Threshold Stochastic Unit Root Models

Jun Yi Peng   Jesús Gonzalo   Raquel Montesinos

Universidad Carlos III de Madrid*

October 26, 2018

Abstract

In this paper we introduce a new class of stochastic unit root ($STUR$) processes, where a threshold variable drives the randomness of the autoregressive unit root allowing us to explain the existence of unit roots. This new model, the threshold autoregressive stochastic unit root ($TARSUR$) process is strictly stationary, but if we do not consider the threshold effect can mislead to conclude that the process has a unit root. $TARSUR$ models are not only an alternative to fix unit root model but present interpretation, estimation and testing advantages with respect to the existent $STUR$ models. The paper analyzes the properties of the $TARSUR$ models and proposes two simple tests to identify these type of process. The first test will allow us to detect the presence of unit roots, which can be fixed or stochastic, the asymptotic distribution (AD) of this test will present a distribution discontinuity depending if the unit root is fixed or stochastic. The second test we propose a simple t-statistic (or the supremum of a sequence of t-statistics) for testing the null hypothesis of a fixed unit root versus a stochastic unit root hypothesis. It is shown that it's asymptotic distribution (AD) depends if the threshold value is identified under the null hypothesis or not. When the threshold parameter is known, the AD is a standard Normal distribution, while in the case of an unknown threshold value, the AD is a functional of Brownian Bridges. Monte Carlo simulation show that the proposed tests behave very well in finite sample and the Dickey-Fuller test cannot easily distinguish between exact unit root and a threshold stochastic unit roots. The paper concludes with applications to U.S stock prices, U.S house prices, U.S interest rates, and USD/Pound exchange rates.

Key Words: Dickey-Fuller test; Difference Stationary; Near unit root; Non-stationary time series; Random coefficients; Stochastic difference equation; Stochastic Unit Roots; Time varying coefficients; Threshold models; Unit Roots.

*Address for correspondence: Jesús Gonzalo, Department of Economics, U. Carlos III de Madrid, 28903 Getafe, Madrid (Spain). E-mails: jesus.gonzalo@uc3m.es, jpeng@eco.uc3m.es

1 Introduction

It is well established that many economic series contain dominant, smooth components, even after the removal of simple deterministic trends. Since the influential work of Nelson and Plosser (1982), this characteristic has been adequately captured by unit root (*UR*) models and unit roots have became a "stylized fact" for most macroeconomic and financial time series. This has produced an extensive literature on econometric issues related to unit root models (see Phillips and Xiao (1998) for a survey).

Trying to get away from the very tight constrains that an exact unit root imposes in a process, and to be able to generate mode flexible models and of a more realistic kind, the research has recently evolved in two directions. The first generalizes *UR* models by allowing for fractional roots: *ARFIMA* models (see Granger and Joujeux (1980), Beran (1994), Robinson (1944), Baillie (1966), Dolado, Gonzalo and Mayoral (2002), etc.) The second one makes the UR model more flexible by allowing the unit root be stochastic (see Leybourne, McCabe and Tremayne (1996), Leybourne, McCabe and Mills (1996), McCabe and Tremayne (1995), Granger and Swanson (1997), Gourieroux and Robert (2006), Distaso (2008), Lieberman and Phillips (2014), etc.) instead of a fixed parameter. With both extensions, more general form of non stationary are allowed that the one implied by standard exact unit root autoregressive models. This paper forms part of the second line of research.

Stochastic unit root models (*STUR*) are seen to arise naturally in economic theory, as well as in many macroeconomic and finance applications (see Leybourne, McCabe and Mills (1996), Granger and Swanson (1997) and Lieberman and Phillips (2014)). *STUR* models can be stationary for some periods or regimes, and mildly explosive for others. This characteristic makes them not to be difference stationary. If a series shows evidence of a non stationary not removable by differencing, it is inappropriate to estimate conventional *ARIMA* or cointegration/error-correction model because the properties of the estimators and the test involved are not the same as those in the standard difference stationary case. For instance, two series generated by two independent *STUR* model will be wrongly detect to be cointegrated according to some of the most used cointegration test (see Gonzalo and Lee (1998)). This problem is not detected with standard unit root tests, such as the Dickey-Fuller test, because they cannot easily distinguish between exact unit roots and stochastic unit roots. In order to obtain a better statistic distintion between those two type of unit roots, McCabe and Tremayne (1995) proposed a locally best invariant test (assuming gaussianity) for the null hypothesis of difference stationary versus a stochastic unit root. The application of this constancy parameter test to the macroeconomic variables analyzed in Nelson and Plosser (1982) suggest that about half of them are not difference stationary, opposite to what it has bee widely believed (see Leybourne, McCabe and Tremayne (1996)). Hence, the notion that some economic time series are non stationary in a rather more general way needs to be considered and consequently more elaborate techniques of modeling and estimation need to be explored.

From a statistical pointy of view, a suitable justification for using time varying parameter models to approximate or represent non stationary process is provided by Cramer's (1961) extension of the Wold's theorem (see Granger and Newbold (1986), page 38). This extension implies that any non stationary stochastic process, with finite second order moments, may be written as an $ARMA$ process with coefficients that are allowed to vary with time. Most of the literature previously cited above considers that the time varying unit root varies as a sequence of independent and identically distributed ($i.i.d.$) random variables. This assumption is not necessarily the most appropriated in economics, because it implies that the model structure will change too often between states corresponding to stationary and explosive roots, whereas in reality, we might suppose that the transition between those two states occurs in a more gradual fashion. One way of introducing this gradual behavior is by allowing the unit autoregressive root itself to follow a random walk (see Leybourne, McCabe and Mills (1996)). In this case the change is smoother than in the $i.i.d.$ case, but it has gain the inconvenience that it occurs regularly at every moment in time. In this paper it is assumed that the economy stays in a "good" or "bad" state for a number of period of time until certain determining variable overpass some key values. This assumption is perfectly captured by modeling the evolution of economic variables via threshold models. In particular to model the random behavior of the largest root of an $ARMA$ process, we propose a threshold autoregressive (TAR) model where the largest root is less than one in some regimes and larger than one in others, in such a way that on average is equal to one. These threshold autoregressive stochastic unit root ($TARSUR$) model present several advantages with respect the previously mentioned approaches. First its computational simplicity. The estimation of all the parameters is done by least squares (LS) regression. Second, the t-statistic used to test the hypothesis of non threshold effect versus the hypothesis of threshold effect, in some cases follows asymptotically a standard distribution and therefore there is not need to generate new critical values. Third, the threshold variable is suggested by economic theory and it will be providing a possible explanation or cause for the existence of an unit root, something that to the best of our knowledge it is still absent in the econometric literature. And fourth, in many situation threshold models are easier to use for forecasting than random coefficient models. This is the case when the threshold variable is an observable variable with past time dependency.

The rest of the paper is organized as follows. In Section 2 we present economic condition which asset prices follow a TARSUR process. In section 3 we define the $TARSUR$ model and examine its properties: strict stationary, covariance stationary, geometric ergodic and impulse response function. In Section 4, we present two different test for identifying this type of process, the first one check the presence of unit roots which can be, fixed or stochastic, the second test checks for the presence of threshold effect. The asymptotic distribution of this test is developed under two different situation: when the threshold value is known and when the threshold value is unknown and unidentified. The finite sample performance (size and power) of the tests developed in this paper are analyzed in Section 5. Section 6 briefly discusses some practical issues present in all threshold models. Section 7 shows four empirical applications of our model: U.S stock prices, U.S house prices, U.S interest rates and U.S/Pound exchange rates. Finally, Section 8 draws some concluding remarks. Proofs are provided in the Appendix.

2 Predictability of Return and TARSUR

Since the work of Samuelson (1965) asset prices have been modeled as a martingale process considering returns to be unpredictable. Following Leroy (1973) and Lucas (1978) the martingale property is obtained from the Euler equation that describes the optimal behavior of the representative consumer:

$$p_t U_t' = E\left[(1+\rho)^{-1}(p_{t+1}+d_t)U_{t+1}\middle|\mathbf{F}_t\right] \tag{1}$$

where the information set $\mathbf{F}_t$ contains all past and current information available, p_t is the stock price at time t, d_t is the dividends, ρ is a constant discount factor and $U^{'}$ is the marginal utility. The simplest way to derive the martingale equivalence for asset pricing and the stochastic difference equation (1) is to assume that the asset has a zero dividend payment, with risk neutrality and $\rho = 0$. This setup is unrealistic for many assets and only can be appealing for intrinsically worthless assets like money. For a non-zero dividend payment, under risk neutrality, Samuelson (1973) shows that the martingale property holds if the discount factor is the dividend-price ratio $\rho = \frac{d_t}{p_t}$.

$$E(p_{t+1}|\mathbf{F}_t) = p_t$$

In order to generalize the martingale property, we propose a stochastic unit root specification which can be derived from an inter-temporal optimization framework. Suppose a two-period lived representative agent at time t which maximize his expected utility function

$$\underset{c_t, c_{t+1}}{\text{Max}} \quad E\Big(U(c_t) + \beta(z_t)U(c_{t+1})\Big|\mathbf{F}_t\Big)$$

where $\beta(z_t) > 0$ represents the individual time preference and will depend on the perception of the individual about the state of the world (expansion and recession, high or low unemployment, etc.). The individual has the opportunity to buy the amount h_t of a risky asset at the beginning of the period t at a known price p_t and sells it in the next period at an unknown price p_{t+1}. The considered asset yields a dividend d_t at the end of period t increasing the possibility of consumption possibility at time $t+1$. Given an exogenous steam of income w_t, the budget constraints are

$$c_t = w_t - p_t h_t$$
$$c_{t+1} = h_t(p_{t+1} + d_t)$$

Then the equilibrium condition for this model is:

$$p_t U_t' = E\left[(1+\rho(z_t))^{-1}(p_{t+1}+d_t)U_{t+1}\middle|\mathbf{F}_t\right] \tag{2}$$

where $\rho(z_t)$ is the state dependent discount factor. Following the work of Samuelson (1978), under risk neutrality and assuming that the state dependent discount rate can be represented as dividend-price ratio with a state dependent premium $\tilde{\delta}(z_t)$ with zero mean, $\rho(z_t) = \frac{p_t}{d_t} + \tilde{\delta}(z_t)$, we can establish the stochastic unit root specification. If we further assume that $\tilde{\delta}(z_t) = \tilde{\rho}_1 I(z_t \leq r) + \tilde{\rho}_2 I(z_t > r)$ have a threshold structure we can get the TARSUR process:

$$E(p_{t+1}|\mathbf{F}_t) = (1+\tilde{\delta}(z_t))p_t = \delta(z_t)p_t \tag{3}$$

where $\delta(z_t) = \rho_1 I(z_t \leq r) + \rho_2 I(z_t > r)$ with $E(\delta_t) = 1$. Under rational expectation

$$p_{t+1} = \delta(z_t)p_t + \varepsilon_t \tag{4}$$

3 *TARSUR* model

Consider the following threshold first order autorregresive (*TAR*) model

$$\begin{aligned} Y_t &= \left[\rho_1 I(Z_{t-d} \leq r_1) + \cdots + \rho_n I(Z_{t-d} > r_{n-1})\right] Y_{t-1} + \varepsilon_t = \\ &= \delta_t Y_{t-1} + \varepsilon_t, \qquad\qquad t = 1, \cdots, T \end{aligned} \tag{5}$$

where $\delta_t = \rho_1 I(Z_{t-d} \leq r_1) + \cdots + \rho_n I(Z_{t-d} > r_{n-1})$, $I(\cdot)$ is an indicator function, and ε_t is an innovation term. Z_t is the threshold variable and in this paper will be a predetermined variable ($E(\varepsilon_{t+j}|Z_t) = 0, \forall j > 0$). d is the delay parameter, and $r_1 < r_2 < \cdots < r_{n-1}$ are the threshold values.

Definition 1 *A TARSUR process is defined by equation (5) with* $E(\delta_t) = \sum_{i=1}^{n} \rho_i p_i = 1$, $\forall t$, *where* p_i *is the probability of* Z_{t-d} *being in regime* i, *and* $V(\delta_t) > 0$.

For simplicity and without loss of generality, in this section, where the properties of the *TARSUR* model are analyzed, we will not introduce any deterministic terms. They will be taken into account in the testing section.

The variables $\{\varepsilon_t\}$ and $\{Z_t\}$ satisfy the following assumptions.

Assumptions

(A.1) $\{\varepsilon_t, Z_t\}$ is strictly stationary, ergodic, adapted to the *sigma-field* $\Im_t \overset{def}{=} \{(\varepsilon_j, Z_j), j \leq t\}$.

(A.2) $\{Z_t\}$ is strong mixing with mixing coefficients α_m satisfying $\sum_{m=1}^{\infty} \alpha_m^{1/2-1/\tau} < \infty$ for some $\tau > 2$.

(A.3) ε_t is independent of $\Im_{t-1}$, $E(\varepsilon_t) = 0$ and $E\left|\varepsilon_t\right|^w = k < \infty$ with $w = 4$.

(A.4) Z_t has a continuous and increasing distribution function.

(A.5) ε_1 admits a positive continuous probability density function.

(A.6) $E(\max(0, \log|\varepsilon_1|)) < \infty$.

(A.7) $ess. \sup|\varepsilon_1| < \infty$[1].

(A.8) For $i = 1, 2, \cdots, n$ the coefficients ρ_i have the following form, $\rho_i = exp\{\frac{c_i}{T}\}$ where $c_1, c_2, \cdots, c_n$ are constants.

Assumptions (A.1) and (A.3) specify that the error term is a conditionally homoskedastic martingale difference sequence. (A.3) also bounds the extent of heterogeneity in the conditional distribution of ε_t. (A.1), (A.2), (A.3), (A.4) and (A.8) are needed to obtain the asymptotic distributions of the statistics proposed in this paper. (A.3) is the most restrictive assumption but is essential for our inference purpose. We need it in order to prove the tightness of a particular partial sum process. Assumptions (A.1) and (A.6) are required to show strict stationarity of Y_t, and (A.7) is needed for weak stationarity of $\{Y_t\}$. In many cases, (A.7) can be relaxed . For instance, if $\{\varepsilon_t\}$ and $\{Z_t\}$ are mutually independent, (A.7) can be replaced by $||\varepsilon||_p = [E|\varepsilon_1|^p]^{1/p} < \infty$, $\forall p < \infty$ (see Karlsen (1990)). Finally (A.8) restrict the autoregressive coefficients for the different regimes moves around unity. This assumption is required to solve the asymptotic distribution discontinuity in one of the test, proposed in this paper to identify these type of model.

It is important to notice that if we limit the analysis to self exciting threshold autoregressive models ($Z_t = Y_t$), then it is not possible to handle the issue of stochastic unit roots (unless we introduce deterministic components with size and sign constrains). This is so because if any of the parameters ρ_i is larger than one, the process Y_t will not be stationary and ergodic (see Petrucelly and Woolford (1984)) and therefore assumption (A.1) will not hold.

3.1 Stationary Properties

Equation (5) represents a particular case of a stochastic difference equation, where δ_t is a discrete random variable that takes different values depending on the location of the threshold variable Z_{t-d}. Iterating backwards the stochastic difference equation (5),

[1] The essential supremum of X is $ess\ sup\ X = \inf\{x : P(|X| > x) = 0\} = ||x||_\infty$.

$$
\begin{aligned}
Y_t &= \varepsilon_t + \sum_{j=1}^{m-1} \left(\prod_{i=0}^{j-1} \delta_{t-i} \right) \varepsilon_{t-j} + \left(\prod_{i=0}^{m-1} \delta_{t-i} \right) Y_{t-m} \\
&= C_{1,t}(m) + C_{2,t}(m),
\end{aligned}
\tag{6}
$$

where $C_{1,t}(m) = \varepsilon_t + \sum_{j=1}^{m-1} \left(\prod_{i=0}^{j-1} \delta_{t-i} \right) \varepsilon_{t-j}$, and $C_{2,t}(m) = \left(\prod_{i=0}^{m-1} \delta_{t-i} \right) Y_{t-m}$. From (5) and (6) the following results are obtained:

(a) If $C_{1,t}(m)$ converges, as $m \to \infty$ in L^p for $p \in [0, \infty]$[2], the $C_{1,t}(m) = \varepsilon_t + \sum_{j=1}^{m-1} \left(\prod_{i=0}^{j-1} \delta_{t-j} \right) \varepsilon_{t-j}$ is a strictly stationary solution of the stochastic difference equation defined by (5).

(b) If $C_{2,t}(m)$ converges in probability to zero, then the above solution is unique.

(c) If $p > 0$ in result (a), then $\{Y_t\}$ has a finite pth order moment.

The problem of finding conditions on $(\{\delta_t, \varepsilon_t\})$ such that $\{Y_t\}$ has a strictly or second-order stationary solution has been studied by several authors. Vervaat (1979) and Nicholls and Quinn (1982) assumes $(\{\delta_t, \varepsilon_t\})$ to be *i.i.d.* and mutually independent. Pourahmadi (1986, 1988) and Tjϕstheim (1986) allow δ_t to be dependent process. More general conditions are given in the following theorem based on Brandt (1986) and Karlsen (1990).

Theorem 1 *If the sequence $\{\varepsilon_t, Z_t\}$ satisfies assumptions (A.1), (A.6), and*

$$
-\infty < E \log |\delta_1| < 0 \tag{7}
$$

holds, then the process (5) is strictly stationary. Moreover, if (A.7) is satisfied and

$$
\sum_{j=0}^{\infty} \left(E|\psi_{t,j}|^2 \right)^{\frac{1}{2}} < \infty, \tag{8}
$$

with $\psi_{t,0} = 1$ and $\psi_{t,j} = \prod_{i=0}^{j-1} \delta_i$ for $j \geq 1$, then the process (5) is second order stationary.

Theorem 1 provides sufficient conditions for (a) and (b) to hold when $p = 0, 1$, or 2. It show that strictly and covariance stationary will depend on the type of convergence of the infinite sequence $\{\psi_{t,j}\}_{j=0}^{\infty}$. In fact, if condition (7) is satisfied, $\{\psi_{t,j}\}$ will converge absolutely almost sure to zero as j goes to infinity, and this implies the strict stationarity of process (5) (see Brandt (1986)). Mean square convergence of $\{\psi_{t,j}\}_{j=0}^{\infty}$ is obtained provided condition (8) holds, and in this case, process (5) is also second order stationary.

