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 the Wald or Lagrange Multiplier tests when the change is assumed to be in the middle of the sample, that is in the interval \([\lfloor \tau n \rfloor, n - \lfloor \tau n \rfloor]\) for some \(\tau > 0\) and where \(n\) denotes the sample size. Recently there has been some work to allow the possibility that the break lies at the end of the sample, i.e. when \(t \in (n - \ell, n)\) for some finite number \(\ell\). However, the previous two setups do not include the important intermediate case when \(t \in (1, \lfloor \tau n \rfloor) \cup (n - \lfloor \tau n \rfloor, n)\), or more generally when we do not assume any prior knowledge on the possible location of the break. The aim of the paper is to extend existing results on stability tests in the latter 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 also it avoids the need to choose either \(\ell\) or \(\tau\). In addition we show that contrary to conventional tests such as the CUSUM or the \(\sup_{\lfloor \tau n \rfloor < t < n - \lfloor \tau n \rfloor} \text{Wald}\) or \(\text{Lagrange Multiplier}\) tests, the tests described and examined in the paper are consistent irrespectively 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 researcher assumes that the possible break lies in the “middle” of the sample, that is in the interval \([\lfloor \tau n \rfloor, n - \lfloor \tau n \rfloor]\) for some \(\tau > 0\), and where \(n\) denotes herewith the sample size. This is equivalent to say that the hypothesized break belongs to an interval \(\Pi\) whose closure is in \((0, 1)\). See for example Andrews (1993) or the latest review article by Perron (2006). In this scenario, it has been shown that the supremum of, for instance, the Wald (\(W\)) or Lagrange Multiplier (\(LM\)) statistics converge to the supremum of a Gaussian process. More recently, there has been some interest to learn what it would be the behaviour of the tests for stability when the practitioner assumes that, if a break exists, this occurs at the end of the sample, that is among the last \(\ell\) observations for some finite value \(\ell\). The work in the latter scenario is less proliﬁc than in the former one, although we can cite the works by Andrews (2003) or Andrews and Kim (2006) and references therein. In this situation, we know that, although the tests

Date: 2 May 2008.

are not consistent and their distribution depends on \( \hat{t} \), it is still possible to make inferences as it was shown by Andrews (2003). Between the two aforementioned setups, there is however an important gap in the theory. More specifically, when the possible break may lie in the intervals \((\hat{t}, [n\tau])\) or \((n - [n\tau], n - \hat{t})\), or more generally when we do not wish to impose any prior knowledge on the location of the possible break.

The paper considers thus the problem of testing for structural stability over the whole sample span \( t = 1, \ldots, n \). That is, when no previous information about the location of the break is available, or that we do not restrict our search for the break to any particular segment of the sample. We will examine tests for breaks on 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 and allowing for heterogeneity or heteroscedasticity of the disturbance error term. In addition, we should mention that we will not assume that the data, for instance the exogenous and disturbance error term in a regression model, to be covariance stationary. More specifically, we might allow for heteroscedasticity or some heterogeneity in the in the error term and exogenous variables. In this way, we substantially extend Horvath’s (1993) results who only examines this problem for a break in the mean of otherwise independent normally distributed random variables. It is worth mentioning that Andrews (1993) looked briefly at our setup. He signalled that the standard Wald, Lagrange Multiplier or Likelihood Ratio tests for breaks will diverge to infinity if the supremum is taken over \( t = 1, \ldots, n \) or equivalently in the interval \((0, 1)\). Thus, the restriction to take the supremum (or other functionals) over a set \( \Pi \) whose closure is in \((0, 1)\) is not made only for technical convenience but it is a crucial assumption to obtain a proper asymptotic distribution. In this way, the paper shows that (a) the reason for Andrews (1993) findings is because the normalization for the test to have a proper asymptotic distribution is different, and (b) that the asymptotic distribution of the tests is completely different than that obtained when we focus our test for stability in the interval \( \Pi \). More specifically, we show that, after appropriate normalization, the sup Wald or sup Lagrange Multiplier tests 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 (or not too close) from the beginning or from the end of the sample. Indeed, looking at the power of the test in Andrews (1993), or those in Andrews and Ploberger (1994), and the power of our tests described in Sections 2 and 3, they indicate and suggest that the assumptions made in the former works are not innocuous. More specifically, we show in Section 2.3 that the conventional tests described in Andrews (1993) or optimal tests of Andrews and Ploberger (1994) are not consistent when the break occurs at time \( t \leq [n^{1/2}] \) or \( n - [n^{1/2}] < t < n \), whereas our tests are. In addition, we show that when the break falls in the region \( t \in \left([n^{1/2}], n / (\log \log n)^{1/2}\right)\), the conventional tests have zero asymptotic efficiency compared to ours, in the sense that our tests are able to detect local alternatives that, for instance, the “optimal tests” are not able to detect. It is worth mentioning that our tests are similar as that of Brown, Durbin and Evans (1975) in the sense that they do not restrict the search of the break to \( \Pi \), but effectively to \((0, 1)\). However, the latter work suffers from the
same lack of power just described for the standard sup—, \textit{Mean}— or \textit{Exp}— tests discussed in Andrews and Ploberger (1994).

We finish this section discussing some theoretical and practical issues regarding our tests in Sections 2 or 3 below when they are compared with more conventional tests presented in Andrews (1993) or Andrews and Ploberger (1994). From a practical point of view, our tests have the benefit that the practitioner does not need to choose the rather artificial parameters $\hat{t}$ and/or $\tau$ when performing the test. Moreover, we avoid the rather undesirable outcome that using the same data set, by choosing two different values of $\tau$, two practitioners may lead to contradictory conclusions. This is confirm in the small Monte Carlo experiment described in Section 4 indicating that the choice of $\tau$ is not irrelevant. In fact, we have that the power and size can be affected but that choice. In fact, if we followed the recommendation given by some authors, see for instance Andrews (1993), when $\tau = .15$, the power of the conventional test is much smaller when it is compared to our test or when we choose $\tau = 0.05$. On the other hand, the Monte Carlo experiment suggests that the results in terms of size the test with $\tau = .15$ behaves better than when $\tau = 0.05$, although it is similar to the empirical size of our tests. From a theoretical point of view, our test is always consistent irrespectively of the location of the break, whereas conventional test they are not always consistent.

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 small Monte Carlo experiment to examine the finite sample performance of the test and how they compared with the standard $\sup_{t} \left[ n^{-1} \right] \left( n^{-1} \right) \text{Wald}$ test for some value of $\tau$. 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

\begin{equation}
    y_t = \alpha + \beta' x_t + \delta' z_t + u_t, \quad t = 1, \ldots, n
\end{equation}

where

\begin{equation}
    z_t = z_t(s) = \begin{cases} 
        x_{t1}; & t \leq s \\
        0; & t > s 
    \end{cases}
\end{equation}

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 error terms. Our null hypothesis $H_0$ of interest is that the parameter $\delta$ is zero for all $s$. That is,

\begin{equation}
    H_0 : \delta = 0 \quad \forall s : p^* < s \leq n - p^*,
\end{equation}

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

\begin{equation}
    H_1 : \exists s : p^* < s \leq n - p^*; \quad \delta \neq 0.
\end{equation}

Notice that when $p_1 = p$, we have that (2.3) corresponds to the so-called pure structural stability 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 change 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 will follow if we also allow the intercept to be subject to possible change over time. The only difference is in the computation of the test and more specifically on the estimation of the asymptotic covariance matrix of the estimators of the parameters subject to possible change.

We now describe the estimators and present some notation to be used throughout this section. As usual, for a fixed \( s \), we denote the least squares estimate of \((\beta', \delta')'\) in (2.1) by

\[
\begin{pmatrix}
\hat{\beta}(s) \\
\hat{\delta}(s)
\end{pmatrix} = 
\left( X_1'(s)X_1(s) \quad X_1'(s)X(n) \right) ^{-1} 
\left( X_1'(s)Y(n) \quad X'(n)Y(s) \right), \quad p^* < s \leq n-p^*,
\]

where from now on, we write

\[
W(s) = (w_1'-\overline{w}(s), \ldots, w_s'-\overline{w}(s))'
\]

for a generic sequence \( \{w_t\}_{t \in \mathbb{Z}} \), with

\[
\overline{w}(s) = \frac{1}{n} \sum_{t=1}^{s} w_t.
\]

It should be mentioned that if, for example, the matrix

\[
X_1'(s)X_1(s)
\]

were singular, in the computation of our estimates and/or tests, we would use the generalized inverse instead of the inverse. The latter will not affect any of the conclusions of the paper. Similar comments apply elsewhere below.

Next, let’s denote the variance of \( k^{-1/2} \sum_{t=1}^{k} (x_t - \mu_x) u_t \) by

\[
\Xi_k := \frac{1}{k} \mathbb{E} \left( \sum_{t_1, t_2=1}^{k} (x_{t_1} - \mu_x) (x_{t_2} - \mu_x)' u_{t_1} u_{t_2} \right) > 0, \quad k = 1, \ldots, n,
\]

where \( \mu_x = \mathbb{E}x_t \).

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

\[
x_t = \sum_{i=0}^{\infty} \chi_i \varepsilon_{t-i}, \quad \sum_{i=0}^{\infty} \|\chi_i\|^{1/\varsigma} < \infty \quad \text{and} \quad \chi_0 = I_{p \times p}
\]

\[
u_t = \sum_{i=0}^{\infty} \phi_i \eta_{t-i}, \quad \sum_{i=0}^{\infty} \|\phi_i\|^{1/\varsigma} < \infty \quad \text{and} \quad \phi_0 = 1,
\]

for some \( \varsigma \geq 3/2 \), and where \( \{\varepsilon_t\}_{t \in \mathbb{Z}} \) and \( \{\eta_t\}_{t \in \mathbb{Z}} \) are mutually independent zero mean independent distributed sequences of random variables such that

\[
\mathbb{E}(\eta_t^2) = \sigma_\eta \quad \text{and} \quad \mathbb{E}(\varepsilon_t^2) = \Sigma_x \quad \text{and} \quad \sup_t \mathbb{E}\|x_t\|^4 < \infty \quad \text{and} \quad \sup_t \mathbb{E}\|u_t\|^4 < \infty,
\]

where \( \|A\| \) denotes the norm of the matrix \( A \).

Denote \( M_s = X'(s)X(s) \). Then,

**A2:** \( s^{-1}M_s \to_{s \to \infty} \Sigma > 0 \).
we employed the least squares estimator in 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 as it stands it implies that \( \chi_i \) and \( \phi_i \) are \( o(i^{-3/2}) \). Although we have assumed that the sequences \( \{x_t\}_{t \in \mathbb{Z}} \) and \( \{u_t\}_{t \in \mathbb{Z}} \) are covariance stationary, the assumption of identical second moments is not strictly essential for our results below to follow. In fact, we can allow the sequences \( \{x_t\}_{t \in \mathbb{Z}} \) and \( \{u_t\}_{t \in \mathbb{Z}} \) to exhibit heterogeneity and relax the assumption of constant variance of \( \{\varepsilon_t\}_{t \in \mathbb{Z}} \) and \( \{\eta_t\}_{t \in \mathbb{Z}} \), or that the sequences \( \{x_t\}_{t \in \mathbb{Z}} \) and \( \{u_t\}_{t \in \mathbb{Z}} \) are mutually independent, so that we can allow for heteroscedastic errors \( E[u_t^2|x_t] = \sigma^2(x_t) \). In the latter scenario, we might replace the assumption by some type of asymptotic covariance stationarity. In particular, all that we would need is an assumption which guaranteed that \( \lim n^{-1} \text{Var} \left( \sum_{t=n^2}^{n^2+n} x_t u_t \right) \) and a consistent estimator exist.

More explicitly, see Andrews (1993) or Bai and Perron (1998) among others, it would suffice to assume that

\[
(2.7) \quad \text{Var} \left( n^{-1/2} \sum_{t=[n^2]+1}^{[n^2]} x_t u_t \right) \to (\tau_2 - \tau_1) \Xi.
\]

We keep nevertheless in this section Assumption A1 as it stands for pedagogical reasons although it will be relaxed in the next section. Observe nevertheless that we shall emphasize that we do not assume anywhere that the sequences \( \{x_t\}_{t \in \mathbb{Z}} \) and \( \{u_t\}_{t \in \mathbb{Z}} \) are identically distributed.

The reason for which the assumption of identically distributed sequences of variables is not strictly needed is because the proof of Proposition 1 relies on results for independent nonidentically distributed random variables. The motivation to keep A1 as it stands is merely to make the proof of the key Proposition 1 below clearer while it involves the main steps for more general type of data and hence models as examined in Section 3.

We notice that A1 and A2 imply that the “long-run” covariance matrix \( \Xi = \lim_{n \to \infty} \Xi_n \),

is a finite and positive definite matrix, which, following Robinson (1998), we can consistently estimate by \( \hat{\Xi}^{(n)}_n \), where

\[
(2.8) \quad \hat{\Xi}^{(n)}_n = \gamma_{x}(0) \gamma_u(0) + \sum_{j=1}^{s-1} \left( \gamma_{x}(j) + \gamma_u(0) \right) \gamma_u(0) = \frac{1}{s} \sum_{t=1}^{s-j} \tilde{u}_t \tilde{u}_{t+j},
\]

being \( \gamma_{x}(j) \) and \( \gamma_u(0) \) respectively the estimators of \( \mathbb{E} \{x_t x_{t+j}\} \) and \( \mathbb{E} \{u_t u_{t+j}\} \) given by the expressions

\[
\gamma_{x}(j) = \frac{1}{s} \sum_{t=1}^{s-j} (x_t - \bar{x}(n))(x_{t+j} - \bar{x}(n))', \quad \gamma_u(0) = \frac{1}{s} \sum_{t=1}^{s-1} \tilde{u}_t \tilde{u}_{t+j}.
\]

(Notice that we have assumed for simplicity that \( \mathbb{E} x_t = 0 \).) Here or elsewhere \( \{\tilde{u}_t\}_{t=1}^{n} \) is a sequence of residuals which will depend on the estimator that we have employed to estimate the parameters of the model (2.1). For instance, if we employed the least squares estimator in (2.4), we would compute the residuals \( \{\tilde{u}_t\}_{t=1}^{n} \) as \( \tilde{u}_t = \beta' (x_t - \bar{x}(n)) + \delta'(z_t - \bar{z}(s)) \). Following Robinson (1998), we
know that \(\hat{\Sigma}_n^{(s)}\) is a consistent estimator of \(\Sigma\) under A1. However, it is true that A1 is stronger than we need for such a result. Indeed, Robinson (1998) showed that for the consistency of \(\hat{\Sigma}_n^{(s)}\), it suffices to assume that

\[
E(\varepsilon_i \varepsilon_i' | F_{t-1} \cup G_t) \quad E(\eta_i^2 | F_t \cup G_{t-1})
\]

are constant, where respectively \(F_t\) and \(G_t\) are the sigma-algebra 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 becomes

\[
E(\varepsilon_i \varepsilon_i' | G_t), \quad E(\eta_i^2 | F_t),
\]

so that A1 would not be much stronger than assuming the latter.

