

TESTING FOR STRUCTURAL STABILITY IN THE WHOLE SAMPLE

JAVIER HIDALGO AND MYUNG HWAN SEO

ABSTRACT. Testing for structural stability has attracted a lot of attention in theoretical and applied research. Oftentimes the test is based on the supremum of, for example, the Wald statistic when the break is assumed to be in the interval $[\tau n] < s < n - [\tau n]$ for some $\tau > 0$ and where n denotes the sample size. More recently there has been some work to allow the possibility that the break lies at the end of the sample, i.e. when $s \in (n - \hat{s}, n)$ for some finite number \hat{s} . However, the previous setups do not include the important intermediate case when $s \in (\hat{s}, [\tau n]) \cup (n - [\tau n], n - \hat{s})$, or more generally when we do not wish to assume any prior knowledge on the location of the break. The aim of the paper is thus to extend existing results on stability tests in the later scenario for models useful in economics such as nonlinear simultaneous equations and transformation models. Letting the time of the break to be anywhere in the sample might not only be more realistic in applied research, but it avoids also the unpleasant need to choose either τ or \hat{s} . In addition we show that, contrary to the conventional tests, the tests described and examined in the paper are consistent irrespective of the location of the break.

JEL Classification: C21, C23.

1. INTRODUCTION

Since the work of Chow (1960) and Quandt (1960), testing for structural stability has been a very active topic of theoretical and applied research. The bulk of the research has focused on the situation when the exact time of the break s is not known but the researcher assumes that s lies in the “middle” of the sample. That is, $s \in ([n\tau], n - [n\tau])$ for some trimming quantity $\tau > 0$, and where herewith n denotes the sample size. See Andrews (1993) or the latest review article by Perron (2006). In this scenario, the location of the break is often parameterized as the fraction $\pi = s/n \in (\tau, 1 - \tau)$, and it has been shown that the *supremum* of, for instance, the *Wald* (W) or *Lagrange Multiplier* (LM) statistics (denoted herewith as conventional statistics) converge to the *supremum* of a *Gaussian* process. More recently, there has been some interest in the case where the break occurs at the “end” of the sample. For example, Andrews (2003) or Andrews and Kim (2006), and references therein, examined the case when $s \in (n - \hat{s}, n)$ for some finite value \hat{s} . In this situation, we know that, although the tests are not consistent and their distributions depend on \hat{s} , it is still possible to make inferences as it was shown by Andrews (2003). What has not been studied is the case when $s \in (\hat{s}, [n\tau])$ or $(n - [n\tau], n - \hat{s})$, or more importantly, when we do not wish to impose any prior knowledge on the location of the break, avoiding the need to choose τ or \hat{s} .

The paper thus considers the problem of testing for structural stability over the whole sample span $s = 1, \dots, n$. That is, when no previous information about the location of the break is available. The test does not involve a trimming quantity

Date: 11 November 2008.

Key words and phrases. Structural Stability. Nonlinear and transformation models. GMM estimation. Strong approximation. Extreme value distributions. Ornstein-Uhlenbeck process.

and it can be applied to models useful in economics, such as nonlinear simultaneous equations and transformation models under general conditions on the dependence structure of the variables of the model. In particular, we do not need to assume that the data, for instance the regressors and error term in a regression model, are covariance stationary. In this way, we substantially extend Horvath's (1993) results who only examines this problem for the mean of otherwise independent and normally distributed random variables.

In our setup, Andrews (1993) showed that the conventional tests diverge to infinity without trimming, signalling that the trimming was not only imposed for technical convenience but it was also crucial to obtain a proper asymptotic distribution. We show that the reason for his finding is because the normalization that we require for the statistic is different to that for the conventional tests. We also show that the asymptotic distribution of our tests is different to that of the conventional test. More specifically, we show that, after appropriate normalization, which is only a simple function of the sample size n and the number of parameters subject to break, the $\sup_{s=1,\dots,n} W(s)$ and $\sup_{s=1,\dots,n} LM(s)$ test statistics converge to the Type I *Extreme Value Distribution* or *Gumbel* distribution.

It is also worth mentioning that as Andrews and Ploberger (1994) discussed, to obtain their optimality results, we need to be away from (or not too close to) the beginning or the end of the sample. Indeed, the Monte-Carlo experiment in Section 4 suggests that when we compare the power of the conventional tests against the power of our tests in Sections 2 and 3, the assumptions made in Andrews and Ploberger (1994) were not innocuous. More specifically, as Section 2.3 shows, the conventional tests are not consistent when the break occurs at time $s \leq n^{1/2}$ or $n - n^{1/2} < s$, whereas our tests are always consistent irrespective of the location of the break. In addition, we show that when the break falls in the region $s \in (n^{1/2}, n/(\log \log n)^{1/2})$, the conventional tests have zero asymptotic relative efficiency compared to ours, in the sense that our tests are able to detect local alternatives that, for instance, the "optimal conventional tests" would not do. It is worth mentioning that our tests are similar as that of Brown, Durbin and Evans (1975), in the sense that they do not trim. However, the latter work suffers from the same lack of power just described for the conventional tests.

We finish this section discussing some theoretical and practical issues regarding our tests in Sections 2 or 3 below when they are compared with conventional tests. From a practical point of view, our tests have the benefit that the practitioner does not need to choose the rather artificial quantities τ or \hat{s} when performing the test. Moreover, we avoid the rather undesirable outcome that, even when using the same data set, two practitioners may lead to contradictory conclusions by choosing two different values of τ . This is confirmed in the small Monte-Carlo experiment described in Section 4, which suggests that the choice of τ is not irrelevant by observing that the size and power of the test vary with the choice of τ . In fact, if we followed the recommendation given by some authors of taking $\tau = .15$, see for instance Andrews (1993), the power of the conventional test is much lower when it is compared to our test or when we choose $\tau = .05$, although when the break is towards the end of the sample our test is preferable.

The remainder of the paper is organized as follows. For exposition purposes, next section describes and establishes the asymptotic distribution of the tests in a linear regression model, whereas Section 3 extends the results to more general models useful in econometrics such as nonlinear simultaneous equation systems and transformation models. Section 4 describes a Monte-Carlo experiment to examine

the finite sample performance of our tests and how they compare with the conventional tests for some values of τ . Finally, Section 5 gives the proofs of our main results in Sections 2 and 3.

2. TESTS FOR BREAKS

This section examines, for exposition purposes, tests for breaks in the linear regression model

$$(2.1) \quad y_t = \alpha + \beta' x_t + \delta' z_t(s) + u_t, \quad t = 1, \dots, n$$

where, denoting $\mathbf{1}(\cdot)$ as the indicator function,

$$(2.2) \quad z_t(s) = x_{t1} \mathbf{1}(t \leq s)$$

being x_{t1} a p_1 subvector of the p -dimensional vector $x_t = (x'_{t1}, x'_{t2})'$ and where $\{u_t\}_{t \in \mathbb{Z}}$ is a zero mean sequence of errors. Our null hypothesis H_0 of interest is that the parameter δ is zero for all s . That is,

$$(2.3) \quad H_0 : \delta = 0 \quad \forall s : p^* < s \leq n - p^*,$$

where $p^* = p + p_1 + 1$, being the alternative hypothesis H_1 the negation of the null, that is

$$H_1 : \exists s : p^* < s \leq n - p^*; \quad \delta \neq 0.$$

Notice that when $p_1 = p$ (and z_t also includes the intercept), we have that (2.3) corresponds to the so-called pure structural break hypothesis testing. We consider the situation under the alternative of a one-time structural break, although as we will see in Section 2.3 below, the tests have non-trivial power when the break is gradual and it takes some periods for the model or parameters to reach its new regime. In addition, we have assumed for simplicity that the intercept is constant. It goes without saying that our results follow if we allow the intercept to be subject to a possible break. The only difference lies in the computation of the test and more specifically on the estimation of the asymptotic covariance matrix of the estimator of the intercept. The same comments would apply if the regression model had a time trend subject to a possible break.

We now describe the estimators and present some notation to be used throughout the paper. Herewith C denotes a generic finite positive constant. For a generic sequence $\{w_t\}_{t=1}^n$, we write $\{\tilde{w}_t\}_{t=1}^n =: \{w_t - \bar{w}\}_{t=1}^n$, where $\bar{w} = n^{-1} \sum_{t=1}^n w_t$. Also,

$$(2.4) \quad W_n = (\tilde{w}_1, \dots, \tilde{w}_n)',$$

which may have a partition $W_n = (W_{1n} : W_{2n})$. For a fixed $s \in (p^*, n - p^*]$, $W_n(s)$ indicates that \tilde{w}_t is obtained from $w_t \mathbf{1}(t \leq s)$ and we denote the least squares estimator of $(\beta', \delta')'$ in (2.1) by

$$(2.5) \quad \begin{pmatrix} \hat{\delta}^{(s)} \\ \hat{\beta}^{(s)} \end{pmatrix} = \begin{pmatrix} Z'_n(s) Z_n(s) & Z'_n(s) X_n \\ X'_n Z_n(s) & X'_n X_n \end{pmatrix}^{-1} \begin{pmatrix} Z'_n(s) Y_n \\ X'_n Y_n \end{pmatrix}.$$

We should mention that if, for example, the matrix $Z'_n(s) Z_n(s)$ were singular, in the computation of our estimates and/or tests, we would use the generalized inverse instead of the inverse. This will not affect any of the conclusions of the paper. Similar comments apply elsewhere below.

We now introduce the following assumptions.

A1: $\{x_t\}_{t \in \mathbb{Z}}$ and $\{u_t\}_{t \in \mathbb{Z}}$ are two linear processes defined by

$$\begin{aligned} x_t &= \mu_x + \sum_{i=0}^{\infty} \chi_i \varepsilon_{t-i}, \quad \sum_{i=0}^{\infty} \|\chi_i\|^{1/\varsigma} < \infty \text{ and } \chi_0 = I_p \\ u_t &= \sum_{i=0}^{\infty} \phi_i \eta_{t-i}, \quad \sum_{i=0}^{\infty} |\phi_i|^{1/\varsigma} < \infty \text{ and } \phi_0 = 1, \end{aligned}$$

for some $\varsigma \geq 3/2$, where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ and $\{\eta_t\}_{t \in \mathbb{Z}}$ are mutually independent sequences of independently distributed zero mean random variables such that $\mathbb{E}(\eta_t^2) = \sigma_\eta^2$ and $\mathbb{E}(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$ and $\sup_t \mathbb{E} \|x_t\|^4 + \sup_t \mathbb{E} |u_t|^4 < \infty$, where $\|A\|$ denotes the Euclidean norm of the matrix A and I_p is the p -dimensional identity matrix.

Denote $M_s = \sum_{t=1}^s \tilde{x}_t \tilde{x}_t'$. Then,

A2: $s^{-1} M_s \rightarrow_{s \rightarrow \infty} \Sigma > 0$.

A1 is restrictive in the assumption of linearity, although as we shall see in the next section, we can relax this requirement to allow for more general sequences of random variables. The condition on the rate of convergence to zero of $\{\chi_i\}_{i \geq 0}$ and $\{\phi_i\}_{i \geq 0}$ is minimal and it implies that χ_i and ϕ_i are $o(i^{-3/2})$. The assumption of identical second moments is not essential for the results to follow, but for expositional purposes we have assumed that the sequences $\{x_t\}_{t \in \mathbb{Z}}$ and $\{u_t\}_{t \in \mathbb{Z}}$ are covariance stationary. In fact, we can allow the sequences $\{x_t\}_{t \in \mathbb{Z}}$ and $\{u_t\}_{t \in \mathbb{Z}}$ to exhibit heterogeneity, so that we could allow for heteroscedastic errors $E[u_t^2 | x_t] = \sigma^2(x_t)$, and hence we do not need to assume that the sequences $\{x_t\}_{t \in \mathbb{Z}}$ and $\{u_t\}_{t \in \mathbb{Z}}$ are mutually independent. More explicitly, see Andrews (1993) or Bai and Perron (1998) among others, it would suffice to assume that

$$(2.6) \quad \text{Var} \left(n^{-1/2} \sum_{t=[n\tau_1]+1}^{[n\tau_2]} (x_t - \mu_x) u_t \right) \rightarrow (\tau_2 - \tau_1) \Xi,$$

where $\Xi = \lim_{n \rightarrow \infty} n^{-1} \text{var} \left(\sum_{t=1}^n (x_t - \mu_x) u_t \right)$. We keep nevertheless in this section Assumption A1 as it stands for pedagogical reasons to make the proof of Proposition 1 below clearer while, at the same time, keeping the main steps for more general type of data and models examined in Section 3. We shall nevertheless emphasize that we do not assume anywhere that the sequences $\{x_t\}_{t \in \mathbb{Z}}$ or $\{u_t\}_{t \in \mathbb{Z}}$ are stationary.

Before we present the test for the null hypothesis H_0 in (2.3), we put forward a proposition which plays a key role in the proof of Theorem 1 below.

Proposition 1. *Under A1 and A2, we can construct on a probability space a p -dimensional Wiener process $\mathcal{B}(k)$ with independent components such that*

$$(2.7) \quad \Pr \left\{ \sup_{1 \leq k \leq n} \left\| \sum_{t=1}^k (x_t - \mu_x) u_t - \Xi_k^{1/2} \mathcal{B}(k) \right\| > a \right\} \leq C a^{-4} n^{\frac{\varsigma+2}{\varsigma+1}},$$

for some $1 < \zeta < 2$, and where $\Xi_k = k^{-1} \text{var} \left(\sum_{t=1}^k (x_t - \mu_x) u_t \right)$.

Proposition 1 extends previous results by Einmahl (1989) who considered partial sums of a vector sequence of independent identically distributed random variables or those in Götze and Zaitsev (2007) for nonidentically distributed sequences of independent random variables. The latter work is an extension to vector sequences of results due to Sakhanenko, see for instance Shao (1995). Observe that the rate of the approximation in (2.7) is worse than the “standard” $n^{1/4}$ for linear sequences of random variables or scalar nonlinear sequences of random variables, as shown

respectively by Wang, Lin and Gulati (2003) and Wu (2007). It is worth indicating that if the sequence $\{x_t\}_{t \in \mathbb{Z}}$ were deterministic, we would have the standard conclusion of the order of approximation to be $n^{1/4}$.

We now comment on the key role that Proposition 1 plays on our results. Take for example the Wald statistic in (2.10) or (2.12) below. Inspecting its formulation and using the notation $S_s = \sum_{t=1}^s (x_t - \mu_x) u_t$, the asymptotic distribution depends on that of $(n/(n-s))^{1/2} s^{-1/2} \tilde{\delta}^{(s)}$, which is governed by the behaviour of

$$(2.8) \quad \begin{aligned} & \left(\frac{n}{n-s} \right)^{1/2} \frac{1}{s^{1/2}} \left(I_{p_1} : 0_{p_1 \times p-p_1} \right) \left\{ S_s - \frac{s}{n} S_n \right\} \\ & = \left(I_{p_1} : 0_{p_1 \times p-p_1} \right) \left\{ \left(\frac{n-s}{ns} \right)^{1/2} S_s - \left(\frac{s}{n(n-s)} \right)^{1/2} (S_n - S_s) \right\}. \end{aligned}$$

Next, noticing that

$$\left(\frac{s}{n(n-s)} \right)^{1/2} (S_n - S_s) \stackrel{d}{=} \left(\frac{n-\hat{s}}{n\hat{s}} \right)^{1/2} \sum_{t=1}^{\hat{s}} (x_t^* - \mu_x) u_t^*$$

where $(x_t^* - \mu_x) u_t^* = (x_{n-t+1} - \mu_x) u_{n-t+1}$ and $\hat{s} = n - s$, we conclude that the asymptotic distribution of the Wald test in (2.12) below is a continuous functional of the asymptotic distribution of

$$\mathfrak{S}_n = \max_{1 \leq s \leq n} \left\| \left(I_{p_1} : 0_{p_1 \times p-p_1} \right) \frac{1}{s^{1/2}} S_s \right\|,$$

which is much more delicate to obtain than the (asymptotic) distribution of

$$\tilde{\mathfrak{S}}_n = \max_{1 \leq s \leq n} \left\| \left(I_{p_1} : 0_{p_1 \times p-p_1} \right) \frac{1}{n^{1/2}} S_s \right\|.$$

One of the reasons is that \mathfrak{S}_n attains its maximum for relatively small values of s and the usual crude application of the central limit theorem will not work. On the other hand, Proposition 1 suggests that the distribution of \mathfrak{S}_n , and thus that of the tests, will be governed by the asymptotic distribution of

$$\mathfrak{Z} = \sup_{p^* < s \leq n-p^*} \left\| \frac{1}{s^{1/2}} \sum_{t=1}^s \nu_t \right\|,$$

where $\{\nu_t\}_{t \in \mathbb{Z}}$ is a p_1 -dimensional vector of independent normally distributed random variables. Notice that for scalar $\{\nu_t\}_{t \in \mathbb{Z}}$, the distribution of \mathfrak{Z} was examined by Darling and Erdős (1956). Finally, we shall draw to our attention that when we assume that the break may occur in the ‘‘middle’’ of the sample, the (asymptotic) distribution of the conventional tests are a functional of (Brownian) $\tilde{\mathfrak{S}}_n$.

