Statistica Sinica Preprint No: SS-2022-0125

Title	Testing for Threshold Regulation in Presence of Measurement Error
Manuscript ID	SS-2022-0125
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0125
Complete List of Authors	Kung-Sik Chan, Simone Giannerini, Greta Goracci and Howell Tong
Corresponding Authors	Kung-Sik Chan
E-mails	kung-sik-chan@uiowa.edu
Notice: Accepted version subject to English editing.	

Statistica Sinica

Testing for threshold regulation in presence of measurement error

Kung-Sik Chan[1], Simone Giannerini[2], Greta Goracci[3], Howell Tong[4,5,6]

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

3. Free University of Bozen/Bolzano, Italy

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

5. Tsinghua University, China; 6. London School of Economics and Political Science, U.K.

Abstract: Regulation is an important feature characterising many dynamical phenomena and is commonly tested within the threshold autoregressive setting, with the null hypothesis being a global non-stationary process. Nonetheless, this setting is debatable since data are often corrupted by measurement errors. Thus, it is more appropriate to consider a threshold autoregressive moving-average model as the general hypothesis. We implement this new setting with the integrated moving-average model of order one as the null hypothesis. We derive a Lagrange multiplier test which has an asymptotically similar null distribution and provide the first rigorous proof of tightness pertaining to testing for threshold nonlinearity against difference stationarity, which is of independent interest. Simulation studies show that the proposed approach enjoys less bias and higher power in detecting threshold regulation than existing tests, especially when there are measurement errors. We apply the new approach to the time series of real exchange rates of a panel of European countries.

Key words and phrases: Lagrange multiplier test, Threshold autoregressive moving-average model, Purchasing power parity.

1. Introduction

Regulation plays a fundamental role in various fields including economics, finance, biological growth and population fluctuations, etc. Growth processes are generally regulation-free until they enter into extreme phases. For instance, real exchange rates should be regulated through a threshold that triggers the mean reversion towards zero. However, existing tests fail to reject the null hypothesis of a random walk resulting in the so called Purchasing Power Parity (PPP) puzzle, see, e.g., Taylor and Taylor (2004).

The random walk is a simple model for regulation-free dynamics. On the other hand, regulation from above (below) may be captured by a first-order threshold autoregressive model (TAR) which follows a random walk until the process crosses a certain threshold above (below) where mean-reversion takes place, while the process as a whole is *stationary*. A nonlinear stationary process generally renders nonlinear and state-dependent the impulse response to a random shock, which is consequential and could be leveraged in economic regulation. Hitherto, a standard approach to test for dynamic regulation is to adopt the preceding threshold model as the general model and test whether it reduces to a *global* random walk. It has received much attention in the literature (Enders and Granger, 1998; Caner and Hansen, 2001; Bec et al., 2004; Kapetanios and Shin, 2006; Seo, 2008; Park and Shintani, 2016; de Jong et al., 2007; Giordano et al., 2017). However, data are almost always corrupted by measurement error and yet, as far as we are aware, this important issue has not been addressed in the literature. In this case, the TAR model is not appropriate and the null hypothesis should be a global exponential smoothing model instead, i.e., the integrated moving-average IMA(1,1) model, rather than the IMA(1,0) model. Then, the general hypothesis may be taken as the first-order threshold autoregressive moving-average model, i.e., TARMA(1,1), which is driven by an IMA(1,1) model in one of its two regimes. See Section S1 of the Supplementary Material for further justification, which shows that the TARMA$(1,1)$ model is approximately invariant w.r.t. data corruption by independent measurement errors while the IMA$(1,1)$ model is exactly invariant w.r.t. adding measurement errors. Above all, we cannot over-emphasize the critical importance of the role of the moving average term for practical applications.

Just as ARMA models provide a parsimonious approximation to some long AR models, so may TARMA models well approximate some high-order

TAR models parsimoniously Goracci (2020, 2021). Thus, the TARMA model holds substantial promise as a class of nonlinear time series models for exploring nonlinear dynamics. Yet, the TARMA model has been under-explored, partly because of a lack of progress in obtaining conditions on stationarity and ergodicity. Unlike the AR-ARMA analogy, the incorporation of a moving-average part in a nonlinear framework poses major theoretical challenges and has non-trivial implications on the probabilistic structure of the process. Recent work by Chan and Goracci (2019) provides, for the first time, a breakthrough in deriving a set of necessary and sufficient conditions for the (multi-regime) TARMA(1,1) model to admit an irreducible and invertible state-space representation and for its stationarity and ergodicity.

By leveraging on the recent results of Chan and Goracci (2019), we develop a supremum Lagrange Multiplier test (supLM) for threshold regulation, with the TARMA(1,1) model as the general framework. We specify an IMA$(1,1)$ model as the null hypothesis and a TARMA$(1,1)$ with a unit-root regime as the alternative. A difficulty arising from testing for a unit-root against a TARMA model is that the threshold parameters are absent under the null hypothesis. This non-standard situation, in the nonlinear time series context, is well recognized both in the TAR setting (Chan, 1990;

Hansen, 1996; Giannerini et al., 2022) and in the TARMA setting Li and Li (2011); Goracci et al. (2021). The supLM framework overcomes this problem. We derive its asymptotic distribution both under the null hypothesis and local alternatives. We prove that the test is consistent and asymptotically similar in that its asymptotic null distribution does not depend on the value of the MA parameter. Moreover, we provide the first rigorous proof of tightness pertaining to testing for threshold nonlinearity against difference stationarity, which is of independent interest and constitutes a general theoretical framework for ARIMA versus TARMA testing. We also introduce a wild bootstrap version of the supLM statistic that, for finite samples, possesses good properties and robustness against heteroskedasticity. We perform a large scale simulation study to compare our tests with existing tests, in which the alternative hypothesis is that of a threshold model. In general, the size of the latter tests is severely biased in a number of cases to the extent that their use in practical applications remains questionable unless additional information on the data generating process is available. In addition, the comparison includes some of the best performing unit-root tests to date, where the alternative hypothesis does not specify explicitly a nonlinear process.

The paper is structured as follows. In Section 2 we present some fundamentals of the first-order TARMA model and a parametrization that reduces to the IMA(1,1) process under the null hypothesis. In Section 3 we present a supremum Lagrange Multiplier test, including the theoretical framework based on Brownian local time. In Section 4 we develop the asymptotic distribution of the supLM test statistic under the null hypothesis and show that it is nuisance-parameter-free and depends only on the search range of the threshold. The results concerning the local power of the proposed test are summarized in Section 5. In Section 6 we perform a large scale simulation study to show the performance of the supLM test and a wild bootstrap version of it and compare them with numerous existing tests in the recent literature. Section 7 contains an empirical illustration in which we apply the new tests to the pre-Euro monthly real exchange rates of a set of European countries. All the proofs are collected in the Supplementary Material, which contains further results from the Monte Carlo study and from the real data application.

2. Threshold autoregressive moving-average model

Consider the following first-order threshold autoregressive moving-average (TARMA) model:

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1} X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{if } X_{t-d} \leq r \\ \phi_{2,0} + \phi_{2,1} X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} & \text{otherwise,} \end{cases} \tag{2.1}$$

where $\phi_{2,1}$ is fixed at 1 unless stated otherwise, the innovations $\{\varepsilon_t\}$ are independent and identically distributed random variables with zero mean and variance σ^2, ε_t is independent of $X_{t-j}, j \geq 1$, the delay d is a positive integer which, for simplicity, is taken to be 1 henceforth, r is the real-valued threshold parameter, and the ϕ's and θ's are unknown coefficients. The assumption of independence and identical distribution of the innovations will be relaxed later to a martingale difference sequence. The preceding (constrained) TARMA model assumes that the sub-model in the upper regime is a first-order IMA model while the lower regime specifies a general first-order ARMA model. Statistical inference with a TARMA model hinges on whether the model is invertible. We assume $|\theta| < 1$ since it is a necessary and sufficient condition for the invertibility of Model (2.1) (Chan and Tong, 2010). By assuming that the innovations admit a positive, continuous probability density function with finite absolute first moment, Chan and Goracci

(2019) showed that Model (2.1) is an ergodic Markov chain if and only if $\phi_{2,0} < 0$ and either (i) $\phi_{1,1} < 1$, or (ii) $\phi_{1,1} = 1$, $\phi_{1,0} > 0$; ergodicity then implies that the first-order TARMA model admits a unique stationary distribution. Furthermore, under the stronger condition that the innovations admit a finite absolute kth moment for some $k > 2$, Chan and Goracci (2019) provides a complete classification of the parametric regions of Model (2.1) into sub-regions of ergodicity, null recurrence and transience. In particular, the (constrained) first-order TARMA model defined by Model (2.1) is null-recurrent if any of the following holds: (iii) $\phi_{1,1} = 1, \phi_{2,0} = 0, \phi_{1,0} \geq 0$; (iv) $\phi_{1,1} = 1, \phi_{2,0} < 0, \phi_{1,0} = 0$; (v) $\phi_{1,1} < 1, \phi_{2,0} = 0$. If none of the conditions (i)–(v) holds, then the model is transient. Therefore, Model (2.1) encompasses both linear and nonlinear processes spanning a wide spectrum of long-run behaviors including ergodicity, null recurrence and transience.

