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

Vanessa Berenguer-Rico* and Bent Nielsen†

16 August 2016

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

Non-linear models with non-stationary regressors are gaining increasing attention. In particular, parametric models of the form

$$y_t = g(x_t, \theta) + \varepsilon_t, \tag{1.1}$$

where x_t is a vector of possibly non-stationary regressors are of interest for economists and econometricians. On the one hand, for a univariate x_t, the specification $g(x_t, \theta) = \alpha|x_t|^\beta + \gamma$ has been considered by several economists when modelling relationships among macroeconomic time series that exhibit explosive behavior –see, for instance, the intrinsic bubbles model by Froot and Obstfeld (1991) or the hyperinflation analysis by Petrović and Mladenović (2000). On the other hand, the econometrics literature on non-linear models with non-stationary regressors has progressively advanced during the last two decades. Specifically, asymptotic theory, estimation methods and testing procedures have been developed –see for instance Park and Phillips (1999, 2001), Pötscher (2004), de Jong (2004), de Jong and Wang (2005), Berkes and Horváth (2006), Karlsen, Myklebust, and Tjøstheim(2007), Schienle (2008), Kasparis (2008, 2011), de Jong and Hu (2009), Christopeit (2009), Wang and Phillips (2009, 2012), Choi and Saikkonen (2010), Kristensen and Rahbek (2010), Berenguer-Rico and Gonzalo (2014), Chan and Wang (2015).

In statistical modelling, testing for misspecification is of primary importance. Specification tests based on the cumulated sum of residuals have a long tradition in econometrics going back to Brown, Durbin and Evans (1975). Even though these tests were originally designed to test for structural changes, they can be used more generally to test for the validity of a particular model from an omnibus testing point of view –see, for example, the cumulated sum of residuals test for cointegration of Xiao and Phillips (2002). In the context of non-linear and non-stationary regressors Kasparis (2008), Choi and Saikkonen (2010), or Berenguer-Rico and Gonzalo (2014), for instance, propose tests which are based on the cumulated sum of residuals. Given the non-stationary nature of the regressors in this framework, the asymptotic distribution of these tests depends on nuisance

*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.

quantities coming from the estimation error. The problem is easily illustrated by considering a simple linear regression model with a random walk regressor –although the same issue is present when dealing with important classes of non-linear transformations, such as homogenous functions, of non-stationary processes. In the simplest linear model with a random walk regressor, it is easy to see that under correct specification

$$\frac{1}{\sqrt{n}}\sum_{s=1}^{t}\hat{\varepsilon}_{s,n} = \frac{1}{\sqrt{n}}\sum_{s=1}^{t}\varepsilon_s + \mathrm{O}_{\mathsf{P}}(1),$$

where the $\mathrm{O}_{\mathsf{P}}(1)$ term on the right hand side is not $\mathrm{o}_{\mathsf{P}}(1)$. To overcome the difficulties that these nuisance quantities pose, tests statistics based on the cumulated sum of residuals require some kind of adjustment such as modified residuals and simulations for each specification or resampling techniques. This complicates the implementation as well as the theoretical analysis of such tests.

In this paper, we propose a simple test statistic based on the cumulated sum of squares residuals, which, under quite general assumptions, converges to a well defined distribution –the supremum of the absolute value of a Brownian Bridge for which critical values are readily available. Specifically, we show that even when the regressors in non-linear regressions are non-stationary

$$\frac{1}{\sqrt{n}}\left(\sum_{s=1}^{t}\hat{\varepsilon}^2_{s,n} - \frac{t}{n}\sum_{s=1}^{n}\hat{\varepsilon}^2_{s,n}\right) = \frac{1}{\sqrt{n}}\left(\sum_{s=1}^{t}\varepsilon^2_{s} - \frac{t}{n}\sum_{s=1}^{n}\varepsilon^2_{s}\right) + \mathrm{o}_{\mathsf{P}}(1),$$

and the Brownian Bridge result follows, for instance, under a martingale difference assumption for ε_t. We also show the Brownian Bridge result holds for a recursive version of the test statistic.

The martingale difference sequence assumption on ε_t, which is a plausible condition in several relevant empirical applications –such as the predictive regressions case– may sound too restrictive in some cases. To deal with this issue several alternative approaches are in place. One option would be to allow ε_t to follow a general linear process. The standard Brownian Bridge result would still follow under an appropriate standarization of the test. However, it is difficult to contol size of residual based specification tests under the linear process assumption since the autocorrelation structure can be arbitrarily close to a random walk behaviour, see for instance Xiao and Phillips (2002), Kasparis (2008), or Choi and Saikkonen (2010). The approach we take in this paper, however, is to think of model (1.1) as a conditional mean model where any unmodelled autocorrelation or correlation between ε_t and x_t will be regarded as misspecification. Nonetheless, the test, which is aimed at non-constancy in the variance, will not be consistent against certain types of autocorrelation in the error term. When the martingale difference specification is refined as an iid normal assumption, size can be controlled quite well in finite samples for a range of models. Therefore, the martingale difference assumption will define, in what follows, our null hypothesis of correct specification.

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). What we show in this paper is that the Brownian Bridge result holds under much more general assumptions on the regressors and the regression function. Then, we demonstrate that these assumptions are satisfied in several different scenarios dealing with linear or non-linear regression functions involving stationary or non-stationary regressors. Special emphasis is given to the following cases: (i) the linear model; (ii) the power function; and more generally (iii) integrable and asymptotically homogenous transformations of unit root processes as in Park and Phillips (2001); and (iv) non-linear error correction models as in Kristensen and Rahbek (2010). Finally, the finite sample performance of the test is studied through several Monte Carlo experiments. These simulations reveal that the test has good properties in terms of size and power.

The paper is organized as follows. In Section 2, the model and test statistics are put forward. Section 3 builds up a general framework under which the Brownian bridge result is obtained. Then, Section 4 shows that the assumptions in Section 3 are satisfied in various models of interest. In

Section 5 the performance of the test in terms of size and power is investigated through Monte Carlo experiments. Section 6 contains some concluding remarks. 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 and the non-linear regression model

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

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 $\mathcal{F}_{t-1}$-adapted, and the parameter θ is a q-vector varying in a parameter space $\Theta \subset \mathbb{R}^q$.

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

$$CUSQ_n = \frac{1}{\hat{\varphi}_n} \max_{1 \le t \le n} \left| \frac{1}{\sqrt{n}} \left(\sum_{s=1}^{t} \hat{\varepsilon}_{s,n}^2 - \frac{t}{n} \sum_{s=1}^{n} \hat{\varepsilon}_{s,n}^2 \right) \right|, \tag{2.3}$$

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

$$\hat{\varphi}_n^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{t,n}^4 - \left(\frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_{t,n}^2 \right)^2. \tag{2.4}$$

We will argue that under quite general assumptions,

$$CUSQ_n \overset{\mathsf{D}}{\to} \sup_{0 \le u \le 1} \left| \mathcal{B}_u^0 \right|, \tag{2.5}$$

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 90%, 95%, 99% quantile are 1.22, 1.36, 1.63; see Smirnov (1948). Edgerton and Wells (1994) developed response surfaces for finite sample quantiles. For the 95% critical value this is

$$1.358 - 0.670 n^{-1/2} - 0.886 n^{-1}. \tag{2.6}$$

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

$$RCUSQ_n = \frac{1}{\hat{\varphi}_n} \max_{n_0 \le t \le n} \left| \frac{1}{\sqrt{n}} \left(\sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2 - \frac{t}{n} \sum_{s=1}^{n} \hat{\varepsilon}_{s,n}^2 \right) \right|, \tag{2.7}$$

where $\hat{\varepsilon}_{s,t} = y_s - g(x_s, \hat{\theta}_t)$ and $\hat{\theta}_t$ is a recursive non-linear least squares estimator. If the sequence of estimators $\hat{\theta}_t$ converges strongly, we can show that also

$$RCUSQ_n \overset{\mathsf{D}}{\to} \sup_{0 \le u \le 1} \left| \mathcal{B}_u^0 \right|. \tag{2.8}$$

Thus, the same limiting distribution applies as in (2.5). Sohkanen (2011) developed a response surface for the finite sample 95% critical value

$$1.358(1 - 0.68 n^{-1/2} + 3.13 n^{-1} - 33.9 n^{-3/2} + 93.9 n^{-2}). \tag{2.9}$$

3 Main Results

We show that the results described in the previous section apply under very mild assumptions. We start by describing these assumptions before stating the theorems. In §4 we discuss how to check these assumptions in particular models.

The first assumption, which defines the null hypothesis of correct specification, is a martingale difference assumption to the innovations ε_t. In other words, the temporal dependence has to be modelled. Since the test is based on the square residuals a fourth moment assumption is needed when estimating the variance of the squared residuals, which is used to standardize the statistic.

