Unit Root Tests and Threshold Adjustment: The Yield Spread Dynamics Revisited *

Frédérique Bec †

Alain Guay ‡

Emmanuel Guerre §

February 2002

Abstract

We consider modelling the yield spread between 10-year and 3-months interest rates with a stationary three-regime threshold autoregressive (TAR) model with possibly a unit root in the middle regime. This is indeed suggested by arbitrage behavior in presence of small transaction costs. To this end, we develop a test of unit root against such TAR alternative. More precisely, we propose a new data-driven choice of the set of admissible thresholds which improves the power of existing tests as shown by simulation experiments. When applying this unit-root test on post-1980 monthly data for France, Germany, New-Zealand and US, the null of unit-root is rejected. These results provide support to the expectations theory of the term structure.

JEL Classification: C12, C22, C32, E43

Keywords: Term structure, threshold autoregressive model, unit root test.

*Financial supports from FCAR and CREST are gratefully acknowledged.

†CREST–ENSAE, Timbre J120 — 3, av. Pierre Larousse, 92245 Malakoff Cedex, France. Email: bec@ensae.fr

‡CREFE, Université du Québec à Montréal, Canada. Email: guay.alain@uqam.ca

§LSTA,Université Paris 6, France. Email: eguerre@ccr.jussieu.fr

1 Introduction

The well-know expectations theory of the term structure asserts that under costless and instantaneous portfolio adjustment assumption, equilibrium interest rates are such that the investor is indifferent between holding a bond which has k period left to maturity and investing in a sequence of one-period bond for k successive periods. This non-arbitrage condition in turn implies that the long-term and short-term interest rates are cointegrated with weights $(1 - 1)$, or equivalently that the yield spread — defined as the difference between the $k-$period and the one-period interest rates — is stationary[1]. However, the empirical evidence of cointegration between yields of different maturities is still not clear-cut[2] : for instance, Campbell and Shiller [1987], Stock and Watson [1988] or Anderson [1997] find a stationary US spread whereas more recent studies by Enders and Siklos [2001] or Bohl and Siklos [2001] fail to reject the null of no-cointegration between long and short term interest rates for the US and Germany respectively. Letting alone the datasets discrepancies, one possible explanation for this quite mixed evidence may be the linear adjustment assumption underlying standard unit-root or cointegration tests. Indeed, as underlined by Anderson [1997], a non-linear adjustment is what one should expect in presence of even small transaction costs. These include the brokers commissions, the price impact of the trades, the costs of delay and missed trades, the costs of collecting relevant information as well as the tax liabilities. In this case, the non-arbitrage argument described above is expected to hold only when the magnitude of the yield spread exceeds these costs. Consequently, transaction costs imply the existence of an inaction band, within which interest rates with different maturities may depart from their equilibrium values. Therefore, this theoretical setup suggests a three-regime threshold autoregressive specification for the yield spread, the middle regime corresponding to this inaction band. The issue arising from this non-linear conjecture is that existing linearity tests require the stationarity of the variable governing the switch between regimes, here the yield spread[3]. So, one has to check first the stationarity of the spread in order to avoid fallacious inference from the threshold model estimates.

Yet, recent studies by e.g. Balke and Fomby [1997], Enders and Granger [1998], Taylor

[1] This will be precisely stated in the next section.

[2] See Pagan, Hall and Martin [1996] for an overview of this literature.

[3] Anderson [1997] actually estimates a three-regime non-linear error correction model using weekly observations of 13-week and 26-week US Treasury Bills. Since she finds that the two yield to maturity series are I(1) and cointegrated with weights $(1, -1)$, standard nonlinear inference applies.

[2001] and Enders and Siklos [2001] pointed out that standard unit-root and cointegration tests all have low power in presence of asymmetric adjustment, which may explain the frequent failure to find a stationary yield spread or a cointegration relationship between long-term and short-term interest rates. Consequently, a proper unit-root or cointegration test must allow for asymmetric adjustment under the alternative. This is precisely the issue tackled by Enders and Granger [1998] and Enders and Siklos [2001], who propose to extend the DF unit-root test and the Engle-Granger cointegration test respectively by permitting a two-regime threshold auto-regressive (TAR) specification under the alternative. However, if non-linearity arises from transaction costs as advocated by Anderson [1997], the relevant alternative hypothesis should be a three-regime threshold auto-regressive specification, so as to account for the "inaction band" in the middle regime. A few recent studies, like the ones by Caner and Hansen [2001], Shin and Lee [2001] and Gonzalez and Gonzalo [1998], have examined models where the threshold variable is stationary and differs from the dependent variable under the alternative. For the topic addressed in our paper, testing unit-root versus a threshold alternative where the threshold variable is the same as the dependent variable is in fact much more appealing. Such a test is proposed by Bec, Ben Salem and Carrasco [2001]. It is based upon the maximum of standard LR-type test statistics, where the maximum is taken with respect of the threshold level. The test we develop in this paper is similar on that point, but it departs from the previous one by allowing the threshold levels to remain bounded. Indeed, our main contribution is to derive analytically the asymptotic distribution of the proposed unit-root test without assuming that the threshold goes asymptotically to infinity under the null, an assumption retained in existing tests (Bec et al. [2001] and Berben and van Dijk [1999]) as well as in the structural change literature. This choice aims at increasing the power of the test. Another contribution of our study is to propose a data-driven choice of the admissible threshold levels used to compute the test statistics. This generally improves the power of the test as shown by a small sample simulation experiment. When applying this unit-root test on post-1980 monthly data for French, German, New Zealander and US yield spreads, the null of unit-root is rejected against the 3-regime SETAR alternative, whereas the ADF test fails to reject it. Moreover, the null of linearity is strongly rejected for the German and US data, while this result is less clear-cut for French and New Zealander data.

The remainder of the paper is organized as follows. Section 2 briefly exposes the implication of transaction costs for the yield spread process according to the expectations theory of the term structure. Section 3 presents our unit-root test statistics and their asymptotic distributions. The empirical results are reported in Section 4 and Section 5 concludes.

2 A threshold adjustment specification for the yield spread

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

$$R(k,t) = \frac{1}{k}\left[\sum_{j=1}^{k} E_t[R(1,t+j-1)]\right] + L(k,t), \tag{2.1}$$

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. This in turn implies the stationarity of the yield spread between longer-term and shorter-term interest rates. Indeed, by rearranging (2.1), the spread may be expressed as :

$$S(k,1,t) = R(k,t) - R(1,t) = \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j=1}^{i} E_t[\Delta R(1,t+j)] + L(k,t) = \mu_t, \tag{2.2}$$

where the right-hand side is stationary since it is a finite sum of stationary variables as soon as interest rates are integrated of order one — which is now a well-established stylized fact — plus the risk premium usually considered as stationary in both theoretical and empirical literature. Hence, as noticed by Hall, Anderson and Granger [1992] and Anderson [1997], arbitrage behavior guarantees that equation (2.2) acts as an attractor as soon as $S(k,1,t) \neq \mu_t$.