Note that there is a trade off between (A.7) and (8). For instance, assumption (A.7) can be relaxed by imposing $||\varepsilon||_p < \infty$, $\forall p < \infty$; but in this case, we need to modify (8) requiring a stronger condition

[2] L^0 is equivalent to convergence in probability.

$$\sum_{j=0}^{\infty}\left(E|\psi_{t,j}|^{2+\kappa}\right)^{\frac{1}{2+\kappa}} < \infty, \ for\ a\ \kappa > 0. \tag{9}$$

Also, as it mentioned before, it is assumed that $\{\varepsilon_t\}$ and $\{Z_t\}$ are mutually independent with $||\varepsilon_1||_p < \infty$, $\forall p < \infty$, then condition (8) is sufficient condition for second-order stationary.

Corollary 1 *A TARSUR process with $\rho_i > 0$, for $i = 1, \cdots, n$, is strictly stationary.*

Corollary 1 follows from Theorem 1 and establishes sufficient and easy to check conditions for a *TARSUR* process to be strictly stationary. It covers the most appealing *TARSUR* model from an empirical point of view, that is to say, the model with ρ_i values around unity: stationary for some regimes and mildly explosive for others. Notice that fixed unit root models are not stationary, but if we allow the root to be stochastic around unity we can achieve strict stationarity.

Theorem 1 produces explicit conditions for strictly stationary. However, no moments needs to exist and to the best of our knowledge, they are not explicit conditions for second order stationary, and therefore each particular case must be studied. In order to obtain explicit expressions, we work with the following representative case:

$\{\delta_t\}$ *is a 1^{st}-order stationary Markov Chain with two regimes or states (ρ_1 and ρ_2)*

This case can be generalized to an *N-order stationary Markov Chain* with $N > 1$, and to more than two regimes, but nothing is gained on the understanding of the process and the algebra became very tedious.

Sufficient conditions for second order stationary are presented in the following Proposition.

Proposition 1 *Suppose $\{Y_t\}$ is generated by (5) and $\{\delta_t\}$ is a 1^{st}-order stationary Markov Chain with two regimes (ρ_1 and ρ_2). Define the following 2×2 matrix*

$$F_2 = \begin{pmatrix} \rho_1^2 p_{11} & \rho_1^2 p_{21} \\ \rho_2^2 p_{12} & \rho_2^2 p_{22} \end{pmatrix}$$

where p_{ij} denotes the conditional probability $P(\delta_t = \rho_j|\delta_{t-1} = \rho_i)$, $i, j = 1, 2$. If the spectral radius of F_2, $\rho(F_2)$, is less than one, $\{Y_t\}$ is covariance stationary.

Notice that if we consider $\{\delta_t\}$ to be *i.i.d.* process, the above proposition became the necessary and sufficient condition established by Nicholls and Quinn (1982) for second order stationary of random coefficient autoregressive models (RCA):

$$\rho(F_2) < 1 \Longleftrightarrow E(\delta_t^2) = \rho_1^2 p_1 + \rho_2^2 p_2 < 1 \tag{10}$$

From this inequality, it is straightforward to conclude that the *TARSUR* process with an *i.i.d.* threshold variable is not covariance stationary, since $E(\delta_t^2) > 1$.

Proposition 1 determines that the covariance stationary of a *TARSUR* process depends on the transition probabilities p_{12} and p_{22}, and on the parameter values ρ_1 and ρ_2. For instance, for values of the parameters $\rho_1 = 0.9$, $\rho_2 = 1.1$, $p_{12} = 0.8$ and $p_{22} = 0.2$ the *TARSUR* process is covariance stationary. More general, it is straightforward to show that a necessary condition for $\rho(F_2) < 1$ is $p_{12} > p_{22}$ (or equivalent $p_{21} > p_{11}$). In other words, the transition probability of being in the same regime has to be smaller than the probability of changing regimes. The idea behind this condition is to avoid staying in the explosive regime for too long.

3.2 Geometric Ergodicity

From the work of Chan (1993) geometric ergodicity is required to obtain consistency (rate T) for the estimator of the threshold value ($\hat{r}$).

Finding conditions for $\{Y_t\}$ to be geometric ergodic have been studied by several authors. Chang and Tong (1985), Chang (1989) and Chen and Tsay (1991) gives conditions on the coefficients for the self exciting threshold autoregressive models. Gonzalo and Gonzalez (1997) and Gourieroux and Robert (2006) shows geometric ergodicity for the threshold autoregressive model assuming that one of the states follows a unit root process. Basrak, Davis and Mikosch (2002), Cline (2007) and Fraq, Marakova and Zakoian (2008) shows geometric ergodicity for the stochastic unit root process assuming that the sequence $\{\delta_t, \varepsilon_t\}$ are independent and identically distributed. More general condition are given in the following result based on the work of Yao and Attali (2000).

Theorem 2 *If the sequence $\{Z_t, \varepsilon_t\}$ satisfy (A.1), (A.3), (A.5) and $\{\delta_t\}$ is a positive recurrent Markov chain on a finite set $E = \{1, 2...n\}$ with transition matrix F and invariant measure η, then if:*

$$\mathbb{E}(log(\delta_t)) = \eta_1 log(\rho_1) + \eta_2 log(\rho_2) + \cdots + \eta_n log(\rho_n) < 0 \tag{11}$$

then there is a $\gamma_0 \in (0, w]$ for $w = 4$ such that the chain $X_t = \{\delta_t, Y_t\}$ is a V-uniform ergodic with $V(z, y) = |y|^{\gamma_0} + 1$.

Note that here we show a V-uniform ergodicity, which imply geometric ergodicity (Meyn and Tweedie, 2005 chapter 16).

Corollary 2 *A TARSUR process with positive recurrent Markov chain $\{\delta_t\}$ equipped with $\rho_i > 0$ for $i \in E = \{1, 2..., n\}$, is V-uniform ergodic.*

Corollary 2 follows from Theorem 2, which establish sufficient condition for the *TARSUR* process to be geometric ergodic. Note that exact unit root process are not ergodic, but if we allows the root be stochastic varying around unity and imposing conditions on the behavior of the stochastic unit root, we can archive a stronger form of geometric ergodicity.

3.3 Impulse Response Function

In order to obtain the impulse response function (*IRF*) of $\{Y_t\}$, we need to derive its $MA(\infty)$ representation. From the conditions of the first part of Theorem 1, this representation exists and can be written as

$$Y_t = \varepsilon_t + \sum_{j=1}^{\infty}\left(\prod_{i=0}^{j-1}\delta_{t-i}\right)\varepsilon_{t-j} = \sum_{j=0}^{\infty}\psi_{t,j}\varepsilon_{t-j}. \tag{12}$$

The response of Y_t to a shock, $\frac{\partial Y_{t+h}}{\partial \varepsilon_t} = \psi_{t,h}$ opposite to the fixed root case, becomes now stochastic. For this reason, the impulse response function (*IRF*) is defined as

$$\xi_h = E\left(\frac{\partial Y_{t+h}}{\varepsilon_t}\right) = E(\psi_{t,h}) = E\left(\prod_{i=0}^{h-1}\delta_{t-i}\right), \quad h = 1, 2, \cdots , \tag{13}$$

Proposition 2 *Under the conditions of Proposition 1, the IRF of the process $\{Y_t\}$ defined by (5) is given by*

$$\xi_h = \begin{pmatrix} 1 & 1 \end{pmatrix} F_1^h \begin{pmatrix} \rho_1 p_1 \\ \rho_2 p_2 \end{pmatrix}, \quad h = 1, 2, \cdots , \tag{14}$$

where $F_1 = \begin{pmatrix} \rho_1 p_{11} & \rho_2 p_{21} \\ \rho_2 p_{12} & \rho_2 p_{22} \end{pmatrix}$. *Shocks have transitory effect* $(\lim_{h\to\infty}\xi_h = 0)$ *if and only if the spectral radius of F_1, $\rho(F_1)$ is less than one.*

Proposition 2 establishes that depending on the transition probabilities, shocks can have transitory or permanent effects. It is easy to check that for *TARSUR* process, the following implications hold:

1. If $p_{22} > p_{12}$: $\lim_{h\to\infty}\xi_h = \infty$, as it happens in an explosive model.
2. If $p_{22} = p_{12}$: $\lim_{h\to\infty}\xi_h = 1\ \forall h$, as it happens in a random walk model. Note that in this case $\{Z_t\}$ is an *i.i.d.* process.
3. If $p_{22} < p_{12}$: $\lim_{h\to\infty}\xi_h = 0$ as it happens in a stationary model.

Proposition 1 together with Proposition 2 shows that *TARSUR* model are more flexible than fixed unit roots, in the sense of being able to produce a richer set of plausible scenarios. If $p_{22} \geq p_{12}$ the process is not covariance stationary and shocks have permanent and even increasing effect in mean; but if $p_{22} < p_{12}$, shocks will have only transitory effects in mean and depending on the parameter value, it can be stationary or not. This latter case of no covariance stationary but transitory shocks resembles, in the *IRF* sense, the *ARFIMA* model with a long memory parameter between 0.5 and 1. (see Dolado, Gonzalo and Mayoral (2002)).

Figure 1, a-c, displays simulated realizations from *TARSUR* and Random Walk (RW) models. The *TARSUR* series are generated by model (1) with two regimes , ε_t is an *i.i.d.* $Normal$ $(0,1)$. The random walk series is generated from the same set of innovation. The first 50 observations of each series have been disregarded to avoid any initial conditions dependency. For comparison reason, each figure shows a random walk versus three different type of TARSUR process: $p_{22} > p_{12}$, $p_{22} = p_{12}$ and $p_{22} < p_{12}$. Each figure differs by the value of the variance of the stochastic unit root coefficient. More specifically, in figure 1a, ρ_1=0.99 and $\rho_2 = 1.01$ ($V(\delta_t) = 0.0001$), in figure 1b, ρ_1=0.97 and $\rho_2 = 1.03$ ($V(\delta_t) = 0.0009$), and in figure 1c ρ_1=0.9 and $\rho_2 = 1.1$ ($V(\delta_t) = 0.001$). For small values of $V(\delta_t)$ the RW and *TARSUR* are indistinguishable. As $V(\delta_t)$ increases the *TARSUR* series became more volatile than its corresponding RW. It is worth mention that even in the most unstable case (see figure 1c), the "explosive" *TARSUR* series ($p_{22} > p_{12}$) does not look like a standard $AR(1)$ with a fixed explosive root.

[Figure 1 enters here]

3.4 Differencing a TARSUR process

Differencing model (5) we obtain

$$\Delta Y_t = (\delta_t - 1)Y_{t-1} + \varepsilon_t \tag{15}$$

Proposition 3 *Assume that $\{Y_t\}$ is generated by model (5). If δ_t has a strictly positive variance, $\{\Delta Y_t\}$ is strictly (covariance) stationary if and only if $\{Y_t\}$ is strictly (covariance) stationary.*

In contrast to fixed unit root models, stochastic unit root models are not difference stationary, in the sense that if the process is not stationary in levels, its difference will not be stationary either. Alternatively, if the process is strictly stationary (i.e., conditions of the first part of Theorem 1 are satisfied), its first difference will also be strictly stationary. In this case we can express model (15) as a $MA(\infty)$

$$\Delta Y_t = \sum_{j=0}^{\infty} \Psi_{t,j}\varepsilon_{t-j} \tag{16}$$

where $\Psi_{t,0} = 1$ and $\Psi_{t,j} = (\delta_t - 1)\psi_{t-1,j-1}$, $j \geq 1$. From (16) the IRF of $\{\Delta Y_t\}$ can be easily obtained.

4 Testing for TARSUR

Since the *TARSUR* model require that both conditions, $E(\delta_t) = 1$ and $V(\delta_t) > 0$ holds. In this section we propose a testing strategy to check both conditions. We present two independent test, in one hand we test the null of $E(\delta_t) = 1$ without any knowledge about $V(\delta_t)$, and in the other hand we test the null of no threshold effect $V(\delta_t) = 0$ without imposing any restriction on $E(\delta_t)$.

To simplify the notation, as in Caner and Hansen (2001) and Gonzalo and Pitarakis (2002), etc., from (A.4) we can replace the threshold variable with a uniform distributed variable using the following equality:

$$I(Z_{t-d} \leq r) = I(P(Z_{t-d}) \leq P(r)) = I(U_{t-d} \leq \lambda) \tag{17}$$

where $P(.)$ is the marginal distribution of $\{Z_t\}$, U_{t-d} denotes an uniform distributed random variable on $[0, 1]$ and $\lambda = P(r)$. Using the suggested transformation we can rewrite the *TARSUR* process defined in (5) as follows:

$$Y_t = \rho_1 I(U_{t-d} \leq \lambda) Y_{t-1} + \rho_2 I(U_{t-d} > \lambda) Y_{t-1} + \varepsilon_t \tag{18}$$

Since our objective is to test the conditions $E(\delta_t) = 1$ and $V(\delta_t) > 0$, is important to rewrite the model above in a way where these two conditions are expressed in terms of parameters in an equivalent regression. To do this, we add and subtract in the right hand side of model (18), $E(\delta_t)Y_{t-1} = [\rho_1\lambda + \rho_2(1-\lambda)]Y_{t-1}$, then we can rewrite the *TARSUR* model as:

$$Y_t = E(\delta_t)Y_{t-1} + (\rho_1 - \rho_2)[I(U_{t-d} \leq \lambda) - \lambda]Y_{t-1} + \varepsilon_t \tag{19}$$

subtracting in both sides Y_{t-1}

$$\Delta Y_t = [E(\delta_t) - 1]Y_{t-1} + (\rho_1 - \rho_2)[I(U_{t-d} \leq \lambda) - \lambda]Y_{t-1} + \varepsilon_t \tag{20}$$

rearranging the different terms

$$\Delta Y_t = \phi Y_{t-1} + \gamma H_t(\lambda) Y_{t-1} + \varepsilon_t \tag{21}$$

where $H_t(\lambda) = I(U_{t-d} \leq \lambda) - \lambda$, $\gamma = (\rho_1 - \rho_2)$ and $\phi = E(\delta_t) - 1$.

Both condition of the *TARSUR* process can be characterized by the parameters ϕ and γ in model (21) because:

- The parameter γ captures the variability of the coefficients, since for all $\lambda \in (0, 1)$, $V(\delta_t) = \gamma^2\lambda(1-\lambda)$ is non-zero, unless $\gamma = 0$.

- The parameter ϕ by construction captures the condition $E(\delta_t) = 1$.

As it occurs with the Dickey-Fuller (DF) t-test, in order to obtain asymptotic distributions that are invariant to the deterministic terms contained in the DGP, the regression model to implement the test will contain state dependent constant:

$$\Delta Y_t = \mu_1 I(U_{t-d} \leq \lambda) + \mu_2 I(U_{t-d} > \lambda) + \phi Y_{t-1} + \gamma H_t(\lambda) Y_{t-1} + \varepsilon_t \tag{22}$$

4.1 Testing for $E(\delta_t) = 1$

For testing the null of $E(\delta_t) = 1$ against the alternative $E(\delta_t) < 1$ without having any knowledge on $V(\delta_t)$, which can be zero or positive, this can be tested by testing in regression model (22):

$$\begin{aligned} H_0 &: \phi = 0 \\ H_1 &: \phi < 0 \end{aligned} \tag{23}$$

Under H_0, the asymptotic distribution of $t_{\phi=0}$ statistic shows a distribution discontinuity, similar to the case when we test for the autoregressive coefficient in an AR(1) process. This distribution discontinuity will depend if $V(\delta_t) > 0$, or $V(\delta_t) = 0$.

- For the case when $V(\delta_t) = 0$ this imply that $\gamma = 0$. Under H_0 the DGP (21) became:

$$Y_t = Y_{t-1} + \varepsilon_t \tag{24}$$

  which is the random walk (RW) process.

- For the case when $V(\delta_t) > 0$, $\gamma \neq 0$. Under H_0 the DGP (21) became:

$$\Delta Y_t = \gamma H_t(\lambda) + \varepsilon_t \tag{25}$$

  which is the *TARSUR* process from Definition 1.

The distribution discontinuity came from the fact that the random walk is a non-stationary process and the *TARSUR* process, from the first part of Theorem 1 and Corollary 1 is strictly stationary also possibly covariance stationary under the conditions on Proposition 1.

Lemma 1 *Suppose that $V(\delta_t) = 0$ and assumptions (A.1), (A.2), (A.3) and (A.4) hold.*

1. *Consider DGP (21), and regression model (22) with no deterministic terms. Then under $H_0 : \phi = 0$ the $t_{\phi=0}$ statistic has the following asymptotic distribution:*

$$t_{\phi=0} \Rightarrow \frac{\frac{1}{2}[W(1)^2 - 1]}{\{\int_0^1 W(r)^2 dr\}^{1/2}} \tag{26}$$

2. *Consider DGP (21), and regression model (22) with a threshold constant term. Then under* $H_0 : \phi = 0$ *the* $t_{\phi=0}$ *statistic has the following asymptotic distribution:*

$$t_{\phi=0} \Rightarrow \frac{\frac{1}{2}[W(1)^2 - 1] - W(1)\int_0^1 W(r)dr}{\{\int_0^1 W(r)^2 dr - [\int_0^1 W(r)dr]^2\}^{1/2}} \tag{27}$$

where $W(.)$ *is the standard Brownian motion.*

Note that in for the case when $V(\delta_t) = 0$, the asymptotic distribution of the $t_{\phi=0}$ statistic is the same as the case when we test for unit roots.

Lemma 2 *Suppose that* $V(\delta_t) > 0$*, under the conditions in Proposition 1 the TARSUR process is covariance stationary, then the* $t_{\phi=0}$ *statistic has the following distribution:*

$$t_{\phi=0} \Rightarrow \mathcal{N}(0, 1) \tag{28}$$

Since we do not know if $V(\delta_t)$ is positive or zero, we do not how is the asymptotic distribution of the $t_{\phi=0}$. Furthermore, even if the $V(\delta_t) > 0$ we do not know if the *TARSUR* process is covariance stationary or not. To overcome these problems we will assume that the coefficients of the *TARSUR* process move around unity, similar to the work of Phillips (1987) and Chan and Wei (1987) for the autoregressive parameter of AR(1).