If we had allowed the intercept \(\alpha\) in (2.1) to have a break, \(\Xi\) would become \(\lim_{n \to \infty} \sum_{n=1}^{n} \mathbb{E}(w_{t_1} w_{t_2} u_{t_1} u_{t_2}) = \Xi\) where \(w_t = (1, x_t)\), so that the analogue estimator of (2.8) corresponding to \(\Xi(1, 1)\), i.e. the long run variance of the least squares estimator of the intercept, is

\[
\hat{\Sigma}_n^{(s)}(1, 1) = \hat{\gamma}_u(0) + \sum_{j=1}^{n-1} \hat{\gamma}_u(j).
\]

However, contrary to \(\hat{\Sigma}_n^{(s)}\) in (2.8), \(\hat{\Sigma}_n^{(s)}(1, 1)\) is not a consistent estimator of \(\Xi(1, 1)\). So, in this situation, we would employ \(\hat{\Sigma}_n^{(m)} = \hat{\gamma}_u(0) + \sum_{j=1}^{m-1} \hat{\gamma}_u(j)\) for some \(m = o(n)\), which is a consistent estimator under Assumptions A1 and A2 as it has been shown by Brillinger (1981) or Andrews (1991) among others. Notice that the estimator \(\hat{\Sigma}_n^{(s)}\) in (2.8) is the time domain formulation of Robinson’s (1998) estimator of the long run variance \(\Xi\). So that, we do not need any kernel spectral density estimator to obtain a consistent estimator of \(\Xi\) nor to choose a bandwidth parameter to perform the test. It is worth mentioning that instead of \(\hat{\Sigma}_n^{(s)}\) given in (2.8) we might have been tempted to employ

\[
\hat{\Sigma}_n^{(s)} = \hat{\gamma}_u(0) + \sum_{j=1}^{s-1} \left( \hat{\gamma}_u^{(s)}(j) + \hat{\gamma}_u^{(s')} (j) \right),
\]

where \(\hat{\gamma}_u^{(s)}(j) = s^{-1} \sum_{t=1}^{s-j} \hat{u}_t \hat{u}_{t+j}\), with \(\hat{u}_t = x_t \hat{\gamma}_u\). However, in this case, contrary to \(\hat{\Sigma}_n^{(s)}\), \(\hat{\Sigma}_n^{(s)}\) would not be a consistent estimator for \(\Xi\), although the standard kernel estimator is, that is

\[
\hat{\Sigma}_n^{(m)} = \hat{\gamma}_u(0) + \sum_{j=1}^{m-1} \left( \hat{\gamma}_u^{(s)}(j) + \hat{\gamma}_u^{(s')} (j) \right).
\]

Notice that if the errors and regressors were not mutually independent, i.e. \(\{u_t\}_{t \in \mathbb{Z}}\) are heteroscedastic but satisfying (2.7), then the estimator (2.8) would not be consistent for \(\Xi\). In this case, we shall employ \(\hat{\Sigma}_n^{(m)}\) to consistently estimate \(\Xi\).

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 \(B(k)\) with independent components such that, for some \(1 < \zeta < 2\),

\[
\sup_{1 \leq k \leq n} \left\| \sum_{t=1}^{k} x_t u_t - \Xi_k^{1/2} B(k) \right\| = O_p \left( n^{-\zeta/2} \right).
\]
where \( \Xi_k \) is given in (2.6).

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 Götte and Zeitsev (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 (2.9) is worst than the “standard” \( O_p(n^{1/4}) \) 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). However, if the sequence \( x_t \) were deterministic, we would have the standard conclusion of the order of approximation to be \( O_p(n^{1/4}) \).

We now comment on the key role that Proposition 1 plays on our results. Its importance lies in the fact that the tests based on both the Wald and Lagrange Multiplier principle are functionals of partial sums of the type \( S_n = \sum_{t=1}^{s} x_t u_t \). For instance, inspecting the formulation of, say the Wald test in (2.11) or (2.13) below, the asymptotic distribution depends on the behaviour of \( (n/(n-s))^{1/2} s^{-1/2} (S_n - s/n S_n) \) which is governed by the behaviour of

\[
\left( \frac{n}{n-s} \right)^{1/2} \frac{1}{s^{1/2}} \left( I_{p_1 \times p_1 \cdot 0_{p_1 \times p-p_1}} \right) \left\{ S_n - \frac{s}{n} S_n \right\}
\]

(2.10)

\[
= \left( I_{p_1 \times p_1 \cdot 0_{p_1 \times p-p_1}} \right) \left\{ \left( \frac{n-s}{n} \right)^{1/2} \frac{1}{s^{1/2}} S_n - \left( \frac{s}{n} \right) \right\} \left( \frac{1}{(n-s)^{1/2}} \right) \left( S_n - s \right).
\]

Then, after noticing that

\[
\left( \frac{s}{n} \right)^{1/2} \frac{1}{(n-s)^{1/2}} (S_n - s) = \left( \frac{s}{n} \right)^{1/2} \frac{1}{(n-s)^{1/2}} \sum_{t=s+1}^{n} x_t u_t
\]

\[
d \equiv \left( \frac{n-s}{n} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=1}^{\hat{s}} x_t^* u_t^*
\]

where \( x_t^* u_t^* = x_{n-t+1} u_{n-t+1} \) and \( \hat{s} = n - s \), we conclude that the Wald test in (2.13) below is a functional of \( S_n = \max_{1 \leq s \leq n} \left\{ \left( I_{p_1 \times p_1 \cdot 0_{p_1 \times p-p_1}} \right) S_n / s^{1/2} \right\} \), whose (asymptotic) distribution is much more delicate to obtain than the well known (asymptotic) distribution of \( S_n = \max_{1 \leq s \leq n} \left\{ \left( I_{p_1 \times p_1 \cdot 0_{p_1 \times p-p_1}} \right) S_n / n^{1/2} \right\} \). One of the reasons is that \( S_n \) 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 \( S_n \), and thus that of the tests, will be governed by the asymptotic distribution of

\[
\sup_{p^* \leq \hat{s} \leq n-p^*} \left\| \frac{1}{s^{1/2}} \sum_{t=1}^{\hat{s}} \nu_t \right\|
\]

where \( \{ \nu_t \}_{t \in \mathbb{Z}} \) is a \( p_1 \)-dimensional vector of independent normally distributed random variables. Notice that for scalar and identically distributed sequences \( \{ \nu_t \}_{t \in \mathbb{Z}} \), the distribution of the last displayed statistic was examined by Darling and Erdös (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 standard tests for breaks are a functional of (Brownian) $s_n$.

We now describe the tests statistics.

2.1. The Wald statistic.

Suppose that we are first interested to test $H_0$ against the alternative hypothesis $H_1(s)$ defined as

$$H_1(s) : \delta \neq 0 \text{ for some } p^* < s \leq n - p^*.$$ 

In this case the Wald statistic is based on whether the estimate of $\delta$ in (2.1) is significantly different than zero. Recalling our notation in (2.5) and using that $(X'(s)X(n)) = M_s$ with $X(s) = (X_1(s), X_2(s))$, standard algebra implies that

$$\hat{\delta}^{(s)} = B_n^{-1}(s) \left( I_{p_1 \times p_1} \cdot 0_{p_1 \times p - p_1} \right) \left\{ X'(s)Y(s) - M_sM_n^{-1}X'(n)Y(n) \right\},$$

where

$$B_n(s) = \left( I_{p_1 \times p_1} \cdot 0_{p_1 \times p - p_1} \right) \left\{ M_s - M_sM_n^{-1}M_s \right\} \left( I_{p_1 \times p_1} \cdot 0_{p_1 \times p - p_1} \right)' .$$

Notice that under the null hypothesis $H_0$, we have that

$$\hat{\delta}^{(s)} = B_n^{-1}(s) \left( I_{p_1 \times p_1} \cdot 0_{p_1 \times p - p_1} \right) \left\{ X'(s)U(s) - M_sM_n^{-1}X'(n)U(n) \right\} .$$

Therefore, defining

$$\hat{\delta}^{(s)} = B_n(s) \hat{\delta}^{(s)} ,$$

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

$$W(s) = \hat{\delta}^{(s)'r} \left( I_{p_1 \times p_1} \cdot 0_{p_1 \times p - p_1} \right) \hat{V}_n(s) \left( I_{p_1 \times p_1} \cdot 0_{p_1 \times p - p_1} \right)' \hat{\delta}^{(s)} ,

$$

where

$$\hat{V}_n(s) = s \hat{\gamma}^{(s)}_{n} + nM_sM_n^{-1}\hat{\gamma}^{(n)}_{n}M_n^{-1}M_s - sM_sM_n^{-1}\hat{\gamma}^{(s)}_{n} - s\hat{\gamma}^{(s)}_{n}M_n^{-1}M_s$$

is an estimator of $V_n(s) = \mathbb{E} \left( \begin{array}{c} \delta' \\ \delta \end{array} \right) \cdot \hat{\gamma}^{(s)}_{n}$ with $\hat{\gamma}^{(s)}_{n}$ given in (2.8). We should bear in mind that in the computation of $\hat{V}_n(s)$, we have ignored an estimator of the expression $\mathbb{E} \left( \sum_{t=1}^{s} (x_t - \pi(n))u_t \sum_{t=1+s}^{n} (x_t - \pi(n))u_t \right)$. The reason is because the latter quantity is asymptotically negligible when we compare it with either $s\hat{\gamma}^{(s)}_{n}$ or $n\hat{\gamma}^{(n)}_{n}$.

On the other hand, if we wish to test for a break when we treat the time of the break “$s$” as an unknown parameter, we employ the standard union-interception principle. That is, the alternative hypothesis becomes $H_1 = \cup_{s=p+1}^{n-p-1} H_1(s)$, and hence our hypothesis testing is

$$H_0 \text{ against } H_1 .$$

In this case, we might consider the statistic

$$\hat{\mathcal{W}} = \max_{p^* < s \leq n - p^*} W(s) .$$
The motivation to look at the sup from \( p^* + 1 \) up to \( n - p^* \) is because the dimension of the parameter vectors \( \beta \) and \( \delta \) are \( p \) and \( p_1 \) respectively. Now, if we use the same estimator of \( \Sigma \), say \( \hat{\Sigma}_n^{(n)} \), and we replace \( M_s \) by its “limit” \( s \Sigma \), then \( \hat{V}_n (s) \) becomes \( \hat{V}_n = s \left( \frac{n-s}{n} \right) \hat{\Sigma}_n^{(n)} \), and hence the Wald statistic \( \hat{W} \) could be written as

\[
\hat{W} = \max_{p^* < s \leq n-p^*} \left( \frac{n}{n-s} \right) \left( \frac{1}{s} \right)^{1/2} \left( \left( I_{p_1 \times p_1 ; 0_{p_1 \times p-p_1}} \right) \hat{\Sigma}_n^{(n)} \left( I_{p_1 \times p_1 ; 0_{p_1 \times p-p_1}} \right)^{1/2} \right)^{-1} \frac{s}{\hat{\Sigma}_n^{(n)}}
\]

which is a rather simpler statistic to compute than \( \hat{W} \).

2.2. The LM statistic.

We now discuss and present the Lagrange Multiplier test. As we did with the Wald statistic, let’s suppose for a moment that we wish to test \( H_0 \) against the alternative hypothesis \( H_1 (s) \). In this case, the test would be based on whether the first order conditions

\[
\hat{F}_n^{(s)} = \sum_{t=1}^{s} \hat{x}_{t1} \left( \tilde{y}_t - \hat{x}_t \hat{\beta}^{(n)} \right)
\]

are not significantly different than zero, with \( \{ \tilde{w}_t \}_{t=1}^{n} \) stands for a generic sequence \( \{ w_t \}_{t=1}^{n} \) centered around its sample mean and where \( \hat{\beta}^{(s)} \) is the least squares estimator of \( \{ \tilde{y}_t \}_{t=1}^{s} \) on \( \{ \tilde{x}_t \}_{t=1}^{s} \).

\[
\hat{\beta}^{(s)} = M_s^{-1} X' \left( \frac{s}{\hat{\Sigma}_n^{(n)}} \right) Y (s) .
\]

Next, observing that

\[
\hat{F}_n^{(s)} = B_n (s) \frac{s}{\hat{\Sigma}_n^{(n)}},
\]

and that one typically employs the restricted estimator to obtain an estimator of the covariance of \( \hat{F}_n^{(s)} \), we obtain the Lagrange Multiplier as

\[
\mathcal{L}M (s) = \left( \frac{n}{n-s} \right) \left( \frac{1}{s} \right)^{1/2} \left( \left( I_{p_1 \times p_1 ; 0_{p_1 \times p-p_1}} \right) \hat{\Sigma}_n^{(n)} \left( I_{p_1 \times p_1 ; 0_{p_1 \times p-p_1}} \right)^{1/2} \right)^{-1} \hat{F}_n^{(s)},
\]

where now we employ the least squares residuals \( \hat{u}_t = \tilde{y}_t - \hat{x}_t \hat{\beta}^{(n)} \) to obtain \( \hat{\gamma}_u (j) \), \( j = 0, \ldots, n-1 \), when computing \( \hat{\Sigma}_n^{(n)} \) in (2.8).

Next for the hypothesis testing in (2.12), the Lagrange Multiplier statistic will be

\[
\mathcal{L}M = \max_{p^* < s \leq n-p^*} \mathcal{L}M (s).
\]

The \( \mathcal{L}M \) statistic only requires estimation under the null hypothesis and hence do not require reestimation of the model for each point of the sample split. So, in situations where the computation of estimates of the parameters of the model is demanding or intensive (such as the GMM estimator in the next section), it appears to be more appropriate to employ the \( \mathcal{L}M \) statistic. More importantly, in our hypothesis testing context, the \( \mathcal{L}M \) statistic can be of more interest compared to the \( \mathcal{W} \) statistic as the latter required the estimation of \( \delta \) and \( \beta \). However, estimates of the latter two parameters may be not very accurate when computed at either the end or beginning of the sample, due to the lack of observations, which turns out to be the most important stretch of the data regarding the properties of the test. Recall our comments after Proposition 1.
Denote $\log_2 x = \log \log x$ and $\log_3 x = \log \log \log x$ and also let $T$ be a random variable distributed as a (double) Gumbel random variable, i.e.

$$(2.16) \quad \Pr \{ T \leq x \} = \exp (-2e^{-x}).$$

Then, we have the following theorem.

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

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

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

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

**Remark 1.** The proof of the theorem indicates that if instead of looking for the maximum in the region $p^* < s \leq n - p^*$, we considered the maximum in either $n/2 \leq s \leq n - p^*$ or $p^* < s < n/2$, the asymptotic distribution of the test would be $T_1$, where

$$\Pr \{ T_1 \leq x \} = \exp (-e^{-x})$$

is the Gumbel distribution.

### 2.3. Power of the Test

In this section, we examine the behaviour of the tests under either fixed or local alternatives. For that purpose, we consider the sequence of models

$$(2.17) \quad y_t = \alpha + \beta' x_t + g_n \delta \left( \frac{t}{n} \right) z_t + u_t, \quad t = 1, \ldots, n$$

where $\{ z_t \}_{t \in \mathbb{Z}}$ is as defined in (2.2) and $\{ gn \}_{n \in \mathbb{N}}$ is some sequence depending only on $n$ and to be made more precise below. Notice that a one-time change at time $s_0$ corresponds to the function $\delta (t/n) = I (t \leq s_0)$. In addition, we shall assume that $\delta (v)$ is a $p_1$-vector function on $[0, 1]$ such that it is uniformly Riemann summable in that

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \delta \left( \frac{t}{n} \right) - \int_0^{n/n} \delta (v) \, dv \right\| = o (1)$$

uniformly in $s$ as $n \to \infty$ and $0 < \int_0^{s/n} \| \delta (v) \| \, dv < C < \infty$.

