ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Adaptive consistent unit-root tests based on autoregressive threshold model

Frédérique Bec[a], Alain Guay[b,c], Emmanuel Guerre[d,*]

[a]*THEMA, Université de Cergy-Pontoise, France and CREST-ENSAE, Timbre J120—3, av. Pierre Larousse, 92245 Malakoff Cedex, France*
[b]*CIRPEE, Université du Québec à Montréal, Canada*
[c]*CIREQ, Canada*
[d]*Department of Economics, Queen Mary, University of London, Mile End, E1 4NS London, UK*

Received 12 February 2004; accepted 30 March 2007

Abstract

This paper proposes SupWald tests from a threshold autoregressive model computed with an adaptive set of thresholds. Simple examples of adaptive threshold sets are given. A second contribution of the paper is a general asymptotic null limit theory when the threshold variable is a level variable. We obtain a pivotal null limiting distribution under some simple conditions for bounded or asymptotically unbounded thresholds. Our general approach is flexible enough to allow a choice of the auxiliary threshold model or of the threshold set involved in the test specifically designed for nonlinear stationary alternatives relevant for macroeconomic and financial topics involving arbitrage in presence of transaction costs. A Monte-Carlo study and an application to the interest rates spread for French, German, New-Zealander and US post-1980 monthly data illustrate the ability of the adaptive SupWald tests to reject unit-root when the ADF does not.

JEL classification: C12; C22; C32; E43

Keywords: Unit-root test; Threshold autoregressive model; Term structure of interest rates

1. Introduction

A debate on the performance of linear unit-root tests to detect nonlinear stationary alternatives has recently grown in the econometric literature. Indeed, the presence of fixed adjustment costs, transaction costs or arbitrage boundaries can create nonlinear adjustments in economic variables quite close to nonstationarity. Economic policy characterized by discrete intervention to manage exchange rate, target zone or inflation–output targets could also induce such nonlinear dynamics. Empirical studies as Anderson (1997), Michael et al. (1997), Obstfeld and Taylor (1997) or Sollis et al. (2002) also argued for nonlinear dynamics. On the other hand, the simulation studies of Balke and Fomby (1997), Pippenger and Goering (1993) and Taylor (2001) have risen doubts about the power of standard linear unit-root tests against nonlinear stationary

*Corresponding author.
E-mail addresses: bec@ensae.fr (F. Bec), guay.alain@uqam.ca (A. Guay), e.guerre@qmul.ac.uk (E. Guerre).

0304-4076/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2007.05.011

alternatives. As a consequence, a fast developing branch of the econometric literature has proposed as a remedy to use an auxiliary nonlinear dynamic model in place of a linear autoregression to build a unit-root test. This includes among others a Threshold Autoregressive (TAR) specification as in Bec et al. (2004), Berben and van Dijk (1999), Caner and Hansen (2001), Enders and Granger (1998), Gonzalez and Gonzalo (1998), Kapetanios and Shin (2006), Seo (2003), Shin and Lee (2001, 2003), or a smooth transition autoregressive specification as in Kapetanios et al. (2003). A substantial difficulty is then that the threshold parameter is not identified under the null. Consequently, much attention has been focused on the null limiting distribution of threshold unit-root tests but, as seen from the previous references, consistency studies are limited to restricted classes of threshold alternatives. This contrasts with the augmented Dickey–Fuller (ADF) test which is consistent against general ergodic alternatives and is somehow paradoxical in view of the claimed power superiority of the nonlinear approach.

A first contribution of the present paper is to examine the construction of threshold unit root tests toward consistency and power comparison issues. In place of the linear autoregression of the ADF statistic, a general threshold specification is considered to serve as an auxiliary model to build a unit-root test. However, in such model, the true threshold is unknown. Our unit-root testing strategy is based on an adaptive set of thresholds which behaves differently under the null and the alternatives. As many of the references above, we propose a SupWald test $\text{SupWald}_T(\Lambda_T)$ which maximizes the Wald statistic over a set of thresholds Λ_T, T being the sample size. In previous works, a quantile choice of Λ_T ensuring a minimal percentage of observations in each regime was considered, see e.g. Caner and Hansen (2001). But, due to this restrictive quantile choice, nothing ensures consistency since such Λ_T does not necessarily contain a threshold associated with a diverging Wald statistic. Therefore, a more general construction of Λ_T should be considered to achieve consistency. Under the alternative, the set Λ_T should allow as many thresholds as possible, including diverging thresholds corresponding to not identified regimes that should typically be avoided under the null. We refer to this property as adaptation, a behavior that can be achieved by defining the boundaries of Λ_T as function of consistent unit-root test statistics. The claimed benefits of adaptation are twofold. First, this gives consistency against any (nonlinear) stationary ergodic alternatives. This finding clarifies in particular early critics on the possible inconsistency of threshold unit-root tests, see e.g. Balke and Fomby (1997). As a by-product of adaptation, we obtain bounds showing that the $\text{SupWald}_T(\Lambda_T)$ is asymptotically larger than the squared ADF statistic under the alternative, indicating so potential power improvements. Second, reconsidering usual quantile threshold sets can be useful to obtain test statistics with smaller critical values which would have better power properties. We give examples of adaptive, asymptotically unbounded or bounded, sets of thresholds Λ_T with boundaries depending upon the consistent ADF statistic. The unbounded example is a modification of the quantile-based Λ_T while the bounded example is new.

A second contribution is a general asymptotic theory under the null. Such a theory must cope in particular with random threshold sets Λ_T and give conditions ensuring a finite null pivotal limiting distribution. We consider a general 3-regime TAR specification as a baseline model. Following Bec et al. (2004), Berben and van Dijk (1999), Enders and Granger (1998), Kapetanios and Shin (2006) and Seo (2003), the lagged level variable is chosen as the threshold variable, which is therefore nonstationary under the null. This differs from the choice of Caner and Hansen (2001), Gonzalez and Gonzalo (1998) and Shin and Lee (2003) who consider an *ad hoc* stationary threshold. By contrast, our approach is in line with many macroeconomic or financial models involving arbitrage behavior in presence of transaction costs. Moreover, it yields a pivotal null limit distribution which simplifies the implementation of the test. Finding the null limiting distribution of such a SupWald test requires to establish a new functional version of the limit results of Park and Phillips (2001) which can be useful for other nonlinear specifications.

Finally, we compare small sample properties of the existing unit-root tests with the ones proposed in simulation experiments that illustrates the interest of adaptive SupWald tests compared to the linear ADF. An application to the yield spread dynamics illustrates the ability of adaptive SupWald tests to detect stationarity when the ADF does not.

The remainder of the paper is as follows. Section 2 introduces adaptation and provides examples of adaptive threshold. The consistency and the null limiting distribution results of the SupWald tests for a simple autoregression of order one is also presented. Section 3 extends those results to more general autoregression of order p and to more general auxiliary models. Section 4 is devoted to simulation experiments and Section 5

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

applies our proposed SupWald tests to the yield spread dynamics. Section 6 gives some final remarks and proofs are gathered in Section 7 and in an Appendix.

2. Adaptation, consistency and null limit distributions

Consider first the basic case of a centered random walk null hypothesis given by

$$H_0 : \Delta y_t = y_t - y_{t-1} = \varepsilon_t,$$

where $y_0 = 0$ and $\{\varepsilon_t\}$ is a sequence of i.i.d. centered random variables with variance σ^2. Assume that $T+1$ observations $y_0, \ldots, y_T$ are available to test H_0 against

$$H_1 : \{y_t\} \text{ is a nonconstant stationary ergodic process with a finite nonvanishing variance.}$$

A well-known example of a linear test of H_0 against H_1 is the Dickey–Fuller (DF) test which uses the auxiliary model

$$\Delta y_t = \mu + \rho y_{t-1} + v_t. \tag{2.1}$$

and the associated t statistic DF for the null hypothesis $\rho = 0$. Indeed, for any alternative in H_1, the limit ρ of the OLS estimate $\widehat{\rho}$ captures a mean reverting effect which ensures that $\rho < 0$ and yields consistency of the DF test. Unfortunately, simulation studies by e.g. Pippenger and Goering (1993) or Taylor (2001) have shown that although consistent, the DF test lacks power against nonlinear stationary alternatives. Hence, subsequent research has focused on developing unit-root tests based on a nonlinear auxiliary model instead of the linear one given by Eq. (2.1). Among the possible nonlinear candidates, the TAR specification aims to explicitly account for mean reversion and to allow for local unit-root in a regime where asymptotic adjustment does not hold. For instance, motivated by the type of nonlinear behavior generated by transaction costs in general equilibrium models, an illustration is the following symmetric mirroring 3-regime TAR specification previously considered in Taylor (2001) and Bec et al. (2004)

$$\Delta y_t = u_t + \begin{cases} \mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \in (-\infty, \lambda_1] = I_1(\lambda) \\ \mu_2 + \rho_2 y_{t-1} & \text{if } y_{t-1} \in [\lambda_1, \lambda_2] = I_2(\lambda) \\ -\mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \in (\lambda_2, +\infty] = I_3(\lambda) \end{cases} \quad \text{with } \lambda_2 = -\lambda_1 = \lambda. \tag{2.2}$$

This specification may be rewritten as the dynamic linear regression model:

$$\Delta y_t = \widetilde{x}_t(\lambda)\beta + u_t \text{ with } \widetilde{x}_t'(\lambda) = \begin{bmatrix} \mathbb{I}(y_{t-1} \in I_1(\lambda)) - \mathbb{I}(y_{t-1} \in I_3(\lambda)) \\ y_{t-1}(\mathbb{I}(y_{t-1} \in I_1(\lambda)) + \mathbb{I}(y_{t-1} \in I_3(\lambda))) \\ \mathbb{I}(y_{t-1} \in I_2(\lambda)) \\ y_{t-1}\mathbb{I}(y_{t-1} \in I_2(\lambda)) \end{bmatrix} \text{ and}$$

$$\beta = \begin{bmatrix} \mu_1 \\ \rho_1 \\ \mu_2 \\ \rho_2 \end{bmatrix}.$$

In such a setup, for a given value of the threshold λ, the OLS estimators of β and $\mathrm{Var}(u_t)$ are given by

$$\widehat{\beta}_T(\lambda) = \left(\sum_{t=1}^{T} \widetilde{x}_t'(\lambda)\widetilde{x}_t(\lambda) \right)^{-1} \sum_{t=1}^{T} \widetilde{x}_t'(\lambda)\Delta y_t, \quad \widehat{\sigma}_T^2(\lambda) = \frac{1}{T-k} \sum_{t=1}^{T} (\Delta y_t - \widetilde{x}_t(\lambda)\widehat{\beta}_T(\lambda))^2,$$

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

where k is the dimension of $\widetilde{x}_t(\lambda)$. Furthermore, the Wald statistic to test that there is a unit-root in each regime, $\rho_1 = \rho_2 = 0$, writes

$$\mathrm{Wald}_T(\lambda) = (R\widehat{\beta}_T(\lambda))' \left(\widehat{\sigma}_T^2(\lambda) R \left(\sum_{t=1}^{T} \widetilde{x}_t'(\lambda)\widetilde{x}_t(\lambda) \right)^{-1} R' \right)^{-1} (R\widehat{\beta}_T(\lambda)), \tag{2.3}$$

where R is a selection matrix such that $(R\beta)' = [\rho_1, \rho_2]$. A large $\mathrm{Wald}_T(\lambda)$ favors rejection of H_0.[1]

In practice, an important issue in building such a Wald threshold unit-root test is the choice of a suitable threshold level λ when its true value is unknown. To overcome this issue, the most widespread approach (Bec et al., 2004; Caner and Hansen, 2001; Gonzalez and Gonzalo, 1998; de Jong et al., 2005; Kapetanios and Shin, 2006; Park and Shintani, 2005; Seo, 2003; Shin and Lee, 2001, 2003) builds on the structural change literature, see Andrews (1993) among others, and uses a SupWald test statistic of the form:

$$\mathrm{SupWald}_T(\Lambda_T) = \sup_{\lambda \in \Lambda_T} \mathrm{Wald}_T(\lambda).$$

This amounts to choose for the test the threshold value which maximizes $\mathrm{Wald}_T(\cdot)$ over Λ_T.

Following Andrews (1993), the common use in the papers cited above consists in considering a quantile-based threshold set such as

$$\Lambda_T = [|y|_{(\pi T)}, |y|_{((1-\pi)T)}], \quad \pi \in (0, 1/2), \tag{2.4}$$

where the $|y|_{(t)}$, $t = 0, \ldots, T-1$, are the ordered $|y_{t-1}|$ and, for $x \in \mathbb{R}^+$ with integer part $[x]$, $|y|_{(x)} = y_{([x])}$, so that $y_{(\pi T)}$ is the empirical quantile of order π. For such thresholds λ, the inequality $|y|_{(\pi T)} \leqslant \lambda$ ensures that the proportion of observations in the inner regime $I_2(\lambda)$ is at least π. Symmetrically, $\lambda \leqslant |y|_{((1-\pi)T)}$ gives a minimal proportion of π observations in the outer regime $I_1(\lambda) \cup I_3(\lambda)$. As a consequence, the parameters of each regime are correctly estimated and $\mathrm{SupWald}(\Lambda_T)$ remains finite. Following Andrews's (1993) suggestion, the usual choice of π is 15%.

However, this SupWald approach does not really tackle the consistency issue. Under the stationary alternative, a Λ_T as defined in (2.4) converges to $[Q(\pi), Q(1-\pi)]$, where $Q(\pi)$ is the πth quantile of $|y_t|$ that solves $\mathbb{P}(|y_t| \leqslant Q(\pi)) = \pi$. But nothing ensures that there is a threshold in $[Q(\pi), Q(1-\pi)]$ that gives a diverging Wald statistic, even in the case of a correctly specified TAR.[2]

2.1. Introducing adaptation to achieve consistency

The cornerstone of our strategy to build a consistent test is the choice of the threshold set Λ_T. We argue that the asymptotic behavior of Λ_T should differ accordingly to the hypotheses at hand. Such a suitable *adaptation* property can be described in the following requirements:

R0. Under H_0, Λ_T should remain "as small as possible" so that the test statistic has a finite null limit distribution, with moderate critical values if possible. Indeed, small critical values z_α would increase the power of the test which has a rejection region $\mathrm{SupWald}_T(\Lambda_T) > z_\alpha$.

R1. Under H_1, Λ_T should be "as large as possible" so that the $\mathrm{SupWald}_T(\Lambda_T)$ test would be more powerful by considering many $\mathrm{Wald}_T(\lambda)$ statistics.[3]

In other words, it is desirable that the boundaries of the threshold set adapt to the hypothesis of interest, making the threshold set wider under H_1 than under H_0. One intuitive way to achieve this feature is to index these boundaries with a consistent unit-root test statistic. A natural candidate is the absolute value of the DF statistic, hereafter denoted $|DF_T|$. Roughly speaking, by defining the lower boundary of λ_1, say $\underline{\lambda}_T$, as a

[1] Observe that $\widehat{\beta}_T(\lambda)$ and $\mathrm{Wald}_T(\lambda)$ may not be defined properly if $\sum_{t=1}^{T} \widetilde{x}_t'(\lambda)\widetilde{x}_t(\lambda)$ has no inverse, in particular if there is no observation in a regime $I_j(\lambda)$. In this case, the Wald statistic can be set to its infinite limit value or a Moore–Penrose pseudo-inverse can be used.

[2] The consistency issue is hardly considered in the literature and existing consistency results often build on assumptions that are difficult to check. For instance, Kapetanios and Shin (2006) and Park and Shintani (2005) assume that the true threshold value is in $[Q(\pi), Q(1-\pi)]$ while de Jong et al. (2005) consider stationary alternatives with $\mathbb{E}[\Delta y_{t-1}(y_{t-p} - m)\mathbb{I}(y_{t-p} > m)] < 0$ for all p where m is the median.

[3] Note that the usual threshold set (2.4) is asymptotically $[Q(\pi), Q(1-\pi)]$ and cannot be adaptive since $\pi > 0$. Actually, (2.4) also contradicts R0 and R1. Indeed, such a Λ_T has a length of order $\sqrt{T}$ under H_0 and remains bounded under H_1.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

decreasing function of $|DF_T|$ and the upper boundary of λ_2, say $\overline{\lambda}_T$, as an increasing function of $|DF_T|$, the desired type of threshold set $\Lambda_T = [\underline{\lambda}_T, \overline{\lambda}_T]$ would obtain. Indeed, from the consistency property of the DF statistic, it follows that $|DF_T|$ is bounded under H_0 but diverges under H_1. Consequently, under the null $\underline{\lambda}_T$ (resp. $\overline{\lambda}_T$) would be relatively large (resp. relatively small), implying a narrow threshold set. By contrast, under the alternative, the threshold set widens as $|DF_T|$ diverges. The next theorem shows that fulfilling conditions as R1 is sufficient to get consistency against ergodic alternatives.

Theorem 1. *Consider the TAR specification* (2.2). *Assume that* Λ_T *is such that, for any* $\{y_t\}$ *in* H_1, *there is a* λ_T *converging to* $Q(1)$ *which is in* Λ_T *with a probability tending to* 1. *Then, under* H_1, $\mathrm{SupWald}_T(\Lambda_T)$ *diverges in probability, with*

$$\mathrm{SupWald}_T(\Lambda_T) \geqslant DF_T^2(1 + o_{\mathbb{P}}(1)). \tag{2.5}$$

Importantly, it follows from Theorem 1 that adaptation is a sufficient condition for consistency against *any* ergodic stationary alternative. Then, the inequality (2.5) indicates that the SupWald test can be more powerful than a DF test provided its critical values are close enough to the squared critical values of the DF_T statistic. The intuition behind (2.5) is that the TAR specification (2.2) is asymptotically equivalent to the autoregressive linear model (2.1) when the threshold is $\lambda_T = |y|_{(T)}$. In this case, the central regime diverges and $\mathrm{Wald}_T(\lambda_T)$ is asymptotically equivalent to DF_T^2.

2.2. Examples of adaptive threshold sets and null limit distributions

Two examples of adaptive threshold sets will be considered here. The first one is an asymptotically unbounded set directly derived from the usual threshold set (2.4). This adaptive set is asymptotically unbounded in the sense that the boundaries grow with the sample size. More precisely, we show how the latter may be amended to satisfy the adaptation property. As will be stressed below, this version of (2.4) is not entirely satisfactory in that it does not match the requirement R0, since its length diverges with the sample size under the null. The second example belongs to the bounded class of threshold sets and gives up any reference to an arbitrary proportion of observations in the definition of the threshold set boundaries. It is shown to match both R0 and R1 requirements.

2.2.1. A class of asymptotically unbounded threshold sets

A first example of adaptive threshold set follows from a modification of the quantile-based threshold set (2.4) that changes $1-\pi$ into a random proportion of the sample. Let π and $\delta > 0$ be proportion and length parameters to be chosen by the econometrician, and define

$$\Lambda_T^U = [\sqrt{T}\underline{\lambda}_T, \sqrt{T}\overline{\lambda}_T] \quad \text{with } \sqrt{T}\underline{\lambda}_T = |y|_{(\pi_T T)} \quad \text{and} \quad \sqrt{T}\overline{\lambda}_T = y_{((1-\pi_T)T)}, \tag{2.6}$$

where

$$1 - \pi_T = \min\left(1 - \pi + \frac{\delta |DF_T|}{\sqrt{T}}, \frac{T-2}{T}\right)$$

which parallels (2.4). The introduction of the term $(T-2)/T$ in the definition of $1-\pi_T$ ensures that there is at least two observations in the outer regime so that $\mathrm{SupWald}_T(\Lambda_T^U)$ is finite. To describe the null behavior, recall that the Donsker line $\{y_{[Tv]}/\sqrt{T}\}_{v\in[0,1]}$ converges in distribution to $\{\sigma W(v)\}_{v\in[0,1]}$ where $W(\cdot)$ is a standard Brownian Motion. For any π in [0, 1], let $Q_{|W|}(\pi)$ be the random variable that solves $\int_0^1 \mathbb{I}(|W(v)| \leqslant Q)\,\mathrm{d}v = \pi$. Since DF_T is bounded under the null, π_T converges to π so that

$$(\underline{\lambda}_T, \overline{\lambda}_T) \xrightarrow{\mathrm{d}} (\sigma Q_{|W|}(\pi), \sigma Q_{|W|}(1-\pi))$$
$$\text{which gives } \frac{\Lambda_T^U}{\sqrt{T}} \xrightarrow{\mathrm{d}} \Lambda^U = [\sigma Q_{|W|}(\pi), \sigma Q_{|W|}(1-\pi)], \tag{2.7}$$

showing that Λ_T^U has the same asymptotic behavior than the threshold set (2.4) and asymptotically contains the same percentage $(1-2\pi)\%$ of observations. Note the standardization of Λ_T^U with $\sqrt{T}$ implies that the

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

thresholds of Λ_T are of order $\sqrt{T}$: in what follows, (asymptotically) *unbounded* thresholds refer to thresholds with this order. Under a fixed alternative, the DF_T statistic diverges with the order T,[4] so that $1-\pi_T$ has the limit of $\min(1-\pi+\delta\sqrt{T},(T-2)/T)$ which goes to 1. Hence $\sqrt{T}\underline{\lambda}_T$ converges to $Q(1)$ in probability and Λ_T^U is adaptive. The $\text{SupWald}_T(\Lambda_T^U)$ thus inherits of the consistency of the DF_T statistic against any ergodic stationary alternative by Theorem 1.

The next theorem shows that $\text{SupWald}_T(\Lambda_T^U)$ has a pivotal null limit distribution. Define, for each regime $j=1,2,3$,

$$\xi_{j,U}(\lambda)=\frac{\int_0^1 W(v)\mathbb{I}_{I_j(\lambda)}(W(v))\,\mathrm{d}W(v)-\dfrac{\int_0^1 W(v)\mathbb{I}_{I_j(\lambda)}(W(v))\,\mathrm{d}v}{\int_0^1 \mathbb{I}_{I_j(\lambda)}(W(v))\,\mathrm{d}v}\int_0^1 \mathbb{I}_{I_j(\lambda)}(W(v))\,\mathrm{d}W(v)}{\left[\int_0^1 W^2(v)\mathbb{I}_{I_j(\lambda)}(W(v))\,\mathrm{d}v-\dfrac{(\int_0^1 W(v)\mathbb{I}_{I_j(\lambda)}(W(v))\,\mathrm{d}v)^2}{\int_0^1 \mathbb{I}_{I_j(\lambda)}(W(v))\,\mathrm{d}v}\right]^{1/2}},\tag{2.8}$$

$$\xi_{O,U}(\lambda)=\frac{\int_0^1 W(v)\mathbb{I}_{I_{1,3}(\lambda)}(W(v))\,\mathrm{d}W(v)-\dfrac{\int_0^1 W(v)\mathbb{I}_{I_{1,3}(\lambda)}(W(v))\,\mathrm{d}v}{\int_0^1 \mathbb{I}_{I_{1,3}(\lambda)}(W(v))\,\mathrm{d}v}\int_0^1 (\mathbb{I}_{I_1(\lambda)}-\mathbb{I}_{I_3(\lambda)})(W(v))\,\mathrm{d}W(v)}{\left[\int_0^1 W^2(v)\mathbb{I}_{I_{1,3}(\lambda)}(W(v))\,\mathrm{d}v-\dfrac{(\int_0^1 W(v)\mathbb{I}_{I_{1,3}(\lambda)}(W(v))\,\mathrm{d}v)^2}{\int_0^1 \mathbb{I}_{I_{1,3}(\lambda)}(W(v))\,\mathrm{d}v}\right]^{1/2}},$$

where $I_{1,3}(\lambda)=I_1(\lambda)\cup I_3(\lambda)$ and $\xi_{O,U}$ is for the outer regimes 1 and 3.

Theorem 2. *Consider the TAR specification* (2.2). *Let* Λ_T^U *and* Λ^U *be as in* (2.6) *and* (2.7) *and assume that Assumption E(s) given in Section 7 for* $s>4$ *holds. Then, under* H_0, $\text{SupWald}_T(\Lambda_T^U)$ *converges in distribution to* $\sup_{\lambda\in\Lambda^U}(\xi_{O,U}^2(\lambda/\sigma)+\xi_{2,U}^2(\lambda/\sigma))$, *which has a pivotal distribution.*

2.2.2. *A class of bounded threshold sets*

An alternative is to use *bounded* thresholds as also considered in Kapetanios and Shin (2006) and Seo (2003).[5] According to the asymptotic theory developed in Park and Phillips (1999), the number of thresholds $|y_{t-1}|$ in a bounded interval is of order $\sqrt{T}$ only, therefore yielding an asymptotically vanishing percentage of observations in a bounded threshold set. Hence bounded threshold sets can be used to produce lower SupWald critical values than asymptotically unbounded ones.

We now give an example of an adaptive bounded threshold set Λ_T^B. The estimated variance of the noise $\{\varepsilon_t\}$ in the linear specification (2.1), $\widehat{\sigma}_{\varepsilon T}^2=\sum_{t=1}^T(y_t-\widehat{\mu}-(1+\widehat{\rho})y_{t-1})^2/(T-2)$, is used as a scaling factor in Λ_T^B, where $\widehat{\mu}$ and $\widehat{\rho}$ are least-squares estimators. Let δ be a length parameter to be chosen by the econometrician and define the bounded set by

$$\Lambda_T^B=[\underline{\lambda}_T,\overline{\lambda}_T]\quad\text{with }\underline{\lambda}_T=|y|_{(2)}+\frac{\widehat{\sigma}_{\varepsilon T}}{\delta|DF_T|}\quad\text{and}\quad\overline{\lambda}_T=\underline{\lambda}_T+\delta\widehat{\sigma}_{\varepsilon T}|DF_T|.\tag{2.9}$$

The term $|y|_{(2)}$ in the definition of the lower threshold $\underline{\lambda}_T$ ensures that there are at least 2 observations in the inner regime of (2.2) to allow for estimation of μ_2 and ρ_2. The artificial term $\widehat{\sigma}_{\varepsilon T}/(\delta|DF_T|)$ in $\underline{\lambda}_T$ has been added to obtain more observations in the inner regime so as to avoid small values of the Wald statistics. The null behavior of Λ_T^B depends on the limit in distribution of the DF_T statistic:

$$\tau=\frac{\int_0^1 W(v)\,\mathrm{d}W(v)-W(1)\int_0^1 W(v)\,\mathrm{d}v}{[\int_0^1 W^2(v)\,\mathrm{d}v-(\int_0^1 W(v)\,\mathrm{d}v)^2]^{1/2}}.\tag{2.10}$$

[4] As formally established in the Proof section for any arbitrary ergodic alternative.

