STATISTICA, anno LXXX, n. 4, 2020

REVISITING THE CANADIAN LYNX TIME SERIES ANALYSIS THROUGH TARMA MODELS

Greta Goracci [1]
Dipartimento di Scienze Statistiche, Università di Bologna, Bologna, Italia

1. Introduction

In many situations, the complexity of real-world phenomena calls in for non-linear modelling. The class of threshold models has been recognized to be a flexible tool to describe non-linear features such us jumps, limit cycles, time irreversibility, chaos, etc. Indeed, threshold non-linearity offers a quite simple easy-to-face approximation of general complex non-linear dynamics while retaining a good interpretability. Threshold models were introduced by Tong (1978) and are based on the threshold principle: when the phenomenon crosses a certain threshold then it changes qualitatively. In a seminal work, Tong and Lim (1980) presented the threshold autoregressive moving-average (TARMA) model defined by the following difference equation:

$$X_t = \begin{cases} \phi_{1,0} + \sum_{i=1}^{p} \phi_{1,i} X_{t-i} + \varepsilon_{1,t} + \sum_{j=1}^{q} \theta_{1,j} \varepsilon_{1,t-j} & \text{if } X_{t-d} \leq r_1 \\ \phi_{2,0} + \sum_{i=1}^{p} \phi_{2,i} X_{t-i} + \varepsilon_{2,t} + \sum_{j=1}^{q} \theta_{2,j} \varepsilon_{2,t-j} & \text{if } r_1 < X_{t-d} \leq r_2 \\ \vdots & \vdots \\ \phi_{l,0} + \sum_{i=1}^{p} \phi_{l,i} X_{t-i} + \varepsilon_{l,t} + \sum_{j=1}^{q} \theta_{l,j} \varepsilon_{l,t-j} & \text{if } X_{t-d} > r_{l-1}, \end{cases} \tag{1}$$

where (i) l is the number of regimes; (ii) p and q are the autoregressive and moving-average order, respectively; (iii) $\phi_{k,i}$ and $\theta_{k,j}$ with $k = 1,\dots,l$, $i = 1,\dots,p$, and $j = 1,\dots,q$ are the autoregressive and moving-average parameters, respectively; (iv) $d \in \mathbb{N}$ is the delay parameter; (v) $r_1 < r_2 < \dots < r_{l-1}$ are the threshold parameters; (vi) $\{\varepsilon_{k,t}\}$ with $k = 1,\dots,l$ are the error processes (innovations). Clearly, both the orders p and q and the innovation processes can be the same across regimes. Furthermore, note that the threshold variable can be either endogenous, such as X_{t-d} or a transformation of the process e.g. $(X_{t-d} - X_{t-d-1})$, or even exogenous. The class of TARMA models

[1] Corresponding Author. E-mail: greta.goracci@unibo.it

subsumes two families of models that are of independent interest: threshold autoregressive models (TAR) and threshold moving-average models (TMA). From Eq. (1), TAR, TMA and TARMA models are an extension of the corresponding well known AR, MA, ARMA models, allowing the parameters to change across regimes. TAR models have been widely analysed from both a probabilistic and a statistical point of view. For results concerning the long-run probabilistic behaviour of TAR models see, among others, Chan and Tong (1985); Guo and Petruccelli (1991); Tong (1990); Cline (2009); An and Chen (1997); Chen and Tsay (1991); Liu and Li (1997); Li and Ling (2012); whereas for results on the estimation see, for instance, Chan (1993),Chan and Tsay (1998) and Qian (1998). The TAR framework has been widely used to develop statistical tests to detect the presence of non-linearity and/or the presence of non-stationarity. See, among others, Chan (1990); Kapetanios and Shin (2006); Bec *et al.* (2008); Giordano *et al.* (2017). See Hansen (2011) and Chen *et al.* (2011) for a review of the threshold models in economics and finance, respectively. Moreover, see also Tong (2007, 2011, 2017) for more recent reviews and discussions. The investigation of TMA and TARMA models is underdeveloped and presents several gaps. Indeed, contrary to the TAR case, both the TMA and the TARMA model have a non-linear parameterization conditionally on the threshold. The presence of the moving-average component in a non-linear setting produces a very complex dynamics and the theory developed for the TAR case cannot be adapted straightforwardly. As concerns TMA models, Ling *et al.* (2007) and Chan and Tong (2010) provided some results on invertibility and ergodicity. The estimation for TMA models has been faced by Gooijer (1998) that focused upon the maximum likelihood estimation under the assumption of Gaussian innovations, whereas Li *et al.* (2013) proposed the least squares estimation for general TMA models. Moreover, Ling and Tong (2005) developed a test for the presence of threshold non-linearity in MA models. The TARMA framework is the most challenging to handle and few results are available. Recently, Chan and Goracci (2019) derived the long-run probabilistic structure of the TARMA$(1, 1)$ process by solving the long-standing open problem of finding an irreducible Markovian representation for TARMA models. The theory regarding the estimation for TARMA models is underdeveloped in that the only result concerns the least squares estimation derived by Li *et al.* (2011). To the best of our knowledge, there are no results on the maximum likelihood estimation for TARMA models. The TARMA framework has been used to design tests for non-linearity (Goracci and Giannerini, 2020; Li and Li, 2011) and threshold regularization (Chan *et al.*, 2019).

As proved in Chan and Goracci (2019), the long-run probabilistic behaviour depends only on the autoregressive part of the two extreme regimes. Hence TARMA models turn out to be very appealing to describe series that appears as random walk in first place because they can have a unit-root in some regimes but be globally stationary. Moreover, TARMA models can encompass a plethora of long-run behaviors, e.g. (geometric) ergodicity, null recurrence and transience. Note that a TAR process naturally turns into a TARMA process if the time series is subjected to measurement error. Since the data are almost always corrupted my measurement error, the TARMA framework is more appropriate than a pure AR-type setting. Also, in analogy with the AR/ARMA dichotomy in the linear framework, Goracci (2020) showed that TARMA(1,1) models can provide a better approximation with respect to both TAR(p) and AR(p) models even when p is large.

Thus, TARMA models hold an enormous potential to analyse non-linear dynamics in different fields. For instance, Chan *et al.* (2020) used TARMA models to address the purchasing power parity puzzle; Goracci (2020) analyzed the sunspot number and the male US unemployment rate time series whereas Goracci and Giannerini (2020) deployed them for dendrological analysis. The results show that TARMA models can provide a better and more parsimonious fit with respect to other models presented in literature, including TAR models.

In this work, we use the TARMA framework to revisit the analysis of the benchmark *Canadian lynx* time series and we show that TARMA models perform better than the proposed TAR models. To this end, we compare 11 different models: 5 TAR models and 6 TARMA models. Clearly, since there is no *true* model but its choice should reflects the purpose of the analysis we compare the performance of the models from different prospectives. The rest of the paper is organized as follows: in Section 2 we describe the Canadian lynx time series; in Section 3 we apply the tests recently proposed within the TARMA framework to detect the presence of unit-root and non-linearity; Section 4 describes the 11 competitor models; in Section 5 we fit the models to the data and compute the resulting information criteria AIC and BIC; the capability of the models to reproduce cycles and the multimodality of the series is compared in Section 6 and Section 7, respectively; Section 8 focuses upon their prediction accuracy; in Section 9 we carry out the diagnostic analysis of the models; lastly we conclude in Section 10.

2. The Canadian lynx time series and threshold models

In an ecosystem the relationships between populations can be classified into four types: mutualism, parasitism, competition and predation. Indeed, the growth of a population is influenced by several factors, such as the competition for resources, the amount of available food, the presence of predators. Here our focus is upon the predator-prey interaction that is characterized by a complex non-linear dynamics related to food chains and food networks. In this respect, one of the first mathematical models was formalized by Lotka and Volterra that proposed a system of first order non-linear differential equations (see Doob, 1936; Herbert, 1959, and references therein). Several systems of both deterministic and stochastic equations have been proposed to model a predator-prey type population dynamics. See, among others, Rudnicki (2003); Wang *et al.* (2019); Wu and Zhu (2008). The predator-prey dynamics usually has a limit cycle that can be synthesized as follows: when the population of predators decreases, there is a resulting increase of preys and habitat resources which leads to a new increasing phase for predators; on the other hand, when the number of predators becomes too large, the preys reduce and hence the population of predators enters in a new decreasing phase.

The Canadian lynx data set is the annual record of the number of the Canadian lynx trapped in the Mackenzie River district of the North-West Canada for the period 1821-1934 inclusively. The population cycle of these animals has received wide attention especially from biologists due to the regularity in the hunted quantities by the Hudson's Bay Company that has been using them for a long period to produce furs. The data set and further useful materials for the analysis are reported in Elton and Nicholson (1942). These data are the total fur return, or total sales, from the London archives of the aforementioned company and is a proxy of the dynamics of the population size. Note that there is a time lag between the year in which a lynx was trapped and the year in which its fur was sold and this complicates the analysis. See also Kajitani *et al.* (2005) for more recent discussions. This data set has attracted great attention among non-linear time series analysts due to its asymmetric cycle, i.e. the increasing phase is slower than the decreasing phase, that makes the investigation very challenging. Interestingly, it has been noted that the lynx-hare interaction is not instantaneous, but presents a time lag of about two years (see De Gooijer, 2017, Section 7.5). See Stenseth *et al.* (1998) and references therein for a comprehensive analysis of the series.

A plethora of models has been proposed to describe the Canadian lynx time series; see, for instance, Lim (1987); Lin and Pourahmadi (1998); De Gooijer (2017) for a summary of presented models. Among all the candidates, the class of TAR models turned out to be the most appropriate (Tong, 1990, Chapter 7, Section 7.2). The reasons behind their success are diverse. First, threshold models are able to display limit cycles, a peculiar aspect of this time series. Moreover, they admit an easy biological interpretation as each regime reflects a different phase: the lower regime and the upper regime corresponds to the increasing and decreasing phase, respectively (see, for instance, Tong, 1990).

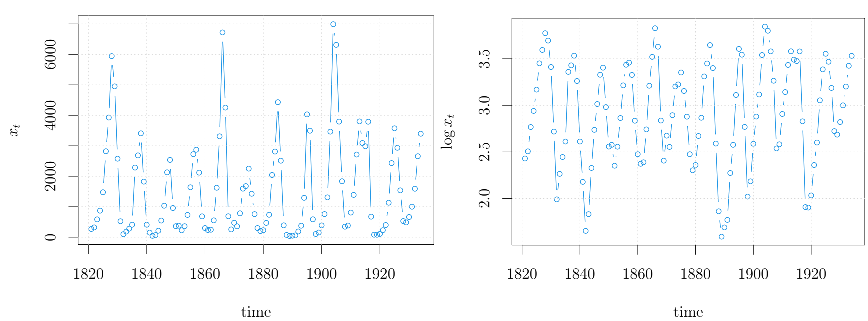

Figure 1 – Time series of the Canadian lynx time series, from 1821 to 1934. (Left) raw time series. (Right) log-transformed series.