Before we describe the tests for the null hypothesis H_0 , we discuss the estimators of the asymptotic covariance matrix $\Xi := \lim_{k \rightarrow \infty} \Xi_k$ of the least squares estimators in (2.5). Notice that A1 and A2 imply that Ξ is a finite and positive definite matrix. An estimator of Ξ_n is $\hat{\Xi}_n$, where

$$(2.9) \quad \hat{\Xi}_m = \hat{\gamma}_m^x(0) \hat{\gamma}_m^u(0) + \sum_{j=1}^{m-1} (\hat{\gamma}_m^x(j) + \hat{\gamma}_m^{x'}(j)) \hat{\gamma}_m^u(j), \quad m = p^* + 1, \dots, n,$$

being $\hat{\gamma}_m^x(j)$ and $\hat{\gamma}_m^u(j)$ the estimators respectively of $\mathbb{E} \{(x_t - \mu_x)(x_{t+j} - \mu_x)'\}$ and $\mathbb{E}(u_t u_{t+j})$ such that

$$\hat{\gamma}_m^x(j) = \frac{1}{m} \sum_{t=1}^{m-j} \tilde{x}_t \tilde{x}_{t+j}', \quad \hat{\gamma}_m^u(j) = \frac{1}{m} \sum_{t=1}^{m-j} \tilde{u}_t^{(s)} \tilde{u}_{t+j}^{(s)}.$$

Here or elsewhere $\{\widehat{u}_t^{(s)}\}_{t=1}^n$ is a sequence of residuals which depends on the estimator of the parameters that we have employed. For instance, if we employed the least squares estimator in (2.5), we would compute the residuals $\{\widehat{u}_t^{(s)}\}_{t=1}^n$ as $\widehat{u}_t^{(s)} = \widetilde{y}_t - \widehat{\beta}^{(s)'} \widetilde{x}_t + \widehat{\delta}^{(s)'} \widetilde{z}_t(s)$. Furthermore, to simplify the notation we have deliberately suppressed the dependence on s of $\widehat{\gamma}_m^u(j)$.

Notice that $\widehat{\Xi}_n$ is the time domain formulation of Robinson's (1998) estimator of Ξ . Following Robinson (1998), we know that $\widehat{\Xi}_n$ is a consistent estimator of Ξ under A1, so that we do not need any kernel spectral density estimator to obtain a consistent estimator of Ξ nor to choose a bandwidth parameter to perform the test. However, it is true that A1 is slightly stronger than we need. Indeed, Robinson (1998) showed that for the consistency of $\widehat{\Xi}_n$, it suffices to assume that

$$\mathbb{E}(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1} \cup \mathcal{G}_t) = \Sigma_\varepsilon; \quad \mathbb{E}(\eta_t^2 | \mathcal{F}_t \cup \mathcal{G}_{t-1}) = \sigma_\eta^2,$$

where \mathcal{F}_t and \mathcal{G}_t are respectively the sigma-algebras generated by $\{\varepsilon_v : v \leq t\}$ and $\{\eta_v : v \leq t\}$. However, because $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ and $\{\eta_t\}_{t \in \mathbb{Z}}$ are sequences of independent random variables, the last displayed expressions become $\mathbb{E}(\varepsilon_t \varepsilon_t' | \mathcal{G}_t) = \Sigma_\varepsilon$ and $\mathbb{E}(\eta_t^2 | \mathcal{F}_t) = \sigma_\eta^2$, so that A1 is not much stronger than assuming the latter.

If we would have allowed the intercept α in (2.1) to have a break, we had that $\Xi = \lim_{n \rightarrow \infty} n^{-1} \sum_{t_1, t_2=1}^n \mathbb{E}(w_{t_1} w_{t_2}' u_{t_1} u_{t_2}')$, where $w_t' = (1, (x_t - \mu_x)')$. In this case, the analogue estimator of (2.9) corresponding to the component $\Xi(1, 1)$, i.e. the asymptotic variance of the least squares estimator of α , is

$$\widehat{\Xi}_n(1, 1) = \widehat{\gamma}_n^u(0) + 2 \sum_{j=1}^{n-1} \widehat{\gamma}_n^u(j).$$

However, contrary to $\widehat{\Xi}_n$ in (2.9), $\widehat{\Xi}_n(1, 1)$ is not a consistent estimator of $\Xi(1, 1)$. So, in this situation, we would employ $\widetilde{\Xi}_m = \widehat{\gamma}_n^u(0) + 2 \sum_{j=1}^{m-1} \widehat{\gamma}_n^u(j)$ for some $m = o(n)$, which is a consistent estimator under Assumptions A1 and A2 as it has been shown by Andrews (1991) among others. It is worth mentioning that instead of $\widehat{\Xi}_n$ given in (2.9) we might have been tempting to employ

$$\widetilde{\Xi}_n = \widehat{\gamma}_n^v(0) + \sum_{j=1}^{n-1} (\widehat{\gamma}_n^v(j) + \widehat{\gamma}_n^{v'}(j)),$$

where $\widehat{\gamma}_n^v(j) = n^{-1} \sum_{t=1}^{n-j} \widehat{v}_t \widehat{v}_{t+j}$, with $\widehat{v}_t = \widetilde{x}_t \widehat{u}_t^{(s)}$. However, in this case, contrary to $\widehat{\Xi}_n$, $\widetilde{\Xi}_n$ would not be a consistent estimator for Ξ , although it is the standard kernel spectral density estimator $\widetilde{\Xi}_m$ for some $m = o(n)$. Observe that if the errors and regressors were not mutually independent, i.e. $\{u_t\}_{t \in \mathbb{Z}}$ is a heteroscedastic sequence but satisfying (2.6), the estimator (2.9) would be inconsistent for Ξ . In this case, we should employ $\widetilde{\Xi}_n$ to estimate Ξ consistently.

We now describe the test statistics.

2.1. The Wald statistic.

Suppose that we are first interested to test H_0 against the alternative hypothesis $H_1(s)$ defined, for some $p^* < s \leq n - p^*$, as

$$H_1(s) : \delta \neq 0.$$

In this case the Wald statistic is based on whether $\widehat{\delta}^{(s)}$ is significantly different than zero. Recalling our notation in (2.4) and denoting $Z_n'(s) X_n = AM_s$ with

$A = \begin{pmatrix} I_{p_1} & 0 \\ 0 & I_{p-p_1} \end{pmatrix}$, using standard algebra we write

$$\begin{aligned}\widehat{\delta}^{(s)} &= B_n^{-1}(s) A \{X'_n(s) Y_n - M_s M_n^{-1} X'_n Y_n\} \\ &= B_n^{-1}(s) A \{X'_n(s) U_n - M_s M_n^{-1} X'_n U_n\},\end{aligned}$$

under H_0 , where

$$B_n(s) = A \{M_s - M_s M_n^{-1} M_s\} A'.$$

Thus, we write the Wald statistic for H_0 against $H_1(s)$ as

$$(2.10) \quad \mathcal{W}(s) = \widetilde{\delta}^{(s)'} \left(A \widehat{V}_n(s) A' \right)^{-1} \widetilde{\delta}^{(s)},$$

where $\widetilde{\delta}^{(s)} = B_n(s) \widehat{\delta}^{(s)}$ and

$$\widehat{V}_n(s) = s \widehat{\Xi}_s + n M_s M_n^{-1} \widehat{\Xi}_n M_n^{-1} M_s - s M_s M_n^{-1} \widehat{\Xi}_s - s \widehat{\Xi}_s M_n^{-1} M_s,$$

with $\widehat{\Xi}_s$ given in (2.9). We should bear in mind that in the computation of $\widehat{V}_n(s)$, we have ignored an estimator of $\mathbb{E} \left(\sum_{t=1}^s \widetilde{x}_t u_t \sum_{t=1+s}^n \widetilde{x}_t u_t \right)$. The reason is because the latter expression is asymptotically negligible when we compare it with either $s \widehat{\Xi}_s$ or $n \widehat{\Xi}_n$.

On the other hand, if we wish to test for a break when the time of the break “ s ” is unknown, we employ the standard union-intersection principle. That is, the alternative hypothesis becomes $H_1 = \cup_{s=p^*+1}^{n-p^*} H_1(s)$, so that the hypothesis testing becomes

$$(2.11) \quad H_0 \text{ against } H_1.$$

In this case, and because the dimension of the parameter vectors β and δ are respectively p and p_1 , we might consider the statistic $\widetilde{\mathcal{W}} = \max_{p^* < s \leq n-p^*} \mathcal{W}(s)$. Now, if we use the same estimator of Ξ , for example $\widehat{\Xi}_n$, and we replace M_s by its “limit” $s \Sigma$, we can simplify $\widetilde{\mathcal{W}}$ as

$$(2.12) \quad \mathcal{W} = \max_{p^* < s \leq n-p^*} \left(\frac{n}{n-s} \right) \frac{1}{s} \widetilde{\delta}^{(s)'} \left(A \widehat{\Xi}_n A' \right)^{-1} \widetilde{\delta}^{(s)}.$$

2.2. The LM statistic.

We now describe the *Lagrange Multiplier* test. As we did with the *Wald* statistic, suppose that we wish to test H_0 against the alternative hypothesis $H_1(s)$. In this case, the test would be based on whether or not the first order derivatives

$$\widehat{F}_n^{(s)} = \sum_{t=1}^s \widetilde{x}_{t1} \left(\widetilde{y}_t - \widetilde{x}'_t \widetilde{\beta} \right)$$

are significantly different than zero, where $\widetilde{\beta}$ is the least squares estimator of $\{\widetilde{y}_t\}_{t=1}^n$ on $\{\widetilde{x}_t\}_{t=1}^n$. Now, observing that

$$\widehat{F}_n^{(s)} = B_n(s) \widehat{\delta}^{(s)},$$

we obtain the *Lagrange Multiplier* as

$$\mathcal{LM}(s) = \left(\frac{n}{n-s} \right) \frac{1}{s} \widehat{F}_n^{(s)'} \left(A \widehat{\Xi}_n A' \right)^{-1} \widehat{F}_n^{(s)},$$

where we now employ the restricted least squares residuals $\widehat{u}_t = \widetilde{y}_t - \widetilde{x}'_t \widetilde{\beta}$ instead of $\widehat{u}_t^{(s)}$ to obtain $\widehat{\gamma}_n^u(j)$, $j = 0, \dots, n-1$, when computing $\widehat{\Xi}_n$ in (2.9).

Next for the hypothesis testing (2.11), the *Lagrange Multiplier* statistic is

$$(2.13) \quad \mathcal{LM} = \max_{p^* < s \leq n-p^*} \mathcal{LM}(s).$$

The \mathcal{LM} statistic only requires estimation under the null hypothesis and hence we do not need to estimate the model for each point of the sample. So, in situations where the computation of estimates of the parameters of the model is computing intensive (such as the *GMM* estimator in the next section), it appears to be more appropriate to employ (2.13).

Let's introduce some notation. We denote $\log_2 x = \log \log x$ and $\log_3 x = \log \log \log x$ and \mathcal{T} a random variable distributed as a (double) Gumbel random variable, i.e.

$$(2.14) \quad \Pr \{ \mathcal{T} \leq x \} = \exp(-2e^{-x}).$$

Theorem 1. *Assuming A1 and A2, under H_0 we have that*

$$(a) \quad a_n \mathcal{W}^{1/2} - b_n \xrightarrow{d} \mathcal{T}; \quad (b) \quad a_n \mathcal{LM}^{1/2} - b_n \xrightarrow{d} \mathcal{T},$$

where $a_n = (2 \log_2 n)^{1/2}$, $b_n = 2 \log_2 n + \frac{p_1}{2} \log_3 n - \log \Gamma(p_1/2)$, and where $\Gamma(\cdot)$ is the gamma function.

Remark 1. *The proof of Theorem 1 indicates that if instead of looking for the maximum in the region $p^* < s \leq n - p^*$, we had considered the maximum in either $n/2 \leq s \leq n - p^*$ or $p^* < s < n/2$, the asymptotic distribution of the test would have been the Gumbel distribution \mathcal{T}_1 , i.e.*

$$\Pr \{ \mathcal{T}_1 \leq x \} = \exp(-e^{-x}).$$

2.3. Power of the Test.

This section examines the behaviour of our tests (2.12) and (2.13) under fixed and local alternatives. For that purpose, we consider a sequence of models

$$(2.15) \quad y_t = \alpha + \beta' x_t + g_n' x_{t1} \mathbf{1}(t \leq s_0) + u_t, \quad t = 1, \dots, n$$

where the sequence $\{g_n\}_{n \in \mathbb{N}}$ only depends on n , to be made more precise below. For simplicity of arguments, we have chosen a one-time structural break model, although the conclusions in Theorems 2 and 3 hold true under more general type of breaks such as when it is a gradual one or there are more than one.

We first examine the behaviour of our tests under fixed alternatives.

Theorem 2. *Assuming A1 and A2, under model (2.15) with $g_n = \delta$, we have that*

$$\Pr \left\{ a_n \mathcal{S}^{1/2} - b_n \leq x \right\} \rightarrow 0; \quad x \in \mathbb{R}$$

if $h_n \leq s_0 \leq n - h_n$, where $h_n^{-1} = o(\log_2^{-1} n)$, a_n and b_n are as in Theorem 1 and \mathcal{S} is either the \mathcal{W} and \mathcal{LM} statistics in (2.12) and (2.13).

Theorem 2 shows that our tests are effectively consistent irrespective of the location of the break since $\log_2 n < p^*$ for the typical samples sizes n that we encounter in real examples. Notice that, for instance when $p^* = 3$, $\log_2 n > p^*$ if $n > 53 \times 10^7$, which is a sample size that we do not encounter even with financial data. This implies that our tests are effectively consistent in contrast with the conventional tests. Indeed, let's consider the conventional Wald statistic

$$(2.16) \quad \check{\mathcal{W}} = \sup_{[\tau n] \leq s \leq n - [\tau n]} \mathcal{W}(s)$$

for some $0 < \tau < 1/2$. Because the noncentrality parameter (function) of $\mathcal{W}(s)$ is

$$(2.17) \quad \Lambda_n(s; s_0) = \left(\frac{n}{n-s} \right)^{1/2} \frac{1}{s^{1/2}} g_n' \left\{ \sum_{t=1}^s \mathbf{1}(t \leq s_0) - \frac{s}{n} \sum_{t=1}^n \mathbf{1}(t \leq s_0) \right\},$$

apart from a constant as seen in the proof of Theorem 2, we have that when $s_0 \in [[\tau n], n - [\tau n]]$,

$$(2.18) \quad \Lambda_n(s; s_0) = O\left(n^{1/2} \|g_n\|\right).$$

So, recalling that we are under fixed alternatives, i.e. $g_n = \delta$, (2.18) implies that \check{W} is a consistent test when the break is in the middle of the sample. On the other hand, when $s_0 < [\tau n]$ or $n - [\tau n] < s_0$, this is not always the case. Indeed consider the case when $s_0 < [\tau n]$. Because $s_0 < s$ as $[\tau n] \leq s \leq n - [\tau n]$, the right side of (2.17) is bounded in absolute value by

$$(2.19) \quad C \frac{1}{s^{1/2}} s_0 \left(\frac{n-s}{n}\right)^{1/2} \mathbf{1}(s_0 \leq s) \leq C \frac{1}{n^{1/2}} s_0$$

and hence if $s_0 = o(n^{1/2})$, the last expression converges to zero uniformly in $[\tau n] \leq s \leq n - [\tau n]$. The latter implies that, if $s_0 = o(n^{1/2})$, the conventional tests have the same asymptotic distribution as under the null hypothesis. By symmetry, it is evident that we can draw the same conclusions when $n - s_0 = o(n^{1/2})$. Hence, the conventional tests are inconsistent when the time of the break s_0 is such that $s_0 = o(n^{1/2})$ or $n - s_0 = o(n^{1/2})$, whereas our tests \mathcal{W} and \mathcal{LM} in (2.12) and (2.13) are still consistent in those regions.

We now investigate the behaviour of our tests under local alternatives.

Theorem 3. *Assuming A1 and A2, under model (2.15) we have that*

- (a) $\Pr\{a_n \mathcal{S}^{1/2} - b_n \leq x\} \rightarrow 0$, if $g_n^{-1} = o\left(s_0^{1/2} \log_2^{-1/2} n\right)$
- (b) $\Pr\{a_n \mathcal{S}^{1/2} - b_n \leq x\} \rightarrow \exp(-2e^{-x})$, if $g_n = o\left(s_0^{-1/2} \log_2^{1/2} n\right)$
- (c) $0 \leq \Pr\{a_n \mathcal{S}^{1/2} - b_n \leq x\} \leq 1$, otherwise,

where a_n and b_n are as in Theorem 1 and \mathcal{S} is either the \mathcal{W} and \mathcal{LM} statistics.