[5] These authors consider 3-regime TAR specifications with $\mu_2=0$ and $\rho_2=0$. This considerably simplifies the derivation of the null limits of the resulting test since the parameters of the central regime are not estimated. Note: however, that their test is not adaptive and that they limit to detection of specific TAR alternatives.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Define

$$\Lambda^B = [\underline{\lambda}, \overline{\lambda}] \text{ with } \underline{\lambda} = \frac{\sigma}{\delta|\tau|} \quad \text{and} \quad \overline{\lambda} = \underline{\lambda} + \sigma\delta|\tau|, \tag{2.11}$$

which is such that $\Lambda_T^B \xrightarrow{d} \Lambda^B$, showing that the thresholds of Λ_T^B are bounded under H_0. Under H_1, the DF_T statistic diverges, so that $\underline{\lambda}_T$ goes to the lower bound $Q(0)$ of the support of the $|y_{t-1}|$'s and $\overline{\lambda}_T$ diverges. Hence Λ_T^B is adaptive, and we now turn to the null limit distribution of SupWald(Λ_T^B). Let $B(\cdot)$ be a standard Brownian Motion independent of $W(\cdot)$. The contribution of the inner regime in the Wald statistic is given here by $\zeta_{2B}(\lambda) = \int_{-\lambda}^{\lambda}(w-\lambda)\,dB(w)/\sqrt{2\lambda^3/3}$.

Theorem 3. *Consider the TAR specification* (2.2). *Let* Λ_T^B *and* Λ^B *be as in* (2.9) *and* (2.11) *and assume that Assumption E(s) given in Section 7 for* $s>4$ *holds. Then, under* H_0, $\text{SupWald}_T(\Lambda_T^B)$ *converges in distribution to* $\xi_{O,U}(0) + \sup_{\lambda\in\Lambda^B}\zeta_{2B}(\lambda)$, *which has a pivotal distribution, and where* $\xi_{O,U}(\cdot)$ *is as in Theorem* 2.

Compared to the null limit distribution of Theorem 2, the contribution of the outer regimes to the limit of $\text{SupWald}_T(\Lambda_T^B)$ is now given by $\xi_{O,U}(0)$ corresponding to the fixed threshold 0, see Kapetanios and Shin (2006) and Seo (2003) for similar results. This can be useful to achieve smaller critical values. The intuition is that the order $\sqrt{T}$ of the y_{t-1}'s in the outer regime dominates the thresholds. The contribution of the inner regime $\sup_{\lambda\in\Lambda^B}\zeta_{2B}(\lambda)$ is given by the bounded values of the y_{t-1}'s and remains finite since $\underline{\lambda}>0$.

3. Extensions

In this section, the consistency and null limit distribution of the SupWald test are extended to the more general case of an autoregression of order p, as well as to more general auxiliary models and threshold sets. The considered null hypothesis is now

$H_0(p)$: $\Delta y_t = a(L)\Delta y_{t-1} + \varepsilon_t$ for $t\geqslant 1$, $y_0 = \cdots = y_{-p-1} = 0$, where $\{\varepsilon_t\}$ is a (strong) white noise sequence with variance σ^2 and $1-a(L)$ is of known order $p\geqslant 0$ with roots outside the unit circle, so that $\sigma_y^2 = \lim_{T\to\infty}\text{Var}(y_T/\sqrt{T})>0$.

To account for the additional lagged polynomial term $a(L)\Delta y_{t-1}$, we extend the TAR specification to include lagged variables. We also allow for asymmetric regimes by considering a two dimensional threshold parameter $\lambda' = [\lambda_1, \lambda_2]$ with $\lambda_1 \leqslant \lambda_2$, noting that $\lambda_1 = \lambda_2$ gives a 2-regimes TAR as considered in Berben and van Dijk (1999), Caner and Hansen (2001), Enders and Granger (1998) and Shin and Lee (2003).

Extensions of the results of the preceding section concerns first consistency. A general definition of adaptation is given, which allow for a general threshold variable s_t that does not need to be the level y_{t-1}. Second, under the null and for $s_t = y_{t-1}$, we derive the functional limit distribution of baseline variables entering the SupWald test for a wide class of TAR specifications, using unbounded or bounded thresholds. We then give general conditions on unbounded or bounded threshold sets that ensures a pivotal limit distribution for the SupWald test.

3.1. Adaptation

Our baseline general TAR model of order p is

$$\Delta y_t = u_t + \begin{cases} \mu_1 + \rho_1 y_{t-1} + a_1(L)\Delta y_{t-1} & \text{if } s_t \in I_1(\lambda), \\ \mu_2 + \rho_2 y_{t-1} + a_2(L)\Delta y_{t-1} & \text{if } s_t \in I_2(\lambda), \\ \mu_3 + \rho_3 y_{t-1} + a_3(L)\Delta y_{t-1} & \text{if } s_t \in I_3(\lambda). \end{cases} \tag{3.1}$$

Bec et al. (2004), Berben and van Dijk (1999), Enders and Granger (1998), Kapetanios and Shin (2006) and Seo (2003) consider a threshold variable $s_t = y_{t-1}$ which is integrated under $H_0(p)$ but stationary under H_1, while Caner and Hansen (2001) and Shin and Lee (2003) use a stationary s_t under $H_0(p)$ and H_1 as for instance $s_t = \Delta y_t$, see also Gonzalez and Gonzalo (1998). Our approach assumes that $\{y_t, s_t\}$ is stationary under H_1,

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

hence allowing for all the choices considered in the references above. Various restrictions of (3.1) have been considered in the literature, as for instance the symmetric mirroring 3-regime TAR specification (2.2). These restriction can be written as

$$\Delta y_t = \widetilde{x}_t(\lambda)\beta + u_t, (\lambda, \beta) \in \Theta_\lambda \times \mathbb{R}^k \quad \text{with } \widetilde{x}_t(\lambda) = x_t(\lambda)r,$$

$$x_t(\lambda) = [(1, y_{t-1})\mathbb{I}(s_t \in I_j(\lambda)), (\Delta y_{t-1}, \ldots, \Delta y_{t-p})\mathbb{I}(s_t \in I_j(\lambda))], \tag{3.2}$$

for $j = 1, 2, 3$ where r is a known selection matrix which is given by the restriction of (3.1) of interest. The parameters set Θ_λ can include various constraints for the thresholds, as $\lambda_1 = -\lambda_2$ which corresponds to a symmetric inner regime as in (2.2). The Wald statistics combined in the SupWald test correspond to the hypothesis $\rho_j = 0, j = 1, 2, 3$, for a threshold λ, and can be computed as in (2.3), with a matrix R such that $R\beta$ gives the ρ coefficients of the model. In what follows, r and R are assumed to be full-rank. As a benchmark, we now consider the augmented Dickey–Fuller (ADF_T) statistic, that is the t-statistic for the autoregressive coefficient ρ of the linear specification

$$\Delta y_t = \mu + \rho y_{t-1} + a(L)\Delta y_{t-1} + v_t. \tag{3.3}$$

An important feature of the symmetric mirroring TAR specification (2.2) was that a growing inner regime $I_2(\lambda)$ gives the dynamic linear model (2.1) as a limit, so that (2.2) asymptotically nests (2.1), as formalized in the next definition.

Definition 1. Consider a restriction (3.2) of the 3-regime threshold autoregressive model (3.1) such that the parameters μ_j, ρ_j and $a_j(L)$ are constant across a subset $\mathcal{J}$ of regime indices. Let $\mathbb{S}$ be the support of the stationary threshold variable s_t and $I_{\mathcal{J}}(\lambda) = \bigcup_{j\in\mathcal{J}} I_j(\lambda)$.

This restricted autoregressive threshold model nests the augmented linear autoregressive model (3.3) through the subset of regimes $\mathcal{J}$ if and only if

(i) There exists a sequence $\lambda_n \in \Theta_\lambda$ such that $I_{\mathcal{J}}(\lambda_n) \to \mathbb{S}$ when n goes to infinity (i.e. $\lim_{n\to\infty}\mathbb{P}(s_t \notin I_{\mathcal{J}}(\lambda_n)) = 0$).
(ii) The covariates $x_t(\lambda)$ admit a partition $x_{\mathcal{J}t}(\lambda), x_{-\mathcal{J}t}(\lambda)$, with $x_{\mathcal{J}t}(\lambda)\mathbb{I}(s_t \in I_{\mathcal{J}}(\lambda)) \neq 0$ and $x_{-\mathcal{J}t} = x_{-\mathcal{J}t}\mathbb{I}(s_t \notin I_{\mathcal{J}}(\lambda))$ (implying $\lim_{I_{\mathcal{J}}(\lambda)\to\mathbb{S}} x_{-\mathcal{J}t}(\lambda) = 0$). In addition, with $\lim_{I_{\mathcal{J}}(\lambda)\to\mathbb{S}} x_{\mathcal{J}t}(\lambda) = [1, y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p}]$.

The introduction of the regime index set $\mathcal{J}$ allows for strongly constrained symmetric TAR specification as

$$\Delta y_t = u_t + \begin{cases} \mu_1 + \rho_1 y_{t-1} + a_1(L)\Delta y_{t-1} & \text{if } s_t \in (-\infty, -\lambda] = I_1(\lambda), \\ \mu_2 + \rho_2 y_{t-1} + a_2(L)\Delta y_{t-1} & \text{if } s_t \in (-\lambda, \lambda) = I_2(\lambda), \\ \mu_1 + \rho_1 y_{t-1} + a_1(L)\Delta y_{t-1} & \text{if } s_t \in [\lambda, +\infty) = I_3(\lambda), \end{cases}$$

where λ is in $\mathbb{R}^+$ here. In this specification $I_1(\lambda)$ and $I_3(\lambda)$ cannot diverge to $\mathbb{R}$ while $I_1(\lambda) \cup I_3(\lambda)$ can, so that this model asymptotically nests (3.3) through the inner regime $j = 2$ and the outer regime $\mathcal{J} = \{1, 3\}$. The next definition introduces adaptation.

Definition 2. Assume that the threshold variable s_t is chosen such that $\{y_t, s_t\}$ is stationary for any alternative $\{y_t\}$ in H_1. Let $\mathbb{S}$ be the support of the stationary s_t. Consider a restriction (3.2) of the TAR model (3.1) which nests the linear augmented autoregressive model (3.3) through the subset of regimes $\mathcal{J}$. Then a random set Λ_T of admissible thresholds is $\mathcal{J}$-adaptive if and only if

(i) The $\text{SupWald}_T(\Lambda_T)$ has a finite null limiting distribution.
(ii) For any alternatives $\{y_t\}$ of H_1, there exists a deterministic sequence λ_T in Θ_λ with $\lim_{T\to\infty} I_{\mathcal{J}}(\lambda_T) = \mathbb{S}$, and λ_T is in Λ_T with a probability tending to 1.

The statistic $\text{SupWald}_T(\Lambda_T)$ is $\mathcal{J}$-adaptive if and only if Λ_T is.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

This gives the following extension for Theorem 1 which similarly suggests that a $\text{SupWald}_T(\Lambda_T)$ can improve on the ADF test.

Theorem 4. *Assume that the threshold variable s_t is chosen such that $\{y_t, s_t\}$ is stationary for any alternative $\{y_t\}$ in H_1. Consider a restriction (3.2) of the TAR model (3.1) which nests the linear augmented autoregressive model (3.3) through the subset of regimes $\mathcal{J}$.*

Then, if Λ_T satisfies Definition 2(ii), $\text{SupWald}_T(\Lambda_T)$ diverges in probability for any $\{y_t\}$ in H_1 and

$$\text{SupWald}_T(\Lambda_T) \geqslant ADF_T^2(1 + o_{\mathbb{P}}(1)).$$

3.2. Pivotal null limit distributions for asymptotically unbounded thresholds

From now on, we consider a level threshold variable, that is $s_t = y_{t-1}$, so that (3.1) becomes

$$\Delta y_t = u_t + \begin{cases} \mu_1 + \rho_1 y_{t-1} + a_1(L)\Delta y_{t-1} & \text{if } y_{t-1} \in I_1(\lambda), \\ \mu_2 + \rho_2 y_{t-1} + a_2(L)\Delta y_{t-1} & \text{if } y_{t-1} \in I_2(\lambda), \\ \mu_3 + \rho_3 y_{t-1} + a_3(L)\Delta y_{t-1} & \text{if } y_{t-1} \in I_3(\lambda). \end{cases} \tag{3.4}$$

We first focus on asymptotically unbounded thresholds

$$\Lambda_T = \{\lambda = [\lambda_1, \lambda_2]' \in \Theta_\lambda; \sqrt{T}\underline{\lambda}_T \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \sqrt{T}\overline{\lambda}_T, \lambda_2 - \lambda_1 \geqslant \sqrt{T}v_T\}, \tag{3.5}$$

where the inequalities $\sqrt{T}\underline{\lambda}_T \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \sqrt{T}\overline{\lambda}_T$ now control for the number of observations in the outer regimes while $\lambda_2 - \lambda_1 \geqslant \sqrt{T}v_T$ deals with the inner regime. Finding the null limit distribution of the SupWald statistic in this context necessitates to introduce a suitable standardization of the baseline covariate $x_t(\lambda)$ in (3.2). Consider $X_{j,t}^U$ and $\Delta X_{j,t}^U(\lambda)$ with,

$$X_{jt}^U(\lambda) = \mathbb{I}\left(\frac{y_{t-1}}{\sigma_y\sqrt{T}} \in I_j(\lambda)\right)\left[\frac{1}{\sqrt{T}}, \frac{y_{t-1}}{T}\right],$$

$$\Delta X_{jt}^U(\lambda) = \mathbb{I}\left(\frac{y_{t-1}}{\sigma_y\sqrt{T}} \in I_j(\lambda)\right)\left[\frac{\Delta y_{t-1}}{\sqrt{T}}, \ldots, \frac{\Delta y_{t-p}}{\sqrt{T}}\right] \tag{3.6}$$

for $j = 1, 2, 3$. Recall that $W(\cdot)$ is a standard Brownian Motion and define

$$M_j(\lambda) = \sigma\begin{bmatrix} \int_0^1 \mathbb{I}(W(v) \in I_j(\lambda))\,\mathrm{d}W(v) \\ \sigma_y \int_0^1 W(v)\mathbb{I}(W(v) \in I_j(\lambda))\,\mathrm{d}W(v) \end{bmatrix}, \quad M_U(\lambda) = \begin{bmatrix} M_1(\lambda) \\ M_2(\lambda) \\ M_3(\lambda) \\ 0_{3p\times 1} \end{bmatrix}, \tag{3.7}$$

$$C_j(\lambda) = \begin{bmatrix} \int_0^1 \mathbb{I}(W(v) \in I_j(\lambda))\,\mathrm{d}v & \sigma_y \int_0^1 W(v)\mathbb{I}(W(v) \in I_j(\lambda))\,\mathrm{d}v \\ \sigma_y \int_0^1 W(v)\mathbb{I}(W(v) \in I_j(\lambda))\,\mathrm{d}v & \sigma_y^2 \int_0^1 W^2(v)\mathbb{I}(W(v) \in I_j(\lambda))\,\mathrm{d}v \end{bmatrix},$$

$$C_U(\lambda) = \text{Diag}[C_1(\lambda), C_2(\lambda), C_3(\lambda), 0_{3p\times 3p}]'. \tag{3.8}$$

The next theorem establishes functional convergence of sums related to the standardized $x_t(\lambda)$'s under $H_0(p)$, for any $p > 1$. Following van der Vaart (1996) and van der Vaart and Wellner (1996), we consider functional convergence in distribution in $\ell^\infty([-a, a]^2)$, the space of bounded functions over $[-a, a]^2$ equipped with the supremum norm.

Theorem 5. *Assume that $H_0(p)$, Assumptions $E(s)$ with $s > 4$ and L given in Section 7 hold, and let $\Omega = \text{Var}[\Delta y_{t-1}, \ldots, \Delta y_{t-p}]'$.*

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Then, for any $a>0$,

$$\left[\sum_{t=1}^{T} X_{jt}^{U}(\cdot)'\varepsilon_t, \sum_{t=1}^{T} X_{jt}^{U}(\cdot)'X_{jt}^{U}(\cdot), \sum_{t=1}^{T} \Delta X_{jt}^{U}(\cdot)'\Delta X_{jt}^{U}(\cdot), j=1,2,3\right]$$

converges in distribution in $\ell^{\infty}([-a,a]^2)$ *to*

$$\left[M_j(\cdot), C_j(\cdot), \Omega \int_0^1 \mathbb{I}(W(v)\in I_j(\cdot))\,\mathrm{d}v, j=1,2,3\right]$$

and $\sup_{\lambda\in[-a,a]^2} \|\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'X_{jt}^{U}(\lambda)\| = \mathrm{o}_{\mathbb{P}}(1)$, *for* $j=1,2,3$. *Moreover,* $\sum_{t=1}^{T}\Delta y_{t-k}\varepsilon_t/\sqrt{T} = \mathrm{O}_{\mathbb{P}}(1)$ *and if* $s>14$, $\sup_{\lambda\in[-a,a]^2}\|\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'\varepsilon_t\| = \mathrm{O}_{\mathbb{P}}(1)$, *for* $j=1,2,3$.

As seen from Bec et al. (2004), a result as Theorem 5 can be used to obtain the null limit distribution of supremum of statistics as the Likelihood ratio, Score and Wald statistics. As shown in the proof section, the fact that Theorem 5 holds for any real number $a>0$ is sufficient to allow for an asymptotically random $\Lambda_T/\sqrt{T}$ in such statistics. The next corollary concerns more specifically the SupWald statistic. Let

$$\Lambda = \{\lambda=[\lambda_1,\lambda_2]'\in\Theta_\lambda; \underline{\lambda}\leqslant\lambda_1\leqslant\lambda_2\leqslant\overline{\lambda}, \lambda_2-\lambda_1\geqslant v\}$$

be the limit in distribution of $\Lambda_T/\sqrt{T}$. The next corollary extends Theorem 2 to a general TAR specification and gives a simple condition on Λ ensuring that a SupWald(Λ_T) has an asymptotic pivotal distribution.

Corollary 1. *Assume that Assumption* $E(s)$ *with* $s>14$ *as in Section* 7 *holds. Assume that* $(\underline{\lambda}_T, \overline{\lambda}_T, v_T)$, *satisfying Assumption* Λ *in Section* 7, *converges in distribution to* $(\underline{\lambda}, \overline{\lambda}, v)$ *with*

$$\inf_{v\in[0,1]} W(v) < \frac{\underline{\lambda}}{\sigma_y} \leqslant \frac{\overline{\lambda}}{\sigma_y} < \sup_{v\in[0,1]} W(v) \quad \text{and } v>0. \tag{3.9}$$

Consider a restriction of the TAR specification (3.4) *with a covariate* $\widetilde{x}_t(\lambda) = x_t(\lambda)r$ *with entries taken from vectors*

$$\widetilde{r}_1\begin{pmatrix}\mathbb{I}_{I_1(\lambda)}(y_{t-1})\\ \mathbb{I}_{I_2(\lambda)}(y_{t-1})\\ \mathbb{I}_{I_3(\lambda)}(y_{t-1})\end{pmatrix},\quad \widetilde{r}_2\begin{pmatrix}y_{t-1}\mathbb{I}_{I_1(\lambda)}(y_{t-1})\\ y_{t-1}\mathbb{I}_{I_2(\lambda)}(y_{t-1})\\ y_{t-1}\mathbb{I}_{I_3(\lambda)}(y_{t-1})\end{pmatrix},\quad \widetilde{r}_3\begin{pmatrix}[\Delta y_{t-1},\ldots,\Delta y_{t-p}]\mathbb{I}_{I_1(\lambda)}(y_{t-1})\\ [\Delta y_{t-1},\ldots,\Delta y_{t-p}]\mathbb{I}_{I_2(\lambda)}(y_{t-1})\\ [\Delta y_{t-1},\ldots,\Delta y_{t-p}]\mathbb{I}_{I_3(\lambda)}(y_{t-1})\end{pmatrix},$$

for some suitable $\widetilde{r}_1,\widetilde{r}_2,\widetilde{r}_3$. *Then, under* $\mathrm{H}_0(p)$ *and if* Λ_T *is as in* (3.5), $\mathrm{SupWald}_T(\Lambda_T)$ *converges in distribution to*

$$\sup_{\lambda\in\Lambda/\sigma_y}\{M_U(\lambda)'r(r'C_U(\lambda)r)^{-1}R'(\sigma^2R(r'C_U(\lambda)r)^{-1}R')^{-1}R(r'C_U(\lambda)r)^{-1}r'M_U(\lambda)\},$$

which is finite and has a pivotal distribution provided that $[\underline{\lambda},\overline{\lambda},v]/\sigma_y$ *has a pivotal distribution.*

The restrictions on the covariates imply in particular that the restricted TAR does not impose constraints linking the autoregression coefficients ρ_j with the mean parameters μ_j or the lags coefficients, $j=1,2,3$. The condition (3.9) entails that each regimes are not empty asymptotically, ensuring that inverse matrix in the limit exists so that the limit variable is finite.

3.3. Pivotal null limit distributions for bounded thresholds

We restrict here to the common dynamic TAR specification

$$\Delta y_t = u_t + a(L)\Delta y_{t-1} + \begin{cases}\mu_1+\rho_1 y_{t-1} & \text{if } y_{t-1}\in I_1(\lambda),\\ \mu_2+\rho_2 y_{t-1} & \text{if } y_{t-1}\in I_2(\lambda),\\ \mu_3+\rho_3 y_{t-1} & \text{if } y_{t-1}\in I_3(\lambda),\end{cases} \tag{3.10}$$

and consider bounded threshold sets as

$$\Lambda_T = \{\lambda=[\lambda_1,\lambda_2]'\in\Theta_\lambda; \underline{\lambda}_T\leqslant\lambda_1\leqslant\lambda_2\leqslant\overline{\lambda}_T, \lambda_2-\lambda_1\geqslant v_T\}. \tag{3.11}$$

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

To study this case, a specific standardization is needed for the inner regime of the baseline covariate $x_t(\lambda)$ in (3.2). Define

$$X_{2t}^B(\lambda) = \frac{\mathbb{I}(y_{t-1} \in I_2(\lambda))}{T^{1/4}}[1, y_{t-1}], \quad X_{jt}^B(\lambda) = \mathbb{I}(y_{t-1} \in I_j(\lambda))\left[\frac{1}{\sqrt{T}}, \frac{y_{t-1}}{T}\right], \quad j = 1, 3,$$

$$\Delta X_t^B = \frac{1}{\sqrt{T}}[\Delta y_{t-1}, \ldots, \Delta y_{t-p}]. \tag{3.12}$$

The study of the inner regime builds here on local-time asymptotics as considered in Park and Phillips (2001). For a Brownian Motion $W(\cdot)$ over the time interval $[0, 1]$, the occupation time $A \mapsto \int_0^1 \mathbb{I}(W(v) \in A)\,\mathrm{d}v$ defines a measure over the Borel subsets of $\mathbb{R}$. This measure has a density $L_W(\cdot)$ with respect to the Lebesgue measure called local time, see Revuz and Yor (1999). In particular, the local time $L_W(0)$ gives the limit distribution of the number of observations in a central interval $I_2(\lambda)$ standardized with $(\lambda_2 - \lambda_1)\sqrt{T}$, see Park and Phillips (2001). Consider a two-sided standard Brownian Motion $B(\cdot)$ independent of $W(\cdot)$.[6] Define, for $M_j(\cdot)$ and $C_j(\cdot)$ as in (3.7) and (3.8),

$$M_2^B(\lambda) = \frac{\sigma L_W^{1/2}(0)}{\sigma_y^{1/2}} \begin{bmatrix} \int_{\lambda_1}^{\lambda_2} \mathrm{d}B(w) \\ \int_{\lambda_1}^{\lambda_2} w\,\mathrm{d}B(w) \end{bmatrix}, \quad M_B(\lambda) = \begin{bmatrix} M_1(0) \\ M_2^B(\lambda) \\ M_3(0) \\ 0_{p\times 1} \end{bmatrix}, \tag{3.13}$$

$$C_2^B(\lambda) = \frac{L_W(0)}{\sigma_y} \begin{bmatrix} \lambda_2 - \lambda_1 & \dfrac{\lambda_2^2 - \lambda_1^2}{2} \\ \dfrac{\lambda_2^2 - \lambda_1^2}{2} & \dfrac{\lambda_2^3 - \lambda_1^3}{3} \end{bmatrix},$$

$$C_B(\lambda) = \mathrm{Diag}[C_1(0), C_2^B(\lambda), C_3(0), 0_{p\times p}]'. \tag{3.14}$$

The next theorem establishes limit results relevant for the bounded case under $\mathrm{H}_0(p)$.

Theorem 6. *Assume that* $\mathrm{H}_0(p)$, *Assumptions* $E(s)$ *with* $s > 4$ *and* L *given in Section* 7 *hold, and let* $\Omega = \mathrm{Var}[\Delta y_{t-1}, \ldots, \Delta y_{t-p}]'$.

Then, for any $a > 0$,

$$\left[\sum_{t=1}^T X_{jt}^B(\cdot)'\varepsilon_t, \sum_{t=1}^T X_{jt}^B(\cdot)' X_{jt}^B(\cdot), \sum_{t=1}^T \Delta X_{jt}^{B\prime} \Delta X_{jt}^B, j = 1, 2, 3\right]$$

converges in distribution in $\ell^\infty([-a, a]^2)$ *to* $[M_j(\cdot), C_j(\cdot), \Omega, j = 1, 2, 3]$ *and*

$$\sup_{\lambda\in[-a,a]^2} \left\| \sum_{t=1}^T \Delta X_{jt}^{B\prime} X_{jt}^B(\lambda) \right\| = \mathrm{o}_{\mathbb{P}}(1), \quad \sum_{t=1}^T \Delta y_{t-k}\varepsilon_t/\sqrt{T} = \mathrm{O}_{\mathbb{P}}(1).$$

for $j = 1, 2, 3$.

Kapetanios and Shin (2006) and Seo (2003) implicitly used a similar result, but that did not include the asymptotic of the inner regime because they impose a central unit root. The fact that the limit variables of the outer regimes ($j = 1, 3$ in (3.13)) are random constant is in line with the results of these authors. As a corollary we derive the limit distribution under $\mathrm{H}_0(p)$ of a SupWald test based on a restriction of the TAR specification. As seen for the specific TAR specification used in Theorem 3, the local time $L_W(0)$ can disappear from the limit variable due to self-standardization of the Wald statistic. As stated below, the SupWald statistics is asymptotically pivotal under conditions that differs from the one of the unbounded case. In particular, the

[6] Recall that a two-sided Brownian Motion is a Brownian Motion defined over $\mathbb{R}$ as $B(w) = B_+(w)$ for $w \geqslant 0$, $B(w) = B_-(-w)$ for $w < 0$, where $\{B_+(w)\}_{w\in\mathbb{R}^+}$ and $\{B_-(w)\}_{w\in\mathbb{R}^+}$ are independent Brownian Motions over $\mathbb{R}^+$.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

parameters of the central regime cannot depend upon the parameters of the inner regime, but μ_2 and ρ_2 can be dependent.