However, as pointed out by Anderson [1997], if one considers homogeneous transaction costs which reduce the investor's yield on a purchased bond by a constant amount λ, then the investor will convert a portfolio of one-period bonds to $k-$period bonds if and only if $\lambda < S(k,1,t) - \mu_t$, or convert $k-$period bonds to 1-period bonds if and only if $S(k,1,t) - \mu_t < -\lambda$. Therefore, in presence of transaction costs, the attraction towards equilibrium (2.2) is inactive when :

$$-\lambda < S(k,1,t) - \mu_t < \lambda. \tag{2.3}$$

Hence, there is no reason for the cointegration relation between long- and short-term rates to hold in this area, or put in other words, for the spread to revert toward μ_t. Yet, 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.

Contrary to Hall et al. [1992] and Anderson [1997] empirical analysis, we cannot derive an error correction model from this theoretical background since standard unit-root tests fail to reject the null for the spread series we consider. Consequently, the first step of our analysis must consist in testing the unit-root against the relevant nonlinear alternative which, according to the discussion above, may have the following 3-regime self-exciting TAR form :

$$\Delta y_t = a(L)\Delta y_{t-1} + \begin{cases} \mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \leq -\lambda \\ \mu_2 + \rho_2 y_{t-1} & \text{if } |y_{t-1}| < \lambda \\ -\mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \geq \lambda \end{cases} + \varepsilon_t \ , \tag{2.4}$$

where $y_t = S(k,1,t) - \mu$ reflects the deviations of the spread from its mean μ, which is defined as the mean of μ_t in (2.2). Since converting a portfolio of $k-$period bonds to 1-period bond or of 1-period bonds to $k-$period bonds is theoretically motivated by the same arbitrage behavior, we assume some kind of “symmetry” around μ in the outer regimes by imposing the autoregressive coefficient ρ_1 to be the same, and the intercept to have the same size but opposite sign.

3 The unit-root test

3.1 The test statistics

Following Bec et al. [2001], Berben and van Dijk [1999], Caner and Hansen [2001] and Enders and Granger [1998], our testing procedure is built for threshold autoregressive alternatives. Consider, for $\lambda > 0$, the following general self exciting Threshold AutoRegressive (TAR) model

$$\Delta y_t = a(L)\Delta y_{t-1} + \begin{cases} \mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \leq -\lambda \\ \mu_2 + \rho_2 y_{t-1} & \text{if } |y_{t-1}| < \lambda \\ \mu_3 + \rho_3 y_{t-1} & \text{if } y_{t-1} \geq \lambda \end{cases} + \varepsilon_t \ , \tag{3.5}$$

assuming both under the null and the alternatives that the roots of $1 - a(x)$ lies outside the unit circle. The null hypothesis of the presence of a unit root writes

$$H_0 : \rho_1 = \rho_2 = \rho_3 = 0 \text{ and } \mu_1 = \mu_2 = \mu_3 = \mu, \text{ i.e. } \Delta y_t = \mu + a(L)\Delta y_{t-1} + \varepsilon_t.$$

Hence, the parameter λ is not identified under H_0. Our TAR model differs from Caner and Hansen [2001] since our threshold variable y_{t-1} is nonstationary under the null.

Bec et al. [2001] derived a sufficient stationarity condition for (3.5). Under the assumption that the i.i.d. ε_t's with $\mathbb{E}\varepsilon_t = 0$ and $\mathrm{Var}(\varepsilon_t) = \sigma^2$ have a positive density over $\mathbb{R}$, Corollary 1 of Bec et al. [2001] states that, if a) the roots of $(1 - a(x))(1 - x) - \rho_i x$, $i = 1, 3$ lie outside the unit circle; b) the roots of $(1 - a(x))(1 - x) - \rho_2 x$ lies outside or *on* the unit circle, then the TAR model (3.5) admits as a solution a stationary process $\{y_t\}$ with β-mixing coefficients decreasing with an exponential rate. A striking fact is that condition b) allows for $\rho_2 = 0$, meaning that the process may behave as a random walk in the inner regime. From a statistical point of view, this shows that the data generated by (3.5) can be much closer to H_0 than observations issued from the linear alternative $\Delta y_t = \mu + a(L)\Delta y_{t-1} + \rho y_{t-1} + \varepsilon_t$. Indeed the latter process is close to H_0 when ρ goes to 0 while in (3.5) is close to H_0 when $\rho_2 = 0$ and $\lambda = 0$: such a sample $y_1, \ldots, y_T$ will be difficult to distinguish from H_0 because much of the observations will be in the inner regime. To detect such alternatives actually requires to base a test on some estimation of the ρ_i's. For such a purpose, we consider the slightly restricted version of the model (3.5) given by (2.4) which is dictated by the theoretical considerations exposed in Section 2. Our alternative hypothesis is

> H_1: the solution $\{y_t\}$ of (2.4) is stationary with $\rho_1 \neq 0$ for an unknown threshold parameter λ_1.

Assume that a sample of size T, $y_1, \ldots, y_T$, is at hand. We base a test of H_0 against H_1 upon the standard Lagrange Multiplier, Likelihood Ratio and Wald statistic, computed in a first step for an arbitrary value of the threshold level λ. The final tests are based upon the maxima of these statistics with respect to λ. To compute these statistics, observe first that (2.4) is a piecewise linear model as soon as λ is given. The corresponding parameters can therefore be estimated by a standard OLS regression. Let $\widehat{\epsilon}(\lambda) = (\widehat{\epsilon}_{p+1}(\lambda), \ldots, \widehat{\epsilon}_T(\lambda))'$ be the estimated residuals associated with the unrestricted model (2.4). Let $\widehat{\epsilon}_0$ be the estimated residuals from (2.4) with $\rho_1 = \rho_2 = 0$. Our test statistics are respectively,

