

# Partial Sum Processes of Residual-Based and Wald-type Break-Point Statistics in Time Series Regression Models\*

Christis Katsouris<sup>†</sup>

Department of Economics, University of Southampton

## Abstract

We revisit classical asymptotics when testing for a structural break in linear regression models by obtaining the limit theory of residual-based and Wald-type processes. First, we establish the Brownian bridge limiting distribution of these test statistics. Second, we study the asymptotic behaviour of the partial-sum processes in nonstationary (linear) time series regression models. Although, the particular comparisons of these two different modelling environments is done from the perspective of the partial-sum processes, it emphasizes that the presence of nuisance parameters can change the asymptotic behaviour of the functionals under consideration. Simulation experiments verify size distortions when testing for a break in nonstationary time series regressions which indicates that the normalized Brownian bridge limit cannot provide a suitable asymptotic approximation in this case. Further research is required to establish the cause of size distortions under the null hypothesis of parameter stability.

**Keywords:** Structural break tests, Wald-type statistic, OLS-CUSUM statistic, Brownian bridge.

## 1 INTRODUCTION

Severe time series fluctuations manifesting as market exuberance are considered by econometricians as early warning signs of upcoming economic recessions which are not usually explained by common boom and bust cycles (see, [Greenwood et al. \(2020\)](#), [Baron et al. \(2021\)](#) and [Katsouris \(2021\)](#)). Furthermore, the economic aspects of prolonged economic policy uncertainty as well as as more recently pandemics can impact the robustness of parameter estimates due to increased model uncertainty. More precisely, such economic phenomena can appear in specification functions as structural breaks to parameter coefficients. Therefore, the identification and estimation of the true break-points can improve forecasts and reduce model uncertainty. Specifically, in the literature there is a plethora of statistical methodologies for structural break detection under different modelling environments (see, [Bai \(1997\)](#), [Bai and Perron \(1998\)](#) and [Pitarakis \(2004\)](#) among others). In this paper we focus on the properties of the partial sum processes when constructing test statistics for testing the null hypothesis of no parameter instability in time series regression models under the assumption of stationary regressors vis-a-vis nonstationary regressors.

---

\*Article History: May 2020. December 2021.

I would like to thank Jose Olmo, Tassos Magdalinos and Jean-Yves Pitarakis for guidance, support and continuous encouragement throughout the PhD programme. Financial support from the VC PhD studentship of the University of Southampton is gratefully acknowledged. The author declares no conflicts of interest. All remaining errors are mine.

<sup>†</sup>Ph.D. Candidate, Department of Economics, University of Southampton, Southampton, SO17 1BJ, UK. *E-mail Address:* [C.Katsouris@soton.ac.uk](mailto:C.Katsouris@soton.ac.uk).

Although we do not introduce any new testing methodologies we review some important asymptotic theory results related to testing for structural breaks in time series regression models. More precisely, we study the asymptotic behaviour of partial-sum processes when these are employed to construct residual-based and Wald-type statistics in time series regression models. Our motivation for revisiting some of these aspects is to add on the discussion regarding the adequacy of finite-distribution approximations to large-sample theory of statistics. In particular, a common misconception when deriving asymptotic theory of test statistics and estimators is the fact that the large sample theory certainly provides good approximations to finite-sample results (see, [Chernoff and Savage \(1958\)](#)). Lastly, we consider two different modelling frameworks that is a stationary time series regression model and a nonstationary time series regression.

In terms of the structural break testing framework, we consider various examples which cover both the case of a known break-point as well as an unknown break-point, under parametric assumptions regarding the distribution of the innovations. Specifically, throughout this paper we assume that the innovations driving the error terms of both the stationary and nonstationary time series models are *i.i.d.*, with a known distribution, that is, are Gaussian distributed which implies that we are within the realm of parametric methods. Moreover, relaxing the particular assumption by assuming that distribution function of these innovations is considered as a nuisance parameter, requires to consider a semiparametric framework for estimation and inference purposes which is beyond the scope of this paper. Therefore, we have that the sequence of innovations,  $\{\epsilon_t\}_{t=1}^{\infty}$  to be  $\epsilon_t \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$  which implies independence and homoscedasticity<sup>1</sup>. Furthermore, the assumption of stationary innovation sequences is also imposed to facilitate the limit theory and this property holds regardless of the time series properties of the regressors and regressand.

In particular, we are interested in obtaining limit results of the following form

$$\sup_{0 \leq s \leq 1} |\mathcal{C}_T(s)| \xrightarrow{d} \sup_{0 \leq s \leq 1} |W(s)| \tag{1.1}$$

where  $\mathcal{C}_T(s)$  is the test statistic and  $W(s)$  is the standard Wiener process for some  $s \in [0, 1]$ . However, the emphasis in this paper is the investigation of the main asymptotic theory aspects based of the regression model under consideration in terms of the properties of regressors. Therefore, the specific comparison allow us to focus on the implementation of the partial-sum processes when deriving asymptotic theory results for test statistics for these two classes of time series regression models, that is, the classical linear regression versus nonstationary time series regression. The investigation of the properties of partial-sum processes for the residuals of non-linear regression models such as ARMA, ARMAX, ARCH or GARCH models can be indeed quite fruitful, but we leave the particular considerations for future research. Related studies include the papers of [Kulperger et al. \(2005\)](#), [Aue et al. \(2006\)](#) and [Gronneberg and Holcblat \(2019\)](#).

All limits are taken as  $T \rightarrow \infty$  where  $T$  is the sample size. The symbol " $\Rightarrow$ " is used to denote the weak convergence of the associated probability measures as  $T \rightarrow \infty$  (as defined in [Billingsley \(2013\)](#)) which implies convergence of sequences of random vectors of continuous cadlag functions on  $[0, 1]$  within a Skorokhod topology. The symbol  $\xrightarrow{d}$  denotes convergence in distribution and  $\xrightarrow{P}$  denotes convergence in probability. The remainder of the paper is organized as follows. Section 2 discusses some examples related to the use of partial-sum processes for residual-based statistics in linear regression models. Section 3 discusses Wald-type statistics in linear regression models. Section 4 discusses some aspects related to the asymptotic theory for the structural break testing framework in nonstationary time series regressions. Section 5 concludes and discusses further aspects for research.

---

<sup>1</sup>Notice that these two assumptions can indeed be quite strong. For example weakly dependent and heteroscedastic innovations can reflect more accurately the properties of aggregate time series and therefore such assumptions can be also included via appropriate modifications.

## 2 RESIDUAL-BASED STATISTICS

The use of partial-sum processes for model residuals when testing the null hypothesis of no structural break using residual-based statistics appears in various applications in the literature (Katsouris, 2021). Firstly, Brown et al. (1975) proposed the OLS-CUSUM test constructed based on cumulated sums of recursive residuals for testing for the presence of a single structural break in coefficients of the linear regression model (see, also Krämer et al. (1988) and Ploberger and Krämer (1992)). Other residual-based statistics found in the literature include the OLS-MOSUM test proposed by Chu et al. (1995a) which is constructed as sums of a fixed number of residuals that move across the whole sample. Therefore, in this case the statistic can be more sensitive in detecting parameter changes in comparison to the cumulated sums of recursive residuals. Furthermore, the literature evolved towards the construction of structural break statistics within an online monitoring framework which includes a window of fixed size (historical period) and an out-of-sample estimation window (monitoring period). In particular, Chu et al. (1995b) and Kuan and Chen (1994) proposed the Moving Estimates and Recursive Estimates statistics for parameter stability respectively. A unified framework has been then proposed by the seminal study of Chu et al. (1996) and also further examined by Leisch et al. (2000) for the case of the generalized fluctuation test.

Some additional considerations include the study of the characteristics of structural change which includes the frequency of breaks in time series such as single vis-a-vis multiple break points (e.g., Bai (1997) and Bai and Perron (1998)) as well as the nature of structural change, which implies detecting structural breaks in the conditional mean vis-a-vis the conditional variance or higher moments of partial sum processes (e.g., Horváth et al. (2001), Kulperger et al. (2005), Pitarakis (2004), Andreou and Ghysels (2006, 2009)). Furthermore, an alternative asymptotic analysis of residual-based statistics is proposed by Andreou and Werker (2012). The current literature extensively studies structural changes in the mean and variance of regression coefficients, however less attention is given to the study of structural break testing due to smooth changes in the persistence properties of regressors as it is defined within the framework of local-to-unity for autoregressive models. Specifically, smooth transitions of stochastic processes from  $I(0)$  to  $I(1)$  and other non-stationarities can be employed for the detection of bubbles in financial markets (see, Horváth et al. (2020)). In this paper, our aim is to provide a discussion of the use of partial sum processes of residual-based and Wald-type statistics for these two different modelling environments.

### 2.1 OLS-CUSUM test

The OLS-CUSUM test statistic (see, Krämer et al. (1988)) belongs to the class of residual based statistics (see, e.g., Stock (1994)) and it involves the partial sum processes of regression residuals based on the model under consideration. Following the literature we define a general class of regression residuals based on the partial sum process as proposed by Kulperger et al. (2005) given by Definition 1 and Theorem 1 below.

**Definition 1.** *The partial sum process of residuals is given by*

$$\widehat{\mathcal{U}}_T(s) = \sum_{t=1}^{\lfloor Ts \rfloor} \widehat{u}_t, \quad \text{for some } s \in [0, 1].$$

*with the partial sum process of the corresponding innovations is similarly defined as*

$$\mathcal{U}_T(s) = \sum_{t=1}^{\lfloor Ts \rfloor} u_t, \quad \text{for some } s \in [0, 1].$$

**Theorem 1.** Suppose that  $\sqrt{T}|\hat{\theta}_T - \theta| = \mathcal{O}_p(1)$  where  $\hat{\theta}_T$  is the set of estimated model parameters and  $\theta$  is in the interior of  $\Theta$ . If  $\mathbb{E}(|u_0|) < \infty$ , then

$$\sup_{s \in [0,1]} \frac{1}{\sqrt{T}} \left| \left( \hat{\mathcal{U}}_T(s) - s\hat{\mathcal{U}}_T(s) \right) - \left( \mathcal{U}_T(s) - s\mathcal{U}_T(s) \right) \right| = o_p(1). \quad (2.1)$$

Then, the invariance principle for partial sums for an i.i.d sequence  $\{u_t\}$  such that

$$\left\{ M_T(s) := \frac{\mathcal{U}_T(s) - s\mathcal{U}_T(s)}{\sigma_u \sqrt{T}}, s \in [0,1] \right\} \quad \text{and} \quad \left\{ \widehat{M}_T(s) := \frac{\hat{\mathcal{U}}_T(s) - s\hat{\mathcal{U}}_T(s)}{\sigma_u \sqrt{T}}, s \in [0,1] \right\} \quad (2.2)$$

where  $\sigma_u^2 = \mathbb{E}(u^2 - \mu_u)^2$  and  $\mu_u = \mathbb{E}(u_0)$ , implies that the functionals  $M_T(s)$  and  $\widehat{M}_T(s)$  both converge weakly in the Skorokhod topology  $\mathcal{D}[0,1]$ , to a Brownian bridge  $\left\{ \mathcal{B}\mathcal{B}(s) = W(s) - sW(1), s \in [0,1] \right\}$ .

**Remark 1.** The functional  $M_T(s)$  represents a partial-sum process and  $W(s)$  is the corresponding standard Wiener process on the interval  $[0,1]$  which preserves the continuous mapping theorem and convergence in a suitable functional space for any continuous function  $g(\cdot)$ , such that  $g(M_T(s)) \xrightarrow{\mathcal{D}} g(W(s))$ . Furthermore, both Definition 1 and Theorem 1 corresponds to partial-sum process of the residual sequence of a time series regression model with a general specification form, under the assumption of stationarity and ergodicity. Similarly, we can generalize these limit results to second order partial-sum processes for square residual sequences. More precisely, the first and second order partial-sum processes represent functionals of the OLS-CUSUM and OLS-CUSUM squared (see, [Deng and Perron \(2008\)](#)).

