Threshold Vector ARMA forecasts under general loss functions (Prévisions pour le model vectoriel ARMA avec seuil générées par une function de perte générale)

Niglio, Marcella
University of Salerno, Department of Economics and Statistics
Via Ponte Don Melillo
84084 Fisciano (SA), Italy
E-mail: mniglio@unisa.it

1 Introduction

In nonlinear time series domain, the introduction of new models is often based on the generalization of linear structures. In this paper we propose a generalization of the well known Vector Autoregressive Moving Average (VARMA) model widely presented, among the others, in Lütkepohl (2006).
Let $\mathbf{y}_t$ a K-variate time series, it is said to follow a VARMA(p, q) process if:

$$(1) \quad \mathbf{y}_t = \mathbf{v} + \mathbf{A}_1\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t + \mathbf{M}_1\mathbf{u}_{t-1} + \ldots + \mathbf{M}_q\mathbf{u}_{t-q}$$

where:

$$\mathbf{y}_t = \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{K,t} \end{bmatrix}, \quad \mathbf{u}_t = \begin{bmatrix} u_{1,t} \\ \vdots \\ u_{K,t} \end{bmatrix}, \qquad \text{with} \quad \mathbf{u}_t \sim WN(\mathbf{0}, \Sigma_{\mathbf{u}}),$$

and coefficients

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_K \end{bmatrix}, \quad \underset{(K\times K)}{\mathbf{A}_i} = [a_{s,r}], \quad \underset{(K\times K)}{\mathbf{M}_i} = [m_{s,r}], \quad \text{with} \quad i = 1, 2, \ldots, p \quad \text{and} \quad j = 1, 2, \ldots, q.$$

As in the univariate case where Tong (1983) proposes the Threshold ARMA model that can be seen as nonlinear generalization of the linear ARMA process (for a wide presentation of nonlinearity in time series analysis see Tong (1990), Fan and Yao (2003) whereas for the linear ARMA model see Box *et al.* (2008)), even in the multivariate context the Threshold VARMA model (TVARMA) can be defined starting from (1). More precisely a time series $\mathbf{y}_t$ follows a TVARMA$(\ell; p, q)$ model if:

$$(2) \quad \mathbf{y}_t = \sum_{k=1}^{\ell} \left[\mathbf{v}^{(k)} + \mathbf{A}_1^{(k)}\mathbf{y}_{t-1} + \ldots + \mathbf{A}_p^{(k)}\mathbf{y}_{t-p} + \mathbf{u}_t + \mathbf{M}_1^{(k)}\mathbf{u}_{t-1} + \ldots + \mathbf{M}_q^{(k)}\mathbf{u}_{t-q}\right] I\left(x_{t-d} \in \mathcal{R}_k\right)$$

where x_t is the univariate and stationary *threshold process*, d is the threshold delay, $\mathcal{R}_k = [r_{k-1}, r_k)$ such that $-\infty = r_0 < r_1 < \ldots < r_\ell = \infty$, $\bigcup_{k=1}^{\ell} \mathcal{R}_k = \mathcal{R}$, r_k is the so called *threshold value* (for $k = 1, 2, \ldots, \ell$) and $\{\mathbf{u}_t\}$ is a sequence of independent white noise with $\mathbb{E}[\mathbf{u}_t] = \mathbf{0}$ and covariance matrix $\Sigma_{\mathbf{u}}$.

Model (2) can be written equivalently as a TVAR$(\ell; 1)$ process:

$$(3) \quad \mathbf{Y}_t = \mathbf{A}^{(k)}\mathbf{Y}_{t-1} + \mathbf{U}_t, \qquad \text{if} \quad x_{t-d} \in \mathcal{R}_k,$$

with

$$\underset{[K(p+q)\times 1]}{\mathbf{Y}_t} = \begin{bmatrix} \mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t-p+1} \\ \mathbf{u}_t \\ \vdots \\ \mathbf{u}_{t-q+1} \end{bmatrix}, \quad \underset{[K(p+q)\times 1]}{\mathbf{U}_t} = \begin{bmatrix} \mathbf{u}_t \\ \underset{[K(p-1)\times 1]}{\mathbf{0}} \\ \mathbf{u}_t \\ \underset{[K(q-1)\times 1]}{\mathbf{0}} \end{bmatrix},$$

