Essays on Threshold Models with Unit Roots

Jun Yi Peng Zhou

A dissertation submitted by in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Economics

Universidad Carlos III de Madrid

Advisor:
Jesús Gonzalo Muñoz

September 2019

Para mis padres WangZhong y ZhiHua, mi hermana Lili
y a mis adorables sobrinas Ana y Beatriz.

Acknowledgements

Doing research is not an easy task and learn how to do exciting research involves struggles, sweat, and tears. This dissertation means the light at the end of this tunnel. Also means the beginning of a new journey in my life. During all the years I spend in UC3M, I met a lot of great people who helped me to be the best of myself, always moving forward and never give up.

First of all, I must thank my advisor, Jesús Gonzalo for his guidance during my Ph.D. He opened my eyes and showed me the right path to be a good and independent researcher. Your advice and guidance will influence me through years to come. I have to thank you for believing in me, on my work and all the time we spend together teaching me the different aspects of academic research. I am very grateful and proud of having you as my advisor.

I am very thankful to the faculty members, the administrative staff, and my fellow Ph.D. college at UC3M. Special thanks to Nazarii, Carlos, Alvaro, Juan Carlos, and Miguel. Your recommendation and comments helped me to improve my work significantly. To Arancha, Angelica, Laura, and Mercedes, the invisible hands that ensure everything works smoothly. Thanks for helping me to navigate through all the complex bureaucratic matters. In all these years, the economic department was my second home with a big family and a lot of great moments. I will never forget these experiences.

I cannot give enough thanks to my Ph.D. colleges, Yuhao, Rui, Beatriz, Tomas, Ursula, Federico Masera, Federico Curci, Meng, Victor, Cleiton, Mianhuang, Jose, Hassin, Salvatore, Miguel, Minghai, Weifeng, Pedro, and Mehdi which accompanied and supported me during this long journey.

I must give special thanks to my parents, which from the day I born, they took care of me, and pushed me to walk in the right path. For all the hardship you suffer the last thirty-five year to carry ahead the family and allow me to have a proper education in Spain. Finally, my dear sister, who I always tease her and stands all my bad jokes. Thank you for everything.

Published and Submitted Content

The first chapter, titled "Threshold Stochastic Unit Root Models," is a project co-authored with Jesús Gonzalo and Raquel Montesinos. Both co-authors started the research of stochastic unit root models with threshold structures. In their initial study, they analyze the stationary properties of the threshold autoregressive stochastic unit root model (TARSUR), propose a test to check the presence of a threshold effect, and analyzes some empirical applications of this model. Their proposed test is not enough to conclude that there is a TARSUR model. For that, we need to test the expected value of the largest AR root to be equal to one. My contributions to this paper are several. First, I show how the TARSUR models can arise in economic theory. Second, I present a testing procedure to asses if the expected value of the time-varying parameter is equal to one and show the asymptotic distribution of the proposed test. The test statistic contains a distribution discontinuity. In order to solve this problem, I extended the asymptotic results used in the near unit root literature to our case. Also, I show that the TARSUR process is geometrically ergodic, a condition needed to confirm the validity of sub-sampling methods used to obtain the critical value for the proposed test. Finally, I extend the number of empirical applications and study the presence of TARSUR structure in U.S. house prices, U.S. interest rates, and U.S. dollar to British pound exchange rate. To the best of our knowledge, the initial version of this study is available in different sources, and it is available here: Source 1.1, Source 1.2, and Source 1.3.

I presented the second chapter, "Multiple Long-run Equilibria Through Cointegration Eyes," at the 43rd Simposio de la Asociación Española de Economía celebrated at UC3M. To be eligible to attend this conference, I submitted a preliminary draft which is available in the following link: Source 2.1.

Other Research Merits

This thesis has been possible thanks to the financial support by the grant BES-2015- 072214 from Spanish Ministry of Economy and Business.

Abstract

Empirical Time Series is too linear. After the 2008 great depression, the board members of the central banks realize that they were unable to foresee the financial meltdown until it was too late, due to the linear structure of the models used for the forecasts, claiming the need for non-linear models. The field of non-linear time series model is too vast, and sometimes these models are too complex to use them for forecasting. Furthermore, most of the economic variables are persistent, viewed as unit roots, adding an extra level of difficulty in the study of non-linear time series models. The challenge is to develop non-linear models with persistent variables.

Threshold models are a class of non-linear model characterized by different regimes, determined by a threshold variable. These regimes can represent the different stages of the economic cycles, for example, economic expansions and recessions, periods with high volatility and low volatility in the stock market, among many other examples. Many of the advantages of the threshold models are the simplicity of estimation using least square estimation, interpretation of the non-linear structure, and testing.

In this dissertation, we study threshold models with unit roots from two different perspectives. In one had we introduce a univariate analysis and on the other hand, a multivariate analysis.

In the first chapter, titled *"Threshold Stochastic Unit Roots Models"* co-authored with Jesús Gonzalo and Raquel Montesinos, we present the univariate analysis by introducing a new class of stochastic unit-root ($STUR$) processes. This new model, namely the threshold autoregressive stochastic unit root ($TARSUR$) process, is strictly stationary, but if we do not consider the threshold effect, it can mislead to conclude that the process has a unit root. The $TARSUR$ models are not only an alternative to fixed unit root models but present interpretation, estimation, and testing advantages to the existent $STUR$ models.

This study analyzes the properties of the $TARSUR$ models and proposes two simple tests to identify this type of processes. The first test will allow us to detect the presence of unit roots, which can be fixed or stochastic, and the asymptotic distribution (AD) of this test presents a distribution discontinuity depending if the unit root is fixed or stochastic. The

second test we propose is 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 its 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 Bridge. A Monte Carlo simulation shows that the proposed tests behave very well in a finite sample, and the Dickey-Fuller test cannot easily distinguish between exact unit roots and threshold stochastic unit roots. The study concludes with applications to U.S. stock prices, U.S. house prices, U.S. interest rates, and USD/Pound exchange rates.

The second chapter, we present the multivariate analysis with *"Multiple Long Run Equilibria Through Cointegration Eyes"*. In this chapter, we introduce threshold effects in the cointegration relation. Cointegration has succeeded in capturing the unique long-run linear equilibrium. Specific non-linearities have been incorporated into cointegrated models but always assuming the existence of a single equilibrium. In this study, we explore the possibility of different long-run equilibria depending on the state of the world (i.e., good and bad times, optimism and pessimism, frictional coordination) in a threshold framework. Starting from the present-value model (PVM) with different discount factors and depending on the state of the economy, we show that this type of PVM implies threshold cointegrated with different long-run equilibria. We present the estimation and inference theory, and the study finishes with two empirical applications where the variables are not linearly cointegrated but threshold cointegrated.

The third chapter, we continue in the multivariate framework and introduce the paper titled *"Quasi-Error Correction Model"*. Cointegration captures single long-run equilibrium relationships between economic variables and the error correction model (ECM) is the mechanism in which the equilibrium is maintained. In this study, we introduce the quasi-error correction model (QECM), derived from the cointegration relation with threshold effects, where each regime represents a different equilibrium relation between the variables. In contrast to the linear ECM, the QECM has a regressor which captures the switching between equilibria, capturing the dynamic structure of the equilibrium change. This regressor will pose a problem similar to the non-linear error correction models, where the model cannot be balanced using the traditional definitions of integration. We present the estimation and the inference theory and finish with an empirical application for U.S. interest rate of instruments with different maturities.

Contents

List of Tables

List of Figures

Chapter 1

Threshold Stochastic Unit Root Models

1.1 Introduction

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

In order to avoid the tight constraints that an exact unit root imposes on a process, and to be able to generate more flexible and realistic models, research has recently evolved in two directions. The first line of research generalizes UR models by allowing for fractional roots: $ARFIMA$ models (see Granger and Joujeux (1980), Beran (1994), Robinson (1994), Baillie (1966), Dolado, Gonzalo and Mayoral (2002)) The second one makes the UR model more flexible by allowing the unit root to 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)) instead of a fixed parameter. With both extensions, a more general form of non-stationary are allowed than those implied by the standard exact unit root autoregressive models. This study forms part of the second line of research.

Stochastic unit root models ($STUR$) arise naturally in economic theories, as well as in many macroeconomic and financial applications (see Leybourne, McCabe and Mills (1996), Granger and Swanson (1997) and Lieberman and Phillips (2014)).The $STUR$ 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 non-stationarity, which is not removable by differencing, it is inappropriate to estimate the conventional

ARIMA or cointegration/error-correction models because the properties of the estimators and the tests involved are not the same as those in the standard difference stationary case. For instance, two series generated by two independent *STUR* models will be wrongly detected to be cointegrated according to some of the most used cointegration tests (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 statistical distinction between these two types 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, as opposed to what has been 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 point of view, a suitable justification for using time varying parameter models to approximate or represent non-stationary processes are provided by Cramer's (1961) extension of the Wold 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 a *ARMA* process with coefficients that can 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 we might assume 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 again it has the inconvenience that it occurs regularly at every moment in time. In this study it is assumed that the economy stays in a "good" or "bad" state for several periods of time until certain determining variables 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 it is equal to one. This threshold autoregressive stochastic unit root (*TARSUR*) model presents several advantages with respect to 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 is used to test the hypothesis of non-threshold effects versus threshold effects, in some cases it follows asymptotically a standard distribution and, therefore, there is no need to generate new critical values. Third, the threshold variable is suggested by economic theory and it will likely provide an explanation or cause for the existence of a unit root, which to the best of our knowledge is still lacking in the econometric literature. Fourth, in many situations, 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 study is organized as follows. In Section 1.2, we present economic conditions when asset prices follow a TARSUR process. In section 1.3, we define the *TARSUR* model and examine its properties: strict stationarity, covariance stationarity, geometric ergodicity and impulse response function. In Section 1.4, we present two different tests for identifying this type of process, the first one checks the presence of unit roots which can be either fixed or stochastic, and the second test checks for the presence of threshold effects. The asymptotic distribution of this test is developed under two different situations: when the threshold value is known and when the threshold value is unknown and unidentified. Section 1.5 analyzes the finite sample performance (size and power) of the tests developed in this study . Section 1.6 briefly discusses some practical issues present in all the threshold models. Section 1.7 shows four empirical applications of the proposed model: U.S. stock prices, U.S. house prices, U.S. interest rates, and U.S./Pound exchange rates. Finally, Section 1.8 provides the concluding remarks. Proofs are provided in the Appendix.

1.2 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:

$$p_t U_t' = E\left[(1+\rho)^{-1}(p_{t+1}+d_t)U_{t+1}\middle|\mathbf{F}_t\right] \tag{1.1}$$

, where the information set $\mathbf{F}_t$ contains all the past and current information available, p_t is the stock price at time t, d_t is the dividend, 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 (1.1) 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 that can be derived from an inter-temporal optimization framework. Assume a two-period lived representative agent at time t, which maximize the 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 depends on the perception of the individual about the state of the world (expansion and recession, high or low unemployment). The individual has the opportunity to buy h_t amount of a risky assets at the beginning of period t at 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 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:

$$p_t U_t' = E\left[(1 + \rho(z_t))^{-1}(p_{t+1} + d_t)U_{t+1}\middle|\mathbf{F}_t\right], \tag{1.2}$$

where $\rho(z_t)$ is the state-dependent discount factor. Following the work of Samuelson (1973), under risk neutrality, and assuming that the state-dependent discount rate can be represented as a 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 $\tilde{\delta}(z_t) = \tilde{\rho}_1 I(z_t \le r) + \tilde{\rho}_2 I(z_t > r)$ have a threshold structure, we can get the TARSUR process:

$$E(p_{t+1}|\mathbf{F}_t) = (1 + \tilde{\delta}(z_t))p_t = \delta(z_t)p_t, \tag{1.3}$$

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

$$p_{t+1} = \delta(z_t) p_t + \varepsilon_t \tag{1.4}$$

1.3 *TARSUR* model

Consider the following threshold first order autoregressive (TAR) model

$$\begin{aligned} Y_t &= [\rho_1 I(Z_{t-d} \leq r_1) + \cdots + \rho_n I(Z_{t-d} > r_{n-1})] Y_{t-1} + \varepsilon_t = \\ &= \delta_t Y_{t-1} + \varepsilon_t, \qquad t = 1, \cdots, T \end{aligned} \tag{1.5}$$

, 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 study, it 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 (1.5) 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 considered 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, and adapted to the *sigma-field* $\Im_t \stackrel{def}{=} \{(\varepsilon_j, Z_j), j \leq t\}$.

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

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

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

conditional distribution of ε_t. (A.1), (A.2), (A.3), (A.4), and (A.8) are used to obtain the asymptotic distributions of the statistics proposed in this study. (A.3) is the most restrictive assumption but is essential for inference purpose. We need it to prove the tightness of a 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) restricts the autoregressive coefficients for different regimes move around unity. This assumption is required to solve the asymptotic distribution discontinuity in one of the tests, proposed in this study to identify these types of models.

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 constraints). This is because if any of the parameters ρ_i is greater 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.

1.3.1 Stationary Properties

Equation (1.5) represents a specific 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 (1.5),

$$\begin{aligned} Y_t &= \varepsilon_t + \sum_{j=1}^{m-1}\left(\prod_{i=0}^{j-1}\delta_{t-i}\right)\varepsilon_{t-j} + \left(\prod_{i=0}^{m-1}\delta_{t-i}\right)Y_{t-m} \\ &= C_{1,t}(m) + C_{2,t}(m), \end{aligned} \tag{1.6}$$

, 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 $C_{2,t}(m) = \left(\prod_{i=0}^{m-1}\delta_{t-i}\right)Y_{t-m}$. The following results are obtained from (1.5) and (1.6):

(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 (1.5).

(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.

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

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) assume ($\{\delta_t, \varepsilon_t\}$) to be *i.i.d.* and mutually independent. Pourahmadi (1986, 1988) and Tjϕstheim (1986) allow δ_t to be a 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*

$$-\infty < E\log|\delta_1| < 0 \tag{1.7}$$

holds, then the process (1.5) is strictly stationary. Moreover, if (A.7) is satisfied and

$$\sum_{j=0}^{\infty}\left(E|\psi_{t,j}|^2\right)^{\frac{1}{2}} < \infty, \tag{1.8}$$

with $\psi_{t,0} = 1$ and $\psi_{t,j} = \prod_{i=0}^{j-1}\delta_i$ for $j \geq 1$, then the process (1.5) is second-order stationary.

Theorem 1 provides sufficient conditions for (a) and (b) to hold when $p = 0, 1,$ or 2. It shows 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 (1.7) 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 (1.5) (see Brandt (1986)). The mean square convergence of $\{\psi_{t,j}\}_{j=0}^{\infty}$ is obtained provided condition (1.8) holds, and in this case, process (1.5) is also second-order stationary.

Note that there is a trade-off between (A.7) and (1.8). For instance, assumption (A.7) can be relaxed by imposing $||\varepsilon||_p < \infty$, $\forall p < \infty$; but in this case, we need to modify (1.8) requiring a stronger condition

$$\sum_{j=0}^{\infty}\left(E|\psi_{t,j}|^{2+\kappa}\right)^{\frac{1}{2+\kappa}} < \infty, \ for\ a\ \kappa > 0. \tag{1.9}$$

Also, as mentioned before, it is assumed that $\{\varepsilon_t\}$ and $\{Z_t\}$ are mutually independent with $||\varepsilon_1||_p < \infty$, $\forall p < \infty$, then condition (1.8) is a 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, the model with ρ_i values around unity: stationary for some regimes and mildly explosive for others. Notice that The fixed unit root models are not stationary, but if we allow the root to be stochastic around unity we can archieve strict stationarity.

Theorem 1 produces explicit conditions for strict stationarity. However, no moments need to exist and, to the best of our knowledge, they are not explicit conditions for second-order stationarity; therefore, each 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; however, nothing is gained on the understanding of the process and the algebra becomes very tedious.

Sufficient conditions for second-order stationarity are presented in the following proposition.

Proposition 1. *Suppose* $\{Y_t\}$ *is generated by (1.5) 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 becomes the necessary and sufficient condition established by Nicholls and Quinn (1982) for second-order stationarity in random coefficient autoregressive models (RCA):

$$\rho(F_2) < 1 \Longleftrightarrow E(\delta_t^2) = \rho_1^2 p_1 + \rho_2^2 p_2 < 1 \tag{1.10}$$

From this inequality, it can be concluded 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 stationarity 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 the 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. Overall, 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 must be smaller than the probability of the changing regimes. The idea underlying this condition is to avoid staying in the explosive regime for too long.

1.3.2 Geometric Ergodicity

From the works of Chan (1993), geometric ergodicity is required to obtain consistency for the estimator of the threshold value ($\hat{r}$), in the case where the TARSUR process is covariance stationary. Also, this condition is needed to apply sub-sampling latter on to obtain valid critical values for one of the proposed tests.

Finding conditions for $\{Y_t\}$ to be geometrically ergodic have been studied by several authors. Chan and Tong (1985), Chan (1989), and Chen and Tsay (1991) give conditions on the coefficients for the self-exciting threshold autoregressive models. Gonzalo and Gonzalez (1998) and Gourieroux and Robert (2006) show 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, Makarova and Zakoian (2008) show geometric ergodicity for the stochastic unit root process assuming that the sequence $\{\delta_t\}$ and $\{\varepsilon_t\}$ are independent and identically distributed. More general conditions are given in the following result based on the works 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 a transition matrix F and invariant measure η, then if:*

$$\mathbb{E}(log(\delta_t)) = \eta_1 log(\rho_1) + \eta_2 log(\rho_2) + \cdots + \eta_n log(\rho_n) < 0 \tag{1.11}$$

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

Note that here we show V-uniform ergodicity, which implies geometrical ergodicity (Meyn and Tweedie, 2005 Chapter 16).

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

Corollary 2 follows from Theorem 2, which establishes a sufficient condition for the *TARSUR* process to be geometrically ergodic. Note that the exact unit root processes are not ergodic, but if we allow the root to be stochastic and vary around unity and impose conditions on the behavior of the stochastic unit root, we can archive a stronger form of geometrical ergodicity.

1.3.3 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:

$$Y_t = \varepsilon_t + \sum_{j=1}^{\infty} \left(\prod_{i=0}^{j-1} \delta_{t-i} \right) \varepsilon_{t-j} = \sum_{j=0}^{\infty} \psi_{t,j} \varepsilon_{t-j}. \tag{1.12}$$

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 stochastic now. For this reason, the impulse response function (IRF) is defined as:

$$\xi_h = E\left(\frac{\partial Y_{t+h}}{\varepsilon_t} \right) = E(\psi_{t,h}) = E\left(\prod_{i=0}^{h-1} \delta_{t-i} \right), \;\; h = 1, 2, \cdots , \tag{1.13}$$

Proposition 2. *Under Proposition 1 conditions, the IRF of the process $\{Y_t\}$ defined by (1.5) is given as*

$$\xi_h = \begin{pmatrix} 1 & 1 \end{pmatrix} F_1^h \begin{pmatrix} \rho_1 p_1 \\ \rho_2 p_2 \end{pmatrix}, \;\; h = 1, 2, \cdots , \tag{1.14}$$

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 the $TARSUR$ 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 and Proposition 2 show that the $TARSUR$ models are more flexible than fixed unit roots, specifically in the sense of being able to produce a rich set of plausible scenarios. If $p_{22} \geq p_{12}$ the process is not covariance stationary and shocks have permanent effects and even increasing effects on the mean; but if $p_{22} < p_{12}$, shocks have only transitory effects on the 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 1.1, a-c, displays simulated realizations from *TARSUR* and random walk (RW) models. The *TARSUR* series are generated by model (1.5) with two regimes , ε_t is an *i.i.d.* *Normal* $(0,1)$. The random walk series is generated from the same set of innovations. The first 50 observations of each series have been disregarded to avoid any initial conditional dependency. For comparison, each figure shows a random walk versus three different types of TARSUR processes: $p_{22} > p_{12}$, $p_{22} = p_{12}$, and $p_{22} < p_{12}$. Each figure differs by the value of the variance of the stochastic unit root coefficient. More specifically, in figure 1.1a, ρ_1=0.99 and $\rho_2 = 1.01$ ($V(\delta_t) = 0.0001$), in figure 1.1b, ρ_1=0.97 and $\rho_2 = 1.03$ ($V(\delta_t) = 0.0009$), and in figure 1.1c 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 become more volatile than the corresponding RW. It is worth mentioning that even in the most unstable case (see figure 1.1c), the "explosive" *TARSUR* series ($p_{22} > p_{12}$) does not look like a standard $AR(1)$ with a fixed explosive root.

[Figure 1.1 enters here]

1.3.4 Differencing a TARSUR process

By differencing model (1.5), we obtain

$$\Delta Y_t = (\delta_t - 1)Y_{t-1} + \varepsilon_t \tag{1.15}$$

Proposition 3. *Assume that $\{Y_t\}$ is generated by model (1.5). 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 (1.15) as a $MA(\infty)$

$$\Delta Y_t = \sum_{j=0}^{\infty} \Psi_{t,j}\varepsilon_{t-j} \tag{1.16}$$

, where $\Psi_{t,0} = 1$ and $\Psi_{t,j} = (\delta_t - 1)\psi_{t-1,j-1}$, $j \geq 1$. From (1.16) the IRF of $\{\Delta Y_t\}$ can be easily obtained.

1.4 Testing for TARSUR

Since the *TARSUR* model requires 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 tests, on one hand, we test the null of $E(\delta_t) = 1$ without any knowledge about $V(\delta_t)$; on 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), from (A.4) we can replace the threshold variable with a uniform distributed variable using the following equality:

$$I(Z_{t-d} \leq r) = I(P(Z_{t-d}) \leq P(r)) = I(U_{t-d} \leq \lambda), \tag{1.17}$$

where $P(.)$ is the marginal distribution of $\{Z_t\}$, and U_{t-d} denotes a uniformly distributed random variable on $[0, 1]$ and $\lambda = P(r)$. Using the suggested transformation, we can rewrite the *TARSUR* process defined in (1.5) as follows:

$$Y_t = \rho_1 I(U_{t-d} \leq \lambda)Y_{t-1} + \rho_2 I(U_{t-d} > \lambda)Y_{t-1} + \varepsilon_t \tag{1.18}$$

Since our objective is to test the conditions $E(\delta_t) = 1$ and $V(\delta_t) > 0$, it is important to rewrite the above model in such a way that these two conditions are expressed in terms of parameters in an equivalent regression. To do this, we add and subtract on the right hand side of model (1.18), $E(\delta_t)Y_{t-1} = [\rho_1\lambda + \rho_2(1-\lambda)]Y_{t-1}$, then we can rewrite the *TARSUR* model as:

$$Y_t = E(\delta_t)Y_{t-1} + (\rho_1 - \rho_2)[I(U_{t-d} \leq \lambda) - \lambda]Y_{t-1} + \varepsilon_t \tag{1.19}$$

subtracting on both sides Y_{t-1}

$$\Delta Y_t = [E(\delta_t) - 1]Y_{t-1} + (\rho_1 - \rho_2)[I(U_{t-d} \leq \lambda) - \lambda]Y_{t-1} + \varepsilon_t \tag{1.20}$$

rearranging the different terms

$$\Delta Y_t = \phi Y_{t-1} + \gamma H_t(\lambda)Y_{t-1} + \varepsilon_t, \tag{1.21}$$

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

Both conditions of the *TARSUR* process can be characterized by the parameters ϕ and γ in model (1.21) 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 data generating process (DGP), the regression model to implement the test will contain the state dependent constant:

$$\Delta Y_t = \mu_1 I(U_{t-d} \leq \lambda) + \mu_2 I(U_{t-d} > \lambda) + \phi Y_{t-1} + \gamma H_t(\lambda) Y_{t-1} + \varepsilon_t \tag{1.22}$$

1.4.1 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 examined by testing in regression model (1.22):

$$\begin{aligned} H_0 &: \phi = 0 \\ H_1 &: \phi < 0 \end{aligned} \tag{1.23}$$

Under H_0, the asymptotic distribution of $t_{\phi=0}$ statistic shows a distribution discontinuity, like 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 $V(\delta_t) = 0$, it implies that $\gamma = 0$. Under H_0, the DGP (1.21) becomes:

  $$Y_t = Y_{t-1} + \varepsilon_t, \tag{1.24}$$

  which is the random walk (RW) process.

- For the case when $V(\delta_t) > 0$, $\gamma \neq 0$. Under H_0, the DGP (1.21) becomes:

  $$\Delta Y_t = \gamma H_t(\lambda) + \varepsilon_t, \tag{1.25}$$

  which is the *TARSUR* process from Definition 1.

The distribution discontinuity is due to 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 and also possibly covariance stationary under the conditions given in 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 (1.21), and regression model (1.22) with no deterministic terms. Then, under $H_0 : \phi = 0$, the $t_{\phi=0}$ statistic has the following asymptotic distribution:*

$$t_{\phi=0} \Rightarrow \frac{\frac{1}{2}[W(1)^2 - 1]}{\{\int_0^1 W(r)^2 dr\}^{1/2}} \tag{1.26}$$

2. *Consider the DGP (1.21), and regression model (1.22) with a threshold constant term. Then, under $H_0 : \phi = 0$, the $t_{\phi=0}$ statistic has the following asymptotic distribution:*

$$t_{\phi=0} \Rightarrow \frac{\frac{1}{2}[W(1)^2 - 1] - W(1)\int_0^1 W(r)dr}{\{\int_0^1 W(r)^2 dr - [\int_0^1 W(r)dr]^2\}^{1/2}} \tag{1.27}$$

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

Note that in 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:*

$$t_{\phi=0} \Rightarrow \mathcal{N}(0, 1) \tag{1.28}$$

Since we do not know if $V(\delta_t)$ is positive or zero, we do not how it is the asymptotic distribution of $t_{\phi=0}$. 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, following 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.*

One may wonder how $V(\delta_t)$ enters the process J_{c_1,c_2}, similar to the 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 (1.21) and the regression model (1.22) with no deterministic term. Then, under* $H_0 : \phi = 0$*, the* $t_{\phi=0}$ *statistic has the following asymptotic distribution:*

$$t_{\phi=0} \Rightarrow \frac{\int_0^1 J_{c_1,c_2}(q)dW(q)}{\{\int_0^1 J^2_{c_1,c_2}(q)dq\}^{1/2}} \tag{1.29}$$

2. *Consider DGP (1.21) and the regression model (1.22) with a threshold constant term. Then, under* $H_0 : \phi = 0$*, the* $t_{\phi=0}$ *statistic has the following asymptotic distribution:*

$$t_{\phi=0} \Rightarrow \frac{\int_0^1 J_{c_1,c_2}(q)dW(q) - W(1)\int_0^1 J_{c_1,c_2}dq}{\{\int_0^1 J^2_{c_1,c_2}(q)dq - [\int_0^1 J_{c_1,c_2}(q)dq]^2\}^{1/2}} \tag{1.30}$$

Note that the 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 must 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 conditions 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 geometrically ergodic and for the case where $V(\delta_t) = 0$, it is proven in Romano and Wolf (2001).

1.4.2 Testing for Threshold Effect

This section attempts to construct a test for the null of no-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 the no-threshold effect ($V(\delta_t) = 0$) versus the alternative of a threshold effect ($V(\delta_t) > 0$) can be tested by testing

$$\begin{aligned} H_0 &: \gamma = 0 \\ H_1 &: \gamma \neq 0 \end{aligned} \tag{1.31}$$

in regression model (1.22)

The proposed test and its asymptotic distribution depend on whether the threshold parameter λ is known or unknown and unidentified under the null.

1.4.2.1 Threshold Value Known

The case of a known threshold value, $\lambda = \bar{\lambda}$, becomes 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 the regression model (1.22), and its asymptotic is shown in the next proposition.

Proposition 5. *Suppose that the threshold value is known, then $\lambda = \bar{\lambda}$, 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*

$$t_{\gamma=0}(\bar{\lambda}) \Rightarrow \mathcal{N}(0,1) \tag{1.32}$$

1.4.2.2 Threshold Value Unknown

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

$$\underset{\lambda \in (0,1)}{\text{argmin}} \quad \hat{\sigma}^2(\lambda) \tag{1.33}$$

, 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 (1.22). This estimate $\hat{\lambda}$ coincides with the one obtained by maximizing the Wald statistics, $W_T(\lambda)$, that test the null hypothesis of no threshold in regression (1.22)

$$W_T = W_T(\hat{\lambda}) = \sup_{\lambda \in (0,1)} \quad W_T(\lambda) \tag{1.34}$$

, 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 (1.21) and regression model (1.22) with no deterministic terms. Then, under the null* $H_0 : \gamma = 0$*, the* W_T *statistic has the following asymptotic distribution:*

$$W_T \Rightarrow \underset{\lambda \in (0,1)}{sup} \frac{\left(\int W(s) dV(s, \lambda)\right)^2}{\lambda(1-\lambda) \int W(s)^2 ds} \equiv \underset{\lambda \in (0,1)}{sup} \frac{[BB(\lambda)]^2}{\lambda(1-\lambda)} \tag{1.35}$$

where $W(.)$ *is the standard Brownian motion and* $V(s, \lambda)$ *is a Kiefer-Muller* [3] *process 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 is due to the fact that* $W(s) = W(s, 1)$ *and* $V(s, \lambda)$ *are independent.*

2. *Consider DGP (1.21) and regression model (1.22) with a threshold constant term. Then, under the null* $H_0 : \gamma = 0$*, the* W_T *statistic has the following asymptotic distribution:*

$$W_T \Rightarrow \underset{\lambda \in (0,1)}{sup} \frac{\left(\int W(s)^* dV(s, \lambda)\right)^2}{\lambda(1-\lambda) \int W^*(s)^2 ds} \equiv \underset{\lambda \in (0,1)}{sup} \frac{[BB(\lambda)]^2}{\lambda(1-\lambda)} \tag{1.36}$$

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

[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 $Cov[B(s_1, t_t), B(s_2, t_2)] = (s_1 \wedge t_1)(s_2 \wedge t_2)$.

From an empirical perspective, we cannot search the threshold parameter λ in the unit interval because as λ approaches zero or one, we do not have enough observations 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

$$\sup_{\lambda \in [\pi_1, \pi_2]} \frac{[BB(\lambda)]^2}{\lambda(1-\lambda)} \tag{1.37}$$

For different π_0, the critical values of the asymptotic distribution (1.37) are tabulated in (Andrews (1993, 2003) and Estrella (2003)).

1.5 A Monte Carlo Experiment and Testing Strategy.

Using the 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 consists of 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, and without loss of generality, a first order Markov process with transition matrix:

$$F = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \tag{1.38}$$

To fill these transition probabilities such that $E(\delta_t) = 1$ holds, first we fix the coefficient ρ_1 and ρ_2, 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 1.1 and 1.4 show the empirical size for the two proposed tests, that is, for the mean $E(\delta_t) = 1$ and for the variance $V(\delta) = 0$, under different sample sizes and dependence levels 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 symmetry restrictions 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 are equal to 0.5, which represents the case where the threshold variable is *i.i.d.*. Table 1.1 summarizes the results by assuming that the threshold parameter is known, and we can see that the empirical size coincides with the nominal size of 5% for both tests. Table 1.4 reports the same results as in Table 1.1 but assumes 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 variables from smallest to the biggest and dropping 15% of the elements on the left and the right side, that is, $\pi_0 = 0.15$.

Tables 1.2 and 1.5 show 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 1.2 are constructed by assuming that the threshold parameter is known, and we can see that independent of the value of $|\gamma|$, the empirical size for the test of the mean ($E(\delta_t) = 1$) coincides with the nominal size of 5%. Also, we can see that as $|\gamma|$ gets bigger the empirical power for the test of the variance, $V(\delta_t)$ changes to one. Tables 1.5 shows the same result under the assumption that the threshold parameter is unknown and unidentified.

Tables 1.3 and 1.6 report the same information as in Tables 1.2 and 1.5, 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 to be asymmetric.

Tables 1.7, 1.8, 1.15, and 1.16 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 to 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 for the matrix F will be 0.5, which means that the threshold variable is *i.i.d.*. Tables 1.7 and 1.8 assume that the threshold value is known, and as shown in Proposition 5, independent of the value of k, the empirical size of the test for $V(\delta_t)$ coincides 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 1.15 and 1.16 show the same results but assume that the threshold parameter is unknown and unidentified.

Table 1.9, 1.10, 1.11, 1.12,1.13, and 1.14 show 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 levels 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 1.17, 1.18, 1.19, 1.20, 1.21, and 1.22 show the same results but assume that the threshold parameter is unknown and unidentified. For illustrative purpose, we present the power of the Dickey-Fuller (DF) unit root test against the same *TARSUR* alternatives previously considered. The $t-$statistic is calculated from the regression

$$\Delta Y_t = b_1 + b_2 Y_{t-1} + v_t \tag{1.39}$$

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

1.6 Some Extra Issues

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

1. Models with higher dynamics. For practical purpose, model (1.5) is rather too simplistic, so it must be replaced as in Leybourne, McCabe and Tremayne (1996) by a more general version of (1.5)

$$Y_t^* = \delta_t Y_{t-1}^* + \varepsilon_t \tag{1.40}$$

where

$$Y_t^* = Y_t - \sum_{i=1}^{p} \omega_i Y_{t-i}, \tag{1.41}$$

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

$$Y_t = \sum_{i=1}^{p+1} \eta_{it} Y_{t-i} + \varepsilon_t, \tag{1.42}$$

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 of the $AR(p+1)$ process. The sum, s_t, of the $p+1$ AR coefficient is given by

$$s_t = \delta_t(1 - \sum_{i=1}^{p} \omega_i) + \sum_{i=1}^{p} \omega_i, \tag{1.43}$$

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 and the asymptotic theory developed in Section 4 still hold. For the latter, the only required modification is to add p lags of ΔY_t in the regression model (1.22). The number of lags can be chosen by some information criteria (see Kapetianos (2001)).

2. Determination of number of regimes. The number of regimes can be determined by sequential testing or by some model selection technique. The first approach consists of running the TARSUR tests sequentially in a similar fashion as it done in Bai and Perron (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 toughest topic in the literature. To the best of our knowledge, the most general solution is given via the use of sub-sampling methods in Gonzalo and Wolf (2004). Extensions of this approach to a TARSUR framework are currently under 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 splits the sample in a similar fashion. In practice, we propose to choose the threshold variable based on some information criteria.

1.7 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 some theoretical and/or empirical controversy exists about the randomness of the unit root in the *AR* representation. The first example models U.S. stock prices, the second investigates the U.S. house prices, the third analyzes the U.S. interest rates, and fourth the nominal exchange rate between the USD/Pound exchange rates.

1.7.1 U.S. Stock Price

Following the economic model presented in Section 1.2, in this application we investigate via our TARSUR model the link between asset prices and real activity, as well as the predictability of asset returns. The data analyzed are the quarterly series of real Standard & Poor's Composite Stock Price Index from 1947:1 to 2016:4. The threshold variable representing the real activity is the increment in 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 the 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 corresponds to changes in the real GDP ($\Delta rgdp_t$). The Dickey-Fuller unit root test suggests that real stock prices as well as the real GDP contain a unit root, therefore, Z_t is $I(0)$.

[Tables 1.23 and 1.24 here]

Table 1.23 summarizes the estimation results for 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 the smallest to the biggest and dropping 15% of the elements of this set in the right and the remaining in the left, in terms of the distribution (1.37), $\pi_1 = 0.15$ 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 the 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 capture a clearly positive relationship between the stock market and real activity but also it finds 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 to 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 are constructed by using the estimated parameters ($\hat{\phi}, \hat{\gamma}, \hat{\lambda}, \hat{\mu}_1, and \hat{\mu}_2$) 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, 2006), the alternative hypothesis considered will be one-sided $H_1 : \sigma^2_{RW} > \sigma^2_{TARSUR}$ because if the process does not follow a RW, we expect forecast from the $TARSUR$ 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 predicts 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:

$$Y_t = \mu_1 I(Z_{t-d} \leq \lambda) + \mu_2 I(Z_{t-d} > \lambda) + \rho_1 I(Z_{t-d} \leq \lambda) Y_{t-1} + \rho_1 I(U_{t-d} \leq \lambda) Y_{t-1} + \varepsilon_t \quad (1.44)$$

The second form is to impose the null of $\phi = 0$ on the regression model (1.22) 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 1.24). When $E(\delta_t) = 1$ holds the estimates of ρ_1 and ρ_2, in both methods it should be the same. The results in Tables 1.23 and 1.24 show that when the increment of real GDP is less than 78.71, the stock price index is in the stationarity 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 is larger than 78.71, prices follow a mildly explosive model (autoregressive parameter is equal to 1.023). This occurs with probability 0.32. Overall, the stochastic root of the autoregressive representation is on average unity.

[Insert figure 1.2 here]

Figure (1.2) presents the plot of the U.S. stock prices, the green dots represent the periods in which the TARSUR model tells us that the stock prices are in the explosive state, and the red dots represent the periods in which the stock prices are in the mean reverting period. The vertical lines represent the 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 are expanding and the periods in which they are 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 a martingale process with respect to 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

$$E_{t-1}\left(\frac{Y_t - Y_{t-1}}{Y_{t-1}}\right) = E_{t-1}(\delta_t - 1) \neq 0 \tag{1.45}$$

From (1.45) and the results in Tables 1.23 and 1.24 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 this way, we find that there exists a positive non-linear relationship between the expected stock returns and the real activity of the economy. The linear links between the stock returns and macroeconomic variables are found widely in the financial literature (Chen et al. (1986), Fama (1990)).

1.7.2 U.S House Price

In this application, we study the link between house prices and real activity using the *TAR-SUR* model. The analyzed data are the quarterly series of the 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 the Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org).

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

Modeling price bubble as an explosive autoregressive process captures 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 to remain above unity for some periods 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 bubbles, and we will also able to find a plausible random variable capable of explaining this behavior change.

As before, the estimated *TARSUR* model for the house prices 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 suggests that the home price index and real GDP per-capita have a unit root but $\Delta RgdpP_t$ is $I(0)$.

[Insert Tables 1.25 and 1.26 here]

Table 1.25 summarizes the estimation results; however, since the threshold parameter is unknown, we search the threshold parameter in a subset generated by dropping 15% of the threshold value of 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 compares 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 the 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 to 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 shows that the *TARSUR* model is slightly superior by predicting 69% correctly against the 65% predicted by the RW.

In Table 1.26 we recover the estimates of ρ_1 and ρ_2 and the transition probabilities $\hat{p}_{22}$ and $\hat{p}_{12}$. The results in Tables 1.25 and 1.26 show 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 1.3 here]

Figure 1.3 presents the plot of the U.S. real house price index, the vertical lines represent the U.S. recessions (www.nber.org/cycles/cyclesmain.html). The green dots represent the periods in which the $TARSUR$ model tells us that the house price is in the explosive state and the red dots represent the periods where the $TARSUR$ model tells us that the house price is in a mean reverting state. Note that the $TARSUR$ model is able to assess something about the 2008 house price bubble, since it is able to capture the explosive behavior of house prices between 2001 to 2008, represented by green dots, and the implosion of house prices between 2008 to 2010, represented by red dots.

