Statistica Sinica Preprint No: SS-2021-0120	
Title	Testing for Threshold Effects in the TARMA Framework
Manuscript ID	SS-2021-0120
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0120
Complete List of Authors	Greta Goracci, Simone Giannerini, Kung-Sik Chan and Howell Tong
Corresponding Author	Simone Giannerini
E-mail	simone.giannerini@unibo.it
Notice: Accepted version subject to English editing.	

Testing for threshold effects in the TARMA framework

Greta Goracci^{1,6}, Simone Giannerini¹, Kung-Sik Chan², Howell Tong^{3,4,5}

1. University of Bologna, Italy; 2. University of Iowa, USA

3. University of Electronic Science and Technology of China, Chengdu, China;

4. Tsinghua University, China; 5. London School of Economics and Political Science, U.K.

6. Free University of Bozen/Bolzano, Italy

Abstract: We present supremum Lagrange Multiplier tests to compare a linear ARMA specification against its threshold ARMA extension. We derive the asymptotic distribution of the test statistics both under the null hypothesis and contiguous local alternatives. Moreover, we prove the consistency of the tests. The Monte Carlo study shows that the tests enjoy good finite-sample properties, are robust against various forms of model mis-specification and have a performance that is not affected by the unknown order of the model. The tests have low computational burden and do not suffer from some of the drawbacks that affect the quasi-likelihood ratio setting. Lastly, we present an application to a time series of standardized tree-ring growth indexes, and the results can lead to new research in climate studies.

Key words and phrases: Threshold autoregressive moving-average models, Lagrange multiplier test, Linearity testing, Marked empirical processes, tree-rings.

1. Introduction

Threshold autoregressive (TAR) models have been very popular and widely applied in Ecology, Economics, Epidemiology, Finance and many other fields, ([Tong, 1990](#); [Chan, 2009](#); [Tong, 2011](#); [Hansen, 2011](#)), and three special issues have been devoted to them: *Statistica Sinica* in 2007 ([Chan and Li, 2007](#)), *Statistics and Its Interface* in 2011 ([Chan et al., 2011](#)) and *Journal of Business and Economic Statistics* in 2017 ([Chan et al., 2017](#)). In particular, threshold autoregressive-moving average (TARMA) models, introduced by Tong (see [Tong, 1990](#)), are non-linear models with a regime-switching mechanism specifying an autoregressive-moving average (ARMA) sub-model in each regime. They include two particular models of independent interest: the threshold autoregressive (TAR) model and the threshold moving-average (TMA) model. By combining them, we have the TARMA model, which is a parsimonious and yet rich model for non-linear time series analysis. See, e.g., [Goracci \(2020, 2021\)](#)). Now, least squares estimation of parameters of a general TARMA model typically assumes stationarity and ergodicity (e.g. [Li et al., 2011](#)), but they are notions that have attracted insufficient attention. An early result concerning the latter was due to [Ling \(1999\)](#), albeit his conditions are quite restrictive. In fact, a full characterization of the long-run probabilistic behaviour of TARMA models is only

available very recently. As far as we know, [Chan and Goracci \(2019\)](#) was the first paper and is the only paper to-date that gives the necessary and sufficient conditions for the ergodicity of the first-order TARMA model. Moreover, they have provided a complete parametric classification of the first-order TARMA into regions of (geometric) ergodicity, null recurrence and transience.

Considerable efforts have been expended to develop tests for a linear model versus its threshold extension. However, most of the contributions have focused on AR-type models. For instance, [Petrucci and Davies \(1986\)](#) developed a portmanteau test based on cumulative sums of standardized residuals from an autoregressive fit. [Tsay \(1998\)](#) studied a variation of it. [Luukkonen et al. \(1988\)](#) proposed a Lagrange Multiplier test for linearity against non-linear models that include TAR models. A Lagrange Multiplier test was also developed in [Wong and Li \(1997, 2000\)](#) for TAR models with conditional heteroscedasticity. Quasi-likelihood ratio tests were studied in [Chan \(1990\)](#), [Chan and Tong \(1990\)](#), [Chan \(1991\)](#) and as recently as [Su and Chan \(2017\)](#) for threshold diffusion. For a review see also [Tong \(2011\)](#).

Threshold models that include moving-average (MA) components have been under-developed, probably due to the mathematical difficulties that arise when MA components are incorporated. However, since data are

almost always collected with measurement errors, a TARMA model is more appropriate than a TAR model. Indeed, it is known that an AR model with measurement errors becomes an ARMA model. Likewise, it can be proved that a TAR model with measurement errors may be approximated by a TARMA model (see e.g. [Chan et al., 2020](#)). It should be stressed that the adoption of an TARMA model is significant because, in the presence of measurement errors, a high order autoregressive model would be necessary if tests are conducted only with respect to TAR models and loss of power will ensue.

For MA-type models, [Ling and Tong \(2005\)](#) investigated a quasi-likelihood ratio test for the MA model against its threshold extension. They proved that, under the null hypothesis, the test statistic converges weakly to a functional of the centered Gaussian process. Their results were extended in [Li and Li \(2008\)](#) to the case with GARCH errors. For a comprehensive development, see [Chan et al. \(2014\)](#). More recently, [Li and Li \(2011\)](#) developed a quasi-likelihood ratio statistic to test the presence of thresholds in an ARMA model setting. They used a stochastic permutation device to build the distribution of the statistic under the null hypothesis.

In this paper we extend the work of [Chan \(1990\)](#) and [Ling and Tong \(2005\)](#) and propose supremum Lagrange multiplier test statistics (supLM)

to determine whether a TARMA model fits a stationary time series significantly better than an ARMA model. One of the main advantages of using a Lagrange Multiplier test over the likelihood-ratio test is that the former avoids estimation under the alternative hypothesis. We prove that under both the null hypothesis and contiguous local alternatives the asymptotic distributions of the test statistics reduce to functionals of Gaussian processes that are centered under the null and non-centered under the alternative. Our results extend the work of [Ling and Tong \(2005\)](#) on the weak convergence of linear marked empirical processes with infinitely many markers to the case where the underlying process is an ARMA model. Moreover, we prove the consistency of our tests and show that they have non-trivial power against local alternatives.

In order to test the null hypothesis of an ARMA model against the alternative hypothesis of a TARMA model, we propose two supLM statistics: in the first one, denoted by sLM, only the AR part is tested for threshold non-linearity whereas in the second statistic, denoted by sLM^{*}, both the AR and MA parts are tested. As it will be clear, the two statistics are different. In particular, the sLM^{*} statistic does not reduce to the sLM when the MA part is either absent or is regime-independent. This is reflected in the different finite sample behaviour of the tests.

We explore systematically the performances of our supLM tests and compare them with the quasi-likelihood ratio test of [Li and Li \(2011\)](#) (qLR from now on). The extensive simulation study shows clearly that our tests have better size and power while they incur a much lower computational burden. Furthermore, the two supLM tests are robust against model misspecification and their performances are not adversely affected if the order of the ARMA model is unknown and selected through the Hannan-Rissanen procedure. Lastly, we apply our tests to the time series of standardized tree-ring growth indexes. We show that a TARMA model can provide a better fit compared to ARMA models commonly adopted in the literature. We suggest that a TARMA model can lead to a deeper understanding of the tree-ring dynamics, possibly leading to some novel directions of research in climate studies, in which non-linear dynamics is incorporated.

The rest of the paper is organized as follows. In [Section 2](#), we present our setting and the two tests. In [Section 3](#), we derive the asymptotic distributions of the supLM test statistics under the null hypothesis. In [Section 4](#), we derive the asymptotic distributions of the statistics under local contiguous alternatives and prove the consistency of the tests. [Section 5](#) contains a Monte Carlo study to assess the finite sample performance of our proposals. We also investigate the behaviour of the tests under model mis-specification

and when the order of the tested model is unknown. In Section 6, we apply our tests to a tree-ring time series. We conclude in Section 7. All the proofs and the tabulated quantiles of the null distribution are detailed in the Supplementary Material, that also contains additional results regarding both the simulation study and the tree-ring data analysis.