and matrix $\mathbf{A}^{(k)}$ so partitioned

$$\underset{[K(p+q)\times K(p+q)]}{\mathbf{A}^{(k)}} = \begin{bmatrix} \mathbf{A}_{11}^{(k)} & \mathbf{A}_{12}^{(k)} \\ \mathbf{A}_{21}^{(k)} & \mathbf{A}_{22}^{(k)} \end{bmatrix}$$

where

$$\underset{(Kp\times Kp)}{\mathbf{A}_{11}^{(k)}} = \begin{bmatrix} \mathbf{A}_1^{(k)} & \dots & \mathbf{A}_{p-1}^{(k)} & \mathbf{A}_p^{(k)} \\ \mathbf{I}_K & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \vdots & \mathbf{I}_K & \mathbf{0} \end{bmatrix}, \quad \underset{(Kq\times Kq)}{\mathbf{A}_{12}^{(k)}} = \begin{bmatrix} \mathbf{M}_1^{(k)} & \dots & \mathbf{M}_{q-1}^{(k)} & \mathbf{M}_q^{(k)} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$$\underset{(Kp\times Kp)}{\mathbf{A}_{21}^{(k)}} = \mathbf{0}, \quad \underset{(Kq\times Kq)}{\mathbf{A}_{22}^{(k)}} = \begin{bmatrix} \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_K & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathbf{I}_K & \mathbf{0} \end{bmatrix}.$$

Starting from model (2), in Section 2 we discuss some properties of the TVARMA model mainly related to the skewness of the data generating process that leads to select asymmetric loss functions when the generation of forecasts is desired. More precisely, in Section 3, under the assumption that the parameters of the model are all known, we show how predictions can be generated from model (2) taking advantage not only of symmetric multivariate loss functions, traditionally used in the linear domain, but even general multivariate loss functions. These functions appear to be not only more flexible but even able to catch some features of the generating process that are neglected form symmetric loss functions. The distribution of the predictors obtained from the selected general loss function are investigated through a Monte Carlo study where symmetric and asymmetric forecasts are compared.

2 Skewness of the TVARMA process

When a new model is proposed, one of the main interest is to evaluate which feature it is able to catch in the generating process.
In the univariate case it is widely known (see among the others Tong (1990)) that the threshold models can the preferred with respect to the linear time series models, in presence of data that show skewness often neglected by other classes of stochastic structures. This property makes the threshold model suitable in presence of business cycles or when the data show asymmetric behavior that is often recognized in financial returns, hydrological time series and in most economic data.

In the multivariate context of the TVARMA model a similar property can be recognized. In this domain, the main difficulty is related to the evaluation of the skewness of data that has been differently discussed in the literature.

Starting from Mardia (1970), several multivariate indices of skewness have been proposed each having desirable properties. Among them we have considered the Kollo (2008) index that can be seen as natural multivariate analogue of the one-dimensional measure of skewness (given by the third moment of a standardized variable). Further it includes the information available in the Mardia (1970) index and in some case it tries to face problems encountered in this last index.

The Kollo (2008) index is based on the definition of a K-vector given respectively by:

$$b(\mathbf{y}_t) = \mathbb{E}\left[\sum_{i=1}^{K}\sum_{j=1}^{K}(z_{i,t}z_{j,t})\mathbf{z}_t\right], \tag{4}$$

where $\mathbf{z}_t = \Sigma_{\mathbf{y}_t}^{-1/2}(\mathbf{y}_t - \mathbb{E}(\mathbf{y}_t))$ whereas $z_{i,t}$ and $z_{j,t}$ are two univariate standardized variables, for $i, j = 1, 2, \ldots, K$.

To empirically evaluate the skewness of the TVARMA process, consider the following Example.

