Supplementary Material

Testing for threshold effects in the TARMA framework

Greta Goracci\(^1,6\), Simone Giannerini\(^1\), Kung-Sik Chan\(^2\), Howell Tong\(^3,4,5\)

1. University of Bologna, Italy; 2. University of Iowa, USA;
3. University of Electronic Science and Technology, Chengdu;
4. Tsinghua University, China;
5. London School of Economics and Political Science, U.K.
6. Free University of Bozen/Bolzano, Italy

This Supplement has five sections. Section S1 contains the results for the tabulated quantiles of the null distribution. Section S2 contains two technical lemmas, whereas Section S3 contains all the proofs. Section S4 contains supplementary results from the simulation study. Lastly, Section S5 presents additional results on the tree-ring data analysis.

S1 Empirical quantiles of the null distribution

In Table 1 we tabulate the empirical quantiles of the null asymptotic distribution of our supLM statistics at levels 90\%, 95\%, 99\% and 99.9\% for autoregressive orders from 1 to 4 and moving-average orders from 1 to 2. The threshold is searched between the 25th and the 75th percentiles of the
sample distribution. For each order, the results have been obtained from 10,000 simulated series of length 1,000 and are shown in Table 1. The quantiles of the asymptotic distribution of the sLM statistic do not depend upon the moving-average parameters and are in good agreement with those of Chan (1991), table 1, that refer to testing the AR against the TAR model (in that case, the length of the series was 200 and the number of replications 1,000). The asymptotic distribution of the sLM statistic is equivalent to that of Chan (1991) since the vector $\nabla_n(r)$ of Eq. (3.12) does not contain the partial derivatives w.r.t. to the moving-average part. The rightmost part of Table 1 contains the quantiles for the sLM* statistic which differ from that of the sLM statistic. In particular, the asymptotic behaviour of the two statistics depends only upon the dimension of the parameter vector $\Psi$, irrespective of its components being either autoregressive or moving-average. Finally, note that the tabulated values match those of table 1 of Andrews (2003) where $\pi_0 = 0.25$.

From Table 1 it is easy to see that the asymptotic distribution of the supLM statistics depends only upon the size of the parameter vector $\Psi$. For instance the 90% percentile of the sLM* statistic for the ARMA(2, 1) (13.48) is the same (up to sampling fluctuations) as that for the ARMA(1, 2) (13.69) since in both cases there are 3 parameters tested. Moreover, such value
S1. EMPIRICAL QUANTILES OF THE NULL DISTRIBUTION

Table 1: Tabulated quantiles for the asymptotic null distribution of the supLM statistics for the threshold range 25th-75th percentiles. The first two columns denote the AR and the MA orders, respectively.

| AR | MA | sLM 90% | sLM 95% | sLM 99% | sLM 99.9% | sLM* 90% | sLM* 95% | sLM* 99% | sLM* 99.9% |
|----|----|---------|---------|---------|-----------|---------|---------|---------|-----------|
| 1  | 1  | 9.61    | 11.37   | 15.19   | 20.38     | 11.64   | 13.44   | 17.42   | 22.83     |
| 2  | 1  | 11.53   | 13.41   | 17.22   | 22.17     | 13.48   | 15.46   | 19.63   | 25.60     |
| 3  | 1  | 13.74   | 15.71   | 19.98   | 25.04     | 15.59   | 17.61   | 21.91   | 27.98     |
| 4  | 1  | 15.65   | 17.68   | 22.25   | 27.44     | 17.42   | 19.52   | 24.02   | 29.94     |
| 1  | 2  | 9.64    | 11.47   | 15.50   | 20.25     | 13.69   | 15.57   | 19.67   | 25.07     |
| 2  | 2  | 11.71   | 13.48   | 17.61   | 22.49     | 15.59   | 17.58   | 21.95   | 28.17     |
| 3  | 2  | 13.46   | 15.35   | 19.33   | 25.06     | 17.13   | 19.18   | 23.54   | 29.78     |
| 4  | 2  | 15.55   | 17.58   | 21.82   | 27.80     | 18.97   | 21.21   | 26.42   | 31.92     |

matches the percentiles of the sLM statistic for the ARMA(3,1) (13.74) and ARMA(3,2) (13.46) since, also in this case, only the 3 autoregressive parameters are tested. This evidence suggests that a unique table could be adopted for both tests, for instance that of Andrews (2003).

The simulations show that the supLM statistics do not depend upon a specific parameters’ choice. In order to investigate further the similarity we simulate from an ARMA(1, 1) model with different parameterizations, some of which present the problem of quasi-cancellation of the AR and MA polynomials. The quantiles are presented in Table 2 and show that the rate of convergence towards the asymptotic distribution of the test statistics might depend upon the true value of the parameters so that some small discrepancies are expected in corner cases. Still, the differences do not affect the practical usability of the tests; in case of small sample sizes, a conservative
Table 2: Tabulated quantiles for the asymptotic null distribution of the supLM statistics as in Table 1 but for different parameterizations of the ARMA(1, 1) process.

| $\phi_1$ | $\theta_1$ | 90% | 95% | 99% | 99.9% | 90% | 95% | 99% | 99.9% |
|----------|------------|-----|-----|-----|-------|-----|-----|-----|-------|
| -0.60    | -0.80      | 9.54 | 11.28 | 14.68 | 19.58 | 11.68 | 13.49 | 17.38 | 22.12 |
| 0.60     | -0.80      | 10.21 | 11.72 | 15.84 | 20.48 | 12.17 | 13.95 | 17.81 | 22.17 |
| -0.60    | -0.40      | 9.49 | 11.19 | 14.95 | 19.90 | 11.64 | 13.42 | 17.54 | 23.45 |
| 0.60     | -0.40      | 9.90 | 11.44 | 14.79 | 19.84 | 11.67 | 13.63 | 17.49 | 21.96 |
| -0.60    | 0.00       | 9.51 | 11.24 | 14.93 | 19.64 | 11.52 | 13.41 | 17.21 | 23.37 |
| 0.60     | 0.00       | 9.76 | 11.41 | 14.98 | 19.89 | 11.61 | 13.54 | 17.57 | 22.18 |
| -0.60    | 0.40       | 9.68 | 11.40 | 15.23 | 19.26 | 11.89 | 13.75 | 17.87 | 22.52 |
| 0.60     | 0.80       | 10.02 | 11.77 | 15.30 | 19.80 | 12.54 | 14.44 | 18.02 | 23.05 |
| -0.60    | 0.80       | 9.99 | 11.81 | 15.51 | 20.82 | 12.17 | 14.03 | 18.02 | 23.03 |

approach could be selecting an ARMA model by means of a consistent order selection procedure. Then, one could simulate the asymptotic null by using the estimated model and sample size in use. The quantiles of the null distribution do not change appreciably when the innovations are not Gaussian. This is shown in Table 3 that tabulates the quantiles when the innovations are Student’s $t$ with 3 degrees of freedom. We have also experimented with skewed innovations with similar outcomes. The results show a close agreement up to the 99% quantile with the quantiles of Table 1, where the innovations are Gaussian. The 99.9% quantiles present some discrepancies that could be due to a slower convergence rate towards the asymptotic null distribution implied by the heavy tails of the innovations. In any case, using the tabulated quantiles up to 99% should be safe in most practical situations even in presence of non-Gaussian innovations. A similar result
Table 3: Tabulated quantiles for the asymptotic null distribution of the supLM statistics for the threshold range 25th-75th percentiles in presence of Student’s $t$ (with 3 df) innovations. The first two columns denote the AR and the MA orders, respectively.

| AR | MA | sLM 90% | sLM 95% | sLM 99% | sLM 99.9% | sLM* 90% | sLM* 95% | sLM* 99% | sLM* 99.9% |
|----|----|---------|---------|---------|-----------|---------|---------|---------|-----------|
| 1  | 1  | 8.79    | 10.63   | 14.65   | 21.30     | 11.08   | 13.04   | 17.47   | 26.24     |
| 2  | 1  | 10.98   | 12.90   | 17.35   | 27.17     | 12.99   | 15.09   | 20.18   | 31.68     |
| 3  | 1  | 12.92   | 15.14   | 19.75   | 30.15     | 14.99   | 17.12   | 22.00   | 35.62     |
| 4  | 1  | 14.91   | 17.15   | 22.56   | 35.15     | 16.74   | 18.93   | 25.41   | 39.39     |
| 1  | 2  | 9.20    | 10.91   | 14.79   | 21.64     | 13.24   | 15.25   | 20.44   | 32.53     |
| 2  | 2  | 11.28   | 13.05   | 17.51   | 26.57     | 15.02   | 17.21   | 22.65   | 35.08     |
| 3  | 2  | 13.06   | 15.30   | 20.32   | 30.55     | 17.06   | 19.43   | 25.10   | 39.59     |
| 4  | 2  | 14.97   | 17.24   | 22.58   | 35.11     | 18.89   | 21.33   | 27.52   | 43.97     |

holds for skewed innovations, we have experimented with chi-squared innovations with 5 degrees of freedom with no noticeable differences from the Gaussian case.