Lemma 3 *Under assumptions (A.2), (A.3) and (A.8), then as* $T \to \infty$*:*

(a) $T^{-\frac{1}{2}} Y_{[Tq]} \Rightarrow \sigma J_{c_1,c_2}(q)$;

(b) $T^{-\frac{3}{2}} \sum Y_t \Rightarrow \sigma \int J_{c_1,c_2}(q)dq$;

(c) $T^{-2} \sum Y_t^2 \Rightarrow \sigma^2 \int J^2_{c_1,c_2}(q)dq$;

(d) $T^{-\frac{3}{2}} \sum Y_t I(U_{t-d} \leq \lambda) \Rightarrow \sigma\lambda \int J_{c_1,c_2}(q)dq$;

(e) $T^{-2} \sum Y_t^2 I(U_{t-d} \leq \lambda) \Rightarrow \sigma^2\lambda \int J^2_{c_1,c_2}(q)dq$;

(f) $T^{-1} \sum Y_{t-1}\varepsilon_t \Rightarrow \sigma^2 \int J_{c_1,c_2}(q)dW(q)$;

(g) $T^{-1} \sum Y_{t-1} I(U_{t-1-d} \leq \lambda)\varepsilon_t \Rightarrow \sigma^2 \int J_{c_1,c_2}(q)dW(q, \lambda)$;

where the integral is over $(0, 1)$ *with* $\sigma^2 = \mathbb{E}(\varepsilon^2)$*, W(.) is the standard Brownian motion and* $J_{c_1,c_2}(q) = [W(q) + (c_1\lambda + c_2(1-\lambda))\int_0^q e^{(q-s)(c_1\lambda + c_2(1-\lambda))} W(s)ds]$ *is the Ornstein-Uhlenbeck process.*

You may wonder how the $V(\delta_t)$ enters in the process J_{c_1,c_2}, similarly to the case of autoregressive process, it is captured by the term $C = c_1\lambda + c_2(1-\lambda)$. Then the asymptotic distribution of the $t_{\phi=0}$ statistic under H_0 using the near unit root set up is

Proposition 4 *Suppose that assumption (A.1), (A.2), (A.3), (A.4) and (A.8) hold.*

1. *Consider DGP (21) and regression model (22) with no deterministic term. Then under $H_0 : \phi = 0$ the $t_{\phi=0}$ statistic has the following asymptotic distribution:*

$$t_{\phi=0} \Rightarrow \frac{\int_0^1 J_{c_1,c_2}(q)dW(q)}{\{\int_0^1 J^2_{c_1,c_2}(q)dq\}^{1/2}} \tag{29}$$

2. *Consider DGP (21) and regression model (22) with a threshold constants term. Then under $H_0 : \phi = 0$ the $t_{\phi=0}$ statistic has the following asymptotic distribution:*

$$t_{\phi=0} \Rightarrow \frac{\int_0^1 J_{c_1,c_2}(q)dW(q) - W(1)\int_0^1 J_{c_1,c_2}dq}{\{\int_0^1 J^2_{c_1,c_2}(q)dq - [\int_0^1 J_{c_1,c_2}(q)dq]^2\}^{1/2}} \tag{30}$$

Note that distribution presented above is a function of the nuisance parameters $C = c_1\lambda + c_2(1-\lambda)$ and this distribution will change depending if the $V(\delta_t)$ is positive or zero.

- If the $V(\delta_t) > 0$, under H_0 of $E(\delta_t) = 1$, the strictly stationary condition in Theorem 1 imposes the restriction $-\infty < \mathbb{E}(log|\delta_t|) < log(\mathbb{E}|\delta_t|) = 0$, which under assumption (A.8) implies that $-\infty < C < 0$.
- If the $V(\delta_t) = 0$, under H_0 of $E(\delta_t) = 1$, this imposes the restriction $\rho_1 = \rho_2 = 1$ and therefore under assumption (A.8) $c_1 = c_2 = 0$, which implies $C = 0$.

Since C is unknown and cannot be estimated, we use sub-sampling to obtain critical values, (Romano and Wolf, 2001 and Berg, McMurry and Politis 2010). Sub-sampling requires knowledge about the rate of convergence of the estimator, $\hat{\phi}$, which in this case can be $\sqrt{T}$ or T depending on $V(\delta_t)$ and $E(\delta_t)$. To overcome this problem we follow the work of Romano and Wolf (2001) by using the studentized statistic.

In order to apply sub-sampling two more conditions have to be checked:

1. Under H_0, the studentized statistic, $t_{\phi=0}$, has a non-degenerated distribution.
2. The sub-sampling statistic is strongly mixing.

For our propose both condition are satisfied since, from Proposition 4, the first condition stated above is satisfied whether $V(\delta_t) = 0$ or $V(\delta_t) > 0$. The second condition is also satisfied because when $V(\delta_t) > 0$, form Theorem 2 and Corollary 2, the process $\{Y_t\}$ is geometric ergodic and for the case where $V(\delta_t) = 0$, it is proven in Romano and Wolf (2001).

4.2 Testing for Threshold Effect

The goal of this section is to construct a test for the null of non-threshold effect versus the alternative of a threshold effect. It is worthwhile to emphasize that we do not make any assumption about $E(\delta_t)$ which can be equal to one or less than one. Assuming that $0 < \lambda < 1$ the null hypothesis of non-threshold effect ($V(\delta_t) = 0$) versus the alternative of a threshold effect ($V(\delta_t) > 0$) can be tested by testing

$$\begin{aligned} H_0 &: \gamma = 0 \\ H_1 &: \gamma \neq 0 \end{aligned} \tag{31}$$

in regression model (22)

The propose test and its asymptotic distribution depends whether the threshold parameter λ is known or unknown and unidentified under the null.

4.2.1 Threshold Value Known

The case of a known threshold value, $\lambda = \bar{\lambda}$, became relevant for pedagogical or explanatory reasons as well as for cases where the regimes are determined by the sign of the threshold value (see Enders and Granger (1998) momentum TAR model). In this situation the proposed is the t-statistic for $\gamma = 0$, $t_{\gamma=0}(\lambda)$ in regression model (22), and its asymptotic is shown in the next proposition.

Proposition 5 *Suppose that the threshold value is known, $\lambda = \bar{\lambda}$, and assumptions (A.1), (A.2), (A.3) and (A.4) hold. Whether $E(\delta_t)$ is equal to one or less than one, under the null of no threshold, $t_{\gamma=0}(\bar{\lambda})$ statistic has the following asymptotic distribution*

$$t_{\gamma=0}(\bar{\lambda}) \Rightarrow \mathcal{N}(0,1) \tag{32}$$

4.2.2 Threshold Value Unknown

When the threshold value λ is unknown it is assumed that this parameter lies in the interval $(0,1)$. The LS estimate of λ is the value that

$$\underset{\lambda \in (0,1)}{\operatorname{argmin}} \quad \hat{\sigma}^2(\lambda) \tag{33}$$

where $\hat{\sigma}^2(\lambda) = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^2$ denotes the residual variance from the LS estimation of model (22). This estimate $\hat{\lambda}$ will coincides with the one obtained by maximizing the Wald statistics, $W_T(\lambda)$, that test the null hypothesis of no threshold in regression (22)

$$W_T = W_T(\hat{\lambda}) = \sup_{\lambda \in (0,1)} \quad W_T(\lambda) \tag{34}$$

where $W_T(\lambda) = t^2_{\gamma=0}(\lambda)$. Then the asymptotic distribution of W_T is

Proposition 6 *Suppose that assumptions (A.1), (A.2), (A.3) and (A.4) hold. Whether $E(\delta_t)$ is equal to one or less than one, under the null of no threshold:*

1. *Consider DGP (21) and regression model (22) with no deterministic term. Then under the null $H_0 : \gamma = 0$, the W_T statistic has the following asymptotic distribution:*

$$W_T \Rightarrow \underset{\lambda \in (0,1)}{sup} \frac{\left(\int W(s) dV(s,\lambda)\right)^2}{\lambda(1-\lambda)\int W(s)^2 ds} \equiv \underset{\lambda \in (0,1)}{sup} \frac{[BB(\lambda)]^2}{\lambda(1-\lambda)} \tag{35}$$

where $W(.)$ is the standard Brownian motion and $V(s,\lambda)$ is a Kiefer-Muller process [3] *on $[0,1]^2$. $BB(\lambda)$ is a standard Brownian bridge (zero mean Gaussian process with covariance $\lambda_1 \wedge \lambda_2 - \lambda_1\lambda_2$). The last equivalence came from the fact that $W(s) = W(s,1)$ and $V(s,\lambda)$ are independent.*

2. *Consider DGP (21) and regression model (22) with a threshold constant term. Then under the null $H_0 : \gamma = 0$, the W_T statistic has the following asymptotic distribution:*

$$W_T \Rightarrow \underset{\lambda \in (0,1)}{sup} \frac{\left(\int W(s)^* dV(s,\lambda)\right)^2}{\lambda(1-\lambda)\int W^*(s)^2 ds} \equiv \underset{\lambda \in (0,1)}{sup} \frac{[BB(\lambda)]^2}{\lambda(1-\lambda)} \tag{36}$$

where $W^*(.) = W(.) - \int_0^1 W(s)ds$.

From the empirical point of view, we cannot search the threshold parameter λ in the unit interval because as λ approaches to zero or one, we do not have enough observation to estimate the parameters of one of the states. As in the structural break literature (Andrews (1993, 2003) and Estrella (2003)) they search for the break in a subset of the unit interval defined by $[\pi_1, \pi_2]$. We propose the same approach by dropping a proportion π_0 of the set of threshold parameter candidates in the right and the left, such that $\pi_1 = \pi_0$ and $\pi_2 = 1 - \pi_0$, then

$$\sup_{\lambda \in [\pi_1, \pi_2]} \frac{[BB(\lambda)]^2}{\lambda(1-\lambda)} \tag{37}$$

For different π_0, the critical values of the asymptotic distribution (37) are tabulated in (Andrews (1993, 2003) and Estrella (2003)).

[3]A *Kiefer-Muller* V on $[0,1]$ is given by $V(t_1,t_2) = B(t_1,t_2) - t_2 B(t_1,1)$ is a *standard Brownian sheet.* The *standard Brownian sheet* $B(t_1,t_2)$ is a zero-mean Gaussian process indexed by $T = [0,1]^2$ and covariance function $Cov[B(s_1,t_t), B(s_2,t_2)] = (s_1 \wedge t_1)(s_2 \wedge t_2)$.

5 A Monte Carlo Experiment and Testing Strategy.

Using Monte Carlo method we examine the performance of the proposed tests, as well as the power of the Dickey Fuller test t-test against different $TARSUR$ alternatives. The Monte Carlos experiment consist on 10,000 replications with sample sizes $T = 200$ and 500. The error term ε_t is generated as $i.i.d.$ $\mathcal{N}(0, 1)$ and the threshold variable follows, without loss of generality, a first order Markov process with transition matrix:

$$F = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \tag{38}$$

To fill these transition probabilities such that $E(\delta_t) = 1$ holds, first we fix the coefficient ρ_1 and ρ_2 we choose a $\lambda = P(r)$ such that $E(\delta_t) = \lambda\rho_1 + (1 - \lambda)\rho_2 = 1$. Second we fix p_{21} and by using the conditional probabilities property we can fill the rest of the transition probabilities since $p_{22} = 1 - p_{21}$, $p_{12} = p_{21}\frac{\lambda}{1-\lambda}$ and $p_{22} = 1 - p_{12}$.

Tables 1 and 4 shows the empirical size for the two proposed tests, for the mean $E(\delta_t) = 1$ and for the variance $V(\delta) = 0$, under different sample sizes and dependence level of the threshold variable. In these simulations we assume that the probability of being in regime $\rho_1 = 1$ is the same as being in regime $\rho_2 = 1$, that is $\lambda = P(r) = 0.5$. This condition imposes a symmetry restriction on the matrix F where $p_{12} = p_{21}$, and also in the case when $p_{21} = 0.5$, all the entries of matrix F is equal to 0.5, which represent the case where the threshold variable is $i.i.d.$. Table 1 summarize the results assuming that the threshold parameter is known and we can see that the empirical size coincide with the nominal size of 5% for both tests. Table 4 reports the same results as in Table 1 but assuming that the threshold parameter is unknown and unidentified. For the latter case we search the threshold parameter in a subset generated by sorting the threshold variable from smallest to the biggest, and dropping 15% of the elements in the left and the right, that is $\pi_0 = 0.15$.

Tables 2 and 5 shows the empirical size of the test for the mean, $E(\delta_t) = 1$ and the power of the tests for the variance, $V(\delta_t)$ under different levels of dependency in the threshold variable $p_{21} = \{0.5, 0.7, 0.9\}$ and different values of $|\gamma| = \{0.02, 0.04, 0.1, 0.2\}$. We choose ρ_1 and ρ_2 such that $\lambda = 0.5$ and $|\rho_1 - \rho_2| = |\gamma|$. The results in Tables 2 is constructed by assuming that the threshold parameter is known, and we can see that independently of the value of $|\gamma|$, the empirical size for the test of the mean $(E(\delta_t) = 1)$ coincide with the nominal size of 5%. Also we can see that as $|\gamma|$ gets bigger the empirical power for the test of the variance, $V(\delta_t)$ goes to one. Tables 5 shows the same result under the assumption where the threshold parameter is unknown and unidentified.

Tables 3 and 6 reports the same information as in Tables 2 and 5 respectively but in these cases we will choose ρ_1 and ρ_2 such that $\lambda = P(r)$ is different from 0.5 which allows the matrix F be asymmetric.

Tables 7, 8, 15 and 16 shows the power for the test of the mean, $E(\delta_t) = 1$, and the size for the test $V(\delta_t) = 0$ under different dependency levels of the threshold variable, $p_{21} = \{0.5, 0.9\}$. Using a local alternative approach we allow the coefficients take the form $\rho_i = 1 - \frac{k}{T}$ for some $k \geq 0$ and $i = 1, 2$. The threshold variable is generated by assuming that the probability of being in regime ρ_1 is the same as being in regime ρ_2, that is $\lambda = 0.5$. As before, for the case where $p_{21} = 0.5$, all the entries of matrix F will be 0.5 which means that the threshold variable is *i.i.d.*. Tables 7 and 8 assumes that the threshold value is known, as we showed in Proposition 5 independently of the value of k the empirical size of the test for $V(\delta_t)$ coincide with the nominal level of 5%. For the test of the mean, $E(\delta_t) = 1$, as $k \to \infty$ the empirical power of the test tends to one. Tables 15 and 16 shows the same results but assuming that the threshold parameter is unknown and unidentified.

Table 9, 10, 11, 12,13 and 14 shows the power for the test of $E(\delta_t) = 1$ under different specifications of $|\gamma| = \{0.02, 0.04, 0.1, 0.2\}$ and dependency level of the threshold variable, $p_{21} = \{0.5, 0.9\}$. We select the coefficients $\rho_1 = a_1 - \frac{k}{T}$ and $\rho_2 = a_2 - \frac{k}{T}$ for some a_1 and a_2 such that $\lambda = 0.5$ and $|\rho_1 - \rho_2| = |\gamma|$. We can see that as $k \to \infty$ the power of the test for the mean $E(\delta_t) = 1$ goes to one. Also as $|\gamma|$ gets bigger the power of the test for the $V(\delta_t)$ tends to one. Tables 17, 18, 19 20, 21 and 22 shows the same results but assuming that the threshold parameter is unknown and unidentified. For an illustration purpose we also present the power of the Dickey-Fuller (DF) unit root test against the same *TARSUR* alternatives previously considered. The $t-$statistic is calculated from regression

$$\Delta Y_t = b_1 + b_2 Y_{t-1} + v_t \tag{39}$$

The conclusion is that the DF unit root test can not easily distinguish between a pure unit root and a threshold stochastic unit root.

6 Some Extra Issues

From the empirical point of view there are four extra issues that are present in all threshold models and need to be discussed. These issues are:

1. Models with higher dynamics. For practical proposes, model (5) is rather too simplistic, so it has to be replaced as in Leybourne, McCabe and Tremayne (1996) by a more general version of (5)

$$Y_t^* = \delta_t Y_{t-1}^* + \varepsilon_t \tag{40}$$

where

$$Y_t^* = Y_t - \sum_{i=1}^{p} \omega_i Y_{t-i}, \tag{41}$$

with all the roots of the lag polynomial $\Phi_p(L) = 1 - \sum_{i=1}^{p} \omega_i L^i$ lying outside of the unit circle. The advantage of this formulation is that when $E(\delta_t) = 1$, under the null the process is integrated of order one and the hypothesis of threshold effect can still be framed in terms of the single parameter γ. The model has the following representation

$$Y_t = \sum_{i=1}^{p+1} \eta_{it} Y_{t-i} + \varepsilon_t, \tag{42}$$

where $\eta_{1t} = (\delta_t + \omega_1)$, $\eta_{it} = (\omega_i - \delta_t \omega_{i-1})$ for $i = 2, \cdots, p$ and $\omega_{p+1,t} = -\delta_t \phi_p$. With $E(\delta_t) = 1$, under the null hypothesis $\gamma = 0$, Y_t is an $AR(p+1)$ process with a non-random unit root because the coefficients η_{it} still sum to unity. Alternatively, when $\gamma \neq 0$, Y_t is a random coefficient $AR(p+1)$ process. The sum, s_t, of the $p+1$ AR coefficient is given by

$$s_t = \delta_t (1 - \sum_{i=1}^{p} \omega_i) + \sum_{i=1}^{p} \omega_i, \tag{43}$$

so that s_t has a mean of unity and variance $V(\delta_t)(1 - \sum_{i=1}^{p} \omega_i)^2$. Thus, when $\gamma \neq 0$ $(V(\delta_t) > 0)$, Y_t represents an $AR(p+1)$ process with a random unit root. It is straightforward to show that the result of Theorem 1 as well as the asymptotic theory developed in Section 3 still hold. For the later, the only required modification is to add p lags of ΔY_t in regression model (22). The number of lags can be chosen by some information criteria (see Kapetianos (2001)).

2. Determination of the number of regimes. The number of regimes can be determined by sequential testing or by some model selection technique. The first approach consist on running the TARSUR tests sequentially in a similar fashion as it is done in Bai and Peroon (1998) for structural breaks. The second approach inherits the spirit of the first one but it uses some information criteria instead of a test. This has been introduced in Gonzalo and Pitarakis (2002). The consistency of both approaches needs to be proved for a TARSUR framework.

3. Inference on the threshold parameter r. This is the hardest topic in the literature. To the best of our knowledge the most general solution is given via the use of subsampling methods in Gonzalo and Wolf (2004). Extensions of this approach to a TARSUR framework is under current investigation by the authors.

4. Misspecification of the threshold forcing variable. This type of misspecification produces, as in the standard omission of a relevant variable case, inconsistency of the parameter estimate, unless the true and wrong threshold variable variable split the sample in a similar fashion. In practice, we propose to choose the threshold variable by some information criteria.

