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

Vanessa Berenguer-Rico* and Bent Nielsen†

6 December 2018

Abstract

We show that the cumulated sum of squares statistic has a standard Brownian bridge-type asymptotic distribution in non-linear regression models with (possibly) non-stationary regressors. This contrasts with cumulated sum statistics which have been previously studied and whose asymptotic distribution has been shown to depend on the functional form and the stochastic properties, such as persistence and stationarity, of the regressors. A recursive version of the test is also considered. A local power analysis is provided and through simulations we show that the test has good size and power properties across a variety 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. 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 (2011), Christopeit (2009), Wang and Phillips (2009, 2012), Choi and Saikkonen (2010), Kristensen and Rahbek (2010), Chan and Wang (2015).

We analyze the cumulated sum of squared residuals test for model (1.1) and show that it has the usual asymptotic distribution when the model is correctly specified. Thus, the test is valid for linear and non-linear models with stationary and non-stationary regressors and it is therefore robust to a wide range of specifications and regressors. 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. For example, Xiao and Phillips (2002) studied the cumulated sum of residuals statistic to test for cointegration. In the context of non-linear and non-stationary regressors Kasparis (2008) and Choi and Saikkonen (2010) 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 quantities coming from the estimation error and the stochastic properties of the regressors. The

*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.

problem is easily illustrated by considering a simple linear regression model –although the same issue is present when dealing with important classes of non-linear transformations, such as homogenous functions, of non-stationary processes. Thus, let $y_t = \theta x_t + \varepsilon_t$, where ε_t are independent standard normal innovations. Let $\hat{\theta}_n$ denote the full sample least squares estimator while $\hat{\varepsilon}_{s,n} = y_s - \hat{\theta}_n x_s$ are the residuals for $s = 1, \dots, n$. Under correct specification we get for the sum of residuals that

$$\frac{1}{\sqrt{n}} \sum_{s=1}^{t} \hat{\varepsilon}_{s,n} = \frac{1}{\sqrt{n}} \sum_{s=1}^{t} \varepsilon_s - (\hat{\theta}_n - \theta) \frac{1}{\sqrt{n}} \sum_{s=1}^{t} x_s.$$

When x_t is stationary with zero mean the second term on the right hand side vanishes so that a Brownian motion theory can be applied to the residual sums. However, when x_t is a random walk we find that the second term converges to a Dickey-Fuller-type distribution that will contribute to the overall asymptotic distribution. The problem extends to non-linear regressions with homogenous functions. To overcome this difficulty Kasparis (2008) uses modified residuals combined with simulations for each specification while Choi and Saikkonen (2010) propose a test involving resampling techniques. This complicates the implementation as well as the theoretical analysis of such tests.

In this paper, we show that the cumulated sum of squared residuals statistic, 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. Hence, the limiting distribution of the test statistic does not depend on the estimation error and is robust to the persistence and stationarity properties of the regressors. To illustrate this consider the simple linear model from above. Under correct specification we get for the sum of squared residuals that

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

both when x_t is stationary and when it is a random walk. Thus, a Brownian Bridge result follows under a martingale difference assumption for ε_t. The Brownian bridge asymptotic result has previosuly been derived in a linear model framework with stationary regressors by for instance Brown, Durbin and Evans (1975), McCabe and Harrison (1980), Ploberger and Krämer (1986), Deng and Perron (2008b) and for non-stationary regressors by Lee, Na and Na (2003) and Nielsen and Sohkanen (2011). Here we show that the Brownian Bridge result extends to non-linear models with non-stationary regressors. This illustrates the robustness of the test to functional form and the type regressors. The proof exploits the results for self-normalized martingales by Lai and Wei (1982).

We also consider a recursive version of the test. For this test the parameter θ is estimated recursively, so that for the sub-sample of the first t observations we get an estimator $\hat{\theta}_t$ and residuals $\hat{\varepsilon}_{s,t} = y_s - \hat{\theta}_t x_s$ for $s = 1, \dots, t$. Under correct specification we can show that the recursive sum of squared recursive residuals has the same expansion as before

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

This again applies to the case when x_t is stationary and when it is a random walk in linear and non-linear regression frameworks. To analyse this statistic one needs to understand the $\hat{\theta}_t$ estimators as a sequence. Deng and Perron (2008b) considered a linear model and required a strong mixing property of the product $x_t \varepsilon_t$, which rules out non-stationary regressors. Nielsen and Sohkanen (2011) observed that if $\hat{\theta}_n$ is strongly consistent, then the sequence $\hat{\theta}_t$ is uniformly convergent in probability by Egorov's theorem. This allows to analyze the recursive test for general regressors. To show that this is applicable in practice we provide some examples of non-linear models with strongly consistent estimators.

