Cumulated sum of squares statistics for non-linear and non-stationary regressions

Vanessa Berenguer-Rico* and Bent Nielsen†

03 August 2015

Abstract

We show that the cumulated sum of squares test has a standard Brownian bridge-type asymptotic distribution in non-linear regression models with non-stationary regressors. This contrasts with cumulated sum tests which have been studied previously and where the asymptotic distribution involves nuisance quantities. Through simulation we show that the power is comparable in a wide of range of situations.

Keywords: Cumulated sum of squares, Non-linear Least Squares, Non-stationarity, Specification tests.

JEL classification: C01; C22.

1 Introduction

An increasing range of non-linear models with non-stationary regressors are available in the literature. We show that the specification of such models can be investigated with ease using a cumulated sum of squares test with a standard Brownian bridge asymptotic distribution.

The Brownian bridge asymptotic result of the cumulated sum of squares test has been derived in a linear model framework with stationary and non-stationary regressors, see for instance Brown, Durbin and Evans (1975), McCabe and Harrison (1980), Ploberger and Krämer (1986), Lee, Na and Na (2003), or Nielsen and Sohkanen (2011). In this paper, we first provide a set of general sufficient assumptions for the Brownian Bridge result to hold. Then, we show that these assumptions are satisfied in several different scenarios dealing with non-linear regression functions involving stationary or non-stationary regressors. In contrast, cumulated sum tests based directly on the residuals rather than on their squares have a more complicated asymptotic theory with nuisance terms when the regressors are nonstationary, see Hao and Inder (1996), Xiao and Phillips (2002), Kasparis (2008), Choi and Saikkonen (2010) or Berenguer-Rico and Gonzalo (2014).

The paper is organized as follows. In Section 2, the model and test statistics are put forward. Sections 3 and 4 provide, respectively, high-level and medium-level sets of sufficient assumptions for the Brownian bridge result. Section 5 shows that the assumptions in Sections 3 and 4 are satisfied in various non-linear models. In Section 6 the performance of the test in terms of size and power is investigated through Monte Carlo experiments. The proofs follow in an Appendix.

2 Model and statistics

Consider data $(y_1, x_1), \ldots, (y_n, x_n)$ where y_t is a scalar and x_t is a p-vector. Consider the non-linear regression model

$$y_t = g(x_t, \theta) + \varepsilon_t \qquad t = 1, \ldots, n, \tag{2.1}$$

*Department of Economics, University of Oxford, Mansfield College, and Programme for Economic Modelling.

†Department of Economics, University of Oxford, Nuffield College, and Programme for Economic Modelling.

where the functional form of g is known. The innovation ε_t is a martingale difference sequence with respect to a filtration $\mathcal{F}_t$ with zero mean, variance σ^2 and fourth moment $\varphi^2 = \mathsf{E}\varepsilon_t^4 - (\mathsf{E}\varepsilon_t^2)^2$, the regressor x_t is a p-vector $\mathcal{F}_{t-1}$-adapted, and the parameter θ is a q-vector varying in a parameter space $\Theta \subset \mathbb{R}^q$. The model is a conditional mean model where any unmodelled autocorrelation or correlation between ε_t and x_t will be regarded as misspecification.

The non-linear least squares estimator $\hat{\theta}_n$ of θ is the minimizer of the least squares criterion

$$Q_n(\theta) = \sum_{t=1}^{n} \{y_t - g(x_t, \theta)\}^2. \tag{2.2}$$

The least squares residuals based on the full sample estimation are then $\hat{\varepsilon}_{t,n} = y_t - g(x_t, \hat{\theta}_n)$.

The cumulated sum of squares statistic, is defined as

$$CUSQ_n = \frac{1}{\hat{\varphi}_n} \max_{1 \le t \le n} \left| \frac{1}{\sqrt{n}} \left(\sum_{s=1}^{t} \hat{\varepsilon}_{s,n}^2 - \frac{t}{n} \sum_{s=1}^{n} \hat{\varepsilon}_{s,n}^2 \right) \right|,$$

where the standard deviation estimator can be chosen as, for instance,

$$\hat{\varphi}_n^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{t,n}^4 - \left(\frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{t,n}^2 \right)^2.$$

We will argue that under quite general assumptions,

$$CUSQ_n \xrightarrow{\mathsf{D}} \sup_{0 \le u \le 1} \left| \mathcal{B}_u^0 \right|,$$

where $\mathcal{B}_u^0$ is a standard Brownian bridge. Billingsley (1999, pp. 101–104) gives an analytic expression for the distribution function. In particular, the 95% quantile is 1.36; see Koziol and Byar (1975, Tab. 1).

We also consider a recursive cumulated sum of squares statistic, where the model (2.1) is estimated recursively. Then define the recursive statistic

$$RCUSQ_n = \frac{1}{\hat{\varphi}_n} \max_{n_0 \le t \le n} \left| \frac{1}{\sqrt{n}} \left(\sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2 - \frac{t}{n} \sum_{s=1}^{n} \hat{\varepsilon}_{s,n}^2 \right) \right|.$$

If the sequence of estimators $\hat{\theta}_n$ converges strongly, we can show that also

$$RCUSQ_n \xrightarrow{\mathsf{D}} \sup_{0 \le u \le 1} \left| \mathcal{B}_u^0 \right|.$$

3 Results under High Level Assumptions

We start by proving the Brownian bridge results under a set of high level assumptions to the residuals and martingale difference innovations.

Assumption 3.1 *Suppose $(\varepsilon_t, \mathcal{F}_t)$ is a martingale difference sequence with respect to a filtration $\mathcal{F}_t$, that is ε_t is $\mathcal{F}_t$-adapted and $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ a.s., so that*
(a) $\mathsf{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$ *a.s.;*
(b) $\mathsf{E}(\varepsilon_t^4 - \sigma^4|\mathcal{F}_{t-1}) = \varphi^2$ *a.s.;*
(c) $\sup_t \mathsf{E}(\varepsilon_t^\psi|\mathcal{F}_{t-1}) < \infty$ *a.s. for some* $\psi > 4$.

The first result shows that the tied down cumulated sum of squared innovations converges to a Brownian bridge. This follows from the standard functional central limit theorem for martingale differences, see for instance Brown (1971).

Theorem 3.1 *Suppose Assumption 3.1 is satisfied. Let $\mathcal{B}_u^0$ be a standard Brownian bridge. Then, as a process on $D[0,1]$, the space of right continuous functions with left limits endowed with the Skorokhod metric,*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\left(\varepsilon_t^2-\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2\right)\xrightarrow{\mathsf{D}}\varphi\mathcal{B}_u^0 \qquad u\in[0,1],$$
$$\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^4-\left(\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2\right)^2\xrightarrow{\mathsf{D}}\varphi^2.$$

We would like to formulate similar results for the cumulated sum of squared residuals. This can be done as long as the squares of residuals and innovations are close. We formulate this as two assumptions.

Assumption 3.2 $\max_{1\le t\le n}\left|n^{-1/2}\sum_{s=1}^{t}(\hat{\varepsilon}_{s,n}^2-\varepsilon_s^2)\right|=\mathrm{o}_{\mathsf{P}}(1)$.

Assumption 3.3 $n^{-1}\sum_{t=1}^{n}(\hat{\varepsilon}_t^4-\varepsilon_t^4)=\mathrm{o}_{\mathsf{P}}(1)$.

We will later show that Assumptions 3.2 and 3.3 are satisfied in a wide range of situations. Under these assumptions we then have the following result.

Theorem 3.2 *If Assumptions 3.1, 3.2, 3.3 are satisfied then $CUSQ_n\xrightarrow{\mathsf{D}}\sup_{0\le u\le 1}\left|\mathcal{B}_u^0\right|$.*

For the recursive version of the result we need to strengthen Assumption 3.2.

Assumption 3.4 $\max_{1\le t\le n}\left|n^{-1/2}\sum_{s=1}^{t}(\hat{\varepsilon}_{s,t}^2-\varepsilon_s^2)\right|=\mathrm{o}_{\mathsf{P}}(1)$.