7 Empirical Applications

In order to provide an empirical illustration of how the estimation and testing of a $TARSUR$ model can be applied in practice, we present four applications where there exist some theoretical and/or empirical controversy about the randomness of the unit root in the AR representation. The first example models U.S stock prices, the second example investigate the U.S house price, the third example analyze the U.S interest rate, and forth the nominal exchange rate between USD/Pound exchange rates.

7.1 U.S Stock Price

Following the economic model presented in Section 2, in this application we investigate via our TARSUR model the link between asset prices and real activity, as well as the predictability of the asset returns. The data analyzed is the quarterly series of real Standard and Poor Composite Stock Price Index from 1947:1 to 2016:4. The threshold variable representing the real activity is the increment of real GDP. More information about the data on stock prices can be found in Shiller (http://www.econ.yale.edu/ shiller/data.htm), and the GDP (S.A.) series in U.S. Bureau of Economic Analysis, retrieved from FRED
(https://fred.stlouisfed.org/series/GDPC1).

The estimated model for the stock prices is the $TARSUR$ model

$$\Delta Y_t = \mu_1 I(Z_{t-d} \leq \lambda) + \mu_2 I(Z_{t-d} > \lambda) + \phi Y_{t-1} + \gamma H_t(r) Y_{t-1} + \varepsilon_t,$$

where Y_t is the real stock price index and Z_t correspond to changes in the real GDP ($\Delta rgdp_t$). Dickey-Fuller unit root test suggest that real stock prices as well as the real GDP contain a unit root, therefore Z_t is $I(0)$.

[Tables 23 and 24 here]

Table (23) summarize the estimation result of the TARSUR model. Since we have to estimate the threshold parameter, we search in the set generated by ordering the observation of $\Delta rgdp_t$ from smallest to the biggest and dropping 15% of the elements of this set in the right and the in left, in terms of the distribution (37), $\pi_1 = 0.15$ and $\pi_2 = 0.85$. Testing for $E(\delta_t) = 1$, we can see that $t_{\phi=0} = -1.398$ and the 5%

critical value obtained using sub-sampling is $CV_{t_{\phi=0}} = -2.43$, therefore we fail to reject the null of $E(\delta_t) = 1$. Testing for $V(\delta_t) > 0$, the null hypothesis of no threshold effect is clearly rejected at 5% significant level since $W_T = 13.76$ versus the critical value of $CV_{t_{\gamma=0}} = 8.86$ tabulated in Estrella (2003) for $\pi_0 = 0.15$.

The *TARSUR* model does not only captures a clearly positive relationship between stock market and the real activity, but it does find a candidate variable Z_t to explain the causes of why stock prices may have a unit root. To evaluate the forecast performance we test the one step-ahead forecast of stock returns, ΔY_t, produced from our *TARSUR* model with respect the RW with drift ($\Delta Y_t = c + u_t$). Since the *TARSUR* process is a nested model of the RW process, we follow the method proposed by Clark and West (2006) where the one step-ahead forecast errors is constructed by using the estimated parameters $(\hat{\phi}, \hat{\gamma}, \hat{\lambda}, \hat{\mu}_1, \hat{\mu}_2)$ from a rolling window regression and construct the mean square prediction adjusted statistic (MSPE-adjusted). We test under the null of equal forecast error variance $H_0 : \sigma^2_{RW} = \sigma^2_{TARSUR}$. Following the argument of Ashley, Granger and Schmalensee (1980), Clark and McCracken (2001, 2003, 2006), the alternative hypothesis considered will be one-sided $H_1 : \sigma^2_{RW} > \sigma^2_{TARSUR}$ because if the the process does not follow a RW, we expect forecast from *TARSUR* model to be superior to those from the RW. The MSPE-adjusted statistic we obtain is $t_{MSPE-adj} = 5.03$ which is greater than the 5% critical value of a standard normal. Also we measure the forecasting performance by counting the number of times that the sign of the returns is predicted correctly. The *TARSUR* model predicts the sign correctly 69% of times whereas the RW model predict 55% of times correctly. From the forecasting point of view the *TARSUR* model also has a good performance.

To recover the estimates of ρ_1 and ρ_2 there are two forms. After failing to reject the null of $H_0 : \phi = 0$, the first method is to estimate the following unrestricted model:

$$Y_t = \mu_1 I(Z_{t-d} \leq \lambda) + \mu_2 I(Z_{t-d} > \lambda) + \rho_1 I(Z_{t-d} \leq \lambda)Y_{t-1} + \rho_1 I(U_{t-d} \leq \lambda)Y_{t-1} + \varepsilon_t \tag{44}$$

The second form is to impose the null of $\phi = 0$ on regression model (17) such that from the maintained hypothesis of unit root ($\rho_1\lambda + \rho_2(1 - \lambda) = 1$) and the estimated parameters, $\hat{\gamma}$ and $\hat{\lambda}$, is straightforward to recover the estimates of ρ_1 and ρ_2 and the transition probabilities $\hat{p}_{22}$ and $\hat{p}_{12}$ (see Table (24)). When $E(\delta_t) = 1$ holds the estimates of ρ_1 and ρ_2 in both methods should be the same. The results is Table (23) and (24) show that when, the increment of real GDP is less than 78.71, the stock price index is in the stationary and mean reverting regime (autoregressive parameter equals to 0.976). The estimated probability of being in this regime is 0.68. On the other hand, when the increment of the real GDP are larger than 78.71, prices follow a mildly explosive model (autoregressive parameter is equal to 1.023). This occurs with probability 0.32. On the overall, the stochastic root of the autoregressive representation is on average unity.

[Insert figure 2 here]

Figure (2) present the plot of the U.S stock prices, the green dots represent the periods in which the TARSUR model tell us that the stock prices is in the explosive state and the red dots represent the periods in which the stock price is the mean reverting period. The vertical lines represents U.S recessions (www.nber.org/cycles/cyclesmain.html). From this plot we can see that the TARSUR model is able to identify the periods in which the stock prices is expanding and the periods in which is contracting.

Given that the estimated value of the delay parameter d is equal to one, at time $t-1$ it is known in which regime we are at period t. Therefore, stock prices will not be martingale process with respect the information set formed by past values of Y_t and $\Delta rgdp_t$. In other words, if $\Delta rgdp_t$ is considered a plausible explanation of the stochastic unit root, future returns could be predictable in the sense that

$$E_{t-1}\left(\frac{Y_t - Y_{t-1}}{Y_{t-1}}\right) = E_{t-1}(\delta_t - 1) \neq 0 \tag{45}$$

From (45) and the results in Table (23) and (24) we conclude that if we were in a "recession" state at time $t-1$ ($\Delta rgdp_t < 78.71$), the expected value of returns at time t would be negative. On the contrary, if we were in an "expansion" state ($\Delta rgdp_t > 78.71$) the expected return would be positive. In that way, we find that there exist a positive non-linear relationship between the expected stock return and the real activity of the economy. Linear links between the stock returns and macroeconomic variables have already found in the finance literature (Chen et al. (1986), Fama (1990)).

7.2 U.S House Price

In this application we study the link between house price and real activity using *TARSUR* model. The analyzed data is the quarterly series of U.S real home price index from 1961:1 to 2016:04. The threshold variable representing real activity is the quarterly growth rate of real GDP per-capita. More information about the U.S real house price index can be found in the website of Shiller (http://www.econ.yale.edu/ shiller/data.htm) and about the real GDP per-capita (S.A) series can be found in Federal Reserve Bank of St.Louis (https://fred.stlouisfed.org).

Price bubbles is not a new phenomena and it was modeled as an explosive autoregressive process. From the historical perspective (Tulipmania, South sea bubble, 1929 stock market crash, Dotcom bubble and the more recent house market bubble) we have learned that bubbles have a peculiar behavior, that is, a period in which the asset price grows sharply followed by a sudden steep drop.

Model price bubble as an explosive autoregressive process capture the period in which the bubble is expanding, but is unable to capture the price drop. The *TARSUR* model solves this problem by allowing some of the autoregressive coefficients be above unity for some period and bellow unity for others, but on average one. This change on the coefficients will be able to capture the explosive and implosive behavior of price bubble, and also we will be able to find a plausible random variable capable to explain this behavior change.

As before the estimated $TARSUR$ model for the house price is:

$$\Delta Y_t = \mu_1 I(Z_{t-d} \leq \lambda) + \mu_2 I(Z_{t-d} > \lambda) + \phi Y_{t-1} + \gamma H_t(r) Y_{t-1} + \varepsilon_t,$$

where Y_t is the real house price index and Z_t is the quarterly growth rate of GDP per-capita ($\Delta RgdpP_t$). The usual Dickey-Fuller test suggest that home price index and real GDP per-capita have a unit root but $\Delta RgdpP_t$ is $I(0)$.

[Insert Tables 25 and 26 here]

Table (25) summarize the estimation results, again since the threshold parameter is unknown, we search the threshold parameter in a subset generated by dropping 15% of the threshold value candidates from the right and the left in the set generated by ordering the observations of ($\Delta RgdpP_t$). Testing for $E(\delta_t) = 1$, the t-statistic is $t_{\phi=0} = 0.551$, which compared with the critical value obtained using sub-sampling $CV_{t_{\phi=0}} = -2.969$, clearly we fail to reject the null of $E(\delta_t) = 1$. Testing for threshold effect, the null hypothesis is clearly rejected at 5% significant level since the Wald test is $W_T = 16.556$, compared to the critical value of $CV_{\gamma=0} = 8.86$ tabulated in Estrella (2003) for $\pi_0 = 0.15$.

In this empirical application we also compare the one-step ahead forecast performance of the estimated $TARSUR$ process with respect the UR with drift ($\Delta Y_t = \mu + u_t$). Again following the procedure proposed by Clark and West (2006), we test the null of equal forecast error variance ($H_0 : \sigma^2_{RW} = \sigma^2_{TARSUR}$). The MSPE-adjusted is $t_{MSE-adj} = 3.06$ which is greater than the 5% critical value of a standard normal, rejecting the null of equal forecast error variance in favor of the $TARSUR$ model. Also we measure the number of times the sign is predicted correctly, which also the $TARSUR$ model is slightly superior by predicting 69% correctly against 65% predicted by the RW.

In Table (26) we recover the estimates of ρ_1 and ρ_2 and also the transition probabilities $\hat{p}_{22}$ and $\hat{p}_{12}$. The results in Table (25) and (26) shows that when the quarterly growth rate of the GDP per capita is less than 0.28%, the real house price is in the stationary regime (with autoregressive parameter of 0.97). The probability of being in this regime is 0.33. If the quarterly growth rate of the GDP per capita is larger than 0.28%, the real house price follows a mildly explosive process (autoregressive parameter 1.02). The probability of being in this regime is 0.67.

[Insert Figure 3 here]

Figure (3) present the plot of the U.S real house price index, the vertical lines represents U.S recessions (www.nber.org/cycles/cyclesmain.html). The green dots represents the periods in which the $TARSUR$ model tell us that the house price is in the explosive state and the red dots represents the periods where the $TARSUR$ model tell us that the house price is in mean reverting state. Note that the $TARSUR$ model is able to asses something about the 2008 house price bubble, since is able to capture the explosive behavior of house price between 2001 to 2008 represented by green dots and the implosion of house price between 2008 to 2010 represented by red dots.

7.3 U.S Interest rates

In his empirical application we analyze the U.S 3-months treasure bill interest rates using our $TARSUR$ model. The series have monthly frequency from January 1949 to December 2016, more information is available in Federal Reserve Bank of St.Louis (https://fred.stlouisfed.org).

Leybourne, McCabe and Mills (1996) performs a similar exercise for the international U.S bond yield data (BUS) but with a higher frequency data on a shorter period (daily close of trade observation from April 1^{st} to December 29^{st} 1989). They find that the null hypothesis of fixed unit root versus the alternative of a stochastic unit root is clearly not rejected.

In order to apply our $TARSUR$ model we need a candidate for a threshold variable. There is an extensive literature showing the negative relation between interest rates and unemployment rate (Sargent, Fand and Goldfeld (1973), Friedman (1977), Blanchard, Wolfers (2000), etc.). Then the threshold variable we use will be annual change of unemployment rate ($Aunrate_t$) available in Federal Reserve Bank of St.Louis.

[Insert Tables 25 and 26 here]

Table (25) shows the estimation results of the $TARSUR$ model. Testing for $E(\delta_t) = 1$, we fail to reject the null hypothesis of $E(\delta_t) = 1$ since the t-statistic $t_{\phi=0} = -0.843$ which is greater than the critical value generated by sub sampling $CV_{\phi=0} = -3.56$. Testing for $Var(\delta_t) > 0$, we reject the null of no threshold effect since the Wald test $W_T = 16.548$ which is greater than $CV_{\gamma=0} = 8.86$ from Estrella (2003) for $\pi_0 = 0.15$.

The $TARSUR$ model captures a negative non-linear relationship between the interest rates and the annual increment of unemployment rates. From Table (25) and (26) we can see that if the annual change of unemployment rate is less than 0.4% the interest rate is in the "explosive" state with autoregressive coefficient of 1.006, which is close to one, and the probability of being in this regime is 0.74. If the annual change of unemployment rate is greater than 0.4%, the interest rates is in the mean reverting state with coefficient 0.968, and the probability of being in this regime is 0.26.

[Insert Figure 4 here]

Figure (4) plots the series of interest rates, the green dots represents the periods in which the the interest rates is in the "explosive" state ($Aunrate_t \leq 0.4\%$) and the red dots are the periods in which the interest rates are in the mean reverting state ($Aunrate_t > 0.4\%$). The vertical lines represent NBER recession (www.nber.org/cycles/cyclesmain.html). As we can see during the recession periods, which unemployment rate increases the interest rate tends to decline, consistent with the economic theory and the $TARSUR$ model is able to capture a non-linear relationship of this phenomena.

Also we evaluate the forecast performance of the TARSUR model against the UR process with drift ($\Delta Y_t = \mu + u_t$). The MSPE-adjusted statistic is $t_{MSPE-adj} = 1.98$ which is rejected at 5% significant level, but it is not rejected at 1% significant level. Furthermore we evaluate the number of times the $TARSUR$ model predicts correctly the sign with respect the RW. In this case the $TARSUR$ model have a very similar performance to the RW with 48% and 47% respectively.

7.4 Dollar/Pound Nominal Exchange Rates

For the last empirical application we try to find a non-linear behavior of the US dollar and British pound nominal exchange rates using our $TARSUR$ model. The data we use is the monthly series of nominal exchange rate of US dollar per Pound from January 1978 to December 2016. More information is available in Federal Reserve Bank of St.Louis (https://fred.stlouisfed.org).

In order to estimate a $TARSUR$ model we need to find a suitable threshold variable. In the work of Messe and Rugoff (1983), Barbara Rossi (2006) they use the first difference of the nominal short-term interest rate differential between countries as one of the explanatory variable suggested by the economic theory. Following their work we use this first difference of the nominal interest differential as a threshold variable. More information about the series of short-term interest rate can be found in OECD database (http://www.oecd.org/std).

Meese and Rogoff (1983) found that economic models used to forecast exchange rates is outperformed by the random walk. One possible explanation of this phenomena is the presence of parameter instability. In order to explore this puzzle and improve the out-of-sample forecast there is a lot of work in time varying parameter models, Engle (1994) and Marsh (2000) uses regime-switching models, but is still unable to beat the random walk. Schinasi and Swamy (1989), Rossi (2006) uses random coefficient models and they are able to have a better out-of-sample forecast than the RW.

[Insert Table 27 and 28 here]

Table (27) Shows the estimation of the $TARSUR$ model. Testing for $E(\delta_t) = 1$, we fail to reject the null hypothesis of $\phi = 0$ since the t-statistic $t_{\phi=0} = -1.79$ which is greater than the critical value obtained using sub-sampling, $CV_{\phi=0} = -2.90$. Testing for $V(\delta_t)$, we clearly do not reject the null of no threshold effect since the Wald statistic $W_T = 7.84$ which is smaller than $CV_{\gamma=0} = 8.86$ from Estrella (2003) for $\pi_0 = 0.15$. The results from the tests, suggests the presence of an unit root and this unit root is fixed.

From the forecast perspective, using the method of Clark and West (2006), we compare the *TARSUR* model with respect the random walk with drift ($\Delta Y_t = \mu + \varepsilon_t$) and clearly we fail to reject the null of equal variance of error forecast. The MSPE-adjusted statistic is $t_{MSPE-adj} = 0.94$. Furthermore the proportion where the *TARSUR* model predicts correctly the sign of the exchange rates is 51.42% which is slightly better than the RW with 47%. This result we obtain is similar to the one obtained by Engle (1993) with a different methodology. The advantage of the TARSUR model is that we can find a reason why both models have a equal out-of-sample forecast performance in terms of mean square error. This is because we are not able to reject the presence of fixed unit roots.

8 Conclusion

This paper introduces a new class of stochastic unit root models (*TARSUR*) where the random behavior of the unit root is driven by an economic threshold variable. By doing that, we do not make only the unit root models more flexible but we find an explanation for the existence of unit roots. Flexibility is obtained because depending on the values of certain parameters, *TARSUR* process can behave like an explosive process, like an exact unit root process, or like a stationary process. Explanatory power is gained because *TARSUR* models, by identifying an economic variable as a threshold variable, can provide a cause for the existence of unit roots.

Empirical applications shows that estimation and testing of *TARSUR* models is not more difficult than the estimation and testing involved fixed unit root models. This is a clear advantage of *TARSUR* models with respect to other stochastic unit root methodologies available in the literature.

9 Proofs

Proof of Theorem 1. The condition for strict stationary follows from Brandt (1986), and the weak stationary from Karlsen (1990).

Proof of Corollary 1. From $V(\delta_t) > 0$ and by Jensen's inequality we get

$$\mathbb{E}log|\delta_1| < log\mathbb{E}|\delta_1| = log\mathbb{E}\delta_1 = 0 \tag{46}$$

Therefore condition (7) holds.

Proof of Proposition 1. The condition for covariance stationary is given by,

$$\sum_{j=0}^{\infty} E\Big(|\psi_{t,j}|^2\Big)^{\frac{1}{2}} = \left[\begin{pmatrix}1 & 1\end{pmatrix}\sum_{j=1}^{\infty} F_2^j \begin{pmatrix}\rho_1 p_1 \\ \rho_2 p_2\end{pmatrix}\right] < \infty, \tag{47}$$

with $F_2 = \begin{pmatrix}\rho_1^2 p_{11} & \rho_1^2 p_{21} \\ \rho_2^2 p_{12} & \rho_2^2 p_{22}\end{pmatrix}$. This infinite sum converges if the spectral radius of F_2 is less than one.