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 next assumption relates to the asymptotic behaviour of $\hat{\theta}_n$. It is worth emphasizing that our intention is to apply the $CUSQ_n$ test to check the specification of models for which the asymptotic properties of $\hat{\theta}_n$ have been already analyzed. In our setup, full knowledge of the convergence rate is not necessary.

Assumption 3.2 *Let* N_{n,θ_0} *be a normalization matrix, possibly stochastic, depending on* n *and* θ_0*, so that* $N_{n,\theta_0}^{-1} = \mathrm{O}(n^\ell)$ *a.s. for some* $\ell > 0$*. 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. and* N_{n,θ_0}^{-1} *non-decreasing under (b).*

The general form of the normalization N_{n,θ_0}^{-1} allows us to consider both stationary and non-stationary regressors, although exponential regressors are ruled out by the requirement that N_{n,θ_0}^{-1} is of polynomial order. Condition $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o}_{\mathsf{P}}(n^\delta)$ allows the rescaled estimator $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0)$ to diverge slightly since $\delta < 1/4$. If $\ell \geq \delta$, then $\hat{\theta}_n$ is consistent but Assumption 3.2, which does not relate ℓ and δ, remains valid for the results that follow even if $\ell < \delta$. For the sake of the present discussion it is not necessary to find the exact convergence rate of $\hat{\theta}_n$. Nonetheless, in many situations of interest it is known that $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{O}_{\mathsf{P}}(1)$. For example, 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, see §4.1. In more general cointegrated models N_{n,θ_0}^{-1} may be block diagonal with different normalizations in different blocks, see §4.5. In non-linear models with non-stationary regressors the normalization depends on the type of regression function under consideration and may also depend on the parameter θ under which we evaluate the distributions, see §4.2, 4.3, 4.4. We use the notation θ_0 to emphasize this choice of parameter.

The next assumption concerns smoothness of the regression function. It 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. In Assumption 3.5 below we introduce slightly stronger conditions that may be easier to check in situations where the non-linear function g is known to satisfy Lipschitz conditions.

Assumption 3.3 *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 3.2 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) $\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq n^\delta\epsilon} \sum_{t=1}^n ||N_{n,\theta_0}'\ddot{g}(x_t, \theta)N_{n,\theta_0}||^2 = \mathrm{O}_{\mathsf{P}}(n^{-4\delta})$.

Finally, we need some technical conditions to ensure invertibility of the matrix of squared first derivatives of the non-linear function g.

Assumption 3.4 *Suppose* $\inf[n : \sum_{t=1}^{n}\{\dot{g}(x_t,\theta_0)\}\{\dot{g}(x_t,\theta_0)\}'$ *is invertible*$] < \infty$ *a.s. with the convention that the empty set has infinite infimum.*

We can now show that the cumulated sums of squares statistic satisfies a standard asymptotic result. Note, that only weak consistency is needed for the estimator, see Assumption 3.2(a).

Theorem 3.1 *If Assumptions 3.1,3.2(a),3.3,3.4 are satisfied then* $CUSQ_n \overset{\mathsf{D}}{\to} \sup_{0\leq u\leq 1}\left|\mathcal{B}_u^0\right|$.

For the recursive cumulated sum of squares statistic the estimator of θ is computed recursively, namely $\hat{\theta}_t$. Therefore, we require uniformity properties over t for the sequence of recursive estimators. If the full sample estimator, $\hat{\theta}_n$, is strongly consistent, then we can get that uniformity in probability for $\hat{\theta}_t$ from Egorov's Theorem, see Davidson (1994, Theorem 18.4) –see also Nielsen and Sohkanen (2011). This then leads to the following result.

Theorem 3.2 *If Assumptions 3.1,3.2(b),3.3,3.4 are satisfied then* $RCUSQ_n \overset{\mathsf{D}}{\to} \sup_{0\leq u\leq 1}\left|\mathcal{B}_u^0\right|$.

The main argument in the proof is to show that $n^{-1/2}\sum_{s=1}^{n}(\hat{\varepsilon}_{s,n}^2 - \varepsilon_s^2) = \mathrm{o}_{\mathsf{P}}(1)$. To do so, 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

$$n^{-1/2}\sum_{s=1}^{n}(\hat{\varepsilon}_{s,n}^2 - \varepsilon_s^2) = -2n^{-1/2}\sum_{s=1}^{n}\varepsilon_s\nabla g(x_s,\hat{\theta}_n) + n^{-1/2}\sum_{s=1}^{n}\{\nabla g(x_s,\hat{\theta}_n)\}^2. \tag{3.1}$$

Due to Assumption 3.2 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 (3.1) vanish uniformly over the local region. These sums are a martingale and its compensator. Now, the compensator vanishes under Assumption 3.3(a). Jennrich (1969, Theorem 6), for instance, 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 3.3(c,d). For this we exploit Lemma 1 of Lai and Wei (1982). The conditions (c,d) 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). This is because the $CUSQ$ test is concerned with variance estimators rather than regression estimators. Finally, Assumption 3.3(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 3.3 to conditions (c,d). 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 3.3.

Assumption 3.5 *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 3.2 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) $\sum_{t=1}^{n}\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq n^{\delta}\epsilon}||N_{n,\theta_0}'\ddot{g}(x_t,\theta)N_{n,\theta_0}||^k = \mathrm{o}_{\mathsf{P}}(n^{(k-2)/2-2k\delta})$.

Theorem 3.3 *Assumption 3.5 implies Assumption 3.3.*

4 Analysis of some particular models

In this section, we illustrate the practical usage of the general assumptions through particular non-linear models that have been discussed in the literature. For all these models we assume Assumption 3.2 has been dealt with elsewhere. Thus, we take the appropriate normalization of the estimators, N_{n,θ_0}^{-1}, and consistency of $\hat{\theta}_n$ as given. The difficulty is therefore to establish the smoothness Assumption 3.5. We will show that this assumption is rather mild.

4.1 The vector autoregressive model

We start by considering a linear model $g(x_t, \theta) = \theta' x_t$, where the joint behaviour of y_t and x_t is described by the vector autoregressive setup in Nielsen 2005, N05 henceforth. Thus, consider a vector autoregression $z_t = \sum_{j=1}^{k} A_j z_{t-1} + \mu d_t + \eta_t$ with non-explosive roots, polynomial deterministic terms d_t, and a martingale difference innovation vector η_t satisfying conditions corresponding to Assumption 3.1. We then get the linear model when y_t is a coordinate of z_t while x_t is the companion vector $(z'_{t-1}, \ldots, z'_{t-k}, d_t)$.

Assumption 3.2 requires weak consistency in part (a) and strong consistency in part (b). For the weak consistency we can work directly with the least squares estimator using a deterministic normalization. For the strong consistency argument it is more convenient to work with a self-normalized estimator, which we achieve by using a stochastic normalization. In subsequent sections we turn to non-linear least squares estimators where only the weak consistency argument has been developed.

Deterministic normalization N_{n,θ_0}. Suppose x_t is univariate. Let $N_{n,\theta_0} = n^{-1/2}$ if x_t is stationary and $N_{n,\theta_0} = n^{-1}$ if x_t is a random walk. With these rates $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0)$ converges in distribution as required in Assumption 3.2(a). Assumption 3.5 reduces to showing $\mathcal{S}_k = \sum_{t=1}^{n} ||N'_{n,\theta_0} x_t||^k = \mathrm{o_P}(n^{k/4 - k\delta})$ for $k = 2, 4$. Indeed, we find $\mathcal{S}_2 = \mathrm{O_P}(1)$ and $\mathcal{S}_4 = \mathrm{O_P}(n^{-1})$. We also note that Assumption 3.4 concerning the invertibility of $\sum_{t=1}^{n}\{\dot{g}(x_t, \theta_0)\}\{\dot{g}(x_t, \theta_0)\}'$ is satisfied by Lemma 8.2 of N05.

Stochastic normalization N_{n,θ_0}. We choose the stochastic normalization matrix N_{n,θ_0} as the square root of the matrix $(\sum_{t=1}^{n} x_t x'_t)^{-1}$ so that $(\sum_{t=1}^{n} x_t x'_t)^{-1} = N_{n,\theta_0} N'_{n,\theta_0}$. As a consequence $N'_{n,\theta_0}(\sum_{t=1}^{n} x_t x'_t) N_{n,\theta_0}$ is the identity. This implies that $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = (\sum_{t=1}^{n} x_t x'_t)^{-1/2} \sum_{t=1}^{n} x_t \varepsilon_t$ is a self-normalized statistic. Such statistics are considered in Lai and Wei (1982), which forms the basis for the results in N05. The strong consistency Assumption 3.2(b) is satisfied due to Theorem 2.4 of N05. Their result shows that $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o}(n^{\delta})$ *a.s.* for some $0 < \delta < 1/4$. This is an example where it is useful to allow $\delta > 0$ in Assumption 3.2. We now need to argue that the normalization is of polynomial order. This is shown in Theorem 7.1 of N05. We need to check the Assumption 3.5 with the new normalization. For part (a) we note $\sum_{t=1}^{n} ||N'_{n,\theta_0} x_t||^k = \mathrm{O}(\log n)$ *a.s.*, for $k = 2, 4$, by Lemma 8.6 of N05. Part (b) is trivially satisfied since $\ddot{g} = 0$.

