**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 non-stationary, 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

\[ y_t = g(x_t, \theta) + \varepsilon_t \quad t = 1, \ldots, n, \]

---

†Department of Economics, University of Oxford, Nuffield College, and Programme for Economic Modelling.
where the functional form of $g$ is known. The innovation $\varepsilon_t$ is a martingale difference sequence with respect to a filtration $\mathcal{F}_t$ with zero mean, variance $\sigma^2$ and fourth moment $\varphi^2 = \mathbb{E}\varepsilon_t^4 - (\mathbb{E}\varepsilon_t^2)^2$, the regressor $x_t$ is a $p$-vector $\mathcal{F}_{t-1}$-adapted, and the parameter $\theta$ 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 $\varepsilon_t$ and $x_t$ will be regarded as misspecification.

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

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

$$\text{CUSQ}_n = \frac{1}{\phi_n} \max_{1 \leq t \leq n} \left| \frac{1}{\sqrt{n}} \left( \sum_{s=1}^t \tilde{\varepsilon}_{s,n}^2 - \frac{t}{n} \sum_{s=1}^n \tilde{\varepsilon}_{s,n}^2 \right) \right|,$$

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

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

We will argue that under quite general assumptions,

$$\text{CUSQ}_n \overset{D}{\to} \sup_{0 \leq u \leq 1} |\mathcal{B}_u^0|,$$

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

$$\text{RCUSQ}_n = \frac{1}{\phi_n} \max_{0 \leq s \leq t \leq n} \left| \frac{1}{\sqrt{n}} \left( \sum_{s=1}^t \tilde{\varepsilon}_{s,t}^2 - \frac{t}{n} \sum_{s=1}^n \tilde{\varepsilon}_{s,n}^2 \right) \right|.$$

If the sequence of estimators $\theta_n$ converges strongly, we can show that also

$$\text{RCUSQ}_n \overset{D}{\to} \sup_{0 \leq u \leq 1} |\mathcal{B}_u^0|.$$

### 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 $\varepsilon_t$ is $\mathcal{F}_t$-adapted and $\mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ a.s., so that

(a) $\mathbb{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma^2$ a.s.;

(b) $\mathbb{E}(\varepsilon_t^4 - \sigma^4 | \mathcal{F}_{t-1}) = \varphi^2$ a.s.;

(c) $\sup_t \mathbb{E}(\varepsilon_t^4 | \mathcal{F}_{t-1}) < \infty$ 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 $\mathcal{B}^0_u$ 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,

$$
\frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} \left( \varepsilon_t^2 - \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2 \right) \overset{D}{\rightarrow} \varphi \mathcal{B}^0_u, \quad u \in [0,1],
$$

$$
\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^4 - \left( \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2 \right)^2 \overset{D}{\rightarrow} \varphi^2.
$$

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 \(\max_{1 \leq t \leq n} n^{-1/2} \sum_{s=1}^{t} (\varepsilon_{s,n}^2 - \varepsilon_s^2) = o_p(1)\).

Assumption 3.3 \(n^{-1} \sum_{t=1}^{n} (\varepsilon_t^4 - \varepsilon_t^4) = o_p(1)\).

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 $\text{CUSQ}_n \overset{D}{\rightarrow} \sup_{0 \leq u \leq 1} |\mathcal{B}^0_u|$.

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

Assumption 3.4 \(\max_{1 \leq t \leq n} n^{-1/2} \sum_{s=1}^{t} (\varepsilon_{s,n}^2 - \varepsilon_s^2) = o_p(1)\).

Theorem 3.3 If Assumptions 3.1, 3.3, 3.4 are satisfied then $\text{RCUSQ}_n \overset{D}{\rightarrow} \sup_{0 \leq u \leq 1} |\mathcal{B}^0_u|$.

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 $N^{-1}_{n,\theta_0}(\hat{\theta}_n - \theta_0)$ is either (a) $o_p(n^\delta)$ or (b) $o(n^\delta)$ a.s.

The normalization $N^{-1}_{n,\theta_0}$ allows both stationary and non-stationary regressors. In linear models $N^{-1}_{n,\theta_0} = n^{1/2}$ for stationary regressors and $N^{-1}_{n,\theta_0} = n$ for random walk regressors. In more general cointegrated models $N^{-1}_{n,\theta_0}$ 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 $\theta$ under which we evaluate the distributions. We use the notation $\theta_0$ 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 $\theta$. These are the $q$-vector $\tilde{g}(x,t,\theta) = \partial g(x,t,\theta)/\partial \theta$ and the $q \times q$ square matrix $\tilde{g}(x,t,\theta) = \partial g(x,t,\theta)/\partial \theta \partial \theta'$. 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 \( \theta \). Let \( \delta < 1/4 \) be the consistency rate in Assumption 4.1 and let \( \epsilon > 0 \). Suppose
\[
\begin{align*}
(a) \; & \sup_{\theta_0} \frac{1}{n} \sum_{i=1}^{n} \left( g(x_i, \theta) - g(x_i, \theta_0) \right)^2 = o_p(n^{1/2}); \\
(b) \; & \sup_{\theta_0} \frac{1}{n} \sum_{i=1}^{n} \left( g(x_i, \theta) - g(x_i, \theta_0) \right)^4 = o_p(n); \\
(c) \; & \sum_{i=1}^{n} \left| \frac{N^{-1}_{n, \theta_0} \tilde{g}(x_i, \theta_0)}{\sum_{i=1}^{n} \left( g(x_i, \theta) - g(x_i, \theta_0) \right)^2} \right|^2 = O_p(n^{1-2\delta}) \quad \text{for some } \eta > 0; \\
(d) \; & \sum_{i=1}^{n} \left| \frac{N^{-1}_{n, \theta_0} \tilde{g}(x_i, \theta_0)}{N_{n, \theta_0} \sum_{i=1}^{n} \left( g(x_i, \theta) - g(x_i, \theta_0) \right)^2} \right|^2 = O_p(n^{1-4\delta}) \quad \text{for some } \eta > 0; \\
(e) \; & \sup_{\theta_0} \frac{1}{n} \sum_{i=1}^{n} \left( \tilde{g}(x_i, \theta) - \tilde{g}(x_i, \theta_0) \right)^2 N_{n, \theta_0}^2 = O_p(n^{-4\delta}).
\end{align*}
\]