Theorem 3.3 *If Assumptions 3.1, 3.3, 3.4 are satisfied then $RCUSQ_n\xrightarrow{\mathsf{D}}\sup_{0\le u\le 1}\left|\mathcal{B}_u^0\right|$.*

For a linear model it is possible to analyze Assumptions 3.2, 3.3 and 3.4 directly. In this way Nielsen and Sohkanen (2011) consider the case of the linear autoregressive distributed lag model with non-stationary (possibly explosive) regressors. Their Lemma 4.2 and Theorem 4.5 show that Assumptions 3.3 and 3.4 are satisfied under the martingale difference Assumption 3.1. For non-linear models it is useful to formulate a set of intermediate level assumptions that imply Assumptions 3.2, 3.3 and 3.4. We do this in the following.

4 Intermediate Level Results

In the non-linear regression model (2.1) we can replace the high level Assumptions 3.2 and 3.3 by local consistency of $\hat{\theta}_n$ and smoothness of the criterion function.

Assumption 4.1 *Let $\delta<1/4$. Suppose $N_{n,\theta_0}^{-1}(\hat{\theta}_n-\theta_0)$ is either (a) $\mathrm{o}_{\mathsf{P}}(n^{\delta})$ or (b) $\mathrm{o}(n^{\delta})$ a.s.*

The normalization N_{n,θ_0}^{-1} allows both stationary and non-stationary regressors. In linear models $N_{n,\theta_0}^{-1}=n^{1/2}$ for stationary regressors and $N_{n,\theta_0}^{-1}=n$ for random walk regressors. In more general cointegrated models N_{n,θ_0}^{-1} may be block diagonal with different normalizations in different blocks, see Kristensen and Rahbek (2010). In non-linear models the normalization may depend on the parameter θ under which we evaluate the distributions. We use the notation θ_0 to emphasize this choice of parameter.

The following smoothness assumption involves normalized sums of the first two derivatives of the known function g with respect to θ. These are the q-vector $\dot{g}(x_t,\theta)=\partial g(x_t,\theta)/\partial\theta$ and the $q\times q$ square matrix $\ddot{g}(x_t,\theta)=\partial g(x_t,\theta)/\partial\theta\partial\theta'$. We will need a matrix norm. In the proof we use the spectral norm, but at this point any equivalent matrix norm can be used.

Assumption 4.2 *Suppose x_t is $\mathcal{F}_{t-1}$-measurable and $g(x_t,\theta)$ is twice differentiable with respect to θ. Let $\delta < 1/4$ be the consistency rate in Assumption 4.1 and let $\epsilon > 0$. Suppose*
(a) $\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq n^{\delta}\epsilon} \sum_{t=1}^{n}\{g(x_t,\theta)-g(x_t,\theta_0)\}^2 = \mathrm{o}_{\mathsf{P}}(n^{1/2})$;
(b) $\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq n^{\delta}\epsilon} \sum_{t=1}^{n}\{g(x_t,\theta)-g(x_t,\theta_0)\}^4 = \mathrm{o}_{\mathsf{P}}(n)$;
(c) $\sum_{t=1}^{n} ||N'_{n,\theta_0}\dot{g}(x_t,\theta_0)||^2 = \mathrm{O}_{\mathsf{P}}(n^{1-2\delta-\eta})$ *for some* $\eta > 0$;
(d) $\sum_{t=1}^{n} ||N'_{n,\theta_0}\ddot{g}(x_t,\theta_0)N_{n,\theta_0}||^2 = \mathrm{O}_{\mathsf{P}}(n^{1-4\delta-\eta})$ *for some* $\eta > 0$;
(e) $\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq n^{\delta}\epsilon} \sum_{t=1}^{n} ||N'_{n,\theta_0}\left\{\ddot{g}(x_t,\theta)-\ddot{g}(x_t,\theta_0)\right\}N_{n,\theta_0}||^2 = \mathrm{O}_{\mathsf{P}}(n^{-4\delta})$.

Finally, we need some technical conditions to ensure invertibility of certain matrices.

Assumption 4.3 *Suppose* $\inf[n : \sum_{t=1}^{n} w_t w_t'$ *is invertible*$] < \infty$ *a.s. for* $w_t = \dot{g}(x_t,\theta_0)$ *and* $w_t = \mathrm{vec}\{\ddot{g}(x_t,\theta_0)\}$ *with the convention that the empty set has infinite infimum. Moreover, suppose* $N_{n,\theta_0}^{-1} = \mathrm{O}(n^{\ell})$ *for some* $\ell > 0$.

We can now show that Assumptions 3.2, 3.3, 3.4 are satisfied. Subsequently, we return to a discussion of the assumptions.

Theorem 4.1 *Assumptions 3.1,4.1(a),4.2,4.3 imply Assumptions 3.2,3.3 so Theorem 3.2 applies.*

For the recursive cumulated sum of squares statistic we require strong uniformity properties. If the estimator is strongly consistent we can get that uniformity from Egorov's Theorem, see Davidson (1994, Theorem 18.4).

Theorem 4.2 *Assumptions 3.1,4.1(b),4.2,4.3 imply Assumptions 3.3,3.4 so Theorem 3.3 applies.*

In the proof we analyze $n^{-1/2}\sum_{s=1}^{n}(\hat{\varepsilon}_{s,n}^2-\varepsilon_s^2)$ through a martingale decomposition. Noting that $\hat{\varepsilon}_{s,n}-\varepsilon_s = \nabla g(x_s,\hat{\theta}_n) = g(x_s,\hat{\theta}_n)-g(x_s,\theta_0)$ and expanding $(\varepsilon-\nabla)^2-\varepsilon^2 = -2\varepsilon\nabla+\nabla^2$ we get

$$n^{-1/2}\sum_{s=1}^{n}(\hat{\varepsilon}_{s,n}^2-\varepsilon_s^2) = -2n^{-1/2}\sum_{s=1}^{n}\varepsilon_s\nabla g(x_s,\hat{\theta}_n)+n^{-1/2}\sum_{s=1}^{n}\{\nabla g(x_s,\hat{\theta}_n)\}^2. \tag{4.1}$$

Due to Assumption 4.1 the estimator $\hat{\theta}_n$ varies in a local region around θ_0. Thus, it suffices to replace $\hat{\theta}_n$ with a deterministic value θ and show that the sums in (4.1) vanish uniformly over the local region. These sums are a martingale and its compensator. Now, the compensator vanishes under Assumption 4.2(a). Jennrich (1969, Theorem 6) uses a similar condition when proving consistency of non-linear least squares, with the difference that he takes supremum over a non-vanishing set. In the proof the main bulk of the work is to show that the martingale part vanishes under Assumption 4.2$(c)-(e)$. For this we exploit Lemma 1 of Lai and Wei (1982). The conditions $(c)-(e)$ are somewhat weaker than the usual conditions for deriving the asymptotic distribution of non-linear least squares estimators, see for instance Amemiya (1985, page 111). Finally, Assumption 4.2(b) is used for showing the consistency of the fourth moment estimator $\hat{\varphi}_n^2$.

In many applications the non-linear function g and its derivatives satisfy a Lipschitz condition. In that case one can easily relate condition (a) of Assumption 4.2 to conditions $(c)-(e)$. To do this, one just needs to second order Taylor expand $g(x_t,\theta)-g(x_t,\theta_0)$ around θ_0, square it, and take supremum before cumulating. A similar argument applies to condition (b). This gives a somewhat shorter set of assumptions that imply Assumption 4.2.

Assumption 4.4 *Suppose x_t is $\mathcal{F}_{t-1}$-measurable and $g(x_t,\theta)$ is twice differentiable with respect to θ. Let $\delta < 1/4$ be the consistency rate in Assumption 4.1 and let $\epsilon > 0$. Suppose, the following conditions hold, for $k = 2, 4$,*
(a) $\sum_{t=1}^{n} ||N'_{n,\theta_0}\dot{g}(x_t,\theta_0)||^k = \mathrm{o}_{\mathsf{P}}(n^{k/4-k\delta})$;
(b) $\sum_{t=1}^{n} ||N'_{n,\theta_0}\ddot{g}(x_t,\theta_0)N_{n,\theta_0}||^k = \mathrm{o}_{\mathsf{P}}(n^{k/4-2k\delta})$;
(c) $\sum_{t=1}^{n}\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq n^{\delta}\epsilon} ||N'_{n,\theta_0}\left\{\ddot{g}(x_t,\theta)-\ddot{g}(x_t,\theta_0)\right\}N_{n,\theta_0}||^k = \mathrm{o}_{\mathsf{P}}(n^{-4\delta})$.