The proof of Theorem 1 is omitted which demonstrates a weak invariance principle; a stronger version of Donsker's classical functional central limit theorem (see, [Kulperger et al. \(2005\)](#) and ([Csörgő et al., 2003](#))). In particular, the weakly convergence of the asymptotic distribution of the OLS-CUSUM statistic for the classical regression model is studied by [Aue and Horváth \(2013\)](#). Related limit results can be found in the book of [Csorgo and Horvath \(1997\)](#). The development of the asymptotic theory for the residual-based and Wald-type statistics when testing for structural breaks is based on the validity of Theorem 1. For the remaining of this section we consider some standard examples from the literature to demonstrate the use of the residual-based statistics and their corresponding limit results when conducting statistical inference.

**Example 1.** Consider the location model formulated as below

$$y_t = \mu_1 \mathbf{1}\{t \leq k\} + \mu_2 \mathbf{1}\{t > k\} + \epsilon_t, \quad t = 1, \dots, T \quad (2.3)$$

where  $\mu_1$  and  $\mu_2$  are both deterministic and the break point  $k = \lfloor Ts \rfloor$  for some  $s \in [0,1]$  is an unknown fixed fraction of the full sample. Suppose a parametric assumption on the error term holds, such that  $\epsilon_t \sim_{i.i.d} \mathcal{N}(0,1)$  for  $t = 1, \dots, T$  and the following moment conditions apply

$$(i). \quad \mathbb{E}(\epsilon_t | \mathcal{F}_{t-1}) = 0 \quad (ii). \quad \mathbb{E}(\epsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\epsilon^2 \quad (iii). \quad \mathbb{E}(|\epsilon_t|)^{2+d} < \infty, \quad d > 0. \quad (2.4)$$

Then, the OLS estimator  $\hat{\mu} = T^{-1} \sum_{t=1}^T y_t$  of  $\mu$  is a  $\sqrt{T}$ -consistent estimator and asymptotically normal. Furthermore, based on the location model formulation given by (2.3) the objective is to conduct a two-sided test of the testing hypothesis  $\mathbb{H}_0 : \mu_1 = \mu_2$ . Under the null hypothesis we obtain a consistent estimator  $\hat{\sigma}_\epsilon^2$  of  $\sigma_\epsilon^2 < \infty$  such that  $\hat{\sigma}_\epsilon^2 \xrightarrow{p} \sigma_\epsilon^2$  as  $T \rightarrow \infty$  as well as the OLS-residuals defined as  $\hat{\epsilon}_t = (y_t - \hat{\mu})$ .

Thus, we have that

$$\hat{\epsilon}_t \equiv (y_t - \bar{y}_T) = \mu + \epsilon_t - \frac{1}{T} \sum_{t=1}^T (\mu + \epsilon_t) \quad (2.5)$$

Based on Example 1, we consider that the following functional central limit theorem (FCLT) holds

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} \epsilon_t \Rightarrow \sigma_\epsilon W(s), \quad s \in [0, 1]. \quad (2.6)$$

which applies to the unobservable innovation terms of the above time series model that includes only a model intercept, where  $W(\cdot)$  is the standard Brownian motion such that  $W(s) \sim N(0, s)$ . For notational convenience we also denote with  $\sigma_\epsilon W(s) \equiv B(s)$  for some  $s \in [0, 1]$ .

Consider the standardized innovations defined as  $\epsilon_t^o = \left( \epsilon_t - \frac{1}{T} \sum_{t=1}^T \epsilon_t \right) \equiv (\epsilon_t - \bar{\epsilon})$  and suppose that  $\epsilon_t$  is an *i.i.d* sequence of innovations. Then the centered partial sum process is defined as below

$$M_T(s) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} (\epsilon_t - \bar{\epsilon}), \quad s \in [0, 1]. \quad (2.7)$$

and the corresponding residual centered partial sum process is defined as below

$$\widehat{M}_T(s) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} (\hat{\epsilon}_t - \bar{\epsilon}), \quad s \in [0, 1]. \quad (2.8)$$

Moreover, let  $\hat{\sigma}_m^2 = \widehat{M}_T^2(1)/T$  to be the sample variance estimator of the above partial sum process. Then,

$$\sup_{0 \leq s \leq 1} \frac{1}{\sqrt{T}} \left| \frac{\widehat{M}_T(s)}{\hat{\sigma}_m^2} - \frac{M_T(s)}{\sigma_m^2} \right| = o_p(1) \quad (2.9)$$

which shows that the self-normalized centered partial sum process  $\{\widehat{M}_T(s)/\hat{\sigma}_m^2, 0 \leq s \leq 1\}$  behaves as the residuals  $\{\hat{\epsilon}_t\}_{t=1}^T$  are asymptotically the same as the unobservable innovations  $\{\epsilon_t\}_{t=1}^T$ .

Therefore, the OLS-CUSUM statistic based on the standardized residuals is expressed as below

$$\mathcal{C}_T(k) = \max_{0 \leq k \leq T} \left\{ \frac{\sum_{t=1}^k \hat{\epsilon}_t - \frac{k}{T} \sum_{t=1}^T \hat{\epsilon}_t}{\hat{\sigma}_\epsilon \sqrt{T}} \right\} \equiv \sup_{0 \leq s \leq 1} \left\{ \frac{1}{\hat{\sigma}_\epsilon} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} \hat{\epsilon}_t - \frac{s}{\sqrt{T}} \sum_{t=1}^T \hat{\epsilon}_t \right) \right\} \quad (2.10)$$

Then, it can be shown that  $\mathcal{C}_T(k) \Rightarrow [W(s) - sW(1)]$  which is Brownian bridge. In particular, the weakly convergence of the OLS-CUSUM statistic is based on  $\max_{0 \leq k \leq T} \mathcal{C}_T(k) \Rightarrow \sup_{0 \leq s \leq 1} [W(s) - sW(1)]$ .

**Remark 2.** Notice that Example 1 considers a linear time series model with no covariates, thus the structural break statistic can be considered as a difference in means for the two sub-samples, under the null hypothesis. The partial sum process for the innovation sequences of the model follows standard Brownian motion limit results. Furthermore, in Example 2 we consider the OLS-CUSUM statistic for the classical regression model as proposed by the seminal study of [Ploberger and Krämer \(1992\)](#).

**Example 2.** Consider the following time series model

$$y_t = x_t' \beta_1 \mathbf{1}\{t \leq k\} + x_t' \beta_2 \mathbf{1}\{t > k\} + \epsilon_t, \quad t = 1, \dots, T \quad (2.11)$$

where  $y_t$  is the dependent variable,  $x_t = [1, x_{2,t}, \dots, x_{p,t}]' = [1, \tilde{x}_t']'$  is a  $(p+1)$ -dimensional vector and the error term  $\epsilon_t$  are assumed to be i.i.d  $(0, \sigma_\epsilon^2)$ . We define  $x_{1,t} \equiv x_t' \mathbf{1}\{t \leq k\}$  and  $x_{2,t} \equiv x_t' \mathbf{1}\{t > k\}$  for  $k = [Ts]$  with  $s \in [0, 1]$ . Under the null hypothesis of no structural break  $\mathbb{H}_0 : \beta_1 = \beta_2$  with  $\hat{\beta}_T$  the  $\sqrt{T}$ -consistent estimator for  $\beta$  such that  $\sqrt{T}(\hat{\beta}_T - \beta) = \mathcal{O}_p(1)$  is bounded in probability.

Based on Example 2 the OLS-CUSUM statistic,  $\mathcal{C}_T(k)$ , is obtained using the OLS residuals under the null hypothesis, defined as  $\hat{\epsilon}_t = y_t - \hat{\beta}_T x_t = \epsilon_t - x_t'(\hat{\beta}_T - \beta)$ . Then, we obtain that

$$\mathcal{C}_T(k) = \frac{1}{\hat{\sigma}_\epsilon} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} \hat{\epsilon}_t - \frac{[Ts]}{T} \frac{1}{\sqrt{T}} \sum_{t=1}^T \hat{\epsilon}_t \right) \quad (2.12)$$

In particular, we aim to show that  $\mathcal{C}_T(k) \Rightarrow [W(s) - sW(1)]$ , which shows weakly convergence of the statistic to the Brownian bridge process. To do this, we consider the OLS residuals which can be expressed as below

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} \hat{\epsilon}_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} \epsilon_t - \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} x_t' (\hat{\beta}_T - \beta) \quad (2.13)$$

Furthermore, the following result holds

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} x_t' (\hat{\beta}_T - \beta) = \frac{r}{\sqrt{T}} \sum_{t=1}^T \epsilon_t + o_p(1). \quad (2.14)$$

A short proof on the asymptotic result above is provided here. We can express the left side of (2.14) as an inner product since our framework allows such representation

$$\left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} x_t' (\hat{\beta}_T - \beta) \right] = \left[ \frac{1}{T} \sum_{t=1}^{[Ts]} x_t' \right] \cdot \left[ \sqrt{T} (\hat{\beta}_T - \beta) \right] \quad (2.15)$$

Notice that it holds that

$$\left( \frac{1}{T} \sum_{t=1}^{[Ts]} x_t' \right) \xrightarrow{p} [s, 0, \dots, 0], \quad \text{since } \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^{[Ts]} \tilde{x}_t = 0 \quad \text{and} \quad \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T 1 = s \quad (2.16)$$

Then, the second term with a matrix decomposition for  $Q = \left( \frac{1}{T} \sum_{t=1}^T x_t x_t' \right)$  where  $x_t = [1, \tilde{x}_t']'$  can be expressed as below

$$\sqrt{T}(\hat{\beta}_T - \beta) = \left( \frac{1}{T} \sum_{t=1}^T x_t x_t' \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T x_t \epsilon_t \right) \equiv \frac{1}{\sqrt{T}} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \tilde{Q} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T \epsilon_t \\ \sum_{t=1}^T \tilde{x}_t \epsilon_t \end{bmatrix} + o_p(1) \quad (2.17)$$

since  $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^{[Ts]} \tilde{x}_t \tilde{x}_t' = \tilde{Q}$ . Also, note that  $\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \tilde{Q} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \tilde{Q}^{-1} \end{bmatrix}$ .

Therefore, we obtain

$$\left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} x_t' (\hat{\beta}_T - \beta) \right] \xrightarrow{p} [s \mathbf{0}] \frac{1}{\sqrt{T}} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \tilde{Q}^{-1} \end{bmatrix} \begin{bmatrix} \sum_{t=1}^T \epsilon_t \\ \sum_{t=1}^T \tilde{x}_t \epsilon_t \end{bmatrix} = \frac{s}{\sqrt{T}} \sum_{t=1}^T \epsilon_t + o_p(1) \quad (2.18)$$

where  $\mathbf{0}$  is  $(p-1)$  dimensional column vectors of zeros. Then, using the asymptotic result given by (2.14) as well as the expression for the OLS residuals as in (2.13) the OLS-CUSUM statistic can be expressed as

$$\begin{aligned} \mathcal{C}_T(k) &= \frac{1}{\hat{\sigma}_\epsilon} \frac{1}{\sqrt{T}} \left\{ \left( \sum_{t=1}^k \epsilon_t - \sum_{t=1}^k x_t' (\hat{\beta}_T - \beta) \right) - s \left( \sum_{t=1}^T \epsilon_t - \sum_{t=1}^T x_t' (\hat{\beta}_T - \beta) \right) \right\} \\ &= \frac{1}{\hat{\sigma}_\epsilon} \frac{1}{\sqrt{T}} \left\{ \left( \sum_{t=1}^k \epsilon_t - s \sum_{t=1}^T \epsilon_t \right) - \left( \sum_{t=1}^k x_t' (\hat{\beta}_T - \beta) - s \sum_{t=1}^T x_t' (\hat{\beta}_T - \beta) \right) \right\} \end{aligned}$$

Notice that the second term above converges to zero in probability such that

$$s \left( \sum_{t=1}^T \epsilon_t - \sum_{t=1}^T \hat{\epsilon}_t \right) = o_p(1) \quad (2.19)$$

then the limit result follows as given below

$$\mathcal{C}_T(k) = \sup_{s \in [0,1]} \left\{ \frac{1}{\hat{\sigma}_\epsilon} \left( \sum_{t=1}^k \frac{\epsilon_t}{\sqrt{T}} - s \sum_{t=1}^T \frac{\epsilon_t}{\sqrt{T}} \right) \right\} \Rightarrow [W(s) - sW(1)], \quad s \in [0,1]. \quad (2.20)$$

Thus, the OLS-CUSUM statistic weakly converges to the Brownian bridge limit uniformly for  $s \in [0,1]$ .