S2 Technical Lemmas

Lemma 1. Under Assumption A1 and under $H_0$, the following holds. Let $\alpha_h$, $h = 0, 1, 2, \ldots$, satisfy the difference equation

$$\alpha_0 = 1, \quad \alpha_h - \sum_{j=1}^{q} \theta_j \alpha_{h-j} = 0,$$

whose initial conditions are $\alpha_h = 0$, for $h < 0$. Then, for each $\eta$,

1. the $(k+1)$-th entry of $\partial \varepsilon_t(\eta, r) / \partial \phi$ is $-\sum_{h=0}^{t-1} \alpha_h$, if $k = 0$; and

$$-\sum_{h=0}^{t-k} \alpha_h X_{t-k-h}, \text{ if } 1 \leq k \leq p;$$

2. the $k$-th entry of $\partial \varepsilon_t(\eta, r) / \partial \theta$ is $\sum_{h=0}^{t-k} \alpha_h \varepsilon_{t-k-h}(\eta, r)$;
3. the \((k+1)\)-th entry of \(\partial \varepsilon_t(\eta, r)/\partial \Psi\) is 
\[-\sum_{h=0}^{t-1} \alpha_h I_r(X_{t-d-h}), \text{ if } k = 0;\]
\[-\sum_{h=0}^{t-k} \alpha_h X_{t-k-h} I_r(X_{t-d-h}), \text{ if } 1 \leq k \leq p;\]
\[-\sum_{h=0}^{t-k} \alpha_h \varepsilon_{t-k-h}(\eta, r) I_r(X_{t-d-h}), \text{ if } p+1 \leq k \leq p+q+1,\]
where, in the preceding equation, the components corresponding to \(p+1 \leq k \leq p+q+1\) are absent in the sLM test.

**Lemma 2.** Under Assumption A1 and under \(H_0\), as \(n \to \infty\), it holds that

\[
\mathcal{I}_n(\eta_0, r) = \left( \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \left( \frac{\partial \varepsilon_t(\eta_0, r)}{\partial \Psi_1} \right)^\top \left( \frac{\partial \varepsilon_t(\eta_0, r)}{\partial \Psi_1} \right) + \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \left( \frac{\partial \varepsilon_t(\eta_0, r)}{\partial \Psi_2} \right)^\top \left( \frac{\partial \varepsilon_t(\eta_0, r)}{\partial \Psi_2} \right) \right) + o_p(n). \tag{S2.1}
\]

**S3 Proofs**

**Proof of Lemma 1**

We prove the following general result: let \(\{a_t, t = 0, 1, 2, \ldots \}\) be a sequence of real numbers and define \(A_t\) as follows:

\[A_t = a_{t-1} + \sum_{j=1}^{q} \theta_j A_{t-j}, \text{ for each } t \geq 1 \text{ and } A_t = 0 \text{ otherwise.}\]

Then it holds that

\[A_t = \sum_{h=0}^{t-1} \alpha_h a_{t-1-h}, \tag{S3.2}\]
where

\[
\alpha_h = \begin{cases} 
0 & \text{if } h < 0 \\
1 & \text{if } h = 0 \\
\sum_{j=1}^q \theta_j \alpha_{h-j} & \text{if } h > 0.
\end{cases}
\]

We proceed by induction.

\[
A_{t+1} = a_t + \sum_{j=1}^q \theta_j A_{t+1-j} = a_t + \sum_{j=1}^q \theta_j \sum_{h=0}^{t-j} \alpha_h a_{t-j-h}
\]

\[
= a_t + \sum_{j=1}^q \theta_j \sum_{h=-j+1}^{t-j} \alpha_h a_{t-j-h} = a_t + \sum_{j=1}^q \theta_j \sum_{h=0}^{t-1} \alpha_{h-j+1} a_{t-1-h}
\]

\[
= a_t + \sum_{h=0}^{t-1} \left[ \sum_{j=1}^q \theta_j \alpha_{h-j+1} \right] a_{t-1-h} = a_t + \sum_{h=1}^t \left[ \sum_{j=1}^q \theta_j \alpha_{h-j} \right] a_{t-h}
\]

\[
= a_t + \sum_{h=1}^t \alpha_h a_{t-h} = \sum_{h=0}^t \alpha_h a_{t-h}.
\]

This results can be applied componentwise to \( \partial \varepsilon_t(\eta, r) / \partial \Psi \) so that the proof is complete.

**Proof of Lemma 2**

Under the null hypothesis, the true parameters are \( \eta_0 = (\eta_{0,1}, 0, \sigma_0^2) \) and we have:

\[
- \frac{\partial^2 \ell_n(\eta_0, r)}{\partial \Psi} = \frac{1}{\sigma_0^2} \sum_{t=1}^n \left( \frac{\partial \varepsilon_t(\eta_0, r)}{\partial \Psi} \right) \left( \frac{\partial \varepsilon_t(\eta_0, r)}{\partial \Psi} \right)^\top + \frac{1}{\sigma_0^2} \sum_{t=1}^n \varepsilon_t \frac{\partial^2 \varepsilon_t(\eta_0, r)}{\partial \Psi \partial \Psi^\top}.
\]
From Lemma 1, the first derivative $\partial (\eta, r) / \partial \Psi$ depends only upon $\theta$, hence it suffices to prove that

$$\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t \frac{\partial^2 \varepsilon_t(\eta_0, r)}{\partial \Psi \partial \theta} = o_p(1) \text{ uniformly on } r. \quad (S3.3)$$

For the sake of presentation, we detail below one case since identical arguments work for the remaining cases. Consider $\partial^2 \varepsilon_t(\eta_0, r)/(\partial \theta_1 \partial \phi_1)$. Lemma 1 implies that:

$$\frac{\partial^2 \varepsilon_t(\eta_0, r)}{\partial \theta_1 \partial \phi_1} = \sum_{h=0}^{t-1} \alpha_h \frac{\partial \varepsilon_{t-h}(\eta_0, r)}{\partial \phi_1}$$

$$= \sum_{h=0}^{t-1} \alpha_h \sum_{k=0}^{t-2-h} \alpha_k X_{t-2-h-k} := B_{t-1}.$$

Hence, (S3.3) reduces to show that

$$\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t B_{t-1} = o_p(1) \text{ uniformly on } r,$$

which is the case upon noting that $n^{-1} \sum_{t=1}^{n} \varepsilon_t B_{t-1}$ is a martingale difference sequence whose variance converges to zero as $n$ increases. Indeed:

$$E \left[ \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t B_{t-1} \right] = \frac{1}{n} \sum_{t=1}^{n} E \left[ B_{t-1} E \left[ \varepsilon_t | \mathcal{F}_{t-1} \right] \right] = 0.$$

On the other hand, Jensen’s inequality implies that

$$B_{t-1}^2 = \left( \sum_{h=0}^{t-1} \alpha_h \sum_{k=0}^{t-2-h} \alpha_k X_{t-2-h-k} \right)^2 \leq K \sum_{h=0}^{t-1} |\alpha_h| \sum_{k=0}^{t-2-h} |\alpha_k| |X_{t-2-h-k}|^2,$$
with $K$ being a constant that depends only on $\theta$. Therefore, it follows that:

$$E \left[ \frac{1}{n^2} \left( \sum_{t=1}^{n} \varepsilon_t B_{t-1} \right)^2 \right] = \frac{1}{n^2} \sum_{t=1}^{n} E \left[ \varepsilon_t^2 B_t^2 \right]$$

$$\leq \frac{K}{n^2} \sum_{t=1}^{n} E \left[ \varepsilon_t^2 \sum_{h=0}^{t-1} |\alpha_h| \sum_{k=0}^{t-2-h} |\alpha_k| |X_{t-2-h-k}|^2 \right]$$

$$= \frac{K}{n^2} \sum_{t=1}^{n} E \left[ \sum_{h=0}^{t-1} |\alpha_h| \sum_{k=0}^{t-2-h} |\alpha_k| |X_{t-2-h-k}|^2 E \left[ \varepsilon_t^2 |F_{t-1}| \right] \right]$$

$$= \frac{K}{n^2} \sum_{t=1}^{n} \sum_{h=0}^{t-1} |\alpha_h| \sum_{k=0}^{t-2-h} |\alpha_k| E \left[ |X_{t-2-h-k}|^2 \right]$$

which converges to zero because $\{X_t\}$ is strictly stationary and ergodic. By applying Markov's inequality the proof is complete.