**Theorem 2.** Assuming A1 and A2, under model (2.17) we have that

(a) $Pr \left\{ a_n S^{1/2} - b_n \leq x \right\} \to 0$ if $gn^{-1} = o \left( n^{1/2} \log_2^{-1/2} \right)$

(b) $Pr \left\{ a_n S^{1/2} - b_n \leq x \right\} \to \exp (-2e^{-x})$, if $gn = o \left( n^{-1/2} \log_2^{1/2} \right)$

(c) $0 \leq Pr \left\{ a_n S^{1/2} - b_n \leq x \right\} \leq \exp (-2e^{-x+C})$, otherwise,

where $a_n$ and $b_n$ are as in Theorem 1 and $S$ is either the $W$ or the $LM$ statistic in (2.13) and (2.15) respectively.

We now comment on the results of Theorem 2. First the main conclusion that we draw is that the tests $W$ or the $LM$ are consistent irrespectively on the location of the break. Moreover, parts (a) and (b) indicate that the tests have non-trivial power against local alternatives of the type

$$H_a: \quad gn = C \left( \frac{\log_2^{1/2} n}{n^{1/2}} \right).$$
(Observe that when \( g_n = Cn^{-1/2} \), we have the standard local alternatives considered in the literature of structural break models.) That is, as long as the size of the jump of the break \( g_n \) satisfies that \( g_n^{-1} = o\left(n^{1/2} \log_2^{-1/2} n\right) \), the test rejects with probability 1 as the sample size increases to infinity. This is in clear contrast to tests which focus on the “middle” or at the end of the sample as we now discuss. First Andrews (2003) discusses that tests when we assume that the break happens to be at the end of the sample are not consistent.

On the other hand, regarding standard tests, e.g. \( W^* = \sup_{[\tau n] \leq s < n - [\tau n]} W(s) \), we now describe the lack of consistency. To that end, let’s suppose for simplicity that the break is a one-time one. Next, suppose that the break occurs at time \( s_0 \in ([\tau n], n - [\tau n]) \) for some \( 0 < \tau \leq 1/2 \). Then, proceeding as in the proof of Theorem 2, cf. (5.14), we obtain that \( W(s) \) will have a “noncentrality function” of order

\[
O\left(\frac{n-s_0}{n-s}\right)^{1/2} \left(\frac{n}{s}\right)^{1/2} \left(\frac{n-s_0}{n}\right) n^{1/2} g_n \right) \text{I} (s < s_0) \\
+ O\left(\frac{n-s}{n}\right)^{1/2} \left(\frac{s_0}{s}\right)^{1/2} \left(\frac{s_0}{n}\right) n^{1/2} g_n \right) \text{I} (s_0 \leq s)
\]

which is \( O\left(n^{1/2} g_n\right) \) because \( 0 < s_0/n < 1 \). So, if the break occurs in the “middle” of the sample we have the standard results that \( W^* \) has non-trivial power against local alternatives of order \( O\left(g_n = n^{-1/2}\right) \), whereas the power of the test converges to 1 if the size of the break satisfies that \( g_n^{-1} = o\left(n^{1/2}\right) \). However, when the break occurs at time \( s_0 < [\tau n] \), the previous result does not follow as we now discuss. (The treatment and conclusions when the break is at \( s_0 > n - [\tau n] \) is identical and so it is not explicitly discussed.) In this scenario we have from (2.18) that the “noncentrality function” of \( W(s) \) is proportional to

\[
C \left(\frac{n-s}{n}\right)^{1/2} \left(\frac{s_0}{n}\right) n^{1/2} g_n = C \left(\frac{s_0}{n}\right) n^{1/2} g_n = G_n(s_0)
\]

for some \( 0 < C < \infty \) because \( s_0 < s \). Now if \( s_0 = o\left(n^{1/2}\right) \), we have that \( W^* \) has trivial power as the “noncentrality function” converges to zero uniformly in \([\tau n] < s < n - [\tau n]\). That is, the distribution of \( W^* \) is the same as under the null hypothesis of no break. Next, when \( s_0 \) satisfies that \( C^{-1} n^{1/2} < s_0 < C n^{1/2} \), the “noncentrality function” \( G_n(s_0) \) is bounded, so that it indicates that \( W^* \) is not consistent in this region either. Finally, when the break \( s_0 \) lies in the interval \((n^{1/2}, [\tau n])\), we have that \( W^* \) has nontrivial power for local alternatives of order

\[
g_n = O\left(n^{1/2} s_0^{-1}\right).
\]

The latter means that when the time of the break satisfies \( s_0 \leq C n \log_2^{-1/2} n \), the tests defined in (2.13) and (2.15) have better local power than the conventional tests when the supremum is taken in the interval \(([\tau n], n - [\tau n])\). That is, the latter suggests that the standard tests are only able to detect local alternatives of order \( O\left(n^{1/2-\gamma}\right) \), whereas Theorem 2 indicates that ours detect local alternatives of order \( O\left(n^{-1/2} \log_2^{1/2} n\right) \), and thus the “standard” tests have zero relative efficiency compares to ours.
Finally, Theorem 2 indicates that the tests have power against more general type of alternatives and breaks. But more importantly it indicates that the penalty to pay to allow the break to be anywhere in the sample instead of being in the “middle” is very negligible. However, this loss is negligible when we compare it to the fact that there is no need to choose the value of $\varepsilon$ or $\hat{t}$ or the consistency of the test, apart from the undesirable property that with the same data set, two different practitioners may conclude differently by choosing two different values of $\tau$ in the definition of $W^*$.

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 also the explained or endogenous variables of the model are subject to nonlinearities. The latter is the case with nonlinear simultaneous equation systems and with nonlinear transformation models. More specifically, we have that the vector of observables $\{z_t\}_{t \in \mathbb{Z}}$ is partitioned as $\{y_t, x_t\}_{t \in \mathbb{Z}}$, where $y_t$ is a $G \times 1$ vector of endogenous variables and $x_t$ is a $N \times 1$ vector of explanatory variables, and we have that there exists a $G \times 1$ vector of functions $u(y_t, x_t; \theta, \varphi)$ such that

$$u_t = u(y_t, x_t; \theta_0, \varphi_0), \quad t = 1, \ldots, n,$$

where respectively $\theta$ and $\varphi$ are a $p_1$- and $p_2$-dimensional vectors of unknown parameters and $u_t$ is the disturbance error term. As usual a subscript $0$ to a parameter vector, say $\theta_0$, indicates the true value whereas $\theta$ is any admissible value of the parameter in its compact parameter set. By (nonlinear) transformation models we mean that $u(y_t, x_t; \theta, \varphi)$ takes the form

$$u(y_t, x_t; \theta, \varphi) = g(y_t; \theta, \varphi) - h(x_t; \theta, \varphi).$$

For instance, one of these transformations is the well-known Box-Cox model,

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

where $\lambda$ is a component of the vector $(\theta', \varphi')$. Another transformation of interest is given in Burbidge, Magee and Robb (1988), where now $g(y_t; \theta, \varphi)$ is

$$g(y_t; \theta, \varphi) = \arcsinh(\lambda y_t)/\lambda,$$

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

We are interested in the hypothesis of constancy of the parameters $\theta$ against a one-time structural break. Again if $p_2 = 0$, we would have the pure structural change hypothesis. More specifically, denoting $\{\theta(s)\}_{s \geq 1}$ the parameter vector subject to possible one-time structural break as

$$\theta(s) = \begin{cases} \theta + \delta & 1 < t < s \\ \theta & s \leq t < n, \end{cases}$$

the null hypothesis $H_0$ becomes

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

being the alternative hypothesis $H_1$ the negation of the null, that is

$$H_1 : \text{For some } s : p^* < s \leq n - p^*, \quad \delta_0 \neq 0$$
with \( p^* = 2p_1 + p_2 \), that is the dimension of the vector \((\delta', \theta', \varphi')'\).

A common procedure to obtain estimators of the parameters is to exploit the assumption that the disturbance error term \( u = (y_t, x_t; \theta_0, \varphi_0) \), are orthogonal to a set of variables. That is, \( \mathbb{E}(u (y_t, x_t; \theta_0 + \delta_0, \varphi_0) \otimes P(x_t)) = 0 \) for \( 1 \leq t \leq s \), whereas for \( s < t \leq n \), \( \mathbb{E}(u (y_t, x_t; \theta_0, \varphi_0) \otimes P(x_t)) = 0 \) with \( P(\cdot) \) a \( M \times 1 \) column vector of known functions and \( M \geq p^* \). Let us call \( \otimes \) denoting the Kronecker product. As elsewhere, see Hansen (1982), Chamberlain (1985) or Robinson (1988) among others, for a given \( s \), we estimate the parameters as

\[
(3.3) \quad \left( \hat{\delta}^{(s)}, \hat{\theta}^{(s)}, \hat{\varphi}^{(s)} \right)' = \arg \min_{(\delta', \theta', \varphi')' \in \Theta} Q_n^{(s)}(\delta, \theta, \varphi),
\]

where \( \Theta \) is a compact parameter set and \( Q_n^{(s)}(\delta, \theta, \varphi) \) is the GMM objective function

\[
(3.4) \quad Q_n^{(s)}(\delta, \theta, \varphi) = \left( \sum_{t=1}^{n} f_t^{(s)}(\delta, \theta, \varphi) \right) \mathcal{Y}_n \left( \sum_{t=1}^{n} f_t^{(s)}(\delta, \theta, \varphi) \right),
\]

where \( \mathcal{Y}_n \) is some positive and symmetric \( 2GM \times 2GM \) definite matrix and

\[
(3.5) \quad f_t^{(s)}(\delta, \theta, \varphi) = \begin{cases} u_t^{(s)}(\delta, \theta, \varphi) \otimes P(x_t) & \text{if } 1 \leq t \leq s \\ 0 & \text{if } s < t \leq n. \end{cases}
\]

We shall indicate that although different weighting matrices \( \mathcal{Y}_n \) and instruments \( P(x_t) \) lead to different GMM estimators with different asymptotic covariance structures, in the paper we will not examine how to compute the optimal GMM estimator. This is done for expositional purposes as the interest is to examine the behaviour of tests when there is no prior information about the location of the break, although the results follow when the optimal weighting matrix and instruments, as discussed in Robinson (1988), are chosen.

We now introduce the following regularity assumptions and the extension of Proposition 1.

\[ \text{A3:} \{x_t\}_{t \in \mathbb{Z}} \text{ and } \{u_t\}_{t \in \mathbb{Z}} \text{ are two sequences of random variables such that} \]
\[ x_t = h_x(\bar{x}_t) \text{ and } u_t = \sum_{i=0}^{\infty} h_v(\bar{u}_t), \text{ where } \{\bar{x}_t\}_{t \in \mathbb{Z}} \text{ and } \{\bar{u}_t\}_{t \in \mathbb{Z}} \text{ are given by} \]
\[ \bar{x}_t = \sum_{i=0}^{\infty} \chi_i \xi_{t-i}, \sum_{i=0}^{\infty} \|\chi_i\|^{1/\zeta} < \infty \text{ and } \chi_0 = I_{N \times N}, \]
\[ \bar{u}_t = \sum_{i=0}^{\infty} \phi_i \eta_{t-i}, \sum_{i=0}^{\infty} \|\phi_i\|^{1/\zeta} < \infty \text{ and } \phi_0 = I_{G \times G}, \]

for some \( \zeta \geq 3/2 \) and the functions \( g_x(\cdot) \) and \( g_v(\cdot) \) are two smooth differentiable vector-functions everywhere in their domain of definition such that \( \{x_t\}_{t \in \mathbb{Z}} \) and \( \{u_t\}_{t \in \mathbb{Z}} \) are \( \mathbf{L}^4 \) – NED (Near Epoch Dependent) sequences of size \( t > 1. \) That is,
\[
\|x_t - \mathbb{E}(x_t | \xi_t, ..., \xi_{t-m})\|_4 = O(m^{-\alpha}) \quad \text{and} \quad \|u_t - \mathbb{E}(u_t | \eta_t, ..., \eta_{t-m})\|_4 = O(m^{-\alpha}) ,
\]
where \( \| \xi_t \|_4 = \left( \mathbb{E} \| \xi_t \|^4 \right)^{1/4} \). In addition, the sequences \( \{ \xi_t \}_{t \in \mathbb{Z}} \) and \( \{ \eta_t \}_{t \in \mathbb{Z}} \) are mutually independent.

In what follows, for given \( s \), we abbreviate \((\delta', \theta', \varphi')'\) by \( \psi \) and we write \( \hat{\psi}^{(s)} = (\hat{\delta}^{(s)}, \hat{\theta}^{(s)}, \hat{\varphi}^{(s)})' \) as the estimator of \( \psi_0 \) given in (3.3).

**A4:** \( \hat{\psi}^{(s)} \) is a consistent estimator of \( \psi_0 \) and

(a) \( \text{Var} \left( \frac{1}{n^{1/2}} \sum_{t=\lfloor nt_1 \rfloor+1}^{\lfloor nt_2 \rfloor} g_t^{(s)} (\psi_0) \right) \to (\tau_2 - \tau_1) \Delta \)

(b) \( \frac{1}{n^{1/2}} \sum_{t=\lfloor nt_1 \rfloor+1}^{\lfloor nt_2 \rfloor} g_t^{(s)} (\psi_0) \to \mathcal{N} \left( 0, (\tau_2 - \tau_1) \Delta \right) \)

(c) \( \bar{F}_{\delta} ([\lfloor nt_1 \rfloor, \lfloor nt_2 \rfloor]) - F_{\delta} (\psi_0) \overset{P}{\to} 0 \)

(d) \( \bar{F}_{\theta} ([\lfloor nt_1 \rfloor, \lfloor nt_2 \rfloor]) - F_{\theta} (\psi_0) \overset{P}{\to} 0 \),

for any consistent estimator \( \bar{\psi}^{(s)} = (\bar{\delta}, \bar{\theta})' \) of \( \psi_0 \) and where \( \bar{F}_{\delta} (\cdot, \cdot) \) is as given in (3.9) but with \( \bar{\psi}^{(s)} \) replaced by \( \psi \) there.

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 for the error \( \{u_t\}_{t \in \mathbb{Z}} \) to be not only serially dependent but also heteroscedastic. That is, we allow \( \mathbb{E} (u_t u'_t | x_t) = \Sigma (x_t) \). A4 is satisfied under primitive conditions. However, as the literature is full of references for which A4 holds true, see for instance Andrews (1993), we keep A4 for simplicity. In addition, A4 implies that \( \bar{\psi}^{(s)} \) satisfies the central limit theorem, as we state in the next proposition for easy reference.