The innovations ε_t are martingale difference sequences so that the model (1.1) is a conditional mean model. A common alternative modeling approach is to allow ε_t to follow a general linear process. The standard Brownian Bridge result would still follow under an appropriate standardization of the test. However, it is difficult to control 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), Choi and Saikkonen (2010), or Pitarakis (2017). When following the conditional mean approach, as here, the temporal dependence has to be modelled and any unmodelled autocorrelation or correlation between ε_t and x_t will be regarded as misspecification. The test will not be consistent against such alternatives so we will need to complement the test with tests for temporal dependence just as in linear time series analysis. In choosing the conditional mean approach we follow the recommendation of Deng and Perron (2008a, p. 229). They compare $CUSQ_n$ (cumulated sum of squared residuals) and $CUSUM_n$ (cumulated sum of residuals) tests in the presence of shifts and recommend to model the dynamics and apply the $CUSQ_n$ test.

We provide three examples of models covered by the analysis. First, we analyze the autoregressive distributed lag model. This is a linear model with non-stationary regressors previously analysed by Nielsen and Sohkanen (2011). Second, we consider separable models of the type $y_t = \theta g(x_t) + \varepsilon_t$ with a random walk regressor. Third, we consider the model $y_t = (x_t + \theta)^2 + \varepsilon_t$ suggested by Wu (1981). This is a toy model that illustrates some of the complexities in analyzing models that are non-linear in parameters. Wu considered a linear trend regressor so that $x_t = t$, while Park and Phillips (2001) and Wang (2015) considered a random walk regressor. These authors assumed a compact parameter space. We provide a global analysis without assuming compactness of the parameter space. The model has the interesting feature that it can have two local minimizers so that one minimizer is consistent while the other minimizer diverges. We give conditions that ensure that the consistent minimizer is the global minimizer. In practice one will typically use an iterative algorithm that may find a local minimizer. The possibility of a diverging minimizer indicates that an assumption of a compact parameter space can be severely misleading in practice.

The local power of the test to detect functional form misspecifications is investigated for a separable model involving homogeneous functions. This complements the results on consistency against fixed alternatives given by for instance Xiao and Phillips (2002) and Kasparis (2008) for the cumulated sum of residuals test and by Deng and Perron (2008a) and Pitarakis (2017) for the cumulated sum of squared residuals test. 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 local power of the test to detect functional form misspecifications is analyzed. In Section 6 the performance of the test in terms of size and power is investigated through Monte Carlo experiments. Section 7 contains some concluding remarks. The proofs follow in an Appendix.

2 Model and statistics

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

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

where the functional form of g is known, the innovations ε_t are martingale difference sequence with respect to a filtration $\mathcal{F}_t$, with zero mean, variance σ^2, and fourth moment $\varphi^2 = E\varepsilon_t^4 - (E\varepsilon_t^2)^2$, while x_t is $\mathcal{F}_{t-1}$-adapted. The parameter θ is a q-vector varying in a parameter space $\Theta \subset \mathbb{R}^q$. Detailed assumptions follow in the next section.

The departures from model (2.1) we have in mind are of the form

$$y_t = g(x_t, \theta) + v_t \qquad \text{with} \qquad v_t = \varepsilon_t + \gamma h(z_t), \tag{2.2}$$

where the term $\gamma h(z_t)$ refers to the misspecified part of the model for some unknown function h. Hence, our null hypothesis is $\gamma = 0$. A formal local power analysis will be carried out in Section 5. Until then we will work under the maintained model with $\gamma = 0$.