**Proof of Proposition 1**

**Part (i)** The proof for sLM* can be found in Li and Li (2011) so that we do not repeat it here. In the following we prove the proposition for the sLM statistic, by showing that

$$\Lambda(\eta, r) = E \left[ \frac{\partial \varepsilon_t(\eta, r)}{\partial \Psi} \left( \frac{\partial \varepsilon_t(\eta, r)}{\partial \Psi} \right)^T \right]$$

is positive definite, where $\Psi = (\phi^T, \varphi^T) \in \mathbb{R}^{2(p+1)}$ Since the matrix is symmetric it is sufficient to show that if

$$E \left[ c^T \frac{\partial \varepsilon_t(\eta, r)}{\partial \Psi} \left( \frac{\partial \varepsilon_t(\eta, r)}{\partial \Psi} \right)^T c \right] = 0$$
then $c$ is the $2(p + 1)$-length zero vector. This holds if and only if

$$c^\top \frac{\partial \varepsilon_t(\eta, r)}{\partial \Psi} = 0 \text{ a.s..}$$

Hereafter, $c = (c_{10}, c_{11}, \ldots, c_{1p}, c_{20}, c_{21}, \ldots, c_{2p})^\top$. Routine algebra implies that

$$\begin{bmatrix} c_{10} + \sum_{i=1}^p c_{1i} X_{t-i} \end{bmatrix} I(X_{t-d} > r) = 0 \text{ a.s.} \quad (S3.4)$$

$$\begin{bmatrix} (c_{10} + c_{20}) + \sum_{i=1}^p (c_{1i} + c_{2i}) X_{t-k} \end{bmatrix} I(X_{t-d} \leq r) = 0 \text{ a.s.} \quad (S3.5)$$

We proceed by contradiction: we assume that $c$ is not the zero vector and prove that equalities (S3.4) and (S3.5) do not hold. Let

$$\mathcal{C} = \left\{ c_{10} + \sum_{i=1}^p c_{1i} X_{t-i} = 0 \right\}.$$  

Hence, under $H_0$, $\mathcal{C} = \{ \varepsilon_{t-1} = \Upsilon_{t-2} \}$ with

$$\Upsilon_{t-2} = \sum_{j=1}^q \theta_{0,j} \varepsilon_{t-1-j} - \phi_{0,0} - \sum_{i=1}^p \phi_{0,i} X_{t-1-i} - c_{10} - \sum_{i=2}^p c_{1i} X_{t-i}.$$  

$\Upsilon_{t-2}$ belongs to the sigma-algebra $\mathcal{F}_{t-2}$ generated by $\varepsilon_{t-2}, \varepsilon_{t-3}, \ldots$ and, therefore, since $\varepsilon_{t-1}$ is independent of $\Upsilon_{t-2}$ and admits a density function, the law of iterated expectations implies that $P(\mathcal{C}) = E[I(\varepsilon_{t-1} = \Upsilon_{t-2})] = E[E[I(\varepsilon_{t-1} = \Upsilon_{t-2})|\mathcal{F}_{t-2}]] = 0$. Since in $\mathcal{C}^c$ it holds that $[c_{10} + \sum_{i=1}^p c_{1i} X_{t-i}] \neq 0$ and $P(\mathcal{C}^c) = 1$, we have $P (\{ [c_{10} + \sum_{i=1}^p c_{1i} X_{t-i}] I(X_{t-d} > r) = 0 \}) = P (\{ X_{t-d} \leq r \}|\mathcal{C}^c)$, which is positive because the density of $\varepsilon_t$ is positive.
everywhere implying that the stationary distribution of \( \{X_t\} \) is positive everywhere. This contradicts equality (S3.4) and hence \( c_{11} \) must be zero. If \( \iota > 1 \), the event \( \mathcal{C} \) reduces to \( \{\varepsilon_{t-\iota} = \Upsilon_{t-\iota-1}\} \), with \( \Upsilon_{t-\iota-1} \in \mathcal{F}_{t-\iota-1} \) therefore the same argument shows that \( c_{10} = c_{12} = \ldots = c_{1p} = 0 \). Lastly, by using (S3.5) instead of (S3.4), it is easy to prove that \( c_{20} = c_{21} = \ldots = c_{2p} = 0 \) and this completes the proof.

**Part (ii)** In this case the proofs are the same for the two supLM statistics. For the sake of presentation and without loss of generality, we focus on the TARMA(1,1) case. Let \( \hat{\varepsilon}_t \) be the function \( \varepsilon_t(\hat{\eta}_1) \) evaluated at \( \hat{\eta}_1 \), i.e.:

\[
\hat{\varepsilon}_t = X_t - \hat{\phi}_0 - \hat{\phi}_1 X_{t-1} + \hat{\theta}_1 \hat{\varepsilon}_{t-1}.
\]

Routine algebra implies that

\[
\varepsilon_t - \hat{\varepsilon}_t = (\hat{\phi}_0 - \phi_{0,0}) \sum_{h=0}^{t-1} \hat{\theta}_1^h + (\hat{\phi}_1 - \phi_{0,1}) \sum_{h=0}^{t-1} \hat{\theta}_1^h X_{t-1-h} + (\theta_{0,1} - \hat{\theta}_1) \sum_{h=0}^{t-1} \hat{\theta}_1^h \varepsilon_{t-1-h} \tag{S3.6}
\]

Since \( \hat{\theta}_1 \) is consistent and the true value \( |\theta_{0,1}| < 1 \), there exists \( 0 < \gamma < 1 \) such that the event \( \mathcal{E}_n = \{|\theta_{0,1}| < \gamma, |\hat{\theta}_1| < \gamma\} \) holds with probability approaching 1 as \( n \to \infty \). Thus, with no loss of generality, \( \mathcal{E}_n \) is assumed to hold. Consequently, there exists a positive constant \( K \) such that for all
By using (S3.6), (S3.7) and Lemma 2, it holds that

\[ n^{-1} \hat{I}_n(r) = n^{-1} I_n(\eta_0, r) + o_p(1) \]

By proving that sup_{r \in [a, b]} \|M_n(r)\| = o_p(1), where

\[ M_n(r) = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(r)}{\partial \Psi} \left( \frac{\partial \varepsilon_t(r)}{\partial \Psi} \right)^T - \Lambda(r). \]

To begin with, we prove that, for any fixed \( r \), \( M_n(r) \to 0 \) in probability entrywise. We present the proof for a specific case, which can be easily adapted to the other entries. Consider

\[ n^{-1} \sigma_0^{-2} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(r)}{\partial \varphi_0} \frac{\partial \varepsilon_t(r)}{\partial \varphi_1}: \]

\[ - \frac{1}{\sigma_0^2} n \sum_{t=1}^{n} \sum_{h_1=0}^{t-1} \theta_{0,1}^{h_1} I_r(X_{t-1-h_1}) \sum_{h_2=0}^{t-1} \theta_{0,2}^{h_2} X_{t-1-h_2} I_r(X_{t-1-h_2}). \]

By using Cesaro’s means, it is not hard to prove that

\[ \left| \frac{1}{\sigma_0^2} n \sum_{t=1}^{n} \sum_{h_1=0}^{t-1} \theta_{0,1}^{h_1} I_r(X_{t-1-h_1}) \sum_{h_2=0}^{t-1} \theta_{0,2}^{h_2} X_{t-1-h_2} I_r(X_{t-1-h_2}) \right| \]

\[ - \left| \frac{1}{\sigma_0^2} n \sum_{t=1}^{n} \sum_{h_1=0}^{\infty} \theta_{0,1}^{h_1} I_r(X_{t-1-h_1}) \sum_{h_2=0}^{\infty} \theta_{0,2}^{h_2} X_{t-1-h_2} I_r(X_{t-1-h_2}) \right| \]

is a \( o_p(1) \) uniformly on \( r \). Hence, the ergodicity of \( \{X_t\} \) implies that \( M_n(r) \) converges to zero in probability for each \( r \). Now, fix \( a < b \) and consider a grid \( a = r_0 < r_1 < \ldots < r_m = b \) with equal mesh size, i.e. \( r_i - r_{i-1} \equiv c \), for
some $c > 0$. It holds that $\sup_{r \in [r_{i-1}, r_i]} \| M_n(r) - M_n(r_{i-1}) \| \leq C_n$, for all $i$.