3. Lagrange multiplier test for threshold regulation

We first formulate a framework for testing for threshold regulation from below. Let $\{X_t, t = 0, 1, \ldots\}$, be a time series and assume that, for $t \geq 1$, X_t satisfies the equation

$$H: \quad X_t = \phi_0 + X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} + (\phi_{1,0} + \phi_{1,1}X_{t-1}) \times I(X_{t-1} \leq r), \quad (3.2)$$

which is a re-parameterization of Model (2.1) with $\phi_0 = \phi_{2,0}$ and by an abuse of notation, $\phi_{1,0}$ and $\phi_{1,1}$ represent, respectively, the difference of intercept and slope of the lower regime relative to their upper-regime counterparts; the initial value X_0 can be fixed at, say, 0. Our interest is in testing whether $\phi_{1,0} = \phi_{1,1} = 0$, in which case the data are generated by the IMA(1,1) model

$$H_0: \quad X_t = \phi_0 + X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, \tag{3.3}$$

where $|\theta| < 1$. If the intercept $\phi_0 \neq 0$, then the IMA(1,1) process has a linear trend. If no such linear trend is apparent in the data, it is reasonable to omit the intercept. Henceforth, we assume that $\phi_0 = 0$ under H_0. The case for $\phi_0 \neq 0$ requires non-trivial modification of the test and its asymptotic distribution so it will be studied elsewhere. However, the intercept terms on the two regimes of any competing stationary first-order TARMA model will be required to model the mean of the data. Indeed, even for mean-deleted data, the intercept terms of the first-order TARMA model are not necessarily zero. Thus, the intercept terms are essential and retained in the constrained TARMA model under H. Testing for threshold regulation from above can be conducted by applying the test to $\{-X_t\}$.

Under the null hypothesis, the threshold parameter is absent thereby complicating the test (Chan, 1990; Hansen, 1996; Li and Li, 2011; Goracci et al., 2021). Our approach is to develop a Lagrange multiplier test statistic for H_0 initially with the threshold parameter fixed at some r. Denote the test statistic as $T_n(r)$. Since r is unknown and indeed absent under H_0, we shall compute $T_n(r)$ for all r over some data-driven interval, say, $[a, b]$ with the end points being some percentiles of the observed data. For instance, a could be the 20-th percentile and b the 80-th percentile. Then the overall test statistic results in $T_n = \sup_{r\in[a,b]} T_n(r)$. Besides taking the supremum, other approaches including integration can be employed to derive an overall test statistic.

The Lagrange multiplier test is developed based on the Gaussian likelihood conditional on X_0:

$$\ell = -\log(2\pi\sigma^2) \times n/2 - \sum_{t=1}^{n} \varepsilon_t^2/(2\sigma^2), \tag{3.4}$$

where, by an abuse of notation, $\forall t \geq 1$,

$$\varepsilon_t = X_t - \{\phi_0 + X_{t-1} + (\phi_{1,0} + \phi_{1,1}X_{t-1}) \times I(X_{t-1} \leq r)\} + \theta\varepsilon_{t-1}, \tag{3.5}$$

with the unknown ε_0 set to be zero; ε_t in the preceding formula is a function of $\phi_0, \phi_{1,0}, \phi_{1,1}, \theta$ and r, but the arguments are generally suppressed for simplicity. Let $\boldsymbol{\psi} = (\phi_0, \theta, \sigma^2,\ \phi_{1,0}, \phi_{1,1})^\intercal$, with its components denoted by $\psi_j, j = 1, 2, \ldots, 5$, and let it be partitioned into $\boldsymbol{\psi}_1 = (\phi_0, \theta, \sigma^2)^\intercal$ and $\boldsymbol{\psi}_2 = (\phi_{1,0}, \phi_{1,1})^\intercal$. The null hypothesis can be succinctly expressed as $H_0 : \boldsymbol{\psi}_2 = 0$.

First, consider the case of known threshold r. Partition the Fisher information matrix according to $\boldsymbol{\psi}_i, i = 1, 2$ into

$$I_n(r) = \begin{pmatrix} I_{1,1,n}(r) & I_{1,2,n}(r) \\ I_{2,1,n}(r) & I_{2,2,n}(r) \end{pmatrix}. \tag{3.6}$$

The Lagrange multiplier test statistic is an asymptotic approximation of twice the Gaussian likelihood ratio statistic, based on a second-order Taylor expansion. For fixed r, it equals

$$T_n(r) = \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2^{\intercal}}(r) \left\{ \hat{I}_{2,2,n}(r) - \hat{I}_{2,1,n}(r) \hat{I}_{1,1,n}^{-1}(r) \hat{I}_{1,2,n}(r) \right\}^{-1} \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2}(r), \tag{3.7}$$

where $\partial \hat{\ell}/\partial \boldsymbol{\psi}_2(r)$ is equal to $\partial \ell/\partial \boldsymbol{\psi}_2$ evaluated at the constrained estimate $\boldsymbol{\psi}_1 = \hat{\boldsymbol{\psi}}_1$ given $\boldsymbol{\psi}_2 = 0$ and the threshold parameter fixed at r. Similarly defined are $\hat{I}_{i,j,n}(r), 1 \leq i, j, \leq 2$; see Subsection 3.1 for the formulas. Because the threshold r is unknown, the overall supLM statistic is $T_n = \sup_{r \in [a,b]} T_n(r)$ with a and b, for instance, being some pre-specified percentiles of the observed data; see Subsection 3.2. for further discussion

3.1 Gaussian likelihood estimation

In this sub-section, we provide some details on the estimation of the model parameters under the null hypothesis, as well as the computation of the score vector and Fisher information matrix. The score vector is

$$\frac{\partial \ell}{\partial \psi_j} = -\sum_{t=1}^{n} \frac{\varepsilon_t}{\sigma^2}\frac{\partial \varepsilon_t}{\partial \psi_j},\ 1 \le j \le 5, j \ne 3, \quad \frac{\partial \ell}{\partial \psi_3} = \frac{\partial \ell}{\partial \sigma^2} = \sum_{t=1}^{n} \frac{\varepsilon_t^2 - \sigma^2}{2\sigma^4}$$

where for $t > 1$,

$$\frac{\partial \varepsilon_t}{\partial \phi_0} = -1 + \theta \frac{\partial \varepsilon_{t-1}}{\partial \phi_0} = -\sum_{j=0}^{t-1} \theta^j, \tag{3.8}$$

$$\frac{\partial \varepsilon_t}{\partial \theta} = \varepsilon_{t-1} + \theta \frac{\partial \varepsilon_{t-1}}{\partial \theta} = \sum_{j=0}^{t-1} \theta^j \varepsilon_{t-1-j}, \tag{3.9}$$

$$\frac{\partial \varepsilon_t}{\partial \phi_{1,0}} = -I(X_{t-1} \le r) + \theta \frac{\partial \varepsilon_{t-1}}{\partial \phi_{1,0}} = -\sum_{j=0}^{t-1} \theta^j I\left(X_{t-1-j} \le r\right), \tag{3.10}$$

$$\frac{\partial \varepsilon_t}{\partial \phi_{1,1}} = -X_{t-1} I(X_{t-1} \le r) + \theta \frac{\partial \varepsilon_{t-1}}{\partial \phi_{1,1}} = -\sum_{j=0}^{t-1} \theta^j X_{t-1-j} I\left(X_{t-1-j} \le r\right), \tag{3.11}$$