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

Following Andrews and Ploberger [1994], Bec et al. [2001], Berben and van Dijk [1999], Caner and Hansen [2001], Davies [1987] and Hansen [1994] which consider testing issues when a nuisance parameter appears under the null, we shall eliminate λ by considering maxima of these statistics. But a major difference with Bec et al. [2001], Berben and van Dijk [1999] is that we do not consider values of λ of order $\sqrt{T}$. The motivation of the approach of Bec et al. [2001], Berben and van Dijk [1999] (and also of Caner and Hansen [2001] although these authors consider the stationary variable Δy_{t-1} as a threshold variable) follows the structural change literature which wants to make sure that a reasonable number of observations lies in the inner and outer regimes. But if λ were known under H_1, the same λ should be used to derive the test. Moreover, such a choice of a threshold depending on T may induce a loss of power because λ should be fixed under the alternative. We propose instead a data-driven choice of the set Λ_T over which the maximum is taken. Our approach is power oriented. The principle of our choice is to take a Λ_T which is small when the null is true, in order to obtain small critical values of the test statistics $\sup_{\lambda\in\Lambda_T} LM_T(\lambda)$, $\sup_{\lambda\in\Lambda_T} LR_T(\lambda)$ and $\sup_{\lambda\in\Lambda_T} W_T(\lambda)$. When the alternative is true, it is desirable that the length of Λ_T increases in order to obtain large values of the test statistic which will give a powerful test under a large class of alternatives. We therefore use in the simulation experiment and in the application a set Λ_T with length proportional to a test statistic of H_0 against H_1.

3.2 Asymptotic distributions of the test statistics

We begin with the case of an arbitrarily fixed $\lambda > 0$. Under H_0 and $\mu = 0$,

$$\left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t, \frac{y_{[Tr]}}{\sqrt{T}} \right\}_{r\in[0,1]} \xrightarrow{d} \{(\sigma, \delta)W(r)\}_{r\in[0,1]} \ , \tag{3.1}$$

where $\{W(r)\}_{r\in[0,1]}$ is a standard Brownian Motion, $\sigma^2 = \mathrm{Var}(\varepsilon_t)$ and $\delta = \sigma/(1-a(1))$. It then follows that the number of y_t's in the inner regime should be small since $|y_{[Tr]}|$ diverges when T grows. More precisely, we know from Akonom (1993) that for any $-\infty < a < b < +\infty$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbb{I}\left(y_t \in [a,b]\right) \xrightarrow{d} \frac{b-a}{\delta} \ell_W(0,1) \ ,$$

where the positive random variable $\ell_W(0,1)$ is the local time at 0 of $\{W(r)\}_{r\in[0,1]}$.[4] This shows indeed that the number of y_t's in the inner regime is of exact order $\sqrt{T}$. Consequently, one may thing that using a fixed λ will give a poor estimate $\widehat{\rho}_2(\lambda)$ which is useless in the test. This is not true because the detection power of the test must be considered under H_1 and should be due to $\widehat{\rho}_1(\lambda) \neq 0$ under H_1. This point actually argues for a fixed λ which will give a large number of observations in the outer regime and therefore a better estimation of ρ_1 under the alternative. However, under the null, the test statistics depend indeed upon a studentization of the estimated parameters. Under H_0 our three statistics are asymptotically equivalent and it is well known that the Wald statistic $W_T(\lambda)$ writes as

$$W_T(\lambda) = [\widehat{\rho}_1(\lambda), \widehat{\rho}_2(\lambda)]\, V_T^{-1}(\lambda) \begin{bmatrix} \widehat{\rho}_1(\lambda) \\ \widehat{\rho}_2(\lambda) \end{bmatrix}, \tag{3.2}$$

where $V_T(\lambda)$ is an estimate of the asymptotic variance of $(\widehat{\rho}_1(\lambda), \widehat{\rho}_2(\lambda))$. It is intuitively clear that the estimates of the two regimes will be asymptotically independent so that $V_T(\lambda)$ is asymptotically block diagonal, and the inner and outer regimes will have asymptotically the same importance in the limit distribution of $W_T(\lambda)$ under the null.

The next proposition derives, under the null, the limit distribution of our test statistic for a given λ. In what follows, $\{B(r)\}_{r\in\mathbb{R}_+}$ is a standard Brownian Motion independent of $\{W(r)\}_{r\in[0,1]}$, and therefore of $\ell_W(0,1)$.

Proposition 1 *Consider a fixed $\lambda > 0$. Under H_0 and if $\mu = 0$, $LM_T(\lambda)$, $LR_T(\lambda)$ and $W_T(\lambda)$ jointly converge in distribution to $\zeta_1^2 + \zeta_2^2(\lambda)$ with*

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

where the equality in distribution above holds for a fixed λ.

[4]The local time $\{\ell_W(w,1)\}_{w\in\mathbb{R}}$ is the occupation density of the Brownian Motion $\{W(r)\}_{r\in[0,1]}$ and can be defined by $\int_0^1 \mathbb{I}(W(r) \in [a,b])dr = \int_{-\infty}^{+\infty} \mathbb{I}(w \in [a,b])\ell_W(w,1)dw$. The distribution of $\ell_W(0,1)$ is the distribution of the absolute value of a standard normal. For further definitions on the local time, see Revuz and Yor [1999]. For applications in econometrics see Burridge and Guerre [1996], Park and Phillips [2001].

The variable ζ_1 is the asymptotic estimation error in the outer regime while $\zeta_2(\lambda)$ is the asymptotic estimation error of the inner regime. Note that ζ_1 and $\zeta_2(\lambda)$ are, as intuitively expected, independent.

Due to the fixed choice of λ, the limit ζ_1 and the distribution of $\zeta_2(\lambda)$ do not depend upon λ.[5] Note that the expression of ζ_1 is quite similar to the expression of the limit of the ADF statistic but is affected by the nonlinearities of the model (2.4) used to build the statistics, as shown by the appearance of the absolute function in ζ_1. Observe also that our test statistics are asymptotically pivotal, in the sense that the limit distribution of Proposition 1 does not depend upon the parameters of the null model. This contrasts with Bec et al. [2001] who consider $\lambda = \sqrt{T}\pi$ and find a limit distribution depending upon δ, see their Proposition 2.

The proof of Proposition 1 shows the joint limit distribution of $W_T(\lambda_1), \ldots, W_T(\lambda_k)$ is described by the Proposition (without taking into account the equality in distribution). The next proposition is a preliminary step to extend Proposition 1 to convergence in distribution of stochastic processes indexed by λ.