Theorem 4.3 *Assumption 4.4 implies Assumption 4.2.*

5 Analysis of some particular models

In this section, we consider some particular non-linear models that have been discussed in the literature. For these models it is relevant to test their validity using a cumulated sum of squares test. We will assume that the consistency Assumption 4.1 has been dealt elsewhere. Thus, we know the appropriate normalization of the estimator. The difficulty is therefore to establish the smoothness Assumption 4.4. We will show that this assumption is rather mild.

5.1 The linear model

In the linear model $g(x_t, \theta) = \theta' x_t$ so that $\dot{g}(x_t, \theta) = x_t$ and $\ddot{g}(x_t, \theta) = 0$. Thus, Assumption 4.4 reduces to showing $\sum_{t=1}^{n} ||N'_{n,\theta_0} x_t||^k = \mathrm{o_P}(n^{k/4-k\delta})$ for $k = 2, 4$. Suppose x_t is univariate and stationary then $N_{n,\theta_0} = n^{-1/2}$ whereas $N_{n,\theta_0} = n^{-1}$ if x_t is a random walk. In both cases $N_{n,\theta_0}^k \sum_{t=1}^{n} |x_t|^k = \mathrm{O_P}(1) = \mathrm{o_P}(n^{k/4-k\delta})$.

For the recursive statistic we would need to establish that $\hat{\theta}_n$ is strongly consistent. For non-stationary models this is not always so easy. To our knowledge this has not been proved for a first order autoregressive model with an intercept and where the autoregressive coefficient is unity. Nielsen and Sohkanen (2011) therefore work directly with the high level Assumption 3.4.

5.2 The power function model

As a first non-linear case we consider the power function $g(x_t, \theta) = |x_t|^\theta$ to illustrate where the difficulties lie in the arguments. The model equation is then

$$y_t = |x_t|^\theta + \varepsilon_t \qquad t = 1, \ldots, n, \tag{5.1}$$

with $\theta > 0$ and where x_t is either stationary or a random walk. We will suppose that Assumption 4.1 is satisfied and show that Assumption 4.2 holds.

The properties of the regressor x_t are reflected in the choice of the normalization N_{n,θ_0}. Hence, if x_t is stationary with finite $|x_t|^{4\theta_0} \log^8 |x_t|$ moments we let $N_{n,\theta_0}^{-1} = \sqrt{n}$ and apply techniques from Wooldridge (1994). If x_t is a random walk we let $N_{n,\theta_0}^{-1} = n^{(1+\theta_0)/2} \log n^{1/2}$ and apply techniques from Park and Phillips (2001). These techniques go back to Cramér and involve smoothness conditions that are similar but also somewhat stronger than Assumption 4.2. Here, we take $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta) = \mathrm{O_P}(1)$ as given. Hence Assumption 4.1 follows for any $\delta > 0$.

To prove Assumption 4.4 we differentiate g and get

$$g(x, \theta) = |x|^\theta, \qquad \dot{g}(x, \theta) = |x|^\theta \log |x|, \qquad \ddot{g}(x, \theta) = |x|^\theta \log^2 |x|.$$

These functions are continuous in x when $\theta > 0$ and $|x| > 0$ and they can be extended continuously to all $x \in \mathbb{R}$ because the power function dominates the logarithm at the origin.

We now look at Assumption 4.4 (a) in some detail. As in the linear case we show

$$\mathcal{S} = \sum_{t=1}^{n} |N_{n,\theta_0} \dot{g}(x_t, \theta_0)|^k = \mathrm{O_P}(1) = \mathrm{o_P}(n^{k/4-k\delta}).$$

In the stationary case we use Theorems 3.5.3, 3.5.7 of Stout (1974) to get

$$\mathcal{S} = \frac{1}{n^{k/2}} \sum_{t=1}^{n} |x_t|^{k\theta_0} \log^k |x_t| = \mathrm{O}(n^{1-k/2}) = \mathrm{O}(1) \qquad a.s.$$

In the random walk case we get

$$\mathcal{S} = \frac{1}{n^{(1+\theta_0)k/2} \log^k n^{1/2}} \sum_{t=1}^{n} |x_t|^{k\theta_0} \log^k |x_t| = \frac{1}{n^{k/2}} \sum_{t=1}^{n} |x_t/n^{1/2}|^{k\theta_0} \left(\frac{\log |x_t/n^{1/2}|}{\log n^{1/2}} + 1 \right)^k = \mathrm{O_P}(1),$$

where the second equality follows by noting that $\log|x| = \log|x/n^{1/2}| + \log n^{1/2}$. For the last bound note that $x_{\text{integer}(nu)}/n^{1/2}$ converges to a Brownian motion as a function on $D[0,1]$. The functions $|y|^{2\theta_0}\log|y|$ and $|y|^{2\theta_0}$ are continuous and therefore the integrals $\int_0^1 |y|^{2\theta_0}\log|y|dy$ and $\int_0^1 |y|^{2\theta_0}dy$ are continuous mappings from $D[0,1]$ to $\mathbb{R}$. The Continuous Mapping Theorem, see Billingsley (1999, Theorem 2.7) then shows that the normalized sum converges in distribution.

For Assumption 4.4 (b) a similar argument shows $\sum_{t=1}^n |N_{n,\theta_0}^2 \ddot{g}(x_t,\theta_0)|^k = \mathrm{O_P}(n^{1-k})$.

For Assumption 4.4 (c) we apply a Lipschitz argument. The second derivative of g satisfies

$$|\ddot{g}(x_t,\theta) - \ddot{g}(x_t,\theta_0)| \leq (|x|^\nu + |x|^{-\nu})|x|^{\theta_0}\log^2|x|$$

for all θ so $|\theta-\theta_0| \leq \nu$ for some $0 < \nu < \theta_0$. The result is proved by analyzing the function $|x|^{\theta-\theta_0} - 1$ for all four sign combinations of $|x|-1$ and $\theta-\theta_0$. Applying this to condition (c) gives

$$\sum_{t=1}^n \sup_{\theta:|\theta-\theta_0|\leq\nu} |N_{n,\theta_0}^2\{\ddot{g}(x_t,\theta) - \ddot{g}(x_t,\theta_0)\}|^k$$
$$\leq \sum_{t=1}^n \{N_{n,\theta_0}^2(|x_t|^\nu + |x_t|^{-\nu})|x_t|^{\theta_0}\log^2|x_t|\}^k = \mathrm{O_P}(n^{1-k}) = \mathrm{o_P}(1), \quad (5.2)$$

where the second bound follows by the same argument as above.

5.3 Cointegration with non-linear error correction

In the model of Kristensen and Rahbek (2010) x_t is a p-dimensional time series satisfying

$$\Delta x_t = g(\beta' x_{t-1},\gamma) + \Phi_1\Delta x_{t-1} + \cdots + \Phi_k\Delta x_{t-k} + \varepsilon_t.$$

In specification analysis we consider the coordinates of the residual vector $\hat{\varepsilon}_t$ separately. Their Theorem 1 gives conditions ensuring that $\beta' x_{t-1}, \Delta x_{t-1}, \ldots, \Delta x_{t-k}$ are geometrically ergodic and that x_t satisfies a Granger–Johansen-type representation. With this and some further conditions their Theorem 5 provides the normalization $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{O_P}(1)$ that is required in our Assumption 4.1. Their Assumption A.5 requires that the first, second and third derivatives of $g(z,\gamma)$ with respect to z or γ are of order $\mathrm{O}(|z|)$. With these boundedness conditions our Assumption 4.4 can be proved. The proof is slightly involved as one will have to keep track of the various components in the Granger-Johansen-type representation and how they interact with the derivatives of g.

5.4 Non-linear models with random walk regressors

Park and Phillips (2001) consider a triangular system with a univariate random walk regressor:

$$y_t = g(x_t,\theta) + \varepsilon_t \qquad t = 1,\ldots,n, \quad (5.3)$$
$$x_t = x_{t-1} + v_t, \quad (5.4)$$

where ε_t is an $\mathcal{F}_t$-martingale difference sequence, $(\varepsilon_t, v_t)'$ satisfies a functional central limit theorem, x_t is $\mathcal{F}_{t-1}$-adapted, and g is in one of two main classes of functions: integrable and asymptotically homogeneous. For recent developments see Chan and Wang (2015).