Moreover, $E(C_n) \to 0$ as $c \to 0$. Because for any $r \in [a, b]$, there exists an $i$ such that $r_{i-1} \leq r \leq r_i$ and hence $M_n(r) = M_n(r) - M_n(r_{i-1}) + M_n(r_{i-1})$
and $\sup_{r \in [a, b]} \| M_n(r) \| \leq \max_{i=0, \ldots, m} M_n(r_i) + C_n$. The proof is complete since for fixed $m$, $\max_{i=0, \ldots, m} M_n(r_i) \to 0$ in probability and $E(C_n) \to 0$ as $c \to 0$ in probability.

**Part (iii)** For this part the proofs for the two statistics are similar and we show that for $sLM^\ast$. For the sake of presentation and without loss of generality, we focus on the TARMA(1, 1) case where $\Psi_1 = (\phi_0, \phi_1, \theta_1)$ and $\Psi_2 = (\varphi_0, \varphi_1, \vartheta_1)$. We use $\partial \ell_n / \partial \Psi$ to indicate the score function evaluated at the true parameters under the null hypothesis $H_0$. Within this proof, all the $o_p(1)$ terms hold uniformly on $r \in [a, b]$. We need to prove that:

$$
\sup_{r \in [a, b]} \left\| \frac{1}{\sqrt{n}} \frac{\partial \hat{\ell}_n(r)}{\partial \Psi_2} - (\nabla_{n,2}(r) - \Lambda_{21}(r) \Lambda_{11}^{-1} \nabla_{n,1}) \right\| = o_p(1).
$$

Since $\sqrt{n}(\Psi_1 - \Psi_{0,1}) = \Lambda_{11}^{-1} n^{-1/2} \partial \ell_n / \partial \Psi_1 + o_p(1)$ and $\Lambda_{21}(r) = O_p(1)$ uniformly in $r \in [a, b]$, then it is sufficient to prove that

$$
\sup_{r \in [a, b]} \left\| \frac{1}{\sqrt{n}} \frac{\partial \hat{\ell}_n(r)}{\partial \Psi_2} - \frac{1}{\sqrt{n}} \frac{\partial \ell_n(r)}{\partial \Psi_2} + \Lambda_{21}(r) \sqrt{n}(\Psi_1 - \Psi_{0,1}) \right\| = o_p(1). \quad (S3.8)
$$
We prove (S3.8) componentwise; below we detail the argument for the first component. Thence, we show that

\[
\sup_{r \in [a, b]} \left\| \frac{1}{\sqrt{n}} \frac{\partial \hat{\ell}_n(r)}{\partial \varphi_0} - \frac{1}{\sqrt{n}} \frac{\partial \ell_n(r)}{\partial \varphi_0} + \Lambda_{21}(r)_{1,1} \sqrt{n}(\hat{\phi}_0 - \phi_{0,0}) \right\| + \Lambda_{21}(r)_{1,2} \sqrt{n}(\hat{\phi}_1 - \phi_{0,1}) + \Lambda_{21}(r)_{1,3} \sqrt{n}(\hat{\theta}_1 - \theta_{0,1}) = o_p(1),
\]

with \( \Lambda_{21}(r)_{i,j} \) being the \((i, j)\)-th component of matrix \( \Lambda_{21}(r) \). Let \( \hat{\varepsilon}_t \) denote \( \varepsilon_t(\eta_1) \) evaluated at \( \hat{\eta}_1 \), i.e.:

\[
\hat{\varepsilon}_t = X_t - \hat{\phi}_0 - \hat{\phi}_1 X_{t-1} + \hat{\theta}_1 \hat{\varepsilon}_{t-1},
\]

and define

\[
\frac{\partial \varepsilon_t(r)}{\partial \Psi} = \frac{\partial \varepsilon_t(\eta, r)}{\partial \Psi} \bigg|_{\eta = \eta_0} \quad \frac{\partial \hat{\varepsilon}_t(r)}{\partial \Psi} = \frac{\partial \varepsilon_t(\eta, r)}{\partial \Psi} \bigg|_{\eta = \hat{\eta}},
\]

with \( \frac{\partial \varepsilon_t(\eta, r)}{\partial \Psi} \) being defined in (2.8).

Routine algebra implies that:

\[
\frac{\partial \varepsilon_t(r)}{\partial \varphi_0} - \frac{\partial \hat{\varepsilon}_t}{\partial \varphi_0} = (\theta_{0,1} - \hat{\theta}_1) \sum_{h=0}^{t-1} \hat{\theta}_1^h \frac{\partial \varepsilon_{t-h}(r)}{\partial \varphi_0}.
\]

Since \( 1/\hat{\sigma}^2 - 1/\sigma_0^2 = O_p(n^{-1/2}) \), by omitting a negligible additive term, it follows that

\[
\frac{1}{\sqrt{n}} \frac{\partial \hat{\ell}_n(r)}{\partial \varphi_0} = -\frac{1}{\sqrt{n} \hat{\sigma}^2} \sum_{t=1}^{n} \hat{\varepsilon}_t \frac{\partial \hat{\varepsilon}_t(r)}{\partial \varphi_0} \approx -\frac{1}{\sqrt{n} \sigma_0^2} \sum_{t=1}^{n} \hat{\varepsilon}_t \frac{\partial \hat{\varepsilon}_t(r)}{\partial \varphi_0} = -\frac{1}{\sqrt{n} \hat{\sigma}^2} \sum_{t=1}^{n} \varepsilon_t \frac{\partial \varepsilon_t(r)}{\partial \varphi_0} + \frac{1}{\sqrt{n} \sigma_0^2} \sum_{t=1}^{n} \hat{\varepsilon}_t \frac{\partial \hat{\varepsilon}_t(r)}{\partial \varphi_0} \cdot
\]
Hence Eq (S3.12) follows upon proving that:

\[
\sup_{r \in [a,b]} \left| \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \varepsilon_t \frac{\partial \xi_t(r)}{\partial \varphi_0} - \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \hat{\varepsilon}_t \frac{\partial \hat{\xi}_t(r)}{\partial \varphi_0} - \Lambda_{21}(r)_{1,1} \sqrt{n}(\hat{\phi}_0 - \phi_0,0) 
- \Lambda_{21}(r)_{1,2} \sqrt{n}(\hat{\phi}_1 - \phi_{0,1}) - \Lambda_{21}(r)_{1,3} \sqrt{n}(\hat{\theta}_1 - \theta_{0,1}) \right| = o_p(1). \tag{S3.12}
\]

By using (S3.6) and (S3.11), we have that:

\[
\frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \varepsilon_t \frac{\partial \xi_t(r)}{\partial \varphi_0} - \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \hat{\varepsilon}_t \frac{\partial \hat{\xi}_t(r)}{\partial \varphi_0}
= \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \hat{\varepsilon}_t \left( \frac{\partial \xi_t(r)}{\partial \varphi_0} - \frac{\partial \hat{\xi}_t(r)}{\partial \varphi_0} \right) + \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^{n} (\varepsilon_t - \hat{\varepsilon}_t) \frac{\partial \xi_t(r)}{\partial \varphi_0}
= \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \hat{\varepsilon}_t \left\{ (\theta_{0,1} - \hat{\theta}_1) \sum_{h=0}^{t-1} \hat{\theta}_1^h \frac{\partial \varepsilon_{t-1-h}(r)}{\partial \varphi_0} \right\}
+ \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \left\{ (\hat{\phi}_0 - \phi_{0,0}) \sum_{h=0}^{t-1} \hat{\theta}_1^h + (\hat{\phi}_1 - \phi_{0,1}) \sum_{h=0}^{t-1} \hat{\theta}_1^h X_{t-1-h} + (\theta_{0,1} - \hat{\theta}_1) \sum_{h=0}^{t-1} \hat{\theta}_1^h \varepsilon_{t-1-h} \right\} \frac{\partial \xi_t(r)}{\partial \varphi_0}.
\]