**Proposition 2.** For given \( s \), under A3 and A4, we have that

\[
\frac{1}{n^{1/2}} \left( \hat{\psi}^{(s)} - \psi_0 \right) \overset{d}{\to} \mathcal{N} \left( 0, B^{(\tau)} - 1 A^{(\tau)} B^{(\tau) - 1} \right),
\]

where \( \tau = \lim_{n \to \infty} s/n \) and

\[
B^{(\tau)} = \begin{pmatrix}
\tau F_{\delta} (\psi_0) & \tau F_{\theta} (\psi_0) \\
0 & (1 - \tau) F_{\theta} (\psi_0)
\end{pmatrix}
\]

\[
A^{(\tau)} = \begin{pmatrix}
\tau F_{\delta} (\psi_0) & \tau F_{\theta} (\psi_0) \\
0 & (1 - \tau) F_{\theta} (\psi_0)
\end{pmatrix}
\]

\[
\times \begin{pmatrix}
\tau \Delta & \\
0 & (1 - \tau) \Delta
\end{pmatrix}
\]

\[
\times \begin{pmatrix}
\tau F_{\delta} (\psi_0) & \tau F_{\theta} (\psi_0) \\
0 & (1 - \tau) F_{\theta} (\psi_0)
\end{pmatrix}
\].

**Proof.** The proof of the proposition is standard and so it is omitted. \( \square \)

We now give an extension of Proposition 1. To that end, we denote the second moment of \( k^{-1/2} \sum_{t=1}^{k} u_t \otimes P (x_t) \), where \( u_t = u_t^{(s)} (\psi_0) \), by \( \Delta_k \), whereas the “long-run” covariance matrix \( \Delta \) is given by its limit, that is

\[
\lim_{n \to \infty} \Delta_n = \sum_{j=-\infty}^{\infty} \mathbb{E} ((u_{t+j} \otimes P (x_{t+j})) (u_t \otimes P (x_t)))' = : \Delta.
\]

Observe that under \( H_0, u_t^{(s)} (\psi_0) = u (y_t, x_t, \theta_0, \varphi_0) \) given in (3.1) for all \( s = 1, \ldots, n \).
**Proposition 3.** Assuming A3, we can construct on a probability space a GM-dimensional Wiener process $B(k)$ with independent components such that, for some $1 < \zeta < 2$,

$$
(3.7) \quad \sup_{1 \leq k \leq n} \left\| \sum_{t=1}^{k} (u_t \otimes P(x_t)) - \Delta_k^{1/2} B(k) \right\| = O_p \left( n^{\frac{\zeta+2}{\zeta+4}} \right).
$$

Before we present the test, let’s introduce some notation and definitions. Write

$$
F_\delta (\psi) = \lim_{n \to \infty} \frac{1}{s} \sum_{t=1}^{s} \mathbb{E} \left\{ \frac{\partial}{\partial \theta^t} g_t^{(s)} (\psi) \right\}
$$

and

$$
F_\theta (\psi) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E} \left\{ \frac{\partial}{\partial \psi^t} g_t^{(s)} (\psi) \right\},
$$

where we have abbreviated $(\theta^t, \varphi^t)$ by $\theta^t$ and

$$
g_t^{(s)} (\psi) = n_t^{(s)} (\psi) \otimes P(x_t)
$$

with $n_t^{(s)} (\psi)$ defined in (3.5). On the other hand, the sample analogues of $F_\delta (\psi)$ and $F_\theta (\psi)$ are obtained respectively by $\hat{F}_\delta (0, s)$ and $\hat{F}_\theta (0, n)$, where

$$
\hat{F}_\delta (\ell_1, \ell_2) = \frac{1}{\ell_2 - \ell_1} \sum_{t=\ell_1+1}^{\ell_2} \left\{ \frac{\partial}{\partial \theta^t} g_t^{(s)} \left( \gamma^{(s)} \right) \right\}
$$

and

$$
\hat{F}_\theta (\ell_1, \ell_2) = \frac{1}{\ell_2 - \ell_1} \sum_{t=\ell_1+1}^{\ell_2} \left\{ \frac{\partial}{\partial \psi^t} g_t^{(s)} \left( \gamma^{(s)} \right) \right\}
$$

and $\hat{\psi} = \left( \hat{\gamma}^{(s)}, \hat{\varphi}^{(s)}, \hat{\theta}^{(s)} \right)'$ is the estimator of the parameters in (3.3). Notice that when $t \leq s$,

$$
\frac{\partial}{\partial \psi^t} g_t^{(s)} (\psi) = \frac{\partial}{\partial \theta^t} g_t^{(s)} (\psi).
$$

Next, an estimator of $\Delta$ is computed as

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

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

$$
\hat{\gamma} (\ell) = \frac{1}{n} \sum_{t=1}^{n-\ell} \left\{ g_t^{(s)} \left( \gamma^{(s)} \right) g_{t+\ell}^{(s)} \left( \hat{\psi}^{(s)} \right) \right\}.
$$

By A3 and A4, we know that the estimator in (3.10) is consistent for $\Delta$, see Andrews (1991) among others. Notice that because we do not assume that $\mathbb{E} (u_t u_t' | x_t) = \Sigma (x_t) = \Sigma$, we cannot then compute $\Delta$, say, as $\hat{\Xi}_n \Delta$ in (2.8). Similarly, we notice that estimates of $\hat{F}_\delta$ and $\hat{F}_\theta$ are respectively $\hat{F}_\delta (0, s)$ and $\hat{F}_\theta (0, n)$ given in (3.9).

Now, if in the definition of $\hat{\psi}^{(s)}$ in (3.3), we replaced the weighting matrix $Y_n$ by $\text{diag} \left( (s\Delta_n)^{-1}, (n-s\Delta_n)^{-1} \right)$, we would obtain that $B(\tau) = A(\tau)$, and hence that

$$
n^{1/2} \left( \hat{\psi}^{(s)} - \psi_0 \right) \overset{d}{\to} \mathcal{N} \left( 0, B^{(\tau)} \right).
$$
Finally, observe that

\[
\hat{\psi}(s) = \psi_0 + \left( s \hat{F}_\delta'(0, s) \hat{\Delta}^{-1} \hat{F}_\delta(0, s) \right) * \left( \frac{s \hat{F}_\delta'(0, s) \hat{\Delta}^{-1} \hat{F}_\delta(0, s)}{s \hat{F}_\delta'(0, s) \hat{\Delta}^{-1} \hat{F}_\delta(0, s) + (n - s) \hat{F}_\delta'(s, n) \hat{\Delta}^{-1} \hat{F}_\delta(s, n)} \right)^{-1}
\]

(3.11) \quad \times \left( \frac{\hat{F}_\delta'(0, s) \hat{\Delta}^{-1}}{\hat{F}_\delta'(0, s) \hat{\Delta}^{-1} \hat{F}_\delta(s, n) \hat{\Delta}^{-1}} \right) \left( \sum_{t=1}^{s} g_t(s) (\psi_0) - \frac{s}{n} \sum_{t=1}^{n} g_t(s) (\psi_0) \right),

where for example \( \hat{F}_\delta'(0, s) \) is as in (3.9) but with \( \hat{\psi}(s) \) being replaced by an intermediate point between \( \psi_0 \) and \( \hat{\psi}(s) \).


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

\[ H_1(s) : \text{For some } s : p^* < s \leq n - p^* \quad \delta_0 \neq 0. \]

In this case the Wald statistic is based on whether \( \hat{\delta}(s) \) in (3.3) is significantly different than zero.

Recalling (3.11) and standard arguments imply that replacing \( \hat{F}_\delta'(0, s) \) and \( \hat{F}_\delta'(s, n) \) by \( \hat{F}_\delta'(0, n) = \hat{F}_\theta \), we have that

\[
\hat{\delta}(s) - \delta_0 = C_n^{-1} \hat{\psi}(s) \left( I_{p_1 \times p_1; 0_{p_1 \times p_2}} \hat{F}_\delta' \hat{\Delta}^{-1} \left( \sum_{t=1}^{s} g_t(s) (\psi_0) - \frac{s}{n} \sum_{t=1}^{n} g_t(s) (\psi_0) \right) \right) \times (1 + o_p(1)),
\]

where the \( o_p(1) \) is uniformly in \( s \) and

\[
C_n \left( \hat{\psi}(s) \right) = s \left( \hat{F}_\delta' \hat{\Delta}^{-1} \hat{F}_\delta - \frac{s}{n} \hat{F}_\delta' \hat{\Delta}^{-1} \hat{F}_\delta \left( \hat{F}_\delta' \hat{\Delta}^{-1} \hat{F}_\delta \right)^{-1} \hat{F}_\delta' \hat{\Delta}^{-1} \hat{F}_\delta \right) = s \left( \frac{n - s}{n} \right) \left( I_{p_1 \times p_1; 0_{p_1 \times p_2}} \left( \hat{F}_\delta' \hat{\Delta}^{-1} \hat{F}_\delta \right) \left( I_{p_1 \times p_1; 0_{p_1 \times p_2}} \right) \right).
\]

Therefore, defining

\[
\tilde{\delta}(s) = C_n \left( \hat{\psi}(s) \right) \hat{\delta}(s),
\]

we write the Wald statistic for \( H_0 \) against \( H_1(s) \) as

\[
W(s) = \frac{n - s}{n - s} \hat{\delta}(s) \left( I_{p_1 \times p_1; 0_{p_1 \times p_2}} \left( \hat{F}_\delta' \hat{\Delta}^{-1} \hat{F}_\delta \right) \left( I_{p_1 \times p_1; 0_{p_1 \times p_2}} \right) \right)^{-1} \tilde{\delta}(s).
\]

On the other hand, as in Section 2.1, for the hypothesis testing in (2.12) we employ the statistic

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

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

$$
(3.13) \quad \left( I_{p_1 \times p_1} : \theta_{p_1 \times p_2} \right) \hat{F}_{\theta} (0, s) \hat{\Delta}^{-1} \sum_{t=1}^{s} g^{(s)}_{t} (0, \hat{\theta}),
$$

where $\hat{\theta}$ is the GMM estimator defined as

$$
(3.14) \quad \hat{\theta} = \arg \min_{\theta} \left( \sum_{t=1}^{n} u(y_t, x_t; \theta) \otimes P(x_t) \right) \hat{\Delta}^{-1} \left( \sum_{t=1}^{n} u(y_t, x_t; \theta) \otimes P(x_t) \right)
$$

with $\hat{\Delta}$ being a consistent estimator of $\Delta$ defined as in (3.10) but with

$$
\hat{\gamma}(\ell) = \frac{1}{n} \sum_{t=1}^{n-\ell} \left\{ g^{(s)}_{t} (\hat{\theta}) g^{(s)}_{t+\ell} (\hat{\theta})' \right\}
$$

there and $\hat{\theta}$ an estimator similar to $\theta$ but with $\hat{\Delta}$ replaced by the identity matrix there. So, from the definition of $\hat{\theta}$ in (3.14), we have that (3.13) becomes

$$
\text{ln} (s) = \left( I_{p_1 \times p_1} : \theta_{p_1 \times p_2} \right)' \hat{F}_{\theta} (0, s) \hat{\Delta}^{-1} \left\{ \sum_{t=1}^{s} g^{(s)}_{t} (0, \theta_0) - \frac{s}{n} \hat{F}_{\theta} (0, s) \left( \hat{F}^{'}_{\theta} (0, n) \hat{\Delta}^{-1} \hat{F}_{\theta} (0, n) \right)^{-1} \hat{F}_{\theta} (0, n) \hat{\Delta}^{-1} \sum_{t=1}^{n} g^{(s)}_{t} (0, \theta_0) \right\}
$$

where

$$
\hat{F}_{\theta} (0, r) = \frac{1}{r} \sum_{t=1}^{r} \left\{ \frac{\partial}{\partial \theta} g^{(s)}_{t} (0, \theta) \right\}.
$$

So, the $LM$ test becomes

$$
\mathcal{L}M(s) = \frac{n}{n-s} \frac{1}{s} \text{ln} (s)' \left( \hat{F}^{'}_{\theta} (0, n) \hat{\Delta}^{-1} \hat{F}_{\theta} (0, n) \right)^{-1} \text{ln} (s).
$$

Now for the hypothesis testing in (2.12), we will compute

$$
(3.15) \quad \mathcal{L}M = \max_{p^* < s \leq n-p^*} \mathcal{L}M(s).
$$

We have now the following result.

**Theorem 3.** Assuming $A3 - A5$, under $H_0$, we obtain that

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

(b) $a_n \mathcal{L}M^{1/2} - b_n \xrightarrow{d} \mathcal{T}$,

where $a_n$ and $b_n$ are as defined in Theorem 1 and $\mathcal{T}$ is defined in (2.16).

4. MONTE CARLO EXPERIMENT

This section examines the finite sample properties of the proposed test statistics compared to some well-known test statistics. In particular, we compare the statistic $\mathcal{W}$ with other sup Wald tests with trimming factor 0.05 and 0.15, which are explored by Andrews (1993) and are denoted by $\mathcal{W}_{0.05}$ and $\mathcal{W}_{0.15}$, respectively, and with the CUSUM test of Brown, Durbin, and Evans (1975). With some abuse of notation, we use $\mathcal{W}$ for $a_n \mathcal{W}^{1/2} - b_n$ in this section.
We report the results based on the Wald statistics $W(s)$, $s = p^* + 1, \ldots, n - p^*$, with $\hat{V}_n(s)$ defined as in (2.11) where $\hat{\mathbb{E}}(s) = \frac{1}{n} \sum_{t=1}^{s} x_t x_t' \hat{\delta}_t$. This corresponds to the standard Wald statistic for each $s$ with heteroskedasticity-robust variance estimator as proposed by White (1980). The same $W(s)$s are used to construct $W$, $W_{0.05}$ and $W_{1.15}$.

The proof of Theorem 1 reveals that the test $W$ comprises two components, one of which is asymptotically negligible compared to the other. In particular, if $p^* < \log n$, we have that

$$W = \max \{ W_1, W_2 \}$$

where

$$W_1 = a_n \sup_{p^* < s \leq \log n, n - \log n < s \leq n - p^*} W(s)^{1/2} - b_n \quad \text{and} \quad W_2 = a_n \sup_{\log n < s \leq n - \log n} W(s)^{1/2} - b_n,$$

and $W_1$ diverges to negative infinity while $W_2$ converges in distribution to $T$. In practical applications with moderate sample sizes, $W = W_2$, since the number $p^*$ of parameters in the model would be at least $[\log n]$, which is 6 even for $n = 500$. For instance, a regression model with a structural break that contains only two regressors and a constant will have 6 parameters and $W = W_2$. However, $W_1$ diverges at a very slow rate of $\log \log n$.

Thus, we first examine the impact of the presence of $W_1$ on the finite sample performance of $W$. For that purpose, we estimate a regression model only with a constant

$$y_t = \beta + \delta 1 \{ t \leq \tau n \} + u_t,$$

and compute both $W$ and $W_2$. Note that $p^* = 2$ for this model and thus $p^*$ is smaller than $\log n$ for all the considered $n = 50, 100, 200, 400$ and 800. The data are generated with $u_t \sim iid N(0, 1)$, and $\beta = \delta = 0$.

Table 1 reports the results of empirical rejection frequencies out of 1000 repetition for the critical levels of 10%, 5% and 1%. As it might be expected from the $\log_2 n$ rate, the small sample size distortion of $W$ is severe due to $W_1$. The rejection frequencies of $W$ increases significantly until the sample size becomes 400, at which they become more than four times of the nominal levels. The performance does not improve even when $n = 800$. On the other hand, $W_2$ keeps reasonable empirical sizes for all the sample sizes considered and it improves on the size as the sample size increases. It also makes good comparison with the other two sup-tests. While it slightly over-rejects than the statistic $W_{1.15}$, it does perform better than $W_{0.05}$, which shows severe over-rejection tendency when the sample size is small. As repeatedly reported in the literature, the CUSUM test seems very conservative.