**Example 3.** Similarly to Examples 1 and 2 we can derive limit theory results for the OLS-CUSUM squared test (see detailed proofs in Deng and Perron (2008)). Following Definition 1 and Theorem 1 the corresponding test statistic is defined as below

$$\mathcal{C}_T^{(2)}(k) = \max_{0 \leq k \leq T} \left( \frac{\sum_{t=1}^k \hat{\epsilon}_t^2 - \frac{k}{T} \sum_{t=1}^T \hat{\epsilon}_t^2}{\hat{\sigma}_\epsilon \sqrt{T}} \right) = \sup_{0 \leq s \leq 1} \left\{ \frac{1}{\hat{\sigma}_\epsilon} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} \hat{\epsilon}_t^2 - \frac{s}{\sqrt{T}} \sum_{t=1}^T \hat{\epsilon}_t^2 \right) \right\} \quad (2.21)$$

Furthermore, the following two sufficient conditions hold

$$\max_{1 \leq j \leq n} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^j \epsilon_t x_t' (\hat{\beta}_T - \beta) \right| \xrightarrow{p} 0 \quad (2.22)$$

$$\max_{1 \leq j \leq n} \frac{1}{\sqrt{T}} (\hat{\beta}_T - \beta)' \sum_{t=1}^j x_t x_t' (\hat{\beta}_T - \beta) \xrightarrow{p} 0 \quad (2.23)$$

Therefore, it can be proved that the weakly convergence result to a Brownian bridge follows

$$\mathcal{C}_T^{(2)}(k) \Rightarrow \sup_{0 \leq s \leq T} [W(s) - sW(1)]. \quad (2.24)$$



### 3 WALD TYPE STATISTICS

In this section we examine the implementation of Wald-type statistics when constructing an equivalent structural break test for the conditional mean of a regression model. The particular methodology has been advanced by the seminal work of [Hawkins \(1987\)](#) and [Andrews \(1993\)](#).

**Definition 2.** We define the following  $p$ -dimensional limiting distribution for some  $\pi \in (0, 1)$  such that

$$\mathcal{Q}_p(\pi) := \frac{\left[ W_p(\pi) - \pi W_p(1) \right]' \left[ W_p(\pi) - \pi W_p(1) \right]}{\pi(1 - \pi)} \quad (3.1)$$

where  $W_p(\cdot)$  is  $p \times 1$  vector of independent standard Brownian motions for some  $p \geq 1$ , where  $p$  represents the number of parameters that the null hypothesis is testing for stability.

**Remark 3.** The limit process  $\mathcal{Q}_p(\pi)$  is referred to as the square of a standardized tied-down Bessel process of order  $p$  (see, [Andrews \(1993\)](#)). In particular, the limit process given by Definition 2 is employed when deriving the asymptotic distribution of sup-Wald statistics, such that  $\sup_{\pi \in (0,1)} W_T(\pi) \xrightarrow{d} \sup_{\pi \in (0,1)} \mathcal{Q}_p(\pi)$ .

We mainly consider univariate regression models but the framework can be also generalized to time series regressions with multiple regressors. Furthermore, when applying the supremum functional to Wald-type statistics we assume that the unknown break-point  $\pi \in (0, 1)$  lies in a symmetric subset of the particular unit set, to ensure that inference is not at the boundary of the parameter space (see, [Andrews \(2001\)](#)).

#### 3.1 Unconditional Mean test

Consider the following time series regression model

$$y_t = \theta_1 \mathbf{1}\{t \leq k\} + \theta_2 \mathbf{1}\{t > k\} + \epsilon_t, \quad t = 1, \dots, T \quad (3.2)$$

where  $\theta_1$  and  $\theta_2$  are both deterministic and the break point  $k = \lfloor Ts \rfloor$  for some  $s \in (0, 1)$  is an unknown fixed fraction of the full sample period. The following assumption hold:

**Assumption 1.** Let  $\epsilon_t$  be a sequence of random variables. Then, the following moment conditions hold.

A1.  $\mathbb{E}(\epsilon_t | \mathcal{F}_{t-1}) = 0$  with an asymptotic variance as  $T \rightarrow \infty$  given as below

$$\text{var} \left( T^{-1/2} \sum_{t=1}^{\lfloor Ts \rfloor} \epsilon_t \right) = \frac{\lfloor Ts \rfloor}{T} \sigma_\epsilon^2 \xrightarrow{p} s \sigma_\epsilon^2, \quad \sigma_\epsilon^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \left( \sum_{t=1}^T \epsilon_t \right)^2 \right] \quad (3.3)$$

A2.  $\sup_t \mathbb{E}(|\epsilon_t|)^{2+d} < \infty$  for some  $d > 2$ .

A3.  $\{\epsilon_t\}_{t=1}^T$  satisfies the following Functional Central Limit Theorem (FCLT),

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} \epsilon_t \xrightarrow{\mathcal{D}} \sigma_\epsilon W(s) \equiv B(s) \quad (3.4)$$



**Remark 4.** Assumption A1 gives the moment conditions of the classical regression model with only intercept. In particular,  $\text{Var}\left(T^{-1/2} \sum_{t=1}^{\lfloor Ts \rfloor} \epsilon_t\right)$  gives the asymptotic variance of the partial-sum of innovations. Since the limiting variance is bounded then convergence in distribution follows.

The null and alternative hypothesis are as below

$$\mathbb{H}_0 : \theta_1 = \theta_2 \text{ versus } \mathbb{H}_A : \theta_1 \neq \theta_2 \quad (3.5)$$

and we consider the following test statistic for detecting structural change in the unconditional mean of the simple linear regression with only intercept.

$$\mathcal{Z}_T(s) = \max_{0 \leq k \leq T} \left\{ T \frac{(\bar{y}_1 - \bar{y}_2)^2}{\hat{\sigma}_z^2} \right\} \equiv \sup_{s \in (0,1)} \left\{ T \frac{(\bar{y}_1 - \bar{y}_2)^2}{\hat{\sigma}_z^2} \right\} + o_p(1) \quad (3.6)$$

$$\bar{y}_1 = \frac{1}{k} \sum_{t=1}^k y_t \quad \text{and} \quad \bar{y}_2 = \frac{1}{T-k} \sum_{t=k+1}^T y_t \quad \text{where} \quad \hat{\sigma}_z^2 = \frac{\hat{\sigma}_\epsilon^2}{\frac{k}{T} \left(1 - \frac{k}{T}\right)} \xrightarrow{p} \frac{\sigma_\epsilon^2}{s(1-s)} \quad (3.7)$$

We can show that the asymptotic variance of the normalized statistical distance measure  $\sqrt{T}(\bar{y}_1 - \bar{y}_2)$  is given by  $\text{Avar}\left[\sqrt{T}(\bar{y}_1 - \bar{y}_2)\right] = \frac{\sigma_\epsilon^2}{s(1-s)} \equiv \sigma_z^2$  by noting that  $\lim_{T \rightarrow \infty} \frac{T}{\lfloor Ts \rfloor} = \frac{1}{s}$  and  $\lim_{T \rightarrow \infty} \frac{T}{T - \lfloor Ts \rfloor} = \frac{1}{1-s}$ . According to [Aue and Horváth \(2013\)](#) testing the null hypothesis of equal means across a  $p$ -dimensional multivariate time series is equivalent to constructing a  $p$ -dimensional CUSUM process and testing for a structural break at an unknown break point  $k$ . Due to the fact that the partial-sum process representing the CUSUM statistic has a weak convergence to a  $p$ -dimensional Brownian Motion process, then the quadratic form of the partial-sum process weakly convergence to the sum of squared independent Brownian bridges. The specific property permits to establish an equivalence between a supremum Wald-type statistic and a CUSUM-type statistic. In particular, we show that the asymptotic distribution of the OLS-CUSUM statistic can be deduced from the asymptotic distribution of  $\mathcal{Z}_T$  and vice-versa.

**Theorem 2.** Let  $\mathcal{Z}_T(s)$  be the test statistic for testing the null hypothesis  $\mathbb{H}_0 : \theta_1 = \theta_2$  of no structural change in the unconditional mean as defined by (3.6) for some  $0 \leq k \leq T$ . Assume the conditions A1 to A3 given by Assumption 2 hold and let  $\mathcal{Z}_T^o(s)$  be an OLS-CUSUM type statistic. Then, both statistics weakly converge in the Skorokhod space  $\mathcal{D}[0,1]$  to functionals of a Brownian bridge such as  $\{\mathcal{B}\mathcal{B}(s) : s \in [0,1]\}$  where  $\mathcal{B}\mathcal{B}(s) = [W(s) - sW(1)]$ .

$$\mathcal{Z}_T(r) = \sup_{s \in [\nu, 1-\nu]} \left\{ T \frac{(\bar{y}_1 - \bar{y}_2)^2}{\hat{\sigma}_z^2} \right\} \Rightarrow \sup_{s \in [\nu, 1-\nu]} \frac{[W(s) - sW(1)]^2}{s(1-s)} \quad (3.8)$$

$$\mathcal{Z}_T^o(s) = \sup_{s \in [\nu, 1-\nu]} \left\{ \frac{1}{\sqrt{T}} \frac{1}{\sqrt{\hat{\sigma}_\epsilon}} \left( \sum_{t=1}^{\lfloor Ts \rfloor} y_t - s \sum_{t=1}^T y_t \right) \right\} \Rightarrow \sup_{s \in [\nu, 1-\nu]} [W(s) - sW(1)] \quad (3.9)$$

**Remark 5.** Theorem 2 implies that the  $\mathcal{Z}_T(s)$  statistic of equality of means in a SLR model with only intercept across the unknown break point has an equivalent weak convergence to the CUSUM statistic of the time series under consideration using a suitable normalization constant.

Next, we consider the corresponding supremum Wald-type statistic for testing for structural change in the unconditional mean of the classical regression model with only intercept. The test is formulated as below

$$\mathcal{W}_T = \frac{1}{\hat{\sigma}_\epsilon^2} (\mathbf{R}\Theta)' \left[ \mathbf{R} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{R}' \right]^{-1} (\mathbf{R}\Theta) \quad (3.10)$$

where  $\hat{\sigma}_\epsilon^2 = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^2(k)$  the residual variance under the null hypothesis and is a consistent estimator of  $\sigma_\epsilon$  such that  $\hat{\sigma}_\epsilon^2 \xrightarrow{P} \sigma_\epsilon^2$ . Thus, by substituting the restriction matrix  $\mathbf{R} = [\mathbf{I} - \mathbf{I}]$ ,  $\mathbf{Z} = [\mathbf{X}_1 \ \mathbf{X}_2]'$  and the estimator  $\hat{\Theta} = [\hat{\theta}_1 \ \hat{\theta}_2]'$  of  $\Theta$  into the above formulation of the Wald statistic we obtain the expression

$$\mathcal{W}_T(s) = \frac{1}{\hat{\sigma}_\epsilon^2} (\hat{\theta}_1 - \hat{\theta}_1)' \left[ (X_1' X_1)^{-1} + (X_2' X_2)^{-1} \right]^{-1} (\hat{\theta}_1 - \hat{\theta}_1) \quad (3.11)$$

We observe that for the unconditional mean model it holds that

$$(X_1' X_1) = \sum_{t=1}^T \mathbf{1}\{t \leq k\} = k \quad \text{and} \quad (X_2' X_2) = \sum_{t=1}^T \mathbf{1}\{t > k\} = T - k$$

Since in this section we consider mean shifts,  $X_1$  stacks the elements of  $x_t$  for which  $t \leq k$ , that is,  $\mathbf{1}\{t \leq k\}$  and  $X_2$  stacks the elements of  $x_t$  for which  $t > k$ , that is,  $\mathbf{1}\{t > k\}$ . Therefore, using the corresponding matrix notation the OLS estimators can be written as below

$$\hat{\theta}_1 = (X_1' X_1)^{-1} X_1' y = \frac{1}{k} \sum_{t=1}^k y_t, \quad \hat{\theta}_2 = (X_2' X_2)^{-1} X_2' y = \frac{1}{T-k} \sum_{t=1}^T y_t$$