Corollary 2. *Assume that Assumption E(s) with $s>14$ as in Section 7 holds. Assume that $(\underline{\lambda}_T, \overline{\lambda}_T, v_T)$, satisfying Assumption Λ in Section 7, converges in distribution to $(\underline{\lambda}, \overline{\lambda}, v)$ with $v>0$ almost surely.*

Consider a restriction of the TAR specification (3.4) *with a covariate $\widetilde{x}_t(\lambda) = x_t(\lambda)r$ with entries taken from vectors*

$$\widetilde{r}_1\begin{pmatrix}\mathbb{I}_{I_1(\lambda)}(y_{-1})\\ \mathbb{I}_{I_3(\lambda)}(y_{-1})\end{pmatrix}, \quad \widetilde{r}_2\begin{pmatrix}y_{t-1}\mathbb{I}_{I_1(\lambda)}(y_{-1})\\ y_{t-1}\mathbb{I}_{I_3(\lambda)}(y_{-1})\end{pmatrix}, \quad \mathbb{I}_{I_2(\lambda)}(y_{-1}), \quad y_{t-1}\mathbb{I}_{I_2(\lambda)}(y_{-1}), \quad \widetilde{r}_4\begin{pmatrix}\Delta y_{t-1}\\ \vdots\\ \Delta y_{t-p}\end{pmatrix},$$

for some suitable $\widetilde{r}_1, \widetilde{r}_2, \widetilde{r}_3$. Then, under $H_0(p)$ and if Λ_T is as in (3.11), *the $\mathrm{SupWald}_T(\Lambda_T)$ statistic converges in distribution to*

$$\sup_{\lambda\in\Lambda}\{M_B(\lambda)'r(r'C_B(\lambda)r)^{-1}R'(\sigma^2R(r'C_B(\lambda)r)^{-1}R')^{-1}R(r'C_B(\lambda)r)^{-1}r'M_B(\lambda)\},$$

which is finite and has a pivotal distribution provided that $[\underline{\lambda}/\overline{\lambda}, v/\overline{\lambda}]$ has a pivotal distribution.

4. Simulation experiments

Comparing the adaptive quantile threshold set (2.6) with its nonadaptive counterpart (2.4) suggests that the practical choice of an adaptive procedure may be more delicate due to a higher number of parameters. But adaptation allows for drastically different behaviors under the null and the alternative as seen from requirements R0 and R1. Hence a possible practical benefit of adopting an adaptive approach is a better trade-off between the length of Λ_T under the null and the alternatives as permitted by these additional parameters, resulting in a more powerful test.[7]

In this section, we propose a practical methodology to devise adaptive threshold sets accordingly. As suggested in Balke and Fomby (1997) and Taylor (2001), we use a set of stationary TAR alternatives exhibiting various features for calibration of the adaptive sets (2.6) and (2.9), see Table 2 which reports the behavior of the resulting test against some of the considered TAR alternatives. To ease calibration, we change $|DF_T|$ into $\max(1, |DF_T|)$ in (2.6) and (2.9). As a result of the comparison of several values, the retained length parameter in (2.9) is $\delta = 6$. For (2.6), we fix π to .85 and retain $\delta = 10$. Because limiting to TAR alternative would ignore that adaptive SupWald tests can detect a larger class of ergodic alternatives, we study the power of the test against Autoregressive Conditional Root models (ACR, see Bec et al., 2005 and Gouriéroux and Robert, 2006). As a benchmark for comparison, we consider the ADF_T test and a nonlinear unit-root test t_{NL} introduced by Kapetanios et al. (2003).[8]

4.1. Critical values

Table 1 gives the critical values based on 40,000 simulations of different sample sizes. Note that these critical values are much higher than the squared ones of ADF test, which is $(-2.88)^2 = 8.2944$ at the 5% level. 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 close to linear DGPs. Note also that the critical values of the test based upon unbounded Λ_T^U are larger in small and medium samples than the ones associated to bounded Λ_T^B,

[7] Caner and Hansen (2001), in analogy with a discussion in Andrews (1993) concerning trimming in tests for a structural change, have already argued that an ideal choice of a quantile threshold set (2.4) should be based on a trade-off between the null and the alternative: a small proportion π decreases the power of the test by increasing its critical values while a large π limits the power by decreasing the set of admissible thresholds. We extend this approach to the case of adaptive threshold sets.

[8] These last authors derived the model $\Delta y_t = \rho y_{t-1}^3 + v_t$ as an approximation of a smooth transition autoregressive model. The t_{NL} test is based on the Student statistic for ρ. For comparison sake with our TAR that includes mean parameters, we consider the demeaned version of the t_{NL} test. From 40,000 replications of simulations with $T = 200$ and 300, it follows that the critical value at the 5%-level is -2.94.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Table 1
Critical values (40,000 simulations)

Sample size	$\mathscr{W}_U^{\mathrm{Sup}}$					$\mathscr{W}_B^{\mathrm{Sup}}$				
	15%	10%	5%	1%	% in Λ_T^U	15%	10%	5%	1%	% in Λ_T^B
100	12.00	13.20	15.24	19.83	97.78	10.82	12.04	14.01	18.64	73.69
150	11.96	13.17	15.11	19.30	93.43	10.84	12.02	13.89	17.99	67.79
200	11.80	12.99	14.84	19.33	88.79	10.80	11.98	13.83	18.05	62.89
250	11.73	12.92	14.76	19.22	87.88	10.90	12.04	13.90	17.92	61.91
300	11.57	12.77	14.74	19.27	82.61	10.90	12.01	13.82	18.14	56.21
500	11.41	12.58	14.54	18.67	77.56	11.05	12.22	14.13	18.09	48.31
1000	11.35	12.49	14.47	18.43	73.74	11.14	12.28	14.20	18.28	37.90

suggesting that the length of Λ_T is larger in mean than the one for the bounded Λ_T. The last column for $\mathscr{W}_U^{\mathrm{Sup}}$ and $\mathscr{W}_B^{\mathrm{Sup}}$ contains the percentage of $y_{t-1} \in \Lambda_T$ and confirms this conjecture. For instance, with a sample size of 200, the percentage of observations in Λ_T^U is 20% greater than the one for Λ_T^B. Moreover, as expected, the percentage of y_{t-1} in the inner regime for Λ_T^B decreases sharply with the sample size. The critical values of the two tests become closer when T increases, suggesting that the maximum of the $\mathrm{Wald}_T(\lambda)$ statistic is achieved for moderate thresholds λ.

4.2. TAR alternatives

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

$$\Delta y_t = a\Delta y_{t-1} + \varepsilon_t + \begin{cases} \mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \leqslant -\lambda \\ \rho_2 y_{t-1} & \text{if } |y_{t-1}| < \lambda \\ -\mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \geqslant \lambda \end{cases}$$

$$\text{with } \mu_1 = 1.3 \times |\rho_1| \times \lambda, \quad \rho_2 = 0.$$

and ε_t is an i.i.d. $\mathscr{N}(0,1)$. The choice of the parameters values follows Bec et al. (2004) analysis of real exchange rate data. Table 2 reports the 5%-level rejection rates of the ADF, $\mathscr{W}_U^{\mathrm{Sup}}$ and $\mathscr{W}_B^{\mathrm{Sup}}$ for $T = 200$ and 300, using 1,000 replications.

The percentages of $|y_t|$ contained in Λ_T^U and Λ_T^B, given into brackets, are greater than the ones under the null (see Table 1) especially for Λ_T^B. This illustrates the adaptive behavior of Λ_T^U and Λ_T^B. It can also be seen that the percentage of data in the stationary regimes depends crucially on the thresholds size (see the third column of Table 2).

As expected, the power of all the tests increases with the sample size. The tests based on $\mathscr{W}_U^{\mathrm{Sup}}$ and $\mathscr{W}_B^{\mathrm{Sup}}$ generally outperform the standard ADF except for close to linear alternatives, that is when the percentage in the stationary regimes is more important. However, for these cases the power of adaptive tests is close or equal to the power of the standard ADF especially for $\mathscr{W}_B^{\mathrm{Sup}}$. For processes characterized by a low percentage of data in the stationary regimes, the gain of the adaptive tests can be as high as 70% compared to the ADF. The power of t_{NL} is always dominated by both SupWald tests. Beyond the lack of adaptation, the relative failure of the t_{NL} test may come from the fact that it is specifically devised for smooth transition threshold autoregressive models under the alternative.

Finally, the test based on $\mathscr{W}_B^{\mathrm{Sup}}$ outperforms the one based on $\mathscr{W}_U^{\mathrm{Sup}}$ for all the cases. This gain in power by the bounded interval compared to unbounded interval is due to the fact that the critical values of $\mathscr{W}_B^{\mathrm{Sup}}$ are relatively small since the percentage of observations in Λ_T^U and Λ_T^B are quite close.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Table 2
Empirical power of the unit-root tests against TAR alternatives

(λ, a, ρ_1)	T	%	ADF	t_{NL}	$\mathscr{W}_U^{\mathrm{Sup}}$	$\mathscr{W}_B^{\mathrm{Sup}}$
$(10, 0, -0.1)$	200	2.8	14.5	22.6	58.5 [91.9]	61.6 [93.6]
	300	2.7	18.4	32.8	74.9 [86.4]	81.8 [95.8]
$(10, 0, -0.3)$	200	1.5	18.3	66.3	84.9 [93.5]	88.6 [97.8]
	300	1.4	23.4	84.1	93.9 [88.7]	96.5 [98.5]
$(10, 0.3, -0.1)$	200	3.9	17.1	38.3	82.2 [93.8]	84.5 [97.0]
	300	3.8	24.4	67.5	95.6 [88.9]	98.0 [98.8]
$(10, 0.3, -0.3)$	200	2.1	36.4	94.2	98.3 [97.2]	98.6 [98.4]
	300	2.0	77.3	98.3	99.8 [93.5]	100 [98.7]
$(2, 0.3, -0.1)$	200	41.5	100	90.3	94.5 [98.5]	97.6 [98.3]
	300	41.4	100	97.4	100 [99.0]	100 [98.7]
$(2, 0.3, -0.3)$	200	23.9	100	100	100 [98.5]	100 [98.3]
	300	23.8	100	100	100 [99.0]	100 [98.7]

Note: The column labeled % reports the percentage of data in the stationary regimes. The figures into brackets are percentages of $|y_t|$ contained in Λ_T^U and Λ_T^B.

4.3. Autoregressive conditional root alternatives

Let us now check consistency of our SupWald test against a broader set of stationary alternatives. We consider here an ACR model proposed by Bec et al. (2005) and Gouriéroux and Robert (2006). This alternative is given by

$$y_t = (1+\rho)^{s_t} y_{t-1} + a\Delta y_{t-1} + \varepsilon_t,$$

where the transition variable s_t is binomial given the past, and specified by its conditional probability $\mathbb{P}(s_t = 1|y_{t-1}, \varepsilon_t) = [1 + \exp(-(\alpha + \beta|y_{t-1}|^{1/2}))]^{-1}$, ρ is a real number, β is non-negative and α and β are finite. In this model, ε_t is an i.i.d. $\mathcal{N}(0, \sigma^2)$. The Markov ACR model exhibits local nonstationarity when $s_t = 0$, which is more likely to arise if $\alpha + \beta|y_{t-1}|^{1/2}$ 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 (2.2). 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 (column '%' in Table 3) for interpretation's sake. The parameters values considered for this power analysis are motivated by Bec et al. (2005). In line with their ACR estimates for real exchange rate data, σ is set to 0.009, and our benchmark calibration sets a, α, β and ρ to 0.3, -10, 30 and -0.3, respectively. The results are reported in Table 3.

Again, the ADF test slightly dominates the other tests in the case where the time spent in the stationary regime ($s_t = 1$) is important. In the other cases, the unit-root tests based on the threshold specification do remarkably well while the ADF test has poor power. For instance, with the parameter values of the first DGP reported in Table 3, the time spent in the stationary regime is equal to 4.4%, and the rejection rate of $\mathscr{W}_B^{\mathrm{Sup}}$ is 75.1% compared to 27.2% for the ADF test. Finally, the power of $\mathscr{W}_B^{\mathrm{Sup}}$ slightly dominates the power of $\mathscr{W}_U^{\mathrm{Sup}}$ for 13 cases out of 16. The t_{NL} test slightly outperforms the $\mathscr{W}_B^{\mathrm{Sup}}$ test in two cases.

5. The yield spread dynamics revisited

We propose here an application of our adaptive approach to the analysis of the yield spread dynamics. Under costless and instantaneous portfolio adjustment assumption, arbitrage arguments often augmented by

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Table 3
Empirical power of the unit-root tests against ACR alternatives

(α, β, a, ρ)	T	%	ADF	t_{NL}	$\mathcal{W}_U^{\mathrm{Sup}}$	$\mathcal{W}_B^{\mathrm{Sup}}$
$(-10, 30, 0.3, -0.3)$	200	4.4	27.2	70.9	63.8 [96.5]	75.1 [98.4]
	300	4.4	61.9	91.5	86.0 [92.2]	93.3 [98.8]
$(-10, 30, 0.3, -0.1)$	200	7.9	14.3	20.9	17.7 [93.7]	27.0 [94.9]
	300	7.9	22.8	38.9	30.0 [88.5]	44.4 [97.9]
$(-15, 30, 0.3, -0.3)$	200	1.3	11.6	27.4	36.6 [91.7]	29.8 [79.4]
	300	1.2	14.4	37.0	39.8 [85.8]	41.0 [80.8]
$(-15, 30, 0.3, -0.1)$	200	2.1	8.1	13.8	11.9 [90.9]	10.7 [67.7]
	300	2.1	12.8	15.1	13.5 [85.0]	13.1 [66.9]
$(-20, 120, 0.3, -0.3)$	200	19.4	100	99.9	100 [98.5]	100 [98.4]
	300	19.4	100	100	100 [99.0]	100 [98.7]
$(-20, 120, 0.3, -0.1)$	200	35.0	89.3	81.6	64.4 [98.4]	72.0 [98.4]
	300	35.1	100	95.0	93.1 [98.1]	97.1 [98.8]
$(-10, 30, 0, -0.3)$	200	3.5	14.4	31.7	32.6 [93.5]	45.2 [97.5]
	300	3.5	24.4	53.4	47.6 [88.5]	68.2 [98.5]
$(-10, 30, 0, -0.1)$	200	6.3	10.0	11.2	10.4 [91.7]	14.6 [92.3]
	300	6.2	15.8	17.2	16.0 [86.1]	24.0 [94.5]

Note: The column labeled % reports the percentage of data in the stationary regimes. The figures into brackets are percentages of $|y_t|$ contained in Λ_T^U and Λ_T^B.

risk considerations leads to a very general relationship between yields of different maturities, i.e.

$$R(k,t) = \frac{1}{k}\left[\sum_{j=1}^{k} \mathrm{E}_t[R(1,t+j-1)]\right] + L(k,t), \tag{5.1}$$

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.[9] This in turn implies the stationarity of the yield spread between longer-term and shorter-term interest rates. Indeed, by rearranging (5.1), the spread which prevails may be expressed as

$$S^*(k,1,t) = R(k,t) - R(1,t) = \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j=1}^{i} \mathrm{E}_t[\Delta R(1,t+j)] + L(k,t), \tag{5.2}$$

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 et al. (1992) and Anderson (1997), Eq. (5.2) acts as an attractor as soon as the actual spread $S(k,1,t)$ differs from the equilibrium spread $S^*(k,1,t)$.

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 $\lambda < S(k,1,t) - S^*(k,1,t)$, or convert k-period bonds to 1-period bonds if and only if $S(k,1,t) - S^*(k,1,t) < -\lambda$. Therefore, in presence of transaction costs, the attraction toward Eq. (5.2) is inactive when

$$-\lambda < S(k,1,t) - S^*(k,1,t) < \lambda. \tag{5.3}$$

[9]For instance, the pure expectation hypothesis implies that $L(k,t)$ is zero while some other versions of the expectations hypothesis assert that the premia are constant over time.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Table 4
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.

Table 5
SupWald unit-root tests

	SupWald(Λ_T^B)	λ	Λ_T^B	Λ_T^U
	(1)	(2)	(3)	(4)
S_F	10.96	1.98	[0.09;7.86]	[0.29;2.77]
S_G	15.42	1.03	[0.11;3.63]	[0.55;2.54]
S_{NZ}	52.16	7.98	[0.06;11.76]	[0.39;9.40]
S_{US}	30.07	1.69	[0.04;5.98]	[0.29;2.68]

Note: See text.

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 $S^*(k,1,t)$. 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 4, 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 fails to reject the null of stationarity for the French spread while the ADF test leads to reject the unit-root for S_{NZ}. The values obtained for the SupWald(Λ_T) statistics—and reported in Table 5—have to be compared with the corresponding critical values given in Table 1. The lag order of the $a(L)$ polynomial in model (2.2) is chosen according to the BIC and Ljung–Box statistics which suggest $p=1$ for the European spreads, and $p=4$ for the remainders. For each series considered here, the same threshold value maximizes both SupWald(Λ_T^B) and SupWald(Λ_T^U)—see column (2) in Table 5. Hence, both test statistics reach exactly the same value which is reported in column (1). However, the SupWald(Λ_T^B) and SupWald(Λ_T^U) statistics depart from each other by the set of thresholds considered, as can be seen from columns (3) and (4).

[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).

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

The SupWald(Λ_T^B) and SupWald(Λ_T^U) statistics lead to the same conclusion in three cases out of four: the null is rejected by both tests at the 1%-level for S_{NZ} and S_{US} and at the 5%-level for S_{G}. Nevertheless, the French data provide an illustration of the potential discrepancy between the bounded and unbounded intervals approaches. Whereas the null can be rejected for S_{F} at the 15%-level according to the SupWald(Λ_T^B) critical values reported in Table 1, the unit root cannot be rejected at this level on the basis of the SupWald(Λ_T^U) statistic. Hence, this empirical application confirms the gain in power implied by the bounded interval—which has already been emphasized by the simulation experiments in the previous section. It is also worth noting that, as expected, the SupWald(Λ_T^B) test seems more adaptive that the SupWald(Λ_T^U) test, since there are always more observations lying in Λ_T^B than in Λ_T^U. The percentage of observations lying in Λ_T^B ranges from 97.2 for the US spread to 99.5 for German and New-Zealander data, whereas it never exceeds 86.2% for Λ_T^U. Finally, this empirical investigation of yield spread data provides support to the so-called expectations hypothesis, once the transaction costs are accounted for—i.e. once the nonlinear feature of the stationary alternative is allowed for.

6. Conclusion

This paper develops adaptive threshold SupWald unit-root tests as an alternative to linear ones. Adaptive threshold SupWald unit-root tests are consistent against stationary ergodic alternatives, therefore including a large variety of nonlinear processes relevant in macroeconomic or financial applications. A power bound indicates that these new unit-root tests can outperform linear ones, hence justifying the quite recent strand in econometrics literature which searches for such an improvement through the use of a more complex nonlinear specification to build a unit-root test. This theoretical finding is mostly confirmed by our simulation experiment. It is also illustrated by an empirical analysis of yield spread data: 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. But the power bound suggests that the order of the improvement crucially depends on the magnitude of the critical values of the SupWald test and our simulations reveal that this is especially true for alternatives close to linearity, against which linear tests better perform. Regarding the choice of a threshold set, the simulation experiments show that using bounded ones give a more powerful test, by producing smaller critical values under the null, and because the retained choice of the bounded Λ_T is larger under the alternative. In addition, considering such bounded set of thresholds may also improve the accuracy of the null limiting distribution, because they are smaller under the null.

7. Main assumptions and proofs

The proof section is organized as follows. We first state our main assumptions. We then prove Theorem 4, which implies Theorem 1 in Section 7.2. Section 7.3 contains a general functional extension of Park and Phillips (2001) which is used in Section 7.4 to establish Theorems 5 and 6. Our main limit results for the SupWald statistics (Theorems 2 and 3, Corollaries 1 and 2) are proved in Section 7.5, and an Appendix groups some useful lemmas and the proof of and intermediate result.

In what follows, $\xrightarrow{\mathbb{P}}$ denotes convergence in probability, $\xrightarrow{\mathrm{d}}$ stands for convergence in distribution and $\overset{\mathrm{d}}{=}$ is equality in distribution. Depending on the context, $\|\cdot\|$ denotes vector, matrix, or function norm, but in case of a vector, $\|z\|$ is the Euclidean norm of z. C is a generic constant that may vary from line to line.

7.1. Assumptions

Let us first introduce a limit theorem for sums of transformations of the y_{t-1}'s due to Park and Phillips (2001), see also Park and Phillips (1999). The next definitions are from Park and Phillips (2001). A map $f(\cdot)$ from $\mathbb{R}$ to $\mathbb{R}$ is *regular* if it is continuous in a neighborhood of infinity, and, for any compact subset C of $\mathbb{R}$, there exist some continuous functions $\underline{f}_\varepsilon(\cdot)$ and $\overline{f}_\varepsilon(\cdot)$ with $\lim_{\varepsilon\to 0}\int(\overline{f}_\varepsilon-\underline{f}_\varepsilon)(w)\,\mathrm{d}w=0$, and $\delta_\varepsilon>0$ such that $\underline{f}_\varepsilon(w')\leqslant f(w)\leqslant\overline{f}_\varepsilon(w')$ for all $|w-w'|\leqslant\delta_\varepsilon$ on C. A finite-dimensional vector of functions is regular if each entry is regular. Typical examples of such functions are the indicators $\mathbb{I}(w\in I_j(\lambda))$ $j=1,3$, of the lower and upper

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

regimes of the TAR model (3.4). A map $f(\cdot)$ from $\mathbb{R}$ to $\mathbb{R}$ is *I-regular* if it is integrable, square integrable, and satisfies the Lipschitz condition $|f(w)-f(w')|\leqslant K|w-w'|$ on its support. A finite-dimensional vector of functions is *I*-regular if each entry is *I*-regular. A typical example of such functions is the indicator $\mathbb{I}(w\in I_2(\lambda))$ of the central regime of (3.4) Following Park and Phillips (2001), we shall assume that

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

Assumption L. For $t\geqslant 1$ $y_t-y_{t-1}=\sum_{i=0}^{\infty}\pi_i\varepsilon_{t-i}$ with $y_0=0$, and where $\pi_0=1$, $\sum_{i=0}^{\infty}\pi_i\neq 0$ and $\sum_{i=1}^{\infty}i|\pi_i|<\infty$.

Note that $\mathrm{H}_0(p)$ is a special case of Assumption L. The next theorem combines Theorems 3.1 and 3.2 in Park and Phillips (2001).

Theorem 7 (*Park and Phillips (2001)*). *Let* $\sigma^2=\mathrm{Var}(\varepsilon_t)$, $\sigma_y^2=\sigma^2(\sum_{i=0}^{\infty}\pi_i)^2$, $\{W(v)\}_{v\in[0,1]}$ *and* $\{B(w)\}_{w\in\mathbb{R}}$ *be two independent standard Brownian Motions. Let* $\mathscr{F}_1$ *and* $\mathscr{F}_2$ *be collections of regular maps and I-regular maps, respectively. Then, under Assumptions L and E*(*s*), $s>4$, *the finite dimensional marginal distributions of the process indexed by* (f_1,f_2),

$$\left[\frac{1}{T}\sum_{t=1}^{T}f_1\left(\frac{y_{t-1}}{\sqrt{T}}\right),\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_1\left(\frac{y_{t-1}}{\sqrt{T}}\right)\varepsilon_t,\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_2\left(y_{t-1}\right),\frac{1}{T^{1/4}}\sum_{t=1}^{T}f_2\left(y_{t-1}\right)\varepsilon_t\right]',$$

$(f_1,f_2)\in\mathscr{F}_1\times\mathscr{F}_2$, *converge to the ones of*

$$\left[\int_0^1 f_1(\sigma_y W(v))\,\mathrm{d}v,\sigma\int_0^1 f_1(\sigma_y W(v))\,\mathrm{d}W(v),\frac{L_W(0,1)}{\sigma_y}\int f_2(w)\,\mathrm{d}w,\frac{\sigma L_W^{1/2}(0,1)}{\sigma_y^{1/2}}\int f_2(w)\,\mathrm{d}B(w)\right]'.$$

We now give our main assumption on the boundaries $\underline{\lambda}_T$, $\overline{\lambda}_T$ and v_T of Λ_T.

Assumption Λ. Set $\mathscr{F}_1=\{[1,w,w^2]\mathbb{I}(w\leqslant\lambda_1),[1,w,w^2]\mathbb{I}(\lambda_1\leqslant w<\lambda_2),[1,w,w^2]\mathbb{I}(w\geqslant\lambda_2),\lambda\in\mathbb{R}^2\}$ and $\mathscr{F}_2=\{[1,w,w^2]\mathbb{I}(\lambda_1<w<\lambda_2),\lambda\in\mathbb{R}^2\}$. Under H_0, the vector $[\underline{\lambda}_T,\overline{\lambda}_T,v_T]'$ converges in distribution to the finite vector $[\underline{\lambda},\overline{\lambda},v]'$ jointly with the convergence in distribution of Theorem 7. The random vector $[\underline{\lambda},\overline{\lambda},v]'$is measurable with respect to the σ field generated by the Brownian Motion $\{W(v)\}_{v\in[0,1]}$. Moreover, $v\geqslant 0$ and $\underline{\lambda}+v\leqslant\overline{\lambda}$ almost surely, i.e. Λ_T is nonempty asymptotically.

7.2. Proof of Theorem 4

We first introduce some related notations. M^{-1} denotes the Moore–Penrose pseudo-inverse of M. $\preceq$ denotes the usual order of symmetric matrix, i.e. $M_1\preceq M_2$ if and only if M_2-M_1 is nonnegative. We use the following generalization of the Cauchy-Schwarz Inequality. Let Y and X be random column vectors and $Y=BX+U$ with $B=\mathbb{E}[YX']\mathbb{E}^{-1}[XX']$. Then $\mathbb{E}[X'U]=0$ and $0\preceq\mathbb{E}[UU']$ yields that

$$0\preceq\mathbb{E}[YX']\mathbb{E}^{-1}[XX']\mathbb{E}[Y'X]\preceq\mathbb{E}[YY']. \tag{7.1}$$

Changing expectations into empirical means yields an empirical version of (7.1).

Let us stack the autoregressive coefficients $\rho_j, j=1,2,3$ of the TAR model into a vector $\overrightarrow{\rho}$. Under H_1 and stationarity of $\{y_t,s_t\}$, the limit $\rho_j(\lambda)$ of $\widehat{\rho}_j(\lambda)$ writes, for each λ

$$\overrightarrow{\rho}(\lambda)=R(\mathbb{E}[x_t'(\lambda)x_t(\lambda)])^{-1}\mathbb{E}[x_t'(\lambda)\Delta y_t] \text{ where } R \text{ is a selection matrix with } R\beta=\overrightarrow{\rho}. \tag{7.2}$$

The next lemma studies the $\rho_j(\lambda)$'s under H_1 when a regime grows.