with initial values given by $\partial\varepsilon_1/\partial\phi_0 = -1, \partial\varepsilon_1/\partial\theta = 0, \partial\varepsilon_1/\partial\phi_{1,0} = -I(X_0 \le r)$ and $\partial\varepsilon_1/\partial\phi_{1,1} = -X_0 I(X_0 \le r)$. Below, we sometimes write, as a typical example, $\partial\varepsilon_t/\partial\phi_{1,1} = -(1-\theta B)^{-1}\{X_{t-1}I(X_{t-1} \le r)\}$, where B is the backshift operator that shifts the indices backward by 1 unit. The IMA(1,1) model under the null hypothesis can be estimated by solving the score equation $\partial\ell/\partial\boldsymbol{\psi}_1 = 0$, yielding $\hat{\boldsymbol{\psi}}_1 = \hat{\boldsymbol{\psi}}_{1,n} = (\hat{\phi}_{0,n}, \hat{\theta}_n, \hat{\sigma}_n^2)^\intercal$. Thus, the overall estimator of $\boldsymbol{\psi}$ under H_0 is $\hat{\boldsymbol{\psi}} = (\hat{\phi}_{0,n}, \hat{\theta}_n, \hat{\sigma}_n^2, 0, 0)^\intercal$, with the residuals given by

$$\hat{\varepsilon}_t = X_t - X_{t-1} - \hat{\phi}_0 + \hat{\theta}\hat{\varepsilon}_{t-1}, \quad \forall t \ge 1, \tag{3.12}$$

where $\hat{\varepsilon}_0 = 0$. The observed Fisher information (excluding the threshold parameter) is given by $I_n = -\frac{\partial^2 \ell}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^\intercal}$, whose (i, j)-th element with $i, j \neq 3$ is given by

$$\sum_{t=1}^{n} \frac{1}{\sigma^2} \frac{\partial \varepsilon_t}{\partial \psi_i} \frac{\partial \varepsilon_t}{\partial \psi_j} + \sum_{t=1}^{n} \frac{\varepsilon_t}{\sigma^2} \frac{\partial^2 \varepsilon_t}{\partial \psi_i \partial \psi_j} = (1 + o_p(1)) \times \sum_{t=1}^{n} \frac{1}{\sigma^2} \frac{\partial \varepsilon_t}{\partial \psi_i} \frac{\partial \varepsilon_t}{\partial \psi_j}, \quad (3.13)$$

its $(3, i)$-th element with $i \neq 3$ equal to $\sum_{t=1}^{n} \frac{\varepsilon_t}{\sigma^4} \frac{\partial \varepsilon_t}{\partial \psi_i} = o_p(n)$, and the $(3, 3)$-th element equal to $\sum_{t=1}^{n} \left(\frac{1}{2\sigma^4} - \frac{\varepsilon_t^2}{\sigma^6} \right)$, where the $o_p(1)$ and $o_p(n)$ terms hold uniformly in r, when the expressions are evaluated at the true parameter value under the null hypothesis; hence they are asymptotically negligible (via arguments similar to those in the proof of Theorem 1), and omitted in all numerical work reported herein. Below, we sometimes write, e.g., $\partial \ell / \partial \psi_j(\boldsymbol{\psi}; r)$, to highlight the role of the arguments; we further simplify the notation, for example, from $\partial \ell / \partial \psi_j(\boldsymbol{\psi}_0; r)$ to $\partial \ell / \partial \psi_j(r)$, with $\boldsymbol{\psi}_0$ denoting the true value under H_0. Moreover, $I_{1,1,n}(\boldsymbol{\psi}_0; r)$ and $\partial \ell / \partial \boldsymbol{\psi}_1(\boldsymbol{\psi}_0; r)$ are further simplified as $I_{1,1,n}$ and $\partial \ell / \partial \boldsymbol{\psi}_1$ as they do not depend on r. By an abuse of notation, the true values of the moving-average coefficient and the innovation variance under H_0 are simply denoted by θ and σ^2; no confusion should arise as the context will make clear whether they stand for the generic parameters or their true values.

3.2 The choice of the threshold range

For theoretical analysis, the threshold range is specified as $R_n = (n^{1/2}(1-\theta)\sigma \times r_L, n^{1/2}(1-\theta)\sigma \times r_U)$ where $r_L < r_U$ are two fixed finite numbers. We now justify this choice of the threshold range. First, some heuristics will be employed. Under the null hypothesis (with $\phi_0 = 0$),

$$X_t = \varepsilon_t + (1-\theta)\sum_{s=1}^{t-1}\varepsilon_s - \theta\varepsilon_0 + X_0.$$

Hence, $\{n^{-1/2}X_{[sn]}, 0 \le s \le 1\}$, where $X_{[sn]} = \sum_{t=1}^{[sn]} X_t$ and $[sn]$ is the largest integer less than or equal to sn, converges in distribution to $\{(1-\theta)\sigma W_s\}$ where $\{W_s\}$ is the standard Brownian motion. It is well known (Björk, 2019, Theorems 3.1 and 3.2) that the Brownian local time $\{L_t^x, t \ge 0, -\infty < x < \infty\}$ defined as follows:

$$L_t^x = |W_t - x| - |x| - \int_0^t \text{sign}(W_s - x)ds,$$

where sign(x) denotes the sign of x, is essentially the probability density function of the Brownian realization in the sense that for any bounded real-valued Borel function f,

$$\int_0^1 f(W_s)ds = \int_{-\infty}^{\infty} f(x)L_1^x dx. \tag{3.14}$$

Thus, any quantile of $\{X_t, t = 0, \dots, n\}$ is asymptotically equal to $n^{1/2}(1-\theta)\sigma$ times the corresponding quantile of $\{W_s, 0 \le s \le 1\}$. Since the Brownian local time process is a random process, so the quantiles are realization specific! This motivates us to set the threshold to be of the form $r_n = (1-\theta)\tau\sigma n^{1/2}$ for some fixed τ, in which case

$$n^{-1/2}\frac{\partial \ell}{\partial \phi_{1,0}}(r_n) = n^{-1/2}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\frac{1}{1-\theta B}\left\{I\left(\frac{X_{t-1}}{n^{1/2}(1-\theta)\sigma}\le \tau\right)\right\}. \quad (3.15)$$

The right side of (3.15) is a Riemann-Stieltjes sum over $[0,1]$, with a step integrator jumping at t/n with jump size $(n\sigma^2)^{-1/2}\varepsilon_t$ and the integrand is a piecewise constant function which equals $\sum_{j=0}^{t-1}\theta^j I\left(\{n^{1/2}(1-\theta)\sigma\}^{-1}X_{t-1-j}\le\tau\right)$ over the interval $[n^{-1}(t-1), n^{-1}t]$, for $t=1,2,\ldots,n$. The integrator converges weakly to the standard Brownian motion whereas the integrand to $(1-\theta)^{-1}I(W_s\le\tau)$ as $t,n\to\infty$ such that $t/n\to s$ in $[0,1]$. Thus, heuristically, $n^{-1/2}\partial\ell/\partial\phi_{1,0}(r_n)$ converges in distribution to $(1-\theta)^{-1}\sigma^{-1}\int_0^1 I(W_s\le \tau)dW_s$ under H_0 and as $n\to\infty$, or in symbol,

$$n^{-1/2}\frac{\partial \ell}{\partial \phi_{1,0}}(r_n) \rightsquigarrow \frac{1}{(1-\theta)\sigma}\int_0^1 I(W_s\le\tau)dW_s. \quad (3.16)$$

This asymptotic result and other heuristic results stated below can be essentially justified using Theorem 7.10 in Kurtz and Protter (1996). Similarly,

$$n^{-1}\frac{\partial \ell}{\partial \phi_{1,1}}(r_n) = n^{-1/2}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma}\frac{1}{1-\theta B}\left[\frac{X_{t-1}}{n^{1/2}\sigma}I\left\{\frac{X_{t-1}}{n^{1/2}(1-\theta)\sigma}\le\tau\right\}\right]$$
$$\rightsquigarrow \int_0^1 W_s I(W_s\le\tau)dW_s \quad (3.17)$$

$$n^{-1/2}\frac{\partial \ell}{\partial \phi_0} = n^{-1/2}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\frac{1}{1-\theta B}(1) \rightsquigarrow \frac{1}{(1-\theta)\sigma}\int_0^1 dW_s = \frac{W_1}{(1-\theta)\sigma}. \quad (3.18)$$

Note the different rates of normalization. Let K_n be the 5×5 diagonal matrix with the last diagonal elements being n and other diagonal elements all being $n^{1/2}$. We can also show that $K_n^{-1}I_n(r_n)K_n^{-1}$ converges in probability to a matrix denoted by $\mathcal{I}(\tau)$ which can be blocked as I_n; see Eq. (3.6). In particular, $\mathcal{I}_{1,1}$ is a diagonal matrix comprising $(1-\theta)^{-2}\sigma^{-2}, (1-\theta^2)^{-1}, (4\sigma^4)^{-1}$ as its diagonal elements,

$$\mathcal{I}_{2,2}(\tau) = \begin{pmatrix} \frac{1}{(1-\theta)^2\sigma^2}\int_0^1 I(W_s\le\tau)ds & \frac{1}{(1-\theta)\sigma}\int_0^1 W_s I(W_s\le\tau)ds \\ \frac{1}{(1-\theta)\sigma}\int_0^1 W_s I(W_s\le\tau)ds & \int_0^1 W_s^2 I(W_s\le\tau)ds \end{pmatrix};$$

$$\mathcal{I}_{2,1}(\tau) = \begin{pmatrix} \frac{1}{(1-\theta)^2\sigma^2}\int_0^1 I(W_s\le\tau)ds & 0 & 0 \\ \frac{1}{(1-\theta)\sigma}\int_0^1 W_s I(W_s\le\tau)ds & 0 & 0 \end{pmatrix}.$$

Note that $\mathcal{I}_{1,1}$ does not depend on τ. Thus, θ and σ^2 are locally orthogonal to the other parameters around the true parametric value under H_0.