Thus, the Wald statistic can be expressed as

$$\mathcal{W}_T(s) = \frac{1}{\hat{\sigma}_\epsilon^2} (\bar{y}_1 - \bar{y}_2)^2 \left[ \frac{1}{k} + \frac{1}{T-k} \right]^{-1} = \frac{k(T-k)}{T\hat{\sigma}_\epsilon^2} (\bar{y}_1 - \bar{y}_2)^2$$

Under the null hypothesis  $\mathbb{H}_0 : \theta_1 = \theta_2$ , we have that since  $(\bar{y}_1 - \bar{y}_2) \equiv (\hat{\theta}_1 - \hat{\theta}_2)$  we obtain the expression

$$(\hat{\theta}_1 - \hat{\theta}_2) = \frac{T}{k(T-k)} \left( \sum_{t=1}^k \epsilon_t - \frac{k}{T} \sum_{t=1}^T \epsilon_t \right) = \frac{T\sqrt{T}}{k(T-k)} \left( \sum_{t=1}^k \frac{\epsilon_t}{\sqrt{T}} - \frac{k}{T} \sum_{t=1}^T \frac{\epsilon_t}{\sqrt{T}} \right)$$

Therefore, the Wald test is equivalently written as below

$$\begin{aligned} \mathcal{W}_T(s) &= \frac{k(T-k)}{T\hat{\sigma}_\epsilon^2} \left\{ \frac{T\sqrt{T}}{k(T-k)} \left( \sum_{t=1}^k \frac{\epsilon_t}{\sqrt{T}} - \frac{k}{T} \sum_{t=1}^T \frac{\epsilon_t}{\sqrt{T}} \right) \right\}^2 \\ &= \frac{T}{\frac{k}{T} \left(1 - \frac{k}{T}\right) \hat{\sigma}_\epsilon^2} \left( \sum_{t=1}^k \frac{\epsilon_t}{\sqrt{T}} - \frac{k}{T} \sum_{t=1}^T \frac{\epsilon_t}{\sqrt{T}} \right)^2. \end{aligned}$$

A simple application of the WLLN for the variance of the OLS estimator implies that  $\hat{\sigma}_\epsilon^2 \xrightarrow{P} \sigma_\epsilon$  as  $T \rightarrow \infty$ . Moreover since it holds that  $\frac{k}{T} \frac{T-k}{T} \xrightarrow{P} s(1-s)$ , we obtain the weak convergence for the sup-Wald statistic

$$\mathcal{W}_T^*(s) := \sup_{s \in [\nu, 1-\nu]} \mathcal{W}_T(s) \Rightarrow \sup_{s \in [\nu, 1-\nu]} \frac{[W(s) - sW(1)]' [W(s) - sW(1)]}{s(1-s)}. \quad (3.12)$$

The above asymptotic result indeed verifies the weakly convergence of the sup-Wald statistic into the supremum of a normalized squared Brownian bridge, specifically for the unconditional mean specification of the regression model. Then, statistical inference is conducted based on the null hypothesis,  $\mathbb{H}_0 : \theta_1 = \theta_2$ , which is rejected for large values of the sup-Wald statistic with a significance level  $\alpha \in (0, 1)$ . Thus, the exact form of the limiting distribution of the statistic is employed to obtain associated critical values, denoted with  $c_\alpha$  such that  $\mathbb{P}(\mathcal{W}_T^*(s) > c_\alpha) > 0$  with  $\lim_{T \rightarrow \infty} \mathbb{P}(\mathcal{W}_T^*(s) > c_\alpha) = 1$ .

### 3.2 Conditional Mean test

Consider the following model

$$y_t = x_t' \theta_1 \mathbf{1}\{t \leq k\} + x_t' \theta_2 \mathbf{1}\{t > k\} + \epsilon_t, \quad t = 1, \dots, T \quad (3.13)$$

where  $y_t$  is the dependent variable,  $x_t$  is a  $p \times 1$  vector of regressors,  $\epsilon_t$  is an unobservable disturbance term with  $\mathbb{E}(\epsilon_t | x_t) = 0$  almost surely and  $\theta_1$  and  $\theta_2$  the regression coefficients formulated under the null hypothesis of no structural break. Define with  $x_{1,t} \equiv x_t' \mathbf{1}\{t \leq k\}$  and  $x_{2,t} \equiv x_t' \mathbf{1}\{t > k\}$  where  $k = \lfloor Ts \rfloor$  for some  $s \in (0, 1)$ . Furthermore, notice that the regressor vector  $x_t$  can contain exogenous regressors and lagged dependent variables with unknown integration order. Moreover, under suitable regularity assumptions one can consider a static, a dynamic time series regression model as well as models with integrated regressors or cointegrated regressors.

The null hypothesis of interest is

$$\mathbb{H}_0 : \theta_1 = \theta_2 \equiv \theta_0 \in \mathbb{R}^{p \times 1} \text{ for all } t. \quad (3.14)$$

The alternative hypothesis  $\mathbb{H}_A$  is that  $\mathbb{H}_0$  is false, which implies that the regression coefficient has a structural break at some unknown break point in the sample. Under  $\mathbb{H}_0$ , the unknown constant parameter vector  $\theta$  can be consistently estimated using the OLS estimator such that

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^{p \times 1}} \sum_{t=1}^T (y_t - x_t' \theta)^2 \quad (3.15)$$

Under the alternative  $\mathbb{H}_A$ ,  $\theta_0 \equiv \theta_t$  can be considered as a time-varying parameter vector, which implies,  $\theta_t \equiv \theta_1$  for  $1 \leq t \leq k$  and  $\theta_t \equiv \theta_2$  for  $k + 1 \leq t \leq T$ . To facilitate the estimation and inference based on the above regression model we impose the following regularity conditions.

**Assumption 2.** *Let  $\epsilon_t$  be a sequence of random variables. Then the following moment conditions and FCLT hold, which allow  $x_t$  and  $\epsilon_t$  to be weakly dependent.*

- B1.  $\{\epsilon_t\}_{t=1}^T$  is a homoskedastic martingale difference sequence (m.d.s) such that  $\mathbb{E}(\epsilon_t | \mathcal{F}_{t-1}) = 0$  and  $\mathbb{E}(\epsilon_t^2) = \sigma^2$ , where  $\mathcal{F}_{t-1} = \{x_t', x_{t-1}', \dots, \epsilon_{t-1}, \epsilon_{t-2}, \dots\}$ .
- B2.  $\{x_t\}_{t=1}^T$  has at least a finite second moment, that is,  $\sup_t \sum_{t=1}^T \|x_t\|^{2+d} < \infty$  where  $d > 0$  and the following weak law of large numbers (WLLN) holds

$$\frac{1}{T} \sum_{t=1}^T x_t x_t' \xrightarrow{p} \mathbf{Q}_T \text{ as } T \rightarrow \infty \text{ such that } \sup_{s \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{\lfloor Ts \rfloor} x_t x_t' - s \mathbf{Q}_T \right| = o_p(1) \quad (3.16)$$

where  $\mathbf{Q}_T$  is a  $(p \times p)$  non-stochastic finite and positive definite matrix.

- B3.  $\{x_t \epsilon_t\}_{t=1}^T$  satisfies the following Functional Central Limit Theorem<sup>2</sup> (FCLT),

$$\mathcal{S}_T(s) = \left( T^{-1/2} \mathbf{\Omega}_T^{-1/2} \sum_{t=1}^{\lfloor Ts \rfloor} x_t \epsilon_t \right) \Rightarrow \mathbf{W}_p(s), \quad s \in (0, 1) \text{ and } \mathbf{\Omega}_T = \mathbb{E}(x_t x_t' \epsilon_t^2) > 0. \quad (3.17)$$

---

<sup>2</sup>The multivariate FCLT is examined in the studies of [Phillips and Durlauf \(1986\)](#) and [Wooldridge and White \(1988\)](#).

In other words, the general moment conditions given by Assumption 2 permits to consider both residual-based statistics as well as Wald-type statistics as suitable detectors when testing for a break-point within the full sample. In particular, [Chen and Hong \(2012\)](#) consider the implementation of generalized Hausman-type tests using nonparametric estimation and inference techniques.

### 3.2.1 Asymptotic Convergence of sup-Wald test

Next, we derive the asymptotic convergence of the Wald test when detecting a single structural break in the regression for the classical regression model with a conditional mean specification form, under Assumptions B1 to B3. Furthermore, under the null hypothesis the break-point is unidentified, thus to facilitate statistical inference we consider the corresponding supremum Wald-type statistic based on the OLS estimator which is expressed as  $\mathcal{W}_T^*(s) := \sup_{s \in [\nu, 1-\nu]} \mathcal{W}_T(s)$  for some  $\nu \in (0, 1)$ .

Using the null hypothesis  $\mathbb{H}_0 : \theta_1 = \theta_2$  of no structural break we obtain an equivalent representation using the linear restriction matrix  $\mathbf{R}$  of rank  $q$ , which implies that  $\mathbb{H}_0 : \mathbf{R}\boldsymbol{\Theta} = \mathbf{0}$ . We define  $\boldsymbol{\Theta} = [\theta_1 \ \theta_2]'$ ,  $\mathbf{Z} = [\mathbf{X}_1 \ \mathbf{X}_2]'$  and  $\mathbf{R} = [\mathbf{I} - \mathbf{I}]$  and prove that the asymptotic distribution of the corresponding sup-Wald statistic weakly converges to the supremum of a normalized squared Brownian bridge.

**Theorem 3.** *Under the null hypothesis  $\mathbb{H}_0 : \theta_1 = \theta_2$  with  $0 < \nu < 1$  we define  $\mathcal{W}_T^*(s)$*

$$\mathcal{W}_T^*(s) := \sup_{s \in [\nu, 1-\nu]} \mathcal{W}_T(s) \Rightarrow \sup_{s \in [\nu, 1-\nu]} \frac{[\mathbf{W}_p(s) - s\mathbf{W}_p(1)]' \mathbf{Q}_T^{-1} [\mathbf{W}_p(s) - s\mathbf{W}_p(1)]}{s(1-s)}. \quad (3.18)$$

where  $\mathcal{W}_T(s)$  denotes the Wald statistic for testing the null hypothesis  $\mathbb{H}_0 : \theta_1 = \theta_2$  and is expressed as

$$\mathcal{W}_T = \frac{1}{\hat{\sigma}_\epsilon^2} (\mathbf{R}\boldsymbol{\Theta})' \left[ \mathbf{R} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{R}' \right]^{-1} (\mathbf{R}\boldsymbol{\Theta}) \quad (3.19)$$

with  $\hat{\sigma}_\epsilon^2 = \frac{1}{T} \sum_{t=1}^T (y_t - x_t' \hat{\theta}_T)$  a consistent estimate of  $\sigma_\epsilon^2$  the OLS variance under the null hypothesis.

**Remark 6.** *The above asymptotic theory analysis demonstrates that CUSUM-type statistics for detecting a structural break in the classical regression model with a single unknown break-point typically weakly converge to Brownian bridge functionals, such that  $\mathcal{B}\mathcal{B}(s) := [\mathbf{W}_p(s) - s\mathbf{W}_p(1)]$  while the corresponding Wald-type tests weakly converge to a normalized version of the CUSUM test with a normalized constant given by  $k_T := \left\{ \frac{k}{T} \left( 1 - \frac{k}{T} \right) \right\}^{1/2}$  where  $k = \lfloor Ts \rfloor$  for some  $s \in [0, 1]$ .*

Monte Carlo simulation experiments can be used to compare the asymptotic validity and performance of OLS-CUSUM type statistics and Wald-type tests by obtaining associated empirical size and power results. An important criticism of CUSUM-type statistics is that these tests are based on residuals under the null hypothesis which implies that the test is not designed with a specific alternative under consideration. Therefore, although the use of these tests can lead to a monotonically increasing power, in practise it can be stochastically dominated by the power of a Wald type test (see, also [Andreou \(2008\)](#)). On the other hand the advantage of using a Wald-type statistic when testing for a structural break, is that the construction of the test allows to incorporate residuals obtained either under the null or under the alternative hypothesis, providing this way superior power performance. Intuitively, using the residual variance of the unrestricted model leads to better finite-sample power since the Wald-type statistic contains information from the alternative model (alternative hypothesis).