The asymptotic results for the cumulated sum of squares tests also apply for explosive processes, albeit with a different proof, see Nielsen and Sohkanen (2011).

4.2 A simple power function model

We consider a simple power function model as a first illustration of the workings of the assumptions for non-linear models. The model equation is then

$$y_t = |x_t|^{\theta} + \varepsilon_t \qquad t = 1, \ldots, n, \tag{4.1}$$

with $\theta \in \Theta \in (0, \infty)$ and where x_t is univariate and either stationary or a random walk. For both cases we need the non-linear functions:

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

These functions are continuous in x when $\theta > 0$ and $|x| > 0$. It is important to notice that even though $\log|x|$ is not continuous at zero, $\dot g$ and $\ddot g$ can be extended continuously to all $x \in \mathbb{R}$, by setting $\dot g(0,\theta) = \ddot g(0,\theta) = 0$, because the power function dominates the logarithm at the origin. Hence g, $\dot g$ and $\ddot g$ are continuous in $x \in \mathbb{R}$ for $\theta > 0$.

Next, we discuss the stationary and random walk cases separately.

The stationary regressor case. Suppose x_t is (strictly) stationary with

$$\mathsf{E}(|x_t|^{8\theta_0}\log^8|x_t|) < \infty. \tag{4.2}$$

We show that the smoothness Assumption 3.5 holds. This requires knowledge of the normalization N_{n,θ_0}^{-1}. Results on non-linear least squares estimation in the case of stationary and ergodic regressors imply $N_{n,\theta_0}^{-1} = \sqrt{n}$, see, for instance, Wooldridge (1994) (Theorem 4.4) and the references therein in relation to the stationary regressors case or Pötscher and Prucha (1997). These arguments involves proving first that $\hat\theta_n$ is tight or assuming a compact parameter space, then showing consistency, and then finding some information on the consistency rate.

For Assumption 3.5 (a) we write, for $k = 2, 4$,

$$\mathcal{S}_k = \sum_{t=1}^n |N_{n,\theta_0}\dot g(x_t,\theta_0)|^k = \frac{1}{n^{k/2}}\sum_{t=1}^n |x_t|^{k\theta_0}\log^k|x_t|.$$

By (strict) stationarity and the moment condition (4.2), Theorems 3.5.3, 3.5.7 of Stout (1974) imply that $\mathcal{S}_k = \mathrm{O}(n^{1-k/2}) = \mathrm{O}(1)$ *a.s.*

Assumption 3.5 (b) involves a supremum over θ so $n^{1/2}|\theta-\theta_0| \le n^{\delta}\epsilon$. It suffices to take supremum over a larger set $|\theta-\theta_0| \le \nu$ for some ν so $0 < \nu < \theta_0$. We then apply the inequality $|a+b|^k \le C(|a|^k + |b|^k)$ to get

$$N_{n,\theta_0}^{2k}\sum_{t=1}^n \sup_{\theta:|\theta-\theta_0|\le\nu}|\ddot g(x_t,\theta)|^k \le \frac{C}{n^k}\sum_{t=1}^n \sup_{\theta:|\theta-\theta_0|\le\nu}|\ddot g(x_t,\theta)-\ddot g(x_t,\theta_0)|^k + \frac{C}{n^k}\sum_{t=1}^n|\ddot g(x_t,\theta_0)|^k. \tag{4.3}$$

Again, by the strict stationarity assumption and the moment condition (4.2), the second term in (4.3) is $\mathrm{O}_{\mathsf{P}}(n^{1-k})$ by Theorems 3.5.3, 3.5.7 of Stout (1974). For the first term in (4.3) we apply a Lipschitz argument. The second derivative of g satisfies

$$|\ddot g(x_t,\theta)-\ddot g(x_t,\theta_0)| = ||x|^{\theta-\theta_0}-1||x|^{\theta_0}\log^2|x| \le (|x|^{\nu}+1)|x|^{\theta_0}\log^2|x|,$$

for all θ so $|\theta-\theta_0| \le \nu$ by the triangle inequality. Thus, the first term in (4.3) is bounded by

$$\frac{C}{n^k}\sum_{t=1}^n\{(|x_t|^{\nu}+1)|x_t|^{\theta_0}\log^2|x_t|\}^k, \tag{4.4}$$

so that we can apply Theorems 3.5.3, 3.5.7 of Stout (1974) again to show that this is $\mathrm{O}_{\mathsf{P}}(n^{1-k}) = \mathrm{o}_{\mathsf{P}}(1)$ for $k = 2, 4$, using (strict) stationarity and the moment condition (4.2).

The random walk regressor case. Again our objective is to show that the smoothness Assumption 3.5 holds, which requires knowledge of N_{n,θ_0}^{-1}. Theorem 5.3 of Park and Phillips (2001), see also their Example 5.1, gives $N_{n,\theta_0}^{-1} = n^{(1+\theta_0)/2}\log n^{1/2}$ where $\theta_0 > 0$ as before. De Jong and Hu (2011) give an extended discussion of the necessary assumptions for the asymptotic theory of $\hat\theta_n$. For the present purpose we take the convergence rate as given.

For Assumption 3.5 (a) write

$$\mathcal{S}_k = \sum_{t=1}^n |N_{n,\theta_0}\dot g(x_t,\theta_0)|^k = \frac{1}{n^{(1+\theta_0)k/2}\log^k n^{1/2}}\sum_{t=1}^n |x_t|^{k\theta_0}\log^k|x_t|.$$

Noting that $\log|x| = \log|x/n^{1/2}| + \log n^{1/2}$ this can be rewritten as

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

where we find the last bound as follows. 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 $\int_0^1 |y_t|^{2\theta_0}\log|y_t|dt$ and $\int_0^1 |y_t|^{2\theta_0}dt$ are continuous mappings from $D[0,1]$ to $\mathbb{R}$ when $\theta_0 > 0$ as assumed here. The Continuous Mapping Theorem, see Billingsley (1999, Theorem 2.7) then shows that the normalized sum converges in distribution.

For Assumption 3.5 (b) we expand $\sum_{t=1}^n \sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\le n^{\delta}\epsilon} ||N_{n,\theta_0}' \ddot{g}(x_t,\theta) N_{n,\theta_0}||^k$ as in (4.3) with $N_{n,\theta_0}^{-1} = n^{(1+\theta_0)/2}\log n^{1/2}$. The second term in the decomposition can be bounded using the Continuous Mapping Theorem in a similar fashion as we did when dealing with Assumption 3.5 (a). That is,

$$\frac{C}{n^{(1+\theta_0)k}\log^{2k} n^{1/2}} \sum_{t=1}^{n} |\ddot{g}(x_t,\theta_0)|^k = \frac{1}{n^{k(1+\theta_0/2)}} \sum_{t=1}^{n} |x_t/n^{1/2}|^{k\theta_0} \left(\frac{\log|x_t/n^{1/2}|}{\log n^{1/2}} + 1 \right)^{2k} = \mathrm{o_P}(1),$$

for $k = 2, 4$ by the continuous mapping theorem and the fact that $\theta_0 > 0$. The first term in the decomposition can be bounded using (4.4) and then applying again the continuous mapping –note that $0 < \nu < \theta_0$.

4.3 Power function model with intercept

We now augment the simple power function model with a regression coefficient and an intercept giving the model equation

$$y_t = \alpha|x_t|^\beta + \gamma + \varepsilon_t \qquad t = 1, \ldots, n, \tag{4.5}$$

with $\theta = (\alpha, \beta, \gamma)' \in \Theta = (0,\infty)^2 \times \mathbb{R}$ and where x_t is univariate and stationary or a random walk. This model is the basis for the empirical illustration in §6. The analysis of this model is by and large a generalization of the analysis of the simple power function model. The model also generalizes that in Example 5.1 of Park and Phillips (2001), PP01 henceforth. PP01 are concerned with proving consistency and we follow them in constraining the parameters α, β to be positive. In the empirical illustration α is always negative. Constraining α to be negative rather than positive should not change any argument, as one can scale the data by minus unity. But in some examples α or β is close to zero in which case further analysis is needed. This is, however, beyond the scope of the paper and left for future research –see also Shi and Phillips (2012). In any case, we now consider the non-linear function $g(x,\theta) = \alpha|x|^\beta + \gamma$ with derivatives:

$$\dot{g}(x,\theta) = \begin{pmatrix} |x|^\beta \\ \alpha|x|^\beta \log|x| \\ 1 \end{pmatrix}, \quad \ddot{g}(x,\theta) = \begin{pmatrix} 0 & |x|^\beta\log|x| & 0 \\ |x|^\beta\log|x| & \alpha|x|^\beta\log^2|x| & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

These functions have a version that is continuous in x when $\beta > 0$ and $|x| \ge 0$ since the power function dominates the logarithm at the origin.