Proposition 2 *Consider $0 < a \leq b < \infty$. Then, under H_0 and if $\mu = 0$, the processes $\{LM_T(\lambda)\}_{\lambda \in [a,b]}$, $\{LR_T(\lambda)\}_{\lambda \in [a,b]}$ and $\{W_T(\lambda)\}_{\lambda \in [a,b]}$ are tight.*

`Tight` Note that we require $a > 0$ because the limit process $\{\zeta_1(\lambda)\}_{\lambda \in [a,b]} \overset{d}{=} \{B(\lambda)/\sqrt{\lambda}\}_{\lambda \in [a,b]}$ is not bounded at 0. The next theorem is a direct consequence of the proof of Proposition 1 and of Proposition 2.

Theorem 1 *Consider a random subset $\Lambda_T = [\underline{\lambda}_T, \overline{\lambda}_T]$ of $\mathbb{R}_+$ where $\underline{\lambda}_T = \underline{\lambda}_T(Y_1, \ldots, Y_T)$ and $\overline{\lambda}_T = \overline{\lambda}_T(Y_1, \ldots, Y_T)$ with, under H_0 and $\mu = 0$, $\mathbb{P}(\underline{\lambda}_T > 0) \to 1$, $\mathbb{P}(\overline{\lambda_T} < \infty) \to 1$ and*

$$\left(\underline{\lambda}_T, \overline{\lambda}_T\right) \xrightarrow{d} \left(\underline{\lambda}, \overline{\lambda}\right) ,$$

jointly with (3.1).

Then, under H_0 and $\mu = 0$, $\sup_{\lambda \in \Lambda_T} LM_T(\lambda)$, $\sup_{\lambda \in \Lambda_T} LR_T(\lambda)$ and $\sup_{\lambda \in \Lambda_T} W_T(\lambda)$ jointly converge in distribution to

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

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

[5] Observe however that the expression of $\zeta_2(\lambda)$ depends explicitly upon λ.

Note that the variable ζ_1 of the outer regime is not affected by the maximum operation since it is independent of λ, at the difference of $\zeta_2(\lambda)$. This is due to the fact that most of the y_t's are asymptotically in the outer regime. Observe also that the conditions on Λ_T holds if the extremities $\underline{\lambda}_T = \underline{\lambda}$ and $\overline{\lambda}_T = \overline{\lambda}$ are fixed real numbers with $0 < \underline{\lambda} \leq \overline{\lambda} < \infty$. Theorem 1 also allows for random sets Λ_T of admissible λ to compute our test statistics. The interest of such Λ_T is detailed below.

The fact that $\sup_{\lambda \in \Lambda_T} LM_T(\lambda)$, $\sup_{\lambda \in \Lambda_T} LR_T(\lambda)$ and $\sup_{\lambda \in \Lambda_T} W_T(\lambda)$ are asymptotically pivotal is essential to derive a test, and we fully this point here. Because ζ_1^2 is pivotal, we have to show that the limit distribution of $\sup_{\lambda \in \Lambda} \zeta_2^2(\lambda)$ is parameter free. We have

$$\sup_{\lambda \in \Lambda} \zeta_2^2(\lambda) \overset{d}{=} \sup_{\lambda \in [\underline{\lambda}, \overline{\lambda}]} \frac{B^2(\lambda)}{\lambda} = \sup_{t \in [1, \overline{\lambda}/\underline{\lambda}]} \frac{B^2(\underline{\lambda} t)}{\underline{\lambda} t} \overset{d}{=} \sup_{t \in [1, \overline{\lambda}/\underline{\lambda}]} \frac{B^2(t)}{t} ,$$

which has a parameter free distribution.

We now give an example of set Λ_T satisfying the requirement of Theorem 1. Let $\widehat{s}^2$ be an estimate of $\sigma^2 = \mathrm{Var}(\varepsilon_t)$ which is consistent under H_0 and remains bounded under H_1. Denote as $|Y|_{(p+1)}, \ldots, |Y|_{(T-1)}$ the ordered values of $Y_{p+1}, \ldots, Y_T$ and observe that $|Y|_{(p+1)}$ goes to 0 in probability under H_0 and H_1. Let $\widehat{\lambda}_{1/2}$ be the median of $|Y|_{(p+1)}, \ldots, |Y|_{(T-1)}$, and $\ell > 0$ be a length parameter. Bec et al. [2001] showed that $W_T(\widehat{\lambda}_{1/2})$ is asymptotically pivotal, so that $\Lambda_T = [\underline{\lambda}_T, \overline{\lambda}_T]$ with

$$\underline{\lambda}_T = |Y|_{(p+1)} + \frac{\widehat{s}}{\ell \max\left(1, W_T^{1/2}(\widehat{\lambda}_{1/2})\right)} \; , \; \overline{\lambda}_T = \underline{\lambda}_T + \ell \widehat{s} \max\left(1, W_T^{1/2}(\widehat{\lambda}_{1/2})\right) \; , \qquad (3.3)$$

satisfies the condition of Theorem 1. In particular, the length $\overline{\lambda}_T - \underline{\lambda}_T$ of Λ_T remains bounded under the null.

A striking fact is that the length $\overline{\lambda}_T - \underline{\lambda}_T$ diverges in probability under H_1 because $W_T(\widehat{\lambda}_{1/2}) \overset{\mathbb{P}}{\to} +\infty$ as shown in Bec et al. [2001], Proposition 2. This implies that the test statistic $\sup_{\lambda \in \Lambda_T} W_T(\lambda)$ will be large under H_1 especially if $\lambda_1 \in \Lambda_T$ with a high probability as shown from (3.2) since $\widehat{\rho}_1(\lambda_1)$ should be large. As a consequence, the choice (3.3) of Λ_T should increase the power of the test compared to the case where $\Lambda_T = \Lambda$ is fixed, as well as compared to other choices proposed in the literature. For instance, choices of Λ_T based upon the estimated quantiles of the $|Y_t|$'s are such that the length of Λ_T is large under H_0 as seen from (3.1), and small under H_1: this should give large critical values and a smaller test statistic under H_1, which is somehow undesirable. Our choice (3.3) has exactly the opposite behavior and should be preferable.

The next proposition studies our test statistic under H_1.