Hence Eq (S3.12) follows upon proving that:

\[
\sup_{r \in [a,b]} \left| \frac{1}{\sqrt{n}} (\hat{\phi}_0 - \phi_{0,0}) \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \frac{\partial \xi_t(r)}{\partial \varphi_0} \sum_{h=0}^{t-1} \hat{\theta}_1^h - \Lambda_{21}(r)_{1,1} \sqrt{n}(\hat{\phi}_0 - \phi_{0,0}) \right| = o_p(1), \tag{S3.13}
\]

\[
\sup_{r \in [a,b]} \left| \frac{1}{\sqrt{n}} (\hat{\phi}_1 - \phi_{0,1}) \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \frac{\partial \xi_t(r)}{\partial \varphi_0} \sum_{h=0}^{t-1} \hat{\theta}_1^h X_{t-1-h} - \Lambda_{21}(r)_{1,2} \sqrt{n}(\hat{\phi}_1 - \phi_{0,1}) \right| = o_p(1), \tag{S3.14}
\]

\[
\sup_{r \in [a,b]} \left| \frac{1}{\sqrt{n}} (\theta_{0,1} - \hat{\theta}_1) \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \frac{\partial \xi_t(r)}{\partial \varphi_0} \sum_{h=0}^{t-1} \hat{\theta}_1^h \varepsilon_{t-1-h} - \Lambda_{21}(r)_{1,3} \sqrt{n}(\hat{\theta}_1 - \theta_{0,1}) \right| = o_p(1) \tag{S3.15}
\]

\[
\sup_{r \in [a,b]} \left| \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \hat{\varepsilon}_t \left\{ (\theta_{0,1} - \hat{\theta}_1) \sum_{h=0}^{t-1} \hat{\theta}_1^h \frac{\partial \varepsilon_{t-1-h}(r)}{\partial \varphi_0} \right\} \right| = o_p(1). \tag{S3.16}
\]
In the following, we prove Eq (S3.13). Since it holds that

\[
\frac{1}{\sqrt{n}} (\hat{\phi}_0 - \phi_{0,0}) \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(r)}{\partial \phi_0} \frac{1}{\sqrt{n}} (\hat{\phi}_0 - \phi_{0,0}) \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(r)}{\partial \phi_0} \frac{\partial \varepsilon_t(r)}{\partial \phi_0} \sum_{h=0}^{t-1} \theta_1^h = \sqrt{n} (\hat{\phi}_0 - \phi_{0,0}) \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(r)}{\partial \phi_0} \frac{\partial \varepsilon_t(r)}{\partial \phi_0} + \frac{1}{\sqrt{n}} (\hat{\phi}_0 - \phi_{0,0}) \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(r)}{\partial \phi_0} \sum_{h=0}^{t-1} \left\{ \theta_1^h - \theta_{0,1}^h \right\}
\]

and, by point (ii)

\[
\sup_{r \in [a,b]} \left| \sqrt{n} (\hat{\phi}_0 - \phi_{0,0}) \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(r)}{\partial \phi_0} \frac{\partial \varepsilon_t(r)}{\partial \phi_0} - \sqrt{n} (\hat{\phi}_0 - \phi_{0,0}) \Lambda_{21}(r)_{1,1} \right| = o_p(1),
\]

it remains to prove that

\[
\sup_{r \in [a,b]} \left| \frac{1}{\sqrt{n}} (\hat{\phi}_0 - \phi_{0,0}) \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(r)}{\partial \phi_0} \sum_{h=0}^{t-1} \left\{ \theta_1^h - \theta_{0,1}^h \right\} \right| = o_p(1).
\]

By using (S3.7) it follows that:

\[
\sup_{r \in [a,b]} \left| \frac{1}{\sqrt{n}} (\hat{\phi}_0 - \phi_{0,0}) \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \frac{\partial \varepsilon_t(r)}{\partial \phi_0} \sum_{h=0}^{t-1} \left\{ \theta_1^h - \theta_{0,1}^h \right\} \right| \leq K |\hat{\phi}_0 - \phi_{0,0}| |\hat{\theta}_1 - \theta_{0,1}| \frac{n}{\sqrt{n} (1 - \gamma)} = o_p(1).
\]

By using the same argument we can handle Eq (S3.14) and Eq. (S3.15).

Hence it remains to prove Eq. (S3.16). In the following we use $K$ to refer to a generic constant that can change across lines. Since $\theta_{0,1} - \hat{\theta}_1 = O(n^{-1/2})$, \[\]
we have that

\[
\frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} \sum_{t=1}^{n} \hat{\varepsilon}_t \left\{ (\theta_{0,1} - \hat{\theta}_1) \sum_{h=0}^{t-1} \hat{\theta}_1^h \frac{\partial \varepsilon_{t-1-h}(r)}{\partial \varphi_0} \right\} = K \sum_{t=1}^{n} \varepsilon_t \sum_{h=0}^{t-1} \left\{ (\theta_{0,1} - \hat{\theta}_1) \sum_{h=0}^{t-1} \hat{\theta}_1^h \frac{\partial \varepsilon_{t-1-h}(r)}{\partial \varphi_0} \right\} + K \sum_{t=1}^{n} (\hat{\varepsilon}_t - \varepsilon_t) \sum_{h=0}^{t-1} \left\{ (\theta_{0,1} - \hat{\theta}_1) \sum_{h=0}^{t-1} \hat{\theta}_1^h \frac{\partial \varepsilon_{t-1-h}(r)}{\partial \varphi_0} \right\},
\]

which is an \(o_p(1)\) by using (S3.6), (S3.7) and the fact that both \(\sum_{h=0}^{\infty} |\hat{\theta}_1|^h\) and \(\partial \varepsilon_{t-1-j}(r)/\partial \varphi_0\) are uniformly bounded by \((1 - \gamma)^{-1}\) in magnitude. Hence, Eq. (S3.14) is verified and the whole proof is complete.

**Proof of Theorem 1**

Proposition 1 implies that:

\[
\sup_{r \in [a,b]} \left\| T_n(r) - \nabla_n(r)^T \left( -\Lambda_{21}(r)\Lambda_{11}^{-1}, \mathbb{I} \right)^T \left( \Lambda_{22}(r) - \Lambda_{21}(r)\Lambda_{11}^{-1}\Lambda_{12}(r) \right)^{-1} \left( -\Lambda_{21}(r)\Lambda_{11}^{-1}, \mathbb{I} \right) \nabla_n(r) \right\| = o_p(1),
\]

with \(\mathbb{I}\) being the \((p + 1) \times (p + 1)\) identity matrix for sLM and the \((p + q + 1) \times (p + q + 1)\) identity matrix for sLM*. The result follows by applying the same proof of Theorem 2.1 in **Ling and Tong (2005)** and Theorem 1 of **Li and Li (2011)**. The proof is the same for both supLM statistics.
Proof of Proposition 2

Under assumptions A1–A3, the proof for the two supLM statistics is similar to that in Li and Li (2011) and is omitted. Here we show the proof for the sLM statistic without assumption A3. From (2.4) and (4.13) and Lemma 1 it follows that:

under $H_0$: $\varepsilon_t = \sum_{h=0}^{t-1} \alpha_h \Delta \phi_0(X_{t-1-h}) - \phi_{0,0} \sum_{h=0}^{t-1} \alpha_h$

under $H_{1,n}$: $\varepsilon_t = \sum_{h=0}^{t-1} \alpha_h \Delta \phi_0(X_{t-1-h}) - \phi_{0,0} \sum_{h=0}^{t-1} \alpha_h - \sum_{h=0}^{t-1} \alpha_h \Delta h_n(X_{t-1-h})$,

where $\Delta \phi_0(X_t) = X_t - \sum_{i=1}^p \phi_{0,i} X_{t-i}$ and

$\Delta h_n(X_t) = \frac{1}{\sqrt{n}} \left( h_{10} + \sum_{i=1}^p h_{1i} X_{t-i} \right) I_{r_0}(X_{t-d})$.