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

Assumption 4.3 Suppose \( \inf \{ n : \sum_{i=1}^{n} w_i w_i^T \text{ is invertible} \} < \infty \) a.s. for \( w_i = \tilde{g}(x_i, \theta_0) \) and \( w_i = \text{vec} \{ \tilde{g}(x_i, \theta_0) \} \) with the convention that the empty set has infinite infimum. Moreover, suppose \( N^{-1}_{n, \theta_0} = o(n^\epsilon) \) for some \( \epsilon > 0 \).

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 \( n^{-1/2} \sum_{s=1}^{n} (\tilde{\varepsilon}_{s,n} - \varepsilon_s^2) \) through a martingale decomposition. Noting that \( \tilde{\varepsilon}_{s,n} - \varepsilon_s = \nabla g(x_s, \theta_n) = g(x_s, \theta_n) - g(x_s, \theta_0) \) and expanding \( (\varepsilon - \nabla)^2 - \varepsilon^2 = -2\varepsilon \nabla + \nabla^2 \) we get
\[
n^{-1/2} \sum_{s=1}^{n} (\tilde{\varepsilon}_{s,n}^2 - \varepsilon_s^2) = -2n^{-1/2} \sum_{s=1}^{n} \varepsilon_s \nabla g(x_s, \theta_n) + n^{-1/2} \sum_{s=1}^{n} \{ \nabla g(x_s, \theta_n) \}^2.
\] (4.1)

Due to Assumption 4.1 the estimator \( \hat{\theta}_n \) varies in a local region around \( \theta_0 \). Thus, it suffices to replace \( \hat{\theta}_n \) with a deterministic value \( \theta \) 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 \( \varphi_n^2 \).

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 \( \theta_0 \), 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 \( \theta \). Let \( \delta < 1/4 \) be the consistency rate in Assumption 4.1 and let \( \epsilon > 0 \). Suppose, the following conditions hold, for \( k = 2, 4, \)
\[
\begin{align*}
(a) \; & \sum_{i=1}^{n} \left| \frac{N^{-1}_{n, \theta_0} \tilde{g}(x_i, \theta_0)}{\sum_{i=1}^{n} \left( g(x_i, \theta) - g(x_i, \theta_0) \right)^2} \right|^k = o_p(n^{k/4 - \delta}); \\
(b) \; & \sum_{i=1}^{n} \left| \frac{N^{-1}_{n, \theta_0} \tilde{g}(x_i, \theta_0)}{N_{n, \theta_0}} \right|^k = o_p(n^{k/4 - \delta}); \\
(c) \; & \sum_{i=1}^{n} \sup_{\theta_0} \frac{1}{N^{-1}_{n, \theta_0}(\theta - \theta_0)} \left| \frac{N^{-1}_{n, \theta_0} \tilde{g}(x_i, \theta) - \tilde{g}(x_i, \theta_0)}{N_{n, \theta_0}} \right|^k = o_p(n^{-4\delta}).
\end{align*}
\]

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 \( \sum_{t=1}^{n} |N_{n, \delta_0} x_t|^k = \text{Op}(n^{k/4-k\delta}) \) for \( k = 2, 4 \). Suppose \( x_t \) is univariate and stationary then \( N_{n, \delta_0} = n^{-1/2} \) whereas \( N_{n, \delta_0} = n^{-1} \) if \( x_t \) is a random walk. In both cases \( N_{n, \delta_0} \sum_{t=1}^{n} |x_t|^k = \text{Op}(1) = \text{op}(n^{k/4-k\delta}) \).

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

\[
y_t = |x_t|^\theta + \varepsilon_t \quad t = 1, \ldots, n, \tag{5.1}
\]

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 \( N_{n, \delta_0} \). Hence, if \( x_t \) is stationary with finite \( |x_t|^{4\delta_0} \log^8 |x_t| \) moments we let \( N_{n, \delta_0} = \sqrt{n} \) and apply techniques from Wooldridge (1994). If \( x_t \) is a random walk we let \( N_{n, \delta_0} = n^{(1+\delta_0)/2} \log n \) 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 \( N_{n, \delta_0} (\hat{\theta}_n - \theta) = \text{Op}(1) \) as given. Hence Assumption 4.1 follows for any \( \delta > 0 \).