Proposition 3 *Assume that, under H_1, there is a fixed $\lambda^* > 0$ such that $\mathbb{P}(\lambda^* \in \Lambda_T) \to 1$. Then $\sup_{\lambda\in\Lambda_T} LM_T(\lambda)$, $\sup_{\lambda\in\Lambda_T} LR_T(\lambda)$ and $\sup_{\lambda\in\Lambda_T} W_T(\lambda)$ diverge under H_1.*

Theorem 1 describes the critical values of an asymptotic α-level test of H_0 against H_1 based on $\sup_{\lambda\in\Lambda_T} LM_T(\lambda)$, $\sup_{\lambda\in\Lambda_T} LR_T(\lambda)$ or $\sup_{\lambda\in\Lambda_T} W_T(\lambda)$, while Proposition 3 establishes consistency of such tests. The proof of Proposition 3 follows from Proposition 2 of Bec et al. [2001] which shows that $LM_T(\lambda^*)$, $LR_T(\lambda^*)$ and $W_T(\lambda^*)$ diverge under the alternatives. Note that the choice (3.3) of Λ_T obeys the condition of Proposition 3 because $\Lambda_T \xrightarrow{\mathbb{P}} \mathbb{R}_+$ as shown above, and then $\mathbb{P}(\lambda_1 \in \Lambda_T) \to 1$.

3.3 Simulation experiments

According to the empirical study below, we use the following TAR model with $p = 1$ and $\mu_2 = 0$ for simulation purposes :

$$\Delta y_t = a\Delta y_{t-1} + \begin{cases} \mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \leq -\lambda, \\ \rho_2 y_{t-1} & \text{if } |y_{t-1}| < \lambda, \\ -\mu_1 + \rho_1 y_{t-1} & \text{if } y_{t-1} \geq \lambda, \end{cases} + \varepsilon_t \ . \tag{3.1}$$

The ε_t's are $\mathcal{N}(0, 1)$ random variables in the simulation experiment. The model considered under H_0 is

$$\Delta y_t = a\Delta y_{t-1} + \varepsilon_t \ .$$

The sample size is $T = 325$, model (3.1) being computed for 600 observations in order that the last 300 observations achieve approximate stationarity[6]. In our simulation study, we focus on the Wald statistic $\sup_{\lambda\in\Lambda_T} W_T(\lambda)$ where Λ_T is as in (3.3) with $\ell = 4$ and

$$\widehat{s}^2 = \frac{1}{T-3}\sum_{t=2}^{T}\left(y_t - \widehat{\beta}_0 - \widehat{\beta}_1 y_{t-1} - \widehat{\beta}_2 y_{t-2}\right)^2 \ ,$$

where $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are the OLS regression coefficients of y_t on 1, y_{t-1} and y_{t-2}. In what follows, this test statistic is denoted $\mathrm{SupWald}(\Lambda_T)$

[6] We retain here the same sample size as Bec et al. [2001] for comparison sakes.

As a baseline for our comparison, we consider the supWald test statistic of Bec et al. [2001] who consider $\mathrm{SupWald}(\Lambda_T^0) = \sup_{\lambda \in \Lambda_T^0} W_T(\lambda)$ with

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

Note that both $\mathrm{SupWald}(\Lambda_T)$ and $\mathrm{SupWald}(\Lambda_T^0)$ are asymptotically pivotal, see Bec et al. [2001] for the latter. All the figures related to $\mathrm{SupWald}(\Lambda_T^0)$ and the Augmented Dickey Fuller (ADF) statistic are from Bec et al. [2001].

Table 1 compares the critical values of the two tests under the null. The critical values of the two tests are computed for $a = 0.3$, but taking $a = 0$ does not seems to affect the estimated critical values, see also Table 4. As expected, the estimated critical values of $\mathrm{SupWald}(\Lambda_T)$ are

Table 1: Empirical critical values of the unit root test ($a = 0.3$, $T = 325$, 10.000 simulations)

	15 %	10%	5%	1%
$\mathrm{SupWald}(\Lambda_T^0)$	13.2	14.5	16.5	21.1
$\mathrm{SupWald}(\Lambda_T)$	10.5	11.7	13.7	18.0

significantly below the ones of $\mathrm{SupWald}(\Lambda_T^0)$. Bec et al. [2001] also reported the estimated level of $\mathrm{SupWald}(\Lambda_T^0)$ for $\mu \neq 0$, that is in presence of a drift. They found that their test is conservative, as desirable.

We now report in Table 2 some simulations of the test under the alternative. We consider alternatives (3.1) such that the process behaves like an integrated process in the inner regime, that is with $\rho_2 = 0$. Such alternatives can indeed be very difficult to detect especially if the threshold λ_1 is large. The choice of the values of λ_1 as well as of $\mu_1 = 1.3 \times |\rho_1| \times \lambda_1$ is from Bec et al. [2001] and is guided by the estimation of model (3.1) for their specific dataset. The simulations are made using the 5% critical value of Table 1.

The power of the tests $\mathrm{SupWald}(\Lambda_T)$ and $\mathrm{SupWald}(\Lambda_T^0)$ is increasing in a, λ_1 and $|\rho_1|$. The power of all the tests falls dramatically as ρ_1 decreases to 0. The power of $\mathrm{SupWald}(\Lambda_T)$ exceeds the power of $\mathrm{SupWald}(\Lambda_T^0)$ and ADF except in cases (a) and (b).

The two tests $\mathrm{SupWald}(\Lambda_T)$ and $\mathrm{SupWald}(\Lambda_T^0)$ generally outperform the ADF test, except in case (b). This can be interpreted as follows. Case (b) corresponds to an alternative far from the null, with a large $|\rho_1|$ and a small $\lambda_1 = 2$. It is then likely that the nonlinearities induced

Table 2: Empirical power of the unit root test ($\alpha = 5\%$,$T = 325$, 1.000 simulations)

$(a, \rho_1, \rho_2, \lambda_1)$	SupWald(Λ_T)	SupWald(Λ_T^0)	ADF
$(0, -0.30, 0, 10)$	89.5	89.2	26.2
$(0, -0.10, 0, 10)$	31.8	26.9	11.8
$(0, -0.05, 0, 10)$	13.0	10.2	07.7
$(0.3, -0.30, 0, 10)^{(a)}$	91.3	97.0	30.4
$(0.3, -0.1, 0, 10)$	41.9	39.1	14.9
$(0, -0.30, 0, 5)$	85.4	84.0	18.1
$(0, -0.10, 0, 5)$	30.5	22.6	13.2
$(0, -0.05, 0, 5)$	13.1	09.4	10.0
$(0.3, -0.30, 0, 5)$	96.3	95.4	25.6
$(0.3, -0.10, 0, 5)$	45.0	35.4	16.8
$(0, -0.30, 0, 2)^{(b)}$	95.2	89.5	97.6
$(0, -0.10, 0, 2)$	44.0	22.8	39.8
$(0, -0.05, 0, 2)$	18.8	09.4	15.4

by the inner regime is of minor importance because the process rapidly escapes from the inner regime due to the random walk behavior. Therefore the most important part of the data generating process is the linear outer regime and there is a slight loss when trying to capture some nonlinearities with additional parameters. But note that the the loss is only 2% for the test SupWald(Λ_T). The SupWald(Λ_T) outperforms the ADF test for the other alternatives with $\lambda_1 = 2$.

