Statistica Sinica **34** (2024), 1413-1434
doi:https://doi.org/10.5705/ss.202022.0125

TESTING FOR THRESHOLD REGULATION IN PRESENCE OF MEASUREMENT ERROR

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

[1] *University of Iowa,* [2] *University of Bologna,* [3] *Free University of Bozen/Bolzano,* [4] *University of Electronic Science and Technology of China,* [5] *Tsinghua University and* [6] *London School of Economics and Political Science*

Abstract: Regulation is an important feature of dynamic phenomena, and is commonly tested within the threshold autoregressive setting, with the null hypothesis being a global nonstationary process. Nonetheless, this setting is debatable, because 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 that has an asymptotically similar null distribution, and provide the first rigorous proof of tightness in the context of 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 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 fields such as economics, finance, biological growth, and population fluctuations, among others. Growth processes are generally regulation-free until they enter extreme phases. For instance, real exchange rates should be regulated through a threshold that triggers the mean reversion toward 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, for example, Taylor and Taylor (2004).

The random walk is a simple model for regulation-free dynamics. On the other hand, regulation from above (below) can be captured using a first-order threshold autoregressive (TAR) model that follows a random walk until the process crosses a certain threshold, above (below) which, mean-reversion takes place, while the process as a whole is *stationary.* In general, a nonlinear

*Corresponding author.

stationary process renders the impulse response to a random shock nonlinear and state-dependent, which is consequential, and could be leveraged in economic regulation. Hitherto, a standard approach for testing for dynamic regulation is to adopt the preceding threshold model as the general model, and then to test whether it reduces to a *global* random walk. This approach has received much attention in the literature (Enders and Granger (1998); Caner and Hansen (2001); Bec, Ben Salem and Carrasco (2004); Kapetanios and Shin (2006); Seo (2008); Park and Shintani (2016); de Jong, Wang and Bae (2007); Giordano, Niglio and Vitale (2017)). However, despite data almost always being corrupted by measurement errors, to the best of our knowledge, 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, that is, 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, TARMA(1,1), which is driven by an IMA(1,1) model in one of its two regimes. In Section S1 of the Supplementary Material we show that the TARMA$(1,1)$ model is approximately invariant with respect to data corruption by independent measurement errors, whereas the IMA$(1,1)$ model is exactly invariant with respect to the addition of measurement errors. Above all, we cannot over-emphasize the importance of the role of the moving average term in practical applications.

Just as ARMA models provide a parsimonious approximation to some long AR models, TARMA models may do so for some high-order TAR models, as noted in Goracci (2020, 2021). Thus, although the TARMA model holds substantial promise as a class of nonlinear time series models for exploring nonlinear dynamics it remains underexplored, partly because of a lack of progress in obtaining conditions on stationarity and ergodicity. Unlike in the AR-ARMA analogy, incorporating a moving-average part in a nonlinear framework poses major theoretical challenges, and has nontrivial implications for the probabilistic structure of the process. Chan and Goracci (2019) derive 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 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 nonstandard situation in the nonlinear time series context is well recognized, both in the TAR setting (Chan (1990); Hansen (1996); Giannerini, Goracci and Rahbek (2024)) and in the TARMA setting (Li and Li (2011); Goracci et al. (2023)). The supLM framework overcomes this problem. We derive its asymptotic distribution under both 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 in the context of 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 remainder of 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 the proposed supLM 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 related to 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 asymptotic supLM test and its wild bootstrap version, and compare them with that of numerous existing tests. Section 7 contains an empirical illustration, in which we apply the proposed tests to the pre-euro monthly real exchange rates of a set of European countries. All proofs are collected in the Supplementary Material, which also 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 TARMA model:

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

where $\phi_{2,1}$ is fixed at one, unless stated otherwise, the innovations $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) random variables with mean zero and variance σ^2, ε_t is independent of X_{t-j}, for $j \ge 1$, the delay d is a positive integer, taken as one, for simplicity, r is the real-valued threshold parameter, and the ϕ and θ are unknown coefficients. Later, we relax the assumption of i.i.d. innovations 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, and 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 $|\theta| < 1$, because this 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) show 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) $\phi_{1,1} = 1$ and $\phi_{1,0} > 0$. 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) provide 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 one of the following hold: (iii) $\phi_{1,1} = 1, \phi_{2,0} = 0, \phi_{1,0} \geq 0$; (iv) $\phi_{1,1} = 1, \phi_{2,0} < 0, \phi_{1,0} = 0$; and (v) $\phi_{1,1} < 1, \phi_{2,0} = 0$. If none of conditions (i)–(v) hold, 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

$$H: \quad X_t = \phi_0 + X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} + (\phi_{1,0} + \phi_{1,1}X_{t-1}) \times I(X_{t-1} \leq r), \tag{3.1}$$

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

$$H_0: \quad X_t = \phi_0 + X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, \tag{3.2}$$

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 a nontrivial modification of the test and its asymptotic distribution and is left for further research. However, the intercept terms on the two regimes of any competing stationary first-order TARMA model are 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 so are retained in the constrained TARMA model under H_0. We can test for threshold regulation from above 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. (2023)). Our approach is to develop a Lagrange multiplier test statistic for H_0, with the threshold parameter fixed initially at some r. Denote the test statistic as $T_n(r)$. Because r is unknown and absent under H_0, we 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 20th percentile and b the 80th percentile. Then, the overall test statistic results in $T_n = \sup_{r\in[a,b]} T_n(r)$. In addition to taking the supremum, other approaches, including integration, can be used to derive an overall test statistic.

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

$$\ell = -\log(2\pi\sigma^2) \times \frac{n}{2} - \sum_{t=1}^{n} \frac{\varepsilon_t^2}{2\sigma^2}, \tag{3.3}$$

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

$$\varepsilon_t = X_t - \{\phi_0 + X_{t-1} + (\phi_{1,0} + \phi_{1,1}X_{t-1}) \times I(X_{t-1} \leq r)\} + \theta\varepsilon_{t-1}, \tag{3.4}$$

with the unknown ε_0 set to zero. Note that 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 usually suppressed for simplicity. Let $\boldsymbol{\psi} = (\phi_0, \theta, \sigma^2, \phi_{1,0}, \phi_{1,1})^\intercal$, with its components denoted by ψ_j, for $j = 1, 2, \ldots, 5$, 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 a known threshold r. Partition the Fisher information matrix according to $\boldsymbol{\psi}_i$, for $i = 1, 2$ into

$$I_n(r) = \begin{pmatrix} I_{1,1,n}(r) & I_{1,2,n}(r) \\ I_{2,1,n}(r) & I_{2,2,n}(r) \end{pmatrix}. \tag{3.5}$$

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 is equal to

$$T_n(r) = \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2^\intercal}(r) \left\{ \hat{I}_{2,2,n}(r) - \hat{I}_{2,1,n}(r)\hat{I}_{1,1,n}^{-1}(r)\hat{I}_{1,2,n}(r) \right\}^{-1} \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2}(r), \tag{3.6}$$

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 $\hat{I}_{i,j,n}(r)$, for $1 \le i, j, \le 2$; 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 prespecified percentiles of the observed data; see Subsection 3.2 for further discussion

3.1. Gaussian likelihood estimation

In this sub-section, we describe estimation of the model parameters under the null hypothesis, and 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 $t > 1$,

$$\frac{\partial \varepsilon_t}{\partial \phi_0} = -1 + \theta \frac{\partial \varepsilon_{t-1}}{\partial \phi_0} = -\sum_{j=0}^{t-1} \theta^j, \tag{3.7}$$

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

$$\frac{\partial \varepsilon_t}{\partial \phi_{1,0}} = -I(X_{t-1} \le r) + \theta \frac{\partial \varepsilon_{t-1}}{\partial \phi_{1,0}} = -\sum_{j=0}^{t-1} \theta^j I\left(X_{t-1-j} \le r\right), \tag{3.9}$$