The stationary regressor case. Suppose x_t is (strictly) stationary, let $N_{n,\theta_0}^{-1} = \sqrt{n}$ and assume $\hat{\theta}_n$ is $\sqrt{n}$-consistent and the moment condition (4.2) holds. Our main concern is with Assumption 3.5. The argument is essentially the same as in §4.2. For instance, for Assumption 3.5 (a) we bound the norm by the maximum of element-wise norms to get $||\dot{g}(x_t,\theta_0)|| \le |x_t|^\beta(1+\alpha\log|x_t|)+1$. Using the inequality $|y+z|^k \le C(|y|^k + |z|^k)$ we see that $\mathsf{E}||\dot{g}(x_t,\theta_0)||^k$ is finite whenever $\mathsf{E}(|x_t|^\beta\log|x_t|)^k$ is finite. We can therefore proceed exactly as in §4.2.

The random walk regressor case. We now generalize Example 5.1 of PP01. They analyze the model without intercept, while essentially assuming consistency. Following their argument we can decompose $\dot{g}(\lambda x,\theta) = \dot{\kappa}(\lambda,\theta)H(x,\beta)$ for $\lambda > 0$ where

$$\dot{\kappa}(\lambda,\theta) = \begin{pmatrix} \lambda^{\beta} & 0 & 0 \\ \alpha\lambda^{\beta}\log\lambda & \alpha\lambda^{\beta} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad h(x,\beta) = \begin{pmatrix} |x|^{\beta} \\ |x|^{\beta}\log|x| \\ 1 \end{pmatrix},$$

and h is H-regular according to PP01. Following PP01 we choose $N_{n,\theta}^{-1} = n^{1/2}\dot{\kappa}(n^{1/2},\theta)'$ and get that $N_{n,\theta_0}^{-1}(\hat{\theta}-\theta_0)$ converges in distribution. Once again, that result assumes consistency. Our main concern is with Assumption 3.5. The argument is essentially the same as in §4.2. For Assumption 3.5 (a) we find

$$\begin{aligned} N_{n,\theta_0}\dot{g}(x_t,\theta) &= \begin{pmatrix} n^{-(\beta_0+1)/2} & 0 & 0 \\ -n^{-(\beta_0+1)/2}\log n^{1/2} & \alpha^{-1}n^{-(\beta_0+1)/2} & 0 \\ 0 & 0 & n^{-1/2} \end{pmatrix} \begin{pmatrix} |x_t|^{\beta} \\ \alpha|x_t|^{\beta}\log|x_t| \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} n^{-(\beta_0+1)/2}|x_t|^{\beta} \\ n^{-(\beta_0+1)/2}|x_t|^{\beta}\log|n^{-1/2}x_t| \\ n^{-1/2} \end{pmatrix}. \end{aligned}$$

Bounding the Euclidean vector norm by the maximum of the element-wise norms and using the inequality $|a+b+c|^k \le C(|a|^k + |b|^k + |c|^k)$ for some $C > 0$ shows, for $k = 2, 4$,

$$\sum_{t=1}^{T} ||N_{n,\theta_0}\dot{g}(x_t,\theta_0)||^k \le \frac{C}{n^{k/2}} \sum_{t=1}^{T} \{|n^{-1/2}x_t|^{k\beta_0} + \left(|n^{-1/2}x_t|^{\beta_0}\log|n^{-1/2}x_t|\right)^k + 1\} = \mathrm{O}_{\mathsf{P}}(1),$$

where the bound is found as in §4.2. For Assumption 3.5 (b) we write $\ddot{g}(x_t,\theta) = \ddot{g}(x_t,\theta_0) + \{\ddot{g}(x_t,\theta) - \ddot{g}(x_t,\theta_0)\}$ and expand each element of the matrix $N_{n,\theta_0}\ddot{g}(x_t,\theta)N_{n,\theta_0}'$ as above. We then bound the norm by the trace and apply arguments as in §4.2.

4.4 Non-linear models with random walk regressors

Park and Phillips (2001), PP01 henceforth, consider a triangular system with a univariate random walk regressor:

$$y_t = g(x_t,\theta) + \varepsilon_t \qquad t = 1,\ldots,n, \tag{4.6}$$

$$x_t = x_{t-1} + v_t, \tag{4.7}$$

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. The results build on work in Park and Phillips (1999), PP99 henceforth. For recent developments see Chan and Wang (2015).

The class of *integrable functions* is defined as follows. Let $h(x,\theta)$ be either of $g, \dot{g}, \ddot{g}$ where x is univariate and differentiation is with respect to θ as before. We require that for each $\theta \in \Theta$, h is bounded, so that $\sup_{x\in\mathbb{R}}|h(x,\theta)| < \infty$, and integrable, so that $\int_{\mathbb{R}}|h(x,\theta)|dx < \infty$. Moreover, following Definition 3.3 of PP01, for all $\theta_0 \in \Theta$ there exists a ball N_0 so $\theta \in N_0 \subset \Theta$ and a function $T : \mathbb{R} \mapsto \mathbb{R}_+$ which is bounded, so that $\sup_{x\in\mathbb{R}} T(x) < \infty$, and integrable, so that $\int_{\mathbb{R}} T(x)dx < \infty$. Further, it must hold that $|h(x,\theta) - h(x,\theta_0)| \le ||\theta-\theta_0||T(x)$ and $|h(x,\theta) - h(y,\theta)| \le c|x-y|^3$.

The class of integrable functions includes transformations $g(x_t,\theta)$ such as $1/(1+\theta x^2)$ and $e^{-\theta x^2}$. These functions are integrable over $x \in \mathbb{R}$. We can check the integrability conditions through differentiation and the mean-value theorem.

As before, we would like to show that Assumption 3.5 holds, which requires knowledge of N_{n,θ_0}^{-1}. Theorem 5.1 of PP01 shows that $n^{1/4}(\hat{\theta}_n - \theta_0)$ converges in distribution in the integrable functions case –see also de Jong and Hu (2011). Hence, we choose the normalization $N_{n,\theta_0}^{-1} = n^{1/4}$.

Assumption 3.5 (a) bounds sums of the first derivative. Hence, for $k = 2, 4$ we get

$$\sum_{t=1}^{n} ||\frac{1}{n^{1/4}}\dot{g}(x_t, \theta_0)||^k = \sup_{x \in \mathbb{R}} ||\dot{g}(x_t, \theta)||^{k-1}\{\frac{1}{n^{k/4}}\sum_{t=1}^{n} ||\dot{g}(x_t, \theta_0)||\} = \mathrm{O}_{\mathsf{P}}(n^{(2-k)/4}) = \mathrm{O}_{\mathsf{P}}(1),$$

since $\dot{g}$ is bounded, while the last sum converges by Theorem 5.1 of PP99, see also Lemma A.6 of PP01.

Assumption 3.5 (b) concerns the second derivative. Decompose

$$\sum_{t=1}^{n} ||\frac{1}{n^{1/2}}\ddot{g}(x_t, \theta)||^k \leq \sum_{t=1}^{n} ||\frac{1}{n^{1/2}}\ddot{g}(x_t, \theta_0)||^k + \sum_{t=1}^{n} ||\frac{1}{n^{1/2}}\{\ddot{g}(x_t, \theta) - \ddot{g}(x_t, \theta_0)\}||^k, \tag{4.8}$$

The first term is $\mathrm{O}_{\mathsf{P}}(n^{(2-k)/4-k/4}) = \mathrm{O}_{\mathsf{P}}(n^{(1-k)/2}) = \mathrm{O}_{\mathsf{P}}(1)$ by the argument used when dealing with Assumption 3.5 (a). Then, by the local Lipschitz property of $\ddot{g}$ we get $\sup_{\theta \in N_0} ||\ddot{g}(x_t, \theta) - \ddot{g}(x_t, \theta_0)|| \leq \sup_{\theta \in N_0} ||\theta - \theta_0|| T(x) \leq CT(x)$. We can then apply Theorem 5.1 of PP99 to the integrable function $T(x)$ so that the second term in (4.8) is of the same order of magnitude as the first term.