Lemma 1. *Assume that the threshold variable* s_t *is chosen such that* $\{y_t,s_t\}$ *is stationary for any alternative* $\{y_t\}$ *in* H_1. *Let* $\mathbb{S}$ *be the support of the stationary* s_t. *Consider a restriction* (3.2) *of the TAR model* (3.1) *which nests the linear augmented autoregressive model through regimes* $\mathscr{J}$. *Then, for any* $\{y_t\}$ *in* H_1 *and* $j\in\mathscr{J}$, $\lim_{I_{\mathscr{J}}(\lambda)\to\mathbb{S}}\rho_j(\lambda)<0$. *Hence there exists a threshold parameter* λ^j *such that* $\rho_j(\lambda^j)<0$.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Proof of Lemma 1. Denote H_1 as $H_{1,p+1}$ to recall that $p+1$ lags are used in the autoregressive model (3.3). Write similarly $v_t = v_{t,p+1}$ and rewrite (3.3) as

$$A_{p+1}(L)(y_t - m) = v_{t,p+1} = y_t - \mu - (1+\rho)y_{t-1} - a(L)\Delta y_{t-1} \quad \text{so that } A_{p+1}(1) = 1 - (1+\rho) = -\rho,$$

where $A_{p+1}(L) = 1 + A_{1,p+1}L + \cdots + A_{p+1,p+1}L^{p+1}$ is given by the linear regression of y_t on $y_{t-1}, \ldots, y_{t-p-1}$, which is uniquely defined since $\mathrm{Var}([y_{t-1}, \ldots, y_{t-p-1}]')$ has an inverse under $H_{1,p+1}$.[13] In the equation above, m is the mean of y_t which can be set to 0 as done from now on, by changing y_t into $y_t - m$. We first show that, for all $p \geqslant 1$, the proposition $\mathscr{P}(p)$: $A_p(1) > 0$ for any $\{y_t\}$ in $H_{1,p}$. is true. We first show that $\mathscr{P}(1)$ is true. For $p = 1$, we have $A_1(L)y_t = y_t - A_{1,1}y_{t-1} = v_{t,1}$ with

$$A_{1,1} = \frac{\mathrm{Cov}(y_t, y_{t-1})}{\mathrm{Var}(y_{t-1})} = \frac{\mathrm{Cov}(y_t, y_{t-1})}{\mathrm{Var}^{1/2}(y_t)\mathrm{Var}^{1/2}(y_{t-1})} = \mathrm{Corr}(y_t, y_{t-1})$$

by stationarity. The Cauchy–Schwarz Inequality then yields that

$$-1 \leqslant A_{1,1} \leqslant 1 \text{ with } |A_{1,1}| = 1 \quad \text{if and only if } y_t = y_{t-1} \text{ or } y_t = -y_{t-1}.$$

Note that $A_{1,1} = 1$ is impossible, since it would give $y_t = y_0$ which is a process excluded by H_1. Then $A_1(1) = 1 - A_{1,1}$ is in $(0, 2]$ and $\mathscr{P}(1)$ is true.

We now show that $\mathscr{P}(p)$ is true for any $p > 1$ by a contradiction argument. Assume that $A_p(1) \leqslant 0$. Then, since $A_p(0) = 1$, $A_p(\cdot)$ has a root r in $(0, 1]$ by the Mean Value Theorem. Write $A_p(L) = (1 - L/r)\widetilde{A}_p(L)$ and consider the stationary process $\widetilde{y}_t = \widetilde{A}_p(L)y_t$. Note that $\{\widetilde{y}_t\}$ satisfies $H_{1,1}$ because $\{y_t\}$ is in $H_{1,p}$ and $\widetilde{y}_t = \widetilde{A}(L)y_t = \widetilde{A}_{0,p-1}y_t + \cdots + \widetilde{A}_{p-1,p-1}y_{t-p+1}$, so that $\mathrm{Var}[\widetilde{y}_t] = 0$ would contradict $H_{1,p}$ which implies that $\mathrm{Var}([y_t, \ldots, y_{t-p+1}]')$ has an inverse. Observe moreover that $\widetilde{y}_{t-1} = \widetilde{A}(L)y_{t-1} = \widetilde{A}_{0,p-1}y_{t-1} + \cdots + \widetilde{A}_{p-1,p-1}y_{t-p}$ is uncorrelated with $v_{t,p}$ since $A_p(L)y_t = v_{t,p}$ corresponds to the regression of y_t on $y_{t-1}, \ldots, y_{t-p}$. Because $A_p(L)y_t = (1 - L/r)\widetilde{A}_p(L)y_t = \widetilde{y}_t - \widetilde{y}_{t-1}/r = v_{t,p}$, $\widetilde{y}_t - \widetilde{y}_{t-1}/r = v_{t,p}$ is then the regression of $\widetilde{y}_t$ on $\widetilde{y}_{t-1}$. But $1 - 1/r \leqslant 0$ since r is in $(0, 1]$. This contradicts $\mathscr{P}(1)$, so that $A_p(1) > 0$ necessarily, and $\mathscr{P}(p)$ is true for any $p > 1$.

We now return to the proof of Lemma 1. Observe that $\rho_j = \rho_{\mathscr{J}}$ for all j in $\mathscr{J}$ by Definition 1. Let $\beta(\lambda) = \mathbb{E}^{-1}[x_t'(\lambda)x_t(\lambda)]\mathbb{E}[x_t'(\lambda)\Delta y_t]$. Split $\beta(\lambda)$ into $\beta_{\mathscr{J}}(\lambda)$ and $\beta_{-\mathscr{J}}(\lambda)$ such that $x_t(\lambda)\beta(\lambda) = x_{\mathscr{J}t}(\lambda)\beta_{\mathscr{J}}(\lambda) + x_{-\mathscr{J}t}(\lambda)\beta_{-\mathscr{J}}(\lambda)$, where $x_{\mathscr{J}t}(\lambda)$ and $x_{-\mathscr{J}t}(\lambda)$ are from Definition 1(ii). Note that $\rho_{\mathscr{J}}$ is one of the entries of $\beta_{\mathscr{J}}(\lambda)$. Let $x_t = [1, y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p}]$ and $\beta = \mathbb{E}^{-1}[x_t'x_t]\mathbb{E}[x_t\Delta y_t]$, noticing that $\mathbb{E}[x_t'x_t]$ has an inverse under H_1. The rest of the proof shows that $\lim_{I_{\mathscr{J}}(\lambda)\to\mathbb{S}} \beta_{\mathscr{J}}(\lambda) = \beta$ so that $\lim_{I_{\mathscr{J}}(\lambda)\to\mathbb{S}} \rho_{\mathscr{J}}(\lambda) = \rho = -A_{p+1}(1) < 0$ by $\mathscr{P}(p+1)$, which is the statement of Lemma 1. The Frish–Waugh Theorem yields

$$\begin{aligned}\beta_{\mathscr{J}}(\lambda) = {} & (\mathbb{E}[x_{\mathscr{J}t}'(\lambda)x_{\mathscr{J}t}(\lambda)] - \mathbb{E}[x_{\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}^{-1}[x_{-\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}[x_{-\mathscr{J}t}'(\lambda)x_{\mathscr{J}t}(\lambda)])^{-1} \\ & \times (\mathbb{E}[x_{\mathscr{J}t}'(\lambda)\Delta y_t] - \mathbb{E}[x_{\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}^{-1}[x_{-\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}[x_{-\mathscr{J}t}'(\lambda)\Delta y_t]).\end{aligned}$$

by Definition 1(ii), $\mathbb{E}[x_{-\mathscr{J}t}'(\lambda)z_t] = \mathbb{E}[x_{-\mathscr{J}t}'(\lambda)z_t\mathbb{I}(s_t \notin I_{\mathscr{J}}(\lambda))]$. We first show that

$$\lim_{I_{\mathscr{J}}(\lambda)\to\mathbb{S}} \mathbb{E}[x_{\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}^{-1}[x_{-\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}[x_{-\mathscr{J}t}'(\lambda)\Delta y_t] = 0. \tag{7.3}$$

Let a be a row vector. The Cauchy–Schwarz Inequality and (7.1) yields

$$\begin{aligned}&|a\mathbb{E}[x_{\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}^{-1}[x_{-\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}[x_{-\mathscr{J}t}'(\lambda)\Delta y_t]a'| \\ &\quad \leqslant [a\mathbb{E}[x_{\mathscr{J}t}'(\lambda)\mathbb{I}(s_t \notin I_{\mathscr{J}}(\lambda))x_{-\mathscr{J}t}(\lambda)]\mathbb{E}^{-1}[x_{-\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}[x_{-\mathscr{J}t}'(\lambda)x_{\mathscr{J}t}(\lambda)\mathbb{I}(s_t \notin I_{\mathscr{J}}(\lambda))]a']^{1/2} \\ &\qquad \times [a\mathbb{E}[\Delta y_t\mathbb{I}(s_t \notin I_{\mathscr{J}}(\lambda))x_{-\mathscr{J}t}(\lambda)]\mathbb{E}^{-1}[x_{-\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}[x_{-\mathscr{J}t}'(\lambda)\mathbb{I}(s_t \notin I_{\mathscr{J}}(\lambda))\Delta y_t]a']^{1/2} \\ &\quad \leqslant [a\mathbb{E}[x_{\mathscr{J}t}'(\lambda)x_{\mathscr{J}t}(\lambda)\mathbb{I}(s_t \notin I_{\mathscr{J}}(\lambda))]a']^{1/2}[a\mathbb{E}[(\Delta y_t^2)\mathbb{I}(s_t \notin I_{\mathscr{J}}(\lambda))]a']^{1/2} = o(1) \quad \text{when } I_{\mathscr{J}}(\lambda) \to \mathbb{S}.\end{aligned}$$

Definition 1(ii) gives, when $I_{\mathscr{J}}(\lambda) \to \mathbb{S}$,

$$\begin{aligned}0 &\preceq \mathbb{E}[x_{\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}^{-1}[x_{-\mathscr{J}t}'(\lambda)x_{-\mathscr{J}t}(\lambda)]\mathbb{E}[x_{-\mathscr{J}t}'(\lambda)x_{\mathscr{J}t}(\lambda)] \\ &\preceq \mathbb{E}[x_{\mathscr{J}t}'(\lambda)x_{\mathscr{J}t}(\lambda)\mathbb{I}(s_t \notin I_{\mathscr{J}}(\lambda))] \to 0.\end{aligned}$$

[13] If not $y_t = b_0 + b_1y_{t-1} + \cdots + y_{t-p-1}$ so that, as a solution of such linear recurrence equation, y_t is a product of polynomial and exponential functions of t. Hence, under stationarity, y_t must be a constant process, which is excluded by definition of H_1.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

As well, $\lim_{I_{\mathscr{J}}(\lambda)\to\mathbb{S}} \mathbb{E}[x'_{\mathscr{J}t}(\lambda)x_{\mathscr{J}t}(\lambda)] = \mathbb{E}[x'_t x_t]$ and $\lim_{I_{\mathscr{J}}(\lambda)\to\mathbb{S}} \mathbb{E}[x'_{\mathscr{J}t}(\lambda)\Delta y_t] = \mathbb{E}[x'_t \Delta y_t]$. Therefore, the continuity of $M \mapsto M^{-1}$ at $\{M; \mathrm{Det}(M) \neq 0\}$, the approximations above together with the expression of $\beta_{\mathscr{J}}(\lambda)$ show that $\lim_{I_{\mathscr{J}}(\lambda)\to\mathbb{S}} \beta_{\mathscr{J}}(\lambda) = \beta$. □

Proof of Theorem 4. We give first a suitable expression of the Wald statistics $\mathrm{Wald}_T(\lambda)$ and ADF_T^2, and collect important facts. Let $\widehat{U}(\lambda)$ and $\widehat{U}_0(\lambda)$ be the column vector of estimated residuals $\widehat{u}_t(\lambda) = \Delta y_t - x_t(\lambda)\widehat{\beta}_T(\lambda)$ and $\widehat{u}_{0t}(\lambda) = \Delta y_t - x_t(\lambda)\widehat{\beta}_{0T}(\lambda)$, $t = 1, \ldots, T$ from the TAR model (3.2), where $\widehat{\beta}_{0T}(\lambda)$ is the OLS estimator of $\beta(\lambda)$ under the constraint $\overrightarrow{\rho}(\lambda) = 0$. Let similarly $\widehat{V}$ and $\widehat{V}_0$ be the vector of estimated residuals of the linear model (3.3), where $\widehat{V}_0$ is computed under the restriction $\rho = 0$. The Wald statistics write (see e.g. Gouriéroux and Monfort, 1995)

$$\mathrm{Wald}_T(\lambda) = T\left(1 - \frac{\widehat{U}'(\lambda)\widehat{U}(\lambda)}{\widehat{U}_0'(\lambda)\widehat{U}_0(\lambda)}\right) \quad \text{and} \quad ADF_T^2 = T\left(1 - \frac{\widehat{V}'\widehat{V}}{\widehat{V}_0'\widehat{V}_0}\right). \tag{7.4}$$

Note that $\widehat{V}'\widehat{V}$ and $\widehat{V}_0'\widehat{V}_0$ are continuous functions of sums of empirical covariance matrices. Therefore, under H_1, the Ergodic Law of Large Numbers yields $\widehat{V}'\widehat{V}/T = \sigma_v^2 + o_{\mathbb{P}}(1)$ and $\widehat{V}_0'\widehat{V}_0/T = \sigma_{v0}^2 + o_{\mathbb{P}}(1)$ where σ_v^2 and σ_{v0}^2 are the innovations variance of the general and constrained linear models (3.3). Under H_1, $\sigma_v^2 > \sigma_{v0}^2 > 0$ since $\rho \neq 0$ as shown by $\mathscr{P}(p+1)$ in the proof of Lemma 1, so that ADF_T^2 diverges with the exact order T in probability.

Let $\Delta Y = [\Delta y_T, \ldots, \Delta y_1]'$, $X_{\mathscr{J}}(\lambda) = [x'_{\mathscr{J}T}(\lambda), \ldots, x'_{\mathscr{J}1}(\lambda)]'$, $X_{-\mathscr{J}}(\lambda) = [x'_{-\mathscr{J}T}(\lambda), \ldots, x'_{-\mathscr{J}1}(\lambda)]'$, and $X = [x'_T, \ldots, x'_1]'$ with $x_t = [1, y_{t-1}, \Delta y_{t-1}, \ldots \Delta y_{t-p}]$. Let

$$P_\lambda = \mathrm{Id} - X_{-\mathscr{J}}(\lambda)(X'_{-\mathscr{J}}(\lambda)X_{-\mathscr{J}}(\lambda))^{-1}X'_{-\mathscr{J}}(\lambda)$$

be the orthogonal projection on the linear span orthogonal to the columns of $X_{-\mathscr{J}}(\lambda)$. Let λ_T be the deterministic sequence of Definition 2. Note that $\widehat{U}(\lambda_T)$ is ΔY minus its orthogonal projection on the linear subspace spanned by $[X_{\mathscr{J}}(\lambda_T), X_{-\mathscr{J}}(\lambda_T)]$, which is also the space spanned by $[P_{\lambda_T}X_{\mathscr{J}}(\lambda_T), X_{-\mathscr{J}}(\lambda_T)]$. The definition of P_{λ_T} gives $X'_{-\mathscr{J}}(\lambda_T)P_{\lambda_T}X_{\mathscr{J}} = 0$ and then

$$\frac{\widehat{U}'(\lambda_T)\widehat{U}(\lambda_T)}{T} = \frac{\Delta Y' P_{\lambda_T}X_{\mathscr{J}}(\lambda_T)}{T}(X'_{\mathscr{J}}(\lambda_T)P_{\lambda_T}X_{\mathscr{J}}(\lambda_T))^{-1}\frac{X'_{\mathscr{J}}(\lambda_T)P_{\lambda_T}\Delta Y}{T} \tag{7.5}$$

$$+ \frac{\Delta Y' X_{-\mathscr{J}}(\lambda_T)}{T}(X'_{-\mathscr{J}}(\lambda_T)X_{-\mathscr{J}}(\lambda_T))^{-1}\frac{X'_{-\mathscr{J}}(\lambda_T)\Delta Y}{T}. \tag{7.6}$$

We first study (7.5) and begin by showing that $T^{-1}\Delta Y' P_{\lambda_T}X_{\mathscr{J}}(\lambda_T) = T^{-1}\Delta Y' X + o_{\mathbb{P}}(1)$. Note that

$$\frac{\Delta Y' P_{\lambda_T}X_{\mathscr{J}}(\lambda_T)}{T} = \frac{\Delta Y' X_{\mathscr{J}}(\lambda_T)}{T} + \frac{\Delta Y' X_{-\mathscr{J}}(\lambda_T)(X'_{-\mathscr{J}}(\lambda_T)X_{-\mathscr{J}}(\lambda_T))^{-1}X'_{-\mathscr{J}}(\lambda_T)X_{\mathscr{J}}(\lambda_T)}{T}.$$

We show that $T^{-1}\Delta Y' X_{\mathscr{J}}(\lambda_T) = T^{-1}\Delta Y' X + o_{\mathbb{P}}(1)$. We have by stationarity and Definition 1(ii)

$$\mathbb{E}\left\|\frac{\Delta Y'(X_{\mathscr{J}}(\lambda_T) - X)}{T}\right\| = \mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T}\Delta y_t(x_{\mathscr{J}t}(\lambda_T) - x_t)\right\| \leqslant \mathbb{E}[|\Delta y_t|\|x_{\mathscr{J}t}(\lambda_T) - x_t\|]$$

$$\leqslant \mathbb{E}^{1/2}[(\Delta y_t)^2] \times \mathbb{E}^{1/2}[\|x_{\mathscr{J}t}(\lambda_T) - x_t\|^2] = o(1). \tag{7.7}$$

Therefore, $T^{-1}\Delta Y' X_{\mathscr{J}}(\lambda_T) = T^{-1}\Delta Y' X + o_{\mathbb{P}}(1)$ and we now show that the second item in $T^{-1}\Delta Y' P_{\lambda_T}X_{\mathscr{J}}(\lambda_T)$ can be neglected. Let a be a column vector of dimension $p+2$. Then the Cauchy–Schwarz Inequality,

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Definition 1(ii) and (7.1) yield

$$\begin{aligned}
&\left|\frac{\Delta Y' X_{-\mathscr{J}}(\lambda_T)(X'_{-\mathscr{J}}(\lambda_T)X_{-\mathscr{J}}(\lambda_T))^{-1}X'_{-\mathscr{J}}(\lambda_T)X_{\mathscr{J}}(\lambda_T)a}{T}\right| \\
&\quad\leqslant \left|\frac{\Delta Y' X_{-\mathscr{J}}(\lambda_T)(X'_{-\mathscr{J}}(\lambda_T)X_{-\mathscr{J}}(\lambda_T))^{-1}X'_{-\mathscr{J}}(\lambda_T)\Delta Y}{T}\right|^{1/2} \\
&\qquad\times \left|\frac{a'X'_{\mathscr{J}}(\lambda_T)X_{-\mathscr{J}}(\lambda_T)(X'_{-\mathscr{J}}(\lambda_T)X_{-\mathscr{J}}(\lambda_T))^{-1}X'_{-\mathscr{J}}(\lambda_T)X_{\mathscr{J}}(\lambda_T)a}{T}\right|^{1/2} \\
&\quad= \left|\frac{(\sum_{t=1}^T \Delta y_t \mathbb{I}(s_t\notin I_{\mathscr{J}}(\lambda_T))x_{-\mathscr{J}}(\lambda_T))(X'_{-\mathscr{J}}(\lambda_T)X_{-\mathscr{J}}(\lambda_T))^{-1}(\sum_{t=1}^T \Delta y_t \mathbb{I}(s_t\notin I_{\mathscr{J}}(\lambda_T))x'_{-\mathscr{J}}(\lambda_T))}{T}\right|^{1/2} \\
&\qquad\times \left|\frac{a'X'_{\mathscr{J}}(\lambda_T)X_{-\mathscr{J}}(\lambda_T)(X'_{-\mathscr{J}}(\lambda_T)X_{-\mathscr{J}}(\lambda_T))^{-1}X'_{-\mathscr{J}}(\lambda_T)X_{\mathscr{J}}(\lambda_T)a}{T}\right|^{1/2} \\
&\quad\leqslant \left(\frac{1}{T}\sum_{t=1}^T (\Delta y_t)^2 \mathbb{I}(s_t\notin I_{\mathscr{J}}(\lambda_T))\right)^{1/2}\left(\sum_{t=1}^T a'x'_{\mathscr{J}t}(\lambda_T)x_{\mathscr{J}t}(\lambda_T)\mathbb{I}(s_t\notin I_{\mathscr{J}}(\lambda_T))a\right)^{1/2}.
\end{aligned}$$

As in (7.7),

$$\mathbb{E}\left|\frac{1}{T}\sum_{t=1}^T z_t \mathbb{I}(s_t\notin I_{\mathscr{J}}(\lambda_T))\right| \leqslant \mathbb{E}|z_t \mathbb{I}(s_t\notin I_{\mathscr{J}}(\lambda_T))| = \mathrm{o}(1), \quad z_t = (\Delta y_t)^2, a'x'_{\mathscr{J}t}(\lambda_T)x_{\mathscr{J}t}a, \tag{7.8}$$

for any a, so that the second item of $T^{-1}\Delta Y' P_{\lambda_T} X_{\mathscr{J}}(\lambda_T)$ can be neglected, and $T^{-1}\Delta Y' P_{\lambda_T} X_{\mathscr{J}}(\lambda_T) = T^{-1}\Delta Y' X + \mathrm{o}_{\mathbb{P}}(1)$. Arguing similarly with (7.7) and (7.8) yields $T^{-1}X'_{\mathscr{J}}(\lambda_T)P_{\lambda_T}X_{\mathscr{J}}(\lambda_T) = T^{-1}X'X + \mathrm{o}_{\mathbb{P}}(1)$. Note that $T^{-1}\Delta Y'X = \mathbb{E}[\Delta y_t x_t] + \mathrm{o}_{\mathbb{P}}(1)$ and $T^{-1}X'X = \mathbb{E}[x'_t x_t] + \mathrm{o}_{\mathbb{P}}(1)$, $\mathbb{E}[x'_t x_t]$ having an inverse, so that (7.5) equals $T^{-1}\Delta Y'X(X'X)^{-1}X'\Delta Y + \mathrm{o}_{\mathbb{P}}(1)$, with $T^{-1}\Delta Y'X(X'X)^{-1}X'\Delta Y = T^{-1}\widehat{V}'\widehat{V}$.Note that arguing as above with (7.1) and (7.8) yields that (7.6) is $\mathrm{o}_{\mathbb{P}}(1)$, so that $T^{-1}\widehat{U}'(\lambda_T)\widehat{U}(\lambda_T) = T^{-1}\widehat{V}'\widehat{V} + \mathrm{o}_{\mathbb{P}}(1)$. Repeating the same steps for the restricted models yields $T^{-1}\widehat{U}'_0(\lambda_T)\widehat{U}_0(\lambda_T) = T^{-1}\widehat{V}'_0\widehat{V}_0 + \mathrm{o}_{\mathbb{P}}(1)$. Substituting into (7.4) now yields

$$\mathrm{SupWald}_T(\Lambda_T) \geqslant \mathrm{Wald}_T(\lambda_T) + \mathrm{o}_{\mathbb{P}}(1) = T\left(1 - \frac{T^{-1}\widehat{V}'\widehat{V} + \mathrm{o}_{\mathbb{P}}(1)}{T^{-1}\widehat{V}'_0\widehat{V}_0 + \mathrm{o}_{\mathbb{P}}(1)}\right) + \mathrm{o}_{\mathbb{P}}(1) = ADF_T^2(1 + \mathrm{o}_{\mathbb{P}}(1)). \qquad \square$$

7.3. *Functional limit distribution for integrated processes*

Theorems 5 and 6 will be derived from a functional version of Theorem 7 from Park and Phillips (2001). Stating this more general result requires additional notations from Empirical Processes Theory that we introduce now. For a collection of functions $\mathscr{F}$ and a functional $z(\cdot)$ over $\mathscr{F}$ (as for instance empirical sums), $\|z\|_{\mathscr{F}} = \sup_{f\in\mathscr{F}}\|z(f)\|$ is the uniform norm of $z(\cdot)$ over $\mathscr{F}$. The functional space $\ell^\infty(\mathscr{F}) = \{z(\cdot); \|z\|_{\mathscr{F}} < \infty\}$ is the space of bounded functionals over $\mathscr{F}$ and is equipped with the norm $\|\cdot\|_{\mathscr{F}}$. More specifically, let $Z_T(f)$ be a sum from Theorem 7 and $\mathscr{F}$ be the associated collection $\mathscr{F}_j$, $j = 1, 2$. $\{Z_T(f)\}_{f\in\mathscr{F}}$ can be viewed as a stochastic process indexed by the function $f(\cdot)$. It is convenient to consider that $\mathscr{F}$ is a subset of a larger collection $\mathscr{E}$, as for instance the set of bounded measurable functions. An *envelope* of $\mathscr{F}$ is a function $F(\cdot)$ such that, for any $f(\cdot)$ in $\mathscr{F}$, $|f(w)| \leqslant F(w)$ for all w in $\mathbb{R}$. Let $\|\cdot\|$ be a norm over $\mathscr{E}$. For $f(\cdot)$ and$f'(\cdot)$ in $\mathscr{E}$, the *bracket* $[f, f']$ is a set of functions $g(\cdot)$ of $\mathscr{E}$ with $f(w) \leqslant g(w) \leqslant f'(w)$ for all w. An ε-*bracket* is a bracket $[f, f']$ with $\|f - f'\| \leqslant \varepsilon$. The *bracketing number* $N_{[]}(\varepsilon, \mathscr{F}, \|.\|)$ is the minimum number of ε-brackets of $\mathscr{E}$ needed to cover $\mathscr{F}$. For vector-valued functions, envelope and brackets can be defined components by components. In view of the form of the limits in Theorem 7, we use the L_1 norm with respect to the Lebesgue measure,

assuming that it also controls the L_q norm, $q = 2, 3, 4$. Recall that the L_q norm $\|\cdot\|_q$ with respect to Lebesgue measure is $\|f\|_q = (\int \|f(w)\|^q \, dw)^{1/q}$, $1 \leqslant q < \infty$.