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

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 one time 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

$$\hat{\varepsilon}_t = X_t - X_{t-1} - \hat{\phi}_0 + \hat{\theta} \hat{\varepsilon}_{t-1}, \quad \forall t \ge 1, \tag{3.11}$$

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

$$\sum_{t=1}^{n} \frac{1}{\sigma^2} \frac{\partial \varepsilon_t}{\partial \psi_i} \frac{\partial \varepsilon_t}{\partial \psi_j} + \sum_{t=1}^{n} \frac{\varepsilon_t}{\sigma^2} \frac{\partial^2 \varepsilon_t}{\partial \psi_i \partial \psi_j} = (1 + o_p(1)) \times \sum_{t=1}^{n} \frac{1}{\sigma^2} \frac{\partial \varepsilon_t}{\partial \psi_i} \frac{\partial \varepsilon_t}{\partial \psi_j}, \tag{3.12}$$

its $(3, i)$th element with $i \neq 3$ is equal to $\sum_{t=1}^{n} (\varepsilon_t/\sigma^4)(\partial \varepsilon_t/\partial \psi_i) = o_p(n)$, and the $(3, 3)$th element is equal to $\sum_{t=1}^{n} \left((1/(2\sigma^4)) - (\varepsilon_t^2/\sigma^6)\right)$, 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 (from arguments similar to those in the proof of Theorem 1), and thus are omitted henceforth. We sometimes write, for example, $\partial \ell/\partial \psi_j(\boldsymbol{\psi}; r)$, to highlight the role of the arguments, and 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 $\partial \ell/\partial \boldsymbol{\psi}_1$, respectively, because they do not depend on r. With an abuse of notation, the true values of the moving-average coefficient and the innovation variance under H_0 are denoted simply by θ and σ^2, respectively. There should be no confusion because the context will make clear whether they represent 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 are 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, $\{n^{-1/2} X_{[sn]}, 0 \leq s \leq 1\}$, 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, Thm. 3.1 and 3.2)) that the Brownian local time $\{L_t^x, t \geq 0, -\infty < x < \infty\}$, defined as:

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

where $\operatorname{sign}(x)$ 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,

$$\int_0^1 f(W_s) ds = \int_{-\infty}^{\infty} f(x) L_1^x dx. \tag{3.13}$$

Thus, any quantile of $\{X_t, t = 0, \ldots, n\}$ is asymptotically equal to $n^{1/2}(1 - \theta)\sigma$ times the corresponding quantile of $\{W_s, 0 \leq s \leq 1\}$. Because the Brownian local time process is a random process, 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,

$$n^{-1/2}\frac{\partial \ell}{\partial \phi_{1,0}}(r_n) = n^{-1/2}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\frac{1}{1-\theta B}\left\{I\left(\frac{X_{t-1}}{n^{1/2}(1-\theta)\sigma} \le \tau\right)\right\}. \tag{3.14}$$

The right side of (3.14) 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 equal to $\sum_{j=0}^{t-1}\theta^j I\left(\{n^{1/2}(1-\theta)\sigma\}^{-1}X_{t-1-j} \le \tau\right)$ 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 converges 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 $n \to \infty$,

$$n^{-1/2}\frac{\partial \ell}{\partial \phi_{1,0}}(r_n) \rightsquigarrow \frac{1}{(1-\theta)\sigma}\int_0^1 I(W_s \le \tau)dW_s. \tag{3.15}$$

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

$$\begin{aligned} n^{-1}\frac{\partial \ell}{\partial \phi_{1,1}}(r_n) &= n^{-1/2}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma}\frac{1}{1-\theta B}\left[\frac{X_{t-1}}{n^{1/2}\sigma}I\left\{\frac{X_{t-1}}{n^{1/2}(1-\theta)\sigma} \le \tau\right\}\right] \\ &\rightsquigarrow \int_0^1 W_s I(W_s \le \tau)dW_s \end{aligned} \tag{3.16}$$

$$n^{-1/2}\frac{\partial \ell}{\partial \phi_0} = n^{-1/2}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\frac{1}{1-\theta B}(1) \rightsquigarrow \frac{1}{(1-\theta)\sigma}\int_0^1 dW_s = \frac{W_1}{(1-\theta)\sigma}. \tag{3.17}$$

Note the different rates of normalization. Let K_n be a 5×5 diagonal matrix with the last diagonal elements being n, and the 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. (3.5). 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,

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

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

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 shown below.