2. Notation and preliminaries

Let the time series $\{X_t : t = 1, \dots, n\}$ follow the threshold autoregressive moving-average model (TARMA(p, q)) defined by the difference equation:

$$X_t = \phi_0 + \sum_{i=1}^p \phi_i X_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t + \left(\varphi_0 + \sum_{i=1}^p \varphi_i X_{t-i} - \sum_{j=1}^q \vartheta_j \varepsilon_{t-j} \right) I(X_{t-d} \leq r). \quad (2.1)$$

In the case where the moving-average parameters are regime-independent, the rightmost summation involving ϑ_j , $j = 1, \dots, q$ is absent. The innovation process $\{\varepsilon_t\}$ is assumed to be iid with zero mean and $E[\varepsilon_t^2] = \sigma^2 < \infty$. The positive integers p and q are the autoregressive and moving-average orders, respectively; d is the delay parameter that takes positive integer values. We assume p, q, d to be known. Moreover, $I(\cdot)$ is the indicator function and $r \in \mathbb{R}$ is the threshold parameter. For notational convenience, we abbreviate $I(X_t \leq r)$ by $I_r(X_t)$. Let $\boldsymbol{\eta}$ be the vector of all the parameters

(excluding the threshold r) in Model (2.1), i.e.

$$\boldsymbol{\eta} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top, \sigma^2)^\top \in \Theta_\eta = \Theta_\phi \times \Theta_\theta \times \Theta_\varphi \times \Theta_\vartheta \times (0, +\infty),$$

$$\boldsymbol{\eta}_1 = (\boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top)^\top, \quad \boldsymbol{\eta}_2 = (\boldsymbol{\varphi}^\top, \boldsymbol{\vartheta}^\top)^\top,$$

where

$$\begin{aligned} \boldsymbol{\phi} &= (\phi_0, \phi_1, \dots, \phi_p)^\top \in \Theta_\phi; & \boldsymbol{\theta} &= (\theta_1, \dots, \theta_q)^\top \in \Theta_\theta; \\ \boldsymbol{\varphi} &= (\varphi_0, \varphi_1, \dots, \varphi_p)^\top \in \Theta_\varphi; & \boldsymbol{\vartheta} &= (\vartheta_1, \dots, \vartheta_q)^\top \in \Theta_\vartheta, \end{aligned} \quad (2.2)$$

with Θ_ϕ and Θ_φ (Θ_θ and Θ_ϑ) being subsets of $\mathbb{R}^{p+1}$ ($\mathbb{R}^q$). Moreover, let

$\boldsymbol{\Psi}$ be the vector that contains the parameters to be tested, partitioned as

$\boldsymbol{\Psi} = (\boldsymbol{\Psi}_1^\top, \boldsymbol{\Psi}_2^\top)^\top$, where

$$\boldsymbol{\Psi}_1 = \boldsymbol{\phi}, \quad \boldsymbol{\Psi}_2 = \boldsymbol{\varphi}, \quad \text{if only the AR part is tested (sLM),} \quad (2.3)$$

$$\boldsymbol{\Psi}_1 = (\boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top)^\top, \quad \boldsymbol{\Psi}_2 = (\boldsymbol{\varphi}^\top, \boldsymbol{\vartheta}^\top)^\top, \quad \text{if both the AR and the MA parts are tested (sLM*).$$

All true parameters have a subscript zero while unknown ones have not.

Specifically, the true parameters are denoted by

$$\begin{aligned} \boldsymbol{\eta}_0 &= (\boldsymbol{\eta}_{0,1}^\top, \boldsymbol{\eta}_{0,2}^\top, \sigma_0^2)^\top; & \boldsymbol{\Psi}_0 &= (\boldsymbol{\Psi}_{0,1}^\top, \boldsymbol{\Psi}_{0,2}^\top)^\top; \\ \boldsymbol{\phi}_0 &= (\phi_{0,0}, \phi_{0,1}, \dots, \phi_{0,p})^\top; & \boldsymbol{\theta}_0 &= (\theta_{0,1}, \dots, \theta_{0,q})^\top; \\ \boldsymbol{\varphi}_0 &= (\varphi_{0,0}, \varphi_{0,1}, \dots, \varphi_{0,p})^\top; & \boldsymbol{\vartheta}_0 &= (\vartheta_{0,1}, \dots, \vartheta_{0,q})^\top. \end{aligned}$$

Also, we assume $\boldsymbol{\eta}_0$ to be an interior point of the parameter space.

We test whether a TARMA(p, q) model provides a significantly better fit than the linear ARMA(p, q) model. To this end, we develop two Lagrange multiplier test statistics for the hypotheses

$$\begin{cases} H_0 & : \Psi_{0,2} = \mathbf{0}, \\ H_1 & : \Psi_{0,2} \neq \mathbf{0}, \end{cases}$$

where $\mathbf{0}$ is the vector with all zeroes. The statistic for testing the threshold effect in the AR parameters is denoted as sLM, whereas sLM * is the statistic for the general test where both the AR and the MA parameters change across regimes. Under H_0 , the time series follows a linear ARMA(p, q) model:

$$X_t = \phi_{0,0} + \sum_{i=1}^p \phi_{0,i} X_{t-i} - \sum_{j=1}^q \theta_{0,j} \varepsilon_{t-j} + \varepsilon_t. \quad (2.4)$$

To study the asymptotic properties of the tests, we assume the model to be ergodic and invertible under the null hypothesis and under the alternative hypothesis. Let the autoregressive and moving-average polynomials be defined as follows:

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p; \quad \theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q.$$

We assume the following:

Assumption A1. $\phi(z) \neq 0$ and $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$

and they do not share common roots. Moreover, $\{X_t\}$ is ergodic and invertible both under H_0 and under H_1 .

Note that the ergodicity of TARMA models was investigated in Ling (1999), Chan and Goracci (2019), whereas the invertibility of TARMA models was studied in Chan and Tong (2010). A discussion on the invertibility of TMA models can also be found in Ling and Tong (2005) and Ling et al. (2007).

From now on we fully develop the theory for the general statistic sLM*. Unless otherwise stated, the results also hold for the statistic sLM. Suppose we observe $X_1, \dots, X_n$. The Gaussian log-likelihood, conditional on the initial values $X_0, X_{-1}, \dots$, is given by

$$\ell_n(\boldsymbol{\eta}, r) = -\frac{n}{2} \ln(\sigma^2 2\pi) - \frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\eta}, r), \quad (2.5)$$

$$\begin{aligned} \varepsilon_t(\boldsymbol{\eta}, r) = X_t - & \left\{ \phi_0 + \sum_{i=1}^p \phi_i X_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j}(\boldsymbol{\eta}, r) \right\} \\ & - \left\{ \varphi_0 + \sum_{i=1}^p \varphi_i X_{t-i} - \sum_{j=1}^q \vartheta_j \varepsilon_{t-j}(\boldsymbol{\eta}, r) \right\} I_r(X_{t-d}) \end{aligned} \quad (2.6)$$

and the q initial values are set to zero. The rightmost summation involving ϑ_j , $j = 1, \dots, q$ is absent for the sLM test. Also,

$$\varepsilon_t(\boldsymbol{\eta}_1) = \varepsilon_t(\boldsymbol{\eta}, -\infty) = X_t - \left\{ \phi_0 + \sum_{i=1}^p \phi_i X_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j}(\boldsymbol{\eta}_1) \right\}. \quad (2.7)$$

Consider the partial derivative of $\ell_n(\boldsymbol{\eta}, r)$ with respect to $\boldsymbol{\Psi}$.