Theorem 8. *Assume that Assumptions E(s), $s > 4$, and L hold. Let $\mathscr{F}_1$ be a class of regular functions with a continuous envelope, and $\mathscr{F}_2$ be a class of I-regular functions with a continuous envelope.*

Assume that there exist some sets $\mathscr{E}_j$ with $\mathscr{F}_j \subset \mathscr{E}_j$, $j = 1, 2$, and some constant C_q, $q = 2, 3, 4$ such that, for any f and f' in $\mathscr{E}_j$, $j = 1, 2$,

$$\|f - f'\|_2^2 \leqslant C_2 \|f - f'\|_1, \quad \|f - f'\|_3^3 \leqslant C_3 \|f - f'\|_1 \quad \text{and} \quad \|f - f'\|_4^4 \leqslant C_4 \|f - f'\|_1. \tag{7.9}$$

Assume moreover that, for $C_5 > 0$

$$N_{[]}(\varepsilon, \mathscr{F}_j, \|\cdot\|_1) \leqslant C_5 \varepsilon^{-\zeta} \quad \text{for some } \zeta \text{ in } (0, 1] \text{ and any } \varepsilon > 0, \ j = 1, 2. \tag{7.10}$$

Then the convergence in distribution of Theorem 7 holds in $\ell^\infty(\mathscr{F}_1) \times \ell^\infty(\mathscr{F}_2)$.

Given the finite dimensional convergence in distribution stated in Theorem 7, proving Theorem 8 works by showing *asymptotic stochastic equicontinuity*. Let $Z_T(f)$ be a sum of Theorem 7 and $\mathscr{F} \subset \mathscr{E}$ be the associated collection $\mathscr{F}_j$, $j = 1, 2$. Asymptotic stochastic equicontinuity of $\{Z_T(f)\}_{f \in \mathscr{F}}$ means that, for every $\delta_0, \delta_1 > 0$, there exists a finite covering $F_1, \ldots, F_n$ of $\mathscr{F}$ such that

$$\limsup_{T \to +\infty} \mathbb{P}\left(\sup_{1 \leqslant i \leqslant n} \sup_{(f, f') \in F_i^2} |Z_T(f') - Z_T(f)| \geqslant \delta_0 \right) \leqslant \delta_1.$$

see Theorem 18.14 in van der Vaart (1996), or Theorem 1.5.6 in van der Vaart and Wellner (1996). As discussed by these authors, finding a covering $F_1, \ldots, F_n$ can be done through any arbitrary norm $\|\cdot\|$ on the index set $\mathscr{F}$. In what follows, we say that $\{Z_T(f)\}_{f \in \mathscr{F}}$ is *asymptotically stochastically $\|\cdot\|$-equicontinuous* if for any $\delta_0, \delta_1 > 0$ there is a δ such that

$$\limsup_{T \to +\infty} \mathbb{P}\left(\sup_{(f, f') \in \mathscr{F}^2; \|f' - f\| \leqslant \delta} |Z_T(f') - Z_T(f)| \geqslant \delta_0 \right) \leqslant \delta_1. \tag{7.11}$$

We specifically have

Lemma 2. *Assume that Assumptions E(0) and L hold. Let $\mathscr{E}$ and $\mathscr{F} \subset \mathscr{E}$ be some collections of functions such that (7.9) holds over $\mathscr{E}$, and that the bracketing entropy numbers of $\mathscr{F}$ satisfy (7.10). Then*

$$\left\{ \frac{1}{T} \sum_{t=1}^{T} f\left(\frac{y_{t-1}}{\sqrt{T}} \right) \right\}_{f \in \mathscr{F}}, \quad \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f\left(\frac{y_{t-1}}{\sqrt{T}} \right) \varepsilon_t \right\}_{f \in \mathscr{F}},$$

$$\left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(y_{t-1}) \right\}_{f \in \mathscr{F}} \quad \text{and} \quad \left\{ \frac{1}{T^{1/4}} \sum_{t=1}^{T} f(y_{t-1}) \varepsilon_t \right\}_{f \in \mathscr{F}}$$

are asymptotically stochastically $\|\cdot\|_1$-equicontinuous.

Assume moreover that E(s) holds with $s > 14$, and that, for $q = (s+4)/(s+1), 2(s+4)/(s+1), 3(s+4)/(s+1)$, $q = 2(s+4)/s$, $4(s+4)/s$, there are some constants C_q such that for all $f(\cdot)$, $f'(\cdot)$ in $\mathscr{E}$

$$\|f - f'\|_q^q \leqslant C_q \|f - f'\|_1. \tag{7.12}$$

Then, for any integer number $k \geqslant 1$, $\{1/\sqrt{T} \sum_{t=1}^{T} \Delta y_{t-k} f(y_{t-1}/\sqrt{T}) \varepsilon_t\}_{f \in \mathscr{F}}$ is asymptotically stochastically $\|\cdot\|_1$-equicontinuous.

Proof of Lemma 2. See Appendix.

Proof of Theorem 8. Let $\mathscr{F}$ be $\mathscr{F}_1$ or $\mathscr{F}_2$. The convergence statement of Theorem 8 follows from Theorem 7 and asymptotic stochastic equicontinuity given by Lemma 2, provided that each sums of the theorem are in $\ell^\infty(\mathscr{F})$. This is due to the fact that the theorem assumes that $\mathscr{F}_1$ and $\mathscr{F}_2$ have continuous envelopes which give, respectively, finite norms $\|\cdot\|_{\mathscr{F}_1}$ or $\|\cdot\|_{\mathscr{F}_2}$ for each sums. □

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

7.4. Proof of Theorems 5 and 6

We use here lemmas from the Appendix. The proofs of Theorems 5 and 6 are also based on the following preliminary result.

Lemma 3. *Let $f(w;\theta)$ be $\mathbb{I}(w\leqslant\theta)$, $\mathbb{I}(w\geqslant\theta)$, $w\mathbb{I}(w\leqslant\theta)$, $w\mathbb{I}(w\geqslant\theta)$, $w^2\mathbb{I}(w\leqslant\theta)$, $w^2\mathbb{I}(w\geqslant\theta)$. Then, under Assumptions L and E(0), for any $a>0$,*

$$\left\{\frac{1}{T}\sum_{t=1}^{T} f\left(\frac{y_{t-1}}{\sqrt{T}};\theta\right)\right\}_{\theta\in[-a,a]}, \quad \left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f\left(\frac{y_{t-1}}{\sqrt{T}};\theta\right)\varepsilon_t\right\}_{\theta\in[-a,a]},$$

$$\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(y_{t-1};\theta)\right\}_{\theta\in[-a,a]} \quad \textit{and} \quad \left\{\frac{1}{T^{1/4}}\sum_{t=1}^{T} f(y_{t-1};\theta)\varepsilon_t\right\}_{\theta\in[-a,a]}$$

satisfy Conditions (7.9) *and* (7.10) *of Theorem* 8 *and are asymptotically stochastically* $|\cdot|$*-equicontinuous.*

Assume in addition that E(s) holds for $s>14$. Then, if $f(w;\theta)$ denotes $\mathbb{I}(w\leqslant\theta)$ or $\mathbb{I}(w\geqslant\theta)$, $\{T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-k} f(y_{t-1}/\sqrt{T};\theta)\varepsilon_t\}_{\theta\in[-a,a]}$, $k\geqslant 1$, is asymptotically stochastically $|\cdot|$-equicontinuous.

Proof of Lemma 3. To prove the first asymptotic stochastic equicontinuity result, we satisfy the conditions (7.9) and (7.10) of Theorem 8 and apply Lemma 2. Take $\mathscr{E}=\mathscr{F}=\{f(\cdot;\theta)\}_{\theta\in[-a,a]}$. Writting $f(w;\theta)$ as $f(w)\mathbb{I}(w\leqslant\theta)$ or $f(w)\mathbb{I}(w\geqslant\theta)$, we have for any integer numbers q

$$\begin{aligned}\|f(\cdot;\theta_2)-f(\cdot;\theta_1)\|_q^q &= \left|\int_{\theta_2}^{\theta_1}|f(w)|^q\,\mathrm{d}w\right| \\ &= \begin{cases} |\theta_2-\theta_1| & \text{if } f(w)=1,\\ \dfrac{|\theta_2^{q+1}-\theta_1^{q+1}|}{q+1}=\dfrac{|\theta_2^q+\theta_1\theta_2^{q-1}+\cdots+\theta_1^q|}{q+1}|\theta_2-\theta_1| & \text{if } f(w)=w,\\ \dfrac{|\theta_2^{2q+1}-\theta_1^{2q+1}|}{2q+1}=\dfrac{|\theta_2^{2q}+\theta_1\theta_2^{2q-1}+\cdots+\theta_1^{2q}|}{2q+1}|\theta_2-\theta_1| & \text{if } f(w)=w^2,\end{cases}\end{aligned} \tag{7.13}$$

so that (7.9) holds with $C_q=1$ if $f(w)=1$,

$$C_q=\max_{(\theta_1,\theta_2)\in[-a,a]^2}\left|\frac{\theta_2^q+\theta_1\theta_2^{q-1}+\cdots+\theta_1^q}{q+1}\right|, \quad C_q=\max_{(\theta_1,\theta_2)\in[-a,a]^2}\left|\frac{\theta_2^{2q}+\theta_1\theta_2^{q-1}+\cdots+\theta_1^{2q}}{2q+1}\right|,$$

for $f(w)=w, f(w)=w^2$, respectively. For $f(w;\theta)=f(w)\mathbb{I}(w\leqslant\theta)$ (the other case being symmetric) brackets can be taken here as $[f(\cdot;\theta_i),f(\cdot;\theta_{i+1})]=\{f(\cdot;\theta);\theta_i\leqslant\theta\leqslant\theta_{i+1}\}$. Eq. (7.13) yields that there is a $C>0$ such that $\|f(\cdot;\theta_2)-f(\cdot;\theta_1)\|_1\leqslant C|\theta_2-\theta_1|$ for all θ_1, θ_2 in the compact interval $[-a,a]$, so that (7.10) holds with $\zeta=1$. Due to the latter inequality, asymptotic stochastic $\|\cdot\|_1$-equicontinuity implies here asymptotic stochastic $|\cdot|$-equicontinuity. The asymptotic stochastic equicontinuity of $T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-k}f(T^{-1/2}y_{t-1};\theta))\varepsilon_t$ on $[-a,a]$ follows from Lemma 2 because, for $f(w)=1$, (7.13) holds for any $q>0$, implying (7.12) in Lemma 2. □

Proof of Theorem 5. Assume for brevity that $\sigma_y=1$. Observe that the processes of the theorem are in $\ell^\infty([-a,a]^2)$. Note that $T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-k}\varepsilon_t$, $k\leqslant 1$, is a centered martingale with $\mathbb{E}(\Delta y_{t-k}\varepsilon_t)^2<\infty$ (see Lemma A.2 in the Appendix), so that $T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-k}\varepsilon_t=\mathrm{O}_{\mathbb{P}}(1)$.

We first show that $\sup_{\lambda\in[-a,a]^2}\|\sum_{t=1}^{T}\Delta X_{jt}^U(\lambda)\varepsilon_t\|=\mathrm{O}_{\mathbb{P}}(1)$, $j=1,2,3$. Since $\Delta X_{1t}^U(\lambda)+\Delta X_{2t}^U(\lambda)+\Delta X_{3t}^U(\lambda)=T^{-1/2}[\Delta y_{t-1},\ldots,\Delta y_{t-k}]$ and $T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-k}\varepsilon_t=\mathrm{O}_{\mathbb{P}}(1)$, $1\leqslant k\leqslant p$, it is sufficient to take $j=1,3$, and to prove the result for $j=1$ by symmetry. Because the entries $T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-k}\mathbb{I}(y_{t-1}/\sqrt{T}\leqslant\lambda_1)\varepsilon_t$, $1\leqslant k\leqslant p$, of

$\sum_{t=1}^{T}\Delta X_{1t}^{U}(\lambda)\varepsilon_t$ are centered martingales with a variance bounded by $\mathbb{E}(\Delta y_{t-k}\varepsilon_t)^2$, $T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-k}\mathbb{I}(y_{t-1}/\sqrt{T}\leqslant\lambda_1)\varepsilon_t = \mathrm{O}_{\mathbb{P}}(1)$ for all λ_1. That

$$\sup_{\lambda_1\in[-a,a]}\left|T^{-1/2}\sum_{t=1}^{T}\Delta y_{t-k}\mathbb{I}(y_{t-1}/\sqrt{T}\leqslant\lambda_1)\varepsilon_t\right| = \mathrm{O}_{\mathbb{P}}(1)$$

follows from the asymptotic stochastic equicontinuity established in Lemma 3.

For the rest of the theorem, we firstly derive the finite-dimensional limit distribution. Note that

$$\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'\Delta X_{jt}^{U}(\lambda) = \frac{\Omega}{T}\sum_{t=1}^{T}\mathbb{I}\left(\frac{y_{t-1}}{\sqrt{T}}\in I_j(\lambda)\right) + \frac{1}{T}\sum_{t=1}^{T}\{[\Delta y_{t-k}\Delta y_{t-k'}]_{1\leqslant k,k'\leqslant p} - \Omega\}\mathbb{I}\left(\frac{y_{t-1}}{\sqrt{T}}\in I_j(\lambda)\right) \tag{7.14}$$

and consider first the process

$$\left\{\sum_{t=1}^{T}X_{jt}^{U}(\lambda)'\varepsilon_t, \sum_{t=1}^{T}X_{jt}^{U}(\lambda)'X_{jt}^{U}(\lambda), \frac{\Omega}{T}\sum_{t=1}^{T}\mathbb{I}\left(\frac{y_{t-1}}{\sqrt{T}}\in I_j(\lambda)\right),\ j=1,2,3\right\}_{\lambda\in[-a,a]^2}. \tag{7.15}$$

Let $\mathscr{F} = \{[1, w, w^2]\mathbb{I}(w\in I_j(\lambda)), j=1,2,3, \lambda\in[-a,a]^2\}$ be the class of functions appearing in the process (7.15). Observe that $\mathscr{F}$ is a class of regular functions, so that Theorem 7 yields that the finite dimensional limit distribution of (7.15) is the limit distribution of the lemma. To check for the asymptotic stochastic equicontinuity of (7.15), note that $\mathbb{I}(w\in I_2(\lambda)) = 1-\mathbb{I}(w\in\leqslant\lambda_1)-\mathbb{I}(w\geqslant\lambda_2)$, so that Lemma 3 establishes that (7.15) is asymptotic stochastic equicontinuity as sum of such processes. It follows that (7.15) converges in distribution in $\ell^{\infty}([-a,a]^2)$ to the limit of Theorem 5.

To complete the proof of the theorem, it remains to show that the remainder term (7.14) and $\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'X_{jt}^{U}(\lambda)$ are uniformly negligible over $[-a,a]$. This is done in the next two steps which show (1) the asymptotic stochastic equicontinuity of $\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'X_{jt}^{U}(\lambda)$ and $\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'\Delta X_{jt}^{U}(\lambda)$ (since, $\Omega T^{-1}\sum_{t=1}^{T}\mathbb{I}(y_{t-1}/\sqrt{T}\in I_j(\lambda))$ is asymptotically stochastically equicontinuous as shown above); (2) pointwise convergence in probability to 0 of $\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'X_{jt}^{U}(\lambda)$ and of the remainder term (7.14).

Step 1: *Asymptotic stochastic equicontinuity*. We begin with $\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'X_{jt}^{U}(\lambda)$. Because

$$\Delta X_{1t}^{U}(\lambda)'X_{1t}^{U}(\lambda)+\Delta X_{2t}^{U}(\lambda)'X_{2t}^{U}(\lambda)+\Delta X_{3t}^{U}(\lambda)'X_{3t}^{U}(\lambda) = \frac{1}{T}[\Delta y_{t-1},\ldots,\Delta y_{t-p}]'\left[1,\frac{y_{t-1}}{\sqrt{T}}\right]$$

does not depend upon λ, it is sufficient to study $j=1$. The absolute values of entries of the increments of $\sum_{t=1}^{T}\Delta X_{1t}^{U}(\lambda_1)'X_{1t}^{U}(\lambda_1)$ are such that, with $f_T(w)=1$ or $f_T(w)=w/\sqrt{T}$,

$$\left|\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-k}f_T(y_{t-1})\mathbb{I}\left(\lambda_1\leqslant\frac{y_{t-1}}{\sqrt{T}}<\lambda_1'\right)\right| \leqslant \left(\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{t-k})^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}f_T^2(y_{t-1})\mathbb{I}\left(\lambda_1\leqslant\frac{y_{t-1}}{\sqrt{T}}<\lambda_1'\right)\right)^{1/2} = \mathrm{O}_{\mathbb{P}}(1)\left(\frac{1}{T}\sum_{t=1}^{T}f_T^2(y_{t-1})\mathbb{I}\left(\lambda_1'<\frac{y_{t-1}}{\sqrt{T}}<\lambda_1\right)\right)^{1/2},$$

$k=1,\ldots,p$, by the Cauchy–Schwarz Inequality and Lemma A.2. Then the asymptotic stochastic equicontinuity of $\sum_{t=1}^{T}\Delta X_{1t}^{U}(\lambda)'X_{1t}^{U}(\lambda)$ follows from Lemma 3 and the definition (7.11) of asymptotic stochastic equicontinuity.

For $\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'\Delta X_{jt}^{U}(\lambda)$, it is also sufficient to restrict to $j=1$, and the bound

$$\left|\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-k}\Delta y_{t-k'}\mathbb{I}\left(\lambda_1\leqslant\frac{y_{t-1}}{\sqrt{T}}<\lambda_1'\right)\right| \leqslant \left(\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{t-k}\Delta y_{t-k'})^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}\left(\lambda_1\leqslant\frac{y_{t-1}}{\sqrt{T}}<\lambda_1'\right)\right)^{1/2} = \mathrm{O}_{\mathbb{P}}(1)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}\left(\lambda_1'<\frac{y_{t-1}}{\sqrt{T}}<\lambda_1\right)\right)^{1/2},$$

$1\leqslant k,k'\leqslant p$ similarly gives asymptotic stochastic equicontinuity.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Step 2: $\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'X_{jt}^{U}(\lambda)=o_{\mathbb{P}}(1)$ *and* $T^{-1}\sum_{t=1}^{T}(\Delta y_{t-k}\Delta y_{t-k'}-\mathbb{E}[\Delta y_{t-k}\Delta y_{t-k'}])\mathbb{I}(y_{t-1}/\sqrt{T}\in I_j(\lambda))=o_{\mathbb{P}}(1)$ *for all* λ, $1\leqslant k,k'\leqslant p$, $j=1,2,3$. Since, for $1\leqslant k,k'\leqslant p$,

$$\frac{1}{T}\sum_{t=1}^{T}[\Delta y_{t-1},\ldots,\Delta y_{t-p}]'\left[1,\frac{y_{t-1}}{T}\right]=o_{\mathbb{P}}(1),\quad \frac{1}{T}\sum_{t=1}^{T}(\Delta y_{t-k}\Delta y_{t-k'}-\mathbb{E}[\Delta y_{t-k}\Delta y_{t-k'}])=o_{\mathbb{P}}(1)$$

as seen from Hamilton (1994), it is again sufficient to study $j=1$. We begin with $\sum_{t=1}^{T}\Delta X_{1t}^{U}(\lambda)'X_{1t}^{U}(\lambda)=o_{\mathbb{P}}(1)$. The entries of this matrix write, for $k=1,\ldots,p$,

$$\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-k}f_T(y_{t-1})\mathbb{I}\left(\frac{y_{t-1}}{\sqrt{T}}\leqslant\lambda_1\right)\quad\text{with } f_T(y)=1 \text{ or } f_T(y)=y/\sqrt{T}.$$

Note that $\max_{1\leqslant t\leqslant T}|f_T(y_{t-1})|=O_{\mathbb{P}}(1)$, using the Donsker Theorem if $f_T(y)=y/\sqrt{T}$. Since $\Delta y_{t-k}=\sum_{i=0}^{\infty}\pi_i\varepsilon_{t-k-i}$, we have for any integer I,

$$\left|\frac{1}{T}\sum_{t=1}^{T}\left(\Delta y_{t-k}-\sum_{i=0}^{I}\pi_i\varepsilon_{t-k-i}\right)f_T(y_{t-1})\mathbb{I}\left(\frac{y_{t-1}}{\sqrt{T}}\leqslant\lambda_1\right)\right|\leqslant O_{\mathbb{P}}(1)\sum_{i=I+1}^{+\infty}|\pi_i|\frac{1}{T}\sum_{t=1}^{T}|\varepsilon_{t-k-i}|$$
$$\text{with }\mathbb{E}\left[\sum_{i=I+1}^{+\infty}|\pi_i|\frac{1}{T}\sum_{t=1}^{T}|\varepsilon_{t-k-i}|\right]\leqslant\mathbb{E}|\varepsilon_t|\sum_{i=I+1}^{+\infty}|\pi_i| \tag{7.16}$$

which can be made arbitrarily small taking I large enough. Therefore, it is sufficient to show that $T^{-1}\sum_{t=1}^{T}\varepsilon_{t-k}f_T(y_{t-1})\mathbb{I}(y_{t-1}\leqslant\lambda_1\sqrt{T})=o_{\mathbb{P}}(1)$, for any $k\geqslant 1$. For that purpose, note that

$$\begin{aligned}
&|\mathbb{I}(y_{t-1}\leqslant\lambda_1\sqrt{T})-\mathbb{I}(y_{t-k-1}\leqslant\lambda_1\sqrt{T})|\\
&\quad\leqslant\mathbb{I}(y_{t-1}>\lambda_1\sqrt{T})\mathbb{I}(y_{t-k-1}\leqslant\lambda_1\sqrt{T})+\mathbb{I}(y_{t-1}\leqslant\lambda_1\sqrt{T})\mathbb{I}(y_{t-k-1}>\lambda_1\sqrt{T})\\
&\quad=\mathbb{I}(y_{t-k-1}-y_{t-1}<y_{t-k-1}-\lambda_1\sqrt{T}\leqslant 0)+\mathbb{I}(0<y_{t-k-1}-\lambda_1\sqrt{T}\leqslant y_{t-k-1}-y_{t-1})\\
&\quad\leqslant\mathbb{I}\left(|y_{t-k-1}-\sqrt{T}\lambda_1|\leqslant\max_{1\leqslant t\leqslant T}|\Delta y_{t-1}+\cdots+\Delta y_{t-k}|\right)\\
&\quad\leqslant\mathbb{I}\left(\left|\frac{y_{t-k-1}}{\sqrt{T}}-\lambda_1\right|\leqslant CT^{1/(4+s)-1/2}\right)\quad\text{with an arbitrary large probability}
\end{aligned}\tag{7.17}$$

provided that C is taken large enough, see Lemma A.2. Reasoning on that event gives

$$\begin{aligned}
&\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t-k}(f_T(y_{t-1})\mathbb{I}(y_{t-1}\leqslant\lambda_1\sqrt{T})-f_T(y_{t-k-1})\mathbb{I}(y_{t-k-1}\leqslant\lambda_1\sqrt{T}))\right|\\
&\quad\leqslant\frac{1}{T}\sum_{t=1}^{T}|\varepsilon_{t-k}|\max_{1\leqslant t\leqslant T}|f_T(y_{t-1})-f_T(y_{t-k-1})|\\
&\qquad+\frac{\max_{1\leqslant t\leqslant T}|f_T(y_{t-1})|}{T}\sum_{t=1}^{T}|\varepsilon_{t-k}|\mathbb{I}\left(\left|\frac{y_{t-k-1}}{\sqrt{T}}-\lambda_1\right|\leqslant CT^{1/(4+s)-1/2}\right)\\
&\quad=O_{\mathbb{P}}\left(\frac{T^{1/(4+s)}}{\sqrt{T}}\right)+\frac{O_{\mathbb{P}}(1)}{T}\sum_{t=1}^{T}|\varepsilon_{t-k}|\mathbb{I}\left(\left|\frac{y_{t-k-1}}{\sqrt{T}}-\lambda_1\right|\leqslant CT^{1/(4+s)-1/2}\right),
\end{aligned}\tag{7.18}$$

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

by Lemma A.2 for the first remainder term, which is due to the case $f_T(y) = y/\sqrt{T}$, and since $\max_{1\leqslant t\leqslant T}|f_T(y_{t-1})| = \mathrm{O}_{\mathbb{P}}(1)$. Now Lemma A.1 and arguing as in (A.11) yield

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}|\varepsilon_{t-k}|\mathbb{I}\left(\left|\frac{y_{t-k-1}}{\sqrt{T}}-\lambda_1\right|\leqslant CT^{1/(4+s)-1/2}\right)\right]\\
&\quad=\frac{\mathbb{E}|\varepsilon_1|}{T}\sum_{t=1}^{T}\int_{\lambda_1\sqrt{T}-CT^{1/(4+s)}}^{\lambda_1\sqrt{T}+CT^{1/(4+s)}}\varphi_{t-k-1}(v)\,\mathrm{d}v\leqslant C\left(\frac{kT^{1/(4+s)}}{T}+\frac{T^{1/(4+s)}}{T}\sum_{t=k+2}^{T}\frac{1}{\sqrt{t-k+1}}\right)\\
&\quad=\mathrm{O}(T^{1/(4+s)-1/2})=\mathrm{o}(1)\quad\text{since } s>4. \qquad (7.19)
\end{aligned}$$

Therefore, (7.18) with $s>4$ and (7.19) shows that

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t-k}f_T(y_{t-1})\mathbb{I}\left(\frac{y_{t-1}}{\sqrt{T}}\leqslant\lambda_1\right)=\frac{1}{T}\sum_{t=1}^{T}f_T(y_{t-k-1})\mathbb{I}\left(\frac{y_{t-k-1}}{\sqrt{T}}\leqslant\lambda_1\right)\varepsilon_{t-k}+\mathrm{o}_{\mathbb{P}}(1).$$

Observe now that the leading term is $\mathrm{o}_{\mathbb{P}}(1)$ since this sum is a martingale with a variance bounded by $\sigma^2T^{-1}\sum_{t=1}^{T}\mathbb{E}f_T^2(y_{t-k-1})=\mathrm{o}(1)$. Therefore, $T^{-1}\sum_{t=1}^{T}\varepsilon_{t-k}f_T(y_{t-1})\mathbb{I}(y_{t-1}\leqslant\lambda_1\sqrt{T})=\mathrm{o}_{\mathbb{P}}(1)$ as desired and then $\sum_{t=1}^{T}\Delta X_{jt}^{U}(\lambda)'X_{jt}^{U}(\lambda)=\mathrm{o}_{\mathbb{P}}(1)$, $j=1,2,3$.