Remark 1. 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. Our 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 that delivers valid inferences under heteroskedastic disturbances (Liu (1988); Mammen (1993); Davidson and Flachaire (2008)). As 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 nonstationary 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 using 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 is equal to ± 1 with equal probability.
4. Obtain the bootstrap resample $\hat{X}_t^* = \sum_{j=1}^{t}(\hat{e}_j^* - \hat{\theta}\hat{e}_{j-1}^*)$, and compute the supLM statistic T_n^* upon it.
5. Repeat steps 3–4 B times to obtain the bootstrap test statistic, T_n^{*b}, for $b = 1, \ldots B$, 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 a zero intercept. Using the second-order Taylor expansion and after some routine algebra, it holds that

$$\frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2}(r_n) \approx \frac{\partial \ell}{\partial \boldsymbol{\psi}_2}(r_n) - I_{2,1,n}(r_n) I_{1,1,n}^{-1} \frac{\partial \ell}{\partial \boldsymbol{\psi}_1}. \tag{4.1}$$

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 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,

$$\begin{aligned} Q_n \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2}(r_n) &= Q_n \frac{\partial \ell}{\partial \boldsymbol{\psi}_2}(r_n) - I_{2,1}(\tau) I_{1,1}^{-1} P_n \frac{\partial \ell}{\partial \boldsymbol{\psi}_1} + o_P(1) \\ &= Q_n \frac{\partial \ell}{\partial \boldsymbol{\psi}_2}(r_n) - \tilde{I}_{2,1}(\tau) \tilde{I}_{1,1}^{-1} P_n \frac{\partial \ell}{\partial \phi_0} + o_P(1), \end{aligned} \tag{4.2}$$

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

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

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

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

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

$$Q_n \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2}(r_n) = Q_n \frac{\partial \ell}{\partial \boldsymbol{\psi}_2}(r_n) - \tilde{I}_{2,1}(\tau) P_n^{-1}(\hat{\phi}_{0,n} - \phi_0) + o_p(1). \tag{4.3}$$

Let

$$H(\tau) = \left(\int_0^1 dW_s, \int_0^1 I(W_s \le \tau) dW_s, \int_0^1 W_s I(W_s \le \tau) dW_s \right)^{\mathsf{T}} \tag{4.4}$$

and

$$\Lambda(\tau) = \begin{pmatrix} 1 & \int_0^1 I(W_s \le \tau) ds & \int_0^1 W_s I(W_s \le \tau) ds \\ \int_0^1 I(W_s \le \tau) ds & \int_0^1 I(W_s \le \tau) ds & \int_0^1 W_s I(W_s \le \tau) ds \\ \int_0^1 W_s I(W_s \le \tau) ds & \int_0^1 W_s I(W_s \le \tau) ds & \int_0^1 W_s^2 I(W_s \le \tau) ds \end{pmatrix}. \tag{4.5}$$

Let $\Lambda(\tau)$ be partitioned into a 2×2 block matrix with the $(2,2)$th block being 2×2. Then, $H(\tau) = (H_1(\tau), H_2(\tau))^{\mathsf{T}}$ is partitioned similarly. It follows from Eq. (4.2) and Eqs. (3.15)–(3.17) that the asymptotic null distribution of $T_n(r_n)$ is 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, under the following assumption:
(**A1**): Let $r_L < r_U$ be two fixed real numbers. Let

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

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

$$E\left\{ |\mathcal{T}_n(\tau_2) - \mathcal{T}_n(\tau_1)|^4 \right\} \leq C\left(|\tau_2 - \tau_1|^{3/2} + \frac{|\tau_2 - \tau_1|}{n} \right), \tag{4.6}$$

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

$$|\mathcal{T}_n(\tau_2) - T_n(\tau_1)| \leq K \times L(n)(n \log\log n)^{1/2} |\tau_2 - \tau_1| + o_p(1) \tag{4.7}$$

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; that is, for any $\lambda > 0, L(\lambda x)/L(x) \to 1$ as $x \to \infty$.