Hence, their estimates are expected to be asymptotically independent of the proposed test statistic, as will be shown below to be the case.

Remark: In practice, the choice of r_L and r_U must ensure adequate data for the asymptotic distribution of T to be valid, which requires adequate data in the left and right tails beyond the threshold range. Simulation results in Section 6 suggest a rough guideline that for normal innovations, there should be at least 25 data points below r_L (above r_U).

3.3 A wild bootstrap approach

In this section we introduce a wild bootstrap version of our supLM statistic. This bootstrap scheme has proved to deliver valid inference under heteroskedastic disturbances (Liu, 1988; Mammen, 1993; Davidson and Flachaire, 2008). As also shown in Cavaliere and Taylor (2008) in the context of unit-root testing, the wild bootstrap is capable of correctly reproducing the first-order limiting null distribution of the statistics in the case of non-stationary volatility. The algorithm has the following structure:

1. Compute $\tilde{X}_t = X_t - \hat{\boldsymbol{\beta}}^\intercal \mathbf{d_t}$, where $\mathbf{d_t}$ is a vector of deterministic components and $\hat{\boldsymbol{\beta}}$ is obtained through either OLS or GLS detrending;

2. Obtain $\hat{\theta}$, the maximum likelihood estimate for θ, and the residuals $\hat{e}_t$ from the following IMA(1,1) model: $\tilde{X}_t = \varepsilon_t - \theta\varepsilon_{t-1}$;

3. Compute wild bootstrap errors $\hat{e}_t^* = \hat{e}_t\eta_t$, where η_t is a random variable such that $E(\eta_t) = 0$ and $E(\eta_t^2) = 1$. Henceforth, we use the Rademacher scheme: η_t equals ± 1 with equal probability.

4. Obtain the bootstrap resample $\hat{X}_t^* = \sum_{j=1}^{t}\left(\hat{e}_j^* - \hat{\theta}\hat{e}_{j-1}^*\right)$, and compute the supLM statistic T_n^* upon it.

5. Repeat steps 3–4 B times so as to obtain the bootstrap test statistic, $T_n^{*b},\ b = 1, \ldots B$ and compute the bootstrap p-value as the relative frequency that T_n^{*b} is not less than the observed T_n.

4. The null distribution

We now derive the asymptotic distribution of $T_n(r)$ under the null hypothesis of an IMA(1,1) model with zero intercept. Using second-order Taylor expansion and after some routine algebra, it holds that

$$\frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2}(r_n) \approx \frac{\partial \ell}{\partial \boldsymbol{\psi}_2}(r_n) - I_{2,1,n}(r_n) I_{1,1,n}^{-1} \frac{\partial \ell}{\partial \boldsymbol{\psi}_1}. \tag{4.19}$$

More rigorously, letting

$$Q_n = \begin{pmatrix} n^{-1/2} & 0 \\ 0 & n^{-1} \end{pmatrix}, \quad P_n = n^{-1/2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

we shall prove below that uniformly for $r_n = n^{1/2}(1-\theta)\sigma\tau \in R_n = (n^{1/2}(1-\theta)\sigma \times r_L, n^{1/2}(1-\theta)\sigma \times r_U)$, where $r_L < r_U$ are fixed numbers,

$$Q_n \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2}(r_n) = Q_n \frac{\partial \ell}{\partial \boldsymbol{\psi}_2}(r_n) - I_{2,1}(\tau) I_{1,1}^{-1} P_n \frac{\partial \ell}{\partial \boldsymbol{\psi}_1} + o_P(1)$$
$$= Q_n \frac{\partial \ell}{\partial \boldsymbol{\psi}_2}(r_n) - \tilde{I}_{2,1}(\tau) \tilde{I}_{1,1}^{-1} P_n \frac{\partial \ell}{\partial \phi_0} + o_P(1), \qquad (4.20)$$

where, owing to the form of $I_{2,1}(\tau)$, $\tilde{I}_{1,1} = (1-\theta)^{-2}\sigma^{-2}$ and

$$\tilde{I}_{2,1} = \begin{pmatrix} \frac{1}{(1-\theta)^2\sigma^2} \int_0^1 I(W_s \le \tau) ds \\ \frac{1}{(1-\theta)\sigma} \int_0^1 W_s I(W_s \le \tau) ds \end{pmatrix}.$$

The intercept $\hat{\phi}_{0,n}$ admits the asymptotic representation under H_0 (Brockwell and Davis, 2001, c.f. Eqn. (8.11.5))

$$P_n^{-1}(\hat{\phi}_{0,n} - \phi_0) = (\tilde{I}_{1,1})^{-1} P_n \frac{\partial \ell}{\partial \phi_0} + o_P(1).$$

A key step in deriving the limiting null distribution of the proposed test is then to demonstrate that uniformly for $r_n = n^{1/2}(1-\theta)\sigma\tau \in R_n$

$$Q_n \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2}(r_n) = Q_n \frac{\partial \ell}{\partial \boldsymbol{\psi}_2}(r_n) - \tilde{I}_{2,1}(\tau) P_n^{-1}(\hat{\phi}_{0,n} - \phi_0) + o_p(1). \qquad (4.21)$$

Let

$$H(\tau) = \left(\int_0^1 dW_s, \int_0^1 I(W_s \le \tau) dW_s, \int_0^1 W_s I(W_s \le \tau) dW_s \right)^\intercal \qquad (4.22)$$

and

$$\Lambda(\tau) = \begin{pmatrix} 1 & \int_0^1 I(W_s \le \tau)ds & \int_0^1 W_s I(W_s \le \tau)ds \\ \int_0^1 I(W_s \le \tau)ds & \int_0^1 I(W_s \le \tau)ds & \int_0^1 W_s I(W_s \le \tau)ds \\ \int_0^1 W_s I(W_s \le \tau)ds & \int_0^1 W_s I(W_s \le \tau)ds & \int_0^1 W_s^2 I(W_s \le \tau)ds \end{pmatrix}. \tag{4.23}$$

Let $\Lambda(\tau)$ be partitioned into a 2×2 block matrix with the $(2,2)$-th block being 2×2. Similarly partitioned is $H(\tau) = (H_1(\tau), H_2(\tau))^\intercal$. It follows from Eq. (4.20) and Eqs. (3.16)–(3.18) that the asymptotic null distribution of $T_n(r_n)$ can be shown to be the same as that of

$$\left\| (\{\Lambda^{-1}(\tau)\}_{2,2})^{1/2} \left(H_2(\tau) - \Lambda_{2,1}(\tau) H_1(\tau)\right) \right\|^2,$$

where $\|\cdot\|^2$ is the squared Euclidean norm of the enclosed vector. It is readily shown that $\{\Lambda^{-1}(\tau)\}_{2,2} = \{\Lambda_{2,2}(\tau) - \Lambda_{2,1}(\tau)\Lambda_{1,2}(\tau)\}^{-1}$. The asymptotic null distribution of T_n is derived in Theorem 1 below, under the following assumption:

(**A1**): Let $r_L < r_U$ be two fixed real numbers. Let

$$\mathcal{T}_n(\tau) = n^{-1/2} \sum_{t=2}^{n} \frac{\varepsilon_t}{\sigma} \sum_{j=0}^{t-2} \theta^j I\left\{ r_L < \frac{X_{t-1-j}}{n^{1/2}(1-\theta)\sigma} \le \tau \right\},$$

for $r_L \le \tau \le r_U$. Suppose (i) there exists a constant $C > 0$ such that, for any fixed $r_L \le \tau_1 < \tau_2 \le r_U$,

$$E\left\{ |\mathcal{T}_n(\tau_2) - \mathcal{T}_n(\tau_1)|^4 \right\} \le C(|\tau_2 - \tau_1|^{3/2} + |\tau_2 - \tau_1|/n), \tag{4.24}$$

and (ii) uniformly for $a \leq \tau_1 < \tau_2 \leq b$,

$$|\mathcal{T}_n(\tau_2) - T_n(\tau_1)| \leq K \times L(n)(n \log\log n)^{1/2}|\tau_2 - \tau_1| + o_p(1) \qquad (4.25)$$

as $n \to \infty$ where the $o_p(1)$ term holds uniformly, K is a constant that may depend on θ, and $L(\cdot)$ is some slowly varying function, i.e., for any $\lambda > 0, L(\lambda x)/L(x) \to 1$ as $x \to \infty$.