## 4 STRUCTURAL BREAK TESTING UNDER NONSTATIONARITY

Although the purpose of this paper is to explain the main challenges one faces when testing for a structural break in nonstandard econometric problems we discuss the main intuition when developing associated asymptotic theory using some examples. More specifically, in order to accommodate the nonstationary aspect in time series models<sup>3</sup> we need to consider a suitable probability space in which weakly convergence arguments holds. Due to the fact that the asymptotic terms of sample moments of nonstationary time series models often involve non-standard limit results such as convergence to stochastic integrals, care is needed when applying related convergence arguments. In particular, the stochastic integrals of the form  $\int_0^1 WdW(s)$  can be shown to converge weakly, under the null hypothesis, to the associated stochastic integral with the limiting Brownian motions, under the assumption of the existence of cadlag functionals in the unit interval equipped with the Skorokhod topology (see, Müller (2011)).

### 4.1 Linear Restrictions Testing

We begin our analysis by discussing the main limit theory which is employed in nonstationary time series models, by providing two examples: (i) a time series regression with integrated regressors and (ii) a predictive regression with persistent regressors. Although, such asymptotics are commonly used when considering the asymptotic behaviour of t-tests and Wald-type tests based on linear restrictions on the parameter coefficients, these are applicable when constructing the corresponding structural break statistics.

#### 4.1.1 Time Series Regression with Integrated Regressors

A class of nonstationary time series models include the linear regressions with integrated regressors as proposed by the studies of Cavanagh et al. (1995) and Jansson and Moreira (2006) among others.

**Example 4.** (Cointegrating Regression, Banerjee et al. (1993))

Consider the following bivariate system of co-integrated variables  $\{y_t\}_{t=1}^\infty$  and  $\{x_t\}_{t=1}^\infty$  such that

$$y_t = \beta x_t + u_t \tag{4.1}$$

$$x_t = x_{t-1} + \epsilon_t \tag{4.2}$$

with  $u_t \sim N(0, \sigma_u^2)$ ,  $\epsilon_t \sim N(0, \sigma_\epsilon^2)$  and  $\mathbb{E}(u_t \epsilon_s) = \sigma_{u\epsilon} \forall t \neq s$ .

The OLS estimator of  $\beta$  is given by  $\hat{\beta} = \left(\sum_{t=1}^T x_t^2\right)^{-1} \left(\sum_{t=1}^T y_t x_t\right)$  which implies

$$T(\hat{\beta} - \beta) = \left(T^{-2} \sum_{t=1}^T x_t^2\right)^{-1} \left(T^{-1} \sum_{t=1}^T x_t u_t\right) \tag{4.3}$$

Thus, for this regression model due to the presence of the integrated regressor the limit is expressed as

$$\left(\frac{1}{T^2} \sum_{t=1}^T x_t^2\right) \Rightarrow \sigma_\epsilon^2 \int_0^1 W_\epsilon(r)^2 dr. \tag{4.4}$$

---

<sup>3</sup>Standard regularity conditions for estimation and inference in nonstationarity time series models, is the asymptotic theory developed for nearly unstable autoregressive processes. The limit theory of time series models such as the asymptotic inference of AR(1) processes first examined by Mann and Wald (1943) has been extended to the non-stationary asymptotic case as it captured via the local to unity framework which allows for the autoregressive coefficient of a univariate AR(1) to be expressed in the form  $\rho = 1 + c/n$ , where  $n$  the sample size and  $c$  the unknown degree of persistence of the data generating process.

Next, in order to derive the limiting distribution of the model parameter for the case of integrated regressors, we first consider the limiting distribution of the sample moment  $\left(T^{-1} \sum_{t=1}^T x_t u_t\right)$ . To do this, we assume the existence of a conditional distribution for  $u_t$  given  $\epsilon_t$ , which implies the conditional mean form

$$u_t = \phi \epsilon_t + v_t, \quad \phi = \frac{\sigma_{u\epsilon}}{\sigma_\epsilon^2} \quad \text{and} \quad \sigma_v^2 = \sigma_u^2 - \frac{\sigma_{u\epsilon}^2}{\sigma_\epsilon^2} \quad (4.5)$$

Furthermore, we define with  $W_\epsilon(r)$  and  $W_v(r)$  to be two independent Wiener processes on  $\mathcal{C}[0, 1]$ . Therefore,

$$\left(T^{-1} \sum_{t=1}^T x_t u_t\right) = T^{-1} \sum_{t=1}^T x_t (\phi \epsilon_t + v_t) = \phi \left(T^{-1} \sum_{t=1}^T x_t \epsilon_t\right) + \left(T^{-1} \sum_{t=1}^T x_t v_t\right) \quad (4.6)$$

Substituting  $x_t = x_{t-1} + \epsilon_t$  into the above expression we obtain

$$\begin{aligned} \left(T^{-1} \sum_{t=1}^T x_t u_t\right) &= \phi \left(T^{-1} \sum_{t=1}^T (x_{t-1} + \epsilon_t) \epsilon_t\right) + \left(T^{-1} \sum_{t=1}^T (x_{t-1} + \epsilon_t) v_t\right) \\ &= \phi \left(T^{-1} \sum_{t=1}^T x_{t-1} \epsilon_t\right) + \phi \left(T^{-1} \sum_{t=1}^T \epsilon_t^2\right) \\ &\quad + \left(T^{-1} \sum_{t=1}^T x_{t-1} v_t\right) + \left(T^{-1} \sum_{t=1}^T \epsilon_t v_t\right) \end{aligned}$$

Most importantly, within this setting the following asymptotic results hold

$$T^{-1} \sum_{t=1}^T \epsilon_t^2 \xrightarrow{p} \sigma_\epsilon^2, \quad T^{-1} \sum_{t=1}^T \epsilon_t v_t \xrightarrow{p} 0 \quad (4.7)$$

and it has been also proved in the seminal study of [Phillips \(1987\)](#) (see, also [Phillips and Durlauf \(1986\)](#))

$$T^{-1} \sum_{t=1}^T x_{t-1} v_t \Rightarrow \sigma_\epsilon \sigma_v \int_0^1 W_\epsilon(r) dW_v(r) \equiv \int_0^1 B_\epsilon(r) dB_v(r) \quad (4.8)$$

$$T^{-1} \sum_{t=1}^T x_{t-1} \epsilon_t \Rightarrow \frac{\sigma_\epsilon^2}{2} \left[W_\epsilon^2(1) - 1\right] \quad (4.9)$$

Putting the above together we obtain that

$$T^{-1} \sum_{t=1}^T x_t u_t \Rightarrow \left\{ \phi \left( \frac{\sigma_\epsilon^2}{2} \left[W_\epsilon^2(1) - 1\right] \right) + \phi \sigma_\epsilon^2 + \sigma_\epsilon \sigma_v \int_0^1 W_\epsilon(r) dW_v(r) \right\} \quad (4.10)$$

Furthermore, [Phillips and Park \(1988\)](#) proved the following limit result

$$\int_0^1 W_\epsilon(r) dW_v(r) \Rightarrow \mathcal{N} \left( 0, \int_0^1 W_\epsilon(r) dr \right) \quad (4.11)$$

Therefore, under the null hypothesis  $\mathbb{H}_0 : \beta = 0$ , it follows that

$$T\hat{\beta} \Rightarrow \left\{ \phi \frac{\sigma_\epsilon^2}{2} \left[W_\epsilon^2(1) + 1\right] + \sigma_\epsilon \sigma_v \left[ \int_0^1 W_\epsilon(r) dW_v(r) \right] \right\} \left( \sigma_\epsilon^2 \int_0^1 W_\epsilon(r)^2 dr \right)^{-1}. \quad (4.12)$$

Thus, the  $t$ -statistic, denoted as  $\mathcal{T}_{\beta=0}$  for testing the null hypothesis,  $\mathbb{H}_0 : \beta = 0$ , is written as

$$\mathcal{T}_{\beta=0} = \frac{\widehat{\beta}}{\widehat{\sigma}_u^2 \left( \sum_{t=1}^T x_t^2 \right)^{-\frac{1}{2}}} = T \frac{\widehat{\beta}}{\widehat{\sigma}_u^2 \left( T^{-2} \sum_{t=1}^T x_t^2 \right)^{-\frac{1}{2}}} \quad (4.13)$$

has the following asymptotic distribution

$$\mathcal{T}_{\beta=0} \Rightarrow \left\{ \phi \frac{\sigma_\epsilon^2}{2} \left[ W_\epsilon^2(1) + 1 \right] + \sigma_\epsilon \sigma_v \left[ \int_0^1 W_\epsilon(r) dW_v(r) \right] \right\} \left( \sigma_\epsilon^2 \int_0^1 W_\epsilon(r)^2 dr \right)^{-\frac{1}{2}} \times \frac{1}{\sigma_u^2} \quad (4.14)$$

$$\equiv \frac{\phi}{2} \frac{\sigma_\epsilon}{\sigma_u} \left[ W_\epsilon^2(1) + 1 \right] \left( \int_0^1 W_\epsilon(r)^2 dr \right)^{-\frac{1}{2}} + \frac{\sigma_v}{\sigma_u} \mathcal{N}(0, 1). \quad (4.15)$$

**Remark 7.** Notice that the derived limiting distribution of the Student- $t$  statistic indicates that the  $t$ -ratio of  $\widehat{\beta}$  does not follow a standard normal distribution unless  $\phi = 0$ ; in which case the structure of the model implies that the regressor  $x_t$  is exogenous for the estimation of the model parameter  $\beta$  which is the main parameter of interest for inference purposes. In particular, when  $\phi \neq 0$  then the first term of the above limiting distribution gives rise to second-order or endogeneity bias, which although asymptotically negligible in estimating  $\beta$  due to super consistency, can appear in finite-samples causing size distortions when obtaining the empirical size of the test.

#### 4.1.2 Predictive Regression Model

Predictive regression models are extensively used in time series econometrics and the empirical finance literature for examining the stock return predictability puzzle as proposed by [Campbell and Yogo \(2006\)](#). A standard predictive regression has the following econometric specification (see, [Kostakis et al. \(2015\)](#))

$$y_t = \mu + \beta x_{t-1} + \epsilon_t, \quad t = 1, \dots, T \quad (4.16)$$

$$x_t = \rho x_{t-1} + u_t \quad (4.17)$$

The innovation sequence  $(\epsilon_t, u_t)$  is generated such that  $(\epsilon_t, u_t) \sim_{i.i.d} \mathcal{N}(0, \Sigma)$  where  $\Sigma = \begin{bmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon u} \\ \sigma_{\epsilon u} & \sigma_u^2 \end{bmatrix}$ .

Similarly to the previous example, the null hypothesis of interest using linear restrictions on the model parameter  $\beta$ , is formulated such that  $\mathbb{H}_0 : \beta = 0$ . The main econometric challenges when conducting statistical inference using the predictive regression model includes the problem of embedded endogeneity due to the innovation structure of the system as well as the nuisance parameter of persistence,  $c$ , when the autocorrelation coefficient of the model is expressed with the local-to-unity specification. As a result, depending on the value of the autocorrelation coefficient, the asymptotics for the parameter of the predictive regression model take a different form which makes statistical inference challenging. Specifically, when  $|\rho| < 1$ , then  $x_t$  is known to be stationary, when  $\rho = 1$  then  $x_t$  is unit root or integrated and when  $c < 0$  is assumed to follow a local-to-unity or nearly integrated process. The literature has proposed various methodologies for conducting statistical inference robust to the nuisance parameter of persistence. For instance, [Phillips and Magdalinos \(2007\)](#) study the limit theory of time series models which includes regressors that are close to the unit root boundary<sup>4</sup>.

<sup>4</sup>The authors consider the limit distribution theory in both the near-stationary ( $c < 0$ ) and the near-explosive cases ( $c > 0$ ).



In terms of the asymptotic theory that correspond to the predictive regression model, we consider the partial-sum process for the integrated regressor. The particular aspect is important especially in comparison to when constructing residual-based statistics in which case the main quantity of interest is the partial-sum process of the residuals corresponding to stationary innovations. Denote with  $\mathcal{F}_{T,t-1}$  the  $\sigma$ -algebra generated by the random variables and with  $X_{\lfloor Tr \rfloor}$  the partial-sum process of interest. Then, under the assumption that the  $x_t$  is generated as a local-unit-root process the weakly convergence of the partial-sum functional corresponds to a uniform convergence to an *Ornstein-Uhlenbeck* (OU) process<sup>5</sup> rather to the standard Wiener process (see, [Phillips \(1987\)](#) and related limit theory in [Durrett and Resnick \(1978\)](#)) as

$$\frac{1}{\sqrt{T}}X_{\lfloor Tr \rfloor} \Rightarrow J_c(r) := \int_0^r e^{(r-s)c} dB_c(s), \quad r \in [0, 1]. \quad (4.18)$$

In other words, the assumptions we impose regarding the parametrization of the autocorrelation coefficient  $\rho$  can change the asymptotic behaviour of the stochastic difference equation. Generally, statistical inference is nonstandard in the sense that when  $\rho_T = (1 + \frac{c}{T})$  for some nuisance parameter  $c$ , then the testing problem concerning the parameter  $\beta$  exhibit nonstandard large-sample properties under local-to-unity asymptotics.