Theorem 1. *Suppose H_0 holds so that $\{X_t, t = 0, 1, \ldots,\}$ is an IMA(1,1) process satisfying Eq.* (3.2)*, with the intercept $\phi_0 = 0$, $|\theta| < 1$, and the innovations are i.i.d. 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*

$$F(W; r_L, r_U) = \sup_{\tau \in [r_L, r_U]} \left\| \left[\{\Lambda^{-1}(\tau)\}_{2,2} \right]^{1/2} \{H_2(\tau) - \Lambda_{2,1}(\tau) H_1(\tau)\} \right\|^2, \tag{4.8}$$

which has a parameter-free distribution, although it depends on the search range of the threshold.

Note that the assumption of i.i.d. 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 ε_{t-s}, for $s \leq 0$; the proof is essentially the same.

Remark 2. Conditions (4.6)–(4.7) 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 ideal 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 a deterministic threshold search interval. It can be extended readily to the case that the end points are fixed quantiles of the data, which are realization specific. We omit the proof, because it is based on routine analysis that builds on Theorem 1 and that, for any fixed $0 < p < 1$, (i) the p-quantile of $\{W_s, 0 \le s \le 1\}$ is $O_p(1)$, following 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 $\{W_s, 0 \le s \le 1\}$ times $n^{1/2}(1-\theta)\sigma$; see the discussion below (3.13).

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

Theorem 2. *Conditions* (4.6) *and* (4.7) *hold if* (i) $|\theta| < 1$ *and* (ii) $\{\varepsilon_t\}$ *are independent and identically normally distributed with zero mean and finite positive variance.*

Because 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 one with increasing departure in some direction from the null hypothesis. The mathematical framework is as follows. For each positive integer n, the system of hypotheses 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

$$X_t = \begin{cases} n^{-1/2}h_{1,0} + (1 + n^{-1}h_{1,1})\, X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} & \text{if } \dfrac{X_{t-1}}{\sigma n^{1/2}(1-\theta)} \le \tau_0 \\ n^{-1/2}h_{2,0} + (1 + n^{-1}h_{2,1})\, X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} & \text{if } \dfrac{X_{t-1}}{\sigma n^{1/2}(1-\theta)} > \tau_0, \end{cases} \tag{5.1}$$

where $\mathbf{h} = (h_{1,0}, h_{2,0}, h_{1,1}, h_{2,1})^\intercal$ is a fixed vector, with $h_{i,1} \le 0$, for $i = 1, 2$, 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 a sufficiently large n. In order to derive the local power, we henceforth impose the following mild regularity conditions:

C1: The innovations are assumed to be i.i.d., with a zero mean, finite positive standard deviation, σ, and probability density function $f(\cdot/\sigma)/\sigma$, where (i) f is a bounded function, and $\log(f(x))$ 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 zero, 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: $-\pi/2 < h_{1,1}, h_{2,1} \leq 0$, 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 Student's 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 $(\dot{f}/f)(\varepsilon_t)$, that is, $\rho = \int x\dot{f}(x)dx/\sqrt{\mathcal{I}_f}$, 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 $n \to \infty$, $T_n = \sup\{T_n(r)$, for $r \in [n^{1/2}(1-\theta)\sigma r_L, n^{1/2}(1-\theta)\sigma r_U]\}$, where r_L, and r_U are two fixed numbers, converges in distribution to $F(W; r_L, r_U)$ defined in Eq.* (4.8) *but with W now being a threshold diffusion process satisfying the following stochastic differential equation (SDE):*

$$dW_s = dW_s^\dagger + \begin{cases} \rho\sqrt{\mathcal{I}_f}\,[h_{1,0}/\{\sigma(1-\theta)\} + h_{1,1}W_s]\,ds, & \text{if } W_s \leq \tau_0, \\ \rho\sqrt{\mathcal{I}_f}\,[h_{2,0}/\{\sigma(1-\theta)\} + h_{2,1}W_s]\,ds, & \text{otherwise}, \end{cases} \tag{5.2}$$

