Adaptive Consistent Unit Root Tests Based on Autoregressive Threshold Model [*]

Frédérique Bec [†]
Alain Guay [‡]
Emmanuel Guerre [§]

November 2002 (first version: February 2002)

[*]Financial supports from FCAR, CREST and LSTA are gratefully acknowledged. Previous versions were presented at the SCSE 2002 conference, macroeconomic seminar of CREST, ESEM 2002 and EMM first annual conference. We benefited from discussions with D. Andrews, M. Carrasco, G. Laroque.

[†]CREST–ENSAE, Timbre J120 — 3, av. Pierre Larousse, 92245 Malakoff Cedex, France. Email: bec@ensae.fr
[‡]CIRPEE, Université du Québec à Montréal, Canada. Email: guay.alain@uqam.ca
[§]CREST-LS and LSTA,Université Paris 6, France. Email: eguerre@ccr.jussieu.fr

Abstract. We develop adaptive consistent unit root tests based on a three-regime threshold autoregression specification as auxiliary model. Under the null, the regimes are not identified and we eliminate the threshold by considering maximum statistics over a set of admissible parameters. The originality of the approach consists in the treatment of this set. We allow the threshold level to remain bounded under the null and unbounded under the alternative. Compared to previous approaches, this adaptive choice improves the power of the test by producing smaller critical values under the null and inspecting a wider set of thresholds under the alternative. We derive the null limit distribution of the procedures and show then that the tests are consistent against a wide class of stationary process. Hence, our tests are specifically designed to detect processes which are globally stationary but locally nonstationary. A Monte-Carlo study compares the finite sample properties of the proposed test to some existing unit root tests. We apply the unit root test to yield spread of interest rates for France, Germany, New-Zealand and US post-1980 monthly data.

JEL Classification: C12, C22, C32, E43

Keywords:Unit root test, threshold autoregressive model, term structure.

Résumé. On propose un test de racine unité, adaptatif et consistant, basé sur un modèle à seuil à trois régimes. Sous l'hypothèse nulle, les régimes ne sont pas identifiés, et le paramètre de seuil est éliminé en prenant le maximum de statistiques. L'originalité de l'approche développée ici tient à l'ensemble aléatoire sur lequel ce maximum est calculé: les seuils doivent rester borné sous la nulle mais peuvent diverger sous l'alternative. On obtient la loi limite de ces statistiques de tests et on établit leur consistance contre une large classe d'alternatives stationnaires. Les tests proposés sont tout particulièrement adaptés à des alternatives stationnaires qui peuvent être localement non stationnaire. Les propriétés de ces tests sont étudiés par simulation. On applique ce nouveau test de racine unité à l'écart de taux d'intérêt pour la France, l'Allemagne, la Nouvelle-Zélande et les Etats-Unis, sur des données mensuelles postérieures à 1980.

1 Introduction

Several economic phenomena can potentially imply nonlinear dynamics. The presence of fixed costs of adjustment or transaction costs create asymmetric adjustment in economic variables. Economic policy characterized by discrete intervention to manage for instance exchange rate target zone, inflation or output gap targeting could also induce nonlinear dynamic in economic variables. Recent studies by e.g. Enders and Granger [1998], Enders and Siklos [2001] and Taylor [2001] pointed out that standard unit-root and cointegration tests all have low power in presence of nonlinear adjustment which may explain the frequent failure to reject nonstationarity for series such that the real exchange rates or yield spread of interest rates. A proper unit-root or cointegration test must allow for asymmetric adjustment under the alternative. This is precisely the issue tackled by Enders and Granger [1998] and Enders and Siklos [2001], who propose to extend the DF unit-root test and the Engle-Granger cointegration test respectively by permitting a two-regime threshold auto-regressive (TAR) specification under the alternative. However, if non-linearity arises from, for instance, transaction costs as advocated by Anderson [1997], the relevant alternative specification should be a three-regime threshold auto-regressive model, so as to account for the "inaction band" in the middle regime. A few recent studies, like the ones by Caner and Hansen [2001], Gonzalez and Gonzalo [1998] and Shin and Lee [2001] have examined models where the threshold variable is stationary and differs from the dependent variable under the alternative. In general, testing unit-root versus a threshold alternative where the switching variable is the same as the dependent variable is in fact much more appealing.

In this paper, we develop adaptive consistent unit root tests based on a parsimonious three-regime threshold autoregressive specification as auxiliary model. As often proposed when a nuisance parameter appears under the null, we eliminate the threshold parameter by considering the maximum of standard statistics over a set of possible threshold values Λ_T, where T is the sample size. Previous works use a quantile based set of thresholds which can diverge under the null of a unit root and remain bounded under the alternative. The originality of our approach comes from the treatment of these threshold values through an adaptive choice of Λ_T: we depart from the previous studies by allowing the threshold levels in the random interval Λ_T to remain bounded under the null and unbounded under the alternative and build such a test. Construction of such threshold sets can be done via a consistent test statistic entering in the upper and lower bounds of Λ_T. Because the test relies on a maximum over Λ_T, the critical values of our procedure are expected to be smaller, with a larger set of thresholds inspected under the alternative, thus increasing the power compared to usual quantile based choices. An important consequence of our choice of the threshold values Λ_T is that we can show the consistency of the tests against a wide class of stationary processes, an issue which is seldom studied in the literature. Hence, the choice of a three-regime threshold autoregressive model yields a test specifically designed to detect stationary processes which may behave locally as nonstationary ones, as considered in e.g. Anderson [1997], Bec, Ben Salem and Carrasco [2001],Taylor [2001], Gouriéroux and Robert [2002], or Rahbek and Shephard [2002] .

One important contribution of the paper is to derive the asymptotic distribution without assuming that the threshold goes asymptotically to infinity under the null. The finite dimensional limit distribution is derived through the machinery of Park and Phillips [2001] introduced to study sums of nonlinear integrable function of integrated variables. The asymptotic equicontinuity of the statistics indexed by the threshold parameter is proven in order to find the limit distribution of our maximum test statistics. The asymptotic equicontinuity follows from moment bounds for the increments of some piecewise continuous processes indexed by the threshold parameter and a maximal inequality. This new result is central to derive the asymptotic distribution of unit root test based on threshold specification such as the ones developed by Bec et al. [2001], Berben and van Dijk [1999] and Enders and Granger [1998] when the threshold is randomly chosen.

A Monte-Carlo study compares the finite sample behavior of the consistent unit root test based on the threshold adjustment specification with the linear specification retained by the augmented Dickey-Fuller (ADF) test. First, the alternative model corresponds to the auxiliary model used to perform the unit root test e.g. the threshold adjustment specification. For this experiment, we also compare our test with the test developed by Bec et al. [2001] who consider a supremum over a quantile set. Second, we assess the finite sample consistency of the test against other stationary processes. A linear stationary model, a piecewise stationary model with a changing mean and an autoregressive conditional model proposed by Rahbek and Shephard [2002] (see also Gouriéroux and Robert [2002]) are considered under the alternative. The results are really encouraging for the proposed nonlinear unit root tests. For a vast majority of the cases, the supremum Wald (supWald) version of our tests outperforms the ADF test.

The supWald test is applied to the yield spread of interest rates. The well-know expectations theory of the term structure asserts that under costless and instantaneous portfolio adjustment assumption, equilibrium interest rates are such that the investor is indifferent between holding a bond which has k periods left to maturity and investing in a sequence of one-period bond for k successive periods. This non-arbitrage condition in turn implies that the long-term and short-term interest rates are cointegrated with weights $(1 - 1)$, or equivalently that the yield spread — defined as the difference between the $k-$period and the one-period interest rates — is stationary. However, the empirical evidence of cointegration between yields of different maturities is still not clear-cut, as surveyed in Pagan, Hall and Martin [1996]: for instance, Campbell and Shiller [1987], Stock and Watson [1988] or Anderson [1997] find a stationary US spread whereas more recent studies by Enders and Siklos [2001] or Bohl and Siklos [2001] fail to reject the null of no-cointegration between long and short term interest rates for the US and Germany respectively. When applying our unit-root test on post-1980 monthly data for French, German, New Zealander and US yield spreads, the null of unit-root is rejected whereas the ADF test fails to reject it.

The remainder of the paper is organized as follows. Section 2 presents our test statistics and groups our main theoretical findings. Section 3 is devoted to Monte-Carlo illustrations of the properties of the SupWald test. Section 4 briefly exposes the implication of transaction costs for the yield spread process according to the expectations theory of the term structure. This section also reports our empirical results and Section 5 concludes. Proofs are gathered in the appendix.

2 The Unit-root tests

First, we set out some important features of three-regime threshold autoregressive models and present our test statistics. The null hypothesis of interest is

$$H_0 : \Delta y_t = a(L)\Delta y_{t-1} + \varepsilon_t$$

where $1-a(L)$ is a lag polynomial function with a known order p and roots lying outside the unit circle, and where $\{\varepsilon_t\}$ is a sequence of centered i.i.d. random variables with variance σ^2. It is now widely documented that tests of H_0 based upon an auxiliary linear model such as the ADF test statistic suffer from a lack of power against some families of nonlinear alternatives. We therefore propose a test based on a *parsimonious auxiliary* self exciting Threshold AutoRegressive (TAR) model

$$\Delta y_t = a(L)\Delta y_{t-1} + \varepsilon_t + \begin{cases} \mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \leq -\lambda \\ \mu_2 + \rho_2 y_{t-1} & \text{if } |y_{t-1}| < \lambda \\ -\mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \geq \lambda \end{cases} . \qquad (2.1)$$

This model differs from Caner and Hansen [2001] who consider a TAR model with three regimes and a stationary switching variable. Berben and van Dijk [1999] and Enders and Granger [1998] consider a two-regime TAR model with switching variable y_{t-1}. Such a choice also differs from three regime models considered in previous works. Our model is more parsimonious than Balke and Fomby [1997] who consider a different dynamics in the upper and lower regimes. Indeed, the introduction of extra parameters may increase, for finite sample size, the critical values of the tests by enlarging the source of randomness. This may therefore affect the power of the test. On the other hand, the TAR model (2.1) is richer than Kapetanios and Shin [2002] who set ρ_2 to 0, as μ_1, μ_2 and $a(L)$. The parsimonious threshold model (2.1) nests H_0 in various ways, such as $\rho_1 = \rho_2 = 0$ and $\mu_1 = \mu_2 = 0$ for any non identified λ, or for $\lambda = \infty$, with $\mu_2 = 0$ and $\rho_2 = 0$.

Due to the existence of the middle regime $|y_{t-1}| < \lambda$, our auxiliary model (2.1) is able to isolate a local unit-root behavior of the process which cannot be taken into account by a two-regime TAR. Indeed, Bec et al. [2001] derive a sufficient stationarity condition for model (2.1) which allows for $\rho_2 = 0$ in the middle regime as soon as the outer dynamics is stationary enough. In the case $p = 0$, Chan, Petruccelli, Tong and Woolford [1985] give necessary and sufficient conditions for stationarity in (2.1). In particular, the presence of unit roots in the outer and inner regimes ($\rho_1 = \rho_2 = 0$) together with $\mu_1 > 0$ still yields a stationary process.[1] Therefore, as noted by Balke and Fomby [1997], just examining the autoregressive parameters ρ_1 and ρ_2 is not sufficient to characterize stationarity in (2.1). However, we aim to limit the source of randomness in our test statistics in order to obtain relatively small critical values and our tests are based on the *parsimonious auxiliary hypotheses* $\rho_1 = \rho_2 = 0$ in (2.1) with λ varying in a random set Λ_T. Although this is not equivalent to stationarity in (2.1) for a given λ, we derive the consistency of the test against a wider class of stationary alternatives when the random Λ_T asymptotically covers the set of all admissible threshold parameters. The underlying intuition

[1] As far as we know, necessary and sufficient conditions are not yet available for (2.1) when the condition $p = 0$ is relaxed. This is one of the main problems in using (2.1) to test for stationarity. Moreover, even for $p = 0$ and a known λ, the sufficient and necessary condition for stationarity of Chan et al. [1985] yields five different conditions so that a corresponding test would be difficult to build.

is that taking a large λ in (2.1) gives approximately a linear model for which the unit root hypothesis is equivalent to $\rho_2 = 0$.

Assume that a sample $y_1, \ldots, y_T$ is at hand and consider first a given threshold parameter λ. Let $\widehat{\epsilon}(\lambda) = (\widehat{\epsilon}_{p+1}(\lambda), \ldots, \widehat{\epsilon}_T(\lambda))'$ be the estimated residuals associated with the unrestricted model (2.1). Let $\widehat{\epsilon}_0(\lambda)$ be the estimated residuals from (2.1) with $\rho_1 = \rho_2 = 0$. Our testing procedures are based upon the following Wald, Lagrange Multiplier and Likelihood Ratio statistics,

$$\begin{aligned}
W_T(\lambda) &= T\left(\frac{\widehat{\varepsilon}_0'(\lambda)\widehat{\varepsilon}_0(\lambda) - \widehat{\epsilon}'(\lambda)\widehat{\epsilon}(\lambda)}{\widehat{\varepsilon}_0'(\lambda)\widehat{\varepsilon}_0(\lambda)}\right) = T\left(1 - \frac{\widehat{\epsilon}'(\lambda)\widehat{\epsilon}(\lambda)}{\widehat{\varepsilon}_0'(\lambda)\widehat{\varepsilon}_0(\lambda)}\right) , \\
LM_T(\lambda) &= T\left(\frac{\widehat{\varepsilon}_0'(\lambda)\widehat{\varepsilon}_0(\lambda) - \widehat{\epsilon}'(\lambda)\widehat{\epsilon}(\lambda)}{\widehat{\epsilon}'(\lambda)\widehat{\epsilon}(\lambda)}\right) = \frac{W_T(\lambda)}{1 - \frac{W_T(\lambda)}{T}} , \\
LR_T(\lambda) &= T\ln\left(\frac{\widehat{\varepsilon}_0'(\lambda)\widehat{\varepsilon}_0(\lambda)}{\widehat{\epsilon}'(\lambda)\widehat{\epsilon}(\lambda)}\right) = -T\ln\left(1 - \frac{W_T(\lambda)}{T}\right) .
\end{aligned}$$

Note that all these statistics are functions of $W_T(\lambda)$, which rewrites as, using linearity with respect to the parameters $a_1, \ldots, a_p$, μ_1, ρ_1 and μ_2, ρ_2,

$$W_T(\lambda) = [\widehat{\rho}_1(\lambda), \widehat{\rho}_2(\lambda)]\, V_T^{-1}(\lambda) \begin{bmatrix} \widehat{\rho}_1(\lambda) \\ \widehat{\rho}_2(\lambda) \end{bmatrix} \tag{2.2}$$

where $V_T(\lambda)$ is an estimate of the variance matrix of $(\widehat{\rho}_1(\lambda), \widehat{\rho}_2(\lambda))'$.[2] As to build the test, observe that λ is a nuisance parameter under the null. Following Bec et al. [2001], Berben and van Dijk [1999], Caner and Hansen [2001], Davies [1987] and Hansen [1994] who consider testing issues when a nuisance parameter appears under the null, we eliminate λ by considering the maximum of these statistics. Let $\Lambda_T = [\underline{\lambda}_T, \overline{\lambda}_T]$ be a random set of admissible threshold parameters. Our test statistics are

$$\sup_{\lambda \in \Lambda_T} LM_T(\lambda) \ , \ \sup_{\lambda \in \Lambda_T} LR_T(\lambda) \text{ and } \sup_{\lambda \in \Lambda_T} W_T(\lambda) \ , \tag{2.3}$$

The choice of Λ_T affects the power of the tests in two ways. First, under H_0, if Λ_T is small, then the critical values of the test will be small so that a better power is obtained under the alternative. Second, under the alternative, it would be desirable to have a large Λ_T in order to

[2]The matrix $V_T(\lambda)$ has no inverse when there is no observation in a regime. In this case, we ignore this regime when computing the statistics. As seen from the limit result (2.5) below, this should not happen for T large enough when λ remains bounded. Note also that the statistics are constant over intervals defined through the ordered values of $|y_1|, \ldots, |y_{t-1}|$, so that it is sufficient to compute the statistics at this value.

detect a large class of alternatives: think for instance of TAR alternatives with $\rho_1 \neq 0$ for which it is suitable to have $\lambda \in \Lambda_T$, so that Λ_T should asymptotically cover the whole set of possible thresholds. Hereafter, we refer to these behaviors of a random Λ_T under H_0 and the alternative by saying that Λ_T is *adaptive*. The next section derives the limit distribution of the test statistics under H_0 and studies the consistency of our procedure. We will also discuss how to build such a set Λ_T through consistent test statistics of H_0 and give some examples of such a construction.