We now comment on the results of Theorem 3. The theorem indicates that the tests have non-trivial power against local alternatives of the type

$$(2.20) \quad H_a : g_n = \delta\left(\log_2^{1/2} n / s_0^{1/2}\right).$$

That is, as long as the size of the break g_n satisfies $g_n^{-1} = o\left(s_0^{1/2} \log_2^{-1/2} n\right)$, our tests (2.12) and (2.13) reject with probability 1 as $n \rightarrow \infty$. This is in clear contrast with conventional tests. We already know by looking at (2.17) that when the break occurs in the ‘‘middle’’ of the sample, the conventional tests have non-trivial power against local alternatives of order $O(n^{-1/2})$. However, when the break occurs at time $s_0 < [n\tau]$, the latter does not follow as we now discuss. (The treatment and conclusions when $s_0 > n - [n\tau]$ are identical and so they are not explicitly discussed.)

Indeed, when $s_0 < [n\tau]$, because $s_0 < s$, (2.17) implies that

$$C^{-1} \left(\frac{s_0}{n^{1/2}}\right) g_n \leq |\Lambda_n(s; s_0)| \leq C \left(\frac{s_0}{n^{1/2}}\right) g_n.$$

So, when the time of the break satisfies $Cn^{1/2} < s_0 < [n\tau]$ (we already know that when $s_0 = o(n^{1/2})$ the conventional tests are not consistent), we easily conclude that \check{W} in (2.16) has nontrivial power when $g_n = O(n^{1/2}/s_0)$, see (2.19). The latter implies that the conventional tests have zero relative efficiency compared to ours when $s_0 = o(n/\log_2 n)$ as the right side of (2.20) is $o(s_0^{-1} n^{1/2})$. On the other hand, conventional tests are more efficient when $[n\tau] \leq s_0 \leq n - [n\tau]$, although Theorems 2 and 3 indicate that the penalty, if at all, to pay to allow the break to be anywhere in the sample is very negligible. Indeed, as mentioned after Theorem 2, our tests

are effectively consistent irrespective of the location of the break. Moreover, and perhaps more importantly, there is no need to choose the value of τ (or \hat{s}) to compute the test. The latter point is relevant for practical purposes as we avoid the undesirable property that with the same data set, two different practitioners may conclude differently by choosing two different values of τ in the definition of $\check{\mathcal{W}}$, as the Monte-Carlo experiment in Section 4 illustrates.

3. TESTING FOR BREAKS IN NONLINEAR MODELS

It is often the case that the relationship among economic variables occur in nonlinear form. When the nonlinearity decomposes into a function of the explanatory variables plus an additive disturbance error term, we have the standard nonlinear regression model. It is however common that the endogenous variables of the model are subject to nonlinearities. The latter is the case with nonlinear simultaneous equation systems and with nonlinear transformation models. To be specific, for $t = 1, \dots, n$, let y_t be a $G \times 1$ vector of endogenous variables, x_t a $N \times 1$ vector of explanatory variables and u_t a $G \times 1$ vector of disturbance terms such that

$$(3.1) \quad u_t = u(y_t, x_t; \theta, \varphi),$$

for a vector-valued function $u(y, x; \theta, \varphi)$, where θ and φ are a p_1 - and p_2 -dimensional vectors of unknown parameters, respectively. By (nonlinear) transformation models we mean that $u(y_t, x_t; \theta, \varphi)$ takes the form

$$u(y_t, x_t; \theta, \varphi) = u_y(y_t; \theta, \varphi) - u_x(x_t; \theta, \varphi).$$

One of these transformations is the well-known Box-Cox model, for which

$$u_y(y_t; \theta, \varphi) = (y_t^\lambda - 1) / \lambda,$$

where λ is a component of the vectors θ and φ . Another transformation of interest is given in Burbidge, Magee and Robb (1988), where now

$$u_y(y_t; \theta, \varphi) = \operatorname{arcsinh}(\lambda y_t) / \lambda,$$

whereas when $u_y(y_t; \theta, \varphi) = y_t$ we have the standard nonlinear regression model.

As in Section 2, we consider the one-time structural break model, where

$$\theta_t = \theta_0 + \delta_0 \mathbf{1}(t \leq s),$$

for some $s \in (1, n)$. We are interested in the constancy of the parameter θ_t , that is $\theta_t = \theta_0$ for all $t \geq 1$, so that the null hypothesis of interest becomes

$$(3.2) \quad H_0 : \delta_0 = 0 \quad \text{for all } s: p^* < s \leq n - p^*,$$

where $p^* = 2p_1 + p_2$. Again if $p_2 = 0$, we would have the pure structural break hypothesis. The alternative hypothesis H_1 is the negation of the null, that is

$$H_1 : \text{For some } s : p^* < s \leq n - p^*, \quad \delta_0 \neq 0.$$

We now introduce the following regularity conditions.

A3: $\{x_t\}_{t \in \mathbb{Z}}$ and $\{u_t\}_{t \in \mathbb{Z}}$ are two sequences of random variables such that $x_t = h_x(\hat{x}_t)$ and $u_t = \Sigma^{1/2}(x_t) h_v(\hat{v}_t)$, where

$$\begin{aligned} \hat{x}_t &= \sum_{i=0}^{\infty} \chi_i \varepsilon_{t-i}, & \sum_{i=0}^{\infty} \|\chi_i\|^{1/\varsigma} < \infty \text{ and } \chi_0 = I_N \\ \hat{v}_t &= \sum_{i=0}^{\infty} \phi_i \eta_{t-i}, & \sum_{i=0}^{\infty} \|\phi_i\|^{1/\varsigma} < \infty \text{ and } \phi_0 = I_G, \end{aligned}$$

for some $\varsigma \geq 3/2$, where the sequences $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ and $\{\eta_t\}_{t \in \mathbb{Z}}$ satisfy Assumption A1. $h_x(\cdot)$ and $h_v(\cdot)$ are differentiable vector-valued functions everywhere in their domain of definition such that $\{x_t\}_{t \in \mathbb{Z}}$ and $\{h_v(\hat{v}_t)\}_{t \in \mathbb{Z}}$

are L^4 -*NED* (Near Epoch Dependent) sequences of size $\iota > 1$. That is, for a generic random variable w_t , denoting $\|w_t\|_4 = (\mathbb{E} \|w_t\|^4)^{1/4}$,

$$\begin{aligned} \|x_t - \mathbb{E}(x_t | \varepsilon_t, \dots, \varepsilon_{t-m})\|_4 &= O(m^{-\iota}) \\ \|h_v(\hat{v}_t) - \mathbb{E}(h_v(\hat{v}_t) | \eta_t, \dots, \eta_{t-m})\|_4 &= O(m^{-\iota}). \end{aligned}$$

Assumption A3 relaxes the assumption of linearity of $\{x_t\}_{t \in \mathbb{Z}}$ and $\{u_t\}_{t \in \mathbb{Z}}$ in A1, allowing for very general type of dependence. Notice that A3 allows $\{u_t\}_{t \in \mathbb{Z}}$ to be not only serially dependent but also heteroscedastic. A3 implies that $\mathbb{E}(u_t \otimes P(x_t)) = 0$, where $P(\cdot)$ is an $M \times 1$ column vector of known functions for $M \geq p^*$, and “ \otimes ” denotes the Kronecker product. Moreover, it is also true that provided enough finite moments on $P(x_t)$, (2.6) holds true under standard arguments.

We now give an extension of Proposition 1.

Proposition 2. *Assuming A3, we can construct on a probability space a GM -dimensional Wiener process $\mathcal{B}(k)$ with independent components such that*

$$(3.3) \quad \Pr \left\{ \sup_{1 \leq k \leq n} \left\| \sum_{t=1}^k (u_t \otimes P(x_t)) - \Delta_k^{1/2} \mathcal{B}(k) \right\| > a \right\} \leq Ca^{-2} n^{\frac{\zeta+2}{2(\zeta+1)}},$$

for some $1 < \zeta < 2$, and where $\Delta_k = k^{-1} \text{var} \left(\sum_{t=1}^k u_t \otimes P(x_t) \right)$.

A common procedure to obtain estimators of the parameters is to exploit the assumption that the disturbance error term u_t is orthogonal to $P(x_t)$. That is,

$$\begin{aligned} \mathbb{E}(u(y_t, x_t; \theta_0 + \delta_0, \varphi_0) \otimes P(x_t)) &= 0, \quad t = 1, \dots, s, \\ \mathbb{E}(u(y_t, x_t; \theta_0, \varphi_0) \otimes P(x_t)) &= 0, \quad t = s+1, \dots, n. \end{aligned}$$

As elsewhere, see Hansen (1982), Chamberlain (1987) or Robinson (1988) among others, we estimate the parameters by *GMM* for each s . To describe the estimation procedure, some notation is useful. Let $\vartheta = (\theta', \varphi')'$, $\psi = (\delta', \vartheta')' \in \Theta \subset \mathbb{R}^{p^*}$,

$$(3.4) \quad g_t^{(s)}(\psi) = \begin{cases} u(y_t, x_t; \theta + \delta, \varphi) \otimes P(x_t) & \text{if } 1 \leq t \leq s \\ u(y_t, x_t; \theta, \varphi) \otimes P(x_t) & \text{if } s < t \leq n, \end{cases}$$

with the convention that $g_t(\vartheta) = u(y_t, x_t; \theta, \varphi) \otimes P(x_t)$ and

$$\bar{g}_n^{(s)}(\psi) = \sum_{t=1}^n \begin{bmatrix} \mathbf{1}(t \leq s) \\ \mathbf{1}(s < t) \end{bmatrix} \otimes g_t^{(s)}(\psi).$$

We define for a fixed s , the *GMM* estimator of ψ , $\hat{\psi}^{(s)}$, as

$$(3.5) \quad \hat{\psi}^{(s)} = \arg \min_{\psi \in \Theta} \bar{g}_n^{(s)'}(\psi) \Upsilon_n^{(s)} \bar{g}_n^{(s)}(\psi),$$

where $\Upsilon_n^{(s)}$ is a $2GM \times 2GM$ symmetric and positive definite matrix

It is well-known how we can choose the weighting matrix $\Upsilon_n^{(s)}$ optimally and then to construct the test statistics such as *Wald* and *LM* statistics for the null hypothesis. However for expositional arguments, we follow conventional methods as in e.g. Andrews (1993) to construct our tests. Indeed with the notation in (3.1), let $\hat{\Delta}^{(s)}$ be an estimator of the “long-run” variance of $g_t(\vartheta_0)$,

$$(3.6) \quad \Delta = \lim_{k \rightarrow \infty} \Delta_k = \sum_{j=-\infty}^{\infty} \mathbb{E}((u_{t+j} \otimes P(x_{t+j})) (u_t \otimes P(x_t))').$$

Then, we shall use in our analysis

$$\Upsilon_n^{(s)} = \text{diag} \left(\left(\frac{s}{n} \hat{\Delta}^{(s)} \right)^{-1}, \left(\frac{n-s}{n} \hat{\Delta}^{(s)} \right)^{-1} \right),$$

which $\Upsilon_n^{(s)} \rightarrow_P \text{diag} \left(\left((\pi \Delta)^{-1}, ((1 - \pi) \Delta)^{-1} \right) \right)$. We compute $\widehat{\Delta}^{(s)}$ as

$$(3.7) \quad \widehat{\Delta}^{(s)} = \widehat{\gamma}^{(s)}(0) + \sum_{\ell=1}^{m-1} \left(\widehat{\gamma}^{(s)}(\ell) + \widehat{\gamma}^{(s)'}(\ell) \right),$$

where m is a bandwidth parameter such that $m^{-1} + m/n \rightarrow 0$ as $n \rightarrow \infty$ and

$$\widehat{\gamma}^{(s)}(\ell) = \frac{1}{n} \sum_{t=1}^{n-\ell} \left\{ g_t^{(s)}(\widetilde{\psi}) g_{t+\ell}^{(s)}(\widetilde{\psi})' \right\},$$

where $\widetilde{\psi}$ is a consistent estimator of ψ_0 . For instance that given in (3.5) but with $\Upsilon_n^{(s)}$ replaced by the identity matrix. Notice that because we do not assume that $\mathbb{E}(u_t u_t' | x_t) = \Sigma(x_t) = \Sigma$ and/or $\mathbb{E}P(x_t) = 0$, we cannot compute $\widehat{\Delta}^{(s)}$ as in (2.9) to obtain a consistent estimator of Δ .

Next, let

$$G_v = \mathbb{E} \left\{ \frac{\partial}{\partial v'} g_t^{(s)}(\psi_0) \right\} \quad \text{for } v = \theta, \varphi, \delta, \psi$$

and observe that when $t \leq s$, we have that $\frac{\partial}{\partial \delta'} g_t^{(s)}(\psi) = \frac{\partial}{\partial \theta'} g_t^{(s)}(\psi)$ so that $G_\delta = G_\theta$. Also, for any s , consider three different estimators of G_v . One is given by

$$(3.8) \quad \widehat{G}_v^0 = \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\partial}{\partial v'} g_t^{(s)}(\check{\vartheta}) \right\},$$

where $\check{\vartheta}$ is the restrictive estimator of ϑ_0 ,

$$(3.9) \quad \check{\vartheta} = \arg \min_{\vartheta \in \Theta} \left(\sum_{t=1}^n u(y_t, x_t; \vartheta) \otimes P(x_t) \right)' \widehat{\Delta}^{-1} \left(\sum_{t=1}^n u(y_t, x_t; \vartheta) \otimes P(x_t) \right).$$

The others are

$$\widehat{G}_v^1 = \frac{1}{s} \sum_{t=1}^s \left\{ \frac{\partial}{\partial v'} g_t^{(s)}(\widehat{\psi}^{(s)}) \right\}; \quad \widehat{G}_v^2 = \frac{1}{n-s} \sum_{t=s+1}^n \left\{ \frac{\partial}{\partial v'} g_t^{(s)}(\widehat{\psi}^{(s)}) \right\}.$$

Standard arguments yield that the asymptotic variance of $\widehat{\psi}^{(s)}$ can be estimated by $\left(\left(\frac{s}{n} \right)^{-1} \widehat{G}_\psi^1 \widehat{\Delta}^{-1} \widehat{G}_\psi^1 + \left(\frac{n-s}{n} \right)^{-1} \widehat{G}_\psi^2 \widehat{\Delta}^{-1} \widehat{G}_\psi^2 \right)^{-1}$, whose submatrix corresponding to δ is denoted by $\widehat{V}^{(s)}$.

Now, we present some regularity conditions on $\{g_t\}_{t=1}^n$ and thus on $\{u_t\}_{t=1}^n$.

- A4:** (i) For $0 \leq \tau_1 < \tau_2 \leq 1$, $\text{var} \left(n^{-1/2} \sum_{t=[\tau_1 n]+1}^{[\tau_2 n]} g_t(\vartheta_0) \right) \rightarrow (\tau_2 - \tau_1) \Delta$
and $n^{-1/2} \sum_{t=[\tau_1 n]+1}^{[\tau_2 n]} g_t(\vartheta_0) \xrightarrow{d} \mathcal{N}(0, (\tau_2 - \tau_1) \Delta)$.
(ii) $g_t(\vartheta)$ is continuously differentiable with probability one and
 $\mathbb{E} \left[\sup_{\vartheta \in \Theta} \|\partial g_t(\vartheta) / \partial \vartheta\|^{1+\varepsilon} \right] < \infty$ for some $\varepsilon > 0$.

Assumption A4 is common in the literature, see for instance Andrews (1993). This follows by A3 together with some smoothness conditions on $u(y_t, x_t; \theta + \delta, \varphi)$ and moment conditions on $P(x_t)$.

3.1. The Wald and LM statistics.

We begin with the *Wald* test. Suppose that we were first interested to test H_0 in (3.2) against $H_1(s)$ defined as

$$H_1(s) : \text{For a given } s : p^* < s \leq n - p^* \quad \delta_0 \neq 0.$$

In this case the Wald statistic is based on whether $\widehat{\delta}^{(s)}$ in (3.5) is significantly different than zero. We write the Wald statistic for H_0 against $H_1(s)$ as $\mathcal{W}(s) =$

$n\widehat{\delta}^{(s)'}\widehat{V}^{(s)-1}\widehat{\delta}^{(s)}$. Now, as in Section 2.1, for the hypothesis testing in (2.11) we employ the statistic

$$\mathcal{W} = \max_{p^* < s \leq n-p^*} \mathcal{W}(s).$$

We now discuss and present the *Lagrange Multiplier* test. As we did with the *Wald* statistic, suppose that we wish to test H_0 against $H_1(s)$ for a given s . In this case, the test is based on the first order conditions

$$(3.10) \quad \left\{ \frac{\partial}{\partial \delta} \bar{g}_n^{(s)'}(\check{\psi}) \right\} \Upsilon_n^{(s)} \bar{g}_n^{(s)}(\check{\psi}),$$

where $\check{\psi} = (0, \check{\vartheta}')$ and $\check{\vartheta}$ is defined in (3.9). Since $\frac{\partial}{\partial \delta} g_t^{(s)}(\check{\psi}) = 0$ for $t > s$, replacing \widehat{G}_ϑ^1 by \widehat{G}_ϑ^0 in the left side of (3.10), and setting $lm(s) = A\widehat{G}_\vartheta^0\widehat{\Delta}^{-1}\sum_{t=1}^s g_t(\check{\vartheta})$, the *LM* statistic becomes $\mathcal{LM}(s) = \frac{n}{n-s} \frac{1}{s} lm(s)' \left(A\widehat{G}_\vartheta^0\widehat{\Delta}^{-1}\widehat{G}_\vartheta^{0'}A' \right)^{-1} lm(s)$.