Theorem 1. *Suppose H_0 holds so that $\{X_t, t = 0, 1, \dots,\}$ is an IMA(1,1) process satisfying Eq. (3.3), with the intercept $\phi_0 = 0$, $|\theta| < 1$ and the innovations are independent and identically distributed with zero mean and finite positive variance. Suppose there exist two real numbers $r_L < r_U$ such that (**A1**) holds. Then as $n \to \infty$, $T_n = \sup\{T_n(r), r \in [n^{1/2}(1-\theta)\sigma r_L, n^{1/2}(1-\theta)\sigma r_U]\}$ converges in distribution to*

$$F(W; r_L, r_U) = \sup_{\tau\in[r_L, r_U]} \left\| \left[\{\Lambda^{-1}(\tau)\}_{2,2}\right]^{1/2} \{H_2(\tau) - \Lambda_{2,1}(\tau)H_1(\tau)\} \right\|^2, \qquad (4.26)$$

whose distribution is parameter-free, although it depends on the search range of the threshold.

We remark that the assumption of independence and identical distribution of the innovations in the preceding theorem can be relaxed to $\{\varepsilon_t\}$ being a stationary, ergodic, martingale difference sequence with respect to the σ-algebra $\mathcal{F}_t$ generated by $\varepsilon_{t-s}, s \leq 0$; the proof is essentially the same.

Remark 1. Conditions (4.24)–(4.25) provide a new set of general sufficient conditions for the tightness of a sequence of stochastic processes; specifically the tightness of $\{T_n(n^{1/2}(1-\theta)\tau), r_L \leq \tau \leq r_U\}$. These sufficient conditions are motivated by the approach taken by Billingsley (1968), Theorem 22.1, for studying the tightness of empirical processes for stationary, mixing data, and are tailor made for coping with nonstationarity under the null. To the best of our knowledge, this is the first rigorous proof of tightness for testing threshold nonlinearity against difference stationarity and constitutes a general theoretical framework that can be used in different settings.

The preceding theorem assumes deterministic threshold search interval. It can be readily extended to the case that the end points are fixed quantiles of the data, which are realization specific. We omit the proof as it is based on routine analysis that builds on Theorem 1 and the facts that for any fixed $0 < p < 1$, (i) the p-quantile of $\{W_s, 0 \leq s \leq 1\}$ is $O_p(1)$, which follow from Björk (2019), Proposition 3.2, and the Markov inequality, and (ii) under H_0, the p-quantile of $\{X_t, t = 0, \ldots, n\}$ is asymptotically equal to its counterpart of $\{W_s, 0 \leq s \leq 1\}$ times $n^{1/2}(1-\theta)\sigma$; see the discussion just below (3.14).

The following result shows that Theorem 1 holds for normally distributed innovations.

Theorem 2. *Conditions (4.24) and (4.25) hold if (i) $|\theta| < 1$ and (ii) $\{\varepsilon_t\}$ are independent and identically normally distributed with zero mean and finite positive variance.*

Since the null distribution of T_n is asymptotically similar, its quantiles can be derived numerically. The tabulated quantiles of the null distribution for different threshold ranges can be found in Section S4 of the Supplementary Material.

5. Local Power

In this section we derive the asymptotic distribution of the supLM statistic under a sequence of local threshold alternatives and prove its consistency in having power approaching 1 with increasing departure in some direction from the null hypothesis. The mathematical framework is as follows. For each positive integer n, the system of hypothesis is:

$H_{0,n}$: $(X_0, \ldots, X_n)$ follow the IMA(1,1) model: $X_t = X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$.

$H_{1,n}$: $(X_0, \ldots, X_n)$ follow the TARMA(1,1) model:

$$X_t = \begin{cases} n^{-1/2}h_{1,0} + \left(1 + n^{-1}h_{1,1}\right)X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} & \text{if } \frac{X_{t-1}}{\sigma n^{1/2}(1-\theta)} \leq \tau_0 \\ n^{-1/2}h_{2,0} + \left(1 + n^{-1}h_{2,1}\right)X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} & \text{if } \frac{X_{t-1}}{\sigma n^{1/2}(1-\theta)} > \tau_0, \end{cases} \tag{5.27}$$

where $\mathbf{h} = (h_{1,0}, h_{2,0}, h_{1,1}, h_{2,1})^\intercal$ is a fixed vector with $h_{i,1} \le 0, i = 1, 2$ and τ_0 is a fixed threshold. Note that if $h_{1,1} < 0$ ($h_{2,1} < 0$), then the model is locally stable in the lower (upper) regime, for all n sufficiently large. In order to derive the local power, we henceforth impose the following mild regularity conditions:

C1: The innovations are assumed to be independent and identically distributed, with zero mean, finite positive standard deviation, σ, and probability density function $f(\cdot/\sigma)/\sigma$, where (i) f is a bounded function, $\log(f(x))$ is twice differentiable with Lipschitz continuous first and second derivatives over the support of the probability density function, (ii) the moment generating function of the innovations exists and is finite over some open interval around 0, and (iii) $\mathcal{I}_f = -\int(\ddot{f}f - \dot{f}^2/f^2)(x) \times f(x)dx$ is a finite positive number, where the first (second) derivative of f is denoted by $\dot{f}$ ($\ddot{f}$).

C2: $-\pi/2 < h_{1,1}, h_{2,1} \le 0$ and $h_{1,1} + h_{2,1} < 0$.

Note that $\mathcal{I}_f$ is the Fisher information for the location model $f(\cdot - \mu)$ where μ is the location parameter. Let $P_{0,n}$ and $P_{1,n}$ be the probability measures induced by $(X_0, \ldots, X_n)$ under $H_{0,n}$ and $H_{1,n}$, respectively. Condition (C1) holds for many commonly used innovation distributions including the normal distribution and the t-distribution. Condition (C2) ensures that the local alternative first-order TARMA model is asymptotically locally stable in at least one regime. These two conditions are imposed to ensure that $\{P_{1,n}\}$ is contiguous to $\{P_{0,n}\}$. Finally, let ρ be the correlation between ε_t and $(\dot{f}/f)(\varepsilon_t)$, i.e., $\rho = \int x\dot{f}(x)dx/\sqrt{\mathcal{I}_f}$ where $\mathcal{I}_f$ is the Fisher information of the innovation distribution with unit σ, as defined in condition (C1).

Theorem 3. *Suppose all the conditions stated in Theorem 1 hold. Assume (C1) and (C2) hold. Under $H_{1,n}$ and as $n \to \infty$, $T_n = \sup\{T_n(r), r \in [n^{1/2}(1-\theta)\sigma r_L, n^{1/2}(1-\theta)\sigma r_U]\}$, where r_L, r_U are two fixed numbers, converges in distribution to $F(W; r_L, r_U)$ defined in Eq. (4.26) but with W now being a threshold diffusion process satisfying the following stochastic differential equation (SDE):*

$$dW_s = dW_s^\dagger + \begin{cases} \rho\sqrt{\mathcal{I}_f}\left\{\frac{h_{1,0}}{\sigma(1-\theta)} + h_{1,1}W_s\right\}ds, & \text{if } W_s \le \tau_0, \\ \rho\sqrt{\mathcal{I}_f}\left\{\frac{h_{2,0}}{\sigma(1-\theta)} + h_{2,1}W_s\right\}ds, & \text{otherwise,} \end{cases} \tag{5.28}$$

where $W_0 = 0$ almost surely and $\{dW_s^\dagger, s \ge 0\}$ is a standard Brownian motion.