The class of *asymptotically homogenous functions* has a related setup. Let $h(x, \theta)$ be either of $g, \dot{g}, \ddot{g}$ where x is univariate. Suppose h satisfies

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

where κ is a normalization, R is a remainder term, which is of lower order than κ, while H is 'regular' for each θ, so that $n^{-1}\sum_{t=1}^{n} h(n^{-1/2}x_t, \theta)$ and $n^{-1/2}\sum_{t=1}^{n} h(n^{-1/2}x_t, \theta)\varepsilon_t$ converge in distribution, see Lemma A.2 of PP01 and Theorem 3.3 of PP99. The conditions can be expressed as a local integrability conditions, see Pötscher (2004). We notice that if h_1, h_2 are regular then $h_1 + h_2$ and $h_1 h_2$ are regular, see Lemma A.1 of PP01. As a consequence, if h is a vector of regular functions then $h'h$ is regular and $(h'h)^2$ are regular. The class of asymptotically homogeneous functions include the power function in §4.2 as well as logistic, threshold-like or logarithmic transformations.

Some further assumptions are needed for the distribution theory for $\hat{\theta}_n$ in PP01. These assumptions are listed in Theorem 5.3 of PP01 and include that for all $\bar{s} > 0$ and some $\delta > 0$ then

$$\mathcal{N}(\lambda) = ||\{\dot{\kappa}(\lambda, \theta_0)\}^{-1} \sup_{|s| < \bar{s}} \sup_{\theta : ||\dot{\kappa}'(\lambda, \theta_0)(\theta - \theta_0)|| \leq \lambda^{\delta - 1}} ||\ddot{g}(\lambda s, \theta)||_{element} \{\dot{\kappa}'(\lambda, \theta_0)\}^{-1}|| = \mathrm{o}(\lambda^{1-\delta}), \tag{4.9}$$

where $|| \cdot ||_{element}$ is the matrix of absolute values –see also de Jong and Hu (2011).

As before, we would like to show that Assumption 3.5 holds, which requires knowledge of N_{n,θ_0}^{-1}. Theorem 5.3 of PP01 shows that $n^{1/2}\dot{\kappa}(n^{1/2}, \theta_0)'(\hat{\theta} - \theta_0)$ converges in distribution, where $\dot{\kappa}$ is the normalization of the derivative $\dot{g}$. Thus, we choose $N_{n,\theta_0}^{-1} = n^{1/2}\dot{\kappa}\left(n^{1/2}, \theta_0\right)'$. For instance, in the power function model (4.1) with random walk regressor we have $\dot{\kappa}(n^{1/2}, \theta_0) = n^{\theta_0/2}\log n^{1/2}$, see Example 5.1 of PP01.

For Assumption 3.5 (a) note that $\{\dot{\kappa}(\lambda, \theta_0)\}^{-1}\dot{g}(x_t, \theta_0)$ is regular up to a remainder term, which is dominated and therefore ignored in the following calculation. Using the properties for regular functions outlined above the leading term of $||\{\dot{\kappa}(\lambda, \theta_0)\}^{-1}\dot{g}(x_t, \theta)||^k$ is regular for $k = 2, 4$. Thus, by the summation property for regular functions,

$$\sum_{t=1}^{n} ||\frac{1}{n^{1/2}}\{\dot{\kappa}(\lambda, \theta_0)\}^{-1}\dot{g}(x_t, \theta_0)||^k = \mathrm{O}_{\mathsf{P}}(n^{1-k/2}),$$

see PP01 Theorem 3.3.

Assumption 3.5 (b) is now satisfied as follows. The paths of the random walk x_t are of order $\mathsf{O_P}(n^{1/2})$. Thus, by (4.9), on a set with large probability, there exists an $\bar{s}$ so that, for $k = 2, 4$, then

$$\sum_{t=1}^{n} ||\frac{1}{n}\{\dot{\kappa}(\lambda, \theta_0)\}^{-1}\ddot{g}(x_t, \theta)\{\dot{\kappa}'(\lambda, \theta_0)\}^{-1}||^k \leq n^{1-k}\{\mathcal{N}(n^{1/2})\}^k + \mathsf{o_P}(1) = \mathsf{o_P}(1).$$

4.5 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 $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathsf{O_P}(1)$ that is required in our Assumption 3.2. 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 3.5 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 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. When studying the size of the test a (possibly non-linear) correctly specified model with iid innovations and a highly persistent regressor is considered. For the power, and given the emphasis on the non-linear model in previous sections, we focus here on functional form misspecifications. Specifically, 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 4 focus on stationary and random walk models the theory does extend to other types of non-stationarity, 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. Non-reported simulations indicate that the size distortion can be removed almost entirely by applying the correction (2.6) due to Edgerton and Wells (1994).

Next, we turn to the power of the test. Before putting forward the results obtained in our simulation exercises, it will be worth mentioning that results from Deng and Perron (2008) and Kasparis (2008) suggest that the $CUSQ$ test diverges under the alternative hypothesis of functional form misspecification of an asymptotically homogeneous form and that the rate of divergence is of order $n^{1/2}$. Indeed, if the true model is say $y_t = f(x_t, \theta_0) + \varepsilon_t$ and we estimate $\hat{y}_t = g(x_t, \hat{\theta}_n)$, then $\hat{\varepsilon}_{t,n} = \varepsilon_t - \{g(x_t, \hat{\theta}_n) - f(x_t, \theta_0)\}$. Recalling the definition of the statistic $CUSQ_n$ in (2.3) and the standarization in (2.4) it can be seen that the cumulated sum of the discrepancy $\{g(x_t, \hat{\theta}_n) - f(x_t, \theta_0)\}^2$ will dominate in the numerator, while the sum of $\{g(x_t, \hat{\theta}_n) - f(x_t, \theta_0)\}^4$ will dominate the denominator of the statistic. In a context of homogeneous transformation of random walks, see Kasparis (2008) for instance, this provides a heuristic argument for the consistency of the $CUSQ_n$ test. We explore this heuristic argument through simulations in what follows.

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 parameter instabilities conducted by McCabe and Harrison (1980), Ploberger and Krämer (1990), Deng and Perron (2008), or Turner (2010). It is worth emphasizing that given the omnibus nature of the $CUSQ$ test, the statistic will have power to detect other types of departures from the null hypothesis other than functional form misspecification. For instance, simulations not reported here show that the $CUSQ$ statistic has power to detect autocorrelated errors as well.

It is also worth mentioning that the $CUSQ$ also has power to detect some misspecifications involving integrable functions of persistent processes. As an example consider the data generating process $y_t = \theta_1/(1+\theta_2 x_t^2)+\varepsilon_t$, while the regression model is polynomial. Simulations not reported here show that power arises as long as the signal from the integrable function component $\theta_1/(1+\theta_2 x_t^2)$ 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.

6 Empirical Illustration

We apply the CUSQ tests to data from the Yugoslavian hyperinflation. The data are taken from Petrović and Mladenović (2000) and are available from the online data archive of Journal of Money, Credit and Banking, see also Petrović and Vujošević (1996), Engsted (1998), Petrović, Bogetić and Vujošević (1999) and Nielsen (2008, 2010). We consider monthly series of changes in the log Dmark exchange rate, $\Delta s_t = \Delta \log S_t$, and of log real narrow money deflated by the exchange rate, $m_t - s_t = \log(M_t/S_t)$, for the period 1990:12–1994:1. The series are shown in Figure 1 (a, b). It is evident that real money, $m_t - s_t$, is falling in a random walk-like way, while the changes in exchange rates, Δs_t grow in a near exponential fashion. This results in a non-linear relationship as seen in the cross-plot of $m_t - s_t$ versus Δs_t in panel (c).

The institutional background is that the Federal Republic of Yugoslavia was falling apart in 1991 and the civil war started. A United Nations embargo was introduced in 1992:5 and the inflation started. The hyperinflation ended in 1994:1 after prices had risen by a factor of 1.6×10^{21} over 24 months.

Petrović and Mladenović (2000), PM henceforth, used a variable semi-elasticity schedule. They first formulate this schedule in a static model and subsequently they embed it in a dynamic model. In the following we consider the static model in order to focus on the choice of functional form. The schedule is captured by a power function model of the form

$$m_t - s_t = \alpha|\Delta s_t|^\beta + \gamma + \varepsilon_t. \tag{6.1}$$

The idea is that for $0 < \beta < 1$ then exponential features of Δs_t get dampened through the transformation $|\Delta s_t|^\beta$, noting that $\Delta s_t > 0$ throughout the sample. Hence, the analysis in PM involves a non-linear model with a constant and a non-stationary regressor. This model was analyzed in §4.3 –see also §5 for the finite sample performance of the $CUSQ$ test when applied to a power function model with a constant term.