Let $\{x_t\}$ be the Canadian lynx time series and denote with $\{y_t\}$ its $\log_{10}$ transformation. In Figure 1 we show both the raw (left) and the $\log_{10}$ transformed series (right). The $\{y_t\}$ time series clearly shows an asymmetric cycle where the increasing phase and the decreasing phase take about 6 and 3 years, respectively. Figure 2 shows the 2-dimensional lag plots of y_t versus y_{t-1} (left) and y_{t-2} (right), whereas in Figure 3 we present the 3-d lag plot of $(y_{t-1}, y_{t-1}, y_{t-2})$. These plots highlight the non-linear oscillatory nature of the lynx time series.

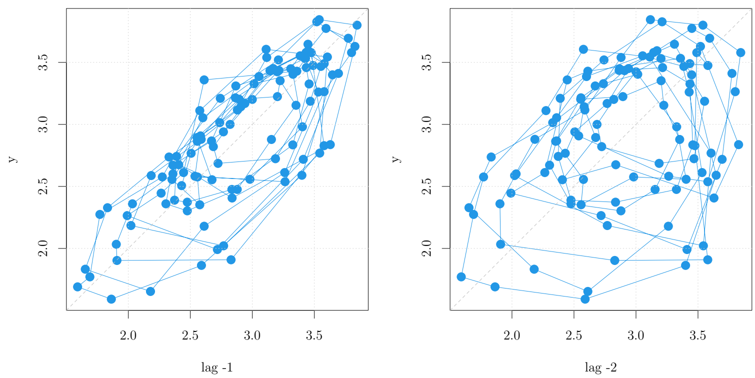

Figure 2 – Lag plot of the time series of the Canadian lynx, from 1821 to 1934. (Left) Lag 1. (Right) Lag 2.

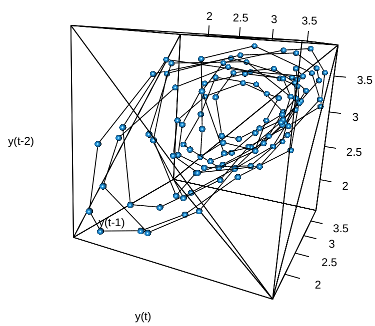

Figure 3 – Three-dimensional lag plot of the time series of the Canadian lynx, from 1821 to 1934.

Figure 4 reports the spectral density of the Canadian lynx time series. As can be seen, there are three main peaks that reveal a periodic behaviour with three main cycles. The highest peak corresponds to a periodicity of about 9-10 years, consistently with the underlying population dynamics that presents an asymmetric cycle. Lastly, Figure 5 shows the histogram and the corresponding non-parametric kernel density estimation of $\{y_t\}$ that point at a bimodal behaviour of the series. Hence, the main features of the series are: (i) the presence of an asymmetric cycle where the increasing phase is faster than the decreasing phase; (ii) multiple peaks for its spectral function; (iii) bimodality. In this work, we will compare the capability of models to reproduce these features as well as their forecasting performance. We start the analysis by detecting the possible presence of unit-roots and non-linearity in the Canadian lynx time series.

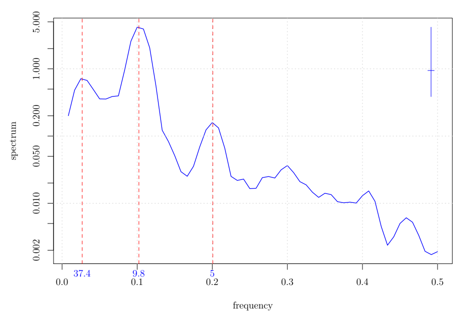

Figure 4 – Spectral density of the Canadian lynx, from 1821 to 1934. There are three main peaks corresponding to a period of 37.4, 9.8 and 5 respectively. The highest one is related to the underlying population dynamic that presents an asymmetric cycle where the increasing phase and the decreasing take about 6 and 3 years, respectively.

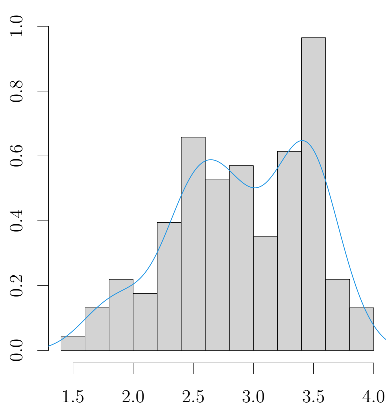

Figure 5 – Histogram and the non-parametric density estimation (blue line) of the time series of the Canadian lynx, from 1821 to 1934.

3. Stationarity and non-linearity

In this section we apply tests recently proposed within the TARMA framework to investigate the presence of unit-roots and non-linearity in the Canadian lynx time series. Testing for the presence of a unit-root in time series has important practical implications as it is attested by vast amount of literature devoted to the problem (see, e.g., Patterson, 2010, 2011, 2012; Choi, 2015). Indeed, the exercise presents several non-trivial theoretical and practical issues (see also Haldrup and Jansson, 2006, and references therein). Early unit-root tests are based upon the framework of ARIMA models and are affected by power loss when the alternative is non-linear and severe size distortion in presence of MA components. To cope with such issues, some unit-root tests have been developed within the framework of TAR models (Enders and Granger, 1998; Caner and Hansen, 2001; Bec *et al.*, 2004; Kapetanios and Shin, 2006; Bec *et al.*, 2008; Seo, 2008; Park and Shintani, 2016; de Jong *et al.*, 2007; Giordano *et al.*, 2017). These tests have good power in most situations but are still affected by severe size distortion in presence of dependent errors, especially of the moving-average kind (for a detailed account see Patterson, 2011, Chapters 6 and 9). Only recently, Chan *et al.* (2020) used TARMA models to develop a novel Lagrange multiplier unit-root test with good size and power, allowing for a wide and flexible non-linear alternative and robust against heteroskedasticity. The test has been designed to detect the presence of regulation, a phenomenon that plays an important role in biological growth and population fluctuations. It specifies an IMA(1, 1) model as the null hypothesis and a TARMA(1, 1) with a unit-root regime as the alternative. The key aspect that distinguishes this test from other proposals is that, for the first time, the moving-average component is directly included both in the null and the alternative hypothesis. The associated theoretical derivations are highly non-standard and more challenging with respect to tests based upon TAR models. Moreover, note that the IMA(1,1) and the TARMA(1,1) models are capable of encompassing a wide range of stationary and non-stationary linear and non-linear dynamics.

The model under the null hypothesis is

$$X_t = \phi_0 + X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1},$$

whereas, under the alternative hypothesis we have

$$X_t = \phi_0 + X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} + \{\Psi_0 + \Psi_1 X_{t-1}\} \times I(X_{t-1} \leq r).$$

Hence, letting $\boldsymbol{\Psi} = (\Psi_0, \Psi_1)^\intercal$ and $\mathbf{0}$ be the zero vector, the system of hypotheses reduces to:

$$H_0 : \boldsymbol{\Psi} = \mathbf{0} \quad \text{vs} \quad H_1 : \boldsymbol{\Psi} \neq \mathbf{0}.$$

The threshold parameter r is absent under the null, hence the test statistic is obtained as the supremum Lagrange multiplier test statistic:

$$T_n = \sup_{r \in [r_L, r_U]} T_n(r),$$

where $T_n(r)$ is the Lagrange multiplier test statistic for a fixed r. The points r_L and r_U are taken to be some percentile of the data. Chan *et al.* (2020) proved that the test is consistent and asymptotically similar in that its asymptotic null distribution does not depend on the value of the MA parameter. Moreover, they introduced a wild bootstrap version of the supLM statistic that possesses good properties in finite samples and is robust against heteroskedasticity.

In this application we take r_L and r_U to be the 25th and 75th percentile of the data, respectively. The value of the test statistic results 26.543, whereas the bootstrap p-value is 0.001. Hence, we reject the null hypothesis of presence of unit-root and the series can be treated as stationary.

Having ascertained the stationarity of the series we focus on testing whether it is stationary and linear or stationary with threshold non-linearity. This kind of tests are different from unit-root tests, not only from the practical/usage point of view, but also from the theoretical prospective. Indeed, under the null hypothesis, the model is stationary so that the theoretical framework is different from that of unit-root tests where, under H_0, the process is not stationary. Also in the family of tests for linearity, the threshold framework has been largely exploited. Chan (1990) developed a test for AR models against their threshold extension, whereas Ling and Tong (2005) and Li and Li (2008) focused on threshold non-linearity in MA models. Li and Li (2011) extended their works for ARMA against TARMA models. All these tests are quasi-likelihood ratio tests and require an estimated model under the alternative through the maximum likelihood approach. As mentioned, there are no results on the maximum likelihood estimators for TARMA models. Indeed, the test proposed by Li and Li (2011) is oversized even for large sample sizes so that its practical application is questionable. In order to overcome these issues, Goracci and Giannerini (2020) proposed two Lagrange multiplier tests to compare a linear ARMA specification against its TARMA extension that only require estimation of the null model. They showed that the tests enjoy very good finite-sample properties, are robust against model mis-specification and their performance is not affected if the order of the model tested is unknown and a consistent model selection procedure is adopted. The two tests are denoted as sLM and sLM*: in the sLM test only the autoregressive part is tested for threshold non-linearity whereas in the sLM* test, both the autoregressive and the moving average part are tested. Here, we assume $d = 2$ and determine the autoregressive and moving-average orders through the Hannan-Rissanen model selection procedure. The procedure selects the ARMA$(2, 1)$ as the optimal model:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t - \theta\varepsilon_{t-1},$$

whereas under the alternative hypothesis we have:

$$
\begin{aligned}
X_t &= \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t - \theta\varepsilon_{t-1} \\
&\quad + \left\{\Psi_{10} + \Psi_{11} X_{t-1} + \Psi_{12} X_{t-2}\right\} \times I(X_{t-2} \le r), \quad \text{for sLM;}
\end{aligned}
$$

$$
\begin{aligned}
X_t &= \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t - \theta\varepsilon_{t-1} \\
&\quad + \left\{\Psi_{10} + \Psi_{11} X_{t-1} + \Psi_{12} X_{t-2} - \Psi_{21}\varepsilon_{t-1}\right\} \times I(X_{t-2} \le r), \quad \text{for sLM}^\star .
\end{aligned}
$$

By setting $\boldsymbol{\Psi}$ to be the vector containing all the Ψ's and $\mathbf{0}$ the zero vector, the hypotheses reduce to:

$$
\begin{cases} H_0 : \boldsymbol{\Psi} = \mathbf{0} \\ H_1 : \boldsymbol{\Psi} \neq \mathbf{0} \end{cases}
$$

As for the unit-root test, the threshold parameter r is absent under the null and hence the test statistic is obtained as the supremum Lagrange multiplier test statistic over the data-driven interval $[r_L, r_U]$. Goracci and Giannerini (2020) derived the asymptotic distribution of the statistics under the null hypothesis and local contiguous alternatives and prove the consistency of the tests. Moreover, they showed empirically their similarity. The sLM and sLM* test statistics result 30.258 and 31.929, respectively. The corresponding critical value are reported in Table 1. The two tests reject the null hypothesis of linear ARMA also at the nominal level 99.9%. The results motivate the use of threshold model to the Canadian lynx time series.