To prove Assumption 4.4 we differentiate \( g \) and get

\[
g(x, \theta) = |x|^\theta, \quad \dot{g}(x, \theta) = |x|^\theta \log |x|, \quad \ddot{g}(x, \theta) = |x|^\theta \log^2 |x|.
\]

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

\[
S = \sum_{t=1}^{n} |N_{n, \delta_0} \dot{g}(x_t, \theta_0)|^k = \text{Op}(1) = \text{op}(n^{k/4-k\delta}).
\]

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

\[
S = \frac{1}{n^{k/2}} \sum_{t=1}^{n} |x_t|^{k\delta_0} \log^k |x_t| = O(n^{1-k/2}) = O(1) \quad \text{a.s.}
\]

In the random walk case we get

\[
S = \frac{1}{n^{1+\delta_0}k/2} \sum_{t=1}^{n} |x_t|^{k\delta_0} \log^k |x_t| = \frac{1}{n^{k/2}} \sum_{t=1}^{n} |x_t/n^{1/2}|^{k\delta_0} \left( \log |x_t/n^{1/2}| \log n^{1/2} + 1 \right)^k = \text{Op}(1),
\]
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 \( x_{\text{integer}(u_n)}/n^{1/2} \) 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 \( \sum_{t=1}^n |N_{n\theta_00} \tilde{g}(x_t, \theta_0)|^k = O_p(n^{-k}) \).

For Assumption 4.4 (c) we apply a Lipschitz argument. The second derivative of \( g \) satisfies
\[
|\tilde{g}(x_t, \theta) - \tilde{g}(x_t, \theta_0)| \leq (|x|^{\theta} + |x|^{-\theta}) |x|^{\theta_0} \log^2 |x|
\]
for all \( \theta \) 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
\[
\sum_{t=1}^n \sup_{|\theta|,|\theta_0| \leq \nu} |N_{n\theta_0}^2 \{ \tilde{g}(x_t, \theta) - \tilde{g}(x_t, \theta_0) \}|^k \\
\leq \sum_{t=1}^n \{ N_{n\theta_0}^2(|x_t|^\theta + |x_t|^{-\theta}) |x_t|^{\theta_0} \log^2 |x_t| \}^k = O_p(n^{-k}) = o_p(1), \quad (5.2)
\]
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 \( \varepsilon_t \) 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 \( N_{n\theta_0}^{-1} (\theta_n - \theta_0) = O_p(1) \) 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 \( \gamma \) are of order \( O(|z|) \). 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:
\[
y_t = g(x_t, \theta) + \varepsilon_t \quad t = 1, \ldots, n, \quad (5.3)
\]
\[
x_t = x_{t-1} + v_t, \quad (5.4)
\]
where \( \varepsilon_t \) is an \( F_t \)-martingale difference sequence, \( (\varepsilon_t, v_t)' \) satisfies a functional central limit theorem, \( x_t \) is \( F_{t-1} \)-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 \( \theta \). 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 \( n^{1/4} (\theta_n - \theta) \) 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 \( \kappa \) 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 \( g, \hat{g} \) and \( \bar{g} \) are assumed to be asymptotically homogeneous and satisfy conditions that have the same flavour as those in Assumption 4.4. It then follows that \( n^{1/2}\kappa(\sqrt{n}, \theta_0) (\theta_n - \theta) \) converges in distribution. Thus, we can choose the normalization \( N_{n,\theta_0}^{-1} = n^{1/2}\kappa(\sqrt{n}, \theta_0) \). For instance, in the power function model (5.1) with random walk regressor we have \( \kappa(\sqrt{n}, \theta_0) = n^{3/2}\log n^{1/2} \).

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 presented for various asymptotically homogeneous models. First, we check size and power for a

\[
\text{CUSQ test with the power of a cumulated sum (CUSUM) test reported by Kasparis (2008). We find that the two tests have power of similar magnitude, so there is no apparent advantage in using the more complicated CUSUM test.}
\]

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 \( \Delta^r x_t = \text{ iid } N(0,1) \) 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 1.36–0.67\(n^{-1/2} - 0.89n^{-1} \) suggested by Edgerton and Wells (1994). Similarly, for the recursive test Sohkanen (2011) suggests the 95% critical value 1.36(1 – 0.68\(n^{-1/2} + 3.13n^{-1} - 33.9n^{-3/2} + 93.9n^{-2} \)).

Table 2, DGPs 5-9, reports the power of the CUSQ test for a range of asymptotically homogeneous 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 \( y_t = \theta_1/(1 + \theta_2 x_t^2) + \epsilon_t \), while the regression model is polynomial. Simulations not reported here show that power arises as long as the signal from the integrable function component \( \theta_1/(1 + \theta_2 x_t^2) \) dominates the noise \( \epsilon_t \).

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

\[
\|m\| = \sqrt{\max \text{eigen}(m'm)}.
\]

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

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

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, \( \|m\|_F = \|\text{vec}(m)\|_F \).