2.1 Asymptotic behavior of the tests

An important issue when deriving the limit distribution of the test statistics under the null is to understand the behavior of the estimators of the central regime parameters. Indeed, under H_0, the Donsker Theorem yields that

$$\left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t, \frac{y_{[Tr]}}{\sqrt{T}} \right\}_{r\in[0,1]} \xrightarrow{d} \{\sigma W(r), \delta W(r)\}_{r\in[0,1]} \ , \tag{2.4}$$

where $\{W(r)\}_{r\geq 0}$ is a standard Brownian motion, σ^2 the variance of the ε_t's and $\delta = \sigma/(1-a(1))$. It then follows that the number of y_t's in the inner regime $(-\lambda, \lambda)$ should be small for any finite $\lambda > 0$ since $y_{[Tr]}$ diverges with the order $\sqrt{T}$.[3] The Donsker Theorem is therefore appropriate to study the estimators of the inner regime for a threshold parameter diverging with the order $\sqrt{T}$ of the variables, as for instance in Bec et al. [2001]. When the threshold parameter remains bounded under H_0 as in our adaptive approach, the central regime appears in the estimation through transformations of the y_t's which vanish outside $(-\lambda, \lambda)$ and requires a deeper study of the bounded values of the variables. As known from Park and Phillips [2001], this can be done by introducing the local time $\{\ell_W(w, r)\}_{w\in\mathbb{R}, r\geq 0}$ of the Brownian motion $\{W(r)\}_{r\geq 0}$. The local time $\ell_W(\cdot, T)$ can be defined as the density of the occupation time of a measurable subset A of $\mathbb{R}$ by the Brownian motion between 0 and T, i.e.

$$\int_0^T \mathbb{I}(W(r) \in A)dr = \int_{-\infty}^{+\infty} \mathbb{I}(w \in A)\ell_W(w, T)dw \ ,$$

[3] As a consequence, one may think to remove the contribution of the inner regime from our test statistics, i.e. to base the tests on the auxiliary hypothesis $\rho_1 = 0$ in (2.1) as in Kapetanios and Shin [2002]. This may affect the power of the tests as reported in Balke and Fomby [1997], footnote 13. See also the discussion following Theorem 2 below.

see Revuz and Yor [1999] and Park and Phillips [2001] for further properties and applications of the local time. Theorems 3.1 and 3.2. of Park and Phillips [2001] characterize the asymptotic behavior of means over the inner regime appearing in $\widehat{\mu}_2(\lambda)$ and $\widehat{\rho}_2(\lambda)$, establishing that

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \mathbb{I}\left[y_{t-1} \in (-\lambda, \lambda)\right] f(y_{t-1}) \\ \frac{1}{T^{1/4}} \sum_{t=2}^{T} \mathbb{I}\left[y_{t-1} \in (-\lambda, \lambda)\right] f(y_{t-1})\varepsilon_t \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{\ell_W(0,1)}{|\delta|} \int_{-\lambda}^{\lambda} f(y)dy \\ B\left(\frac{\ell_W(0,1)}{|\delta|} \int_{-\lambda}^{\lambda} f^2(y)dy\right) \end{pmatrix}, \tag{2.5}$$

jointly with (2.4), where $\{B(r)\}_{r\geq 0}$ is a standard Brownian motion independent of $\{W(r)\}_{r\geq 0}$ and then of its local time $\{\ell_W(w,r)\}_{w\in\mathbb{R}, r\geq 0}$. This shows in particular that the number of y_t's in the inner regime is of order $\sqrt{T}$ and can be used to find the limit distribution of $T^{1/4}\widehat{\rho}_2(\lambda)$ and $T^{1/4}\widehat{\mu}_2(\lambda)$. This limit result is therefore a key tool to derive the asymptotic contribution of the inner regime to the Wald statistic (2.2) for a fixed λ. In order to ensure the validity of the fundamental results of Park and Phillips [2001], we assume that under H_0,

> **Assumption E**. *the i.i.d. ε_t's are such that $\mathbb{E}\varepsilon_t = 0$ and $\mathbb{E}|\varepsilon_t|^{4+s} < \infty$ for some $s > 0$. The ε_t's have a bounded density and $\lim_{y\to\infty} y^{\gamma}\mathbb{E}\exp(iy\varepsilon_1) = 0$ for some $\gamma > 0$.*

A first step to derive the limit distribution of our test statistics is the finite-dimensional convergence in distribution of the LM_T, LR_T and W_T processes under H_0 as stated in the next proposition. In what follows $\text{sgn}(x)$ is the sign of x, i.e. $\text{sgn}(x) = 1$ if $x \geq 0$, $\text{sgn}(x) = -1$ if $x < 0$.

Proposition 1 *Consider some fixed $\lambda_j > 0$, $j = 1, \ldots, J$. If Assumption E holds, then under H_0 $(LM_T(\lambda_j), j = 1, \ldots, J)$, $(LR_T(\lambda_j), j = 1, \ldots, J)$ and $(W_T(\lambda_j), j = 1, \ldots, J)$ jointly converge in distribution to $(\zeta_1^2 + \zeta_2^2(\lambda_j), j = 1, \ldots, J)$ with, for $\lambda > 0$,*

$$\begin{aligned} \begin{bmatrix} \zeta_1^2 \\ \{\zeta_2^2(\lambda)\}_{\lambda>0} \end{bmatrix} &= \begin{bmatrix} \frac{\left(\int_0^1 W(r)dW(r) - \int_0^1 |W(r)|dr \int_0^1 \text{sgn}(W(r))dW(r)\right)^2}{\int_0^1 W^2(r)dr - \left(\int_0^1 |W(r)|dr\right)^2} \\ \left\{ \frac{B^2\left(\frac{2\lambda^3}{3|\delta|}\ell_W(0,1)\right)}{\frac{2\lambda^3}{3|\delta|}\ell_W(0,1)} \right\}_{\lambda>0} \end{bmatrix} \\ &\stackrel{d}{=} \begin{bmatrix} \frac{\left(\int_0^1 W(r)dW(r) - \int_0^1 |W(r)|dr \int_0^1 \text{sgn}(W(r))dW(r)\right)^2}{\int_0^1 W^2(r)dr - \left(\int_0^1 |W(r)|dr\right)^2} \\ \left\{ \frac{B^2(\lambda^3)}{\lambda^3} \right\}_{\lambda>0} \end{bmatrix}. \end{aligned}$$

The variable ζ_1 is the contribution of the outer regime, and is somehow similar to the limit of the squared ADF statistic. Note that ζ_1 does not depend on λ because the most part of the y_t's in the outer regime diverge as can be seen from the Donsker Theorem and therefore dominate λ which can be set to 0. This is not the case for the contribution $\{\zeta_2(\lambda)\}_{\lambda>0}$ of the inner regime which varies with λ as it can be expected from (2.5). Note that $\{\zeta_2(\lambda)\}_{\lambda>0}$ does not depend on the local time due to standardization of the Wald statistic. The contributions of the two regimes are asymptotically independent and $(\zeta_1, \zeta_2(\lambda))$ has a parameter-free distribution.

Proposition 1 can be easily extended in several directions. First, different lag polynomial functions can be considered for the three regimes which can be defined through two thresholds $\lambda_1 < \lambda_2$. Different mean and ρ coefficients can as well be introduced for the upper and lower regimes. Second, the auxiliary hypothesis $\rho_1 = \rho_2 = 0$ used to build the statistics can include a mean restriction $\mu_1 = \mu_2 = 0$. But this will induce an additional term $\zeta_{1\mu}^2 + \zeta_{2\mu}(\lambda)$ in the limit distribution due to estimation of the means, which will increase the critical values of the test.

The variable $\zeta_2(\lambda)$ diverges when λ goes to 0 due to the law of iterated logarithm which yields that $B(\lambda^3)$ goes to 0 slower than $\sqrt{\lambda^3}$ almost surely, see e.g. Revuz and Yor [1999]. This can be explained more intuitively by noting that taking $\lambda = 0$ gives a non identified inner regime in model (2.1). Therefore, in order to avoid large values of the test statistics, we restrict the following analysis to a $\underline{\lambda}_T$ asymptotically bounded away from 0 under H_0.

To derive the limit distribution of the LM_T, LR_T and W_T statistics considered as processes indexed by λ, we now establish the asymptotic equicontinuity of some intermediary processes involved in the expression (2.2) of the Wald statistics $W_T(\lambda)$.

Proposition 2 *Let*

$$S_T(\lambda) = T^{-1/2} \sum_{t=2}^{T} f(y_{t-1}) \mathbb{I}\left(y_{t-1} \in \Lambda\right) \ , \ M_T(\lambda) = T^{-1/4} \sum_{t=2}^{T} f(y_{t-1}) \mathbb{I}\left(y_{t-1} \in \Lambda\right) \varepsilon_t \ ,$$

where Λ *is* $(-\infty, -\lambda]$, $(-\lambda, \lambda)$ *or* $[\lambda, +\infty)$ *and* $f(\cdot)$ *is bounded on the compact interval* $[a, b]$. *Then under* H_0 *and Assumption E, the processes* $\{S_T(\lambda)\}_{\lambda \in [a,b]}$ *and* $\{M_T(\lambda)\}_{\lambda \in [a,b]}$ *are asymptotically equicontinuous.*

Proposition 2 can also be used to extend our framework in several directions. First the auxiliary model (2.1) can be generalized by allowing different parameters in the upper, middle and lower regimes as well as an asymmetric central band $[\lambda_1, \lambda_2]$, and the tests may include a mean restrictions $\mu_1 = \mu_2 = 0$. Second, the lower and upper threshold values $\underline{\lambda}_T$ and $\overline{\lambda}_T$ may diverge with T as in Bec et al. [2001] without affecting asymptotic equicontinuity of the (renormalized) processes in Proposition 2. Indeed, the proof of Proposition 2 relies on a bound of the increments of the process $s_T(\lambda) = T^{-1/2}\sum_{t=2}^{T}\mathbb{I}(y_{t-1} \leq \lambda)$ based on a study of the densities of the y_t's by Akonom [1993]. Lemma 3 in the proof appendix states that, for some $C > 0$ any λ_1, λ_2,[4]

$$\mathbb{E}\left(s_T(\lambda_2) - s_T(\lambda_1)\right)^2 \leq C\left((\lambda_2-\lambda_1)^2 + 2\frac{|\lambda_2-\lambda_1|}{\sqrt{T}}\right).$$

Changing λ into $\pi\sqrt{T}$ as in Bec et al. [2001] now yields, for any bounded map $f(\cdot)$

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{T}\sum_{t=2}^{T} f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\left(\mathbb{I}[y_{t-1}\leq \pi_2\sqrt{T}] - \mathbb{I}[y_{t-1}\leq \pi_1\sqrt{T}]\right)\right]^2 \\
\leq\ & \frac{\sup_{y\in[a,b]} f^2(y)}{T}\mathbb{E}\left(s_T(\pi_2\sqrt{T}) - s_T(\pi_1\sqrt{T})\right)^2 \\
\leq\ & C\sup_{y\in[a,b]} f^2(y)\left((\pi_2-\pi_1)^2 + 2\frac{|\pi_2-\pi_1|}{T}\right) \text{ for } \pi_1,\pi_2 \text{ in } [a,b],
\end{aligned}$$

which is similar to the bound above and can be used to study the asymptotic equicontinuity of processes such as $T^{-1}\sum f(y_{t-1}/\sqrt{T})\mathbb{I}[y_{t-1}\leq \pi\sqrt{T}]$ or $T^{-1/2}\sum f(y_{t-1}/\sqrt{T})\mathbb{I}[y_{t-1}\leq \pi\sqrt{T}]\varepsilon_t$. Such a result is actually a preliminary step to derive the limit distribution of maximum statistics when the maximum is taken over a set of threshold values delimited by sample quantiles as in Bec et al. [2001] or Berben and van Dijk [1999]. Third, slight modifications of Proposition 2 would allow to tackle smooth transition threshold models.

The asymptotic equicontinuity of the LM, LR and Wald statistics over finite intervals bounded away from 0 is obtained as a corollary.

Corollary 1 *Consider $0 < a \leq b < \infty$. Then, under H_0 and Assumption E, the processes $\{LM_T(\lambda)\}_{\lambda\in[a,b]}$, $\{LR_T(\lambda)\}_{\lambda\in[a,b]}$ and $\{W_T(\lambda)\}_{\lambda\in[a,b]}$ are asymptotically equicontinuous.*

As in the null limit distribution of Proposition 1, the inner and outer regimes differ when studying asymptotic equicontinuity. Here, the technical difficulty is to obtain the asymptotic

[4]The dependence of this bound upon $|\lambda_2-\lambda_1|/\sqrt{T}$ prevents here from applying usual asymptotic equicontinuity criterion as the Chentchov-Komolgorov one, see Theorem 15.6 and below in Billingsley [1968]. We circumvent this technical issue via a maximal inequality of Van der Vaart and Wellner [1996], see the proof of Proposition 2.

equicontinuity of the inner regime contribution to the statistics. Indeed, the outer regime is easier to study because its limit contribution is constant with respect to λ[5].

To complete the study of the test statistics under the null, we now derive their common asymptotic distribution from Proposition 1 and Corollary 1. Theorem 1 is a consequence of Proposition 1 and Corollary 1 which can also be used to study other functionals of $\{LM_T(\lambda)\}_{\lambda>0}$, $\{LR_T(\lambda)\}_{\lambda>0}$ or $\{W_T(\lambda)\}_{\lambda>0}$, as, for instance, the exponential average considered in Andrews and Ploberger (1994).

Theorem 1 *Consider a random subset* $\Lambda_T = [\underline{\lambda}_T, \overline{\lambda}_T]$ *of* $\mathbb{R}_+$ *where* $\underline{\lambda}_T = \underline{\lambda}_T(y_1, \ldots, y_T)$ *and* $\overline{\lambda}_T = \overline{\lambda}_T(y_1, \ldots, y_T)$ *with, under* H_0,

$$\left(\underline{\lambda}_T, \overline{\lambda}_T\right) \xrightarrow{d} \left(\underline{\lambda}, \overline{\lambda}\right) \text{ jointly with the limit of Proposition 1 and } \mathbb{P}\left(0 < \underline{\lambda} \leq \overline{\lambda} < \infty\right) = 1.$$

Then, under H_0 *and Assumption E,* $\sup_{\lambda\in\Lambda_T} LM_T(\lambda)$, $\sup_{\lambda\in\Lambda_T} LR_T(\lambda)$ *and* $\sup_{\lambda\in\Lambda_T} W_T(\lambda)$ *jointly converge in distribution to*

$$\zeta_1^2 + \sup_{\lambda\in\Lambda} \zeta_2^2(\lambda) \text{ where } \Lambda = [\underline{\lambda}, \overline{\lambda}].$$

If, moreover, $(\overline{\lambda}, \underline{\lambda})$ *is independent of* $\{B(r)\}_{r>0}$ *and the distribution of* $\overline{\lambda}/\underline{\lambda}$ *is parameter-free, then the test statistics* $\sup_{\lambda\in\Lambda_T} LM_T(\lambda)$, $\sup_{\lambda\in\Lambda_T} LR_T(\lambda)$ *and* $\sup_{\lambda\in\Lambda_T} W_T(\lambda)$ *are asymptotically pivotal.*

The variable $\zeta_1^2 + \sup_{\lambda\in\Lambda} \zeta_2^2(\lambda)$ remains finite since ζ_2 is a.s. continuous over $\mathbb{R}_+^*$ and $\underline{\lambda}$, $\overline{\lambda}$ are bounded away from 0 and infinity with probability 1. A choice of Λ_T satisfying the conditions of Theorem 1 is a deterministic finite interval $[a, b]$, $a > 0$, which would be appropriate if some *a priori* information on λ were available. However, such a choice is somehow arbitrary and may affect the power of the test: think for instance of a stationary threshold alternative (2.1) with a threshold parameter outside $[a, b]$. We now study the tests under general stationary alternatives and give a sufficient condition on Λ_T ensuring that the tests are consistent.