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

Henceforth, in this section, W denotes the threshold diffusion satisfying Eq. (5.2). Note that if $h_{i,0} = h_{i,1} = 0$, for $i = 1, 2$, then we get the limiting null distribution for T_n. Otherwise, W is a threshold diffusion process (Su and Chan (2015)). Thus, the building block W that determines 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 has the power to detect local threshold alternatives. Because the functional $F(\cdot; r_L, r_U)$ is quite complex, in Section S2.4 of the Supplementary Material, we provide an example that demonstrates the consistency of the proposed test.

6. Finite-Sample Performance

To better approximate the finite-sample distribution of T_n, we simulated the null distributions for the sample sizes in use. Moreover, because the finite-sample distribution of T_n changes appreciably only when $|\theta|$ is close to one, we adopted the following conservative approach: if $|\hat{\theta}| > 0.3$, we use the quantiles of the

Table 1. Rejection percentages from the TARMA model of Eq.(6.1), with nominal size at $\alpha = 5\%$. Sizes over 15% are highlighted in bold.

	asymptotic									bootstrap	
θ	sLM	$\bar{\mathrm{M}}^{\mathrm{g}}$	M^{g}	$\mathrm{MP_T}$	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**

simulated null with $\theta = \mathrm{sign}(\hat{\theta}) \cdot 0.9$. Furthermore, we add a wild bootstrap scheme (see Section 3.3) to improve the empirical size of the test. We denote our asymptotic test and its wild bootstrap version as sLM and sLMb, respectively.

We compare the empirical performance of the proposed test with that of several competing tests, namely, those designed for threshold alternatives, and those without a specific nonlinear alternative. The former tests include those proposed by Kapetanios and Shin (2006) (KS), Enders and Granger (1998) (EG), and Bec, Ben Salem and Carrasco (2004) (BBC), with their bootstrap variants (if implemented) denoted as KSb, and so on. 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 $\bar{\mathrm{MP}}_{\mathrm{T}}^{\mathrm{GLS}}$ test of Ng and Perron (2001) ($\mathrm{MP_T}$) 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}). Note that we report only the results for the best performing tests.

The sample sizes considered are 100, 300, and 500. The rejection percentages are derived with a nominal size $\alpha = 5\%$, and are based upon 10,000 replications. In order to reduce the computational burden, for the bootstrap tests, we use 1,000 replications and $B = 1000$ bootstrap resamples. The threshold search ranges from 25% to 75% of the sample distribution. We simulate data from the following first-order TARMA model:

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

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1}X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{if } X_{t-1} \leq 0, \\ \phi_{2,0} + \phi_{2,1}X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{otherwise}, \end{cases} \tag{6.1}$$

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 zero to 1.5 with increments of 0.5. When $\tau = 0$, the model is an IMA(1,1) model with a zero intercept. When $\tau > 0$, the model becomes a stationary first-order TARMA model that becomes increasingly distant from the IMA(1,1) model with increasing τ. 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 nine are asymptotic, and the last two are bootstrap tests. Clearly, the ADF, KS, BBC and EG tests are severely oversized as θ approaches unity. Moreover, the wild bootstrap sLMb test is the only test to show a correct size in all 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 models of the KS, BBC, EG tests are specified correctly and their sizes are correct; however, when θ becomes positive, their sizes are severely biased, raising concerns about their practical utility.