$$\frac{\partial \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}} = \left(\left(\frac{\partial \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_1} \right)^\top, \left(\frac{\partial \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_2} \right)^\top \right)^\top = -\frac{1}{\sigma^2} \sum_{t=1}^n \varepsilon_t(\boldsymbol{\eta}, r) \frac{\partial \varepsilon_t(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}},$$

where $\partial \varepsilon_t(\boldsymbol{\eta}, r)/\partial \boldsymbol{\Psi}$ is the partial derivative of $\varepsilon_t(\boldsymbol{\eta}, r)$ with respect to $\boldsymbol{\Psi}$, namely

$$\frac{\partial \varepsilon_t(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}} = \left(\left(\frac{\partial \varepsilon_t(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_1} \right)^\top, \left(\frac{\partial \varepsilon_t(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_2} \right)^\top \right)^\top. \quad (2.8)$$

Moreover, let $\partial \varepsilon_t(\boldsymbol{\eta}_1)/\partial \boldsymbol{\Psi}_1$ be the derivative of the function $\varepsilon_t(\boldsymbol{\eta}_1)$ defined in Eq. (2.7). Lastly, define the block matrix $\mathcal{I}_n(\boldsymbol{\eta}, r)$ as follows:

$$\mathcal{I}_n(\boldsymbol{\eta}, r) = \begin{pmatrix} \mathcal{I}_{n,11}(\boldsymbol{\eta}) & \mathcal{I}_{n,12}(\boldsymbol{\eta}, r) \\ \mathcal{I}_{n,21}(\boldsymbol{\eta}, r) & \mathcal{I}_{n,22}(\boldsymbol{\eta}, r) \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_1 \partial \boldsymbol{\Psi}_1^\top} & -\frac{\partial^2 \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_1 \partial \boldsymbol{\Psi}_2^\top} \\ -\frac{\partial^2 \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_2 \partial \boldsymbol{\Psi}_1^\top} & -\frac{\partial^2 \ell_n(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}_2 \partial \boldsymbol{\Psi}_2^\top} \end{pmatrix}. \quad (2.9)$$

Let $\ell_n(\boldsymbol{\eta}_1) = \ell_n(\boldsymbol{\eta}, -\infty)$. Let $\hat{\boldsymbol{\eta}}_1 = (\hat{\boldsymbol{\phi}}^\top, \hat{\boldsymbol{\theta}}^\top)^\top = \arg \min_{\boldsymbol{\eta}_1} \ell_n(\boldsymbol{\eta}_1)$ be the maximum likelihood estimator (hereafter MLE) of the ARMA coefficients in Eq. (2.4) and $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^n \varepsilon_t(\hat{\boldsymbol{\eta}}_1)$. We define $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\eta}}_1^\top, \mathbf{0}^\top, \hat{\sigma}^2)^\top$ to be the so-called *restricted* MLE, i.e. under the null hypothesis. We write $\partial \hat{\ell}_n(r)/\partial \boldsymbol{\Psi}_2$ and $\hat{\mathcal{I}}_n(r)$ to refer to $\partial \ell_n(\boldsymbol{\eta}, r)/\partial \boldsymbol{\Psi}$ and $\mathcal{I}_n(\boldsymbol{\eta}, r)$ evaluated at the restricted MLE, i.e.

$$\frac{\partial \hat{\ell}_n(r)}{\partial \boldsymbol{\Psi}_2} = \frac{\partial \ell_n(\hat{\boldsymbol{\eta}}, r)}{\partial \boldsymbol{\Psi}_2}; \quad \hat{\mathcal{I}}_n(r) = \mathcal{I}_n(\hat{\boldsymbol{\eta}}, r) = \begin{pmatrix} \hat{\mathcal{I}}_{n,11} & \hat{\mathcal{I}}_{n,12}(r) \\ \hat{\mathcal{I}}_{n,21}(r) & \hat{\mathcal{I}}_{n,22}(r) \end{pmatrix}.$$

Under the null hypothesis, the threshold parameter r is absent and so standard asymptotic theory is not applicable. To cope with this issue, we firstly develop the Lagrange multiplier test statistic as a function of r ranging over a set $\mathcal{R}$. Then, we take the overall test statistic as the supremum on $\mathcal{R}$. We set $\mathcal{R} = [r_L, r_U]$, with r_L and r_U being some percentiles of the sample.

This approach has become widely used in the time series literature for tests involving threshold models. It could be dated back to [Chan \(1990\)](#) and was followed by others such as [Ling and Tong \(2005\)](#), [Li and Li \(2011\)](#), [Chan et al. \(2020\)](#). Our test statistic is

$$T_n = \sup_{r \in [r_L, r_U]} T_n(r), \quad (2.10)$$

where

$$T_n(r) = \left(\frac{\partial \hat{\ell}_n(r)}{\partial \Psi_2} \right)^\top \left(\hat{\mathcal{I}}_{n,22}(r) - \hat{\mathcal{I}}_{n,21}(r) \hat{\mathcal{I}}_{n,11}^{-1} \hat{\mathcal{I}}_{n,12}(r) \right)^{-1} \frac{\partial \hat{\ell}_n(r)}{\partial \Psi_2}. \quad (2.11)$$

3. The null distribution

In this section we derive the asymptotic distribution of T_n under the null hypothesis that $\{X_t\}$ follows an ARMA(p, q) model of AR order p and MA order q , and the true data generating process (DGP) is defined in Eq. (2.4). Unless stated otherwise, all expectations are taken with respect to the true probability distribution under H_0 . Also, $o_p(1)$ denotes the convergence in probability to zero as n increases and $\|\cdot\|$ is the $\mathcal{L}^2$ matrix norm (the Frobenius' norm, i.e. $\|A\| = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2}$, where A is a $n \times m$ matrix). Moreover, let $\mathcal{D}_{\mathbb{R}}(a, b)$, $a < b$, be the space of functions from (a, b) to $\mathbb{R}$ that are right continuous with left-hand limits. $\mathcal{D}_{\mathbb{R}}(a, b)$ is equipped with the topology of uniform convergence on compact sets; see [Billingsley \(1968\)](#)

for more details.

In Lemma 1 of the Supplementary Material, we rewrite $\partial\varepsilon_t(\boldsymbol{\eta}, r)/\partial\boldsymbol{\Psi}$ as a function of the roots of the moving-average characteristic polynomial. It is an alternative way of representing the derivatives that simplifies theoretical derivations. Under the null hypothesis and $\boldsymbol{\eta} = \boldsymbol{\eta}_0$, we obtain the same expansion for the partial derivatives of $\varepsilon_t(\boldsymbol{\eta}_0, r)$ described by Eqs. (6.3) and (6.4) in Ling and Tong (2005). In Lemma 2 of the Supplementary Material we derive an approximation for the matrix of second derivatives of the log-likelihood function under H_0 .

Define the vector $\nabla_n(r) = (\nabla_{n,1}^\top, \nabla_{n,2}^\top(r))^\top$, with

$$\nabla_{n,1} = -\frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^n \varepsilon_t \frac{\partial \varepsilon_t(\boldsymbol{\eta}_{0,1})}{\partial \boldsymbol{\Psi}_1}, \quad \nabla_{n,2}(r) = -\frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^n \varepsilon_t \frac{\partial \varepsilon_t(\boldsymbol{\eta}_0, r)}{\partial \boldsymbol{\Psi}_2}, \quad (3.12)$$

and the matrices

$$\Lambda(\boldsymbol{\eta}, r) = \begin{pmatrix} \Lambda_{11}(\boldsymbol{\eta}) & \Lambda_{12}(\boldsymbol{\eta}, r) \\ \Lambda_{21}(\boldsymbol{\eta}, r) & \Lambda_{22}(\boldsymbol{\eta}, r) \end{pmatrix} = E \left[\frac{1}{\sigma^2} \left(\frac{\partial \varepsilon_t(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}} \right) \left(\frac{\partial \varepsilon_t(\boldsymbol{\eta}, r)}{\partial \boldsymbol{\Psi}} \right)^\top \right],$$

$$\Lambda(r) = \begin{pmatrix} \Lambda_{11} & \Lambda_{12}(r) \\ \Lambda_{21}(r) & \Lambda_{22}(r) \end{pmatrix} = E \left[\frac{1}{\sigma_0^2} \left(\frac{\partial \varepsilon_t(\boldsymbol{\eta}_0, r)}{\partial \boldsymbol{\Psi}} \right) \left(\frac{\partial \varepsilon_t(\boldsymbol{\eta}_0, r)}{\partial \boldsymbol{\Psi}} \right)^\top \right],$$

where the sub-matrices Λ 's are $(p+1) \times (p+1)$ matrices in the test sLM, and $(p+q+1) \times (p+q+1)$ matrices in the test sLM*.