The test SupWald(Λ_T) improves SupWald(Λ_T^0) with gain higher than 5% for 5 alternatives. The test SupWald(Λ_T^0) does substantially better than SupWald(Λ_T) for alternative (a). Inspection of the simulation reveals that the condition $\lambda_1 \in \Lambda_T^0$ with a higher frequency than $\lambda_1 \in \Lambda_T$ for this specific alternative. We do not have clear explanation for this fact.

4 Empirical results

Since the tests of linearity versus threshold proposed by Andrews and Ploberger [1994] and Hansen [1996] assume that the series is stationary and $\beta-$mixing, we must perform the unit-root test in a first step. Then, the linearity hypothesis will be tested along the lines suggested by Hansen [1996] for the series rejecting the null of unit-root.

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[7]. For the New Zealand[8], 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 3, performing the standard ADF unit-root test and KPSS stationarity test[9] 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 statis-

Table 3: 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.

tics fails to reject the null of stationarity for the US spread while the ADF fails to reject the unit-root for S_{NZ}. So, if these yield spreads were I(1) as suggested by this preliminary analysis, estimating models like (2.4) would lead to fallacious inference since the switching variable has

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

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

[9]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].

to be stationary in order to make sure that each regime is visited infinitely often asymptotically. Therefore, our dataset requires testing for unit-root against the SETAR alternative before conducting statistical inference from (2.4). We first estimate the critical values of the test for our sample size $T = 250$ using model (3.1). It is worth noting that the estimated critical values for

Table 4: Empirical critical values of the unit root test ($T = 250$, $a = 0$, 10.000 simulations)

	20 %	15 %	10%	5%	1%
SupWald(Λ_T)	10.0	10.9	12.1	14.2	18.5

$T = 250$ given in Table 4 do not seem to differ significantly from the ones for $T = 325$ except for $\alpha = 1\%$, see Table 1. We also compute the critical values of the lag polynomial $a(L)$ (increasing also the number of lags to 4) but we do not note significant variations. The values obtained for the SupWald(Λ_T) statistics are reported in Table 5. The lag order of the $a(L)$ polynomial in

Table 5: SupWald(Λ_T) unit-root test

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

model (2.4) is chosen according to the BIC and Ljung-Box statistics which suggest $p = 1$ for the European spreads, and $p = 4$ for the remainders. The SupWald(Λ_T) statistic strongly rejects the null for S_G, S_{NZ} and S_{US}, but only at the 15% level for S_F. The corresponding SETAR non-linear least squares estimates are given in Table 6. The numbers in parentheses are the Wald statistics of the null that the corresponding coefficient is zero[10]. Since White's test finds evidence of heteroskedasticity in all the cases but Germany, we present the heteroskedasticity robust version of the Wald statistics when required. Bold and italic characters denote the rejection of the null at the 5% and 10% level respectively. The W_{lin} statistics is the SupWald linearity test proposed by Hansen [1996] for testing the null hypothesis of a stationary autoregressive model that is $H_0 : \mu_1 = \mu_2 = 0$ and $\rho_1 = \rho_2$. Since the distribution of this test depends on nuisance

[10]It is distributed as a $\chi^2(1)$.

parameters (namely the values of μ, ρ, $a(L)$ and σ under H_0), the p-values are computed by simulations using the method of Hansen [1996].

As can be seen from Table 6, the linearity hypothesis is strongly rejected for the US and German data, whereas it is only rejected at the 18% and 32% level in the New Zealander and French cases respectively. Nevertheless, the point estimates of the intercepts and of the autoregressive roots seem clearly different across regimes for the French spread. On the whole, these results confirm that taking transaction costs into account may reconcile the expectations theory of the term structure with the data.

Indeed, the rather large absolute values obtained for ρ_1 — ranging from 0.13 to 0.45 — reveal a quite strong reversion mechanism towards the middle regime, especially for the US yield spread. On the contrary, the null of a unit root in the middle regime is never rejected, which suggests a high degree of persistence in the inaction band.

Then, according to our point estimates, these transaction costs range from 1.21% in New Zealand to 1.978% in France. A direct evaluation of the actual average transaction costs is rather difficult to obtain, since these costs include many components. However, the values reported in Table 6 lie in the interval of magnitudes reported in the literature. For instance, Keim and Madhavan [1997] find that the average cost of a buy (resp. sell) order is 0.49% (resp.0.55%) for exchange-listed stocks, and that the average cost of buy (resp. sell) orders is 1.23% (resp. 1.43%) for Nasdaq stocks. These authors retain a narrow definition of the transaction costs including brokers commissions and the price impact of the trade only. Their estimates for exchange-listed stocks coincide with the Plexus Group estimates of "narrow" transaction costs which are 0.32% for small cap stocks and 0.55% for large cap stocks. However, taking the "invisible" costs (which include costs of delay and costs of missed trades) into account, the Plexus Group concludes that (see Wagner [1998], p.1) "The difference in performance must be at least 2% to make the trade worthwhile for large cap stocks. [...] The average cost of trading small caps is $4,5$ times as large as trading large cap stocks." Finally, the empirical study by Grundy and Martin [2001] implies an average transaction cost of 1.5%.

5 Concluding remarks

In this paper, we proposed a test of unit root against a self-exciting TAR alternative inspired by Bec et al. [2001]. The computation of our test statistic relies on a new data-driven choice of

Table 6: TAR estimates

	S_F	S_G	S_{NZ}	S_{US}
a_1	**0.066**	**0.210**	**0.539**	**0.486**
	(13.25)	(5.61)	(45.14)	(29.93)
a_2	—	—	-0.108	*-0.211*
			(1.492)	(3.66)
a_3	—	—	0.023	*0.112*
			(0.235)	(2.91)
a_4	—	—	-0.024	*-0.090*
			(0.066)	(1.26)
μ_1	**-0.529**	-0.169	-0.361	**-0.863**
	(12.55)	(1.16)	(2.36)	(7.72)
μ_2	-0.008	-0.023	0.025	0.006
	(0.04)	(0.57)	(0.46)	(0.10)
ρ_1	**-0.254**	**-0.132**	*-0.166*	**-0.452**
	(13.23)	(4.14)	(2.77)	(8.55)
ρ_2	-0.055	0.040	-0.088	-0.028
	(1.99)	(1.64)	(2.20)	(1.64)
λ	1.978	1.700	1.211	1.591
σ	0.474	0.297	0.579	0.335
p_1	10%	15%	18%	6.6%
p_2	81%	72%	47%	83%
n	228	228	205	259
W_{lin}	4.28	15.05	3.81	7.85
	(0.32)	(0.01)	(0.18)	(0.03)

Notes: The data are centered.

p_1 = % of observations in the lower regime.

p_2 = % of observations in the inner regime.

the set of admissible thresholds which succeeds in increasing its power. This test is relevant for all economic time series driven by an underlying behavior involving an inaction band — such as the real exchange rates in presence of transaction costs, the nominal exchange rates managed by a target-zone mechanism or central banks interventions more generally — that are found to be integrated of order one in linear frameworks.