### 4.1.3 Predictive tests

To provide some further clarity regarding the effect of expressing the autocorrelation coefficient in terms of moderate deviations from unity, to the validity of conventional inference methods, we consider as an example the stationary autoregressive model AR(1),  $y_t = \rho_T y_{t-1} + u_t$  where  $u_t \stackrel{i.i.d}{\sim} (0, \sigma^2)$  and  $|\rho| < 1$ . In this case, it is a well-known fact that the limit distribution of the t-test for testing the null hypothesis  $\mathbb{H}_0 : \rho_T = 0$  with  $\mathcal{T}_T(\rho_T) = \frac{\hat{\rho}_T - \rho}{\hat{\sigma}}$  converges to a standard normal distribution. In addition, the asymptotic distribution of  $\mathcal{T}_T(\rho_T)$  is invariant even under the assumption of conditional heteroscedasticity which implies  $\mathbb{E}(u_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2$  and  $\sup_{t \in \mathbb{Z}} |\hat{\sigma}_t^2 - \sigma_t^2| = o_p(1)$ , a condition for consistent estimation.

On the other hand, the limiting distribution of the model parameter  $\beta$  of the predictive regression model as well as the associated t-test for testing the null hypothesis,  $\mathbb{H}_0 : \beta = 0$ , appears to be challenging due to the fact that it is found to be nonstandard and the corresponding t-test is non-pivotal since it depends on the nuisance parameter  $c$  (see, [Cavanagh et al. \(1995\)](#) and [Campbell and Yogo \(2006\)](#)). Consequently, given the focus of our study to the asymptotic behaviour of partial-sum processes when constructing test statistics in nonstationary time series models, we illustrate the related asymptotic theory with some examples.

It is worth mentioning that the partial-sum processes,  $X_{\lfloor Tr \rfloor}(r)$ , are considered to be (maximally) invariant with respect to the presence of the model intercept  $\mu$ . Therefore under the null hypothesis,  $\mathbb{H}_0 : \beta = 0$ , joint weak convergence of observation processes to their Brownian motion counterparts holds. Specifically, an application of the invariance principle proposed by [Phillips \(1987\)](#) such that  $\frac{x_{\lfloor Tr \rfloor}}{\sqrt{T}} \Rightarrow J_c(r)$ , where  $J_c(r) = \int_0^r e^{(r-s)c} dB_c(s)$  is a standard Ornstein-Uhlenbeck process, implies that

$$\frac{1}{T\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} x_t \Rightarrow \int_0^r J_c(s) ds, \quad (4.19)$$

$$\frac{1}{T} \sum_{t=1}^{\lfloor Tr \rfloor} x_t \epsilon_t \Rightarrow \int_0^r J_c(s) dB_\epsilon(s). \quad (4.20)$$

---

<sup>5</sup>The continuous time OU diffusion process given by  $dy_t = \theta y_t dt + \sigma dw_t$ ,  $y_0 = b, t > 0$ , where  $\theta$  and  $\sigma > 0$  are unknown parameters and  $w_t$  is the standard Wiener process, has a unique solution to  $\{y_t\}$  which is expressed as  $y_t = \exp(\theta t) b + \sigma \int_0^t \exp[\theta(t-s)] dw_s \equiv \exp(\theta t) + \sigma J_\theta(t)$  (see, e.g., ([Perron, 1991](#))).

Since  $T(\widehat{\beta}_T - \beta) = \mathcal{O}_p(1)$  is bounded in probability, then we can establish the usual mode of converges in distribution in the same probability space such that

$$T(\widehat{\beta}_T - \beta) = \frac{\frac{1}{T} \sum_{t=1}^T x_{t-1} \epsilon_t}{\frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2} \Rightarrow \frac{\int_0^1 J_c(s) dB_\epsilon(s)}{\int_0^1 J_x^2(s) ds}. \quad (4.21)$$

Consider the stationary case such that  $|\rho| < 1$ , then by partitioning the covariance matrix  $\Sigma$  similar to the regression model with the integrated regressor, we use the decomposition  $\epsilon_{1,2t} = \epsilon_t - \frac{\sigma_{\epsilon u}}{\sigma_\epsilon} u_t$  which implies

$$T(\widehat{\beta}_T - \beta) = \frac{\frac{1}{T} \sum_{t=1}^T x_{t-1} \epsilon_t}{\frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2} = \frac{\frac{1}{T} \sum_{t=1}^T x_{t-1} \epsilon_{1,2t}}{\frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2} + \frac{\sigma_{\epsilon u}}{\sigma_\epsilon} \frac{\frac{1}{T} \sum_{t=1}^T x_{t-1} u_t}{\frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2} \quad (4.22)$$

The first term of expression (4.22) since it includes a conditional error term then the corresponding limit distribution converges to a mixed normal limit. Moreover, we consider the joint convergence of the martingale sequences  $\{\sum_{t=1}^n x_{t-1} \epsilon_t\}$  and  $\{\sum_{t=1}^n u_t\}$  are defined on the same probability space

$$\xi_{Tt} = \left( \frac{1}{T} x_{t-1} \epsilon_t, \frac{1}{\sqrt{T}} u_t \right)' \quad (4.23)$$

Thus, the conditional covariance matrix of the martingale vector  $\xi_{Tt}$  is given by

$$\sum_{t=1}^T \mathbb{E} \left( \xi_{Tt} \xi_{Tt}' | \mathcal{F}_{t-1} \right) = \begin{bmatrix} \left( \frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2 \right) \sigma_\epsilon^2 & \left( \frac{1}{T\sqrt{T}} \sum_{t=1}^T x_{t-1} \right) \sigma_{\epsilon u} \\ \left( \frac{1}{T\sqrt{T}} \sum_{t=1}^T x_{t-1} \right) \sigma_{u\epsilon} & \sigma_{uu} \end{bmatrix} \quad (4.24)$$

Using the partition matrix identity,  $\Sigma_{1,2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ , to the predictive regression model we obtain the relation  $\frac{\sigma_{1,2}}{\sigma_\epsilon} = 1 - \frac{\sigma_{u\epsilon}}{\sigma_\epsilon \sigma_u}$ . Therefore, the following mixed normal limit convergence holds

$$\frac{\frac{1}{T} \sum_{t=1}^T x_{t-1} \epsilon_{1,2t}}{\frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2} \Rightarrow \mathcal{MN} \left( 0, \sigma_{1,2} \left( \int_0^1 J_c(r) dr \right)^{-1} \right) \quad (4.25)$$

Hence, using expressions (4.22) and (4.25) and the limit result below

$$T\widehat{\sigma}_\beta = \widehat{\sigma}_\epsilon \left( \frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2 \right)^{-1/2} \Rightarrow \sigma_\epsilon \left( \int_0^1 J_c^2(r) dr \right)^{-1} \quad (4.26)$$

for some  $r \in (0, 1)$ , we obtain an analytical expression for the asymptotic distribution of the t-statistic  $\mathcal{T}_T(\beta_T) = \frac{\widehat{\beta}_T - \beta}{\widehat{\sigma}_\beta} \Rightarrow \phi \widehat{\mathcal{M}}(c) + (1 - \phi^2)^{1/2} \mathcal{Z}$ , where  $\mathcal{Z} \sim \mathcal{N}(0, 1)$  is independent of the random quantity  $\widehat{\mathcal{M}}(c)$  and  $c$  denotes the nuisance parameter of persistence.

More precisely, it holds that

$$\mathcal{T}_T(\beta_T) = \frac{\widehat{\beta}_T - \beta}{\widehat{\sigma}_\beta} \Rightarrow \frac{1}{\sigma_\epsilon \left( \int_0^1 J_x^2(r) dr \right)^{-1}} \left\{ \mathcal{MN} \left( 0, \sigma_{1.2} \left( \int_0^1 J_c(r) dr \right)^{-1} \right) + \frac{\sigma_{\epsilon u} \int_0^1 J_c(s) dB_x(s)}{\sigma_\epsilon \int_0^1 J_c^2(s) ds} \right\},$$

where  $\phi = \frac{\sigma_{\epsilon u}}{\sigma_\epsilon \sigma_u}$ , and  $\widehat{\mathcal{M}}(c) = \frac{\int_0^1 J_c(s) dB_x(s)}{\sigma_u^2 \int_0^1 J_c^2(s) ds}$ .

In summary, the t-statistic for the predictive regression coefficient has been proved to have a non-standard limiting distribution which implies that normal or chi-square based inference is not available in practise, due to the endogeneity problem as well as the existence of persistence regressors. Therefore, the particular non-standard testing problem makes it difficult to conduct inference without prior knowledge regarding the exact value of the coefficient of persistence and in practise cannot be consistently estimated. Suggested solutions to overcome this problem include the Bonferroni confidence interval proposed by [Cavanagh et al. \(1995\)](#) and [Elliott et al. \(1996\)](#), the conditional likelihood approach that uses sufficient statistics proposed by [Jansson and Moreira \(2006\)](#) and the control function approach proposed by [Elliott \(2011\)](#).

## 4.2 Structural Break Testing

Setting against the background described in details in Section 4.1 we now discuss the main challenges for the development of the structural break testing framework.

### 4.2.1 Asymptotic Distributions of Test Statistics

In this section we consider the implementation of OLS-CUSUM and Wald-type statistics within the local-to-unity framework when detecting instabilities in the parameters of a predictive regression model with persistent regressors. Although in this paper we consider an in-sample monitoring scheme, under suitable modifications an on-line (sequential) monitoring scheme as proposed by [Chu et al. \(1996\)](#) can provide an early warning mechanism for risk management purposes based on macroeconomic and financial conditions. The implementation of such a framework can be interpreted as a dynamic methodology for testing for parameter instability under the assumption of time-varying persistence properties.

Therefore, we are interested in proposing suitable testing methodologies for detecting structural change in the vector of regression coefficients  $\theta$  of the following time series regression model

$$y_t = x'_{t-1} \beta_1 \mathbf{1}\{t \leq k\} + x'_{t-1} \beta_2 \mathbf{1}\{t > k\} + \epsilon_t, \quad t = 1, \dots, T, \quad (4.27)$$

where  $x_t = R_T x_{t-1} + u_t$  with  $R_T = \left(1 - \frac{C}{T}\right)$  and  $C = \text{diag}\{c_1, \dots, c_p\}$  (see, [Kostakis et al. \(2015\)](#)). Under the null hypothesis of no structural break  $\mathbb{H}_0 : \beta_1 = \beta_2$ . Our proposition aims to incorporate neglected non-linearities such as structural breaks to the current LUR framework. For example, certain non-linear functions<sup>6</sup> of  $I(1)$  processes can wrongly behave like stationary long memory processes (see, e.g., [Kasparis et al. \(2014\)](#)). A recent approach which considers structural breaks under such conditions is presented by [Berenguer-Rico and Nielsen \(2020\)](#). In this paper, we consider the weak dependence assumption.

<sup>6</sup>For example, [Wang et al. \(2012\)](#) consider a specification test for nonlinear nonstationary models within the LUR framework of cointegrating regression system.