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 the results for $n = 100, 500$. The rows for $\tau = 0$ correspond to the size, and the other rows give the size-corrected power. The size correction for the bootstrap tests is achieved by calibrating the p-values. In some cases, the corrected size deviates from the nominal 5% because of discretization effects on the empirical distribution of the 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 the 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 in which 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, owing 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 errors and heteroskedasticity on the behavior of the tests. We simulate from the following IMA(1,1) model:

$$X_t = X_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t, \tag{6.2}$$

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

$$Y_t = X_t + \eta_t, \tag{6.3}$$

where the measurement error $\eta_t \sim N(0, \sigma_\eta^2)$ is such that the signal-to-noise ratio SNR $= \sigma_X^2/\sigma_\eta^2$ is equal to $\{+\infty, 50, 10, 5\}$. Here, σ_X^2 is the variance of X_t computed by means of simulation. Because the variance in the non-stationary case depends on the sample size n, we compute it on simulated trajectories for varying values of n 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 S6 of the Supplementary Material. Clearly, the measurement noise has little effect on the size of the supLM tests. In contrast, the size bias of the KS, BBC, and EG tests 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 the supLM tests are well behaved in the presence of heteroskedasticity and measurement errors, particularly the sLMb wild bootstrap test. The KS, BBC, and EG tests are severely affected by the combined presence of heteroskedasticity and measurement errors and their size bias gets worse as the sample size increases.

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	$\mathrm{ADF^g}$	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

In addition, the sLM and sLMb tests, are affected nontrivially. For instance, in Tables 6–8 of the Supplementary Material for Model M7 (integrated AR-GARCH), the two tests present a size that varies with both the sample size and the SNR. However, overall, the tests are well behaved. The class of M tests is also robust in this respect, but 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, in general, unit-root tests fail to reject the null hypothesis of a random walk. As noted 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. Threshold models provide a solution to the problem by allowing a "band of inaction" random walk regime, where arbitrage does not occur, and other regimes in which mean reversion takes place so that the model is globally stationary; see Bec, Ben Salem and Carrasco (2004) and references therein for further discussion. For a review on how TAR models

Table 4. Results for the set of unit-root tests applied to the eight monthly series of real exchange rates. The first two rows report the p-values for the supLM tests; the remaining rows show a checkmark ✓if the test are 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
$\bar{M}^g$	.	.	.	.	.	.	.	.
M^g	.	.	.	.	.	.	.	.
MP_T	.	.	.	.	.	.	.	.
ADF	.	.	.	.	.	.	.	.
ADF^g	.	.	.	.	.	.	.	.
KS	.	.	.	.	.	.	.	.
BBC	.	.	.	.	.	.	.	✓
EG	.	.	.	.	.	.	.	.

are used to analyze exchange rates dynamics, see also Hansen (2011). Caner and Kilian (2001) provide a critical investigation on the practical usefulness of combining unit-root tests and other stationarity tests in the PPP debate.

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

Table 4 reports the results of applying the battery of unit-root tests described in the previous section on the eight monthly series of real exchange rates. The first two rows show the p-values from our supLM tests, where the threshold search ranges from the 15th to the 85th quantile of the data. Furthermore, for the sLMb test, we chose 9,999 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 the 1% level. Based on our tests, we can reject the null hypothesis with some confidence for Germany (DE) (p-values in bold). Interestingly, the other tests all fail to reject the null hypothesis, a finding that is somehow consistent with that of Bec, Ben Salem and Carrasco (2004), where the authors rejected the null hypothesis for the pairwise real exchange rates of Germany versus France, Italy, Belgium, the Netherland and Portugal. The BBC also rejects for Italy but our tests do not, possibly because of the oversize of the former. Moreover, as shown in Table 2, the M tests have

Table 5. Parameter estimates from the TARMA(1,1) fit of Eq. (7.1) 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)

very little power against some TARMA alternatives, which explains their failure to reject the null hypothesis. This result suggests that we should determine whether a TARMA model is plausible for the series for Germany. Hence, we fit the following TARMA(1,1) model:

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1}X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{if } X_{t-1} > r \\ \phi_{2,0} + \phi_{2,1}X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{if } X_{t-1} \leq r. \end{cases} \tag{7.1}$$