Now for the hypothesis testing in (2.11), we compute

$$\mathcal{LM} = \max_{p^* < s \leq n-p^*} \mathcal{LM}(s).$$

We have now the following result.

Theorem 4. *Assuming A3 – A4, under H_0 ,*

$$(a) \quad a_n \mathcal{W}^{1/2} - b_n \xrightarrow{d} \mathcal{T}, \quad (b) \quad a_n \mathcal{LM}^{1/2} - b_n \xrightarrow{d} \mathcal{T},$$

where a_n and b_n are as defined in Theorem 1 and \mathcal{T} is defined in (2.14).

4. MONTE-CARLO EXPERIMENT

This section examines the finite sample properties of our tests as well as their relative performance when they are compared with the conventional statistics. For the Monte-Carlo experiment we have chosen the same design of Krämer and Sonnberger (1986) or Andrews (1992). That is, the linear regression model

$$(4.1) \quad y_t = \beta' x_t + \delta' x_{t1} \mathbf{1}\{t \leq s\} + \sigma(w_{t3}) u_t, \quad t = 1, \dots, n,$$

where $\{u_t\}_{t \in \mathbb{Z}}$ is a sequence of independent $\mathcal{N}(0, 1)$ random variables, $x'_{t1} = (1, (-1)^t)$ and $x_t = (x'_{t1}, w'_t)'$ with $\{w_t\}_{t \in \mathbb{Z}}$ a 5-dimensional vector sequence of independent $\mathcal{N}(0, I_5)$ random variables. The function $\sigma(\cdot)$ is the same as that employed in Hidalgo (1992), that is $\sigma(w_{t3}) = 1 + \exp(a + bw_{t3})$. The purpose of the latter is to examine the influence of the heteroscedasticity into the finite sample performance of the tests. The sample sizes for all models in the experiment were $n = 100, 200, 300, 500, 1000$. Finally, to keep the number of tables manageable, we only present the results for the Wald statistic and $(a, b) = (0, 0)$ and $(0.125, 0.4)$, which we will denote respectively by σ_0 and σ_1 in Tables 4.1 to 4.4. below.

Although our main interest is to compare the performance of the statistic \mathcal{W} given in (2.10) to the conventional statistic

$$\mathcal{W}_\tau = \max_{[n\tau] < s \leq n - [n\tau]} \mathcal{W}(s)$$

we have also considered the popular Brown, et al.'s (1975) CUSUM test and the “optimal” Chow test as a benchmark. We have chosen two trimming quantities $\tau = .05$ and $.15$, which we shall denote herewith as $\mathcal{W}_{.05}$ and $\mathcal{W}_{.15}$. In the definition of $\widehat{V}_n(s)$ we compute $\widehat{\Xi}_s$ as $\widehat{\Xi}_s = s^{-1} \sum_{t=1}^s x_t x'_t \widehat{u}_t^2$, where $\{\widehat{u}_t\}_{t=1}^n$ are the least squares residuals. The estimator $\widehat{\Xi}_n$ corresponds to the White's (1980) standard heteroskedasticity-robust variance estimator.

Table 4.1 describes the performance of the test using the asymptotic critical values under the null hypothesis H_0 at the 5% nominal level. The value chosen for β in (4.1) was $\beta = 0$.

TABLE 4.1 ABOUT HERE

The results presented in Table 4.1 indicate that the statistics \mathcal{W} , $\mathcal{W}_{.05}$ and $\mathcal{W}_{.15}$ tend to over-reject the null hypothesis even for moderately large sample sizes. Also Table 4.1 suggests that the choice of the trimming quantity τ can have the undesirable consequence that a researcher can draw different conclusions about H_0 even with the same data set. On the other hand, the CUSUM test tends to under-reject the null hypothesis, whereas for the “optimal” Chow test, we observe the same type of reservations as with the conventional tests, that is, the outcome of the test appears to depend greatly on the choice of the break point when computing the statistic. For example, when the assumed time of the break is $n^{1/2}$, the empirical size is much higher than when the break is assumed at time $n/2$.

Because of the poor performance of the tests using the asymptotic critical values, we decided to perform a bootstrap algorithm to compute the tests. Because we allow for heteroscedasticity in our regression model (4.1), we shall distinguish the scenarios $(a, b) = (0, 0)$ and $(a, b) = (0.125, 0.4)$ when computing the bootstrap. In the case of homoscedasticity, that is $(a, b) = (0, 0)$, we compute the standard residual based Efron’s bootstrap. That is,

STEP 1: Compute the least squares residuals $\hat{u}_t = y_t - \hat{\beta}' x_t - \hat{\delta}' x_{t1} \mathbf{1}\{t \leq \hat{s}\}$, $t = 1, \dots, n$, where \hat{s} is the value that minimizes the residual sum of squares.

STEP 2: Obtain a random sample of size n from the empirical distribution function of $\{\hat{u}_t\}_{t=1}^n$. (Observe that as the regression model has an intercept, $\sum_{t=1}^n \hat{u}_t = 0$. Denote the random sample by $\{u_t^*\}_{t=1}^n$.)

STEP 3: Compute the regression model

$$y_t^* = \hat{\beta}' x_t + u_t^*, \quad t = 1, \dots, n.$$

STEP 4: Compute the bootstrap statistics \mathcal{W}^* , $\mathcal{W}_{.05}^*$, $\mathcal{W}_{.15}^*$, CUSUM* and Chow test as with the original data, but using instead of (4.1),

$$y_t^* = \hat{\beta}' x_t + \delta' x_{t1} \mathbf{1}\{t \leq s\} + u_t^*, \quad t = 1, \dots, n.$$

With heteroscedastic errors, that is when $(a, b) = (0.125, 0.25)$, we need to modify STEPS 2 to 4 as follows.

STEP 2': Obtain a random sample of size n , $\{\nu_t^*\}_{t=1}^n$, from the distribution

$$\Pr\{\nu_t^* = 1\} = \Pr\{\nu_t^* = -1\} = 1/2$$

STEP 3': Setting $h_t = x_t'(X_n'X_n)^{-1}x_t$, we compute the regression model

$$y_t^* = \hat{\beta}' x_t + \frac{|\hat{u}_t|}{(1-h_t)} \nu_t^*, \quad t = 1, \dots, n.$$

STEP 4': Compute the bootstrap statistics \mathcal{W}^* , $\mathcal{W}_{.05}^*$, $\mathcal{W}_{.15}^*$, CUSUM* and Chow test as with the original data, but using instead of (4.1)

$$y_t^* = \hat{\beta}' x_t + \delta' x_{t1} \mathbf{1}\{t \leq s\} + \frac{|\hat{u}_t|}{(1-h_t)} \nu_t^*, \quad t = 1, \dots, n.$$

Because the distribution of, for instance, \mathcal{W}^* is not known, we approximate it using the *WARP* algorithm as proposed by Giacomini, Politis and White (2007), which we now describe. For each of the $k = 1, \dots, K$ Monte-Carlo samples, we draw one bootstrap sample and compute the statistic, for instance \mathcal{W}_k^* , $k = 1, \dots, K$. Then the critical value at the 5% level is obtained from the empirical distribution of \mathcal{T}_k^* , where \mathcal{T}_k^* is the statistic under consideration, for instance \mathcal{W}_k^* . We set

$K = 1000$ in this Monte-Carlo study. Note that we do not need to renormalize and recenter the statistics \mathcal{W} and \mathcal{W}^* by a_n and b_n . The results are given in Table 4.2.

TABLE 4.2 ABOUT HERE

Table 4.2 illustrates that there is a substantial improvement in the empirical size, being now much closer to the nominal one of 5%. This improvement is uniform across all the sample sizes, models and statistics employed to test the null hypothesis. Because of that, we decided to examine the power for the different statistics using the bootstrap algorithm. As in Andrews (1992), the coefficient(s) subject to change are those corresponding to the trend regressors x_{t1} . That is, we examine the power in the model

$$y_t = \beta' x_t + \delta' x_{t1} \mathbf{1}\{t \leq s\} + \sigma(w_{t3}) u_t, \quad t = 1, \dots, n,$$

as in (4.1) but with $\delta \neq 0$. We consider three scenarios, namely $\delta = (\delta_1, 0)'$, $(\delta_1, \delta_1)'$ and $(0, \delta_1)'$ for $\delta_1 = 1.0$ and 1.5 . We denote the three scenarios respectively as $\theta = 0, \pi/4$ and $\pi/2$ in Tables 4.3 and 4.4 below. Moreover, to gain some light about the performance of \mathcal{W} relative to, for example, $\mathcal{W}_{.05}$ or $\mathcal{W}_{.15}$, we have explored two situations regarding the time of the break s . In the first one, the break lies in the “middle” of the sample, whereas in the second one, the break occurs towards the end of the sample. In particular, in Table 4.3 we present the results when $s = \lceil n\tau \rceil$ for $\tau = 1/2, 3/4, 9/10$ and $19/20$, whereas in Table 4.4 the break occurs at time $s = n - cn^{1/2}$, with $c = 1, 2/3, 1/2$ and $1/3$.

TABLES 4.3 AND 4.4 ABOUT HERE

We now comment on Tables 4.3 and 4.4. Both tables appear to indicate that the power function is independent of θ although the power seems to be affected by the heteroscedasticity, being lower when $(a, b) = (0.125, 0.4)$. Also, the tables suggest that the power increases with the size of the break as one would expect that the power should increase as the alternative hypothesis becomes far apart from the null. In particular, Table 4.3 suggests that our test does not perform worse than the conventional ones when the break is in the “middle” and the power is very comparable to that of $\mathcal{W}_{.05}$ for all “ n ” and values of τ , although when the break is towards the end of the sample our test performs better than $\mathcal{W}_{.05}$. On the other hand, \mathcal{W} seems to perform better than $\mathcal{W}_{.15}$ in moderate sample sizes when $\tau \geq 9/10$, being the deterioration of the power of the latter even bigger when $\tau = 19/20$. Observe that when $\tau = 9/10$, the time of the break is beyond the interval at which we compute the statistic $\mathcal{W}_{.15}$. The latter might indicate that $\mathcal{W}_{.15}$, and in general conventional tests, is not a very useful statistic to detect a break when it occurs towards the end (beginning) of the sample.

Next, Table 4.4 suggests that \mathcal{W} outperforms $\mathcal{W}_{.15}$ and also $\mathcal{W}_{.05}$ when the break is at the “end” of the sample although in less degree with $\mathcal{W}_{.05}$. Notice also that the power function is smaller with $\mathcal{W}_{.15}$ than with $\mathcal{W}_{.05}$. But more importantly, as we can expect from the results of Section 2.3, the power function of the conventional statistics does not appear to converge to 1, but on the contrary the power seems to remain constant with the sample size. On the other hand, the power of \mathcal{W} increases with the sample size which corroborates the consistency of \mathcal{W} even when the break occurs towards the end (or beginning) of the sample. So, the main conclusion that we could draw from Tables 4.3 and 4.4 is that the statistic \mathcal{W} appears to be more desirable than conventional tests, not only because its power behaviour appears to be superior but also as we do not need to choose τ to compute the statistic, we avoid the unpleasant feature that depending on τ we might obtain different conclusions with the same data set.

5. PROOFS OF RESULTS

5.1. Proof of Proposition 1.

Let $q = 6\zeta - 3$ and assuming, without loss of generality, that $\mu_x = 0$, we define the sequences $\{\dot{x}_t\}_{t \in \mathbb{Z}}$ and $\{\dot{u}_t\}_{t \in \mathbb{Z}}$ as

$$\dot{x}_t = \sum_{0 \leq i \leq Ct^{1/q}} \chi_i \varepsilon_{t-i}; \quad \dot{u}_t = \sum_{0 \leq i \leq Ct^{1/q}} \phi_i \eta_{t-i},$$

where C is a large enough positive constant but fixed. Note that $\{\dot{x}_t\}_{t \in \mathbb{Z}}$ and $\{\dot{u}_t\}_{t \in \mathbb{Z}}$ behave as $MA(Ct^{1/q})$ sequences. So, abbreviating $\{\dot{x}_t \dot{u}_t\}_{t \in \mathbb{Z}}$ by $\{\dot{v}_t\}_{t \in \mathbb{Z}}$, \dot{v}_t is independent of \dot{v}_s if $s < t$ and $t - s > Ct^{1/q}$. However, contrary to $\{x_t\}_{t \in \mathbb{Z}}$ or $\{u_t\}_{t \in \mathbb{Z}}$, $\{\dot{v}_t\}_{t \in \mathbb{Z}}$ is not covariance stationary as $\mathbb{E}\|\dot{v}_t\|^2$ depends on t .

We first show that

$$(5.1) \quad \mathbb{E} \sup_{1 \leq k \leq n} \left\| \sum_{t=1}^k v_t - \sum_{t=1}^k \dot{v}_t \right\| = o\left(n^{\frac{\zeta+2}{4(\zeta+1)}}\right),$$

where $\{x_t u_t\}_{t \in \mathbb{Z}} =: \{v_t\}_{t \in \mathbb{Z}}$. (5.1) implies that it suffices to show (2.7) with $\{\dot{v}_t\}_{t \in \mathbb{Z}}$ replacing $\{v_t\}_{t \in \mathbb{Z}}$ there. To that end, denote $\{v_t - \dot{v}_t\}_{t \in \mathbb{Z}}$ by $\{\ddot{v}_t\}_{t \in \mathbb{Z}}$, and likewise $\{\ddot{x}_t\}_{t \in \mathbb{Z}}$ and $\{\ddot{u}_t\}_{t \in \mathbb{Z}}$. Assuming, without loss of generality, that $n = 2^d$, Wu's (2007) Proposition 1 implies that the left side of (5.1) is bounded by

$$(5.2) \quad \sum_{p=0}^d \left[\sum_{r=1}^{2^{d-p}} \mathbb{E} \left\| \sum_{t=2^p(r-1)+1}^{2^p r} \ddot{v}_t \right\|^2 \right]^{1/2} \leq C \sum_{p=0}^d 2^{\frac{p}{3}} \left[\sum_{r=1}^{2^{d-p}} r^{-1/3} \right]^{\frac{1}{2}} = O\left(n^{1/3} \log n\right)$$

as we now show. Because $\ddot{v}_t = -\ddot{x}_t \ddot{u}_t + x_t \ddot{u}_t + \ddot{x}_t u_t$, by standard inequalities, it suffices to show (5.2) with \ddot{v}_t replaced by, for instance, $x_t \ddot{u}_t$. Now, for $t_1 \leq t_2$, A1 implies that

$$\|\mathbb{E}(x_{t_1} \ddot{u}_{t_1} x_{t_2} \ddot{u}_{t_2})\| = \|\mathbb{E}(\ddot{u}_{t_1} \ddot{u}_{t_2}) \mathbb{E}(x_{t_1} x_{t_2})\| \leq Ct_1^{-1/3} \sum_{j=1}^{\infty} j^{-\zeta} (t_2 - t_1 + j)^{-\zeta}$$

because $|\mathbb{E}(\ddot{u}_{t_1} \ddot{u}_{t_2})| \leq C \sum_{j=1}^{\infty} j^{-\zeta} (t_2 - t_1 + j)^{-\zeta}$. So, because $q = 6\zeta - 3$ and $\zeta > 3/2$,

$$\mathbb{E} \left\| \sum_{t=2^p(r-1)+1}^{2^p r} x_t \ddot{u}_t \right\|^2 \leq C \sum_{t=2^p(r-1)+1}^{2^p r} t^{-1/3} \leq C 2^{2p/3} r^{-1/3},$$

and (5.2) follows because $n = 2^d$. Notice that $\text{Var}(\ddot{u}_t)$ and $\text{Var}(\ddot{x}_t)$ are $O(t^{-1/3})$.