1.7.3 U.S Interest rates

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

Leybourne, McCabe and Mills (1996) perform a similar exercise for the international U.S. bond yield data (BUS) but with 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 body of literature showing the negative relation between interest rates and unemployment rates (Sargent, Fand and Goldfeld (1973), Friedman (1977), Blanchard and Wolfers (2000)). Then, the threshold variables we use will be the annual changes in the unemployment rate ($Aunrate_t$) available in the Federal Reserve Bank of St. Louis.

[Insert Tables 1.27 and 1.28 here]

Table 1.27 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. We can see from Tables 1.27 and 1.28 that if the annual change in unemployment rate is less than 0.4%, the interest rate is in the "explosive" state with the 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 in unemployment rate is greater than 0.4%, the interest rate is in the mean reverting state with the coefficient 0.968, and the probability of being in this regime is 0.26.

[Insert Figure 1.4 here]

Figure 1.4 plots the series of interest rates, the green dots represent the periods in which the interest rates are 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 a recession as determined by the National Bureau of Economic Research (NBER)(www.nber.org/cycles/cyclesmain.html). As we can see during the recession periods, during which the unemployment rate increases and 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 the 5% significant level but not rejected at the 1% significant level. Furthermore, we evaluate the number of times the *TARSUR* model predicts correctly the sign with respect to the RW. In this case, the *TARSUR* model has a similar performance to the RW with 48% and 47%, respectively.

1.7.4 Dollar/Pound Nominal Exchange Rates

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

In order to estimate a *TARSUR* model, we need to find a suitable threshold variable. In their work, Messe and Rugoff (1983)and Barbara Rossi (2006) use the first difference of the nominal short-term interest rate differential between countries as one of the explanatory variables 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 rates can be found in the OECD database (http://www.oecd.org/std).

Meese and Rogoff (1983) show that economic models used to forecast exchange rates are outperformed by the random walk. A possible explanation for this phenomenon 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 the time-varying parameter models, Engle (1994) and Marsh (2000) use regime-switching models, but it is still unable to beat the random walk. Schinasi and Swamy (1989) and Rossi (2006) use random coefficient models and they can have a better out-of-sample forecast than the RW.

[Insert Tables 1.29]

Table 1.29 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 suggest the presence of a unit root that is fixed.

From the forecast perspective, using the method of Clark and West (2006), we compare the *TARSUR* model with respect to 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 at 47%. This result that we obtain is like the one obtained by Engle (1994), but with a different methodology. The advantage of the TARSUR model is that we can find a reason why both models have an equal out-of-sample forecast performance in terms of mean squared error. This is because we are not able to reject the existence of fixed unit roots.

1.8 Conclusion

This study 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 not only make the unit root models more flexible but also find an explanation for the existence of unit roots. Flexibility is obtained because depending on the values of certain parameters, the *TARSUR* process can behave like an explosive process, an exact unit root process, or 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 show that estimation and testing of *TARSUR* models is not more difficult than the estimation and testing involved in fixed-unit root models. This is a clear advantage of the *TARSUR* models with respect to other stochastic unit root methodologies available in the literature.

Appendix

1.A 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

$$\mathbb{E}log|\delta_1| < log\mathbb{E}|\delta_1| = log\mathbb{E}\delta_1 = 0 \tag{1.46}$$

Therefore condition (1.7) holds.

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

$$\sum_{j=0}^{\infty} E\Big(|\psi_{t,j}|^2\Big)^{\frac{1}{2}} = \left[\begin{pmatrix}1 & 1\end{pmatrix} \sum_{j=1}^{\infty} F_2^j \begin{pmatrix}\rho_1 p_1 \\ \rho_2 p_2\end{pmatrix}\right] < \infty, \tag{1.47}$$

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

$$\xi_h = \begin{pmatrix}1 & 1\end{pmatrix} \sum_{j=1}^{\infty} F_1^h \begin{pmatrix}\rho_1 p_1 \\ \rho_2 p_2\end{pmatrix}, \quad h = 1, 2, \cdots, \tag{1.48}$$

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 (1.15),

$$\Delta Y_t = \varepsilon_t + (\delta_t - 1)\sum_{j=1}^{m-1}\left(\prod_{i=1}^{j-1}\delta_{t-i}\right)\varepsilon_{t-j} + (\delta_t - 1)\left(\prod_{i=1}^{m-1}\delta_{t-i}\right)Y_{t-m}. \tag{1.49}$$

Subtracting (1.15) from equation (1.49)

$$\Delta Y_t(Y_{t-m}) - \Delta Y_t = (\delta_t - 1)(Y_{t-1}(Y_{t-m}) - Y_{t-1}), \tag{1.50}$$

where $\Delta(Y_{t-m})$ correspond to equation (1.49) and ΔY_t to equation (1.15). 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

$$W_T(s,\lambda) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} I(U_{t-d} \leq \lambda)\varepsilon_t, \tag{1.51}$$

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

$$W_T(s,\lambda) \Rightarrow \sigma W(s,\lambda), \tag{1.52}$$

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

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Y_t I(U_{t-d} \leq \lambda)\varepsilon_t \Rightarrow \sigma \int_0^1 W(s) dW(s,\lambda), \tag{1.53}$$

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

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} Y_t I(U_{t-d} \leq \lambda) \Rightarrow \lambda\sigma \int_0^1 W(s) ds \tag{1.54}$$

$$\frac{1}{T^2} \sum_{t=1}^{T} Y_t^2 I(U_{t-d} \leq \lambda) \Rightarrow \lambda\sigma^2 \int_0^1 W^2(s) ds \tag{1.55}$$

The proofs are divided into two parts depending if the deterministic components are included in the regression model (1.22): (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 (1.22) as follows,

$$\Delta Y_t = X_{t-1}\beta + \varepsilon_t \tag{1.56}$$

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,

$$\hat{\beta} = \left(\sum_{t=1}^{T} X'_{t-1}X_{t-1}\right)^{-1}\left(\sum_{t=1}^{T} X'_{t-1}\Delta Y_t\right), \tag{1.57}$$

equivalently

$$\hat{\beta} - \beta = \left(\sum_{t=1}^{T} X'_{t-1}X_{t-1}\right)^{-1}\left(\sum_{t=1}^{T} X'_{t-1}\varepsilon_t\right), \tag{1.58}$$

Now,

$$\sum_{t=1}^{T} X'_{t-1}X_{t-1} = \begin{pmatrix} \sum_{t=1}^{T} Y_{t-1}^2 & \sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) \\ \sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) & \sum_{t=1}^{T} Y_{t-1}^2 H_t^2(\lambda) \end{pmatrix} \tag{1.59}$$

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 (1.58) we get

$$\Gamma_b(\hat{\beta} - \beta) = \begin{pmatrix} T^{-2b}\sum_{t=1}^{T} Y_{t-1}^2 & T^{-2b}\sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) \\ T^{-2b}\sum_{t=1}^{T} Y_{t-1}^2 H_t(\lambda) & T^{-2b}\sum_{t=1}^{T} Y_{t-1}^2 H_t^2(\lambda) \end{pmatrix}^{-1} \begin{pmatrix} T^{-2b}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \\ T^{-2b}\sum_{t=1}^{T} Y_{t-1}H_t(\lambda)\varepsilon_t \end{pmatrix} \tag{1.60}$$

Equation (1.60) is key since we derive the asymptotic distribution of the tests from here, for the case in which the regression model (1.22) 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 (1.22). We can estimate both parameters from the following regression,

$$[I(U_{t-d} \leq \lambda)\Delta Y_t^I + I(U_{t-d} > \lambda)\Delta Y_t^{II}] = \phi[I(U_{t-d} \leq \lambda)Y_{t-1}^I + I(U_{t-d} > \lambda)Y_{t-1}^{II}] + \gamma[(1-\lambda)I(U_{t-d} \leq \lambda)Y_{t-1}^I + \lambda I(U_{t-d} > \lambda)Y_{t-1}^{II}] \tag{1.61}$$

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 (1.61) as follows

$$[I(U_{t-d} \leq \lambda)\Delta Y_t^I + I(U_{t-d} > \lambda)\Delta Y_t^{II}] = \tilde{X}_{t-1}'\beta + \varepsilon_t \tag{1.62}$$

where $\beta = \begin{pmatrix} \phi \\ \gamma \end{pmatrix}$ and

$$\tilde{X}_{t-1}' = \begin{pmatrix} I(U_{t-d} \leq \lambda)Y_{t-1}^I + I(U_{t-d} > \lambda)Y_{t-1}^{II} & (1-\lambda)I(U_{t-d} \leq \lambda)Y_{t-1}^I + \lambda I(U_{t-d} > \lambda)Y_{t-1}^{II} \end{pmatrix} \tag{1.63}$$

Then as before the least square estimate

$$\tilde{\beta} = \left(\sum_{t=1}^T \tilde{X}_{t-1}'\tilde{X}_{t-1}\right)^{-1}\left(\sum_{t=1}^T \tilde{X}_{t-1}'\Delta Y_t\right), \tag{1.64}$$

equivalently

$$\tilde{\beta} - \beta = \left(\sum_{t=1}^T \tilde{X}_{t-1}'\tilde{X}_{t-1}\right)^{-1}\left(\sum_{t=1}^T \tilde{X}_{t-1}'\varepsilon_t\right), \tag{1.65}$$

Now,

$$\sum_{t=1}^T \tilde{X}_{t-1}'\tilde{X}_{t-1} = \begin{pmatrix} \tilde{x}_1 & \tilde{x}_2 \\ \tilde{x}_3 & \tilde{x}_4 \end{pmatrix} \tag{1.66}$$

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

For $\Gamma_1 = \begin{pmatrix} T & 0 \\ 0 & T \end{pmatrix}$ and multiplying both sides of (1.65) we get,

$$\Gamma_1(\tilde{\beta}-\beta) = \begin{pmatrix} T^{-2}\tilde{x}_1 & T^{-2}\tilde{x}_2 \\ T^{-2}\tilde{x}_3 & T^{-2}\tilde{x}_4 \end{pmatrix}^{-1} \begin{pmatrix} T^{-1}\sum_{t=1}^{T}[I(U_{t-1}\leq\lambda)Y_{t-1}^{I} + I(U_{t-1}>\lambda)Y_{t-1}^{II}]\varepsilon_t \\ T^{-1}\sum_{t=1}^{T}[(1-\lambda)I(U_{t-d}\leq\lambda)Y_{t-1}^{I} + \lambda I(U_{t-d}>\lambda)Y_{t-1}^{II}]\varepsilon_t \end{pmatrix} \tag{1.67}$$

Expression (1.67) 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 (1.21) 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 (1.60), for $b=1$, that is

$$\Gamma_1(\hat{\beta}-\beta) = \begin{pmatrix} T^{-2}\sum_{t=1}^{T}Y_{t-1}^2 & T^{-2}\sum_{t=1}^{T}Y_{t-1}^2H_t(\lambda) \\ T^{-2}\sum_{t=1}^{T}Y_{t-1}^2H_t(\lambda) & T^{-2}\sum_{t=1}^{T}Y_{t-1}^2H_t^2(\lambda) \end{pmatrix}^{-1} \begin{pmatrix} T^{-1}\sum_{t=1}^{T}Y_{t-1}\varepsilon_t \\ T^{-1}\sum_{t=1}^{T}Y_{t-1}H_t(\lambda)\varepsilon_t \end{pmatrix} \tag{1.68}$$

Note that since Y_t is a RW we have that

$$T^{-2}\sum_{t=1}^{T}Y_{t-1}^2 \Rightarrow \sigma^2\int_0^1 W^2(s)ds \tag{1.69}$$

by construction of $H_t(\lambda)=I(U_{t-d}\leq\lambda)-\lambda$ and (1.55) with (1.69) we know that

$$T^{-2}\sum_{t=1}^{T}Y_{t-1}^2H_t(\lambda) = T^{-2}\sum_{t=1}^{T}I(U_{t-d}\leq\lambda)Y_{t-1}^2 - \lambda T^{-2}\sum_{t=1}^{T}Y_{t-1}^2 \to 0 \tag{1.70}$$

Finally from (1.55) and (1.69) we have

$$T^{-2}\sum_{t=1}^{T}Y_{t-1}^2H_t^2(\lambda) \Rightarrow \sigma^2\lambda(1-\lambda)\int_0^1 W^2(s)ds. \tag{1.71}$$

From (1.69), (1.70) and (1.71) the matrix

$$\begin{pmatrix} T^{-2}\sum_{t=1}^{T}Y_{t-1}^2 & T^{-2}\sum_{t=1}^{T}Y_{t-1}^2H_t(\lambda) \\ T^{-2}\sum_{t=1}^{T}Y_{t-1}^2H_t(\lambda) & T^{-2}\sum_{t=1}^{T}Y_{t-1}^2H_t^2(\lambda) \end{pmatrix}^{-1} \Rightarrow \left(\sigma^2\int_0^1 W^2(s)ds\right)^{-1}\begin{pmatrix} 1 & 0 \\ 0 & [\lambda(1-\lambda)]^{-1} \end{pmatrix} \tag{1.72}$$

From the usual unit root asymptotic we known that

$$T^{-1}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \Rightarrow \sigma^2\frac{1}{2}[W(1)^2-1] \tag{1.73}$$

and

$$T^{-1}\sum_{t=1}^{T} Y_{t-1}H_t(\lambda)\varepsilon_t \Rightarrow \sigma^2\int_0^1 W(s)dV(s,\lambda), \tag{1.74}$$

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

$$\begin{pmatrix} T^{-1}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \\ T^{-1}\sum_{t=1}^{T} Y_{t-1}H_t(\lambda)\varepsilon_t \end{pmatrix} \Rightarrow \sigma^2 \begin{pmatrix} \frac{1}{2}[W(1)^2-1] \\ \int_0^1 W(s)dV(s,\lambda) \end{pmatrix} \tag{1.75}$$

Putting all together we have

$$\Gamma_1(\hat{\beta}-\beta) \Rightarrow \begin{pmatrix} \frac{\frac{1}{2}[W(1)^2-1]}{\int_0^1 W^2(s)ds} \\ \frac{\int_0^1 W(s)dV(s,\lambda)}{\lambda(1-\lambda)\int_0^1 W^2(s)ds} \end{pmatrix} \tag{1.76}$$

From (1.76) 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 $\tilde{\beta}$.

$$\begin{pmatrix} T^{-2}\tilde{x}_1 & T^{-2}\tilde{x}_2 \\ T^{-2}\tilde{x}_3 & T^{-2}\tilde{x}_4 \end{pmatrix}^{-1} \Rightarrow \frac{1}{\left(\int_0^1 W(s)^2ds - [\int_0^1 W(s)ds]^2\right)} \begin{pmatrix} 1 & 0 \\ 0 & [(1-\lambda)\lambda]^{-1} \end{pmatrix} \tag{1.77}$$

$$\begin{pmatrix} T^{-1}\sum_{t=1}^{T}\left[I(U_{t-d}\le\lambda)Y_{t-1}^{I} + I(U_{t-d}>\lambda)Y_{t-1}^{II}\right]\varepsilon_t \\ T^{-1}\sum_{t=1}^{T}\left[(1-\lambda)I(U_{t-d}\le\lambda)Y_{t-1}^{I} - \lambda I(U_{t-d}>\lambda)Y_{t-1}^{II}\right]\varepsilon_t \end{pmatrix} \Rightarrow \sigma^2 \begin{pmatrix} \int_0^1 W(s)dB(s) - W(1)\int_0^1 W(s)ds \\ \int_0^1 W(s)dV(s,\lambda) - V(1,\lambda)\int_0^1 W(s)ds \end{pmatrix} \tag{1.78}$$

Putting all together we have that:

$$\Gamma_1(\tilde{\beta}-\beta) \Rightarrow \begin{pmatrix} \frac{\int_0^1 W(s)dW(s) - W(1)\int_0^1 W(s)ds}{\left(\int_0^1 W(s)^2ds - [\int_0^1 W(s)ds]^2\right)} \\ \frac{\int_0^1 w(s)dV(s,\lambda) - V(1,\lambda)\int_0^1 W(s)ds}{\lambda(1-\lambda)\left(\int_0^1 W(s)^2ds - [\int_0^1 W(s)ds]^2\right)} \end{pmatrix} \tag{1.79}$$

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 (1.60) 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

$$\ln(\delta_t) = \ln(\rho_1 I(U_{t-d} \leq \lambda) + \rho_2 I(U_{t-d} > \lambda)) = \ln(\rho_1) I(U_{t-d} \leq \lambda) + \ln(\rho_2) I(U_{t-d} > \lambda) \quad (1.80)$$

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

$$X_T(q) = T^{-1/2}\sigma^{-1}S_{[Tq]} = T^{-1/2}\sigma^{-1}S_{j-1}, \quad \frac{j-1}{T} \leq q < \frac{j}{T} \quad (1.81)$$

we have that,

$$X_T(q) \Rightarrow W(q) \quad (1.82)$$

Iterating backward the TARSUR model (1.5) we have that:

$$Y_{[Tq]} = \sum_{i=1}^{[Tq]} \Big(\prod_{j=1}^{[Tq]-i} \delta_{[Tq]-j+1} \Big) \varepsilon_i + \Big(\prod_{j=1}^{[Tq]} \delta_j \Big) Y_0 \quad (1.83)$$

Taking logs and exponential in the product of δ_t

$$Y_{[Tq]} = \sum_{i=1}^{[Tq]} e^{\sum_{j=1}^{[Tq]-i} \ln(\delta_{t-j+1})} \varepsilon_i + \Big(\prod_{j=1}^{[Tq]} \delta_j \Big) Y_0 \quad (1.84)$$

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

$$Y_{[Tq]} = \sum_{i=1}^{[Tq]} e^{([Tq]-i)E(\ln(\delta_t))} e^{\sum_{j=1}^{[Tq]-i} [\ln(\delta_{t-j+1}) - E(\ln(\delta_t))]} \varepsilon_i + \Big(\prod_{j=1}^{[Tq]} \delta_j \Big) Y_0 \quad (1.85)$$

First focus on the term $e^{([Tq]-i)E(\ln(\delta_t))}$ in equation (1.85). From assumption (A.8) we have that

$$e^{([Tq]-i)E(\ln(\delta_t))} = e^{\frac{[Tq]-i}{T}[c_1\lambda + c_2(1-\lambda)]} = e^{\frac{[Tq]-i}{T}C} \quad (1.86)$$

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

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

$$
\begin{aligned}
& e^{\sum_{j=1}^{[Tq]-i}[\ln(\rho_1)I(U_{[Tq]-d-j+1}\leq\lambda)+\ln(\rho_2)I(U_{[Tq]-d-j+1}>\lambda)-\ln(\rho_1)\lambda-\ln(\rho_2)(1-\lambda)]} \\
&= e^{\sum_{j=1}^{[Tq]-i}[\ln(\rho_1)(I(U_{[Tq]-d-j+1}\leq\lambda)-\lambda)+\ln(\rho_2)(I(U_{[Tq]-d-j+1}>\lambda)-(1-\lambda))]} \qquad (1.87)\\
&= e^{(\ln(\rho_1)-\ln(\rho_2))\sum_{j=1}^{[Tq]-i}[I(U_{[Tq]-d-j+1}\leq\lambda)-\lambda]}
\end{aligned}
$$

From assumption (A.8)we have that:

$$= e^{\frac{c_1-c_2}{T}\sum_{j=1}^{[Tq]-i}[I(U_{[Tq]-d-j+1}\leq\lambda)-\lambda]} \qquad (1.88)$$

Note that as $T \to \infty$

$$\frac{1}{[Tq]-i}\sum_{j=1}^{[Tq]-i}[I(U_{[Tq]-d-j+1}\leq\lambda)-\lambda] \to_p 0 \qquad (1.89)$$

such that expression (1.88) can be written as

$$= e^{(c_1-c_2)\frac{[Tq]-i}{T}o_p(1)} \qquad (1.90)$$

From (1.86) and (1.90) we can rewrite (1.85) as follows

$$T^{-1/2}Y_{[Tq]} = \sum_{i=1}^{[Tq]} e^{\frac{[Tq-i]}{T}C+o_p(1)}\varepsilon_i + O(T^{-1/2}) \qquad (1.91)$$

Then

$$
\begin{aligned}
T^{-1/2}Y_{[Tq]} &= \sigma\sum_{i=1}^{[Tq]} e^{\frac{[Tq-i]}{T}C+o_p(1)}\int_{\frac{i-1}{T}}^{\frac{i}{T}} dX_T(s) + O(T^{-1/2}) \\
&= \sigma\sum_{i=1}^{[Tq]}\int_{\frac{i-1}{T}}^{\frac{i}{T}} e^{\frac{[Tq-i]}{T}C+o_p(1)}dX_T(s) + O(T^{-1/2}) \qquad (1.92)\\
&= \sigma\int_0^q e^{(q-s)C+o_p(1)}dX_T(s) + O(T^{-1/2})
\end{aligned}
$$

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 (1.82) and the continuous mapping theorem as $T \to \infty$,

$$\sigma\{X_T(q)+(C+o_p(1))\int_0^q e^{(q-s)C+o_p(1)}X_T(s)ds\}+O(T^{-1/2})\Rightarrow\sigma\{W(q)+C\int_0^q e^{(q-s)C}W(s)ds\} \tag{1.93}$$

The proofs of (b) and (c) are similar. To prove (d) we follow the results of Gonzalo and Pitarakis (2012). We have to show the strong approximation

$$\underset{q\in[0,1]}{\text{Sup}} \quad \Big|\frac{Y_{[Tq]}}{\sqrt{T}} - J_{c_1,c_2}(q)\Big| = o_{a.s}(1) \tag{1.94}$$

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:

$$\underset{q\in[0,1]}{\text{Sup}} \quad \Big|\frac{\sum_{i=1}^{[Tq]}\varepsilon_i}{\sqrt{T}} - W(q)\Big| = o_{a.s}(1) \tag{1.95}$$

Then

$$\begin{aligned} T^{-1/2}Y_{[Tq]} &= \sigma\int_0^q e^{(q-s)C+\{([Tq/T]-q)-(i/T-s)\}C+o_{a.s}(1)}dX_T(s) \\ &= \sigma\int_0^q e^{(q-s)C}dX_T(s)(1+o_{a.s}(1)) \\ &= \sigma\int_0^q e^{(q-s)C}dX_T(s)+o_{a.s}(1) \end{aligned} \tag{1.96}$$

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 (1.96)

$$T^{-1/2}Y_{[Tq]} = \sigma\{X_T(q)+C\int_0^q e^{(q-s)C}X_T(s)ds\}+o_{a.s}(1) \tag{1.97}$$

$$\begin{aligned} \underset{q\in[0,1]}{\text{Sup}}\Big|\frac{Y_{[Tq]}}{\sqrt{T}} - J_{c_1,c_2}(q)\Big| &\le \underset{q\in[0,1]}{\text{Sup}}|X_T(q)-W_T(q)| \\ &+ \underset{q\in[0,1]}{\text{Sup}}\Big|\int_0^q e^{(q-s)C}\Big|\underset{s\in[0,1]}{\text{Sup}}|X_T(a)-W_T(a)|+o_{a.s}(1) = o_{a.s} \end{aligned} \tag{1.98}$$

The rest of the proof of part (d) follows from Gonzalo and Pitarakis (2012). (e) follows identical lines to the proof of (d).

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

$$T^{-1}Y_T^2 = 2c_1\frac{1}{T^2}\sum_{t=1}^{T} I(U_{t-d} \leq \lambda)Y_{t-1}^2 + 2c_2\frac{1}{T^2}\sum_{t=1}^{T} I(U_{t-d} > \lambda)Y_{t-1}^2 + 2\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t + O(T^{-1/2}) \tag{1.99}$$

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

$$2T^{-1}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \Rightarrow \sigma^2\{J_{c_1,c_2}(1)\}^2 - 2\sigma^2 C\int_0^1 \{J_{c_1,c_2}(s)\}^2 ds - \sigma^2 = 2\sigma^2\int_0^1 J_{c_1,c_2}(s)dW(s) \tag{1.100}$$

The last inequality came form

$$\{J_{c_1,c_2}(1)\}^2 = 1 + 2C\int_0^1 \{J_{c_1,c_2}(s)\}^2 ds + \int_0^1 J_{c_1,c_2}(s)dW(s) \tag{1.101}$$

Our result in (g) follows along the same lines as in Lemma 1 in Gonzalo and Pitarakis (2012) 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 (1.60) 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 (1.67).

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 (1.21) 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 (1.22) with $\mu_1 = \mu_2 = 0$. From (1.78) we can see that

$$T^{-1}(\hat{\gamma} - \gamma) \Rightarrow \frac{\int_0^1 W(s)dV(s,\lambda)}{\lambda(1-\lambda)\int_0^1 W^2(s)ds} \tag{1.102}$$

From the continuous mapping theorem

$$t_{\gamma=0} \Rightarrow \frac{\int_0^1 W(s)dV(s,\lambda)}{\sqrt{\lambda(1-\lambda)\int_0^1 W^2(s)ds}} \tag{1.103}$$

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

$$\frac{\int_0^1 W(s)dV(s,\lambda)}{\sqrt{\int_0^1 W(s)^2 ds}} \equiv \mathcal{N}(0,\sigma_\lambda^2), \tag{1.104}$$

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

Case 2: Regression model (1.22) with state dependent constants. From equation (1.79) we have that:

$$T^{-1}(\tilde{\gamma}-\gamma) \Rightarrow \frac{\int_0^1 B(s)dV(s,\lambda) - V(1,\lambda)\int_0^1 B(s)ds}{\lambda(1-\lambda)\Big(\int_0^1 W(s)^2 ds - [\int_0^1 W(s)ds]^2\Big)} \equiv \frac{\int_0^1 W^*(s)dV(s,\lambda)}{\lambda(1-\lambda)\Big(\int_0^1 W^*(s)^2 ds\Big)} \tag{1.105}$$

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

$$t_{\gamma=0} \Rightarrow \frac{\int_0^1 W^*(s)dV(s,\lambda)}{\sqrt{\lambda(1-\lambda)\int_0^1 W^*(s)^2 ds}} \tag{1.106}$$

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

$$\underset{\lambda\in(0,1)}{\text{Sup}}\; t_{\gamma=0}(\lambda)^2. \tag{1.107}$$

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

$$W_T \Rightarrow \underset{\lambda\in(0,1)}{\text{Sup}}\; t(\lambda)^2. \tag{1.108}$$

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

1.B Tables and Figures

Table 1.1: Empirical size of test for $E(\delta_t)$ and the $V(\delta_y)$. Threshold parameter known $\lambda = 0.5$.

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 1.2: Empirical size of test for $E(\delta_t)$ and power of the test for $V(\delta_t)$. Threshold parameter known and $\lambda = 0.5$.

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 1.3: Empirical size of test for $E(\delta_t)$ and empirical power of the test for $V(\delta_t)$. Threshold parameter known.

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 1.4: Empirical size of test for $E(\delta_t)$ and the variance, $V(\delta_y)$. Threshold parameter unknown and $\lambda = 0.5$.

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 1.5: Empirical size of test for $E(\delta_t)$ and empirical power for $V(\delta_t)$. Threshold parameter Unknown $\lambda = 0.5$

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 1.6: Empirical size of test for $E(\delta_t)$ and empirical power for $V(\delta_t)$. Threshold parameter Unknown.

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 1.7: Local power of the test $E(\delta_t)$, compared with Dickey-Fuller test. Empirical size for $V(\delta_t)$. Threshold parameter is known with $i.i.d.$ threshold variable.

$\rho_1 = 1 - \frac{k}{T},\ \rho_2 = 1 - \frac{k}{T},\ (\lvert\gamma\rvert = 0)$						
	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 1.8: Local power of the test $E(\delta_t)$ and compared with Dickey-Fuller. Empirical size for $V(\delta_t)$. Threshold parameter is known generated by $p_{12} = 0.9$ and $\lambda = 0.5$.

$\rho_1 = 1 - \frac{k}{T}$, $\rho_2 = 1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0)$						
	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 1.9: Local power of the test for $E(\delta_t)$ compared withDickey-Fuller test. Empirical power for $V(\delta_t)$. Assumed threshold parameter is known with *i.i.d.* threshold variable.

$\rho_1 = 0.99 - \frac{k}{T}$, $\rho_2 = 1.01 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.02)$						
	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 1.10: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical power for$V(\delta_t)$. Threshold parameter known with $p_{12} = 0.9$ and $\lambda = 0.5$.

$\rho_1 = 0.99 - \frac{k}{T}$, $\rho_2 = 1.01 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.02)$						
	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 1.11: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical power for $V(\delta_t)$. Threshold parameter known with $i.i.d.$ threshold variable.

$\rho_1 = 0.95 - \frac{k}{T}$, $\rho_2 = 1.05 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.1)$						
	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 1.12: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical power for $V(\delta_t)$. Threshold parameter known with $p_{12} = 0.9$ and $\lambda = 0.5$.

$\rho_1 = 0.95 - \frac{k}{T}$, $\rho_2 = 1.05 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.1)$						
	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 1.13: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical power for $V(\delta_t)$. Threshold parameter is known with *i.i.d.* threshold variable.

$\rho_1 = 0.9 - \frac{k}{T}$, $\rho_2 = 1.1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.2)$						
	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 1.14: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical power for $V(\delta_t)$. Threshold parameter known with $p_{12} = 0.9$ and $\lambda = 0.5$.

$\rho_1 = 0.9 - \frac{k}{T}$, $\rho_2 = 1.1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.1)$						
	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 1.15: Local power for$E(\delta_t)$ compared with Dickey-Fuller. Empirical size for $V(\delta_t)$. Threshold parameter is Unknown with $i.i.d.$ threshold variable.

$\rho_1 = 1 - \frac{k}{T}$, $\rho_2 = 1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0)$						
	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 1.16: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical size for $V(\delta_t)$. Threshold parameter is Unknown with $p_{12} = 0.9$ and $\lambda = 0.5$.

$\rho_1 = 1 - \frac{k}{T}$, $\rho_2 = 1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0)$						
	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 1.17: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical power for $V(\delta_t)$. Assumed threshold parameter is Unknown with *i.i.d.* threshold variable.

$\rho_1 = 0.99 - \frac{k}{T}$, $\rho_2 = 1.01 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.02)$						
	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 1.18: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical power for $V(\delta_t)$. Threshold parameter Unknown with $p_{12} = 0.9$ 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 1.19: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical power for $V(\delta_t)$. Threshold parameter Unknown with *i.i.d.* threshold variable.

$\rho_1 = 0.95 - \frac{k}{T}$, $\rho_2 = 1.05 - \frac{k}{T}$, ($\lvert\gamma\rvert = 0.1$)						
	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 1.20: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical power for $V(\delta_t)$. Threshold parameter Unknown with $p_{12} = 0.9$ 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 1.21: Local power for $E(\delta_t)$ compared with Dickey-Fuller. Empirical power for $V(\delta_t)$. Threshold parameter Unknown with *i.i.d.* threshold variable.

$\rho_1 = 0.9 - \frac{k}{T}$, $\rho_2 = 1.1 - \frac{k}{T}$, $(\lvert\gamma\rvert = 0.2)$						
	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 1.22: Local power for $E(\delta_t)$ compared Dickey-Fuller. Empirical power for $V(\delta_t)$. Threshold parameter is Unknown with $p_{12} = 0.9$ 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 1.23: TARSUR model for U.S. Stock Prices.

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\gamma}$	$\hat{\phi}$	$t_{\phi=0}$	$CV_{t_{\phi=0}}$	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 1.24: U.S. Stock Prices 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 1.25: TARSUR model for U.S. real house prices.

$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\gamma}$	$\hat{\phi}$	$t_{\phi=0}$	$CV_{t_{\phi=0}}$	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 1.26: U.S. house prices 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 1.27: TARSUR model for U.S. interest rates.

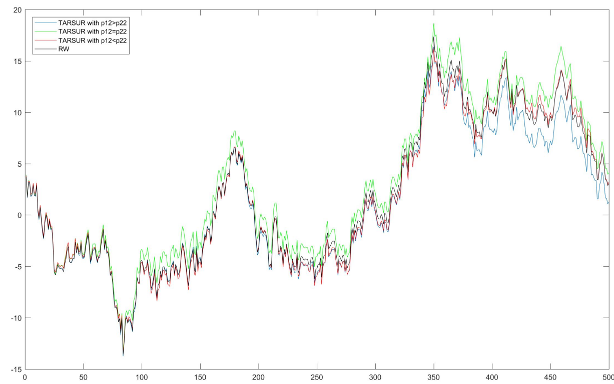
$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\gamma}$	$\hat{\phi}$	$t_{\phi=0}$	$CV_{t_{\phi=0}}$	d	$\hat{r}$	W_T	CV_{W_T}
0.012	0.029	0.038	-0.004	-0.844	-3.557	1	0.4	16.548	8.86
(0.023)	(0.042)	(0.009)	(0.005)						

Table 1.28: U.S. interest rates TARSUR regime roots.

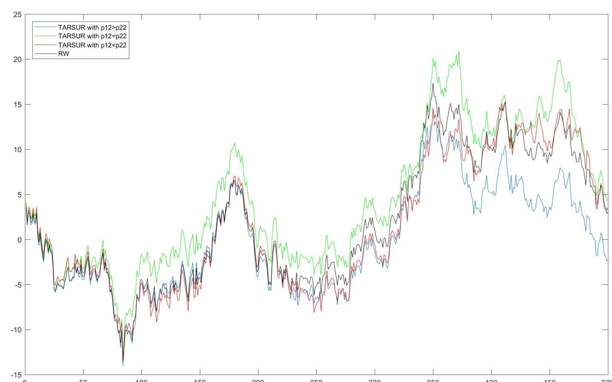
Z_{t-d}	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{P}(r)$	p_{22}	p_{12}
$\Delta gdppc_{t-d}$	1.006	0.968	0.74	0.920	0.028

Table 1.29: TARSUR model for U.S.Dollar/Pound.

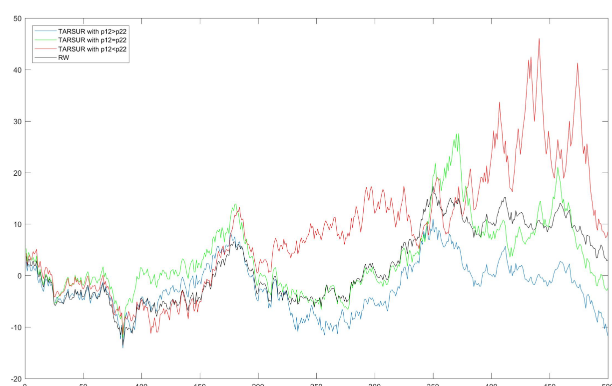
$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\gamma}$	$\hat{\phi}$	$t_{\phi=0}$	$CV_{t_{\phi=0}}$	d	$\hat{r}$	W_T	CV_{W_T}
-0.013	0.058	0.047	-0.015	-1.787	-2.90	1	0	7.78	8.86
(0.020)	(0.019)	(0.017)	(0.008)						



(a)



(b)



(c)

Figure 1.1: Random Walk versus different TARSUR series. For different $V(\delta_t)$.

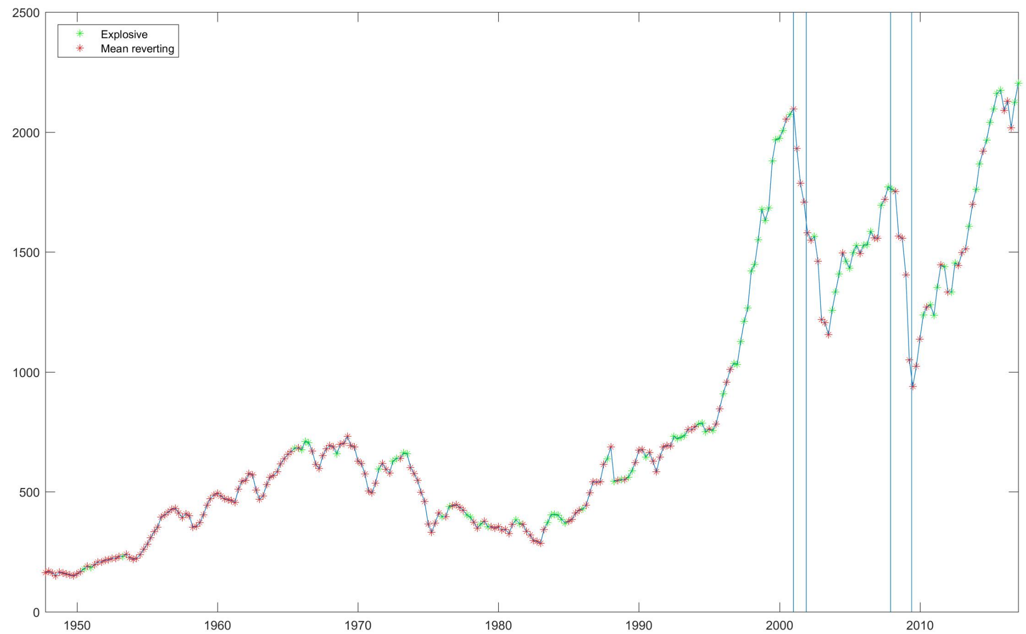


Figure 1.2: Regimes selected by TARSUR model for U.S. stock prices.

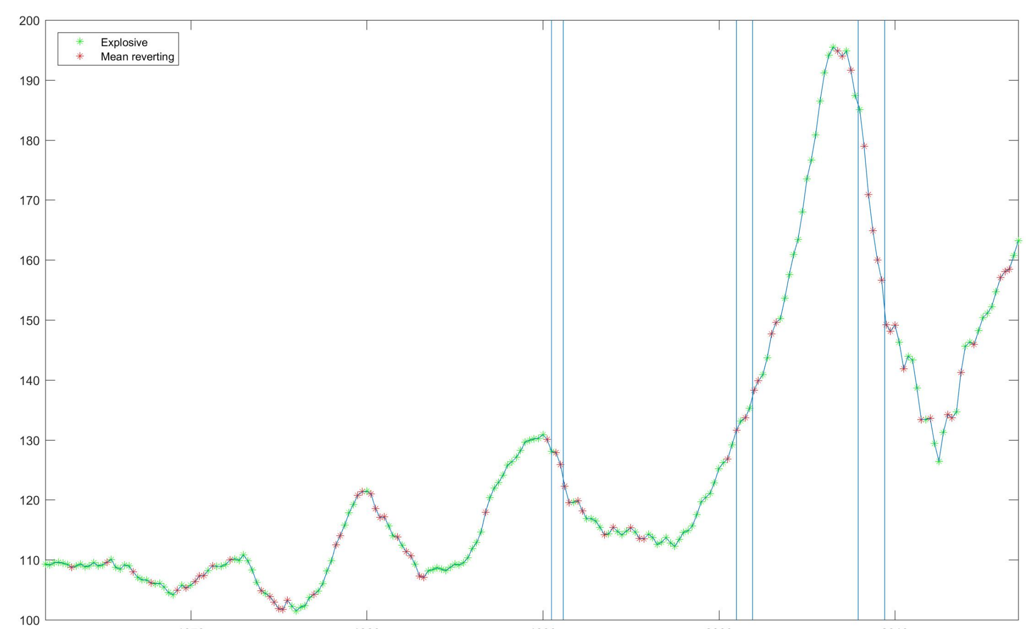


Figure 1.3: Regimes selected by TARSUR model for U.S. house prices.