Theorem 2 *Assume that the alternative process* $\{y_t\}_{t\geq1}$ *is ergodic stationary with finite second-order moments and has a continuous marginal distribution of order* $p+1$. *Assume also* $\underline{\lambda}_T$ *is bounded in probability and that* $\overline{\lambda}_T \xrightarrow{\mathbb{P}} +\infty$ *for the alternative at hand. Then the test statistics*

[5]Kapetanios and Shin [2002] go around this issue by considering a restricted version of model (2.1), in which the inner regime is arbitrarily constrained to be a random walk.

$\sup_{\lambda \in \Lambda_T} LM_T(\lambda)$, $\sup_{\lambda \in \Lambda_T} LR_T(\lambda)$ *and* $\sup_{\lambda \in \Lambda_T} W_T(\lambda)$ *diverge and the corresponding tests are therefore consistent.*

The divergence of $\overline{\lambda}_T$ can be weakened by assuming that $\overline{\lambda}_T$ is asymptotically larger than the upper bound of the common support of the y_t's. The proof of Theorem 2 is based on the fact that, for a large $\lambda > 0$, the coefficients of the inner regime are approximately given by the regression of Δy_t on $1, y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p}$. We show that this gives, for a stationary alternative as in Theorem 2, a coefficient ρ_2 of the inner regime which differs from 0, so that the test statistics diverge as the ADF statistic.[6] But other consistency arguments also apply, suggesting that our test can be more powerful in presence of local non stationarity. For instance, in case of a stationary threshold alternative (2.1) with $\rho_1 \neq 0$ or $\rho_2 \neq 0$, it is sufficient for consistency that the threshold parameter is in Λ_T with a probability going to 1, so that convergence of $\underline{\lambda}_T$ to the lower bound of the support of the $|y_t|$'s seems to be desirable in practice. As discussed above and as will be seen from our simulation experiments, the ADF test is less powerful against such alternatives, especially if $\rho_2 = 0$ and the threshold is large.

The comparison of the conditions of Λ_T required by Theorems 1 and 2 therefore suggests that the behavior of this random set should dramatically differ under H_0 and the alternative. A set Λ_T which is bounded under H_0 and obeys the conditions of Theorem 2 can easily be built by choosing a length $\overline{\lambda}_T - \underline{\lambda}_T$ depending on a test statistic of H_0 which is bounded under the null and diverges under the alternative. Choosing a lower threshold $\underline{\lambda}_T$ which is bounded away from 0 under H_0 and asymptotically vanishes under the alternative gives a Λ_T which satisfies the requirements of both Theorems 1 and 2. This will give a test with a potential higher power compared to existing ones.

2.2 Examples of adaptive threshold sets

We now give two examples of Λ_T satisfying the requirements of Theorems 1 and 2. Let $\widehat{s}^2$ be an estimate of $\sigma^2 = \mathrm{Var}(\varepsilon_t)$ which is consistent under H_0 and remains bounded under the

[6] Allowing for a large λ and estimating the coefficient ρ_2 of the inner regime is therefore crucial in this consistency argument. From this point of view, our approach differs from Kapetanios and Shin [2002] who do not take in consideration the inner regime by imposing $\rho_2 = 0$. This may also explain the loss of power mentioned in the footnote 13 of Balke and Fomby [1997] when removing the contribution of the inner regime from the test statistics.

alternative, as for instance

$$\widehat{s}^2 = \frac{1}{T-(p+2)} \sum_{t=}^{T} \left(y_t - \widehat{\beta}_0 - \widehat{\beta}_1 y_{t-1} - \cdots - \widehat{\beta}_{p+1} y_{t-(p+1)} \right)^2 \tag{2.6}$$

where $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_{p+1}$ are the OLS estimates of the regression coefficients of y_t on $1, y_{t-1}, \ldots, y_{t-(p+1)}$. We use $\widehat{s}$ as a scaling factor in the construction of Λ_T. Denote $|y|_{(1)}, \ldots, |y|_{(T-1)}$ the ordered values of $|y_1|, \ldots, |y_{T-1}|$. Observe that taking $\lambda \in [|y|_{(3)}, |y|_{(T-2)}]$ gives a Wald statistic (2.2) which is well defined for processes with finite continuous distributions of order $p+1$. We shall then base our choice of $\underline{\lambda}_T$ on a test statistic and $|y|_{(3)}$.[7] Under H_0, $|y|_{(3)}$ goes to 0 due to the recurrence property of $\{y_t\}_{t\geq 1}$ as given in (2.5). Under an ergodic stationary alternative, $|y|_{(3)}$ goes to the lower bound of the support of the $|y_t|$'s as shown by the Law of Large Number. We now propose two examples of adaptive Λ_T based upon different choices of $\overline{\lambda}_T - \underline{\lambda}_T$ proportional to a test statistic of H_0.

Our first example builds on the t-statistic $\widehat{t}(ADF)$ of the ADF test. Let ℓ be a length parameter to be specified in our simulation experiment. The set $\Lambda_T = [\underline{\lambda}_T, \overline{\lambda}_T]$ with

$$\underline{\lambda}_T = |y|_{(3)} + \frac{\widehat{s}}{\ell \max\left(1, |\ \hat{t}(ADF)\ |\right)}\ ,\ \overline{\lambda}_T = \underline{\lambda}_T + \ell \widehat{s} \max\left(1, |\ \hat{t}(ADF)\ |\right)\ , \tag{2.7}$$

satisfies the conditions of Theorem 1 and Theorem 2. Indeed, under H_0, the limit distribution of $\overline{\lambda}_T/\underline{\lambda}_T$ is given by the one of $\ell^2 \max(1, \widehat{t}^2(ADF))$ which holds jointly with Proposition 1 and only depends upon the Brownian motion $\{W(r)\}_{r\in[0,1]}$. As a consequence, a test statistic (2.3) using (2.7) has an asymptotic pivotal distribution. Under the alternative, $|\widehat{t}(ADF)|$ diverges as seen from the discussion following Theorem 2. As a consequence, $\underline{\lambda}_T$ and $\overline{\lambda}_T$ remain bounded away from 0 and infinity under H_0, while, under a stationary ergodic alternative, $\overline{\lambda}_T$ diverges and $\underline{\lambda}_T$ converges to the lower bound of the support of the $|y_t|$'s in probability. The next consistency result is therefore a direct corollary of Theorem 2 by consistency of the ADF test.

Corollary 2 *Let $\{y_t\}_{t\geq 1}$ be as in Theorem 2. Then $\sup_{\lambda\in\Lambda_T} LM_T(\lambda)$, $\sup_{\lambda\in\Lambda_T} LR_T(\lambda)$ and $\sup_{\lambda\in\Lambda_T} W_T(\lambda)$ diverge for a choice of Λ_T corresponding to (2.7).*

[7] In the construction of the examples below, we therefore propose a lower bound $\underline{\lambda}_T$ with $\underline{\lambda}_T \geq |y|_{(3)}$ holding by construction but do not impose a similar restriction on the upper bound. We do not encounter any existence problem due to this unrestricted choice of $\overline{\lambda}_T$ in our simulation experiments. Otherwise the convention of footnote 2 can be used.

Our second example is more specific to threshold alternatives and builds on Bec et al. [2001]. Let $\widehat{\lambda}_{1/2}$ be the median of $|y|_{(1)}, \ldots, |y|_{(T-1)}$ and consider now a Λ_T with

$$\underline{\lambda}_T = |y|_{(3)} + \frac{\widehat{s}}{\ell \max\left(1, W_T^{1/2}(\widehat{\lambda}_{1/2})\right)} \ , \ \overline{\lambda}_T = \underline{\lambda}_T + \ell \widehat{s} \max\left(1, W_T^{1/2}(\widehat{\lambda}_{1/2})\right) \ . \tag{2.8}$$

Proposition 2 of Bec et al. [2001] showed that this choice of (2.8) yields a test statistic (2.3) with a pivotal limit distribution under the null. The study of this choice (2.8) under a general alternative is slightly more difficult. Bec et al. [2001] established that $W_T^{1/2}(\widehat{\lambda}_{1/2})$ diverges for some specific TAR alternatives, see Proposition 2 therein.

Corollary 3 *Let $\{y_t\}_{t\geq 1}$ be as in Theorem 2, assuming moreover that $W_T^{1/2}(\widehat{\lambda}_{1/2})$ diverges in probability. Then $\sup_{\lambda\in\Lambda_T} LM_T(\lambda)$, $\sup_{\lambda\in\Lambda_T} LR_T(\lambda)$ and $\sup_{\lambda\in\Lambda_T} W_T(\lambda)$ diverge for a choice of Λ_T corresponding to (2.8).*

Table 1 gives the critical values based on 40,000 simulations of different sample sizes for admissible sets (2.7) and (2.8) with $\ell = 4$. Note that these critical values are much higher than the squared ones of ADF test. For instance, at the 5% level, the squared critical values of the ADF test is $(-2.88)^2 = 8.2944$ which is smaller than the one of our test. As shown in Proposition 1 and Theorem 1, the limit distribution of our test statistics has two components. The first one, ζ_1^2, is comparable to the limit of a squared ADF statistic. The second component, $\sup_{\lambda\in\Lambda} \zeta_2^2(\lambda)$, comes from the introduction of an unknown threshold parameter in the test statistic. Higher critical values is here a price to pay for the introduction of a threshold parameter. As shown later on in the simulation experiments, this will have some consequences on the relative power of our tests with respect to the ADF test for linear or close to linear DGP. Note also that the critical values of the test based upon $W^{1/2}(\widehat{\lambda}_{1/2})$ are slightly larger in general than the ones associated to $|\widehat{t}(ADF)|$, suggesting that the length of Λ_T is larger in mean than the one of Λ_T^{ADF}.

Other sets Λ_T of admissible threshold parameters can be proposed by considering various test statistics of H_0. Our construction of Λ_T can also be iterated. In the next section, simulations experiments under several alternatives are presented for a choice of Λ_T given by (2.7) and (2.8).

3 Simulation experiments

We now propose a Monte-Carlo investigation of the power properties of the test. In a first set of experiments, we analyze the effect of the choice of the threshold values used to compute our

Table 1: Critical values (40,000 simulations)

	$W^{1/2}(\hat{\lambda}_{1/2})$				$\hat{t}(ADF)$			
Sample size	15 %	10%	5%	1%	15 %	10%	5%	1%
100	10.93	12.21	14.32	19.01	10.49	11.70	13.68	18.11
150	10.88	12.14	14.12	18.34	10.47	11.68	13.57	17.64
200	10.76	12.00	13.93	18.28	10.42	11.59	13.48	17.67
250	10.85	12.05	14.00	18.26	10.52	11.66	13.58	17.70
300	10.83	11.99	13.88	18.33	10.50	11.63	13.45	17.79
500	10.90	12.07	14.04	18.14	10.66	11.84	13.77	17.76
1000	10.87	12.08	14.10	18.06	10.66	11.88	13.84	17.82

supremum test statistics. In a second set we study the power properties of our test against various nonlinear alternatives. In our simulation study, we focus on the Wald statistic $\sup_{\lambda \in \Lambda_T} W_T(\lambda)$ where Λ_T is as in (2.7) and (2.8) with $\ell = 4$. We fix the lag parameter p to 1 and use the variance estimate (2.6). In what follows, the test statistics are respectively denoted SupWald(Λ_T^{ADF}) and SupWald(Λ_T). In the experiments we let $T = 325$.[8]

3.1 The choice of the threshold values

As a baseline for our comparison, we consider SupWald(Λ_T^0) $= \sup_{\lambda \in \Lambda_T^0} W_T(\lambda)$ with

$$\Lambda_T^0 = \left[|y|_{([0.15T])}, |y|_{([0.85T])} \right] .$$

The set Λ_T^0 is typical of percentile based choices used in the literature, see Balke and Fomby [1997], Bec et al. [2001], Berben and van Dijk [1999] and Caner and Hansen [2001]. The length

$$|y|_{([0.85T])} - |y|_{([0.15T])} = \sqrt{T} \left(\frac{|y|_{([0.85T])}}{\sqrt{T}} - \frac{|y|_{([0.15T])}}{\sqrt{T}} \right)$$

of Λ_T^0 diverges under H_0 as seen from (2.4) and remains bounded under a stationary alternative. The sets Λ_T based on (2.7) and (2.8) of our SupWald tests behave at the opposite and are thus expected to yield a more powerful test.

[8] We retain here the same sample size as Bec et al. [2001] for the sake of comparison.

We first estimate the finite sample critical values of our test under the null. The model considered under H_0 is

$$\Delta y_t = a\Delta y_{t-1} + \varepsilon_t \ , \ a = 0.3 \ , \ \varepsilon_t \rightsquigarrow \mathcal{N}(0,1) \ , \ T = 325.$$

Table 2 compares the critical values of the three tests under the null.[9] The critical values of the

Table 2: Empirical critical values of the unit root tests ($a = 0.3$, $T = 325$, 10,000 simulations)

	15 %	10%	5%	1%
SupWald(Λ_T^0)	13.2	14.5	16.5	21.1
SupWald(Λ_T)	10.8	12.0	13.9	18.1
SupWald(Λ_T^{ADF})	10.6	11.7	13.6	17.4

three tests are computed for $a = 0.3$, but taking $a = 0$ does not seems to affect the estimated critical values of our test. Indeed, the critical values are close to ones reported in Table 1. The estimated critical values of SupWald(Λ_T^{ADF}) and SupWald(Λ_T) are significantly below the ones of SupWald(Λ_T^0). Indeed, further investigation of our simulation datasets shows that Λ_T contains around 26% of the $|y_t|$'s while this percentage reaches 70% for Λ_T^0 by definition.

In order to investigate the effect of the choice of the threshold values on the power of the test, we consider the TAR alternatives with an integrated inner regime

$$\Delta y_t = a\Delta y_{t-1} + \varepsilon_t + \begin{cases} \mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \leq -\lambda, \\ \rho_2 y_{t-1} & \text{if } |y_{t-1}| < \lambda, \\ -\mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \geq \lambda, \end{cases} \text{, with } \mu_1 = 1.3 \times |\rho_1| \times \lambda,\ \rho_2 = 0. \quad (3.1)$$

A preliminary estimation of the 15% and 75% quantiles of the data generating processes of this experiment shows that λ should be in Λ_T^0 with a high probability for the data generating process of the experiment. Table 3 also reports the power of the ADF test. The estimated power of SupWald(Λ_T^0) is computed using the 5% critical value of Table 1. The values in parenthesis are percentages of $|y_t|$ contains in Λ_T^{ADF} and Λ_T. As expected by Corollary 2 and Corollary 3, these values are substantially greater than the ones under the null (around 26 %). This illustrates the adaptive behavior of Λ_T^{ADF} and Λ_T. Finally, % denotes the percentage of data in the stationary regime.

[9]We thank Bec et al. [2001] to supply us the critical values for Λ_T^0.