We turn to a more general set-up

$$y_t = x_t' \beta + z_t' \delta 1 \{ t \leq \tau n \} + u_t,$$

where $t = 1, \ldots, n$, $u_t \sim iid N(0, 1)$, $z_t = \left( 1, (-1)^t \right)'$, and $x_t = (z_t', x_{2t})'$. This design, in which the coefficients of $\left( 1, (-1)^t \right)$ subject to a change, has been used frequently in the previous literature, e.g. by Andrews (1992) and Krämer and Sonnberger (1986) among others. We tried different distributions for $x_{2t}$ and report the results from the case where $x_t$ is a vector of 5-dimensional multivariate standard normal. The presence of $x_{2t}$ prevents $W_1$ from contributing to $W$, which makes the comparisons more reasonable and realistic. Empirical size properties of the tests
are reported in Table 2 for \( n = 50, 100 \) and 200, \( \beta = 0 \) and \( \delta = 0 \). The table does not include \( \mathcal{W}_2 \) since \( \mathcal{W} = \mathcal{W}_2 \). The three sup-Wald tests show higher rejection rates than previous table, while the CUSUM test under-rejects even further. For smaller sample sizes (\( n = 50, 100 \)), the two sup-tests \( \mathcal{W}_{05} \) and \( \mathcal{W}_{15} \) show marked over-rejection and \( \mathcal{W} \) performs much better than the two. The severe over-rejection tendency of \( \mathcal{W}_{05} \) remains even when \( n = 200 \), while that of \( \mathcal{W}_{15} \) has decreased. Overall, the size property of \( \mathcal{W} \) seems to be reasonably good in small samples.

Next, we consider the power properties of the four tests using similar alternatives considered in Andrews (1992). In particular, \( \beta = 0 \) while \( \delta \) varies in a way that

\[
\delta = b \left( \frac{\log_2 n}{n} \right)^{1/2} (\cos \theta, \sin \theta)'.
\]

We use the \( \log_2 n \) factor as discussed in Section 2.3. We report the results for \( b = 4.8, 9.6, 12; \theta = 0, \pi/4, \pi/2; \tau = 0.05, 0.25, 0.5, 0.75, 0.95 \). The rejection frequencies are for 5% significance levels. The results with asymptotic critical values are reported in Tables 3 - 5 for \( n = 50, 100 \) and 200, and those with empirical critical values, which are obtained from 10,000 repetition, are reported in Tables 6 - 8 for the same sample sizes. That is, Tables 6 - 8 report size-corrected powers.

We first confirm well-known features of power properties of some of the tests. The CUSUM test has greater power against early breaks than late ones, while the sup-tests appear invariant to the direction of time. The CUSUM test also exhibits

\[
\text{the same reasoning as above, we compare}
\]

\( \mathcal{W} \) to the direction of time. The CUSUM test also exhibits fluctuations in power across the angle \( \theta \). It has larger rejection rates when the angle is zero as opposed to when it is 90°, at which the test has trivial power. On the other hand, all the other tests have similar powers against the angle \( \theta \).

The tests are compared each other now. In terms of powers based on asymptotic critical values, \( \mathcal{W}_{15} \) dominates the others for all parameter values when \( n = 50 \) and \( \mathcal{W}_{05} \) overpowers the others for most parameterization when \( n = 200 \). This is much expected due to the severe over-rejection tendencies of those two tests for given sample sizes. However, \( \mathcal{W}_{15} \) has similar power as \( \mathcal{W}_{05} \) against the break at \( \tau = 0.5 \), even though the empirical size of \( \mathcal{W}_{05} \) is 0.192 compared to 0.076 of \( \mathcal{W}_{15} \) at the 5% significance level when \( n = 200 \).

We move to the size-adjusted powers. When \( n = 50 \), the three tests based on Wald statistics are equivalent when we correct the critical values, since the number of trimming at the factor 0.15 is smaller than the number of the parameters. For this reason, the tests have only trivial power at \( \tau = 0.05 \) and 0.95 while the CUSUM test has some power at \( \tau = 0.05 \) and the angle \( \theta = 0 \). Otherwise, \( \mathcal{W} \) mostly dominates CUSUM except when \( \tau = 0.25 \) and \( \theta = 0 \). Since \( \mathcal{W} = \mathcal{W}_{05} \) when \( n = 100 \) by the same reasoning as above, we compare \( \mathcal{W} \) with \( \mathcal{W}_{15} \) and CUSUM tests. For the breaks happening around the middle of the sample (\( \tau = 0.25, 0.5 \) and 0.75), \( \mathcal{W}_{15} \) overpowers the others, although the CUSUM test does slightly better when \( \tau = 0.25 \) and \( \theta = 0 \). For the break at the beginning of the sample (\( \tau = 0.05 \)), the CUSUM and \( \mathcal{W} \) outperform. The CUSUM test does better for \( \theta = 0, \pi/4 \) and \( \mathcal{W} \) for \( \theta = \pi/2 \). And \( \mathcal{W} \) performs more uniformly over the angle \( \theta \) than CUSUM. For the break at the end of the sample with \( \tau = 0.95 \), \( \mathcal{W} \) dominates the others.

When \( n = 200 \), all four tests are separated and neither dominates or is dominated. The CUSUM test has some advantages for small \( \theta \) and \( \tau \), that is, for the early break. The other tests are not much variant against whether the break is early or late. Thus, the comparison is made on how much portion of the sample
subject to change. Clearly, $W_{15}$ outperforms the others when the break is in the middle of the sample, say, when $\tau = 0.5, 0.25,$ and $0.75,$ followed by $W_{05}$ and then by $W$. As the change time moves toward both ends of the sample, either $\tau = 0.05$ or $0.95$, $W$ and $W_{05}$ come to dominates $W_{15}$, which exhibits only trivial powers when $b = 4.8$. Between the two, there is not much difference, although $W$ does slightly better than $W_{05}$ when $\tau = 0.05$ and vice versa when $\tau = 0.95$. It is also worthwhile to note that the minimum rejection rates of $W$ and $W_{05}$ over all the scenarios considered are $0.09$ and $0.087$, respectively, compared to $0.059$ and $0.021$ of $W_{15}$ and CUSUM.

In sum, either $W$ or $W_{15}$ is recommended in practice depending on whether there is a prior that the break is in the middle of the sample or not. The statistic $W_{05}$ has a serious problem with size in small samples, even for $n = 200$. The CUSUM test has an issue with power as well-documented in the literature. Thus, $W$ and $W_{15}$ appear more attractive in practice as they exhibit less severe size distortions and have relatively uniform power properties across various parameterization. It is the reason why Andrews recommends $W_{15}$ to practitioners over $W_{05}$. As we have seen here, however, when the break is suspected to occur for a relatively small proportion of the sample, $W_{15}$ is dominated by $W$ in terms of power. When we are concerned with a possible break not in the middle of the sample but anywhere, $W$ becomes more attractive. The uniform performance of $W$ is better than $W_{15}$ in the sense that the former has bigger minimum power than the latter.
5. PROOFS OF RESULTS

5.1. Proof of Proposition 1.

Let \( q = 6\zeta - 3 \) and define the sequences \( \{\tilde{x}_t\}_{t \in \mathbb{Z}} \) and \( \{\tilde{u}_t\}_{t \in \mathbb{Z}} \) as

\[
\tilde{x}_t = \sum_{0 \leq i < Ct^{1/\zeta}} \psi_i x_{t-i} ; \quad \tilde{u}_t = \sum_{0 \leq i < Ct^{1/\zeta}} \varphi_i \eta_{t-i},
\]

where \( C \) is a large enough positive constant but fixed. Note that the sequences \( \{\tilde{x}_t\}_{t \in \mathbb{Z}} \) and \( \{\tilde{u}_t\}_{t \in \mathbb{Z}} \) behave as MA \((Ct^{1/\zeta})\). More specifically, abbreviating \( \{\tilde{x}_t, \tilde{u}_t\}_{t \in \mathbb{Z}} \) by \( \{\tilde{v}_t\}_{t \in \mathbb{Z}} \), \( \tilde{v}_t \) is independent of \( v_s \) if \( s < t \) and \( t - s > Ct^{1/\zeta} \). However, contrary to \( \{x_t\}_{t \in \mathbb{Z}} \) or \( \{u_t\}_{t \in \mathbb{Z}} \), it is not covariance stationary as \( \mathbb{E} \|\tilde{v}_t\|^2 \) is a function of \( t \).

We first show that

\[
(5.1) \quad \sup_{1 \leq k \leq n} \left| \sum_{t=1}^{k} v_t - \sum_{t=1}^{k} \tilde{v}_t \right| = o_p \left( n^{\frac{\zeta+2}{\zeta+3}} \right),
\]

where we abbreviate \( \{x_t, u_t\}_{t \in \mathbb{Z}} \) by \( \{v_t\}_{t \in \mathbb{Z}} \). (5.1) implies that it suffices to show (2.9) with \( \{\tilde{v}_t\}_{t \in \mathbb{Z}} \) replacing \( \{v_t\}_{t \in \mathbb{Z}} \) there. For that purpose, denote \( \{v_t - \tilde{v}_t\}_{t \in \mathbb{Z}} \) by \( \{\tilde{v}_t\}_{t \in \mathbb{Z}} \) and likewise \( \{\tilde{x}_t\}_{t \in \mathbb{Z}} \) and \( \{\tilde{u}_t\}_{t \in \mathbb{Z}} \). Assuming, without loss of generality, that \( n = 2^d \), Wu (2007) Proposition 1 implies that the expectation of the left side of (5.1) is bounded by

\[
(5.2) \quad \sum_{p=0}^{d} \left[ \sum_{m=1}^{2^{d-p}} \mathbb{E} \left( \sum_{t=2^p(m-1)+1}^{2^p m} \tilde{v}_t \right)^2 \right]^{1/2}.
\]

Because \( \tilde{v}_t = \tilde{x}_t \tilde{u}_t + x_t \tilde{u}_t + \tilde{x}_t u_t \), by standard inequalities, it suffices to show (5.2) when \( \tilde{v}_t \) is replaced by, say, \( x_t \tilde{u}_t \). To that end, we first notice that because A1 implies that, for \( t_1 \leq t_2 \leq t_3 \leq t_4 \),

\[
|\mathbb{E}(x_t \tilde{u}_t x_{t_2} \tilde{u}_{t_2})| = |\mathbb{E}(\tilde{u}_t \tilde{u}_{t_2})\mathbb{E}(x_t x_{t_2})| \leq Ct_1^{-1/3} \sum_{j=1}^{\infty} j^{-\zeta} (t_2 - t_1 + j)^{-\zeta}
\]

because \( |\mathbb{E}(\tilde{u}_t \tilde{u}_{t_2})| \leq C \sum_{j=Ct_1^{1/\zeta}}^{\infty} j^{-\zeta} (t_2 - t_1 + j)^{-\zeta} \), we then have that

\[
\mathbb{E} \left( \sum_{t=2^p(m-1)+1}^{2^p m} x_t \tilde{u}_t \right)^2 \leq C \sum_{t=2^p(m-1)+1}^{2^p m} t^{-1/3} \leq C 2^{2p/3} m^{-1/3}.
\]

Notice that \( \text{Var}(\tilde{u}_t) \) and \( \text{Var}(\tilde{x}_t) \) are \( O(t^{-1/3}) \).

Hence, because \( q = 6\zeta - 3 \) and \( n = 2^d \), the right side of (5.2) is bounded by

\[
C \sum_{p=0}^{d} 2^{d-p} \left[ \sum_{m=1}^{2^{d-p} m^{-1/3}} \right]^{1/2} = O \left( n^{1/3} \log n \right),
\]

and (5.1) follows by Markov’s inequality.

To show that (2.9) holds true for \( \{\tilde{v}_t\}_{t \in \mathbb{Z}} \), 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
\[ B_\ell = \left\{ t : n_{\ell-1} + \ell^{1/q} \leq t \leq n_\ell \right\}, \] where \( n_\ell = n_{\ell-1} + \ell^{1/\zeta} + \ell^{1/q}; \ell \geq 1, \) for some \( 1 < \zeta < 2, \) and consider the sequences

\[ \xi_\ell = \sum_{t \in A_\ell} \dot{\nu}_t; \quad \epsilon_\ell = \sum_{t \in B_\ell} \dot{\nu}_t. \]

Let \( \ell \) be the value such that \( n_{\ell-1} < n \leq n_\ell, \) respectively. We first show that

\begin{equation}
\sup_{1 \leq \ell \leq \ell} \left\| \sum_{i=1}^n \dot{\nu}_t - \mathbb{E}^{1/2} \mathcal{B}(n_\ell) \right\| = O_p \left( n^{\frac{\zeta \epsilon + 2}{4(\zeta + 1)}} \right), \tag{5.3} \end{equation}

where \( \mathcal{B}(t) \) denotes the standard Wiener process. Observe that \( \sum_{t=1}^n \dot{\nu}_t = \sum_{j=1}^\ell (\xi_j + \epsilon_j), \) and that \( 0 < C_1 \leq \ell n^{-\zeta/(\zeta+1)} \leq C_2 < \infty. \)

Now by construction and A1, \( \{\xi_j\}_{j \geq 1} \) is a sequence of nonidentically independent random variables with finite 4th moments. So, by Götze and Zaitsev’s (2007) Theorem 4, we can find a sequence of iid normally distributed random variables \( \{\nu_t\}_{t \geq 1} \) such that

\begin{equation}
\sup_{1 \leq \ell \leq \ell} \left\| \sum_{j=1}^\ell \mathbb{E}^{1/2} (\xi_j^2) \nu_j \right\| = O_p \left( \left( \frac{\mathbb{E} \|\xi_j\|^4}{\ell} \right)^{1/4} \right) = O_p \left( n^{\frac{\zeta \epsilon + 2}{4(\zeta + 1)}} \right), \end{equation}

because \( \mathbb{E} \|\xi_j\|^4 \equiv O \left( j^{2/\zeta} \right) \) and \( \ell \leq C_2 n^{\zeta/(\zeta+1)}. \) 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), \( \sup_{1 \leq \ell \leq \ell} \left| \sum_{j=1}^\ell \epsilon_j / \ell^{(q+1)/2q} \log_2 \ell \right| = 1 \) almost surely by Shao (1995). So, \( \sup_{1 \leq \ell \leq \ell} \left| \sum_{j=1}^\ell \epsilon_j \right| = o_p \left( n^{\frac{\zeta \epsilon + 2}{4(\zeta + 1)}} \right) \) and we conclude that

\begin{equation}
\left\| \sum_{i=1}^n \dot{\nu}_t - \mathbb{E}^{1/2} (\xi_j^2) \nu_j \right\| = O_p \left( n^{\frac{\zeta \epsilon + 2}{4(\zeta + 1)}} \right). \end{equation}