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 also minimizes the AIC over the same grid. In the right panel of the figure, we plot the time series of the monthly real exchange rates for Germany, where we indicate the selected threshold using a red line. The gray shaded area indicates the months associated with the upper regime. The parameter estimates are presented in Table 5, pointing to a lower regime with a possible unit-root, and an upper regime in which the slope is strictly smaller than one. 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, Ben Salem and Carrasco (2004), obtained on the real exchange rate series of the French franc against the 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. Furthermore, 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. To the best of our knowlwdge, this is the first time that a TARMA specification has been used in the present context, although it has been used for linearity testing under stationarity (Li and Li (2011); Goracci et al. (2023)). We derive a Lagrange multiplier test that is asymptotically similar, given the threshold search range. Empirical studies confirm that the proposed approach enjoys much higher power in terms of detecting regulation in dynamics than that of existing tests that do not address measurement errors. The surprisingly good size property of our tests may be because of the versatility of the IMA(1,1) model in approximating general nonseasonal difference stationary processes. In particular, the empirical results reported in Chan et al. (2020) and in the Supplementary Material indicate that, owing to the wild bootstrap scheme, our new tests perform well under heteroskedasticity, in general, even when the null hypothesis entails a nonstationary process different from the IMA(1,1) model, and remain powerful for other forms of regulation. Finally, an application of our proposed tests to real exchange rates shows that TARMA models could represent a modest step toward a positive resolution of the PPP puzzle.

Supplementary Material

The online Supplementary Material contains all proofs, further results from the real-data analysis, the tabulated quantiles of the null distribution, and additional Monte Carlo investigations. The routines for the sLM tests are included in the R package `tseriesTARMA` (Giannerini and Goracci (2023)), publicly available at `https://cran.r-project.org/package=tseriesTARMA`.

References

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

Billingsley, P. (1968). *Convergence of Probability Measure.* Wiley, New York.

Björk, T. (2019). *The Pedestrian's Guide to Local Time*, 43–67. World Scientific.

Brockwell, P. and Davis, R. (2001). *Time Series: Theory and Methods.* Springer.

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

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

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

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

Chan, K.-S., Giannerini, S., Goracci, G. and Tong, H. (2020). Unit-root test within a threshold ARMA framework. Technical report. University of Iowa, University of Bologna, London School of Economics.

Chan, K.-S. and Goracci, G. (2019). On the ergodicity of first-order threshold autoregressive moving-average processes. *J. Time Series Anal.* **40**, 256–264.

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

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

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

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

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

Giannerini, S. and Goracci, G. (2023). *tseriesTARMA: Analysis of Nonlinear Time Series Through TARMA Models*. R package version 0.3-2.

Giannerini, S., Goracci, G. and Rahbek, A. (2024). The validity of bootstrap testing in the threshold framework. *J. Econometrics* **239**, 105379. Web: `https://doi.org/10.1016/j.jeconom.2023.01.004`.

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

Goracci, G. (2020). Revisiting the Canadian Lynx time series analysis through TARMA models. *Statistica* **80**, 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., Giannerini, S., Chan, K.-S. and Tong, H. (2023). Testing for threshold effects in the TARMA framework. *Statist. Sinica* **33**, 1879–1901.

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

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

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

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

Kurtz, T. and Protter, P. (1996). *Weak Convergence of Stochastic Integrals and Differential Equations*, 1–41. Springer, Berlin, Heidelberg.

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

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

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

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

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

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

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

Su, F. and Chan, K.-S. (2015). Quasi-likelihood estimation of a threshold diffusion process. *J. Econometrics* **189**, 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**, 473–498.

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

Kung-Sik Chan

Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242-1409, USA.

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

Simone Giannerini

Department of Statistical Sciences, University of Bologna, 40126, Bologna, Italy.

E-mail: simone.giannerini@unibo.it

Greta Goracci

Faculty of Economics and Management, Free University of Bozen/Bolzano, 39100, Bolzano, Italy.

E-mail: greta.goracci@unibz.it

Howell Tong

Department of Statistics, London School of Economics and Political Science, London WC2A 2AE, UK.

Center for Statistical Science, Tsinghua University, Beijing 100084, China.

E-mail: howell.tong@gmail.com

(Received April 2022; accepted November 2022)