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

**Lemma A.1** Let \( N_{n,\theta_0} \) be a \( q \times q \) normalizing matrix where \( N_{n,\theta_0}^{-1} = O(n^\ell) \) 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 \( \theta \): \( \dot{g}, \ddot{g} \). Also let Assumption 3.1(a) hold. Let \( w_t \) be \( \mathcal{F}_{t-1} \)-measurable and given as either of

\[(i) \ w_t = N_{n,\theta_0}^{-1} \dot{g}(x_t, \theta_0); \]
\[(ii) \ w_t = N_{n,\theta_0}^{-1} \ddot{g}(x_t, \theta_0) N_{n,\theta_0}. \]

Suppose \( n_0 = \inf \{n : \sum_{i=1}^n \{\text{vec}(w_i)\} \{\text{vec}(w_i)\}' \} \) is invertible \( < \infty \) a.s. Then, for all \( \zeta > 0 \),

\[
\max_{n_0 \leq s \leq n} \|\sum_{i=1}^s w_i e_i\| \overset{a.s.}{=} o \left( n^\zeta \left\| \sum_{i=1}^n \{\text{vec}(w_i)\} \{\text{vec}(w_i)\}' \right\|^{1/2+\zeta} \right) + O(1)
\]

\[
= o \left\{ n^\zeta \left( \sum_{i=1}^n \|w_i\|^2 \right)^{1/2+\zeta} \right\} + O(1).
\]

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

\[
S_{g\bar{g},u} = \sum_{i=1}^{[nu]} \{\dot{g}(x_t, \theta_0)\} \dot{g}(x_t, \theta_0)',
\]

so that

\[
N_{n,\theta_0} S_{g\bar{g},u} = \sum_{i=1}^{[nu]} \{\dot{g}(x_t, \theta_0)\} \ddot{g}(x_t, \theta_0) N_{n,\theta_0}.
\]

Notice that, for \( n_0 < [nu] \),

\[
\|N_{n,\theta_0}' S_{g\bar{g},u} \| = \left( \|N_{n,\theta_0}' S_{g\bar{g},u} N_{n,\theta_0}^{-1} \|^2 \right)^{1/2} \|S_{g\bar{g},u} \|^{1/2} \|S_{g\bar{g},u}^{-1/2} S_{g\bar{g},u} \| \overset{A.1}{=} \left( \|S_{g\bar{g},u} \|^2 \right)^{1/2} + O(1),
\]

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

\[
\left( \|S_{g\bar{g},u} \|^{2\zeta} \right)^{1/2 \zeta} = o \left( \|N_{n,\theta_0}' S_{g\bar{g},1} N_{n,\theta_0} \|^\zeta \right) + O(1) = o \left( n^\zeta \|N_{n,\theta_0}' S_{g\bar{g},1} N_{n,\theta_0} \|^\zeta \right) + O(1),
\]

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

\[
\sup_u \|N_{n,\theta_0}' S_{g\bar{g},u} \| = o(n^\zeta \|N_{n,\theta_0}' S_{g\bar{g},1} N_{n,\theta_0} \|^1) + O(1),
\]

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
\[
\left\| \sum_{t=1}^{nu} w_t \varepsilon_t \right\| \leq \left\| \sum_{t=1}^{nu} w_t \varepsilon_t \right\|_F = \left\| \sum_{t=1}^{nu} \text{vec} \left( w_t \varepsilon_t \right) \right\|_F.
\]
Now argue as in (i) with \( w_t \) replaced by \( \text{vec} \left( w_t \right) \) to get
\[
\sup_u \left\| \sum_{t=1}^{nu} w_t \varepsilon_t \right\| \leq o \left( n^\alpha \left\| \sum_{t=1}^{n} \{ \text{vec} \left( w_t \right) \} \right\|^{1/2+\epsilon} \right) \quad a.s.,
\]
which is the first desired expression. To get the second expression notice that
\[
\left\| \sum_{t=1}^{n} \{ \text{vec} \left( w_t \right) \} \right\| \leq \sum_{t=1}^{n} \left\| \{ \text{vec} \left( w_t \right) \} \right\|
\]
and \( \left\| \{ \text{vec} (w_t) \} \right\|^2 = \| w_t \|^2 \leq q \| w_t \|^2 \) as desired. \( \Box \)