Under the null hypothesis, we have that

$$\log \frac{dP_{1,n}}{dP_{0,n}} = h^T \nabla n_2(r_0) - \frac{1}{2} h^T \left[ \frac{1}{n \sigma_0^2} \sum_{t=1}^n \frac{\partial \varepsilon_t(\eta_0, r_0)}{\partial \Psi_2} \left( \frac{\partial \varepsilon_t(\eta_0, r_0)}{\partial \Psi_2} \right)^T \right] h.$$

Since the matrix inside square brackets converges to $\Lambda_{22}(r_0)$ and the martingale central limit theorem implies that $\nabla n_2(r_0)$ converges to $\nabla_2(r_0)$, the proof of point (i) is complete. To prove point (ii) note that such limiting random variable is Gaussian distributed with mean $-(h^T \Lambda_{22}(r_0) h)/2$ and variance $h^T \Lambda_{22}(r_0) h$. The contiguity readily follows by applying Le Cam’s first lemma (see example 6.5 pag 89 of van der Vaart (1998)).
Proof of Theorem 2

Under assumptions A1–A3, the proof for the two supLM statistics is similar to that in Li and Li (2011) and is omitted. Here we show the proof for the sLM statistic without assumption A3. In order to prove point (i) note that the tightness of \( \nabla_n, 2(r) - \Lambda_{21}(r) \Lambda_{11} \nabla_{n,1} \) under \( H_{1,n} \) follows from its tightness under \( H_0 \) due to the contiguity. Hence, it suffices to show that, under \( H_{1,n} \), \( \nabla_{n,2}(r) - \Lambda_{21}(r) \Lambda_{11}^{-1} \nabla_{n,1} \) converges to \( (\xi_r + \gamma_r) \) when \( r \) is fixed. By virtue of the third Le Cam’s Lemma, it suffices to prove that

\[
\left( \nabla_{n,2}(r) - \Lambda_{21}(r) \Lambda_{11}^{-1} \nabla_{n,1}, \log \frac{dP_{1,n}}{dP_{0,n}} \right)
\]

converges to a multivariate Gaussian random vector whose variance-covariance matrix is

\[
\begin{pmatrix}
\Lambda_{22}(r) - \Lambda_{21}(r) \Lambda_{11}^{-1} \Lambda_{12}(r) & \gamma_r \\
\gamma_r^\top & h^\top \Lambda_{22}(r_0) h
\end{pmatrix}
\]

which is the case since:

\[
\text{Cov} \left( \nabla_{n,2}(r) - \Lambda_{21}(r) \Lambda_{11}^{-1} \nabla_{n,1}, \log \frac{dP_{1,n}}{dP_{0,n}} \right) = \left\{ \Lambda_{22}(\min\{r, r_0\}) - \Lambda_{21}(r) \Lambda_{11}^{-1} \Lambda_{12}(r_0) \right\} h.
\]

The proof of point (ii) is omitted since it follows directly from (i).
Proof of Theorem 3

The proof is the same for both statistics. It suffices to show that for each $q$

$$\Pr \left( (\xi_{r_0} + \gamma_{r_0})^\top \left( \Lambda_{22}(r_0) - \Lambda_{21}(r_0)\Lambda_{11}^{-1}\Lambda_{12}(r_0) \right)^{-1} (\xi_{r_0} + \gamma_{r_0}) > q \right) \overset{p}{\longrightarrow} 1.$$

(S3.17)

For simplicity, we set $\Sigma_{r_0} = (\Lambda_{22}(r_0) - \Lambda_{21}(r_0)\Lambda_{11}^{-1}\Lambda_{12}(r_0))$. Routine algebra implies that:

$$(\xi_{r_0} + \gamma_{r_0})^\top \left( \Lambda_{22}(r_0) - \Lambda_{21}(r_0)\Lambda_{11}^{-1}\Lambda_{12}(r_0) \right)^{-1} (\xi_{r_0} + \gamma_{r_0})$$

$$= \xi_{r_0}^\top \Sigma_{r_0}^{-1} \xi_{r_0} + 2h^\top \xi_{r_0} + h^\top \Sigma_{r_0} h,$$

For any fixed $r_0$, $\xi_{r_0}$ is a Gaussian random vector with zero mean and variance-covariance matrix $\Sigma_{r_0}$ and hence the Markov inequality implies that

$$\frac{h^\top \xi_{r_0}}{h^\top \Sigma_{r_0} h} = o_p(1), \text{ as } ||h|| \text{ increases.}$$

Therefore, we have

$$\xi_{r_0}^\top \Sigma_{r_0}^{-1} \xi_{r_0} + 2h^\top \xi_{r_0} + h^\top \Sigma_{r_0} h = \xi_{r_0}^\top \Sigma_{r_0}^{-1} \xi_{r_0} + 2(h^\top \Sigma_{r_0} h) \left( \frac{h^\top \xi_{r_0}}{h^\top \Sigma_{r_0} h} + \frac{1}{2} \right)$$

$$= O_p(1) + C_h o_p(1) + 1/2 \overset{p}{\longrightarrow} +\infty,$$

where $C_h = 2(h^\top \Sigma_{r_0} h) \to \infty$ as $||h|| \to +\infty$ and hence Eq. (S3.17) follows and the proof is complete.
S4 Supplementary Monte Carlo results

S4.1 Power when only the MA parameters change across regimes

The case \( \text{ii}) \) where only the moving-average parameter changes across regimes is studied by simulating from the following model:

\[
X_t = \phi_0 + \phi_1 X_{t-1} - \theta_1 \varepsilon_{t-1} + (\vartheta_1 \varepsilon_{t-1}) I(X_{t-1} \leq 0) + \varepsilon_t. \tag{S4.18}
\]

where \( \theta_1 = -0.6 \), whereas \( \phi_0, \phi_1 \) and \( \vartheta_1 \) are as in Table 4, that shows the size-corrected power (percent). In this case the behaviour of the tests depends upon different factors. When the departure from the null hypothesis is mild, the qLR test has an advantage over supLM tests. The situation is reversed when \( \vartheta_1 \) is large (e.g. \( \vartheta_1 = 1.2 \)): in such case both supLM tests are more powerful than the qLR test. When \( n = 500 \) and \( \phi_0 = 0.6, \phi_1 = 0.7 \) the sLM* test is always more powerful than the qLR test, whereas in the remaining cases there is not a clear winner and the results are comparable.

S4.2 Power when both the AR and MA parameters change across regimes

As concerns case \( \text{iii}) \) we simulate from the following model:

\[
X_t = \phi_0 + \phi_1 X_{t-1} - \theta_1 \varepsilon_{t-1} + (\varphi_0 + \varphi_1 X_{t-1} + \vartheta_1 \varepsilon_{t-1}) I(X_{t-1} \leq 0) + \varepsilon_t. \tag{S4.19}
\]
Table 4: Size-corrected power for the TARMA(1,1) model of Eq. (S4.18), case \( ii \): only the moving-average parameters change across regimes. Sample size \( n = 100, 200, 500 \), \( \alpha = 5\% \).

| \( \phi_0 \) | \( \phi_1 \) | \( \theta_1 \) | \( n = 100 \) | \( n = 200 \) | \( n = 500 \) |
|-----|-----|-----|-----|-----|-----|
| \( \phi_0 \) | \( \phi_1 \) | \( \theta_1 \) | sLM | sLM* | qLR | sLM | sLM* | qLR | sLM | sLM* | qLR |
| 0.2 | 0.1 | 0.4 | 9.1 | 12.6 | 14.9 | 23.1 | 29.1 | 43.8 | 62.6 | 77.4 | 83.1 |
| 0.2 | 0.1 | 0.6 | 23.8 | 24.8 | 29.9 | 50.2 | 55.6 | 69.6 | 95.1 | 97.8 | 99.2 |
| 0.2 | 0.1 | 0.8 | 39.2 | 38.9 | 46.8 | 81.8 | 81.8 | 90.9 | 100.0 | 100.0 | 100.0 |
| 0.2 | 0.1 | 1.2 | 71.2 | 67.7 | 75.2 | 96.4 | 96.6 | 94.5 | 99.9 | 99.9 | 99.7 |
| 0.4 | 0.4 | 0.4 | 8.3 | 10.3 | 15.5 | 23.3 | 34.2 | 47.0 | 62.6 | 85.1 | 91.4 |
| 0.4 | 0.4 | 0.6 | 17.2 | 20.4 | 31.4 | 52.3 | 62.9 | 73.6 | 95.4 | 98.8 | 99.6 |
| 0.4 | 0.4 | 0.8 | 38.3 | 39.1 | 47.6 | 83.8 | 87.3 | 91.6 | 100.0 | 100.0 | 99.9 |
| 0.4 | 0.4 | 1.2 | 77.5 | 72.4 | 75.5 | 99.4 | 99.3 | 98.7 | 100.0 | 100.0 | 100.0 |
| 0.6 | 0.7 | 0.4 | 8.1 | 8.2 | 10.5 | 11.1 | 14.9 | 21.6 | 42.6 | 57.2 | 56.8 |
| 0.6 | 0.7 | 0.6 | 12.2 | 14.8 | 17.8 | 31.2 | 36.4 | 38.8 | 82.0 | 90.5 | 85.3 |
| 0.6 | 0.7 | 0.8 | 23.2 | 25.5 | 27.2 | 58.3 | 61.8 | 55.9 | 98.5 | 98.8 | 93.4 |
| 0.6 | 0.7 | 1.2 | 48.6 | 48.4 | 38.4 | 90.6 | 89.5 | 76.5 | 100.0 | 100.0 | 99.1 |