Table 3: Empirical power of the unit root tests ($\alpha = 5\%$,$T = 325$, 1,000 simulations)

a	ρ_1	ρ_2	λ	%	ADF	SupW(Λ_T^0)	SupW(Λ_T)	SupW(Λ_T^{ADF})
0	-0.10	0	10	1.0	14.7	67.2	80.0 (92.9)	83.6 (85.3)
0	-0.05	0	10	2.3	14.6	47.5	61.4 (87.4)	58.6 (80.3)
0	-0.02	0	10	9.4	13.6	13.9	29.7 (78.3)	27.0 (71.1)
0.3	-0.10	0	10	3.2	24.8	90.0	97.4 (97.0)	99.3 (95.1)
0.3	-0.05	0	10	5.8	22.8	74.2	85.7 (93.6)	89.9 (89.4)
0.3	-0.02	0	10	12.7	17.3	27.1	47.3 (82.2)	47.7 (79.24)
0	-0.10	0	5	8.7	73.2	95.0	99.2 (98.3)	99.1 (98.0)
0	-0.05	0	5	14.3	46.9	57.8	82.1 (97.7)	81.8 (98.1)
0	-0.02	0	5	27.8	20.8	10.3	33.2 (87.0)	34.2 (94.6)
0.3	-0.10	0	5	11.3	100.0	99.7	100.0 (98.7)	100.0 (98.5)
0.3	-0.05	0	5	19.0	96.6	76.8	95.7 (98.1)	96.4 (98.5)
0.3	-0.02	0	5	34.3	45.2	13.4	38.0 (82.6)	43.6 (95.5)
0	-0.10	0	2	32.9	100.0	85.8	99.8 (96.3)	100.0 (97.8)
0	-0.05	0	2	46,3	93.3	20.9	72.1 (91.7)	81.3 (97.8)
0	-0.02	0	2	63.2	27.5	3.9	17.5 (91.7)	22.6 (97.8)
0.3	-0.10	0	2	39.8	100.0	94.5	100.0 (96.3)	100.0 (98.3)
0.3	-0.05	0	2	53.5	99.4	37.4	79.6 (89.1)	95.3 (98.3)
0.3	-0.02	0	2	69.3	44.6	6.4	22.6 (74.86)	35.30 (93.0)

In all cases, the tests SupWald(Λ_T^{ADF}) and SupWald(Λ_T) outperform the test SupWald(Λ_T^0). The adaptive property greatly increases the power of the test. For instance, the increase in power exceeds 20 % in six case and is close to 60 % in two cases ($a = 0, \rho_1 = -0.05, \lambda = .2$ and $a = .3, \rho_1 = -0.05, \lambda = .2$). As discussed above, the gain in power is due to *i)* the fact that the critical values of SupWald(Λ_T) are relatively small and *ii)* the fact that Λ_T should be larger than Λ_T^0 under the alternatives. It is rather difficult to discuss which one of these two effects has the most important impact on the observed gain of power. But, since λ is in Λ_T^0 with a high probability so that SupWald(Λ_T^{ADF}), SupWald(Λ_T) and SupWald(Λ_T^0) should be large, one may suspect that the behavior of Λ_T under H_0 which yields a smaller critical value provides the most important contribution to the improvement of the power.

The tests SupWald(Λ_T^{ADF}) and SupWald(Λ_T) generally outperform the standard ADF except for cases where the percentage in the stationary regimes is more important. However, for these cases the power of adaptive tests is close to the power of the standard ADF and substantially higher than the power of SupWald(Λ_T^0). For processes characterized by a percentage of data in the stationary regimes below 10%, the gain of the adaptative tests compared to the standard ADF can be as high as 70%. Finally, the power of the tests SupWald(Λ_T^{ADF}) and SupWald(Λ_T) is really close.

3.2 Power of the test against various nonlinear alternatives

Let us now proceed to simulation experiments so as to check the consistency of our SupWald test against a broader set of stationary alternatives, either linear or not. More precisely, we will consider below a stationary autoregressive process (H_{AR}^a), a piecewise stationary model with a changing mean (H_{SC}^a), and an Autoregressive Conditional Root model (H_{ACR}^a) proposed by Rahbek and Shephard [2002] (see also Gouriéroux and Robert [2002]). Those alternatives are respectively given by:

$$H_{AR}^a : \Delta y_t = \mu + \rho y_{t-1} + a\Delta y_{t-1} + \varepsilon_t, \tag{3.2}$$

$$H_{SC}^a : \Delta y_t = \mu_1 \mathbb{I}(t \leq \tau T) + \mu_2 \mathbb{I}(t > \tau T) + \rho y_{t-1} + a\Delta y_{t-1} + \varepsilon_t, \tag{3.3}$$

where $\mathbb{I}(\cdot)$ is the indicator function which is equal to 1 when the condition into parenthesis holds, and to 0 elsewhere, and τT is the breaking time. Finally, the Markov ACR alternative writes:

$$H_{ACR}^a : y_t = \rho^{s_t} y_{t-1} + \varepsilon_t, \tag{3.4}$$

where s_t is binomial given the past with $p_t = \mathbb{P}(s_t = 1 | y_{t-1}, \varepsilon_t) = \left[1 + \exp\left(-(\alpha + \beta y_{t-1}^2)\right)\right]^{-1}$, ρ is a real number, β is non-negative and α and β are finite. In the three models considered here, ε_t is an i.i.d. $\mathcal{N}(0, \sigma^2)$. The sample size is $T = 325$.

Table 4: Empirical power of the unit root tests for linear model ($\alpha = 5\%$,$T = 325$, 1,000 simulations)

μ	ρ	ADF	SupWald(Λ_T)	SupWald(Λ_T^{ADF})
1	-0.02	25.1	17.9	17.8
0	-0.02	27.0	17.2	23.0
-1	-0.02	28.9	17.3	17.2
1	-0.05	92.6	73.8	73.9
0	-0.05	92.8	58.7	73.6
-1	-0.05	92.9	73.9	73.4
1	-0.10	100	99.8	99.8
0	-0.10	100	98.2	99.5
-1	-0.10	100	99.9	99.9

Tables 4 and 5 report the results for H^a_{AR} and H^a_{SC} alternatives. As expected, the ADF test outperforms the unit root test based on the threshold specification in the linear stationary case. Indeed, the ADF test is specifically designed for this alternative. As mentioned after Table 1, our test implies the computation of the supremum over an interval involving a loss of power compared to the ADF test for a linear alternative. However, the performance of our tests seems to be reasonable. Here again, the power of the tests SupWald(Λ_T^{ADF}) and SupWald(Λ_T) is really close.

Table 5: Empirical power of the unit root tests for changing mean model ($\alpha = 5\%$,$T = 325$, 1,000 simulations)

μ_1	μ_2	ρ	τ	ADF	SupWald(Λ_T)	SupWald(Λ_T^{ADF})
0	0.5	-0.05	0.5	5.6	11.5	12.9
0	1	-0.05	0.5	0.0	37.5	7.3
0	2	-0.05	0.5	0.0	84.0	4.3
0	1	-0.02	0.5	0.0	14.1	1.9
0	1	-0.10	0.5	4.7	62.5	61.8
0	1	-0.05	0.05	83.4	80.7	82.3
0	1	-0.05	0.25	0.0	56.8	41.7
0	1	-0.05	0.75	0.0	18.5	5.1
0	1	-0.05	0.95	28.3	21.0	36.8

It is well known that the ADF test has poor power to detect piecewise stationary model with a changing mean. As expected, the unit root tests based on the threshold specification with the set Λ_T outperforms greatly the linear ADF test and the test SupWald(Λ_T^{ADF}). The important power gain of the SupWald(Λ_T) probably comes from the threshold specification of Λ_T which better seperates the two regimes of the process. The power is similar for the cases where the breakpoint is at the beginning and at the end of the sample. For those cases, the process is almost linear stationary. Finally, as pointed by previous studies, the power of the ADF test is below the nominal size (5%) for several cases.

The Markov ACR model exhibits local non stationarity when $s_t = 0$, which is more likely to arise if $\alpha + \beta y_{t-1}^2$ is small. When $\beta > 0$ as in our simulation experiment, this source of local stationarity corresponds to a central regime, but with a less precise delimitation than for the TAR model (2.1). Indeed, due to the randomness of s_t, local nonstationarity may also hold outside a central zone. Even though the degree of local nonstationarity of the ACR model is related to the parameters (α, β), it is worth computing the percentage of time spent in the stationary regime (% in Table 6) for interpretation's sake. The values of the parameters are motivated by the example given by Rahbek and Shephard [2002]. The results are reported in

Table 6.

Table 6: Empirical power of the unit root tests for the Markov ACR alternative ($\alpha = 5\%$, $T = 325$, 1,000 simulations)

α	β	ρ	%	ADF	SupWald(Λ_T)	SupWald(Λ_T^{ADF})
0	1.1	0.9	100.0	99.7	91.6	94.2
0	1.1	0.7	100.0	100	100	100
-10	1.1	0.9	52.0	99.4	87.5	91.5
-10	1.1	0.7	33.0	100	100	100
-25	1.1	0.9	19.0	75.2	84.9	84.4
-25	1.1	0.7	9.2	100	100	76.9
-50	1.1	0.9	10.6	25.6	76.5	76.4
-50	1.1	0.7	4.4	44.1	99.8	99.7
-100	1.1	0.9	5.5	16.9	61.5	61.9
-100	1.1	0.7	2.4	18.9	93.8	95.9
-200	1.1	0.9	2.7	13.8	34.3	23.9
-200	1.1	0.7	1.3	17.7	61.8	56.7
-300	1.1	0.9	1.9	13.1	25.5	20.0
-300	1.1	0.7	0.9	12.2	42.2	29.8
-100	0.5	0.9	2.6	13.7	34.2	27.1
-100	0.5	0.7	1.3	15.6	60.4	54.7
-100	1.5	0.9	7.0	18.0	72.9	75.1
-100	1.5	0.7	3.0	21.6	97.5	98.1

The ADF test slightly dominates the supWald tests in the case where the time spent in the stationary regime ($s_t = 1$) exceeds 50 %. In the other cases, the unit root tests based on the threshold specification does remarkably well while the ADF test has poor power. In particular, for processes characterized by a time spent in the stationary regime below 10 %, the supWald tests clearly dominate the ADF test. For instance, for parameters values equal to $\alpha = -100$, $\beta = 1.1$ and $\rho = .9$ and a time spent in the stationary regime equal to 5.5%, the rate of rejection for the supWald tests is around 62 percent compared to 16.9 percent for the ADF test. Finally, the power of the tests SupWald(Λ_T^{ADF}) and SupWald(Λ_T) is again really close.

According to these simulation results, the tests proposed in this paper seem to be able to detect globally stationary process with a relatively important nonstationary component. This simulation study based on four alternative specifications clearly reveals the usefulness of the proposed tests based on an adaptive choice of the band. We then suggest to practitioners to perform our tests in conjunction with the traditional ADF test.

4 The yield spread dynamics revisited

Under costless and instantaneous portfolio adjustment assumption, the expectations theory of the term structure implies the following non-arbitrage condition :

$$R(k,t) = \frac{1}{k}\left[\sum_{j=1}^{k} E_t[R(1,t+j-1)]\right] + L(k,t), \tag{4.5}$$

where $R(k,t)$ denotes the k-period interest rate, E_t is the expectation operator conditional on time t information, and $L(k,t)$ represents the term premium, accounting for risk and liquidity premia. This in turn implies the stationarity of the yield spread between longer-term and shorter-term interest rates. Indeed, by rearranging (4.5), the spread may be expressed as :

$$S(k,1,t) = R(k,t) - R(1,t) = \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j=1}^{i} E_t[\Delta R(1,t+j)] + L(k,t) = \mu_t, \tag{4.6}$$

where the right-hand side is stationary as soon as interest rates are integrated of order one and the risk premium is stationary. Hence, as noticed by Hall, Anderson and Granger [1992] and Anderson [1997], arbitrage behavior guarantees that equation (4.6) acts as an attractor as soon as $S(k,1,t) \neq \mu_t$.

However, as pointed out by Anderson [1997], if one considers homogeneous transaction costs which reduce the investor's yield on a purchased bond by a constant amount λ, then the investor will convert a portfolio of one-period bonds to $k-$period bonds if and only if $\lambda < S(k,1,t) - \mu_t$, or convert $k-$period bonds to 1-period bonds if and only if $S(k,1,t) - \mu_t < -\lambda$. Therefore, in presence of transaction costs, the attraction towards equilibrium (4.6) is inactive when :

$$-\lambda < S(k,1,t) - \mu_t < \lambda. \tag{4.7}$$

Hence, there is no reason for the cointegration relation between long- and short-term rates to hold in this area, or put in other words, for the spread to revert toward μ_t. This arbitrage behavior clearly suggests a stationary but nonlinear dynamics for the yield spread, which should be well captured by our parsimonious auxiliary model. Moreover, recent empirical evidence — see e.g. Keim and Madhavan [1997] or Wagner [1998] — displays transaction costs estimates ranging roughly from 0.5% to more than 2% depending on the types of costs included in the calculation.

The interest rates data used in this study are monthly averages spanning from 1980:01 to 1998:12 for France and Germany since the Euro was introduced in January 1999, and to 2001:08 for the US[10]. For the New Zealand[11], the available data span from 1985:01 to 2002:01. For France, Germany, the New Zealand and the US, the short term interest rate is respectively the 3-month PIBOR, the 3-month FIBOR, the 90-day Bank Bill yield and the 3-month Treasury Bill rate, while the long term is the 10-year public and semi-public sector bonds rate, the 9 to 10-year Bd listed federal securities rate, the 10-year secondary market government bond yield and the 10-year Treasury constant maturity rate. The yield spreads are defined as the difference between the long and the short-term rates, and are denoted S_F, S_G, S_{NZ} and S_{US}.

As can be seen from Table 7, performing the standard ADF unit-root test and KPSS stationarity test[12] reveals that the US and German spreads are well characterized by a unit-root process, whereas no clear-cut conclusion emerges for S_F and S_{NZ}. Indeed, the KPSS statistics

Table 7: ADF and KPSS tests

Stat.	k,ℓ	S_G	k,ℓ	S_{US}	k,ℓ	S_F	k,ℓ	S_{NZ}
ADF(k)	1	-1.889	4	-2.726	1	-2.672	4	**-3.211**
KPSS(ℓ)	3	1.671	4	0.602	2	**0.101**	4	1.691

The critical values at the 5 % level are -2.88 for ADF and 0.463 for KPSS.

fails to reject the null of stationarity for the US spread while the ADF fails to reject the unit-root for S_{NZ}. The critical values for the test SupWald(Λ_T) are given in Table 1. The values obtained

[10]European and US data come respectively from Datastream and FRED databanks.

[11]These data come from the Reserve Bank of New Zealand.

[12]The lag length for the ADF(k) is chosen according to the Ljung-Box statistic. The size of the Bartlett windows for KPSS(ℓ) is obtained following Andrews [1991].

for the SupWald(Λ_T) statistics are reported in Table 8. The lag order of the $a(L)$ polynomial in

Table 8: SupWald unit-root test

	S_F	S_G	S_{NZ}	S_{US}
SupWald(Λ_T)	10.96	15.42	52.16	30.07
SupWald(Λ_T^{ADF})	10.96	15.42	32.98	30.07
p-value	15%	5%	1%	1%

model (2.1) is chosen according to the BIC and Ljung-Box statistics which suggest $p = 1$ for the European spreads, and $p = 4$ for the remainders. The SupWald(Λ_T) statistic strongly rejects the null for S_G, S_{NZ} and S_{US}, but only at the 15% level for S_F.

5 Concluding remarks

In this paper, we develop adaptive consistent unit root tests based on a three-regime threshold autoregression specification as auxiliary model. We retain this parsimonious auxiliary model for its potential ability to isolate local nonstationarity. The originality of the approach consists in the treatment of the unknown threshold in a random set Λ_T, over which a maximum of the Lagrange multiplier, Likelihood ratio or Wald statistic is computed to build the test. At the opposite of previous approaches, we consider that the threshold levels remain bounded under the null and allow for unbounded ones under the alternative. This adaptive random choice of thresholds improves the power of the test, by producing smaller critical values and, under the alternative, increasing the value of the test statistic. Such a methodology can be applied to various models and testing issues. For instance in the unit-root context, an alternative test statistic would be an average over Λ_T of some exponential average statistics, as derived by Kapetanios and Shin [2002] from Andrews and Ploberger [1994]. In specification testing, Cramer-von Mises or Kolmogorov type test statistics, or average of non parametric estimators, can be computed over an adaptive Λ_T as done in trimming for technical purposes. In addition of increasing the power of a test, considering a smaller set Λ_T may also improve the accuracy of the null limit distribution, as well as providing a more robust statistic. This new approach therefore suggests various promising directions for future research.