Proof of Theorem 3.2: The statistic of interest is \( CUSQ_n = A_n / \hat \varphi_n \), where
\[
A_n = \max_{1 \leq t \leq n} \left| n^{-1/2} \sum_{s=1}^{t} \left( \hat \varepsilon_{s,n}^2 - n^{-1} \sum_{r=1}^{n-1} \hat \varepsilon_{r,n}^2 \right) \right|.
\]
Expand \( A_n = B_n + (A_n - B_n) \), where
\[
B_n = \max_{1 \leq t \leq n} \left| n^{-1/2} \sum_{s=1}^{t} \left( \varepsilon_{s}^2 - n^{-1} \sum_{r=1}^{n-1} \varepsilon_{r}^2 \right) \right|.
\]
Noting that \( \hat \varepsilon_{s,n}^2 = (\hat \varepsilon_{s,n}^2 - \varepsilon_s^2) + \varepsilon_s^2 \) we get
\[
A_n - B_n = \max_{1 \leq t \leq n} \left| n^{-1/2} \sum_{s=1}^{t} \left( \hat \varepsilon_{s,n}^2 - n^{-1} \sum_{r=1}^{n} \hat \varepsilon_{r,n}^2 \right) - n^{-1/2} \sum_{s=1}^{t} \left( \hat \varepsilon_{s,n}^2 - \varepsilon_s^2 \right) \right|
\]
By the triangle inequality \( A_n - B_n \leq C_n - B_n = C_n \) where
\[
C_n = \max_{1 \leq t \leq n} \left| n^{-1/2} \sum_{s=1}^{t} \left( \hat \varepsilon_{s,n}^2 - \varepsilon_s^2 \right) - n^{-1} \sum_{s=1}^{n} \hat \varepsilon_{s,n}^2 \right|
\]
By Assumption 3.2,
\[
A_n - B_n \leq C_n \leq 2 \max_{1 \leq t \leq n} \left| n^{-1/2} \sum_{s=1}^{t} \left( \hat \varepsilon_{s,n}^2 - \varepsilon_s^2 \right) \right| = \text{op}(1). \tag{A.2}
\]
Thus, by Theorem 3.1 and the Continuous Mapping Theorem applied to the maximum, we have
\[
A_n = B_n + \text{op}(1) \xrightarrow{D} \varphi \sup_{0 \leq u \leq 1} |B_u^0|.
\]
Consider now \( \hat \varphi_n^2 = n^{-1} \sum_{t=1}^{n} \hat \varepsilon_{t,n}^4 - \left( n^{-1} \sum_{t=1}^{n} \hat \varepsilon_{t,n}^2 \right)^2 \). Further, \( n^{-1} \sum_{t=1}^{n} (\varepsilon_{t,n}^2 - \varepsilon_{t}^2) = \text{op}(1) \) for \( k = 2, 4 \) by Assumptions 3.2, 3.3. Therefore,
\[
\hat \varphi_n^2 = n^{-1} \sum_{t=1}^{n} \varepsilon_{t,n}^4 - \left( n^{-1} \sum_{t=1}^{n} \varepsilon_{t,n}^2 \right)^2 + \text{op}(1).
\]
By Theorem 3.1, under Assumption 3.1, we have \( \hat \varphi_n^2 = \varphi^2 + \text{op}(1) \).
All together, \( CUSQ_n \) converges in distribution to \( \sup_{0 \leq u \leq 1} |B_u^0| \) as desired. \( \Box \)

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). \( \Box \)
Proof of Theorem 4.1: Part I: Assumption 3.2.

1. The problem. Let \( \tilde{S}_{t, \theta} = n^{-1/2} \{ Q_t(\theta) - Q_t(0) \} \) so that \( S_{t, \hat{\theta}_n} = n^{-1/2} \sum_{s=1}^{t} \epsilon_s^2 - \epsilon_s^2 \). We show that \( S_{t, \hat{\theta}_n} = o_P(1) \) uniformly in \( 1 \leq t \leq n \). From (4.1) we have \( \tilde{S}_{t, \theta} = -2 \tilde{S}_{t, \theta} + \bar{S}_{t, \theta} \), where

\[
\tilde{S}_{t, \theta} = n^{-1/2} \sum_{s=1}^{t} \epsilon_s \nabla g_s(\theta), \quad \bar{S}_{t, \theta} = n^{-1/2} \sum_{s=1}^{t} (\nabla g_s(\theta))^2.
\]

2. Expand the martingale \( \tilde{S}_{t, \theta} \). We use a second order mean value result. To simplify the expression we introduce the notation

\[
\hat{h}_s(\theta) = N'_{n, \theta_0} g(x_s, \theta), \quad \bar{h}_s(\theta) = N'_{n, \theta_0} \hat{g}(x_s, \theta) N_{n, \theta_0},
\]

\[
\bar{\var} = N_{n, \theta_0}^{-1} (\theta - \theta_0), \quad \nabla \bar{h}_s(\theta) = N_{n, \theta_0} \{ \bar{g}(x_s, \theta) - \hat{g}(x_s, \theta_0) \} N_{n, \theta_0}.
\]

With this notation we get, for instance, that

\[
(\theta - \theta_0) \hat{g}(x_s, \theta_0) = \{ N_{n, \theta_0} (\theta - \theta_0) \} N_{n, \theta_0} \hat{g}(x_s, \theta) = \bar{\var} \hat{h}_s(\theta_0).
\]

Overall, we can expand \( \tilde{S}_{t, \hat{\theta}_n} = n^{-1/2} \sum_{s=1}^{t} \epsilon_s \nabla g_s(\hat{\theta}_n) \) as

\[
\tilde{S}_{t, \hat{\theta}_n} = n^{-1/2} \sum_{s=1}^{t} \epsilon_s \hat{h}_s(\theta_0) + \frac{1}{2} n^{-1/2} \sum_{s=1}^{t} \epsilon_s \hat{h}_s(\theta_0) \bar{\var} \hat{h}_s(\theta_0).
\]

\[
\tilde{S}_{t, \hat{\theta}_n} = n^{-1/2} \sum_{s=1}^{t} \epsilon_s \hat{h}_s(\theta_0) + \frac{1}{2} n^{-1/2} \sum_{s=1}^{t} \epsilon_s \bar{h}_s(\theta_0) \bar{\var} \hat{h}_s(\theta_0).
\]

for an intermediate point \( \theta_s \) depending on the summation limit \( t \) and \( \hat{\theta}_n \) so \( ||\theta_s - \theta_0|| \leq ||\hat{\theta}_n - \theta_0|| \). Note that the first two terms only depend on \( \hat{\theta}_n \) through the factor \( \bar{\var} \). For simplicity we write (A.3) as \( \tilde{S}_{t, \hat{\theta}_n} = \tilde{S}_{t, 1} + (\tilde{S}_{t, 2} + \tilde{S}_{t, 3})/2 \).