When applied to post-1980 French, German, New-Zealander and US monthly data, our test rejects the null of unit-root whereas ADF or KPSS give mixed evidence at best. Moreover, the transaction costs estimated from the SETAR specification are in accordance with the figures typically reported in existing studies. These results provide support to the expectations theory of the term structure which predicts a stationary process for the yield spread, and a non-linear behavior of this series in presence of transaction costs.

References

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

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

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

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

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

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

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

Burridge, P. and E. Guerre, The Limit Distribution of Level Crossings of a Random Walk, and a Simple Unit Root Test, *Econometric Theory*, 1996, *12*, 705–723.

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

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

Davies, R. B., Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, 1987, *74*, 33–43.

Enders, W. and C.W.J. Granger, Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates, *Journal of Business and Economic Statistics*, 1998, *16* (3), 304–11.

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

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

Grundy, B. and J. Martin, Understanding the Nature of the Risks and the Source of the Rewards to Momentum Investing, *The Review of Financial Studies*, 2001, *14* (1), 29–78.

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

Hansen, B.E., Testing for parameter instability in linear models, in N. Ericsson and J. Irons, editors, *Testing exogeneity*, Oxford University Press, 1994, chapter 15, pp. 389–406.

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

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

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

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

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

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

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

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

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

Appendix: Proofs of the main results

Let $u_t = \Delta y_t$ and for $\lambda \in \mathbb{R}_+^*$ under H_0, $\lambda = \lambda_1$ under H_1, define

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

where $\text{sgn}_\lambda(y_{t-1}) = \mathbb{I}(y_{t-1} \geq \lambda) - \mathbb{I}(y_{t-1} \leq -\lambda)$, $\beta = \beta(\lambda) = (a_1, \ldots, a_p, \mu_1, \rho_1, \mu_2, \rho_2)'$, $\beta = \beta_0 = \beta(0) = (a_1, \ldots, a_p, \mu, 1, \mu, 1)'$ under H_0. Model (2.4) gives

$$u_t = x_t'\beta + \varepsilon_t \ .$$

Consider also the $(p+4) \times (p+4)$ diagonal scaling matrix

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

We have, under H_0,

$$\Gamma\left(\widehat{\beta} - \beta_0\right) = \Gamma\left[\Gamma^{-1} \sum x_t x_t' \Gamma^{-1}\right]^{-1} \Gamma^{-1} \sum x_t \varepsilon_t \ . \tag{A.1}$$

Define also,

$$C = C(\lambda) = \Gamma^{-1} \textstyle\sum x_t x_t' \Gamma^{-1} = \begin{bmatrix} C_{11} & C_{21}' \\ C_{21} & C_{22} \end{bmatrix} ,$$

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

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

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

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

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

$$M = M(\lambda) = \Gamma^{-1}\sum x_t\varepsilon_t = \begin{bmatrix} M_1 \\ M_2^o \\ M_2^i \end{bmatrix} ,$$

$$M_1 = \begin{bmatrix} \frac{1}{T}\sum u_{t-1}\varepsilon_t \\ \frac{1}{T}\sum u_{t-2}\varepsilon_t \\ \vdots \\ \frac{1}{T}\sum u_{t-p}\varepsilon_t \end{bmatrix} , \; M_2^o = \begin{bmatrix} -\sum \mathrm{sgn}_\lambda(y_{t-1})\frac{\varepsilon_t}{\sqrt{T}} \\ \sum \frac{y_{t-1}}{\sqrt{T}}\mathbb{I}\left(|y_{t-1}| \geq \lambda\right)\frac{\varepsilon_t}{\sqrt{T}} \end{bmatrix} , \; M_2^i = \begin{bmatrix} \frac{1}{T^{1/4}}\sum \mathbb{I}(|y_{t-1}| < \lambda)\varepsilon_t \\ \frac{1}{T^{1/4}}\sum y_{t-1}\mathbb{I}(|y_{t-1}| < \lambda)\varepsilon_t \end{bmatrix} .$$

Sketch of the proof of Proposition 1. Under H_0, $C_{11} = O_{\mathbb{P}}(1)$ and $M_1 = O_{\mathbb{P}}(1)$.

We consider first the entries of C_{21} of the inner regime. We have, using Theorem 3.2 of Park and Phillips [2001], $k = 1, \ldots, p$,

$$\left|\frac{1}{T^{3/4}}\sum h(y_{t-1})\mathbb{I}(|y_{t-1}| < \lambda)u_{t-k}\right| \leq \frac{\max|u_t|}{T^{1/4}}\left|\frac{1}{\sqrt{T}}\sum h(y_{t-1})\mathbb{I}(|y_{t-1}| < \lambda)\right| = o_{\mathbb{P}}(1) \; .$$

For the entries of C_{21} of the outer regime, observe

$$\begin{aligned} &\frac{1}{T^{3/2}}\sum y_{t-1}\mathbb{I}(|y_{t-1}| \geq \lambda)u_{t-k} \\ &= \frac{1}{T^{3/2}}\sum y_{t-1}u_{t-k} - \frac{1}{T^{3/2}}\sum y_{t-1}\mathbb{I}(|y_{t-1}| < \lambda)u_{t-k} \\ &= \frac{1}{T^{3/2}}\sum u_{t-1}u_{t-k} + \cdots + \frac{1}{T^{3/2}}\sum u_{t-k}^2 + \frac{1}{\sqrt{T}}\sum \frac{y_{t-k-1}}{\sqrt{T}}\frac{u_{t-k}}{\sqrt{T}} + o_{\mathbb{P}}(1) \\ &= o_{\mathbb{P}}(1) \; . \end{aligned}$$

Mimicking arguments in Burridge and Guerre [1996] also yields that $T^{-1}\sum \mathrm{sgn}_\lambda(y_{t-1})u_{t-1} = o_{\mathbb{P}}(1)$. This shows that $C_{21} = o_{\mathbb{P}}(1)$.