TABLE 1

Critical values for the asymptotic null distribution of the sLM and sLM* *statistics when the autoregressive order and moving-average order are 2 and 1, respectively. The threshold range is* $[r_L, r_U]$ *with* r_L *and* r_U *being the 25th and 75th percentile, respectively.*

		sLM				sLM*			
AR	MA	90%	95%	99%	99.9%	90%	95%	99%	99.9%
2	1	11.53	13.41	17.22	22.17	13.48	15.46	19.63	25.60

4. The models

In this section we present the panel of 11 models that we compare from different prospectives. We consider 5 two-regime TAR models and 6 two-regime TARMA models. First, we include the following 3 TAR models presented in Tong (1983, 1990)

$$\text{M1: } X_t = \begin{cases} 0,546 + 1,032X_{t-1} - 0,173X_{t-2} \\ +0,171X_{t-3} - 0,431X_{t-4} + 0,332X_{t-5} \\ -0,284X_{t-6} + 0,210X_{t-7} + \varepsilon_{1,t}, & \text{if } X_{t-2} \leq 3,116 \\ 2,632 + 1,492X_{t-1} - 1,324X_{t-2} + \varepsilon_{2,t}, & \text{if } X_{t-2} > 3,116. \end{cases}$$

$$\text{M2: } X_t = \begin{cases} 0,768 + 1,064X_{t-1} - 0,200X_{t-2} + 0,164X_{t-3} \\ -0,428X_{t-4} + 0,181X_{t-5} + \varepsilon_{1,t}, & \text{if } X_{t-2} \leq 3,05 \\ 2,254 + 1,474X_{t-1} - 1,202X_{t-2} + \varepsilon_{2,t} & \text{if } X_{t-2} > 3,05. \end{cases}$$

$$\text{M3: } X_t = \begin{cases} 0,62 + 1,25X_{t-1} - 0,43X_{t-2} + \varepsilon_{1,t} & \text{if } X_{t-2} \leq 3,25 \\ 2,25 + 1,52X_{t-1} - 1,24X_{t-2} + \varepsilon_{2,t} & \text{if } X_{t-2} > 3,25. \end{cases}$$

The variances of the innovation terms are reported in Tong (1990, chapter 7). Model M3 is the simplest and has a nice biological interpretation (see Tong, 1990; Stenseth *et al.*, 1998; Fan and Yao, 2003, for comments). On the other hand, as we will show, the other models represents an improvement in terms of statistical fitting.

TABLE 2

Models used to analyse the Canadian lynx time series. The "AR lags" column indicates the subset of lags whose parameters are fixed across regimes, whereas the "TAR lags" column refers to the subset of lags whore parameters are regime-dependent. Moreover, d is the delay parameter and "MA ord" is the moving-average order. The last column indicates whether the intercept is common between regimes or it is regime-dependent.

Model	AR lags	TAR lags	MA ord	d	intercept
M4	4,12	1,2,9	0	2	common
M5		1:7	0	2	regime-dependent
M6	4,12	1,2,9	1	2	common
M7	4,9,12	1,2	1	2	common
M8		1:4,6,7,10	1	2	regime-dependent
M9	4,9,12,16	1,2	1	2	common
M10	4,12,16	1,2,9	1	2	common
M11		1,2	1	2	regime-dependent

The remaining models M4 – M11 are synthesized in Table 2. The "AR lags" column indicates the subset of lags whose parameters are fixed across regimes, whereas the "TAR lags" column refers to the subset of lags whose parameters are regime-dependent. Moreover, d is the delay parameter and "MA ord" is the moving-average order, the moving-average part is assumed to be fixed across regimes. The last column indicates whether the intercept is common between regimes or it is regime-dependent. For instance, M4 is described by the following difference equation:

$$X_t = \phi_0 + \phi_4 X_{t-4} + \phi_{12} X_{t-12} + \varepsilon_t \\ + \begin{cases} \phi_{1,1} X_{t-1} + \phi_{1,2} X_{t-2} + \phi_{1,9} X_{t-9}, & \text{if } X_{t-2} \leq r \\ \phi_{2,1} X_{t-1} + \phi_{2,2} X_{t-2} + \phi_{2,9} X_{t-9}, & \text{if } X_{t-2} > r. \end{cases}$$

Models M4 and M5 are TAR processes whereas models M6 – M11 are TARMA. All these models have common features consistent with the underlying biology, e.g. (i) the delay parameter d is always 2 and this is related to the time lag between the dynamic of the lynx population and that of the hares; (ii) there is a TAR effect at lags 1 and 2 (see Tong, 1990; Chan and Tong, 2001; Cryer and Chan, 2008; De Gooijer, 2017, and references therein for more details).

5. Parameter estimation and goodness-of-fit

In this section we fit models M4 – M11 to the Canadian lynx time series. We adopted the estimation procedure presented in Giannerini and Goracci (2020) that deploys the well-tested maximum likelihood framework available for ARMA models where the likelihood is derived via a state-space representation of the ARMA process by means of the Kalman filter. The idea relies on considering the autoregressive parts as regressors and, hence, reducing the problem to estimating a regression with ARMA errors. The implementation is flexible as it is possible to have (i) an autoregressive part common among regimes; (ii) regime dependent autoregressive orders; (iii) specific subset of lags to be included in each autoregressive part; (iv) arbitrary number of regimes. Note that since the MA parameters do not change across regimes, the maximum likelihood framework is well established and can be exploited. The estimated thresholds are reported in Table 3, whereas Tables 4 and 5 show the estimates for the other parameters from models M4 – M11.

TABLE 3
Estimated thresholds from models M4 – M11 of Table 2 on the Canadian lynx data.

	M4	M5	M6	M7	M8	M9	M10	M11
r	3.31	3.31	3.31	3.31	3.36	3.35	3.35	3.31

First, we compare the models with respect to their goodness-of-fit by means of the AIC and the BIC. Information criteria are based on the derivation of the likelihood function. Since, given the threshold r and the delay parameter d, TAR models have a linear parametrization, the profile likelihood function can be factorized as the sum of the likelihood functions associated to the two regimes. Hence the AIC and BIC can be derived separately for each regime of and the overall information criterion is the sum of those

TABLE 4
Parameter estimates from models M4 – M7 of Table 2 on the Canadian lynx data.

	MA	int.	AR lags			TAR lags							
MODEL M4		ϕ_0 1.02 (0.25)	ϕ_4 -0.14 (0.04)		ϕ_{12} -0.18 (0.04)	$\phi_{1,1}$ 0.93 (0.09)	$\phi_{1,2}$ -0.04 (0.10)						$\phi_{1,9}$ 0.11 (0.05)
						$\phi_{2,1}$ 1.18 (0.12)	$\phi_{2,2}$ -0.43 (0.12)						$\phi_{2,9}$ 0.20 (0.07)
MODEL M5		$\phi_{1,0}$ 0.56 (0.29)				$\phi_{1,1}$ 1.05 (0.10)	$\phi_{1,2}$ -0.19 (0.17)	$\phi_{1,3}$ 0.07 (0.17)	$\phi_{1,4}$ -0.28 (0.16)	$\phi_{1,5}$ 0.17 (0.16)	$\phi_{1,6}$ -0.2 (0.15)	$\phi_{1,7}$ 0.21 (0.10)	
		$\phi_{2,0}$ 0.93 (0.89)				$\phi_{2,1}$ 1.54 (0.14)	$\phi_{2,2}$ -1.01 (0.36)	$\phi_{2,3}$ 0.39 (0.34)	$\phi_{2,4}$ -0.45 (0.32)	$\phi_{2,5}$ 0.11 (0.38)	$\phi_{2,6}$ -0.09 (0.38)	$\phi_{2,7}$ 0.18 (0.17)	
MODEL M6	θ 0.37 (0.14)	ϕ_0 1.24 (0.31)	ϕ_4 -0.21 (0.06)		ϕ_{12} -0.21 (0.05)	$\phi_{1,1}$ 0.739 (0.11)	$\phi_{1,2}$ 0.147 (0.11)						$\phi_{1,9}$ 0.13 (0.06)
						$\phi_{2,1}$ 0.98 (0.14)	$\phi_{2,2}$ -0.25 (0.13)						$\phi_{2,9}$ 0.26 (0.07)
MODEL M7	θ 0.33 (0.14)	ϕ_0 1.10 (0.29)	ϕ_4 -0.19 (0.06)	ϕ_9 0.18 (0.05)	ϕ_{12} -0.21 (0.05)	$\phi_{1,1}$ 0.72 (0.11)	$\phi_{1,2}$ 0.15 (0.11)						
						$\phi_{2,1}$ 1.08 (0.13)	$\phi_{2,2}$ -0.26 (0.14)						

TABLE 5
Parameter estimates from models M8 – M11 Table 2 on the Canadian lynx data.

	MA	int.	AR lags				TAR lags							
Model M8	θ	$\phi_{1,0}$					$\phi_{1,1}$	$\phi_{1,2}$	$\phi_{1,3}$	$\phi_{1,4}$	$\phi_{1,6}$	$\phi_{1,7}$		$\phi_{1,10}$
	0.25	0.93					0.91	-0.04	-0.01	-0.21	-0.12	0.23		-0.06
	(0.15)	(0.37)					(0.12)	(0.15)	(0.12)	(0.09)	(0.09)	(0.09)		(0.07)
		$\phi_{2,0}$					$\phi_{2,1}$	$\phi_{2,2}$	$\phi_{2,3}$	$\phi_{2,4}$	$\phi_{2,6}$	$\phi_{2,7}$		$\phi_{2,10}$
		-0.02					1.24	-1.05	0.87	-0.35	-0.11	0.02		0.27
		(1.08)					(0.18)	(0.38)	(0.37)	(0.26)	(0.27)	(0.18)		(0.09)
Model M9	θ	ϕ_0	ϕ_4	ϕ_9	ϕ_{12}	ϕ_{16}	$\phi_{1,1}$	$\phi_{1,2}$						
	0.65	1.66	-0.21	0.21	-0.28	-0.09	0.57	0.25						
	(0.12)	(0.48)	(0.06)	(0.06)	(0.06)	(0.06)	(0.11)	(0.11)						
							$\phi_{2,1}$	$\phi_{2,2}$						
							0.89	-0.11						
							(0.15)	(0.15)						
Model M10	θ	ϕ_0	ϕ_4		ϕ_{12}	ϕ_{16}	$\phi_{1,1}$	$\phi_{1,2}$					$\phi_{1,9}$	
	0.65	1.68	-0.22		-0.27	-0.08	0.59	0.24					0.18	
	(0.13)	(0.47)	(0.06)		(0.06)	(0.06)	(0.11)	(0.11)					(0.07)	
							$\phi_{2,1}$	$\phi_{2,2}$					$\phi_{2,9}$	
							0.86	-0.13					0.24	
							(0.15)	(0.15)					(0.07)	
Model M11	θ	$\phi_{1,0}$					$\phi_{1,1}$	$\phi_{1,2}$						
	-0.20	0.59					1.34	-0.51						
	(0.15)	(0.12)					(0.08)	(0.09)						
		$\phi_{2,0}$					$\phi_{2,1}$	$\phi_{2,2}$						
		0.61					1.66	-0.92						
		(0.91)					(0.11)	(0.26)						

(for more details see Tong, 1990; Cryer and Chan, 2008). On the contrary, TARMA models still present a non-linear parameterization given the threshold and hence we derive a global likelihood function. Moreover, the estimation procedure for TARMA models consists in estimating regression models with MA errors and, hence, it is not possible to consider each regime separately. For more details see Giannerini and Goracci (2020). In order to achieve comparability, we compute the AIC and BIC as follows:

$$\mathrm{AIC} = (n-p)(1+\log 2\pi) + (n-p)\log\left(\frac{\mathrm{RSS}}{n-p}\right) + 4(p+1),$$
$$\mathrm{BIC} = (n-p)(1+\log 2\pi) + (n-p)\log\left(\frac{\mathrm{RSS}}{n-p}\right) + 2(p+1)\log(n-p),$$

where RSS indicates the residual sum of squares derived from the global model.