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

$$Q_n(\theta) = \sum_{t=1}^{n} \{y_t - g(x_t, \theta)\}^2. \tag{2.3}$$

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

$$CUSQ_n = \frac{1}{\hat{\varphi}_n} \max_{1 \leq t \leq n} \left| \frac{1}{\sqrt{n}} \left(\sum_{s=1}^{t} \hat{\varepsilon}^2_{s,n} - \frac{t}{n} \sum_{s=1}^{n} \hat{\varepsilon}^2_{s,n} \right) \right|, \tag{2.4}$$

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

$$\hat{\varphi}^2_n = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}^4_{t,n} - \left(\frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}^2_{t,n} \right)^2. \tag{2.5}$$

We will argue that under quite general assumptions, under the null hypothesis of $\gamma = 0$,

$$CUSQ_n \overset{\mathsf{D}}{\to} \sup_{0 \leq u \leq 1} \left| \mathcal{B}^0_u \right|, \tag{2.6}$$

where $\mathcal{B}^0_u$ is a standard Brownian bridge. Billingsley (1999, pp. 101–104) gives an analytic expression for the distribution function. In particular, the 90%, 95%, 99% quantiles 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

$$1.358 - 0.670 n^{-1/2} - 0.886 n^{-1}. \tag{2.7}$$

We also consider a recursive cumulated sum of squares statistic. Estimating model (2.1) recursively for expanding samples $(y_1, x_1), \ldots, (y_t, x_t)$ gives estimators $\hat{\theta}_t$ for $n_0 \leq t \leq n$ where n_0 is chosen so large that $\sum_{t=1}^{n_0} \{\dot{g}(x_t, \theta_0)\}\{\dot{g}(x_t, \theta_0)\}'$ is invertible and $\dot{g}(x_t, \theta) = \partial g(x_t, \theta)/\partial\theta$ is a q-vector. From the recursive estimators $\hat{\theta}_t$, we compute recursive residuals $\hat{\varepsilon}_{s,t} = y_s - g(x_s, \hat{\theta}_t)$ for $s = 1, \ldots, t$. The recursive cumulative sum of squares test statistic is then defined as

$$RCUSQ_n = \frac{1}{\hat{\varphi}_n} \max_{n_0 \leq t \leq n} \left| \frac{1}{\sqrt{n}} \left(\sum_{s=1}^{t} \hat{\varepsilon}^2_{s,t} - \frac{t}{n} \sum_{s=1}^{n} \hat{\varepsilon}^2_{s,n} \right) \right|. \tag{2.8}$$

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

$$RCUSQ_n \overset{\mathsf{D}}{\to} \sup_{0 \leq u \leq 1} \left| \mathcal{B}^0_u \right|. \tag{2.9}$$

Thus, the same limiting distribution applies as in (2.6). Sohkanen (2011) developed a response surface for the finite sample 95% critical value of $RCUSQ_n$, which is

$$1.358(1 - 0.68 n^{-1/2} + 3.13 n^{-1} - 33.9 n^{-3/2} + 93.9 n^{-2}). \tag{2.10}$$

3 Main results

We show that the above convergence results apply under very mild assumptions. First we consider the $CUSQ_n$ statistic and then we consider the recursive statistic $RCUSQ_n$. Finally, we present some stronger assumptions that may be easier to check in applications.

3.1 The statistic $CUSQ_n$

We start by describing the assumptions before stating the theorems. In Section 4 we discuss how to check these assumptions in particular models.

The first assumption is a martingale difference condition on 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 $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ a.s., so that*
(a) $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2 > 0$ *a.s.;*
(b) $E(\varepsilon_t^4 - \sigma^4|\mathcal{F}_{t-1}) = \varphi^2 < \infty$ *a.s.;*
(c) $\sup_t E(|\varepsilon_t|^\psi|\mathcal{F}_{t-1}) < \infty$ *a.s. for some* $\psi > 4$.

The next assumption relates to the asymptotic behaviour of $\hat{\theta}_n$. We introduce a normalization N_{n,θ_0}^{-1} that allows us to consider both stationary and non-stationary regressors. In many situations of interest it is known that $N_{n,\theta_0}^{-1}(\hat{\theta} - \theta_0)$ converges in distribution, where the notation θ_0 emphasizes the choice of parameter under which we evaluate the distributions. The normalization N_{n,θ_0}^{-1} can be chosen in many ways. In linear models we have $N_{n,\theta_0}^{-1} = n^{1/2}$ for stationary regressors, $N_{n,\theta_0}^{-1} = n$ for random walk or near unit root regressors, while in more general cointegrated models N_{n,θ_0}^{-1} may be block diagonal with different normalizations in different blocks. 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_0. We may also choose a stochastic normalization, for instance $N_{n,\theta_0}^{-1} = (\sum_{t=1}^n x_t x_t')^{1/2}$ in the linear model, so that $N_{n,\theta_0}^{-1}(\hat{\theta} - \theta_0)$ is self-normalized, see Section 4.1, 4.2, 4.4 for examples. In any case the normalization cannot grow faster than at a polynomial rate which rules out explosive regressors.