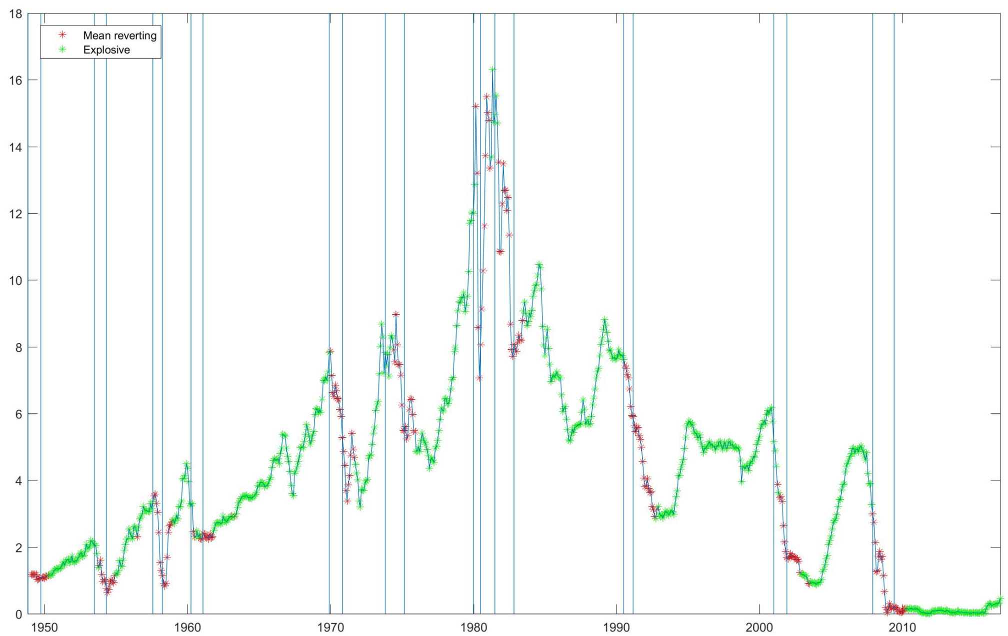


Figure 1.4: Regimes selected by TARSUR model for U.S. interest rates.

Chapter 2

Multiple Long-run Equilibria Through Cointegration Eyes

2.1 Introduction

Most of the financial and macroeconomic time series show persistent behavior (Granger, 1966), such that the unit root (UR) become stylized facts. The economic theory assesses the interrelation between economic variables with unit roots via long-run equilibrium relationships. When the relationship between variables is linear, we can test the existence of such relationships through the concept of cointegration (Granger, 1981 and Engel and Granger 1987). Indeed, when two or more economic variables are in equilibrium, then they must be cointegrated. For example, the literature suggests links between short-term interest rates and long-term interest rates (Froot 1989, Campbell and Shiller 1991, Hall, Anderson and Granger 1992, Choi and Wohar 1995), and also links between price and dividends via the present value models (PVM)derived by Campbell and Shiller (1987).

Nearly all the economic models in macroeconomics are highly non-linear, and this gives us good reasons to think that the actual data-generating process of the macroeconomic series is non-linear; for instance, the DSGE models predict a complicated non-linear relationship between the variables and between the past and future. Many other examples are the non-linear Taylor rules, environmental Kuznets curve, models for financial bubbles, and non-linear growth models. The concept of non-linear cointegration captures persistence with non-linear behavior of economic variables, and the research has moved in two directions. One line of research focuses on the short-term dynamic where the non-linearity arose from the adjustment mechanism toward a single linear equilibrium, Balke and Fomby (1997), Hansen and Seo (2002), Seo (2006), Kapetanios, and Shin and Snell (2006). Another line of research attempts to introduce non-linear behavior in the long-run equilibrium relation, see, for instance, Park and Phillips (2001) Chang, Park, and Phillips (2001), and Saikkonen and Choi (2004, 2010).

Economic theory has developed models with the presence of multiple equilibria, that is, Azariadis and Drazen (1990) propose a Diamond-type model that allows multiple, locally stable equilibria, while Benhabib, Schmitt-Grohé and Uribe, (2001) explore the condition where the Taylor rule generates multiple equilibria, but Time Series econometrics has not considered this type of non-linear cointegration with persistent variables.

The goal of this study is to analyze the presence of multiple long-run equilibria via a threshold cointegration framework where the non-linearity arises from introducing state-dependent behavior in the long-run equilibrium relationship. Specifically, we introduce threshold effects in the long-run equilibrium relationships to capture different relations between non-stationary variables during different stages of the business cycle. Also, we introduce methods to test for the presence of threshold cointegration and inference about the presence of multiple equilibria.

Our analysis focuses on the threshold effects induced by observable factors dictated by the economic theory (e.g., economic growth, unemployment growth), which are assumed to be stationary. The advantages such models offer, with respect to other non-linear models, are a straightforward estimation by the least-squares method and an intuitive economic interpretation of the non-linear relation. Very often, the economic theory does not specify the type of non-linearity that links different economic variables, but a threshold specification can be viewed as an approximation to a more general class of non-linear processes, see Petrucelli (1992).

Inference tools to assess both the presence of non-linear cointegration and threshold effects within the long-run equilibrium regression are essential in applied work, since omitting the presence of non-linear components in the long-run equilibrium relationship produces an inconsistent estimation of the cointegrating vector and leads to misinterpretation of the long-run equilibrium.

In related work, Saikkonen and Choi (2004) analyze the statistical properties of the test for detecting the presence of non-linearities in a cointegrating regression with a smooth transition functional form. Choi and Saikkonen (2010) test the null hypothesis of cointegration in the non-linear regression with $I(1)$ variables using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test proposed by Kwiatkowski, Phillips, Schmidt, and Shin (1992). Earlier work that analyzes the possibility of regime change in the cointegrating vector can be found in Seo (1998), as they test for structural breaks in the cointegrating vector and the adjustment mechanism. Pitarakis and Gonzalo (2006) present a test to detect for threshold effects in the long-run equilibrium equation when the threshold parameter, which determines the different regimes, is unknown and cannot be identified under the null hypothesis of no-threshold effects, but assuming the existence of the cointegration relation.

In Section 2.2, we show an example stating the conditions where the PVM can generate a threshold cointegration. Section 2.3 introduces the statistical model formally and the assumption used in the study. Section 2.4 proposes tests to identify these types of processes and its asymptotic distribution. Section 2.5 shows a relevant extension of the basic framework presented in Section 2.4, generalizing the testing procedure. Section 2.6 shows the finite sample performance of the tests proposed in this study. Section 2.7 illustrates two relevant empirical applications where multiple equilibria may arise. The first application analyzes the presence of multiple cointegration relationships between U.S. interest rates of different maturities. The second application analyzes the existence of multiple equilibrium relationships between prices and dividends. Section 2.8 concludes.

2.2 The Economic Model

The present value models are one of the simplest stochastic dynamic models in economics that define price as a linear function of the expected discounted dividend.

$$P_t = \mathbb{E}_t\left[\sum_{j=0}^{\infty}\left(\frac{1}{1+R}\right)^{j+1} d_{t+j}\right] \tag{2.1}$$

, where $\mathbb{E}_t(.)$ is the conditional expectation given this information up to time t, $\{d_t\}$ is the dividend, and R is the implicit discount rate and is assumed to be constant. Campbell and Shiller (1987) show that if $\{d_t\}$ is an I(1) process, $\{P_t\}$ must be also an $I(1)$ process, and they have to be cointegrated with the cointegrating vector $\begin{pmatrix}1 & \frac{1}{R}\end{pmatrix}'$, that is,

$$(P_t - \frac{1}{R}d_t) = \frac{1}{R}\mathbb{E}_t(\Delta P_t) \tag{2.2}$$

Assuming that the discount rate is a state-dependent variable driven by the business cycle, that is GDP growth and industrial production index growth.

$$R_{t-1} = R_1 I(z_{t-1} \leq r) + R_2 I(z_{t-1} > r) \tag{2.3}$$

, where $\{z_{t-1}\}$ is the threshold variable, and r is the threshold value that determines the different regimes (expansions and recessions, good times, and bad times). Then, the PVM can be written as

$$P_t = \mathbb{E}\Big[\sum_{j=0}^{\infty}\Big(\prod_{i=0}^{j}\frac{1}{1+R_{t+i-1}}\Big)d_{t+j}\Big] \tag{2.4}$$

Reordering the different terms, we get:

$$(P_t - \frac{1}{R_{t-1}}d_t) = \mathbb{E}_t(\Delta P_t) \tag{2.5}$$

From the structure of the discount factor in (2.3) and assuming that dividends follow RW, we can write (2.5) as follows:

$$\begin{aligned} P_t &= \frac{1}{R_1}I(z_{t-1} \leq r)d_t + \frac{1}{R_2}I(z_{t-1} > r)d_t + \tilde{e}_t \\ d_t &= d_{t-1} + \tilde{\varepsilon}_t \end{aligned} \tag{2.6}$$

, which is a triangular type representation that will allow us to test for threshold cointegration and the presence of multiple equilibrium relations between prices and dividends.

2.3 The Econometric Model

Consider the following non-linear cointegration regression with a threshold effect:

$$\begin{aligned} y_t &= \beta_1 I(z_{t-h} \leq r_0)x_t + \beta_2 I(z_{t-h} > r_0)x_t + e_t, \\ x_t &= x_{t-1} + \varepsilon_t, \end{aligned} \tag{2.7}$$

, where e_t and ε_t are the scalar stationary disturbance terms. z_{t-h} is the threshold variable, r_0 is the threshold parameter, h is the delay where we observe the threshold variable, which is not essential for our analysis and we set up $h = 1$, and $I(z_{t-1} \leq r_0)$ is an indicator function that takes value one when $z_{t-1} \leq r_0$ and zero otherwise. If e_t is an $I(0)$ process, then (2.7) is a cointegration relation with the cointegrating vector $\begin{pmatrix}1 & -\beta_1\end{pmatrix}'$ if $I(z_{t-1} \leq r_0)$, and $\begin{pmatrix}1 & -\beta_2\end{pmatrix}'$ if $I(z_{t-1} > r_0)$.

In the linear framework, the definition of cointegration says that two or more $I(1)$ variables are cointegrated if there exists a linear combination, which is $I(0)$. Overall, one of the biggest problems to define cointegration in the threshold framework and the non-linear world are derived from the concept of integration, which helps to classify linear processes as $I(0)$ and $I(1)$, depending on its stochastic properties but is unable to establish the properties of the non-linear processes. For instance, x_t defined in system (2.7) is an $I(1)$ process, but when it is multiplied with the indicator function, $I(z_{t-1} \leq r_0)x_t$, it is not a $I(1)$ process any more since taking the first difference does not make the series an $I(0)$ process; indeed, we can consider the many differences and never will be $I(0)$. Due to these difficulties, we follow the work of Gonzalo and Pitarakis (2006), Berenger-Rico and Gonzalo (2014a, 2014b), and use the concept of summability, which is the generalization of the concept of I(.)ness, to characterize the stochastic properties of the non-linear process.

Definition 1. *A stochastic process $y_t : t \in \mathbb{N}$ is said to be summable of order δ, or $S(\delta)$, if there exist a slowly varying function $L(T)$ and a deterministic sequence k_t such that*

$$S_T = \frac{1}{T^{\frac{1}{2}+\delta}} L(T) \sum_{t=1}^{T} (y_t - k_t) = O_p(1), \tag{2.8}$$

, where δ is the minimum real number that makes S_T bounded in probability.

Definition 1 shares the same spirit as the definition of $I(0)$, presented in Müller (2008) and Davison (2009), where they define a process to be an $I(0)$ if it satisfies the functional central limit theorem (FLCT). Once the generalization of the order of integration for the non-linear process is available, it is easy to extend the concept of cointegration for non-linear relationships and this can be done through the concept of co-summability developed by Berenguer-Rico and Gonzalo (2014b)

Definition 2. *Two summable stochastic processes, $y_t \sim S(\delta_y)$ and $x_t \sim S(\delta_y)$, are said to be co-summable if there exists $f(x_t) \sim S(\delta_y)$ such that $m_t = y_t - f(x_t)$ is $S(\delta_m)$, with $\delta_m = \delta_y - \delta$ for $\delta > 0$.*

Proposition 1. *(Berenger-Rico and Gonzalo, 2014). An I(d) with $d \geq 0$ is $S(d)$.*

Proposition 1 shows that any integrated linear process of order d is summable of order d, for example, a random walk with i.i.d innovations is an $I(1)$ process; then, it also must be $S(1)$ since the partial sum $\frac{1}{T^{3/2}} \sum_{t=1}^{T} x_t$, is convergent for $\delta = 1$.

To establish the order of summability of $\{I(z_{t-1} \leq r)x_t\}$, $\{I(z_{t-1} > r)x_t\}$ and y_t and the limiting distribution of the tests required to identify non-linear cointegration. In this section and Section 2.4 we will work under the following set of assumptions on $\{\varepsilon_t\}$, $\{e_t\}$ and $\{z_t\}$, which in the extension section we will relax A.3

Assumptions

(A.1) $\{\varepsilon_t, e_t, z_t\}$ is strictly both stationary and ergodic.

(A.2) $\{\varepsilon_t, e_t, z_t\}$ 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) e_t is independent of $\mathcal{F}_{t-1} = \{(\varepsilon_{t-j}, e_{t-j}, z_{t-j}), j \geq 1\}$, $E(e_t) = 0$ and $E\left|e_t\right|^4 = k < \infty$.

(A.4) z_t has a continuous and increasing distribution function $P(.)$.

(A.5) The threshold value r is in a closed and bounded subset of the space of the threshold variable, that is, $r \in [r_L, r_H]$.

(A.6) $E(\varepsilon_t) = 0$ with $E|\varepsilon_t|^{2+\rho} < \infty$ for some $\rho > 0$.

Assumptions A.1, A.2, and A.3 are similar to the assumptions proposed by Caner and Hansen (2001) to establish the convergence of the partial sum $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} I(z_{t-1} \leq r)e_t$, which will allow us to derive the asymptotic distribution of the tests for the presence of non-linear cointegration. Assumption A.1 requires that the threshold variable is a stationary process, ruling out the possibility of $\{z_t\}$ being an $I(1)$ process, but general enough to admit a rich class of stochastic processes. A.3 is very restrictive in the sense it rules out the possibility of e_t being serially correlated. The finite fourth moment assumption is not necessary for the invariance principle, but it is required to establish the tightness properties of the above empirical process. In the following sections, we are going to abandon this assumption and allow it to follow a stationary linear process. A.5 restricts the parameter space of r ensuring that there are enough observations in each regime and assures the existence of non-degenerated limiting distribution for the test statistic of interest. We choose r_L and r_H such that $P(z_{t-1} \leq r_L) = \theta > 0$ and $P(z_{t-1} > r_H) = 1 - \theta$, where θ is commonly selected in the threshold literature to be 10% of 15% (see Hansen (2000), Caner and Hansen (2001), Gonzalo and Pitarakis (2006)).

Proposition 2 establishes the order of summability of $\{I(z_{t-1} \leq r)x_t\}$, $\{I(z_{t-1} > r)x_t\}$ and y_t.

Proposition 2. *Under assumption A.1, A.2, and A.3, the processes $\{I(z_{t-1} \leq r)x_t\}$, $\{I(z_{t-1} > r)x_t\}$ and $\{y_t\}$ are $S(1)$ and $\{e_t\}$ is $S(0)$, $(I(0))$ and, therefore, $\{y_t\}$ and $\{x_t\}$ are co-summable (non-linear cointegrated).*

2.4 Inference

Let us rewrite the system (2.7) as follows:

$$\begin{aligned} y_t &= \beta_2 x_t + \gamma I(z_{t-1} \leq r_0)x_t + e_t, \\ x_t &= x_{t-1} + \varepsilon_t \end{aligned} \tag{2.9}$$

, where $\gamma = (\beta_1 - \beta_2)$. Also, we can define the model with state dependent constants, namely α_1 and α_2.

$$
\begin{aligned}
y_t &= \alpha_1 I(z_{t-1} \leq r_0) + \alpha_2 I(z_{t-1} > r_0) + \beta_2 x_t + \gamma I(z_{t-1} \leq r_0) x_t + e_t, \\
x_t &= x_{t-1} + \varepsilon_t
\end{aligned}
\tag{2.10}
$$

Threshold cointegration requires two conditions that must be tested:

- The first condition is that the residuals e_t must be an $I(0)$. If this condition is not satisfied, neither we have linear cointegration nor threshold cointegration.
- The second condition is the presence of a threshold effect. If there is no-threshold effect in the long-run equilibrium equation, then the cointegration relation is linear.

In this section, we present a testing procedure to check both conditions.

2.4.1 Residual Based Test for Non-linear Cointegration

We test for the presence of threshold cointegration by testing if e_t is an $I(0)$ process, using the residual based KPSS test proposed by Shin (1994). Testing for cointegration in a non-linear framework is not new, for example, Choi and Saikkonen (2004, 2010) present the residual based KPSS test to detect the presence of cointegration assuming that the non-linear functions are continuous, ruling out the possibility of threshold structures. The proposed KPSS test uses specifications (2.9) and (2.10), which are very general in the sense that we will be able to detect both, threshold cointegration and linear cointegration. If e_t is an $I(0)$ process and $\gamma \neq 0$, then y_t and x_t are threshold cointegrated. If e_t is an $I(0)$ process but $\gamma = 0$, then y_t and x_t are linearly cointegrated.

To set up the test, define $m_t = m_{t-1} + u_t$ and let $v_{1t} = m_t + e_t$; our aim is to test if the variance of $\{u_t\}$ is zero, that is, $\sigma_u^2 = 0$. Under the null of cointegration, $m_t = m_0$, where m_0 is a constant that produces $v_{1t} = m_0 + e_t$ and v_{1t} will be an $I(0)$ process. Testing the null hypothesis of threshold cointegration versus the alternative of no-threshold cointegration can be done by testing

$$
\begin{aligned}
H_0 &: \sigma_u^2 = 0 \\
H_1 &: \sigma_u^2 > 0
\end{aligned}
\tag{2.11}
$$

To perform this test, we must

1. Recover $\hat{e}_{\alpha,t}$ and $\hat{e}_t$, which are the CLS residuals from the threshold cointegrating regressions (2.10) and (2.9), respectively. Construct $S_{\alpha,t} = \sum_{j=1}^{t} \hat{e}_{\alpha,j}$ and $S_t = \sum_{j=1}^{t} \hat{e}_j$ be the partial sum process based on these residuals.

2. Then, the KPSS statistics are

$$CI = n^{-2}\sum_{t=1}^{n} S_t^2/\hat{\sigma}_e^2(l) \tag{2.12}$$

$$CI_\alpha = n^{-2}\sum_{t=1}^{n} S_{\alpha,t}^2/\hat{\sigma}_{\alpha,e}^2(l) \tag{2.13}$$

where

$$\hat{\sigma}_e^2(l) = \frac{1}{n}\sum_{j=1}^{n}\hat{e}_t^2 + \frac{2}{n}\sum_{i=1}^{l} L(i,l)\sum_{t=i+i}^{n}\hat{e}_t\hat{e}_{t-i} \tag{2.14}$$

$$\hat{\sigma}_{\alpha,e}^2(l) = \frac{1}{n}\sum_{j=1}^{n}\hat{e}_{\alpha,t}^2 + \frac{2}{n}\sum_{i=1}^{l} L(i,l)\sum_{t=i+i}^{n}\hat{e}_{\alpha,t}\hat{e}_{\alpha,t-i} \tag{2.15}$$

and $L(i,1) = 1 - i/(l+i)$ is the Barlett window.

Following the work of Choi and Saikonnen (2004, 2010) and Shin (1994), under the null of threshold cointegration, we derive the asymptotic distribution of the KPSS test, namely CI and CI_α. Note that even under assumption A.3, estimation of the single-equation LS estimator involves second-order bias due to the presence of correlations between x_t and e_t. To simplify this problem, we can assume strict exogeneity between x_t and e_t or use an efficient estimation proposed in the next section.

Proposition 3. *Suppose that assumptions A.1, A.2, A.3, A.4, and A.6 hold and assume that x_t is strictly exogenous w.r.t e_t, then the test statistic CI_α and CI have the following limiting distribution.*

$$\begin{aligned} CI &\Rightarrow \int_0^1 Q^2 \\ CI_\alpha &\Rightarrow \int_0^1 Q_\alpha^2 \end{aligned} \tag{2.16}$$

where

$$Q = W_e - \Big(\int_0^s W_x\Big)\Big(\int_0^1 W_x^2\Big)^{-1}\Big(\int_0^1 W_x dW_e\Big) \tag{2.17}$$

$$Q_\alpha = V_e - \Big(\int_0^s W_x^\alpha\Big)\Big(\int_0^1 (W_z^\alpha)^2\Big)^{-1}\Big(\int_0^1 W_x^\alpha dW_e\Big) \tag{2.18}$$

$$\tag{2.19}$$

, where W is the standard Brownian process, $W_x^\alpha = W_x - \int_0^1 W_x$ is the standard demeaned Browninan motion, and $V_e = W_e - sW_e$ is a standard Browninan bridge.

Proposition 3 shows that under strict exogeneity, the asymptotic distribution of the test does not depend on nuisance parameters and is the same distribution for testing the null of linear cointegration using KPSS (Shin, 1994). Note that this distribution in the linear case depends on the number of regressors included in the regression, but as observed in Proposition 3, the threshold regression depends on the number of non-threshold regressors included in the regression.

Now, we show the consistency of the KPSS test and the limiting distribution of statistics under the alternative of no cointegration ($\sigma_u^2 > 0$). The limited distribution is based on the same functionals of Brownian as in the tests presented by Phillips and Ouliaris (1990), but the form of the limiting distributions is different.

Proposition 4. *Under the alternative hypothesis of no cointegration $\sigma_u^2 > 0$, the statistics CI and CI_α, normalized by l/n have the following distributions*

$$\frac{l}{n} CI \Rightarrow \int_0^1 \Big(\int_0^s Q_p \Big)^2 / \int_0^1 Q_p^2 \tag{2.20}$$

$$\frac{l}{n} CI_\alpha \Rightarrow \int_0^1 \Big(\int_0^s Q_p^\alpha \Big)^2 / \int_0^1 (Q_p^\alpha)^2 \tag{2.21}$$

(2.22)

, where

$$Q_p = W_m + W_x \Big(\int_0^1 W_x^2 \Big)^{-1} \Big(\int_0^1 W_x W_m \Big) \tag{2.23}$$

$$Q_p^\alpha = W_m^\alpha + W_x^\alpha \Big(\int_0^1 (W_x^\alpha)^2 \Big)^{-1} \Big(\int_0^1 W_x^\alpha W_m^\alpha \Big) \tag{2.24}$$

(2.25)

Proposition 4 shows that the tests CI and CI_α are consistent and diverge at rate (n/l) under the alternative of no cointegration. As observed in Kwiatkowsky, Phillips, Schmidt and Shin (1992) and Shin (1994), the test depends on our choice of the lag truncation of the long-run variance estimation l, and this choice is critical for the test to have good power.

2.4.2 Testing for Threshold Effect

Once we obtain cointegration in the first stage, we proceed to test for the threshold effect, and the goal of this section is to construct a test for analyzing the null of no-threshold effects in the long run equation versus the alternative of a threshold effect. Assuming that $r \in [r_L, r_H]$, this can be examined by testing:

$$\begin{aligned} H_0 &: \gamma = 0 \\ H_1 &: \gamma \neq 0 \end{aligned} \tag{2.26}$$

in the first equation of system (2.9) and (2.10), such that if $\gamma = 0$, the long-run equation becomes $y_t = \beta_2 x_t + e_t$, which is the linear cointegration, and $y_t = \alpha_1 I(z_{t-1} \leq r_0) + \alpha_2 I(z_{t-1} > r_0) + \beta_2 x_t + e_t$, which is also the linear cointegration case with a possible state dependent drift, whether $\alpha_1 = \alpha_2$ or $\alpha_1 \neq \alpha_2$.

The asymptotic distribution of the test depends on whether the threshold parameter r is known or unknown, and in the latter case, whether it can be identified or unidentified under the null hypothesis.

2.4.2.1 Threshold Parameter is Known

The case where the threshold parameter is known, $r = r_0$, is relevant for explanatory reasons as well as for the cases in which the different regimes are predetermined, for example, from the economic theory or the sign of the threshold variable. In this case, the proposed $t-$statistic for $\gamma = 0$, $t_{\gamma=0}(r_0)$, has the following asymptotic distribution

Proposition 5. *Suppose that the threshold value is known, $r = r_0$, and assumptions A.1, A.2, A.3, A.4, and A.6 hold. Under the null of no-threshold effects, $t_{\gamma=0}(r_0)$, the statistic has the following asymptotic distribution:*

$$t_{\gamma=0}(r_0) \Rightarrow \mathcal{N}(0,1) \tag{2.27}$$

2.4.2.2 Threshold Parameter is Unknown but Identified

In the case where the DGP has a threshold effect in the drift term, we can estimate by using the LS threshold value, $\hat{r}_n$, in the first stage, before testing for the threshold effect. This is possible because under the null $H_0 : \gamma = 0$, we can estimate the super-consistency of the threshold parameter (T-consistent) and it can be taken as known.

Proposition 6. *Suppose that assumptions A.1-A.6 hold, under $H_0 : \gamma = 0$, as $n \to \infty$ (i) $\hat{r}_n \to_p r_0$, and (ii) $n|\hat{r}_n - r_0| = O_p(1)$.*

In this case, we can use the t-statistics for $\gamma = 0$ evaluated at the estimated threshold parameter $\hat{r}_n$, which takes us back to Proposition 5.

Proposition 7. *Under assumption A.1-A.6, under $H_0 : \gamma = 0$ and $\alpha_1 \neq \alpha_2$ the $t_{\gamma=0}(\hat{r}_n)$ statistic has the following asymptotic distribution:*

$$t_{\gamma=0}(\hat{r}_n) \Rightarrow \mathcal{N}(0,1) \tag{2.28}$$

2.4.2.3 Threshold Parameter is Unknown and Unidentified

Under assumption A.3, testing for the presence of threshold effects, when the threshold value is unknown and unidentified, is studied extensively by Gonzalo and Pitarakis (2006). For completeness, we include their results in this section.

When the threshold value r_0 is unknown and unidentified under the null of no-threshold effect, the proposed test is the Supremum of the Wald statistics, W_N,

$$W_N = \sup_{r \in [r_L, r_H]} W_n(r) \tag{2.29}$$

, where $W_n(r) = t^2_{\gamma=0}(r)$. Then, the asymptotic distribution of the Wald statistics is given as:

Proposition 8. *Suppose that assumptions A.1-A.6 hold. Consider the long-run equation (2.7) under the null $H_0 : \gamma = 0$, the W_n statistics has the following asymptotic distribution:*

$$W_n \Rightarrow \underset{r \in [r_L, r_H]}{sup} \frac{\left(\int W_\varepsilon(s) dV_e(s,\lambda)\right)^2}{\lambda(1-\lambda)\int W_\varepsilon(s)^2 ds} \equiv \underset{r \in [r_L, r_H]}{sup} \frac{[BB(\lambda)]^2}{\lambda(1-\lambda)} \tag{2.30}$$

, where $\lambda = P(z_{t-1} \leq r)$, $W_\varepsilon(.)$ is the Brownian motion and $V_e(s,\lambda)$ is a Kiefer-Muller process 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 is due to the fact that $W_\varepsilon(s) = W_\varepsilon(s,1)$ and $V_e(s,\lambda)$ are independent.

Note that the asymptotic distribution presented in Proposition 8 is the same as the one for testing structural breaks, according to Andrews (1993, 2003), and this distribution is tabulated by Estrella (2003) for different values of θ.

2.5 Extensions to $I(0)$ Cointegrating Errors

Assuming that e_t is an independent process with respect to its own past $\{e_{t-1}, e_{t-2}, \ldots\}$ in A.3 is a strong assumption, ruling out the possibility of $\{e_t\}$ being serially correlated and this will pose a problem for example, when testing for cointegration, the null of cointegration possibly is rejected because e_t is autocorrelated. Also, this assumption does not allow for short-term dynamics, where any disequilibria are instantly corrected.

As discussed by Gonzalo and Pitarakis (2006), there is a non-natural extension for the weak convergence of the partial sum $G_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]} e_t I(z_{t-1} \leq r)$ as shown in the work of Caner and Hansen (2001, Theorem 1, assuming that e_t follows an i.i.d process) in which both the marks e_t as well as z_t are the general stationary process. To have a tractable limit theorem for elements like G_n, in this section, we relax our earlier assumption A.3 by allowing e_t to follow a linear process, more formally

(B.3): Let $e_t = \sum_{j=0}^{\infty} a_j v_{t-j}$, where $\sum_{t=1}^{\infty} \frac{1}{\sqrt{t}}\Big(\sum_{j=t}^{\infty} a_j^2\Big)^{1/2} < \infty$ with $a_0 = 1$ and $\{v_t\}$ satisfy the following conditions $E(v_t) = 0$, $E(v_t^2) = \sigma_v^2$, $E|v_t|^4 < \infty$, and v_t is independent with respect $F_{t-1} = \sigma\{v_{t-j} : j \geq 1\}$ and independent of z_{t-j} for $j = \pm 1, \pm 2, \ldots$.

The assumption $\sum_{t=1}^{\infty} \frac{1}{\sqrt{t}}\Big(\sum_{j=t}^{\infty} a_j^2\Big)^{1/2} < \infty$ and the independence between v_t and z_j are needed to derive the invariance principle of $G_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]} e_t I(z_{t-1} \leq r)$. Note that the requiring $\sum_{t=1}^{\infty} \frac{1}{\sqrt{t}}\Big(\sum_{j=t}^{\infty} a_j^2\Big)^{1/2} < \infty$ is slightly stronger than assuming $\sum_{j=0}^{\infty} |a_n| < \infty$, as pointed out by Wu (2002, Lemma 1). Assuming independence between v_t and z_j can be strong but it is used, for example, in the study of Caner and Hansen (2001), which requires the independence of v_t and z_{t-j} for $j = 1, 2 \ldots$, also the work of Gonzalo and Pitarakis (2006) requires that v_t is independent of z_{t+q-j} for $j = 1, 2 \ldots$, when e_t follows a moving average of finite order q.

The functional central limit theorem (CLT) result for G_n is derived using the results in Peligrad and Utev (2005) and Mervelede, Peligrad and Utev (2006),

Proposition 9. *Under assumptions A.1, A.2, and B.3,*

$$G_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]} e_t I(z_{t-1} \leq r) \Rightarrow \sqrt{\eta(\lambda)} W_{eI}(s) \equiv G_e(s, \lambda) \tag{2.31}$$

, where $\eta(\lambda) = \lim_{n\to\infty} \frac{E(\sum_{t=1}^{n} e_t I(z_{t-1} \leq r))^2}{n}$.

Proposition 9 is an extension of the result from the work of Gonzalo and Pitarakis (2006, Proposition 3), where they derive the convergence of the partial sum G_n by allowing e_t to be a finite moving average process. Also, the proposition specializes the result from Caner and Hansen (2001) by setting $a_j = 0$ for all $j \geq 1$ such that $e_t = v_t$, then $\eta(\lambda) = \lambda\sigma_v^2$ and $G_e(s, \lambda) = \sqrt{\sigma_v \lambda} W_e(s) = \sqrt{\sigma_v} W_e(s, \lambda)$.

Then, it is easy to show that under assumptions A.1, A.2, A.4, A.6, and B.3, for $\xi_t = \begin{pmatrix} \varepsilon_t & e_t & e_t I(z_{t-1} \leq r) \end{pmatrix}'$ we have the following result

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{[ns]}\xi_t \Rightarrow \mathbb{B}(s) \equiv \Big(B_x(s) \quad B_e(s) \quad G_e(s,\lambda)\Big)' \tag{2.32}$$

with covariance matrix

$$\Omega = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon,e} & \sigma_{\varepsilon,e,\lambda} \\ \sigma_{\varepsilon,e} & \sigma_e^2 & \sigma_{e,e,\lambda} \\ \sigma_{\varepsilon,e,\lambda} & \sigma_{e,e,\lambda} & \eta(\lambda) \end{pmatrix} \text{ such that } \Omega = LL' \text{ we have } \mathbb{B}(s) = L\mathbb{W}(s) \tag{2.33}$$

, where $\mathbb{W}(s)$ is the vector of the Wiener process. Since e_t is a linear process and assuming B.3, we can write $\sigma_e^2 = \sigma_v^2 C(1)^2$, $\sigma_{e,e,\lambda} = \lambda\sigma_v^2 C(1)^2$ and $\eta(\lambda) = \lambda^2\sigma_v^2 C(1)^2 + \mathbb{G}$, where $C(1) = \sum_{j=0}^{\infty} a_j$ and $\mathbb{G} = lim_{n\to\infty}\frac{1}{n}E(\sum_{t=1}^{n} e_t(I(z_{t-1} \le r) - \lambda))^2$.

The second step is to show the limiting distribution of the process $\frac{1}{n}\sum_{t=1}^{n} x_t I(z_{t-1} \le r)e_t$. It is well known that in certain cases when G_n is not a martingale process, then $\frac{1}{n}\sum_{t=1}^{n} x_t I(z_{t-1} \le r)e_t$ does not converge to $\int_0^1 B_x(s)dG_e(s,\lambda)$, see for example Phillips (1987). Using the martingale approximation proposed in Hansen (1992), we can derive the following result:

Theorem 1. *Under assumption A.1., A.2, and B.3,*

$$\frac{1}{n}\sum_{t=1}^{n} x_t I(z_{t-1} \le r)e_t \Rightarrow \int_0^1 B_x(s)dG_e(s,\lambda) + \lambda E(\varepsilon_i e_i) + \lambda\Lambda_1 \tag{2.34}$$

, where $\Lambda_1 = lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{\infty} E(\varepsilon_i e_j)$.

From Theorem 4.1 in Hansen (1992), we know that $\frac{1}{n}\sum_{t=1}^{n} x_t e_t \Rightarrow \int_0^1 B_x(s)dB_e(s) + E(\varepsilon_i e_i) + \Lambda_1$. Equipped with Theorem 1 and Theorem 3 in Caner and Hansen (2001), we can show the convergence of the LS estimate for the parameter of interest β_2 and γ:

Lemma 1. *Under assumptions A.1, A.2, and B.3, when no drift is considered*

$$T(\hat{\Gamma} - \Gamma) \Rightarrow \begin{pmatrix} \frac{\lambda\left(\int_0^1 B_x(s)dB_e(s) - \int_0^1 B_x(s)dG_e(s,\lambda) + (1-\lambda)[\Lambda + E(\varepsilon_i e_i)]\right)}{\lambda(1-\lambda)\int_0^1 B_x(s)^2 ds} \\ \frac{\int_0^1 B_x(s)dG_e(s,\lambda) - \lambda\int_0^1 B_x(s)dB_e(s)}{\lambda(1-\lambda)\int_0^1 B_x(s)^2 ds} \end{pmatrix} \tag{2.35}$$

when considering a state-dependent drift

$$T(\hat{\Gamma}^\alpha - \Gamma^\alpha) \Rightarrow \begin{pmatrix} \frac{\lambda\left(\int_0^1 B_x^\alpha(s)dB_e(s) - \int_0^1 B_x^\alpha(s)dG_e(s,\lambda) + (1-\lambda)[\Lambda + E(\varepsilon_i e_i)]\right)}{\lambda(1-\lambda)\int_0^1 (B_x^\alpha(s))^2 ds} \\ \frac{\int_0^1 B_x^\alpha(s)dG_e(s,\lambda) - \lambda\int_0^1 B_x^\alpha(s)dB_e(s)}{\lambda(1-\lambda)\int_0^1 (B_x^\alpha(s))^2 ds} \end{pmatrix} \tag{2.36}$$

2.5.1 Testing for Non-Linear Cointegration

In the case where the regressors are strictly exogenous, that is, $\sigma_{\varepsilon,e} = \sigma_{\varepsilon,e,\lambda} = 0$, as it happens in the previous section, the KPSS test is free of nuisance parameters and the distribution of CI and CI_α are the same as in Proposition 3.

In general, assuming $\sigma_{\varepsilon,e} = \sigma_{\varepsilon,e,\lambda} = 0$ is very restrictive in time series modeling. Note that as it happens in linear cointegration, the regressors x_t may be correlated with e_t, and the asymptotic result developed for the KPSS test derived previously is not robust to the problem of endogenous regressors, since it would involve nuisance parameters. In general, when e_t and ε_t is serially correlated it is not enough to consider only the contemporaneous relationship between e_t and ε_t and, therefore, we have to consider the past and future values of Δx_t as additional regressors. Following the work of Shin (1994) and Saikkonen (1991), we require the following conditions:

Condition 1: The spectral density matrix $f_{uu}(\omega)$ is bounded away from zero $f_{uu}(\omega) > a$, $\omega \in [0, \pi]$ and $a > 0$, where $u_t = \begin{pmatrix} \varepsilon_t & e_t \end{pmatrix}'$.

Condition 2: The covariance function of u_t is absolutely summable, $\sum_{j=-\infty}^{\infty} ||E(u_t u'_{t+j})|| < \infty$, where $||.||$ is the Euclidean norm.