TABLE 6
Information criteria for models M1 – M11 presented in Table 2. The corresponding parameter estimation are reported in Table 4 and Table 5.

	AIC	BIC
M1	-25.53	6.54
M2	-27.58	-0.67
M3	-25.22	-6.20
M4	-54.24	-27.99
M5	-23.66	21.77
M6	-57.27	-31.02
M7	-57.70	-28.83
M8	-37.35	10.25
M9	-62.03	-33.59
M10	-60.75	-29.73
M11	-31.46	-9.71

Table 6 reports the information criteria for the eleven models M1 – M11. According to both information criteria, the best model is the TARMA model M9. The moving-average parameter is significantly positive and less than 1. The threshold effect involves only the first and second lags. The values of X_{t-1} and X_{t-2} influence X_t in different ways according to the phase, whereas the lagged values X_{t-4}, X_{t-9}, X_{t-12} and X_{t-16} seem to exert a global effect independent from the phase. The difference between the autoregressive parameters in the two regimes reflects the so-called phase-dependence and density-dependence (see, e.g., Fan and Yao, 2003, Chapter 4 and references therein for more details). Note that the AIC and BIC have to be treated with great care, especially in a non-linear setting. The crucial point is that model selection procedures based upon AIC and BIC require the true data generating process to be among the candidates, which is not always guaranteed.

To cope with this issue, other information criteria have been proposed. The most recently contribution is given by Hsu *et al.* (2019) that proposed a misspecification-resistant information criterion (MRIC) to perform consistent model selection without assuming the presence of the true model among the panel of candidates. This information criterion relies upon the computation of the mean squared predictor error, which has still not been derived for TARMA models and hence we cannot include it in our analysis.

In the framework of threshold models several specific information criteria have been developed. For instance, Kapetanios (2001) extended the theoretical results concerning the information criteria in selecting the lag order available only for AR models to TAR models; Galeano and Peña (2007) proposed a modified model selection criteria for the class of TAR models. To date, it is an open issue whether such criteria can be further extended to the general TARMA framework. See also Stone (1977) and Konishi and Kitagawa (1996) for more references.

6. Spectral analysis

The spectral density of the series is reported in Figure 4. There are three prominent peaks; the major peak occurs at frequency 0.102 that corresponds to a period of about 9-10 years. The other two peaks point to a shorter period of 5 years and a longer period of 37 years. We compare the capability of models M1 – M11 to reproduce cycles presented in the series. To this end, we simulate time series of length 10000 from the fitted models and compute their spectral density, see Figures 6 and 7. The dashed lines indicate the prominent peaks and the corresponding periods are also reported. All the models are able to describe the 9-10-year cycle. The TAR models M1, M2 and M4 cannot reproduce the other two periods. All the other models capture also the peak corresponding to the 5-year period. The spectra of models M3, M5, M6, M8 and M9 also present a peak at a period of around 30 years.

To complete the analysis, in Figures 18 – 20 of Appendix A, we show the time plots of the simulated time series and their skeletons, obtained removing the noise terms. All the skeletons capture the (6+3)-year asymmetric cycle. On the other hand, it is recognized that the behaviour of a process (e.g. its stationarity) is not implied by the behaviour of its skeleton (e.g. its stability) (see Tong, 1990). Moreover it is questionable whether considering the skeleton of a TARMA model is meaningful since it reduces to the skeleton of the corresponding TAR model. Hence, we consider the simulated trajectories: they appear qualitatively similar to the observed time series and still present the main cycle.

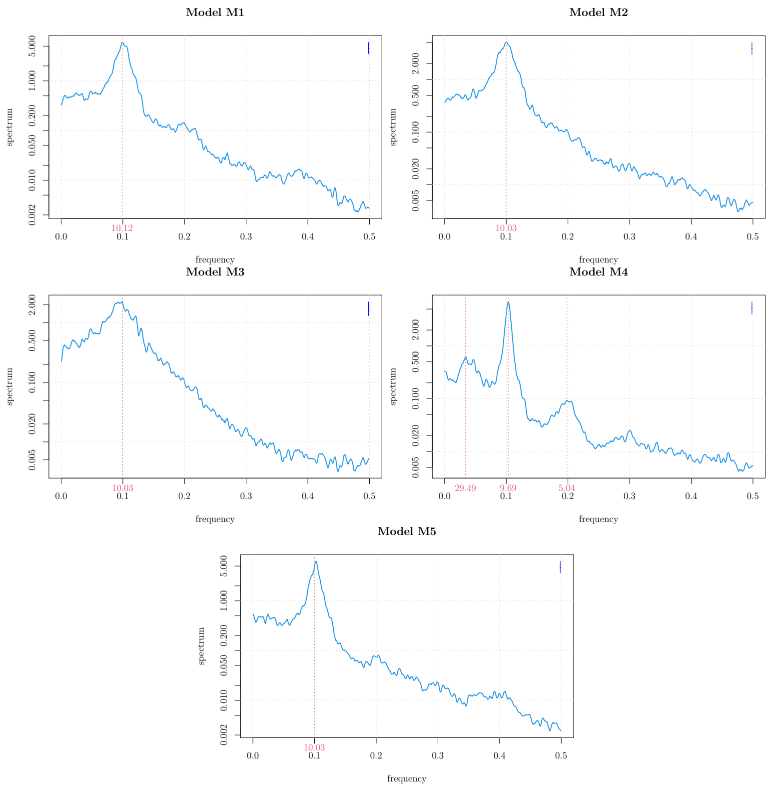

Figure 6 – Spectral density of time series of length 10000 simulated from TAR models. The dashed line indicate the prominent pikes and the corresponding periods are reported.

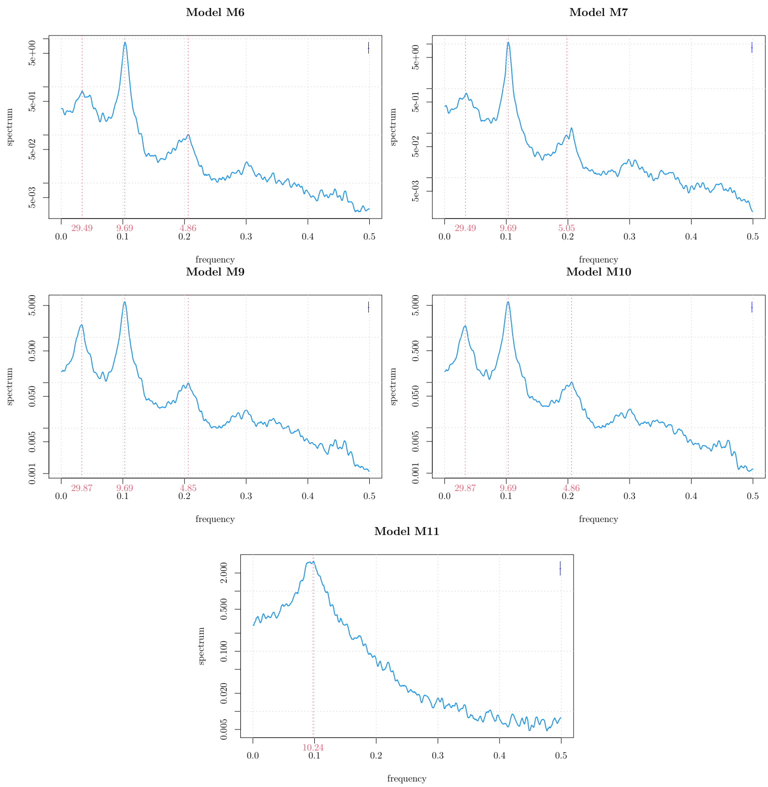

Figure 7 – Spectral density of time series of length 10000 simulated from TARMA models. The dashed line indicate the prominent pikes and the corresponding periods are reported.

7. Bimodality

In this section we study the capability of models M1 – M11 to reproduce the multimodality of the series. Figures 8 – 10 exhibit the histogram of the Canadian lynx time series and superimpose the non-parametric kernel density estimation of simulated trajectories from models M1 – M11. TAR models capture the small peaks in the left tail of the distribution better than TARMA models. Among TAR models, model M3 seems to have the best performance in terms of reproducing the bimodality of the time series, whereas its TARMA counterpart, i.e. model M11 is the most representative of the bimodality among TARMA models.

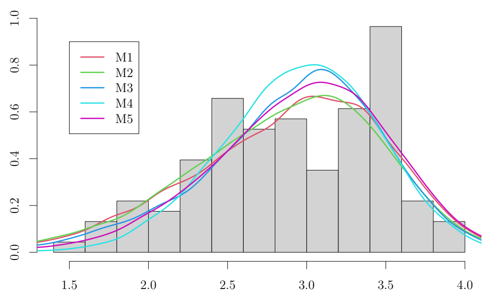

Figure 8 – Histogram of the Canadian lynx time series and the non-parametric kernel density estimation of 10000-length simulated trajectories from the TAR models M1 – M5.

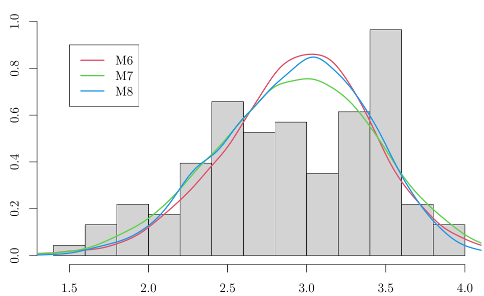

Figure 9 – Histogram of the Canadian lynx time series and the non-parametric kernel density estimation of 10000-length simulated trajectories from the TARMA models M6 – M8.

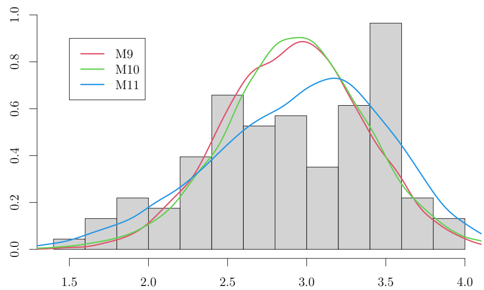

Figure 10 – Histogram of the Canadian lynx time series and the non-parametric kernel density estimation of 10000-length simulated trajectories from the TARMA models M9 – M11.