Assumption 3.2 *Let N_{n,θ_0} be a normalization matrix, possibly stochastic, depending on n and θ_0. Suppose* $\inf(n : N_{n,\theta_0}$ *is invertible*$) < \infty$ *a.s. with the convention that the empty set has infinite infimum. Suppose also that* $N_{n,\theta_0}^{-1} = \mathrm{O}(n^\ell)$ *a.s. for some* $\ell > 0$.

In the subsequent theory it suffices that the normalized statistic $N_{n,\theta_0}^{-1}(\hat{\theta} - \theta_0)$ is $\mathrm{o_P}(n^\delta)$ for some $0 < \delta < 1/4$. As an example consider a linear regression so that $g(x_t, \theta) = \theta x_t$ with stationary regressor x_t. In this case $n^{1/2}(\hat{\theta} - \theta_0) = \mathrm{O_P}(1)$, but it suffices to know that $n^{1/2}(\hat{\theta} - \theta_0) = \mathrm{o_P}(n^\delta)$. The condition gives some freedom, for instance, when dealing with the slowly varying functions that often appear in non-linear analysis, see for instance Phillips (2007). Moreover, in some situations Assumption 3.3 can be established directly by the strong martingale bound of Lai and Wei (1982) presented as Lemma A.7 in the Appendix.

Assumption 3.3 *Suppose* $N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o_P}(n^\delta)$ *for some* $0 < \delta < 1/4$.

The next assumption concerns the 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 3.9 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 3.4 *Suppose x_t is $\mathcal{F}_{t-1}$-measurable and $g(x_t, \theta)$ is twice θ-differentiable. Let* $0 < \delta < 1/4$ *be the consistency rate in Assumption 3.3 and let* $\epsilon, \eta > 0$. *Suppose*
(a) $\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)|| \le n^\delta\epsilon} \sum_{t=1}^n \{g(x_t, \theta) - g(x_t, \theta_0)\}^2 = \mathrm{o_P}(n^{1/2})$;
(b) $\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)|| \le n^\delta\epsilon} \sum_{t=1}^n \{g(x_t, \theta) - g(x_t, \theta_0)\}^4 = \mathrm{o_P}(n)$;

(*c*) $\sum_{t=1}^{n}||N_{n,\theta_0}'\dot{g}(x_t,\theta_0)||^2 = \mathsf{o}_\mathsf{P}(n^{1-2\delta-\eta})$;
(*d*) $\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq n^\delta\epsilon}\sum_{t=1}^{n}||N_{n,\theta_0}'\ddot{g}(x_t,\theta)N_{n,\theta_0}||^2 = \mathsf{o}_\mathsf{P}(n^{-4\delta})$.

Finally, we need invertibility of the matrix of squared first θ-derivatives of g.

Assumption 3.5 *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 the main result for cumulated sums of squares statistics.

Theorem 3.6 *If Assumptions 3.1, 3.2, 3.3, 3.4, 3.5 are satisfied then* $CUSQ_n \overset{\mathsf{D}}{\to} \sup_{0\leq u\leq 1}|\mathcal{B}_u^0|$.

The key argument in the proof is to show that $n^{-1/2}\sum_{s=1}^{n}(\hat{\varepsilon}_{s,n}^2-\varepsilon_s^2)$ vanishes. For this we apply 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 write

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

Due to Assumption 3.3 the rescaled estimator $\hat{\vartheta}_n = N_{n,\theta_0}^{-1}(\hat{\theta}_n-\theta_0)$ varies in the region $||\hat{\vartheta}_n|| \leq \epsilon n^\delta$ with large probability. Thus, it suffices to replace $\hat{\vartheta}_n$ 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 3.4(a). Jennrich (1969, Theorem 6), for instance, uses a similar argument 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 3.4(c,d). For this we modify Lai and Wei (1982, Lemma 1) in Lemma A.8. Finally, Assumption 3.4(b) is used to show consistency of the fourth moment estimator $\hat{\varphi}_n^2$.