A simulation study illustrates the good detection properties of our procedure. Compared to previous approaches, it is suggested that the improvement of the power comes from a reduction of the critical values due to the adaptive choice of the threshold parameters involved in the computation of the maximum. However, the critical values of the test remain large, so that the traditional ADF test is more powerful against alternatives that are not affected by local nonstationarity. In particular, the proposed tests should be applied in conjunction with the traditional ADF test if some a priori information favors a stationary linear alternative. When applied to post-1980 French, German, New-Zealander and US monthly data, our test rejects the null of unit root whereas ADF and KPSS tests give mixed evidence at best.

References

Akonom, J., Comportement asymptotique du temps d'occupation du processus des sommes partielles, *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 1993, *29*, 57–81.

Anderson, H., Transaction Costs and Non-Linear Adjustment Towards Equilibrium in the U.S. Treasury Bill Market, *Oxford Bulletin of Economics and Statistics*, 1997, *59* (4), 465–484.

Andrews, D.W.K., Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 1991, *59* (3), 817–858.

—— and W. Ploberger, Optimal Tests When A Nuisance Parameter Is Present Only Under The Alternative, *Econometrica*, 1994, *62* (6), 1383–1414.

Balke, N.S. and T.B. Fomby, Threshold Cointegration, *International Economic Review*, 1997, *38*, 627–45.

Bec, F., M. Ben Salem, and M. Carrasco, *Test for Unit-root Versus Threshold Specification with an Application to the PPP*, Working Paper 2001-72, EUREQua, University of Paris 1 2001.

Berben, R. P. and D. van Dijk, *Unit Root Tests and Asymmetric Adjustment : A Reassessment*, Research Report EI-9902/A, Econometric Institute 1999.

Billingsley, P., *Convergence of Probability Measures*, Wiley, 1968.

Bohl, M. and P. Siklos, *The Bundesbank's Inflation Policy and Asymmetric Behavior of the German Term Structure*, manuscript, Dept of economics, European University Viadrina Frankfurt, Germany 2001.

Campbell, J. and R. Shiller, Cointegration and Tests of Present Value Models, *Journal of Political Economy*, 1987, *95* (5), 1062–1088.

Caner, M. and B. Hansen, Threshold Autoregression With a Unit Root, *Econometrica*, 2001, *69*, 1555–1596.

Chan, K. S., J. Petruccelli, H. Tong, and S. Woolford, A Multiple-Threshold AR(1) Model, *Journal of Applied Probability*, 1985, *22*, 267–279.

Chow, Y.S. and H. Teicher, *Probability theory. Independence, interchangeability, Martingales*, Springer-Verlag, 1988. Second edition.

Darling, D. A. and M. Kac, On Occupation Times for Markoff Processes, *Transactions of the American Mathematical Society*, 1957, *84*, 444–458.

Davies, R. B., Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative, *Biometrika*, 1987, *74*, 33–43.

der Vaart, A. Van and J.A. Wellner, *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer-Verlag, 1996.

Enders, W. and C.W.J. Granger, Unit-root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates, *Journal of Business and Economic Statistics*, 1998, *16* (3), 304–11.

—— and P.L. Siklos, Cointegration and Threshold Adjustment, *Journal of Business and Economic Statistics*, 2001, *19* (2), 166–176.

Gonzalez, M. and J. Gonzalo, *Threshold Unit Root Models*, Working Paper, U. Carlos III 1998.

Gouriéroux, C. and C. Robert, *Tails and Extrema Behaviour of Stochastic Unit Root Models*, Working Paper, CREST 2002.

Hall, A., H. Anderson, and C. Granger, A Cointegration Analysis of Treasury Bill Yields, *The Review Of Economics and Statistics*, 1992, *LXXIV*, 116–26.

Hamilton, J., *Time Series Analysis*, Princeton, New Jersey: Princeton University Press, 1994.

Hansen, B.E., Testing for Parameter Instability in Linear Models, in N. Ericsson and J. Irons, editors, *Testing exogeneity*, Oxford University Press, 1994, chapter 15, pp. 389–406.

Kapetanios, G. and Y. Shin, *Unit Root Tests in Three-Regime SETAR Models*, Mimeo, University of Edinburgh 2002.

Keim, D. and A. Madhavan, Transaction Costs and Investment Style: An Inter-Exchange Analysis of Institutional Equity Trades, *Journal of Financial Economics*, 1997, *46*, 265–292.

Pagan, A., A. Hall, and V. Martin, Modelling the Term Structure, in G. Maddala and C. Rao, editors, *Handbook of Statistics*, Amsterdam: North-Holland, 1996, chapter 14.

Park, J. and P. Phillips, Nonlinear Regressions with Integrated Time Series, *Econometrica*, 2001, *69*, 117–161.

Rahbek, A. and N. Shephard, *Autoregressive Conditional Root Model*, Working Paper, Nuffield College, Oxford University 2002.

Revuz, D. and M. Yor, *Continuous Martingales and Brownian Motion*, Springer-Verlag, 1999.

Shin, D.W. and O. Lee, Tests for Asymmetry in Possibly Nonstationary Time Series Data, *Journal of Business and Economic Statistics*, 2001, *19* (2), 233–244.

Stock, J. and M. Watson, Testing for Common Trends, *Journal of the American Statistical Association*, 1988, *83*, 1097–1107.

Taylor, A., Potential Pitfalls for the PPP Puzzle ? Sampling and Specification Biases in Mean-Reversion Tests of the LOOP, *Econometrica*, 2001, *69*, 473–498.

Wagner, W., *The Official Icebergs of Transaction Costs*, Commentary 54, Plexus Group 1998.

Appendix: Proofs of the main results

We first introduce our main notations. Let $u_t = \Delta y_t$ and for $\lambda \in \mathbb{R}_+$ define the column vector x_t as

$$x_t = x_t(\lambda) = \left(u_{t-1}, \ldots, u_{t-p}, -\mathrm{sgn}_\lambda(y_{t-1}), y_{t-1}\mathbb{I}(|y_{t-1}| \geq \lambda), \mathbb{I}(|y_{t-1}| < \lambda), y_{t-1}\mathbb{I}(|y_{t-1}| < \lambda)\right)' ,$$

where $\mathrm{sgn}_\lambda(y_{t-1}) = \mathbb{I}(y_{t-1} \geq \lambda) - \mathbb{I}(y_{t-1} \leq -\lambda)$. Let $\beta = \beta(\lambda) = (a_1, \ldots, a_p, \mu_1, \rho_1, \mu_2, \rho_2)'$ be a column vector so that

$$u_t = x_t'\beta + \varepsilon_t \text{ with } \beta = \beta_0 = (a_1, \ldots, a_p, 0, 1, 0, 1)' \text{ under } H_0.$$

We now aim to give a more explicit expression of the Wald statistic $W_T(\lambda)$ in (2.2). Let $\widehat{\beta} = \widehat{\beta}_T$ be the ordinary least squares estimate of β. Consider also the $(p+4) \times (p+4)$ diagonal scaling matrix $\Gamma = \Gamma_T$

$$\Gamma = \begin{bmatrix} \sqrt{T} & & & & & & \\ & \ddots & & & & & \\ & & \sqrt{T} & & & & \\ & & & \sqrt{T} & & & \\ & & & & T & & \\ & & & & & T^{1/4} & \\ & & & & & & T^{1/4} \end{bmatrix} = \begin{bmatrix} \Gamma_{11}\ (p \times p) & 0 & 0 \\ 0 & \Gamma_{22}^o\ (2 \times 2) & 0 \\ 0 & 0 & \Gamma_{22}^i\ (2 \times 2) \end{bmatrix}.$$

The scaling matrix Γ corresponds to the standardization of the estimation error $\widehat{\beta} - \beta$ under H_0. This is done through

$$\Gamma\left(\widehat{\beta} - \beta_0\right) = \Gamma\left[\Gamma^{-1}\sum_{t=p+1}^{T} x_t x_t' \Gamma^{-1}\right]^{-1} \Gamma^{-1} \sum_{t=p+1}^{T} x_t \varepsilon_t \text{ under } H_0. \tag{A.1}$$

We rewrite $\Gamma\left(\widehat{\beta} - \beta_0\right)$ as $C^{-1}(\lambda)M(\lambda)$ where $C(\lambda) = C$ and $M(\lambda) = M$ are defined as

$$C = C(\lambda) = \Gamma^{-1}\sum_{t=p+1}^{T} x_t x_t' \Gamma^{-1} = \begin{bmatrix} C_{11} & C_{21}' \\ C_{21} & C_{22} \end{bmatrix} \text{ with}$$

$$C_{11} = \frac{1}{T}\begin{bmatrix} \sum_{t=p+1}^{T} u_{t-1}^2 & \sum_{t=p+1}^{T} u_{t-1}u_{t-2} & \cdots & \sum_{t=p+1}^{T} u_{t-1}u_{t-p} \\ \sum_{t=p+1}^{T} u_{t-1}u_{t-2} & \sum_{t=p+1}^{T} u_{t-2}^2 & \cdots & \sum_{t=p+1}^{T} u_{t-2}u_{t-p} \\ \vdots & \vdots & & \vdots \\ \sum_{t=p+1}^{T} u_{t-1}u_{t-p} & \sum_{t=p+1}^{T} u_{t-2}u_{t-p} & \cdots & \sum_{t=p+1}^{T} u_{t-p}^2 \end{bmatrix},$$

$$
C_{21} = \begin{bmatrix}
-\frac{1}{T}\sum_{t=p+1}^{T} \operatorname{sgn}_\lambda(y_{t-1})u_{t-1} & \cdots & -\frac{1}{T}\sum_{t=p+1}^{T} \operatorname{sgn}_\lambda(y_{t-1})u_{t-p} \\
\frac{1}{T^{3/2}}\sum_{t=p+1}^{T} y_{t-1}\mathbb{I}(|y_{t-1}| \geq \lambda)u_{t-1} & \cdots & \frac{1}{T^{3/2}}\sum_{t=p+1}^{T} y_{t-1}\mathbb{I}(|y_{t-1}| \geq \lambda)u_{t-p} \\
\frac{1}{T^{3/4}}\sum_{t=p+1}^{T} \mathbb{I}(|y_{t-1}| < \lambda)u_{t-1} & \cdots & \frac{1}{T^{3/4}}\sum_{t=p+1}^{T} \mathbb{I}(|y_{t-1}| < \lambda)u_{t-p} \\
\frac{1}{T^{3/4}}\sum_{t=p+1}^{T} y_{t-1}\mathbb{I}(|y_{t-1}| < \lambda)u_{t-1} & \cdots & \frac{1}{T^{3/4}}\sum_{t=p+1}^{T} y_{t-1}\mathbb{I}(|y_{t-1}| < \lambda)u_{t-p}
\end{bmatrix},
$$

$$
C_{22} = \begin{bmatrix} C_{22}^o & 0 \\ 0 & C_{22}^i \end{bmatrix} \text{ with}
$$

$$
C_{22}^o = \begin{bmatrix}
\frac{1}{T}\sum_{t=p+1}^{T} \mathbb{I}\left(|y_{t-1}| \geq \lambda\right) & -\frac{1}{T}\sum_{t=p+1}^{T} \frac{y_{t-1}}{\sqrt{T}}\operatorname{sgn}_\lambda(y_{t-1}) \\
-\frac{1}{T}\sum_{t=p+1}^{T} \frac{y_{t-1}}{\sqrt{T}}\operatorname{sgn}_\lambda(y_{t-1}) & \frac{1}{T}\sum_{t=p+1}^{T} \left(\frac{y_{t-1}}{\sqrt{T}}\right)^2 \mathbb{I}(|y_{t-1}| \geq \lambda)
\end{bmatrix},
$$

$$
C_{22}^i = \frac{1}{\sqrt{T}} \begin{bmatrix}
\sum_{t=p+1}^{T} \mathbb{I}(|y_{t-1}| < \lambda) & \sum_{t=p+1}^{T} y_{t-1}\mathbb{I}(|y_{t-1}| < \lambda) \\
\sum_{t=p+1}^{T} y_{t-1}\mathbb{I}(|y_{t-1}| < \lambda) & \sum_{t=p+1}^{T} y_{t-1}^2\mathbb{I}(|y_{t-1}| < \lambda)
\end{bmatrix},
$$

$$
M = M(\lambda) = \Gamma^{-1}\sum_{t=p+1}^{T} x_t\varepsilon_t = \begin{bmatrix} M_1 \\ M_2^o \\ M_2^i \end{bmatrix} \text{ with } M_1 = \begin{bmatrix}
\frac{1}{T}\sum_{t=p+1}^{T} u_{t-1}\varepsilon_t \\
\frac{1}{T}\sum_{t=p+1}^{T} u_{t-2}\varepsilon_t \\
\vdots \\
\frac{1}{T}\sum_{t=p+1}^{T} u_{t-p}\varepsilon_t
\end{bmatrix},
$$

$$
M_2^o = \begin{bmatrix}
-\sum_{t=p+1}^{T} \operatorname{sgn}_\lambda(y_{t-1})\frac{\varepsilon_t}{\sqrt{T}} \\
\sum_{t=p+1}^{T} \frac{y_{t-1}}{\sqrt{T}}\mathbb{I}\left(|y_{t-1}| \geq \lambda\right)\frac{\varepsilon_t}{\sqrt{T}}
\end{bmatrix}, \; M_2^i = \begin{bmatrix}
\frac{1}{T^{1/4}}\sum_{t=p+1}^{T} \mathbb{I}(|y_{t-1}| < \lambda)\varepsilon_t \\
\frac{1}{T^{1/4}}\sum_{t=p+1}^{T} y_{t-1}\mathbb{I}(|y_{t-1}| < \lambda)\varepsilon_t
\end{bmatrix}.
$$

Let R be a selection matrix with entries 0 or 1 such that $R\beta = (\rho_1, \rho_2)'$. Under H_0, (A.1) yields that the Wald statistic $W_T(\lambda)$ in (2.2) writes as

$$
\begin{aligned}
W_T(\lambda) &= \frac{1}{\widehat{\sigma}^2(\lambda)}\left(RC^{-1}(\lambda)M(\lambda)\right)'\left(RC^{-1}(\lambda)R'\right)^{-1}\left(RC^{-1}(\lambda)M(\lambda)\right) \\
&\quad \text{with } \widehat{\sigma}^2(\lambda) = \widehat{\sigma}^2 = \frac{1}{T}\sum_{t=p+1}^{T}\varepsilon_t^2 - \frac{M'(\lambda)C^{-1}(\lambda)M(\lambda)}{T}\,. \qquad \text{(A.2)}
\end{aligned}
$$

Preliminary results

In the results below, C denotes a constant which varies from line to line.

Lemma 1 *Under H_0 and Assumption E,* $\max_{t\leq T}|u_t| = O_{\mathbb{P}}(T^{1/(4+s)})$.