8. Forecasting performance

Giannerini and Goracci (2020) analysed the forecasting performance within the TARMA framework and showed the asymmetric behaviour of the one-step-ahead predictor of the TARMA(1,1) model. In this section, we exploit their implementation to compare the set of 11 models in terms of their forecasting accuracy. We consider a *training set*, composed by the first T observations, i.e. $\{y_1,\dots,y_T\}$ and a *test set* defined as $\{y_{T+1},\dots,y_{T+h}\}$, with h being the forecasting horizon. We fit models M1 – M11 on the training set and obtain the h-step ahead prediction. The forecasting performance depends upon both the last state of the system y_T and the forecasting horizon h. In general, besides the impact of h, we expect different performance if y_T is in a trough rather than in a peak. In this spirit, we consider 10 different training sets, identified by their end point $y_T \in \mathcal{F} := \{y_{1915},\dots,y_{1924}\}$. For each of the 10 training sets, we compute the h-step-ahead prediction with $h \in \mathcal{H} := \{1,\dots,10\}$ and derive the mean absolute percentage error (MAPE), defined as follows:

$$\text{MAPE} = \frac{1}{\#(\mathcal{F})}\frac{1}{\#(\mathcal{H})}\sum_{y_T\in\mathcal{F}}\sum_{h\in\mathcal{H}}|e(y_T,h)|\times 100, \tag{2}$$

$$e(y_T,h) = \frac{\hat{y}_{T+h}-y_{T+h}}{y_{T+h}}$$

and $\#(A)$ is the cardinality of the set A. In order to investigate the dependence of the forecasting performance upon the position of the last available observation y_T and the forecasting horizon h we consider the following two measures:

$$\text{MAPE}(y_T) = \frac{1}{\#(\mathcal{H})}\sum_{h\in\mathcal{H}}|e(y_T,h)|\times 100, \tag{3}$$

$$\text{MAPE}(h) = \frac{1}{\#(\mathcal{F})}\sum_{y_T\in\mathcal{F}}|e(y_T,h)|\times 100. \tag{4}$$

TABLE 7

Mean absolute percentage error MAPE(h), Eq. (4), as a function of the forecasting horizon h, for models M1 – M11. Each row corresponds to a different forecasting horizon $h = 1, \dots, 10$ and we have highlighted the best and the worst forecasting performance in green and red, respectively.

h	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
1	9.58	8.40	10.76	6.51	9.82	6.45	6.84	7.19	5.74	6.05	7.18
2	13.59	13.48	16.02	10.14	17.62	10.27	11.47	9.84	10.76	10.31	9.31
3	16.54	19.07	19.11	12.95	24.51	13.58	14.42	14.01	13.61	13.13	9.91
4	18.39	19.96	13.87	12.11	23.60	12.54	13.79	12.64	14.39	13.93	9.16
5	18.33	19.82	15.46	11.31	19.34	11.47	13.12	10.29	15.50	14.29	8.51
6	14.95	18.07	18.53	8.85	15.24	9.32	11.27	8.80	14.92	13.88	7.66
7	14.83	18.40	20.79	6.36	12.03	6.81	8.78	6.44	13.79	12.96	6.43
8	10.37	17.91	28.29	4.25	9.80	4.62	6.23	4.26	13.00	12.41	6.01
9	9.79	20.37	30.43	4.45	8.82	4.76	5.58	4.38	12.72	11.94	6.75
10	18.79	24.81	27.35	4.92	7.24	4.96	6.82	4.60	12.53	11.36	7.86

TABLE 8

Mean absolute percentage error MAPE(y_T), Eq. (3), as a function of the end point y_T, for models M1 – M11. Each row corresponds to a different training set identified by its end point $y_T = y_{1915}, \ldots, y_{1924}$ and we have highlighted the best and the worst forecasting performance in green and red, respectively.

y_T	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
y_{1915}	29.50	36.58	20.05	14.75	29.83	15.46	16.01	23.59	13.03	12.43	20.41
y_{1916}	25.52	18.33	15.61	15.67	43.01	17.82	20.19	18.64	17.00	15.78	15.14
y_{1917}	15.43	9.09	6.54	12.06	22.49	12.59	13.29	6.13	11.03	11.97	6.38
y_{1918}	7.25	18.19	36.59	4.90	8.06	5.13	7.29	7.51	9.67	8.83	4.08
y_{1919}	7.77	20.39	30.06	4.77	7.77	5.05	6.57	4.28	11.04	10.57	6.03
y_{1920}	20.55	20.88	18.05	4.78	7.77	5.07	6.56	4.85	12.38	11.79	5.34
y_{1921}	6.28	11.27	12.79	5.01	8.56	5.04	6.75	5.55	13.00	12.09	6.97
y_{1922}	9.36	19.17	20.38	6.77	4.44	7.15	8.49	3.64	15.17	13.80	4.15
y_{1923}	11.76	13.68	27.54	4.89	7.72	3.50	5.01	3.90	10.64	10.59	5.15
y_{1924}	11.73	12.71	13.00	8.23	8.37	7.98	8.15	4.37	14.00	12.39	5.14

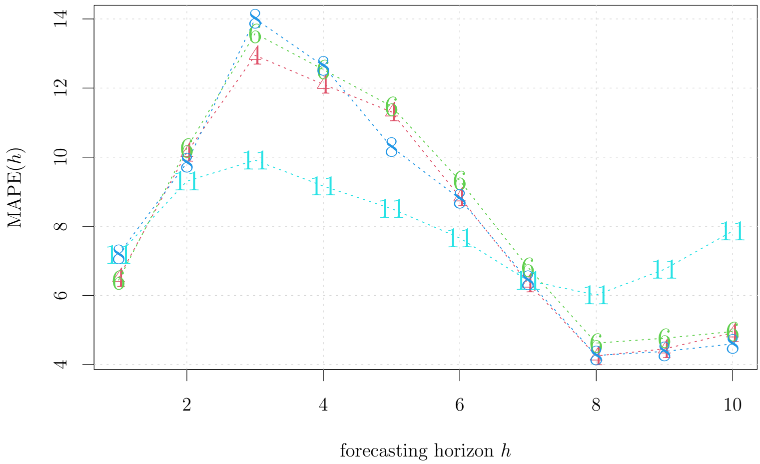

Figure 11 – Plot of MAPE(h) versus h for the models M4, M6, M8 and M11. The forecasting horizon h ranges from 1 to 10.

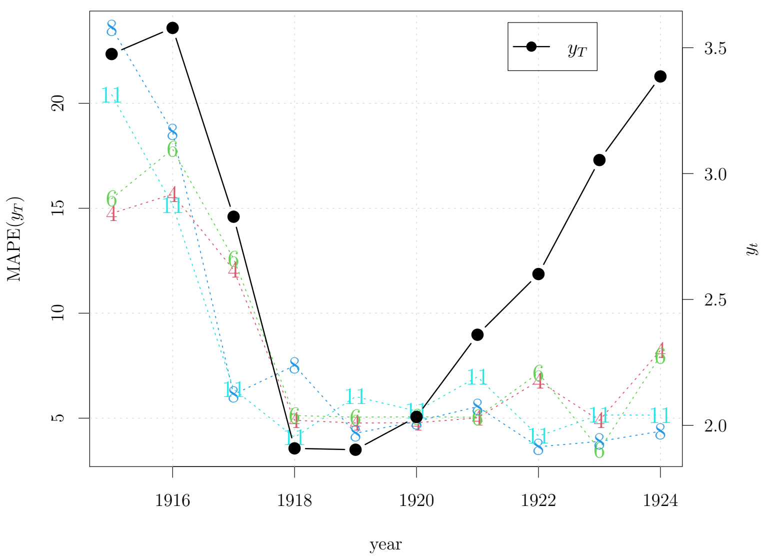

Figure 12 – Plot of MAPE(y_T) versus y_T for the models M4, M6, M8 and M11. The black line is the observed series and each bullet represents y_T, $T = 1915, \dots, 1924$, i.e. the end point of each training set.

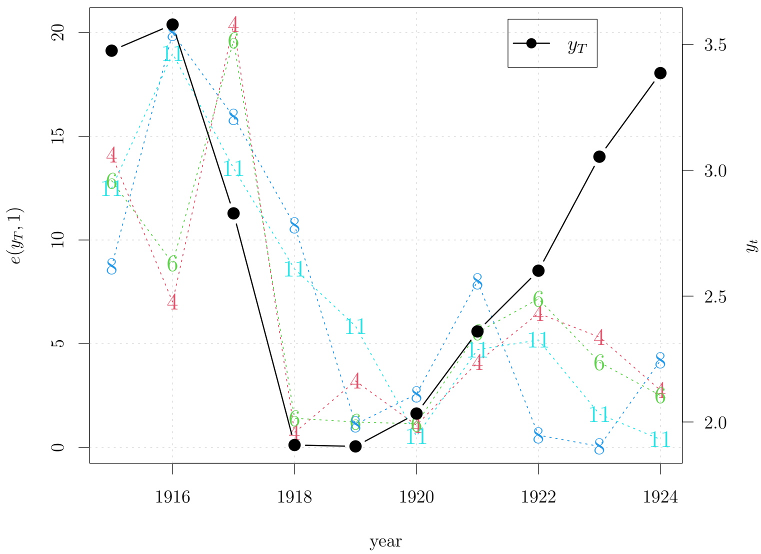

Figure 13 – Plot of $e(y_T, h = 1)$ versus y_T when the forecasting horizon is fixed to 1 for the models M4, M6, M8 and M11. The black line is the observed series and each bullet represents y_T, $T = 1915, \ldots, 1924$, i.e. the end point of each training set.

Tables 7 and 8 contain the values of MAPE(h) and MAPE(y_T) computed for models M1 – M11, respectively. For each row, we have highlighted the best and the worst forecasting performance in green and red, respectively. From Tables 7 model M11 results the best of the panel as it presents the lowest MAPE in 5 out of 10 values of h, and it is not far from the best model for the other values of h. Among TARMA models, also models M6 and M8 enjoy a good performance. Among TAR models, model M4 performs well and turns out to be the best for $h = 8, 9$. Table 8 clearly shows that the performance is indeed influenced by the value of y_T and it seems that this impinges more on the class of TAR models. Also in this setting, the best models are M4, M6, M8 and M11. For these, in Figures 11 and 12, we report the plots of the MAPE(h) versus h and MAPE(y_T) versus y_T, respectively. As expected, the prediction error does not increase monotonically with h and this is a typical feature of non-linear prediction. If we consider the MAPE(h), M11 is the best model, except for $h = 8, 9, 10$. The only case where a TAR model performs better that the TARMA ones, is when $h = 8$ with M4 enjoying the best performance. The analysis of the MAPE(y_T), shows that the forecasting performance is better if the end point y_T belongs to the ascending phase of the cycle. This is a consequence of the asymmetric cycle where the decreasing phase is faster and so it is generally harder to predict. Overall, TARMA models tend to attain a superior forecasting accuracy, the only exception being when $y_T = y_{1915}$ for which the TAR model M4 is superior. To complete the analysis, in Figure 13 we report the prediction error $e(y_T, h)$ as function of the end point y_T for the forecasting horizon h fixed to 1. The error is higher when y_T is at the beginning of the descendent phase, where the best model is confirmed to be the TAR M4, whereas models M8 and M11 that in general have the best performance, here turn out to be the worse models. Remarkably, contrarily to the situation in Figure 12, model M11 has the best performance only in two instances. The TAR model M4 appears to be preferable for the one-step-ahead prediction.

9. Diagnostics

In this section, we report the diagnostics for the fits of models M4 – M11. Figures 14 and 15 show the global and partial correlograms for the residuals and these do not show any structures. The results are reinforced by the entropy measure S_ρ (Giannerini *et al.*, 2015) computed up to lag 12 and reported in Figures 16 and 17.