We get a linear model when imposing $\beta = 1$. We fitted the models using both Matlab (2014) and OxMetrics (Doornik and Hendry, 2013). Estimates are reported in the first column of Table 5. The linear model is rejected by the $RCUSQ$ test, albeit for the $CUSQ$ the decision is marginal at a 5% level.

The estimated power function is reported in the second column of Table 5. This model is not rejected by neither the $CUSQ$ nor the $RCUSQ$ test. Our estimate of β is $\hat{\beta} = 0.128$. Using that estimation we find that $|\Delta s_t|^{0.128}$ is increasing in a random walk-like way as shown in Figure 1 (d). Thereby we achieve a balanced regression between real money $m_t - s_t$ and the transformed exchange rate depreciation $|\Delta s_t|^{\beta}$.

The probabilistic interpretation of the power function model (6.1) is not fully developed as yet. The intention of the model is to stabilize the regressor Δs_t so that we get a balanced regression between the dependent variable $m_t - s_t$ and the transformed variable $|\Delta s_t|^{\beta}$. This clearly works in some empirical sense, but the economic theory remains somewhat vague as in the varying semi-elasticity interpretation of PM. In the power function model we apply the econometric trick of transforming the regressor using a parametrized function in order to balance the regressand and the regressor. In many situations this trick may be the best that can be done to achieve balancedness. However, in the case of hyperinflation we can find an alternative transformation that does not depend on parameters and which has a neat economic interpretation.

The purpose of regression (6.1) is to fit a money demand equation under hyperinflation. The real money stock is steadily falling through the hyperinflation because there is an increasing cost to holding money. The real money stock can be thought of as a co-explosive relation between money and exchange rates, which are both increasing explosively during hyper-inflation, see Nielsen (2008, 2010). The cost of holding money can be measured through $d_t = (S_t - S_{t-1})/S_t$ which tells how much we loose in Dmark terms by waiting one period. With modest inflation we have $d_t \approx \Delta s_t$, whereas in extreme hyperinflation, where Δs_t explodes, the depreciation cost d_t is bounded by unity. Nielsen (2008) used this transformation to analyze the Yugoslavia data and found that $m_t - s_t$ and d_t have random walk-like behaviour and can be analyzed using a cointegrated vector regression where the cointegrating relation between $m_t - s_t$ and d_t can be interpreted as money demand. The depreciation measures d_t and $|\Delta s_t|^{0.128}$ have a rather similar appearance apart from a location-scale transformation. This analysis effectively replaces $|\Delta s_t|^{\beta}$ in (6.1) with d_t. Indeed, that regression will give nearly the same fit as seen in Table 5.

The power function model remains a useful econometric tool. The linear model between $m_t - s_t$ and Δs_t is not balanced, whereas the power function model used by PM fares much better. In fact, in econometric terms it works just as well as the model involving d_t, and in this case the PM study inspired the development of the cost of holding money measure.

7 Concluding Remarks

We have shown that by using the cumulated sum of squares residuals we get rid of the nuisance quantities that show up in tests based on the cumulated sum of residuals in a non-stationary context. The cumulated sum of squares test statistic has a well defined limiting distribution, for which asymptotic critical values and finite sample corrections are readily available. The test has good power against a variety of non-linear misspecifications within a non-stationary environment.

8 Acknowledgements

Comments from the referees are gratefully acknowledged.

A Appendix: Proofs

A.1 High Level Result

We first prove a set of high level results for the $CUSQ_n$ statistics where we make assumptions directly on the squared residuals. When proving the main theorems, we then need to check those assumptions. For an autoregressive model it is possible to demonstrate those assumptions directly under the martingale difference Assumption 3.1, see Lemma 4.2 and Theorem 4.5 of Nielsen and Sohkanen (2011). For non-linear models we need to formulate the intermediate level assumptions stated in Assumptions 3.2, 3.3 and 3.4.

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).

Lemma A.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,*

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

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

Assumption A.2 $n^{-1}\sum_{t=1}^{n}(\hat\varepsilon_t^4-\varepsilon_t^4)=\mathrm{o}_{\mathsf{P}}(1)$.

Lemma A.2 *If Assumptions 3.1, A.1, A.2 are satisfied then* $CUSQ_n\overset{\mathsf{D}}{\to}\sup_{0\le u\le 1}\left|\mathcal{B}_u^0\right|$.

Proof of Lemma A.2: The statistic of interest is $CUSQ_n=\max_{1\le t\le n}|\mathcal{A}_{nt}|/\hat\varphi_n$, where

$$\mathcal{A}_{nt}=n^{-1/2}\sum_{s=1}^{t}\left(\hat\varepsilon_{s,n}^2-n^{-1}\sum_{r=1}^{n}\hat\varepsilon_{r,n}^2\right).$$

Expand $\mathcal{A}_{nt}=\mathcal{B}_{nt}+\mathcal{C}_{nt}$, where

$$\mathcal{B}_{nt}=n^{-1/2}\sum_{s=1}^{t}\left(\varepsilon_s^2-n^{-1}\sum_{r=1}^{n}\varepsilon_r^2\right),$$
$$\mathcal{C}_{nt}=n^{-1/2}\sum_{s=1}^{t}\left\{\left(\hat\varepsilon_{s,n}^2-\varepsilon_s^2\right)-n^{-1}\sum_{r=1}^{n}\left(\hat\varepsilon_{r,n}^2-\varepsilon_r^2\right)\right\}.$$

By the triangle inequality, $\max_{1\le t\le n}|\mathcal{A}_{nt}|\le\max_{1\le t\le n}|\mathcal{B}_{nt}|+\max_{1\le t\le n}|\mathcal{C}_{nt}|$. By Assumption A.1,

$$\max_{1\le t\le n}|\mathcal{C}_{nt}|\le 2\max_{1\le t\le n}\left|n^{-1/2}\sum_{s=1}^{t}\left(\hat\varepsilon_{s,n}^2-\varepsilon_s^2\right)\right|=\mathrm{o}_{\mathsf{P}}(1). \qquad (A.1)$$

Thus, by Lemma A.1 and the Continuous Mapping Theorem applied to the maximum, we have

$$\max_{1\le t\le n}|\mathcal{A}_{nt}|=\max_{1\le t\le n}|\mathcal{B}_{nt}|+\mathrm{o}_{\mathsf{P}}(1)\overset{\mathsf{D}}{\to}\varphi\sup_{0\le u\le 1}\left|\mathcal{B}_u^0\right|.$$

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 A.1, A.2. 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 Lemma A.1, under Assumption 3.1, we have $\hat{\varphi}_n^2 = \varphi^2 + \mathsf{o}_\mathsf{P}(1)$. All together, $CUSQ_n$ converges in distribution to $\sup_{0\leq u\leq 1}\left|\mathcal{B}_u^0\right|$ as desired. □

For the recursive version of the result we need to strengthen Assumption A.1.

Assumption A.3 $\max_{1\leq t\leq n}\left|n^{-1/2}\sum_{s=1}^t(\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)\right| = \mathsf{o}_\mathsf{P}(1)$.

Lemma A.3 *If Assumptions 3.1, A.2, A.3 are satisfied then* $RCUSQ_n \overset{\mathsf{D}}{\to} \sup_{0\leq u\leq 1}\left|\mathcal{B}_u^0\right|$.

Proof of Lemma A.3: Follow the proof of Theorem A.2 replacing $\hat{\varepsilon}_{s,n}^2$ by $\hat{\varepsilon}_{s,t}^2$ and using Assumption A.3 instead of Assumption A.1 when evaluating (A.1). □

A.2 Some Martingale Results

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.

We need a triangular array modification of the martingale results in Lemma 1 of Lai and Wei (1982).