Under the conditions stated above, we can write $e_t = \sum_{j=-\infty}^{\infty} \pi_j \varepsilon_{t-j} + \tilde{e}_t$, where $\sum_{j=-\infty}^{\infty} |\pi_j| < \infty$ and $\tilde{e}_t$ is a stationary process such that $E(\varepsilon_{t+j}\tilde{e}_t) = 0$ for $j = \pm 1, \pm 2, \ldots$. As discussed in Shin (1994) and Saikkonen (1991), we cannot regress a model with an infinite number of lags and leads of $\Delta x_t = \varepsilon_t$. Since $\{\pi_j\}$ is absolutely summable, that is, $\pi_j \approx 0$ for $|j| > K$, and K is large enough, we can truncate the regression for using K lags and leads. The choice of K must satisfy the following condition, as $n \to \infty$ and $K \to \infty$

$$K^3/n \to 0, \quad \text{and} \quad n^{1/2} \sum_{|j|>K}^{\infty} |\pi_j| \to 0 \tag{2.37}$$

For further details see Saikkonen (1991). The assumption given in (2.37) is sufficient to obtain the asymptotic distribution of the KPSS test, therefore, for a chosen lag truncation K we can rewrite (2.9),

$$\begin{aligned} y_t &= \beta_2 x_t + \gamma I(z_{t-1} \leq r_0) x_t + \sum_{j=-K}^{K} \pi_j \Delta x_{t-j} + \tilde{e}_t^* \\ x_t &= x_{t-1} + \varepsilon_t \end{aligned} \tag{2.38}$$

, where $\tilde{e}_t^* = \sum_{|j|>K}^{\infty} \pi_j + \tilde{e}_t$, and similarly (2.10),

$$\begin{aligned} y_t &= \alpha_1 I(z_{t-1} \le r_0) + \alpha_2 I(z_{t-1} > r_0) + \beta_2 x_t + \gamma I(z_{t-1} \le r_0) x_t + \sum_{j=-K}^{K} \pi_{\alpha,j} \Delta x_{t-j} + \tilde{e}_t^* \\ x_t &= x_{t-1} + \varepsilon_t \end{aligned} \tag{2.39}$$

We now proceed to construct the stochastic process $\tilde{B}_n$ by $\tilde{B}_n = \frac{1}{n}\sum_{t=1}^{[ns]} \tilde{\xi}_t$, where $\tilde{\xi}_t = \begin{pmatrix}\varepsilon_t & \tilde{e}_t & \tilde{e}_t I(z_{t-1} \le r)\end{pmatrix}'$. $\tilde{B}_n$ converges weakly to $\tilde{B}$ as $n \to \infty$, where $\tilde{B}$ is the vector of Brownian motion with the following block diagonal covariance matrix, $\tilde{\Omega} = \operatorname{diag}\{\sigma_\varepsilon, \Omega_{e.\varepsilon}\}$. The elements of the vector $\tilde{B}$ are $\tilde{B} = \begin{pmatrix}B_x & B_{e.\varepsilon} & G_{e.\varepsilon}\end{pmatrix}'$, where $B_{e,\varepsilon} = B_e - \sigma_{\varepsilon,e}\sigma_\varepsilon^{-2} B_x$ and $G_{e,\varepsilon} = G_e - \lambda \sigma_{\varepsilon,e}\sigma_\varepsilon^{-2} B_x$. By construction, $B_{e.\varepsilon}$ and $G_{e.\varepsilon}$ are independent of B_x.

Lemma 2. *Let $\tilde{\beta}_2$, $\tilde{\gamma}$ and $\tilde{\pi}_j$, $\tilde{\beta}_2^\alpha$, $\tilde{\gamma}^\alpha$ and $\tilde{\pi}_j{}^\alpha$ be the ordinary least squares (OLS) estimators obtained from (2.38) and (2.39). Then,*

$$T(\tilde{\Gamma} - \Gamma) \Rightarrow \frac{1}{\lambda(1-\lambda)\int_0^1 B_x^2} \begin{pmatrix} \lambda\Big(\int_0^1 B_x dB_{e,x} - \int_0^1 B_x dG_{e,x}\Big) \\ \int_0^1 B_x dG_{e,x} - \lambda \int_0^1 B_x dB_{e,x} \end{pmatrix} \tag{2.40}$$

$$T(\tilde{\Gamma}^\alpha - \Gamma^\alpha) \Rightarrow \frac{1}{\lambda(1-\lambda)\int_0^1 (B_x^\alpha)^2} \begin{pmatrix} \lambda\Big(\int_0^1 B_x^\alpha dB_{e,x} - \int_0^1 B_x^\alpha dG_{e,x}\Big) \\ \int_0^1 B_x^\alpha dG_{e,x} - \lambda \int_0^1 B_x^\alpha dB_{e,x} \end{pmatrix} \tag{2.41}$$

Also,

$$\left(\frac{n}{K}\right)^{1/2} \sum_{j=-K}^{K} (\tilde{\pi}_j - \pi_j) = O_p(1), \qquad \left(\frac{n}{K}\right)^{1/2} \sum_{j=-K}^{K} (\tilde{\pi}_j^\alpha - \pi_j) = O_p(1), \tag{2.42}$$

From the result above, let $\hat{\tilde{e}}_t^*$ and $\hat{\tilde{e}}_{\alpha,t}^*$ be the OLS residuals from (2.38) and (2.39) respectively, and $\tilde{S}_t = \sum_{j=1} \hat{\tilde{e}}_t^*$ and $\tilde{S}_{\alpha,t} = \sum_{j=1} \hat{\tilde{e}}_{\alpha,t}^*$. Let $\tilde{\sigma}_e^2(l)$ and $\tilde{\sigma}_{\alpha,e}^2(l)$ be the estimators defined in (2.14) and (2.15) based on $\hat{\tilde{e}}_t^*$ and $\hat{\tilde{e}}_{\alpha,t}^*$. Then, the modified statistics for cointegration are defined ass

$$\tilde{CI} = n^{-2} \sum_{t=1}^{n} \tilde{S}_t^2 / \tilde{\sigma}_e^2(l), \qquad \tilde{CI}_\alpha = n^{-2} \sum_{t=1}^{n} \tilde{S}_{\alpha,t} / \tilde{\sigma}_{\alpha,e}^2(l) \tag{2.43}$$

Theorem 2. *The limiting distribution of the KPSS test obtained using the modified statistic, $\tilde{CI}$ and $\tilde{CI}_\alpha$ are the same as in Proposition 3.*

2.5.2 Testing for Threshold Effect

From (2.33) and assumption B.3 we can see that

$$\mathbb{B}(s) = \begin{pmatrix} \sigma_\varepsilon W_x(s) \\ \frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon} W_x(s) + \left[C(1)^2\sigma_v^2 - \left(\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon}\right)^2\right]^{1/2} W_e(s) \\ \lambda\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon} W_x(s) + \lambda\left[C(1)^2\sigma_v^2 - \left(\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon}\right)^2\right]^{1/2} W_e(s) + \sqrt{\mathbb{G}} W_{eI}(s,\lambda) \end{pmatrix} \tag{2.44}$$

Testing for threshold effects, we are interested in the distribution of $\hat{\gamma}$ and $\hat{\gamma^\alpha}$ from Lemma 1.

$$\begin{aligned} T(\hat{\gamma}-\gamma) &\Rightarrow \frac{\int_0^1 B_x(s)dG_e(s,\lambda) - \lambda\int_0^1 B_x(s)dB_e(s)}{\lambda(1-\lambda)\int_0^1 (B_x(s))^2 ds} = \frac{\sqrt{\mathbb{G}}}{\sigma_\varepsilon\lambda(1-\lambda)}\frac{\int_0^1 W_x(s)dW_{eI}(s)}{\int_0^1 W_x^2(s)ds} \\ T(\hat{\gamma}^\alpha-\gamma^\alpha) &\Rightarrow \frac{\int_0^1 B_x^\alpha(s)dG_e(s,\lambda) - \lambda\int_0^1 B_x^\alpha(s)dB_e(s)}{\lambda(1-\lambda)\int_0^1 (B_x^\alpha(s))^2 ds} = \frac{\sqrt{\mathbb{G}}}{\sigma_\varepsilon\lambda(1-\lambda)}\frac{\int_0^1 W_x^\alpha(s)dW_{eI}(s)}{\int_0^1 (W_x^\alpha)^2(s)ds} \end{aligned} \tag{2.45}$$

Since W_x and W_{eI} are independent, it is well known that $\frac{\int_0^1 W_x(s)dW_{eI}(s)}{\sqrt{\int_0^1 W_x^2(s)ds}} \equiv N(0,1)$. Testing for the null of $\gamma = 0$ in this framework, which is free of nuisance parameters, is feasible when $\mathbb{G}$ can be estimated under the null hypothesis, which are cases where the threshold value r is known or identified under the null.

2.5.2.1 When r is Known

When the threshold value is known, $r = r_0$, we can recover the residuals $\hat{e}_t$ from the model (2.9), such that $\mathbb{G}$ can be estimated as in Phillips (1987) and Phillips and Perron (1988) by

$$\hat{\mathbb{G}} = \frac{1}{n}\sum_{j=1}^{n} \hat{e}_t^2(I(z_{t-1} \le r_0) - \bar{\lambda}_0)^2 + \frac{2}{n}\sum_{i=1}^{l} L(i,l)\sum_{t=i+1}^{n} \hat{e}_t\hat{e}_{t-i}(I(z_{t-1} \le r_0) - \bar{\lambda}_0)(I(z_{t-1-i} \le r_0) - \bar{\lambda}_0) \tag{2.46}$$

, where $\bar{\lambda}_0 = \frac{1}{n}\sum I(z_{t-1} \le r_0)$ is a consistent estimator of $P(z_{t-1} \le r_0) = \lambda_0$. Now we can define a simple transformation of the conventional test statistic for testing $\gamma = 0$, which eliminates the nuisance parameters in the distribution:

$$\tilde{t}_{\gamma=0}(r_0) = \hat{\gamma}(r_0)\sqrt{\frac{\bar{\lambda}_0(1-\bar{\lambda}_0)}{\hat{\mathbb{G}}((X(r_0)'X(r_0))^{-1})_{22}}} \tag{2.47}$$

, where $((X(r_0)'X(r_0))^{-1})_{22}$ is the element (2,2) in the following matrix

$$(X(r_0)'X(r_0))^{-1} = \begin{pmatrix} \frac{1}{n^2}\sum_{t=1}^{n} x_t^2 & \frac{1}{n^2}\sum_{t=1}^{n} x_t^2 I(z_{t-1} \leq r_0) \\ \frac{1}{n^2}\sum_{t=1}^{n} x_t^2 I(z_{t-1} \leq r_0) & \frac{1}{n^2}\sum_{t=1}^{n} x_t^2 I(z_{t-1} \leq r_0) \end{pmatrix}^{-1} \tag{2.48}$$

Proposition 10. *When the threshold values is known, that is, $r = r_0$, with assumptions A.1, A.2, B.3, A.4, and A.6, under the null $H_0 : \gamma = 0$, the test statistic $\tilde{t}_{\gamma=0}(r_0)$ has the following distribution*

$$\tilde{t}_{\gamma=0}(r_0) \Rightarrow \mathcal{N}(0,1) \tag{2.49}$$

2.5.2.2 When r is Unknown but Identified

Again, when there is a threshold effect in the drift, we can estimate super-consistently the threshold parameter under $H_0 : \gamma = 0$; then, we can use $\hat{r}_n$ as if we know r_0. We can recover $\hat{e}_{\alpha,t}$ from model (2.10) and estimate $\mathbb{G}^\alpha$ as follows.

$$\hat{\mathbb{G}}^\alpha = \frac{1}{n}\sum_{j=1}^{n} \hat{e}_{\alpha,t}^2(I(z_{t-1} \leq \hat{r}_n)-\hat{\lambda})^2 + \frac{2}{n}\sum_{i=1}^{l} L(i,l) \sum_{t=i+1}^{n} \hat{e}_{\alpha,t}\hat{e}_{\alpha,t-i}(I(z_{t-1} \leq \hat{r}_n)-\hat{\lambda})(I(z_{t-1-i} \leq \hat{r}_n)-\hat{\lambda}) \tag{2.50}$$

, where $\hat{\lambda} = \frac{1}{n}\sum I(z_{t-1} \leq \hat{r}_n)$, which will also be a consistent estimator of $P(z_{t-1} \leq r_0) = \lambda_0$, then the modified statistic for testing $\gamma = 0$ is

$$\tilde{t}^\alpha_{\gamma=0}(\hat{r}_n) = \hat{\gamma}(\hat{r}_n)\sqrt{\frac{\hat{\lambda}(1-\hat{\lambda})}{\hat{\mathbb{G}}^\alpha((X(\hat{r}_n)'X(\hat{r}_n))^{-1})_{22}}} \tag{2.51}$$

Proposition 11. *When the threshold value is unknown but identified under the null of $\gamma = 0$, with assumptions A.1, A.2, B.3, A.4, A.5, and A.6., the test statistic $\tilde{t}^\alpha_{\gamma=0}(\hat{r})$ has the following distribution*

$$\tilde{t}^\alpha_{\gamma=0}(\hat{r}_n) \Rightarrow \mathcal{N}(0,1) \tag{2.52}$$

Assuming $\sigma_{\varepsilon,e} = \sigma_{\varepsilon,e,\lambda} = 0$ is relevant for the distribution of the KPSS test to be free of nuisance parameters, but it is not relevant for testing for threshold effects since independently if e_t and ε_t are serially correlated or not, we can construct $\tilde{t}_{\gamma=0}(\bar{r})$ or $\tilde{t}^\alpha_{\gamma=0}(\hat{r}_n)$ as in (2.47) and (2.51) such that their distribution under the null of $\gamma = 0$ is free of nuisance parameters and is the same as in Proposition 5 and Proposition 7. Gonzalo and Pitarakis (2006) found a similar situation, where the proposed test for threshold effects was robust under the problem of endogeneity.

2.6 Simulations

In this section, we illustrate the key features of the different tests presented under different scenarios, when e_t is independent of past realizations, when e_t is a linear process but x_t is strictly exogenous and, finally, when e_t is a linear process and x_t is endogenous.

Our data-generating process (DGP) is given by $y_t = \beta_2 x_t + \gamma I(z_{t-1} \leq r)x_t + e_t$ and $y_t = \alpha_1 I(z_{t-1} \leq r) + \alpha_2 I(z_{t-1} > r)\beta_2 x_t + \gamma I(z_{t-1} \leq r)x_t + e_t$, respectively, where $\Delta x_t = \varepsilon_t$. We take z_t as an AR(1) process $z_t = \rho_z z_{t-1} + \eta_t$ with $\eta_t = n.i.d(0,1)$. e_t also is constructed as an AR(1) process $e_t = \rho e_{t-1} + v_t$, where $v_t = n.i.d(0,1)$, and by changing the value of ρ we can control the dependence structure of the shocks in the long run equation. We also consider the cases where the threshold parameter r is known and the threshold value is estimated $\hat{r}_n$. All the experiments are based on 10000 replication and setting $\beta_2 = 1$, $\alpha_1 = 1$ and $\alpha_2 = 2$ throughout.

First, we evaluate the behavior of the KPSS test under different scenarios. In these simulations, we choose different values of the bandwidth parameters as a function of the sample size n, $l0 = 0$, $l4 = 4Integer\left[n/100\right]^{1/4}$ and $l12 = 12integer\left[n/100\right]^{1/4}$ in the estimation of the long run variance $\hat{\sigma}_e^2(l)$ and $\hat{\sigma}_{\alpha,e}^2(l)$.

Tables 2.11 to 2.19 show the size of the KPSS under assumption A.3, that is, $\{e_t\}$, is an i.i.d process for different choices of the bandwidth parameter l, under different levels of persistence of the threshold variable $\rho_z = \{0.5, 0.9\}$ with different values of $\gamma = \{0,1\}$. When $\gamma = 0$, it is the linear specification, and when $\gamma = 1$, it is the threshold specification. Also, we perform the test including state dependent drifts and without drift, whether the threshold parameter is known or unknown.

As we can see when the DGP is linear, $\gamma = 0$, and we perform the KPSS test including the regressors with a threshold effect, the size of the test is correct since the empirical size approaches the nominal size of 5%. Also, the size is correct for the different levels of persistence of the threshold variable ρ_z, whether the threshold parameter is known or estimated. Note that, independent of the choice of the bandwidth parameter, the estimation of the long run variance does not have any effect on the size of the test.

Tables 2.20 to 2.28 show the power of the KPSS test when $\{e_t\}$ is an i.i.d process, considering different values of $\sigma_u^2 = \{0.01, 0.1, 1\}$, and for different choices of $l0$, $l4$, and $l12$. As we have shown in Proposition 4, the choice of the bandwidth parameter l is relevant since choosing a large l will cost power, for example, as observed in Tables 2.20 to 2.22, when the threshold parameter is assumed to be known and the DGP does not have state-dependent constants, by fixing the sample size n and fixing σ_u^2, we can see a decrease of power as the bandwidth parameter increases. Choosing $l4$ and $l12$ and increasing σ_u^2 but fixing the sample size, the power approaches to a limit which is not necessarily one. The power increases as n increases, thereby reflecting the consistency of the test. We can observe similar results when the threshold parameter is estimated, see Tables 2.26 and 2.28.

Tables 2.29 to 2.35, show the size of the KPSS test, when e_t follows an AR(1) process with different values of $\rho = \{-0.8, -0.5, -0.2, 0.2, 0.5, 0.8\}$ and the different specifications of the DGP. As we can see, the KPSS test for testing threshold cointegration have the same problem pointed out by Kwiatkowsky, Phillips, Schmidt y Shin (1992), where under the null of $\sigma_u^2 = 0$, as $\rho \to 1$ e_t become a random walk, and the test will tend to over-reject the null of threshold cointegration, also when $\rho < 0$, the KPSS tends to be conservative. The over-rejection problem is severe for the case of $l = 0$, which is not a valid test even asymptotically. For $l4$, when $\rho = 0.5$, the test presents a moderate size distortion but a severe size distortion for $\rho = 0.8$. Finally, for $l12$, the test has the correct size for $\rho = 0.5$ but a slight over-rejection for $\rho \geq 0.8$.

In the last experiment, we use dynamic OLS estimation to control for the endogeneity between x_t with e_t. In this case, we create η_t and ε as a bivariate normal with the covariance matrix.

$$\begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\eta,\varepsilon} \\ \sigma_{\eta,\varepsilon} & \sigma_\eta^2 \end{pmatrix} \tag{2.53}$$

we set $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$, and we allow $\sigma_{\eta,\varepsilon} = \{0.5, 0.8\}$. As we can see, we have the same results as in the case where x_t is exogenous, in which we have over-rejection when $\rho \to 1$ and the test is conservative when $\rho < 0$, and as in the previous case, this problem aggravates when $l = 0$. When $\rho = 0$, x_t, and e_t are only correlated contemporaneously, and the dynamic ordinary least squares (DOLS) estimation helps to control the second-order bias, and the empirical size of the test approaches the nominal size of 5%. We only have a contemporaneous correlation between x_t and e_t as the second-order bias, and we can see in this case the size of the KPSS test is under control. See Tables 2.36 and 2.39.

In the next experiments, we show the behavior of the test to identify the threshold effect in the long run equation. Table 2.40 shows the size of the test when $\{e_t\}$ is an i.i.d process. In the case where the state-dependent constant is included, we can see that the empirical size is close to the nominal size of 5% for different persistence levels of the threshold variable $\rho_z = \{0.5, 0.9\}$, or if the threshold value is known or estimated. Table 2.41 shows the same result in the case where the state-dependent drift is not considered.

In Table 2.42, we show the power of the test when $\{e_t\}$ follows an i.i.d process considering different values of $\gamma = \{0.01, 0.05, 0.1, 0.5\}$. The DGP considered has a state-dependent drift, and we can see that the tests have correct power since both, as sample size increases and the value of γ deviates from zero, the power of the test approaches unity. Table 2.43 shows the same results when the DGP does not have state-dependent constants and the threshold parameter is assumed to be known.

In the last experiment for the test for threshold effect, we show the performance of the test when e_t follows an AR(1) process with autoregressive coefficients $\rho = \{-0.8, -0.5, -0.2, 0.2, 0.5, 0.8\}$ and different values of $\sigma_{\varepsilon,\eta} = \{0, 0.5, 0.8\}$. The bandwidth parameter for the estimation of $\mathbb{G}$ is chosen as $l0$, $l4$, and $l12$. In the case where $l = 0$, the test rejects too often when $\rho > 0$ and too seldom when $\rho < 0$ and this problem aggravates when $\rho_z = 0.9$. In the cases where the choices are $l4$ and $l12$, and $\rho_z \leq 0.5$ for any values of ρ, the empirical size of the test is correct. When $\rho_z = 0.9$ for $\rho \geq 0.8$, the test tends to over reject the null of no-threshold effect for small sample sizes, but as the sample size increases the empirical size of the test approaches the nominal size of 5%. The over-rejection of the null in small samples is due to the estimation of the long-run variance of $\mathbb{G}$, which is poorly approximated in small samples. See tables 2.44 to 2.50.

2.7 Empirical Applications

2.7.1 Term Structure of U.S Interest Rates

In this application, we analyze the existence of multiple equilibria between interest rates of instruments with different maturities. It is widely known that the interest rate series are $I(1)$, and interest rates with different maturities must be cointegrated, as proposed by Stock and Watson (1988). There is a vast amount of work that studies the cointegration relation between interest rates with different maturities, but all of them assume the existence of a single equilibrium relationship with a mixed conclusion, for example, the Johansen (1996) procedure is unable to find cointegration at the usual significance levels. In their work, Enders and Siklos (2001) extend the analysis by allowing a threshold structure in the short-term dynamics, however, assuming the existence of a unique equilibrium relation. They conclude that when the adjustment follows an autoregressive threshold structure (TAR) the series of U.S. federal funds rate and the U.S. 10-year rates on government bonds are not cointegrated, however, when the short-term dynamics follows a momentum-TAR, they find a cointegration relation.

Following the work of Enders and Siklos (2001), we use monthly observations of the federal fund rates and 10-year government bonds from January 1960 to March 2019. The data are daily averages and are available in the Federal Reserve Bank of St. Louis database. We use the annual growth rate of the U.S. production index as the threshold variable, which will determine the different periods of the business cycles, economic expansions, and recessions.

As discussed in Shin (1994) and Sikkonen (1991), the choice of the number of lags and leads, K, and the bandwidth parameter, l, for the estimation of the long run variance are critical, especially on l. In this application, we choose K using the Akaike information criterion (AIC) information criteria, and for the bandwidth parameter, we present the results of the tests using different values of l. For illustration purpose, we estimate the linear long-run equilibrium relationship between short- and long-term interest rates using the DOLS estimation. The optimal number of lags and leads included in the regression is $K = 3$

Table 2.1: Linear cointegration, estimation result.

α	β
-2.88	1.166
(0.137)	(0.020)

We test for linear cointegration using the KPSS test as in Shin (1994), and note that for different values of l we reject at 5% significance the null hypothesis of linear cointegration, since the KPSS statistic is higher than the critical value (CV) of 0.314 tabulated in Shin (1994) .

Table 2.2: KPSS test using linear regression model.

$l = 7$	$l = 8$	$l = 9$	$l = 10$	$1 = 11$	$l = 12$
0.585	0.533	0.492	0.4583	0.431	0.4076

Now we test for threshold cointegration, and in this case, we assume that there is a threshold effect in the drift, which is relevant for estimating the threshold parameter that is T-consistent. First, we test for cointegration including a non-linear component in the long-run equation, as in the linear case, the optimal choice of lags and leads in the DOLS estimation is $K = 3$, and the estimated long-run equation is given as:

Table 2.3: Cointegration with threshold effect, estimation result.

α_1	α_2	β_1	β_2	r
-2.573	-0.354	1.244	0.909	0.048
(0.151)	(0.023)	(0.288)	(0.041)	

We perform the KPSS test on the estimated residual for different values of l, and since we find cointegration at the 5% significance level, the KPSS statistic is lower than the critical value of 0.314.

Table 2.4: KPSS test using threshold regression model.

$l = 7$	$l = 8$	$l = 9$	$l = 10$	$1 = 11$	$l = 12$
0.286	0.262	0.242	0.227	0.214	0.203

As we have argued before, constructing the KPSS test using a threshold specification in the regression is unable to differentiate between linear and threshold cointegration. The second step is to test for the presence of a threshold effect in the long-run equilibrium equation by testing if $\gamma = (\beta_1 - \beta_2) = 0$. For different values of l, we reject the null of no-threshold effect since the statistic is higher than the 5% CV of a standard normal.

Table 2.5: Testing for threshold effect.

$l = 7$	$l = 8$	$l = 9$	$l = 10$	$1 = 11$	$l = 12$
3.621	3.547	3.499	3.468	3.452	3.448

Since we reject the null of $\gamma = 0$, we can conclude the presence of two equilibrium relations between both short- and long-term interest rates. When the annual growth of industrial production (IP) is above 4.8%, we are in an expansion period, and the cointegrating vector is $\begin{pmatrix} 1 & 0.909 \end{pmatrix}'$, implying that the long-term interest rate is higher than the short-term interest rate. When the industrial activity slows down, the annual growth of IP is under 4.8%, and the cointegrating vector is $\begin{pmatrix} 1 & 1.244 \end{pmatrix}$, which indicates that the short-term interest rate is higher than the long-term interest rate.

2.7.2 Empirical Application: US Stock Price and Dividend

In this application, we investigate via our threshold cointegration model the non-linear link between price and dividends using the Volatility Index (VIX) as a threshold variable.

The data analyzed are the monthly series of real Standard and Poor&s Composite Stock Price Index and the real dividend from 1960:1 to 2018:7. The threshold variable representing the different regimes is the VIX Index, which generates periods where the perceived volatility is high and periods with low volatility. More information about the data on stock prices and dividends can be found in Shiller (http://www.econ.yale.edu/ shiller/data.htm) and information on the VIX Index series can be obtained from FRED (https://fred.stlouisfed.org).

As an illustration porpoise, we estimate the linear cointegration equation and perform the KPSS test on the estimated residuals. The optimal choice of lags and leads using AIC is $K = 0$,

Table 2.6: Linear cointegration, estimation result.

c	β	K
121.89	47.25	0
(65.50)	(2.21)	

Testing for linear cointegration, we can see that for different values of the bandwidth parameter l, we reject the null of cointegration at 5% significance since the KPSS test is higher than the CV of 0.314 tabulated in Shin (1994).

Table 2.7: KPSS test using linear regression model.

$l = 10$	$l = 11$	$l = 12$	$l = 13$	$1 = 14$	$l = 15$
0.46	0.43	0.40	0.34	0.35	0.33

Now we estimate the threshold specification including the non-linear regression and test for cointegration using the estimated residuals. The optimal number of lags and leads indicated by AIC is $K = 11$

Table 2.8: Cointegration with threshold effect, estimation result.

α_1	α_2	β_1	β_2	r	K
-185.20	1357	52.14	8.22	19.57	11
(71.34)	(141.71)	(2.48)	(5.21)		

Checking for cointegration, we can see for each value if l fails to reject the null of cointegration at 5% significant level.

Table 2.9: KPSS test using threshold regression model.

$l = 10$	$l = 11$	$l = 12$	$l = 13$	$1 = 14$	$l = 15$
0.31	0.29	0.27	0.25	0.24	0.23

Once we find cointegration in the threshold regression estimation, as in the previous application, we proceed to test for the presence of threshold effects by testing if $\beta_1 = \beta_2$. We perform the test using different choices of the bandwidth parameter in the estimation of the long run variance of $\mathbb{G}$. As we can see in the Table 2.10 for each value of l, we reject the null of no-threshold effect at 5% significance level, thereby concluding the presence of multiple cointegration relations.

Table 2.10: Testing for threshold effect.

$l = 10$	$l = 11$	$l = 12$	$l = 13$	$1 = 14$	$l = 15$
2.79	2.63	2.61	2.54	2.47	2.44

Since we have concluded that there is a threshold effect, the real price and real dividend present two equilibrium relations driven by the perceived volatility in the market. When the volatility is high, the VIX index is above 19.57, and the implicit discount rate is $R_2 = \frac{1}{\beta_2} = 12.17\%$. When the volatility is low, the VIX index is below 19.57, and the implicit discount rate is $R_1 = \frac{1}{\beta_1} = 1.92\%$. This result is consistent with the economic theory because the return of a risky asset must be higher in periods when the volatility is higher than in periods with lower volatility.

2.8 Conclusion

Many economic relations between persistent variables are not linear, and this is captured by the concept of non-linear cointegration. Extensions of linear cointegration to a non-linear framework have always considered the existence of a single long-run equilibrium where the non-linearity comes from the adjustment mechanism towards it. In this study, we analyze non-linear cointegration with multiple long-run equilibria via threshold cointegration. We present a test to assess the presence of non-linear cointegration, and an inference procedure to detect threshold structures. Two empirical applications are shown, between U.S. stock prices and dividends and U.S. interest rates from instruments with different maturities, where cointegration with different equilibrium relations is not rejected whereas standard linear cointegration is rejected.

Appendix

2.A Proofs

Proof of Proposition 2

In order to show that $\{I(z_{t-d} \leq r)x_t\}$ is summable of order one we have to prove that:

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta_x}} L(n) \sum_{t=1}^{n} I(z_{t-d} \leq r)x_t = O_p(1), \tag{2.54}$$

From the asymptotic results from Canner and Hansen (2001), for $\delta_x = 1$ we have that;

$$S_n = \frac{1}{n^{\frac{3}{2}}\lambda\sigma_\varepsilon} \sum_{t=1}^{n} I(z_{t-d} \leq r)x_t \rightarrow \int_0^1 W_x(s)ds, \tag{2.55}$$

where $L(n) = \frac{1}{\lambda\sigma_\varepsilon}$, $\lambda = Pr(z_{t-d} \leq r)$ and $W_n(.)$ is the standard Brownian motion. For $\{I(z_{t-d} > r)x_t\}$, the proof is similar with $L(n) = \frac{1}{(1-\lambda)\sigma_\varepsilon}$.

Also, we have to show that:

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta_y}} L(n) \sum_{t=1}^{n} y_t = O_p(1), \tag{2.56}$$

by construction we know that:

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta_y}} L_1(n) \Big[\beta_1 \sum_{t=1}^{n} I(z_{t-d} \leq r)x_t + \beta_2 \sum_{t=1}^{n} I(z_{t-d} > r)x_t\Big] + \frac{L_1(n)}{L_2(n)} \frac{1}{n^{\frac{1}{2}+\delta_y}} L_2(n) \sum_{t=1}^{n} I(z_{t-d} \leq r)e_t \tag{2.57}$$

Note that from the work of Caner and Hansen (2001) we know that for $L_2(n) = \frac{1}{\sigma_e}$, it is easy to show that $\frac{1}{n^{\frac{1}{2}}\sigma_e} \sum_{t=1}^{n} I(z_{t-d} \leq r)e_t \rightarrow B_e(s,\lambda)$ then for $\delta_y = 1$ we have $\frac{1}{n^{\frac{1}{2}+\delta_y}\sigma_e} \sum_{t=1}^{n} I(z_{t-d} \leq r)e_t = o_p(1)$. Then for $L_1(n) = \frac{1}{[\beta_1\lambda+\beta_2(1-\lambda)]\sigma_x}$, from corollary 1 we have

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta_y}[\beta_1\lambda + \beta_2(1-\lambda)]\sigma_\varepsilon} \sum_{t=1}^{n} y_t \rightarrow \int_0^1 W_x(s)ds \tag{2.58}$$

and is easy to see that $\frac{L_1(n)}{L_2(n)} = O(1)$.

Proof of Proposition 3

The proof of proposition 3 we split into two parts. The first part we show the asymptotic distribution of the LS estimate of β_2 and γ under the null of $\sigma_u = 0$. The second part we show the convergence of $n^{-1/2}S_{[ns]}$ under the null of $\sigma_u = 0$. Note that under the null of cointegration $\sigma_u = 0$ then $v_t = m_0 + e_t$. Without loss of generality, set $m_0 = 0$, then $v_t = e_t$

To show the asymptotic distribution of the LS estimate of β_2 and γ, write equation (2.9) as follows

$$y_t = X_t'\Gamma + v_t \tag{2.59}$$

where $X_t = \begin{pmatrix} x_t \\ x_t I(z_{t-1} \le r) \end{pmatrix}$ and $\Gamma = \begin{pmatrix} \beta_2 \\ \gamma \end{pmatrix}$

Then the LS estimate of Γ is

$$\hat{\Gamma} = \Big(\sum_{t=1}^{n} X_t X_t'\Big)^{-1}\Big(\sum_{t=1}^{n} X_t y_t\Big) = \Gamma + \Big(\sum_{t=1}^{n} X_t X_t'\Big)^{-1}\Big(\sum_{t=1}^{n} X_t v_t\Big) \tag{2.60}$$

Under the null of cointegration:

$$\hat{\Gamma} = \Gamma + \Big(\sum_{t=1}^{n} X_t X_t'\Big)^{-1}\Big(\sum_{t=1}^{n} X_t e_t\Big) \tag{2.61}$$

we can rewrite equation (2.61) as follows

$$n(\hat{\Gamma} - \Gamma) = \underbrace{\Big(\frac{1}{n^2}\sum_{t=1}^{n} X_t X_t'\Big)^{-1}}_{A}\underbrace{\Big(\frac{1}{n}\sum_{t=1}^{n} X_t e_t\Big)}_{B} \tag{2.62}$$

Lets see the convergence of A.

$$\frac{1}{n^2}\sum_{t=1}^{n} X_t X_t' = \begin{pmatrix} \frac{1}{n^2}\sum_{t=1}^{n} x_t^2 & \frac{1}{n^2}\sum_{t=1}^{n} x_t^2 I(z_{t-1} \le r) \\ \frac{1}{n^2}\sum_{t=1}^{n} x_t^2 I(z_{t-1} \le r) & \frac{1}{n^2}\sum_{t=1}^{n} x_t^2 I(z_{t-1} \le r) \end{pmatrix} \tag{2.63}$$

Since x_t is a random walk, from Caner and Hansen (2001) we know that:

$$\begin{aligned} \frac{1}{n^2}\sum_{t=1}^{n} X_t X_t' &\Rightarrow \begin{pmatrix} \sigma_\varepsilon^2 \int_0^1 W_x^2(s)ds & \lambda\sigma_\varepsilon^2 \int_0^1 W_x^2(s)ds \\ \lambda\sigma_\varepsilon^2 \int_0^1 W_x^2(s)ds & \lambda\sigma_\varepsilon^2 \int_0^1 W_x^2(s)ds \end{pmatrix} \\ &\equiv \sigma_\varepsilon^2 \int_0^1 W_x^2(s)ds \begin{pmatrix} 1 & \lambda \\ \lambda & \lambda \end{pmatrix} \end{aligned} \tag{2.64}$$

Then by the continuous mapping theorem we have:

$$\Big(\frac{1}{n^2}\sum_{t=1}^{n} X_t X_t'\Big)^{-1} \Rightarrow \frac{1}{\sigma_\varepsilon^2 \lambda(1-\lambda)\int_0^1 W_x^2(s)ds} \begin{pmatrix} \lambda & -\lambda \\ -\lambda & 1 \end{pmatrix} \tag{2.65}$$

Lets see what converges the term B,

$$\left(\frac{1}{n}\sum_{t=1}^{n} X_t e_t\right) = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} x_t e_t \\ \frac{1}{n}\sum_{t=1}^{n} x_t e_t I(z_{t-1} \leq r) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} x_{t-1} e_t + \frac{1}{n}\sum_{t=1}^{n} e_t \varepsilon_t \\ \frac{1}{n}\sum_{t=1}^{n} x_{t-1} e_t I(z_{t-1} \leq r) + \frac{1}{n}\sum_{t=1}^{n} I(z_{t-1} \leq r) e_t \varepsilon_t \end{pmatrix} \tag{2.66}$$

Again from Caner and Hansen (2001) we have

$$\left(\frac{1}{n}\sum_{t=1}^{n} X_t e_t\right) \Rightarrow \begin{pmatrix} \sigma_e \sigma_\varepsilon \int_0^1 W_x(s) dW_e(s) + w_{12} \\ \sigma_e \sigma_\varepsilon \int_0^1 W_x(s) dW_e(s,\lambda) + \lambda w_{12} \end{pmatrix} \equiv \begin{pmatrix} \sigma_e \sigma_\varepsilon \int_0^1 W_x(s) dW_e(s) \\ \sigma_e \sigma_\varepsilon \int_0^1 W_x(s) dW_e(s,\lambda) \end{pmatrix} \tag{2.67}$$

From the assumption where ε_t and e_{2t} are independent, we can forget about the second order bias w_{12}.