Henceforth in this section, W denotes the threshold diffusion satisfying Eq. (5.28). Note that if $h_{i,0} = h_{i,1} = 0, i = 1, 2$, then we get back the limiting null distribution for T_n. Otherwise, W is a threshold diffusion process (Su and Chan, 2015). Thus, the building block W determining the limiting distribution of the supLM statistic changes from a standard Brownian motion under $H_{0,n}$ to a threshold diffusion under $H_{1,n}$, if $\rho \neq 0$. Consequently, the proposed test would have power to detect the local threshold alternatives. Since the functional $F(\cdot; r_L, r_U)$ is quite complex, in Section S2.4 of the Supplementary Material we detail an example to demonstrate the consistency of the proposed test.

6. Finite sample performance

To better approximate the finite sample distribution of T_n, we have simulated the null distributions for the sample sizes in use. Moreover, since we have found that the finite sample distribution of T_n changes appreciably only when $|\theta|$ is close to one, we have adopted the following, conservative, approach: if $|\hat{\theta}| > 0.3$, we use the quantiles of the simulated null with $\theta = \text{sign}(\hat{\theta}) \cdot 0.9$. Furthermore, wild bootstrap (see § 3.3) is also added to improve the empirical size of the test. We denote our asymptotic test and its wild bootstrap version by sLM and sLMb, respectively.

We compare the empirical performance of the proposed test with several competing tests – those designed for threshold alternatives and those without specific nonlinear alternative. The former tests include those proposed by Kapetanios and Shin (2006) (KS), Enders and Granger (1998) (EG) and Bec et al. (2004) (BBC), with their bootstrap variants (if implemented) denoted as KSb, etc. The latter set includes the ADF test of Dickey and Fuller (1979), the class of M tests of Ng and Perron (2001) ($\bar{\mathrm{M}}^{\mathrm{g}}$), the $\bar{\mathrm{M}}\mathrm{P}_{\mathrm{T}}^{\mathrm{GLS}}$ test of Ng and Perron (2001) (MP_{T}) and the GLS detrended version of the ADF test ($\mathrm{ADF}^{\mathrm{g}}$), and the test $\mathrm{M}^{\mathrm{GLS}}$ of Perron and Qu (2007) (M^{g}). We have obtained the results for these preceding tests, although only the best performing ones are reported.

Table 1: Rejection percentages from the TARMA model of Eq.(6.29), with nominal size at $\alpha = 5\%$. Sizes over 15% are highlighted in bold font.

	asymptotic									bootstrap	
θ	sLM	$\bar{\mathrm{M}}^{\mathrm{g}}$	M^{g}	MP_{T}	ADF	$\mathrm{ADF}^{\mathrm{g}}$	KS	BBC	EG	sLMb	KSb
$n = 100$											
-0.9	2.2	7.7	7.0	7.1	2.5	3.8	8.1	11.2	7.1	5.1	4.9
-0.5	1.6	6.3	6.1	5.8	4.8	5.1	7.0	6.1	5.7	5.0	5.6
0.0	1.6	5.1	5.1	4.6	5.3	5.6	8.1	2.7	5.0	4.5	5.3
0.5	1.7	5.6	5.9	5.1	6.7	7.4	**64.5**	10.2	**57.5**	5.2	**58.6**
0.9	11.3	6.5	**17.7**	6.4	**77.9**	**17.8**	**100.0**	**92.4**	**100.0**	5.7	**99.8**
$n = 300$											
-0.9	5.5	6.7	6.3	6.1	3.3	4.2	6.3	14.0	6.5	5.3	3.8
-0.5	4.7	5.2	5.1	4.8	4.5	4.5	5.1	8.5	5.4	5.0	4.5
0.0	2.9	4.9	4.9	4.4	5.1	4.6	6.9	3.2	4.4	5.6	4.3
0.5	2.3	5.5	5.4	5.1	5.4	5.8	**74.5**	**19.0**	**61.1**	4.9	**67.7**
0.9	4.9	1.9	2.4	1.9	**86.0**	**15.8**	**100.0**	**99.7**	**100.0**	4.9	**100.0**
$n = 500$											
-0.9	8.1	6.4	6.1	6.0	7.4	4.7	5.7	**16.0**	6.1	5.5	4.0
-0.5	5.3	5.5	5.3	5.0	5.1	4.8	5.2	9.2	5.4	4.7	4.2
0.0	3.5	4.9	4.8	4.5	4.9	4.6	7.3	3.5	5.0	3.8	4.5
0.5	2.5	5.2	5.1	4.8	5.1	5.3	**78.4**	**23.7**	**62.3**	4.5	**71.7**
0.9	3.3	1.3	1.4	1.4	**83.2**	14.5	**100.0**	**99.9**	**100.0**	5.4	**100.0**

The sample sizes considered are 100, 300 and 500. The rejection percentages are derived with a nominal size $\alpha = 5\%$ and based upon 10000 replications. In order to reduce the computational burden, for the bootstrap tests we select 1000 replications and $B = 1000$ bootstrap resamples. The threshold search ranges from the 25% to the 75% of the sample distribution. We simulate data from the following first-order TARMA model

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1}X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{if } X_{t-1} \leq 0, \\ \phi_{2,0} + \phi_{2,1}X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{otherwise,} \end{cases} \tag{6.29}$$

where $(\phi_{1,0}, \phi_{1,1}, \phi_{2,0}, \phi_{2,1}) = \tau \times (0, 0.7, -0.02, 0.99) + (1 - \tau) \times (0, 1, 0, 1)$ with τ increasing from 0 to 1.5 with increments 0.5. When $\tau = 0$, the model is an IMA(1,1) model with zero intercept. When $\tau > 0$, the model becomes a stationary first-order TARMA model that is increasingly distant from the IMA(1,1) model with increasing τ. As for the MA parameter we set $\theta = -0.9, -0.5, 0, 0.5, 0.9$. The empirical sizes of the tests are displayed in Table 1. Note that we have partitioned the set of 11 tests according to their nature: the first 9 are asymptotic and the last 2 are bootstrap tests. Clearly, the ADF, the KS, the BBC and the EG tests are severely oversized as θ approaches unity. Moreover, the wild bootstrap sLMb test is the only test that shows a correct size in all the settings, whereas both the sLM and the M class of tests show some bias, albeit small. Note that, when $\theta = 0$

the TARMA model reduces to a TAR model. In this case, the auxiliary model of the KS, BBC, EG tests is correctly specified and their size is correct; however, when θ becomes positive their size is severely biased and this raises issues concerning their practical utility.

Table 2: Size-corrected power of the asymptotic and bootstrap tests at nominal size $\alpha = 5\%$

$n = 300$	asymptotic									bootstrap	
τ ; θ	sLM	$\bar{\mathrm{M}}^{\mathrm{g}}$	M^{g}	$\mathrm{MP_T}$	ADF	$\mathrm{ADF}^{\mathrm{g}}$	KS	BBC	EG	sLMb	KSb
0.0;-0.9	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
0.5;-0.9	25.7	17.6	17.8	18.2	10.2	19.4	5.1	16.5	1.6	23.7	8.3
1.0;-0.9	52.5	26.7	26.9	27.7	15.8	30.4	17.4	31.9	3.8	54.3	27.0
1.5;-0.9	77.1	33.5	34.0	35.1	22.1	38.2	36.9	50.1	8.2	75.6	45.5
0.0;-0.5	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.1
0.5;-0.5	21.7	22.8	22.8	22.7	11.6	22.4	11.2	15.6	3.3	23.5	9.1
1.0;-0.5	48.3	34.5	34.9	34.9	18.2	35.0	32.3	31.1	8.2	47.8	29.1
1.5;-0.5	72.6	45.0	45.8	46.1	25.9	45.7	55.5	50.1	17.1	74.9	53.6
0.0;0.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.1	5.1
0.5;0.0	22.4	25.8	26.1	26.9	11.0	26.6	37.9	15.2	22.6	22.5	40.7
1.0;0.0	50.5	41.3	42.0	41.7	18.0	42.0	66.7	33.6	43.5	46.9	69.7
1.5;0.0	75.3	54.7	55.7	55.7	27.7	55.7	84.8	55.8	65.8	73.8	84.7
0.0;0.5	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.1	0.0
0.5;0.5	20.6	25.0	25.1	24.5	12.9	25.6	42.9	18.8	35.3	21.8	0.0
1.0;0.5	50.1	39.5	40.1	39.2	26.2	40.9	70.9	45.1	65.8	49.2	0.0
1.5;0.5	76.9	49.8	51.9	49.9	43.1	53.1	88.9	72.5	88.0	77.3	0.0
0.0;0.9	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	4.9	0.0
0.5;0.9	24.8	16.2	19.9	16.0	14.9	18.2	6.4	34.6	29.6	14.3	0.0
1.0;0.9	62.8	22.9	34.5	22.5	32.8	26.5	12.4	63.6	52.4	36.1	0.0
1.5;0.9	86.3	25.3	44.9	25.0	47.8	29.5	23.5	77.2	65.8	61.7	0.0

The size-corrected power of the tests is presented in Table 2. Here, the sample size is 300; see Section S5 of the Supplementary Material for results for $n = 100, 500$. The rows for $\tau = 0$ correspond to the size and other rows give size-corrected power. The size correction for bootstrap tests is achieved by calibrating the p-values. In some cases, the corrected size deviates from the nominal 5% due to discretization effects on the empirical distribution of bootstrap p-values. Clearly, the supLM tests are almost always more powerful than the other tests, especially as τ increases. For instance, when $\tau = 1.5$ the sLM test has almost double the power of M tests in several instances. As mentioned before, the case $\theta = 0$ (central panel) corresponds to a TAR model and this is one of two instances where the KS tests are slightly more powerful than the supLM tests. The power of the bootstrap version of the KS test is zero in three cases, due to its 100% oversize. See the Supplementary Material for further simulation results.