### 4.2.2 OLS-CUSUM test statistic

Next, we focus on the limit theory of the residual-based statistic, CUSUM test, constructed using the OLS residuals of the predictive regression under the null hypothesis,  $\mathbb{H}_0 : \beta_1 = \beta_2$ . The OLS residuals of the predictive regression are given by  $\hat{\epsilon}_t^{ols} = y_t - x_t' \hat{\beta}^{ols}$  and the corresponding OLS-CUSUM statistic is expressed in the usual way as below for  $s \in [0, 1]$

$$\tilde{\mathcal{C}}_T(k) = \max_{0 \leq k \leq T} \frac{\sum_{t=1}^k \hat{\epsilon}_t^{ols} - \frac{k}{T} \sum_{t=1}^T \hat{\epsilon}_t^{ols}}{\hat{\sigma}_\epsilon \sqrt{T}} = \sup_{0 \leq r \leq 1} \frac{\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{\epsilon}_t^{ols} - \frac{r}{\sqrt{T}} \sum_{t=1}^T \hat{\epsilon}_t^{ols}}{\hat{\sigma}_\epsilon \sqrt{T}} \quad (4.28)$$

The OLS residuals of the predictive regression can be expressed as  $\hat{\epsilon}_t^{ols} = y_t - x_t' \hat{\beta}^{ols} \equiv \epsilon_t - x_t' (\hat{\beta}_T - \beta)$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{\epsilon}_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \epsilon_t - T (\hat{\beta}_T - \beta) \frac{1}{T\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} x_t \quad (4.29)$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{\epsilon}_t \Rightarrow_d B_\epsilon(r) - \left\{ \frac{\int_0^r J_c(s) dB_u(s)}{\int_0^1 J_c^2(s) ds} \right\} \times \int_0^r J_c(s) ds. \quad (4.30)$$

Therefore, using expression (4.28) the OLS-CUSUM statistic within the LUR framework becomes

$$\mathcal{C}_T(k) = \frac{1}{\hat{\sigma}_\epsilon} \left\{ \left( B_u(r) - \frac{\int_0^r J_c(s) dB_u(s)}{\int_0^1 J_c^2(s) ds} \int_0^r J_c(s) ds \right) - r \left( B_u(1) - \frac{\int_0^r J_c(s) dB_u(s)}{\int_0^1 J_c^2(s) ds} \int_0^1 J_c(s) ds \right) \right\}$$

$$\mathcal{C}_T(k) = \frac{1}{\hat{\sigma}_\epsilon} \left\{ \left[ B_u(r) - r B_u(1) \right] - \frac{\int_0^r J_c(s) dB_u(s)}{\int_0^1 J_c^2(s) ds} \times \left( \int_0^r J_c(s) ds - r \int_0^1 J_c(s) ds \right) \right\}. \quad (4.31)$$

As we can clearly observe from the second term of the above expression that corresponds to the limiting distribution of the OLS-CUSUM statistic in a predictive regression model with persistent regressors the dependence on the nuisance parameter of persistence,  $c$ , makes the limit result non-standard and non-pivotal. In other words, the implementation of a residual-based statistic in a predictive regression model using OLS residuals is considered to be problematic in the derivation of the asymptotic distribution due to its dependence on the nuisance degree of persistence of the autoregressive specification of the model. Furthermore, for  $k = \lfloor Tr \rfloor$  we define the following term for notation simplicity

$$\tilde{\mathcal{J}}_\infty(c; r) := \frac{\int_0^r J_c(s) dB_u(s)}{\int_0^1 J_c^2(s) ds} \times \int_0^r J_c(s) ds. \quad (4.32)$$

Then, the weakly convergence result can be written as below

$$\mathcal{C}_T(k) \Rightarrow \left( \left[ W(r) - rW(1) \right] - \left[ \tilde{\mathcal{J}}(c; r) - r\tilde{\mathcal{J}}(c, 1) \right] \right). \quad (4.33)$$

In summary, the weakly convergence of the in-sample OLS-CUSUM statistic includes the component  $\tilde{J}_\infty(c, r)$  which depends on the nuisance parameter  $c$  and thus can affect the true size of the test under the null hypothesis of no parameter instability given the fact that we cannot consistently estimate the coefficient of persistence. However, for example the term  $\int_0^r J_c(s)ds - r \int_0^1 J_c(s)ds$  it is likely to be quite small and therefore can be considered not to be contributing to huge size distortions. An extensive Monte Carlo study can shed light on the particular aspect of the proposed test for detecting structural change in predictive regression models with persistent regressors. Thus, we have demonstrated that when testing for a structural break in linear time series regression models, the partial-sum processes of conventional test such as those of residual-based and Wald-type statistics have different properties when information regarding the integration order of regressors is available in the form of the LUR specification form.

## 5 CONCLUSION

In this paper we establish the Brownian Bridge limiting distributions in a fairly standard settings, that is linear regression models under the assumption of stationarity and ergodicity when constructing residual-based and Wald-type statistics for testing the null hypothesis of no parameter instability. In particular, in all those cases we have demonstrated that the normalized Brownian bridge limit holds for both test statistics. Additionally, we investigate whether this property also holds for nonstationary time series regression models with integrated or persistent regressors. Our asymptotic theory analysis has demonstrated that while in the classical linear regression model with stationary regressors the convergence of the test statistics to brownian bridge limit results hold, in the case of the nonstationary time series model it appears to be the case that the limiting distribution is non-standard and non-pivotal due to the dependence of the distribution to the nuisance parameter of persistence.

Based on the general assumption that the underline stochastic processes are mean-reverting, then we can establish adequate approximations to finite-sample moments for the model under consideration regardless of the econometric environment operates under the assumption of stationarity or we consider the settings of a nonstationary time series model. On the other hand, since the main feature of nonstationary time series models is the parametrization of the autocorrelation coefficient with respect to the nuisance parameter of persistence, this implies that limiting distributions are non-standard and non-pivotal which makes inference difficult. In particular, despite the large availability of macroeconomic and financial variables which can be included as regressors in predictive regressions (such as financial ratios, diffusion indices, fundamentals), practitioners have no prior knowledge regarding the persistence properties of predictors so conventional estimation and inference methods for model parameters, such as predictability tests, forecast evaluation tests as well as structural break testing require to handle the nuisance parameter of persistent.

Therefore, further research is needed to propose suitable statistical methodologies that take into consideration these challenges, especially when testing for the presence of a structural break in nonstationary time series models. Practically, the extension of structural break tests in nonstationary time series models, such as predictive regressions, which are particularly useful when information regarding the time series properties of regressors is not lost by taking the first difference for instance, is crucial for both theoretical and empirical studies. In a subsequent paper, we propose a formal econometric framework and develop the associated asymptotic theory for Wald-type statistics under regressors nonstationarity.

## 6 APPENDIX

### 6.1 Appendix A: Technical Proofs

#### Proof of Theorem 2

*Proof.* Consider the functional  $\hat{\sigma}_z^2 \mathcal{Z}_T(s)$  given by

$$\hat{\sigma}_z^2 \mathcal{Z}_T(s) = T (\bar{y}_1 - \bar{y}_2)^2 = T \left( \frac{1}{[Ts]} \sum_{t=1}^{[Ts]} y_t - \frac{1}{T - [Ts]} \sum_{t=[Ts]+1}^T y_t \right)^2.$$

Under the null hypothesis  $\mathbb{H}_0 : \theta_1 = \theta_2$ , we have that the following expression holds

$$\begin{aligned} (\bar{y}_1 - \bar{y}_2) &= \left( \frac{1}{k} \sum_{t=1}^k \epsilon_t - \frac{1}{T-k} \sum_{t=k+1}^T \epsilon_t \right) = \frac{1}{k} \sum_{t=1}^k \epsilon_t - \frac{1}{T-k} \left( \sum_{t=1}^T \epsilon_t - \sum_{t=1}^k \epsilon_t \right) \\ &= \frac{T\sqrt{T}}{k(T-k)} \left( \sum_{t=1}^k \frac{\epsilon_t}{\sqrt{T}} - \frac{k}{T} \sum_{t=1}^T \frac{\epsilon_t}{\sqrt{T}} \right) \end{aligned}$$

Therefore we obtain that

$$\hat{\sigma}_z^2 \mathcal{Z}_T(s) = T (\bar{y}_1 - \bar{y}_2)^2 = \frac{1}{\left(\frac{k}{T}\right)^2 \left(1 - \frac{k}{T}\right)^2} \left( \sum_{t=1}^k \frac{\epsilon_t}{\sqrt{T}} - \frac{k}{T} \sum_{t=1}^T \frac{\epsilon_t}{\sqrt{T}} \right)^2. \quad (6.1)$$

Thus, equality (6.1) shows that  $\hat{\sigma}_z^2 \mathcal{Z}_T(s) \stackrel{\text{plim}}{=} \frac{1}{s^2(1-s)^2} \hat{\sigma}_\epsilon^2 [\mathcal{Z}_T^o(s)]^2$ , due to the continuous mapping theorem. Since  $\hat{\sigma}_\epsilon^2 = \hat{\sigma}_z^2 \times s(1-s)$  and using the weakly convergence result for  $\mathcal{Z}_T^o(s)$  deduced in Example 2 (see, [Aue and Horváth \(2013\)](#)), which implies that  $\mathcal{Z}_T^o(s) \Rightarrow \sup_{s \in [\nu, 1-\nu]} \mathcal{B}\mathcal{B}(s)$  gives the following asymptotic result for the test statistic  $\mathcal{Z}_T(s)$  provided that  $0 < \nu < 1$

$$\mathcal{Z}_T(s) \Rightarrow \sup_{s \in [\nu, 1-\nu]} \frac{[W(s) - sW(1)]^2}{s(1-s)}. \quad (6.2)$$

□

#### Proof of Theorem 3

*Proof.* Let  $\hat{\theta}_1$  and  $\hat{\theta}_2$  be the OLS estimator for  $\{1, \dots, k\}$  and  $\{k+1, \dots, T\}$  respectively, then it follows that

$$\begin{aligned} \hat{\theta}_1 &= \left( \frac{1}{T} \sum_{t=1}^k x_{1,t} x'_{1,t} \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^k x_{1,t} y_t \right) \\ \hat{\theta}_2 &= \left( \frac{1}{T} \sum_{t=k+1}^T x_{2,t} x_{2,t}' \right)^{-1} \left( \frac{1}{T} \sum_{t=k+1}^T x'_{2,t} y_t \right) \end{aligned}$$

Using the restriction matrix and the coefficient matrix we formulate the Wald statistic as below

$$\mathcal{W}_T(s) = \frac{1}{\hat{\sigma}_\epsilon^2} (\hat{\theta}_1 - \hat{\theta}_1)' \left[ (X_1' X_1)^{-1} + (X_2' X_2)^{-1} \right]^{-1} (\hat{\theta}_1 - \hat{\theta}_1)$$

Let  $A$  and  $B$  be matrices whose dimensions are such that the addition  $(A + B)$  is well-defined, we can use the binomial matrix identity

$$(A + B)^{-1} = A^{-1} - A^{-1} (B^{-1} + A^{-1})^{-1} A^{-1}$$

and since the following property holds due to the orthogonality condition (and no model intercept)

$$(X' X) = (X_1' X_1) + (X_2' X_2)$$

then we get that

$$\left[ (X_1' X_1)^{-1} + (X_2' X_2)^{-1} \right] = \left[ X_1' X_1 - X_1' X_1 (X' X)^{-1} X_1' X_1 \right]$$

Therefore, keeping the matrix notation for simplicity the Wald test is expressed as below

$$\mathcal{W}_T(s) = \frac{1}{\hat{\sigma}_\epsilon^2} (\hat{\theta}_1 - \hat{\theta}_1)' \left[ X_1' X_1 - X_1' X_1 (X' X)^{-1} X_1' X_1 \right]^{-1} (\hat{\theta}_1 - \hat{\theta}_1)$$

which according to [Pitarakis \(2008\)](#) has an equivalent representation given by

$$\mathcal{W}_T(s) = \frac{1}{\hat{\sigma}_\epsilon^2} \left[ \epsilon' X_1 - \epsilon' X (X' X)^{-1} X_1' X_1 \right] \left[ X_1' X_1 - X_1' X_1 (X' X)^{-1} X_1' X_1 \right]^{-1} \left[ X_1' \epsilon - X_1' X_1 (X' X)^{-1} X_1' \epsilon \right]$$

Then, the econometric specification of the model with when testing for a single structural break at an unknown break- point  $k = \lfloor Ts \rfloor, s \in (0, 1)$  is given by

$$y = X_1 \theta_1 + X_2 \theta_2 + \epsilon \stackrel{\mathbb{H}_0}{=} X \theta_1 + \epsilon \quad (6.3)$$