Example 1.

Let $\mathbf{y}_t$ a stationary TVARMA(2; 1) model:

$$\mathbf{y}_t = \begin{cases} \mathbf{v}^{(1)} + \begin{bmatrix} 0.7 & 0.3 \\ 0 & 0.4 \end{bmatrix} \mathbf{y}_{t-1} + \mathbf{u}_t + \begin{bmatrix} -0.4 & 0.6 \\ 0.34 & 0 \end{bmatrix} \mathbf{u}_{t-1} & \text{if} \quad x_{t-1} \leq 0 \\ \mathbf{v}^{(2)} + \begin{bmatrix} -0.5 & 0.1 \\ -0.4 & 0.5 \end{bmatrix} \mathbf{y}_{t-1} + \mathbf{u}_t + \begin{bmatrix} 0 & 0 \\ -0.8 & 0.3 \end{bmatrix} \mathbf{u}_{t-1} & \text{if} \quad x_{t-1} > 0 \end{cases} \tag{5}$$

with $\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{u}})$, $\Sigma_{\mathbf{u}} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ and $x_t \sim AR(1)$ with parameter $\phi_1 = 0.8$.

In the univariate case, it is shown that the skewness of the threshold model is strongly related to the value of the intercepts in each regime (Niglio and Vitale, 2010). This property seems to be confirmed even in the multivariate case. To give evidence of this statement, in Figure 1 are presented the time plot of the first 100 observations and the contour-plots of two traces (of length 500) generated from model (5), where in the first model (first line of Figure 1) the two intercepts are $\mathbf{v}^{(1)} = (0, 0)$ and $\mathbf{v}^{(2)} = (0, -1)$ whereas the second model (second line of Figure 1) has intercepts $\mathbf{v}^{(1)} = (0, 0)$ and $\mathbf{v}^{(2)} = (-2, -1)$. The more asymmetric distribution of the second model, with respect to the first, can be clearly graphically appreciated and further confirmed after the computation of the index (4) which has values $b(\mathbf{y}_t) = (0.7155, -0.1739)$ for the first model and $b(\mathbf{y}_t) = (0.7742, -0.2506)$ for the second model.

3 The forecasts generation

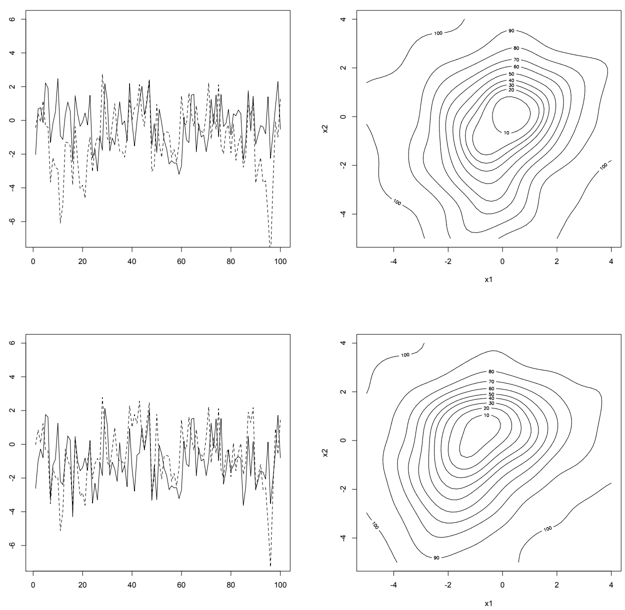
In time series analysis the generation of forecasts is based on the selection of a loss function $\mathcal{L}(\cdot)$ such that the optimal point forecast $\hat{\mathbf{y}}_{t+h}$ is obtained from

$$\arg\min_{\hat{\mathbf{y}}_{t+h}} \mathbb{E}\left[\mathcal{L}\left(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}\right) | \mathcal{F}_t\right] \tag{6}$$

where $\mathcal{F}_t = [\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots; x_t, x_{t-1}, \ldots]$ is the information set (available up to time t) of model (2) and $h \in \mathcal{N}^+$ is the forecast horizon.