Proof of Theorem 2. The proof is in the paper of Yao and Attali (2000), Theorem 1, with $|f_k(y)| = a_k|y| + b_k$ for $k \in E = \{1, 2..., n\}$ where $\{a_k, b_k\}$ are positive constants.

Proof of Corollary 2. The proof is the same as Corollary 1.

Proof of Proposition 2. The IRF can be expressed as

$$\xi_h = \begin{pmatrix}1 & 1\end{pmatrix}\sum_{j=1}^{\infty} F_1^h \begin{pmatrix}\rho_1 p_1 \\ \rho_2 p_2\end{pmatrix}, \quad h = 1, 2, \cdots, \tag{48}$$

where $F_1 = \begin{pmatrix}\rho_1 p_{11} & \rho_1 p_{21} \\ \rho_2 p_{12} & \rho_2 p_{22}\end{pmatrix}$. Therefore $lim_{h\to\infty}\xi_h$ converges to zero if and only if the spectral radius of F_1 is less than one.

Proof of Proposition 3. Iterating backwards equation (15),

$$\Delta Y_t = \varepsilon_t + (\delta_t - 1)\sum_{j=1}^{m-1}\left(\prod_{i=1}^{j-1}\delta_{t-i}\right)\varepsilon_{t-j} + (\delta_t - 1)\left(\prod_{i=1}^{m-1}\delta_{t-i}\right)Y_{t-m}. \tag{49}$$

Subtracting (15) from equation (49)

$$\Delta Y_t(Y_{t-m}) - \Delta Y_t = (\delta_t - 1)(Y_{t-1}(Y_{t-m}) - Y_{t-1}), \tag{50}$$

where $\Delta(Y_{t-m})$ correspond to equation (49) and ΔY_t to equation (15). As long as $V(\delta_t) > 0$, $\Delta(Y_{t-m})$ converges almost sure (in mean square) to ΔY_t as $m \to \infty$, if and only if $Y_{t-1}(Y_{t-m})$ converges almost sure (in mean square) to Y_{t-1}.

In order to derive the asymptotic distribution of the proposed tests we need to use some of the asymptotic tools developed in Caner and Hansen (2001).

Define the partial-sum process

$$W_T(s,\lambda) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} I(U_{t-d} \leq \lambda)\varepsilon_t, \tag{51}$$

with $\lambda = P(Z_{t-d} \leq r) = P(r)$. Theorem 1 in Caner and Hansen (2001) establishes that

$$W_T(s,\lambda) \Rightarrow \sigma W(s,\lambda), \tag{52}$$

on $(s,\lambda) \in [0,1]^2$ as $T \to \infty$, where $W(s,\lambda)$ is a *standard Brownian sheet* on $[0,1]^2$, and $\sigma^2 = E(\epsilon_1^2)$.

Definition 2 *A standard Brownian sheet S indexed by $R^+ \times [0,1]$ is a zero-mean Gaussian process with continuous sample paths and covariance function,*

$$Cov[S(s,u), S(t,v)] = (s \wedge t)(u \wedge v).$$

Following Theorem 2 in Caner and Hansen (2001) if $Y_t = Y_{t-1} + \varepsilon_t$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_t I(U_{t-d} \leq \lambda)\varepsilon_t \Rightarrow \sigma \int_0^1 W(s)dW(s,\lambda), \tag{53}$$

where $W(.)$ is a standard Brownian motion. Finally from Theorem 3 in Caner and Hansen (2001)

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} Y_t I(U_{t-d} \leq \lambda) \Rightarrow \lambda\sigma \int_0^1 W(s)ds \tag{54}$$

$$\frac{1}{T^2} \sum_{t=1}^{T} Y_t^2 I(U_{t-d} \leq \lambda) \Rightarrow \lambda\sigma^2 \int_0^1 W^2(s)ds \tag{55}$$

The proofs are divided into two parts depending if the deterministic components are included in the regression model (22): (1) no deterministic components included $\mu_1 = \mu_2 = 0$, and (2) including state dependent constant terms $\mu_1 \neq \mu_2$.

Let's start writing a close form of the estimator of ϕ and γ for the case in which no deterministic components are considered and allow us to rewrite model (22) as follows,

$$\Delta Y_t = X_{t-1}\beta + \varepsilon_t \tag{56}$$

where $X_t = \begin{pmatrix} Y_{t-1} & H_t(\lambda)Y_{t-1} \end{pmatrix}$ and $\beta = \begin{pmatrix} \phi \\ \gamma \end{pmatrix}$. Then the least square estimate of β is,

$$\hat{\beta} = \left(\sum_{t=1}^{T} X'_{t-1}X_{t-1}\right)^{-1}\left(\sum_{t=1}^{T} X'_{t-1}\Delta Y_t\right), \tag{57}$$

equivalently

$$\hat{\beta} - \beta = \left(\sum_{t=1}^{T} X'_{t-1}X_{t-1}\right)^{-1}\left(\sum_{t=1}^{T} X'_{t-1}\varepsilon_t\right), \tag{58}$$

Now,

$$\sum_{t=1}^{T} X'_{t-1}X_{t-1} = \begin{pmatrix} \sum_{t=1}^{T} Y_{t-1}^2 & \sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) \\ \sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) & \sum_{t=1}^{T} Y_{t-1}^2 H_t^2(\lambda) \end{pmatrix} \tag{59}$$

Define $\Gamma_b = \begin{pmatrix} T^b & 0 \\ 0 & T^b \end{pmatrix}$ for $b = \{\frac{1}{2}, 1\}$ depending if the process Y_t is covariance stationary or not, and multiplying both sides of (58) we get

$$\Gamma_b(\hat{\beta} - \beta) = \begin{pmatrix} T^{-2b}\sum_{t=1}^{T} Y_{t-1}^2 & T^{-2b}\sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) \\ T^{-2b}\sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) & T^{-2b}\sum_{t=1}^{T} Y_{t-1}^2 H_t^2(\lambda) \end{pmatrix}^{-1} \begin{pmatrix} T^{-2b}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \\ T^{-2b}\sum_{t=1}^{T} Y_{t-1}H_t(\lambda)\varepsilon_t \end{pmatrix} \tag{60}$$

Equation (60) is key since we derive the asymptotic distribution of the tests from here, for the case in which the regression model (22) we do not considere deterministic terms.

Let's write the least square estimate of ϕ and γ when state dependent constants are introduced in the regression model (22). We can estimate both parameters from the following regression,

$$\begin{aligned} [I(U_{t-d} \leq \lambda)\Delta Y_t^I + I(U_{t-d} > \lambda)\Delta Y_t^{II}] &= \phi[I(U_{t-d} \leq \lambda)Y_{t-1}^I + I(U_{t-d} > \lambda)Y_{t-1}^{II}] \\ &\quad + \gamma[(1-\lambda)I(U_{t-d} \leq \lambda)Y_{t-1}^I + \lambda I(U_{t-d} > \lambda)Y_{t-1}^{II}] \end{aligned} \tag{61}$$

where $\Delta Y_t^I = \left(\Delta Y_t - \frac{\sum_{t=1}^{T} I(U_{t-d}\leq\lambda)\Delta Y_t}{\sum_{t=1}^{T} I(U_{t-d}\leq\lambda)}\right)$, $\Delta Y_t^{II} = \left(\Delta Y_t - \frac{\sum_{t=1}^{T} I(U_{t-d}>\lambda)\Delta Y_t}{\sum_{t=1}^{T} I(U_{t-d}>\lambda)}\right)$, $Y_{t-1}^I = \left(Y_{t-1} - \frac{\sum_{t=1}^{T} I(U_{t-d}\leq\lambda)Y_{t-1}}{\sum_{t=1}^{T} I(U_{t-d}\leq\lambda)}\right)$ and $Y_{t-1}^{II} = \left(Y_{t-1} - \frac{\sum_{t=1}^{T} I(U_{t-d}>\lambda)Y_{t-1}}{\sum_{t=1}^{T} I(U_{t-d}>\lambda)}\right)$.

Let us rewrite model (61) as follows

$$[I(U_{t-d} \leq \lambda)\Delta Y_t^I + I(U_{t-d} > \lambda)\Delta Y_t^{II}] = \tilde{X}'_{t-1}\beta + \varepsilon_t \tag{62}$$

where $\tilde{X}'_{t-1} = \Big(I(U_{t-d} \leq \lambda)Y^I_{t-1} + I(U_{t-d} > \lambda)Y^{II}_{t-1} \quad (1-\lambda)I(U_{t-d} \leq \lambda)Y^I_{t-1} + \lambda I(U_{t-d} > \lambda)Y^{II}_{t-1}\Big)$ and $\beta = \begin{pmatrix} \phi \\ \gamma \end{pmatrix}$. Then as before the least square estimate

$$\tilde{\beta} = \left(\sum_{t=1}^{T} \tilde{X}'_{t-1}\tilde{X}_{t-1}\right)^{-1}\left(\sum_{t=1}^{T} \tilde{X}'_{t-1}\Delta Y_t\right), \tag{63}$$

equivalently

$$\tilde{\beta} - \beta = \left(\sum_{t=1}^{T} \tilde{X}'_{t-1}\tilde{X}_{t-1}\right)^{-1}\left(\sum_{t=1}^{T} \tilde{X}'_{t-1}\varepsilon_t\right), \tag{64}$$

Now,

$$\sum_{t=1}^{T} \tilde{X}'_{t-1}\tilde{X}_{t-1} = \begin{pmatrix} \tilde{x}_1 & \tilde{x}_2 \\ \tilde{x}_3 & \tilde{x}_4 \end{pmatrix} \tag{65}$$

where $\tilde{x}_1 = \sum_{t=1}^{T}\Big[I(U_{t-d} \leq \lambda)(Y^I_{t-1})^2 + I(U_{t-d} > \lambda)(Y^{II}_{t-1})^2\Big]$, $\tilde{x}_2 = \tilde{x}_3 = \sum_{t=1}^{T}\Big[(1-\lambda)I(U_{t-d} \leq \lambda)(Y^I_{t-1})^2 + \lambda I(U_{t-d} > \lambda)(Y^{II}_{t-1})^2\Big]$ and $\tilde{x}_4 = \sum_{t=1}^{T}\Big[(1-\lambda)^2 I(U_{t-d} \leq \lambda)(Y^I_{t-1})^2 + \lambda^2 I(U_{t-d} > \lambda)(Y^{II}_{t-1})^2\Big]$.

For $\Gamma_1 = \begin{pmatrix} T & 0 \\ 0 & T \end{pmatrix}$ and multiplying both sides of (64) we get,

$$\Gamma_1(\tilde{\beta} - \beta) = \begin{pmatrix} T^{-2}\tilde{x}_1 & T^{-2}\tilde{x}_2 \\ T^{-2}\tilde{x}_3 & T^{-2}\tilde{x}_4 \end{pmatrix}^{-1} \begin{pmatrix} T^{-1}\sum_{t=1}^{T}[I(U_{t-1} \leq \lambda)Y^I_{t-1} + I(U_{t-1} > \lambda)Y^{II}_{t-1}]\varepsilon_t \\ T^{-1}\sum_{t=1}^{T}[(1-\lambda)I(U_{t-d} \leq \lambda)Y^I_{t-1} + \lambda I(U_{t-d} > \lambda)Y^{II}_{t-1}]\varepsilon_t \end{pmatrix} \tag{66}$$

Expression (66) is important since we will use to derive the asymptotic distribution of the tests when we include in the regression model state dependent constants.

Proof of Lemma 1. For the case in which $V(\delta_t) = 0$, under the null of $\phi = 0$ the DGP (21) became a Random Walk process $Y_t = Y_{t-1} + \varepsilon_t$.

To prove paragraph (1), we use the close form of the estimator of β in equation (62), for $b = 1$, that is

$$\Gamma_1(\hat{\beta} - \beta) = \begin{pmatrix} T^{-2}\sum_{t=1}^{T} Y^2_{t-1} & T^{-2}\sum_{t=1}^{T} Y^2_{t-1}H_t(\lambda) \\ T^{-2}\sum_{t=1}^{T} Y^2_{t-1}H_t(\lambda) & T^{-2}\sum_{t=1}^{T} Y^2_{t-1}H^2_t(\lambda) \end{pmatrix}^{-1} \begin{pmatrix} T^{-1}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \\ T^{-1}\sum_{t=1}^{T} Y_{t-1}H_t(\lambda)\varepsilon_t \end{pmatrix} \tag{67}$$

Note that since Y_t is a RW we have that

$$T^{-2}\sum_{t=1}^{T} Y^2_{t-1} \Rightarrow \sigma^2 \int_0^1 W^2(s)ds \tag{68}$$

by construction of $H_t(\lambda) = I(U_{t-d} \leq \lambda) - \lambda$ and (55) with (68) we know that

$$T^{-2}\sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) = T^{-2}\sum_{t=1}^{T} I(U_{t-d} \le \lambda) Y_{t-1}^2 - \lambda T^{-2}\sum_{t=1}^{T} Y_{t-1}^2 \to 0 \tag{69}$$

Finally from (55) and (68) we have

$$T^{-2}\sum_{t=1}^{T} Y_{t-1}^2 H_t^2(\lambda) \Rightarrow \sigma^2 \lambda(1-\lambda)\int_0^1 W^2(s)ds. \tag{70}$$

From (68), (69) and (70) the matrix

$$\begin{pmatrix} T^{-2}\sum_{t=1}^{T} Y_{t-1}^2 & T^{-2}\sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) \\ T^{-2}\sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) & T^{-2}\sum_{t=1}^{T} Y_{t-1}^2 H_t^2(\lambda) \end{pmatrix}^{-1} \Rightarrow \left(\sigma^2 \int_0^1 W^2(s)ds\right)^{-1} \begin{pmatrix} 1 & 0 \\ 0 & [\lambda(1-\lambda)]^{-1} \end{pmatrix} \tag{71}$$

From the usual unit root asymptotic we known that

$$T^{-1}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \Rightarrow \sigma^2 \frac{1}{2}[W(1)^2 - 1] \tag{72}$$

From (52) and (53)

$$T^{-1}\sum_{t=1}^{T} Y_{t-1} H_t(\lambda)\varepsilon_t \Rightarrow \sigma^2 \int_0^1 W(s)dV(s,\lambda), \tag{73}$$

where $V(s,\lambda)$ is a Kiefer-Muller process on $[0,1]^2$, then

$$\begin{pmatrix} T^{-1}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \\ T^{-1}\sum_{t=1}^{T} Y_{t-1} H_t(\lambda)\varepsilon_t \end{pmatrix} \Rightarrow \sigma^2 \begin{pmatrix} \frac{1}{2}[W(1)^2 - 1] \\ \int_0^1 W(s)dV(s,\lambda) \end{pmatrix} \tag{74}$$

Putting all together we have

$$\Gamma_1(\hat{\beta} - \beta) \Rightarrow \begin{pmatrix} \frac{\frac{1}{2}[W(1)^2-1]}{\int_0^1 W^2(s)ds} \\ \frac{\int_0^1 W(s)dV(s,\lambda)}{\lambda(1-\lambda)\int_0^1 W^2(s)ds} \end{pmatrix} \tag{75}$$

From (75) the distribution of the $t_{\phi=0}$ is the same as the Dickey-Fuller test, and is free of the threshold parameter λ.

The proof for paragraph (2) is done in the same way as the paragraph (1) by using the closed form of the estimator $\tilde{\beta}$.

$$\begin{pmatrix} T^{-2}\tilde{x}_1 & T^{-2}\tilde{x}_2 \\ T^{-2}\tilde{x}_3 & T^{-2}\tilde{x}_4 \end{pmatrix}^{-1} \Rightarrow \frac{1}{\left(\int_0^1 W(s)^2 ds - [\int_0^1 W(s)ds]^2\right)} \begin{pmatrix} 1 & 0 \\ 0 & [(1-\lambda)\lambda]^{-1} \end{pmatrix} \tag{76}$$

$$\begin{pmatrix} T^{-1}\sum_{t=1}^{T}\Big[I(U_{t-d}\leq\lambda)Y_{t-1}^{I}+I(U_{t-d}>\lambda)Y_{t-1}^{II}\Big]\varepsilon_t \\ T^{-1}\sum_{t=1}^{T}\Big[(1-\lambda)I(U_{t-d}\leq\lambda)Y_{t-1}^{I}-\lambda I(U_{t-d}>\lambda)Y_{t-1}^{II}\Big]\varepsilon_t \end{pmatrix} \Rightarrow \sigma^2 \begin{pmatrix} \int_0^1 W(s)dB(s)-W(1)\int_0^1 W(s)ds \\ \int_0^1 W(s)dV(s,\lambda)-V(1,\lambda)\int_0^1 W(s)ds \end{pmatrix} \tag{77}$$

Putting all together we have that:

$$\Gamma_1(\tilde{\beta}-\beta) \Rightarrow \begin{pmatrix} \frac{\int_0^1 W(s)dW(s)-W(1)\int_0^1 W(s)ds}{\left(\int_0^1 W(s)^2ds-[\int_0^1 W(s)ds]^2\right)} \\ \frac{\int_0^1 w(s)dV(s,\lambda)-V(1,\lambda)\int_0^1 W(s)ds}{\lambda(1-\lambda)\left(\int_0^1 W(s)^2ds-[\int_0^1 W(s)ds]^2\right)} \end{pmatrix} \tag{78}$$

This complete the proof of Lemma 1.

Proof of Lemma 2. The proof of this Lemma is straightforward, since the TARSUR process is covariance stationary, from equation (60) with $b=\frac{1}{2}$, we apply the ergodic stationary martingale differences central limit theorem.