3. The martingale term \( \tilde{S}_{t, 1} \). The norm inequality and the bound to \( \bar{\var} \) in Assumption 4.1 (a) give

\[
|\tilde{S}_{t, 1}| \leq n^{-1/2} ||\bar{\var}|| ||\sum_{s=1}^{t} \epsilon_s \hat{h}_s(\theta_0)|| \leq o_P(n^{\delta-1/2}) ||\sum_{s=1}^{t} \epsilon_s \bar{h}_s(\theta_0)||.
\]

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

\[
\max_{n_0 \leq t \leq n} \tilde{S}_{t, 1} = o_P(n^{\delta-1/2}) + o_P(n^{\delta-1/2}) O_{a.s.} (1).
\]

By Assumption 4.2 (c), we have that \( \sum_{t=1}^{n} ||\hat{h}_t(\theta_0)||^2 = o_P(n^{1-2\delta}) \) for some \( \eta > 0 \) while \( \delta < 1/4 \).

We then get, when choosing \( 2\zeta \leq \eta/(2 - 2\delta - \eta) \),

\[
\max_{n_0 \leq t \leq n} \tilde{S}_{t, 1} = o_P(n^{\delta-1/2}) o_P(n^{(1-2\delta-\eta)(1/2+\zeta)}) + o_P(n^{\delta-1/2}) = o_P(1).
\]

4. The martingale term \( \tilde{S}_{t, 2} \). Argue as in item 3. First, the norm inequality gives

\[
|\tilde{S}_{t, 2}| \leq n^{-1/2} ||\bar{\var}|| ||\sum_{s=1}^{t} \epsilon_s \bar{h}_s(\theta_0)|| \leq o_P(n^{\delta-1/2}) ||\sum_{s=1}^{t} \epsilon_s \bar{h}_s(\theta_0)||.
\]

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

\[
\max_{n_0 \leq t \leq n} |\tilde{S}_{t, 2}| = o_P(n^{2\delta-1/2}) o_P(n^{(1-4\delta-\eta)(1/2+\zeta)}) + o_P(n^{2\delta-1/2}) = o_P(1),
\]

when \( \delta < 1/4 \) and \( \zeta > 0 \) is chosen sufficiently small.

5. The term \( \tilde{S}_{t, 3} \). Apply the norm and triangle inequalities to get

\[
|\tilde{S}_{t, 3}|| \leq ||\bar{\var}|| ||\sum_{s=1}^{t} \epsilon_s || ||\bar{h}_s(\theta_0) || - \tilde{h}_s(\theta_0)||.
\]

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

\[
|\tilde{S}_{t, 3}|| \leq ||\bar{\var}|| ||\sum_{s=1}^{t} \epsilon_s || ||\bar{h}_s(\theta_0) || - \tilde{h}_s(\theta_0)||.
\]
where \( \theta_* \) remains dependent on \( t \) and \( \hat{\theta}_n \). Apply the Hölder inequality to get

\[
|\tilde{S}_{t, \hat{\theta}_n, \beta}| \leq ||\hat{\theta}_n||^2 (n^{-1}\sum_{s=1}^{n} \varepsilon_s^2)^{1/2} \left( \sum_{s=1}^{n} ||\tilde{h}_s(\theta_s) - \tilde{h}_s(\theta_0)||^2 \right)^{1/2}.
\]

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

\[
\sum_{s=1}^{n} ||\tilde{h}_s(\theta_s) - \tilde{h}_s(\theta_0)||^2 \leq \sup_{\theta \in ||N^{-1}_{n, \theta_0}(\theta - \theta_0)|| \leq \epsilon n^\delta} \sum_{s=1}^{n} ||\tilde{h}_s(\theta) - \tilde{h}_s(\theta_0)||^2,
\]

which depends neither on \( t \) nor \( \hat{\theta}_n \). Then Assumption 4.2 (e) implies \( |\tilde{S}_{t, \hat{\theta}_n, \beta}| = o_p(n^{2\delta - 2\delta}) = o_p(1) \) uniformly in \( t \).