Now, proceeding as with the proof of (5.2), we obtain that

\begin{equation}
\left| \mathbb{E} (\xi_j^2) - \mathbb{E} \left( \sum_{t \in A_j} \nu_t \right) \right|^2 \leq C \sum_{t=n_{j-1} + 1}^{n_j} t^{-1/3} \leq C j^{(2/\zeta - 1)/3} \end{equation}

because \( n_j = \sum_{k=1}^j (h^{1/\zeta} + h^{1/q}) \leq C j^{(\zeta + 1)/\zeta}. \) So, using that \( (a-b)^2 \leq a^2 - b^2 \) for \( a > b > 0, \) LIL implies that

\begin{equation}
\sup_{1 \leq \ell \leq \ell} \left\| \sum_{j=1}^\ell \left\{ \mathbb{E}^{1/2} (\xi_j^2) - \mathbb{E}^{1/2} \left( \sum_{t \in A_j} \nu_t \right) \right\} \nu_j \right\| = O_p \left( n^{\frac{\zeta \epsilon + 2}{4(\zeta + 1)}} \right) \end{equation}

Likewise, as we proceed above, because by LIL, \( \sup_{1 \leq \ell \leq \ell} \left| \sum_{j=1}^\ell \mathbb{E}^{1/2} (\xi_j^2) \nu_j \right| = O_p \left( n^{\frac{\zeta \epsilon + 2}{4(\zeta + 1)}} \right) \) by (5.1), we conclude that

\begin{equation}
\left\| \sum_{t=1}^n \nu_t - \left( \mathbb{E} \left( \sum_{t \in A_j} \nu_t \right)^2 + \mathbb{E} \left( \sum_{t \in B_j} \nu_t \right)^2 \right)^{1/2} \right\| = O_p \left( n^{\frac{\zeta \epsilon + 2}{4(\zeta + 1)}} \right). \end{equation}
But,

$$\sum_{j=1}^{\ell} \left\{ \left( \mathbb{E} \left( \sum_{t \in A_j} v_t \right)^2 + \mathbb{E} \left( \sum_{t \in B_j} v_t \right)^2 \right)^{1/2} - \left( \sup_{j \leq j_n} j_n^{1/2} \right) \right\} v_j$$

is $O_p \left( \left( \mathbb{E} \left( \sum_{t \in A_j} v_t \right)^2 + \mathbb{E} \left( \sum_{t \in B_j} v_t \right)^2 \right) \right)^{1/2} \log_2 n$ by LIL and standard algebra after observing that, say $\mathbb{E} \left( \sum_{t \in A_j} v_t \right)^2 = j^{1/\zeta} \mathbb{E}_{j_{1/\zeta}}$, and $|\Xi_{n_j} - \Xi_{j_{1/\zeta}}| \leq C \sum_j j^{-1/3}$ and $n_j = \sum_{h=1}^j (h^{1/\zeta} + h^{1/\eta})$. So we have finish the proof of (5.3), and that of (2.9), when the supremum is taken over those $k$ for which there exists $n$ satisfying $n_k = k$.

Hence, to complete the proof we have to examine the approximation when $n_j < k < n_{j+1}$. But by Theorem 1.2. of Csörgö and Révész (1981),

$$\max_{n_{k-1} \leq j \leq n_k} |B(n_{k-1}) - B(j)| = O \left( \ell^{1/2} \log^{1/2} \ell \right)$$

a.s. This concludes the proof.

5.2. Proof of Theorem 1.

We begin with part (a). First observe that by A1, LIL and Proposition 1,

$$\sup_{p^* < s} \left\| \frac{M_s - \sum_{t=1}^s u_t}{s^{1/2} \log_2^{1/2} s} \right\| = O_p(1).$$

We first examine the behaviour of $\left( \frac{n_{n-s}}{n-s} \right)^{1/2} s^{-1/2} \delta(s)$, which is

$$\left( \frac{n}{n-s} \right)^{1/2} \left( I_{p^*} \mathbb{I}_{p^*} \right) \left( 1 - \sum_{t=1}^n u_t M_n - \frac{1}{n} \sum_{t=1}^n u_t \right)$$

$$- \left( \frac{n}{n-s} \right)^{1/2} \left( I_{p^*} \mathbb{I}_{p^*} \right) \left( 1 - \sum_{t=1}^n u_t M_n - \frac{1}{n} \sum_{t=1}^n u_t \right).$$

Because Proposition 1, and then LIL, implies that $\sup_{s} \left\| s^{-1/2} \log_2^{1/2} s \right\| s \sum_{t=1}^n u_t \right\| = O_p(1)$, for $a_t = x_t$ or $u_t$, it is clear that it suffices to examine the behaviour of the first term of the last displayed expression.

Next, we have that

$$a_n \sup_{p^* < s < \log n} \left\| \left( \frac{n}{n-s} \right)^{1/2} \frac{1}{s^{1/2}} \left( 1 - \sum_{t=1}^n u_t M_n - \frac{1}{n} \sum_{t=1}^n u_t \right) \right\| - b_n \to -\infty$$

holds true because by (5.4), $\sup_{p^* < s < \log n} \left\| \frac{M_s}{s} \left( \frac{M_s}{n} \right)^{-1} - I \right\| = o_p \left( \log^{1/2} n \right)$ and because by LIL and Proposition 1, $\sup_{p^* < s < \log n} \left\| s^{-1/2} \sum_{t=1}^n u_t \right\| = o_p \left( \log^{1/2} n \right)$,
we conclude that (5.5) holds true. Next,

(5.6) \[ a_n \sup_{n - \log n < s \leq n - p^*} \left\| \left( \frac{n}{n-s} \right)^{1/2} \frac{1}{s^{1/2}} \left\{ \sum_{t=1}^{s} v_t - M_t M_n^{-1} \sum_{t=1}^{n} v_t \right\} \right\| - b_n \overset{P}{\to} -\infty. \]

Indeed (5.6) follows easily after observing that by (5.4) and standard algebra,

(5.7) \[ \sup_{n - \log n < s \leq n - p^*} \left\| \frac{M_s}{s} - \frac{\Sigma}{\sqrt{n}} \right\| \leq \sup_{n - \log n < s \leq n - p^*} \left\{ \frac{n}{s} \left\| \frac{M_n}{n} - \frac{\Sigma}{\sqrt{n}} \right\| + \frac{n-s}{n} \left\| \frac{M_n - M_s}{n-s} - \frac{\Sigma}{\sqrt{n}} \right\| \right\} = o_p \left( \log^{1/2} n \right) \]

and then because LIL and Proposition 1 imply that

\[ \sup_{n - \log n < s \leq n - p^*} \left\| \left( \frac{n}{n-s} \right)^{1/2} \left( \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t - \frac{s}{n} \right) \right\| \leq \sup_{n - \log n < s \leq n - p^*} \left\{ \left( \frac{n}{s} \right)^{1/2} \frac{1}{(n-s)^{1/2}} \sum_{t=s+1}^{n} v_t \right\} \]

after using that, with \( v_t = v_{n-t+1} \),

(5.8) \[ a_n \sup_{\log n \leq s \leq n - \log n} \left\| \left( \frac{n}{n-s} \right)^{1/2} \left( \sum_{t=1}^{s} v_t - \frac{s}{n} \right) \right\| - b_n = a_n \sup_{\log n \leq s \leq n - \log n} \left\| \left( \frac{n}{n-s} \right)^{1/2} \left( \frac{n-s}{ns} \sum_{t=1}^{s} v_t - \frac{s}{n} \sum_{t=s+1}^{n} v_t \right) \right\| - b_n. \]

When the supremum is taken in \( \log n \leq s < n - [\epsilon n] \) for any \( \epsilon > 0 \), by standard functional central limit theorems (FCLT) and continuous mapping theorem,

(5.9) \[ a_n \sup_{\log n \leq s < n - [\epsilon n]} \left\| \left( \frac{n}{n-s} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=s+1}^{n} v_t \right\| - b_n \overset{P}{\to} -\infty \]

whereas for \( [\epsilon n] < s \leq n - \log n \)

(5.10) \[ a_n \sup_{[\epsilon n] < s \leq n - \log n} \left\| \left( \frac{n}{n-s} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t \right\| - b_n \overset{P}{\to} -\infty. \]

Now we consider the interval \( n/\log n < s \leq [\epsilon n] \). Denoting by \( U(s^*) \) the Ornstein-Uhlenbeck process, by Proposition 1 and the change of time \( s \to e^{s^*} \),
we obtain that
\[
\alpha_n \sup_{n/\log n < s \leq [n\epsilon]} \left( \frac{n-s}{n} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t - b_n
\]
\[
= \alpha_n \sup_{n/\log n < s \leq [n\epsilon]} \left( \frac{\log n - s^*}{\log n} \right) U(s^*) - b_n + o_p(1).
\]

So, gathering (5.1)–(5.6), (5.9)–(5.10) and the last displayed expression, we have that
\[
\sup_{p^* < s \leq n-p^*} \left( \frac{n-s}{n} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t - b_n
\]
and that by Proposition 1, we can consider
\[
\sup_{p^* < s \leq n-p^*} \left( \frac{n-s}{n} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t - b_n
\]

Now consider the interval \( n/\log^\beta n < s \leq [n\epsilon] \) for some \( 1/2 < \beta < 1 \). Then
\[
\mathbb{P}_{n/\log^\beta n < s < n/\log^\beta n} \left( \frac{n-s}{n} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t - b_n \overset{p}{\to} -\infty.
\]
Moreover, because
\[
\alpha_n \sup_{n/\log^\beta n < s < n/\log^\beta n} \left( \frac{n-s}{n} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t - b_n \overset{p}{\to} -\infty.
\]
by Theorem 1.1.1 of Csörgő and Révész (1981). See also their remark 1.1.2. So, the supremum must be in the region Likewise \( \alpha_n \sup_{n/\log^\beta n < s < n/\log^\beta n} \left( \frac{n-s}{n} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t \overset{p}{\to} -\infty \).

So, gathering (5.11)–(5.13), we have that \( \sup_{p^* < s \leq n-p^*} \left( \frac{n-s}{n} \right)^{1/2} \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t \ overset{p}{\to} -\infty \) is governed by the distribution of
\[
\max \left\{ \alpha_n \sup_{n/\log n < s < [n\epsilon]} \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t - b_n ; \alpha_n \sup_{n/\log n < s < [n\epsilon]} \frac{1}{s^{1/2}} \sum_{t=1}^{s} v_t - b_n \right\}
\]
From here the proof follows by Lemma 2.2 of Horváth (1993), after we employ (5.7) and that by Proposition 1, we can consider \( \{v_t\}_{t \in \mathbb{Z}} \) as a sequence of independent normal random variables. Observe that \( \hat{\Xi}_n(a) - \Xi = O_p(n^{-1/2}) \) by Robinson.
independent standard normal random variables.

Next part (b). The proof proceeds as that of part (a) after observing that \( \hat{F}_n^{(s)} = \delta^{(s)} \) and \( \hat{F}_n^{(n)} - \delta = o_p(1) \) by Robinson (1998). This concludes the proof.

5.3. Proof of Theorem 2.

Denoting

\[
U_n(s) = \sum_{i=1}^{s} v_i - \frac{s}{n} \sum_{i=1}^{n} v_i; \quad F(s/n) = \frac{1}{n} \sum_{i=1}^{s} \delta(t/n) - \frac{s}{n} \frac{1}{n} \sum_{i=1}^{n} \delta(t/n),
\]

it is immediate to observe that \( \delta^{(s)} = U_n(s) + n^{1/2} g_n \mathbb{E}(\bar{x}_i\bar{x}_j) F(s/n)(1 + o_p(1)) \), so that proceeding as in the proof of Theorem 1, it suffices to examine the behaviour of

\[
Z_n(s) = \left( \frac{n}{n-s} \right)^{1/2} \left( U_n(s) + F(s/n) \right)^{1/2} V_{11}^{-1} \left( U_n(s) + F(s/n) \right)^{2}
\]

and

\[
V_{11} = \left( I_{p_1 \times p_1} : 0_{p_1 \times p-1} \right) V \left( I_{p_1 \times p_1} : 0_{p_1 \times p-1} \right)^t.
\]

Let us begin with part (a). Because for any sequence \( \{c(s)\} \leq 1 \), \( |c(n/2)| \leq \sup_{1 \leq s \leq n} |c(s)| \), we conclude that

\[
\Pr\{Z_n < x\} \leq \Pr\left\{ a_n Z_n^{1/2}(s) - b_n \leq x \right\}.
\]

But, we know that \( (a_n g_n n^{1/2})^{-1} = o\left( (a_n \log_2^{1/2} n)^{-1} \right) = o(b_n^{-1}) \), so with probability approaching 1, we obtain that \( a_n Z_n^{1/2}(s) - b_n \) increases to infinity. From here the conclusion of part (a) is standard. Next we show part (b). First, we observe that when we consider the situation where \( n/\log n \leq s \leq n - n/\log n \), we have that

\[
a_n \left( \frac{n}{n-s} \right)^{1/2} \frac{1}{s^{1/2}} n^{1/2} g_n F(s/n) = o(b_n).
\]

Indeed, if \( s < s_0 \), then the left side of (5.14) is in absolute value bounded by

\[
C \left( \frac{n-s_0}{n} \right)^{1/2} \left( \frac{s_0}{s} \right)^{1/2} \left( \frac{n-s_0}{n-s} \right)^{1/2} o\left( a_n \log_2^{1/2} n \right) = o(b_n),
\]

whereas when \( s_0 \leq s \), we have that the left side of (5.14) is bounded in absolute value by

\[
C \left( \frac{n-s_0}{n} \right)^{1/2} \left( \frac{s_0}{s} \right)^{1/2} \left( \frac{s_0}{n} \right)^{1/2} o(\log n) = o(b_n).
\]

So, arguing as in the proof of Theorem 1, (5.14) implies that the supremum must be in the interval \( \log n \leq s \leq n/\log n \) or \( n - n/\log n < s < n - \log n \). But, in the latter intervals we have that

\[
a_n \left( \frac{n}{n-s} \right)^{1/2} \frac{1}{s^{1/2}} n^{1/2} g_n F(s/n) = o(1),
\]
so that we conclude that \( \Pr \{ Z_n < x \} \) is
\[
\Pr \left\{ a_n \left( \sup_{p^* \leq s \leq n - p^*} \left( \frac{n}{n - s} \frac{1}{s} U_n(s)' V_{11}^{-1} U_n(s) \right)^{1/2} - b_n < x + o(1) \right) \right\}.
\]
So the asymptotic distribution of the test is the same than that under the null hypothesis. This concludes the proof of part (b). Finally part (c) follows because here \( C^{-1} b_n < a_n g_a n^{1/2} < C b_n \) and because the results in parts (a) and (b).