Proof of Lemma 3 To show the convergence of $T^{-1/2}Y_{[Tq]} \Rightarrow J_{c_1,c_2}(q)$, first note that

$$\ln(\delta_t)=\ln(\rho_1 I(U_{t-d}\leq\lambda)+\rho_2 I(U_{t-d}>\lambda))=\ln(\rho_1)I(U_{t-d}\leq\lambda)+\ln(\rho_2)I(U_{t-d}>\lambda) \tag{79}$$

Let define $S_t=\sum_{i=1}^{t}\varepsilon_i$, from this sequence of partial sum construct.

$$X_T(q)=T^{-1/2}\sigma^{-1}S_{[Tq]}=T^{-1/2}\sigma^{-1}S_{j-1}, \quad \frac{j-1}{T}\leq q<\frac{j}{T} \tag{80}$$

we have that,

$$X_T(q)\Rightarrow W(q) \tag{81}$$

Iterating backward the TARSUR model (1) we have that:

$$Y_{[Tq]}=\sum_{i=1}^{[Tq]}\Big(\prod_{j=1}^{[Tq]-i}\delta_{[Tq]-j+1}\Big)\varepsilon_i+\Big(\prod_{j=1}^{[Tq]}\delta_j\Big)Y_0 \tag{82}$$

Taking logs and exponential in the product of δ_t

$$Y_{[Tq]}=\sum_{i=1}^{[Tq]}e^{\sum_{j=1}^{[Tq]-i}\ln(\delta_{t-j+1})}\varepsilon_i+\Big(\prod_{j=1}^{[Tq]}\delta_j\Big)Y_0 \tag{83}$$

by adding and subtracting inside the exponential $([Tq]-j)E(\ln(\delta_t))$ and reordering the terms

$$Y_{[Tq]}=\sum_{i=1}^{[Tq]}e^{([Tq]-i)E(\ln(\delta_t))}e^{\sum_{j=1}^{[Tq]-i}[\ln(\delta_{t-j+1})-E(\ln(\delta_t))]}\varepsilon_i+\Big(\prod_{j=1}^{[Tq]}\delta_j\Big)Y_0 \tag{84}$$

First focus on the term $e^{([Tq]-i)E(\ln(\delta_t))}$ in equation (84). From assumption (A.8) and (79) we have that

$$e^{([Tq]-i)E(\ln(\delta_t))} = e^{\frac{[Tq]-i}{T}[c_1\lambda + c_2(1-\lambda)]} = e^{\frac{[Tq]-i}{T}C} \tag{85}$$

where $C = [c_1\lambda + c_2(1-\lambda)]$

Second, focus on the term $e^{\sum_{j=1}^{[Tq]-i}[\ln(\delta_{[Tq]j+1}) - E(\ln(\delta_t))]}$, again from (79) we can write as follows

$$\begin{aligned} & e^{\sum_{j=1}^{[Tq]-i}[\ln(\rho_1)I(U_{[Tq]-d-j+1}\le\lambda) + \ln(\rho_2)I(U_{[Tq]-d-j+1}>\lambda) - \ln(\rho_1)\lambda - \ln(\rho_2)(1-\lambda)]} \\ &= e^{\sum_{j=1}^{[Tq]-i}[\ln(\rho_1)(I(U_{[Tq]-d-j+1}\le\lambda)-\lambda) + \ln(\rho_2)(I(U_{[Tq]-d-j+1}>\lambda)-(1-\lambda))]} \\ &= e^{(\ln(\rho_1)-\ln(\rho_2))\sum_{j=1}^{[Tq]-i}[I(U_{[Tq]-d-j+1}\le\lambda)-\lambda]} \end{aligned} \tag{86}$$

From assumption (A.8)we have that:

$$= e^{\frac{c_1-c_2}{T}\sum_{j=1}^{[Tq]-i}[I(U_{[Tq]-d-j+1}\le\lambda)-\lambda]} \tag{87}$$

Note that as $T \to \infty$

$$\frac{1}{[Tq]-i}\sum_{j=1}^{[Tq]-i}[I(U_{[Tq]-d-j+1} \le \lambda) - \lambda] \to_{a.s} 0 \tag{88}$$

such that expression (86) can be written as

$$= e^{(c_1-c_2)\frac{[Tq]-i}{T}o_{a.s}(1)} \tag{89}$$

From (85) and (87) we can rewrite (84) as follows

$$T^{-1/2}Y_{[Tq]} = \sum_{i=1}^{[Tq]} e^{\frac{[Tq-i]}{T}C + o_{a.s}(1)}\varepsilon_i + O(T^{-1/2}) \tag{90}$$

Then

$$\begin{aligned} T^{-1/2}Y_{[Tq]} &= \sigma\sum_{i=1}^{[Tq]} e^{\frac{[Tq-i]}{T}C + o_{a.s}(1)} \int_{\frac{i-1}{T}}^{\frac{i}{T}} dX_T(s) + O(T^{-1/2}) \\ &= \sigma\sum_{i=1}^{[Tq]} \int_{\frac{i-1}{T}}^{\frac{i}{T}} e^{\frac{[Tq-i]}{T}C + o_{a.s}(1)} dX_T(s) + O(T^{-1/2}) \\ &= \sigma\int_0^q e^{(q-s)C + o_{a.s}(1)} dX_T(s) + O(T^{-1/2}) \end{aligned} \tag{91}$$

We use integration by parts on the first term which is valid since $e^{(q-s)C}$ is continuous and $X_T(s)$ is increasing and of bounded variation. From (81) and the continuous mapping theorem as $T \to \infty$,

$$\sigma\{X_T(q) + (C + o_{a.s}(1))\int_0^q e^{(q-s)C+o_{a.s}(1)} X_T(s)ds\} + O(T^{-1/2}) \Rightarrow \sigma\{W(q) + C\int_0^q e^{(q-s)C} W(s)ds\} \quad (92)$$

The proofs of (b) and (c) are similar. To prove (d) we follow the results of Gonzalo and Pitarakis (2011). We have to show the strong approximation

$$\operatorname*{Sup}_{q\in[0,1]} \Big|\frac{Y_{[Tq]}}{\sqrt{T}} - J_{c_1,c_2}(q)\Big| = o_{a.s}(1) \quad (93)$$

Following the steps in Phillips (1998) lemma A.3 we use the Hungarian strong approximation to the partial sum process , $\sum_{i=1}^{t}\varepsilon_i$ and construct an expanded probability space that contains $\{\varepsilon_t, Y_t\}$ and the Brownian motion $W(.)$ for which the following strong approximation holds:

$$\operatorname*{Sup}_{q\in[0,1]} \Big|\frac{\sum_{i=1}^{[Tq]}\varepsilon_i}{\sqrt{T}} - W(q)\Big| = o_{a.s}(1) \quad (94)$$

Then

$$\begin{aligned} T^{-1/2}Y_{[Tq]} &= \sigma\int_0^q e^{(q-s)C+\{([Tq/T]-q)-(i/T-s)\}C+o_{a.s}(1)} dX_T(s) \\ &= \sigma\int_0^q e^{(q-s)C} dX_T(s)(1+o_{a.s}(1)) \\ &= \sigma\int_0^q e^{(q-s)C} dX_T(s) + o_{a.s}(1) \end{aligned} \quad (95)$$

since $e^{\{([Tq/T]-q)-(i/T-s)\}} = e^{O(T^{-1})} = 1 + o(1)$ uniformly in $q \in [0,1]$ and $s \in [\frac{j-1}{T}, \frac{j}{T}]$ uniformly over $j = 1, \cdots, T$. Since $e^{(q-s)C}$ is continuous and $X_T(s)$ is increasing and bounded variations we can integrate by parts (95)

$$T^{-1/2}Y_{[Tq]} = \sigma\{X_T(q) + C\int_0^q e^{(q-s)C} X_T(s)ds\} + o_{a.s}(1) \quad (96)$$

It follows from (93) and (96),

$$\begin{aligned} \operatorname*{Sup}_{q\in[0,1]}\Big|\frac{Y_{[Tq]}}{\sqrt{T}} - J_{c_1,c_2}(q)\Big| &\leq \operatorname*{Sup}_{q\in[0,1]}|X_T(q) - W_T(q)| \\ &\quad + \operatorname*{Sup}_{q\in[0,1]}\Big|\int_0^q e^{(q-s)C}\Big| \operatorname*{Sup}_{s\in[0,1]}|X_T(a) - W_T(a)| + o_{a.s}(1) = o_{a.s} \end{aligned} \quad (97)$$

The rest of the proof of part (d) follows from Gonzalo and Pitarakis. (e) follows identical lines to the proof of (d).

To prove (f) note that by squaring (1) and summing over t we have

$$T^{-1}Y_T^2 = 2c_1\frac{1}{T^2}\sum_{t=1}^{T} I(U_{t-d}\leq\lambda)Y_{t-1}^2 + 2c_2\frac{1}{T^2}\sum_{t=1}^{T} I(U_{t-d}>\lambda)Y_{t-1}^2 + 2\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t + O(T^{-1/2}) \tag{98}$$

From the strong law of large numbers for weakly dependent sequence $T^{-1}\sum\varepsilon_t \rightarrow \sigma^2$ almost surely. From (a) and (d) with the continuous mapping theorem, as $T\rightarrow\infty$,

$$2T^{-1}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \Rightarrow \sigma^2\{J_{c_1,c_2}(1)\}^2 - 2\sigma^2 C\int_0^1\{J_{c_1,c_2}(s)\}^2 ds - \sigma^2 = 2\sigma^2\int_0^1 J_{c_1,c_2}(s)dW(s) \tag{99}$$

The last inequality came form

$$\{J_{c_1,c_2}(1)\}^2 = 1 + 2C\int_0^1\{J_{c_1,c_2}(s)\}^2 ds + \int_0^1 J_{c_1,c_2}(s)dW(s) \tag{100}$$

Our result in (g) follows along the same lines as in Lemma 1 in Gonzalo and Pitarakis (2011) and Theorem 2 of Caner and Hansen (2001).

Proof of Proposition 4.

To prove the first part of Proposition 4, we use the close form of the estimators presented in (57) with $b=1$, and the results in Lemma 3 with the continuous mapping theorem. The second part of Proposition 4 is proven similarly, by using in this case equation (66).

Proof of Proposition 5.

For the cases where $E(\delta_t)<1$ the proof can be found in Gonzalez and Gonzalo (1997).

For the case where $E(\delta_t)=1$, under the null of $H_0:\gamma=0$, note that DGP (21) became a random walk process. For this case, whether the regression model does not have deterministic components or have state dependent constants is already proven in Lemma 1.

Case 1: Regression model (22) with $\mu_1=\mu_2=0$. From (82) we can see that

$$T^{-1}(\hat{\gamma}-\gamma) \Rightarrow \frac{\int_0^1 W(s)dV(s,\lambda)}{\lambda(1-\lambda)\int_0^1 W^2(s)ds} \tag{101}$$

From the continuous mapping theorem

$$t_{\gamma=0} \Rightarrow \frac{\int_0^1 W(s)dV(s,\lambda)}{\sqrt{\lambda(1-\lambda)\int_0^1 W^2(s)ds}} \tag{102}$$

Since $V(s,\lambda)$ and $B(s)\equiv B(s,1)$ are independent, it can be proved for a fixed λ,

$$\frac{\int_0^1 W(s)dV(s,\lambda)}{\sqrt{\int_0^1 W(s)^2 ds}} \equiv \mathcal{N}(0,\sigma_\lambda^2), \tag{103}$$

where $\sigma_\lambda^2 = V(H_t(\lambda)\varepsilon/\sigma) = \lambda(1-\lambda)$.

Case 2: Regression model (22) with state dependent constants. From equation (78) we have that:

$$T^{-1}(\tilde{\gamma}-\gamma) \Rightarrow \frac{\int_0^1 B(s)dV(s,\lambda) - V(1,\lambda)\int_0^1 B(s)ds}{\lambda(1-\lambda)\Big(\int_0^1 W(s)^2 ds - [\int_0^1 W(s)ds]^2\Big)} \equiv \frac{\int_0^1 W^*(s)dV(s,\lambda)}{\lambda(1-\lambda)\Big(\int_0^1 W^*(s)^2 ds\Big)} \tag{104}$$

where $W(.)^* = W(.) - \int_0^1 W(s)ds$. From the continuous mapping theorem we have that:

$$t_{\gamma=0} \Rightarrow \frac{\int_0^1 W^*(s)dV(s,\lambda)}{\sqrt{\lambda(1-\lambda)\int_0^1 W^*(s)^2 ds}} \tag{105}$$

Again note that $W^*(s)$ and $V(s,\lambda)$ are independent, we get the desired result.

Proof of Proposition 6. Since the threshold value is unknown and unidentified,the test statistic proposed is

$$\operatorname*{Sup}_{\lambda\in(0,1)} t_{\gamma=0}(\lambda)^2. \tag{106}$$

All the cases considered in Proposition 6 are examined in Proposition 5. Applying the continuous mapping theorem we have that

$$W_T \Rightarrow \operatorname*{Sup}_{\lambda\in(0,1)} t(\lambda)^2. \tag{107}$$

where $t(\lambda)$ is the asymptotic distribution of the $t-$statistic obtained in Proposition 5.

10 References

Andrews, D.W.K., (1993), "Test for parameter instability and structural change with unknown change point", *Econometrica*, 61, 821-856.

——, (2003), "Test for parameter instability and structural change with unknown change point: A corrigendum", *Econometrica*, 71, 395-3976.

Bai, J., and P. Perron (1998), "Estimating and testing linear models with multiple structural breaks", *Econometrica*, 66, 47-78.

Baillie, R.T. (1996), "Long memory processes and fractional integration in econometrics" *Journal of Econometrics,* 73, 5-59.

Basrak, B., R.A. Davis, T. Mikosch (2002), "Regular variation of GARCH processes", *Stochastic processes and their applications*, 99, 95-115.

Beran, J. (1994), *Statistics for long memory processes.* New York: Chapman & Hall.

Berg, A., T.L. McMurry, D.N. Politis (2010), "Subsampling p-values", Working paper.

Brandt, A. (1986), "The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients", *Advances in Applied Probability,* 18, 211-220.

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

Chan, K.S. (1989), "A note on geometric ergodicity of a Markov chain", *Advances in applied probability*, 21, 702-704.

Chan, K.S., H. Tong (1990), "On likelihood ratio test for threshold autoregression ", *Journal of the Royal Statistical Society. Series B (Methodological)*, 52, 469-476.

Chan, K.S (1993), "Consistency and limiting distribution of the least square estimator of a threshold autoregressive model", *The annals of statistics*, 21, 520-533.

Chen, N., R. Roll and S. Ross (1986), "Economics forces and the stock market", *Journal of Business*, 56, 383-403.

Chen, R., R.S. Tsay (1991),"On the ergodicity of TAR(1) process", *The annals of applied probability*, 1, 613-634.

Clark, T.E., K.D.West (2006) "Using out-of-sample mean square prediction errors to test the martingale difference hypothesis", *Journal of econometrics*, 135, 155-186

Cline, D.B.H. (2007), "Regular variation of order 1 nonlinear AR-ARCH models", *Stochastic process and their applications*, 117, 810-861.

Cramer, H. (1961), "On some classes of nonstationary stochastic processes", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol.2), Berkeley, University of California

Press, 57-78.

Distaso, W. (2008), “Testing for unit root process in random coefficient autoregressive models”, *Journal of econometrics*, 142, 581-609.

Dolado, J., J. Gonzalo and L. Mayoral (2002), “A Fractional Dickey-Fuller Test”, *Econometrica,* 70, 1963-2006.

Enders, W., and C.W.J. Granger (1998), “Unit root test and asymmetric adjustment with an example using the term structure of interest rates”, *Journal of Business* & *Economic Statistics*, 16, 304-311.

Engel, C. (1994), “Can the Markov switching model forecast exchange rate?”, *Journal of international economics*, 36, 151-165.

Estrella, A. (2003), “Critical values and P values of Bessel process distributions: Computation ans application to structural break tests”, *Econometric Theory*, 19, 1128-1143.

Fama, F. E. (1990), “Stock returns, expected returns, and real activity”, *Journal of Finance,* 45, 1089-1108.

Francq, C., S. Makarova, J.M. Zakoian (2008), “A class of stochastic unit-root bilinear process: Mixing properties and unit-root test”, *Journal of econometrics*, 142, 312-326.

Friedman, M. (1977), “Nobel lecture: Inflation and unemployment”, *Journal of political economy*, 85, 3, 451-472.

Giacomini, R., and H.White (2006), “Test of conditional predictive ability”, *Econometrica*, 74, 1545-1578.

Gonzalo, J., and T. Lee (1998), “Pitfalls in testing for long run relationships”, *Journal of Econometrics,* 86, 129-154.

Gonzalo, J., and J.Pitarakis (2002), “Estimation and model selection based inference in single and multiple threshold models”, *Journal of Econometrics*, 110, 319-352.

—– (2012), “Regime-specific predictability in predictive regression”, *Journal of business & economic statistics*, 30:2,229-241.

Gonzalo, J., and M. Wolf (2004), “Subsampling inference in threshold models”, *Journal of Econometrics*, forthcoming.

González, M., and J. Gonzalo (1998), “Threshold unit root models”, Working Paper, U. Carlos III de Madrid.

Gourieroux, G., C.Y. Robert (2006), “Stochastic unit root models”, *Economic theory*, 22, 1052-1090.

Granger, C., and J. Jouyeux (1980), “An introduction to long memory time series models and fractional differencing”, *Journal of Time Series Analysis,* 1, 15-39.

Granger, C.W.J, and Newbold, P. (1986), *Forecasting economic time series* (2nd ed.). Orlando, FL: Academic Press.

Granger, C., and N.R. Swanson (1997), "An introduction to stochastic unit-root processes", *Journal of Econometrics,* 80, 35-62.

Kapetianos, G. (2001), "Model selection in threshold models", *Journal of Time Series Analysis*, 22, 733-754.

Karlsen, H.A. (1990), "Existence of moments in a stationary difference equation", *Advances in Applied Probability,* 22, 129-146.

LeRoy, S.F. (1973), "Risk aversion and the martingale property of stock prices", *International Economic Review,* 14, 436-446.

Leybourne, S.J., B. McCabe and T. Mills (1996), "Randomized unit root processes for modeling and forecasting financial time series: Theory and Applications", *Journal of Forecasting*, 15, 253-270.

Leybourne, S.J., B. McCabe and A.R. Tremayne (1996), "Can economic time series be differenced to stationarity?", *Journal of Business and Economic Statistic,* 14, 435-446.

Lieberman, O., P.C.B. Phillips (2014), "A multivariate stochastic unit root model with an application to derivative pricing", Cowles foundation discussion paper No.1964.

Lucas, R.E. (1978), "Asset prices in an exchange economy", *Econometrica,* 49, 1426-1445.

McCabe,B., and A.R. Tremayne (1995), "Testing a time series for difference stationarity", *The Annals of Statistics*, 23, 1015-1028.

Meese, R.A., K. Rogoff (1983),"Empirical exchange rate models of the seventies: Do they fit out-of-sample?", *Journal of International Economics*, 14, 3-24.