6.1 Measurement error and heteroskedasticity

In this section we assess the effect of measurement error and heteroskedasticity on the behaviour of the tests. We simulate from the following IMA(1,1) model

$$X_t = X_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \tag{6.30}$$

where $\theta = -0.9$ (model M1), -0.5 (model M2), 0.5 (model M3), 0.9 (model M4). We add measurement noise as follows

$$Y_t = X_t + \eta_t, \tag{6.31}$$

where the measurement error $\eta_t \sim N(0, \sigma^2_\eta)$ is such that the signal to noise ratio SNR $= \sigma^2_X/\sigma^2_\eta$ is equal to $\{+\infty, 50, 10, 5\}$. Here, σ^2_X is the variance of X_t computed by means of simulation. Since the variance in the non-stationary case depends upon the sample size n we have computed it on simulated trajectories for varying values of n as to be able to replicate it for the sample size in use. The case without noise (SNR $= +\infty$) is taken as the benchmark. The empirical sizes (rejection percentages) for models M1–M4 are presented in Table 3 for $n = 300$ and the results for $n = 100$ and 500 can be found in Section S5 of the Supplementary Material. Clearly, the measurement noise has little effect upon the size of the supLM tests. On the contrary, the size bias of the tests KS, BBC and EG increases appreciably when θ is positive (Models M3–M4). Worst still, the bias does not reduce when the sample size increases. The results shown in Section S6 of the Supplementary Material show that supLM tests are well behaved in the presence of heteroskedasticity and measurement error, with the sLMb wild bootstrap test being more so. The KS, BBC, and EG tests are severely affected by the combined presence of heteroskedasticity and measurement error and their size bias gets worse with increasing sample size. Also the sLM and sLMb are affected non-trivially. For instance, in Tables 6–8 of the Supplementary Material for Model M7 (integrated AR-GARCH) the two

Table 3: Empirical size (rejection percentage) at nominal $\alpha = 5\%$ and $n = 300$ for the IMA(1,1) models M1–M4 with increasing levels of measurement error.

		asymptotic									bootstrap	
	SNR	sLM	$\bar{\mathrm{M}}^{\mathrm{g}}$	M^{g}	$\mathrm{MP_T}$	ADF	$\mathrm{ADF}^{\mathrm{g}}$	KS	BBC	EG	sLMb	KSb
M1	∞	4.4	7.1	6.7	6.8	3.4	4.8	6.2	12.5	6.7	5.0	5.3
	50	3.6	6.4	6.4	5.4	4.8	4.9	5.8	10.4	6.5	3.8	4.7
	10	2.8	5.0	5.0	4.2	6.1	4.7	4.9	5.7	5.1	5.0	4.0
	5	5.5	5.3	5.0	4.8	5.2	4.8	4.9	3.1	3.1	5.3	3.8
M2	∞	4.0	6.4	6.2	5.6	5.6	5.4	4.2	5.5	5.6	5.2	3.4
	50	4.7	6.3	5.9	5.9	5.8	5.4	4.0	4.8	5.4	6.1	3.2
	10	5.9	6.3	6.1	5.1	6.6	5.3	3.6	4.1	4.6	6.4	2.3
	5	5.4	5.5	5.3	5.1	6.3	5.2	5.4	2.5	5.6	5.4	3.8
M3	∞	2.8	5.8	5.8	4.4	5.6	6.3	67.8	14.1	59.9	5.2	62.0
	50	3.4	5.6	5.7	4.2	5.7	5.9	68.2	15.7	60.9	5.0	62.7
	10	2.4	6.2	6.0	5.0	5.8	7.0	74.2	19.7	67.4	4.8	66.7
	5	2.5	5.6	5.5	4.2	5.2	6.7	84.3	28.4	77.6	5.3	76.7
M4	∞	6.1	1.2	2.1	0.9	86.6	15.4	100.0	98.5	100.0	4.3	99.5
	50	5.8	1.2	1.8	1.1	87.8	15.6	100.0	98.8	100.0	3.6	99.6
	10	4.2	2.5	2.7	1.4	89.8	17.1	100.0	99.3	100.0	2.2	99.9
	5	6.7	3.5	4.5	2.9	94.9	19.5	100.0	99.8	100.0	3.6	100.0

tests present a size that varies both with sample size and SNR but, overall, the tests are well behaved. The class of M tests is also robust in this respect but they can display low power in a number of instances, especially when the DGP is nonlinear. See also Chan et al. (2020).

7. A real application: testing the PPP hypothesis

In this section we apply our supLM tests to the post-Bretton Woods and pre-euro real exchange rates of a panel of European countries. Based on macroeconomic theory, there is some consensus on the fact that price gaps (measured in a common currency) for the same goods in different countries should rapidly disappear. However, empirical evidence points to a strong persistence and unit root tests generally fail to reject the null hypothesis of a random walk. As also pointed out in Taylor (2001) this can be ascribed to the incorrect linear specification for the price dynamics. The presence of trading costs implies that the mechanisms governing price adjustments are nonlinear and threshold autoregressive models provide a solution to the problem by allowing a "band of inaction" random walk regime, where arbitrage does not occur, and other regimes where mean reversion takes place so that the model is globally stationary (see Bec et al., 2004, and references therein for further discussion). For a review on how TAR models are used to analyse the exchange rates dynamics see also Hansen (2011). A critical investigation on the practical usefulness of combining unit-root tests and other stationarity tests in the PPP debate is put forward in Caner and Kilian (2001).

We consider the monthly $\log_{10}$ real exchange rates for the following countries: Portugal (PT), Germany (DE), France (FR), Belgium (BE), Austria (AT), Great Britain (GB), Netherland (NL), Italy (IT). The series range from 1973:09 to 1998:12 ($n = 304$) and are produced by the Bank of International Settlements (BIS) by taking the geometric weighted average

Table 4: Results for the set of unit root tests applied to the 8 monthly series of real exchange rates. The first two rows report the p-value for the supLM tests; the remaining rows show the checkmark ✓ if the test results significant at 1%.

	PT	DE	FR	BE	AT	GB	NL	IT
sLM	0.167	**0.002**	0.126	0.900	0.329	0.318	0.900	0.874
sLMb	0.384	**0.009**	0.292	0.833	0.417	0.259	0.802	0.836
$\bar{\mathrm{M}}^{\mathrm{g}}$	.	.	.	.	.	.	.	.
M^{g}	.	.	.	.	.	.	.	.
MP_{T}	.	.	.	.	.	.	.	.
ADF	.	.	.	.	.	.	.	.
$\mathrm{ADF}^{\mathrm{g}}$	.	.	.	.	.	.	.	.
KS	.	.	.	.	.	.	.	.
BBC	.	.	.	.	.	.	.	✓
EG	.	.	.	.	.	.	.	.

of a basket of bilateral exchange rates (27 economies), adjusted with the corresponding relative consumer prices. Such weights are constructed from manufacturing trade flows so as to encompass both third-market competition and direct bilateral trade through a double-weighting scheme. See Klau and Fung (2006) and https://www.bis.org/ for more details on the construction of the indexes.