We now show that $T^{-1}\sum_{t=1}^{T}(\Delta y_{t-k}\Delta y_{t-k'}-\mathbb{E}[\Delta y_{t-k}\Delta y_{t-k'}])\mathbb{I}(y_{t-1}/\sqrt{T}\in I_1(\lambda))=\mathrm{o}_{\mathbb{P}}(1)$ for all λ, $1\leqslant k,k'\leqslant p$. Let $\varepsilon_{t-k,t-k'}=\varepsilon_{t-k}\varepsilon_{t-k'}-\mathbb{E}[\varepsilon_{t-k}\varepsilon_{t-k'}]$ so that $\Delta y_{t-k}\Delta y_{t-k'}-\mathbb{E}[\Delta y_{t-k}\Delta y_{t-k'}]=\sum_{0\leqslant i,i'\leqslant+\infty}\pi_i\pi_{i'}\varepsilon_{t-k-i,t-k'-i'}$. This gives, as in (7.16), for any integer I,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left(\Delta y_{t-k}\Delta y_{t-k'}-\mathbb{E}[\Delta y_{t-k}\Delta y_{t-k'}]-\sum_{0\leqslant i,i'\leqslant I+1}\pi_i\pi_{i'}\varepsilon_{t-k-i,t-k'-i'}\right)\mathbb{I}\left(\frac{y_{t-1}}{\sqrt{T}}\leqslant\lambda_1\right)\\
&\leqslant\sum_{I+1\leqslant i,i'\leqslant\infty}|\pi_i||\pi_{i'}|\frac{1}{T}\sum_{t=1}^{T}|\varepsilon_{t-k-i,t-k'-i'}|\\
&\quad\text{with }\mathbb{E}\left[\sum_{I+1\leqslant i,i'\leqslant\infty}|\pi_i||\pi_{i'}|\frac{1}{T}\sum_{t=1}^{T}|\varepsilon_{t-k-i,t-k'-i'}|\right]\leqslant2\mathbb{E}\varepsilon_t^2\left(\sum_{i=I+1}^{\infty}|\pi_i|\right)^2\overset{I\to\infty}{\longrightarrow}0,
\end{aligned}$$

so that it is sufficient to show that $T^{-1}\sum_{t=1}^{T}\varepsilon_{t-k,t-k'}\mathbb{I}(y_{t-1}/\sqrt{T}\leqslant\lambda_1)=\mathrm{o}_{\mathbb{P}}(1)$ for all $k,k'\geqslant1$. Assume that $k'=k-n\leqslant k$. Then arguing as in (7.18) and (7.19) gives

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t-k,t-k'}\mathbb{I}\left(\frac{y_{t-1}}{\sqrt{T}}\leqslant\lambda_1\right)=\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t-k,t+n-k}\mathbb{I}\left(\frac{y_{t-k-1}}{\sqrt{T}}\leqslant\lambda_1\right)+\mathrm{o}_{\mathbb{P}}(1),$$

where the leading term is again an $\mathrm{o}_{\mathbb{P}}(1)$ martingale by definition of $\varepsilon_{t-k,t+n-k}$. This ends the proof of Theorem 5. □

Proof of Theorem 6. Assume that $\sigma_y=1$. That $\sum_{t=1}^{T}\Delta X_t^{B\prime}\Delta X_t^{B}=\Omega+\mathrm{o}_{\mathbb{P}}(1)$ has been proven in Theorem 5 since $\Delta X_t^{B\prime}\Delta X_t^{B}=\Delta X_{1t}^{U}(\lambda)'\Delta X_{1t}^{U}(\lambda)+\Delta X_{2t}^{U}(\lambda)'\Delta X_{2t}^{U}(\lambda)+\Delta X_{3t}^{U}(\lambda)'\Delta X_{3t}^{U}(\lambda)$. Note that all the processes of the lemma are in $\ell^{\infty}([-a,a]^2)$. The rest of the proof is divided in three steps.

Step 1: *Asymptotic stochastic equicontinuity of* $\sum_{t=1}^{T}X_{jt}^{B}(\lambda)'\varepsilon_t$, $\sum_{t=1}^{T}X_{jt}^{B}(\lambda)'X_{jt}^{B}(\lambda)$, $\sum_{t=1}^{T}\Delta X_t^{B\prime}X_{jt}^{B}(\lambda)$, $j=1,2,3$. Let

$$X_{jt}(\lambda)=\frac{\mathbb{I}(y_{t-1}\in I_j(\lambda))}{T^{1/4}}[1,y_{t-1}],\quad j=1,3,\quad X_t=\frac{1}{T^{1/4}}[1,y_{t-1}],$$

with $X_{1t}(\lambda)=X_{1t}(\lambda_1)$ and $X_{3t}(\lambda)=X_{3t}(\lambda_2)$. The asymptotic stochastic equicontinuity over λ_1 or λ_2 in $[-a,a]$ of $\sum_{t=1}^{T}X_{jt}'(\lambda)\varepsilon_t$ and $\sum_{t=1}^{T}X_{jt}'(\lambda)X_{jt}'$, $j=1,3$, directly follows from Lemma 3. This gives the asymptotic stochastic equicontinuity of $\sum_{t=1}^{T}X_{jt}^{B}(\lambda)\varepsilon_t$, $\sum_{t=1}^{T}X_{jt}^{B}(\lambda)'X_{jt}^{B}(\lambda)$, $j=1,3$, which uses a higher standardization ($\sqrt{T}$ for 1 and T for y_{t-1} instead of $T^{1/4}$). For $j=2$, note that $X_{2t}^{B}(\lambda)=X_t-X_{1t}(\lambda)-X_{3t}(\lambda)$, $X_{2t}^{B}(\lambda)'X_{2t}^{B}(\lambda)=X_t'X_t-X_{1t}'(\lambda)X_{1t}(\lambda)-X_{3t}'(\lambda)X_{3t}(\lambda)$ so that $\sum_{t=1}^{T}X_{2t}^{B}(\lambda)'\varepsilon_t$, $\sum_{t=1}^{T}X_{2t}^{B}(\lambda)'X_{2t}^{B}(\lambda)$ are also asymptotically equicontinuous.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

The asymptotic stochastic equicontinuity of $\sum_{t=1}^{T}\Delta X_t^{B\prime}X_{jt}^{B}(\lambda)$, $j=1,2,3$ similarly follows from the one of $\sum_{t=1}^{T}\Delta X_t^{B\prime}X_{jt}(\lambda)$, $j=1,3$, that we establish now. Consider $j=1$, $j=3$ being similar. The entries of the increments of $\sum_{t=1}^{T}\Delta X_t^{B\prime}X_{1t}(\lambda)$ between $\lambda_1<\lambda_1'$, are such that, with $f(w)=1$ or $f(w)=w$, $k=1,\ldots,p$

$$\left|\frac{1}{T^{1/2+1/4}}\sum_{t=1}^{T}\Delta y_{t-k}f(y_{t-1})\mathbb{I}(\lambda_1<y_{t-1}\leqslant\lambda_1')\right|\leqslant\left(\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{t-k})^2\right)^{1/2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f^2(y_{t-1})\mathbb{I}(\lambda_1<y_{t-1}\leqslant\lambda_1')\right)^{1/2}$$
$$=\mathrm{O}_{\mathbb{P}}(1)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f^2(y_{t-1})\mathbb{I}(\lambda_1<y_{t-1}\leqslant\lambda_1')\right)^{1/2},$$

by the Cauchy–Schwartz Inequality and Lemma A.2. Then the asymptotic stochastic equicontinuity of $\sum_{t=1}^{T}\Delta X_t^{B\prime}X_{1t}^{B}(\lambda)$ follows from Lemma 3.

Step 2: $\|\sum_{t=1}^{T}\Delta X_t^{B\prime}X_{jt}^{B}(\lambda)\|=\mathrm{o}_{\mathbb{P}}(1)$ *for all* λ, $j=1,2,3$. For $j=2$, the absolute value of the entries of $\sum_{t=1}^{T}\Delta X_t^{B\prime}X_{2t}^{B}(\lambda)$ write as, with $f(w)=1$ or $f(w)=w$ and $k=1,\ldots,p$,

$$\left|\frac{1}{\sqrt{T}T^{1/4}}\sum_{t=1}^{T}\Delta y_{t-k}f(y_{t-1})\mathbb{I}(\lambda_1<y_{t-1}<\lambda_2)\right|\leqslant\frac{\max_{1\leqslant t\leqslant T}|\Delta y_{t-k}|}{T^{1/4}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|f(y_{t-1})|\mathbb{I}(\lambda_1<y_{t-1}<\lambda_2)=\mathrm{O}_{\mathbb{P}}(T^{1/(4+s)-1/4})\mathrm{O}_{\mathbb{P}}(1)=\mathrm{o}_{\mathbb{P}}(1),$$

because $s>0$ and by Lemma A.2, and since $T^{-1/2}\sum_{t=1}^{T}|f(y_{t-1})|\mathbb{I}(\lambda_1<y_{t-1}<\lambda_2)$ converges in distribution to $L_W(0,1)\int_{\lambda_1}^{\lambda_2}|f(w)|\,\mathrm{d}w/\sigma_y$ by Theorem 7, $f(w)\mathbb{I}(\lambda_1<w<\lambda_2)$ being I-regular.

Consider now $j=1$, $j=3$ being similar. The entries of $\sum_{t=1}^{T}\Delta X_t^{B}X_{1t}^{B}(\lambda)$ write, with $f_T(w)=1$ or $f_T(w)=w/\sqrt{T}$,

$$\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-k}f_T(y_{t-1})\mathbb{I}(y_{t-1}\leqslant\lambda_1),\quad k=1,\ldots,p.$$

Up to λ_1 changed into $\sqrt{T}\lambda_1$, this is exactly the items studied in Step 2 of Theorem 5. Changing $\sqrt{T}\lambda_1$ into λ_1 in Eqs. (7.16), (7.17), (7.18) and (7.19) give that these items are $\mathrm{o}_{\mathbb{P}}(1)$.

Step 3: *Convergence of the finite dimensional distributions and conclusion*. Because, for $j=1,2,3$, $\sum_{t=1}^{T}\Delta X_t^{B\prime}X_{jt}^{B}(\lambda)=\mathrm{o}_{\mathbb{P}}(1)$ and from asymptotic stochastic equicontinuity of these processes, we get that $\sup_{\lambda\in[-a,a]^2}\|\sum_{t=1}^{T}\Delta X_t^{B\prime}X_{jt}^{B}(\lambda)\|=\mathrm{o}_{\mathbb{P}}(1)$. Asymptotic stochastic equicontinuity of Step 1 and Theorem 7 yield the convergence in distribution of $\sum_{t=1}^{T}X_{2t}^{B}(\lambda)'X_{2t}^{B}(\lambda)$ and $\sum_{t=1}^{T}X_{2t}^{B}(\lambda)'\varepsilon_t$, $[1,w,w^2]\mathbb{I}(\lambda_1<w<\lambda_2)$ being I-regular. For $j=1$ (the case $j=3$ being similar), note that the entries of $\sum_{t=1}^{T}X_{1t}^{B}(\lambda)'X_{1t}^{B}(\lambda)$ and $\sum_{t=1}^{T}X_{1t}^{B}(\lambda)'\varepsilon_t$ write

$$\frac{1}{T}\sum_{t=1}^{T}f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\mathbb{I}(y_{t-1}\leqslant\lambda_1),\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\mathbb{I}(y_{t-1}\leqslant\lambda_1)\varepsilon_t,$$

with $f(w)=1$, $f(w)=w$, or $f(w)=w^2$. Therefore, (A.3) and (A.6) in Lemma A.3 yield

$$\mathbb{E}^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}f\left(\frac{y_{t-1}}{\sqrt{T}}\right)(\mathbb{I}(y_{t-1}\leqslant\lambda_1)-\mathbb{I}(y_{t-1}\leqslant0))\right]^2$$
$$\leqslant\frac{C}{\sqrt{T}}\left[\left|\int_0^{\lambda_1}|f(w/\sqrt{T})|\,\mathrm{d}w\right|+\left|\frac{\int_0^{\lambda_1}f^2(w/\sqrt{T})\,\mathrm{d}w}{\sqrt{T}}\right|^{1/2}\right]=\mathrm{o}(1),$$

$$\mathbb{E}^{1/4}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f\left(\frac{y_{t-1}}{\sqrt{T}}\right)(\mathbb{I}(y_{t-1}\leqslant\lambda_1)-\mathbb{I}(y_{t-1}\leqslant0))\varepsilon_t\right)^4$$
$$\leqslant\frac{C}{T^{1/4}}\left[\left|\int_0^{\lambda_1}f^2(w/\sqrt{T})\,\mathrm{d}w\right|^{1/2}+\left|\int_0^{\lambda_1}f^4(w/\sqrt{T})\,\mathrm{d}w\right|^{1/4}\right]=\mathrm{o}(1).$$

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

This gives that $\sum_{t=1}^{T} X_{1t}^{B}(\lambda)' X_{1t}^{B}(\lambda) = \sum_{t=1}^{T} X_{1t}^{B}(0)' X_{1t}^{B}(0) + o_{\mathbb{P}}(1)$ and $\sum_{t=1}^{T} X_{1t}^{B}(\lambda)' \varepsilon_t = \sum_{t=1}^{T} X_{1t}^{B}(0)' \varepsilon_t + o_{\mathbb{P}}(1)$. Asymptotic equicontinuity and Theorem 7 yield the convergence in distribution of Theorem 6 for $j = 1, 3$ because $\mathbb{I}(y_{t-1} \leqslant 0) = \mathbb{I}(y_{t-1}/(\sqrt{T} \leqslant 0)$ and because the functions $\mathbb{I}(w \leqslant 0)[1, w, w^2]$ are regular. □

7.5. Proof of Corollaries 1 and 2, Theorems 2 and 3

Proof of Corollary 1. The proof is divided in three steps.

Step 1: *Choice of the covariates and additional notations*. Let $X_{jt}(\lambda) = X_{jt}^{U}(\lambda), \Delta X_{jt}(\lambda) = \Delta X_{jt}^{U}(\lambda)$ be as in (3.6), and $X_t(\lambda) = [X_{1t}(\lambda), X_{2t}(\lambda), X_{3t}(\lambda), \Delta X_{1t}(\lambda), \Delta X_{2t}(\lambda), \Delta X_{3t}(\lambda)]$. Since the μ's and ρ's coefficients in (3.2) vanish under the null, the definition (3.6) of the covariates $X_t(\lambda)$ shows that the TAR specification (3.2) writes as

$$\Delta y_t = X_t(\lambda) r \beta_{0T} + \varepsilon_t \quad \text{with } \beta_{0T} = \sqrt{T}\beta_0, \tag{7.20}$$

where r is from (3.2) and β_0 is given by the coefficients of $a(L)$ in (3.3). For the matrix R such that $R\beta$ gives the ρ coefficient of the TAR specification, define now

$$\widetilde{\mathrm{Wald}}_T(\lambda) = \frac{1}{\widetilde{\sigma}_T^2(\lambda)} (R\widetilde{\beta}_T(\lambda))' \left(R \left(r' \sum_{t=1}^{T} X_t'(\lambda) X_t(\lambda) r \right)^{-1} R' \right)^{-1} (R\widetilde{\beta}_T(\lambda))$$

where

$$\widetilde{\beta}_T(\lambda) = \left(r' \sum_{t=1}^{T} X_t'(\lambda) X_t(\lambda) r \right)^{-1} r' \sum_{t=1}^{T} X_t'(\lambda) \Delta y_t, \quad \widetilde{\sigma}_T^2(\lambda) = \frac{1}{T} \sum_{t=1}^{T} (\Delta y_t - X_t(\lambda) r \widetilde{\beta}_T(\lambda))^2.$$

Observe that the $X_t(\lambda/(\sigma_y\sqrt{T}))$, $t = 1, \ldots, T$ and the $x_t(\lambda)$, $t = 1, \ldots, T$ in (3.2) generate the same linear span by (3.6), so that regressing on the former or the latter gives the same residuals. Therefore, the residuals-based formula (7.4) for the Wald statistic show that $\mathrm{Wald}_T(\lambda) = \widetilde{\mathrm{Wald}}_T(\lambda/(\sigma_y\sqrt{T}))$, and

$$\mathrm{SupWald}_T(\Lambda_T) = \sup_{\lambda \in \widetilde{\Lambda}_T} \widetilde{\mathrm{Wald}}_T(\lambda), \quad \widetilde{\Lambda}_T = \frac{\Lambda_T}{\sigma_y\sqrt{T}} = \left\{ \lambda; \frac{\underline{\lambda}_T}{\sigma_y} \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \frac{\overline{\lambda}_T}{\sigma_y}, \lambda_2 - \lambda_1 \geqslant \frac{v_T}{\sigma_y} \right\}. \tag{7.21}$$

Define also

$$\mathrm{Wald}_U(\lambda) = M_U(\lambda)' r (r' C_U(\lambda) r)^{-1} R' (\sigma^2 R (r' C_U(\lambda) r)^{-1} R')^{-1} R (r' C_U(\lambda) r)^{-1} r' M_U(\lambda).$$

We show that this statistic only depends upon the Brownian Motion $W(\cdot)$ and not on σ_y, so that $\mathrm{Wald}(\cdot)$ is pivotal, showing that the limit variable $\sup_{\lambda \in \Lambda/\sigma_y}$ has a pivotal distribution when the distribution of $[\underline{\lambda}, \overline{\lambda}, v]/\sigma_y$ is pivotal under Assumption Λ. Let $m_0(y_{t-1}, \lambda)$ and $m_1(y_{t-1}, \lambda)$ be the linear functions of the $\mathbb{I}(y_{t-1} \in I_j(\lambda))$ and $y_{t-1}\mathbb{I}(y_{t-1} \in I_j(\lambda))$, $j = 1, 2, 3$, entering in the TAR specification, written in columns. Due to the choice of covariate imposed in the corollary, the Partitioned Inverse Formula yields that

$$\mathrm{Wald}_U(\lambda) = \widetilde{M}_1'(\lambda)(\widetilde{C}^{11}(\lambda))^{-1} \widetilde{M}_1(\lambda)$$

$$\text{with, for } \widetilde{C}_{ij}(\lambda) = \sigma_y^{i+j} \int_0^1 m_i(W(v), \lambda) m_j'(W(v), \lambda)\, \mathrm{d}v,$$

$$\widetilde{M}_1(\lambda) = \widetilde{C}^{10}(\lambda) \int_0^1 m_0(W(v), \lambda)\, \mathrm{d}W(v) + \widetilde{C}^{11}(\lambda)\sigma_y \int_0^1 m_1(W(v), \lambda)\, \mathrm{d}W(v),$$

$$\widetilde{C}^{10}(\lambda) = -[\widetilde{C}_{00}(\lambda) - \widetilde{C}_{01}(\lambda)(\widetilde{C}_{11}(\lambda))^{-1} \widetilde{C}_{10}(\lambda)]^{-1} \widetilde{C}_{10}(\lambda)(\widetilde{C}_{11}(\lambda))^{-1},$$

$$\widetilde{C}^{11}(\lambda) = [\widetilde{C}_{11}(\lambda) - \widetilde{C}_{10}(\lambda)(\widetilde{C}_{00}(\lambda))^{-1} \widetilde{C}_{11}(\lambda)]^{-1}. \tag{7.22}$$

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Since $\widetilde{M}_1(\lambda)$ is proportional to $1/\sigma_y$ and $\widetilde{C}^{11}(\lambda)$ to $1/\sigma_y^2$, σ_y cancels out from the expression of $\mathrm{Wald}_U(\lambda)$. Note also that the inverse matrices are well defined for all λ in Λ/σ_y under (3.9).

Step 2: *A truncation of* Λ_T. Consider $a>0$ and define

$$\widetilde{\Lambda}_T(a)=\{\lambda;\underline{\lambda}_T(a)\leqslant\lambda_1\leqslant\lambda_2\leqslant\overline{\lambda}_T(a),\lambda_2-\lambda_1\geqslant v_T(a)\}\text{ with }\underline{\lambda}_T(a)=\max\left(\frac{\underline{\lambda}_T}{\sigma_y},-a\right),\quad\overline{\lambda}_T(a)=\min\left(\frac{\overline{\lambda}_T}{\sigma_y},a\right),$$

$$v_T(a)=\min\left(\frac{v_T}{\sigma_y},\overline{\lambda}_T(a)-\underline{\lambda}_T(a)\right),$$

$$\widetilde{\Lambda}(a)=\{\lambda;\underline{\lambda}(a)\leqslant\lambda_1\leqslant\lambda_2\leqslant\overline{\lambda}(a),\lambda_2-\lambda_1\geqslant v(a)\}\text{ with }\underline{\lambda}(a)=\max\left(\frac{\underline{\lambda}}{\sigma_y},-a\right),\quad\overline{\lambda}(a)=\min\left(\frac{\overline{\lambda}}{\sigma_y},a\right),$$

$$v(a)=\min\left(\frac{v}{\sigma_y},\overline{\lambda}(a)-\underline{\lambda}(a)\right).$$

Observe that $\widetilde{\Lambda}_T(a)$ and $\widetilde{\Lambda}(a)$ are subsets of $[-a,a]^2$. Because $\underline{\lambda}/\sigma_y\leqslant\underline{\lambda}(a)\leqslant\overline{\lambda}(a)\leqslant\overline{\lambda}/\sigma_y$ and $v(a)/\sigma_y\geqslant\min(v,2a)$, (3.9) yields that

$$\inf_{v\in[0,1]}W(v)\leqslant\underline{\lambda}(a)\leqslant\overline{\lambda}(a)\leqslant\sup_{v\in[0,1]}W(v)\text{ and }v(a)>0\quad\text{a.s.}\tag{7.23}$$

Note that

$$\mathbb{P}\left(\sup_{\lambda\in\Lambda}\mathrm{Wald}_U\left(\frac{\lambda}{\sigma_y}\right)\neq\sup_{\lambda\in\widetilde{\Lambda}(a)}\mathrm{Wald}_U(\lambda)\right)\leqslant\mathbb{P}\left(\frac{\underline{\lambda}}{\sigma_y}\leqslant-a\text{ or }a\leqslant\frac{\overline{\lambda}}{\sigma_y}\right)\overset{a\to\infty}{\longrightarrow}0.$$

Because $[\underline{\lambda}_T,\overline{\lambda}_T,v_T]\overset{\mathrm{d}}{\to}[\underline{\lambda},\overline{\lambda},v]$ by Assumption Λ, we similarly have

$$\sup_{T\geqslant1}\mathbb{P}\left(\sup_{\lambda\in\widetilde{\Lambda}_T}\widetilde{\mathrm{Wald}}_T(\lambda)\neq\sup_{\lambda\in\widetilde{\Lambda}_T(a)}\widetilde{\mathrm{Wald}}_T(\lambda)\right)\leqslant\sup_{T\geqslant1}\mathbb{P}\left(\frac{\underline{\lambda}_T}{\sigma_y}\leqslant-a\text{ or }a\leqslant\frac{\overline{\lambda}_T}{\sigma_y}\right)\overset{a\to\infty}{\longrightarrow}0.$$

Therefore, (7.21) and noting that $\sup_{\lambda\in\widetilde{\Lambda}(a)}\mathrm{Wald}_U(\lambda)$ converges by a.s. continuity to $\sup_{\lambda\in\Lambda/\sigma}\mathrm{Wald}_U(\lambda)$ when $a\to+\infty$ show that the limit result of Corollary 1 holds if, for all a,

$$\sup_{\lambda\in\widetilde{\Lambda}_T(a)}\widetilde{\mathrm{Wald}}_T(\lambda)\overset{\mathrm{d}}{\to}\sup_{\lambda\in\widetilde{\Lambda}(a)}\mathrm{Wald}_U(\lambda).\tag{7.24}$$

Step 3: Proof of (7.24) *and conclusion*. Recall that $(\sum_{t=1}^T X_t'(\lambda)X_t(\lambda))^{-1}$ stands for the pseudo-inverse of $\sum_{t=1}^T X_t'(\lambda)X_t(\lambda)$. We first show that, with a probability tending to 1, $\sum_{t=1}^T X_t'(\lambda)X_t(\lambda)$ has an inverse for all λ in $\widetilde{\Lambda}_T(a)$, i.e. that $\lim_{T\to\infty}\mathbb{P}(\inf_{\lambda\in\widetilde{\Lambda}_T(a)}\mathrm{Det}(\sum_{t=1}^T X_t'(\lambda)X_t(\lambda))>0)=1$. Orthogonality of the regimes, Theorem 5 and Assumption Λ yield that

$$\left[\underline{\lambda}_T(a),\overline{\lambda}_T(a),v_T(a),\left\{\sum_{t=1}^T X_t'(\lambda)X_t(\lambda)\right\}_{\lambda\in[-a,a]^2}\right]\overset{\mathrm{d}}{\to}[\underline{\lambda}(a),\overline{\lambda}(a),v(a),\{C_U(\lambda)\}_{\lambda\in[-a,a]^2}]$$

in $\mathbb{R}^3\times\ell^\infty([-a,a]^2)$, where $C_U(\lambda)$ is as in (3.8). Note that $\mathrm{Det}(\Omega)>0$. It then follows by (7.23) and continuity of $\{W(v)\}_{v\in[0,1]}$ which, moreover, cannot be constant over any nonempty $I_j(\lambda)$, that $\mathrm{Det}(C_U(\lambda))>0$ for all $\lambda\in\widetilde{\Lambda}(a)$ a.s. Moreover, $\lambda\mapsto\mathrm{Det}(C_U(\lambda))$ is a.s. continuous over $\widetilde{\Lambda}(a)$ which is a.s. compact. Therefore, $\inf_{\lambda\in\widetilde{\Lambda}(a)}\mathrm{Det}(C_U(\lambda))>0$ a.s. In addition, because $\widetilde{\Lambda}_T(a)\subset[-a,a]^2$ and $\widetilde{\Lambda}_T(a)\subset T[-a,a]^2$, the map from $\mathbb{R}^3\times\ell^\infty([-a,a]^2)$ to $\mathbb{R}$

$$\left[\underline{\lambda}_T(a),\overline{\lambda}_T(a),v_T(a),\left\{\sum_{t=1}^T X_t'(\lambda)X_t(\lambda)\right\}_{\lambda\in T(-a,a)}\right]\mapsto\inf_{\lambda\in\widetilde{\Lambda}_T(a)}\mathrm{Det}\left(\sum_{t=1}^T X_t'(\lambda)X_t(\lambda)\right),$$

is continuous at $[\underline{\lambda}(a), \overline{\lambda}(a), v(a), \{C_U(\lambda)\}_{\lambda\in[-a,a]^2}]$ by continuity of Det(·) and $C_U(\cdot)$. Therefore, the Functional Continuous Mapping Theorem (see e.g. van der Vaart and Wellner, 1996, Theorem 1.3.6) yields that

$$\lim_{T\to\infty} \mathbb{P}\left(\inf_{\lambda\in\widetilde{\Lambda}_T(a)} \mathrm{Det}\left(\sum_{t=1}^{T} X_t'(\lambda)X_t(\lambda)\right) > 0\right) = \mathbb{P}\left(\inf_{\lambda\in\widetilde{\Lambda}(a)} \mathrm{Det}(C_U(\lambda)) > 0\right) = 1.$$