3.2 The statistic $RCUSQ_n$

For the recursive cumulated sum of squares statistic the estimator of θ is computed recursively, namely $\hat{\theta}_t$. Thus, we require uniformity properties over t for the sequence of recursive estimators. We follow Nielsen and Sohkanen (2011) and require strong consistency of $\hat{\theta}_n$ and get that the sequence $\hat{\theta}_t$ is uniformly convergent in probability by Egorov's Theorem, see Davidson (1994, Theorem 18.4). Thus, we will need a strong version of Assumption 3.3. Likewise we will need a strong version of Assumption 3.4(a,c,d).

Assumption 3.7 *Suppose Assumptions 3.3, 3.4*(a,c,d) *hold a.s. for some* $0<\delta<1/4$.

Theorem 3.8 *If Assumptions 3.1, 3.2, 3.4*(b)*, 3.5, 3.7 are satisfied then* $RCUSQ_n \overset{\mathsf{D}}{\to} \sup_{0\leq u\leq 1}|\mathcal{B}_u^0|$.

3.3 Some stronger, sufficient assumptions

In many applications we can replace Assumption 3.4 with a set of slightly stronger assumptions formulated in a weak convergence mode and therefore aimed at the non-recursive test. These are useful when working with a fairly generally formulated model satisfying Lipschitz conditions and where weak consistency is often assumed. For the recursive test we would need almost sure properties. Proving those would typically require a more tightly formulated model, in which case one can just as well work with Assumption 3.4.

Assumption 3.9 *Suppose* x_t *is* $\mathcal{F}_{t-1}$*-measurable and* $g(x_t,\theta)$ *is twice* θ*-differentiable. Let* $0<\delta<1/4$ *be the consistency rate in Assumption 3.3 and let* $\epsilon>0$. *Suppose, for* $k=2,4$,
(*a*) $\sum_{t=1}^{n}||N_{n,\theta_0}'\dot{g}(x_t,\theta_0)||^k = \mathsf{o}_\mathsf{P}(n^{k/4-k\delta})$;
(*b*) $\sum_{t=1}^{n}\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)||\leq n^\delta\epsilon}||N_{n,\theta_0}'\ddot{g}(x_t,\theta)N_{n,\theta_0}||^k = \mathsf{o}_\mathsf{P}\{n^{(k-2)/2-2k\delta}\}$.

Since $0 < \delta < 1/4$ it suffices that the rates in (a) and (b) are $\mathrm{O}_\mathsf{P}(1)$ and $\mathrm{O}_\mathsf{P}(n^{-1})$, respectively.

Theorem 3.10 *Assumption 3.9 implies Assumption 3.4.*

4 Analysis of some particular models

We illustrate the practical use of the general assumptions through particular models that have been discussed in the literature. We first consider the autoregressive distributed lag model. We then turn to separable non-linear models with random walk regressors considering models that are linear in parameters. Finally, as an example of a model that is not linear in parameters we consider Wu's (1981) power-curve model.

4.1 Autoregressive distributed lag model

We consider an autoregressive distributed lag model involving time series that are stationary or integrated. This is a linear model so estimation reduces to ordinary least squares. It provides a relatively simple setting for appreciating the n^δ rate appearing in the consistency Assumption 3.3 as well as the benefits of choosing a stochastic normalization matrix N_n. Thus this linear model serves as a first illustration of the developed theory, noting that it has been previously studied by Nielsen and Sohkanen (2011).

Consider data y_t, z_t, where z_t has dimension m, and the autoregressive distributed lag model

$$y_t = \sum_{j=1}^{k} \alpha_j y_{t-j} + \sum_{j=1}^{k} \beta_j' z_{t-j} + \nu + \varepsilon_t. \tag{4.1}$$

Thus, in terms of the notation in equation (2.1) we have a regressor $x_t = (y_{t-1}, z_{t-1}', \ldots, y_{t-k}, z_{t-k}', 1)'$ of dimension $p = k(m+1)+1$ and a parameter $\theta = (\alpha_1, \beta_1', ..., \alpha_k, \beta_k', \nu)'$ of dimension $q = p$ estimated by the ordinary least squares estimator $\hat{\theta}_n$.