Proof of Lemma 1. It immediately follows from the Chebychev inequality and the fact that $\max \mathbb{E}|u_t|^{4+s} < \infty$ under Assumption E using the mean average expression of the u_t's. □

The next results are essential to establish asymptotic equicontinuity. Let us first introduce some additional notations. Let $\{\pi_i\}_{i\geq 0}$ be the coefficients of the Wold representation of $u_t = y_t - y_{t-1}$, which are such that $u_t = \sum_{i=0}^{\infty}\pi_i\varepsilon_{t-i}$. As a consequence, we have $y_t = \pi_0\varepsilon_t + (\pi_0+\pi_1)\varepsilon_{t-1} + \cdots + (\pi_0 + \cdots + \pi_{t-1})\varepsilon_1 + \sum_{i=0}^{\infty}(\pi_{i+1} + \cdots + \pi_{i+t})\varepsilon_{-i}$. Define

$$\psi_i = \sum_{j=0}^{i}\pi_j \text{ so that } y_t = \sum_{i=0}^{t-1}\psi_i\varepsilon_{t-i} + \sum_{i=0}^{\infty}(\psi_{i+t} - \psi_i)\,\varepsilon_{-i}\ ,$$

$$\widetilde{y}_{t,i} = \sum_{j=t+1}^{i}\psi_{i-j}\varepsilon_j \text{ and } \overline{y}_{t,i} = y_i - \widetilde{y}_{t,i} = \sum_{j=1}^{t}\psi_{i-j}\varepsilon_j + \sum_{j=0}^{\infty}(\psi_{j+t} - \psi_j)\,\varepsilon_{-j} \text{ for } t < i.$$

Let $f_t(\cdot)$ and $\widetilde{f}_{t,i}(\cdot)$ be the Lebesgue densities of y_t and $\widetilde{y}_{t,i}$. The bound (3.7) of Akonom [1993] write as:

Lemma 2 (Akonom [1993]) *Under H_0 and Assumption E, there exists a constant $C > 0$ such that for all $1 \leq i$ and $1 \leq t < i$,*

$$\sup_{y\in\mathbb{R}} f_t(y) \leq \frac{C}{\sqrt{t+1}} \text{ and } \sup_{y\in\mathbb{R}} \widetilde{f}_{t,i}(y) \leq \frac{C}{\sqrt{i-t}}\ .$$

The density bound of Lemma 2 would be straightforward in case of i.i.d. normal ε_t. We now give a moment bound for the number of y_t's between two thresholds $\lambda_1 \leq \lambda_2$.[13]

Lemma 3 *Let $s_T(\lambda)$ be $T^{-1/2}\sum_{t=2}^{T}\mathbb{I}(y_{t-1}\in\Lambda)$ where Λ is $(-\infty,-\lambda]$, $(-\lambda,\lambda)$ or $[\lambda,+\infty)$. Then, under H_0 and Assumption E, there is a constant $C > 0$ such that*

$$\mathbb{E}^{1/2}\left(s_T(\lambda_2) - s_T(\lambda_1)\right)^2 \leq C\left[|\lambda_2-\lambda_1| + 2\left(\frac{|\lambda_2-\lambda_1|}{\sqrt{T}}\right)^{1/2}\right] \text{ for any } \lambda_1, \lambda_2 \text{ and } T\geq 2.$$

[13] A continuous-time version of this bound can be deduced from a more general moment computation in Darling and Kac [1957]. The discrete-time version given by these authors is not correct since it ignores the contribution of (A.3) in $\mathbb{E}(s_T(\lambda_2) - s_T(\lambda_1))^2$, see the proof of Lemma 3. This item gives the term $(|\lambda_2-\lambda_1|/\sqrt{T})^{1/2}$ in the bound of the lemma.

Proof of Lemma 3. We consider $s_T(\lambda) = T^{-1/2} \sum_{t=2}^{T} \mathbb{I}\left(y_{t-1} \geq \lambda\right)$ and $\lambda_1 < \lambda_2$, the other cases being similar. This gives

$$
\begin{aligned}
\mathbb{E}\left(s_T(\lambda_2) - s_T(\lambda_1)\right)^2 &= \frac{1}{T}\sum_{t=2}^{T} \mathbb{E}\mathbb{I}(\lambda_1 \leq y_{t-1} < \lambda_2) && \text{(A.3)}\\
&\quad + \frac{2}{T}\sum_{t=2}^{T-1} \mathbb{E}\left[\mathbb{I}(\lambda_1 \leq y_{t-1} < \lambda_2) \sum_{i=t+1}^{T} \mathbb{I}(\lambda_1 \leq y_{i-1} < \lambda_2)\right]. && \text{(A.4)}
\end{aligned}
$$

Lemma 2 gives for (A.3)

$$
\begin{aligned}
\frac{1}{T}\sum_{t=2}^{T} \mathbb{E}\mathbb{I}(\lambda_1 \leq y_{t-1} < \lambda_2) &= \frac{1}{T}\sum_{t=2}^{T} \int_{\lambda_1}^{\lambda_2} f_{t-1}(y)dy \leq \frac{C(\lambda_2 - \lambda_1)}{T} \sum_{t=2}^{T} \frac{1}{\sqrt{t}} \\
&\leq \frac{C(\lambda_2 - \lambda_1)}{\sqrt{T}} \text{ since } \textstyle\sum_{t=2}^{T} \frac{1}{\sqrt{t}} \leq \int_0^T \frac{dx}{\sqrt{x}} = \frac{\sqrt{T}}{2}.
\end{aligned}
$$

For (A.4), write $y_{i-1} = \widetilde{y}_{t-1,i-1} + \overline{y}_{t-1,i-1}$ and note that $\widetilde{y}_{t-1,i-1}$ is independent of the sigma-field $\mathcal{F}_{t-1} = \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$, $i \geq t+1$. Observe also that y_{t-1}, $\overline{y}_{t-1,i-1}$, $i \geq t+1$, are in $\mathcal{F}_{t-1}$. Therefore Lemma 2 yields

$$
\begin{aligned}
\sum_{i=t+1}^{T} \mathbb{E}\left[\mathbb{I}\left(\lambda_1 \leq y_{i-1} < \lambda_2\right) | \mathcal{F}_{t-1}\right] &= \sum_{i=t+1}^{T} \int_{\lambda_1 - \overline{y}_{t-1,i-1}}^{\lambda_2 - \overline{y}_{t-1,i-1}} \widetilde{f}_{t-1,i-1}(\widetilde{y})d\widetilde{y} \leq C(\lambda_2 - \lambda_1) \sum_{i=t+1}^{T} \frac{1}{\sqrt{i-t}} \\
&\leq C(\lambda_2 - \lambda_1)\sqrt{T-t} \leq C\sqrt{T}(\lambda_2 - \lambda_1).
\end{aligned}
$$

Combining the bounds above gives for the item (A.4),

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=2}^{T-1} \mathbb{E}\left[\mathbb{I}(\lambda_1 \leq y_{t-1} < \lambda_2) \sum_{i=t+1}^{T} \mathbb{I}(\lambda_1 \leq y_{i-1} < \lambda_2)\right] \\
&= \frac{1}{T}\sum_{t=2}^{T-1} \mathbb{E}\left[\mathbb{I}(\lambda_1 \leq y_{t-1} < \lambda_2) \sum_{i=t+1}^{T} \mathbb{E}\left[\mathbb{I}(\lambda_1 \leq y_{i-1} < \lambda_2) | \mathcal{F}_{t-1}\right]\right] \\
&\leq \frac{C(\lambda_2 - \lambda_1)}{\sqrt{T}} \sum_{t=2}^{T} \mathbb{E}\left[\mathbb{I}(\lambda_1 \leq y_{t-1} < \lambda_2)\right] \leq C(\lambda_2 - \lambda_1)^2 .
\end{aligned}
$$

Combining the bounds of the items in (A.3) and (A.4) yields that

$$
\begin{aligned}
\mathbb{E}^{1/2}\left(s_T(\lambda_2) - s_T(\lambda_1)\right)^2 &\leq \left(C\frac{|\lambda_2 - \lambda_1|}{\sqrt{T}} + 2C(\lambda_2 - \lambda_1)^2\right)^{1/2} \leq C\left[\sqrt{2}|\lambda_1 - \lambda_2| + \left(\frac{|\lambda_2 - \lambda_1|}{\sqrt{T}}\right)^{1/2}\right] \\
&\leq \sqrt{2}C\left[|\lambda_1 - \lambda_2| + 2\left(\frac{|\lambda_2 - \lambda_1|}{\sqrt{T}}\right)^{1/2}\right],
\end{aligned}
$$

since $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$ for positive real numbers, showing that Lemma 3 is proven. The form of the last bound above slightly simplifies the proof of Proposition 2. □

We now extend the preceding lemma to martingale and more general functions of the lagged variable y_{t-1}. The following moment bounds are the main tool to establish the asymptotic equicontinuity of Proposition 2 using the maximal inequality.

Lemma 4 *Let $[a,b]$ be a compact interval of $\mathbb{R}$ and $f(\cdot)$ be a bounded function from $[a,b]$ to $\mathbb{R}$. Then under H_0 and Assumption E,*

i. Let $S_T(\lambda)$ be $T^{-1/2}\sum_{t=2}^{T} f(y_{t-1})\mathbb{I}\left(y_{t-1}\in\Lambda\right)$ where Λ is $(-\infty,-\lambda]$, $(-\lambda,\lambda)$ or $[\lambda,+\infty)$. There is a constant C such that, for all $\lambda_1,\lambda_2\in[a,b]$ and all $T\geq 2$,

$$\mathbb{E}^{1/2}\left(S_T(\lambda_2)-S_T(\lambda_1)\right)^2\leq d_{2T}(\lambda_1,\lambda_2)=C\left[|\lambda_2-\lambda_1|+2\left(\frac{|\lambda_2-\lambda_1|}{\sqrt{T}}\right)^{1/2}\right]$$

ii. Let $M_T(\lambda)$ be $T^{-1/4}\sum_{t=2}^{T} f(y_{t-1})\mathbb{I}\left(y_{t-1}\in\Lambda\right)\varepsilon_t$ where Λ is $(-\infty,-\lambda]$, $(-\lambda,\lambda)$ or $[\lambda,+\infty)$. There is a constant C such that, for all $\lambda_1,\lambda_2\in[a,b]$ and all $T\geq 2$,

$$\mathbb{E}^{1/4}\left(M_T(\lambda_2)-M_T(\lambda_1)\right)^4\leq d_{4T}(\lambda_1,\lambda_2)=C\left[|\lambda_2-\lambda_1|^{1/2}+2\left(\frac{|\lambda_2-\lambda_1|}{\sqrt{T}}\right)^{1/4}\right].$$

Proof of Lemma 4. Lemma 4-i is a direct consequence of Lemma 3 and of the fact that $f(\cdot)$ is bounded on $[a,b]$. To establish the second part of the Lemma, consider $M_T(\lambda)=\sum_{t=2}^{T} f(y_{t-1})\mathbb{I}\left(y_{t-1}\geq\lambda\right)\varepsilon_t$ and $\lambda_1<\lambda_2$, the other cases being similar. Note that $M_T(\lambda_2)-M_T(\lambda_1)$ has a martingale structure with respect to the sigma-field $\mathcal{F}_t=\sigma(\varepsilon_t,\varepsilon_{t-1},\ldots)$. The Burkholder inequality (see e.g. Chow and Teicher [1988]) therefore yields, y_{t-1} and ε_t being independent,

$$\begin{aligned}
&\mathbb{E}\left(M_T(\lambda_2)-M_T(\lambda_1)\right)^4\\
\leq\;& 6(4)^{3/2}\mathbb{E}\left(\frac{1}{T^{1/2}}\sum_{t=2}^{T} f^2(y_{t-1})\mathbb{I}(\lambda_1\leq y_{t-1}<\lambda_2)\varepsilon_t^2\right)^2\\
\leq\;& 6(4)^{3/2}\sup_{y\in[a,b]} f^4(y)\mathbb{E}\left(\frac{1}{T^{1/2}}\sum_{t=2}^{T}\mathbb{I}(\lambda_1\leq y_{t-1}<\lambda_2)\varepsilon_t^2\right)^2\\
=\;& \frac{C}{T}\sum_{t=2}^{T}\mathbb{E}\left[\mathbb{I}(\lambda_1\leq y_{t-1}<\lambda_2)\varepsilon_t^4\right]+\frac{2C}{T}\sum_{t=2}^{T-1}\mathbb{E}\left[\mathbb{I}(\lambda_1\leq y_{t-1}<\lambda_2)\varepsilon_t^2\sum_{i=t+1}^{T}\mathbb{I}(\lambda_1\leq y_{i-1}<\lambda_2)\varepsilon_i^2\right]\\
\leq\;& \frac{C}{T}\sum_{t=2}^{T}\mathbb{E}\left[\mathbb{I}(\lambda_1\leq y_{t-1}<\lambda_2)\right]\left(\mathbb{E}\varepsilon_t^4+2\mathbb{E}^2\varepsilon_t^2\right) \qquad\text{(A.5)}\\
&+\frac{2C}{T}\sum_{t=2}^{T-1}\mathbb{E}\left[\mathbb{I}(\lambda_1\leq y_{t-1}<\lambda_2)\varepsilon_t^2\sum_{i=t+2}^{T}\mathbb{E}\left[\mathbb{I}(\lambda_1\leq y_{i-1}<\lambda_2)\,|\mathcal{F}_t\right]\mathbb{E}\varepsilon_i^2\right]. \qquad\text{(A.6)}
\end{aligned}$$

(A.5) can be bounded as (A.3), and (A.6) as (A.4) using the decomposition $y_{i-1}=\widetilde{y}_{t,i-1}+\overline{y}_{t,i-1}$, observing that $y_{t-1},\varepsilon_t,\overline{y}_{t,i-1}\in\mathcal{F}_t$ and are independent of $\widetilde{y}_{t,i-1}$. It then follows that $\mathbb{E}\left(M_T(\lambda_2)-M_T(\lambda_1)\right)^4\leq C\mathbb{E}\left(S_T(\lambda_2)-S_T(\lambda_1)\right)^2$ and the Lemma is proven. □

We are now able to prove Proposition 2.

Proof of Proposition 2

Assume without loss of generality that $[a,b] = [0,1]$ and $\Lambda = [\lambda, +\infty)$. Recall that the asymptotic equicontinuity of $\{X_T(\lambda)\}_{\lambda\in[0,1]}$ means that

$$\lim_{\delta\downarrow 0}\limsup_{T\to\infty}\mathbb{P}\left(\sup_{|\lambda_1-\lambda_2|\le\delta}|X_T(\lambda_2)-X_T(\lambda_1)|\ge\epsilon\right)=0 \text{ for any } \epsilon>0,$$

where the sup above is over all λ_1,λ_2 in $[0,1]$.

We begin with $\{S_T(\lambda)\}_{\lambda\in[0,1]}$. Observe that for λ_1 and λ_2 in $[0,1]$, $|S_T(\lambda_2)-S_T(\lambda_1)|\le \sup_{y\in[0,1]}|f(y)||s_T(\lambda_2)-s_T(\lambda_1)|$ so that it suffices to establish the stochastic equicontinuity of $\{s_T(\lambda)\}_{\lambda\in[0,1]}$. Let $\delta>0$ be such that $1/\delta$ is an integer number and consider the piecewise constant process over deterministic intervals $\widetilde{s}_T(\lambda)=\widetilde{s}_T(\lambda;\delta)=s_T(i\delta)$ if $\lambda\in[(i-1)\delta,i\delta)$ for $i=1,\dots,1/\delta-1$, $\lambda\in[(1/\delta-1)\delta,1]$ for $i=1/\delta$. Because $s_T(\cdot)$ is monotonous, we have $\sup_{\lambda\in[0,1]}|s_T(\lambda)-\widetilde{s}_T(\lambda)|\le\max_{1\le i\le 1/\delta}|s_T(i\delta)-s_T[(i-1)\delta]|$ and then

$$\begin{aligned}
\sup_{|\lambda_2-\lambda_1|\le\delta}|s_T(\lambda_2)-s_T(\lambda_1)| &= \sup_{|\lambda_2-\lambda_1|\le\delta}|\widetilde{s}_T(\lambda_2)-\widetilde{s}_T(\lambda_1)+(s_T(\lambda_2)-\widetilde{s}_T(\lambda_2))-(s_T(\lambda_1)-\widetilde{s}_T(\lambda_1))|\\
&\le \sup_{|\lambda_2-\lambda_1|\le\delta}|\widetilde{s}_T(\lambda_2)-\widetilde{s}_T(\lambda_1)|+2\sup_{\lambda\in[0,1]}|s_T(\lambda)-\widetilde{s}_T(\lambda)|\\
&= 3\max_{1\le i\le 1/\delta}|s_T(i\delta)-s_T[(i-1)\delta]|\ .
\end{aligned}$$

Therefore the Chebychev inequality and Lemma 3 yield

$$\begin{aligned}
\mathbb{P}\left(\sup_{|\lambda_2-\lambda_1|\le\delta}|s_T(\lambda_2)-s_T(\lambda_1)|\ge\epsilon\right) &\le \mathbb{P}\left(3\max_{1\le i\le 1/\delta}|s_T(i\delta)-s_T[(i-1)\delta]|\ge\epsilon\right)\\
&\le \sum_{i=1}^{1/\delta}\mathbb{P}\left(3\max_{1\le i\le 1/\delta}|s_T(i\delta)-s_T[(i-1)\delta]|\ge\epsilon\right)\\
&\le \frac{9}{\epsilon^2}\frac{C}{\delta}\left(\delta+\frac{\sqrt{\delta}}{\sqrt{T}}\right)^2=\frac{C}{\epsilon^2}\left(\sqrt{\delta}+\frac{1}{\sqrt{T}}\right)^2
\end{aligned}$$

and the process $\{S_T(\lambda)\}_{\lambda\in[0,1]}$ is asymptotically equicontinuous.