In the following proposition we provide a uniform approximation of the test statistics that will allow us to derive the asymptotic null distribution of the test statistic T_n in Theorem 1.

Proposition 1. *Under Assumption A1 and under H_0 , we have the following results.*

(i) *For each $\boldsymbol{\eta}$, the matrix $\Lambda(\boldsymbol{\eta}, r)$ is positive definite.*

(ii)

$$\sup_{r \in [a, b]} \left\| \left(\frac{\hat{\mathcal{I}}_{n,22}(r)}{n} - \frac{\hat{\mathcal{I}}_{n,21}(r)}{n} \left(\frac{\hat{\mathcal{I}}_{n,11}}{n} \right)^{-1} \frac{\hat{\mathcal{I}}_{n,12}(r)}{n} \right)^{-1} - \left(\Lambda_{22}(r) - \Lambda_{21}(r) \Lambda_{11}^{-1} \Lambda_{12}(r) \right)^{-1} \right\| = o_p(1).$$

(iii)

$$\sup_{r \in [a, b]} \left\| \frac{1}{\sqrt{n}} \frac{\partial \hat{\ell}_n(r)}{\partial \boldsymbol{\Psi}_2} - \left(\nabla_{n,2}(r) - \Lambda_{21}(r) \Lambda_{11}^{-1} \nabla_{n,1} \right) \right\| = o_p(1).$$

Note that in the above proposition, $(\nabla_{n,2}(r) - \Lambda_{21}(r) \Lambda_{11}^{-1} \nabla_{n,1})$ is a linear marked empirical process with infinitely many markers. Similar to [Ling and Tong \(2005\)](#) and [Li and Li \(2011\)](#), we rely on Assumption A2 stated below.

A2. ε_t has a continuous and strictly positive density on the real line and

$E[\varepsilon_t^4]$ is finite.

Theorem 1. *Let $\{\xi(r), r \in \mathbb{R}\}$ be a centered Gaussian process with covariance kernel $\Sigma(r_1, r_2) = \Lambda_{22}(r_1 \wedge r_2) - \Lambda_{21}(r_1)\Lambda_{11}^{-1}\Lambda_{12}(r_2)$. Then, under H_0 and Assumptions A1 and A2, T_n converges weakly to*

$$\sup_{r \in [r_L, r_U]} \xi(r)^\top (\Lambda_{22}(r) - \Lambda_{21}(r)\Lambda_{11}^{-1}\Lambda_{12}(r))^{-1} \xi(r).$$

In Table 1 of the Supplementary Material we tabulate the empirical quantiles of the null asymptotic distribution of our supLM statistics for autoregressive orders from 1 to 4 and moving-average orders from 1 to 2. The quantiles do not depend upon the moving-average parameters and are in good agreement with those of table 1 of Andrews (2003), where $\pi_0 = 0.25$, and those in table 1 of Chan (1991). The asymptotic behaviour of the two statistics depends only upon the dimension of the parameter vector Ψ , irrespective of its components being either autoregressive or moving-average. As shown in Section S1 of the Supplementary Material, the results do not depend upon the parameters' values of the data generating process and are robust with respect to deviations from normality of the innovation process.

4. The distribution under local alternatives and consistency of the test

In this section we derive the asymptotic distribution of T_n under a sequence of local alternatives and prove the consistency of the associated tests. For each n , the null hypothesis H_0 states that $\{X_t, t = 1, \dots, n\}$ follows the model

$$X_t = \phi_{0,0} + \sum_{i=1}^p \phi_{0,i} X_{t-i} - \sum_{j=1}^q \theta_{0,j} \varepsilon_{t-j} + \varepsilon_t.$$

The sequence of alternative hypotheses $H_{1,n}$ states that $\{X_t, t = 1, \dots, n\}$ follows the model

$$\begin{aligned} X_t = & \phi_{0,0} + \sum_{i=1}^p \phi_{0,i} X_{t-i} - \sum_{j=1}^q \theta_{0,j} \varepsilon_{t-j} + \varepsilon_t \\ & + \left[\frac{h_{10}}{\sqrt{n}} + \sum_{i=1}^p \frac{h_{1i}}{\sqrt{n}} X_{t-i} - \sum_{j=1}^q \frac{h_{2j}}{\sqrt{n}} \varepsilon_{t-j} \right] I_{r_0}(X_{t-d}). \end{aligned} \quad (4.13)$$

Here, $\mathbf{h} = (h_{10}, h_{11}, \dots, h_{1p}, h_{21}, \dots, h_{2q})^\top \in \mathbb{R}^{p+q+1}$ is a vector of constants and r_0 is a fixed scalar. In the case of the sLM statistic, the rightmost summation term within square brackets is absent so that $\mathbf{h} = (h_{10}, h_{11}, \dots, h_{1p})^\top \in \mathbb{R}^{p+1}$.

Let $P_{0,n}$ and $P_{1,n}$ be the probability measure of $(X_1, \dots, X_n)$ under H_0 and $H_{1,n}$, respectively. In the following proposition we prove the asymptotic normality of the log-likelihood ratio and the contiguity of $P_{1,n}$ to $P_{0,n}$. As

in C5 of [Chan \(1990\)](#), 3.1 of [Ling and Tong \(2005\)](#), A5 of [Li and Li \(2011\)](#),

we assume

A3. The density f of ε_t is absolutely continuous with derivative f' almost everywhere and $\int (f'(x)/f(x))^2 f(x) dx < \infty$.

Proposition 2. *Under Assumptions A1-A3, the following assertions hold.*

(i) *Let $\nabla_2(r)$ be a Gaussian distributed random vector with zero mean and covariance matrix equal to $\Lambda_{22}(r)$. Under the null hypothesis, the log-likelihood ratio $\log(dP_{1,n}/dP_{0,n})$ converges to the Gaussian random variable $\mathbf{h}^\top \nabla_2(r_0) - 2^{-1} \mathbf{h}^\top \Lambda_{22}(r_0) \mathbf{h}$.*

(ii) *$\{P_{1,n}\}$ is contiguous to $\{P_{0,n}\}$.*

Next, we have the following asymptotic distribution for T_n .

Theorem 2. *Assume that A1-A3 hold. Under $H_{1,n}$, the following assertions hold.*

(i) *T_n converges weakly in $\mathcal{D}(-\infty, \infty)$ to*

$$(\xi_r + \gamma_r)^\top (\Lambda_{22}(r) - \Lambda_{2,1}(r) \Lambda_{11}^{-1} \Lambda_{12}(r))^{-1} (\xi_r + \gamma_r),$$

where $\gamma_r = \{\Lambda_{22}(\min\{r, r_0\}) - \Lambda_{21}(r) \Lambda_{11}^{-1} \Lambda_{12}(r_0)\} \mathbf{h}$.

(ii) $\sup_{r \in [r_L, r_U]} T_n$ converges weakly to

$$\sup_{r \in [r_L, r_U]} (\xi_r + \gamma_r)^\top (\Lambda_{22}(r) - \Lambda_{21}(r) \Lambda_{11}^{-1} \Lambda_{12}(r))^{-1} (\xi_r + \gamma_r).$$

Finally, we prove the consistency of our tests.

Theorem 3. *Under $H_{1,n}$, as $\|\mathbf{h}\| \rightarrow \infty$, the test statistic T_n has power approaching 100%.*

Note that Proposition 2 and Theorem 2 can be proved for the sLM statistic without assumption A3. The two proofs are reported separately in the Supplementary Material.