5.4. Proof of Proposition 2.

First we notice that we can assume without loss of generality that \( E (u_i u_i' | x_t) = \Sigma \). Indeed, if \( E (u_i u_i' | x_t) = \Sigma (x_t) \), it implies that \( u_i = \Sigma^{1/2} (x_t) u_i^* \), where \( \{ u_i^* \}_{i \in \mathbb{Z}} \) satisfies the same assumptions than \( \{ u_i \}_{i \in \mathbb{Z}} \). Then, we have that
\[
u_i \otimes P_\lambda (x_t) = u_i^* \Sigma^{1/2} (x_t) \otimes P_\lambda (x_t) = u_i^* \otimes \bar{P}_\lambda (x_t)
\]
where under suitable regularity conditions on \( \Sigma (x_t) \) and \( P_\lambda (x_t) \), we have that \( \bar{P}_\lambda (x_t) \) is also a \( L^1 - NED \) sequence on size \( t > 1 \). See for instance Davidson (1994, Sec. 17.3). So, from now on, we just assume that \( \Sigma (x_t) = \Sigma \) for all \( t \). The proof proceeds similarly to that of Proposition 1. Denote \( \tilde{g} (\eta_t, \ldots, \eta_{t-Ct^{1/4}}) = \mathbb{E} (u_t | \eta_t, \ldots, \eta_{t-Ct^{1/4}}) \) and \( \bar{P} (\varepsilon_t, \ldots, \varepsilon_{t-Ct^{1/4}}) = \mathbb{E} (P_\lambda (x_t) | \varepsilon_t, \ldots, \varepsilon_{t-Ct^{1/4}}) \) and let \( \{ \varepsilon_t \}_{t \in \mathbb{Z}} = \{ u_t \otimes P_\lambda (x_t) \}_{t \in \mathbb{Z}} \) and \( \{ \varepsilon_t \}_{t \in \mathbb{Z}} = \{ \varepsilon_t - \varepsilon_t' \}_{t \in \mathbb{Z}} \) with \( \{ \varepsilon_t' \}_{t \in \mathbb{Z}} = \{ \tilde{g} (\eta_t, \ldots, \eta_{t-Ct^{1/4}}) \otimes \bar{P} (\varepsilon_t, \ldots, \varepsilon_{t-Ct^{1/4}}) \}_{t \in \mathbb{Z}} \). Now, because by A4 we know that
\[
\mathbb{E} \left\| \sum_t \varepsilon_t \right\|^2 \leq C \sum_t t^{-2s/q},
\]
proceeding as in Proposition 1, cf. (5.1) – (5.2), we have that
\[
\mathbb{E} \sup_{1 \leq k \leq n} \left\| \sum_{t=1}^k \varepsilon_t \right\|^2 \leq \sum_{p=0}^d \left[ \sum_{m=1}^{2^m} \mathbb{E} \left( \sum_{t=2^p(m-1)+1}^{2^p m} \varepsilon_t \right)^2 \right]^{1/2} \leq O \left( n^{1/3} \log n \right).
\]
So, it suffices to consider the strong approximation of
\[
\sum_{t=1}^k \tilde{g}_u (\eta_t, \ldots, \eta_{t-Ct^{1/4}}) \otimes \bar{P} (\varepsilon_t, \ldots, \varepsilon_{t-Ct^{1/4}}) = \sum_{t=1}^k \varepsilon_t.
\]
But the latter follows proceeding as we did in the proof of Proposition 1 after we observe that because \( \tilde{g}_u (\eta_t, \ldots, \eta_{t-Ct^{1/4}}) \) and \( \bar{P} (\varepsilon_t, \ldots, \varepsilon_{t-Ct^{1/4}}) \) are MA (\( Ct^{1/4} \)) we then have that this observation together with A3 implies that \( \{ \varepsilon_t \}_{t \in \mathbb{Z}} \) satisfies the same conditions as \( \{ \varepsilon_t \}_{t \in \mathbb{Z}} \) in Proposition 1.

5.5. Proof of Theorem 3.

Because the proof of parts (a) and (b) are similarly handled, we shall explicitly proof part (b). For that purpose, we first notice that
\[
(5.15) \quad \hat{\Delta} - \Delta = O_p \left( m^{-1/2} \right) = o_p (1)
\]
and

\[
(5.16) \quad \sup_{p^* < s} \left\| \frac{\sum_{t=1}^{s} \frac{\partial}{\partial \vartheta} g_t (\tilde{\vartheta}) - sF_\vartheta}{s^{1/2} \log_2^{1/2} s} \right\| = O_p (1)
\]

as we now prove. (5.16) holds true because

\[
\frac{1}{s} \sum_{t=1}^{s} \frac{\partial}{\partial \vartheta} g_t (\tilde{\vartheta}) = \frac{1}{s} \sum_{t=1}^{s} \frac{\partial}{\partial \vartheta} g_t (\vartheta_0) + \frac{1}{s} \sum_{t=1}^{s} \left( \frac{\partial}{\partial \vartheta} g_t (\tilde{\vartheta}) - \frac{\partial}{\partial \vartheta} g_t (\vartheta_0) \right)
\]

\[
= \frac{1}{s} \sum_{t=1}^{s} \frac{\partial}{\partial \vartheta} g_t (\vartheta_0) + O_p \left( n^{-1/2} \right).
\]

But A3, LIL and Proposition 2 imply that

\[
\sup_{p^* < s} \left\| \frac{\sum_{t=1}^{s} \frac{\partial}{\partial \vartheta} g_t (\vartheta_0) - sF_\vartheta}{s^{1/2} \log_2^{1/2} s} \right\| = O_p (1).
\]

From here, (5.16) follows by straight arguments.

We now examine the behaviour of

\[
\left( \frac{n - s}{n} \right)^{1/2} s^{-1/2} \text{Im} (s)
\]

which by (5.15) it suffices to examine that of

\[
\text{Im}^* (s) = \left( \frac{n - s}{n} \right)^{1/2} s^{-1/2} \left( I_{p_1 \times p_1} \Delta^{-1} \sum_{t=1}^{s} g_t (0, \vartheta_0) \right. \nabla \tilde{F}_\vartheta (0, s) \Delta^{-1} \sum_{t=1}^{s} g_t (0, \vartheta_0) \nabla \tilde{F}_\vartheta (0, n) \right)
\]

\[
= - \frac{s}{n} \nabla \tilde{F}_\vartheta (0, s) \left( \tilde{F}_\vartheta (0, n) \Delta^{-1} \tilde{F}_\vartheta (0, n) \right. \nabla \tilde{F}_\vartheta (0, n) \Delta^{-1} \sum_{t=1}^{s} g_t (0, \vartheta_0) \right).
\]

Now proceeding as with the proof of Theorem 1,

\[
a_n \sup_{p^* < s < \log n} \| \text{Im}^* (s) \| - b_n \rightarrow -\infty
\]

because by (5.16), \( \sup_{p^* < s < \log n} \left\| \tilde{F}_\vartheta (0, s) - F_\vartheta \right\| = o_p \left( \log_2^{1/2} n \right) \) and \( \sup_{p^* < s < \log n} \left\| s^{-1/2} \sum_{t=1}^{s} g_t (0, \vartheta_0) \right\| = o_p \left( \log_2^{1/2} n \right) \) by LIL and Preposition 2.

Next, we examine the behaviour in the region \( n - \log n < s \leq n - p^* \). But here, (5.16) implies that \( a_n \sup_{n - \log n < s \leq n - p^*} \| \text{Im}^* (s) \| - b_n \) is governed by

\[
a_n \sup_{n - \log n < s \leq n - p^*} \left\| \left( \frac{n - s}{n} \right)^{1/2} s^{-1/2} F_\vartheta \Delta^{-1} \left( \sum_{t=1}^{s} g_t (0, \vartheta_0) - \frac{s}{n} \sum_{t=1}^{s} g_t (0, \vartheta_0) \right) \right\| - b_n
\]

\[
\rightarrow -\infty
\]

proceeding with the same arguments as ion (5.6). Now, in the region \( \log n \leq s \leq n - \log n \). But the proof proceeds as in Theorem 1 because (5.16) indicates that it
suffices to examine
\[
\sup_{\log n \leq s \leq n - \log n} \left\| \left( \frac{n - s}{n} \right)^{1/2} s^{-1/2} \left( I_{p_1 \times p_1}^{p_1 \times p_1} \right) F_0^{-1} \Delta^{-1} \left\{ \sum_{t=1}^{s} g_t(s) (0, \theta_0) - \frac{s}{n} \sum_{t=1}^{n} g_t(s) (0, \theta_0) \right\} \right\| - b_n,
\]
which is essentially the same as (5.8). This concludes the proof of the theorem.
References


<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( W )</td>
<td>0.296</td>
<td>0.214</td>
<td>0.112</td>
</tr>
<tr>
<td>( W_{05} )</td>
<td>0.197</td>
<td>0.109</td>
<td>0.040</td>
</tr>
<tr>
<td>( W_{15} )</td>
<td>0.385</td>
<td>0.309</td>
<td>0.213</td>
</tr>
<tr>
<td>CUSUM</td>
<td>0.158</td>
<td>0.098</td>
<td>0.034</td>
</tr>
<tr>
<td>( W )</td>
<td>0.045</td>
<td>0.017</td>
<td>0.004</td>
</tr>
<tr>
<td>( W_{05} )</td>
<td>0.179</td>
<td>0.109</td>
<td>0.040</td>
</tr>
<tr>
<td>( W_{15} )</td>
<td>0.116</td>
<td>0.074</td>
<td>0.025</td>
</tr>
<tr>
<td>CUSUM</td>
<td>0.063</td>
<td>0.027</td>
<td>0.007</td>
</tr>
<tr>
<td>( W )</td>
<td>0.357</td>
<td>0.288</td>
<td>0.184</td>
</tr>
<tr>
<td>( W_{05} )</td>
<td>0.140</td>
<td>0.089</td>
<td>0.026</td>
</tr>
<tr>
<td>( W_{15} )</td>
<td>0.187</td>
<td>0.127</td>
<td>0.064</td>
</tr>
<tr>
<td>CUSUM</td>
<td>0.116</td>
<td>0.074</td>
<td>0.025</td>
</tr>
<tr>
<td>( W )</td>
<td>0.05</td>
<td>0.385</td>
<td>0.309</td>
</tr>
<tr>
<td>( W_{05} )</td>
<td>0.357</td>
<td>0.288</td>
<td>0.184</td>
</tr>
<tr>
<td>( W_{15} )</td>
<td>0.140</td>
<td>0.089</td>
<td>0.026</td>
</tr>
<tr>
<td>CUSUM</td>
<td>0.116</td>
<td>0.074</td>
<td>0.025</td>
</tr>
<tr>
<td>( W )</td>
<td>0.200</td>
<td>0.123</td>
<td>0.071</td>
</tr>
<tr>
<td>( W_{05} )</td>
<td>0.139</td>
<td>0.078</td>
<td>0.026</td>
</tr>
<tr>
<td>( W_{15} )</td>
<td>0.097</td>
<td>0.050</td>
<td>0.018</td>
</tr>
<tr>
<td>CUSUM</td>
<td>0.083</td>
<td>0.039</td>
<td>0.007</td>
</tr>
<tr>
<td>( W )</td>
<td>0.421</td>
<td>0.342</td>
<td>0.249</td>
</tr>
<tr>
<td>( W_{05} )</td>
<td>0.281</td>
<td>0.203</td>
<td>0.105</td>
</tr>
<tr>
<td>( W_{15} )</td>
<td>0.121</td>
<td>0.062</td>
<td>0.025</td>
</tr>
<tr>
<td>CUSUM</td>
<td>0.072</td>
<td>0.034</td>
<td>0.005</td>
</tr>
<tr>
<td>( W )</td>
<td>0.182</td>
<td>0.103</td>
<td>0.026</td>
</tr>
<tr>
<td>( W_{05} )</td>
<td>0.116</td>
<td>0.070</td>
<td>0.015</td>
</tr>
<tr>
<td>( W_{15} )</td>
<td>0.101</td>
<td>0.048</td>
<td>0.012</td>
</tr>
<tr>
<td>CUSUM</td>
<td>0.083</td>
<td>0.036</td>
<td>0.010</td>
</tr>
</tbody>
</table>

**Empirical size I**

**Table 1**

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Test Statistic</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 50 )</td>
<td>( W )</td>
<td>0.163</td>
<td>0.094</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>( W_{05} )</td>
<td>0.238</td>
<td>0.170</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>( W_{15} )</td>
<td>0.293</td>
<td>0.204</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>CUSUM</td>
<td>0.003</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>( n = 100 )</td>
<td>( W )</td>
<td>0.182</td>
<td>0.103</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>( W_{05} )</td>
<td>0.281</td>
<td>0.203</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>( W_{15} )</td>
<td>0.208</td>
<td>0.140</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>CUSUM</td>
<td>0.024</td>
<td>0.012</td>
<td>0.003</td>
</tr>
<tr>
<td>( n = 200 )</td>
<td>( W )</td>
<td>0.172</td>
<td>0.100</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>( W_{05} )</td>
<td>0.287</td>
<td>0.192</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td>( W_{15} )</td>
<td>0.130</td>
<td>0.076</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>CUSUM</td>
<td>0.045</td>
<td>0.017</td>
<td>0.004</td>
</tr>
</tbody>
</table>

**Empirical size II**

**Table 2**
Testing for Structural Breaks

<table>
<thead>
<tr>
<th>$b / \theta$</th>
<th>$0$</th>
<th>$\pi/4$</th>
<th>$\pi/2$</th>
<th>$0$</th>
<th>$\pi/4$</th>
<th>$\pi/2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$W$</td>
<td></td>
<td></td>
<td>$W_{0.05}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>4.8</td>
<td>0.478</td>
<td>0.470</td>
<td>0.475</td>
<td>0.769</td>
<td>0.768</td>
</tr>
<tr>
<td>9.6</td>
<td>0.976</td>
<td>0.968</td>
<td>0.983</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>12</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.75</td>
<td>4.8</td>
<td>0.385</td>
<td>0.413</td>
<td>0.413</td>
<td>0.679</td>
<td>0.693</td>
</tr>
<tr>
<td>9.6</td>
<td>0.936</td>
<td>0.916</td>
<td>0.945</td>
<td>0.988</td>
<td>0.989</td>
<td>0.994</td>
</tr>
<tr>
<td>12</td>
<td>0.990</td>
<td>0.991</td>
<td>0.994</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.95</td>
<td>4.8</td>
<td>0.094</td>
<td>0.113</td>
<td>0.111</td>
<td>0.311</td>
<td>0.324</td>
</tr>
<tr>
<td>9.6</td>
<td>0.120</td>
<td>0.122</td>
<td>0.117</td>
<td>0.366</td>
<td>0.368</td>
<td>0.357</td>
</tr>
<tr>
<td>12</td>
<td>0.104</td>
<td>0.097</td>
<td>0.108</td>
<td>0.362</td>
<td>0.354</td>
<td>0.365</td>
</tr>
<tr>
<td>$\mathcal{W}_{15}$</td>
<td>CUSUM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>4.8</td>
<td>0.806</td>
<td>0.814</td>
<td>0.811</td>
<td>0.012</td>
<td>0.003</td>
</tr>
<tr>
<td>9.6</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td>0.127</td>
<td>0.017</td>
</tr>
<tr>
<td>12</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.266</td>
<td>0.026</td>
</tr>
<tr>
<td>0.75</td>
<td>4.8</td>
<td>0.744</td>
<td>0.755</td>
<td>0.752</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>9.6</td>
<td>0.990</td>
<td>0.992</td>
<td>0.997</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>12</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>0.95</td>
<td>4.8</td>
<td>0.374</td>
<td>0.386</td>
<td>0.422</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>9.6</td>
<td>0.437</td>
<td>0.444</td>
<td>0.427</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>12</td>
<td>0.429</td>
<td>0.432</td>
<td>0.424</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Empirical Power with Asym c.v. (level .05) for n=50