The class of integrable functions includes transformations $g(x_t,\theta)$ such as $1/(1+\theta x^2)$, $e^{-\theta x^2}$, or $\theta 1(0 \leq x \leq 1)$ which are integrable over $x \in \mathbb{R}$ and satisfy a Lipschitz condition over θ. In Theorem 5.1 of Park and Phillips (2001) the asymptotic distribution of the non-linear least squares estimator for the integrable functions case is derived, showing that $n^{1/4}(\hat{\theta}_n - \theta)$ converges in distribution. Thus, we can choose $N_{n,\theta_0}^{-1} = n^{1/4}$ and otherwise proceed as in the power function example.

The class of asymptotically homogenous functions includes transformations $g(x,\theta)$ which asymptotically behave like homogeneous functions; they include the power function in Section 5.2

as well as logistic, threshold-like or logarithmic transformations. Specifically, an asymptotically homogeneous function $f(x,\theta)$ is a function such that

$$f(\lambda x,\theta)=\kappa(\lambda,\theta)H(x,\theta)+R(\lambda,x,\theta),$$

where κ is a normalization, H satisfies some regularity conditions (such as local integrability – see also Pötscher, 2004) and R is a lower order remainder term. In Theorem 5.3 of Park and Phillips (2001) each of the functions $g,\dot{g}$ and $\ddot{g}$ are assumed to be asymptotically homogenous and satisfy conditions that have the same flavour as those in Assumption 4.4. It then follows that $n^{1/2}\dot{\kappa}\left(\sqrt{n},\theta_0\right)'\left(\hat{\theta}_n-\theta\right)$ converges in distribution. Thus, we can choose the normalization $N_{n,\theta_0}^{-1}=n^{1/2}\dot{\kappa}\left(\sqrt{n},\theta_0\right)'$. For instance, in the power function model (5.1) with random walk regressor we have $\dot{\kappa}\left(\sqrt{n},\theta_0\right)=n^{\theta_0/2}\log n^{1/2}$.

6 Finite Sample Performance

In this section, we study the finite sample performance of the *CUSQ* test through simulation. We use the exact asymptotic 95% critical value of 1.36 and 10000 replicas. Two sets of results are presented for various asymptotically homogeneous models. First, we check size and power for a set of models that are either linear or non-linear in parameters. Next, we consider a set of models suggested by Kasparis (2008). For these we compare the power of the *CUSQ* test with the power of a cumulated sum (CUSUM) test reported by Kasparis (2008). We find that the two tests have power of similar magnitude, so there is no apparent advantage in using the more complicated CUSUM test.

Table 1 contains the first set of data generating processes (DGPs). Four correctly specified (CS) DGPs and five misspecified (M) DGPs are analyzed. The regressor x_t is (fractionally) integrated so that $\Delta^{\tau}x_t$ is iid $\mathsf{N}(0,1)$ with $x_t=0$ for $t\leq 0$ and with $\tau=0.7,\ 1,\ 2$. While the models in Section 5 focus on stationary and random walk models the theory does extend to other types of nonstationarity, see Chan and Wang (2015).

Table 2, DGPs 1-4, reports the size of the *CUSQ* test. The size control is fairly uniform across the DGPs. This is in correspondence with the results for linear autoregressions in Nielsen and Sohkanen (2011). The test is, however, slightly undersized in small samples. The size distortion can be removed by applying the finite sample 95% critical value $1.36-0.67n^{-1/2}-0.89n^{-1}$ suggested by Edgerton and Wells (1994). Similarly, for the recursive test Sohkanen (2011) suggests the 95% critical value $1.36(1-0.68n^{-1/2}+3.13n^{-1}-33.9n^{-3/2}+93.9n^{-2})$.

Table 2, DGPs 5-9, reports the power of the *CUSQ* test for a range of asymptotically homogenous functions. The power increases with sample size in all cases. The power also tends to increase with the order of integration of the regressors. This is in line with the power analysis for linear models conducted by McCabe and Harrison (1980), Ploberger and Krämer (1990), Deng and Perron (2008), or Turner (2010).

The *CUSQ* also has power to detect misspecification involving integrable functions of persistent processes. As an example consider the data generating process $y_t=\theta_1/(1+\theta_2x_t^2)+\varepsilon_t$, while the regression model is polynomial. Simulations not reported here show that power arises as long as the signal from the integrable function component $\theta_1/(1+\theta_2x_t^2)$ dominates the noise ε_t.

Next, we compare the power of the *CUSQ* test with the CUSUM test of Kasparis (2008). Table 3 reports his ten DGPs. In all cases a linear model for y_t and x_t is fitted, which is therefore misspecified. The results are reported in Table 4. Kasparis' test uses a long run variance estimator to standardize the statistic; hence, the power of the test depends on a bandwidth choice. Kasparis reports power for different bandwidths and we report the highest of these. Table 4 shows that no test dominates in all cases but both tests perform in a similar way. We note that the CUSUM test involves nuisance terms depending on the functional form of the model whereas the *CUSQ* has a Brownian bridge theory quite generally.

A Appendix: Proofs

In most places we use the spectral norm for matrices, so that for a matrix m then

$$\|m\| = \sqrt{\max \operatorname{eigen}(m'm)}.$$

The spectral norm reduces to the Euclidean norm for vectors. It is compatible with the Euclidean norm in the sense that $||mv|| = ||m||||v||$ for a matrix m and a vector v. It satisfies the norm inequality $||mn|| \leq ||m||||n||$ for matrices m, n. Occasionally, we will use the Frobenius norm

$$||m||_F = (\textstyle\sum_{i,j} |m_{ij}|^2)^{1/2}.$$

Note that $||m|| \leq ||m||_F$ with equality when m is a vector, while $||m||_F \leq q||m||$ where q is the column dimension of m. Further, $||m||_F = ||vec(m)||_F$.

We start with a modification of the martingale result by Lai and Wei (1982).

Lemma A.1 *Let N_{n,θ_0} be a $q \times q$ normalizing matrix where $N_{n,\theta_0}^{-1} = \mathrm{O}(n^\ell)$ for some $\ell > 0$. Further, let $g(x_t, \theta_0)$ be a function $g : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$, with derivatives with respect to θ: $\dot{g}$, $\ddot{g}$. Also let Assumption 3.1(a) hold. Let w_t be $\mathcal{F}_{t-1}$-measurable and given as either of*
(i) $w_t = N'_{n,\theta_0}\dot{g}(x_t, \theta_0)$;
(ii) $w_t = N'_{n,\theta_0}\ddot{g}(x_t, \theta_0)N_{n,\theta_0}$.
Suppose $n_0 = \inf[n : \sum_{t=1}^n \{\mathrm{vec}(w_t)\}\{\mathrm{vec}(w_t)\}'$ is invertible$] < \infty$ a.s. Then, for all $\varsigma > 0$,

$$\begin{aligned}\max_{n_0 \leq s \leq n} \|\textstyle\sum_{t=1}^s w_t\varepsilon_t\| &\overset{a.s.}{=} \mathrm{o}\left[n^\varsigma \left\|\sum_{t=1}^n \{\mathrm{vec}(w_t)\}\{\mathrm{vec}(w_t)\}'\right\|^{1/2+\varsigma}\right] + \mathrm{O}(1) \\ &= \mathrm{o}\left\{n^\varsigma \left(\textstyle\sum_{t=1}^n \|w_t\|^2\right)^{1/2+\varsigma}\right\} + \mathrm{O}(1).\end{aligned}$$

Proof of Lemma A.1: Part (i): Introduce the notation

$$S_{\dot{g}\varepsilon,u} = \textstyle\sum_{t=1}^{[nu]} \dot{g}(x_t, \theta_0)\varepsilon_t \qquad \text{and} \qquad S_{\dot{g}\dot{g},u} = \sum_{t=1}^{[nu]} \dot{g}(x_t, \theta_0)\dot{g}(x_t, \theta_0)',$$

so that

$$N'_{n,\theta_0}S_{\dot{g}\varepsilon,u} = \textstyle\sum_{t=1}^{[nu]} w_t\varepsilon_t \qquad \text{and} \qquad N'_{n,\theta_0}S_{\dot{g}\dot{g},u}N_{n,\theta_0} = \sum_{t=1}^{[nu]} w_t w_t'.$$