To show that (2.7) holds true for $\{\dot{v}_t\}_{t \in \mathbb{Z}}$, we employ standard blocking arguments. For that purpose, consider blocks $A_\ell = \{t : n_{\ell-1} < t \leq n_{\ell-1} + \ell^{1/\zeta}\}$ and $B_\ell = \{t : n_{\ell-1} + \ell^{1/\zeta} < t \leq n_\ell\}$, where by definition $n_\ell = n_{\ell-1} + \ell^{1/\zeta} + \ell^{1/q}$; $\ell \geq 1$, for some $1 < \zeta < 2$, and the sequences

$$\xi_\ell = \sum_{t \in A_\ell} \dot{v}_t; \quad \epsilon_\ell = \sum_{t \in B_\ell} \dot{v}_t.$$

Notice that $\sum_{t=1}^{n_\ell} \dot{v}_t = \sum_{j=1}^{\ell} (\xi_j + \epsilon_j)$. Let $\bar{\ell}$ be the value such that $n_{\bar{\ell}-1} < n \leq n_{\bar{\ell}}$, so $C^{-1} \leq \bar{\ell} n^{-\zeta/(\zeta+1)} \leq C$. We first show that

$$(5.3) \quad \text{Pr} \left\{ \sup_{1 \leq \ell \leq \bar{\ell}} \left\| \sum_{t=1}^{n_\ell} \dot{v}_t - \Xi_{n_\ell}^{1/2} \mathcal{B}(n_\ell) \right\| > z \right\} \leq Cz^{-4} n^{\frac{\zeta+2}{\zeta+1}}.$$

By construction and A1, $\{\xi_\ell\}_{\ell \geq 1}$ is a sequence of independently distributed random variables with finite 4th moments. So, by Götze and Zaitsev's (2007) Theorem 4 (Proposition 1), we can find a sequence of *iid* normal random variables $\{\nu_j\}_{j \geq 1}$ such that

$$\Pr \left\{ \sup_{1 \leq \ell \leq \bar{\ell}} \left\| \sum_{j=1}^{\ell} \xi_j - \sum_{j=1}^{\ell} \mathbb{E}^{1/2}(\xi_j \xi'_j) \nu_j \right\| > z \right\} \leq Cz^{-4} \sum_{j=1}^{n^{\zeta/(\zeta+1)}} \mathbb{E} \|\xi_j\|^4 \\ = Cz^{-4} n^{\frac{\zeta+2}{\zeta+1}}$$

because $\mathbb{E} \|\xi_j\|^4 = O(j^{2/\zeta})$ and $\bar{\ell} \leq Cn^{\zeta/(\zeta+1)}$. Next, because $\{\epsilon_j\}_{j \geq 1}$ is a sequence of independent random variables with finite fourth moments, by the law of iterated logarithms (LIL), $\overline{\lim}_{1 \leq \ell \leq \bar{\ell}} \left\| \sum_{j=1}^{\ell} \epsilon_j / \ell^{(q+1)/2q} \log_2 \ell \right\| = 1$ a.s. by Shao's (1995) Theorem 3.2. So, $\sup_{1 \leq \ell \leq \bar{\ell}} \left\| \sum_{j=1}^{\ell} \epsilon_j \right\| = o\left(n^{\frac{\zeta+2}{4(\zeta+1)}}\right)$ a.s. and we conclude that

$$(5.4) \quad \Pr \left\{ \sup_{1 \leq \ell \leq \bar{\ell}} \left\| \sum_{t=1}^{n_\ell} \dot{v}_t - \sum_{j=1}^{\ell} \mathbb{E}^{1/2}(\xi_j \xi'_j) \nu_j \right\| > z \right\} \leq Cz^{-4} n^{\frac{\zeta+2}{\zeta+1}}.$$

Now using the notation $\mathbb{E} \left(\sum_{t \in A_j} v_t \sum_{t \in A_j} v'_t \right) = \mathbb{E} \left(\sum_{t \in A_j} v_t \right)^2$, because proceeding as with the proof of (5.2),

$$\left| \mathbb{E}(\xi_j \xi'_j) - \mathbb{E} \left(\sum_{t \in A_j} v_t \right)^2 \right| \leq C \sum_{t=n_{j-1}+1}^{n_{j-1}+j^{1/\zeta}} t^{-1/3} \leq Cj^{(2-\zeta)/3\zeta}$$

as $n_j = \sum_{h=1}^j (h^{1/\zeta} + h^{1/q}) \leq Cj^{(\zeta+1)/\zeta}$, we have that

$$\Pr \left\{ \sup_{1 \leq \ell \leq \bar{\ell}} \left\| \sum_{j=1}^{\ell} \left\{ \mathbb{E}^{1/2}(\xi_j \xi'_j) - \mathbb{E}^{1/2} \left(\sum_{t \in A_j} v_t \right)^2 \right\} \nu_j \right\| > z \right\} \leq Cz^{-4} n^{\frac{\zeta+2}{\zeta+1}}$$

using that $(a-b)^2 \leq a^2 - b^2$ for $a > b > 0$ and Levy's inequality, which also implies that $\Pr \left\{ \sup_{1 \leq \ell \leq \bar{\ell}} \left\| \sum_{j=1}^{\ell} \mathbb{E}^{1/2}(\epsilon_j \epsilon'_j) \nu_j \right\| > z \right\} \leq Cz^{-4} n^{\frac{\zeta+2}{\zeta+1}}$. So, we conclude that

in (5.4) we can replace $\mathbb{E}^{1/2}(\xi_j \xi'_j)$ by $\left(\mathbb{E} \left(\sum_{t \in A_j} v_t \right)^2 + \mathbb{E} \left(\sum_{t \in B_j} v_t \right)^2 \right)^{1/2}$ and standard arguments imply that

$$\Pr \left\{ \sup_{1 \leq \ell \leq \bar{\ell}} \left\| \sum_{t=1}^{n_\ell} \dot{v}_t - \sum_{j=1}^{\ell} \left(\mathbb{E} \left(\sum_{t \in A_j \cup B_j} v_t \right)^2 \right)^{1/2} \nu_j \right\| > z \right\} \leq Cz^{-4} n^{\frac{\zeta+2}{\zeta+1}}.$$

However because A1 implies that $\mathbb{E} \left(\sum_{t \in A_j \cup B_j} v_t \right)^2 = \Xi_{j^{1/q} + j^{1/\zeta}} (j^{1/q} + j^{1/\zeta})$ and $\|\Xi_{n_j} - \Xi_{j^{1/q} + j^{1/\zeta}}\| = O(j^{-1/2\zeta})$, by Levy's inequality again,

$$\Pr \left\{ \sup_{1 \leq \ell \leq \bar{\ell}} \left\| \sum_{j=1}^{\ell} \left\{ \left(\mathbb{E} \left(\sum_{t \in A_j \cup B_j} v_t \right)^2 \right)^{1/2} - \Xi_{n_\ell}^{1/2} (j^{1/q} + j^{1/\zeta})^{1/2} \right\} \nu_j \right\| > z \right\} \\ \leq Cz^{-4} n^{\frac{\zeta+2}{\zeta+1}}$$

and because $n_j = \sum_{h=1}^j (h^{1/\zeta} + h^{1/q})$, we obtain that

$$\Xi_{n_\ell}^{1/2} \sum_{j=1}^{\ell} \left(j^{1/q} + j^{1/\zeta} \right)^{1/2} \nu_j \stackrel{d}{=} \Xi_{n_\ell}^{1/2} \sum_{j=1}^{n_\ell} \nu_j,$$

where “ $\stackrel{d}{=}$ ” denotes “distributed as”. This concludes the proof of (5.3), and that of (2.7), when the supremum is taken over those values of k for which there exists a \bar{k} satisfying $n_{\bar{k}} = k$.

Hence, to finish the proof we need to examine the approximation when $n_\ell < k < n_{\ell+1}$. But by Csörgo and Révész's (1981) Theorem 1.2, we know that

$$\max_{n_{\ell-1} \leq j \leq n_\ell} |\mathcal{B}(n_{\ell-1}) - \mathcal{B}(j)| = O\left(\ell^{1/2\zeta} \log_2^{1/2} \ell\right) \quad \text{a.s.}$$

and because $\mathbb{E} \left\| \sum_{s \leq t \leq r} v_t \right\|^4 = O\left((r-s)^2\right)$, the Borel Cantelli's theorem implies that $\max_{n_{\ell-1} \leq j \leq n_\ell} \left\| \sum_{j \leq t \leq n_\ell} v_t \right\| = O\left(n^{\frac{\zeta+2}{4(\zeta+1)}}\right)$ a.s.. This concludes the proof. ■

5.2. Proof of Theorem 1.

We begin with part (a). First, recall that $\tilde{\delta}^{(s)}$ is

$$A \left\{ \sum_{t=1}^s v_t - M_s M_n^{-1} \sum_{t=1}^n v_t \right\} - A \left\{ \frac{1}{s} \sum_{t=1}^s x_t \sum_{t=1}^s u_t - M_s M_n^{-1} \frac{1}{n} \sum_{t=1}^n x_t \sum_{t=1}^n u_t \right\}.$$

Because *LIL* implies that $\sup_{s \geq 1} \left| (s \log_2 s)^{-1/2} \sum_{t=1}^s w_t \right| = O(1)$ a.s., for either $w_t = x_t$ or $w_t = u_t$, and also that

$$\sup_{p^* < s} \left\| \frac{M_s - s\Sigma}{(2s \log_2 s)^{1/2}} \right\| = O(1) \quad \text{a.s.},$$

implies that $\sup_{p^* < s} \frac{s^{1/2}}{\log_2 s} \left\| \frac{M_s}{s} \left(\frac{M_n}{n} \right)^{-1} - I_p \right\| = O(1)$ a.s.. So uniformly in s ,

$$(5.5) \quad \tilde{\delta}^{(s)} = A \left\{ \sum_{t=1}^s v_t - \frac{s}{n} \sum_{t=1}^n v_t \left(1 + O\left(\frac{\log_2^{1/2} s}{s^{1/2}} \right) \right) \right\} \quad \text{a.s..}$$

Next, because Proposition 1 and then *LIL* implies that $\sup_{p^* < s \leq \log n} \left\| s^{-1/2} \sum_{t=1}^s v_t \right\| = o_p\left(\log_2^{1/2} n\right)$, we obtain that

$$(5.6) \quad a_n \sup_{p^* < s < \log n} \left\| \left(\frac{n}{(n-s)s} \right)^{1/2} \left\{ \sum_{t=1}^s v_t - \frac{s}{n} \sum_{t=1}^n v_t \right\} \right\| - b_n \xrightarrow{P} -\infty,$$

whereas

$$(5.7) \quad a_n \sup_{n - \log n < s \leq n - p^*} \left\| \left(\frac{n}{(n-s)s} \right)^{1/2} \left\{ \sum_{t=1}^s v_t - \frac{s}{n} \sum_{t=1}^n v_t \right\} \right\| - b_n \xrightarrow{P} -\infty,$$

because the expression inside the absolute value is $(s(n-s)/n)^{-1/2} \sum_{t=s+1}^n v_t + (sn/(n-s))^{-1/2} \sum_{t=1}^n v_t$ and by *LIL* with $v_t^* = v_{n-t+1}$,

$$(5.8) \quad \sup_{n - \log n < s \leq n - p^*} \left\| \frac{1}{(n-s)^{1/2}} \sum_{t=s+1}^n v_t \right\| = \sup_{p^* < s < \log n} \left\| \frac{1}{s^{1/2}} \sum_{t=1}^s v_t^* \right\| = o_p\left(\log_2^{1/2} n\right).$$

So, proceeding as above and recalling that Robinson (1998) implies that $\widehat{\Xi}_n - \Xi_n = O_p\left(n^{-1/2} \log_2^{1/2} n\right)$, we observe that the asymptotic distribution of $a_n \mathcal{W}^{1/2} -$

b_n is that of

$$(5.9) \quad \hat{\mathcal{W}} = a_n \max_{\log n \leq s \leq n - \log n} \hat{\mathcal{W}}(s)^{1/2} - b_n,$$

where

$$\hat{\mathcal{W}}(s) = \left(\frac{n}{(n-s)s} \left\{ \sum_{t=1}^s v_{t1} - \frac{s}{n} \sum_{t=1}^n v_{t1} \right\}' \hat{\Xi}_n^{-1} \left\{ \sum_{t=1}^s v_{t1} - \frac{s}{n} \sum_{t=1}^n v_{t1} \right\} \right),$$

and where $Av_t = v_{t1}$, that is the first p_1 components of v_t and $\hat{\Xi}_n = A\Xi_n A'$.

However, after noticing that by A1, $\|\hat{\Xi}_n - \hat{\Xi}_s\| = O(s^{-1/2})$, Proposition 1 implies that

$$\hat{\Xi}_n^{-1/2} \frac{1}{s^{1/2}} \left\{ \sum_{t=1}^s v_{t1} - \frac{s}{n} \sum_{t=1}^n v_{t1} \right\} = \frac{1}{s^{1/2}} \left\{ \sum_{t=1}^s \nu_t - \frac{s}{n} \sum_{t=1}^n \nu_t \right\} \left(1 + o_p \left(\frac{1}{\log^{1/2} n} \right) \right),$$

uniformly in $\log n \leq s \leq n - \log n$. Thus, we conclude that the asymptotic distribution of $\hat{\mathcal{W}}$ in (5.9), and that of $a_n \mathcal{W}^{1/2} - b_n$, is governed by that of

$$(5.10) \quad a_n \max_{\log n \leq s \leq n - \log n} \left(\frac{n}{(n-s)s} \left\{ \sum_{t=1}^s \nu_t - \frac{s}{n} \sum_{t=1}^n \nu_t \right\}' \left\{ \sum_{t=1}^s \nu_t - \frac{s}{n} \sum_{t=1}^n \nu_t \right\} \right)^{1/2} - b_n.$$

On the other hand, by standard functional central limit theorem (FCLT) and the continuous mapping theorem, for any $\rho > 0$,

$$(5.11) \quad a_n \sup_{\log n \leq s \leq n - [n\rho]} \left\| \left(\frac{s}{n} \right)^{1/2} \frac{1}{(n-s)^{1/2}} \sum_{t=s+1}^n \nu_t \right\| - b_n \xrightarrow{P} -\infty$$

$$(5.12) \quad a_n \sup_{[n\rho] < s \leq n - \log n} \left\| \left(\frac{n-s}{n} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=1}^s \nu_t \right\| - b_n \xrightarrow{P} -\infty.$$

So, abbreviating $\left(\frac{n-s}{ns} \right)^{1/2}$ by $\varkappa_n(s)$, (5.11) – (5.12) imply that the (asymptotic) distribution of (5.10) is that of

$$a_n \max \left\{ \sup_{\log n \leq s < [n\rho]} \left| \varkappa_n(s) \sum_{t=1}^s \nu_t \right|^2; \sup_{\log n \leq s < [n\rho]} \left| \varkappa_n(s) \sum_{t=1}^s \nu_t^* \right|^2 \right\}^{1/2} - b_n$$

using (5.8) and where the sequences $\{\nu_t\}_{t \geq 1}$ and $\{\nu_t^*\}_{t \geq 1}$ are independent sequences of mutually independent $\mathcal{N}(0, 1)$ random variables. However, because $\{\nu_t\}_{t \geq 1}$ is a sequence of iid Gaussian random variables, we know that

$$(5.13) \quad \sup_{n/\log n \leq s < [n\rho]} \left| \varkappa_n(s) \sum_{t=1}^s \nu_t \right| = O_p(\log_3 n),$$

which implies that $\hat{\mathcal{W}}$ in (5.9) is

$$a_n \max \left\{ \sup_{\log n \leq s < \frac{n}{\log n}} \left| \varkappa_n(s) \sum_{t=1}^s \nu_t \right|^2; \sup_{\log n \leq s < \frac{n}{\log n}} \left| \varkappa_n(s) \sum_{t=1}^s \nu_t^* \right|^2 \right\}^{1/2} - b_n + o_p(1).$$

But, $\sup_{\log n \leq s < n/\log n} \left| \left(\frac{n-s}{n} \right)^{1/2} - 1 \right| = O(\log^{-1} n)$, so we can conclude that

$$\hat{\mathcal{W}} = a_n \max \left\{ \sup_{\log n \leq s < \frac{n}{\log n}} \left| \frac{1}{s^{1/2}} \sum_{t=1}^s \nu_t \right|^2; \sup_{\log n \leq s < \frac{n}{\log n}} \left| \frac{1}{s^{1/2}} \sum_{t=1}^s \nu_t^* \right|^2 \right\}^{1/2} - b_n + o_p(1)$$

Denoting by $U(s^*)$ the Ornstein-Uhlenbeck process, by the change of time $s \rightarrow e^{s^*}$, we obtain that

$$\sup_{\log n \leq s < n/\log n} \frac{1}{s^{1/2}} \sum_{t=1}^s \nu_t = \sup_{\log_2 n < s^* \leq \log n - \log_2 n} U(s^*).$$

From here the proof follows by Lemma 2.2 of Horváth (1993).