In order to characterize the asymptotic distribution of the test statistic we specify a joint model for the time series $\mathbf{x}_t = (y_t, z_t')'$. Suppose $\mathbf{x}_t$ satisfies the vector autoregression

$$\mathbf{x}_t = \sum_{j=1}^{k} A_j \mathbf{x}_{t-j} + \mu + \xi_t, \tag{4.2}$$

so that model (4.1) is the first equation of the vector autoregression (4.2). Note that $x_t = (\mathbf{x}_{t-1}', ..., \mathbf{x}_{t-k}', 1)'$. The vector autoregression $\mathbf{x}_t$ has companion form $x_{t+1} = Bx_t + \omega_t$ where

$$B = \begin{pmatrix} A_1 & \cdots & A_{k-1} & A_k & \mu \\ I & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & I & 0 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix}, \qquad \omega_t = \begin{pmatrix} \xi_t \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}.$$

The companion matrix B has a unit root corresponding to the constant term. If the remaining eigenvalues of B have absolute value less than unity then $\mathbf{x}_t$ can be given a stationary initial distribution. The random walk case arises when B has more than one unit root. We need the following assumption.

Assumption 4.1 *The vector $\mathbf{x}_t$ satisfies (4.2) so that*
(a) *all linear combinations of ξ_t satisfy the martingale Assumption 3.1;*
(b) *the companion matrix B has eigenvalues so that* $\max|\mathrm{eigen}(B)| \leq 1$.

We choose a stochastic normalization matrix $N_n^{-1} = (\sum_{t=1}^n x_t x_t')^{1/2}$ so that $N_n N_n' = (\sum_{t=1}^n x_t x_t')^{-1}$. This yields the self-normalized least squares statistic $N_n^{-1}(\hat{\theta}_n - \theta_0) = (\sum_{t=1}^n x_t x_t')^{-1/2} \sum_{t=1}^n x_t \varepsilon_t$. When ε_t is a martingale difference sequence it can be shown that $N_n^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{O}\{(\log n)^{1/2}\}$ *a.s.* by appealing to Lai and Wei (1982, Lemma 1), see also Lemma A.8 in the Appendix. Thus, $N_n^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o}(n^\delta)$ *a.s.* for any $\delta > 0$ as required in Assumption 3.7 since $\log n$ is dominated by n^δ for any $\delta > 0$.

Theorem 4.2 *Consider model (4.1) with $\theta \in \Theta = \mathbb{R}^p$ and Assumption 4.1. Then $CUSQ_n$ and $RCUSQ_n$ converge in distribution to* $\sup_{0\le u\le 1} |\mathcal{B}_u^0|$.

The CUSQ and RCUSQ tests have been previously studied for this model in Nielsen and Sohkanen (2011). Their result also allows contemporaneous regressors z_t as well as explosive roots for the companion vector.

4.2 Separable models with homogenous functions

As a first non-linear model we consider the separable model

$$y_t = \theta g(x_t) + \varepsilon_t \qquad t = 1, \ldots, n, \tag{4.3}$$

where g is a known scalar function, so that non-linear least squares for θ simplifies to ordinary least squares. Moreover, g is continuous and homogenous, so that $g(x_t/n^{1/2}) = g(x_t)/g(n^{1/2})$. Polynomial regressions are a classical example of such functions. The more general class of asymptotically homogeneous functions discussed in Park and Phillips (1999) will be explored in the next section. We consider the random walk type regressor. We explore the robustness of the results to other types of regressors through simulation in Section 6. Our assumptions are as follows.

Assumption 4.3 *The regressor satisfies $x_t = \sum_{s=1}^{t-1} \eta_s + \alpha' \mathbf{x}_{t-1}$ where η_t is an $\mathcal{F}_t$-martingale difference sequence satisfying Assumption 3.1(a, c) with $\sigma_\eta^2 = E(\eta_t^2|\mathcal{F}_{t-1})$ and for some $\psi > 2$. Moreover, $\alpha \in \mathbb{R}^{m+1}$ and $\mathbf{x}_t$ is a vector autoregressive process of the form (4.2) satisfying Assumption 4.1 so that the companion matrix B has eigenvalues so that* $\max|\mathrm{eigen}(B)| < 1$.