Lemma A.4 *Let $\mathcal{F}_t$ be a filtration so that w_t is $\mathcal{F}_{t-1}$ adapted and m_t is $\mathcal{F}_t$ adapted with $\mathsf{E}(m_t|\mathcal{F}_{t-1}) = 0$ and $\sup_t \mathsf{E}(m_t^2|\mathcal{F}_{t-1}) < \infty$ a.s. Let N_n be a $q\times q$ normalization matrix, possibly stochastic, where $N_n^{-1} = \mathrm{O}(n^\ell)$ a.s. for some $\ell > 0$. Suppose $n_0 = \inf\{n : \sum_{t=1}^n w_t w_t'$ is invertible$\} < \infty$ a.s. Then, for all $\varsigma > 0$,*

$$\max_{n_0\leq s\leq n}\left\|\sum_{t=1}^s N_n' w_t m_t\right\| \overset{a.s.}{=} \mathrm{o}\left(n^\varsigma \left\|\sum_{t=1}^n N_n' w_t w_t' N_n\right\|^{1/2+\varsigma}\right) + \mathrm{O}(1).$$

Proof of Lemma A.4: Introduce the notation

$$S_{wm,u} = \sum_{t=1}^{[nu]} w_t m_t \qquad \text{and} \qquad S_{ww,u} = \sum_{t=1}^{[nu]} w_t w_t'.$$

For a positive definite matrix $M = RR'$ then $M^{1/2} = R$ and $M^{-1/2} = R^{-1}$. In particular, if $N'S_{ww,u}N = (N'R)(R'N)$ then $N'S_{wm,u} = N'RR^{-1}S_{wm,u} = (N'S_{ww,u}N)^{1/2}S_{ww,u}^{-1/2}S_{wm,u}$. Then, by the norm inequality, for $n_0 < [nu]$,

$$\left\|N_n' S_{wm,u}\right\| = \left\|(N_n' S_{ww,u} N_n)^{1/2} S_{ww,u}^{-1/2} S_{wm,u}\right\| \leq \left\|N_n' S_{ww,u} N_n\right\|^{1/2}\left\|S_{ww,u}^{-1/2} S_{wm,u}\right\|. \qquad \text{(A.2)}$$

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

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

for all $\tilde{\varsigma} > 0$. Since $\|S_{ww,u}\|$ is non-decreasing in u then $\|S_{ww,u}\| \leq \|S_{ww,1}\|$.

A.3 Proof of Main Results

Proof of Theorem 3.1: Part I: Assumption A.1.
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 $\mathcal{S}_{t,\hat{\theta}_n} = \mathrm{o}_{\mathsf{P}}(1)$ uniformly in $1 \leq t \leq n$. From (3.1) 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

$$\dot{h}_s(\theta) = N_{n,\theta_0}'\dot{g}(x_s,\theta), \qquad \ddot{h}_s(\theta) = N_{n,\theta_0}'\ddot{g}(x_s,\theta)N_{n,\theta_0},$$
$$\vartheta = N_{n,\theta_0}^{-1}(\theta - \theta_0), \qquad \nabla\ddot{h}_s(\theta) = N_{n,\theta_0}'\{\ddot{g}(x_s,\theta) - \ddot{g}(x_s,\theta_0)\}N_{n,\theta_0}.$$

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

$$\widetilde{\mathcal{S}}_{t,\hat{\theta}_n} = n^{-1/2}\textstyle\sum_{s=1}^{t}\varepsilon_s\hat{\vartheta}_n'\dot{h}_s(\theta_0) + \frac{1}{2}n^{-1/2}\textstyle\sum_{s=1}^{t}\varepsilon_s\hat{\vartheta}_n'\ddot{h}_s(\theta_*)\hat{\vartheta}_n, \tag{A.3}$$

for an intermediate point θ_* depending on the summation limit t and $\hat{\theta}_n$ so $||\theta_* - \theta_0|| \leq ||\hat{\theta}_n - \theta_0||$. Note that the first term only depends on $\hat{\theta}_n$ through the factor $\hat{\vartheta}_n$. For simplicity we write (A.3) as $\widetilde{\mathcal{S}}_{t,\hat{\theta}_n} = \widetilde{\mathcal{S}}_{t,1} + \widetilde{\mathcal{S}}_{t,2}/2$.
3. *The martingale term $\widetilde{\mathcal{S}}_{t,1}$.* The norm inequality and the bound to $\hat{\vartheta}_n$ in Assumption 3.2 (a) give

$$|\widetilde{\mathcal{S}}_{t,1}| \leq n^{-1/2}||\hat{\vartheta}_n||||\textstyle\sum_{s=1}^{t}\varepsilon_s\dot{h}_s(\theta_0)|| \leq \mathrm{o}_{\mathsf{P}}(n^{\delta-1/2})||\textstyle\sum_{s=1}^{t}\varepsilon_s\dot{h}_s(\theta_0)||.$$

Apply Lemma A.4 using Assumptions 3.1, 3.2, 3.4 to get, for any $\varsigma > 0$,

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

By Assumption 3.3 (c), we have that $\sum_{t=1}^{n}||\dot{h}_t(\theta_0)||^2 = \mathrm{O}_{\mathsf{P}}\left(n^{1-2\delta-\eta}\right)$ for some $\eta > 0$ while $\delta < 1/4$. We then get, when choosing $2\varsigma \leq \eta/(2-2\delta-\eta)$,

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

4. *The term $\widetilde{\mathcal{S}}_{t,2}$.* Apply the norm and triangle inequalities to get

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

Apply the Hölder inequality to get

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

Assumption 3.2 (a) shows $\hat{\vartheta}_n = N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o}_{\mathsf{P}}(n^{\delta})$. The martingale Law of Large Numbers (Chow, 1965, Theorem 5) shows

$$n^{-1}\textstyle\sum_{s=1}^{t}\varepsilon_s^2 \leq n^{-1}\textstyle\sum_{s=1}^{n}\varepsilon_s^2 \overset{a.s.}{=} \mathrm{O}(1).$$

For the third term, note that θ_* is local to θ_0 in probability and depending on t. Let $\varphi > 0$ and note that $\mathsf{P}(\sum_{s=1}^{t}||\ddot{h}_s(\theta_*)||^2 > \varphi)$ can be bounded by

$$\mathsf{P}(\textstyle\sum_{s=1}^{t}||\ddot{h}_s(\theta_*)||^2 > \varphi) \leq \mathsf{P}(\textstyle\sum_{s=1}^{t}||\ddot{h}_s(\theta_*)||^2 > \varphi, ||N_{n,\theta_0}^{-1}(\hat{\theta}_n-\theta_0)|| \leq \epsilon n^{\delta}) + \mathsf{P}(||N_{n,\theta_0}^{-1}(\hat{\theta}_n-\theta_0)|| > \epsilon n^{\delta}).$$

By Assumption 3.2 (a), $\mathsf{P}(||N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0)|| > \epsilon n^{\delta}) = \mathrm{o}(1)$, hence

$$\mathsf{P}(\textstyle\sum_{s=1}^{t}||\ddot{h}_s(\theta_*)||^2 > \varphi) \leq \mathsf{P}(\sum_{s=1}^{t}||\ddot{h}_s(\theta_*)||^2 > \varphi, ||N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0)|| \leq \epsilon n^{\delta}) + \mathrm{o}(1).$$

Therefore, it suffices to show that $\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq\epsilon n^{\delta}} \sum_{s=1}^{t}||\ddot{h}_s(\theta)||^2$ is $\mathrm{o}_{\mathsf{P}}(1)$ uniformly in t. So, taking supremum of θ_*, extending the sum and using Assumption 3.3 (d) gives

$$\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq\epsilon n^{\delta}} \textstyle\sum_{s=1}^{t}||\ddot{h}_s(\theta)||^2 \leq \sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq\epsilon n^{\delta}} \sum_{s=1}^{n}||\ddot{h}_s(\theta)||^2 = \mathrm{O}_{\mathsf{P}}(n^{-4\delta}).$$

Combine the bounds to see that $|\widetilde{\mathcal{S}}_{t,2}| = \mathrm{o}_{\mathsf{P}}(n^{2\delta})\mathrm{O}(1)\mathrm{O}_{\mathsf{P}}(n^{-4\delta}) = \mathrm{o}_{\mathsf{P}}(1)$ uniformly in t.
5. *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 3.3 (a).

Part II: Assumption A.2.
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,

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

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 3.3 (b). □

Proof of Theorem 3.2. We want to show that $\max_{1\leq t\leq n}|n^{-1/2}\sum_{s=1}^{t}(\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)| = \mathrm{o}_{\mathsf{P}}(1)$. Choose $\nu, \eta, \varphi > 0$ small.

First, we show that on a set with large probability $\sup_{t>t_0}|N_{n,\theta_0}^{-1}(\hat{\theta}_t - \theta_0)| = \mathrm{o}(n^{\delta})$. Since $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o}(n^{\delta})$ *a.s.* by Assumption 3.2 (b), then Egorov's theorem (Davidson 1994, Theorem 18.4) implies for all $\nu, \eta > 0$ there exists a $t_0 > 0$ and a set Ω_ν with probability $\mathsf{P}(\Omega_\nu) \geq 1 - \nu$ so that $\sup_{t\geq t_0}|N_{t,\theta_0}^{-1}(\hat{\theta}_t - \theta_0)| < \eta n^{\delta}$ on Ω_ν. Since N_{t,θ_0}^{-1} is non-decreasing in t by Assumption then $\sup_{t_0\leq t\leq n}|N_{n,\theta_0}^{-1}(\hat{\theta}_t - \theta_0)| < \eta n^{\delta}$ on Ω_ν.