5. Finite-sample performance

In this section we investigate the finite sample performance of our supLM tests (sLM, sLM*) and compare them with the quasi-likelihood ratio test developed in Li and Li (2011) (qLR). Hereafter ε_t , $t = 1, \dots, n$ is generated from a standard Gaussian white noise, the length of the series is $n = 100, 200, 500$, the nominal size is $\alpha = 5\%$ and the number of Monte Carlo replications is 1,000. For our tests we have used the tabulated values of Table 1. For the qLR test we have used $B = 1,000$ resamples. In Section 5.1, we study the size of the tests. Section 5.2 shows the

5.1 Size of the tests

power of the tests in scenarios where *i*) only the autoregressive parameters change across regimes, *ii*) only the moving-average parameters change across regimes and *iii*) both the autoregressive and the moving-average parameters change across regimes. Then, we assess the behaviour of the tests in presence of model mis-specification in Section 5.3 and when the order of the ARMA process tested is treated as unknown and is selected by means of the Hannan-Rissanen method in Section 5.4.

5.1 Size of the tests

We have generated time series from 25 different simulation settings of the following ARMA(1, 1) model:

$$X_t = \phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}, \quad (5.14)$$

where $\phi_1 = 0, \pm 0.3, \pm 0.6$ and $\theta_1 = 0, \pm 0.4, \pm 0.8$. Table 1 shows the rejection percentages for the three sample sizes in use. Note that the case $\theta_1 = 0$ corresponds to testing an AR versus a TAR model. For $n = 100$ the qLR test is biased in almost all settings and reaches 53% of false rejections for the case $\theta_1 = \phi_1 = 0$. The size of the sLM test is always acceptable as it is slightly greater than 10% only in three cases and its maximum value is 15.5%. The size of the sLM* is slightly more biased than that of the sLM test. When $n = 200$ the bias of the sLM test reduces further and its

5.1 Size of the tests

Table 1: Empirical size at nominal level 5% of the supLM tests (sLM, sLM^{*}) and the quasi-likelihood Ratio test (qLR). Rejection percentages from the ARMA(1, 1) model of Eq. (5.14). Sample size $n = 100, 200, 500$.

ϕ_1	θ_1	$n = 100$			$n = 200$			$n = 500$		
		sLM	sLM [*]	qLR	sLM	sLM [*]	qLR	sLM	sLM [*]	qLR
-0.6	-0.8	5.8	10.0	45.0	4.6	6.2	25.2	4.4	5.7	10.2
-0.3	-0.8	3.4	8.4	29.1	3.4	7.0	14.9	4.9	5.3	7.3
0.0	-0.8	4.3	7.4	22.4	4.1	7.2	9.2	4.2	4.2	6.3
0.3	-0.8	4.2	8.7	21.0	5.0	8.0	9.8	4.9	4.6	5.6
0.6	-0.8	4.6	8.4	19.5	5.6	7.9	8.1	4.4	5.8	4.9
-0.6	-0.4	6.1	8.5	41.0	5.1	5.2	26.2	4.7	4.8	13.9
-0.3	-0.4	7.5	8.3	46.1	4.9	5.0	40.7	5.9	5.8	24.0
0.0	-0.4	4.5	4.2	20.4	4.1	5.2	8.2	5.0	4.5	5.1
0.3	-0.4	5.4	6.2	10.8	3.8	4.3	6.0	4.3	4.3	4.7
0.6	-0.4	7.1	6.5	6.7	4.4	5.6	5.8	6.1	5.2	3.2
-0.6	0.0	4.6	4.2	10.0	4.2	4.0	5.8	4.6	4.9	3.1
-0.3	0.0	5.1	5.6	31.8	3.4	3.7	16.0	5.9	6.2	4.8
0.0	0.0	9.4	11.4	53.5	9.2	9.4	42.0	6.7	7.1	34.0
0.3	0.0	7.5	6.9	22.4	5.1	4.6	7.6	3.3	4.2	4.3
0.6	0.0	9.8	8.5	7.5	6.1	5.4	4.1	5.2	5.8	4.6
-0.6	0.4	4.2	4.8	5.9	3.3	4.3	3.7	4.6	4.4	3.1
-0.3	0.4	4.7	5.6	9.7	4.1	5.0	5.1	3.9	4.5	3.8
0.0	0.4	5.0	4.5	24.2	3.7	4.1	8.2	4.5	5.2	5.7
0.3	0.4	6.4	7.7	45.6	6.0	7.1	38.1	5.1	5.4	23.4
0.6	0.4	15.5	14.4	27.6	8.8	9.2	8.8	6.9	7.7	3.2
-0.6	0.8	10.3	16.8	20.2	7.2	10.7	7.0	4.5	7.0	5.3
-0.3	0.8	8.3	14.8	20.2	6.3	10.4	9.7	4.2	6.6	3.3
0.0	0.8	8.2	16.0	20.4	6.0	11.3	8.7	4.9	7.2	5.2
0.3	0.8	7.7	14.4	24.5	7.2	10.0	10.5	4.1	6.5	4.4
0.6	0.8	11.9	16.0	32.0	9.3	11.2	22.2	6.3	8.1	7.6

5.2 Power of the tests

size is not far from the nominal 5% in most situations. This also holds for the sLM* test, except for the case $\theta_1 = 0.8$ where the size is still around 10%. This is not the case for the qLR test whose size is close to 40% in three simulation settings. When $n = 500$ both our supLM tests achieve a size which is close to the nominal 5% level, whereas the qLR test is still severely biased for some cases when near cancellation occurs. This can be due to the sensitivity of the quadratic approximation of the qLR test to near invertibility and/or near cancellation. The bias is particularly severe when $\theta_1 = \phi_1 = 0$. One may argue that it is not appropriate to apply these tests to a realization of a white noise process and it would be more sensible to apply other kinds of tests in the first place. Nevertheless, it may occur that a threshold process resembles white noise in terms of second order structure. Indeed, some sort of mis-specification is always present and this aspect will be investigated in Section 5.3.

5.2 Power of the tests

In this section we study the power of the supLM tests and highlight the differences among them. Note that the parameter vector Ψ_2 (see Eq. (2.3)) represents the departure from the null hypothesis and in all the simulations below we take sequences of increasing distance from H_0 in all of its com-

5.2 Power of the tests

ponents. We simulate three different TARMA(1,1) models where *i*) only the autoregressive parameters change across regimes, *ii*) only the moving-average parameters change across regimes, and *iii*) both the autoregressive and the moving-average parameters change across regimes. As for the first case, we simulate the following model:

$$X_t = -0.5 - 0.2X_{t-1} - \theta_1\varepsilon_{t-1} + (\varphi_0 + \varphi_1X_{t-1})I(X_{t-1} \leq 0) + \varepsilon_t, \quad (5.15)$$

where $(\varphi_0, \varphi_1) = (0.1, 0.4), (0.3, 0.6), (0.5, 0.8), (0.7, 1.0)$. We combine these with $\theta_1 = 0, \pm 0.4, \pm 0.8$ so as to obtain 20 different parameter settings. Figure 1 presents the size-corrected power of the tests (in percentage) for different values of θ_1 . Clearly, the supLM tests outperform the qLR test in all cases except for the single instance $\theta = 0.4$ and $n = 100$. The power depends upon the true value of θ_1 and the case $\theta_1 = 0$ seems to impinge most negatively. In such instances, the qLR test has no power even for $n = 200$ whereas the supLM tests show power loss due to the size correction only for $n = 100$. Moreover, for $\theta = 0.8$ both supLM tests outperform neatly the qLR test. Indeed, their power for $n = 200(100)$ is greater than that of the qLR test for $n = 500(200)$. Overall, starting from $n = 200$ both supLM tests possess good power in almost every situation. As expected, the sLM test is slightly superior to the sLM* test since the moving-average parameter is regime-independent. The results for case *ii*) and *iii*) and for

5.3 Size and power in presence of mis-specification

higher order TARMA models are consistent with the above conclusions and are reported in the Supplement. In particular, they show that when both the AR and the MA parameters change, the sLM^{*} is the most powerful.