Table 3
\begin{table}[h]
\centering
\begin{tabular}{ccccccccc}
\hline
\multicolumn{2}{c}{$b / \theta$} & 0 & $\pi/4$ & $\pi/2$ & 0 & $\pi/4$ & $\pi/2$ \\
\hline
\multirow{3}{*}{0.50} & 4.8 & 0.499 & 0.453 & 0.475 & 0.835 & 0.811 & 0.807 \\
& 9.6 & 0.989 & 0.988 & 0.989 & 1.000 & 1.000 & 0.999 \\
& 12 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 & 1.000 \\
\multirow{3}{*}{0.75} & 4.8 & 0.406 & 0.390 & 0.401 & 0.768 & 0.731 & 0.758 \\
& 9.6 & 0.955 & 0.933 & 0.946 & 0.993 & 0.996 & 0.987 \\
& 12 & 0.992 & 0.994 & 0.998 & 1.000 & 1.000 & 1.000 \\
\multirow{3}{*}{0.95} & 4.8 & 0.130 & 0.119 & 0.131 & 0.412 & 0.405 & 0.397 \\
& 9.6 & 0.194 & 0.181 & 0.226 & 0.505 & 0.483 & 0.539 \\
& 12 & 0.246 & 0.226 & 0.262 & 0.561 & 0.563 & 0.592 \\
\hline
\end{tabular}
\end{table}

Empirical Power with Asym c.v. (level .05) for n=100

\textbf{Table 4}
<table>
<thead>
<tr>
<th>$b / \theta$</th>
<th>0</th>
<th>$\pi/4$</th>
<th>$\pi/2$</th>
<th>0</th>
<th>$\pi/4$</th>
<th>$\pi/2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{W}$</td>
<td>0.493</td>
<td>0.493</td>
<td>0.473</td>
<td>0.850</td>
<td>0.849</td>
<td>0.839</td>
</tr>
<tr>
<td>9.6</td>
<td>0.990</td>
<td>0.991</td>
<td>0.993</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>12</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.75</td>
<td>4.8</td>
<td>0.377</td>
<td>0.396</td>
<td>0.387</td>
<td>0.774</td>
<td>0.769</td>
</tr>
<tr>
<td>9.6</td>
<td>0.964</td>
<td>0.965</td>
<td>0.955</td>
<td>0.998</td>
<td>0.999</td>
<td>0.997</td>
</tr>
<tr>
<td>12</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.95</td>
<td>4.8</td>
<td>0.227</td>
<td>0.209</td>
<td>0.213</td>
<td>0.532</td>
<td>0.503</td>
</tr>
<tr>
<td>9.6</td>
<td>0.537</td>
<td>0.498</td>
<td>0.553</td>
<td>0.803</td>
<td>0.781</td>
<td>0.814</td>
</tr>
<tr>
<td>12</td>
<td>0.687</td>
<td>0.657</td>
<td>0.683</td>
<td>0.903</td>
<td>0.890</td>
<td>0.902</td>
</tr>
<tr>
<td>$\mathcal{W}_{15}$</td>
<td>0.493</td>
<td>0.843</td>
<td>0.853</td>
<td>0.825</td>
<td>0.272</td>
<td>0.107</td>
</tr>
<tr>
<td>9.6</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.881</td>
<td>0.445</td>
<td>0.009</td>
</tr>
<tr>
<td>12</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.981</td>
<td>0.711</td>
<td>0.003</td>
</tr>
<tr>
<td>0.75</td>
<td>4.8</td>
<td>0.752</td>
<td>0.743</td>
<td>0.715</td>
<td>0.053</td>
<td>0.033</td>
</tr>
<tr>
<td>9.6</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
<td>0.223</td>
<td>0.071</td>
<td>0.010</td>
</tr>
<tr>
<td>12</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.404</td>
<td>0.128</td>
<td>0.006</td>
</tr>
<tr>
<td>0.95</td>
<td>4.8</td>
<td>0.262</td>
<td>0.277</td>
<td>0.250</td>
<td>0.017</td>
<td>0.014</td>
</tr>
<tr>
<td>9.6</td>
<td>0.349</td>
<td>0.350</td>
<td>0.367</td>
<td>0.016</td>
<td>0.027</td>
<td>0.013</td>
</tr>
<tr>
<td>12</td>
<td>0.401</td>
<td>0.398</td>
<td>0.396</td>
<td>0.013</td>
<td>0.006</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Empirical Power with Asym c.v. (level .05) for n=200

Table 5
\[ b / \theta = 0 \quad \pi/4 \quad \pi/2 \quad 0 \quad \pi/4 \quad \pi/2 \]

<table>
<thead>
<tr>
<th></th>
<th>( W )</th>
<th>( W_{0.05} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>0.347</td>
<td>0.339 0.338 0.347 0.339 0.338</td>
</tr>
<tr>
<td>9.6</td>
<td>0.957</td>
<td>0.941 0.953 0.957 0.941 0.953</td>
</tr>
<tr>
<td>12</td>
<td>0.997</td>
<td>0.992 0.999 0.997 0.992 0.999</td>
</tr>
<tr>
<td>0.75</td>
<td>0.288</td>
<td>0.308 0.295 0.288 0.308 0.295</td>
</tr>
<tr>
<td>9.6</td>
<td>0.884</td>
<td>0.849 0.895 0.884 0.849 0.895</td>
</tr>
<tr>
<td>12</td>
<td>0.974</td>
<td>0.969 0.975 0.974 0.969 0.975</td>
</tr>
<tr>
<td>0.95</td>
<td>0.054</td>
<td>0.057 0.069 0.054 0.057 0.069</td>
</tr>
<tr>
<td>9.6</td>
<td>0.061</td>
<td>0.060 0.058 0.061 0.060 0.058</td>
</tr>
<tr>
<td>12</td>
<td>0.062</td>
<td>0.056 0.064 0.062 0.056 0.064</td>
</tr>
<tr>
<td>( W_{15} )</td>
<td>CUSUM</td>
<td>0.266 0.136 0.034</td>
</tr>
<tr>
<td>0.50</td>
<td>0.347</td>
<td>0.339 0.338 0.266 0.136 0.034</td>
</tr>
<tr>
<td>9.6</td>
<td>0.957</td>
<td>0.941 0.953 0.727 0.328 0.016</td>
</tr>
<tr>
<td>12</td>
<td>0.997</td>
<td>0.992 0.999 0.884 0.446 0.004</td>
</tr>
<tr>
<td>0.75</td>
<td>0.288</td>
<td>0.308 0.295 0.068 0.037 0.035</td>
</tr>
<tr>
<td>9.6</td>
<td>0.884</td>
<td>0.849 0.895 0.045 0.020 0.017</td>
</tr>
<tr>
<td>12</td>
<td>0.974</td>
<td>0.969 0.975 0.043 0.016 0.010</td>
</tr>
<tr>
<td>0.95</td>
<td>0.054</td>
<td>0.057 0.069 0.038 0.050 0.040</td>
</tr>
<tr>
<td>9.6</td>
<td>0.061</td>
<td>0.060 0.058 0.036 0.036 0.032</td>
</tr>
<tr>
<td>12</td>
<td>0.062</td>
<td>0.056 0.064 0.019 0.021 0.020</td>
</tr>
</tbody>
</table>

Empirical Power with Emp. c.v. (level .05) for n=50

Table 6
Testing for Structural Breaks

<table>
<thead>
<tr>
<th>$b / \theta$</th>
<th>0</th>
<th>$\pi/4$</th>
<th>$\pi/2$</th>
<th>0</th>
<th>$\pi/4$</th>
<th>$\pi/2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$W$</td>
<td></td>
<td></td>
<td>$W_{05}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>4.8</td>
<td>0.299</td>
<td>0.272</td>
<td>0.289</td>
<td>0.299</td>
<td>0.272</td>
</tr>
<tr>
<td>9.6</td>
<td>0.963</td>
<td>0.959</td>
<td>0.955</td>
<td>0.963</td>
<td>0.959</td>
<td>0.955</td>
</tr>
<tr>
<td>12</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
</tr>
<tr>
<td>0.75</td>
<td>4.8</td>
<td>0.260</td>
<td>0.247</td>
<td>0.246</td>
<td>0.260</td>
<td>0.247</td>
</tr>
<tr>
<td>9.6</td>
<td>0.898</td>
<td>0.870</td>
<td>0.891</td>
<td>0.898</td>
<td>0.870</td>
<td>0.891</td>
</tr>
<tr>
<td>12</td>
<td>0.987</td>
<td>0.981</td>
<td>0.990</td>
<td>0.987</td>
<td>0.981</td>
<td>0.990</td>
</tr>
<tr>
<td>0.95</td>
<td>4.8</td>
<td>0.070</td>
<td>0.061</td>
<td>0.078</td>
<td>0.070</td>
<td>0.061</td>
</tr>
<tr>
<td>9.6</td>
<td>0.119</td>
<td>0.086</td>
<td>0.142</td>
<td>0.119</td>
<td>0.086</td>
<td>0.142</td>
</tr>
<tr>
<td>12</td>
<td>0.142</td>
<td>0.133</td>
<td>0.165</td>
<td>0.142</td>
<td>0.133</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>$W_{15}$</td>
<td>CUSUM</td>
<td></td>
<td>$W_{15}$</td>
<td>CUSUM</td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>4.8</td>
<td>0.488</td>
<td>0.446</td>
<td>0.461</td>
<td>0.344</td>
<td>0.159</td>
</tr>
<tr>
<td>9.6</td>
<td>0.993</td>
<td>0.990</td>
<td>0.993</td>
<td>0.889</td>
<td>0.504</td>
<td>0.013</td>
</tr>
<tr>
<td>12</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.980</td>
<td>0.682</td>
<td>0.008</td>
</tr>
<tr>
<td>0.75</td>
<td>4.8</td>
<td>0.364</td>
<td>0.380</td>
<td>0.376</td>
<td>0.074</td>
<td>0.053</td>
</tr>
<tr>
<td>9.6</td>
<td>0.955</td>
<td>0.946</td>
<td>0.946</td>
<td>0.236</td>
<td>0.087</td>
<td>0.019</td>
</tr>
<tr>
<td>12</td>
<td>0.996</td>
<td>0.995</td>
<td>0.997</td>
<td>0.340</td>
<td>0.110</td>
<td>0.018</td>
</tr>
<tr>
<td>0.95</td>
<td>4.8</td>
<td>0.056</td>
<td>0.064</td>
<td>0.058</td>
<td>0.048</td>
<td>0.046</td>
</tr>
<tr>
<td>9.6</td>
<td>0.079</td>
<td>0.074</td>
<td>0.088</td>
<td>0.041</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td>12</td>
<td>0.100</td>
<td>0.098</td>
<td>0.103</td>
<td>0.029</td>
<td>0.038</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Empirical Power with Emp. c.v. (level .05) for n=100

Table 7
<table>
<thead>
<tr>
<th>$b / \theta$</th>
<th>0</th>
<th>$\pi/4$</th>
<th>$\pi/2$</th>
<th>0</th>
<th>$\pi/4$</th>
<th>$\pi/2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>4.8</td>
<td>0.235</td>
<td>0.213</td>
<td>0.207</td>
<td>0.285</td>
<td>0.263</td>
</tr>
<tr>
<td>9.6</td>
<td>0.954</td>
<td>0.943</td>
<td>0.964</td>
<td>0.968</td>
<td>0.968</td>
<td>0.982</td>
</tr>
<tr>
<td>12</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.75</td>
<td>4.8</td>
<td>0.170</td>
<td>0.173</td>
<td>0.178</td>
<td>0.209</td>
<td>0.228</td>
</tr>
<tr>
<td>9.6</td>
<td>0.855</td>
<td>0.843</td>
<td>0.854</td>
<td>0.899</td>
<td>0.893</td>
<td>0.908</td>
</tr>
<tr>
<td>12</td>
<td>0.986</td>
<td>0.985</td>
<td>0.989</td>
<td>0.991</td>
<td>0.994</td>
<td>0.995</td>
</tr>
<tr>
<td>0.95</td>
<td>4.8</td>
<td>0.121</td>
<td>0.100</td>
<td>0.114</td>
<td>0.135</td>
<td>0.111</td>
</tr>
<tr>
<td>9.6</td>
<td>0.345</td>
<td>0.324</td>
<td>0.384</td>
<td>0.362</td>
<td>0.346</td>
<td>0.386</td>
</tr>
<tr>
<td>12</td>
<td>0.494</td>
<td>0.448</td>
<td>0.490</td>
<td>0.512</td>
<td>0.490</td>
<td>0.522</td>
</tr>
<tr>
<td>$\mathcal{W}_{15}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUSUM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>4.8</td>
<td>0.581</td>
<td>0.604</td>
<td>0.565</td>
<td>0.394</td>
<td>0.195</td>
</tr>
<tr>
<td>9.6</td>
<td>0.997</td>
<td>0.995</td>
<td>0.998</td>
<td>0.946</td>
<td>0.602</td>
<td>0.028</td>
</tr>
<tr>
<td>12</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>0.837</td>
<td>0.021</td>
</tr>
<tr>
<td>0.75</td>
<td>4.8</td>
<td>0.449</td>
<td>0.457</td>
<td>0.437</td>
<td>0.116</td>
<td>0.072</td>
</tr>
<tr>
<td>9.6</td>
<td>0.980</td>
<td>0.989</td>
<td>0.978</td>
<td>0.383</td>
<td>0.153</td>
<td>0.032</td>
</tr>
<tr>
<td>12</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.547</td>
<td>0.223</td>
<td>0.021</td>
</tr>
<tr>
<td>0.95</td>
<td>4.8</td>
<td>0.065</td>
<td>0.061</td>
<td>0.074</td>
<td>0.046</td>
<td>0.037</td>
</tr>
<tr>
<td>9.6</td>
<td>0.120</td>
<td>0.105</td>
<td>0.104</td>
<td>0.052</td>
<td>0.057</td>
<td>0.052</td>
</tr>
<tr>
<td>12</td>
<td>0.151</td>
<td>0.120</td>
<td>0.124</td>
<td>0.046</td>
<td>0.032</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Empirical Power with Emp. c.v. (level .05) for n=200

**Table 8**
<table>
<thead>
<tr>
<th>b / theta</th>
<th>0.50</th>
<th>0.75</th>
<th>0.95</th>
<th>0.50</th>
<th>0.75</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>pi/5</td>
<td>4.800</td>
<td>4.800</td>
<td>4.800</td>
<td>4.800</td>
<td>4.800</td>
<td>4.800</td>
</tr>
<tr>
<td>3*pi/10</td>
<td>1.000</td>
<td>1.000</td>
<td>0.140</td>
<td>1.000</td>
<td>0.987</td>
<td>0.087</td>
</tr>
<tr>
<td>pi/2</td>
<td>1.000</td>
<td>1.000</td>
<td>0.146</td>
<td>1.000</td>
<td>0.987</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>12.000</td>
<td>12.000</td>
<td>12.000</td>
<td>12.000</td>
<td>12.000</td>
<td>12.000</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>0.249</td>
<td>1.000</td>
<td>0.987</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>0.257</td>
<td>1.000</td>
<td>0.987</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>0.249</td>
<td>1.000</td>
<td>0.987</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>0.257</td>
<td>1.000</td>
<td>0.987</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Empirical Power with Bootstrap c.v. (level .05) for n=100

Table 9