6. The compensator. As before \( ||N^{-1}_{n, \theta_0}(\theta - \theta_0)|| \leq \epsilon n^\delta \) on a set with large probability. On that set \( \tilde{S}_{t, \theta} \leq \sup_{\theta \in ||N^{-1}_{n, \theta_0}(\theta - \theta_0)|| \leq \epsilon n^\delta} \tilde{S}_{t, \theta} \) which is \( o_p(1) \) 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 \( \mathcal{V}_{n, \theta} = n^{-1}\sum_{t=1}^{n}[(\varepsilon_t^4 - \varepsilon_t^4)] \).
2. Some inequalities: By binomial expansion \( (\varepsilon - \nabla \varepsilon - \varepsilon)^4 = \varepsilon^4 - 4\varepsilon^3 + 6\varepsilon^2 - 4\varepsilon^3 \). Thus, by Hölder’s inequality,

\[
|\mathcal{V}_{n, \theta}| \leq n^{-1}\sum_{t=1}^{n} (\nabla g_s(\theta))^4 - 4n^{-1}\sum_{t=1}^{n} (\nabla g_s(\theta))^4 + 6n^{-1}\sum_{t=1}^{n} (\nabla g_s(\theta))^4 - 4n^{-1}\sum_{t=1}^{n} (\nabla g_s(\theta))^4 \leq 4n^{-1}\sum_{t=1}^{n} (\varepsilon_t^4 - \varepsilon_t^4).
\]

Now, \( n^{-1}\sum_{t=1}^{n} (\varepsilon_t^4) = O_p(1) \) by the martingale Law of Large Numbers and Assumption 3.1 while \( n^{-1}\sum_{t=1}^{n} (\nabla g_s(\theta))^4 = o_p(1) \) by an argument as in part I, item 6 using Assumption 4.2 (b).

Proof of Theorem 4.2. Since \( \hat{\theta}_n = \theta_0 + o(n^\delta) \text{ a.s.} \) by Assumption 4.1 (b) then Egorov’s theorem (Davidson 1994, Theorem 18.4) implies \( \forall \nu > 0 \exists t_0 \) so \( \Omega_n = \{ \sup_{t > t_0} |N^{-1}_{n, \theta_0}(\hat{\theta}_t - \theta_0)| < \nu n^\delta \} \) satisfies \( P(\Omega_n) > 1 - \nu \). On \( \Omega_n \) we bound

\[
\max_{1 \leq t \leq n} n^{-1/2} \sum_{s=1}^{t} \varepsilon_{s,t}^2 \leq n^{-1/2} \max_{1 \leq t \leq n} t \sum_{s=1}^{t} \varepsilon_{s,t}^2 + \max_{0 < t \leq n} t n^{-1/2} \sum_{s=1}^{n} \varepsilon_{s,t}^2.
\]

Since \( t_0 \) is finite, the first term vanishes. For the second term we can follow the proof of Theorem 4.1 replacing \( \varepsilon_{s,t}^2 \) by \( \varepsilon_{s,t}^2 \). When expanding in item 3 the intermediate point \( \theta_* \) 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 \( \theta_0 \) 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

\[
g(x_t, \theta) - g(x_t, \theta_0) = \nabla^T \tilde{h}_t(\theta_0) + \frac{1}{2} \nabla^T \tilde{h}_t(\theta_0) \theta + \frac{1}{2} \nabla^T \tilde{h}_t(\hat{\theta}_t) - \tilde{h}_t(\theta_0) \right] \widetilde{\theta},
\]

where \( \theta_t \) 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

\[
|g(x_t, \theta) - g(x_t, \theta_0)|^k \leq C \left\{ ||\nabla^T||k||\tilde{h}_t(\theta_0)||k + ||\nabla^T||2k||\tilde{h}_t(\theta_0)||k + ||\nabla^T||2k||\tilde{h}_t(\hat{\theta}_t) - \tilde{h}_t(\theta_0)||k \right\}
\]
In Assumption 4.2 (a,b) we only consider \(|\theta| \leq en^\delta\). Thus \(\theta_t\) is local to \(\theta_0\) so that 
\[
|\hat{\theta}_t(\theta_t) - \tilde{\theta}_t(\theta_0)|^k \leq \sup_{\theta:||N_{n,\theta}^{-1}(\theta-\theta_0)|| \leq en^\delta} |\hat{\theta}_t(\theta) - \tilde{\theta}_t(\theta_0)|^k.
\]
Then cumulate to get
\[
\sum_{t=1}^n \{g(x_t, \theta) - g(x_t, \theta_0)\}^k \leq e^k n^{2k} \sum_{t=1}^n |\hat{\theta}_t(\theta_0)|^k
\]
\[
+ e^{2k} n^{25k} \sum_{t=1}^n |\tilde{\theta}_t(\theta_0)|^k + e^{2k} n^{25k} \sup_{\theta:||N_{n,\theta}^{-1}(\theta-\theta_0)|| \leq en^\delta} |\tilde{\theta}_t(\theta) - \tilde{\theta}_t(\theta_0)|^{2k},
\]
which is \(o_p(n^{1/2})\) for \(k = 2\) and \(o_p(n)\) for \(k = 4\) due to Assumption 4.4.

\[\Box\]