Next part (b). The proof proceeds as that of part (a) after observing that by Robinson (1998), $\widehat{F}_n^{(s)} = \widetilde{\delta}^{(s)}$ and $\widehat{\Xi}_n - \Xi = o_p(1)$. This concludes the proof of the theorem. \blacksquare

5.3. Proof of Theorem 2.

First of all, proceeding as in the proof of Theorem 1, it suffices to examine the behaviour of

$$(5.14) \quad Q_n = a_n \sup_{p^* < s \leq n-p^*} Q_n^{1/2}(s) - b_n,$$

where $Q_n(s) = (U_n(s) + \Lambda_n(s; s_0))' (A \Xi A')^{-1} (U_n(s) + \Lambda_n(s; s_0))$ with

$$(5.15) \quad U_n(s) = \left(\frac{n}{n-s} \right)^{1/2} \frac{1}{s^{1/2}} \left(\sum_{t=1}^s \nu_{t1} - \frac{s}{n} \sum_{t=1}^n \nu_{t1} \right),$$

and $\Lambda_n(s; s_0)$ given in (2.17).

Now, taking for simplicity $\mathbb{E}(\widetilde{x}_t \widetilde{x}'_{t1}) = I_p A'$, we have that

$$(5.16) \quad \Lambda_n(s; s_0) = g_n \frac{s_0}{s^{1/2}} \left(\frac{n-s}{n} \right)^{1/2} \mathbf{1}(s_0 \leq s) + g_n \left(\frac{s}{n} \right)^{1/2} \frac{n-s_0}{(n-s)^{1/2}} \mathbf{1}(s < s_0),$$

so that when $s_0^{-1} = o(\log_2^{-1} n)$, because $\|g_n\| < C$ we can easily conclude that

$$|\Lambda_n(s; s_0)| = C \left\{ s_0 s^{-1/2} \mathbf{1}(s_0 \leq s) + s^{1/2} \mathbf{1}(s < s_0) \right\}$$

which implies that $a_n \sup_{p^* < s \leq n-p^*} |\Lambda_n(s; s_0)| - b_n \rightarrow \infty$. Similarly, when $(n-s_0)^{-1} = o(\log_2^{-1} n)$, we also have that $a_n \sup_{p^* < s \leq n-p^*} |\Lambda_n(s; s_0)| - b_n \rightarrow \infty$. From here, the conclusion of the theorem is standard. \blacksquare

Before we give the proof of Theorem 3, we shall give a lemma.

Lemma 1. *Under the assumptions of Theorem 3, if $\sup_{p^* < s < [n/2]} |\Lambda_n(s; s_0)| = o(b_n^{-1/2})$,*

$$\begin{aligned} & \Pr \left\{ a_n \sup_{p^* < s < [n/2]} |U_n(s) + \Lambda_n(s; s_0)| - b_n < x \right\} \\ &= \Pr \left\{ a_n \sup_{p^* < s < [n/2]} |U_n(s)| - b_n < x \right\} + o(1). \end{aligned}$$

Proof. The proof is immediate. Indeed, we have that

$$a_n \sup_{p^* < s < [n/2]} |U_n(s) + \Lambda_n(s; s_0)| - b_n = b_n \left(\sup_{p^* < s < [n/2]} \left| \frac{a_n}{b_n} U_n(s) + o(b_n^{-1}) \right| - 1 \right).$$

Now, using the inequality $|a| - |b| \leq |a + b| \leq |a| + |b|$, the conclusion of the lemma is standard. \square

5.4. Proof of Theorem 3.

By symmetry, we can focus on the situation when the break occurs at time $s_0 \leq [n/2]$. We shall prove parts (a), (b) and (c) simultaneously. Again, as in the proof of Theorem 2, it suffices to examine the behaviour of Q_n in (5.14). The proof will be done in three steps. Namely, when s_0 lies in the interval (i) $[n\tau] \leq s_0 < [n/2]$, (ii) $[n/\log n] \leq s_0 < [n\tau]$ and when (iii) $\log_2 n \leq s_0 < [n/\log n]$.

We begin with case (i). Because by standard arguments, $\sup_{[n\tau] \leq s < [n/2]} |U_n(s)| = O_p(1)$ with $U_n(s)$ defined in (5.15), we conclude that when $g_n^{-1} = o\left(s_0^{1/2}/\log_2^{1/2} n\right)$ and using (5.16),

$$\sup_{[n\tau] \leq s < [n/2]} |U_n(s) + \Lambda_n(s; s_0)| = O_p(1) + O\left(a_n^{1/2} c_n\right),$$

where $c_n^{-1} = o(1)$. So, $Q_n^{-1} \rightarrow_P 0$ which implies that the test rejects with probability 1 as n increases to infinity.

Next, proceeding as above, when $g_n = o\left(\log_2^{1/2} n/s_0^{1/2}\right)$ we obtain that

$$\sup_{[n\tau] \leq s < [n/2]} |U_n(s) + \Lambda_n(s; s_0)| = O_p(1) + o\left(a_n^{1/2}\right)$$

so that $Q_n \rightarrow_P -\infty$, which implies that the “max” in the definition of Q_n is when $s < [n\tau]$. Now because $s < [n\tau]$ and hence $s_0 > s$, we obtain that

$$(5.17) \quad \Lambda_n(s; s_0) = O\left(g_n \left(\frac{s}{n}\right)^{1/2} \frac{n-s_0}{(n-s)^{1/2}}\right) = o\left(\left(\frac{s \log_2 n}{n}\right)^{1/2}\right).$$

So, $a_n \sup_{[n/\log n] < s < [n\tau]} |\Lambda_n(s; s_0)| = o(b_n)$ and proceeding as in the proof of Theorem 1, cf. (5.13), we have that the “max” of Q_n is achieved when $s < [n/\log n]$. But in this region, uniformly in s , (5.17) = $o\left((\log_2 n/\log n)^{1/2}\right) = o\left(b_n^{-1/2}\right)$, so by Lemma 1, we conclude part (b) of the theorem. From parts (a) and (b), we easily conclude part (c) and so it is omitted. This finishes the proof of (i).

Next we examine case (ii). By definition of $\Lambda_n(s; s_0)$, we have that

$$\sup_{p^* < s < [n/\log n]} |\Lambda_n(s; s_0)| \geq C g_n \frac{s_0}{n^{1/2}} \log^{1/2} n.$$

So, when $g_n^{-1} = o\left(s_0^{1/2}/\log_2^{1/2} n\right)$ the last displayed inequality implies that

$$g_n^{-1} \frac{n^{1/2}}{s_0 \log^{1/2} n} = o\left(\left(\frac{n}{s_0 \log n \log_2 n}\right)^{1/2}\right) = o\left(b_n^{-1/2}\right).$$

Hence $|a_n \Lambda_n(s; s_0)|^{-1} = o(b_n^{-1})$, which implies that the test rejects with probability 1 as n increases to infinity. Next, when $g_n = o\left(\log_2^{1/2} n/s_0^{1/2}\right)$, we have that $\sup_{[n\tau] < s < [n/2]} |\Lambda_n(s; s_0)| = o\left(g_n \frac{s_0}{n^{1/2}}\right) = o\left(b_n^{1/2}\right)$ as is $\sup_{[n/\log n] < s \leq [n\tau]} |\Lambda_n(s; s_0)|$ as we now show. Indeed,

$$(5.18) \quad |\Lambda_n(s; s_0)| = o\left(\left(\frac{s_0}{s}\right) \log_2^{1/2} n \mathbf{1}(s_0 < s) + \left(\frac{s}{n}\right)^{1/2} \left(\frac{n-s_0}{s_0}\right)^{1/2} \log_2^{1/2} n \mathbf{1}(s \leq s_0)\right),$$

so that $\sup_{[n/\log n] < s \leq [n\tau]} |\Lambda_n(s; s_0)| = o\left(b_n^{1/2}\right)$. Thus, proceeding as in the proof of Theorem 1, cf. (5.13), we have that the “max” in Q_n is when $p^* < s < [n/\log n]$. But in that region, $\sup_{p^* < s < [n/\log n]} |\Lambda_n(s; s_0)| = o\left(b_n^{-1/2}\right)$ and by Lemma 1, we

conclude part (b) of the theorem. Part (c) follows easily from parts (a) and (b) and so it is omitted.

Finally, we examine case (iii). If $s_0 \leq s$, we know that $|\Lambda_n(s; s_0)| = Cg_n s_0 / s^{1/2}$. So, we have that when $g_n^{-1} = o\left(s_0^{1/2} / \log_2^{1/2} n\right)$,

$$|\Lambda_n(s; s_0)|^{-1} = o\left((s/s_0)^{1/2} \log_2^{-1/2} n\right)$$

which implies that $\inf |\Lambda_n(s; s_0)|^{-1} = o\left(\log_2^{-1/2} n\right) = o\left(b_n^{-1/2}\right)$. So $Q_n^{-1} \rightarrow_P 0$ and hence the test rejects with probability 1 as n increases to infinity. Next, when $g_n = o\left(\log_2^{1/2} n / s_0^{1/2}\right)$, the proof proceeds as in cases (i) or (ii) and so it is omitted. Part (c) as before follows from the results of parts (a) and (b). ■

5.5. Proof of Proposition 2.

First we notice that we can assume without loss of generality that $E(u_t u_t' | x_t) = \Sigma$. Indeed, if $E(u_t u_t' | x_t) = \Sigma(x_t)$, it implies that by A3,

$$u_t \otimes P(x_t) = h_v(\hat{v}_t) \Sigma^{1/2}(x_t) \otimes P(x_t) =: h_v(\hat{v}_t) \otimes \tilde{P}(x_t),$$

where under suitable regularity conditions on $\Sigma(x_t)$ and $P(x_t)$, we have that $\tilde{P}(x_t)$ is also a \mathbf{L}^4 -NED sequence of size $\iota > 1$. See for instance Davidson (1994, Sec. 17.3). So, from now on, we assume that $\Sigma(x_t) = \Sigma$. The proof proceeds similarly to that of Proposition 1. Let $\hat{g}(\eta_t, \dots, \eta_{t-Ct^{1/q}}) = \mathbb{E}(u_t | \eta_t, \dots, \eta_{t-Ct^{1/q}})$, $\hat{P}(\varepsilon_t, \dots, \varepsilon_{t-Ct^{1/q}}) = \mathbb{E}(P(x_t) | \varepsilon_t, \dots, \varepsilon_{t-Ct^{1/q}})$, $\{\check{\varepsilon}_t\}_{t \in \mathbb{Z}} = \{u_t \otimes P(x_t)\}_{t \in \mathbb{Z}}$ and $\{\check{\varepsilon}_t\}_{t \in \mathbb{Z}} = \{\varepsilon_t - \hat{\varepsilon}_t\}_{t \in \mathbb{Z}}$ with $\hat{\varepsilon}_t = \hat{g}(\eta_t, \dots, \eta_{t-Ct^{1/q}}) \otimes \hat{P}(\varepsilon_t, \dots, \varepsilon_{t-Ct^{1/q}})$.

Now, A4 implies that $\mathbb{E} \left\| \sum_{t=1}^k \check{\varepsilon}_t \right\|^2 \leq C \sum_{t=1}^k t^{-2\iota/q}$. So, proceeding as in Proposition 1, cf. (5.1) – (5.2), we obtain that

$$\mathbb{E} \sup_{1 \leq k \leq n} \left\| \sum_{t=1}^k \check{\varepsilon}_t \right\|^2 \leq \sum_{p=0}^d \left[\sum_{r=1}^{2^{d-p}} \mathbb{E} \left\| \sum_{t=2^p(r-1)+1}^{2^p r} \check{\varepsilon}_t \right\|^2 \right]^{1/2} = O\left(n^{1/3} \log n\right).$$

So, it suffices to consider the strong approximation of

$$\sum_{t=1}^k \hat{g}_u(\eta_t, \dots, \eta_{t-Ct^{1/q}}) \otimes \hat{P}(\varepsilon_t, \dots, \varepsilon_{t-Ct^{1/q}}) = \sum_{t=1}^k \check{\varepsilon}_t.$$

But the latter follows proceeding as we did in the proof of Proposition 1 after we observe that $\hat{g}_u(\eta_t, \dots, \eta_{t-Ct^{1/q}})$ and $\hat{P}(\varepsilon_t, \dots, \varepsilon_{t-Ct^{1/q}})$ are $MA(Ct^{1/q})$ and A3 implies that $\{\check{\varepsilon}_t\}_{t \in \mathbb{Z}}$ satisfies the same conditions of $\{\hat{v}_t\}_{t \in \mathbb{Z}}$ in Proposition 1. ■

5.6. Proof of Theorem 4.

Because the proof of parts (a) and (b) are similarly handled, we shall explicitly prove part (b). For that purpose, we first notice that

$$(5.19) \quad \widehat{\Delta}^{(s)} - \Delta = O_p\left(m^{-1/2}\right) = o_p(1)$$

$$(5.20) \quad G(\check{\vartheta}) =: \sup_{p^* < s} \left\| \frac{\sum_{t=1}^s \frac{\partial}{\partial \check{\vartheta}} g_t(\check{\vartheta}) - sG_{\check{\vartheta}}}{s^{1/2} \log_2^{1/2} s} \right\| = O_p(1)$$

as we now show. (5.20) holds true because mean value theorem implies that

$$\begin{aligned} \frac{1}{s} \sum_{t=1}^s \frac{\partial}{\partial \vartheta} g_t(\check{\vartheta}) &= \frac{1}{s} \sum_{t=1}^s \frac{\partial}{\partial \vartheta} g_t(\vartheta_0) + \frac{1}{s} \sum_{t=1}^s \left(\frac{\partial}{\partial \vartheta} g_t(\check{\vartheta}) - \frac{\partial}{\partial \vartheta} g_t(\vartheta_0) \right) \\ &= \frac{1}{s} \sum_{t=1}^s \frac{\partial}{\partial \vartheta} g_t(\vartheta_0) + O_p(n^{-1/2}). \end{aligned}$$

But A3, LIL and Proposition 2 imply that $G(\vartheta_0) = O_p(1)$. From here, (5.20) follows by straight arguments.

We now examine the behaviour of $((n-s)/n)^{1/2} s^{-1/2} lm(s)$. By (5.19) and standard procedures using $(\check{\vartheta} - \vartheta_0)$ and the mean value theorem for $g_t(\check{\vartheta})$,

$$\begin{aligned} lm^*(s) &= \left(\frac{n-s}{ns} \right)^{1/2} A \widehat{G}_{\check{\vartheta}}^0 \Delta^{-1} \sum_{t=1}^s g_t(\check{\vartheta}) \\ &= \left(\frac{n-s}{ns} \right)^{1/2} A \widehat{G}_{\check{\vartheta}}^0 \Delta^{-1} \left\{ \sum_{t=1}^s g_t(\vartheta_0) - \frac{s}{n} \sum_{t=1}^n g_t(\vartheta_0) \right\}. \end{aligned}$$

Now proceeding as with the proof of Theorem 1,

$$a_n \sup_{p^* < s < \log n} \|lm^*(s)\| - b_n \xrightarrow{P} -\infty$$

because by (5.20), LIL and Proposition 2,

$$\sup_{p^* < s < \log n} \left\| \widehat{G}_{\check{\vartheta}}^0 - G_{\check{\vartheta}} \right\| + \sup_{p^* < s < \log n} \left\| s^{-1/2} \sum_{t=1}^s g_t(\vartheta_0) \right\| = o_p(\log_2^{1/2} n).$$

Next, we examine the behaviour in the region $n - \log n < s \leq n - p^*$. But here, (5.20) implies that $a_n \sup_{n - \log n < s \leq n - p^*} \|lm^*(s)\| - b_n$ is governed by

$$a_n \sup_{n - \log n < s \leq n - p^*} \left\| \left(\frac{n-s}{ns} \right)^{1/2} G_{\check{\vartheta}} \Delta^{-1} \left\{ \sum_{t=1}^s g_t(\vartheta_0) - \frac{s}{n} \sum_{t=1}^n g_t(\vartheta_0) \right\} \right\| - b_n$$

which diverges in probability to $-\infty$ arguing as with (5.7). Now, in the region $\log n \leq s \leq n - \log n$ the proof proceeds as in Theorem 1 because (5.20) implies that it suffices to examine

$$\sup_{\log n \leq s \leq n - \log n} \left\| \left(\frac{n-s}{ns} \right)^{1/2} A G_{\check{\vartheta}} \Delta^{-1} \left\{ \sum_{t=1}^s g_t(\vartheta_0) - \frac{s}{n} \sum_{t=1}^n g_t(\vartheta_0) \right\} \right\|,$$

which is essentially the same as (5.5). This concludes the proof of the theorem. ■