Then, the OLS estimators are expressed in matrix form as below

$$\hat{\theta}_1 = (X_1' X_1)^{-1} X_1' y \quad \text{and} \quad \hat{\theta}_2 = (X_2' X_2)^{-1} X_2' y$$

Under  $\mathbb{H}_0 : \theta_1 = \theta_2$ , we have that  $y = X_1 \theta_1 + X_2 \theta_1 + \epsilon = X \theta_1$ , where  $X = X_1 + X_2$  since  $X_1 \perp X_2$ . Thus,  $X_1' X = X_1' X_1$  and  $X_1' X_2 = X_2' X_1 = 0$  due to the orthogonality property. Therefore, the OLS estimators are written in the following form

$$\begin{aligned} \hat{\theta}_1 &= (X_1' X_1)^{-1} X_1' (X \theta_1 + \epsilon) = \theta_1 + (X_1' X_1)^{-1} (X_1' X_2) \theta_1 + (X_1' X_1)^{-1} X_1' \epsilon \\ \hat{\theta}_2 &= (X_2' X_2)^{-1} X_2' (X \theta_1 + \epsilon) = \theta_1 + (X_2' X_2)^{-1} (X_2' X_1) \theta_1 + (X_2' X_2)^{-1} X_2' \epsilon \end{aligned}$$

Consider the expression

$$\begin{aligned} \sqrt{T} (\hat{\theta}_1 - \hat{\theta}_2) &= \left[ (X_1' X_1)^{-1} X_1' \epsilon - (X_2' X_2)^{-1} X_2' \epsilon \right] \\ &\equiv \left\{ \left( \frac{1}{T} \sum_{t=1}^k x_{1,t} x_{1,t}' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^k x_{1,t} \epsilon_t - \left( \frac{1}{T} \sum_{t=k+1}^T x_{2,t} x_{2,t}' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=k+1}^T x_{2,t} \epsilon_t \right\} \end{aligned}$$



Furthermore, the following probability limits hold uniformly for  $s \in (0, 1)$ ,

$$\left( \frac{1}{T} \sum_{t=1}^{\lfloor Ts \rfloor} x_{1,t} x'_{1,t} \right) \xrightarrow{p} s \mathbf{Q}_T, \quad \left( \frac{1}{T} \sum_{t=\lfloor Ts \rfloor+1}^T x_{2,t} x'_{2,t} \right) \xrightarrow{p} (1-s) \mathbf{Q}_T, \quad \left( \frac{X'X}{T} \right) \xrightarrow{p} \mathbf{Q}_T \quad (6.4)$$

which holds due to moment homogeneity. Furthermore, using the following two limit results

Result A:

$$\left[ \left( \frac{X'_1 X_1}{T} \right)^{-1} + \left( \frac{X'_2 X_2}{T} \right)^{-1} \right]^{-1} \stackrel{\text{plim}}{=} \left[ \frac{1}{s} \mathbf{Q}_T^{-1} + \frac{1}{1-s} \mathbf{Q}_T^{-1} \right]^{-1} = \left[ \frac{\mathbf{Q}_T^{-1}}{s(1-s)} \right]^{-1} \quad (6.5)$$

Result B:

$$\frac{1}{\hat{\sigma}_\epsilon^2} \left( \frac{X'_1 \epsilon}{\sqrt{T}} - s \frac{X' \epsilon}{\sqrt{T}} \right)^2 = \left( \frac{1}{\hat{\sigma}_\epsilon \sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} x_t \epsilon_t - s \frac{1}{\hat{\sigma}_\epsilon \sqrt{T}} \sum_{t=1}^T x_t \epsilon_t \right)^2 \Rightarrow [\mathbf{W}_p(s) - s \mathbf{W}_p(1)]^2 \quad (6.6)$$

using the fact that  $X'_2 \epsilon = (X' \epsilon - X'_1 \epsilon)$ .

Therefore, we obtain the following limit result

$$\begin{aligned} \sqrt{T} (\hat{\theta}_1 - \hat{\theta}_2) &\stackrel{\text{plim}}{=} \left\{ \frac{1}{s} \mathbf{Q}_T^{-1} \sigma_\epsilon \mathbf{W}_p(s) - \frac{1}{s(1-s)} \mathbf{Q}_T^{-1} \sigma_\epsilon [\mathbf{W}_p(1) - \mathbf{W}_p(s)] \right\} \\ &= \sigma_\epsilon \mathbf{Q}_T^{-1} \left\{ \frac{1}{s} \mathbf{W}_p(s) - \frac{1}{s(1-s)} [\mathbf{W}_p(1) - \mathbf{W}_p(s)] \right\} \\ &= \sigma_\epsilon \mathbf{Q}_T^{-1} \frac{1}{s(1-s)} \left\{ \mathbf{W}_p(s) - s \mathbf{W}_p(1) \right\} \end{aligned}$$

Then, the limiting distribution of the Wald statistic can be expressed as below

$$\begin{aligned} \mathcal{W}_T(s) &\Rightarrow \frac{1}{\sigma_\epsilon^2} \frac{\sigma_\epsilon}{s(1-s)} \left\{ \mathbf{W}_p(s) - s \mathbf{W}_p(1) \right\}' \mathbf{Q}_T^{-1} \left[ \frac{\mathbf{Q}_T^{-1}}{s(1-s)} \right]^{-1} \frac{\sigma_\epsilon}{s(1-s)} \mathbf{Q}_T^{-1} \left\{ \mathbf{W}_p(s) - s \mathbf{W}_p(1) \right\} \\ \mathcal{W}_T(s) &\Rightarrow \frac{\left[ \mathbf{W}_p(s) - s \mathbf{W}_p(1) \right]' \mathbf{Q}_T^{-1} \left[ \mathbf{W}_p(s) - s \mathbf{W}_p(1) \right]}{s(1-s)} \end{aligned}$$

Therefore, the limiting distribution of the sup-Wald test follows

$$\mathcal{W}_T^*(s) := \sup_{s \in [\nu, 1-\nu]} \mathcal{W}_T(s) \Rightarrow \sup_{s \in [\nu, 1-\nu]} \frac{\left[ \mathbf{W}_p(s) - s \mathbf{W}_p(1) \right]' \mathbf{Q}_T^{-1} \left[ \mathbf{W}_p(s) - s \mathbf{W}_p(1) \right]}{s(1-s)} \quad (6.7)$$

which verifies a weakly convergence to a normalized squared Brownian bridge process.  $\square$

## References

- Andreou, E. (2008). Restoring monotone power in the cusum test. *Economics Letters*, 98(1):48–58.
- Andreou, E. and Ghysels, E. (2006). Monitoring disruptions in financial markets. *Journal of Econometrics*, 135(1-2):77–124.
- Andreou, E. and Ghysels, E. (2009). Structural breaks in financial time series. *Handbook of financial time series*, pages 839–870.
- Andreou, E. and Werker, B. J. (2012). An alternative asymptotic analysis of residual-based statistics. *Review of Economics and Statistics*, 94(1):88–99.
- Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856.
- Andrews, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734.
- Aue, A., Berkes, I., and Horváth, L. (2006). Strong approximation for the sums of squares of augmented garch sequences. *Bernoulli*, 12(4):583–608.
- Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16.
- Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric theory*, 13(3):315–352.
- Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78.
- Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. (1993). *Co-integration, error correction, and the econometric analysis of non-stationary data*. Oxford university press.
- Baron, M., Verner, E., and Xiong, W. (2021). Banking crises without panics. *The Quarterly Journal of Economics*, 136(1):51–113.
- Berenguer-Rico, V. and Nielsen, B. (2020). Cumulated sum of squares statistics for nonlinear and nonstationary regressions. *Econometric Theory*, 36(1):1–47.
- Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.
- Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(2):149–163.
- Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of financial economics*, 81(1):27–60.
- Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric theory*, 11(5):1131–1147.
- Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183.
- Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *The Annals of Mathematical Statistics*, pages 972–994.
- Chu, C.-S. J., Hornik, K., and Kaun, C.-M. (1995a). Mosum tests for parameter constancy. *Biometrika*, 82(3):603–617.
- Chu, C.-S. J., Hornik, K., and Kuan, C.-M. (1995b). The moving-estimates test for parameter stability. *Econometric Theory*, 11(4):699–720.
- Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica: Journal of the Econometric Society*, pages 1045–1065.
- Csorgo, M. and Horvath, L. (1997). Limit theorems in change-point analysis.
- Csörgő, M., Szyszkowicz, B., Wu, Q., et al. (2003). Donsker’s theorem for self-normalized partial sums processes. *The Annals of Probability*, 31(3):1228–1240.
- Deng, A. and Perron, P. (2008). The limit distribution of the cusum of squares test under general mixing conditions. *Econometric Theory*, 24(3):809–822.
- Durrett, R. and Resnick, S. I. (1978). Functional limit theorems for dependent variables. *The Annals of Probability*, pages 829–846.
- Elliott, G. (2011). A control function approach for testing the usefulness of trending variables in forecast models and linear regression. *Journal of econometrics*, 164(1):79–91.
- Elliott, G., Rothenberg, T. J., and Stock, J. (1996). Efficient tests for an autoregressive unit root. *Econometrica*,

- 64(4):813–36.
- Greenwood, R., Hanson, S. G., Shleifer, A., and Sørensen, J. A. (2020). Predictable financial crises. Technical report, National Bureau of Economic Research.
- Gronneberg, S. and Holcblat, B. (2019). On partial-sum processes of armax residuals. *The Annals of Statistics*, 47(6):3216–3243.
- Hawkins, D. (1987). A test for a change point in a parametric model based on a maximal wald-type statistic. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 368–376.
- Horváth, L., Liu, Z., Rice, G., and Wang, S. (2020). Sequential monitoring for changes from stationarity to mild non-stationarity. *Journal of Econometrics*, 215(1):209–238.
- Horváth, L., Teyssière, G., et al. (2001). Empirical process of the squared residuals of an arch sequence. *The Annals of Statistics*, 29(2):445–469.
- Jansson, M. and Moreira, M. J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, 74(3):681–714.
- Kasparis, I., Phillips, P. C., and Magdalinos, T. (2014). Nonlinearity induced weak instrumentation. *Econometric Reviews*, 33(5-6):676–712.
- Katsouris, C. (2021). Sequential break-point detection in stationary time series: An application to monitoring economic indicators. *arXiv preprint arXiv:2112.06889*.
- Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5):1506–1553.
- Krämer, W., Ploberger, W., and Alt, R. (1988). Testing for structural change in dynamic models. *Econometrica: Journal of the Econometric Society*, pages 1355–1369.
- Kuan, C.-M. and Chen, M.-Y. (1994). Implementing the fluctuation and moving-estimates tests in dynamic econometric models. *Economics Letters*, 44(3):235–239.
- Kulperger, R., Yu, H., et al. (2005). High moment partial sum processes of residuals in garch models and their applications. *The Annals of Statistics*, 33(5):2395–2422.
- Leisch, F., Hornik, K., and Kuan, C.-M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory*, 16(6):835–854.
- Mann, H. B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica, Journal of the Econometric Society*, pages 173–220.
- Müller, U. K. (2011). Efficient tests under a weak convergence assumption. *Econometrica*, 79(2):395–435.
- Perron, P. (1991). A continuous time approximation to the unstable first-order autoregressive process: the case without an intercept. *Econometrica: Journal of the Econometric Society*, pages 211–236.
- Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica: Journal of the Econometric Society*, pages 277–301.
- Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53(4):473–495.
- Phillips, P. C. B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1):115–130.
- Phillips, P. C. B. and Park, J. Y. (1988). Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association*, 83(401):111–115.
- Pitarakis, J.-Y. (2004). Least squares estimation and tests of breaks in mean and variance under misspecification. *The Econometrics Journal*, 7(1):32–54.
- Pitarakis, J.-Y. (2008). Comment on: threshold autoregressions with a unit root. *Econometrica*, 76(5):1207–1217.
- Ploberger, W. and Krämer, W. (1992). The cusum test with ols residuals. *Econometrica: Journal of the Econometric Society*, pages 271–285.
- Stock, J. H. (1994). Unit roots, structural breaks and trends. *Handbook of econometrics*, 4:2739–2841.
- Wang, Q., Phillips, P. C., et al. (2012). A specification test for nonlinear nonstationary models. *The Annals of Statistics*, 40(2):727–758.
- Wooldridge, J. M. and White, H. (1988). Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric theory*, 4(2):210–230.