Notice that, for $n_0 < [nu]$,

$$\left\|N'_{n,\theta_0}S_{\dot{g}\varepsilon,u}\right\| = \left\|(N'_{n,\theta_0}S_{\dot{g}\dot{g},u}N_{n,\theta_0})^{1/2}S_{\dot{g}\dot{g},u}^{-1/2}S_{\dot{g}\varepsilon,u}\right\| \leq \left\|N'_{n,\theta_0}S_{\dot{g}\dot{g},u}N_{n,\theta_0}\right\|^{1/2}\left\|S_{\dot{g}\dot{g},u}^{-1/2}S_{\dot{g}\varepsilon,u}\right\|. \tag{A.1}$$

Use Lai and Wei (1982, Lemma 1,i,ii) with Assumption 3.1(a) recalling the definition of the spectral norm, to see that

$$\left\|S_{\dot{g}\dot{g},u}^{-1/2}S_{\dot{g}\varepsilon,u}\right\| = \left\|S_{\varepsilon\dot{g},u}S_{\dot{g}\dot{g},u}^{-1}S_{\dot{g}\varepsilon,u}\right\|^{1/2} \overset{a.s.}{=} \mathrm{o}\left(\|S_{\dot{g}\dot{g},u}\|^{\tilde{\varsigma}}\right) + \mathrm{O}(1),$$

for all $\tilde{\varsigma} > 0$. Since $\|S_{\dot{g}\dot{g},u}\|$ is non-decreasing in u this is bounded by $\|S_{\dot{g}\dot{g},1}\|$. Using that $N_{n,\theta_0}^{-1} = \mathrm{O}\left(n^\ell\right)$ for some $\ell > 0$, we can write

$$\left\|S_{\dot{g}\dot{g},u}^{-1/2}S_{\dot{g}\varepsilon,u}\right\| \overset{a.s.}{=} \mathrm{o}\left(\left\|N_{n,\theta_0}^{-1}\right\|^{2\tilde{\varsigma}}\left\|N'_{n,\theta_0}S_{\dot{g}\dot{g},1}N_{n,\theta_0}\right\|^{\tilde{\varsigma}}\right) + \mathrm{O}(1) = \mathrm{o}\left(n^\varsigma\left\|N'_{n,\theta_0}S_{\dot{g}\dot{g},1}N_{n,\theta_0}\right\|^\varsigma\right) + \mathrm{O}(1),$$

for all $\varsigma > 0$, uniformly in u. Hence, using (A.1),

$$\sup_u \left\|N'_{n,\theta_0}S_{\dot{g}\varepsilon,u}\right\| \overset{a.s.}{=} \mathrm{o}(n^\varsigma\left\|N'_{n,\theta_0}S_{\dot{g}\dot{g},1}N_{n,\theta_0}\right\|^{1/2+\varsigma}) + \mathrm{O}\left(1\right),$$

which is the first desired expression and by the triangle inequality we get the second expression.

Part (ii): By the properties of the Frobenius norm we get that

$$\left\|\sum_{t=1}^{[nu]} w_t \varepsilon_t\right\| \leq \left\|\sum_{t=1}^{[nu]} w_t \varepsilon_t\right\|_F = \left\|\sum_{t=1}^{[nu]} \mathrm{vec}\left(w_t\right) \varepsilon_t\right\|.$$

Now argue as in (i) with w_t replaced by $\mathrm{vec}\left(w_t\right)$ to get

$$\sup_u \left\|\sum_{t=1}^{[nu]} w_t \varepsilon_t\right\| \leq \mathrm{o}\left(n^{\varsigma} \left\|\sum_{t=1}^{n} \left\{\mathrm{vec}\left(w_t\right)\right\} \left\{\mathrm{vec}\left(w_t\right)\right\}'\right\|^{1/2+\varsigma}\right) \qquad a.s.,$$

which is the first desired expression. To get the second expression notice that

$$\left\|\sum_{t=1}^{n} \left\{\mathrm{vec}\left(w_t\right)\right\} \left\{\mathrm{vec}\left(w_t\right)\right\}'\right\| \leq \sum_{t=1}^{n} \left\|\left\{\mathrm{vec}\left(w_t\right)\right\} \left\{\mathrm{vec}\left(w_t\right)\right\}'\right\|$$

and $||\{\mathrm{vec}(w_t)\}\{\mathrm{vec}(w_t)\}'|| = \|\mathrm{vec}\left(w_t\right)\|_F^2 = \|w_t\|_F^2 \leq q\left\|w_t\right\|^2$ as desired. □

Proof of Theorem 3.2: The statistic of interest is $CUSQ_n = \mathcal{A}_n/\hat{\varphi}_n$, where

$$\mathcal{A}_n = \max_{1\leq t\leq n} \left|n^{-1/2}\sum_{s=1}^{t} \left(\hat{\varepsilon}_{s,n}^2 - n^{-1}\sum_{r=1}^{n}\hat{\varepsilon}_{r,n}^2\right)\right|.$$

Expand $\mathcal{A}_n = \mathcal{B}_n + \left(\mathcal{A}_n - \mathcal{B}_n\right)$, where

$$\mathcal{B}_n = \max_{1\leq t\leq n} \left|n^{-1/2}\sum_{s=1}^{t} \left(\varepsilon_s^2 - n^{-1}\sum_{r=1}^{n}\varepsilon_r^2\right)\right|.$$

Noting that $\hat{\varepsilon}_{s,n}^2 = (\hat{\varepsilon}_{s,n}^2 - \varepsilon_s^2) + \varepsilon_s^2$ we get

$$\mathcal{A}_n - \mathcal{B}_n = \max_{1\leq t\leq n} \left|n^{-1/2}\sum_{s=1}^{t} \left(\varepsilon_s^2 - n^{-1}\sum_{r=1}^{n}\varepsilon_r^2\right) + n^{-1/2}\sum_{s=1}^{t} \left\{\left(\hat{\varepsilon}_{s,n}^2 - \varepsilon_s^2\right) - n^{-1}\sum_{r=1}^{n} \left(\hat{\varepsilon}_{r,n}^2 - \varepsilon_r^2\right)\right\}\right|$$
$$- \max_{1\leq t\leq n} \left|n^{-1/2}\sum_{s=1}^{t} \left(\varepsilon_s^2 - n^{-1}\sum_{r=1}^{n}\varepsilon_r^2\right)\right|.$$

By the triangle inequality $\mathcal{A}_n - \mathcal{B}_n \leq \mathcal{B}_n + \mathcal{C}_n - \mathcal{B}_n = \mathcal{C}_n$ where

$$\mathcal{C}_n = \max_{1\leq t\leq n} \left|n^{-1/2}\sum_{s=1}^{t} \left\{\left(\hat{\varepsilon}_{s,n}^2 - \varepsilon_s^2\right) - n^{-1}\sum_{r=1}^{n} \left(\hat{\varepsilon}_{r,n}^2 - \varepsilon_r^2\right)\right\}\right|.$$

By Assumption 3.2,

$$\mathcal{A}_n - \mathcal{B}_n \leq \mathcal{C}_n \leq 2 \max_{1\leq t\leq n} \left|n^{-1/2}\sum_{s=1}^{t} \left(\hat{\varepsilon}_{s,n}^2 - \varepsilon_s^2\right)\right| = \mathrm{o}_{\mathsf{P}}(1). \tag{A.2}$$

Thus, by Theorem 3.1 and the Continuous Mapping Theorem applied to the maximum, we have

$$\mathcal{A}_n = \mathcal{B}_n + \mathrm{o}_{\mathsf{P}}(1) \xrightarrow{\mathsf{D}} \varphi \sup_{0\leq u\leq 1} \left|\mathcal{B}_u^0\right|.$$

Consider now $\hat{\varphi}_n^2 = n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{t,n}^4 - (n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{t,n}^2)^2$. Further, $n^{-1}\sum_{t=1}^{n}(\hat{\varepsilon}_{t,n}^k - \varepsilon_t^k) = \mathrm{o}_{\mathsf{P}}(1)$ for $k = 2, 4$ by Assumptions 3.2, 3.3. Therefore,

$$\hat{\varphi}_n^2 = n^{-1}\sum_{t=1}^{n}\varepsilon_t^4 - \left(n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\right)^2 + \mathrm{o}_{\mathsf{P}}(1).$$

By Theorem 3.1, under Assumption 3.1, we have $\hat{\varphi}_n^2 = \varphi^2 + \mathrm{o}_{\mathsf{P}}(1)$.