Table 4 reports the results of the application of the battery of unit root tests described in the previous section on the 8 monthly series of real exchange rates. The first two rows show the p-values from our supLM tests, where the threshold search ranges from quantiles 15th to 85th of the data. Also, for the sLMb test, we chose 9999 bootstrap resamples and the Rademacher auxiliary distribution. To enhance readability, the remaining rows show a checkmark ✓ if the corresponding test rejects the null hypothesis at level 1%. Based upon our tests, we can reject the null hypothesis with some confidence for Germany (DE) (p-values in bold). Interestingly, all the other tests fail to reject and the finding is somehow consistent with that of Bec et al. (2004) where the authors rejected the null hypothesis for the pairwise real exchange rates of Germany versus France, Italy, Belgium, Netherland and Portugal. The BBC test leads to rejecting also for Italy but our tests do not and this might be due to the oversize of the latter. Moreover, as shown in Table 2, the M tests have very little power against some TARMA alternatives and this explains their failure to reject the null hypothesis. This result suggests an exploration to see if a TARMA model is plausible for the series for Germany. Hence, we fit the following TARMA(1,1) model

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1} X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{if } X_{t-1} > r \\ \phi_{2,0} + \phi_{2,1} X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{if } X_{t-1} \le r \end{cases} \quad (7.32)$$

In Figure 2(left) of the Supplementary Material we plot the values of the LM statistic T_r computed over a threshold grid that ranges from the 15th to the 85th percentiles of the data. The estimated threshold $\hat{r} = 4.700$, that maximizes T_r, is also the value that minimizes the AIC criterion over the same grid. In the right panel of the figure we plot the time series

Table 5: Parameter estimates from the TARMA(1,1) fit of Eq. (7.32) on the monthly real exchange rates for Germany (DE) with $\hat{r} = 4.700$.

	θ	$\phi_{1,0}$	$\phi_{1,1}$	$\phi_{2,0}$	$\phi_{2,1}$
estimate	0.31	-1.25	0.74	-0.15	0.97
s.e.	(0.06)	(0.28)	(0.06)	(0.09)	(0.02)

of the monthly real exchange rates for Germany, where we have indicated the selected threshold with a red line. The gray shaded area indicates the months associated with the upper regime. The parameter estimates are presented in Table 5 and point to a lower regime with a possible unit root and an upper regime where the slope is strictly smaller than 1. This is consistent with the idea of a nonlinear adjustment mechanism that activates when the rate crosses the threshold. Figure 2(right) of the Supplementary Material shows that the intervention regime is visited mostly before 1980 and after 1995. This is in general agreement with the results of Bec et al. (2004), as well as those that Bec et al. (2008) obtained on the real exchange rate series of French Franc against Deutsche Mark. The MA parameter θ greatly enhances the fitting ability of the model while retaining parsimony. This is witnessed by the diagnostics computed on the residuals that do not show any unaccounted dependence or deviation from normality, see Figure 3 and Figure 4, of the Supplementary Material.

8. Conclusion

In this paper, we argue that measurement errors are often neglected in the regulation/unit-root literature with serious consequences, and their ubiquity implies that to test for regulation in dynamics, it is more appropriate and perhaps even crucially important to formulate the test within a TARMA specification. We adopt the TARMA(1,1) model as the general hypothesis and the IMA(1,1) model as the null hypothesis. As far as we know, this is the first time that a TARMA specification is used in the present context although it was previously utilized in a very different context, namely for linearity testing under stationarity (Li and Li, 2011; Goracci et al., 2021). We derive a Lagrange multiplier test which is asymptotically similar given the threshold search range. Empirical studies confirm that the proposed approach enjoys much higher power in detecting regulation in dynamics than existing tests that do not address measurement errors. The surprisingly good size property of our tests may be owing to the versatility of an IMA(1,1) model in approximating general non-seasonal difference stationary processes. In particular, empirical results reported in Chan et al. (2020) and in the Supplementary Material indicate that, thanks to the wild bootstrap scheme, our new tests generally perform well under heteroskedasticity, even when the null hypothesis entails a non-stationary process different from the IMA(1,1) model, and remain powerful for other forms of regulation. Finally, the application of our tests to real exchange rates shows that TARMA models could well represent a modest step towards a positive resolution of the PPP puzzle.

Supplementary Material

The Supplementary Material (pdf format) contains all the proofs, further results from the real data analysis, the tabulated quantiles of the null distribution and further Monte Carlo investigations.

References

Bec, F., M. Ben Salem, and M. Carrasco (2004). Tests for unit-root versus threshold specification with an application to the purchasing power parity relationship. *J. Bus. Econom. Statist. 22*(4), 382–395.

Bec, F., A. Rahbek, and N. Shephard (2008). The ACR model: A multivariate dynamic mixture autoregression. *Oxford Bulletin of Economics and Statistics 70*(5), 583–618.

Billingsley, P. (1968). *Convergence of probability measure.* New York: Wiley.

Björk, T. (2019). *The Pedestrian's Guide to Local Time*, Chapter Chapter 3, pp. 43–67.

Brockwell, P. and R. Davis (2001). *Time series: theory and methods.* Springer.

Caner, M. and B. Hansen (2001). Threshold autoregression with a unit root. *Econometrica 69*(6), 1555–1596.

Caner, M. and L. Kilian (2001). Size distortions of tests of the null hypothesis of stationarity: evidence and implications for the PPP debate. *Journal of International Money and Finance 20*(5), 639 – 657.

Cavaliere, G. and A. Taylor (2008). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory 24*(1), 43–71.

Chan, K.-S. (1990, 12). Testing for threshold autoregression. *Ann. Statist. 18*(4), 1886–1894.

Chan, K.-S., S. Giannerini, G. Goracci, and H. Tong (2020). Unit-root test within a threshold ARMA framework. Technical report.

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. and H. Tong (2010). A note on the invertibility of nonlinear ARMA models. *J. Statist. Plann. Inference 140*(12), 3709–3714.

Davidson, R. and E. Flachaire (2008). The wild bootstrap, tamed at last. *J. Econometrics 146*(1), 162–169.

de Jong, R., C.-H. Wang, and Y. Bae (2007). Correlation robust threshold unit root tests. Mimeo, Ohio State University, Michigan.

Dickey, D. and W. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc. 74*(366a), 427–431.

Enders, W. and C. Granger (1998). Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. *J. Bus. Econom. Statist. 16*(3), 304–311.

Giannerini, S., G. Goracci, and A. Rahbek (2022). The validity of bootstrap testing in the threshold framework. Technical report, University of Bologna and University of Copenhagen.

Giordano, F., M. Niglio, and C. Vitale (2017). Unit root testing in presence of a double threshold process. *Methodol. Comput. Appl. Probab. 19*(2), 539–556.

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.

Goracci, G., S. Giannerini, K.-S. Chan, and H. Tong (2021). Testing for threshold effects in the TARMA framework. *Statistica Sinica in press*.

Hansen, B. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*(2), 413–430.

Hansen, B. (2011). Threshold autoregression in economics. *Stat. Interface 4*(2), 123–127.

Kapetanios, G. and Y. Shin (2006). Unit root tests in three-regime SETAR models. *Econom. J. 9*(2), 252–278.

Klau, M. and S. Fung (2006, March). The new BIS effective exchange rate indices. BIS quarterly review, Bank of International Settlements.

Kurtz, T. and P. Protter (1996). Weak convergence of stochastic integrals and differential equations. *Lecture notes in mathematics-Springer Verlag*, 1–41.

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

Liu, R. (1988, 12). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist. 16*(4), 1696–1708.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist. 21*(1), 255–285.

Ng, S. and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica 69*(6), 1519–1554.

Park, J. and M. Shintani (2016). Testing for a unit root against transitional autoregressive models. *Internat. Econom. Rev. 57*(2), 635–664.

Perron, P. and Z. Qu (2007). A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Econom. Lett. 94*(1), 12–19.

Seo, M. (2008). Unit root test in a threshold autoregression: asymptotic theory and residual-based block bootstrap. *Econometric Theory 24*(6), 1699–1716.

Su, F. and K.-S. Chan (2015). Quasi-likelihood estimation of a threshold diffusion process. *J. Econometrics 189*(2), 473–484.

Taylor, A. (2001). Potential pitfalls for the purchasing-power-parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price. *Econometrica 69*(2), 473–498.

Taylor, A. and M. Taylor (2004, December). The purchasing power parity debate. *Journal of Economic Perspectives 18*(4), 135–158.

- Kung-Sik Chan,
Department of Statistics and Actuarial Science, University of Iowa
263 Schaeffer Hall (SH), 20 East Washington Street, Iowa City, IA 52242-1409
E-mail: (kung-sik-chan@uiowa.edu)

- Simone Giannerini
Department of Statistical Sciences, University of Bologna
Via belle arti 41, 40126, Bologna, Italy
E-mail: (simone.giannerini@unibo.it)

- Greta Goracci
Faculty of Economics and Management, Free University of Bozen/Bolzano
piazza Università, 1, 39100, Bolzano, Italy
E-mail: (greta.goracci@unibz.it)

- 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)