5.3 Size and power in presence of mis-specification

In this section we assess the impact of model mis-specification upon the performance of the tests. The sources of mis-specification can be diverse. As above, we focus on testing the ARMA(1, 1) versus the TARMA(1, 1) specification but the data generating process is not encompassed within the two models. Loosely speaking, we are investigating the capability of the test to detect general departures from linearity beyond the direct comparison of two specific models. Ideally, if the data generating process is linear, we would want the test not to reject the null hypothesis. Likewise, if the data generating process is non-linear in some of its components, then we expect the test to reject the null hypothesis. The list of linear and non-linear data generating processes used is presented in Table 7 of the Supplementary Material. The seven linear processes are not ARMA(1, 1) since they contain higher-order autoregressive/moving-average terms. In the second part of the table we show non-linear processes that cannot be encompassed within the two-regime TARMA(1, 1) specification. In particular, we sim-

5.3 Size and power in presence of mis-specification

Figure 1: Size-corrected power (percentages) at nominal $\alpha = 5\%$ for the TARMA(1,1) model of Eq. (5.15), case i): only the autoregressive parameters change across regimes. The sLM, sLM* and qLR tests are denoted with a blue filled point, an empty green point and a red triangle, respectively. The line type indicates the sample size: $n = 100$ (continuous), $n = 200$ (dashed), $n = 500$ (dotted).

5.3 Size and power in presence of mis-specification

Table 2: Rejection percentages under model mis-specification for the processes of Table 7 of the Supplementary Material for the supLM tests (sLM, sLM*) and the quasi-likelihood ratio test (qLR). The upper panel (linear processes) reflects the empirical size at nominal level 5%. The lower panel (non-linear processes) reflects the empirical power for non-linear processes that are not representable as a TARMA(1, 1) process.

		<i>n</i> = 100			<i>n</i> = 200			<i>n</i> = 500		
		sLM	sLM*	qLR	sLM	sLM*	qLR	sLM	sLM*	qLR
LINEAR	AR5	7.9	15.2	22.7	5.5	9.8	7.8	4.6	6.9	4.6
	AR2.1	6.7	7.1	6.8	4.7	5.5	4.0	5.3	5.5	4.0
	AR2.2	8.7	10.1	6.6	4.1	5.8	3.2	3.4	5.0	1.6
	ARMA21.1	5.9	6.8	48.0	5.5	5.2	39.7	5.5	4.8	27.8
	ARMA21.2	5.8	8.9	15.8	4.5	5.7	8.3	4.0	4.9	4.1
	ARMA22	5.5	13.6	29.1	3.7	12.7	17.4	3.5	7.5	8.5
	MA2	3.5	12.1	29.9	2.3	5.8	17.0	4.4	6.2	6.5
NON-LINEAR	TAR3	34.7	98.1	63.0	61.8	99.7	48.0	96.7	100.0	36.8
	3TAR1	19.2	18.4	8.8	36.1	30.4	8.5	78.5	72.4	16.5
	NLMA.1	85.1	84.0	78.7	97.3	97.2	92.1	99.8	99.9	99.0
	NLMA.2	86.4	86.8	73.6	98.3	98.5	85.2	100.0	99.9	97.0
	BIL.1	12.0	62.1	53.2	13.5	84.8	58.2	14.7	95.0	77.4
	BIL.2	84.0	83.1	86.8	98.7	98.9	96.0	100.0	100.0	99.9
	EXPAR.1	100.0	100.0	32.5	100.0	100.0	30.5	100.0	100.0	80.1
	EXPAR.2	95.8	99.1	45.0	99.6	99.9	59.3	100.0	100.0	96.2
	NLAR	100.0	100.0	63.7	100.0	100.0	47.0	100.0	100.0	27.1

ulate from TAR models with both higher autoregressive order and more than two regimes. Lastly, we generate from seven non-linear models that do not belong to the TARMA class, such as the non-linear moving-average (NLMA), the bilinear (BIL), the exponential autoregressive (EXPAR) and deterministic chaos (NLAR).

The rejection percentages are reported in Table 2. As discussed above,

5.3 Size and power in presence of mis-specification

the first seven rows should reflect the empirical size at nominal level 5% under mis-specification. Consistent with the results of Table 1, the sLM test is well behaved in terms of size even for $n = 100$ whereas the sLM* test has acceptable size starting from $n = 200$. The qLR test presents acceptable size just for $n = 500$, except for the ARMA21.1 case with a 27.8% of false rejections. The lower panel of Table 2 shows the rejection percentages for the 9 non-linear models that do not belong to the TARMA(1, 1) class. Here, the power of the sLM tests is higher than qLR in almost every case, even for $n = 100$, and increases consistently with the sample size. The qLR test has good power in several instances but for the TAR3, the 3TAR1 and the NLAR models the power decreases as the sample size increases. The causes of this phenomenon are not clear and deserve further investigation. As pointed out by one referee, one possible intuitive explanation is the following: consider the case where we test the null hypothesis of an AR(p) model. In practice, the LM statistics test whether $E[\varepsilon_t X_{t-i} I_r(X_{t-d})] = 0$, $i = 1, \dots, p$. When the true model is a linear model, the above result most likely holds, and consequently the sLM tests still have an accurate size. However, when the true model is a non-linear model the result most likely does not hold, so the proposed LM tests are still powerful. The above arguments fail for the qLR test, which needs to consider the model under

5.4 The impact of model selection

Table 3: Empirical size of the supLM tests at nominal level 5% for 6 parameterizations of the ARMA(2, 2) process. The subscript “HR” indicates that the order of the ARMA model has been selected through the Hannan-Rissanen procedure.

ϕ_1	ϕ_2	θ_1	θ_2	$n = 100$		$n = 200$		$n = 500$	
				sLM	sLM _{HR}	sLM	sLM _{HR}	sLM	sLM _{HR}
-0.35	-0.45	0.25	-0.25	5.4	4.4	3.6	3.7	4.9	4.9
0.45	-0.55	0.25	-0.25	5.6	5.8	5.7	4.8	4.6	4.6
-0.90	-0.25	0.25	-0.25	6.1	4.6	5.8	5.9	4.4	5.6
-0.35	-0.45	-0.25	0.25	2.9	3.0	4.3	4.4	5.3	5.3
0.45	-0.55	-0.25	0.25	3.0	3.4	3.9	4.3	5.3	5.0
-0.90	-0.25	-0.25	0.25	7.1	4.7	6.8	4.6	5.7	5.2

ϕ_1	ϕ_2	θ_1	θ_2	$n = 100$		$n = 200$		$n = 500$	
				sLM*	sLM* _{HR}	sLM*	sLM* _{HR}	sLM*	sLM* _{HR}
-0.35	-0.45	0.25	-0.25	6.1	4.3	4.3	3.7	4.5	4.9
0.45	-0.55	0.25	-0.25	7.2	6.4	4.8	5.3	4.5	5.1
-0.90	-0.25	0.25	-0.25	6.7	5.9	4.9	5.3	4.8	5.4
-0.35	-0.45	-0.25	0.25	3.1	3.0	4.9	4.8	4.6	4.2
0.45	-0.55	-0.25	0.25	4.4	4.8	4.1	4.7	4.9	4.8
-0.90	-0.25	-0.25	0.25	11.4	6.5	6.7	4.3	5.3	5.2

both the null and the alternative hypotheses.

5.4 The impact of model selection

In practice, the tests require selecting the model order beforehand. In this section we show that there is virtually no loss incurred when using our supLM tests without prior information on the order, provided a proper model selection procedure is adopted. Our experience favours the use of

5.4 The impact of model selection

the order selection proposed in Hannan and Rissanen (1982). See also Choi (1992).

In Table 3 we present the empirical size of the supLM tests at nominal level 5% for 6 parameterizations of an ARMA(2, 2) model (the first four columns). The upper panel of the table refers to the sLM test whereas the lower panel refers to the sLM* test. In both cases, the subscript HR indicates that the order of the ARMA process has been selected through the Hannan-Rissanen procedure; no subscript is used if the true order has been used. The results indicate that, not only the model selection step does not produce a size bias, but it also seems to reduce it in some instances.