All together, $CUSQ_n$ converges in distribution to $\sup_{0\leq u\leq 1}\left|\mathcal{B}_u^0\right|$ as desired. □

Proof of Theorem 3.3: Follow the proof of Theorem 3.2 replacing $\hat{\varepsilon}_{s,n}^2$ by $\hat{\varepsilon}_{s,t}^2$ and using Assumption 3.4 instead of Assumption 3.2 when evaluating (A.2). □

Proof of Theorem 4.1: Part I: Assumption 3.2.

1. *The problem.* Let $\mathcal{S}_{t,\theta} = n^{-1/2}\{Q_t(\theta) - Q_t(\theta_0)\}$ so that $\mathcal{S}_{t,\hat\theta_n} = n^{-1/2}\sum_{s=1}^t \hat\varepsilon_{s,n}^2 - \varepsilon_s^2$. We show that $\mathcal{S}_{t,\hat\theta_n} = \mathrm{o_P}(1)$ uniformly in $1 \le t \le n$. From (4.1) we have $\mathcal{S}_{t,\theta} = -2\widetilde{\mathcal{S}}_{t,\theta} + \overline{\mathcal{S}}_{t,\theta}$, where

$$\widetilde{\mathcal{S}}_{t,\theta} = n^{-1/2}\textstyle\sum_{s=1}^t \varepsilon_s \nabla g_s(\theta), \qquad \overline{\mathcal{S}}_{t,\theta} = n^{-1/2}\textstyle\sum_{s=1}^t \{\nabla g_s(\theta)\}^2.$$

2. *Expand the martingale* $\widetilde{\mathcal{S}}_{t,\theta}$. We use a second order mean value result. To simplify the expression we introduce the notation

$$\begin{aligned} \dot h_s(\theta) &= N'_{n,\theta_0}\dot g(x_s,\theta), & \ddot h_s(\theta) &= N'_{n,\theta_0}\ddot g(x_s,\theta)N_{n,\theta_0}, \\ \vartheta &= N_{n,\theta_0}^{-1}(\theta-\theta_0), & \nabla\ddot h_s(\theta) &= N'_{n,\theta_0}\{\ddot g(x_s,\theta) - \ddot g(x_s,\theta_0)\}N_{n,\theta_0}. \end{aligned}$$

With this notation we get, for instance, that

$$(\theta-\theta_0)'\dot g(x_s,\theta_0) = \{N_{n,\theta_0}^{-1}(\theta-\theta_0)\}'N'_{n,\theta_0}\dot g(x_s,\theta_0) = \vartheta'\dot h_s(\theta_0).$$

Overall, we can expand $\widetilde{\mathcal{S}}_{t,\hat\theta_n} = n^{-1/2}\sum_{s=1}^t \varepsilon_s \nabla g_s(\hat\theta_n)$ as

$$\begin{aligned} \widetilde{\mathcal{S}}_{t,\hat\theta_n} = n^{-1/2}\textstyle\sum_{s=1}^t \varepsilon_s \hat\vartheta_n'\dot h_s(\theta_0) + \frac{1}{2}n^{-1/2}\textstyle\sum_{s=1}^t \varepsilon_s\hat\vartheta_n'\ddot h_s(\theta_0)\hat\vartheta_n \\ + \frac{1}{2}n^{-1/2}\textstyle\sum_{s=1}^t \varepsilon_s\hat\vartheta_n'\{\ddot h_s(\theta_*) - \ddot h_s(\theta_0)\}\hat\vartheta_n, \end{aligned} \tag{A.3}$$

for an intermediate point θ_* depending on the summation limit t and $\hat\theta_n$ so $||\theta_* - \theta_0|| \le ||\hat\theta_n - \theta_0||$. Note that the first two terms only depend on $\hat\theta_n$ through the factor $\hat\vartheta_n$. For simplicity we write (A.3) as $\widetilde{\mathcal{S}}_{t,\hat\theta_n} = \widetilde{\mathcal{S}}_{t,1} + (\widetilde{\mathcal{S}}_{t,2} + \widetilde{\mathcal{S}}_{t,\hat\theta_n,3})/2$.

3. *The martingale term* $\widetilde{\mathcal{S}}_{t,1}$. The norm inequality and the bound to $\hat\vartheta_n$ in Assumption 4.1 (a) give

$$|\widetilde{\mathcal{S}}_{t,1}| \le n^{-1/2}||\hat\vartheta_n||||\textstyle\sum_{s=1}^t \varepsilon_s\dot h_s(\theta_0)|| \le \mathrm{o_P}(n^{\delta-1/2})||\textstyle\sum_{s=1}^t \varepsilon_s\dot h_s(\theta_0)||.$$

Apply Lemma A.1 (i) using Assumptions 3.1, 4.3 to get, for any $\varsigma > 0$,

$$\max_{n_0\le t\le n}|\widetilde{\mathcal{S}}_{t,1}| = \mathrm{o_P}(n^{\delta-1/2})\mathrm{o}_{a.s.}[n^\varsigma\{\textstyle\sum_{t=1}^n||\dot h_t(\theta_0)||^2\}^{1/2+\varsigma}] + \mathrm{o_P}(n^{\delta-1/2})\mathrm{O}_{a.s.}(1).$$

By Assumption 4.2 (c), we have that $\sum_{t=1}^n ||\dot h_t(\theta_0)||^2 = \mathrm{O_P}\left(n^{1-2\delta-\eta}\right)$ for some $\eta > 0$ while $\delta < 1/4$. We then get, when choosing $2\varsigma \le \eta/(2-2\delta-\eta)$,

$$\max_{n_0\le t\le n}|\widetilde{\mathcal{S}}_{t,1}| = \mathrm{o_P}(n^{\delta-1/2})\mathrm{o_P}\{n^{(1-2\delta-\eta)(1/2+\varsigma)+\varsigma}\} + \mathrm{o_P}(n^{\delta-1/2}) = \mathrm{o_P}(1).$$

4. *The martingale term* $\widetilde{\mathcal{S}}_{t,2}$. Argue as in item 3. First, the norm inequality gives

$$|\widetilde{\mathcal{S}}_{t,2}| \le n^{-1/2}||\hat\vartheta_n||^2||\textstyle\sum_{s=1}^t \varepsilon_s\ddot h_s(\theta_0)|| \le \mathrm{o_P}(n^{2\delta-1/2})||\textstyle\sum_{s=1}^t \varepsilon_s\ddot h_s(\theta_0)||.$$

Then apply Lemma A.1 (ii) using Assumptions 3.1, 4.3 along with Assumption 4.2 (d) to get

$$\max_{n_0\le t\le n}|\widetilde{\mathcal{S}}_{t,2}| = \mathrm{o_P}(n^{2\delta-1/2})\mathrm{o_P}(n^{(1-4\delta-\eta)(1/2+\varsigma)+\varsigma}) + \mathrm{o_P}(n^{2\delta-1/2}) = \mathrm{o_P}(1),$$

when $\delta < 1/4$ and $\varsigma > 0$ is chosen sufficiently small.

5. *The term* $\widetilde{\mathcal{S}}_{t,\hat\theta_n,3}$. Apply the norm and triangle inequalities to get

$$|\widetilde{\mathcal{S}}_{t,\hat\theta_n,3}| \le ||\hat\vartheta_n||^2 n^{-1/2}\textstyle\sum_{s=1}^t |\varepsilon_s|\,||\ddot h_s(\theta_*) - \ddot h_s(\theta_0)||.$$

The summands are positive so that a further bound arises by extending the summation limit

$$|\widetilde{\mathcal{S}}_{t,\hat\theta_n,3}| \le ||\hat\vartheta_n||^2 n^{-1/2}\textstyle\sum_{s=1}^n |\varepsilon_s|\,||\ddot h_s(\theta_*) - \ddot h_s(\theta_0)||,$$

where θ_* remains dependent on t and $\hat{\theta}_n$. Apply the Hölder inequality to get

$$|\widetilde{\mathcal{S}}_{t,\hat{\theta}_n,3}| \leq ||\hat{\vartheta}_n||^2 (n^{-1}\textstyle\sum_{s=1}^n \varepsilon_s^2)^{1/2} (\sum_{s=1}^n ||\ddot{h}_s(\theta_*) - \ddot{h}_s(\theta_0)||^2)^{1/2}.$$

The martingale Law of Large Numbers (Chow, 1965, Theorem 5) shows $n^{-1}\sum_{s=1}^n \varepsilon_s^2 = \mathrm{O}(1)$ *a.s.* By Assumption 4.1 (a) then $\hat{\vartheta}_n = N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o}_{\mathsf{P}}(n^\delta)$. For any $\epsilon > 0$ and large n then $||N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0)|| \leq \epsilon n^\delta$ with large probability. For such $\hat{\theta}_n$ we have that θ_* is also local to θ_0 and we can then bound

$$\textstyle\sum_{s=1}^n ||\ddot{h}_s(\theta_*) - \ddot{h}_s(\theta_0)||^2 \leq \sup_{\theta: ||N_{n,\theta_0}^{-1}(\theta-\theta_0)|| \leq \epsilon n^\delta} \sum_{s=1}^n ||\ddot{h}_s(\theta) - \ddot{h}_s(\theta_0)||^2,$$

which depends neither on t nor $\hat{\theta}_n$. Then Assumption 4.2 (e) implies $|\widetilde{\mathcal{S}}_{t,\hat{\theta}_n,3}| = \mathrm{o}_{\mathsf{P}}(n^{2\delta-2\delta}) = \mathrm{o}_{\mathsf{P}}(1)$ uniformly in t.