When $\mathcal{L}\left(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}\right) = (||\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}||)^2$, with $||\cdot||$ the Euclidean norm, the optimal predictor is

Figure 1. Time plots (of the first 100 observations) and contour-plots of two bivariate time series generated from model (5)



the conditional expectation $\mathbb{E}[\boldsymbol{y}_{t+h}|\mathcal{F}_t]$ that, for model (2) with $h \leq d$, has the following form:

$$\begin{aligned}\hat{\mathbf{y}}_{t+h} &= \mathbb{E}[\mathbf{y}_{t+h}|\mathcal{F}_t] \\ (7)&= \sum_{k=1}^{\ell}\left[\mathbf{v}^{(k)} + \mathbf{A}_1^{(k)}\mathbf{y}_{t+h-1} + \ldots + \mathbf{A}_p^{(k)}\mathbf{y}_{t+h-p} + \mathbf{M}_1^{(k)}\mathbf{u}_{t+h-1} + \ldots + \mathbf{M}_q^{(k)}\mathbf{u}_{t+h-q}\right] I(x_{t+h-d} \in \mathcal{R}_k),\end{aligned}$$

where

$$\mathbf{y}_{t+h-i} = \begin{cases}\mathbf{y}_{t+h-i} & \text{if } h \leq i \\ \mathbb{E}[\mathbf{y}_{t+h-i}|\mathcal{F}_t] & \text{otherwise}\end{cases} \quad \text{and} \quad \mathbf{u}_{t+h-j} = \begin{cases}\mathbf{u}_{t+h-j} & \text{if } h \leq j \\ \mathbf{0} & \text{otherwise,}\end{cases}$$

for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$.
When $h > d$ the threshold variable needs to be predicted as well. In this cases, generalizing what proposed in the univariate context in Amendola *et al.* (2007), the forecast can be obtained as weighted mean of the predictions generated by each regime:

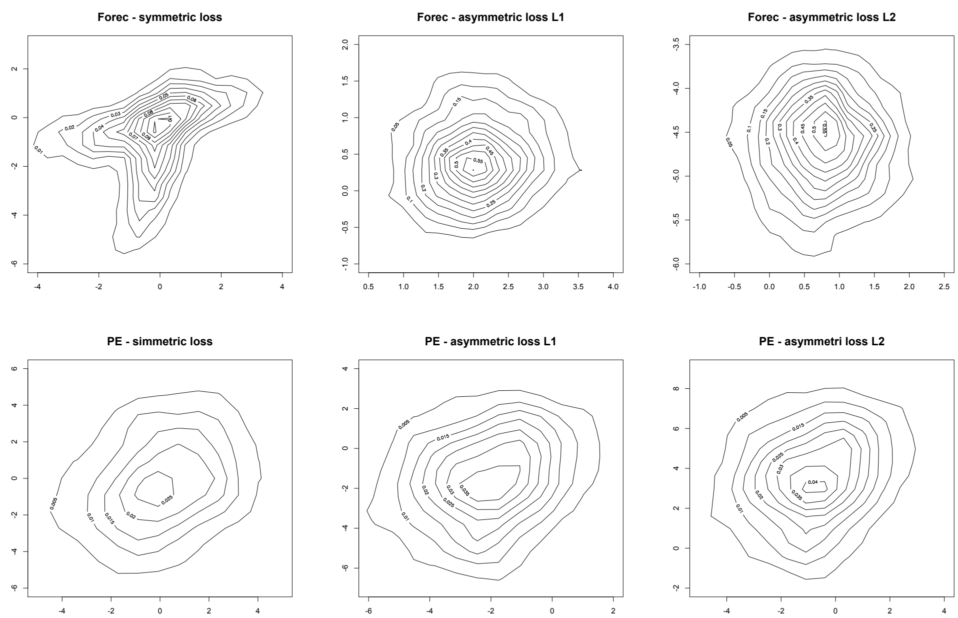
$$(8) \quad \hat{\mathbf{y}}_{t+h} = \sum_{k=1}^{\ell} \hat{\mathbf{y}}_{t+h}^{(k)} p^{(k)}, \quad \text{with } h > d$$