Hence $\sum_{t=1}^{T} X_t'(\lambda)X_t(\lambda)$ has an inverse for all λ in $\widetilde{\Lambda}_T(a)$ with a probability tending to 1. Observe also that

$$\begin{aligned}\widetilde{\sigma}_T^2(\lambda) &= \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 - \frac{1}{T}\left(\sum_{t=1}^{T} X_t(\lambda)r\varepsilon_t\right)\left(r'\sum_{t=1}^{T} X_t'(\lambda)X_t(\lambda)r\right)^{-1}\left(r'\sum_{t=1}^{T} X_t'(\lambda)\varepsilon_t\right) + \varepsilon_T'(\lambda)\\ &= \sigma^2 + \varepsilon_T''(\lambda) \quad \text{with} \sup_{\lambda\in\widehat{\Lambda}_T(a)} (|\varepsilon_T(\lambda)| + |\varepsilon_T'{}'(\lambda)|) = o_{\mathbb{P}}(1),\end{aligned}$$

where the last approximation of $\widetilde{\sigma}_T^2(\lambda)$ comes from Theorem 5 and the Law of Large Numbers. Then the Functional Mapping Theorem yields that (7.24) is proved. □

Proof of Corollary 2. The proof follows the same steps than for Corollary 1, up to the choice of $X_t(\lambda) = [X_{1t}^B(\lambda), X_{2t}^B(\lambda), X_{3t}^B(\lambda), \Delta X_t^B]$ from (3.12) in (7.20) and the use of Theorem 6 in place of Theorem 5. It remains to show that the null limit distribution is pivotal. Due to the restriction on the choice of covariate imposed in the corollary, this limit distribution decomposes as a sum of two independent terms. A first term comes from the outer regimes and, as seen from (3.13) and (3.14), involves $C_1(0)$, $C_3(0)$, $M_1(0)$ and $M_3(0)$ in a quadratic form similar to the variable $\mathrm{Wald}_U(0)$ from (7.22). The second is due to the inner regime and writes as $\sup_{\lambda\in\Lambda}\zeta_{2B}^2(\lambda)$, where the expression of $\zeta_{2B}(\cdot)$ depends upon the central regime variables retained in $\widetilde{x}_t(\lambda)$. If none of these variables appear then $\zeta_{2B}(\lambda) = 0$, while if $y_{t-1}\mathbb{I}(y_{t-1}\in I_2(\lambda))$ only or if $\mathbb{I}(y_{t-1}\in I_2(\lambda))$ and $y_{t-1}\mathbb{I}(y_{t-1}\in I_2(\lambda))$ are retained, $\zeta_{2B}(\lambda)$, respectively, writes as

$$\frac{\int_{\lambda_1}^{\lambda_2} w\,\mathrm{d}B(w)}{\sqrt{(\lambda_2^3-\lambda_1^3)/3}}, \quad \frac{(\lambda_2-\lambda_1)\int_{\lambda_1}^{\lambda_2} w\,\mathrm{d}B(w) - (\lambda_2^2-\lambda_1^2)\int_{\lambda_1}^{\lambda_2}\mathrm{d}B(w)/2}{\sqrt{(\lambda_2-\lambda_1)^2(\lambda_2^3-\lambda_1^3)/3 - (\lambda_2-\lambda_1)(\lambda_2^2-\lambda_1^2)^2/4}}.$$

Since $B(v\cdot) \overset{\mathrm{d}}{=} \sqrt{v}B(\cdot)$, elementary algebra yields that $\zeta_{2B}(v\cdot) = \zeta_{2B}(\cdot)$. Hence $\sup_{\lambda\in\Lambda/\sigma_y}\zeta_{2B}^2(\lambda) \overset{\mathrm{d}}{=} \sup_{\lambda\in\Lambda/v}\zeta_{2B}^2(\lambda)$ which has a pivotal distribution if the distribution of $[\underline{\lambda}/v, \overline{\lambda}/v]$ is pivotal. □

Proof of Theorems 2 and 3. Elementary manipulations based on Theorem 7 shows that the threshold sets used in Theorems 2 and 3 satisfy Assumption Λ and the conditions for achieving a pivotal limit distribution of Corollaries 1 and 2. Hence Theorem 2 directly follows from Corollary 1 and elementary algebra. Theorem 3 similarly follows from Corollary 2. □

Acknowledgment

The third author was with LSTA (Université Paris 6) when completing the first version of this paper. 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, EMM first annual conference and Rencontre d'économétrie et statistique Lille3 Littoral 2003. We have benefited from discussions with D.W.K. Andrews, M. Carrasco, C. Francq, R.M de Jong, G. Laroque, A. Rahbek and J.M. Zakoian. We have also appreciated the comments of the referees which have been very helpful to improve the presentation of the paper.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Appendix A. Proof of Lemma 2

We first introduce some additional notations and preliminary results. Assumption L yields that $y_t = \pi_0\varepsilon_t + (\pi_0+\pi_1)\varepsilon_{t-1} + \cdots + (\pi_0+\cdots+\pi_{t-1})\varepsilon_1 + \sum_{i=0}^{\infty}(\pi_{i+1}+\cdots+\pi_{i+t})\varepsilon_{-i}$. Define

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

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

Let $\varphi_t(\cdot)$ and $\varphi_{t,i}(\cdot)$ be the densities of y_t and $\widetilde{y}_{t,i}$. The bound (3.7) in Akonom (1993) writes as

Lemma A.1 (*Akonom, 1993*). *Under Assumptions L and E(s) with* $s>-2$, *there exists a constant* $C>0$ *such that for all t, i with* $1\leqslant t<i$, $\sup_{y\in\mathbb{R}}\varphi_t(w)\leqslant C/\sqrt{t+1}$ *and* $\sup_{w\in\mathbb{R}}\varphi_{t,i}(w)\leqslant C/\sqrt{i-t}$.

Lemma A.2. *Under Assumptions L and E(s) with* $s>-3$, $\sup_{-p\leqslant t}\mathbb{E}|\Delta y_{t-1}|^{4+s}<\infty$, $\max_{-p\leqslant t\leqslant T}|\Delta y_{t-1}| = \mathrm{O}_{\mathbb{P}}(T^{1/(4+s)})$ *and* $\max_{1\leqslant t\leqslant T}|\varepsilon_t| = \mathrm{O}_{\mathbb{P}}(T^{1/(4+s)})$.

Proof of Lemma A.2. Recall that $\Delta y_t = 0$ for $t\leqslant 0$. The Minkowski Inequality yields, for all t, $\mathbb{E}^{1/(4+s)}|\Delta y_{t-1}|^{4+s}\leqslant\sum_{i=1}^{\infty}|\pi_i|\mathbb{E}^{1/(4+s)}|\varepsilon_t|^{4+s}\leqslant C<\infty$.We have by the Markov Inequality

$$\begin{aligned}\mathbb{P}\left(\max_{-p\leqslant t\leqslant T}|\Delta y_{t-1}|\geqslant MT^{1/(4+s)}\right) &\leqslant \frac{1}{MT}\mathbb{E}\max_{-p\leqslant t\leqslant T}|\Delta y_{t-1}|^{(4+s)}\\ &\leqslant \frac{1}{MT}\sum_{t=-p}^{T}\mathbb{E}|\Delta y_{t-1}|^{4+s}\leqslant\frac{C}{M}\end{aligned}$$

which can be made small by taking M large enough. The order of $\max_{1\leqslant t\leqslant T}|\varepsilon_t|$ is similarly obtained. □

Lemma A.3. *Under Assumptions L and E(0), there exists a constant* $C>0$ *such that, for any measurable map* $f(\cdot)$ *from* $\mathbb{R}$ *to* $\mathbb{R}$,

$$\mathbb{E}^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\right]^2\leqslant C\left[\int|f(w)|\,\mathrm{d}w + 2\left(\frac{\int f^2(w)\,\mathrm{d}w}{T}\right)^{1/2}\right], \tag{A.1}$$

$$\begin{aligned}&\mathbb{E}^{1/3}\left|\frac{1}{T}\sum_{t=1}^{T}f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\right|^3\\ &\quad\leqslant C\left[\left(\int|f(w)|\,\mathrm{d}w\right)^3+\frac{\int|f(w)|\,\mathrm{d}w\int f^2(w)\,\mathrm{d}w}{T}+\frac{\int|f(w)|^3\,\mathrm{d}w}{T^2}\right]^{1/3},\end{aligned} \tag{A.2}$$

$$\mathbb{E}^{1/2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(y_{t-1})\right)^2\leqslant C\left[\int|f(w)|\,\mathrm{d}w + 2\left(\frac{\int f^2(w)\,\mathrm{d}w}{\sqrt{T}}\right)^{1/2}\right], \tag{A.3}$$

$$\begin{aligned}&\mathbb{E}^{1/3}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(y_{t-1})\right|^3\\ &\quad\leqslant C\left[\left(\int|f(w)|\,\mathrm{d}w\right)^3+\frac{\int|f(w)|\,\mathrm{d}w\int f^2(w)\,\mathrm{d}w}{\sqrt{T}}+\frac{\int|f(w)|^3\,\mathrm{d}w}{T}\right]^{1/3},\end{aligned} \tag{A.4}$$

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

$$\mathbb{E}^{1/4}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\varepsilon_t\right]^4 \leqslant C\left[\int f^2(w)\,\mathrm{d}w + 2\left(\frac{\int f^4(w)\,\mathrm{d}w}{T}\right)^{1/2}\right]^{1/2}, \tag{A.5}$$

$$\mathbb{E}^{1/4}\left[\frac{1}{T^{1/4}}\sum_{t=1}^{T} f(y_{t-1})\varepsilon_t\right]^4 \leqslant C\left[\int f^2(w)\,\mathrm{d}w + 2\left(\frac{\int f^4(w)\,\mathrm{d}y}{\sqrt{T}}\right)^{1/2}\right]^{1/2}. \tag{A.6}$$

If $E(s)$ holds with $s>0$, then for any integer number $k\geqslant 1$,

$$\begin{aligned}&\mathbb{E}^{1/3}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left|\Delta y_{t-k} f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\varepsilon_t\right|\right]^3\\ &\quad\leqslant C\sqrt{T}\left[\left(\int |f(w)|^{(s+4)/(s+1)}\,\mathrm{d}w\right)^3 + \frac{\int |f(w)|^{(s+4)/(s+1)}\,\mathrm{d}w \int |f(w)|^{2(s+4)/(s+1)}\,\mathrm{d}w}{T}\right.\\ &\qquad\left. + \frac{\int |f(w)|^{3(s+4)/(s+1)}\,\mathrm{d}w}{T^2}\right]^{(1/3)(s+1)/(s+4)},\end{aligned} \tag{A.7}$$

$$\begin{aligned}&\mathbb{E}^{1/4}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta y_{t-k} f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\varepsilon_t\right]^4\\ &\quad\leqslant C\left[\int |f(w)|^{2(s+4)/s}\,\mathrm{d}w + 2\left(\frac{\int |f(w)|^{4(s+4)/s}\,\mathrm{d}w}{T}\right)^{1/2}\right]^{(1/2)(s/(s+4))}.\end{aligned} \tag{A.8}$$

Proof of Lemma A.3. We begin with (A.1) and (A.3). Let $f_T(w)$ be, respectively, $T^{-1/2}f(T^{-1/2}w)$ or $f(w)$, so that we have to bound $\mathbb{E}^{1/2}(T^{-1/2}\sum_{t=1}^{T} f_T(y_{t-1}))^2$. We have

$$\mathbb{E}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_T(y_{t-1})\right)^2 = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}f_T^2(y_{t-1}) \tag{A.9}$$

$$+\frac{2}{T}\sum_{t=1}^{T-1}\mathbb{E}\left[f_T(y_{t-1})\sum_{i=t+1}^{T} f_T(y_{i-1})\right]. \tag{A.10}$$

Lemma A.1 gives for any measurable $g(\cdot)$ from $\mathbb{R}$ to $\mathbb{R}$,

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}|g(y_{t-1})|\\ &\quad=\frac{1}{T}\sum_{t=1}^{T}\int |g(w)|\varphi_{t-1}(w)\,\mathrm{d}y \leqslant \frac{C}{T}\int |g(w)|\,\mathrm{d}w\sum_{t=1}^{T}\frac{1}{\sqrt{t}}\\ &\quad\leqslant \frac{C}{\sqrt{T}}\int |g(w)|\,\mathrm{d}w \quad \text{since } \sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leqslant \sum_{t=1}^{T}\int_{t-1}^{t}\frac{\mathrm{d}v}{\sqrt{v}} = \int_0^T \frac{\mathrm{d}v}{\sqrt{v}} = \frac{\sqrt{T}}{2}.\end{aligned} \tag{A.11}$$

Therefore, (A.9) is bounded by $C\int f_T^2(w)\,\mathrm{d}w/\sqrt{T}$. For (A.10), 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 $F_{t-1} = \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$, $i \geqslant t+1$. Observe also that y_{t-1}, $\overline{y}_{t-1,i-1}$,

$i \geqslant t+1$, are in F_{t-1}. Therefore, Lemma A.1 yields

$$\begin{aligned}\sum_{i=t+1}^{T} \mathbb{E}[|g(y_{i-1})||F_{t-1}] &= \sum_{i=t+1}^{T} \int |g(\widetilde{y}+\overline{y}_{t-1,i-1})|\varphi_{t-1,i-1}(\widetilde{y})\,\mathrm{d}\widetilde{y} \\ &\leqslant \sum_{i=t+1}^{T} \frac{C}{\sqrt{i-t}} \int |g(\widetilde{y}+\overline{y}_{t-1,i-1})|\,\mathrm{d}\widetilde{y} \\ &= C\int |g(w)|\,\mathrm{d}w \sum_{i=t+1}^{T} \frac{1}{\sqrt{i-t}} \leqslant C\sqrt{T}\int g(w)\,\mathrm{d}w \qquad \text{(A.12)} \\ &\quad \text{since } \sum_{i=t+1}^{T} \frac{1}{\sqrt{i-t}} \leqslant \int_{t}^{T} \frac{\mathrm{d}x}{\sqrt{x-t}} = \frac{\sqrt{T-t}}{2}.\end{aligned}$$

Applying (A.11) and (A.12) gives for (A.10)

$$\begin{aligned}&\frac{1}{T}\left|\sum_{t=1}^{T-1} \mathbb{E}\left[f_T(y_{t-1}) \sum_{i=t+1}^{T} f_T(y_{i-1})\right]\right| \\ &\quad = \frac{1}{T}\sum_{t=1}^{T-1} \mathbb{E}\left[|f_T(y_{t-1})| \sum_{i=t+1}^{T} \mathbb{E}\big[|f_T(y_{i-1})||F_t\big]\right] \\ &\quad \leqslant C\int |f_T(w)|\,\mathrm{d}w \frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \mathbb{E}[|f_T(y_{t-1})|] \leqslant C\left(\int |f_T(w)|\,\mathrm{d}w\right)^2.\end{aligned}$$

It then follows that

$$\begin{aligned}\mathbb{E}^{1/2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_T(y_{t-1})\right)^2 &\leqslant \left[\frac{C}{\sqrt{T}}\int f_T^2(w)\,\mathrm{d}w + 2C\left(\int |f_T(w)|\,\mathrm{d}w\right)^2\right]^{1/2} \\ &\leqslant (2C)^{1/2}\left[\left(\int |f_T(w)|\,\mathrm{d}w\right)^2 + \frac{4}{\sqrt{T}}\int f_T^2(w)\,\mathrm{d}w\right]^{1/2} \\ &\leqslant C\left[\int |f_T(w)|\,\mathrm{d}w + 2\left(\frac{\int f_T^2(w)\,\mathrm{d}w}{\sqrt{T}}\right)^{1/2}\right]\end{aligned}$$

since $(a+b)^{1/2} \leqslant a^{1/2}+b^{1/2}$ for nonnegative a, b. Taking $f_T(\cdot) = f(\cdot)$ yields that (A.3) is proven. For (A.1) take $f_T(w) = T^{-1/2} f(T^{-1/2}w)$ and note that $T^{-1/2}\int |f(T^{-1/2}w)|\,\mathrm{d}w = \int |f(w)|\,\mathrm{d}w$, $T^{-1}\int f^2(T^{-1/2}w)\,\mathrm{d}w = T^{-1/2}\int f^2(w)\,\mathrm{d}w$.

For (A.2) and (A.4), observe that

$$\begin{aligned}\mathbb{E}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_T(y_{t-1})\right|^3 &\leqslant \frac{1}{T^{3/2}}\sum_{t=1}^{T} \mathbb{E}|f_T(y_{t-1})|^3 + \frac{3}{T^{3/2}}\sum_{t_1=1}^{T-1} \mathbb{E}\left[f_T^2(y_{t_1-1}) \sum_{t_2=t_1+1}^{T} |f_T(y_{t_2-1})|\right] \\ &\quad + \frac{6}{T^{3/2}}\sum_{t_1=1}^{T-2} \mathbb{E}\left[|f_T(y_{t_1-1})| \sum_{t_2=t_1+1}^{T-1} |f_T(y_{t_2-1})| \sum_{t_3=t_2+1}^{T} |f_T(y_{t_3-1})|\right] \\ &\leqslant C\left[\frac{\int |f_T(w)|^3\,\mathrm{d}w}{T} + \frac{\int f_T^2(w)\,\mathrm{d}w \int |f_T(w)|\,\mathrm{d}w}{\sqrt{T}} + \left(\int |f_T(w)|\,\mathrm{d}w\right)^3\right],\end{aligned}$$

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

by (A.11) and (A.12), applied twice to bound the last sum. Computing each integrals for the corresponding choice of $f_T(\cdot)$ yields (A.2) and (A.4).

We now turn to (A.5) and (A.6). Let $f_T(w)$ be, respectively, $T^{-1/4}f(T^{-1/2}w)$ or $f(w)$, so that we have to bound $\mathbb{E}^{1/4}(T^{-1/4}\sum_{t=1}^{T}f_T(y_{t-1})\varepsilon_t)^4$. Note that $\sum_{t=1}^{T}f_T(y_{t-1})\varepsilon_t$ is a martingale with respect to F_t. The Burkholder Inequality (see e.g. Chow and Teicher, 1988, Theorem 1, p. 396) yields

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{T^{1/4}}\sum_{t=1}^{T}f_T(y_{t-1})\varepsilon_t\right)^4\\
&\quad\leqslant C\mathbb{E}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_T^2(y_{t-1})\varepsilon_t^2\right)^2\\
&\quad=\frac{C}{T}\left(\sum_{t=1}^{T}\mathbb{E}[f_T^4(y_{t-1})\varepsilon_t^4]+2\sum_{t=1}^{T-1}\mathbb{E}[f_T^2(y_{t-1})\varepsilon_t^2f_T^2(y_t)\varepsilon_{t+1}^2]\right) \qquad \text{(A.13)}\\
&\qquad+\frac{2C}{T}\sum_{t=1}^{T-2}\mathbb{E}\left[f_T^2(y_{t-1})\varepsilon_t^2\sum_{i=t+2}^{T}f_T^2(y_{i-1})\varepsilon_i^2\right]. \qquad \text{(A.14)}
\end{aligned}$$

We first deal with the second item of (A.13). Applying the Cauchy–Schwarz Inequality twice yields

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T-1}\mathbb{E}[f_T^2(y_{t-1})\varepsilon_t^2f_T^2(y_t)\varepsilon_{t+1}^2]&\leqslant\frac{1}{T}\sum_{t=1}^{T-1}\mathbb{E}^{1/2}[f_T^4(y_{t-1})\varepsilon_t^4]\mathbb{E}^{1/2}[f_T^4(y_t)\varepsilon_{t+1}^4]\\
&\leqslant\left(\frac{1}{T}\sum_{t=1}^{T-1}\mathbb{E}[f_T^4(y_{t-1})\varepsilon_t^4]\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-1}\mathbb{E}[f_T^4(y_t)\varepsilon_{t+1}^4]\right)^{1/2}\\
&\leqslant\frac{\mathbb{E}\varepsilon_t^4}{T}\sum_{t=1}^{T+1}\mathbb{E}[f_T^4(y_{t-1})],
\end{aligned}$$

since ε_t and y_{t-1} are independent. Therefore, (A.11) yields for (A.13)

$$0\leqslant\frac{1}{T}\left(\sum_{t=1}^{T}\mathbb{E}[f_T^4(y_{t-1})\varepsilon_t^4]+2\sum_{t=1}^{T-1}\mathbb{E}[f_T^2(y_{t-1})\varepsilon_t^2f_T^2(y_t)\varepsilon_{t+1}^2]\right)\leqslant\frac{C}{\sqrt{T}}\int f_T^4(w)\,\mathrm{d}w.$$

For (A.14), (A.12) and (A.11) yield

$$\begin{aligned}
0&\leqslant\frac{1}{T}\sum_{t=1}^{T-2}\mathbb{E}\left[f_T^2(y_{t-1})\varepsilon_t^2\sum_{i=t+2}^{T}f_T^2(y_{i-1})\varepsilon_i^2\right]\\
&=\frac{\mathbb{E}\varepsilon_t^2}{T}\sum_{t=1}^{T-2}\mathbb{E}\left[f_T^2(y_{t-1})\varepsilon_t^2\sum_{i=t+2}^{T}\mathbb{E}[f_T^2(y_{i-1})|F_{t+1}]\right]\\
&\leqslant\int f_T^2(y)\,\mathrm{d}y\frac{C\mathbb{E}^2\varepsilon_t^2}{\sqrt{T}}\sum_{t=1}^{T-2}\mathbb{E}[f_T^2(y_{t-1})]\\
&\leqslant C\left(\int f_T^2(w)\,\mathrm{d}w\right)^2.
\end{aligned}$$

Combining the bounds above with (A.13) and (A.14) gives

$$\begin{aligned}
\mathbb{E}^{1/4}\left(\frac{1}{T^{1/4}}\sum_{t=1}^{T}f_T(y_{t-1})\varepsilon_t\right)^4&\leqslant C\left[\left(\int f_T^2(w)\,\mathrm{d}w\right)^2+\frac{1}{\sqrt{T}}\int f_T^4(w)\,\mathrm{d}w\right]^{1/4}\\
&\leqslant C\left[\int f_T^2(w)\,\mathrm{d}w+2\left(\frac{\int f_T^4(w)\,\mathrm{d}w}{\sqrt{T}}\right)^{1/2}\right]^{1/2}.
\end{aligned}$$

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Taking $f_T(\cdot)=f(\cdot)$ yields (A.6). For (A.5), take $f_T(w)=T^{-1/4}f(T^{-1/2}w)$ and note that $\int T^{-1/2}f^2(T^{-1/2}w)\,\mathrm{d}w=\int f^2(w)\,\mathrm{d}w$, $\int T^{-1}f^4(T^{-1/2}w)\,\mathrm{d}w=T^{-1/2}\int f^4(w)\,\mathrm{d}w$.

For (A.7), let $q_1=(s+4)/3$ and $q_2=(s+4)/(s+1)$ so that $1/q_1+1/q_2=1$. The Hölder and Minkowski Inequalities together with Lemma A.2 and (A.2) yield

$$\begin{aligned}
&\mathbb{E}^{1/3}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|\Delta y_{t-k}\varepsilon_t|\left|f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\right|\right]^3\\
&\leqslant\sqrt{T}\mathbb{E}^{1/3}\left[\left(\frac{1}{T}\sum_{t=1}^{T}|\Delta y_{t-k}\varepsilon_t|^{q_1}\right)^{3/q_1}\left(\frac{1}{T}\sum_{t=1}^{T}\left|f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\right|^{q_2}\right)^{3/q_2}\right]\\
&\leqslant\sqrt{T}\mathbb{E}^{1/3q_1}\left[\left(\frac{1}{T}\sum_{t=1}^{T}|\Delta y_{t-k}\varepsilon_t|^{q_1}\right)^{3}\right]\times\mathbb{E}^{1/3q_2}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\left|f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\right|^{q_2}\right)^{3}\right]\\
&\leqslant C\sqrt{T}[\mathbb{E}^{1/3}|\varepsilon_t|^{3q_1}\mathbb{E}^{1/3}|\Delta y_{t-k}|^{3q_1}]^{1/q_1}\\
&\quad\times\left[\left(\int|f(w)|^{q_2}\,\mathrm{d}w\right)^3+\frac{\int|f(w)|^{q_2}\,\mathrm{d}w\int|f(w)|^{2q_2}\,\mathrm{d}w}{T}+\frac{\int|f(w)|^{3q_2}\,\mathrm{d}w}{T^2}\right]^{1/3q_2}.
\end{aligned}$$

For (A.8), let $q_3=(s+4)/4$, $q_4=(s+4)/s$ so that $1/q_3+1/q_4=1$. The Burkholder and Hölder Inequalities, Lemma A.2 and (A.1) yield

$$\begin{aligned}
&\mathbb{E}^{1/4}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta y_{t-k}f\left(\frac{y_{t-1}}{\sqrt{T}}\right)\varepsilon_t\right]^4\leqslant\mathbb{E}^{1/4}\left[\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{t-k}\varepsilon_t)^2f^2\left(\frac{y_{t-1}}{\sqrt{T}}\right)\right]^2\\
&\leqslant\mathbb{E}^{1/4}\left[\left(\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{t-k}\varepsilon_t)^{2q_3}\right)^{2/q_3}\left(\frac{1}{T}\sum_{t=1}^{T}f^{2q_4}\left(\frac{y_{t-1}}{\sqrt{T}}\right)\right)^{2/q_4}\right]\\
&\leqslant\mathbb{E}^{1/4q_3}\left(\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{t-k}\varepsilon_t)^{2q_3}\right)^2\times\left[\mathbb{E}^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}f^{2q_4}\left(\frac{y_{t-1}}{\sqrt{T}}\right)\right)^2\right]^{1/2q_4}\\
&\leqslant C[\mathbb{E}^{1/2}|\varepsilon_t|^{4q_3}\mathbb{E}^{1/2}|\Delta y_{t-k}|^{4q_3}]^{1/2q_3}\\
&\quad\times\left[\int|f(w)|^{2q_4}\,\mathrm{d}w+2\left(\frac{\int|f(w)|^{4q_4}\,\mathrm{d}w}{T}\right)^{1/2}\right]^{1/2q_4}.\qquad\square
\end{aligned}$$

Proof of Lemma 2. Let us first recall a Maximal Inequality from Empirical Processes Theory, see van der Vaart and Wellner (1996). Let $\mathscr{G}$ be a subset of $\mathscr{E}$ and consider a distance d over $\mathscr{G}$. Let $\{Z(g)\}_{g\in\mathscr{G}}$ be a stochastic process such that, for some $q\geqslant1$ and $C>0$,

$$\mathbb{E}^{1/q}|Z(g)-Z(g')|^q\leqslant Cd(g,g').$$

Let $N(\varepsilon,\mathscr{G},d)$ be the covering numbers of $\mathscr{G}$, that is the minimal number of d-balls with radius ε needed to cover $\mathscr{G}$. Then Theorem 2.2.4 in van der Vaart and Wellner (1996) gives, for any $\eta,\delta>0$,

$$\mathbb{E}^{1/q}\sup_{(g,g')\in\mathscr{G}:d(g,g')\leqslant\delta}|Z(g)-Z(g')|^q\leqslant K\left[\int_0^{\eta}N^{1/q}(\varepsilon/2,\mathscr{G},d)\,\mathrm{d}\varepsilon+\delta N^{2/q}(\eta/2,\mathscr{G},d)\right],\tag{A.15}$$

where K depends on q and C only.