where \( \phi_0 = -0.5 \), \( \phi_1 = -0.5 \), \( \theta_1 = 0.5 \) and \( \phi_0 \), \( \varphi_1 \) and \( \theta_1 \) are as in Table 5, first three columns. When \( n = 100 \) the sLM* and the qLR tests are comparable and more powerful than the sLM test. However when \( n = 200 \) the sLM* test is more powerful than the qLR and sLM tests, whose power is comparable. When \( n = 500 \) the sLM* test is always the most powerful of the three. Also, on average the sLM test has 6% less power than the qLR test. Note that, in principle, this setting is more favourable to the sLM* and qLR tests since, even if the sequence of departures from \( H_0 \) is monotonically increasing with respect to all the components, the rate is faster along the moving-average component \( \theta_1 \) and slower on the autoregressive part \( \phi_0 \) and \( \varphi_1 \) (see the first three columns of Table 5). When all the components
Table 5: Size-corrected power for the TARMA(1, 1) model of Eq. (S4.19), case iii): both
the autoregressive and the moving-average parameters change across regimes. Sample
size \( n = 100, 200, 500 \), \( \alpha = 5\% \).

| \( \varphi_0 \) | \( \varphi_1 \) | \( \theta_1 \) | \( n = 100 \) | \( n = 200 \) | \( n = 500 \) |
|---|---|---|---|---|---|
| \( \phi \) | \( \phi \) | \( \phi \) | sLM | sLM* | qLR | sLM | sLM* | qLR | sLM | sLM* | qLR |
| 0.02 | 0.02 | -0.10 | 6.2 | 6.0 | 6.5 | 8.8 | 7.9 | 5.0 | 7.8 | 10.9 | 8.9 |
| 0.04 | 0.04 | -0.20 | 7.0 | 7.1 | 9.5 | 10.1 | 13.0 | 8.8 | 18.5 | 32.6 | 21.6 |
| 0.06 | 0.06 | -0.30 | 8.8 | 11.3 | 11.1 | 17.9 | 23.5 | 15.6 | 37.3 | 66.8 | 50.7 |
| 0.08 | 0.08 | -0.40 | 14.2 | 16.8 | 17.0 | 28.9 | 40.9 | 29.9 | 63.5 | 90.9 | 78.9 |
| 0.10 | 0.10 | -0.50 | 18.3 | 23.9 | 23.8 | 43.1 | 57.6 | 44.5 | 85.0 | 97.1 | 92.3 |
| 0.12 | 0.12 | -0.60 | 25.6 | 32.6 | 30.1 | 54.9 | 72.3 | 60.3 | 95.7 | 99.6 | 99.3 |
| 0.14 | 0.14 | -0.70 | 30.9 | 37.3 | 40.4 | 76.0 | 85.3 | 74.4 | 99.4 | 100.0 | 99.8 |

 Kurds are distant from the null hypothesis, then, the power of the tests are either
comparable or the sLM test is even more powerful (results not shown here).

**S4.3 Power: TARMA(2, 1) model**

We simulate from the following TARMA(2, 1) model where only the autore-
gressive parameters change across regimes:

\[
X_t = -0.35X_{t-1} - 0.45X_{t-2} - \theta_1 \varepsilon_{t-1} + (\varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2}) I(X_{t-1} \leq 0) + \varepsilon_t.
\]

(S4.20)

where \( \varphi_0 = \varphi_1 = \varphi_2 = \varphi \) and \( \theta_1 \) are as in Table 6, that presents the size-
corrected power of the tests. The results are consistent with the TARMA(1, 1)
case i) of the main article, where only the autoregressive parameters change.
Also in this case the sLM test is almost uniformly superior to the sLM* test,
that, in turn is more powerful than the qLR in 60% of the cases.
Table 6: Size-corrected power for the TARMA(2, 1) model of Eq. (S4.20). Sample size $n = 100, 200, 500$, $\alpha = 5\%$.

| $\varphi$ | $\theta_1$ | $n = 100$ | $n = 200$ | $n = 500$ |
|----------|------------|-----------|-----------|-----------|
| 0.2      | -0.5       | 16.4      | 13.3      | 14.0      |
| 0.6      | -0.5       | 84.0      | 78.9      | 74.6      |
| 0.8      | -0.5       | 94.5      | 93.7      | 93.5      |
| 0.2      | 0.0        | 10.7      | 8.3       | 10.0      |
| 0.6      | 0.0        | 63.1      | 60.6      | 60.3      |
| 0.8      | 0.0        | 88.5      | 87.7      | 84.1      |
| 0.2      | 0.5        | 15.4      | 13.7      | 7.5       |
| 0.6      | 0.5        | 81.6      | 76.0      | 44.9      |
| 0.8      | 0.5        | 94.1      | 92.8      | 79.5      |

S4.4 The impact of model selection in presence of mis-specification

In Table 7 we present the list of data generating processes used. The first seven processes are linear models that are not encompassed within the ARMA(1,1) model. The lower panel of the table lists nine non-linear processes that cannot be represented in terms of a TARMA(1, 1) model.

In Table 8 we show the results of the Hannan-Rissanen model selection in presence of modelling misspecification. As for the size, the bias produced is small and stays within 3%. The case of power is more interesting. While there is little or no power loss, there might be a substantial gain for the sLM test, see e.g. the TAR3 instance, where the power for $n = 100$ passes from 34.7% to 96.1%, most likely due to the HR criterion picking the dependence upon lag 3.
Table 7: Data generating processes used to investigate the size and power of the tests under model mis-specification. Unless otherwise stated \{\varepsilon_t\} follows a standard Gaussian white noise.

|   |   |   |
|---|---|---|
| 01. AR5 | \(X_t = -0.6X_{t-1} - 0.4X_{t-2} - 0.3X_{t-3} - 0.4X_{t-4} - 0.5X_{t-5} + \varepsilon_t\) |
| 02. AR2.1 | \(X_t = 0.75X_{t-1} - 0.125X_{t-2} + \varepsilon_t\) |
| 03. AR2.2 | \(X_t = 1.35X_{t-1} - 0.55X_{t-2} + \varepsilon_t\) |
| 04. ARMA21.1 | \(X_t = 0.75X_{t-1} - 0.125X_{t-2} - 0.7\varepsilon_{t-1} + \varepsilon_t\) |
| 05. ARMA21.2 | \(X_t = 0.75X_{t-1} - 0.125X_{t-2} + 0.7\varepsilon_{t-1} + \varepsilon_t\) |
| 06. ARMA22 | \(X_t = 0.75X_{t-1} - 0.125X_{t-2} + 0.7\varepsilon_{t-1} - 0.4\varepsilon_{t-2} + \varepsilon_t\) |
| 07. MA2 | \(X_t = 0.7\varepsilon_{t-1} - 0.125\varepsilon_{t-2} + \varepsilon_t\) |
| 08. TAR3 | \(X_t = \begin{cases} 0.3X_{t-1} - 0.7X_{t-2} + 0.6X_{t-3} + \varepsilon_t, & \text{if } X_{t-1} \leq 0 \\ -0.3X_{t-1} + 0.7X_{t-2} - 0.6X_{t-3} + \varepsilon_t, & \text{if } X_{t-1} > 0 \\ 0.3 + 0.5X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} \leq -1 \end{cases}\) |
| 09. 3TAR1 | \(X_t = \begin{cases} 0.3 + X_{t-1} + \varepsilon_t, & \text{if } -1 < X_{t-1} \leq 1 \\ 0.3 + 0.5X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} > 1 \end{cases}\) |
| 10. NLMA.1 | \(X_t = -0.8\varepsilon_t^2_{t-1} + \varepsilon_t\) |
| 11. NLMA.2 | \(X_t = 0.8\varepsilon_t^2_{t-1} + \varepsilon_t\) |
| 12. BIL.1 | \(X_t = 0.5 - 0.4X_{t-1} + 0.4\varepsilon_{t-1}X_{t-1} + \varepsilon_t\) |
| 13. BIL.2 | \(X_t = 0.7\varepsilon_{t-1}X_{t-2} + \varepsilon_t\) |
| 14. EXPAR.1 | \(X_t = 0.3 + 10 \exp(-X^2_{t-1})X_{t-1} + \varepsilon_t\) |
| 15. EXPAR.2 | \(X_t = 0.3 + 100 \exp(-X^2_{t-1})X_{t-1} + \varepsilon_t\) |
| 16. NLAR | \(X_t = 4X_t(1 - X_t)\) |
Table 8: Rejection percentages under model misspecification for the processes of Table 6 of the main article for the supLM tests (sLM, sLM⋆). The subscript “HR” indicates that the order of the ARMA model has been selected through the Hannan-Rissanen procedure, otherwise, an ARMA\((1, 1)\) model is tested. 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.