The impact of model selection on the power of the tests is shown in Table 4 where we simulate 12 parameter settings of the TARMA(1, 1) model of Eq. (5.15). Clearly, the power loss produced by using model selection is minimal and lies within 2% for the sLM test and 3% for the sLM* test. Finally, also in presence of mis-specification the HR model selection step poses no problems. The results are shown in the Supplementary material.

5.4 The impact of model selection

Table 4: Empirical power for the TARMA(1, 1) model of Eq. (5.15). Sample size $n = 100, 200, 500$. The subscript “HR” indicates that the order of the ARMA model has been selected through the Hannan-Rissanen procedure.

φ_0	φ_1	θ_1	$n = 100$		$n = 200$		$n = 500$	
			sLM	sLM _{HR}	sLM	sLM _{HR}	sLM	sLM _{HR}
0.1	0.4	-0.5	13.0	12.9	30.2	29.4	71.2	69.4
0.3	0.6	-0.5	34.8	35.1	71.9	70.6	99.6	98.9
0.5	0.8	-0.5	65.5	65.2	97.1	96.3	100.0	100.0
0.7	1.0	-0.5	90.4	89.6	99.7	99.5	100.0	100.0
0.1	0.4	0.0	15.7	14.9	19.0	19.2	43.0	43.0
0.3	0.6	0.0	30.7	30.4	51.8	51.0	94.0	93.6
0.5	0.8	0.0	54.5	54.6	87.9	87.5	100.0	99.9
0.7	1.0	0.0	81.3	80.5	99.1	99.1	100.0	100.0
0.1	0.4	0.5	17.4	16.0	35.9	33.8	77.9	75.8
0.3	0.6	0.5	41.3	38.7	78.2	73.1	99.9	99.6
0.5	0.8	0.5	70.7	67.2	95.7	93.9	99.9	99.7
0.7	1.0	0.5	75.5	74.6	93.7	93.1	99.5	99.5

φ_0	φ_1	θ_1	$n = 100$		$n = 200$		$n = 500$	
			sLM*	sLM* _{HR}	sLM*	sLM* _{HR}	sLM*	sLM* _{HR}
0.1	0.4	-0.5	13.9	13.9	28.7	28.7	66.7	65.0
0.3	0.6	-0.5	35.5	35.7	69.5	67.0	99.0	98.6
0.5	0.8	-0.5	65.7	65.1	96.4	95.7	100.0	100.0
0.7	1.0	-0.5	89.7	88.8	99.8	99.6	100.0	100.0
0.1	0.4	0.0	15.0	14.4	18.7	19.1	40.3	40.6
0.3	0.6	0.0	27.8	28.2	48.8	48.5	91.7	91.5
0.5	0.8	0.0	52.0	52.1	86.1	85.5	100.0	99.9
0.7	1.0	0.0	80.1	79.7	98.9	98.7	100.0	100.0
0.1	0.4	0.5	16.4	16.0	32.6	31.2	75.4	73.6
0.3	0.6	0.5	40.8	38.7	75.3	71.6	99.6	98.9
0.5	0.8	0.5	68.3	65.8	95.9	94.5	100.0	100.0
0.7	1.0	0.5	74.8	74.3	93.7	93.0	99.6	99.6

5.5 Discussion

The Monte Carlo study has shown that supLM tests have good finite sample properties. They are also robust against model mis-specification and their performance is not affected if the order of the tested process is unknown, provided a consistent order selection procedure is used. In theory, since the asymptotic distribution of supLM statistics is the same as that of the qLR statistic, the tests should display the same asymptotic behaviour. Nevertheless, the Monte Carlo evidence points at clear finite sample differences as supLM tests do not suffer from some of the drawbacks that affect the quasi-likelihood ratio test. The reasons can be diverse. First and foremost, supLM statistics only require fitting an ARMA model whereas the qLR test is bound to estimating a full TARMA model. We are reminded of the fact that there are no theoretical results regarding the sampling properties of the maximum likelihood estimators for the parameters of a TARMA model. Moreover, the qLR test by [Li and Li \(2011\)](#) uses a representation in terms of a quadratic form but this is only valid asymptotically so that it can impinge on the rate of convergence of the statistic towards its asymptotic distribution. This does not occur in our case since such a representation is exact for supLM statistics. The size distortion for the qLR approach is at its greatest when the model is either nearly non-invertible or there is near

cancellation in the MA and AR roots and this can reflect the sensitivity of the quadratic approximation to these two cases.

The statistic sLM tests the $\text{ARMA}(p, q)$ against the $\text{TARMA}(p, q)$ model when only the p autoregressive parameters change across regimes. However, the results show that it has power also when only the moving-average parameters change. This could be ascribed to the duality between MA and AR processes and indicates some capabilities in detecting general departures from linearity. Results of Section 5.3 provide additional evidence. The results also show that, as expected, when either only the q MA parameters or all the $p + q$ parameters of the ARMA model change across regimes, then the sLM* test is more powerful. The price to be paid for this superior power is the increased size bias in small samples. In general, we expect the two tests to behave similarly but in the case of small samples the sLM statistic is recommended and can be used in conjunction with the test based upon sLM*.

6. A real data application: the tree-ring time series

In this section we present an application of our test to the time series of the tree-ring standardized growth index. Tree rings provide a measure of the responses of tree growth to past climatic variation and this information

is very useful in climate studies. Despite the recognition that the climatic factors affecting tree growth form a complex network, according to the literature, the best model adopted is either the AR(1) or the ARMA(1, 1), see table 3 in [Fox et al. \(2001\)](#). Usually, the indices of many trees from the same site are used to cross-date the rings and are averaged into a single index to obtain a chronology that covers a long time span. Here we focus on the tree-ring chronology of a *pinus aristata* var. *longaeva* (California, USA) from year 800 to year 1979 ($n = 1180$). For more details on the data see [Graybill \(2018\)](#).

We test the null hypothesis of an ARMA(1, 1) model against the following TARMA(1, 1) model

$$X_t = \begin{cases} \phi_{10} + \phi_{11}X_{t-1} + \varepsilon_t + \theta_{11}\varepsilon_{t-1}, & \text{if } X_{t-1} \leq r, \\ \phi_{20} + \phi_{21}X_{t-1} + \varepsilon_t + \theta_{11}\varepsilon_{t-1}, & \text{otherwise.} \end{cases} \quad (6.16)$$

With the threshold searched between the 10th to 90th percentiles, the sLM test statistic is 23.45, while the sLM* statistic is 25.21 and both of them correspond to a p -value smaller than 0.001, suggesting that tree-ring growth is regulated by floor. In Figure 2(left) we show the time series whereas the right panel reports the values of the AIC criterion versus the threshold values. The vertical line indicates the estimated threshold $\hat{r} = 0.97$, the same value of r that maximizes the T_n statistic. Table 5 reports a TARMA(1, 1)

Figure 2: Time series of the yearly standardized tree-ring growth index of *Pinus aristata* var. *longaeva*, California, USA (left). AIC vs threshold r (right). The value of r that minimizes the AIC is indicated with a dashed red line.

model parameterized in the form of (6.16) with common moving-average parameter but with unconstrained $\phi_{i,1}, i = 0, 1, 2$, and an ARMA(1, 1) model fitted to the data. The estimated autoregressive parameters point to a threshold effect and the normalized AIC and BIC indicate an improvement with respect to the ARMA(1, 1) model. The estimated TARMA(1, 1) model is invertible and geometrically ergodic. The estimated threshold is $\hat{r} = 0.97$, which is close to 1, the mean of the time series. It identifies an upper regime with increased persistence with respect to the lower regime. The entropy based diagnostics of Giannerini et al. (2015), reported in the Supplementary Material, indicate that the TARMA(1, 1) model provides

Table 5: Parameter estimates for the tree-ring time series

	θ_{11}	ϕ_{10}	ϕ_{11}	ϕ_{20}	ϕ_{21}	r	d	NAIC	NBIC
ARMA	-0.60 (0.09)	1.00 (0.01)	0.76 (0.07)					0.232	0.251
TARMA	-0.44 (0.09)	0.54 (0.10)	0.37 (0.11)	0.29 (0.09)	0.71 (0.09)	0.97	1	0.213	0.240

a good fit to the data whereas an unaccounted dependence structure is present in the residuals of the ARMA(1,1) model. Note that tree-ring growth indices are produced by fitting a growth curve that can be derived from a prey-predator type differential equation describing the growth of a tree in presence of surrounding trees. Threshold models are discrete-time versions of such equations and appear as natural candidates to describe this phenomenon. Moreover, the data are likely to be affected by measurement errors and this is accounted for by the MA parameter.