The proof will be divided in two parts devoted to the two kinds of sums. We first consider the case where the ε_t's do not appear explicitly in the sum (S_T-type sums hereafter) and the case where it does, yielding martingale (M_T-type sums hereafter). Let us now introduce some preliminary notations. Consider $h=h_T>0$ and a minimal covering of $\mathscr{F}$ with h-brackets $[f_i,f_{i+1}]$, $i=1,\ldots,N_{[]}(h,\mathscr{F},\|\cdot\|_1)=N_h$. Choose a $\widetilde{f}_i(\cdot)$ in

$[f_i,f_{i+1}]\cap\mathscr{F}$, and for $f(\cdot)\in[f_i,f_{i+1}]\cap\mathscr{F}$, let $\vartheta_h(f)=\widetilde{f}_i$. Let

$$\mathscr{F}_h=\vartheta_h(\mathscr{F})=\{\widetilde{f}_i,i=1,\ldots,N_h\}\subset\mathscr{F}.$$

We use $h=1/T$ for Donsker asymptotics and $h=1/\sqrt{T}$ for local time asymptotics (in short $h=1/r_T$ where r_T is defined below).

We begin with the sums $S_T(f)$ which write as

$$\frac{1}{r_T}\sum_{t=1}^{T}f_T(y_{t-1})\quad\text{with }(r_T,f_T(w))=\left(T,f\left(\frac{w}{\sqrt{T}}\right)\right)\quad\text{or }(r_T,f_T(w))=(\sqrt{T},f(w)).$$

The condition (7.9) of Theorem 8 and inequalities (A.3) and (A.6) of Lemma A.3 imply that, since $S_T(f)-S_T(f')=S_T(f-f')$

$$\mathbb{E}^{1/2}(S_T(f)-S_T(f'))^2\leqslant C\left[\|f-f'\|_1+2\left(\frac{\|f-f'\|_1}{r_T}\right)^{1/2}\right].\tag{A.16}$$

We now show the asymptotic stochastic $\|\cdot\|_1$-equicontinuity of $\{S_T(\vartheta_h(f))\}_{f\in\mathscr{F}}=\{S_T(f)\}_{f\in\mathscr{F}_h}$. Let $\|\cdot\|_{1T}$ be defined from (A.16), i.e.

$$\|f\|_{1T}=\|f\|_1+2\left(\frac{\|f\|_1}{r_T}\right)^{1/2}=\left(\|f\|_1^{1/2}+\frac{1}{\sqrt{r_T}}\right)^2-\frac{1}{r_T}=\frac{((r_T\|f\|_1)^{1/2}+1)^2-1}{r_T}.$$

Because $\|f-f'\|_1^{1/2}\leqslant(\|f-f''\|_1+\|f''-f'\|_1)^{1/2}\leqslant\|f-f''\|_1^{1/2}+\|f''-f'\|^{1/2}$, $\|\cdot\|_{1T}$ satisfies the Triangular Inequality and defines a distance. From the definition of $\|\cdot\|_{1T}$, we have

$$\|f\|_{1T}\leqslant\varepsilon\quad\text{if and only if }\|f\|_1\leqslant\frac{((r_T\varepsilon+1)^{1/2}-1)^2}{r_T}.\tag{A.17}$$

We now bound the covering number $N(\varepsilon,\mathscr{F}_h,\|\cdot\|_{1T})$. (A.17) yields

$$N(\varepsilon,\mathscr{F}_h,\|\cdot\|_{1T})=N\left(\frac{((r_T\varepsilon+1)^{1/2}-1)^2}{r_T},\mathscr{F}_h,\|\cdot\|_1\right).$$

We now relate $N(\cdot,\mathscr{F}_h,\|\cdot\|_1)$ and $N_{[]}(\cdot,\mathscr{F}_h,\|\cdot\|_1)$. If $|f(w)|\leqslant|f'(w)|$ for all w, then $\|f\|_1\leqslant\|f'\|_1$. It then follows that $\|f-(f_1+f_2)/2\|_1\leqslant\|f_2-f_1\|_1/2$ for $f\in[f_1,f_2]$, so that $[f_1,f_2]$ is a subset of the $\|\cdot\|_1$-ball of radius $\|f_2-f_1\|_1/2$ and center $(f_1+f_2)/2$. This gives $N(\mathscr{F}_h,\varepsilon,\|\cdot\|_1)\leqslant N_{[]}(\mathscr{F}_h,2\varepsilon,\|\cdot\|_1)$ and then, since $\mathscr{F}_h\subset\mathscr{F}$,

$$N(\varepsilon,\mathscr{F}_h,\|\cdot\|_{1T})\leqslant N_{[]}\left(\frac{2((r_T\varepsilon+1)^{1/2}-1)^2}{r_T},\mathscr{F},\|\cdot\|_1\right),$$

$$N(\varepsilon,\mathscr{F}_h,\|\cdot\|_{1T})\leqslant N_{[]}(h,\mathscr{F},\|\cdot\|_1)\quad\text{for }\varepsilon\leqslant h_{1T}=\frac{h}{2}+\sqrt{\frac{2h}{r_T}},\tag{A.18}$$

because, for the latter, $N(\varepsilon,\mathscr{F}_h,\|\cdot\|_{1T})\leqslant N_h$ for $\varepsilon\leqslant h_{1T}$, where h_{1T} is such that $\{f\in\mathscr{E};\|f\|_{1T}\leqslant h_{1T}\}=\{f\in\mathscr{E};\|f\|_1\leqslant h/2\}$, so that h_{1T} is as in (A.18) by definition of $\|\cdot\|_{1T}$. Now (A.15) with $\eta=\delta$, (A.16), (A.17) and (A.18) yield, for T large enough

$$\begin{aligned}&\mathbb{E}^{1/2}\sup_{(f,f')\in\mathscr{F}^2;\|f-f'\|_1\leqslant\delta}|S_T(\vartheta_h(f))-S_T(\vartheta_h(f'))|^2\\&=\mathbb{E}^{1/2}\sup_{(f,f')\in\mathscr{F}_h^2;\|f-f'\|_{1T}\leqslant\delta+2\sqrt{\delta/r_T}}|S_T(f)-S_T(f')|^2\end{aligned}$$

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

$$\leqslant K\left[h_{1T}N_{[]}^{1/2}(h,\mathscr{F},\|\cdot\|_1)+\int_{h_{1T}}^{\delta+2\sqrt{\delta/r_T}}N_{[]}^{1/2}\left(\frac{(2(r_T\varepsilon/2+1)^{1/2}-1)^2}{r_T},\mathscr{F},\|\cdot\|_1\right)\mathrm{d}\varepsilon\right.$$
$$\left.+\delta N_{[]}\left(\frac{2((r_T\delta/2+1)^{1/2}-1)^2}{r_T},\mathscr{F},\|\cdot\|_1\right)\right].$$

Therefore, (7.10) yields since $\zeta<1$

$$\mathbb{E}^{1/2}\sup_{(f,f')\in\mathscr{F}^2;\|f-f'\|_1\leqslant\delta}|S_T(\vartheta_h(f))-S_T(\vartheta_h(f'))|^2$$
$$\leqslant C\left[h^{1-\zeta/2}+\frac{h^{1/2-\zeta/2}}{\sqrt{r_T}}+\int_{h_{1T}}^{\delta+2\sqrt{\delta/r_T}}\left(\frac{r_T}{((r_T\varepsilon/2+1)^{1/2}-1)^2}\right)^{\zeta/2}\mathrm{d}\varepsilon\right.$$
$$\left.+\delta\left(\frac{r_T}{(2(r_T\delta/2+1)^{1/2}-1)^2}\right)^{\zeta}\right]$$
$$\overset{\varepsilon=2v/r_T}{=}C\left[h^{1-\zeta/2}+\frac{h^{1/2-\zeta/2}}{\sqrt{r_T}}+r_T^{\zeta/2-1}\int_{r_Th_{1T}/2}^{(r_T\delta+2\sqrt{r_T\delta})/2}\frac{1}{((v+1)^{1/2}-1)^{\zeta}}\,\mathrm{d}v+\frac{\delta}{\delta^{\zeta}}(1+\mathrm{o}(1))\right]$$
$$=C\left[h^{1-\zeta/2}+\frac{h^{1/2-\zeta/2}}{\sqrt{r_T}}+(\delta^{1-\zeta/2}+\delta^{1-\zeta})(1+\mathrm{o}(1))\right]$$
$$\text{using }\int_{r_Th_{1T}/2}^{(r_T\delta+2\sqrt{r_T\delta})/2}\frac{1}{((v+1)^{1/2}-1)^{\zeta}}\,\mathrm{d}v$$
$$\sim\frac{(r_T\delta/2)^{1-\zeta/2}}{1-\zeta/2}\quad\text{since }\frac{1}{((v+1)^{1/2}-1)^{\zeta}}\overset{v\to\infty}{\sim}v^{-\zeta/2},\tag{A.19}$$

and $r_Th_{1T}\geqslant r_Th/2=1/2>0$.[14] The Chebychev Inequality then gives

$$\mathbb{P}\left(\sup_{(f,f')\in\mathscr{F}^2;\|f-f'\|_1\leqslant\delta}|S_T(\vartheta_h(f))-S_T(\vartheta_h(f'))|\geqslant\delta_0\right)\leqslant\frac{C(\delta^{1-\zeta/2}+\delta^{1-\zeta})^2+\mathrm{o}(1)}{\delta_0^2}$$

so that (7.11) yields that $\{S_T(\vartheta_h(f))\}_{f\in\mathscr{F}}$ is asymptotically stochastically equicontinuous. Showing that, for a suitable choice of h, $\sup_{f\in\mathscr{F}}|S_T(f)-S_T(\vartheta_h(f))|=\mathrm{o}_{\mathbb{P}}(1)$ will give that $\{S_T(f)\}_{f\in\mathscr{F}}$ is asymptotically stochastically equicontinuous. For $f\in[f_i,f_{i+1}]$, $S_T(f_i)\leqslant S_T(f)\leqslant S_T(f_{i+1})$, and since $\vartheta_h(f)\in[f_i,f_{i+1}]$, $\sup_{f\in\mathscr{F}}|S_T(f)-S_T(\vartheta_h(f))|\leqslant\max_{1\leqslant i\leqslant N_h}|S_T(f_{i+1})-S_T(f_i)|$. Repeating the steps leading to (A.19) with f_i in place of $\vartheta_h(f)$ yields that $\sup_{f\in\mathscr{F}}|S_T(f)-S_T(\vartheta_h(f))|=\mathrm{o}_{\mathbb{P}}(1)$, ending the study of the $S_T(f)$-type sums.

We now consider the first martingale sums. Let $M_T(f)$ and $\widetilde{M}_T(f)$ be

$$M_T(f)=\frac{1}{\sqrt{r_T}}\sum_{t=1}^{T}f_T(y_{t-1})\varepsilon_t,\quad\widetilde{M}_T(f)=\frac{1}{\sqrt{r_T}}\sum_{t=1}^{T}f_T(y_{t-1})(|\varepsilon_t|-\mathbb{E}|\varepsilon_t|),$$

with $(r_T,f_T(w))=(T,f(w/\sqrt{T}))$ or $(r_T,f_T(w))=(\sqrt{T},f(w))$, so that (7.9), (A.5) or (A.6) in Lemma A.3 yield

$$\mathbb{E}^{1/4}(M_T(f)-M_T(f'))^4\leqslant C\left[\|f-f'\|_1+2\left(\frac{\|f-f'\|_1}{r_T}\right)^{1/2}\right]^{1/2}$$
$$=C\|f-f'\|_{1T}^{1/2}=C\|f-f'\|_{2T}.\tag{A.20}$$

[14] Applying (A.15) directly to $S_T(f)$ would give diverging integrals at 0. This justifies the introduction of $S_T(\vartheta_h(f))$. Note also that usual tightness criterion based on increment bounds (see e.g. van der Vaart and Wellner, 1996, p. 104) does not apply here.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Note that $\|\cdot\|_{2T}$ defines a distance. Because $N(\varepsilon,\mathscr{F}_h,\|\cdot\|_{2T})=N(\varepsilon,\mathscr{F}_h,\|\cdot\|_{1T}^{1/2})=N(\varepsilon^2,\mathscr{F}_h,\|\cdot\|_{1T})$, we obtain in place of (A.18)

$$N(\varepsilon,\mathscr{F}_h,\|\cdot\|_{2T})\leqslant N_{[]}\left(\frac{2((r_T\varepsilon^2+1)^{1/2}-1)^2}{r_T},\mathscr{F},\|\cdot\|_1\right),$$

$$N(\varepsilon,\mathscr{F}_h,\|\cdot\|_{2T})\leqslant N_{[]}(h,\mathscr{F},\|\cdot\|_1)\quad\text{for }\varepsilon\leqslant h_{2T}=\sqrt{h_{1T}}=\left(\frac{h}{2}+\sqrt{\frac{2h}{r_T}}\right)^{1/2}.$$

Therefore, (A.15) with $\eta=\delta$, (7.10), and the change of variables $\varepsilon=2v/\sqrt{r_T}$ yield here, arguing as in (A.19)

$$\begin{aligned}
&\mathbb{E}^{1/4}\sup_{(f,f')\in\mathscr{F}^2;\|f-f'\|_1\leqslant\delta}|M_T(\vartheta_h(f))-M_T(\vartheta_h(f'))|^4\\
&=\mathbb{E}^{1/4}\sup_{(f,f')\in\mathscr{F}_h^2;\|f-f'\|_{2T}\leqslant(\delta+2\sqrt{\delta/r_T})^{1/2}}|M_T(f)-M_T(f')|^4\\
&\leqslant K\left[h_{2T}N_{[]}^{1/4}(h,\mathscr{F},\|\cdot\|_1)+\int_{h_{2T}}^{C(\delta+2\sqrt{\delta/r_T})^{1/2}}N_{[]}^{1/4}\left(\frac{2((r_T\varepsilon^2/4+1)^{1/2}-1)^2}{r_T},\mathscr{F},\|\cdot\|_1\right)\mathrm{d}\varepsilon\right.\\
&\quad\left.+\delta N_{[]}^{1/2}\left(\frac{2((r_T\delta^2/4+1)^{1/2}-1))^2}{r_T},\mathscr{F},\|\cdot\|_1\right)\right]\\
&\leqslant C\left[\left(h^{1-\zeta/2}+\frac{h^{1/2-\zeta/2}}{r_T^{1/2}}\right)^{1/2}+r_T^{\zeta/4-1/2}\int_{\sqrt{r_Th_{1T}}/2}^{(r_T\delta+2\sqrt{r_T\delta})^{1/2}/2}\frac{\mathrm{d}v}{((v^2+1)^{1/2}-1)^{\zeta/2}}\right.\\
&\quad\left.+\delta\left(\frac{r_T}{((r_T\delta^2/4+1)^{1/2}-1)^2}\right)^{\zeta/2}\right]\\
&\overset{T\to+\infty}{\to}C(\delta^{(1-\zeta/2)/2}+\delta^{1-\zeta})\quad\text{since } r_Th \text{ is bounded away from 0.}
\end{aligned}$$

It follows that $\{M_T(\vartheta_h(f))\}_{f\in\mathscr{F}}$ is asymptotically stochastically equicontinuous, so that $\{M_T(f)\}_{f\in\mathscr{F}}$ is asymptotically stochastically equicontinuous if $\sup_{f\in\mathscr{F}}|M_T(f)-M_T(\vartheta_h(f))|=o_{\mathbb{P}}(1)$. We have

$$\begin{aligned}
\sup_{f\in\mathscr{F}}|M_T(f)-M_T(\vartheta_h(f))|&\leqslant\max_{1\leqslant i\leqslant N_h}\frac{1}{\sqrt{r_T}}\sum_{t=1}^{T}(f_{(i+1)T}(y_{t-1})-f_{iT}(y_{t-1}))|\varepsilon_t|\\
&\leqslant\mathbb{E}|\varepsilon_t|r_T^{1/2}\max_{1\leqslant i\leqslant N_h}S_T(f_{i+1}-f_i)+\max_{1\leqslant i\leqslant N_h}\widetilde{M}_T(f_{i+1}-f_i).
\end{aligned}$$

Observe that $\widetilde{M}_T(\cdot)$ satisfies (A.20), so that under (7.10) and by definition of the f_i's

$$\begin{aligned}
\mathbb{E}\left(\max_{1\leqslant i\leqslant N_h}\widetilde{M}_T(f_{i+1}-f_i)\right)^4&\leqslant\sum_{i=1}^{N_h}\mathbb{E}(\widetilde{M}_T(f_{i+1}-f_i))^4\leqslant CN_h\left[h+\left(\frac{h}{r_T}\right)^{1/2}\right]^2\\
&=ChN_{[]}(h,\mathscr{F},\|\cdot\|_1)[h^{1/2}+r_T^{-1/2}]^2=o(1).
\end{aligned}$$

For $r_T^{1/2}\max_{1\leqslant i\leqslant N_h}S_T(f_{i+1}-f_i)$, (A.4) and (A.2) of Lemma A.3, (7.9) and (7.10) yield

$$\begin{aligned}
\mathbb{E}\left(r_T^{1/2}\max_{1\leqslant i\leqslant N_h}S_T(f_{i+1}-f_i)\right)^3&\leqslant r_T^{3/2}\sum_{i=1}^{N_h}\mathbb{E}(S_T(f_{i+1}-f_i))^3\leqslant Cr_T^{3/2}N_h\left[h^3+\frac{h^2}{r_T}+\frac{h}{r_T^2}\right]\\
&=ChN_{[]}(h,\mathscr{F},\|\cdot\|_1)[r_T^{3/2}h^2+r_T^{1/2}h+r_T^{-1/2}]=o(1).
\end{aligned}$$

The Markov Inequality yields $\sup_{f\in\mathscr{F}}|M_T(f)-M_T(\vartheta_h(f))|=o_{\mathbb{P}}(1)$.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

We now consider the martingale sums in Lemma 2 depending upon the increments Δy_{t-k}, $k \geqslant 1$. We set $h = 1/T$. Define

$$\overline{M}_T(f) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Delta y_{t-k} f\left(\frac{y_{t-1}}{\sqrt{T}}\right) \varepsilon_t, \quad \overline{S}_T(f) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left| \Delta y_{t-k} f\left(\frac{y_{t-1}}{\sqrt{T}}\right) \varepsilon_t \right|.$$

Let $q = (s+4)/s < 2$ since $s > 4$. Then (A.8) of Lemma A.3 and (7.12) yields

$$\begin{aligned} \mathbb{E}^{1/4}(\overline{M}_T(f) - \overline{M}_T(f'))^4 &\leqslant C\left[\|f - f'\|_1 + 2\left(\frac{\|f - f'\|_1}{T}\right)^{1/2}\right]^{1/2q} \\ &= C\|f - f'\|_{1T}^{1/2q} = C\|f - f'\|_{3T}, \end{aligned}$$

taking $r_T = T$ in the definition of $\|\cdot\|_{1T}$. Note that $\|\cdot\|_{3T}$ defines a distance with, in place of (A.18),

$$N(\varepsilon, \mathscr{F}_h, \|\cdot\|_{3T}) \leqslant N_{[]}\left(\frac{2((T\varepsilon^{2q}+1)^{1/2}-1)^2}{T}, \mathscr{F}, \|\cdot\|_1\right),$$

$$N(\varepsilon, \mathscr{F}_h, \|\cdot\|_{3T}) \leqslant N_{[]}(h, \mathscr{F}, \|\cdot\|_1) \quad \text{for } \varepsilon \leqslant h_{3T} = h_{1T}^{1/2q} = \left(\frac{h}{2} + \sqrt{\frac{2h}{T}}\right)^{1/2q}.$$

Therefore (A.15) with $\eta = \delta$, (7.10) and the change of variables $\varepsilon = 2v/T^{1/(2q)}$ yields here

$$\begin{aligned} &\mathbb{E}^{1/4} \sup_{(f,f') \in \mathscr{F}^2; \|f-f'\|_1 \leqslant \delta} |\overline{M}_T(\vartheta_h(f)) - \overline{M}_T(\vartheta_h(f'))|^4 \\ &= \mathbb{E}^{1/4} \sup_{(f,f') \in \mathscr{F}_h^2; \|f-f'\|_{3T} \leqslant (\delta + 2\sqrt{\delta/r_T})^{1/2q}} |\overline{M}_T(f) - \overline{M}_T(f')|^4 \\ &\leqslant C\left[h^{1/2q-\zeta/4} + T^{\zeta/4-1/2q} \int_{(Th_{1T})^{1/2q}/2}^{(T\delta+2\sqrt{T\delta})^{1/2q}/2} \frac{\mathrm{d}v}{((v^{2q}+1)^{1/2}-1)^{\zeta/2}} \right. \\ &\quad \left. + \delta\left(\frac{T}{((T\delta^{2q}/4+1)^{1/2}-1)^2}\right)^{\zeta/2}\right] \\ &\overset{T\to+\infty}{\to} C(\delta^{(1/2q)(1-q\zeta/2)} + \delta^{1-\zeta/q}). \end{aligned}$$

It follows that $\{\overline{M}_T(\vartheta_h(f))\}_{f\in\mathscr{F}}$ is asymptotically stochastically equicontinuous, so that $\{\overline{M}_T(f)\}_{f\in\mathscr{F}}$ is if $\sup_{f\in\mathscr{F}} |\overline{M}_T(f) - \overline{M}_T(\vartheta_h(f))| \leqslant \max_{1\leqslant i\leqslant N_h} \overline{S}_T(f_{i+1} - f_i) = o_{\mathbb{P}}(1)$. But (A.7), (7.12) and (7.10) yield

$$\begin{aligned} \mathbb{E}\left(\max_{1\leqslant i\leqslant N_h} \overline{S}_T(f_{i+1} - f_i)\right)^3 &\leqslant T^{3/2} \sum_{i=1}^{N_h} \mathbb{E}(\overline{S}_T(f_{i+1} - f_i))^3 \leqslant CT^{3/2+\zeta-3((s+1)/(s+4))} \\ &\leqslant CT^{(5/2)-3((s+1)/(s+4))} = o(1) \end{aligned}$$

since $3((s+1)/(s+4)) > \frac{5}{2}$ for $s > 14$. This ends the proof of the lemma. □

References

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

Anderson, H., 1997. Transaction costs and non-linear adjustment towards equilibrium in the U.S. treasury bill market. Oxford Bulletin of Economics and Statistics 59 (4), 465–484.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61 (4), 821–856.

Balke, N.S., Fomby, T.B., 1997. Threshold cointegration. International Economic Review 38, 627–645.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011

Bec, F., Ben Salem, M., Carrasco, M., 2004. Test for unit-root versus threshold specification with an application to the PPP. Journal of Business and Economic Statistics 22, 382–395.
Bec, F., Rahbek, A., Shephard, N., 2005. Autoregressive conditional root model. Manuscript, Nuffield College, Oxford University.
Berben, R.P., van Dijk, D., 1999. Unit root tests and asymmetric adjustment: a reassessment. Research Report EI-9902/A, Econometric Institute.
Caner, M., Hansen, B., 2001. Threshold autoregression with a unit root. Econometrica 69, 1555–1596.
Chow, Y.S., Teicher, H., 1988. Probability Theory. Independence, Interchangeability, Martingales, second ed. Springer, Berlin.
de Jong, R.M., Wang, C.H., Bae, Y., 2005. Correlation robust threshold unit root tests. Mimeo.
Enders, W., Granger, C.W.J., 1998. Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. Journal of Business and Economic Statistics 16 (3), 304–311.
Gonzalez, M., Gonzalo, J., 1998. Threshold Unit Root Models. Working Paper, U. Carlos III.
Gouriéroux, C., Monfort, A., 1995. Statistics and Econometric Models. Cambridge University Press, Cambridge, MA.
Gouriéroux, C., Robert, C., 2006. Stochastic unit root models. Econometric Theory 22 (6), 1052–1090.
Hall, A., Anderson, H., Granger, C., 1992. A cointegration analysis of treasury bill yields. The Review of Economics and Statistics LXXIV, 116–126.
Hamilton, J., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.
Kapetanios, G., Shin, Y., 2006. Unit root tests in three-regime SETAR models. The Econometrics Journal 9, 252–278.
Kapetanios, G., Shin, Y., Snell, A., 2003. Testing for a unit root in the nonlinear framework. Journal of Econometrics 112, 359–379.
Keim, D., Madhavan, A., 1997. Transaction costs and investment style: an inter-exchange analysis of institutional equity trades. Journal of Financial Economics 46, 265–292.
Michael, P., Nobay, A., Peel, D., 1997. Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation. Journal of Political Economy 105 (4), 862–879.
Obstfeld, M., Taylor, A., 1997. Nonlinear aspects of goods–market arbitrage and adjustment: Heckscher's commodity points revisited. Journal of the Japanese and International Economies 11, 441–479.
Park, J.Y., Phillips, P.C.B., 1999. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory 15, 269–298.
Park, J.Y., Phillips, P.C.B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117–161.
Park, J.Y., Shintani, M., 2005. Testing for a unit root against transitional autoregressive model. Mimeo, Vanderbilt University.
Pippenger, M.K., Goering, G.E., 1993. A note on the empirical power of unit root tests under threshold processes. Oxford Bulletin of Economics and Statistics 55 (4), 473–481.
Revuz, D., Yor, M., 1999. Continuous Martingales and Brownian Motion. Springer, Berlin.
Seo, M., 2003. Unit root test in a threshold autoregression: asymptotic theory and residual based block bootstrap. Manuscript, University of Wisconsin-Madison.
Shin, D.W., Lee, O., 2001. Tests for asymmetry in possibly nonstationary time series data. Journal of Business and Economic Statistics 19 (2), 233–244.
Shin, D.W., Lee, O., 2003. An instrumental variable approach for tests of unit-roots and seasonal unit roots in asymmetric time series models. Journal of Econometrics 115, 29–52.
Sollis, R., Leybourne, S., Newbold, P., 2002. Tests for symmetric and asymmetric nonlinear mean reversion in real exchange rates. Journal of Money, Credit and Banking 34 (3), 686–700.
Taylor, A., 2001. Potential pitfalls for the PPP Puzzle? Sampling and specification biases in mean-reversion tests of the LOOP. Econometrica 69, 473–498.
van der Vaart, A.W., 1996. Asymptotic Statistics. Cambridge University Press, Cambridge, MA.
van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes with Applications to Statistics. Springer, Berlin.
Wagner, W., 1998. The official icebergs of transaction costs. Commentary 54, Plexus Group.

Please cite this article as: Bec, F., et al., Adaptive consistent unit-root tests based on autoregressive threshold model. Journal of Econometrics (2007), doi:10.1016/j.jeconom.2007.05.011