Putting all together we have that:

$$n(\hat{\Gamma} - \Gamma) \Rightarrow \frac{\sigma_e}{\lambda(1-\lambda)\sigma_\varepsilon \int_0^1 W_x^2(s) ds} \begin{pmatrix} \lambda & -\lambda \\ -\lambda & 1 \end{pmatrix} \begin{pmatrix} \int_0^1 W_x(s) dW_e(s) \\ \int_0^1 W_x(s) dW_e(s,\lambda) \end{pmatrix} \tag{2.68}$$

The second part of the proof consist to show the convergence $n^{-1/2} S_{[ns]} = n^{-1/2} \sum_{j=1}^{[ns]} \hat{v}_j$. By adding and subtracting v_j and reordering

$$n^{-1/2} S_{[ns]} = n^{-1/2} \sum_{j=1}^{[ns]} (\hat{v}_j - v_j + v_j) = T^{-1/2} \sum_{j=1}^{[ns]} v_j + n^{-1/2} \sum_{j=1}^{[ns]} (\hat{v}_j - v_j) \tag{2.69}$$

Recall that under the null of cointegration $v_j = e_j$ and note that $\hat{v}_j = y_j - X_j' \hat{\Gamma}$ with $v_j = y_j - X_j' \Gamma$, then

$$(\hat{v}_j - v_j) = -X_j'(\hat{\Gamma} - \Gamma) \tag{2.70}$$

then rewritting

$$n^{-1/2} \sum_{j=1}^{[ns]} e_j - n^{-3/2} \sum_{j=1}^{[ns]} X_j' n(\hat{\Gamma} - \Gamma) \tag{2.71}$$

Then as $n \to \infty$, $n^{-1/2} S_{[ns]}$ converges to

$$\sigma_e W_e(s) - \frac{\sigma_e}{\lambda(1-\lambda)\int_0^1 W_x^2(s)ds} \begin{pmatrix} \int_0^s W_x(s)ds & \lambda \int_0^s W_x(s)ds \end{pmatrix} \begin{pmatrix} \lambda & -\lambda \\ -\lambda & 1 \end{pmatrix} \begin{pmatrix} \int_0^1 W_x(s)dW_e(s) \\ \int_0^1 W_x(s)dW_e(s,\lambda) \end{pmatrix} \tag{2.72}$$

Then (2.72) can be written as:

$$\sigma_e \Big(W_e(s) - \frac{\int_0^s W_x(s)ds \int_0^1 W_x(s)dW_e(s)}{\int_0^1 W_x^2(s)ds} \Big) \equiv \sigma_e Q(s) \tag{2.73}$$

From (2.73) we can conclude that:

$$n^{-2} \sum_{t=1}^{n} S_t^2 \Rightarrow \sigma_e^2 \int_0^1 Q^2(s)ds \tag{2.74}$$

The only task left is to show that: $\hat{\sigma}_e^2 \to_p \sigma_e^2$. Under the null of cointegration

$$\hat{\sigma}_e^2 = \frac{1}{n} \sum_{t=1}^{n} e_t^2 + o_p(1) \to_p \sigma_e^2. \tag{2.75}$$

Then we can conclude that

$$CI = n^{-2} \sum_{t=1}^{n} S_t^2 / \hat{\sigma}_e^2 \Rightarrow \int_0^1 Q^2(s)ds \tag{2.76}$$

When state dependent drift is included

$$y_t = \alpha_1 I(z_{t-1} \le r) + \alpha_2 I(z_{t-1} > r) + \beta_1 I(z_{t-1} \le r)x_t + \beta_2 I(z_{t-1} > r)x_t + e_t \tag{2.77}$$

Construct the following stack matrices

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad 1_z = \begin{pmatrix} I(z_0 \le r) & I(z_0 > r) \\ \vdots & \vdots \\ I(z_{n-1} \le r) & I(z_{n-1} > r) \end{pmatrix}, \quad X_z = \begin{pmatrix} x_1 I(z_0 \le r) & x_1 I(z_0 > r) \\ \vdots & \vdots \\ x_n I(z_{n-1} \le r) & x_n I(z_{n-1} > r) \end{pmatrix}$$

$$e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}, \quad C = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Define $M_z = I_n - 1_z(1_z'1_z)^{-1}1_z'$ where I_n is the identity matrix, then we can rewrite the model as

$$M_zY = M_zX\Gamma + M_ze \tag{2.78}$$

Then the LS estimate of Γ is the usual

$$(\hat{\Gamma} - \Gamma) = (X_z'M_zX)^{-1}(X_z'M_ze) \tag{2.79}$$

where

$$(X_z'M_qX_z) = \begin{pmatrix} \sum_{i=1}^n x_i^2 I(z_{i-1} \leq r) - \frac{\left(\sum_{i=1}^n x_i I(z_{i-1}\leq r)\right)^2}{\sum_{i=1}^n I(z_{i-1}\leq r)} & 0 \\ 0 & \sum_{i=1}^n x_i^2 I(z_{i-1} > r) - \frac{\left(\sum_{i=1}^n x_i I(z_{i-1}>r)\right)^2}{\sum_{i=1}^n I(z_{i-1}>r)} \end{pmatrix} \tag{2.80}$$

$$(X_z'M_ze) = \begin{pmatrix} \sum_{i=1}^n I(z_{i-1} \leq r)x_ie_i - \frac{\left(\sum_{i=1}^n x_i I(z_{i-1}\leq r)\right)\left(\sum_{i=1}^n e_i I(z_{i-1}\leq r)\right)}{\sum_{i=1}^n I(z_{i-1}\leq r)} \\ \sum_{i=1}^n I(z_{i-1} > r)x_ie_i - \frac{\left(\sum_{i=1}^n x_i I(z_{i-1}>r)\right)\left(\sum_{i=1}^n e_i I(z_{i-1}>r)\right)}{\sum_{i=1}^n I(z_{i-1}>r)} \end{pmatrix} \tag{2.81}$$

Using the results from Caner and Hansen (2001) we can see that;

$$n(\hat{\Gamma} - \Gamma) \Rightarrow \begin{pmatrix} \frac{\int_0^1 B_x^\alpha(s)dB_e(s,\lambda)}{\lambda\int_0^1(B_z^\alpha(s))^2ds} \\ \frac{\int_0^1 B_x^\alpha(s)dB_e(s) - \int_0^1 B_x^\alpha(s)dB_e(s,\lambda)}{(1-\lambda)\int_0^1(B_z^\alpha(s))^2ds} \end{pmatrix} \tag{2.82}$$

If we write each element of M_zY, M_zX_z and M_ze we can see that

$$M_zY = \begin{pmatrix} y_1 - \left(I(z_0 \le r)\frac{\sum_{i=1}^n y_i I(z_{i-1}\le r)}{\sum_{i=1}^n I(z_{i-1}\le r)} + I(z_0 > r)\frac{\sum_{i=1}^n y_i I(z_{i-1}>r)}{\sum_{i=1}^n I(z_{i-1}>r)}\right) \\ \vdots \\ y_n - \left(I(z_{n-1} \le r)\frac{\sum_{i=1}^n y_i I(z_{i-1}\le r)}{\sum_{i=1}^n I(z_{i-1}\le r)} + I(z_{n-1} > r)\frac{\sum_{i=1}^n y_i I(z_{i-1}>r)}{\sum_{i=1}^n I(z_{i-1}>r)}\right) \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{pmatrix}$$

$$M_zX_z = \begin{pmatrix} I(z_0 \le r)\left(x_1 - \frac{\sum_{i=1}^n x_i I(z_{i-1}\le r)}{\sum_{i=1}^n I(z_{i-1}\le r)}\right) & I(z_0 > r)\left(x_1 - \frac{\sum_{i=1}^n x_i I(z_{i-1}>r)}{\sum_{i=1}^n I(z_{i-1}>r)}\right) \\ \vdots & \vdots \\ I(z_{n-1} \le r)\left(x_n - \frac{\sum_{i=1}^n x_i I(z_{i-1}\le r)}{\sum_{i=1}^n I(z_{i-1}\le r)}\right) & I(z_{n-1} > r)\left(x_n - \frac{\sum_{i=1}^n x_i I(z_{i-1}>r)}{\sum_{i=1}^n I(z_{i-1}>r)}\right) \end{pmatrix} = \begin{pmatrix} \bar{x}_1^- & \bar{x}_1^+ \\ \vdots & \vdots \\ \bar{x}_n^- & \bar{x}_n^+ \end{pmatrix}$$

$$M_ze = \begin{pmatrix} e_1 - \left(I(z_0 \le r)\frac{\sum_{i=1}^n e_i I(z_{i-1}\le r)}{\sum_{i=1}^n I(z_{i-1}\le r)} + I(z_0 > r)\frac{\sum_{i=1}^n e_i I(z_{i-1}>r)}{\sum_{i=1}^n I(z_{i-1}>r)}\right) \\ \vdots \\ e_n - \left(I(z_{n-1} \le r)\frac{\sum_{i=1}^n e_i I(z_{i-1}\le r)}{\sum_{i=1}^n I(z_{i-1}\le r)} + I(z_{n-1} > r)\frac{\sum_{i=1}^n e_i I(z_{i-1}>r)}{\sum_{i=1}^n I(z_{i-1}>r)}\right) \end{pmatrix} = \begin{pmatrix} \bar{e}_1 \\ \vdots \\ \bar{e}_n \end{pmatrix}$$

Such that for each t, it can be written as $\bar{y}_t = \begin{pmatrix} \bar{x}_t^- & \bar{x}_t^+ \end{pmatrix} \Gamma + \bar{e}_t$, then the constructed partial sum $\frac{1}{\sqrt{n}} S_{[ns]}$:

$$\frac{1}{\sqrt{n}} S_{[ns]} = \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \bar{e}_t}_{\text{A}} - \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \begin{pmatrix} \bar{x}_i^- & \bar{x}_i^+ \end{pmatrix} (\hat{\Gamma} - \Gamma)}_{\text{B}} \tag{2.83}$$

In A, from the structure of $\bar{e}_t$ we have that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \bar{e}_t = \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} e_t - \frac{[ns]}{n} \frac{1}{[ns]} \sum_{t=1}^{[ns]} I(z_{t-1} \le r) \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i I(z_{i-1} \le r)}{\frac{1}{n} \sum_{i=1}^n I(z_{i-1} \le r)} \\ - \frac{[ns]}{n} \frac{1}{[ns]} \sum_{t=1}^{[ns]} I(z_{t-1} > r) \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i I(z_{i-1} \le r)}{\frac{1}{n} \sum_{i=1}^n I(z_{i-1} > r)} \Rightarrow B_e(s) - sB_e(1) \end{aligned} \tag{2.84}$$

Now from B we can see that $\begin{pmatrix} \frac{1}{n^{3/2}} \sum_{t=1}^{[ns]} \bar{x}_i^- & \frac{1}{n^{3/2}} \sum_{t=1}^{[ns]} \bar{x}_i^+ \end{pmatrix} n(\hat{\Gamma} - \Gamma)$ such that

$$\frac{1}{n^{3/2}}\sum_{t=1}^{[ns]}\bar{x}_i^- = \frac{1}{n^{3/2}}\sum_{t=1}^{[ns]} I(z_{t-1}\leq r)x_t - \frac{[ns]}{n}\frac{1}{[ns]}\sum_{t=1}^{[ns]} I(z_{t-1}\leq r)\frac{\frac{1}{n^{3/2}}\sum_{t=1}^{n} I(z_{t-1}\leq r)x_t}{\frac{1}{n}\sum_{t=1}^{n} I(z_{t-1}\leq r)} \Rightarrow \lambda\Big(\int_0^s B_x^\alpha(s)ds\Big) \tag{2.85}$$

Also is easy to see that:

$$\frac{1}{n^{3/2}}\sum_{t=1}^{[ns]}\bar{x}_i^+ \Rightarrow (1-\lambda)\Big(\int_0^s B_x^\alpha(s)ds\Big) \tag{2.86}$$

we can see that:

$$\begin{pmatrix}\frac{1}{n^{3/2}}\sum_{t=1}^{[ns]}\bar{x}_i^- & \frac{1}{n^{3/2}}\sum_{t=1}^{[ns]}\bar{x}_i^+\end{pmatrix} n(\hat{\Gamma}-\Gamma) \Rightarrow \int_0^s B_x^\alpha(s)ds\frac{\int_0^1 B_x^\alpha(s)dB_e(s)}{\int_0^1 (B_x^\alpha(s))^2 ds} \tag{2.87}$$

Then putting all the pieces together we have that;

$$\frac{1}{\sqrt{n}}S_{[ns]} \Rightarrow B_e(s)-sB_e(1)-\int_0^s B_x^\alpha(s)ds\frac{\int_0^1 B_x^\alpha(s)dB_e(s)}{\int_0^1 (B_x^\alpha(s))^2 ds} = \sigma_e\left(V_e(s)-\int_0^s W_x^\alpha(s)ds\frac{\int_0^1 W_x^\alpha(s)dW_e(s)}{\int_0^1 (W_x^\alpha(s))^2 ds}\right) = Q_\alpha(s) \tag{2.88}$$

The first equality came from the strong exogeneity assumption between x_t and e_t. The using continous mapping theorem we have:

$$\frac{1}{n^2}\sum_{i=1}^{n} S_i^2 \Rightarrow \sigma_e^2\int_0^1 Q_\alpha^2(s)ds \tag{2.89}$$

Then:

$$CI_\alpha \Rightarrow \int_0^1 Q_\alpha^2(s)ds \tag{2.90}$$

Proof of Proposition 4

Under the alternative $\sigma_u^2 > 0$ the process $m_t = m_{t-1} + u_t$ is a Random Walk, the $e_t = m_t + v_{1t}$ also will be an Random Walk. For simplicity of exposition we consider the case without drift, a similar approach can be used when includes state dependent drift. Define

$$y_t = \begin{pmatrix} x_t & x_t I(z_{t-1}\leq r)\end{pmatrix}\begin{pmatrix}\beta_2\\ \gamma\end{pmatrix} + e_t \tag{2.91}$$

Call $\tilde{X}_t = \begin{pmatrix} x_t \\ x_t I(z_{t-1} \leq r) \end{pmatrix}$ and $\Gamma = \begin{pmatrix} \beta_2 \\ \gamma \end{pmatrix}$. Then is easy to see that

$$\begin{pmatrix} \hat{\beta}_2 - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} = \Big(\frac{1}{n^2}\sum_t \tilde{X}_t \tilde{X}_t'\Big)^{-1}\Big(\frac{1}{n^2}\sum_t \tilde{X}_t v_{1t}\Big) \tag{2.92}$$

As in (65), $\left(\frac{1}{n^2}\sum_t \tilde{X}_t \tilde{X}_t'\right)^{-1} \Rightarrow \frac{1}{\lambda(1-\lambda)\int_0^1 B_x(s)^2} \begin{pmatrix} \lambda & -\lambda \\ -\lambda & 1 \end{pmatrix}$. Let see

$$\Big(\frac{1}{n^2}\sum_t \tilde{X}_t v_{1t}\Big) = \begin{pmatrix} \frac{1}{n^2}\sum_t x_t(m_t + e_t) \\ \frac{1}{n^2}\sum_t x_t I(z_{t-1} \leq r)(m_t + e_t) \end{pmatrix} = \begin{pmatrix} \frac{1}{n^2}\sum_t x_t m_t + o_p(1) \\ \frac{1}{n^2}\sum_t x_t I(z_{t-1} \leq r) m_t + o_p(1) \end{pmatrix} \tag{2.93}$$

From the results in Caner and Hansen (2001) then

$$\Rightarrow \begin{pmatrix} \int_0^1 B_x B_m \\ \lambda \int_0^1 B_x B_m \end{pmatrix} \tag{2.94}$$

Then we can see that:

$$\begin{pmatrix} \hat{\beta}_2 - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} \Rightarrow \begin{pmatrix} \frac{\int_0^1 B_x B_m}{\int_0^1 B_x^2} \\ 0 \end{pmatrix} \tag{2.95}$$

Then the partial sum $\frac{1}{n^{3/2}} S_{[ns]}$ is

$$\begin{aligned} \frac{1}{n^{3/2}} S_{[ns]} =& \frac{1}{n^{3/2}}\sum_{i=1}^{[ns]} e_i - \frac{1}{n^{3/2}}\sum_{i=1}^{[ns]} \tilde{X}_t' \begin{pmatrix} \hat{\beta}_2 - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} = \frac{1}{n^{3/2}}\sum_{i=1}^{[ns]} m_t + o_p(1) - \frac{1}{n^{3/2}}\sum_{i=1}^{[ns]} \tilde{X}_t' \begin{pmatrix} \hat{\beta}_2 - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} \\ &\Rightarrow \int_0^s B_m + \int_0^s B_m \Big(\int_0^1 B_z\Big)^{-1}\Big(\int_0^1 B_x B_m\Big) = \sigma_u \int_0^s Q_p \end{aligned} \tag{2.96}$$

Then we can see that:

$$\frac{1}{n^4}\sum_{t=1}^{n} S_{[ns]}^2 \Rightarrow \sigma_u^2 \int_0^1 \Big(\int_0^s Q_p\Big)^2 \tag{2.97}$$

From Kwiatkowski, Phillips, Schmidt and Shin (1992), for the Barlett window, we obtain the result that $(nl)^{-1}s^2(l) \Rightarrow \sigma_u^2 \int_0^1 Q_p^2$. Therefore combining the results above we have:

$$(l/n)CI = \frac{1}{n^4}\sum_{t=1}^{n} S_{[ns]}^2/(nl)^{-1}s^2(l) \Rightarrow \int_0^1 \Big(\int_0^s Q_p\Big)^2 \Big/ \int_0^1 Q_p^2 \tag{2.98}$$

Proof of Proposition 5

From equation (2.68), Note that

$$n(\hat{\gamma}-\gamma) \Rightarrow \frac{\int_0^1 W_x(s)dW_e(s,\lambda) - \lambda\int_0^1 W_x(s)dW_e(s)}{\lambda(1-\lambda)\int_0^1 W_x(s)^2 ds} \equiv \frac{\int_0^1 W_x(s)dV_e(s,\lambda)}{\lambda(1-\lambda)\int_0^1 W_x(s)^2 ds} \tag{2.99}$$

where that $V_e(s,\lambda) = W_e(s,\lambda) - \lambda W_e(s,1)$ is the Kiefer-Muller process. From the continuous mapping theory, then the distribution of the t-statistic is:

$$t_{\gamma=0}(\lambda) = \frac{\int_0^1 W_x(s)dV_e(s,\lambda)}{\sqrt{\lambda(1-\lambda)\int_0^1 W_x(s)^2 ds}} \tag{2.100}$$

Since $W_e(s)$ and $V_e(s,\lambda)$ are independent, it can be proven for a fixed λ that:

$$\frac{\int_0^1 W_x(s)dV_e(s,\lambda)}{\sqrt{\int_0^1 W_x(s)^2 ds}} \equiv \mathcal{N}(0,\sigma_\lambda^2) \tag{2.101}$$

where $\sigma_\lambda = \lambda(1-\lambda)$.

Proof of Proposition 6

(i) As $T \to \infty$, $\hat{r}_n \to_p r_0$.

Following the work of Hansen (2000) and Chen (2015), to prove consistency of $\hat{r}_n$, we have to show that $Pr(|\hat{r}_n - r_0| > \varepsilon) \to 0$ for any $\varepsilon > 0$. Define $B(\varepsilon) = \{r : |r - r_0| > \varepsilon\}$ and $\bar{B}(\varepsilon) = \{[r_L, r_H] \backslash B(\varepsilon)\}$. From the definition of $\hat{r}_n$ we can see:

$$\begin{aligned} Pr(|\hat{r}_n - r_0| > \varepsilon) &= Pr\Big(\inf_{r\in B(\varepsilon)} SSR_n(r) < \inf_{r\in\bar{B}(\varepsilon)} SSR_n(r)\Big) \\ &\leq Pr\Big(\inf_{r\in B(\varepsilon)} SSR_n(r) < SSR_n(r_0)\Big) \\ &= Pr\Big(\inf_{r\in B(\varepsilon)} n^{-1}(SSR_n(r) - SSR_n(r_0)) < 0\Big) \end{aligned} \tag{2.102}$$

where $SSR_n(r) = \sum_{i=1}^n \tilde{e}^2$ and $SSR_n(r_0) = \sum_{i=1}^n e^2$. Thus, proving $Pr(|\hat{r}_n - r_0| > \varepsilon) \to 0$, is equivalent to prove that $Pr\Big(\inf_{r\in B(\varepsilon)} n^{-1}(SSR_n(r) - SSR_n(r_0)) > 0\Big) \to 1$

First we define the stacking vectors, without loss of generality assume that $h = 1$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad Q = \begin{pmatrix} I(z_0 \leq r) \\ I(z_1 \leq r) \\ \vdots \\ I(z_{n-1} \leq r) \end{pmatrix}$$
$$e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \quad Q_0 = \begin{pmatrix} I(z_0 \leq r_0) \\ I(z_1 \leq r_0) \\ \vdots \\ I(z_{n-1} \leq r_0) \end{pmatrix} \quad I_z = \begin{pmatrix} T & Q \end{pmatrix} \quad \Gamma = \begin{pmatrix} \alpha_2 \\ \gamma \end{pmatrix} \tag{2.103}$$

Define the following regression for any $r \in [r_L, r_H]$

$$Y = I_z\Gamma + X\beta + \tilde{e} \tag{2.104}$$

To prove the result we use the Frisch–Waugh–Lovell Theorem, by projecting Y on X, and I_z on X, to get rid of X. Define the annihilator matrix $M_x = (I_n - X(X'X)^{-1}X')$ such that $M_xX = 0$, then equation (2.104) can we rewrite as:

$$M_xY = M_xI_z\Gamma + M_x\tilde{e} \tag{2.105}$$

Then the LS estimate of Γ is:

$$\hat{\Gamma} = (I_z'M_xI_z)^{-1}(I_z'M_x'M_xY) \tag{2.106}$$

The estimated residuals can be written as:

$$M_x\tilde{e} = M_xY - M_xI_z(I_z'M_xI_z)^{-1}I_z'M_x'M_xY \tag{2.107}$$

Equivalently it can be written as:

$$M_x\tilde{e} = (I_n - M_xI_z(I_z'M_xI_z)^{-1}I_z'M_x')M_xY \tag{2.108}$$

Call $M_q = (I_n - M_xI_z(I_z'M_xI_z)^{-1}I_z'M_x')$, then $SSR(r) = YM_xM_qM_x'Y'$ Note that equation (2.105) can be rewritten as:

$$M_xY = M_xT\alpha_2 + M_xQ_0\delta + M_xe \tag{2.109}$$

We can write the $SSR(r)$ as follows:

$$\begin{aligned} SSR(r) &= \delta^2 Q_0' M_x M_q M_x Q_0 + 2\delta Q_0 M_x M_q M_x e \\ &+ e' M_x M_q M_x e \end{aligned} \tag{2.110}$$

The equality came from the fact that $T'M_xM_q = 0$ since $T'M_x$ is a linear combination of $\begin{pmatrix}(T-Q)'M_x & Q'M_x\end{pmatrix}$.

Note that $SSR(r_0) = e'M_xM_q^0M_xe'$, then:

$$\begin{aligned} \frac{1}{n}(SSR(r) - SSR(r_0)) &= \frac{1}{n}\delta^2 Q_0' M_x M_q M_x Q_0 + \underbrace{2\frac{1}{n}\delta Q_0 M_x M_q M_x}_{A} \\ &+ \frac{1}{n}\Big(\underbrace{e'M_xM_qM_xe - e'M_xM_q^0M_xe}_{B}\Big) \end{aligned} \tag{2.111}$$

We can see that the term B can be written as:

$$\underbrace{e'M_xI_z^0(I_z'^0M_xI_z^0)^{-1}I_z'^0M_xe}_{B.1} - \underbrace{e'M_xI_z(I_z'M_xI_z)^{-1}I_z'M_xe}_{B.2} \tag{2.112}$$

Lets focus on the term (B.2)

$$\left(\frac{1}{n}e'M_xI_z\right)\left(\frac{1}{n}I_z'M_xI_z\right)^{-1}\left(\frac{1}{n}I_z'M_xe\right) \tag{2.113}$$

Note that the first term:

$$\frac{1}{n}e'M_xI_z = \frac{1}{n}e'I_z - \left(\frac{1}{n}e'X\right)\left(\frac{1}{n^2}X'X\right)^{-1}\left(\frac{1}{n^{3/2}}X'I_z\right)\frac{1}{n^{1/2}} \tag{2.114}$$

Then we can see:

$$\begin{aligned} &\frac{1}{n}e'I_z \to \begin{pmatrix}E[e_i] & E[e_iI(q_{i-1} \leq r)]\end{pmatrix} = 0 \\ &\frac{1}{n}e'X \to \int_0^1 B_x dB_e \\ &\frac{1}{n^2}X'X \to \int_0^1 B_x^2 \\ &\frac{1}{n^{3/2}}X'I_z \to \begin{pmatrix}\int_0^1 B_x & \lambda\int_0^1 B_x\end{pmatrix} \end{aligned} \tag{2.115}$$

where $B(.)$ is the Brownian motion $\lambda = Pr(q_i \leq r)$. Then $\frac{1}{n}e'M_xI_z = o_p(1)$

The second step is to show the convergence of $\left(\frac{1}{n}I_z'M_xI_z\right)^{-1}$.

$$\begin{aligned}\frac{1}{n}I_z'M_xI_z &= \frac{I_z'I_z}{n} - \Big(\frac{I_z'X}{n^{3/2}}\Big)\Big(\frac{X'X}{n^2}\Big)^{-1}\Big(\frac{X'I_z}{n^{3/2}}\Big) \\ &\Rightarrow \begin{pmatrix}1 & \lambda\\ \lambda & \lambda\end{pmatrix} - \frac{(\int_0^1 B_x)^2}{\int_0^1 B_x^2}\begin{pmatrix}1 & \lambda\\ \lambda & \lambda^2\end{pmatrix}\end{aligned} \tag{2.116}$$

Then we can see that:

$$(\frac{1}{n}I_z'M_xI_z)^{-1} = \frac{1}{\lambda(1-\lambda)}\begin{pmatrix}\frac{\lambda\int_0^1 B_x^2 - \lambda^2(\int_0^1 B_x)^2}{\int_0^1 B_x^2 - (\int_0^1 B_x)^2} & -\lambda\\ -\lambda & 1\end{pmatrix} \tag{2.117}$$

Then we can see that the term B.2 converges to zero. The same happens with B.1.

Now lets focus on the term A. Note that:

$$M_xM_qM_x = M_x - S_q \tag{2.118}$$

Where $S_q = M_xI_z(I_z'M_xI_z)^{-1}I_z'M_x$, then:

$$2\frac{1}{n}\delta Q_0'M_xM_qM_xe = 2\frac{1}{n}\delta Q_0'(M_x - S_q)e = \underbrace{2\frac{1}{n}\delta Q_0'M_xe}_{A.1} - \underbrace{2\frac{1}{n}\delta Q_0'S_qe}_{A.2} \tag{2.119}$$

Lets focus on A.1

$$A.1 = 2\delta\Big(\frac{1}{n}Q_0'e\Big) - 2\delta\frac{1}{n^{1/2}}\Big(\frac{1}{n^{3/2}}Q_0'X\Big)\Big(\frac{1}{n^2}XX\Big)^{-1}\Big(\frac{1}{n}X'e\Big) \to 0 \tag{2.120}$$

Now focus on A.2

$$A.2 = 2\delta\Big(\frac{Q_0'M_xI_z}{n}\Big)\Big(\frac{I_z'M_xI_z}{n}\Big)^{-1}\Big(\frac{I_z'M_xe}{n}\Big) \to 0 \tag{2.121}$$

Then we can conclude that:

$$\frac{1}{n}(SSR(r) - SSR(r_0)) = \frac{1}{n}\delta^2Q_0'M_xM_qM_xQ_0 + o_p(1) \tag{2.122}$$

The last step is to show the convergence of $\frac{1}{n}\delta^2Q_0'M_xM_qM_xQ_0$. From (18) we can see:

$$\frac{1}{n}\delta^2Q_0'M_xM_qM_xQ_0 = \delta^2\underbrace{\frac{1}{n}Q_0'M_xQ_0}_{D.1} - \delta^2\underbrace{\frac{1}{n}Q_0'S_qQ_0}_{D.2} \tag{2.123}$$

To analyze properly equation (2.123), start by considering all $r \in [r_0, r_h]$. Lets focus on D.1,

$$\frac{1}{n}\delta^2 Q_0' M_x Q_0 = \frac{Q_0' Q_0}{n} - \Big(\frac{Q_0' X}{n^{3/2}}\Big)\Big(\frac{X'X}{n^2}\Big)^{-1}\Big(\frac{X' Q_0}{n^{3/2}}\Big) \\ \Rightarrow \lambda_0\Big(1 - \lambda_0 \frac{(\int_0^1 B_x)^2}{\int_0^1 B_x^2}\Big) \tag{2.124}$$

where $\lambda_0 = Pr(q_i \leq r_0)$. From term D.2, we can see:

$$\frac{1}{n} Q_0' S_q Q_0 = \Big(\frac{Q_0' M_x I_z}{n}\Big)\Big(\frac{I_z M_x I_z}{n}\Big)^{-1}\Big(\frac{I_z M_x Q_0}{n}\Big) \tag{2.125}$$

After some calculations

$$\frac{1}{n} Q_0' S_q Q_0 \Rightarrow \Big(\frac{\lambda_0^2}{\lambda} - \lambda_0 \frac{(\int_0^1 B_x)^2}{\int_0^1 B_x^2}\Big) \tag{2.126}$$

Then we can see that for $r \in [r_0, r_h]$:

$$\frac{1}{n}\delta^2 Q_0' M_x M_q M_x Q_0 \Rightarrow \delta^2 \lambda_0 (1 - \lambda_0 \lambda^{-1}) \tag{2.127}$$

Then $\delta^2 \lambda_0 > 0$, and for any $r \in (r_0, r_h]$ the term $\delta^2\lambda_0(1 - \lambda_0\lambda^{-1}) > 0$ and the minimum is attained at $r = r_0$, where $\delta^2\lambda_0(1 - \lambda_0\lambda^{-1}) = 0$. For the case where $r \in [r_l, r_0]$, doing similar calculation as in (125), (126) and (127) we can show that:

$$\frac{1}{n}\delta^2 Q_0' M_x M_q M_x Q_0 \Rightarrow \delta^2 \frac{(1-\lambda_0)}{(1-\lambda)}(\lambda_0 - \lambda) \tag{2.128}$$

Note again that for any $r \in [r_l, r_0)$ the term $\delta^2 \frac{(1-\lambda_0)}{(1-\lambda)}(\lambda_0 - \lambda) > 0$ and the minimum is attained at $r = r_0$, with $\delta^2 \frac{(1-\lambda_0)}{(1-\lambda)}(\lambda_0 - \lambda) = 0$

We have shown that:

$$\frac{1}{n}(SSR(r) - SSR(r_0)) \rightarrow_p \delta^2 \Big(\lambda_0(1 - \lambda_0\lambda^{-1}) I(r \geq r_0) \\ + \frac{(1-\lambda_0)}{(1-\lambda)}(\lambda_0 - \lambda) I(r \leq r_0)\Big)$$

Which is strictly positive when $r \in B(\varepsilon)$, thus

$$Pr\Big(\underset{r \in B(\varepsilon)}{inf}\, n^{-1}(SSR_n(r) - SSR_n(r_0)) > 0\Big) \rightarrow 1 \tag{2.129}$$

Showing the consistency in the estimate of $\hat{r}_n$.

(ii) As $n \to \infty$, we have $n|\hat{r}_n - r_0| = O_p(1)$

Let $a_n = n$, To prove that $\hat{r}_n$ converges to r_0 with rate a_n, we need to prove that $a_n|\hat{r}_n - r_0| = O_p(1)$ or show that $\exists\, \bar{v} > 0$ s.t $lim_{n\to\infty} Pr(|\hat{r}_n - r_0| \leq \bar{v}/a_n) = 1$. For any $B > 0$ define $V_B = \{r : |r - r_0| < B\}$, when n is large enough we have $\frac{\bar{v}}{a_n} < B$. From (i) we showed that $\hat{r}_n \to r_0$ which implies that $P(\hat{r}_n \in V_B) \to_p 1$, so we need only to examine the behaviour of r in V_B. Define $V_B(\bar{v}) = \{r : \frac{\bar{v}}{a_n} < |r - r_0| < B\}$, note that $V_B(\bar{v}) \subseteq V_B$. To prove $lim_{n\to\infty} Pr(|\hat{r}_n - r_0| \leq \bar{v}/a_n) = 1$ we have to show that $Pr(\hat{r}_n \in V_B(\bar{v})) = 0$. Let $\hat{\alpha}$ and $\hat{\delta}$ be $\hat{\alpha}(\hat{r}_n)$ and $\hat{\delta}(\hat{r}_n)$ and define $SSR^*_n(r) = \sum_{t=1}^n \left(y_t - \hat{\alpha} - \hat{\delta} I(z_{t-1} \leq r) - \hat{\beta} x_t\right)^2$ and $SSR^*_n(r_0) = \sum_{t=1}^n \left(y_t - \hat{\alpha} - \hat{\delta} I(z_{t-1} \leq r_0) - \hat{\beta} x_t\right)^2$. By definition $SSR^*_n(\hat{r}_n) \leq SSR^*_n(r_0)$, hence is sufficient to show that $\forall r \in V_B(\bar{v})$, $SSR^*_n(r) > SSR^*_n(r_0)$ with probability 1. As in (i) we can write:

$$M_x Y = M_x I_z \Gamma + M_x \tilde{e} \tag{2.130}$$

Such that $M_x I_z = \begin{pmatrix} M_x T & M_x Q \end{pmatrix}$ and

$$M_x Y = \begin{pmatrix} y_1 - x_1\hat{\beta} \\ y_2 - x_2\hat{\beta} \\ \vdots \\ y_n - x_n\hat{\beta} \end{pmatrix},\ M_x T = \begin{pmatrix} 1 - x_1(\sum_{i=1}^n x_i^2)^{-1}(\sum_{i=1}^n x_i) \\ 1 - x_2(\sum_{i=1}^n x_i^2)^{-1}(\sum_{i=1}^n x_i) \\ \vdots \\ 1 - x_n(\sum_{i=1}^n x_i^2)^{-1}(\sum_{i=1}^n x_i) \end{pmatrix} \tag{2.131}$$

$$M_x Q = \begin{pmatrix} I(z_0 \leq r) - x_1(\sum_{i=1}^n x_i^2)^{-1}(\sum_{i=1}^n x_i I(z_{i-1} \leq r)) \\ I(z_1 \leq r) - x_2(\sum_{i=1}^n x_i^2)^{-1}(\sum_{i=1}^n x_i I(z_{i-1} \leq r)) \\ \vdots \\ I(z_{n-1} \leq r) - x_n(\sum_{i=1}^n x_i^2)^{-1}(\sum_{i=1}^n x_i I(z_{i-1} \leq r)) \end{pmatrix} \tag{2.132}$$

Call $y_{x_t} = (y_t - x_t\hat{\beta})$, $1_{x_t} = \left(1 - x_t(\sum_{i=1}^n x_i^2)^{-1}(\sum_{i=1}^n x_i)\right)$,
$1_{z_{t-1},x_t}(r) = \left(I(z_{t-1} \leq r) - x_t(\sum_{i=1}^n x_i^2)^{-1}(\sum_{i=1}^n x_i I(z_{i-1} \leq r))\right)$ and
$e_{x_t} = \left(e_t - x_t(\sum_{i=1}^n x_i^2)^{-1})(\sum_{i=1}^n x_i e_i)\right)$, then we can rewrite:

$$\begin{aligned} SSR^*_n(r) &= \sum_{t=1}^n \left(y_{x_t} - 1_{x_t}\hat{\alpha} - 1_{q_{t-1},x_t}(r)\hat{\delta}\right)^2 \\ SSR^*_n(r_0) &= \sum_{t=1}^n \left(y_{x_t} - 1_{x_t}\hat{\alpha} - 1_{q_{t-1},x_t}(r_0)\hat{\delta}\right)^2 \end{aligned} \tag{2.133}$$

Then:

$$\begin{aligned} SSR^*_n(r) - SSR^*_n(r_0) &= 2(\delta-\hat{\delta})\sum_{t=1}^{n}(1_{z_{t-1},x_t}(r) - 1_{z_{t-1},x_t}(r_0))e_{x_t} - 2\delta\sum_{t=1}^{n}(1_{z_{t-1},x_t}(r) - 1_{z_{t-1},x_t}(r_0))e_{x_t} \\ &+ 2(\hat{\alpha}-\alpha)\hat{\delta}\sum_{t=1}^{n}(1_{z_{t-1},x_t}(r) - 1_{z_{t-1},x_t}(r_0))1_{x_t} + \delta^2\sum_{t=1}^{n}(1_{z_{t-1},x_t}(r) - 1_{z_{t-1},x_t}(r_0))^2 \\ &+ 2(\hat{\delta}-\delta)\hat{\delta}\sum_{t=1}^{n}(1_{z_{t-1},x_t}(r) - 1_{z_{t-1},x_t}(r_0))1_{q_{t-1},x_t}(r) - (\hat{\delta}-\delta)^2\sum_{t=1}^{n}(1_{z_{t-1},x_t}(r) - 1_{z_{t-1},x_t}(r_0))^2 \\ &= R_1 + R_2 + R_3 + R_4 + R_5 + R_6 \end{aligned} \tag{2.134}$$

We have to show that $\frac{R_1+R_2+R_3+R_4+R_5+R_6}{a_n(r-r_0)} > 0$. Consider the case where $r \in (r_0, r_H]$, the other case $r \in [r_l, r_0)$ can be shown by symmetry.

Step 1: $\frac{R_1}{n(r-r_0)} = \frac{2(\delta-\hat{\delta})}{(r-r_0)}\Big[\underbrace{\frac{1}{n}\sum_{t=1}^{n}1_{q_{t-1},x_t}(r)e_{x_t}}_{C.1} - \underbrace{\frac{1}{n}\sum_{t=1}^{n}1_{q_{t-1},x_t}(r_0)e_{x_t}}_{C.2}\Big]$

we can show that:

$$\frac{1}{n}\sum_{t=1}^{n}1_{q_{t-1},x_t}(r)e_{x_t} = \frac{1}{n}\sum_{t=1}^{n}I(z_{t-1}\leq r)e_t - (\frac{1}{n^2}\sum_{t=1}^{n}x_t^2)^{-1}(\frac{1}{n^{3/2}}\sum_{t=1}^{n}x_tI(q_{t-1}\leq r))(\frac{1}{n}\sum_{t=1}^{n}x_te_t)\frac{1}{\sqrt{n}} \tag{2.135}$$

From the proof of (i) note that $(\frac{1}{n^2}\sum_{t=1}^{n}x_t^2)^{-1}(\frac{1}{n^{3/2}}\sum_{t=1}^{n}x_tI(q_{t-1}\leq r))(\frac{1}{n}\sum_{t=1}^{n}x_te_t) = O_p(1)$, then

$$\frac{1}{n}\sum_{t=1}^{n}1_{q_{t-1},x_t}(r)e_{x_t} \to_p E(I(z_{t-1}\leq r)e_t) = 0 \tag{2.136}$$

the last equality came from (A.3). The argument for C.2 is the same es C.1, then:

$$\frac{R_1}{n(r-r_0)} = o_p(1) \tag{2.137}$$

Step 2 Using the result from Step 1 is easy to show that $\frac{R_2}{n(r-r_0)} = o_p(1)$.

Step 3: Now consider $\frac{R_4}{n(r-r_0)} = \frac{2(\hat{\alpha}-\alpha)\hat{\delta}}{r-r_0}\Big(\underbrace{\frac{1}{n}\sum_{t=1}^{n}1_{q_{t-1},x_t}(r)1_{x_t}}_{D.1} - \underbrace{\frac{1}{n}\sum_{t=1}^{n}1_{q_{t-1},x_t}(r_0)1_{x_t}}_{D.2}\Big)$. Lets analyze D.1

$$\frac{1}{n}\sum_{t=1}^{n}1_{q_{t-1},x_t}(r)1_{x_t} = \frac{1}{n}\sum_{t=1}^{n}I(z_{t-1}\le r) - \Big(\frac{1}{n^{3/2}}\sum_{t=1}^{n}I(z_{t-1}\le r)x_t\Big)\Big(\frac{1}{n^2}\sum_{t=1}^{n}x_t^2\Big)^{-1}\Big(\frac{1}{n^{3/2}}\sum_{t=1}^{n}x_t\Big)$$
$$\Rightarrow \lambda\Big[1-\Big(\int_0^1 B_x^2\Big)^{-1}\int_0^1 B_x\Big] \tag{2.138}$$

Similarly from D.2 $\Rightarrow \lambda_0\Big[1-\Big(\int_0^1 B_x^2\Big)^{-1}\int_0^1 B_x\Big]$, then $\frac{R_4}{n(r-r_0)} = \frac{2(\hat{\alpha}-\alpha)\hat{\delta}}{(r-r_0)}O_p(1) = o_p(1)$, since $\hat{\delta} = \delta + o_p(1)$ and $(\hat{\alpha}-\alpha) = o_p(1)$.