Assumption 4.4 *The function g is continuous and satisfies*
(i) $g(x/n^{1/2}) = g(x)/g(n^{1/2})$ and $|g(x)| \le C(1 + |x|^\ell)$ for some finite $C, \ell > 0$;
(ii) $n_0 = \inf[n : \sum_{t=1}^n \{g(x_t)\}^2$ is invertible$] < \infty$ a.s.

Once again, we choose a stochastic normalization matrix $N_n = [\sum_{t=1}^n \{g(x_t)\}^2]^{-1/2}$. We can then prove, first, almost sure consistency and subsequently that Theorems 3.6, 3.8 apply.

Theorem 4.5 *Consider model (4.3) with $\theta \in \Theta = \mathbb{R}$ and suppose that Assumptions 3.1, 4.3, 4.4 hold. Then $N_n^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o}\{(\log n)^2\} = \mathrm{o}(n^\delta)$ a.s. for all $\delta > 0$.*

Theorem 4.6 *Consider model (4.3) with $\theta \in \Theta = \mathbb{R}$ and suppose that Assumptions 3.1, 4.3, 4.4 hold. Then $CUSQ_n$ and $RCUSQ_n$ converge in distribution to* $\sup_{0\le u\le 1} |\mathcal{B}_u^0|$.

4.3 Separable models with asymptotically homogeneous functions

We now consider the separable model $y_t = \theta g(x_t) + \varepsilon_t$ as in (4.3) under a more general class of functions g. Park and Phillips (1999) introduced a class of asymptotically homogenous functions, say T, which is dominated by a locally integrable function, say H, see Definition 4.8 below. They develop a limit theory for sums $\sum_{t=1}^n T(x_t)$ when T is asymptotically homogeneous and the dominating function, H, satisfies some additional regularity conditions. Pötscher (2004) developed a more general theory for locally integrable functions H and showed that sums $n^{-1}\sum_{t=1}^n H(n^{-1/2} x_t)$ converge without additional regularity conditions for H. He suggested that a theory for asymptotically homogeneous functions could be developed along the lines of Park and Phillips (1999) where the dominating function is just locally integrable. We formalize this suggestion and use that to analyze the $CUSQ_n$ test. We note that when analyzing the $CUSQ_n$ statistic we do not need a full limit theory for the estimator. We only need a weak consistency bound to $\hat{\theta}_n - \theta$ as required in Assumption 3.3. By focusing on separable functions we can use Lai and Wei (1982, Lemma 1) combined with Pötscher's results to establish a sufficient weak consistency bound to $\hat{\theta}_n - \theta$ for regressions with asymptotically homogeneous functions. Analyzing the recursive $RCUSQ_n$ statistic requires almost sure bounds to sums of asymptotically homogeneous functions which are not available at present, hence we focus on the $CUSQ_n$ statistic in this sub-section.

Pötscher (2004) defines locally integrable functions as follows.

Definition 4.7 *Let H be a real-valued Borel-measurable function on $\mathbb{R}$. We say that H is locally integrable if and only if*

$$\int_{-K}^{K} |H(x)| dx < \infty \qquad \textit{for all } 0 < K < \infty.$$

The class of locally integrable functions includes locally bounded functions in the sense that $\sup_{|x| \leq K} |H(x)| < \infty$ for all $0 < K < \infty$. Pötscher (2004, Remarks 2.1, 2.5) points out that locally unbounded functions like $\log |x|$ can be dealt with by considering functions H on the extended real line.

Park and Phillips (1999, Definition 4.2) define the class of asymptotically homogeneous functions as follows –we minimally adapt their definition to our setup in relation to a normalization.

Definition 4.8 *Let $T : \mathbb{R} \mapsto \mathbb{R}$. We say that T is asymptotically homogeneous if and only if*

$$T(\lambda x) = v(\lambda) H(x) + R(x, \lambda) \qquad \textit{for } \lambda > 0,$$

where $H(x)$ is locally integrable and $v(\lambda)$ is a normalization in the sense that is $v(\lambda)$ is positive, non-decreasing and $\lim_{\lambda \to \infty} \lambda^{-\ell} v(\lambda) \to 0$ for some $\ell > 0$. The remainder R satisfies $|R(x, \lambda)| \leq a(\lambda) P(x)$ where $a(\lambda)$ is a normalization so that $\sup_{\lambda \to \infty} a(\lambda)/v(\lambda) = 0$ and $P(x)$ is locally integrable.