where $p^{(k)} = P(x_{t+h-d} \in \mathcal{R}_k|\mathcal{F}_t)$ is the probability that the threshold variable belongs to regime k, given the information set $\mathcal{F}_t$, and $\hat{\mathbf{y}}_{t+h}^{(k)} = \mathbb{E}[\mathbf{y}_{t+h}^{(k)}|\mathcal{F}_t]$.

The symmetry of the loss function under which the predictor (7) is based could not be of interest when the data generating process shows unquestionable asymmetries. In this case an asymmetric loss should be selected to take into account the features of the process. In empirical domain an asymmetric loss could be selected by agents interested to generate forecasts that differently treat the cost of over or under prediction.

This kind of problem has been largely discussed for univariate time series (see among the others Christoffersen and Diebold (1996, 1997), Granger (1999), Patton and Timmermann (2007a, 2007b,

Figure 2. Contour plots of the forecasts (Forec) and of the prediction errors (PE) generated through a symmetric loss function and the asymmetric loss function (9) with $\boldsymbol{\tau} = (0.5, 0.5)$ ***(L1) and*** $\boldsymbol{\tau} = (0.36, -076)$ ***(L2)***



2010)) but, in our knowledge, in the multivariate domain it has been only recently faced in Alp and Demetrescu (2010) and Komunjer and Owyang (2010).

The approach used in the present paper is based on the loss function $\mathcal{L}(\boldsymbol{\tau}; \mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}) : \mathcal{B}_q^K \times \mathcal{R}^K \to \mathcal{R}$ proposed in Komunjer and Owyang (2010) given by:

$$(9) \quad \mathcal{L}(\boldsymbol{\tau}; \mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}) = \left[||\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}||_p + \boldsymbol{\tau}'(\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}) \right] ||\mathbf{y}_{t+h} - \hat{\mathbf{y}}_{t+h}||_p^{p-1}$$

where $|| \cdot ||_p$ is the l_p norm, $1 \leq p < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$ and $||\boldsymbol{\tau}||_q < 1$, with $\boldsymbol{\tau}$ the K-vector of asymmetry.

In more datail we use the general loss function (9) with $p = 2$ ($q = 2$ consequently) and $h = 1$, assuming that the vector $\boldsymbol{\tau}$ is known. In other words the assumption on $\boldsymbol{\tau}$ means that the asymmetry of the function (9) is defined by agents that generate forecasts.

The minimization (6), under the loss function (9), is carried out using a Monte Carlo procedure based on two main steps:

1. *Estimation of the conditional distribution of* $\mathbf{y}_{t+1}$. After generating n time series of length $t+1$ from model (2) the conditional distribution of $\mathbf{y}_{t+1}$ has been estimated non parametrically using a kernel approach (Venables and Ripley, 2002);

2. *Minimization.* Following Alp and Demetrescu (2010), the optimal forecast $\hat{\mathbf{y}}_{t+1}$ is obtained from

$$\arg\min_{\hat{\mathbf{y}}_{t+1}} \frac{1}{W} \sum_{w=1}^{W} \mathcal{L}(\mathbf{y}_{t+1}^{w} - \hat{\mathbf{y}}_{t+1})$$

where $\mathbf{y}_{t+1}^{w}$ is a pseudo-random number drawn from the conditional distribution of $\mathbf{y}_{t+1}$ and W is the number of replicates that needs to be enough large.