6. *The compensator.* As before $||N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0)|| \leq \epsilon n^\delta$ on a set with large probability. On that set $\overline{\mathcal{S}}_{t,\hat{\theta}_n} \leq \sup_{\theta: ||N_{n,\theta_0}^{-1}(\theta-\theta_0)|| \leq \epsilon n^\delta} \overline{\mathcal{S}}_{t,\theta}$ which is $\mathrm{o}_{\mathsf{P}}(1)$ by Assumption 4.2 (a).

Part II: Assumption 3.3.

1. *The problem.* Let $\mathcal{V}_{n,\theta} = n^{-1}\sum_{t=1}^n [\{\varepsilon_t - \nabla g_s(\theta)\}^4 - \varepsilon_t^4]$ where $\nabla g_s(\theta) = g(x_s,\theta) - g(x_s,\theta_0)$ as before, so that $\mathcal{V}_{n,\hat{\theta}_n} = n^{-1}\sum_{t=1}^n (\hat{\varepsilon}_t^4 - \varepsilon_t^4)$.

2. *Some inequalities*: By binomial expansion $(\varepsilon - \nabla)^4 - \varepsilon^4 = \nabla^4 - 4\nabla^3\varepsilon + 6\nabla^2\varepsilon^2 - 4\nabla\varepsilon^3$. Thus, by Hölder's inequality,

$$\begin{aligned}|\mathcal{V}_{n,\theta}| &\leq n^{-1}\textstyle\sum_{t=1}^n \{\nabla g_s(\theta)\}^4 - 4[n^{-1}\sum_{t=1}^n \{\nabla g_s(\theta)\}^4]^{3/4}(n^{-1}\sum_{t=1}^n \varepsilon_t^4)^{1/4} \\ &\quad + 6[n^{-1}\textstyle\sum_{t=1}^n \{\nabla g_s(\theta)\}^4]^{1/2}(n^{-1}\sum_{t=1}^n \varepsilon_t^4)^{1/2} - 4[n^{-1}\sum_{t=1}^n \{\nabla g_s(\theta)\}^4]^{1/4}(n^{-1}\sum_{t=1}^n \varepsilon_t^4)^{3/4}.\end{aligned}$$

Now, $n^{-1}\sum_{t=1}^n \varepsilon_t^4 = \mathrm{O}_{\mathsf{P}}(1)$ by the martingale Law of Large Numbers and Assumption 3.1 while $n^{-1}\sum_{t=1}^n \{\nabla g_s(\hat{\theta}_n)\}^4 = \mathrm{o}_{\mathsf{P}}(1)$ by an argument as in part I, item 6 using Assumption 4.2 (b). □

Proof of Theorem 4.2. Since $\hat{\theta}_n = \theta_0 + \mathrm{o}(n^\delta)$ *a.s.* by Assumption 4.1 (b) then Egorov's theorem (Davidson 1994, Theorem 18.4) implies $\forall \nu > 0\ \exists t_0$ so $\Omega_\eta = \{\sup_{t>t_0} |N_{n,\theta_0}^{-1}(\hat{\theta}_t - \theta_0)| < \nu n^\delta\}$ satisfies $\mathsf{P}(\Omega_\eta) > 1 - \nu$. On Ω_η we bound

$$\max_{1\leq t\leq n} |n^{-1/2}\sum_{s=1}^t (\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)| \leq n^{-1/2} \max_{1\leq t\leq t_0} |\sum_{s=1}^t (\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)| + \max_{t_0+1\leq t\leq n} |n^{-1/2}\sum_{s=1}^t (\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)|.$$

Since t_0 is finite, the first term vanishes. For the second term we can follow the proof of Theorem 4.1 replacing $\hat{\varepsilon}_{s,n}^2$ by $\hat{\varepsilon}_{s,t}^2$. When expanding in item 3 the intermediate point θ_* will now depend on t through the summation limit and $\hat{\theta}_t$. However, with $t > t_0$ then $\hat{\theta}_t$ is local to θ_0 uniformly in t and the remaining arguments apply. □

Proof of Theorem 4.3. Assumption 4.4 (a,b,c) with $k = 2$ imply Assumption 4.2 (c,d,e).

Now, recall the notation in item 3 in the proof of Theorem 4.1 and expand

$$g(x_t,\theta) - g(x_t,\theta_0) = \vartheta'\dot{h}_t(\theta_0) + \frac{1}{2}\vartheta'\ddot{h}_t(\theta_0)\vartheta + \frac{1}{2}\vartheta'\{\ddot{h}_t(\theta_t) - \ddot{h}_t(\theta_0)\}\vartheta,$$

where θ_t is an intermediate point depending on x_t so $|\theta_t - \theta_0| \leq |\theta - \theta_0|$. Raise this to the power $k = 2$ or $k = 4$ and apply the inequality $(x + y + z)^m \leq C(x^m + y^m + z^m)$ to see that

$$|g(x_t,\theta) - g(x_t,\theta_0)|^k \leq C\{||\vartheta||^k||\dot{h}_t(\theta_0)||^k + ||\vartheta||^{2k}||\ddot{h}_t(\theta_0)||^k + ||\vartheta||^{2k}||\ddot{h}_t(\theta_t) - \ddot{h}_t(\theta_0)||^k\}.$$

In Assumption 4.2 (a,b) we only consider $||\vartheta|| \leq \epsilon n^{\delta}$. Thus θ_t is local to θ_0 so that $||\ddot{h}_t(\theta_t) - \ddot{h}_t(\theta_0)||^k \leq \sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq \epsilon n^{\delta}} ||\ddot{h}_t(\theta) - \ddot{h}_t(\theta_0)||^k$. Then cumulate to get

$$|\sum_{t=1}^{n}\{g(x_t,\theta) - g(x_t,\theta_0)\}^k| \leq \epsilon^k n^{\delta k} \sum_{t=1}^{n} ||\dot{h}_t(\theta_0)||^k$$
$$+ \epsilon^{2k} n^{2\delta k} \sum_{t=1}^{n} ||\ddot{h}_t(\theta_0)||^k + \epsilon^{2k} n^{2\delta k} \sum_{t=1}^{n} \sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq \epsilon n^{\delta}} ||\ddot{h}_t(\theta) - \ddot{h}_t(\theta_0)||^{2k},$$

which is $o_{\mathsf{P}}(n^{1/2})$ for $k = 2$ and $o_{\mathsf{P}}(n)$ for $k = 4$ due to Assumption 4.4. □

B Tables

Table 1: DGPs: Data Generating Processes

*	DGP	y_t	$g(x_t,\theta)$
CS	1	$1 + 0.5x_t + \varepsilon_t$	$\theta_1 + \theta_2 x_t$
CS	2	$1 + 0.5x_t^2 + \varepsilon_t$	$\theta_1 + \theta_2 x_t^2$
CS	3	$1 + 0.9x_t 1(v_t \leq 0) + 0.5x_t 1(v_t > 0) + \varepsilon_t$	$\theta_1 + \theta_2 x_t 1(v_t \leq 0) + \theta_3 x_t 1(v_t > 0)$
CS	4	$1 + 0.3\lvert x_t\rvert^{1.5} + \varepsilon_t$	$\theta_1 + \theta_2 \lvert x_t\rvert^{\theta_3}$
M	5	$y_{t-1} + \varepsilon_t$	$\theta_1 + \theta_2 \lvert x_t\rvert^{\theta_3}$
M	6	$1 + 0.9x_t 1(v_t \leq 0) + 0.5x_t 1(v_t > 0) + \varepsilon_t$	$\theta_1 + \theta_2 x_t$
M	7	$1 + 0.5x_t^2 + \varepsilon_t$	$\theta_1 + \theta_2 x_t$
M	8	$1 + 0.3\lvert x_t\rvert^{1.5} + u_t \quad u_t = x_t + \varepsilon_t$	$\theta_1 + \theta_2 \lvert x_t\rvert^{\theta_3}$
M	9	$1 + 0.5x_t^2 + \varepsilon_t$	$\theta_1 + \theta_2 \ln^2 \lvert x_t\rvert$

CS denotes correct specification and M denotes misspecification. y_t and $g(x_t,\theta)$ are the dependent variable and the estimated regression function, respetively. $x_t \sim I(\tau)$ with $\tau = 0.7, 1, 2$. $\varepsilon_t, v_t \sim i.i.d.N(0,1)$. x_t, ε_t, and v_t are independent of each other.