**B Tables**

Table 1: DGPs: Data Generating Processes

<table>
<thead>
<tr>
<th>*</th>
<th>DGP</th>
<th>(y_t)</th>
<th>(g(x_t, \theta))</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>1</td>
<td>(1 + 0.5x_t + \varepsilon_t)</td>
<td>(\theta_1 + \theta_2x_t)</td>
</tr>
<tr>
<td>CS</td>
<td>2</td>
<td>(1 + 0.5x_t^2 + \varepsilon_t)</td>
<td>(\theta_1 + \theta_2x_t^2)</td>
</tr>
<tr>
<td>CS</td>
<td>3</td>
<td>(1 + 0.9x_t 1(v_t \leq 0) + 0.5x_t 1(v_t &gt; 0) + \varepsilon_t)</td>
<td>(\theta_1 + \theta_2x_t 1(v_t \leq 0) + \theta_3x_t 1(v_t &gt; 0))</td>
</tr>
<tr>
<td>CS</td>
<td>4</td>
<td>(1 + 0.3</td>
<td>x_t</td>
</tr>
<tr>
<td>M</td>
<td>5</td>
<td>(y_t - 1 + \varepsilon_t)</td>
<td>(\theta_1 + \theta_2</td>
</tr>
<tr>
<td>M</td>
<td>6</td>
<td>(1 + 0.9x_t 1(v_t \leq 0) + 0.5x_t 1(v_t &gt; 0) + \varepsilon_t)</td>
<td>(\theta_1 + \theta_2x_t)</td>
</tr>
<tr>
<td>M</td>
<td>7</td>
<td>(1 + 0.5x_t^2 + \varepsilon_t)</td>
<td>(\theta_1 + \theta_2x_t^2)</td>
</tr>
<tr>
<td>M</td>
<td>8</td>
<td>(1 + 0.3</td>
<td>x_t</td>
</tr>
<tr>
<td>M</td>
<td>9</td>
<td>(1 + 0.5x_t^2 + \varepsilon_t)</td>
<td>(\theta_1 + \theta_2</td>
</tr>
</tbody>
</table>

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

Table 2: Size and Power: Finite Sample Performance

<table>
<thead>
<tr>
<th>CUSQ_n</th>
<th>(x_t \sim I(0.7))</th>
<th>(x_t \sim I(1))</th>
<th>(x_t \sim I(2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>(n)</td>
<td>(n)</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>DGP</td>
<td>100 500 1000</td>
<td>100 500 1000</td>
</tr>
<tr>
<td>CS</td>
<td>1</td>
<td>0.031 0.040 0.044</td>
<td>0.032 0.040 0.044</td>
</tr>
<tr>
<td>CS</td>
<td>2</td>
<td>0.031 0.040 0.045</td>
<td>0.031 0.040 0.044</td>
</tr>
<tr>
<td>CS</td>
<td>3</td>
<td>0.030 0.041 0.043</td>
<td>0.033 0.042 0.043</td>
</tr>
<tr>
<td>CS</td>
<td>4</td>
<td>0.031 0.040 0.045</td>
<td>0.031 0.041 0.043</td>
</tr>
<tr>
<td>M</td>
<td>5</td>
<td>0.527 0.975 0.997</td>
<td>0.814 0.999 1.000</td>
</tr>
<tr>
<td>M</td>
<td>6</td>
<td>0.085 0.485 0.708</td>
<td>0.553 0.984 0.999</td>
</tr>
<tr>
<td>M</td>
<td>7</td>
<td>0.096 0.790 0.962</td>
<td>0.479 0.993 1.000</td>
</tr>
<tr>
<td>M</td>
<td>8</td>
<td>0.302 0.854 0.946</td>
<td>0.460 0.846 0.913</td>
</tr>
<tr>
<td>M</td>
<td>9</td>
<td>0.313 0.709 0.775</td>
<td>0.320 0.599 0.759</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>DGP</th>
<th>$y_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>$z_t$</td>
</tr>
<tr>
<td>R2</td>
<td>$\text{sign}(z_t)</td>
</tr>
<tr>
<td>R3</td>
<td>$\text{sign}(x_t)</td>
</tr>
<tr>
<td>R4</td>
<td>$\text{sign}(x_t)</td>
</tr>
<tr>
<td>R5</td>
<td>$\ln(1 +</td>
</tr>
<tr>
<td>R6</td>
<td>$x_t +</td>
</tr>
<tr>
<td>R7</td>
<td>$0.4x_{t1} (x_t \leq 0) + 1.8x_{t1} (x_t \geq 0) + u_t$</td>
</tr>
<tr>
<td>R8</td>
<td>$x_t + 1.8 [x_t/(1 + \exp(-x_t/\sqrt{n} - 2))] + u_t$</td>
</tr>
<tr>
<td>R9</td>
<td>$x_t + z_t + u_t$</td>
</tr>
<tr>
<td>R10</td>
<td>$\text{sign}(x_t) (</td>
</tr>
</tbody>
</table>

$z_t = z_{t-1} + w_t$ 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 $D = [1.2, .1, .3, 2, 0, .1, .2]$

Table 4: Power performance comparison with Kasparis (2008)

| $CUSQ_n$ | \hline
<table>
<thead>
<tr>
<th>$n$</th>
<th>$\text{Kasparis'} \text{best power}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td>R1</td>
<td>0.909</td>
</tr>
<tr>
<td>R2</td>
<td>0.925</td>
</tr>
<tr>
<td>R3</td>
<td>0.093</td>
</tr>
<tr>
<td>R4</td>
<td>0.349</td>
</tr>
<tr>
<td>R5</td>
<td>0.408</td>
</tr>
<tr>
<td>R6</td>
<td>0.514</td>
</tr>
<tr>
<td>R7</td>
<td>0.548</td>
</tr>
<tr>
<td>R8</td>
<td>0.340</td>
</tr>
<tr>
<td>R9</td>
<td>0.882</td>
</tr>
<tr>
<td>R10</td>
<td>0.670</td>
</tr>
</tbody>
</table>
References