Step 4: Now consider $\frac{R_5}{n(r-r_0)} = \frac{\hat{\delta}(\hat{\delta}-\delta)}{r-r_0}\Big(\underbrace{\frac{1}{n}\sum_{t=1}^{n}1_{q_{t-1},x_t}(r)^2}_{f.1} - \underbrace{\frac{1}{n}\sum_{t=1}^{n}1_{q_{t-1},x_t}(r_0)1_{q_{t-1},x_t}(r)}_{f.2}\Big)$

From f.1 we have:

$$\frac{1}{n}\sum_{t=1}^{n}1_{q_{t-1},x_t}(r)^2 = \frac{1}{n}\sum_{t=1}^{n}I(z_{t-1}\le r) - \Big(\frac{1}{n^{3/2}}\sum_{t=1}^{n}I(z_{t-1}\le r)x_t\Big)^2\Big(\frac{1}{n^2}\sum_{t=1}^{n}x_t^2\Big)^{-1}$$
$$\Rightarrow \lambda\Big[1-\lambda\Big(\int_0^1 B_x^2\Big)^{-1}\Big(\int_0^1 B_x\Big)^2\Big] \tag{2.139}$$

Similarly from f.1 we can show that f.2 $\Rightarrow \lambda_0\Big[1-\lambda\Big(\int_0^1 B_x^2\Big)^{-1}\Big(\int_0^1 B_x\Big)^2\Big]$, then we can conclude that $\frac{R_5}{n(r-r_0)} = \frac{\hat{\delta}(\hat{\delta}-\delta)}{r-r_0}O_p(1) = o_p(1)$.

Step 5: Consider $\frac{R_5}{n(r-r_0)} = \frac{(\hat{\delta}-\delta)}{(r-r_0)}\frac{1}{n}\sum_{t=1}^{n}(1_{q_{t-1},x_t}(r) - 1_{q_{t-1},x_t}(r_0))^2$ Then we can see that:

$$\frac{1}{n}\sum_{t=1}^{n}(1_{q_{t-1},x_t}(r) - 1_{q_{t-1},x_t}(r_0))^2 \Rightarrow (\lambda-\lambda_0)\Big[1-(\lambda-\lambda_0)\Big(\int_0^1 B_x^2\Big)^{-1}\Big(\int_0^1 B_x\Big)^2\Big] \tag{2.140}$$

again we conclude that $\frac{R_6}{n(r-r_0)} = \frac{(\hat{\delta}-\delta)}{r-r_0}O_p(1) = o_p(1)$.

Step 6 Finally let see $\frac{R_3}{n(r-r_0)} = \frac{\delta^2}{(r-r_0)}\frac{1}{n}\sum_{t=1}^{n}(1_{q_{t-1},x_t}(r) - 1_{q_{t-1},x_t}(r_0))^2$ then

$$\frac{R_3}{n(r-r_0)} \Rightarrow \frac{\delta^2}{(r-r_0)}(\lambda-\lambda_0)\Big[1-(\lambda-\lambda_0)\Big(\int_0^1 B_x^2\Big)^{-1}\Big(\int_0^1 B_x\Big)^2\Big] \tag{2.141}$$

Note that since $r > r_0$, and by A.5 $\lambda > \lambda_0$, then $\frac{\delta^2}{(r-r_0)}(\lambda-\lambda_0) > 0$. Note that also $(\lambda-\lambda_0)\Big(\int_0^1 B_x^2\Big)^{-1}\Big(\int_0^1 B_x\Big)^2 < 1$ then $\frac{R_3}{n(r-r_0)} \Rightarrow \frac{\delta^2}{(r-r_0)} > 0$ showing the desired result.

Proof of Proposition 7

In the paper of Gonzalo and Pitarakis (2006) Lemma 2, shows that when the least square estimator of the threshold parameter is n-consistent, $n|\hat{r}_n - r_0| = O_p(1)$,

$$\frac{1}{\sqrt{n}}\sum I(z_{t-1}\leq \hat{r}_n)I(z_{t-1}\leq r_0)-\frac{1}{\sqrt{n}}\sum I(z_{t-1}\leq r_0)\rightarrow_p 0 \tag{2.142}$$

and we can use the estimation $\hat{r}_n$ as if it is known, r_0, and the rest is the same as in Proposition 5.

Proof of Proposition 8

The proof is available in Gonzalo and Pitarakis (2006), Proposition 1 and 2.

Since the threshold value is unknown, the test statistic proposed is

$$W_n = \operatorname*{Sup}_{r\in(r_L,r_H)} t_{\gamma=0}(r)^2. \tag{2.143}$$

Applying the continuous mapping theorem, we have that

$$W_n \Rightarrow \operatorname*{Sup}_{r\in(r_L,r_H)} t(r)^2. \tag{2.144}$$

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

To obtain the equivalence in equation (2.30), following the work of Gonzalo and Pitarakis (2012), first is to show that the process $W_\varepsilon(s)$ and $V_e(s,\lambda)$ are independent. Since both processes are Gaussian, it is enough to show that both are uncorrelated

$$E(W_\varepsilon(s_1)V_e(s_2,\lambda)) = E(W_\varepsilon(s_1)[W_e(s_2,\lambda)-W_e(s_2,1)]) = E(W_\varepsilon(s_1)W_e(s_2,\lambda))-\lambda E(W_\varepsilon(s_1)W_e(s_2,1))$$
$$=\sigma_{\varepsilon,e}\lambda(s_1\wedge s_2)-\sigma_{\varepsilon,e}\lambda(s_1\wedge s_2)=0 \tag{2.145}$$

Since $W_\varepsilon(s)$ and $V_e(s,\lambda)$ are independent, equipped with $E[V_e(r_1,\lambda_1),V_e(r_2,\lambda_2)]=\sigma_e^2(s_1\wedge s_2)[(\lambda_1\wedge\lambda_2)-\lambda_1\lambda_2]$, which give us $\int_0^1 W_\varepsilon(s)dV_e(s,\lambda)\equiv\mathcal{N}(0,\sigma_e\lambda(1-\lambda)\int_0^1 W_\varepsilon^2(s)ds)$. Normalizing by $\sigma_e^2\int_0^1 W_\varepsilon^2(s)ds$ we get the desired result.

Proof of Proposition 9

To show the invariance principle in proposition 9, we use the following result from Peligrad and Utev (2005).

Theorem 3. *Let $\{d_i\}$ be a stationary sequence with finite second moment $E(d_i^2)<\infty$. Assume that*

$$\sum_{n=1}^{\infty}\frac{||E(S_n|\mathcal{F}_0)||}{n^{3/2}}<\infty \tag{2.146}$$

Then $\{max_{1\leq k\leq n}S_k^2/n : n\geq 1\}$ is uniformly integrable and $W_n(t)\Rightarrow\sqrt{\eta}W(t)$, where η is a non-negative random variable with finite mean $E(\eta)=\sigma^2$ and independent of $\{W(t):t\geq 0\}$.

Moreover η *is determined by the limit*

$$lim_{n\to\infty}\frac{E(S_n^2|\mathcal{I})}{n} = \eta \tag{2.147}$$

where $\mathcal{I}$ *is the invariant sigma field.*

For the sake of application, Merlevede, Peligrad and Utev (2006) formulates sufficient conditions for the invariance principle in terms of the conditional expectation of an individual summand $\{e_i I(z_{i-1} \leq r)\}$ with respect $\mathcal{F}_0$.

Corollary 1. *If*

$$\sum_{i=1}^{\infty}\frac{1}{\sqrt{i}}||E(e_i I(z_{i-1} \leq r)|\mathcal{F}_0)||_2 < \infty \tag{2.148}$$

Then (2.146) is satisfied, then the conclusion of Theorem 1 holds.

Proof of Corollary 1

We have to check that:

$$\sum_{i=1}^{\infty}\frac{1}{\sqrt{i}}||E(I(z_{i-1} \leq r)e_i|\mathcal{F}_0)||_2 < \infty \tag{2.149}$$

First let see what it is $E(I(z_{i-1} \leq r)e_i|\mathcal{F}_0)$. From the independence assumption between e_t and z_t for all t

$$E(I(z_{i-1} \leq r)e_i|\mathcal{F}_0) = E(I(z_{i-1} \leq r)|\mathcal{F}_0)E(e_i|\mathcal{F}_0) \tag{2.150}$$

Call $E(I(z_{i-1} \leq r)|\mathcal{F}_0) = P_{i-1,0}$, since v_i independent w.r.t $\mathcal{F}_{i-1}$, then

$$= P_{i-1,0}E(\sum_{j=-\infty}^{i} a_{i-j}v_j|\mathcal{F}_0) = P_{i-1,0}\sum_{j=-\infty}^{0} a_{i-j}v_j = P_{i-1,0}\sum_{j=i}^{\infty} a_j v_{i-j} \tag{2.151}$$

Then we can see that:

$$||E(I(z_{i-1} \leq r)e_i|\mathcal{F}_0)||_2 = ||P_{i-1,0}\sum_{j=i}^{\infty} a_j v_{i-j}||_2 \tag{2.152}$$

Since $P_{i-1,0} \in [0,1]$ we can see that:

$$||P_{i-1,0}\sum_{j=i}^{\infty} a_j v_{i-j}||_2 \leq ||\sum_{j=i}^{\infty} a_j v_{i-j}||_2 = ||v_j||_2\Big(\sum_{j=i}^{\infty} a_j^2\Big)^{1/2} \tag{2.153}$$

Then we can see that:

$$\sum_{i=1}^{\infty}\frac{1}{\sqrt{i}}||E(I(z_{i-1}\leq r)e_i|\mathcal{F}_0)||_2 \leq ||v_j||_2\sum_{i=1}^{\infty}\frac{1}{\sqrt{i}}\Big(\sum_{j=i}^{\infty}a_j^2\Big)^{1/2} < \infty \tag{2.154}$$

Proof of Theorem 1

The proof of theorem 1, follows the martingale approximation of Hansen (1992). For our case we want to show the convergence of

$$\frac{1}{n}\sum_{i=1}^{n}x_iI(z_{i-1}\leq r)e_i \tag{2.155}$$

Define the filtration $\mathcal{F}_t^x = \sigma\{x_i, e_i, q_i : i \leq r\}$ and denote $E_t(e) = E(e|\mathcal{F}_t^x)$. Then we can construct the following martingale approximation for $I(z_{i-1}\leq r)e_i$. Start with

$$\eta_i = \sum_{k=0}^{\infty}\Big(E_i(I(z_{z_{i-1+k}}\leq r)e_{i+k}) - E_{i-1}(I(z_{z_{i-1+k}}\leq r)e_{i+k})\Big),\quad q_i = \sum_{k=1}^{\infty}E_i(I(z_{z_{i-1+k}}\leq r)e_{i+k}) \tag{2.156}$$

then $I(z_{i-1}\leq r)e_i = \eta_i + q_{i-1} - q_i$, and note that $E_{i-1}(\eta_i) = 0$. Then

$$\frac{1}{n}\sum_{i=1}^{n}x_iI(z_{i-1}\leq r)e_i = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\varepsilon_iI(z_{i-1}\leq r)e_i}_{A} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}x_{i-1}\eta_i}_{B} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}x_{i-1}(q_{i-1}-q_i)}_{C} \tag{2.157}$$

The term A, $\frac{1}{n}\sum_{i=1}^{n}\varepsilon_iI(z_{i-1}\leq r)e_i \rightarrow_p E(\varepsilon_iI(z_{i-1}\leq r)e_i)$.

From the term B, it is easy to see that under our assumptions $\frac{1}{\sqrt{n}}x_{[ns]} \Rightarrow B_x(s)$, $\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}e_tI(z_{t-1}\leq r) \Rightarrow G_e(s,\lambda)$. Then from Theorem 3.1 in Hansen (1992)

$$\frac{1}{n}\sum_{i=1}^{n}x_{i-1}\eta_i \Rightarrow \int_0^1 B_x(s)dG_e(s,\lambda) \tag{2.158}$$

For the last term C, we add and subtract x_iq_i and rewrite,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - x_{i-1})q_i - \frac{1}{n}x_nq_n \tag{2.159}$$

first, observe that

$$\sup_{t\leq n}\frac{1}{n}|x_tq_t| \leq \sup_{t\leq n}\Big|\frac{1}{\sqrt{n}}x_t\Big|\frac{1}{\sqrt{n}}\sup_{t\leq n}|q_t| \rightarrow_p 0 \tag{2.160}$$

since $\underset{t\leq n}{sup}|x_t| = O_p(\sqrt{n})$ and $\frac{1}{\sqrt{n}}\underset{t\leq n}{sup}|q_t| \rightarrow_p 0$. Applying the Corollary of Theorem 3.3 in Hansen (1992)

$$\underset{t\leq n}{sup}\frac{1}{n}\sum_{i=1}^{t}[\varepsilon_i q_i - E(\varepsilon_i q_i)] \tag{2.161}$$

Finally we can see that

$$E\Big(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i q_i\Big) = \frac{1}{n}\sum_{i=1}^{n}E\Big(\varepsilon_i\sum_{k=1}^{\infty}E_i(I(z_{i-1+k}\leq r)e_{i+k})\Big) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{\infty}E\Big(\varepsilon_i I(z_{i-1+k}\leq r)e_{i+k}\Big) \rightarrow \lambda\Lambda_1 \tag{2.162}$$

as $n\rightarrow\infty$, where $\Lambda_1 = lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{\infty}E\Big(\varepsilon_i e_{i+k}\Big)$. Putting everything together we can see that:

$$\frac{1}{n}\sum_{i=1}^{n}x_i I(z_{i-1}\leq r)e_i \Rightarrow \lambda E(\varepsilon_i e_i) + \int_0^1 B_x(s)dG_e(s,\lambda) + \lambda\Lambda_1 \tag{2.163}$$

Proof of Lemma 1

For the case without drift, we know that

$$n(\hat{\Gamma}-\Gamma) = \left(\frac{1}{n}\sum_{t=1}^{n}X_tX_t'\right)^{-1}\left(\frac{1}{n}\sum_{t=1}^{n}X_te_t\right) \tag{2.164}$$

From Caner and Hansen (2001) we know that

$$\left(\frac{1}{n^2}\sum_{t=1}^{n}X_tX_t'\right)^{-1} \Rightarrow \frac{1}{\lambda(1-\lambda)\int_0^1 B_x(s)^2}\begin{pmatrix}\lambda & -\lambda\\ -\lambda & 1\end{pmatrix} \tag{2.165}$$

From Theorem 1 and Theorem 4.1 in Hansen (1992) we can see that

$$\left(\frac{1}{n}\sum_{t=1}^{n}X_te_t\right) \Rightarrow \begin{pmatrix}\int_0^1 B_X(s)dB_e(s) + \Lambda + E(\varepsilon_t e_t)\\ \int_0^1 B_X(s)dG_e(s,\lambda) + \lambda\Lambda + \lambda E(\varepsilon_t e_t)\end{pmatrix} \tag{2.166}$$

Then putting everything together it is easy to see that:

$$n(\hat{\Gamma}-\Gamma) \Rightarrow \begin{pmatrix}\frac{\lambda\Big(\int_0^1 B_x(s)dB_e(s) - \int_0^1 B_x(s)dG_e(s,\lambda) + (1-\lambda)[\Lambda + E(\varepsilon_i e_i)]\Big)}{\lambda(1-\lambda)\int_0^1 B_x(s)^2ds}\\ \frac{\int_0^1 B_x(s)dG_e(s,\lambda) - \lambda\int_0^1 B_x(s)dB_e(s)}{\lambda(1-\lambda)\int_0^1 B_x(s)^2ds}\end{pmatrix} \tag{2.167}$$

For the case where state dependent constants are considered is the same, and omitted here.

Proof of Proposition 10

Start defining the vector $\xi_t = \Big(\varepsilon_t \quad e_t \quad e_t I(z_{t-1} \leq r)\Big)'$ then under our assumption, the following result holds

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{[ns]} \xi_t \to \mathbb{B}(s) \equiv \begin{pmatrix} B_x(s) \\ B_e(s) \\ G_e(s,\lambda) \end{pmatrix} \tag{2.168}$$

with covariance matrix

$$\Omega = lim_{n\to\infty} \frac{1}{n} \Big\{ \Big(\sum_{i=1}^{n} \xi_t \Big) \Big(\sum_{i=1}^{n} \xi_t \Big)' \Big\} \tag{2.169}$$

We can writte in the matrix form

$$\Omega = lim_{n\to\infty} \frac{1}{n} \begin{pmatrix} \big(\sum_{t=1}^n \varepsilon_t\big)^2 & \big(\sum_{t=1}^n \varepsilon_t\big)\big(\sum_{t=1}^n e_t\big) & \big(\sum_{t=1}^n \varepsilon_t\big)\big(\sum_{t=1}^n e_t I(z_{t-1} \leq r)\big) \\ \big(\sum_{t=1}^n \varepsilon_t\big)\big(\sum_{t=1}^n e_t\big) & \big(\sum_{t=1}^n e_t\big)^2 & \big(\sum_{t=1}^n e_t\big)\big(\sum_{t=1}^n e_t I(z_{t-1} \leq r)\big) \\ \big(\sum_{t=1}^n \varepsilon_t\big)\big(\sum_{t=1}^n e_t I(z_{t-1} \leq r)\big) & \big(\sum_{t=1}^n e_t\big)\big(\sum_{t=1}^n e_t I(z_{t-1} \leq r)\big) & \big(\sum_{t=1}^n e_t I(z_{t-1} \leq r)\big)^2 \end{pmatrix} \tag{2.170}$$

Note that under our assumptions and assumption B.3 we can see that the following

1. $lim_{n\to\infty} \frac{1}{n} E\Big(\sum_{t=1}^n \varepsilon_t \Big)^2 = \sigma_\varepsilon^2$
2. Since e_t is a linear process we can use the results from Phillips and Solo (1992) among others to show that $lim_{n\to\infty} \frac{1}{n} E\Big(\sum_{t=1}^n e_t \Big)^2 = C(1)^2 \sigma_v^2$, where $C(1) = \sum_{j=0}^\infty a_j$
3. $lim_{n\to\infty} \frac{1}{n} E\Big(\sum_{t=1}^n e_t \Big)\Big(\sum_{t=1}^n e_t I(z_{t-1} \leq r) \Big) = \lambda C(1)^2 \sigma_v^2$, since

   $$E\Big(\sum_{t=1}^n e_t\Big)\Big(\sum_{t=1}^n e_t[I(z_{t-1} \leq r) + \lambda - \lambda]\Big) = \lambda E\Big(\sum_{t=1}^n e_t\Big)^2 + E\Big(\sum_{t=1}^n e_t\Big)\Big(\sum_{t=1}^n e_t[I(z_{t-1} \leq r) - \lambda]\Big)$$

   note that the firs term in therm in the sum is equal to previous point and the second term

   $$\begin{aligned} E\Big(\sum_{t=1}^n e_t\Big)\Big(\sum_{t=1}^n e_t[I(z_{t-1} \leq r) - \lambda]\Big) &= \sum_{t=1}^n \sum_{j=1}^n E\Big(e_t e_j [I(z_{j-1} \leq r) - \lambda]\Big) \\ &= \sum_{t=1}^n \sum_{j=1}^n E(e_t e_j) E[I(z_{j-1} \leq r) - \lambda] = 0 \end{aligned} \tag{2.171}$$

   where the last inequality came from the independence between $\{e_j\}$ and $\{z_k\}$, $\forall j,k$ and $E[I(z_{j-1} \leq r) - \lambda] = 0$.

4. $lim_{n\to\infty}\frac{1}{n}E\Big(\sum_{t=1}^{n} e_t I(z_{t-1}\leq r)\Big)^2 = \lambda^2 C(1)^2\sigma_v^2 + \mathbb{G}$, where $\mathbb{G} = lim_{n\to\infty}\frac{1}{n}E\Big(\sum_{t=1}^{n} e_t[I(z_{t-1}\leq r)-\lambda]\Big)^2$. We obtain this result, since

$$
\begin{aligned}
E\Big(\sum_{t=1}^{n} e_t[I(z_{t-1}\leq r)-\lambda+\lambda]\Big)^2 &= \lambda^2 E\Big(\sum_{t=1}^{n} e_t\Big)^2 + E\Big(\sum_{t=1}^{n} e_t[I(z_{t-1}\leq r)-\lambda]\Big)^2 \\
&+ 2\lambda E\Big(\sum_{t=1}^{n} e_t\Big)\Big(\sum_{t=1}^{n} e_t[I(z_{t-1}\leq r)-\lambda]\Big) \\
&= \lambda^2 C(1)^2\sigma_v^2 + E\Big(\sum_{t=1}^{n} e_t[I(z_{t-1}\leq r)-\lambda]\Big)^2
\end{aligned}
$$

5. Finally, under our assumptions we can see that

$$
\Big(\sum_{t=1}^{n}\varepsilon_t\Big)\Big(\sum_{t=1}^{n} e_t I(z_{t-1}\leq r)\Big) = \lambda\Big(\sum_{t=1}^{n}\varepsilon_t\Big)\Big(\sum_{t=1}^{n} e_t\Big) = \lambda\sigma_{\varepsilon,e} \tag{2.172}
$$

Then putting everything toghether

$$
\Omega = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon,e} & \lambda\sigma_{\varepsilon,e} \\ \sigma_{\varepsilon,e} & \sigma_v^2 C(1)^2 & \lambda\sigma_v^2 C(1)^2 \\ \lambda\sigma_{\varepsilon,e} & \lambda\sigma_v^2 C(1)^2 & \lambda^2\sigma_v^2 C(1)^2 + \mathbb{G} \end{pmatrix} \tag{2.173}
$$

We can partition the matrix of variance and covariance as $\Omega = LL'$, where

$$
L = \begin{pmatrix} \sigma_\varepsilon & 0 & 0 \\ \frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon} & \Big[C(1)^2\sigma_v^2 - \Big(\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon}\Big)^2\Big]^{1/2} & 0 \\ \lambda\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon} & \lambda\Big[C(1)^2\sigma_v^2 - \Big(\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon}\Big)^2\Big]^{1/2} & \sqrt{\mathbb{G}} \end{pmatrix} \tag{2.174}
$$

Then we can writte $\mathbb{B}(s) = L\mathbb{W}(s,\lambda)$, where $\mathbb{W}(s,\lambda)' = \Big(Wx(s) \quad W_e(s) \quad W_{eI}(s,\lambda)\Big)'$ such that $\mathbb{B}(s)$ can be written as;

$$
\mathbb{B}(s) = \begin{pmatrix} \sigma_\varepsilon W_x(s) \\ \frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon}W_x(s) + \Big[C(1)^2\sigma_v^2 - \Big(\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon}\Big)^2\Big]^{1/2} W_e(s) \\ \lambda\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon}W_x(s) + \lambda\Big[C(1)^2\sigma_v^2 - \Big(\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon}\Big)^2\Big]^{1/2} W_e(s) + \sqrt{\mathbb{G}}W_{eI}(s,\lambda) \end{pmatrix} \tag{2.175}
$$

Then

$$
\begin{aligned}
n(\hat{\gamma}-\gamma) \Rightarrow & \frac{\int_0^1 B_x(s)dG_e(s,\lambda) - \lambda\int_0^1 B_x(s)dB_e(s)}{\lambda(1-\lambda)\int_0^1 B_x(s)^2 ds} \\
\equiv & \frac{\lambda\sigma_{\varepsilon,e}\int_0^1 W_x(s)dW_x(s) + \lambda\sigma_e\left[C(1)^2\sigma_v^2 - \left(\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon}\right)^2\right]^{1/2}\int_0^1 W_x(s)dW_e(s) + \sigma_\varepsilon\sqrt{\mathbb{G}}\int_0^1 W_x(s)dW_{eI}(s,\lambda)}{\lambda(1-\lambda)\sigma_\varepsilon^2\int_0^1 W_x(s)ds} \\
- & \frac{\lambda\sigma_{\varepsilon,e}\int_0^1 W_x(s)dW_x(s) + \lambda\sigma_e\left[C(1)^2\sigma_v^2 - \left(\frac{\sigma_{\varepsilon,e}}{\sigma_\varepsilon}\right)^2\right]^{1/2}\int_0^1 W_x(s)dW_e(s)}{\lambda(1-\lambda)\sigma_\varepsilon^2\int_0^1 W_x(s)ds} \equiv \frac{\sqrt{\mathbb{G}}\int_0^1 W_x(s)dW_{eI}(s,\lambda)}{\lambda(1-\lambda)\sigma_\varepsilon\int_0^1 W_x(s)ds}
\end{aligned}
\tag{2.176}
$$

Note that it easy to see that $\frac{\int_0^1 W_x(s)dW_{eI}(s)}{\sqrt{\int_0^1 W_x^2(s)ds}} \equiv \mathbb{N}(0,1)$. When the threshold parameter is known we have an consistent estimator for $\hat{\mathbb{G}} \to_p \mathbb{G}$, then

$$
t_{\gamma=0}(r_0) = \hat{\gamma}(r_0)\sqrt{\frac{\bar{\lambda}(1-\bar{\lambda})}{\hat{\mathbb{G}}((X(r_0)'X(r_0))^{-1})_{22}}} \tag{2.177}
$$

where $(X(r_0)'X(r_0))^{-1})_{22}$ is the element 2x2 of the following matrix

$$
(X(r_0)'X(r_0))^{-1} = \begin{pmatrix} \sum_{t=1}^n x_t^2 & \sum_{t=1}^n x_t^2 I(z_{t-1}\le r_0) \\ \sum_{t=1}^n x_t^2 I(z_{t-1}\le r_0) & \sum_{t=1}^n x_t^2 I(z_{t-1}\le r_0) \end{pmatrix}^{-1} \tag{2.178}
$$

Then it is easy to see that

$$
t_{\gamma=0}(r_0) \Rightarrow \mathbb{N}(0,1) \tag{2.179}
$$

Proof of Proposition 11

The proof of Proposition 11 is the same as Proposition 10 but changing r_0 for $\hat{r}_n$.

Proof of Lemma 2

We transform the models (No drift) into matrix form:

$$
y_t = X_t^*\Gamma^* + e_t^* \tag{2.180}
$$

where $X_t^* = \begin{pmatrix} x_t & x_t I(z_{t-1}\le r) & \Delta x_{t+k} & \dots & \Delta x_{t-k}\end{pmatrix}'$, $\Gamma^* = \begin{pmatrix}\beta_2 & \gamma & \pi_{-j} & \dots & \pi_j\end{pmatrix}$, and $e^* = \phi_t + \sum_{|j|>K}\pi_j\Delta x_{t-j}$ such that $E(\phi_j x_t) = 0$, $\forall j,t$. Define the scale matrix $D = diag\{(n-2k)^{-1}, (n-2k)^{-1}, (n-2k)^{-1/2}I_k\}$. Using Conditions 2 and (2.38), it can be shown that $\sum_{|j|>K}\pi_j\Delta x_{t-j} = o_p(n^{-1/2})$, which is also proven in Lemma A5 of Saikkonen (1991). Following the analysis of Said and Dickey (1984) Lemma 5.1 and Saikkonen (1991) Lemma A4, we can show that

$$D^{-1}(\hat{\Gamma}^* - \Gamma^*) = (D\sum_t X_t^* X_t'^*)^{-1}(D\sum_t X_t^* e_t^*) \to R^{-1}(D\sum_t X_t^* \phi_t) \tag{2.181}$$

where

$$R = diag\{n^{-2}\sum_t X_t X_t',\ E(U_t U_t')\}, \quad with \ \ U_t = (\Delta x_{t+k}, \ldots, \Delta x_{t-k})' \tag{2.182}$$

Solving and rearranging we obtain the asymptotic result of Lemma 2. The order of probability for $\sum_{i=-k}^{k}(\tilde{\pi}_j - \pi_j)$ is given in the appendix of Saikkonen (1991).

Proof of Theorem 2

Using Lemma 2 we want to show that

$$n^{-1/2}\tilde{S}_{[ns]} = n^{-1/2}\sum_{j=1}^{[ns]} e_j^* + n^{-1/2}\sum_{j=1}^{[ns]}(\hat{e}_j^* - e_j^*) \tag{2.183}$$

and note that

$$(\hat{e}_t^* - e_t^*) = -x_t(\hat{\beta} - \beta) - x_t I(z_{t-1} \leq r)(\hat{\gamma} - \gamma) - U_t'(\tilde{\Pi} - \Pi) \tag{2.184}$$

Given the structure of e_t^* we can write (2.184) as follows,

$$\begin{aligned} n^{-1/2}\tilde{S}_{[ns]} = {} & n^{-1/2}\sum_{j=1}^{[ns]}\tilde{e}_j + n^{-1/2}\sum_{j=1}^{[ns]}\Big(\sum_{i>|k|}\pi_i \Delta x_{j-i}\Big) - n(\hat{\beta} - \beta_2)\frac{1}{n^{3/2}}\sum_{j=1}^{[ns]} z_j \\ & - n(\hat{\gamma} - \gamma_2)\frac{1}{n^{3/2}}\sum_{j=1}^{[ns]} z_j I(z_{j-1} \leq r) - n^{-1/2}\sum_{j=1}^{[ns]}\Big(\sum_{i=-k}^{k}\Delta x_{j-1}(\tilde{\pi}_i - \pi_i)\Big) \end{aligned} \tag{2.185}$$

Note that the first element of the sum converges to

$$n^{-1/2}\sum_{j=1}^{[ns]}\tilde{e}_j \Rightarrow B_{e,x}(s) \tag{2.186}$$

The third element of the sum

$$n(\hat{\beta} - \beta_2)\frac{1}{n^{3/2}}\sum_{j=1}^{[ns]} z_j \Rightarrow \int_0^s B_x(s)\Big(\frac{\int_0^1 B_x(s)dB_{e,x}(s)}{(1-\lambda)\int_0^1 B_x^2(s)ds} - \frac{\int_0^1 B_x(s)dG_{e,x}(s,\lambda)}{(1-\lambda)\int_0^1 B_x^2(s)ds}\Big) \tag{2.187}$$

For the four element

$$n(\hat{\gamma}-\gamma_2)\frac{1}{n^{3/2}}\sum_{j=1}^{[ns]} z_j I(z_{j-1}\leq r) \Rightarrow \int_0^s B_x(s)\Big(\frac{\int_0^1 B_x(s)dG_{e,x}(s,\lambda)}{(1-\lambda)\int_0^1 B_x^2(s)ds} - \lambda\frac{\int_0^1 B_x(s)dB_{e,x}(s)}{(1-\lambda)\int_0^1 B_x^2(s)ds}\Big) \tag{2.188}$$

Then from (2.187) and (2.188) is easy to see that

$$-n(\hat{\beta}-\beta_2)\frac{1}{n^{3/2}}\sum_{j=1}^{[ns]} z_j - n(\hat{\gamma}-\gamma_2)\frac{1}{n^{3/2}}\sum_{j=1}^{[ns]} z_j I(z_{j-1}\leq r) \Rightarrow -\int_0^s B_x(s)\frac{\int_0^1 B_x(s)dB_{e,x}(s)}{\int_0^1 B_x^2(s)ds} \tag{2.189}$$

We have to show that $n^{-1/2}\sum_{j=1}^{[ns]}\Big(\sum_{i>|k|}\pi_i\Delta x_{j-i}\Big) \to 0$ uniformly in s

$$\begin{aligned}
E\underset{s\leq 1}{Sup}\left|\frac{1}{n^{1/2}}\sum_{j=1}^{[ns]}\Big(\sum_{|i|>k}^{\infty}\varepsilon_{j-i}\pi_i\Big)\right| &\leq E\underset{s\leq 1}{Sup}\frac{1}{n^{1/2}}\sum_{j=1}^{[ns]}\Big(\sum_{|i|>k}^{\infty}|\varepsilon_{j-i}||\pi_i|\Big)\\
&= \frac{1}{n^{1/2}}\sum_{j=1}^{[ns]}\Big(\sum_{|i|>k}^{\infty}E|\varepsilon_{j-i}||\pi_i|\Big)\\
&\leq \underset{t}{Sup}E|\varepsilon_t|\frac{1}{n^{1/2}}\sum_{|i|>k}^{\infty}|\pi_i| \to 0
\end{aligned} \tag{2.190}$$

Then by Markov inequality $\underset{s\leq 1}{Sup}\left|\frac{1}{n^{1/2}}\sum_{j=1}^{[ns]}\Big(\sum_{|i|>k}^{\infty}\varepsilon_{j-i}\pi_i\Big)\right| \to 0$

Finally the proof where $n^{-1/2}\sum_{j=1}^{[ns]}\Big(\sum_{i=-k}^{k}\Delta x_{j-1}(\tilde{\pi}_i-\pi_i)\Big) \to 0$ uniformly in s can be found in Shin (1994).

its easy to see that

$$n^{-1/2}\tilde{S}_{[ns]} \Rightarrow B_{e,x}(s) - \int_0^s B_x(s)\Big(\int_0^1 B_x^2(s)ds\Big)^{-1}\int_0^1 B_x(s)dB_{e,x}(s) \tag{2.191}$$

Since $B_{e,x} = \omega_{e,x}^{1/2}W_e$ and $B_{e,x}$ is independent of B_x

$$Q_{e,x} = \omega_{e,x}^{1/2}W_e(s) - \omega_{e,x}^{1/2}\int_0^s W_x(s)\Big(\int_0^1 W_x^2(s)ds\Big)^{-1}\int_0^1 W_x(s)dW_e(s) = \omega_{e,x}^{1/2}Q(s) \tag{2.192}$$

The estimator of the long run variance of $\tilde{e}_t$, $\tilde{\sigma}_e(l)$ is a consistent estimator of $\omega_{e,x}$, since $\sum_{|j|>k}^{\infty}\Delta x_{t-j}\pi_j = o_p(n^{-1/2})$ and Theorem 3 in Hansen (1992). Therefore

$$\tilde{CI} \Rightarrow \int_0^1 Q^2(s)ds \tag{2.193}$$

For the case where state dependent drift is included is the same the drift-less case.