Second, we show that is suffices to show that $B_n = \max_{t_0\leq t\leq n}|n^{-1/2}\sum_{s=1}^{t}(\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)| = \mathrm{o}_{\mathsf{P}}(1)$ on Ω_ν. Let $A_n = \max_{1\leq t\leq n}|n^{-1/2}\sum_{s=1}^{t}(\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)|$ and note that we can bound

$$\mathsf{P}(A_n > \varphi) \leq \mathsf{P}(A_n > \varphi \cap \Omega_\nu) + \mathsf{P}(\Omega_\nu^c) \leq \mathsf{P}(A_n > \varphi \cap \Omega_\nu) + \nu.$$

Thus is suffices to show that $A_n = \mathrm{o}_{\mathsf{P}}(1)$ on Ω_ν. On Ω_ν, we further bound

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

Since t_0 is finite, the first term is $\mathrm{O}_{\mathsf{P}}(n^{-1/2})$ and vanishes. Thus it suffices to show that $B_n = \mathrm{o}_{\mathsf{P}}(1)$ on Ω_ν.

Third, we analyze B_n on Ω_ν. For this we follow the proof of Theorem 3.1 replacing $\hat{\varepsilon}_{s,n}^2$ by $\hat{\varepsilon}_{s,t}^2$. The only changes are summation indices start in t_0 and that $\hat{\vartheta}_t = N_{n,\theta_0}^{-1}(\hat{\theta}_t - \theta_0)$ now depends on t. But the bounds are multiplicative in $\hat{\vartheta}_t$ so that we can exploit that $||\hat{\vartheta}_t|| < \eta n^{\delta}$ on Ω_ν. In the original proof the intermediate point θ_* depends on t through the summation index. It will now also depend on $\hat{\theta}_t$ rather than $\hat{\theta}_n$, but that does not add further difficulties since $\hat{\theta}_t$ is local to θ_0.□

Proof of Theorem 3.3. Assumption 3.5 (a,b) with $k = 2$ imply Assumption 3.3 (c,d).

For Assumption 3.3 (a,b), recall the notation in item 3 in the proof of Theorem 3.1 and expand

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

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 $(x + y)^m \leq C(x^m + y^m)$ to see that

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

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

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

which is $\mathsf{o}_\mathsf{P}(n^{1/2})$ for $k = 2$ and $\mathsf{o}_\mathsf{P}(n)$ for $k = 4$ due to Assumption 3.5. $\square$

B Tables and Figures

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	$1 + 0.3\|x_t\|^{1.5} + \varepsilon_t$	$\theta_1 + \theta_2 \|x_t\|^{\theta_3}$
M	5	$y_{t-1} + \varepsilon_t$	$\theta_1 + \theta_2 \|x_t\|^{\theta_3}$
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	$1 + 0.3\|x_t\|^{1.5} + u_t \quad u_t = x_t + \varepsilon_t$	$\theta_1 + \theta_2 \|x_t\|^{\theta_3}$
M	9	$1 + 0.5x_t^2 + \varepsilon_t$	$\theta_1 + \theta_2 \ln^2 \|x_t\|$

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	$x_t + 1.8\left[x_t/(1 + \exp(-x_t/\sqrt{n} - 2))\right] + u_t$
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

Table 5: Yugoslavian hyperinflation. Estimated models and specification tests.

Hyperinflations		$m_t - s_t$	$m_t - s_t$	$m_t - s_t$
$\alpha\lvert\Delta s_t\rvert^\beta$	α	-0.48 (0.07)	-8.15 (5.43)	
	β	1	0.128 (0.09)	
1	γ	7.24 (0.19)	14.20 (5.34)	8.56 (0.16)
d_t				-4.41 (0.29)
se	σ	0.997	0.585	0.557
log likelihood	ℓ	-51.35	-31.11	-29.81
$CUSQ_n$		1.181	0.690	0.672
$RCUSQ_n$		3.615	0.897	0.977

Finite Sample Corrected 95% Critical Value for the $CUSQ$ and the $RCUSQ$ is 1.22. $T = 37$.

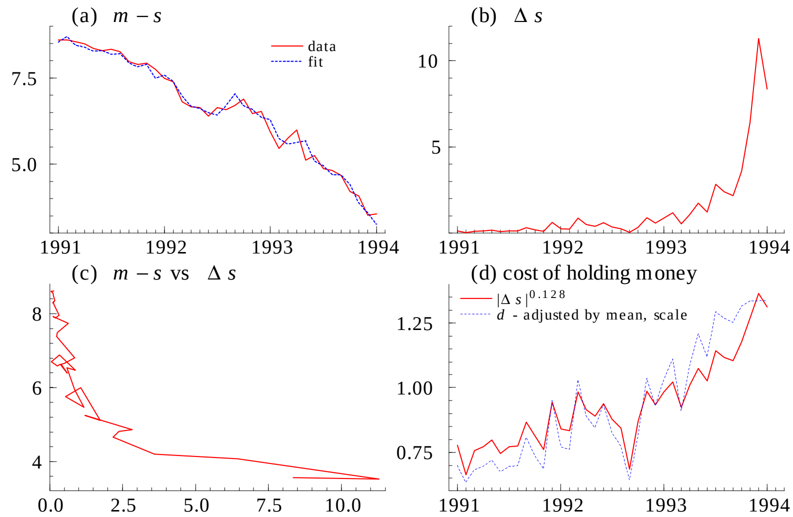

Figure 1: Yugoslavian data and fit.

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.

Berkes I. and L. Horváth (2006) Convergence of integral functionals of stochastic processes. Econometric Theory 22, 304-322.

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.

Christopeit, N. (2009) Weak Convergence of Nonlinear Transformations of Integrated Processes: The Multivariate Case. Econometric Theory 25, 1180-1207.

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

de Jong, R.M. (2004) Addendum to asymptotics for nonlinear transformations of integrated time series. Econometric Theory 20, 627-635.

de Jong, R.M. and C.H. Wang (2005) Further results on the asymptotics for nonlinear transformations of integrated time series. Econometric Theory 21, 413-430.

de Jong, R.M. and L. Hu (2011) A note on nonlinear models with integrated regressors and convergence order results. Economic Letters 111, 23-25.

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.

Engsted, T. (1998) Money demand during hyperinflation: Cointegration, rational expectations, and the importance of money demand shocks. Journal of Macroeconomics 20, 533–552.

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.

Karlsen, H.A., T. Myklebust and D. Tjøstheim (2007) Nonparametric estimation in a non-linear cointegration type model. The Annals of Statistics 35, 252-299.

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

Kasparis, I. (2011) Functional form misspecification in regressions with a unit root. Econometric Theory 27, 285-311.

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. (2005) Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms. Econometric Theory 21, 534-561.

Nielsen, B. (2008) On the Explosive Nature of Hyper-Inflation Data. Economics: The Open-Access, Open-Assessment E-Journal 2, 2008-21.

Nielsen, B. (2010) Analysis of co-explosive processes. Econometric Theory 26, 882-915.

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 (1999) Asymptotics for nonlinear transformations of integrated time series. Econometric Theory 15, 269–298.

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

Petrović, P. and Z. Vujošević (1996) The monetary dynamics in the Yugoslav hyperinflation of 1991–1993: The Cagan money demand. European Journal of Political Economy 12, 467–483. Erratum in (1997), 13, 385–387

Petrović, P., Z. Bogetić and Z. Vujošević (1999) The Yugoslav hyperinflation of 1992–1994: Causes, dynamics, and money supply process. Journal of Comparative Economics 27, 335–353.

Petrović, P. and Z. Mladenović (2000) Money demand and exchange rate determination under hyperinflation: Conceptual issues and evidence from Yugoslavia. Journal of Money, Credit and Banking 32, 785-806.

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.

Pötscher, B.M. and I.R. Prucha (1997) Dynamic nonlinear econometric models: Asymptotic Theory. Springer-Verlag.

Schienle, M. (2011) Nonparametric nonstationary regression with many covariates. Mimeo.

Shi, X. and P.C.B. Phillips (2012) Nonlinear cointegrating regression under weak identification. Econometric Theory 28, 509-547.

Shiller, R.J. (2000) Irrational Exuberance. Princeton, NJ: Princeton University Press.

Smirnov, N. (1948) Tables for estimating the goodness of fit of empirical distributions. Annals of Mathematical Statistics 19, 279-281.

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.

Wang Q. and P.C.B. Phillips (2009) Structural nonparametric cointegrating regression. Econometrica 77, 1901–1948.

Wang Q. and P.C.B. Phillips (2009) A specification test for nonlinear nonstationary models. The Annals of Statistics 40, 727–758.

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.