7. Conclusions

In this paper, we have presented consistent supremum Lagrange Multiplier tests to test a linear ARMA model against its TARMA extension. Our proposal extends previous results, such as [Chan \(1990\)](#); [Ling and Tong \(2005\)](#) and enjoys very good finite-sample properties in terms of size and power. Moreover, being based upon asymptotic theory, it has a low computational

burden. The Monte Carlo study has shown that supLM tests are also robust against various forms of model mis-specification and their performance is not affected even if the order of the tested process is unknown, provided a consistent order selection procedure is used. Our supLM tests do not suffer from some of the shortcomings that affect the quasi-likelihood test and can be used for small samples. In such cases, the sLM statistic has less power than the sLM* statistic but it is better behaved in terms of size so that it is also recommended. For sample sizes from 200 upwards, the two tests can be used in conjunction. The theoretical framework of our supLM tests does not take into account GARCH-type innovations. A possible solution would be to adopt a wild-bootstrap scheme similar to that used in [Chan et al. \(2020\)](#). While the implementation is straightforward, to the best of our knowledge, the validity of the bootstrap in a threshold framework has not been proved, even for TAR models, and constitutes an interesting challenge for future investigations. The analysis of the tree-ring time series shows that TARMA models can provide some new insights into all those problems that make use of dendrochronological data. The TARMA(1,1) fit improves considerably over the commonly accepted linear models because the latter do not account for non-linear effects.

REFERENCES

Supplementary Materials

The supplemental document contains all the proofs and additional results from both the simulation study and the tree-ring data analysis.

Acknowledgements

We thank the Associate Editor and the referees for comments that helped improving the paper. Thanks are also due to Guodong Li for kindly making available the code for the qLR test.

References

- Andrews, D. (2003). Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica* 71(1), 395–397.
- Billingsley, P. (1968). *Convergence of probability measures*. New York: Wiley.
- Chan, K.-S. (1990, 12). Testing for threshold autoregression. *Ann. Statist.* 18(4), 1886–1894.
- Chan, K. S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 53(3), 691–696.
- Chan, K.-S. (2009). *Exploration of a Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics*. World Scientific.
- Chan, K.-S., S. Giannerini, G. Goracci, and H. Tong (2020). Unit-root test within a threshold ARMA framework. Technical report.

REFERENCES

- Chan, K.-S. and G. Goracci (2019). On the ergodicity of first-order threshold autoregressive moving-average processes. *J. Time Series Anal.* 40(2), 256–264.
- Chan, K.-S., B. Hansen, and A. Timmermann (2017). Guest editors' introduction: Regime switching and threshold models. *Journal of Business & Economic Statistics* 35(2), 159–161.
- Chan, K.-S., D. Li, S. Ling, and H. Tong (2014). On conditionally heteroscedastic AR models with thresholds. *Statistica Sinica* 24(2), 625–652.
- Chan, K.-S. and W. Li (2007). *Special issue: Threshold Models and New Developments in Time Series*, Volume 17 (1). Statistica Sinica.
- Chan, K.-S., W. Li, and Q. Yao (2011). *Special issue: Non-linear Time Series: Threshold Modelling and Beyond*, Volume 4 (2). Statistics and Its Interface.
- Chan, K.-S. and H. Tong (1990). On likelihood ratio tests for threshold autoregression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 52(3), 469–476.
- Chan, K.-S. and H. Tong (2010). A note on the invertibility of nonlinear ARMA models. *J. Statist. Plann. Inference* 140(12), 3709–3714.
- Choi, B. (1992). *ARMA Model Identification*. New York, NY: Springer US.
- Fox, J., P. Ades, and H. Bi (2001). Stochastic structure and individual-tree growth models. *Forest Ecol. Manag.* 154(1), 261 – 276.
- Giannerini, S., E. Maasoumi, and E. Bee Dagum (2015). Entropy testing for nonlinear serial dependence in time series. *Biometrika* 102, 661–675.

REFERENCES

- Goracci, G. (2020). Revisiting the canadian lynx time series analysis through TARMA models. *Statistica* 80(4), 357–394.
- Goracci, G. (2021). An empirical study on the parsimony and descriptive power of TARMA models. *Stat. Method Appl.-Ger.* 30, 109–137.
- Graybill, D. (2018). Graybill - Methuselah Walk - PILO - ITRDB CA535. <https://www.ncdc.noaa.gov/paleo/study/3376>, accessed: 2020-03-02.
- Hannan, E. and J. Rissanen (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika* 69(1), 81–94.
- Hansen, B. (2011). Threshold autoregression in economics. *Stat. Interface* 4(2), 123–127.
- Li, D., W. Li, and S. Ling (2011). On the least squares estimation of threshold autoregressive and moving-average models. *Stat. Interface* 4(2), 183–196.
- Li, G. and W. Li (2008, 04). Testing for threshold moving average with conditional heteroscedasticity. *Statist. Sinica* 18, 647–665.
- Li, G. and W. Li (2011, 02). Testing a linear time series model against its threshold extension. *Biometrika* 98(1), 243–250.
- Ling, S. (1999, 09). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *J. Appl. Probab.* 36(3), 688–705.
- Ling, S. and H. Tong (2005, 12). Testing for a linear MA model against threshold MA models. *Ann. Statist.* 33(6), 2529–2552.
- Ling, S., H. Tong, and D. Li (2007). Ergodicity and invertibility of threshold moving-average

REFERENCES

- models. *Bernoulli* 13(1), 161–168.
- Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988). Testing linearity against smooth transition autoregressive models. *Biometrika* 75(3), 491–499.
- Petrucelli, J. and N. Davies (1986, 12). A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika* 73(3), 687–694.
- Su, F. and K.-S. Chan (2017, 04). Testing for threshold diffusion. *J. Bus. Econom. Statist.* 35, 218–227.
- Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Clarendon Press.
- Tong, H. (2011). Threshold models in time series analysis—30 years on. *Stat. Interface* 4(2), 107–118.
- Tsay, R. (1998). Testing and modeling multivariate threshold models. *J. Amer. Statist. Assoc.* 93(443), 1188–1202.
- Wong, C. S. and W. K. Li (1997). Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika* 84(2), 407–418.
- Wong, C. S. and W. K. Li (2000). Testing for double threshold autoregressive conditional heteroscedastic model. *Statist. Sinica* 10(1), 173–189.

Greta Goracci

Department of Statistical Sciences, University of Bologna

Via belle arti 41, 40126, Bologna, Italy

Faculty of Economics and Management, Free University of Bozen/Bolzano

REFERENCES

piazza Università, 1, 39100, Bolzano, Italy

E-mail: (greta.goracci@unibz.it)

Simone Giannerini

Department of Statistical Sciences, University of Bologna

Via belle arti 41, 40126, Bologna, Italy

E-mail: (simone.giannerini@unibo.it)

Kung-Sik Chan,

Department of Statistics and Actuarial Sciences, University of Iowa

263 Schaeffer Hall (SH), 20 East Washington Street, Iowa City, IA 52242-1409

E-mail: (kung-sik-chan@uiowa.edu)

Howell Tong,

School of Mathematical Science, University of Electronic Science and Technology of China,

Chengdu, China

Center for Statistical Science, Tsinghua University, China;

London School of Economics and Political Science, U.K.

E-mail: (howell.tong@gmail.com)