10. Conclusions

In this work, we have deployed the recent theoretical results of Chan and Goracci (2019) on TARMA models to revisit the analysis of the benchmark *Canadian lynx* time series. This data set has attracted attention since it is characterized by an asymmetric limit cycle where the increasing phase is slower than the decreasing phase. Since threshold models can be seen as the discrete time version of continuous time prey-predator models and are able to capture limit cycles, they hold substantial promise to model the lynx time series. Several TAR and TARMA models have been considered and compared from three different prospectives: goodness-of-fit through information criteria; the ability to reproduce both characteristic cycles and multimodality; forecasting performance. We have found models that perform better than TAR models with respect to all these aspects. The reasons behind the superiority of TARMA models can be diverse. In particular, TARMA models can describe complex phenomena parsimoniously and allow different long-run probabilistic behaviours, including non-stationary sub-models in some regimes. Moreover, they naturally account for measurement error. On the other hand, the theory for TARMA models presents several gaps. The long-run probabilistic behaviour is completely characterized only for the TARMA$(1, 1)$, for which it depends only upon the autoregressive part. It could be interesting to investigate whether this holds also for higher-order models. A possible first direction is to assess whether the conditions for the ergodicity of the TARMA(1,1) with general delay parameter $d > 1$ is the same as those obtained in Chen and Tsay (1991). Another interesting aspect is related to the derivation of the mean square prediction error for TARMA models and the application of the MRIC proposed in Hsu *et al.* (2019) in the TARMA framework. Moreover, there are no results on the sampling properties of maximum-likelihood estimators for TARMA models. Another open challenge concerns bootstrap inference for threshold models. Different bootstrap schemes have been proposed in tests within the threshold framework, but there are no results about their validity even for the simplest TAR case.

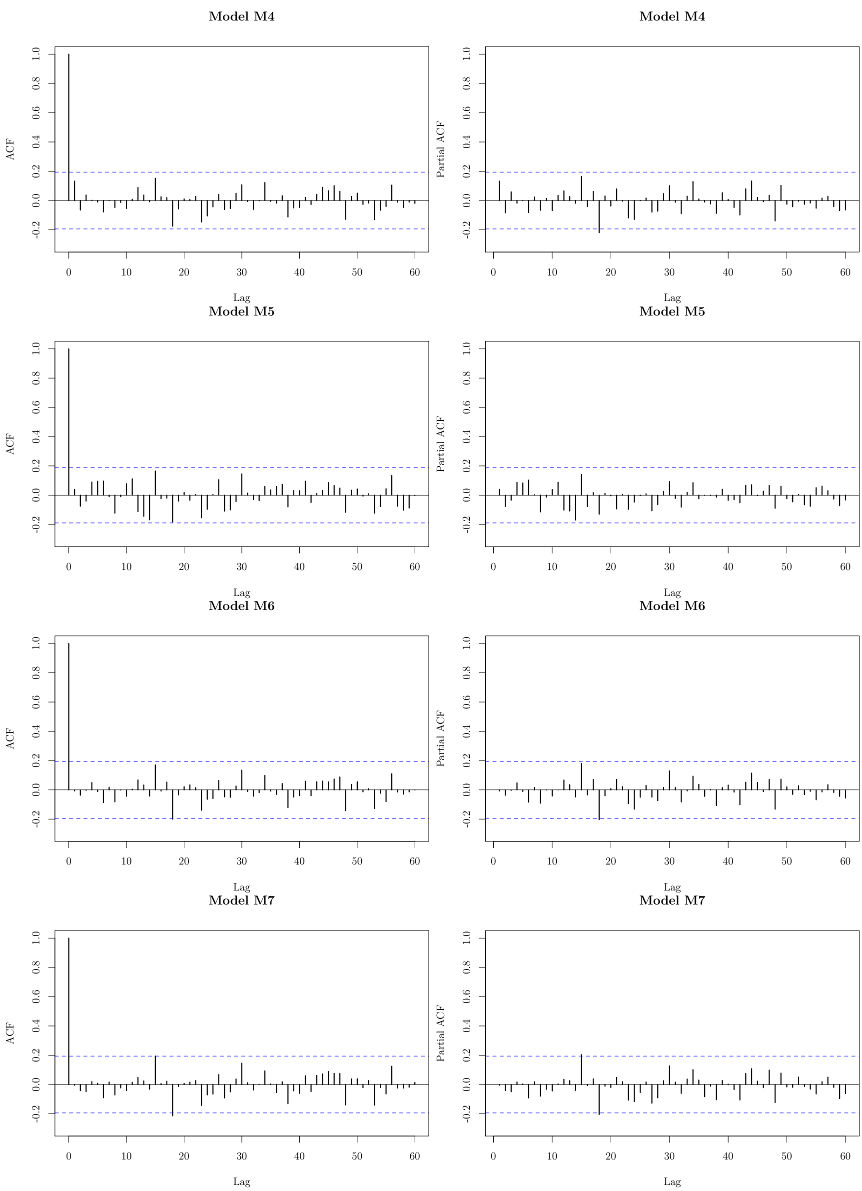

Figure 14 – Correlograms of the residuals from Model M4 – M7 fit for the time series of Canadian lynx.

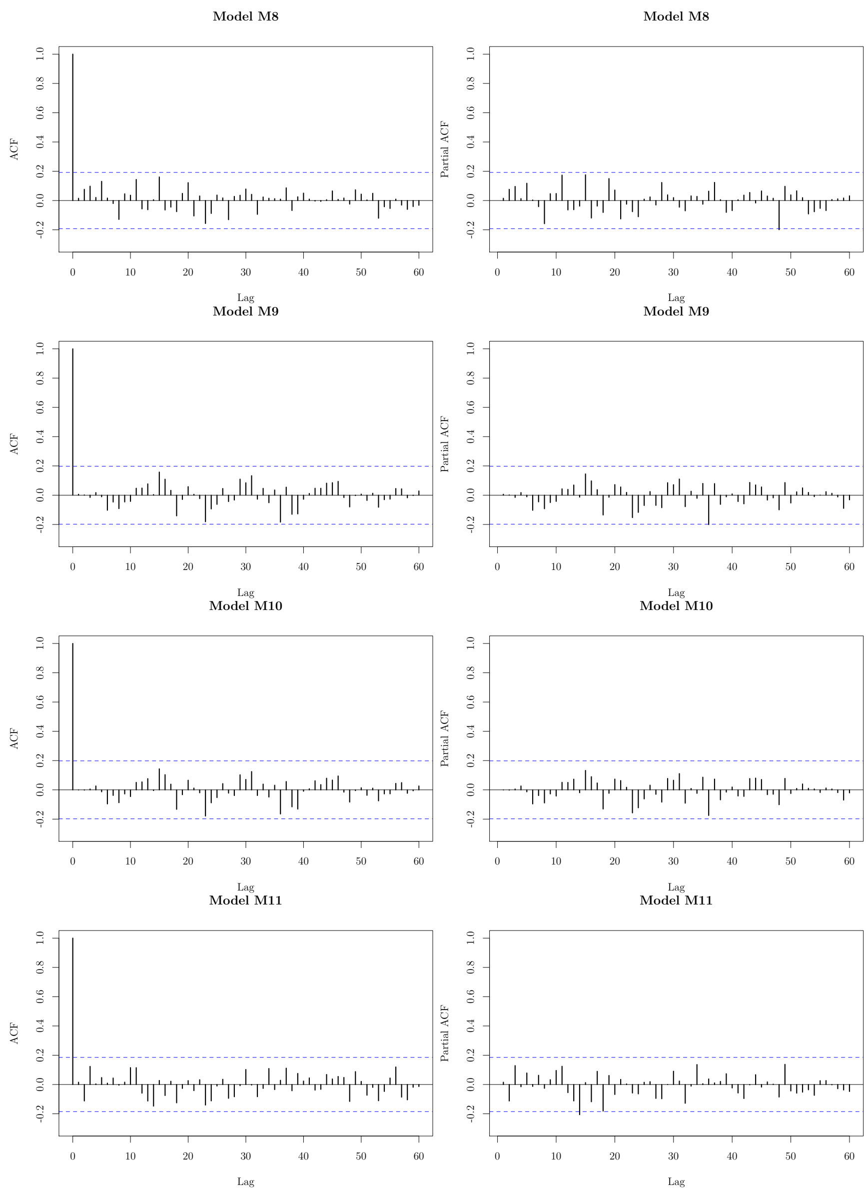

Figure 15 – Correlograms of the residuals from models M8 – M11 fit for the time series of Canadian lynx.

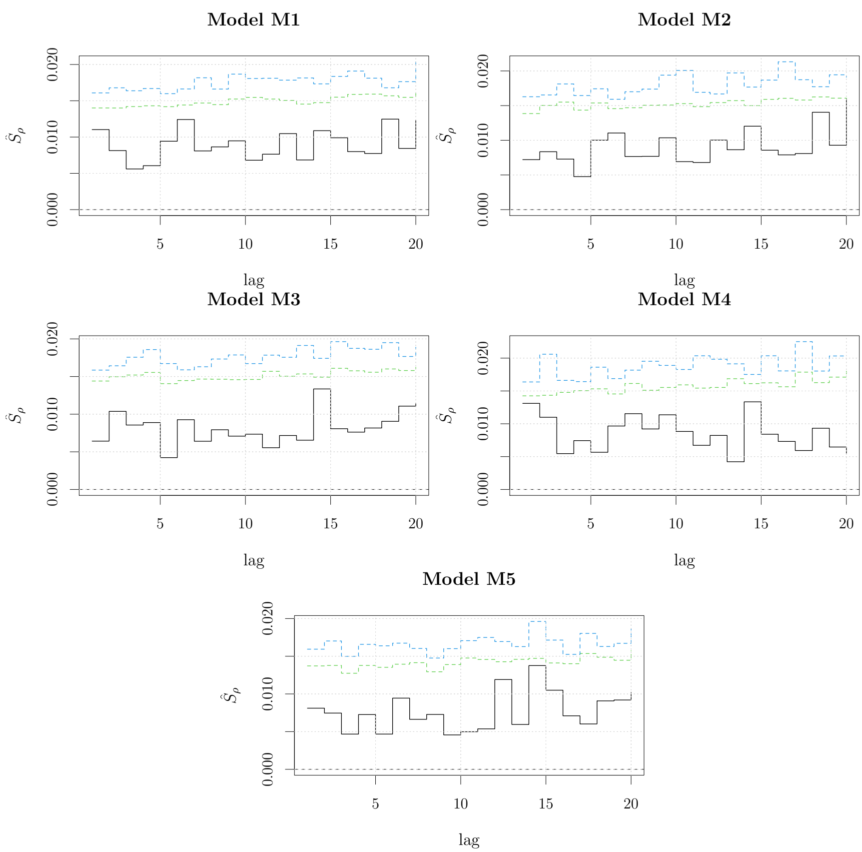

Figure 16 – Entropy measure S_ρ computed on the residuals from the TAR models M1 – M5 fit for the time series of the Canadian lynx. The confidence bands at 95% (green) and 99%(blue) correspond to the null hypothesis of serial independence.