Table 2: Size and Power: Finite Sample Performance

		$x_t \sim I(0.7)$			$x_t \sim I(1)$			$x_t \sim I(2)$		
$CUSQ_n$		n			n			n		
*	DGP	100	500	1000	100	500	1000	100	500	1000
CS	1	0.031	0.041	0.044	0.032	0.040	0.044	0.031	0.040	0.044
CS	2	0.031	0.040	0.045	0.031	0.040	0.044	0.031	0.039	0.044
CS	3	0.030	0.041	0.043	0.033	0.042	0.043	0.033	0.041	0.042
CS	4	0.031	0.040	0.045	0.031	0.041	0.043	0.033	0.040	0.044
M	5	0.527	0.975	0.997	0.814	0.999	1.000	0.957	1.000	1.000
M	6	0.085	0.485	0.708	0.553	0.984	0.999	0.998	1.000	1.000
M	7	0.096	0.790	0.962	0.479	0.993	1.000	0.974	1.000	1.000
M	8	0.302	0.854	0.946	0.460	0.846	0.913	0.935	1.000	1.000
M	9	0.313	0.709	0.775	0.320	0.599	0.759	0.945	0.999	0.999

CS denotes correct specification; hence, size is being analyzed in those cases. M denotes misspecification; hence, power is considered in those cases. 10000 replications are conducted.

Table 3: Power performance comparison with Kasparis (2008)

DGP	y_t
R1	z_t
R2	$\text{sign}(z_t)\,\lvert z_t\rvert^{0.5}$
R3	$\text{sign}(x_t)\,\lvert x_t\rvert^{0.75} + u_t$
R4	$\text{sign}(x_t)\,\lvert x_t\rvert^{1.25} + u_t$
R5	$\ln(1 + \lvert x_t\rvert) + u_t$
R6	$x_t + \lvert x_t\rvert^{0.5} + u_t$
R7	$0.4x_t 1(x_t \leq 0) + 1.8x_t 1(x_t \geq 0) + u_t$
R8	$x_t + 1.8\,[x_t/(1 + \exp(-x_t/\sqrt{n} - 2))] + u_t$
R9	$x_t + z_t + u_t$
R10	$\text{sign}(x_t)\,(\lvert x_t\rvert\,\lvert z_t\rvert)^{0.5} + u_t$

$z_t = z_{t-1} + w_t$ where $w_t = 0.3w_{t-1} + \omega_t$, $x_t = x_{t-1} + \eta_t$, $u_t = \epsilon_t$, $(\epsilon_t, \eta_{t+1}, \omega_{t+1})' = Dr_t$ where $r_t \sim i.i.d.N(0,1)$ and $D = [1\ .2\ .1,\ .3\ 2\ 0,\ 0\ .1\ 1.2]$

Table 4: Power performance comparison with Kasparis (2008)

	$CUSQ_n$			Kasparis' best power		
n	100	200	500	100	200	500
R1	0.909	0.999	1.000	0.762	0.920	0.984
R2	0.925	1.000	1.000	0.790	0.930	0.984
R3	0.093	0.612	0.860	0.180	0.377	0.698
R4	0.349	0.962	0.996	0.430	0.706	0.902
R5	0.408	0.922	0.986	0.706	0.901	0.993
R6	0.514	0.953	0.993	0.626	0.862	0.989
R7	0.548	0.825	0.872	0.485	0.597	0.704
R8	0.340	0.849	0.959	0.327	0.557	0.825
R9	0.882	0.999	1.000	0.753	0.915	0.983
R10	0.670	0.997	1.000	0.411	0.702	0.904

References

Amemiya, T. (1985) Advanced econometrics. Cambridge, MA: Harvard University Press.

Berenguer-Rico, V. and J. Gonzalo (2014) Co-summability: From linear to non-linear co-integration. Mimeo.

Billingsley, P. (1999) Convergence of probability measures. New York, NJ: Wiley.

Brown, B.M. (1971) Martingale central limit theorems. Annals of Mathematical Statistics 42, 59–66.

Brown, R.L., J. Durbin and J.M. Evans (1975) Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society B37, 149-192.

Chan, N. and Q. Wang (2015) Nonlinear regressions with nonstationary time series. Journal of Econometrics 185, 182–195.

Choi, I. and P. Saikkonen (2010) Tests for non-linear cointegration. Econometric Theory 26, 682-709.

Chow, Y.S. (1965) Local convergence of martingales and the law of large numbers. Annals of Mathematical Statistics 36, 552–558.

Davidson, J. (1994) Stochastic limit theory. Oxford: Oxford University Press.

Deng, A. and P. Perron (2008) A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change. Journal of Econometrics 142, 212–240.

Edgerton, D. and C. Wells (1994) Critical values for the CUSUMSQ statistic in medium and large sized samples. Oxford Bulletin of Economics and Statistics 56, 355–365.

Hao, K. and B. Inder (1996) Diagnostic test for structural change in cointegrated regression models. Economic Letters 50, 179–187.

Jennrich, R.I. (1969) Asymptotic properties of non-linear least squares estimators. The Annals of Mathematical Statistics 40, 633–643.

Kasparis, I. (2008) Detection of functional form misspecification in cointegrating relations. Econometric Theory 24, 1373-1403.

Koziol, J.A. and D.P. Byar (1975) Percentage points of the asymptotic distributions of one and two sample K-S statistics for truncated or censored data. Technometrics 17, 507-510.

Kristensen, D. and A. Rahbek (2010) Likelihood-based inference for cointegration with nonlinear error-correction. Journal of Econometrics 158, 78-94.

Lai, T.L. and C.Z. Wei (1982) Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. Annals of Statistics 10, 154-166.

Lee, S., O. Na and S. Na (2003) On the CUSUM of squares test for variance change in non-stationary and non-parametric time series models. Annals of the Institute of Statistical Mathematics 55, 467-485.

McCabe, B.P. M. and M.J. Harrison (1980) Testing the constancy of regression relationships over time using least squares residuals. Applied Statistics 29, 142-148.

Nielsen, B. and J. Sohkanen (2011) Asymptotic behaviour of the CUSUM of squares test under stochastic and deterministic time trends. Econometric Theory 27, 913-927.

Park, J.Y. and P.C.B. Phillips (2001) Non-linear regressions with integrated time series. Econometrica 69, 117-161.

Ploberger, W. and W. Krämer (1986) On studentizing a test for structural change. Economics Letters 20, 341-344.

Ploberger, W. and W. Krämer (1990) The local power of the CUSUM and CUSUM of squares test. Econometric Theory 6, 335-347.

Pötscher, B.M. (2004) Nonlinear functions and convergence to Brownian motion: Beyond the continuous mapping theorem. Econometric Theory 20, 1-22.

Sohkanen, J.S. (2011) Properties of tests for mis-specification in non-stationary autoregressions. DPhil Thesis, University of Oxford.

Stout, W.F. (1974) Almost sure convergence. New York, NJ: Academic Press.

Turner, P. (2010) Power properties of the CUSUM and CUSUMSQ tests for parameter instability. Applied Economic Letters 17, 1049-1053.

Wooldridge, J.M. (1994) Estimation and inference for dependent processes. In R.F. Engle and D.L. McFadden (eds.) Handbook of Econometrics 4, 2639-2738.

Xiao, Z. and P.C.B. Phillips (2002) A CUSUM test for cointegration using regression residuals. Journal of Econometrics 108, 43–61.