Theorems 3.1 and 3.2 of Park and Phillips yields that, jointly

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

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

Observe also that, under H_0

$$\widehat{\sigma}^2(\lambda) = \frac{1}{T}\sum \left(u_t - x_t\widehat{\beta}\right)^2 \xrightarrow{\mathbb{P}} \sigma^2 .$$

Let $e = (0,1)'$. It then follows from (3.2) that

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

where the equality in distribution above comes from standard scaling properties of the Brownian motion and of the sign function, the fact that $\delta = \sigma/(1 - a(1))$ since the roots of $1 - a(x)$ lies outside the unit circle with $1 - a(0) = 1$, together with the independence between $\{W(s)\}_{s\in\mathbb{R}_+}$ and $\{B(s)\}_{s\in\mathbb{R}_+}$. Because the approximations above hold jointly with respect to λ, Proposition 1 describes the joint asymptotic distribution of $W_T(\lambda_1), \ldots, W(\lambda_k)$. It is easily seen that the three tests statistics have the same limit distribution under H_0. □

Sketch of the proof of Proposition 2. Proposition 2 is a straightforward consequence of the tightness of $\{C(\lambda)\}_{\lambda\in[a,b]}$ and $\{M(\lambda)\}_{\lambda\in[a,b]}$. The bounds used to establish Proposition 1 show that the former will follow from the tightness of

$$\left\{ \frac{1}{\sqrt{T}} \sum f(Y_{t-1}) \mathbb{I}(Y_{t-1} \leq \lambda) \right\}_{\lambda \in [a,b]} ,$$

where $f(\cdot)$ is bounded on any compact subset of $\mathbb{R}$. Note that $\{y_t\}$ is a Markov process under H_0. From Darling and Kac (1957), one can expect that, for any $\lambda_1 < \lambda_2$ and an integer $p \geq 2$,

$$\begin{aligned} \mathbb{E}\left| \frac{1}{\sqrt{T}} \sum f(Y_{t-1}) \mathbb{I}(\lambda_1 \leq Y_{t-1} \leq \lambda_2) \right|^p &\leq \sup_{y \in [a,b]} |f(y)| \mathbb{E}\left| \frac{1}{\sqrt{T}} \sum \mathbb{I}(\lambda_1 \leq Y_{t-1} \leq \lambda_2) \right|^p \\ &\leq C\left(\lambda_2 - \lambda_1\right)^p , \qquad \text{(A.2)} \end{aligned}$$

which gives tightness of $\{C(\lambda)\}_{\lambda \in [a,b]}$ using the tightness condition of Billingsley (1999).

The bounds used to establish Proposition 1 show that the tightness of $\{M(\lambda)\}_{\lambda \in [a,b]}$. will follow from the tightness of

$$\left\{ \frac{1}{T^{1/4}} \sum f(Y_{t-1}) \mathbb{I}(Y_{t-1} \leq \lambda) \varepsilon_t \right\}_{\lambda \in [a,b]} .$$

The Burkholder inequality for martingale yields

$$\mathbb{E}\left| \frac{1}{T^{1/4}} \sum f(Y_{t-1}) \mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2) \varepsilon_t \right|^{4p} \leq \frac{18(4p)^{\frac{3}{2}}}{4p-1} \mathbb{E}\left| \frac{1}{T^{1/2}} \sum f^2(Y_{t-1}) \mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2) \varepsilon_t^2 \right|^{2p} .$$

Now, since $(x+y)^{2p} \leq 2^{2p-1}(x^{2p} + y^{2p})$, we have

$$\begin{aligned} &\mathbb{E}\left| \frac{1}{T^{1/4}} \sum f(Y_{t-1}) \mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2) \varepsilon_t \right|^{4p} \\ \leq\;& \frac{18(4p)^{\frac{3}{2}}}{4p-1} \mathbb{E}\left| \frac{1}{T^{1/2}} \sum f^2(Y_{t-1}) \mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2) \left(\mathbb{E}\varepsilon_t^2 + \varepsilon_t^2 - \mathbb{E}\varepsilon_t^2\right) \right|^{2p} \\ \leq\;& C\mathbb{E}\left| \frac{1}{T^{1/2}} \sum f^2(Y_{t-1}) \mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2) \right|^{2p} \qquad \text{(A.3)} \\ &+ C\mathbb{E}\left| \frac{1}{T^{1/2}} \sum f^2(Y_{t-1}) \mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2) \left(\varepsilon_t^2 - \mathbb{E}\varepsilon_t^2\right) \right|^{2p} . \qquad \text{(A.4)} \end{aligned}$$

We will apply (A.2 to (A.3) and now study (A.4). Note that the item in this term has a martingale structure. Then Burkholder, Cauchy-Schwartz and convexity inequalities yields,

$$\begin{aligned} &\mathbb{E}\left| \frac{1}{T^{1/2}} \sum f^2(Y_{t-1}) \mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2) \left(\varepsilon_t^2 - \mathbb{E}\varepsilon_t^2\right) \right|^{2p} \\ \leq\;& \mathbb{E}\left| \frac{1}{T} \sum f^4(Y_{t-1}) \mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2) \left(\varepsilon_t^2 - \mathbb{E}\varepsilon_t^2\right)^2 \right|^{p} \end{aligned}$$

$$
\begin{aligned}
&\leq \quad C\mathbb{E}\left|\sqrt{\frac{1}{T}\mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2)}\sqrt{\frac{1}{T}\left(\varepsilon_t^2 - \mathbb{E}\varepsilon_t^2\right)^4}\right|^p \\
&\leq \quad \frac{C'}{T^{p/2}}\mathbb{E}\left|\frac{1}{\sqrt{T}}\mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2)\right|^{p/2} \\
&\leq \quad \frac{C''}{T^{p/2}}(\lambda_2 - \lambda_1)^{p/2} \, .
\end{aligned} \tag{A.5}
$$

Combining (A.3) with (A.2) and (A.4) with (A.5) yields

$$
\mathbb{E}\left|\frac{1}{T^{1/2}}\sum f^2(Y_{t-1})\mathbb{I}(\lambda_1 < Y_{t-1} \leq \lambda_2)\left(\varepsilon_t^2 - \mathbb{E}\varepsilon_t^2\right)\right|^{2p} \leq C(\lambda_2 - \lambda_1)^{p/2} \, ,
$$

which gives the tightness of $\{M(\lambda)\}_{\lambda\in[a,b]}$ if p is an integer larger than 2. $\square$