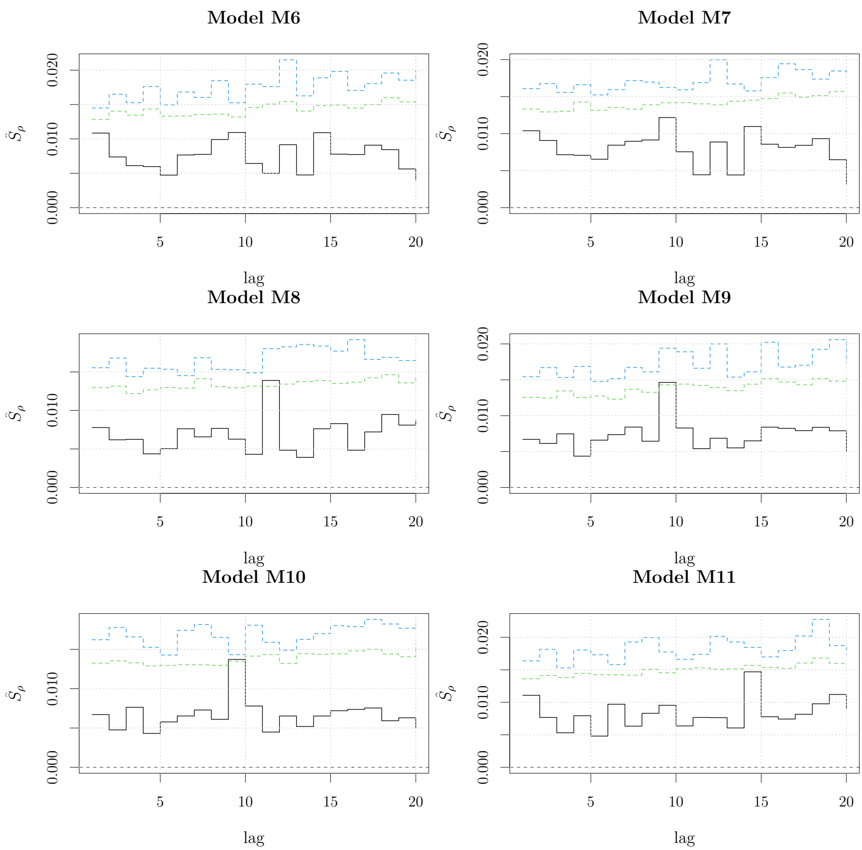

Figure 17 – Entropy measure S_ρ computed on the residuals from the TARMA models M6 – M11 fit for the time series of the Canadian lynx. The confidence bands at 95% (green) and 99%(blue) correspond to the null hypothesis of serial independence.

Lastly, we mention the extension of TARMA models in the continuous time because the available results cover only the continuous-time TAR case (e.g. Brockwell and Williams, 1997).

ACKNOWLEDGEMENTS

The author is grateful to the Editor Simone Giannerini, the Executive Editor Mariagiulia Matteucci and the anonymous referees whose comments helped improving the paper.

APPENDIX

A. SUPPLEMENTARY RESULTS: SPECTRAL ANALYSIS

We report time plots of the simulated time series and their skeletons for TARMA models M1 – M11.

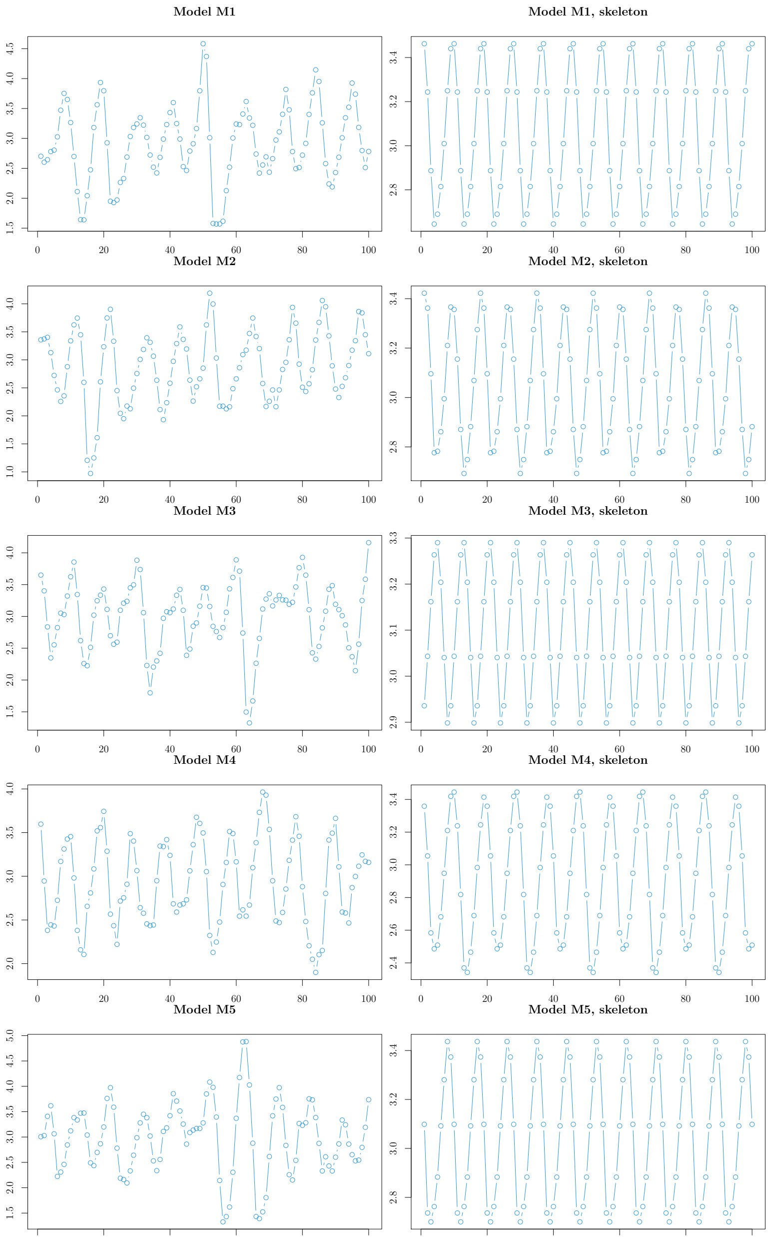

Figure 18 – Time plot of the simulated time series and their skeletons for TAR models M1 – M3.

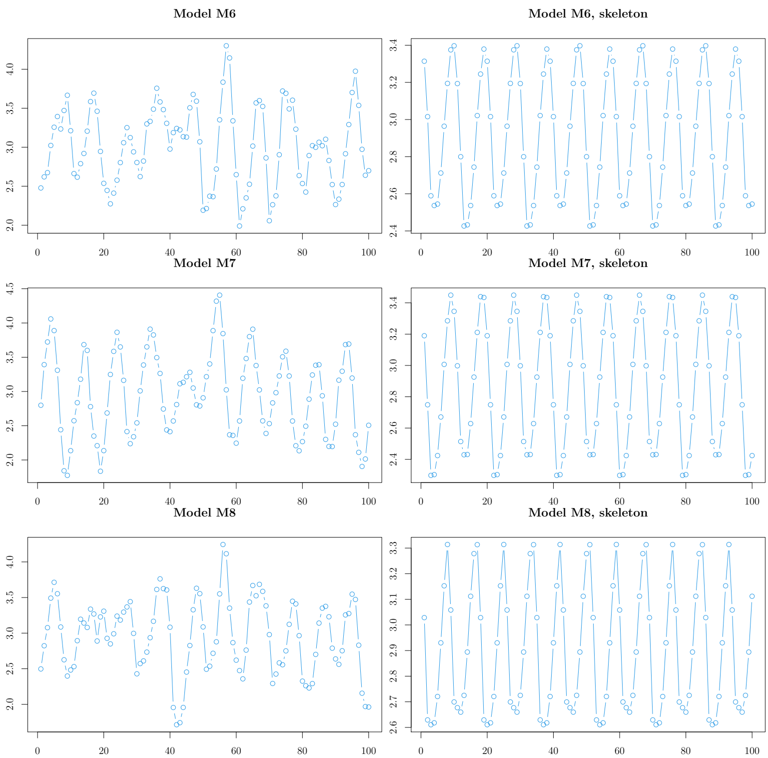

Figure 19 – Time plot of the simulated time series and their skeletons for TARMA models M6 – M8.

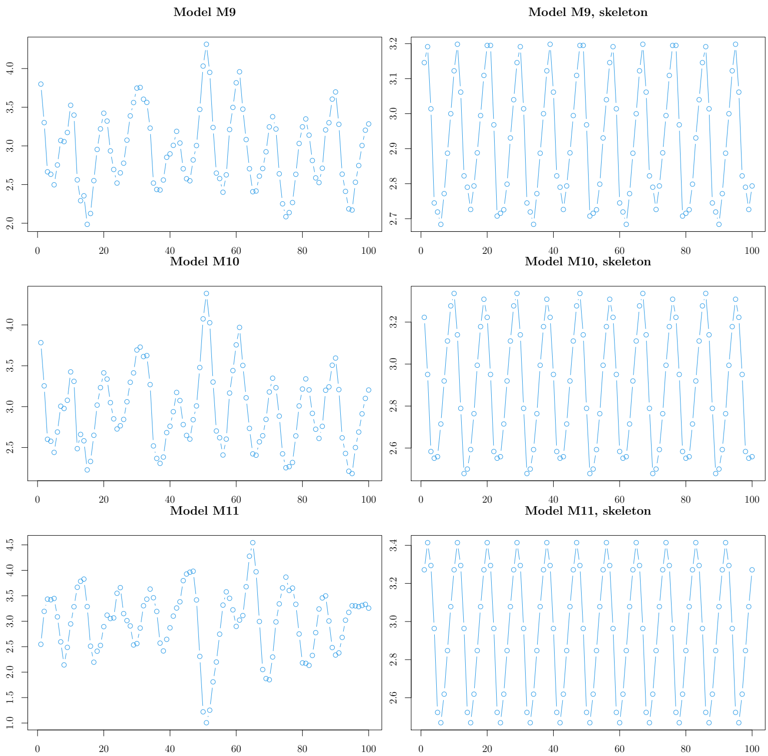

Figure 20 – Time plot of the simulated time series and their skeletons for TARMA models M9 – M11 .

References

H. An, S. Chen (1997). *A note on the ergodicity of non-linear autoregressive model*. Statistics & Probability Letters, 34, no. 4, pp. 365–372.

F. Bec, M. Ben Salem, M. Carrasco (2004). *Tests for unit-root versus threshold specification with an application to the purchasing power parity relationship*. Journal of Business & Economic Statistics, 22, no. 4, pp. 382–395.

F. Bec, A. Guay, E. Guerre (2008). *Adaptive consistent unit-root tests based on autoregressive threshold model*. Journal of Econometrics, 142, no. 1, pp. 94–133.

P. J. Brockwell, R. J. Williams (1997). *On the existence and application of continuous-time threshold autoregressions of order two*. Advances in Applied Probability, 29, no. 1, p. 205–227.

M. Caner, B. Hansen (2001). *Threshold autoregression with a unit root*. Econometrica, 69, no. 6, pp. 1555–1596.

K.-S. Chan (1990). *Testing for threshold autoregression*. The Annals of Statistics, 18, no. 4, pp. 1886–1894.

K.-S. Chan (1993). *Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model*. The Annals of Statistics, 21, no. 1, pp. 520–533.

K.-S. Chan, S. Giannerini, G. Goracci, H. Tong (2019). *Testing for threshold regulation with measurement error*. Tech. rep., University of Iowa and University of Bologna.