We now consider $\{M_T(\lambda)\}_{\lambda\in[0,1]}$ which is more difficult to study. Let $h=h_T$ be such that $1/h$ is an integer number. Define $\widetilde{M}_T(\lambda)=\widetilde{M}_T(\lambda;h)$ as $\widetilde{s}_T(\lambda)=\widetilde{s}_T(\lambda;h)$ and denote $J_i=[(i-1)h,ih)$, $J_{1/h}=[(1/h-1)h,1]$. We have

$$\begin{aligned}
\sup_{\lambda\in[0,1]}\left|M_T(\lambda)-\widetilde{M}_T(\lambda)\right| &= \max_{1\le i\le h}\max_{\lambda\in J_i}\left|\frac{1}{T^{1/4}}\sum_{t=2}^{T}f(y_{t-1})\mathbb{I}\left(\lambda\le y_{t-1}\le ih\right)\varepsilon_t\right|\\
&\le \sup_{y\in[0,1]}|f(y)|T^{1/4}\max_{1\le t\le T}|\varepsilon_t|\max_{1\le i\le 1/h}|s_T(ih)-s_T[(i-1)h]|\\
&= O_{\mathbb{P}}\left(T^{\frac{1}{4}+\frac{1}{4+s}}\right)\max_{1\le i\le 1/h}|s_T(ih)-s_T[(i-1)h]|
\end{aligned}$$

by Lemma 1, with, by Lemma 3,

$$\begin{aligned}\mathbb{E}\left(\max_{1\leq i\leq 1/h}|s_T(ih)-s_T[(i-1)h]|\right)^2 &\leq \sum_{i=1}^{1/h}\mathbb{E}\left(|s_T(ih)-s_T[(i-1)h]|\right)^2\\ &\leq \frac{C}{h}\left(h+2\frac{\sqrt{h}}{\sqrt{T}}\right)^2 = C\left(\sqrt{h}+\frac{2}{\sqrt{T}}\right)^2 .\end{aligned}$$

Therefore, taking $h = h_T$ of exact order $1/T$ yields that $\sup_{\lambda\in[0,1]}\left|M_T(\lambda)-\widetilde{M}_T(\lambda)\right| = o_{\mathbb{P}}(1)$, and the asymptotic equicontinuity of $\{M_T(\lambda)\}_{\lambda\in[0,1]}$ will be a consequence of the one of $\{\widetilde{M}_T(\lambda)\}_{\lambda\in[0,1]}$. We wish to apply a maximal inequality for separable processes, see e.g. Theorem 2.2.4 in Van der Vaart and Wellner [1996]. Observe that $\{\widetilde{M}_T(\lambda)\}_{\lambda\in[0,1]}$ is a piecewise constant cadlag process with, by Lemma 4-*ii*,

$$\mathbb{E}^{1/4}\left(\widetilde{M}_T(\lambda_2)-\widetilde{M}_T(\lambda_1)\right)^4 \leq \widetilde{d}_{4T}(\lambda_1,\lambda_2) = d_{4T}(\widetilde{\lambda}_1,\widetilde{\lambda}_2) \text{ with } \widetilde{\lambda} = ih \text{ if } \lambda\in J_i.$$

Note that $\widetilde{d}_{4T}$ is a semimetric (i.e. vanishes when $\lambda_1 = \lambda_2$, is symmetric and verifies the triangular inequality). For $x > 0$, let $\widetilde{N}_{4T}(x)$ be the $\widetilde{d}_{4T}$-covering number of size x of $[0,1]$, i.e. the minimal number of $\widetilde{d}_{4T}$-balls of radius x needed to cover $[0,1]$. The maximal inequality of Theorem 2.2.4 in Van der Vaart and Wellner [1996] yields for any positive η, δ[14]

$$\mathbb{E}^{1/4}\sup_{\widetilde{d}_{4T}(\lambda_1,\lambda_2)\leq\delta}\left|\widetilde{M}_T(\lambda_2)-\widetilde{M}_T(\lambda_1)\right|^4 \leq C'\left[\int_0^\eta \widetilde{N}_{4T}^{1/4}\left(\frac{x}{2}\right)dx + \delta\widetilde{N}_{4T}^{1/2}\left(\frac{\eta}{2}\right)\right] ,$$

where C' is a constant which can vary from line to line in the inequalities below. We now find a suitable bound for the covering numbers $\widetilde{N}_{4T}$. Let C be from Lemma 4-*ii* and $x_h = Cx'_h$ be such that $\left(\frac{\sqrt{Tx'_h+1}-1}{\sqrt{T}}\right)^4 = h$, that is

$$x_h = C\frac{\left(\sqrt{T}h^{1/4}+1\right)^2-1}{T} = C\left[h^{1/2}+2\left(\frac{h}{\sqrt{T}}\right)^{1/4}\right] \sim C\sqrt{h} \text{ with } h = h_T \text{ of exact order } 1/T.$$

The definitions of x_h and d_{4T} are such that a covering of $\widetilde{d}_{4T}$-balls with radius $x < x_h$ is given by the J_i's, $i = 1,\ldots,1/h$. This gives that $\widetilde{N}_{4T}(x) \leq 1/h$ for $x < x_h$. We consider now $x \geq x_h$. Observe that the d_{4T}-ball of center $\widetilde{\lambda}_0 = i_0 h$ with radius Cx is the set of $\lambda\in[0,1]$ with

$$C\left[\left(|\lambda-\widetilde{\lambda}_0|^{1/4}+\frac{1}{\sqrt{T}}\right)^2-\frac{1}{T}\right] \leq Cx \text{ i.e. } |\lambda-\widetilde{\lambda}_0| \leq \left[\left(x+\frac{1}{T}\right)^{1/2}-\frac{1}{\sqrt{T}}\right]^4 = \left(\frac{\sqrt{Tx+1}-1}{\sqrt{T}}\right)^4 .$$

Since a $\widetilde{d}_{4T}$-covering is a covering with unions of contiguous J_i's, we have

$$\widetilde{N}_{4T}^{1/4}(x) \leq \frac{\sqrt{T}}{\sqrt{Tx/C+1}-1} \text{ for } x\geq x_h,\ \widetilde{N}_{4T}^{1/4}(x) \leq h^{-1/4} \text{ for } x < x_h.$$

[14]Applying this result directly to the process $\{M_T(\lambda)\}_{\lambda\in[0,1]}$ yields a diverging integral at 0 as can be seen from the bounds below.

Since $x_h \sim C\sqrt{h} \to 0$ with $Tx_h \to \infty$ and $Tx_h = o(T)$, the maximal inequality yields for T large enough such that $\eta \geq x_h$,

$$
\begin{aligned}
&\mathbb{E}^{1/4} \sup_{\widetilde{d}_{4T}(\lambda_1,\lambda_2)\leq\delta} \left|\widetilde{M}_T(\lambda_2) - \widetilde{M}_T(\lambda_1)\right|^4 \\
&\quad \leq \quad C'\left[h^{-1/4}x_h + \int_{x_h}^{\eta} \frac{\sqrt{T}}{\sqrt{Tx/(2C)+1}-1}dx + \delta\frac{T}{(\sqrt{T\eta/(2C)+1}-1)^2}\right] \\
&\quad \overset{x=2Cu/T}{=} \quad C'\left[o(1) + \frac{1}{\sqrt{T}}\int_{Tx_h/(2C)}^{T\eta/(2C)} \frac{du}{\sqrt{u+1}-1} + 2C\frac{\delta}{\eta}[1+o(1)]\right] \\
&\quad \overset{T\to\infty}{\to} \quad C'\left((\eta/2C)^{1/2}/2 + 2C\frac{\delta}{\eta}\right), \\
&\qquad\qquad \text{using } \int_{Tx_h/(2C)}^{T\eta/(2C)} \frac{du}{\sqrt{u+1}-1} \overset{T\to\infty}{\sim} \frac{1}{2}\sqrt{T\eta/(2C)} \text{ since } \tfrac{1}{\sqrt{u+1}-1} \overset{u\to\infty}{\sim} u^{-1/2}\ .
\end{aligned}
$$

Recall now that $d_{4T}(\widetilde{\lambda_1,\lambda_2}) \leq \delta$ is equivalent to

$$
\left|\widetilde{\lambda}_2 - \widetilde{\lambda}_1\right| \leq \left(\frac{\sqrt{T\delta/C+1}-1}{\sqrt{T}}\right)^4 \overset{T\to\infty}{\sim} \delta^2/C^2 \text{ for } \delta > 0,
$$

so that the definition of $\widetilde{\lambda}$ yields, for T large enough and taking $\eta = \sqrt{\delta}$,

$$
\mathbb{E}^{1/4} \sup_{|\lambda_2-\lambda_1|^{1/2}\leq\delta/(2C)} \left|\widetilde{M}_T(\lambda_2) - \widetilde{M}_T(\lambda_1)\right|^4 \leq \mathbb{E}^{1/4} \sup_{\widetilde{d}_{4T}(\lambda_1,\lambda_2)\leq\delta} \left|\widetilde{M}_T(\lambda_2) - \widetilde{M}_T(\lambda_1)\right|^4 \overset{T\to T}{\to} C'\left(\delta^{1/4}+\delta^{1/2}\right)\ .
$$

Because $\lim_{\delta\downarrow 0}(\delta^{1/4}+\delta^{1/2}) = 0$, $\{\widetilde{M}_T(\lambda)\}_{\lambda\in[0,1]}$ is asymptotically equicontinuous by the Markov inequality, as $\{M_T(\lambda)\}_{\lambda\in[0,1]}$. □

Proof of Proposition 1

Under H_0, $C_{11} = O_{\mathbb{P}}(1)$ and $M_1 = O_{\mathbb{P}}(1)$.

We first show that $C_{21} = o_{\mathbb{P}}(1)$. We consider first the entries of C_{21} of the inner regime. We have, using Lemma 1 and Theorem 3.2 of Park and Phillips [2001], $k = 1, \ldots, p$,

$$
\left|\frac{1}{T^{3/4}}\sum_{t=p+1}^{T} h(y_{t-1})\mathbb{I}(|y_{t-1}|<\lambda)u_{t-k}\right| \leq \frac{\max_{t\leq T}|u_t|}{T^{1/4}}\left|\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} h(y_{t-1})\mathbb{I}(|y_{t-1}|<\lambda)\right| = o_{\mathbb{P}}(1),
$$

with $h(y) = 1$ or $h(y) = y$ here. For the entries of C_{21} of the outer regime, observe

$$
\begin{aligned}
&\frac{1}{T^{3/2}}\sum_{t=p+1}^{T} y_{t-1}\mathbb{I}(|y_{t-1}|\geq\lambda)u_{t-k} \\
&\quad = \quad \frac{1}{T^{3/2}}\sum_{t=p+1}^{T} y_{t-1}u_{t-k} - \frac{1}{T^{3/2}}\sum_{t=p+1}^{T} y_{t-1}\mathbb{I}(|y_{t-1}|<\lambda)u_{t-k}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{T^{3/2}} \sum_{t=p+1}^{T} y_{t-1}u_{t-k} + o_{\mathbb{P}}(1) \text{ by the equation above,} \\
&= \frac{1}{T^{3/2}} \sum_{t=p+1}^{T} u_{t-1}u_{t-k} + \cdots + \frac{1}{T^{3/2}} \sum_{t=p+1}^{T} u_{t-k}^2 + \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \frac{y_{t-k-1}}{\sqrt{T}} \frac{u_{t-k}}{\sqrt{T}} + o_{\mathbb{P}}(1) \\
&= o_{\mathbb{P}}(1) \text{ see Proposition 17.3-(c,e) in Hamilton [1994].}
\end{aligned}
$$

It remains to study the items

$$
\frac{1}{T} \sum_{t=p+1}^{T} \operatorname{sgn}_\lambda(y_{t-1})u_{t-k} = \sum_{j=0}^{\infty} \frac{\pi_j}{T} \sum_{t=p+1}^{T} \operatorname{sgn}_\lambda(y_{t-1})\varepsilon_{t-k-j}, k = 1, \ldots, p
$$

to find the order of C_{21}. Observe that

$$
\mathbb{E} \left| \sum_{j=J_0+1}^{\infty} \frac{\pi_j}{T} \sum_{t=p+1}^{T} \operatorname{sgn}_\lambda(y_{t-1})\varepsilon_{t-k-j} \right| \leq \sum_{j=J_0+1}^{\infty} \frac{|\pi_j|}{T} \sum_{t=p+1}^{T} \mathbb{E}|\varepsilon_{t-k-j}| \leq C \sum_{j=J_0+1}^{\infty} |\pi_j|
$$

which can be arbitrarily small by taking J_0 large enough. Then $T^{-1}\sum \operatorname{sgn}_\lambda(y_{t-1})u_{t-k} = o_{\mathbb{P}}(1)$ for $k = 1, \ldots, p$ is a consequence of $T^{-1}\sum \operatorname{sgn}_\lambda(y_{t-1})\varepsilon_{t-q} = o_{\mathbb{P}}(1)$ for any finite q. Recall that $\operatorname{sgn}_\lambda(y) = \mathbb{I}(y \geq \lambda) - \mathbb{I}(-y \geq -\lambda)$ and note that

$$
\begin{aligned}
|\mathbb{I}(y_1 + y_2 \geq \lambda) - \mathbb{I}(y_1 \geq \lambda)| &= |\mathbb{I}(y_1 + y_2 - \lambda < 0, y_1 - \lambda \geq 0) + \mathbb{I}(y_1 + y_2 - \lambda \geq 0, y_1 - \lambda < 0)| \\
&\leq \mathbb{I}(|y_1 - \lambda| \leq |y_2|) \,,
\end{aligned}
$$

and then $|\operatorname{sgn}_\lambda(y_1 + y_2) - \operatorname{sgn}_\lambda(y_1)| \leq 2\mathbb{I}(|y_1| \leq |y_2 - \lambda|)$. Therefore, writing $y_t = y_{t-q-1} + y_t - y_{t-q-1}$ with $y_t - y_{t-q-1} = u_t + \cdots + u_{t-q}$, $u_{t-q} = 0$ for $t - q < 1$ and $y_{t-q-1} = 0$ for $t < q + 2$, and Lemma 1 yields

$$
\begin{aligned}
&\left| \frac{1}{T} \sum_{t=p+1}^{T} \operatorname{sgn}_\lambda(y_{t-1})\varepsilon_{t-q} \right| \\
&\leq \left| \frac{1}{T} \sum_{t=p+1}^{T} \operatorname{sgn}_\lambda(y_{t-q-1})\varepsilon_{t-q} \right| + \frac{2}{T} \sum_{t=p+1}^{T} \mathbb{I}(|y_{t-q-1} - \lambda| \leq |y_t - y_{t-q-1}|)|\varepsilon_{t-q}| \\
&\leq \left| \frac{1}{T} \sum_{t=p+1}^{T} \operatorname{sgn}_\lambda(y_{t-q-1})\varepsilon_{t-q} \right| && \text{(A.7)} \\
&\quad + \frac{2}{T} \sum_{t=p+1}^{T} \mathbb{I}\left(|y_{t-q-1} - \lambda| \leq CT^{1/(4+s)}\right) |\varepsilon_{t-q}| && \text{(A.8)}
\end{aligned}
$$

with a probability which can be made arbitrarily close to 1 by choosing C large enough. $\sum_{t=p+1}^{T} \operatorname{sgn}_\lambda(y_{t-q-1})\varepsilon_{t-q}$ is a martingale with a variance $\sum_{t=p+1}^{T} \mathbb{E}[\operatorname{sgn}_\lambda^2(y_{t-q-1}\varepsilon_{t-q}^2] \leq \sigma^2 T$, so that (A.7) is $O_{\mathbb{P}}(1/\sqrt{T})$. For (A.8), Lemma 2 yields

$$
\mathbb{E}\left[\frac{1}{T} \sum_{t=p+1}^{T} \mathbb{I}\left(|y_{t-q-1} - \lambda| \leq CT^{1/(4+s)}\right) |\varepsilon_{t-q}|\right] \leq \frac{1}{T} \sum_{t=q+2}^{T} \mathbb{E}|\varepsilon_{t-q}| \int_{-CT^{1/(4+s)}}^{CT^{1/(4+s)}} f_{t-q-1}(y - \lambda)dy
$$

$$
\begin{aligned}
&\leq \frac{CT^{1/(4+s)}}{T} \sum_{t=q+2}^{T} \frac{1}{\sqrt{t-q}} \\
&= O\left(\frac{T^{1/2+1/(4+s)}}{T}\right) = o(1) ,
\end{aligned}
$$

showing that $T^{-1}\sum_{t=p+1}^{T} \mathrm{sgn}_\lambda(y_{t-q-1})\varepsilon_{t-q} = o_{\mathbb{P}}(1)$. As a consequence, $C_{21} = o_{\mathbb{P}}(1)$.