2.B Tables

Table 2.11: Size of KPSS test when $r = \bar{r}$ and No Drift, long run equation shocks are i.i.d, $l = 0$

		$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0$	5.45	4.83
	$\gamma = 1$	5.28	4.85
$n = 500$	$\gamma = 0$	5.04	5.01
	$\gamma = 1$	5.13	4.89
$n = 1000$	$\gamma = 0$	5.49	4.70
	$\gamma = 1$	4.95	5.22

Table 2.12: Size of KPSS test when $r = \bar{r}$ and No Drift, long run equation shocks are i.i.d, $l = 4(n/100)^{1/4}$

		$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0$	5.37	5.34
	$\gamma = 1$	5.02	4.79
$n = 500$	$\gamma = 0$	5.07	4.75
	$\gamma = 1$	5.36	4.88
$n = 1000$	$\gamma = 0$	4.55	4.96
	$\gamma = 1$	4.69	5.25

Table 2.13: Size of KPSS test when $r = \bar{r}$ and No Drift, long run equation shocks are i.i.d, $l = 12(n/100)^{1/4}$

		$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0$	5.21	5.06
	$\gamma = 1$	5.10	5.12
$n = 500$	$\gamma = 0$	4.67	4.64
	$\gamma = 1$	4.90	4.87
$n = 1000$	$\gamma = 0$	5.15	5.27
	$\gamma = 1$	5.08	5.16

Table 2.14: Size of KPSS test when $r = \bar{r}$ with state dependent Drift, long run equation shocks are i.i.d, $l = 0$

		$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0$	4.83	4.19
	$\gamma = 1$	5.03	4.24
$n = 500$	$\gamma = 0$	5.04	4.88
	$\gamma = 1$	5.39	4.73
$n = 1000$	$\gamma = 0$	4.88	4.28
	$\gamma = 1$	5.31	4.81

Table 2.15: Size of KPSS test when $r = \bar{r}$ with state dependent Drift, long run equation shocks are i.i.d, $l = 4(n/100)^{1/4}$

		$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0$	4.91	4.22
	$\gamma = 1$	4.83	4.75
$n = 500$	$\gamma = 0$	5.42	4.69
	$\gamma = 1$	5.60	5.24
$n = 1000$	$\gamma = 0$	5.03	5.52
	$\gamma = 1$	4.92	4.47

Table 2.16: Size of KPSS test when $r = \bar{r}$ with state dependent Drift, long run equation shocks are i.i.d, $l = 12(n/100)^{1/4}$

		$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0$	5.40	4.84
	$\gamma = 1$	5.35	4.75
$n = 500$	$\gamma = 0$	5.00	5.19
	$\gamma = 1$	4.98	5.10
$n = 1000$	$\gamma = 0$	5.23	5.17
	$\gamma = 1$	5.15	5.2

Table 2.17: Size of KPSS test when threshold value is unknown but can be estimated with state dependent drift, long run equation shocks are i.i.d, $l = 0$

		$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0$	5.50	4.70
	$\gamma = 1$	4.20	4.30
$n = 500$	$\gamma = 0$	4.80	4.20
	$\gamma = 1$	4.20	5.30
$n = 1000$	$\gamma = 0$	5.00	5.80
	$\gamma = 1$	4.80	4.50

Table 2.18: Size of KPSS test when threshold value is unknown but can be estimated with state dependent drift, long run equation shocks are i.i.d, $l = 4(n/100)^{1/4}$

		$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0$	5.19	4.36
	$\gamma = 1$	4.94	5.27
$n = 500$	$\gamma = 0$	5.03	4.82
	$\gamma = 1$	4.87	5.60
$n = 1000$	$\gamma = 0$	4.73	4.72
	$\gamma = 1$	5.13	5.62

Table 2.19: Size of KPSS test when threshold value is unknown but can be estimated with state dependent drift, long run equation shocks are i.i.d, $l = 12(n/100)^{1/4}$

		$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0$	5.53	4.89
	$\gamma = 1$	5.39	4.86
$n = 500$	$\gamma = 0$	5.08	4.97
	$\gamma = 1$	5.03	5.07
$n = 1000$	$\gamma = 0$	4.82	5.32
	$\gamma = 1$	5.30	4.85

Table 2.20: Power of the KPSS for different values of σ_u^2, long run equation shocks are i.i.d, No Drift and threshold parameter known, $l = 0$

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$	$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$
$n = 200$	$\gamma = 0$	9.49	72.43	98.26	9.18	70.71	97.78
	$\gamma = 1$	9.49	71.00	98.02	9.11	71.21	97.92
$n = 500$	$\gamma = 0$	25.67	94.81	99.93	25.28	94.71	99.98
	$\gamma = 1$	24.97	94.62	99.97	25.56	94.43	99.93
$n = 1000$	$\gamma = 0$	48.98	99.34	99.99	48.46	99.41	100.00
	$\gamma = 1$	48.64	99.40	100.00	48.63	99.34	100.00

Table 2.21: Power of the KPSS for different values of σ_u^2, long run equation shocks are i.i.d, No Drift and threshold parameter known, $l = 4(n/100)^{1/4}$

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$	$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$
$n = 200$	$\gamma = 0$	8.81	55.44	69.95	8.77	54.49	69.51
	$\gamma = 1$	9.11	55.81	70.34	8.98	55.66	69.82
$n = 500$	$\gamma = 0$	22.93	82.22	87.11	23.90	81.62	87.88
	$\gamma = 1$	22.78	81.92	87.50	23.19	82.16	87.11
$n = 1000$	$\gamma = 0$	45.70	93.49	56.64	45.49	93.74	95.07
	$\gamma = 1$	46.81	93.30	95.51	45.95	93.00	95.18

Table 2.22: Power of the KPSS for different values of σ_u^2, long run equation shocks are i.i.d, No Drift and threshold parameter known, $l = 12(n/100)^{1/4}$

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$	$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$
$n = 200$	$\gamma = 0$	8.19	40.18	47.45	8.20	40.75	47.64
	$\gamma = 1$	7.90	39.95	46.67	8.05	39.89	46.85
$n = 500$	$\gamma = 0$	21.29	59.92	63.98	21.10	61.19	63.56
	$\gamma = 1$	20.84	61.02	64.81	20.68	61.25	64.23
$n = 1000$	$\gamma = 0$	41.24	75.68	77.17	41.84	76.51	77.49
	$\gamma = 1$	42.12	76.75	77.61	40.77	77.04	78.20

Table 2.23: Power of the KPSS for different values of σ_u^2, long run equation shocks are i.i.d, with state dependent Drift and threshold parameter known, $l = 0$

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$	$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$
$n = 200$	$\gamma = 0$	7.81	74.26	99.88	7.04	70.88	99.82
	$\gamma = 1$	8.05	74.87	99.89	7.06	70.67	99.83
$n = 500$	$\gamma = 0$	21.73	97.69	100.00	20.72	96.89	100.00
	$\gamma = 1$	21.52	97.61	100.00	20.11	96.95	100.00
$n = 1000$	$\gamma = 0$	46.59	99.91	100.00	44.59	99.91	100.00
	$\gamma = 1$	46.68	99.92	100.00	44.58	99.92	100.00

Table 2.24: Power of the KPSS for different values of σ_u^2, long run equation shocks are i.i.d, with state dependent Drift and threshold parameter known, $l = 4(n/100)^{1/4}$

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$	$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$
$n = 200$	$\gamma = 0$	7.51	64.21	87.48	6.82	60.87	86.72
	$\gamma = 1$	8.41	65.15	88.29	6.66	61.24	86.29
$n = 500$	$\gamma = 0$	21.06	92.54	97.92	19.29	91.71	98.05
	$\gamma = 1$	20.84	92.97	98.15	20.57	92.12	97.85
$n = 1000$	$\gamma = 0$	44.31	98.89	99.73	44.15	99.04	99.72
	$\gamma = 1$	44.69	99.02	99.70	43.38	99.07	99.57

Table 2.25: Power of the KPSS for different values of σ_u^2, long run equation shocks are i.i.d, with state dependent Drift and threshold parameter known, $l = 12(n/100)^{1/4}$

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$	$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$
$n = 200$	$\gamma = 0$	8.14	49.83	61.61	7.15	47.28	59.68
	$\gamma = 1$	7.88	48.66	61.84	7.19	48.58	61.44
$n = 500$	$\gamma = 0$	19.71	78.85	83.79	19.49	78.38	83.35
	$\gamma = 1$	19.33	78.34	83.39	19.43	78.07	82.83
$n = 1000$	$\gamma = 0$	41.24	92.69	94.23	41.16	92.24	94.29
	$\gamma = 1$	42.65	92.53	93.86	42.68	91.64	93.48

Table 2.26: Power of the KPSS for different values of σ_u^2, long run equation shocks are i.i.d, with state dependent Drift and threshold parameter Unknown and estimated, $l = 0$

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$	$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$
$n = 200$	$\gamma = 0$	7.87	73.79	99.87	6.64	70.52	99.80
	$\gamma = 1$	7.72	71.57	99.83	10.75	67.83	99.68
$n = 500$	$\gamma = 0$	21.27	97.51	100.00	19.67	97.08	100.00
	$\gamma = 1$	19.62	95.64	100.00	21.79	95.43	100.00
$n = 1000$	$\gamma = 0$	46.55	99.99	100.00	45.08	99.89	100.00
	$\gamma = 1$	42.51	98.96	100.00	44.17	99.32	100.00

Table 2.27: Power of the KPSS for different values of σ_u^2, long run equation shocks are i.i.d, with state dependent Drift and threshold parameter Unknown and estimated, $l = 4(n/100)^{1/4}$

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$	$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$
$n = 200$	$\gamma = 0$	7.93	63.98	88.52	6.39	60.57	85.78
	$\gamma = 1$	7.47	60.76	87.82	7.38	55.26	84.98
$n = 500$	$\gamma = 0$	21.31	92.73	98.07	19.26	92.07	97.77
	$\gamma = 1$	19.01	89.83	98.02	17.57	87.75	97.78
$n = 1000$	$\gamma = 0$	45.28	99.09	99.77	44.88	98.83	99.72
	$\gamma = 1$	40.92	98.15	99.70	40.29	96.46	99.68

Table 2.28: Power of the KPSS for different values of σ_u^2, long run equation shocks are i.i.d, with state dependent Drift and threshold parameter Unknown and estimated, $l = 12(n/100)^{1/4}$

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$	$\sigma_u^2 = 0.01$	$\sigma_u^2 = 0.1$	$\sigma_u^2 = 1$
$n = 200$	$\gamma = 0$	8.41	48.73	60.97	7.08	46.47	59.26
	$\gamma = 1$	7.75	48.83	60.65	7.61	46.75	59.94
$n = 500$	$\gamma = 0$	19.89	78.00	83.22	18.70	76.77	83.08
	$\gamma = 1$	19.44	78.30	83.18	18.74	78.03	83.25
$n = 1000$	$\gamma = 0$	42.45	92.05	94.29	41.67	92.16	93.66
	$\gamma = 1$	43.36	92.27	94.12	40.50	92.34	93.74

Table 2.29: Size of the KPSS test, whe $e_t \approx AR(1)$, for different level of persistence and No Drift. Threshold parameter Known, $l = 0$

		$\rho_z = 0.5$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	0.00	0.14	1.90	10.52	27.99	61.79
	$\gamma = 1$	0.00	0.11	1.87	11.15	27.16	61.52
$n = 500$	$\gamma = 0$	0.00	0.06	1.87	11.65	27.78	63.41
	$\gamma = 1$	0.00	0.08	1.80	10.54	27.98	63.08
$n = 1000$	$\gamma = 0$	0.00	0.01	1.73	10.99	27.37	63.23
	$\gamma = 1$	0.00	0.05	1.51	11.00	27.80	63.73
		$\rho_z = 0.9$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	0.00	0.13	1.86	10.40	27.31	60.53
	$\gamma = 1$	0.00	0.14	1.73	10.74	26.69	60.57
$n = 500$	$\gamma = 0$	0.00	0.01	1.82	10.65	26.70	62.66
	$\gamma = 1$	0.00	0.12	1.69	11.31	27.67	63.60
$n = 1000$	$\gamma = 0$	0.00	0.15	1.47	11.47	27.03	64.61
	$\gamma = 1$	0.00	0.11	1.74	10.71	27.42	64.02

Table 2.30: Size of the KPSS test, whe $e_t \approx AR(1)$, for different level of persistence and No Drift. Threshold parameter Known, $l = 4(n/100)^{1/4}$

		$\rho_z = 0.5$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	1.56	2.96	4.34	5.75	8.06	18.82
	$\gamma = 1$	1.50	2.96	3.91	5.66	8.31	19.73
$n = 500$	$\gamma = 0$	0.97	3.44	4.33	5.61	7.93	16.89
	$\gamma = 1$	1.10	2.08	4.34	6.05	8.02	16.86
$n = 1000$	$\gamma = 0$	1.81	3.27	4.35	5.34	6.95	15.14
	$\gamma = 1$	2.03	3.56	4.57	5.69	7.10	15.53
		$\rho_z = 0.9$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	1.48	3.14	4.24	5.52	7.56	19.93
	$\gamma = 1$	1.59	2.82	4.17	5.70	8.25	18.56
$n = 500$	$\gamma = 0$	1.06	2.89	4.15	5.70	7.80	16.41
	$\gamma = 1$	1.07	3.14	4.65	5.68	7.78	17.51
$n = 1000$	$\gamma = 0$	1.59	3.32	4.57	5.69	7.10	15.53
	$\gamma = 1$	1.88	3.44	4.74	5.90	5.56	15.15

Table 2.31: Size of the KPSS test, whe $e_t \approx AR(1)$, for different level of persistence and No Drift. Threshold parameter Known, $l = 12(n/100)^{1/4}$

		$\rho_z = 0.5$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	1.98	3.55	4.15	4.98	5.89	9.29
	$\gamma = 1$	2.09	3.54	4.41	4.99	5.68	8.82
$n = 500$	$\gamma = 0$	2.80	4.46	4.74	5.32	5.74	8.39
	$\gamma = 1$	2.71	4.14	4.54	5.04	5.55	8.50
$n = 1000$	$\gamma = 0$	3.04	4.31	4.67	5.17	5.72	5.17
	$\gamma = 1$	3.13	4.34	4.90	5.24	5.65	7.71
		$\rho_z = 0.9$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	2.09	3.79	4.26	4.77	5.92	8.51
	$\gamma = 1$	2.07	3.75	4.17	4.85	5.91	9.02
$n = 500$	$\gamma = 0$	2.71	4.07	4.69	5.19	5.33	8.75
	$\gamma = 1$	2.76	4.07	4.69	5.44	5.57	8.62
$n = 1000$	$\gamma = 0$	3.01	4.37	4.42	4.59	5.91	8.05
	$\gamma = 1$	2.69	4.07	4.58	5.43	5.93	8.20

Table 2.32: Size of the KPSS test, whe $e_t \approx AR(1)$, for different level of persistence with state dependent Drift. Threshold parameter Known, $l = 0$

		$\rho_z = 0.5$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	0.00	0.12	1.40	12.42	36.15	84.37
	$\gamma = 1$	0.00	0.11	1.84	11.62	36.17	84.93
$n = 500$	$\gamma = 0$	0.00	0.11	1.69	12.60	36.68	88.13
	$\gamma = 1$	0.00	0.14	1.47	12.01	36.76	88.91
$n = 1000$	$\gamma = 0$	0.00	0.12	1.45	12.25	37.62	88.99
	$\gamma = 1$	0.00	0.12	1.63	12.79	37.89	89.80
		$\rho_z = 0.9$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	0.00	0.09	1.63	9.95	32.17	81.84
	$\gamma = 1$	0.00	0.07	1.26	10.06	32.28	81.68.57
$n = 500$	$\gamma = 0$	0.00	0.01	1.52	11.35	35.58	87.08
	$\gamma = 1$	0.00	0.08	1.48	11.03	34.64	87.14
$n = 1000$	$\gamma = 0$	0.00	0.04	1.57	11.75	36.90	88.95
	$\gamma = 1$	0.00	0.09	1.43	11.90	36.79	88.90

Table 2.33: Size of the KPSS test, whe $e_t \approx AR(1)$, for different level of persistence with state dependent Drift. Threshold parameter Known, $l = 4(n/100)^{1/4}$

		$\rho_z = 0.5$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	1.98	3.20	4.31	5.86	9.01	22.26
	$\gamma = 1$	2.00	3.23	4.44	6.18	8.44	21.67
$n = 500$	$\gamma = 0$	1.15	3.06	4.22	5.34	8.21	20.36
	$\gamma = 1$	1.03	3.33	4.60	5.74	8.34	19.93
$n = 1000$	$\gamma = 0$	1.95	3.42	4.96	5.74	7.66	17.88
	$\gamma = 1$	1.87	3.38	4.28	5.62	7.79	17.48
		$\rho_z = 0.9$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	1.78	2.73	3.88	5.30	7.42	19.14
	$\gamma = 1$	1.96	2.81	4.05	5.37	7.30	19.64
$n = 500$	$\gamma = 0$	1.11	2.95	4.12	5.51	8.11	19.08
	$\gamma = 1$	1.07	3.01	4.11	5.40	8.26	18.75
$n = 1000$	$\gamma = 0$	1.88	3.43	4.25	5.82	7.30	17.36
	$\gamma = 1$	1.91	3.04	4.09	5.39	7.04	17.00

Table 2.34: Size of the KPSS test, whe $e_t \approx AR(1)$, for different level of persistence with state dependent Drift. Threshold parameter Known, $l = 12(n/100)^{1/4}$

		$\rho_z = 0.5$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	3.11	4.75	4.81	5.82	6.11	10.26
	$\gamma = 1$	3.09	4.52	5.03	5.42	6.63	9.74
$n = 500$	$\gamma = 0$	2.87	4.42	4.90	5.12	6.40	9.09
	$\gamma = 1$	3.05	4.33	4.70	5.22	5.93	9.07
$n = 1000$	$\gamma = 0$	3.27	4.38	5.12	5.10	5.62	8.49
	$\gamma = 1$	3.33	3.94	4.77	5.63	6.10	8.60
		$\rho_z = 0.9$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	3.28	4.23	4.29	5.08	5.76	9.56
	$\gamma = 1$	3.09	4.39	4.63	5,29	6.02	9.46
$n = 500$	$\gamma = 0$	2.90	3.88	4.69	5.02	5.86	8.60
	$\gamma = 1$	3.00	4.56	4.54	5.39	5.87	8.60
$n = 1000$	$\gamma = 0$	3.35	4.42	4.97	5.06	8.52	7.82
	$\gamma = 1$	3.16	4.48	4.67	5.46	5.57	8.14

Table 2.35: Size of the KPSS test, when $e_t \approx AR(1)$,for different level of persistence with state dependent Drift. Threshold parameter Estimated, $l = 12(n/100)^{1/4}$

		$\rho_z = 0.5$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	3.26	4.92	4.97	5.40	6.51	10.36
	$\gamma = 1$	3.02	4.15	5.12	5.79	6.51	10.22
$n = 500$	$\gamma = 0$	3.20	4.54	4.69	5.31	5.74	9.34
	$\gamma = 1$	3.22	4.79	5.16	5.86	6.19	9.62
$n = 1000$	$\gamma = 0$	3.29	4.41	4.81	4.99	5.95	8.72
	$\gamma = 1$	3.12	4.53	5.01	5.38	5.91	8.61
		$\rho_z = 0.9$					
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	3.06	4.54	4.46	5.55	5.89	8.62
	$\gamma = 1$	3.07	4.50	4.29	5.08	5.84	9.38
$n = 500$	$\gamma = 0$	3.21	4.38	4.86	5.34	5.63	8.85
	$\gamma = 1$	2.94	4.01	4.59	5.32	6.08	8.47
$n = 1000$	$\gamma = 0$	3.46	4.51	4.48	5.40	6.11	8.36
	$\gamma = 1$	3.11	4.34	4.56	5.16	6.12	8.20

Table 2.36: Size of the KPSS test, whe $e_t \approx AR(1)$ and correlated with x_t, for different level of persistence and No Drift. Threshold parameter known and $l = 0$.

		$\sigma_{\varepsilon,\eta} = 0.5$						
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	0.00	0.11	1.71	4.58	10.18	26.15	55.92
	$\gamma = 1$	0.00	0.13	1.93	4.93	10.60	25.02	56.12
$n = 500$	$\gamma = 0$	0.00	0.06	1.73	5.39	10.80	26.66	59.12
	$\gamma = 1$	0.00	0.01	1.88	4.88	10.90	26.91	59.16
$n = 1000$	$\gamma = 0$	0.00	0.07	1.90	4.85	10.33	28.01	62.27
	$\gamma = 1$	0.00	0.15	1.81	5.13	10.50	27.80	61.85
		$\sigma_{\varepsilon,\eta} = 0.8$						
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	0.00	0.14	1.79	5.34	10.79	26.21	59.16
	$\gamma = 1$	0.00	0.09	1.63	4.78	10.15	25.44	58.01
$n = 500$	$\gamma = 0$	0.00	0.14	1.74	5.08	11.11	26.36	60.45
	$\gamma = 1$	0.00	0.70	1.77	5.14	10.50	26.73	60.92
$n = 1000$	$\gamma = 0$	0.00	0.16	1.74	5.19	11.19	27.22	62.59
	$\gamma = 1$	0.00	0.09	1.97	5.28	11.16	27.68	62.37

Table 2.37: Size of the KPSS test, whe $e_t \approx AR(1)$ and correlated with x_t, for different level of persistence and No Drift. Threshold parameter known and $l = 4(n/100)^{1/4}$.

		$\sigma_{\varepsilon,\eta} = 0.5$						
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	1.51	2.66	4.29	4.73	4.35	7.90	17.11
	$\gamma = 1$	1.35	2.85	4.46	4.63	5.23	7.91	17.23
$n = 500$	$\gamma = 0$	1.14	3.20	4.06	5.48	5.55	8.00	16.57
	$\gamma = 1$	1.00	3.24	4.24	4.84	5.89	7.79	15.68
$n = 1000$	$\gamma = 0$	1.84	3.43	4.31	5.21	5.55	7.38	14.76
	$\gamma = 1$	1.56	3.37	4.36	5.23	5.89	7.03	14.57
		$\sigma_{\varepsilon,\eta} = 0.8$						
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	1.55	2.88	3.94	4.18	5.57	8.07	17.57
	$\gamma = 1$	1.47	3.03	4.38	5.04	5.36	7.40	17.49
$n = 500$	$\gamma = 0$	1.13	3.15	4.21	5.01	5.69	8.08	15.73
	$\gamma = 1$	0.92	3.30	4.24	4.88	5.74	7.21	15.70
$n = 1000$	$\gamma = 0$	1.71	3.57	4.51	4.98	5.83	7.67	15.11
	$\gamma = 1$	1.67	3.21	4.18	4.92	5.57	7.50	14.57

Table 2.38: Size of the KPSS test, whe $e_t \approx AR(1)$ and correlated with x_t, for different level of persistence and No Drift. Threshold parameter known and $l = 12(n/100)^{1/4}$.

		$\sigma_{\varepsilon,\eta} = 0.5$						
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	2.01	3.57	4.09	4.44	5.08	5.48	8.11
	$\gamma = 1$	2.14	3.72	4.22	4.28	4.77	5.75	8.44
$n = 500$	$\gamma = 0$	2.89	4.33	4.99	4.53	4.87	5.77	7.65
	$\gamma = 1$	2.36	3.90	4.85	4.71	4.98	5.34	7.92
$n = 1000$	$\gamma = 0$	3.02	4.21	4.90	4.84	5.13	5.51	7.46
	$\gamma = 1$	3.01	4.08	4.62	4.94	5.08	5.55	7.30
		$\sigma_{\varepsilon,\eta} = 0.8$						
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	2.27	3.60	4.25	4.64	4.68	5.11	7.50
	$\gamma = 1$	2.04	3.55	4.06	4.45	4.84	5.00	7.83
$n = 500$	$\gamma = 0$	2.61	3.93	4.36	4.94	4.76	5.71	7.73
	$\gamma = 1$	2.87	3.86	4.16	4.98	5.21	5.38	7.40
$n = 1000$	$\gamma = 0$	3.17	4.54	4.80	4.92	5.08	5.58	7.47
	$\gamma = 1$	2.95	4.20	4.67	5.05	5.24	5.81	7.64

Table 2.39: Size of the KPSS test, whe $e_t \approx AR(1)$ and correlated with x_t, for different level of persistence and with State dependent Drift. Threshold parameter known, and $l = 12(n/100)^{1/4}$.

		$\sigma_{\varepsilon,\eta} = 0.8$						
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	2.07	3.22	4.86	5.12	5.64	4.71	6.32
	$\gamma = 1$	2.08	3.17	4.90	4.90	5.04	5.14	6.69
$n = 500$	$\gamma = 0$	2.88	3.67	4.73	4.81	4.83	6.02	7.73
	$\gamma = 1$	2.66	3.53	4.79	5.35	4.91	5.01	6.69
$n = 1000$	$\gamma = 0$	3.10	4.25	4.65	5.24	5.14	5.31	7.46
	$\gamma = 1$	3.15	4.05	4.65	4.95	4.81	5.60	7.39
		$\sigma_{\varepsilon,\eta} = 0.5$						
		$\rho = -0.8$	$\rho = -0.5$	$\rho = -0.2$	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
$n = 200$	$\gamma = 0$	2.76	4.31	4.81	5.09	5.19	5.43	7.35
	$\gamma = 1$	2.58	4.19	4.91	4.86	4.28	5.33	7.17
$n = 500$	$\gamma = 0$	2.77	3.92	5.18	5.08	5.33	5.26	7.30
	$\gamma = 1$	2.62	3.92	4.96	5.15	5.44	5.41	6.98
$n = 1000$	$\gamma = 0$	2.95	4.06	5.05	5.29	5.43	5.72	7.23
	$\gamma = 1$	2.96	4.06	4.84	5.20	5.58	5.73	7.12

Table 2.40: Size of the test for threshold effect with state dependent Drift, long run equation shocks are i.i.d

	Threshold.P Known		Threshold.P Unknown	
	$\rho_z = 0.5$	$\rho_z = 0.9$	$\rho_z = 0.5$	$\rho_z = 0.9$
n=200	5.16	4.98	5.48	5.28
n=500	4.84	5.08	4.88	4.75
n=1000	5.22	5.00	4.64	4.73

Table 2.41: Size of the test for threshold effect Without Drift, long run equation shocks are i.i.d

	$\rho_z = 0.5$	$\rho_z = 0.9$
n=200	5.40	5.40
n=500	5.21	5.38
n=1000	5.26	5.19

Table 2.42: Power of the test for threshold effect for different values of γ, long run equation shocks are i.i.d, With state dependent Drift

		Threshold.P Known		Threshold.P Unknown	
		$\rho_z = 0.5$	$\rho_z = 0.9$	$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0.01$	7.16	7.42	6.88	7.46
	$\gamma = 0.05$	45.95	44.03	41.61	30.70
	$\gamma = 0.10$	84.86	82.57	78.16	74.13
	$\gamma = 0.50$	100.00	100.00	99.86	99.14
$n = 500$	$\gamma = 0.01$	17.33	16.96	17.03	16.64
	$\gamma = 0.05$	93.28	92.25	90.29	89.60
	$\gamma = 0.10$	99.96	99.83	98.90	98.26
	$\gamma = 0.50$	100.00	100.00	100.00	100.00
$n = 1000$	$\gamma = 0.01$	45.65	46.49	44.92	44.30
	$\gamma = 0.05$	99.87	99.93	99.46	99.49
	$\gamma = 0.10$	100.00	100.00	99.96	99.91
	$\gamma = 0.50$	100.00	100.00	100.00	100.00

Table 2.43: Power of the test for threshold effect for different values of γ, long run equation shocks are i.i.d, Without state dependent Drift

		$\rho_z = 0.5$	$\rho_z = 0.9$
$n = 200$	$\gamma = 0.01$	10.89	11.62
	$\gamma = 0.05$	70.22	69.30
	$\gamma = 0.10$	94.75	94.11
	$\gamma = 0.50$	100.00	100.00
$n = 500$	$\gamma = 0.01$	36.54	34.74
	$\gamma = 0.05$	97.85	97.51
	$\gamma = 0.10$	99.99	99.99
	$\gamma = 0.50$	100.00	100.00
$n = 1000$	$\gamma = 0.01$	69.74	69.83
	$\gamma = 0.05$	99.98	99.97
	$\gamma = 0.10$	100.00	100.00
	$\gamma = 0.50$	100.00	100.00

Table 2.44: Size of the test for threshold effect Without Drift, long run equation shocks are AR(1) process and the threshold parameter is known, and $l = 0$

	$\rho_z = 0.5$						$\rho_z = 0.9$					
ρ	-0.8	-0.5	-0.2	0.2	0.5	0.8	-0.8	-0.5	-0.2	0.2	0.5	0.8
$n = 200$	1.58	2.97	3.74	6.75	10.60	13.92	0.30	0.84	3.00	9.41	17.97	32.55
$n = 500$	1.73	2.36	3.95	6.87	9.65	14.49	0.15	0.86	2.59	9.20	18.78	33.72
$n = 1000$	1.20	2.41	3.81	6.87	9.99	15.11	0.03	0.66	2.44	9.46	18.13	35.55
$n = 2000$	1.41	2.66	3.90	6.51	10.03	15.60	0.01	0.43	2.40	9.09	18.97	36.75

Table 2.45: Size of the test for threshold effect Without Drift, long run equation shocks are AR(1) process and the threshold parameter is known, and $l = 4(n/100)^{1/4}$

	$\rho_z = 0.5$						$\rho_z = 0.9$					
ρ	-0.8	-0.5	-0.2	0.2	0.5	0.8	-0.8	-0.5	-0.2	0.2	0.5	0.8
$n = 200$	5.36	5.86	5.98	7.02	7.33	7.78	5.19	5.43	6.24	7.18	8.51	12.14
$n = 500$	4.89	5.29	5.66	5.89	6.10	6.85	4.25	5.03	5.43	5.99	7.51	10.55
$n = 1000$	4.88	4.66	5.04	5.64	5.83	6.37	4.03	4.32	5.22	6.01	6.92	10.55
$n = 2000$	4.46	5.24	4.93	5.44	5.85	6.16	3.80	4.44	4.86	5.81	6.58	9.83

Table 2.46: Size of the test for threshold effect Without Drift, long run equation shocks are AR(1) process and the threshold parameter is known and $l = 12(n/100)^{1/4}$

	$\rho_z = 0.5$						$\rho_z = 0.9$					
ρ	-0.8	-0.5	-0.2	0.2	0.5	0.8	-0.8	-0.5	-0.2	0.2	0.5	0.8
$n = 200$	6.69	7.09	7.63	7.97	7.88	7.66	6.51	6.93	7.53	7.77	8.89	9.90
$n = 500$	5.85	6.27	6.18	6.28	6.26	6.62	5.55	5.98	6.08	6.31	7.26	8.37
$n = 1000$	5.45	5.28	5.43	6.00	5.48	5.76	5.37	5.10	5.25	6.12	6.05	7.49
$n = 2000$	5.29	5.33	5.29	5.36	5.62	5.77	4.71	5.00	5.37	5.91	5.69	6.83

Table 2.47: Size of the test for threshold effect With state dependent Drift, long run equation shocks are AR(1) process and the threshold parameter is known, and $l = 0$

	$\rho_z = 0.5$						$\rho_z = 0.9$					
ρ	-0.8	-0.5	-0.2	0.2	0.5	0.8	-0.8	-0.5	-0.2	0.2	0.5	0.8
$n = 200$	1.45	2.85	3.51	6.91	10.28	13.56	0.18	0.85	2.99	9.07	18.07	32.54
$n = 500$	1.32	2.34	4.00	6.70	10.13	14.69	0.10	0.76	2.68	9.29	18.49	35.03
$n = 1000$	1.26	2.33	3.79	6.51	10.69	14.79	0.08	0.62	2.53	8.77	18.74	35.73
$n = 2000$	1.36	2.40	3.86	7.00	10.13	14.93	0.08	0.6	2.48	8.89	19.49	36.26

Table 2.48: Size of the test for threshold effect With state dependent Drift, long run equation shocks are AR(1) process and the threshold parameter is known, and $l = 4(n/100)^{1/4}$

	$\rho_z = 0.5$						$\rho_z = 0.9$					
ρ	-0.8	-0.5	-0.2	0.2	0.5	0.8	-0.8	-0.5	-0.2	0.2	0.5	0.8
$n = 200$	5.44	5.87	6.08	7.25	7.17	7.53	5.23	6.27	6.53	6.56	7.44	11.85
$n = 500$	4.96	5.71	5.71	6.04	6.02	6.99	4.31	4.55	5.82	6.56	7.44	17.85
$n = 1000$	4.65	5.12	5.16	5.72	6.19	6.04	3.75	4.83	5.33	5.96	7.13	10.65
$n = 2000$	4.43	4.82	5.01	5.67	5.57	6.40	3.53	4.19	5.14	5.91	6.30	9.19

Table 2.49: Size of the test for threshold effect With state dependent Drift, long run equation shocks are AR(1) process and the threshold parameter is known, and $l = 12(n/100)^{1/4}$

	$\rho_z = 0.5$						$\rho_z = 0.9$					
ρ	-0.8	-0.5	-0.2	0.2	0.5	0.8	-0.8	-0.5	-0.2	0.2	0.5	0.8
$n = 200$	7.94	8.10	8.43	9.41	9.28	8.88	8.49	9.09	9.27	9.14	9.57	11.18
$n = 500$	6.04	6.48	6.53	6.83	6.63	7.21	6.58	6.90	7.19	6.68	8.02	8.84
$n = 1000$	5.62	5.84	6.22	6.10	5.99	6.25	5.59	5.95	6.15	6.12	6.65	7.37
$n = 2000$	5.76	5.03	5.37	5.77	6.09	5.49	4.98	5.00	5.57	5.68	5.91	6.77

Table 2.50: Size of the test for threshold effect With state dependent Drift, long run equation shocks are AR(1) process and the threshold parameter is Estimated.

	$\rho_z = 0.5$						$\rho_z = 0.9$					
ρ	-0.8	-0.5	-0.2	0.2	0.5	0.8	-0.8	-0.5	-0.2	0.2	0.5	0.8
$n = 200$	7.96	7.99	8.33	8.42	8.52	7.59	8.47	8.83	9.06	8.51	9.69	8.64
$n = 500$	6.28	6.31	6.80	6.28	6.75	6.67	6.24	6.36	6.57	6.90	7.69	8.04
$n = 1000$	5.98	5.54	6.01	5.65	6.04	6.54	5.36	5.92	6.03	6.36	6.34	7.44
$n = 2000$	5.34	5.55	5.21	5.58	5.44	5.57	5.10	5.39	6.12	5.97	6.23	6.83

Chapter 3

Quasi-Error Correction Model

3.1 Introduction

Single economic variables, observed as a time series, move freely in an aimless path and yet we may find some pairs of series moving closely, not too far from each other. Economic theory asses the existence of long-run equilibrium relationships between economic variables, and cointegration is a method to study empirically the forces which keep these variables moving together in the long-run, see Granger (1986) and Engle and Granger (1987). Cointegration has been used, for example, to study the relation between consumption and income, to show the link between prices and dividends through present values models, see Campbell and Shiller (1987). Also has been used to study the relations between the short and long term interest rate (Campbell and Shiller 1991, etc.). When the variables are cointegrated, Granger representation theorem assures the existence of an error correction representation which describes how variables respond to disturbances from the equilibrium. One can see the ECM as an attractor where the long-run equilibrium is maintained.

The development of the ECM has gone into many directions. In one hand we have the linear ECM where the adjustment mechanism is constant. In the other hand introduces the possibility of a threshold effect in the adjustment process, see Balke and Fomby (1997), Hansen and Seo (2002), Gonzalo and Pitrakis (2006). In the work of Granger (2001), Escribano and Mira (2002), Saikonnen (2005), introduces the nonlinear ECM and analyze its properties. In all these cases, they assume the existence of a single long-run equilibrium.

The objective of this study is to analyze the ECM representation theorem when the economic variables present multiple long-run equilibria driven by the business cycle. We study the simplest case where the long-run equilibrium equation presents a threshold effect, indicating the presence of multiple cointegration relations but considering a common adjustment mechanism. This work is very preliminary, and this case is a stepping stone to the most general case, which both, the cointegrating vectors and the adjustment present threshold effect.

Balke and Fomby (1997) are the first to introduce a threshold structure in the adjustment process of the ECM. It attracted much attention since it includes appealing features for economics like the different speed of the adjustment toward the equilibrium depending, for example, on how far the system is away from the long-run equilibrium. Also, it allows the possibility of shutting off the adjustment mechanism over specific regimes, for instance, by the law of one piece the price of an asset traded simultaneously in two different markets must be the same. When the price is different, a profit can be made by buying in the cheapest market and simultaneously sell it in the dearest market. When the price difference is small, market participants may not be interested in taking this arbitrage opportunity due to factors like transaction costs, interest rates, and other barriers. In other words, arbitration occurs when the price deviation is substantial so that the profit is higher than the trading costs.

In Section 3.2 we introduce the single equation quasi-error correction model (QECM) in the presence of multiple equilibria and discuss if the model is balanced, consistency and asymptotic normality of the LS estimate of the adjustment parameter. In Section 3.3, we construct confidence intervals for the adjustment parameter and show that have the correct coverage. In Section 3.4, we introduce an application of the QECM to U.S. interest rate with different maturities. In section 3.5 concludes.

3.2 The Quasi-Error Correction Model Representation

Consider the following triangular representation of the cointegration relation with a threshold effect

$$\begin{aligned} y_t &= \beta_1 I(z_{t-1} \leq r) x_t + \beta_2 I(z_{t-1} > r) x_t + e_t \\ x_t &= x_{t-1} + \varepsilon_t \end{aligned} \tag{3.1}$$

where $\{z_t\}$ is the threshold variable variable, r is the threshold value which determines the different regimes, expansions and recessions, high volatility and low volatility, and $I(.)$ is the indicator function. System (3.1) captures the existence of two cointegrating vectors driven by the threshold variable z_t, that is $\begin{pmatrix} 1 & \beta_1 \end{pmatrix}'$ when $z_{t-1} \leq r$, and $\begin{pmatrix} 1 & \beta_2 \end{pmatrix}'$ when $z_{t-1} > r$ which represent the different long run equilibrium relationships between y_t and x_t. Though all the paper, we are working under the following set of assumptions

Assumptions

- **A.1:** The sequence $\{\varepsilon_t, z_t, v_t\}$ is strictly stationary and ergodic and strong mixing with mixing coefficients α_n satisfying $\sum_{n=1}^{\infty} \alpha_n^{\frac{1}{2}-\frac{1}{r}}$ for some $r > 2$. The threshold variable z_t has a continuous and strictly increasing distribution function $F(.)$.

- **A.2:** $E(\varepsilon_t) = 0$, $E|\varepsilon_t|^4 < \infty$ and ε_t is independent of $\mathcal{F}_{t-1}^{z\varepsilon}$, where $\mathcal{F}_t^{z\varepsilon} = \sigma(z_{t-j}, \varepsilon_{t-j} : j \geq 0)$.

- **A.3:** e_t follows an AR(1) process with autorregresive coefficient $|\rho| < 1$, $e_t = \rho e_{t-1} + v_t$, and v_t satisfy the following conditions, $E(v_t) = 0$, $E|v_t|^4 < \infty$ and independent with respect z_{t-j} for all j, also independent of v_{t-j} for $j = 1, 2, 3...$

- **A.4:** Assume that v_t is strictly exogenous with respect to ε_t.

Assumptions A.1 and A.2 are standard assumptions in the threshold literature in which the regressors has a unit root. A.1 restrict the threshold variable to be stationary ruling out the possibility of $\{z_t\}$ be an $I(1)$ process but general enough to accommodate a wide variety of processes. A.2 is required to obtain the weak convergence of the partial sum $\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varepsilon_t I(z_{t-1} \leq r)$, which is needed to show the weak convergence of the LS estimate of the parameter of interest, $(\rho - 1)$. A.3 assumes that the shocks in the long run equation is linear and follows an AR(1) process, this has an impact on the structure of the QECM representation in which the adjustment speed is the same for all the different regimes. This assumption is very restrictive but is needed to understand the most general case in which both the cointegration relation and the adjustment mechanism has a threshold effect.

A critical aspect of the linear cointegration, i.e., $\beta_1 = \beta_2$ is that taking x_t to be an $I(1)$ process, implies that y_t also is an $I(1)$ process such that after taking first difference in both sides $\Delta y_t = \beta \Delta x_t + \Delta e_t$ the equation is balanced, that is the right-hand side and the left-hand side of the equation have the same order of integration.

When we introduce a threshold effect in the long-run equilibrium relationship linking y_t and x_t, assuming that x_t is difference stationary does not imply that y_t also is difference stationary. It is easy to show that if x_t is an $I(1)$ process then y_t is nonstationary but the nonstationary of y_t cannot be removed by taking first difference, more formally, we can see that by differencing (3.1)

$$\begin{aligned} \Delta y_t &= \tilde{\rho} e_{t-1} + \beta \Delta x_t + \gamma(x_t I(z_{t-1} \leq r) - x_{t-1} I(z_{t-2} \leq r)) + v_t \\ \Delta x_t &= \varepsilon_t \end{aligned} \tag{3.2}$$

, where $\tilde{\rho} = (\rho - 1)$ and $\gamma = (\beta_1 - \beta_2)$. The presence of the term $(x_t I(z_{t-1} \leq r) - x_{t-1} I(z_{t-2} \leq r))$ in the right hand side on the first equation in (3.2) hinder the possibility of the usual ECM representation. In summary, introducing threshold effect in the cointegration relation, does not admit an ECM representation which is balanced, in the sense where the right- and left-hand side are $I(0)$ process. To move away from the thigh constrains that arise from the concept of integration, we use the definition of summability proposed by Berenger-Rico and Gonzalo (2014a) which is the natural extension of the concept of integration. Summability can characterizes the stochastic properties for both, non-linear process and also linear process. Any integrated process $I(d)$ for some $d \geq 0$ also is $S(d)$. Using the results from Caner and Hansen (2001) we can shown that $x_t I(z_{t-1} \leq r)$, y_t are $S(1)$ process and how the first equation in (3.2) can be balanced in terms of summability.

Lemma 1. *Under assumption A.1 and A.2, the process $\{x_t I(z_{t-1} \leq r)\}$ is $S(1)$ and its first difference, $(x_t I(z_{t-1} \leq r) - x_{t-1} I(z_{t-2} \leq r))$ is $S(0)$.*

Lemma 1 show that $\{x_t I(z_{t-1} \leq r)\}$ is $S(1)$ and taking first differences reduces the order of summability to S(0). Viewing the properties of $[x_t I(z_{t-1} \leq r) - x_{t-1} I(z_{t-2} \leq r)]$ in terms of the variance, if z_t is an i.i.d process, the $Var([x_t I(z_{t-1} \leq r) - x_{t-1} I(z_{t-2} \leq r)]) = 2F(r)(1 - F(r))\sigma_\varepsilon(t-1) + F(r)\sigma_\varepsilon$ has a similar behaviour to the variance of a random walk (RW), but the RW is S(1) whereas $[x_t I(z_{t-1} \leq r) - x_{t-1} I(z_{t-2} \leq r)]$ is S(0). With the result in Lemma 1, in terms of summability, the QECM representation in (3.2) is balanced.

Proposition 1. *If the DGP is (3.1) and the regression model is $\Delta y_t = (\rho - 1)e_{t-1} + H_t$ where $H_t = \beta \Delta x_t + \gamma(x_t I(z_{t-1} \leq r) - x_{t-1} I(z_{t-2} \leq r)) + v_t$ the LS estimate of $\tilde{\rho}$ is not consistent*

$$(\hat{\tilde{\rho}} - \tilde{\rho}) = O_p(1) \tag{3.3}$$

Proposition 1 states that in the presence of multiple cointegration relation, the LS estimate of the short term dynamics is inconsistent when the switching effect $(x_t I(z_{t-1} \leq r) - x_{t-1} I(z_{t-2} \leq r))$ is not included in regression model (3.2).