Meyn, S.P., R.L. Tweedie (2005), "Markov chains and stochastic stability", Springer-Verlag, chapter 16.

Mills, T.C. (1993), *The econometric modeling of financial time series.* Cambridge University Press.

Nelson, R.C., and C.I. Plosser (1982), "Trends and random walks in macroeconomic time series: some evidence and implications", *Journal of Monetary Economics,* 10, 130-162.

Nicholls, D.F., and B.G. Quinn (1982), *Random coefficient autoregressive models: an introduction.* Lectures Notes in Statistics 11, New York: Springer-Verlag.

Petruccelli, J., and S. Woolford (1984), "A threshold AR(1) model", *Journal of Applied Probability*, 21, 270-286.

Phillips, P.C.B (1987), "Towards a unified asymptotic theory for autoregression ", *Biometrica*, 74, 535-547.

—— (1998), "New tools for understanding spurious regressions", *Econometrica*,56, 1299-1325.

Phillips, P.C.B., and Z. Xiao (1998),"A primer in unit root testing", *Journal of Economic Surveys,* 12, 423-470.

Pourahmadi, M. (1986), "On stationary of the solution of a doubly stochastic model", *Journal of Time*

Series Analysis, 7, 123-132.

——(1988), "Stationary of the solution of $X_t = A_t X_{t-1} + \varepsilon_t$ and analysis of non-gaussian dependent random variables", *Journal of Time Series Analysis,* 9, 225-239.

Robinson, P.M. (1994), "Time series with strong dependence", *Advances in Econometrics,* Sixth World Congress. Edited by C. Sims, Cambridge University Press.

Romano, J.P., M.Wolf (2001), "Subsampling intervals in autoregressive models with linear time trend", *Econometrica*, 69, 1283-1314.

Rossi, B. (2006), "Are exchange rates really random walks? some evidence robust to parameter instability", *Macroeconomic dynamics*, 10, 20-38.

Samuelson, P.A. (1965), "Proof that properly anticipated prices fluctuate randomly", *Industrial Management Review*, 6, 41-49.

Tjϕstheim, D. (1986), "Some doubly stochastic time series", *Journal of Time Series Analysis,* 7, 51-72.

Vervaat, W. (1979), "On a stochastic difference equation and a representation of non-negative infinitely divisible random variables", *Advances in Applied Probability,* 11, 750-783.

Yao, J.F., J.G. Attali (2000), "On stability of nonlinear AR process with Markov switching", *Advances in applied probability*, 32, 394-407.

Table 1: Empirical size of test for the mean, $E(\delta_t)$ and the variance, $V(\delta_y)$. Threshold parameter known and $P(U_{t-d} \leq \lambda) = 0.5$.

Coefficients	Dependence	T=200		T=500	
		$t_{\phi=0}$	$t_{\gamma=0}$	$t_{\phi=0}$	$t_{\gamma=0}$
$\rho_1 = \rho_2 = 1$ $(\lvert\gamma\rvert = 0)$	$p_{12} = 0.5$ (*i.i.d.*)	4.83	5.00	5.40	5.20
	$p_{12} = 0.7$	5.10	5.24	5.90	5.10
	$p_{12} = 0.9$	6.12	5.44	6.10	5.11

Table 2: Empirical size of test for the mean, $E(\delta_t)$ and empirical power of the test for the variance, $V(\delta_t)$. Threshold parameter known and $P(U_{t-d} \leq \lambda) = 0.5$.

Coefficients	Dependence	T=200		T=500	
		$t_{\phi=0}$	$t_{\gamma=0}$	$t_{\phi=0}$	$t_{\gamma=0}$
$\rho_1 = 0.99$ $\rho_2 = 1.01$ ($\lvert\gamma\rvert = 0.02$)	$p_{12} = 0.5$ $(i.i.d.)$	4.91	12.53	5.07	46.93
	$p_{12} = 0.7$	6.20	13.16	6.10	45.22
	$p_{12} = 0.9$	5.85	13.49	5.94	45.79
$\rho_1 = 0.98$ $\rho_2 = 1.02$ ($\lvert\gamma\rvert = 0.04$)	$p_{12} = 0.5$ $(i.i.d.)$	4.54	35.34	5.00	85.73
	$p_{12} = 0.7$	5.68	34.91	5.61	86.40
	$p_{12} = 0.9$	6.03	34.35	6.02	85.50
$\rho_1 = 0.95$ $\rho_2 = 1.05$ ($\lvert\gamma\rvert = 0.1$)	$p_{12} = 0.5$ $(i.i.d.)$	4.66	88.85	4.50	99.96
	$p_{12} = 0.7$	5.04	87.32	4.72	99.92
	$p_{12} = 0.9$	5.54	85.47	4.98	99.93
$\rho_1 = 0.9$ $\rho_2 = 1.1$ ($\lvert\gamma\rvert = 0.2$)	$p_{12} = 0.5$ $(i.i.d.)$	4.52	99.64	5.57	100.00
	$p_{12} = 0.7$	3.97	99.58	3.60	100.00
	$p_{12} = 0.9$	4.40	99.51	3.26	100.00

Table 3: Empirical size of test for the mean, $E(\delta_t)$ and empirical power of the test for the variance, $V(\delta_t)$. Threshold parameter known and $P(U_{t-d} \leq \lambda) = \lambda$.

Coefficients	λ	F	T=200		T=500	
			$t_{\phi=0}$	$t_{\gamma=0}$	$t_{\phi=0}$	$t_{\gamma=0}$
$\rho_1 = 0.985,\ \rho_2 = 1.01$	0.4	$\begin{pmatrix} 0.4 & 0.6 \\ 0.9 & 0.1 \end{pmatrix}$	5.46	17.39	5.33	58.16
$\rho_1 = 0.95,\ \rho_2 = 1.02$	0.28	$\begin{pmatrix} 0.8 & 0.2 \\ 0.5 & 0.5 \end{pmatrix}$	4.70	66.23	4.98	98.42
$\rho_1 = 0.99,\ \rho_2 = 1.03$	0.75	$\begin{pmatrix} 0.2 & 0.8 \\ 0.27 & 0.73 \end{pmatrix}$	5.29	28.47	5.16	78.85
$\rho_1 = 0.8,\ \rho_2 = 1.08$	0.8	$\begin{pmatrix} 0.4 & 0.6 \\ 0.15 & 0.85 \end{pmatrix}$	4.44	76.27	4.68	99.26
$\rho_1 = 0.95,\ \rho_2 = 1.02$	0.28	$\begin{pmatrix} 0.7 & 0.3 \\ 0.75 & 0.25 \end{pmatrix}$	5.16	63.46	4.99	98.40

Table 4: Empirical size of test for the mean, $E(\delta_t)$ and the variance, $V(\delta_y)$. Threshold parameter Unknown and $P(U_{t-d} \leq \lambda) = 0.5$.

Coefficients	Dependence	T=200		T=500	
		$t_{\phi=0}$	$t_{\gamma=0}$	$t_{\phi=0}$	$t_{\gamma=0}$
$\rho_1 = \rho_2 = 1$ $(\lvert\gamma\rvert = 0)$	$(i.i.d.)$	5.17	3.97	5.33	4.45
	$p_{01} = 0.7$	5.80	3.96	5.62	4.46
	$p_{01} = 0.9$	5.69	3.83	5.53	4.33

Table 5: Empirical size of test for the mean, $E(\delta_t)$ and empirical power of the test for the variance, $V(\delta_t)$. Threshold parameter Unknown and $P(U_{t-d} \leq \lambda) = 0.5$

Coefficients	Dependence	T=200		T=500	
		$t_{\phi=0}$	$t_{\gamma=0}$	$t_{\phi=0}$	$t_{\gamma=0}$
$\rho_1 = 0.99$ $\rho_2 = 1.01$ ($\lvert\gamma\rvert = 0.02$)	$(i.i.d.)$	5.39	8.28	5.11	33.28
	$p_{01} = 0.7$	5.35	7.99	5.32	32.85
	$p_{01} = 0.9$	5.73	8.42	5.58	32.86
$\rho_1 = 0.98$ $\rho_2 = 1.02$ ($\lvert\gamma\rvert = 0.04$)	$(i.i.d.)$	5.29	22.86	5.41	76.31
	$p_{01} = 0.7$	5.78	22.41	5.32	75.63
	$p_{01} = 0.9$	5.99	22.01	5.80	74.39
$\rho_1 = 0.95$ $\rho_2 = 1.05$ ($\lvert\gamma\rvert = 0.1$)	$(i.i.d.)$	5.38	78.85	4.92	99.76
	$p_{01} = 0.7$	5.60	76.35	5.19	99.67
	$p_{01} = 0.9$	6.05	73.89	5.22	99.65
$\rho_1 = 0.9$ $\rho_2 = 1.1$ ($\lvert\gamma\rvert = 0.2$)	$(i.i.d.)$	5.00	100.00	5.62	100.00
	$p_{01} = 0.7$	5.57	98.32	3.88	100.00
	$p_{01} = 0.9$	4.96	98.77	3.46	100.00

Table 6: Empirical size of test for the mean, $E(\delta_t)$ and empirical power of the test for the variance, $V(\delta_t)$. Threshold parameter Unknown and $P(U_{t-d} \leq \lambda) = \lambda$.

Coefficients	$P(r)$	F	T=200		T=500	
			$t_{\phi=0}$	$t_{\gamma=0}$	$t_{\phi=0}$	$t_{\gamma=0}$
$\rho_1 = 0.985,\ \rho_2 = 1.01$	0.4	$\begin{pmatrix} 0.4 & 0.6 \\ 0.9 & 0.1 \end{pmatrix}$	5.60	9.71	5.63	45.34
$\rho_1 = 0.95,\ \rho_2 = 1.02$	0.28	$\begin{pmatrix} 0.8 & 0.2 \\ 0.5 & 0.5 \end{pmatrix}$	5.20	53.26	5.80	96.18
$\rho_1 = 0.99,\ \rho_2 = 1.03$	0.75	$\begin{pmatrix} 0.2 & 0.8 \\ 0.27 & 0.73 \end{pmatrix}$	5.23	12.60	4.87	52.69
$\rho_1 = 0.8,\ \rho_2 = 1.08$	0.8	$\begin{pmatrix} 0.4 & 0.6 \\ 0.15 & 0.85 \end{pmatrix}$	4.17	42.14	3.91	90.21
$\rho_1 = 0.95,\ \rho_2 = 1.02$	0.28	$\begin{pmatrix} 0.7 & 0.3 \\ 0.75 & 0.25 \end{pmatrix}$	5.69	48.91	5.76	95.77

Table 7: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical size for the test of the variance $V(\delta_t)$. Assumed threshold parameter is known with *i.i.d.* threshold variable, $p_{12} = 0.5$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 1 - \frac{k}{T}$, $\rho_2 = 1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k=0$	5.11	5.43	5.10	5.01	4.90	5.09
$k=2$	6.00	6.80	5.20	7.07	7.35	5.14
$k=5$	10.72	12.81	5.59	11.48	11.83	4.95
$k=8$	19.68	23.80	5.72	21.01	22.89	5.16
$k=10$	26.66	33.37	5.12	29.94	32.71	5.11
$k=12$	35.36	44.68	5.24	39.49	44.95	4.99
$k=15$	49.61	64.31	5.44	54.72	62.77	5.38
$k=18$	63.09	80.45	5.15	69.65	78.54	5.27
$k=20$	70.47	87.26	5.30	77.19	86.89	5.06
$k=28$	89.99	99.30	5.58	94.83	99.05	4.91
$k=30$	92.80	99.73	5.08	97.12	99.59	5.34
$k=35$	96.98	99.97	5.29	98.95	99.94	4.75

Table 8: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical size for the test of the variance $V(\delta_t)$. Assumed threshold parameter is known with threshold variable generated by $p_{12} = 0.9$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 1 - \frac{k}{T}$, $\rho_2 = 1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k=0$	6.04	5.37	5.17	6.04	4.89	4.97
$k=2$	8.18	7.13	5.06	8.03	6.78	4.85
$k=5$	13.60	12.81	5.20	13.75	12.51	5.00
$k=8$	24.78	23.80	4.85	24.40	22.86	5.11
$k=10$	33.29	34.35	5.06	34.07	32.93	4.83
$k=12$	43.70	45.58	4.93	45.03	44.65	4.91
$k=15$	58.33	64.01	4.98	61.27	62.82	4.82
$k=18$	70.98	87.40	5.33	75.21	78.55	5.12
$k=20$	79.18	87.40	5.33	82.34	86.82	4.83
$k=28$	94.28	99.17	5.37	96.66	99.06	4.84
$k=30$	96.15	99.70	4.79	98.07	99.69	5.47
$k=35$	97.95	99.99	.37	99.32	99.98	4.67

Table 9: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is known with *i.i.d.* threshold variable, $p_{12} = 0.5$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.99 - \frac{k}{T}$, $\rho_2 = 1.01 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.02)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k=0$	4.88	5.26	12.81	4.55	4.57	47.32
$k=2$	6.07	6.76	10.23	6.63	6.87	34.94
$k=5$	10.62	12.74	8.967	11.44	11.76	25.71
$k=8$	19.75	23.85	8.08	21.38	22.96	19.91
$k=10$	26.95	34.32	7.04	29.71	32.95	17.28
$k=12$	35.45	45.59	6.89	39.53	44.81	16.02
$k=15$	49.80	64.05	6.80	55.12	63.22	14.15
$k=18$	62.35	79.53	6.70	69.73	78.79	12.44
$k=20$	71.13	88.16	6.05	77.48	86.95	11.91
$k=28$	90.16	99.29	6.24	94.97	98.98	10.33
$k=30$	92.90	99.73	5.87	96.70	99.67	9.91
$k=35$	96.62	99.99	5.88	99.01	99.98	9.10

Table 10: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is known with threshold variable generated by $p_{12} = 0.9$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.99 - \frac{k}{T}$, $\rho_2 = 1.01 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.02)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k=0$	6.17	5.66	13.14	5.73	5.06	46.64
$k=2$	7.91	6.87	9.64	7.87	7.03	34.19
$k=5$	13.81	12.97	8.35	13.73	12.62	24.76
$k=8$	24.51	24.01	7.05	24.47	23.16	20.25
$k=10$	33.57	34.34	6.74	34.26	.33.29	16.90
$k=12$	42.97	45.76	7.09	44.98	45.06	15.27
$k=15$	58.47	64.42	6.49	60.96	63.07	13.64
$k=18$	71.83	80.17	5.92	75.21	78.87	12.88
$k=20$	79.14	87.50	6.43	82.21	86.48	11.45
$k=28$	94.41	99.31	6.05	96.56	99.07	10.12
$k=30$	95.62	99.76	5.99	98.06	99.69	10.34
$k=35$	98.27	99.97	6.03	99.42	99.99	8.95

Table 11: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is known with $i.i.d.$ threshold variable, $p_{12} = 0.5$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.95 - \frac{k}{T}$, $\rho_2 = 1.05 - \frac{k}{T}$, ($\lvert\gamma\rvert = 0.1$)						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k=0$	4.65	5.06	88.36	4.76	4.67	99.95
$k=2$	7.04	7.17	80.31	7.74	7.25	99.87
$k=5$	11.75	13.92	68.75	14.25	13.58	99.65
$k=8$	20.41	23.91	58.26	23.99	24.80	99.36
$k=10$	27.77	35.33	51.45	32.52	34.12	98.93
$k=12$	37.14	45.87	46.42	43.27	46.58	88.36
$k=15$	49.88	64.47	40.54	57.32	63.00	96.96
$k=18$	63.66	80.24	36.14	72.01	79.72	95.44
$k=20$	71.22	88.31	34.42	78.19	86.55	94.19
$k=28$	90.57	99.34	26.36	95.34	98.96	88.53
$k=30$	93.38	99.67	25.95	96.76	99.50	86.58
$k=35$	96.43	99.99	23.19	98.89	99.97	82.11

Table 12: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is known with threshold variable generated by $p_{12} = 0.9$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.95 - \frac{k}{T}$, $\rho_2 = 1.05 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.1)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k = 0$	5.44	9.28	86.13	4.80	11.37	99.91
$k = 2$	7.32	8.43	78.65	7.01	10.80	99.85
$k = 5$	13.99	15.33	67.42	12.80	17.04	99.60
$k = 8$	23.62	25.89	57.32	22.82	29.42	99.18
$k = 10$	32.26	36.83	51.75	32.08	40.12	98.81
$k = 12$	41.72	48.03	46.09	42.24	52.34	98.26
$k = 15$	57.84	66.94	40.09	59.05	70.31	97.12
$k = 18$	70.96	82.15	35.75	72.59	84.51	95.34
$k = 20$	77.96	89.70	34.28	80.24	90.80	94.60
$k = 28$	93.74	99.42	27.29	96.79	99.68	88.36
$k = 30$	95.49	99.76	25.76	97.80	99.85	86.81
$k = 35$	.98.13	99.96	23.30	99.30	99.98	82.34

Table 13: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is known with *i.i.d.* threshold variable, $p_{12} = 0.5$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.9 - \frac{k}{T}$, $\rho_2 = 1.1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.2)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k=0$	4.49	5.83	99.66	5.92	7.66	1
$k=2$	8.48	8.34	99.21	10.85	10.46	1
$k=5$	13.48	15.09	98.29	21.00	18.86	1
$k=8$	23.14	26.23	96.23	32.26	30.15	1
$k=10$	30.98	36.67	95.09	41.76	40.10	1
$k=12$	38.90	47.53	93.56	50.04	50.94	1
$k=15$	53.42	65.36	89.63	64.38	66.85	1
$k=18$	66.39	80.26	86.96	75.86	79.83	99.99
$k=20$	73.20	87.70	84.32	82.15	87.02	99.99
$k=28$	90.61	99.12	74.90	95.21	98.46	99.98
$k=30$	93.50	99.60	72.16	96.90	99.33	99.98
$k=35$	96.65	99.99	67.42	99.05	99.88	99.97

Table 14: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is known with threshold variable generated by $p_{12} = 0.9$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.9 - \frac{k}{T}$, $\rho_2 = 1.1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.1)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k=0$	4.23	16.52	99.47	3.31	23.60	1
$k=2$	6.30	13.30	99.09	5.23	24.21	1
$k=5$	11.92	20.49	98.12	11.12	35.02	1
$k=8$	22.42	34.40	96.09	19.75	51.43	1
$k=10$	30.60	45.78	95.07	27.96	62.72	1
$k=12$	40.32	58.51	92.82	37.06	73.83	1
$k=15$	55.74	75.11	89.83	52.56	86.29	1
$k=18$	68.21	87.37	86.37	66.33	94.39	1
$k=20$	75.53	92.47	84.23	74.98	97.47	1
$k=28$	92.99	99.61	74.73	94.30	99.95	99.99
$k=30$	94.96	99.84	71.90	96.54	99.98	1
$k=35$	97.66	1.00	67.08	98.79	1	99.99