We now derive the limit distribution of $(C_{22}(\lambda), M_2(\lambda))$, the study of $(C_{22}(\lambda_j), M_2(\lambda_j), j = 1, \ldots, J)$ being similarly studied using the Cramer-Wold device in (2.5). Due to the standardization by $(1/T, 1/\sqrt{T})$ in $(C_{22}^o(\lambda), M_2^o(\lambda))$ and by $(1/\sqrt{T}, 1/T^{1/4})$ in the processes $(S_T(\lambda), M_T(\lambda))$ of Lemma 3, Lemma 3 yields that $(C_{22}^o(\lambda), M_2^o(\lambda)) = (C_{22}^o(0), M_2^o(0)) + o_{\mathbb{P}}(1)$. Note also that

$$
\int \mathbb{I}(|\delta y| \leq \lambda) dy = 2\frac{\lambda}{|\delta|}, \int \delta y \mathbb{I}(|\delta y| \leq \lambda) dy = 0, \int \delta^2 y^2 \mathbb{I}(|\delta y| \leq \lambda) dy = \frac{2}{3}\frac{\lambda^3}{|\delta|}.
$$

Writing $y_{t-1} = \delta y_{t-1}/\delta$, Theorems 3.1 and 3.2 of Park and Phillips [2001] therefore give that, jointly

$$
C_{22}(\lambda) \xrightarrow{d} \begin{bmatrix} 1 & -\int_0^1 |\delta W(r)|\, dr & 0 & 0 \\ -\int_0^1 |\delta W(r)|\, dr & \int_0^1 \delta^2 W^2(r) dr & 0 & 0 \\ 0 & 0 & \frac{2\lambda}{|\delta|}\ell_W(0,1) & 0 \\ 0 & 0 & 0 & \frac{2\lambda^3}{3|\delta|}\ell_W(0,1) \end{bmatrix},
$$

$$
M_2(\lambda) \xrightarrow{d} \sigma \begin{bmatrix} -\int_0^1 \mathrm{sgn}\left(\delta W(r)\right) dW(r) \\ \int_0^1 \delta W(r) dW(r) \\ B\left(\frac{2\lambda}{|\delta|}\ell_W(0,1)\right) \\ B\left(\frac{2\lambda^3}{3|\delta|}\ell_W(0,1)\right) \end{bmatrix}.
$$

We now derive the asymptotic distribution of the Wald statistic. Let $r = (0,1)'$. Observe that $\widehat{\sigma}^2(\lambda) \xrightarrow{\mathbb{P}} \sigma^2$ since $M'(\lambda)C^{-1}(\lambda)M(\lambda)/T = o_{\mathbb{P}}(1)$, and $C_{21} = o_{\mathbb{P}}(1)$. It then follows from (A.2) that

$$
\begin{aligned}
W_T(\lambda) &= \frac{1}{\sigma^2 + o_{\mathbb{P}}(1)} \left[\frac{\left(r' \left[C_{22}^o(\lambda)\right]^{-1} M_2^o(\lambda)\right)^2}{r' \left[C_{22}^o(\lambda)\right]^{-1} r} + \frac{\left(r' \left[C_{22}^i(\lambda)\right]^{-1} M_2^i(\lambda)\right)^2}{r' \left[C_{22}^i(\lambda)\right]^{-1} r} \right] + o_{\mathbb{P}}(1) \\
&\xrightarrow{d} \frac{\left(\int_0^1 \delta W(r) dW(r) - \int_0^1 |\delta W(r)|\, dr \int_0^1 \mathrm{sgn}\left(\delta W(r)\right) dW(r)\right)^2}{\int_0^1 \delta^2 W^2(r) dr - \left(\int_0^1 |\delta W(r)|\, dr\right)^2} + \frac{B^2\left(\frac{2\lambda^3}{3|\delta|}\ell_W(0,1)\right)}{\frac{2\lambda^3}{3|\delta|}\ell_W(0,1)} \\
&\overset{d}{=} \frac{\left(\int_0^1 W(r) dW(r) - \int_0^1 |W(r)|\, dr \int_0^1 \mathrm{sgn}\left(W(r)\right) dW(r)\right)^2}{\int_0^1 W^2(r) dr - \left(\int_0^1 |W(r)|\, dr\right)^2} + B^2(\lambda^3)/\lambda^3 ,
\end{aligned}
$$

where the equality in distribution above holds jointly with respect to λ in $\mathbb{R}_+^*$, and comes from standard scaling properties of the Brownian motion and of the sign function, the fact that $\delta = \sigma/(1 - a(1)) > 0$ (since the roots of $1 - a(x)$ lies outside the unit circle with $1 - a(0) = 1$, so that $1 - a(1) > 0$ by continuity), together with the independence between $\{W(s)\}_{s \in \mathbb{R}_+}$ and $\{B(s)\}_{s \in \mathbb{R}_+}$. It is easily seen that the three tests statistics have the same limit distribution under H_0. □

Proof of Corollary 1

We first check the asymptotic equicontinuity of the entries of $C(\lambda)$ which depend upon λ. Consider first C_{21} and let $h(y;\lambda)$ be $\mathrm{sgn}(\lambda)$, $y\mathbb{I}(|y| \geq \lambda)$, $\mathbb{I}(|y| < \lambda)$ or $y\mathbb{I}(|y| < \lambda)$. We have, for $k = 1, \ldots, p$,

$$\begin{aligned} \left| \frac{1}{T} \sum_{t=p+1}^{T} h(y_{t-1};\lambda) u_{t-k} \right| &\leq \frac{\max_{t\leq T} |u_t|}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} |h(y_{t-1};\lambda)| \\ &= o_{\mathbb{P}}(1) \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} |h(y_{t-1};\lambda)| \end{aligned}$$

by Lemma 1, and then $\{C_{21}(\lambda)\}_{\lambda\in[a,b]}$, is asymptotically equicontinuous by Proposition 2 due to a standardization smaller or equal to $1/T$ in front of the sums involved in this submatrix. The asymptotic equicontinuity of the other entries of $C(\lambda)$, and then the asymptotic equicontinuity of $\{C(\lambda)\}_{\lambda\in[a,b]}$ also follows from Proposition 2, which also implies the asymptotic equicontinuity of $\{M(\lambda)\}_{\lambda\in[a,b]}$. Since the limit in distribution of $\inf_{\lambda\in[a,b]} \det C(\lambda)$ and $\inf_{\lambda\in[a,b]} \det \left(RC^{-1}(\lambda)R\right)$ are bounded away from 0 with probability 1, (A.2) shows that the asymptotic equicontinuity of the Wald statistic is proven. The other statistics easily follow. □

Proof of Theorem 1

As above, we consider only the Wald statistic which is asymptotically equivalent to the other ones. Because $\mathbb{I}(y_{t-1} \geq \lambda) = 1 - \mathbb{I}(y_{t-1} < \lambda)$, $\{W_T(\lambda)\}_{\lambda\in[a,b]}$ writes as a continuous functional of cadlag functions. It then follows from Proposition 1 and Corollary 1 that

$$\begin{bmatrix} (\underline{\lambda}_T, \overline{\lambda}_T) \\ \{W_T(\lambda)\}_{\lambda\in[a,b]} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} (\underline{\lambda}, \overline{\lambda}) \\ \{\zeta_1^2 + \zeta_2^2(\lambda)\}_{\lambda\in[a,b]} \end{bmatrix}$$

for any compact interval $[a,b]$ with $a > 0$, and where weak convergence is with respect to the Skorohod topology of $D[a,b]$, see Billingsley [1968], p.111 and further. Recall also that convergence in $D[a,b]$ implies uniform convergence when the limit is continuous. Due to the a.s. continuity of the limit process $\{\zeta_1 + \zeta_2(\lambda)\}_{\lambda\in[a,b]}$, we then have

$$\max_{\lambda\in\Lambda_T\cap[a,b]} W_T(\lambda) \xrightarrow{d} \sup_{\lambda\in\Lambda_T\cap[a,b]} \left(\zeta_1^2 + \zeta_2^2(\lambda)\right) .$$

Observe now that

$$\mathbb{P}\left(\max_{\lambda\in\Lambda_T} W_T(\lambda) \neq \max_{\lambda\in\Lambda_T\cap[a,b]} W_T(\lambda) \right) \leq 1 - \mathbb{P}\left(a \leq \underline{\lambda}_T \leq \overline{\lambda}_T \leq b\right)$$

$$\mathbb{P}\left(\sup_{\lambda\in[\underline{\lambda},\overline{\lambda}]} \left(\zeta_1^2 + \zeta_2^2(\lambda)\right) \neq \sup_{\lambda\in[\underline{\lambda},\overline{\lambda}]\cap[a,b]} \left(\zeta_1^2 + \zeta_2^2(\lambda)\right) \right) \leq 1 - \mathbb{P}\left(a \leq \underline{\lambda} \leq \overline{\lambda} \leq b\right) .$$

Note that $\zeta_2(\lambda)$ is a.s. continuous over $\mathbb{R}_+^*$. Therefore, taking an interval $[a,b] \subset \mathbb{R}_+^*$ with $\liminf_{T\to\infty} \mathbb{P}(a \leq \underline{\lambda} \leq \overline{\lambda} \leq b)$ arbitrary close to 1 yield

$$\sup_{\lambda \in [\underline{\lambda}_T, \overline{\lambda}_T]} W_T(\lambda) \xrightarrow{d} \zeta_1^2 + \sup_{\lambda \in [\underline{\lambda}, \overline{\lambda}]} \zeta_2^2(\lambda) \text{ with, since } (\underline{\lambda}, \overline{\lambda}) \text{ and } B(\cdot) \text{ are independent,}$$

$$\left(\zeta_1^2, \sup_{\lambda \in [\underline{\lambda}, \overline{\lambda}]} \zeta_2^2(\lambda)\right) = \left(\zeta_1^2, \sup_{t \in [1, \overline{\lambda}/\underline{\lambda}]} \frac{B^2\left(\underline{\lambda}^3 t^3\right)}{\underline{\lambda}^3 t^3}\right) \stackrel{d}{=} \left(\zeta_1^2, \sup_{t \in [1, \overline{\lambda}/\underline{\lambda}]} \frac{B^2\left(t^3\right)}{t^3}\right).$$

The test statistic is then asymptotically pivotal under the conditions of the Theorem. $\square$

Proof of Theorem 2

Let μ be the common mean of the stationary y_t's. We first show that, for any stationary $\{y_t\}$, for any q

> $\mathcal{P}(q)$: if $(y_1, \ldots, y_q, y_{q+1})$ has a continuous distribution, the (auto)regression $b_q(L)(y_t - \mu) = v_{q,t} = v_t$ of $y_t - \mu$ on $y_{t-1} - \mu, \ldots, y_{t-q} - \mu$ is such that $b_q(1) \neq 0$.

Assume without loss of generality that $\mu = 0$. For $q = 1$, the autoregression writes $y_t - b_1 y_{t-1} = v_t$ with $b_1 = \mathrm{Cov}(y_t, y_{t-1})/\mathrm{Var}(y_{t-1}) = \mathrm{Corr}(y_t, y_{t-1})$ where $\mathrm{Corr}(A,B)$ is the correlation coefficient of A and B, and since $\mathrm{Var}(y_t) = \mathrm{Var}(y_{t-1})$ by stationarity. Therefore $b_1(1) = 0$ yields that y_t and y_{t-1} have a correlation of 1 and then $v_t = 0$ so that $y_t = y_{t-1}$, a contradiction. Therefore $\mathcal{P}(1)$ is true. Let us now show that $\mathcal{P}(q)$ is true for any $q > 1$. Assume that $\mathcal{P}(q)$ is not true so that $b_q(L) = (1-L)\widetilde{b}_q(L)$. Observe now that $\{\widetilde{b}_q(L) y_t\}$ is stationary as a linear transformation of a stationary process, and that $(\widetilde{b}_q(L) y_t, \widetilde{b}_q(L) y_{t-1})$ has a continuous distribution since $\widetilde{b}_q(L)$ is of degree $q-1$. Note that $\{\widetilde{b}_q(L) y_t\}$ admits an autoregression of order 1 given by $\widetilde{b}_q(L) y_t = \widetilde{b}_q(L) y_{t-1} + v_t$ with a unit root 1, contradicting $\mathcal{P}(1)$. Therefore $\mathcal{P}(q)$ is true for any $q \geq 1$.

We now show that $W_T(\lambda^*)$ diverges for a well chosen $\lambda^* > 0$, showing that $\sup_{\lambda \in \Lambda_T} W_T(\lambda)$ diverges since $\lambda^* \in \Lambda_T$ with a probability tending to 1 under the conditions on $(\underline{\lambda}_T, \overline{\lambda}_T)$ of Theorem 2, the case of the other test statistics being similar. For any $\lambda \in \mathbb{R}_+$, let $(a_\lambda(L), \mu_1(L), \rho_1(L), \mu_2(\lambda), \rho_2(\lambda))$ be the coefficients of the TAR model (2.1) of order p associated with $\{y_t\}$ by a theoretical linear regression. Setting $\mu_1(\lambda) = \rho_1(\lambda) = 0$ if λ is above the common support of the y_t's shows that the pseudo TAR model is uniquely defined since $y_1, \ldots, y_{p+1}$ has a continuous distribution. Observe that $\rho_2(\lambda) \to b_p(1) \neq 0$ when $\lambda \to +\infty$ by the Lebesgue Dominated Convergence Theorem. Therefore taking λ^* large enough and ergodicity yield that $\widehat{\rho}_2(\lambda^*) \xrightarrow{\mathbb{P}} \rho_2(\lambda^*) \neq 0$ ensuring that $W_T(\lambda^*)$ will diverge under the convention of footnote 2 since $V_T^{-1}(\lambda^*)$ in (2.2) is of order T under the alternative. $\square$