/j.jeconom.2012.01.014).  $\square$
- Opuchlik, M. (2009). *On Asymptotic Distributions of  $U$ -statistics with Sample-size Dependent Bandwidth*. MPhil Thesis, Tinbergen Institute, Amsterdam.  $\square$



- Pham, D.T. and Tran, L.T. (1981). On the first-order bilinear time series model. *Journal of Applied Probability*, 18(3), 617–627. DOI: [10.2307/3213316](https://doi.org/10.2307/3213316). □
- Patil, G.P. and Seshadri, V. (1964). Characterization theorems for some univariate probability distributions. *Journal of the Royal Statistical Society, B* 26(2), 286–292. □
- Pitarakis, J.-Y. (2006). Model selection uncertainty and detection of threshold effects. *Studies in Nonlinear Dynamics & Econometrics*, 10(1), 1–30. DOI: [10.2202/1558-3708.1256](https://doi.org/10.2202/1558-3708.1256). □
- Powell, J.L. and Stoker, T.M. (1996). Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics*, 75(2), 291–315. DOI: [10.1016/0304-4076\(95\)01761-5](https://doi.org/10.1016/0304-4076(95)01761-5). □
- Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, (2nd edn.). Wiley, New York. □
- Rosanov, Yu.A. (1967). *Stationary Random Processes*. Holden-Day, San Francisco, California. □
- Sangoyomi, T.B., Lall, U., and Abarbanel, H.D.I. (1996). Nonlinear dynamics of the Great Salt Lake: Dimension estimation. *Water Resources Research*, 32(1), 149–159. DOI: [10.1029/95wr02872](https://doi.org/10.1029/95wr02872). □
- Subba Rao, T. and Gabr, M.M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Springer-Verlag, New York. DOI: [10.1007/978-1-4684-6318-7](https://doi.org/10.1007/978-1-4684-6318-7). □
- Tan, L. and Wang, L. (2016). The LASSO method for bilinear time series models. *Communications in Statistics: Simulation and Computation*, 45(3), 1072–1082. DOI: [10.1080/03610918.2014.915036](https://doi.org/10.1080/03610918.2014.915036). □
- Terdik, G. and Subba Rao, T. (1989). On Wiener–Ito representation and the best linear predictors for bilinear time series. *Journal of Applied Probability*, 26(2), 274–286. DOI: [10.2307/3214034](https://doi.org/10.2307/3214034). □
- Tong, H. (1990). *Nonlinear Time Series Analysis: A Dynamical System Approach*. Oxford University Press, London. □
- Tsay, R.S. (2010). *Analysis of Financial Time Series* (3rd edn.). Wiley, New York. DOI: [10.1002/0471264105](https://doi.org/10.1002/0471264105). □
- Urzúa, C.M. (1996). On the correct use of omnibus tests for normality. *Economics Letters*, 53(3), 247–251. DOI: [10.1016/s0165-1765\(96\)00923-8](https://doi.org/10.1016/s0165-1765(96)00923-8). □
- Wecker, W.E. (1981). Asymmetric time series. *Journal of the American Statistical Association*, 76(373), 16–21. Corrigendum: p. 954. DOI: [10.2307/2287034](https://doi.org/10.2307/2287034). □
- Yoshihara, K. (1976). Limiting behavior of  $U$ -statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35(3), 237–252. DOI: [10.1007/bf00532676](https://doi.org/10.1007/bf00532676). □