REFERENCES

- [1] ANDREWS, D.W.K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, **59**, 817-858.
- [2] ANDREWS, D.W.K. (1992): "Tests for parameter instability and structural change with unknown change point," *Preprint*.
- [3] ANDREWS, D.W.K. (1993): "Tests for parameter instability and structural change with unknown change point," *Econometrica*, **61**, 821-856.
- [4] ANDREWS, D.W.K. (2003): "End-of-sample instability tests," *Econometrica*, **71**, 1661-1694.
- [5] ANDREWS, D.W.K. AND KIM, J.-Y. (2006): "End-of-sample cointegration breakdown tests," *Journal of Business and Economic Statistics*, **24**, 379-394.
- [6] ANDREWS, D.W.K. AND PLOBERGER, W. (1994): "Optimal test when a nuisance parameter is present only under the alternative," *Econometrica*, **62**, 1383-1414.
- [7] BAI, J. AND PERRON, P. (1998): "Estimating and testing linear models with multiple structural changes," *Econometrica*, **66**, 47-78.
- [8] 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. Ser. B.*, **37**, 149-192.
- [9] BURBIDGE, J.B., MAGEE, L. AND ROBB, A.L. (1988): "Alternative transformations to handle extreme values of the dependent variable," *Journal of the American Statistical Association*, **83**, 123-127.
- [10] CHAMBERLAIN, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, **34**, 305-334.
- [11] CHOW, G.C. (1960): "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, **28**, 591-605.
- [12] CSÖRGO, M. AND RÉVÉSZ, P. (1981): *Strong approximations in probability and statistics*. Academic Press.
- [13] DAVIDSON, J. (1994): *Stochastic limit theory*. Oxford University Press.
- [14] DARLING, D.A. AND ERDŐS, P. (1956): "A limit theorem for the maximum of normalized sums of independent random variables," *Duke Mathematical Journal*, **23**, 143-155.
- [15] EINMAHL, U. (1989): "Extensions of results of Komlós, Major and Tusnády to the multivariate case," *Journal of Multivariate Analysis*, **28**, 20-68.
- [16] GIACOMINI, R., POLITIS, D.N. AND WHITE, H. (2007): "A Warp-speed method for conducting Monte-Carlo experiments involving bootstrap estimators," *Preprint*.
- [17] GÖTZE, F. AND ZAITSEV, A.Y. (2007): "Bounds for the rate of strong approximation in the multidimensional invariance principle," *Preprint*.
- [18] HANSEN, L.P. (1982): "Large samples properties of generalized method of moments estimators," *Econometrica*, **50**, 1029-1054.
- [19] HIDALGO, J. (1992): "Adaptive estimation in time series regression models with heteroscedasticity of unknown form," *Econometric Theory*, **8**, 161-187.
- [20] HORVÁTH, L. (1993): "The maximum likelihood method for testing change in the parameters of normal observations," *Annals of Statistics*, **21**, 671-680.
- [21] KRÄMER, W. AND SONNBERGER, (1986): *The linear regression model under test*. Physica-Verlag.
- [22] PERRON, P. (2006): "Dealing with structural breaks," Forthcoming in *Pelgrave Handbook of Econometrics*, Vol. 1: *Econometric Theory*.
- [23] QUANDT, R.E. (1960): "Test of the hypothesis that a linear regression system obeys two separate regimes," *Journal of the American Statistical Association*, **55**, 324-330.
- [24] ROBINSON, P. M. (1988): "Best nonlinear three-stage least squares estimation of certain econometric models," *Econometrica*, **56**, 755-786.
- [25] ROBINSON, P. M. (1998): "Inference-without-smoothing in the presence of nonparametric autocorrelation," *Econometrica*, **66**, 1163-1182.
- [26] SHAO, Q.-M. (1995): "Strong Approximation theorems for independent random variables and their applications," *Journal of Multivariate Analysis*, **52**, 107-130.
- [27] WANG, Q., LIN, Y-X. AND GULATI, C.M. (2003): "Strong approximation for long memory processes with applications," *Journal of Theoretical Probability*, **16**, 377-389.
- [28] WHITE, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, **48**, 817-838.
- [29] WU, W. B. (2007): "Strong invariance principles for dependent random variables," *Annals of Probability*, **35**, 2294-2320.

For all the tables below, σ_0 and σ_1 means respectively that $\sigma(w_{t3}) = 1$ and $1 + \exp(.125 + .4w_{t3})$ and bootstrap critical values are computed based on the WARP algorithm.

n	\mathcal{W}		$\mathcal{W}_{.05}$		$\mathcal{W}_{.15}$		CUSUM		$\text{CHOW}_{n^{1/2}}$		$\text{CHOW}_{n/2}$	
	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1
100	0.10	0.09	0.35	0.34	0.28	0.29	0.01	0.00	0.17	0.16	0.08	0.08
200	0.15	0.12	0.40	0.38	0.22	0.21	0.02	0.02	0.13	0.13	0.06	0.06
300	0.13	0.12	0.28	0.28	0.18	0.17	0.02	0.02	0.10	0.10	0.05	0.06
500	0.18	0.16	0.25	0.24	0.18	0.18	0.03	0.03	0.11	0.10	0.06	0.06
1000	0.18	0.16	0.22	0.21	0.18	0.18	0.03	0.04	0.10	0.09	0.05	0.05

Table 4.1

Proportion of rejections, in 1000 Monte-Carlo experiments, under H_0 using the asymptotic critical values

n	\mathcal{W}		$\mathcal{W}_{.05}$		$\mathcal{W}_{.15}$		CUSUM		$\text{CHOW}_{n^{1/2}}$		$\text{CHOW}_{n/2}$	
	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1
100	0.06	0.05	0.06	0.05	0.06	0.05	0.06	0.07	0.04	0.03	0.07	0.06
200	0.07	0.06	0.06	0.07	0.06	0.07	0.05	0.08	0.06	0.04	0.05	0.05
300	0.05	0.05	0.04	0.04	0.04	0.05	0.05	0.07	0.03	0.03	0.04	0.04
500	0.06	0.06	0.04	0.04	0.05	0.04	0.06	0.07	0.05	0.05	0.05	0.05
1000	0.04	0.06	0.06	0.04	0.05	0.04	0.06	0.05	0.06	0.05	0.06	0.04

Table 4.2

Proportion of rejections, in 1000 Monte-Carlo experiments, under H_0 using the bootstrap critical values.

$s = [n\tau]$	δ	\mathcal{W}		$\mathcal{W}_{.05}$		$\mathcal{W}_{.15}$		CUSUM		CHOW	
		σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1
50	1	0.82	0.20	0.82	0.20	0.96	0.28	0.77	0.22	0.99	0.46
	1.5	1.00	0.42	1.00	0.42	1.00	0.57	0.98	0.43	1.00	0.82
75	1	0.69	0.16	0.69	0.16	0.86	0.22	0.16	0.08	0.96	0.38
	1.5	0.99	0.36	0.99	0.36	1.00	0.48	0.35	0.10	1.00	0.70
90	1	0.44	0.12	0.44	0.12	0.25	0.10	0.04	0.06	0.57	0.17
	1.5	0.77	0.22	0.77	0.22	0.47	0.14	0.02	0.05	0.91	0.30
95	1	0.12	0.07	0.12	0.07	0.09	0.07	0.05	0.06	0.28	0.08
	1.5	0.17	0.09	0.17	0.09	0.12	0.07	0.04	0.06	0.50	0.12
100	1	1.00	0.22	1.00	0.26	1.00	0.60	0.99	0.42	1.00	0.81
	1.5	1.00	0.67	1.00	0.78	1.00	0.94	1.00	0.77	1.00	0.99
150	1	0.96	0.21	0.97	0.27	1.00	0.48	0.48	0.14	1.00	0.65
	1.5	1.00	0.57	1.00	0.63	1.00	0.86	0.86	0.25	1.00	0.95
180	1	0.68	0.17	0.70	0.19	0.56	0.14	0.07	0.08	0.95	0.32
	1.5	0.98	0.34	0.99	0.38	0.88	0.27	0.07	0.09	1.00	0.61
190	1	0.43	0.14	0.43	0.14	0.14	0.06	0.04	0.07	0.64	0.17
	1.5	0.78	0.24	0.79	0.25	0.25	0.09	0.04	0.07	0.92	0.31
150	1	1.00	0.45	1.00	0.64	1.00	0.79	1.00	0.58	1.00	0.92
	1.5	1.00	0.93	1.00	0.99	1.00	0.99	1.00	0.93	1.00	1.00
225	1	1.00	0.34	1.00	0.51	1.00	0.69	0.73	0.21	1.00	0.85
	1.5	1.00	0.83	1.00	0.92	1.00	0.97	0.99	0.40	1.00	0.99
270	1	0.82	0.20	0.96	0.26	0.79	0.20	0.08	0.07	0.99	0.50
	1.5	1.00	0.47	1.00	0.63	0.98	0.43	0.15	0.08	1.00	0.86
285	1	0.56	0.15	0.67	0.14	0.16	0.08	0.04	0.07	0.83	0.23
	1.5	0.91	0.29	0.97	0.34	0.33	0.13	0.03	0.07	0.99	0.53
250	1	1.00	0.59	1.00	0.93	1.00	0.97	1.00	0.82	1.00	0.99
	1.5	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
375	1	1.00	0.40	1.00	0.80	1.00	0.90	0.97	0.31	1.00	0.98
	1.5	1.00	0.95	1.00	1.00	1.00	1.00	1.00	0.64	1.00	1.00
450	1	0.95	0.20	1.00	0.43	0.97	0.29	0.20	0.08	1.00	0.74
	1.5	1.00	0.55	1.00	0.89	1.00	0.66	0.43	0.10	1.00	0.98
475	1	0.65	0.17	0.91	0.19	0.33	0.09	0.07	0.06	0.98	0.42
	1.5	0.99	0.35	1.00	0.52	0.63	0.15	0.08	0.06	1.00	0.80
500	1	1.00	0.98	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
	1.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
750	1	1.00	0.86	1.00	0.99	1.00	1.00	1.00	0.59	1.00	1.00
	1.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00
900	1	1.00	0.42	1.00	0.85	1.00	0.60	0.50	0.09	1.00	0.98
	1.5	1.00	0.94	1.00	1.00	1.00	0.95	0.91	0.17	1.00	1.00
950	1	0.96	0.20	1.00	0.53	0.71	0.17	0.10	0.05	1.00	0.75
	1.5	1.00	0.58	1.00	0.91	0.98	0.32	0.19	0.05	1.00	0.97

Table 4.3

Proportion of rejections, in 1000 Monte-Carlo experiments,
under H_1 for middle-of-sample breaks when $\theta = 0$
using the bootstrap critical values.

$s = [n\tau]$	δ	\mathcal{W}		$\mathcal{W}_{.05}$		$\mathcal{W}_{.15}$		CUSUM		CHOW	
		σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1
50	1	0.83	0.17	0.83	0.17	0.94	0.26	0.39	0.12	0.99	0.48
	1.5	1.00	0.38	1.00	0.38	1.00	0.54	0.70	0.22	1.00	0.81
75	1	0.71	0.15	0.71	0.15	0.87	0.21	0.10	0.08	0.97	0.38
	1.5	0.98	0.32	0.98	0.32	1.00	0.46	0.14	0.08	1.00	0.74
90	1	0.39	0.11	0.39	0.11	0.19	0.09	0.04	0.06	0.56	0.13
	1.5	0.74	0.20	0.74	0.20	0.29	0.11	0.03	0.05	0.88	0.26
95	1	0.11	0.08	0.11	0.08	0.10	0.06	0.05	0.07	0.25	0.06
	1.5	0.14	0.09	0.14	0.09	0.11	0.07	0.04	0.06	0.46	0.11
100	1	1.00	0.25	1.00	0.32	1.00	0.58	0.74	0.23	1.00	0.80
	1.5	1.00	0.68	1.00	0.75	1.00	0.95	0.98	0.43	1.00	0.99
150	1	0.95	0.23	0.97	0.27	1.00	0.47	0.21	0.10	1.00	0.64
	1.5	1.00	0.56	1.00	0.63	1.00	0.82	0.42	0.13	1.00	0.95
180	1	0.64	0.17	0.67	0.19	0.49	0.15	0.05	0.07	0.93	0.32
	1.5	0.96	0.32	0.96	0.37	0.74	0.25	0.05	0.08	1.00	0.59
190	1	0.37	0.13	0.38	0.15	0.13	0.07	0.04	0.07	0.57	0.19
	1.5	0.70	0.21	0.71	0.24	0.19	0.07	0.04	0.07	0.89	0.29
150	1	1.00	0.39	1.00	0.62	1.00	0.80	0.93	0.33	1.00	0.91
	1.5	1.00	0.90	1.00	0.97	1.00	0.99	1.00	0.62	1.00	1.00
225	1	1.00	0.30	1.00	0.51	1.00	0.64	0.35	0.12	1.00	0.82
	1.5	1.00	0.77	1.00	0.91	1.00	0.96	0.71	0.21	1.00	0.99
270	1	0.76	0.17	0.92	0.25	0.67	0.17	0.05	0.07	0.99	0.46
	1.5	0.99	0.42	1.00	0.57	0.94	0.34	0.06	0.07	1.00	0.82
285	1	0.48	0.13	0.62	0.14	0.16	0.08	0.04	0.07	0.79	0.22
	1.5	0.86	0.28	0.94	0.31	0.27	0.12	0.03	0.07	0.99	0.47
250	1	1.00	0.56	1.00	0.91	1.00	0.96	1.00	0.46	1.00	0.99
	1.5	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.84	1.00	1.00
375	1	1.00	0.40	1.00	0.81	1.00	0.88	0.67	0.15	1.00	0.98
	1.5	1.00	0.94	1.00	1.00	1.00	1.00	0.97	0.33	1.00	1.00
450	1	0.92	0.22	1.00	0.43	0.95	0.28	0.10	0.07	1.00	0.71
	1.5	1.00	0.57	1.00	0.85	1.00	0.58	0.18	0.07	1.00	0.97
475	1	0.65	0.18	0.89	0.21	0.28	0.07	0.05	0.06	0.97	0.39
	1.5	0.97	0.34	1.00	0.51	0.54	0.13	0.05	0.06	1.00	0.75
500	1	1.00	0.97	1.00	1.00	1.00	1.00	1.00	0.77	1.00	1.00
	1.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
750	1	1.00	0.85	1.00	1.00	1.00	1.00	0.98	0.26	1.00	1.00
	1.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.63	1.00	1.00
900	1	1.00	0.41	1.00	0.86	1.00	0.59	0.21	0.06	1.00	0.96
	1.5	1.00	0.90	1.00	1.00	1.00	0.94	0.50	0.09	1.00	1.00
950	1	0.95	0.21	1.00	0.51	0.63	0.15	0.07	0.05	1.00	0.73
	1.5	1.00	0.59	1.00	0.88	0.92	0.30	0.09	0.05	1.00	0.97

Table 4.3 (Continuation)

Proportion of rejections, in 1000 Monte-Carlo experiments,
under H_1 for middle-of-sample breaks when $\theta = \pi/4$
using the bootstrap critical values.