K.-S. Chan, S. Giannerini, G. Goracci, H. Tong (2020). *Unit-root test within a threshold ARMA framework*. arXiv:2002.09968 [stat.ME]. URL https://arxiv.org/abs/2002.09968.

K.-S. Chan, G. Goracci (2019). *On the ergodicity of first-order threshold autoregressive moving-average processes*. Journal of Time Series Analysis, 40, no. 2, pp. 256–264.

K.-S. Chan, H. Tong (1985). *On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations*. Advances in Applied Probability, 17, no. 3, pp. 666–678.

K.-S. Chan, H. Tong (2001). *Chaos: A Statistical Perspective*. Springer Series in Statistics. Springer-Verlag, New York.

K.-S. Chan, H. Tong (2010). *A note on the invertibility of nonlinear ARMA models*. Journal of Statistical Planning and Inference, 140, no. 12, pp. 3709–3714.

K. S. Chan, R. S. Tsay (1998). *Limiting properties of the least squares estimator of a continuous threshold autoregressive model*. Biometrika, 85, no. 2, pp. 413–426.

C. CHEN, M. K. SO, F.-C. LIU (2011). *A review of threshold time series models in finance*. Statistics and its Interface, 4, no. 2, pp. 167–181.

R. CHEN, R. TSAY (1991). *On the ergodicity of TAR(1) processes*. The Annals of Applied Probability, 1, no. 4, pp. 613–634.

I. CHOI (2015). *Almost All about Unit Roots: Foundations, Developments, and Applications*. Themes in Modern Econometrics. Cambridge University Press, Cambridge.

D. CLINE (2009). *Thoughts on the connection between threshold time series models and dynamical systems*. In K.-S. CHAN (ed.), *Exploration of a Nonlinear World. An Appreciation of Howell Tong's Contributions to Statistics*, World Scientific, Singapore, pp. 165–181.

J. CRYER, K.-S. CHAN (2008). *Time Series Analysis - With Applications in R*. Springer, New York, 2nd ed.

J. G. DE GOOIJER (2017). *Elements of Nonlinear Time Series Analysis and Forecasting*. Springer International Publishing, Cham, Switzerland.

R. DE JONG, C.-H. WANG, . BAE (2007). *Correlation robust threshold unit root tests*. Mimeo, Ohio State University, Michigan.

J. L. DOOB (1936). *Review: Vito Volterra, Leçons sur la Théorie Mathématique de la Lutte pour la Vie*. Bulletin of the American Mathematical Society, 42, no. 5, pp. 304–305.

C. ELTON, M. NICHOLSON (1942). *The ten-year cycle in numbers of the Lynx in Canada*. Journal of Animal Ecology, 11, no. 2, pp. 215–244.

W. ENDERS, C. GRANGER (1998). *Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates*. Journal of Business & Economic Statistics, 16, no. 3, pp. 304–311.

J. FAN, Q. YAO (2003). *Nonlinear Time Series*. Springer-Verlag, New York.

P. GALEANO, D. PEÑA (2007). *Improved model selection criteria for SETAR time series models*. Journal of Statistical Planning and Inference, 137, pp. 2802–2814.

S. GIANNERINI, G. GORACCI (2020). *Forecasting with threshold ARMA models*. Tech. rep., University of Bologna.

S. GIANNERINI, E. MAASOUMI, E. BEE DAGUM (2015). *Entropy testing for nonlinear serial dependence in time series*. Biometrika, 102, pp. 661–675.

F. GIORDANO, M. NIGLIO, C. VITALE (2017). *Unit root testing in presence of a double threshold process*. Methodology and Computing in Applied Probability, 19, no. 2, pp. 539–556.

J. GOOIJER (1998). *On threshold moving-average models*. Journal of Time Series Analysis, 19, no. 1, pp. 1–18.

G. GORACCI (2020). *An empirical study on the parsimony and descriptive power of TARMA models*. Statistical Methods & Applications. URL https://doi.org/10.1007/s10260-020-00516-8.

G. GORACCI, S. GIANNERINI (2020). *Testing for threshold effects in ARMA models*. Tech. rep., University of Bologna.

M. GUO, J. PETRUCCELLI (1991). *On the null recurrence and transience of a first-order SETAR model*. Journal of Applied Probability, 28, no. 3, pp. 584–592.

N. HALDRUP, M. JANSSON (2006). *Improving size and power in unit root testing*. In K. PATTERSON, T. MILLS (eds.), *Palgrave Handbook of Econometrics: Volume 1: Econometric Theory*, Palgrave Macmillan UK, Basingstoke, pp. 252–277.

B. HANSEN (2011). *Threshold autoregression in economics*. Statistics and its Interface, 4, no. 2, pp. 123–127.

A. S. HERBERT (1959). *Book reviews: Elements of Mathematical Biology by Alfred J. Lotka*. Econometrica, 27, no. 3, pp. 493–495.

H.-L. HSU, C.-K. ING, H. TONG (2019). *On model selection from a finite family of possibly misspecified time series models*. The Annals of Statistics, 47, no. 2, pp. 1061–1087.

Y. KAJITANI, K. W. HIPEL, A. I. MCLEOD (2005). *Forecasting nonlinear time series with feed-forward neural networks: A case study of Canadian lynx data*. Journal of Forecasting, 24, no. 2, pp. 105–117.

G. KAPETANIOS (2001). *Model selection in threshold models*. Journal of Time Series Analysis, 22, no. 6, pp. 733–754.

G. KAPETANIOS, Y. SHIN (2006). *Unit root tests in three-regime SETAR models*. The Econometrics Journal, 9, no. 2, pp. 252–278.

S. KONISHI, G. KITAGAWA (1996). *Generalised information criteria in model selection*. Biometrika, 83, no. 4, pp. 875–890.

D. LI, W. LI, S. LING (2011). *On the least squares estimation of threshold autoregressive moving-average models*. Statistics and its Interface, 4, pp. 183–196.

D. LI, S. LING (2012). *On the least squares estimation of multiple-regime threshold autoregressive models*. Journal of Econometrics, 167, no. 1, pp. 240 – 253.

D. LI, S. LING, W. K. LI (2013). *Asymptotic theory on the least squares estimation of threshold moving-average models*. Econometric Theory, 29, no. 3, p. 482–516.

G. LI, W. LI (2008). *Testing for threshold moving average with conditional heteroscedasticity*. Statistica Sinica, 18, pp. 647–665.

G. LI, W. LI (2011). *Testing a linear time series model against its threshold extension*. Biometrika, 98, no. 1, pp. 243–250.

K. S. LIM (1987). *A comparative study of various univariate time series models for Canadian lynx data*. Journal of Time Series Analysis, 8, no. 2, pp. 161–176.

T. C. LIN, M. POURAHMADI (1998). *Nonparametric and non-linear models and data mining in time series: A case-study on the Canadian lynx data*. Journal of the Royal Statistical Society, Series C, 47, no. 2, pp. 187–201.

S. LING, H. TONG (2005). *Testing for a linear MA model against threshold MA models*. The Annals of Statistics, 33, no. 6, pp. 2529–2552.

S. LING, H. TONG, D. LI (2007). *Ergodicity and invertibility of threshold moving-average models*. Bernoulli, 13, no. 1, pp. 161–168.

W. LIU, J.L., C. LI (1997). *On a threshold autoregression with conditional heteroscedastic variances*. Journal of Statistical Planning and inference, 62, no. 2, pp. 279–300.

J. PARK, M. SHINTANI (2016). *Testing for a unit root against transitional autoregressive models*. International Economic Review, 57, no. 2, pp. 635–664.

K. PATTERSON (2010). *A Primer for Unit Root Testing*. Palgrave Texts in Econometrics. Palgrave Macmillan UK, Basingstoke.

K. PATTERSON (2011). *Unit Root Tests in Time Series Volume 1: Key Concepts and Problems*. Palgrave Texts in Econometrics. Palgrave Macmillan UK, Basingstoke.

K. PATTERSON (2012). *Unit Root Tests in Time Series Volume 2: Extensions and Developments*. Palgrave Texts in Econometrics. Palgrave Macmillan UK, Basingstoke.

L. QIAN (1998). *On maximum likelihood estimators for a threshold autoregression.* Journal of Statistical Planning and Inference, 75, no. 1, pp. 21 – 46.

R. RUDNICKI (2003). *Long-time behaviour of a stochastic prey–predator model*. Stochastic Processes and their Applications, 108, no. 1, pp. 93 – 107.

M. SEO (2008). *Unit root test in a threshold autoregression: Asymptotic theory and residual-based block bootstrap*. Econometric Theory, 24, no. 6, pp. 1699–1716.

N. C. STENSETH, W. FALCK, K.-S. CHAN, O. BJORNSTAD, M. O'DONOGHUE, H. TONG, R. BOONSTRA, S. BOUTIN, C. KREBS, N. YOCCOZ (1998). *From patterns to processes: Phase and density dependencies in the Canadian lynx cycle*. Proceedings of the National Academy of Sciences, 95, pp. 15430–15435.

M. STONE (1977). *An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion*. Journal of the Royal Statistical Society, Series B, 39, no. 1, pp. 44–47.

H. TONG (1978). *On a threshold model*. In C. CHEN (ed.), *Pattern Recognition and Signal Processing*, Sijthoff & Noordhoff, Amsterdam, NATO ASI Series E: Applied Sc.(29), pp. 575–586.

H. TONG (1983). *Threshold models in non-linear time series analysis*. Lecture Notes in Statistics. Springer-Verlag, New York.

H. TONG (1990). *Non-linear Time Series: A Dynamical System Approach*. Clarendon Press, Oxford.

H. TONG (2007). *Birth of the threshold time series model*. Statistica Sinica, 17, no. 1, pp. 8–14.

H. TONG (2011). *Threshold models in time series analysis - 30 years on*. Statistics and its Interface, 4, no. 2, pp. 107–118.

H. TONG (2017). *Threshold models in time series analysis - some reflections*. Journal of Econometrics, 189, no. 2, pp. 485 – 491.

H. TONG, K. LIM (1980). *Threshold autoregression, limit cycles and cyclical data*. Journal of the Royal Statistical Society, Series B, 42, no. 3, pp. 245–292.

Z. WANG, Y. XIE, J. LU, Y. LI (2019). *Stability and bifurcation of a delayed generalized fractional-order prey–predator model with interspecific competition*. Applied Mathematics and Computation, 347, pp. 360 – 369.

Y. WU, W. Q. ZHU (2008). *Stochastic analysis of a pulse-type prey-predator model*. Physical Review E, 77, p. 041911.

SUMMARY

The class of threshold autoregressive models has been proven to be a powerful and appropriate tool to describe many dynamical phenomena in different fields. In this work, we deploy the threshold autoregressive moving-average framework to revisit the analysis of the benchmark Canadian lynx time series. This data set has attracted great attention among non-linear time series analysts due to its asymmetric cycle that makes the investigation very challenging. We compare some of the best threshold autoregressive models (TAR) proposed in literature with a selection of threshold autoregressive moving-average models (TARMA). The models are compared under different prospectives: (i) goodness-of-fit through information criteria, (ii) their ability to reproduce characteristic cycles, (iv) their capability to capture multimodality and (iii) forecasting performance. We found TARMA models that perform better than TAR models with respect to all these aspects.

Keywords: Population dynamics; Predator-prey interaction; Canadian lynx time series; Non-linear time series; TARMA processes; Asymmetric cycle.