|       | sLM | sLM\textsubscript{HR} | sLM | sLM\textsubscript{HR} | sLM | sLM\textsubscript{HR} |
|-------|-----|------------------------|-----|------------------------|-----|------------------------|
|       | \(n = 100\) | \(n = 200\) | \(n = 500\) | \(n = 100\) | \(n = 200\) | \(n = 500\) |
|       |     |                        |     |                        |     |                        |
| Linear|     |                        |     |                        |     |                        |
| AR5   | 7.9 | 10.2                   | 5.5 | 5.3                    | 4.6 | 5.7                    |
| AR2.1 | 6.7 | 7.7                    | 4.7 | 4.9                    | 5.3 | 5.0                    |
| AR2.2 | 8.7 | 9.5                    | 4.1 | 5.7                    | 3.4 | 4.9                    |
| ARMA21.1 | 5.9 | 7.2                    | 5.5 | 6.3                    | 5.5 | 6.2                    |
| ARMA21.2 | 5.8 | 6.4                    | 4.5 | 5.0                    | 4.0 | 5.1                    |
| ARMA22 | 5.5 | 6.2                    | 3.7 | 3.7                    | 3.5 | 3.7                    |
| MA2   | 3.5 | 6.9                    | 2.3 | 3.1                    | 4.4 | 5.3                    |
| TAR3  | 34.7| 96.1                   | 61.8| 99.9                   | 96.7| 100.0                  |
| 3TAR1 | 19.2| 19.5                   | 36.1| 35.9                   | 78.5| 78.1                   |
| NLMA.1| 85.1| 84.7                   | 97.3| 97.4                   | 99.8| 99.8                   |
| NLMA.2| 86.4| 84.5                   | 98.3| 98.6                   | 100.0| 100.0                  |
| BIL.1 | 12.0| 21.6                   | 13.5| 29.0                   | 14.7| 28.9                   |
| BIL.2 | 84.0| 82.0                   | 98.7| 97.9                   | 100.0| 99.8                   |
| EXPAR.1| 100.0| 100.0                 | 100.0| 99.9                   | 100.0| 100.0                  |
| EXPAR.2| 95.8 | 97.7                 | 99.6| 99.9                   | 100.0| 100.0                  |
| NLAR  | 100.0| 100.0                 | 100.0| 100.0                 | 100.0| 100.0                  |

|       | sLM⋆ | sLM\textsubscript{HR}⋆ | sLM⋆ | sLM\textsubscript{HR}⋆ | sLM⋆ | sLM\textsubscript{HR}⋆ |
|-------|------|------------------------|------|------------------------|------|------------------------|
|       | \(n = 100\) | \(n = 200\) | \(n = 500\) | \(n = 100\) | \(n = 200\) | \(n = 500\) |
|       |     |                        |     |                        |     |                        |
| Linear|     |                        |     |                        |     |                        |
| AR5   | 15.2| 17.5                   | 9.8 | 10.5                   | 6.9 | 7.2                    |
| AR2.1 | 7.1 | 8.3                    | 5.5 | 5.5                    | 5.5 | 5.8                    |
| AR2.2 | 10.1| 9.9                    | 5.8 | 5.5                    | 5.0 | 5.7                    |
| ARMA21.1 | 6.8 | 8.7                    | 5.2 | 6.7                    | 4.8 | 6.2                    |
| ARMA21.2 | 8.9 | 9.8                    | 5.7 | 6.0                    | 4.9 | 5.6                    |
| ARMA22 | 13.6| 16.8                   | 12.7| 13.0                   | 7.5 | 8.4                    |
| MA2   | 12.1| 14.9                   | 5.8 | 7.2                    | 6.2 | 8.1                    |
| TAR3  | 98.1| 99.6                   | 99.7| 100.0                  | 100.0| 100.0                  |
| 3TAR1 | 18.4| 18.8                   | 30.4| 29.9                   | 72.4| 71.5                   |
| NLMA.1| 84.0| 83.6                   | 97.2| 96.9                   | 99.9| 99.9                   |
| NLMA.2| 86.8| 84.2                   | 98.5| 98.3                   | 99.9| 99.9                   |
| BIL.1 | 62.1| 67.0                   | 84.8| 90.9                   | 95.0| 97.0                   |
| BIL.2 | 83.1| 81.5                   | 98.9| 97.8                   | 100.0| 99.9                   |
| EXPAR.1| 100.0| 100.0                 | 100.0| 99.8                   | 100.0| 100.0                  |
| EXPAR.2| 99.1 | 98.2                 | 99.9| 99.9                   | 100.0| 100.0                  |
| NLAR  | 100.0| 100.0                 | 100.0| 100.0                 | 100.0| 100.0                  |
Figure 1: 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.

**S5 Supplementary results for the tree-ring data analysis**

In Figure 1(left) we show the time series of the yearly standardized tree-ring growth index of *Pinus aristata* var. *longaeva*, California, USA (left) (reconstructed by D.A. Graybill, Graybill (2018)). The series ranges from year 800 to 1979 ($n = 1180$). The vertical line in Figure 1(right) indicates the estimated threshold $\hat{r} = 0.97$, the same value of $r$ that maximizes the $T_n$ statistic. The residual analysis is somehow revealing. The correlograms of the residuals for both the ARMA and the TARMA fit are shown in Figure 2. Clearly, there is no correlation structure left in the residuals of both fits and...
this may lead to conclude in favour of the ARMA(1,1) model. However, if the residuals are tested for independence with the entropy metric $S_\rho$ (Giannerini et al., 2015) then a different picture emerges. The results are shown in Figure 3: the residuals of the ARMA(1,1) fit (left panel) present a significant unaccounted (non-linear) dependence at lag 1 whereas there is no dependence structure left in the residuals of the TARMA(1,1) fit (right panel).

**Bibliography**

Andrews, D. (2003). Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica* 71(1), 395–397.

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.

Giannerini, S., E. Maasoumi, and E. Bee Dagum (2015). Entropy testing for nonlinear serial dependence in time series. *Biometrika* 102, 661–675.

Graybill, D. (2018). Graybill - Methuselah Walk - PILO - ITRDB CA535. [https://www.ncdc.noaa.gov/paleo/study/3376](https://www.ncdc.noaa.gov/paleo/study/3376), accessed: 2020-03-02.
Figure 2: Global and partial correlograms up to 30 years computed over the residuals of the ARMA(1,1)(red/left) and TARMA(1,1)(blue/right) fits. The horizontal dashed lines correspond to the rejection bands of the null hypothesis of absence of correlation at level 99%.
Figure 3: Entropy measure $S_\rho$ up to 30 years computed over the residuals of the ARMA(1,1) and TARMA(1,1) fits. The green and blue dashed lines correspond to the rejection bands of the null hypothesis of independence at level 99% and 99.9%, respectively.

Li, G. and W. Li (2011, 02). Testing a linear time series model against its threshold extension. *Biometrika* 98(1), 243–250.

Ling, S. and H. Tong (2005, 12). Testing for a linear MA model against threshold MA models. *Ann. Statist.* 33(6), 2529–2552.

van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge series in statistical and probabilistic Mathematics, Cambridge University Press.

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
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)