To evaluate the proposed procedure we have simulated $N = 1000$ time series of length $T = 500$ from model (5) with $\mathbf{v}^{(1)} = (0, 0)$ and $\mathbf{v}^{(2)} = (0, -1)$ and for each of them we have generated forecasts

using three predictors based on: 1) the conditional expectation (7); 2) the loss function (9) with $p = 2$, $h = 1$ and $\boldsymbol{\tau} = (0.5, 0.5)$, denoted L1; the loss function (9) with $p = 2$, $h = 1$ and $\boldsymbol{\tau} = (0.36, -076)$, denoted L2. The forecast generation through L1 and L2 has been performed following the two steps of the Monte Carlo procedure described above, with $n = 500$ and $W = 10000$.
It is well known that a way to evaluate the predictors is based on the study of their distributions. In this regard the contour plots of the densities estimated for $\hat{\mathbf{y}}_{t+1}$ and the corresponding prediction errors $\hat{\mathbf{u}}_{t+1} = \mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1}$ are presented in Figure 2 were it can be noted that the shape of the contour plot of $\hat{\mathbf{y}}_{t+1}$ in presence of a symmetric loss, is related to the fact that it can be seen as a mixture of distributions. The contour plots of the prediction errors show a clear dependence between $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$ and further under L1 and L2 it can be noted a bias that is strictly related to the asymmetry of the loss function. This bias, here empirically evaluated, has been largely investigated in the univariate domain in Patton and Timmermann (2007a) and will be object of future research.

REFERENCES (RÉFÉRENCES)

Alp T., Demetrescu M. (2010), Joint forecast of Down Jones stocks under general multivariate loss function, *Computational Statistics and Data Analysis*, 54, 2360-2371.

Amendola A., Niglio M., Vitale C.D. (2007), Least squares predictors for threshold models: properties and forecast evaluation, in *Mathematical and Statistical Methods for Insurance and Finance* (C. Perna and M. Sibillo eds), Springer, 1-9.

Box G.E.P., Jenkins G.M., Reinsel G.C. (2008), *Time Series Analysis. Forecasting and Control*, Wiley, Hoboken, New Jersey

Christoffersen P.F., Diedold F.X. (1996), Further results on forecasting and model selection under asymmetric loss, *Journal of Applied Econometrics*, 11, 561-571.

Christoffersen P.F., Diedold F.X. (1997), Optimal prediction under asymmetric loss, *Econometric Theory*, 13, 808-817.

Fan J., Yao Q. (2003), *Nonlinear Time Series. Nonparametric and Parametric Methods*, Springer, New York.

Granger C.W.J. (1999), Outline of forecast theory using generalized cost functions, *Spanish Economic Review*, 1, 161-173.

Kollo T. (2008), Multivariate skewness and kurtosis measures with an application in ICA, *Journal of Multivariate Analysis*, **99**, 2328-2338.

Komunjer I., Owyang M.T.(2010), Multivariate forecast evaluation and rationality, Mimeo.

Lütkepohl H. (2006), *New Introduction to Multiple Time Series Analysis*, Springer Verlag, N.Y.

Mardia K.V. (1970), Measures of multivariate skewness and kurtosis with applications, *Biometrika*, 57, 519-530.

Niglio M., Vitale C.D. (2010), Quasi-Maximum Likelihood Estimators for Threshold ARMA Models: Theoretical Results and Computational Issues, in *Proceedings of COMPSTAT 2010, 19th International Conference on Computational Statistics* (Y. Lechevallier, G. Saporta eds.), Physica-Verlag, 1461-1468.

Patton A.J., Timmermann A. (2007a), Properties of optimal forecasts under asymmetric loss and nonlinearity, *Journal of Econometrics*, 140, 884-918.

Patton A.J., Timmermann A. (2007b), Testing forecast optimality under unknown loss, *Journal of the American Statistical Association*, 102, 1172-1184.

Patton A.J., Timmermann A. (2010), Generalized forecast errors, a change of measure and forecast optimality, in *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle* (Bollerslev T, Russel JR, Watson MW eds.), Oxford University Press.

Tong H. (1990), *Non-Linear Time Series. A Dynamical System Approach*, Clarendon Press, Oxford.

Tong H. (1983), *Threshold Models in Nonlinear Time Series Analysis*, Springer-Verlag, London.

Venables W. N., Ripley, B. D. (2002), *Modern Applied Statistics with S*, Springer.