Table 15: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical size for the test of the variance $V(\delta_t)$. Assumed threshold parameter is Unknown with $i.i.d.$ threshold variable, $p_{12} = 0.5$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 1 - \frac{k}{T}$, $\rho_2 = 1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k=0$	5.48	5.53	4.10	5.03	4.84	3.79
$k=2$	6.86	7.06	3.72	6.96	7.04	4.19
$k=5$	12.43	13.28	4.13	12.49	12.70	4.36
$k=8$	20.58	23.73	4.00	21.83	23.33	4.17
$k=10$	27.89	33.33	3.8	30.23	33.23	4.13
$k=12$	37.15	45.76	3.58	40.02	44.63	4.03
$k=15$	51.40	64.48	3.70	55.92	62.25	4.30
$k=18$	63.38	79.76	4.06	69.09	79.16	4.06
$k=20$	70.98	87.82	3.67	77.39	87.41	4.25
$k=28$	90.18	99.37	3.75	95.19	99.24	4.11
$k=30$	92.46	99.65	3.50	96.97	99.57	4.08
$k=35$	96.67	99.96	3.95	98.88	99.96	4.25

Table 16: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical size for the test of the variance $V(\delta_t)$. Assumed threshold parameter is Unknown with threshold variable generated by $p_{12} = 0.9$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 1 - \frac{k}{T}$, $\rho_2 = 1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k = 0$	5.71	5.45	3.91	5.33	4.84	4.00
$k = 2$	7.34	7.19	3.90	7.86	7.04	4.04
$k = 5$	12.38	13.03	3.84	13.02	12.70	4.08
$k = 8$	21.26	23.06	3.69	23.18	23.33	4.39
$k = 10$	30.66	34.40	3.91	31.79	33.48	4.00
$k = 12$	39.06	45.59	3.91	41.94	44.85	4.06
$k = 15$	53.51	63.80	3.95	58.17	63.47	4.16
$k = 18$	66.75	80.86	3.58	72.03	78.25	4.44
$k = 20$	73.73	88.01	4.44	80.28	86.67	3.96
$k = 28$	92.00	99.33	3.93	95.98	99.10	3.94
$k = 30$	93.87	99.67	4.12	97.33	99.73	4.32
$k = 35$	97.04	99.96	3.94	99.11	99.96	4.01

Table 17: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is Unknown with $i.i.d.$ threshold variable, $p_{12} = 0.5$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.99 - \frac{k}{T}$, $\rho_2 = 1.01 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.02)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k = 0$	5.45	5.32	7.73	5.17	4.94	33.31
$k = 2$	6.59	6.92	6.75	7.29	6.96	23.23
$k = 5$	12.04	13.24	5.78	12.53	12.79	16.11
$k = 8$	20.77	23.84	4.96	21.60	23.44	12.64
$k = 10$	27.58	33.23	4.55	29.34	32.40	11.25
$k = 12$	36.33	45.64	4.82	40.60	44.62	10.18
$k = 15$	50.64	64.94	4.73	55.87	63.48	8.86
$k = 18$	63.40	79.83	4.82	70.02	78.40	7.93
$k = 20$	71.08	87.70	4.17	78.90	86.89	7.91
$k = 28$	89.86	99.11	4.25	94.92	99.18	6.99
$k = 30$	93.06	99.73	4.31	96.83	99.55	6.69
$k = 35$	96.31	99.96	3.90	98.81	99.95	6.18

Table 18: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is Unknown with threshold variable generated by $p_{12} = 0.9$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.99 - \frac{k}{T}$, $\rho_2 = 1.01 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.02)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k = 0$	5.82	5.49	8.00	6.08	5.78	31.71
$k = 2$	7.63	7.26	6.76	7.03	6.38	22.74
$k = 5$	12.69	13.00	5.78	12.48	12.18	16.37
$k = 8$	21.34	23.37	4.93	22.89	23.68	12.61
$k = 10$	30.43	34.71	4.61	30.88	33.00	11.32
$k = 12$	38.55	45.73	4.,75	42.61	45.67	10.36
$k = 15$	53.94	65.09	4.30	58.60	63.85	8.99
$k = 18$	66.86	80.16	4.43	72.04	78.53	8.62
$k = 20$	75.05	87.66	4.49	80.07	87.58	7.80
$k = 28$	91.55	99.10	4.48	95.88	99.25	6.89
$k = 30$	94.02	99.72	4.79	97.64	99.59	6.92
$k = 35$	97.24	99.96	4.01	98.99	99.95	6.01

Table 19: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is Unknown with $i.i.d.$ threshold variable, $p_{12} = 0.5$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.95 - \frac{k}{T}$, $\rho_2 = 1.05 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.1)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k = 0$	5.46	5.13	79.75	5.13	4.90	99.76
$k = 2$	7.33	7.25	67.11	7.97	7.48	99.45
$k = 5$	12.63	13.38	51.11	14.21	13.95	98.77
$k = 8$	21.52	24.77	40.39	24.22	24.97	97.76
$k = 10$	28.53	34.35	34.70	32.74	35.05	96.06
$k = 12$	38.51	46.81	30.40	42.30	46.66	94.93
$k = 15$	50.52	64.44	26.59	57.92	63.26	91.91
$k = 18$	63.77	79.86	22.26	71.27	78.14	88.70
$k = 20$	71.66	88.02	20.17	78.88	86.43	85.87
$k = 28$	90.15	99.27	16.01	95.16	98.92	75.61
$k = 30$	92.62	99.62	14.62	96.35	99.51	73.34
$k = 35$	96.46	99.95	13.78	98.69	99.93	67.48

Table 20: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is Unknown with threshold variable generated by $p_{12} = 0.9$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.95 - \frac{k}{T}$, $\rho_2 = 1.05 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.1)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k = 0$	6.27	9.43	74.89	5.32	10.99	99.67
$k = 2$	7.67	8.75	64.04	6.76	11.05	99.06
$k = 5$	12.99	14.75	49.89	13.00	17.44	98.58
$k = 8$	22.19	26.76	38.64	22.66	29.57	97.65
$k = 10$	30.01	36.63	34.93	31.65	39.97	95.96
$k = 12$	39.94	48.80	29.66	42.59	53.04	94.57
$k = 15$	54.08	67.32	25.22	57.73	69.33	91.44
$k = 18$	67.57	82.71	22.11	71.17	83.01	87.82
$k = 20$	74.08	89.62	20.39	79.30	90.32	85.41
$k = 28$	92.09	99.38	15.21	95.88	99.49	75.04
$k = 30$	94.03	99.75	14.48	97.31	99.79	71.25
$k = 35$	97.36	99.98	13.44	99.04	99.96	66.98

Table 21: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is Unknown with $i.i.d.$ threshold variable, $p_{12} = 0.5$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.9 - \frac{k}{T}$, $\rho_2 = 1.1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.2)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k=0$	4.99	6.15	99.01	5.79	7.60	100
$k=2$	9.10	7.91	98.00	10.35	10.31	100
$k=5$	14.52	14.88	95.58	20.92	19.11	100
$k=8$	23.39	26.48	91.24	32.05	29.56	100
$k=10$	30.41	36.64	87.49	41.13	39.76	100
$k=12$	40.20	47.97	84.53	50.27	50.28	100
$k=15$	53.03	65.80	78.58	63.31	65.80	100
$k=18$	65.61	80.07	72.97	75.00	80.41	99.99
$k=20$	72.61	87.68	68.75	81.97	86.45	99.99
$k=28$	90.16	98.95	56.08	95.64	98.62	99.84
$k=30$	93.07	99.65	53.60	97.14	99.38	99.87
$k=35$	96.60	99.99	47.75	98.95	99.96	99.78

Table 22: Local power of the test for the mean $E(\delta_t)$ and compared with respect the usual Dickey-Fuller test. Empirical power for the test of the variance $V(\delta_t)$. Assumed threshold parameter is Unknown with threshold variable generated by $p_{12} = 0.9$ and $P(U_{t-d} \leq \lambda) = 0.5$.

$\rho_1 = 0.9 - \frac{k}{T}$, $\rho_2 = 1.1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.2)$						
	T=200			T=500		
	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$	$t_{\phi=0}$	D-F tests	$t_{\gamma=0}$
$k = 0$	5.15	16.80	98.41	3.50	23.47	100
$k = 2$	7.10	14.11	96.94	6.01	24.76	100
$k = 5$	12.11	20.45	94.22	10.74	34.67	100
$k = 8$	22.43	34.61	89.52	20.19	50.95	100
$k = 10$	29.61	45.80	86.80	27.74	62.21	100
$k = 12$	39.08	58.08	82.14	37.52	73.44	100
$k = 15$	53.77	75.55	75.92	51.88	87.05	99.99
$k = 18$	66.10	87.52	71.00	66.36	94.28	100
$k = 20$	74.14	92.92	67.04	74.35	97.08	99.97
$k = 28$	91.57	99.76	55.95	94.21	99.97	99.98
$k = 30$	93.80	99.89	51.54	95.98	99.94	99.91
$k = 35$	96.94	99.98	46.69	98.48	99.99	99.63

Table 23: TARSUR model for US Stock Prices (1947:1 to 2016:4)

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\gamma}$	$\hat{\phi}$	$t_{\phi=0}$	$CV_{t_{\phi=0}}$	d	$\hat{r}$	W_T	CV_{W_t}
15.398	-0.893	-0.0466	-0.0088	-1.3983	-2.4307	1	78.71	13.76	8.86
(6.728)	(11.892)	(0.0125)	(0.0063)						

Table 24: SyP TARSUR regime roots

Z_{t-d}	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{P}(r)$	p_{22}	p_{12}
Δgdp_{t-d}	0.9761	1.0226	0.677	0.528	0.225

Table 25: TARSUR model for US real house prices (1961:1 to 2016:4)

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\gamma}$	$\hat{\phi}$	$t_{\phi=0}$	$CV_{t_{\phi=0}}$	d	$\hat{r}$	W_T	CV_{W_T}
3.374	-1.811	-0.049	0.003	0.551	-2.969	1	0.0028	16.556	8.86
(1.271)	(0.885)	(0.012)	(0.005)						

Table 26: House price TARSUR regime roots

Z_{t-d}	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{P}(r)$	p_{22}	p_{12}
$\Delta gdppc_{t-d}$	0.970	1.019	0.327	0.723	0.569

Table 27: TARSUR model for US interest rates (January 1949 to December 2016)

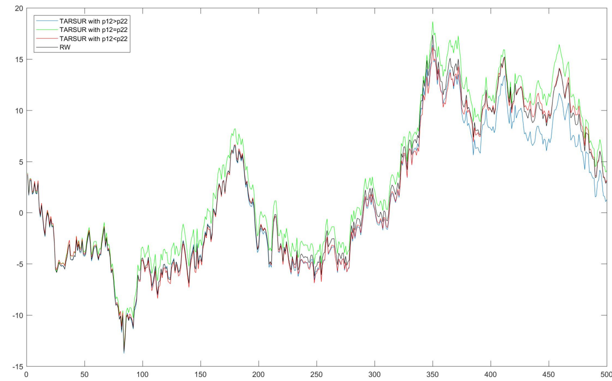
$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\gamma}$	$\hat{\phi}$	$t_{\phi=0}$	$CV_{t_{\phi=0}}$	d	$\hat{r}$	W_T	CV_{W_T}
0.012	0.029	0.038	-0.004	-0.844	-3.557	1	0.4	16.548	8.86
(0.023)	(0.042)	(0.009)	(0.005)						

Table 28: Interest rates TARSUR regime roots

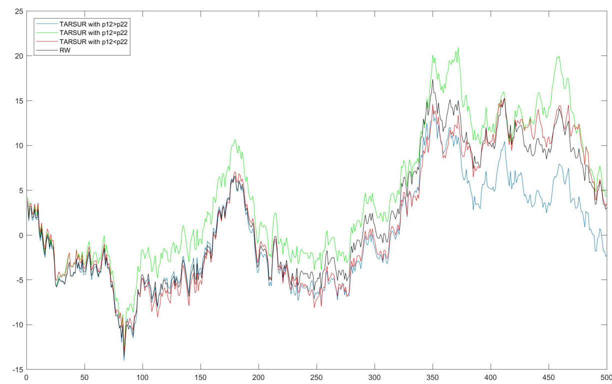
Z_{t-d}	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{P}(r)$	p_{22}	p_{12}
$\Delta gdppc_{t-d}$	1.006	0.968	0.74	0.920	0.028

Table 29: TARSUR model for USD/Pound (January 1978 to December 2016)

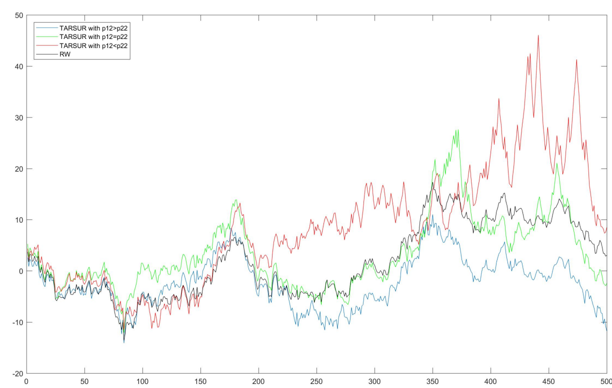
$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\gamma}$	$\hat{\phi}$	$t_{\phi=0}$	$CV_{t_{\phi=0}}$	d	$\hat{r}$	W_T	CV_{W_T}
-0.013	0.058	0.047	-0.015	-1.787	-2.90	1	0	7.78	8.86
(0.020)	(0.019)	(0.017)	(0.008)						



(a)



(b)



(c)

Figure 1: Random Walk versus different TARSUR series. Each figure differs by $V(\delta_t)$.

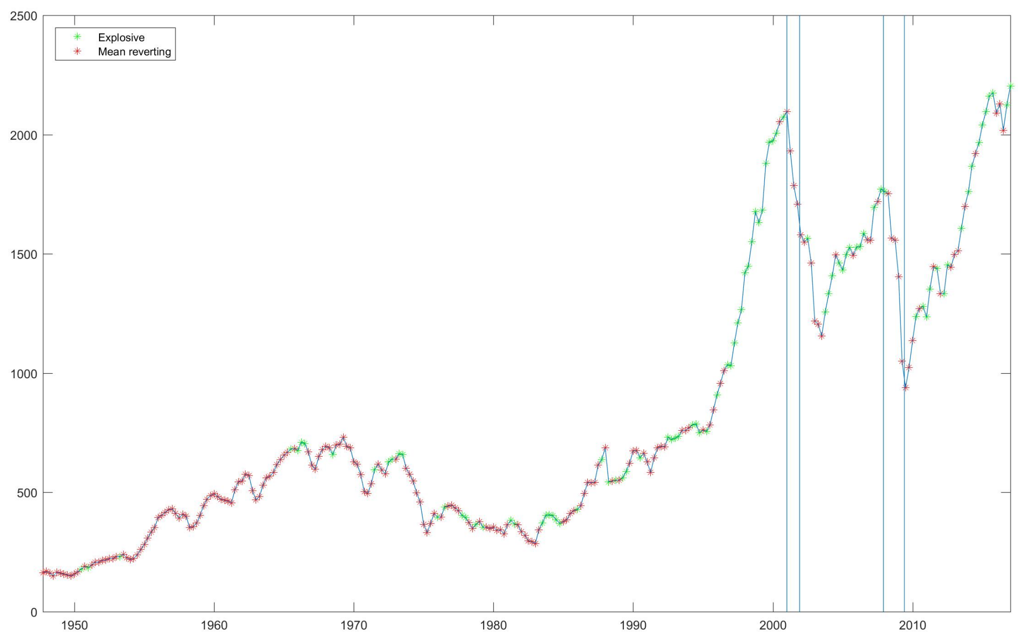


Figure 2: US stock prices and the regimes selected by TARSUR model, $1947:1-2016:4$.

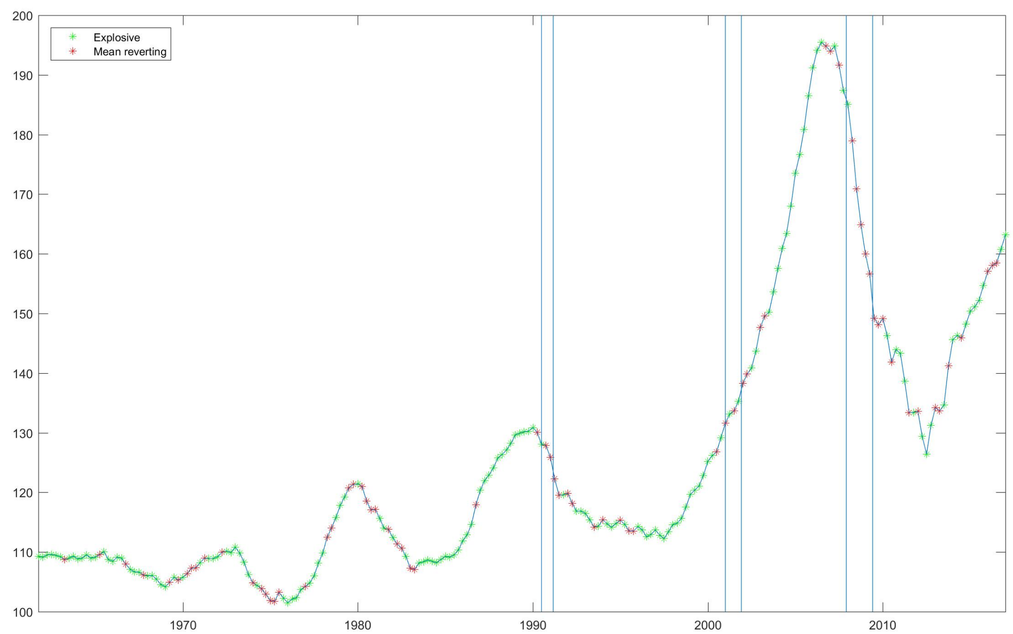


Figure 3: US house prices and the regimes selected by TARSUR model, $1961:1-2016:4$.

Figure 4: US interest rates prices and the regimes selected by TARSUR model, January 1949-December 2016.