Proposition 2. *If the DGP is (3.1) and the regression model is (3.2) the LS estimate of $\tilde{\rho}$ is consistent*

$$(\hat{\tilde{\rho}} - \tilde{\rho}) = o_p(1) \tag{3.4}$$

Proposition 2 shows that including the switching effect in the estimation, the LS estimate of the adjustment mechanism is consistent. Finally, we show the asymptotic normality of the LS estimate of $\tilde{\rho}$

Proposition 3. *Under assumptions A.1-A.4, the LS estimate of $\tilde{\rho}$*

$$\sqrt{n}(\hat{\tilde{\rho}} - \tilde{\rho}) \Rightarrow N(0, \sigma_v^2) \tag{3.5}$$

where $\sigma_v^2 = E(v_i^2)$.

The result in Proposition 3 tells that we can perform inference about the parameter of interest, $\tilde{\rho}$, and construct confidence intervals and testing.

Proposition 4. *Under assumptions A.1-A.4, Proposition 2 and 3 holds using the LS estimated residuals $\{\hat{e}_t\}$ obtained in a first stage using the regression model in (3.1).*

Proposition 4 shows that using estimated residuals obtained in a first step does not affect asymptotically the estimation of $\tilde{\rho}$, and this is possible due to the T-consistency of the LS estimate of β_1 and β_2, which can be used as if it were known.

3.3 Simulations

In this section, we look at the performance of the confidence intervals for $\tilde{\rho}$. In these simulations, we assume that the threshold parameter is known $r = r_0$. The data generating process (DGP) for this experiment is the following threshold cointegration process:

$$\begin{aligned} y_t &= \beta_1 I(z_{t-1} \leq r)x_t + \beta_2 I(z_{t-1} > r)x_t + e_t \\ x_t &= x_{t-1} + \varepsilon_t \end{aligned} \tag{3.6}$$

In the simulation we set up the parameters $\beta_1 = 2$, $\beta_2 = 1$. We assume that the threshold variable, z_t follow an AR(1) process, $z_t = \rho_z z_{t-1} + \eta_t$ with $|\rho_z| < 1$. Also we generate e_t as an AR(1) process, $e_t = \rho_e e_{t-1} + v_t$. The shocks $\{v_t, \varepsilon_t, \eta_t\}$ are generated as a multivariate normal with the following variance and covariance matrix:

$$\Omega = \begin{pmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & \sigma_{\varepsilon,\eta} \\ 0 & \sigma_{\varepsilon,\eta} & \sigma_\eta^2 \end{pmatrix} \tag{3.7}$$

We set up $\sigma_v^2 = \sigma_\varepsilon^2 = \sigma_\eta^2 = 1$ and allow $\sigma_{\varepsilon,\eta} = \{0.5, 0.9\}$. We consider different levels of persistence for the threshold variable $\rho_z = \{0.5, 0.9\}$ and for the shocks in the long run equation $\rho_e = \{0.1, 0.5, 0.9\}$. We perform the simulations with 10000 repetitions, with different sample sizes $n = \{200, 500, 1000\}$ observations.

Table 3.1 shows the coverage of the confidence interval for $\tilde{\rho}$ in the case where β_1 and β_2 is known. We can see that the coverage is correct, since the empirical coverage approaches the nominal coverage of 95%, for different values of ρ_e, and different persistence levels of the threshold variable.

Table 3.2 shows a similar result in the case where β_1 and β_2 are estimated using LS in a first step and using the estimated residual $\{\hat{e}_t\}$ as a regressor to estimate the parameter of interest $\tilde{\rho}$.

Table 3.1: Coverage of the CI for $\tilde{\rho}$, here it assumes that β_1, β_2 is known.

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\rho_e = 0.1$	$\rho_e = 0.5$	$\rho_e = 0.9$	$\rho_z = 0.1$	$\rho_z = 0.5$	$\rho_e = 0.9$
$n = 200$	$\sigma_{\varepsilon,\eta} = 0.5$	95.04	94.97	94.49	94.46	94.43	95.07
	$\sigma_{\varepsilon,\eta} = 0.9$	94.43	94.79	94.98	95.12	94.88	94.82
$n = 500$	$\sigma_{\varepsilon,\eta} = 0.5$	94.50	94.87	94.93	94.49	94.78	95.04
	$\sigma_{\varepsilon,\eta} = 0.9$	94.36	94.98	95.33	94.83	94.47	95.28
$n = 1000$	$\sigma_{\varepsilon,\eta} = 0.5$	94.62	94.77	95.07	94.93	94.97	94.95
	$\sigma_{\varepsilon,\eta} = 0.9$	95.12	95.31	94.93	95.20	94.84	95.45

Table 3.2: Coverage of the CI for $\tilde{\rho}$, with β_1, β_2 estimated in a first stage.

		$\rho_z = 0.5$			$\rho_z = 0.9$		
		$\rho_e = 0.1$	$\rho_e = 0.5$	$\rho_e = 0.9$	$\rho_z = 0.1$	$\rho_z = 0.5$	$\rho_e = 0.9$
$n = 200$	$\sigma_{\varepsilon,\eta} = 0.5$	94.28	94.44	92.62	94.50	94.14	91.74
	$\sigma_{\varepsilon,\eta} = 0.9$	94.59	94.53	92.19	94.45	94.29	91.93
$n = 500$	$\sigma_{\varepsilon,\eta} = 0.5$	94.55	94.70	94.16	94.85	94.60	93.71
	$\sigma_{\varepsilon,\eta} = 0.9$	95.05	94.44	93.79	94.70	94.78	93.66
$n = 1000$	$\sigma_{\varepsilon,\eta} = 0.5$	94.71	94.62	94.55	94.83	94.84	94.26
	$\sigma_{\varepsilon,\eta} = 0.9$	95.11	95.16	94.52	94 75	94.52	94.93

3.4 Empirical Application. U.S. Interest Rate

Application of the error correction models for U.S. interest rates of instruments with different maturities have been studied extensively under different specifications, for example in the work of Bradley and Lumpkin (1992), Mehra (1994) among many others, considers the linear ECM. Siklos and Enders (1998) study the asymmetric behavior of the error correction between interest rates with different maturities.

In this application, we use the QECM to study the adjustment mechanism where the equilibrium between U.S. short term interest rate and the long term interest rate is maintained. Following the result stated above, we assume that for all the different equilibria has the same adjustment speed. The estimated model is

$$\Delta y_t = \tilde{\rho} e_{t-1} + \beta_2 \Delta x_t + \gamma [I(z_{t-1} \leq r) x_t - I(z_{t-2} \leq r) x_{t-1}] + v_t \tag{3.8}$$

We use the Effective federal fund rates as short term interest rate and the ten years government bond yield as the long term interest rate. These data have monthly frequency constructed by averaging daily observation for the sample 1960:1 to 2019:3 retrieved from FRED (at the Federal Reserve Bank of St. Louis). We consider the annual increment of the production index as a threshold variable since it is a crucial indicator of the economic health, expansions, and recession.

Table 3.3: QECM estimation for the U.S. interest rates

$\tilde{\rho}$	β_2	γ
-0.0584	0.4616	0.0649

We have to take this result carefully since it suffers two major problems. In one hand, we estimate the threshold parameter as in Chan (1993) and Canner and Hansen (2001), but the result stated above is based on the assumption where the threshold value is known. On the other hand, the adjustment may be different for each regime, but we restrict to be common for all the states.

3.5 Conclusion

Cointegration is a method to asses empirically the existence of a long-run equilibrium relationship between economic variables, and the error correction models is a process where this long-run equilibrium is maintained.

The QECM representation have a different structure to the ECM in the linear case since it contains an extra term which represents the switching between the different regimes. This representation is balanced using the concept of summability.

In this study, we present the QECM under the assumption where the error correction term is common for all the different regimes. Also, we present the consistency and asymptotic normality of the LS estimate of the adjustment process. We finalize the paper with an empirical application with U.S. interest rates of instruments with different maturities.

Appendix

3.A Proofs

Proof Lemma 1

The summablility order of $x_t I(z_{t-1} \leq r)$ can be found in Berenger-Rico and Gonzalo (2014) and is based in Theorem 3 in Caner and Hensen (2001). Without loss of generality assume that $x_0 = 0$, the second part of Lemma 1

$$\begin{aligned}
&|\frac{1}{n^{1/2+\delta}} \sum_{i=1}^{n} [x_t I(z_{t-1} \leq r) - x_{t-1} I(z_{t-2} \leq r)]| \\
&= |\frac{1}{n^{1/2+\delta}} (x_n I(z_{n-1} \leq r) - x_0 I(z_{-1} \leq r))| = |\frac{1}{n^{1/2+\delta}} x_n I(z_{n-1} \leq r)| \\
&\leq |\frac{1}{n^{1/2+\delta}} x_n| = O_p(1) \text{ for } \delta = 0.
\end{aligned}$$

showing the desired result.

Lemma 2. *Under assumptions A.1 and A.3*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{[ns]} e_{i-1}(I(z_{i-1} \leq r) - I(z_{i-2} \leq r)) \Rightarrow G_{eI}(s, \lambda) \tag{3.9}$$

Proof of Lemma 2

To show this result we use the work of Mervelede, Peligrad and Utev (2006) where we have to check that:

$$\sum_{i=1}^{\infty} \frac{1}{\sqrt{i}} ||E(e_i[I(z_i \leq r) - I(z_{i-1} \leq r)]|\mathcal{F}_0)||_2 < \infty \tag{3.10}$$

First note that by A.3

$$\begin{aligned}
&E(e_i[I(z_i \leq r) - I(z_{i-1} \leq r)]|\mathcal{F}_0) \\
&= \sum_{j=0}^{i-1} \rho^j E(v_{i-j}[I(z_j \leq r) - I(z_{j-1} \leq r)]|\mathcal{F}_0) + \sum_{j=i}^{\infty} \rho^j v_{i-j} E([I(z_j \leq r) - I(z_{j-1} \leq r)]|\mathcal{F}_0) \\
&= \sum_{j=i}^{\infty} \rho^j v_{i-j} E([I(z_j \leq r) - I(z_{j-1} \leq r)]|\mathcal{F}_0)
\end{aligned}$$

The last equality came from the independence between v_t and z_j for all t and j and independence between v_t and $\mathcal{F}_0$.

Since e_t is an AR(1) process

$$= e_0\rho^i(F_{i,0} - F_{i-1,0})$$

where $F_{i,0} = E(I(z_i \leq r)|\mathcal{F}_0) \in (0,1)$. Then it is easy to see that

$$\begin{aligned} &||E(e_i[I(z_i \leq r) - I(z_{i-1} \leq r)]|\mathcal{F}_0)||_2 = ||e_0\rho^i(F_{i,0} - F_{i-1,0})||_2 \\ &= E\Big((e_0\rho^i)^2(F_{i,0} - F_{i-1,0})^2\Big)^{1/2} \leq E\Big((e_0\rho^i)^2\Big)^{1/2} = ||e_0\rho^i||_2 \end{aligned} \tag{3.11}$$

With the result above we can see that

$$\begin{aligned} &\sum_{i=1}^{\infty}\frac{1}{\sqrt{i}}||E(e_i[I(z_i \leq r) - I(z_{i-1} \leq r)]|\mathcal{F}_0)||_2 \leq \sum_{i=1}^{\infty}\frac{1}{\sqrt{i}}||e_0\rho^i||_2 \\ &\leq ||e_0||_2\sum_{i=1}^{\infty}\rho^i < \infty \end{aligned}$$

Lemma 3. *Under assumptions A.1, A.2, A.3 and A.4*

$$\frac{1}{n}\sum_{i=1}^{n}x_ie_i(I(z_i \leq r) - I(z_{i-1} \leq r)) \Rightarrow \int_0^1 B_x(s)dG_{eI}(s,\lambda) \tag{3.12}$$

Proof of Lemma 3

The proof came from using Lemma 2 and the result from Hansen (1992) with assumption A.4.

Lemma 4. *Under assumptions A.1 and A.2*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{[ns]}\varepsilon_iI(z_{i-1} \leq r)I(z_{i-2} \leq r) \Rightarrow G_{\varepsilon I}(s,\lambda) \tag{3.13}$$

Proof of Lemma 4

For all r, $\{\varepsilon_iI(z_{i-1} \leq r)I(z_{i-2} \leq r), \mathcal{F}_{i-1}^{z\varepsilon}\}$ is a strictly stationary ergodic martingale difference sequence with variance $\sigma_\varepsilon^2E(I(z_{i-1} \leq r)I(z_{i-2} \leq r))$. By the central limit theorem for martingale difference sequence $\frac{1}{\sqrt{n}}\sum_{i=1}^{[ns]}\varepsilon_i I(z_{i-1}\leq r)I(z_{i-2}\leq r)\rightarrow^d N(0,\omega_G^2(s,\lambda))$, where $\omega_G^2(s,\lambda)=s\sigma_\varepsilon^2 E(I(z_{i-1}\leq r)I(z_{i-2}\leq r))$. It is straightforward to see that the covariance kernel is $\omega_G^2(s_1,s_2,\lambda_1,\lambda_2)=(s_1\wedge s_2)\sigma_\varepsilon^2 E[I(z_{i-1}\leq r_1)I(z_{i-1}\leq r_2)I(z_{i-2}\leq r_1)I(z_{i-2}\leq r_2)]$. Combined with the Crame-Wold device the fidi convergence follows. Given our assumptions on ε_t and z_t the stochastic continuity of $\frac{1}{\sqrt{n}}\sum_{i=1}^{[ns]}\varepsilon_i I(z_{i-1}\leq r)I(z_{i-2}\leq r)$ follows directly from Theorem 1 of Caner and Hansen (2001).

Lemma 5. *Under assumptions A.1 and A.2*

$$\frac{1}{n}\sum_{i=1}^{n}x_{i-1}\varepsilon_i(I(z_{i-1}\leq r)I(z_{i-2}\leq r))\Rightarrow\int_0^1 B_x(s)dG_{\varepsilon I}(s,\lambda)\tag{3.14}$$

Proof of Lemma 5

For a fixed λ, Lemma 4 and Theorem 2.2 in Kurtz and Potter (1991) $\frac{1}{n}\sum_{i=1}^{n}x_{i-1}\varepsilon_i(I(z_{i-1}\leq r)I(z_{i-2}\leq r))\Rightarrow\int_0^1 B_x(s)dG_{\varepsilon I}(s,\lambda)$, furthermore this result can be extended uniformly for $\lambda\in[0,1]$, see Caner and Hansen (2001) Theorem 2.

Lemma 6. *Under assumptions A.1 and A.2*

$$\frac{1}{n}\sum_{i=1}^{n}x_{i-1}\varepsilon_i(I(z_{i-1}\leq r)-I(z_{i-2}\leq r))=O_p(1)\tag{3.15}$$

Proof of Lemma 6

It is the same as in Lemma 4 and 5.

Define the following matrices

$$\Delta Y=\begin{pmatrix}\Delta y_2\\ \vdots\\ \Delta y_n\end{pmatrix},\quad V=\begin{pmatrix}v_2\\ \vdots\\ v_n\end{pmatrix},\quad e=\begin{pmatrix}e_1\\ \vdots\\ e_{n-1}\end{pmatrix},\quad \Gamma=\begin{pmatrix}\beta_2\\ \gamma\end{pmatrix}$$

$$X_z=\begin{pmatrix}\Delta x_2 & I(z_1\leq r)x_2-I(z_0\leq r)x_1\\ \vdots & \vdots\\ \Delta x_n & I(z_{n-1}\leq r)x_n-I(z_{n-2}\leq r)x_{n-2}\end{pmatrix}$$

Such that model (3.2) can be written as follows:

$$\Delta Y = e(\rho - 1) + X_z\Gamma + V \tag{3.16}$$

Construct the following matrix

$$M_x = I_{n-1} - X_z(X_z'X_z)^{-1}X_z' \tag{3.17}$$

where I_{n-1} is the (n-1)x(n-1) identity matrix.

Then equivalently we can write

$$M_x\Delta Y = M_x e\tilde{\rho} + M_x V \tag{3.18}$$

where $\tilde{\rho} = (1 - \rho)$. Then the LS estimate of $\tilde{\rho}$ is

$$\hat{\tilde{\rho}} = (e'M_x e)^{-1}(eM_x\Delta Y) = \tilde{\rho} + (e'M_x e)^{-1}(eM_x v) \tag{3.19}$$

Proof of Proposition 1

We can write the LS estimate of $\tilde{\rho}$ as

$$(\hat{\tilde{\rho}} - \tilde{\rho}) = \big(\frac{1}{n}\sum_{i=2}^{n} e_{t-1}^2\big)^{-1}\big(\frac{1}{n}\sum_{j=2}^{n} e_{t-1}H_t\big) \tag{3.20}$$

From A.3 by the LLN and Slutsky's theorem $\big(\frac{1}{n}\sum_{i=2}^{n} e_{t-1}^2\big)^{-1} \to_p E(e_{t-1}^2)^{-1}$. Then we can write

$$\begin{aligned}\frac{1}{n}\sum_{j=2}^{n} e_{t-1}H_t = \frac{1}{n}\sum_{j=2}^{n} e_{t-1}v_t + \beta_1\frac{1}{n}\sum_{j=2}^{n} e_{t-1}\varepsilon_t \\ + \gamma\frac{1}{n}\sum_{j=2}^{n} e_{t-1}\varepsilon_t I(z_{t-1} \le r) + \gamma\frac{1}{n}\sum_{j=2}^{n} e_{t-1}x_{t-1}[I(z_{t-1} \le r) - I(z_{t-2} \le r)]\end{aligned}$$

From A.3 we can see that $\frac{1}{n}\sum_{j=2}^{n} e_{t-1}v_t \to_p E(e_{t-1}v_t) = 0$. By the exogeneity assumption $\beta_1\frac{1}{n}\sum_{j=2}^{n} e_{t-1}\varepsilon_t \to_p 0$, and $\gamma\frac{1}{n}\sum_{j=2}^{n} e_{t-1}\varepsilon_t I(z_{t-1} \le r) \to_p 0$.

Finally, from Lemma 2

$$\gamma\frac{1}{n}\sum_{j=2}^{n} e_{t-1}x_{t-1}[I(z_{t-1} \le r) - I(z_{t-2} \le r)] = O_p(1) \tag{3.21}$$

Proof of Proposition 2

Note that:

$$(\hat{\tilde{\rho}} - \tilde{\rho}) = \underbrace{\Big(\frac{e'M_x e}{n}\Big)^{-1}}_{A} \underbrace{\Big(\frac{eM_x v}{n}\Big)}_{B} \tag{3.22}$$

Lets start with the term A and define the following matrix $D = diag\{n^{-1/2}, n^{-1}\}$:

$$\Big(\frac{e'M_x e}{n}\Big) = \frac{e'e}{n} - \Big(\frac{e'X_z D}{n^{1/2}}\Big)(DX_z'X_z D)^{-1}\Big(\frac{DX_z'e}{n^{1/2}}\Big)$$

It is easy to see that from A.3 and the LLN:

$$\frac{e'e}{n} = \frac{\sum_{j=2}^n e_{j-1}^2}{n} \rightarrow_p E(e_{j-1}^2) \tag{3.23}$$

Also

$$\begin{aligned}\Big(\frac{e'X_z D}{n^{1/2}}\Big) &= \begin{pmatrix} \frac{1}{n}\sum_{i=2}^n \varepsilon_i e_{i-1} \\ \frac{1}{n^{3/2}}\sum_{i=2}^n e_{i-1}I(z_{i-1} \leq r)\varepsilon_i + \frac{1}{n^{1/2}}\Big(\frac{1}{n}\sum_{j=2}^n x_{i-1}e_{i-1}[I(z_{i-1} \leq r) - I(z_{i-2} \leq r)]\Big)\end{pmatrix}' \\ &= \begin{pmatrix} E(\varepsilon_i e_{i-1}) + o_p(1) \\ \frac{1}{n^{1/2}}\Big(E(e_{i-1}I(z_{i-1} \leq r)\varepsilon_i) + o_p(1)\Big) + \frac{1}{n^{1/2}}\Big(O_p(1)\Big)\end{pmatrix} \rightarrow 0\end{aligned}$$

The second equality came from the result in Lemma 3, and from assumption A.2, A.3 and A.4 $E(\varepsilon_i e_{i-1}) = 0$ and $E(e_{i-1}I(z_{i-1} \leq r)\varepsilon_i) = 0$. Now lets see

$$(DX_z'X_z D) = \begin{pmatrix} \frac{1}{n}\sum_{i=2}^n \varepsilon_i^2 & \frac{1}{n^{3/2}}\sum_{i=2}^n \varepsilon_i(I(z_{i-1} \leq r)x_i - I(z_{i-2} \leq r)x_{i-1}) \\ - & \frac{1}{n^2}\sum_{i=2}^n (I(z_{i-1} \leq r)x_i - I(z_{i-2} \leq r)x_{i-1})^2 \end{pmatrix} \tag{3.24}$$

Note that the first element of the matrix:

$$\frac{1}{n}\sum_{i=2}^n \varepsilon_i^2 \rightarrow_p E(\varepsilon_i^2) \tag{3.25}$$

$$\begin{aligned}\frac{1}{n^{3/2}}\sum_{i=2}^n \varepsilon_i(I(z_{i-1} \leq r)x_i - I(z_{i-2} \leq r)x_{i-1}) &= \frac{1}{n^{1/2}}\Big(\frac{1}{n}\sum_{i=2}^n \varepsilon_i^2 I(z_{i-1} \leq r)\Big) \\ &+ \frac{1}{n^{1/2}}\Big(\frac{1}{n}\sum_{i=2}^n x_{i-1}\varepsilon_i(I(z_{i-1} \leq r) - I(z_{i-2} \leq r))\Big) \rightarrow 0\end{aligned}$$

The last result came from Lemma 6 and assumption A.2. Finally we see the convergence of

$$\frac{1}{n^2}\sum_{i=2}^{n}(I(z_{i-1}\leq r)x_i - I(z_{i-2}\leq r)x_{i-1})^2 = \frac{1}{n^2}\sum_{i=2}^{n}x_i^2 I(z_{i-1}\leq r) + \frac{1}{n^2}\sum_{i=2}^{n}x_{i-1}^2 I(z_{i-2}\leq r)$$
$$-2\frac{1}{n^2}\sum_{i=2}^{n}x_i x_{i-1}I(z_{i-1}\leq r)I(z_{i-2}\leq r) = 2\lambda\int_0^1 B_x^2(s)ds + o_p(1) - 2\frac{1}{n^2}\sum_{i=2}^{n}x_{i-1}^2 I(z_{i-1}\leq r)I(z_{i-2}\leq r)$$
$$-\frac{1}{n}\underbrace{\Big(\frac{1}{n}\sum_{i=2}^{n}x_{i-1}\varepsilon_i I(z_{i-1}\leq r)I(z_{i-2}\leq r)\Big)}_{O_p(1) by Lemma 5} \Rightarrow 2\Big(\lambda - E(I(z_{i-1}\leq r)I(z_{i-2}\leq r))\Big)\int_0^1 B_x^2(s)ds$$

Then we can conclude that $(DX_z'X_zD) = O_p(1)$ and have an inverse.

Then we can conclude that $\left(\frac{e'M_xe}{n}\right)^{-1} \rightarrow_p E(e_i^2)^{-1}$. Now lets focus on B.

$$\frac{e'M_xV}{n} = \frac{e'V}{n} - \Big(\frac{e'X_zD}{n^{1/2}}\Big)\Big(DX_z'X_zD\Big)^{-1}\Big(\frac{DX_z'V}{n^{1/2}}\Big) \tag{3.26}$$

It is easy to see that

$$\frac{e'V}{n} = \frac{1}{n}\sum_{j=2}^{n}v_ie_{i-1} \rightarrow_p E(v_ie_{i-1}) = 0 \text{ by A.3} \tag{3.27}$$

Finally

$$\frac{DX_z'V}{n^{1/2}} = \begin{pmatrix} \frac{1}{n}\sum_{i=2}^{n}v_i\varepsilon_i \\ \frac{1}{n^{3/2}}\sum_{i=2}^{n}v_i(x_{i-1}I(q_{i-1}\leq r) - x_{i-2}I(q_{i-2}\leq r)) \end{pmatrix} = \begin{pmatrix} E(v_i\varepsilon_i) + o_p(1) \\ \frac{1}{n^{1/2}}O_p(1) \end{pmatrix} \tag{3.28}$$

Then we can see that $\frac{e'M_xV}{n} \rightarrow_p 0$ showing the consistency of the LS estimator of $\tilde{\rho}$

Proof of Proposition 3

As in the proof of Proposition 1, note that

$$\sqrt{n}(\hat{\tilde{\rho}} - \tilde{\rho}) = \underbrace{\left(\frac{e'M_xe}{n}\right)^{-1}}_{A}\underbrace{\left(\frac{e'M_xv}{\sqrt{n}}\right)}_{B} \tag{3.29}$$

From proposition 1 we have shown that $\left(\frac{e'M_xe}{n}\right)^{-1} \rightarrow_p E(e_i^2)^{-1}$. Now lets focus on B. As before define $D = diag\{n^{1/2}, n^{-1}\}$

then

$$\left(\frac{e'M_xv}{\sqrt{n}}\right)=\frac{e'v}{\sqrt{n}}-\left(e'X_zD\right)\left(DX_z'X_zD\right)^{-1}\left(\frac{DX_z'V}{\sqrt{n}}\right) \tag{3.30}$$

Again from proposition 1 we know that:

$$\left(DX_z'X_zD\right)^{-1}\Rightarrow\begin{pmatrix}E(\varepsilon_i^2) & 0\\ 0 & 2\Big(\lambda-E(I(z_{i-1}\leq r)I(z_{i-2}\leq r))\Big)\int_0^1 B_x^2(s)ds\end{pmatrix}^{-1} \tag{3.31}$$

and also

$$\left(\frac{DX_z'V}{\sqrt{n}}\right)\Rightarrow\begin{pmatrix}E(v_i\varepsilon_i)\\ 0\end{pmatrix}\equiv 0 \tag{3.32}$$

The last equivalence came from A.4. Lets see the convergence of

$$\left(e'X_zD\right)=\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=2}^n\varepsilon_ie_{i-1}\\ \frac{1}{n}\sum_{i=2}^n e_{i-1}I(z_{i-1}\leq r)\varepsilon_i+\frac{1}{n}\sum_{i=2}^n x_{i-1}e_{i-1}[I(z_{i-1}\leq r)-I(z_{i-2}\leq r)]\end{pmatrix} \tag{3.33}$$

Lets see the convergence of $\frac{1}{\sqrt{n}}\sum_{i=2}^n\varepsilon_ie_{i-1}$. From assumption A.2, ε_te_{t-1} is a martingale difference sequence w.r.t. $\mathcal{F}_{t-1}=\sigma\{\varepsilon_{t-j},v_{t-j},z_{t-j}: j\geq 1\}$. Since e_{t-1} is a function of $\{v_{t-1},v_{t-2},\ldots\}$ then is $\mathcal{F}_{t-1}$ measurable, such that

$$E(\varepsilon_te_{t-1}|\mathcal{F}_{t-1})=e_{t-1}E(\varepsilon_t|\mathcal{F}_{t-1})=e_{t-1}E(\varepsilon_t)=0 \tag{3.34}$$

The second equality came from the independence between ε_t and $\mathcal{F}_{t-1}$. Then from the ergodic stationary martingale differences CLT

$$\frac{1}{\sqrt{n}}\sum_{i=2}^n\varepsilon_ie_{i-1}\to_d N(0,\sigma_\varepsilon^2E(e_{t-1}^2)) \tag{3.35}$$

Finally from Lemma 3 it is easy to show that:

$$\begin{aligned}&\frac{1}{n}\sum_{i=2}^n e_{i-1}I(z_{i-1}\leq r)\varepsilon_i+\frac{1}{n}\sum_{i=2}^n x_{i-1}e_{i-1}[I(z_{i-1}\leq r)-I(z_{i-2}\leq r)]\\ &\Rightarrow E(e_{i-1}I(z_{i-1}\leq r)\varepsilon_{i-1})+\int_0^1 B_x(s)dG_{eI}(s,\lambda)\end{aligned} \tag{3.36}$$

Then we can see that $\left(e'X_zD\right)=O_p(1)$, concluding that

$$\left(e'X_zD\right)\left(DX_z'X_zD\right)^{-1}\left(\frac{DX_z'V}{\sqrt{n}}\right) \to 0 \tag{3.37}$$

Finally from the martingale difference CLT, it is easy to see that:

$$\frac{e'v}{\sqrt{n}} \Rightarrow N(0, E(e_{t-1}^2 v_t^2)) \tag{3.38}$$

Since v_t is independent of $\mathcal{F}_{t-1}$ it is easy to see that $E(e_{t-1}^2 E(v_t^2|\mathcal{F}_{t-1})) = \sigma_v^2 E(e_{t-1}^2)$.

Bibliography

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

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

Ashley, R., C. W. J. Granger, and R. Schmalensee (1980)."Advertising and Aggregate Consumption: An Analysis of Causality ". *Econometrica*, 48, 1149-1167.

Azariadis, C., and A.Drazen (1990). "Threshold externalities in economic development". *The Quarterly Journal of Economics*, 105(2), 501-526.

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.

Balke, N., and T. Fomby (1997). "Threshold cointegration", *International Economic Review.* 38, 627-645.

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

Benhabib, J., S.Schmitt-Grohé, and M.Uribe (2001). "Monetary Policy and Multiple equilibria". *Journal of Economic Theory*, 96, 40-69.

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

Berenger-Rico, V., and J.Gonzalo (2014a). "Summability of stochastic process- A generalisation of integration for non-linear processes". *Journal of Econometrics*, 178, 331-341.

Berenger-Rico, V., and J.Gonzalo (2014b). "Co-summability ". Working paper.

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

Blanchard, O., and J. Wolfers (2000). “The Role of Shocks and Institutions in the Rise of European Unemployment: the Aggregate Evidence ”. *The Economic Journal*, 110, C1-C33.

Bradley, M.G., and S.A. Lumpkin (1992). “The treasury yield curve as a cointegrated system”. *Journal of Financial and Quantitative Analysis*, 27, 449,463.

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.

Campbell, J.Y., and R.J. Shiller (1987). “Cointegration and tests of present value models”. *Journal of Political Economy*, 95(5), 1062-1088.

Campbell, J.Y., and R.J. Shiller (1991). “Yield spreads and interest rate movements: A bird’s eye view”. *Review of Economic Studies.* 58, 495-514.

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., and H. Tong (1985). “On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations ”. *Advances in Applied Probability*, 17, 666-678.

Chan, K.S., and 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.

Chan, N.H., and C.Z. Wei (1987). “Asymptotic Inference for Nearly Nonstationary AR(1) Processes”. *The annals of statistics*, 15, 1050-1063.

Chang, Y., J. Park, and P.C.B Phillips (2001). “Nonlinear econometric models with cointegrated and deterministically trending regressors”. *Econometrics Journal*, 4, 1-36.

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

Chen, H. (2015). “ Robust estimation and inference for threshold models with integrated regressors”. *Econometric Theory*, 31, 778-810.

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

Choi, I., and P.Saikkonen (2004). “Testing linearity in cointegrating smooth transition regression”. *Econometrics Journal*, 7, 341-365.

Choi, I., and P.Saikkonen (2010). “Test for nonlinear cointegration”. *Econometric Theory*, 26, 682-709.

Choi, S., and M.E. Wohar (1995). “The expectation theory of interest rates: cointegration and factor decomposition”. *International Journal of Forecasting*, 11, 253-262.

Clark, T.E., M. McCracken (2001). “Tests of equal forecast accuracy and encompassing for nested models”. *Journal of Econometrics*, 1051, 85-110.

Clark, T.E., M. McCracken (2006). “The Predictive Content of the Output Gap for Inflation: Resolving In-Sample and Out-of-Sample Evidence”. *Journal of Money, Credit and Banking*, 38, 1127-1148.

Clark, T.E., and 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.

Davidson, J. (2009). “When is a time series I(0)?”. The Methodology and Practice of Econometrics, a Festschrift for David F. Hendry. *Oxford University Press*

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.

Enders, W., and P.L. Siklos (2001). “Cointegration and threshold adjustment”. *Journal of Business & Economic Statistics*, 19, 166-176.

Engel, R., and C.Granger (1987). “ Co-integration and error correction: representation, estimation, and testing”. *Econometrica*, 55, 251-276.

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

Escribano, A., and S. Mira (2002). "Nonlinear error correction modeld". *Journal of Time Series Analysis*, 23.

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, and 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.

Froot, K.A. (1989). "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates". *The Journal of Finance*, 44, 283-305.

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.

Gonzalo, J., and J. Pitarakis (2006). "Threshold effect in cointegrating relationships". *Oxford Bulletin of Economics and Statistics*,68, 813-833.

Gonzalo, J., and J. Pitarakis (2006). "Threshold effect in multivariate error correction models". *Palgrave Handbook of Econometrics*, Chapter 15, 579-609.

Gonzalo, J., and J. Pitarakis (2012). "Regime-specific predictability in predictive regression". *Journal of business & economic statistics*, 30:2,229-241.

Gonzalo, J., and J. Pitarakis (2017). " Inferring the predictability induced by a persistent regressor in a predictive threshold model". *Journal of Business & Economic Statistics*, 35, 202-2017.

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., and C.Y. Robert (2006). "Stochastic unit root models". *Economic theory*, 22, 1052-1090.

Granger, C.W.J. (1966). "The Typical Spectral Shape of an Economic Variable". *Econometrica*, 34, 150-161.

Granger, C.W.J., 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. (1981). "Some properties of time series data and their use in econometric model specification". *Journal of Econometrics*, 16, 121-130.

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

Granger, C.W.J. (1986). "Development in the study of cointegrated economic variables". *Oxford Bulletin of Economics and Statistics*, 48, 213-228.

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

Granger, C.W.J. (2001). "Overview of nonlinear macroeconometric empirical models". *Macroeconomic Dynamics*, 5, 466–81.

Hall, A., H. Anderson, and C.W.J. Granger (1992). "A Cointegration Analysis of Treasury Bill Yields ". *The Review of Economics and Statistics*, 74, 116-126.

Hansen, B.E. (1992). " Convergence to stochastic integrals for dependent heterogeneous processes". *Econometric Theory*, 8, 489-500.

Hansen, B. E. (2000). "Sample Splitting and Threshold Estimation". *Econometrica*, 68, 575-603.

Hansen, B.E., and B. Seo (2002). "Testing for two-regime threshold cointegration in vector error correction models". *Journal of Econometrics*, 110, 293-318.

Hansen, P.R. (2003). "Structural change in the cointegrated vector autoregressive model". *Journal of Econometrics*, 114, 261-295.

Johansen, S. (1996), *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford, U.K.: Oxford University Press

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

Kapetanios, G., Y. Shin, and A. Snell (2006). “Testing for cointegration in nonlinear smooth transition error correction models”. *Econometric Theory*, 22, 279-303.

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

Kurtz, T. G., and P. Protter (1991) : “Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations”. *The Annals of Probability*, 19, 1035-1070.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992). “ Testing the null hypothesis of stationarity against the alternative of a unit root”. *Journal of Econometrics*, 54, 159-178.

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

Mehra, Y.P. (1994). “An error-correction model of the long-term bond rate”. *Economic Quarterly*, 80, 49-68.

Marsh, I.W.(2000). “High-frequency Markov switching models in the foreign exchange market”. *Journal of Forecasting*, 19, 123-134.

Merlévede, F., M. Peligrad, and S. Utev (2006). “ Recent advances in invariance principles for stationary sequences”. *Probability Survey*, 3, 1-36.

Meyn, S.P., and 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.

Müller, U.K. (2008). “The impossibility of consistent discrimination between I(0) and I(1) processes”. *Econometric Theory*, 24, 616–630.

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.

Park, J., and P.C.B. Phillips (2001). “Nonlinear Regressions with Integrated Time Series”. *Econometrica*, 69, 117-161.

Peligrad, M., and S. Utev (2005). “A new maximal inequality and invariance principle for stationary sequences”. *The Annals of Probability*, 33, 798-815.

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

Petruccelli, J.D. (1992). “ On the approximation of time series by threshold autoregressive models”. *Sankhya: The Indian Journal of Statistics*, 54, 106-113.

Phillips, P.C.B. (1986). “Understanding spurious regressions in econometrics”. *Journal of Econometrics*, 33, 311-340.

Phillips, P.C.B. (1987). “Time series regression with a unit root”. *Econometrica*, 55, 277-301.

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

Phillips, P.C.B., and P. Perron (1988). "Testing for a Unit Root in Time Series Regression." *Biometrika*, 75, 335-46.

Phillips, P.C.B., and S. Ouliaris (1990). “ Asymptotic properties of residual based test for cointegration”. *Econometrica*, 58, 165-193.

Phillips, P.C.B., and V. Solo (1992). “ Asymptotics for linear processes”. *The Annals of Statistics*, 20, 971-1001.

Phillips, P.C.B (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.

Pourahmadi, M. (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., and 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.

Said E.S., and D.A. Dickey (1984). "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order". Biometrika, 71, 599-607.

Saikkonen, P. (1991). “Asymptotically efficient estimation of cointegration regressions”. *Econometric Theory*, 7, 1-21.

Saikkonen, P., and I. Choi (2004). “Cointegrating smooth transition regression”. *Econometric Theory*, 20, 301-340.

Saikkonen, P. (2005). “ Stability results for nonlinear error correction models”. *Journal of Econometrics*, 127, 69-81.

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

Samuelson, P. (1973). “Proof that properly discounted present values of assets vibrate randomly”. *Bell Journal of Economics and Management Science*, 4, 369-374.

Sargent, T. J., D. Fand, and S. Goldfeld (1973). “Rational expectations, the real rate of interest, and the natural rate of unemployment”. *Brookings Papers on Economic Activity*, 2, 429-480.

Schinasi, G.J., and P.A.V.B. Swamy (1989). “The out-of-sample forecasting performance of exchange rate models when coefficients are allowed to change”. *Journal of International*

Money and Finance, 8, 375-390.

Seo, B. (1998). "Test for structural change in cointegrated systems". *Econometric Theory*, 14, 222-259.

Seo, M. (2006). "Bootstrap testing for the null of no cointegration in a threshold vector error correction model". *Journal of Econometrics*, 134, 129-150.

Shin, Y. (1994). " A residual-based test of the null of cointegration against the alternative of no cointegration". *Econometric Theory*, 10, 91-115.

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

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.

Wu, W.B. (2002). "Central limit theorems for functional of linear processes and their applications". *Statistica Sinica*, 12, 635-649.

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