$s = [n\tau]$	δ	\mathcal{W}		$\mathcal{W}_{.05}$		$\mathcal{W}_{.15}$		CUSUM		CHOW	
		σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1
50	1	0.84	0.17	0.84	0.17	0.95	0.25	0.02	0.07	0.99	0.49
	1.5	1.00	0.42	1.00	0.42	1.00	0.58	0.01	0.05	1.00	0.84
75	1	0.72	0.15	0.72	0.15	0.87	0.21	0.03	0.07	0.96	0.41
	1.5	0.99	0.37	0.99	0.37	1.00	0.48	0.01	0.05	1.00	0.72
90	1	0.47	0.12	0.47	0.12	0.27	0.08	0.04	0.07	0.60	0.16
	1.5	0.81	0.24	0.81	0.24	0.47	0.14	0.02	0.06	0.92	0.30
95	1	0.11	0.07	0.11	0.07	0.10	0.06	0.05	0.07	0.27	0.07
	1.5	0.16	0.09	0.16	0.09	0.12	0.07	0.04	0.07	0.52	0.09
100	1	1.00	0.23	1.00	0.31	1.00	0.61	0.02	0.05	1.00	0.79
	1.5	1.00	0.69	1.00	0.75	1.00	0.95	0.00	0.05	1.00	0.98
150	1	0.96	0.20	0.97	0.27	1.00	0.48	0.03	0.06	1.00	0.67
	1.5	1.00	0.59	1.00	0.64	1.00	0.85	0.01	0.05	1.00	0.95
180	1	0.66	0.14	0.67	0.19	0.55	0.14	0.04	0.07	0.95	0.28
	1.5	0.98	0.34	0.99	0.41	0.90	0.26	0.03	0.06	1.00	0.61
190	1	0.44	0.14	0.42	0.15	0.14	0.07	0.05	0.07	0.60	0.19
	1.5	0.78	0.23	0.77	0.24	0.24	0.08	0.04	0.07	0.93	0.34
150	1	1.00	0.32	1.00	0.62	1.00	0.78	0.02	0.06	1.00	0.90
	1.5	1.00	0.89	1.00	0.98	1.00	1.00	0.01	0.05	1.00	1.00
225	1	1.00	0.29	1.00	0.46	1.00	0.64	0.02	0.06	1.00	0.81
	1.5	1.00	0.78	1.00	0.91	1.00	0.97	0.01	0.05	1.00	0.99
270	1	0.82	0.16	0.93	0.24	0.74	0.18	0.03	0.06	1.00	0.46
	1.5	1.00	0.39	1.00	0.57	0.99	0.37	0.02	0.06	1.00	0.84
285	1	0.51	0.11	0.62	0.13	0.17	0.07	0.03	0.07	0.79	0.19
	1.5	0.90	0.26	0.97	0.30	0.32	0.10	0.03	0.06	0.99	0.47
250	1	1.00	0.57	1.00	0.89	1.00	0.96	0.02	0.06	1.00	0.99
	1.5	1.00	0.99	1.00	1.00	1.00	1.00	0.01	0.05	1.00	1.00
375	1	1.00	0.37	1.00	0.80	1.00	0.89	0.03	0.06	1.00	0.97
	1.5	1.00	0.94	1.00	1.00	1.00	1.00	0.01	0.05	1.00	1.00
450	1	0.95	0.21	1.00	0.46	0.96	0.28	0.04	0.06	1.00	0.69
	1.5	1.00	0.53	1.00	0.86	1.00	0.64	0.02	0.06	1.00	0.97
475	1	0.64	0.18	0.92	0.20	0.31	0.09	0.06	0.07	0.98	0.41
	1.5	0.98	0.34	1.00	0.51	0.64	0.14	0.05	0.06	1.00	0.76
500	1	1.00	0.97	1.00	1.00	1.00	1.00	0.03	0.04	1.00	1.00
	1.5	1.00	1.00	1.00	1.00	1.00	1.00	0.01	0.03	1.00	1.00
750	1	1.00	0.85	1.00	1.00	1.00	1.00	0.03	0.04	1.00	1.00
	1.5	1.00	1.00	1.00	1.00	1.00	1.00	0.02	0.03	1.00	1.00
900	1	1.00	0.35	1.00	0.85	1.00	0.60	0.04	0.04	1.00	0.97
	1.5	1.00	0.92	1.00	1.00	1.00	0.94	0.03	0.04	1.00	1.00
950	1	0.97	0.21	1.00	0.53	0.70	0.15	0.05	0.05	1.00	0.72
	1.5	1.00	0.58	1.00	0.90	0.97	0.34	0.04	0.04	1.00	0.99

Table 4.3 (Continuation)

Proportion of rejections, in 1000 Monte-Carlo experiments,
under H_1 for middle-of-sample breaks when $\theta = \pi/2$
using the bootstrap critical values.

$s = n - cn^{1/2}$	δ	\mathcal{W}		$\mathcal{W}_{.05}$		$\mathcal{W}_{.15}$		CUSUM		CHOW	
		σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1
90	1	0.44	0.12	0.44	0.12	0.25	0.10	0.04	0.06	0.57	0.17
	1.5	0.77	0.22	0.77	0.22	0.47	0.14	0.02	0.05	0.91	0.30
94	1	0.19	0.08	0.19	0.08	0.12	0.07	0.04	0.06	0.40	0.10
	1.5	0.37	0.10	0.37	0.10	0.20	0.09	0.03	0.06	0.75	0.18
95	1	0.12	0.07	0.12	0.07	0.09	0.07	0.05	0.06	0.28	0.08
	1.5	0.17	0.09	0.17	0.09	0.12	0.07	0.04	0.06	0.50	0.12
97	1	0.09	0.07	0.09	0.07	0.07	0.07	0.05	0.07	0.25	0.10
	1.5	0.12	0.07	0.12	0.07	0.09	0.06	0.04	0.06	0.45	0.12
186	1	0.58	0.16	0.58	0.18	0.31	0.10	0.05	0.07	0.85	0.24
	1.5	0.93	0.31	0.93	0.32	0.57	0.15	0.05	0.07	1.00	0.48
191	1	0.43	0.14	0.43	0.14	0.14	0.06	0.04	0.07	0.64	0.17
	1.5	0.78	0.24	0.79	0.25	0.25	0.09	0.04	0.07	0.92	0.31
193	1	0.31	0.11	0.24	0.09	0.11	0.06	0.04	0.08	0.47	0.13
	1.5	0.57	0.18	0.44	0.15	0.15	0.07	0.04	0.08	0.81	0.24
196	1	0.11	0.06	0.11	0.07	0.07	0.06	0.05	0.07	0.24	0.08
	1.5	0.15	0.08	0.15	0.08	0.09	0.06	0.05	0.07	0.41	0.11
283	1	0.63	0.16	0.78	0.18	0.25	0.10	0.04	0.07	0.91	0.29
	1.5	0.96	0.34	0.99	0.43	0.53	0.16	0.03	0.06	1.00	0.64
289	1	0.45	0.15	0.38	0.10	0.11	0.07	0.04	0.07	0.75	0.17
	1.5	0.83	0.25	0.67	0.20	0.21	0.10	0.03	0.07	0.98	0.40
292	1	0.35	0.12	0.19	0.06	0.07	0.06	0.04	0.07	0.52	0.14
	1.5	0.72	0.20	0.31	0.10	0.11	0.07	0.04	0.07	0.86	0.26
295	1	0.14	0.08	0.09	0.05	0.04	0.06	0.04	0.07	0.30	0.11
	1.5	0.21	0.10	0.13	0.05	0.06	0.06	0.04	0.07	0.54	0.18
478	1	0.63	0.17	0.83	0.16	0.26	0.08	0.06	0.06	0.97	0.41
	1.5	0.98	0.33	0.99	0.40	0.53	0.13	0.06	0.06	1.00	0.73
486	1	0.50	0.15	0.28	0.07	0.12	0.06	0.06	0.07	0.79	0.22
	1.5	0.89	0.28	0.58	0.13	0.21	0.06	0.06	0.06	0.99	0.48
489	1	0.39	0.14	0.16	0.06	0.09	0.05	0.06	0.07	0.61	0.21
	1.5	0.76	0.26	0.33	0.08	0.14	0.06	0.05	0.07	0.95	0.41
493	1	0.24	0.11	0.07	0.04	0.06	0.05	0.06	0.07	0.42	0.13
	1.5	0.44	0.18	0.11	0.05	0.08	0.05	0.06	0.07	0.74	0.27
969	1	0.80	0.17	0.82	0.17	0.30	0.09	0.07	0.05	1.00	0.55
	1.5	1.00	0.40	0.99	0.43	0.58	0.15	0.08	0.05	1.00	0.87
979	1	0.61	0.12	0.43	0.09	0.16	0.04	0.06	0.05	0.98	0.38
	1.5	0.97	0.25	0.73	0.15	0.29	0.08	0.06	0.05	1.00	0.71
985	1	0.46	0.12	0.18	0.07	0.10	0.04	0.06	0.05	0.90	0.23
	1.5	0.89	0.18	0.39	0.09	0.15	0.05	0.06	0.05	1.00	0.48
990	1	0.36	0.11	0.10	0.05	0.07	0.03	0.06	0.05	0.61	0.16
	1.5	0.74	0.18	0.15	0.07	0.10	0.03	0.06	0.05	0.93	0.30

Table 4.4

Proportion of rejections, in 1000 Monte-Carlo experiments,
under H_1 for end-of-sample breaks when $\theta = 0$
using the bootstrap critical values.

$s = n - cn^{1/2}$	δ	\mathcal{W}		$\mathcal{W}_{.05}$		$\mathcal{W}_{.15}$		CUSUM		CHOW	
		σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1
90	1	0.39	0.11	0.39	0.11	0.19	0.09	0.04	0.06	0.56	0.13
	1.5	0.74	0.20	0.74	0.20	0.29	0.11	0.03	0.05	0.88	0.26
94	1	0.21	0.08	0.21	0.08	0.14	0.07	0.04	0.06	0.41	0.10
	1.5	0.30	0.11	0.30	0.11	0.18	0.09	0.03	0.06	0.73	0.20
95	1	0.11	0.08	0.11	0.08	0.10	0.06	0.05	0.07	0.25	0.06
	1.5	0.14	0.09	0.14	0.09	0.11	0.07	0.04	0.06	0.46	0.11
97	1	0.08	0.06	0.08	0.06	0.08	0.06	0.05	0.07	0.20	0.07
	1.5	0.08	0.06	0.08	0.06	0.09	0.05	0.04	0.07	0.43	0.10
186	1	0.54	0.17	0.56	0.18	0.30	0.11	0.04	0.07	0.85	0.24
	1.5	0.90	0.28	0.92	0.34	0.49	0.17	0.04	0.07	1.00	0.46
191	1	0.37	0.13	0.38	0.15	0.13	0.07	0.04	0.07	0.57	0.19
	1.5	0.70	0.21	0.71	0.24	0.19	0.07	0.04	0.07	0.89	0.29
193	1	0.18	0.10	0.20	0.10	0.10	0.06	0.04	0.08	0.44	0.14
	1.5	0.27	0.12	0.28	0.14	0.13	0.08	0.04	0.07	0.75	0.24
196	1	0.10	0.08	0.11	0.08	0.07	0.06	0.04	0.07	0.20	0.07
	1.5	0.11	0.08	0.11	0.09	0.10	0.07	0.04	0.07	0.38	0.12
283	1	0.55	0.14	0.69	0.17	0.21	0.09	0.04	0.07	0.86	0.28
	1.5	0.90	0.31	0.96	0.38	0.36	0.14	0.03	0.07	1.00	0.60
289	1	0.39	0.12	0.25	0.08	0.10	0.06	0.04	0.07	0.67	0.17
	1.5	0.74	0.22	0.42	0.15	0.15	0.07	0.03	0.07	0.96	0.35
292	1	0.34	0.11	0.15	0.06	0.06	0.06	0.04	0.07	0.52	0.12
	1.5	0.67	0.19	0.24	0.09	0.09	0.06	0.03	0.07	0.82	0.22
295	1	0.09	0.06	0.07	0.05	0.05	0.05	0.04	0.07	0.29	0.09
	1.5	0.09	0.07	0.07	0.05	0.05	0.06	0.04	0.07	0.48	0.15
478	1	0.63	0.17	0.78	0.16	0.24	0.07	0.06	0.06	0.97	0.37
	1.5	0.96	0.34	0.98	0.39	0.44	0.12	0.05	0.06	1.00	0.68
486	1	0.46	0.16	0.27	0.07	0.11	0.05	0.06	0.07	0.76	0.21
	1.5	0.84	0.27	0.45	0.12	0.17	0.06	0.05	0.06	0.99	0.45
489	1	0.37	0.15	0.14	0.05	0.08	0.04	0.06	0.07	0.56	0.21
	1.5	0.71	0.23	0.20	0.07	0.10	0.05	0.05	0.06	0.91	0.36
493	1	0.15	0.11	0.06	0.05	0.05	0.05	0.06	0.07	0.40	0.16
	1.5	0.19	0.14	0.07	0.05	0.07	0.05	0.05	0.07	0.69	0.25
969	1	0.75	0.17	0.73	0.18	0.25	0.07	0.06	0.05	1.00	0.51
	1.5	0.99	0.38	0.95	0.39	0.47	0.13	0.06	0.05	1.00	0.86
979	1	0.57	0.13	0.33	0.10	0.14	0.05	0.06	0.05	0.98	0.39
	1.5	0.94	0.29	0.57	0.17	0.23	0.07	0.06	0.05	1.00	0.68
985	1	0.45	0.12	0.16	0.06	0.09	0.05	0.06	0.05	0.87	0.25
	1.5	0.82	0.20	0.26	0.09	0.14	0.06	0.06	0.05	0.99	0.49
990	1	0.37	0.12	0.10	0.05	0.07	0.04	0.06	0.05	0.64	0.17
	1.5	0.71	0.17	0.13	0.07	0.10	0.05	0.06	0.05	0.92	0.36

Table 4.4 (continuation)

Proportion of rejections, in 1000 Monte-Carlo experiments,
under H_1 for end-of-sample breaks when $\theta = \pi/4$
using the bootstrap critical values.

$s = n - cn^{1/2}$	δ	\mathcal{W}		$\mathcal{W}_{.05}$		$\mathcal{W}_{.15}$		CUSUM		CHOW	
		σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1	σ_0	σ_1
90	1	0.47	0.12	0.47	0.12	0.27	0.08	0.04	0.07	0.60	0.16
	1.5	0.81	0.24	0.81	0.24	0.47	0.14	0.02	0.06	0.92	0.30
94	1	0.20	0.09	0.20	0.09	0.14	0.07	0.05	0.07	0.42	0.10
	1.5	0.35	0.13	0.35	0.13	0.23	0.09	0.04	0.06	0.75	0.21
95	1	0.11	0.07	0.11	0.07	0.10	0.06	0.05	0.07	0.27	0.07
	1.5	0.16	0.09	0.16	0.09	0.12	0.07	0.04	0.07	0.52	0.09
97	1	0.09	0.06	0.09	0.06	0.09	0.06	0.05	0.07	0.22	0.08
	1.5	0.12	0.07	0.12	0.07	0.09	0.06	0.04	0.07	0.44	0.10
186	1	0.57	0.14	0.56	0.17	0.31	0.10	0.05	0.07	0.86	0.22
	1.5	0.93	0.32	0.93	0.34	0.58	0.14	0.04	0.06	1.00	0.47
191	1	0.44	0.14	0.42	0.15	0.14	0.07	0.05	0.07	0.60	0.19
	1.5	0.78	0.23	0.77	0.24	0.24	0.08	0.04	0.07	0.93	0.34
193	1	0.32	0.12	0.23	0.10	0.11	0.06	0.05	0.07	0.47	0.15
	1.5	0.57	0.18	0.40	0.14	0.16	0.06	0.04	0.07	0.80	0.26
196	1	0.11	0.08	0.10	0.08	0.07	0.06	0.05	0.07	0.20	0.09
	1.5	0.14	0.08	0.13	0.08	0.08	0.07	0.04	0.07	0.41	0.12
283	1	0.57	0.13	0.73	0.17	0.24	0.09	0.03	0.07	0.89	0.26
	1.5	0.95	0.31	0.99	0.35	0.49	0.14	0.03	0.06	1.00	0.56
289	1	0.46	0.12	0.34	0.08	0.11	0.05	0.04	0.07	0.70	0.17
	1.5	0.84	0.21	0.68	0.18	0.20	0.08	0.03	0.07	0.97	0.33
292	1	0.36	0.10	0.15	0.06	0.07	0.05	0.04	0.07	0.50	0.12
	1.5	0.71	0.18	0.31	0.09	0.12	0.06	0.04	0.07	0.85	0.23
295	1	0.14	0.07	0.08	0.05	0.06	0.05	0.04	0.07	0.29	0.09
	1.5	0.21	0.09	0.10	0.05	0.06	0.05	0.04	0.07	0.57	0.15
478	1	0.63	0.17	0.81	0.18	0.26	0.08	0.06	0.06	0.97	0.38
	1.5	0.98	0.33	0.99	0.41	0.52	0.11	0.05	0.06	1.00	0.70
486	1	0.48	0.13	0.29	0.08	0.12	0.06	0.06	0.07	0.79	0.21
	1.5	0.86	0.28	0.58	0.14	0.21	0.07	0.05	0.06	0.99	0.45
489	1	0.39	0.12	0.16	0.07	0.08	0.06	0.06	0.07	0.58	0.20
	1.5	0.79	0.22	0.33	0.10	0.13	0.07	0.06	0.06	0.94	0.40
493	1	0.23	0.11	0.09	0.05	0.06	0.05	0.06	0.07	0.40	0.14
	1.5	0.46	0.17	0.14	0.07	0.07	0.06	0.06	0.07	0.74	0.24
969	1	0.81	0.18	0.80	0.21	0.30	0.08	0.05	0.05	1.00	0.56
	1.5	1.00	0.41	0.99	0.43	0.60	0.13	0.05	0.04	1.00	0.88
979	1	0.63	0.16	0.40	0.11	0.16	0.05	0.06	0.05	0.98	0.39
	1.5	0.97	0.28	0.73	0.21	0.29	0.07	0.06	0.04	1.00	0.75
985	1	0.50	0.14	0.21	0.07	0.10	0.05	0.06	0.05	0.90	0.24
	1.5	0.90	0.21	0.37	0.11	0.17	0.06	0.05	0.05	1.00	0.54
990	1	0.39	0.13	0.12	0.05	0.07	0.05	0.06	0.05	0.65	0.18
	1.5	0.76	0.22	0.18	0.07	0.09	0.05	0.06	0.05	0.95	0.35

Table 4.4 (continuation)

Proportion of rejections, in 1000 Monte-Carlo experiments,
under H_1 for end-of-sample breaks when $\theta = \pi/2$
using the bootstrap critical values.

LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON WC2A 2AE, UK
E-mail address: f.j.hidalgo@lse.ac.uk

LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON WC2A 2AE, UK
E-mail address: m.seo@lse.ac.uk