Before proceeding we need the following technical Assumption 2.2 from Pötscher (2004).

Assumption 4.9 *Let x_t satisfy Assumption 4.3. For every $n \in \mathbb{N}$ the distribution of $n^{-1/2} x_n$ possesses a density, h_n say, with respect to the Lebesgue measure on $\mathbb{R}$. The density h_n is uniformly bounded, that is $\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}} |h_n(x)| < \infty$.*

We can then generalize Theorem 2.1 of Pötscher (2004) as follows.

Theorem 4.10 *Suppose Assumptions 4.3, 4.9 are satisfied. Then $n^{-1/2} x_{[un]}$, where $[un]$ denotes the integer part of un, converges weakly to $\sigma_\eta W_u$ on the $D[0, 1]$ where W is a standard Brownian motion. If $T : \mathbb{R} \mapsto \mathbb{R}$ is asymptotically homogeneous in the sense of Definition 4.8, then*

$$\{n v(n^{1/2})\}^{-1} \sum_{t=1}^{n} T(x_t) \overset{\mathsf{D}}{\to} \int_0^1 H(\sigma_\eta W_u) du.$$

To analyze the $CUSQ_n$ statistic we make the following assumption on the square of the function g. In contrast to Park and Phillips (1999), no regularity conditions on the function g itself are required as we do not need a full asymptotic theory for the least squares estimator.

Assumption 4.11 *Suppose $g^2(x)$ is asymptotically homogeneous in the sense of Definition 4.8.*

We note that asymptotic homogeneity of g will not in general imply asymptotic homogeneity of g^2. For instance, if $g(x) = |x|^{-1/2}$ then g is locally integrable, but g^2 is not locally integrable. If we had assumed that g is locally bounded then g^2 would also be locally bounded and hence locally integrable.

Finally, as in previous sections, the following technical assumption is needed.

Assumption 4.12 $n_0 = \inf[n : \sum_{t=1}^n \{g(x_t)\}^2$ *is invertible*$] < \infty$ *a.s.*

Theorem 4.13 *Consider model (4.3) with* $\theta \in \Theta = \mathbb{R}$ *and where Assumptions 3.1, 4.3, 4.9, 4.11, 4.12 hold. Let* $N_n^{-1} = \sum_{t=1}^n \{g(x_t)\}^2$. *Then* $N_n^{-1}(\hat{\theta}_n - \theta_0) = \mathsf{o_P}(n^\delta)$ *for all* $\delta > 0$.

Theorem 4.14 *Consider model (4.3) with* $\theta \in \Theta = \mathbb{R}$ *and where Assumptions 3.1, 4.3, 4.9, 4.11, 4.12 hold. Then* $CUSQ_n$ *converges in distribution to* $\sup_{0\le u\le 1} |\mathcal{B}_u^0|$.

4.4 Power curve model

A further development in the literature is to consider non-separable functions $g(x_t, \theta)$ as proposed by Park and Phillips (2001) and subsequently analyzed by for instance Chan and Wang (2015). Their results for the estimator $\hat{\theta}_n$ require a compact parameter space or equivalently an assumption that $\hat{\theta}_n$ is bounded in probability. This assumption is clearly non-trivial. We illustrate this fact by analyzing the following model

$$y_t = (x_t + \theta)^2 + \varepsilon_t, \tag{4.4}$$

where $\theta \in \Theta = \mathbb{R}$ and x_t is either stationary, a random walk or a deterministic power function. When proving consistency of $\hat{\theta}_n$ for a particular value θ_0 we work with the normalization $N_{n,\theta_0}^{-1} = \{\sum_{t=1}^n (x_t + \theta_0)^2\}^{1/2}$.

Model (4.4) was previously considered by Wu (1981) in the linear trend case $x_t = t$, while Park and Phillips (2001) or Wang (2015) considered the random walk case. These authors assumed a compact parameter space when analyzing the properties of the non-linear least squares estimator. Compactness is not so attractive in this case since the objective function is quartic and can have two local minimizers. As we will show, a second local minimizer is not always present. When it is present it can even be diverging. We analyze the objective function and its minimizers in detail in Theorem A.17 in the Appendix. There we show that, asymptotically, the global minimizer is unique. Here, we concentrate on the asymptotic properties of this unique global minimizer and the statistics $CUSQ_n$ and $RCUSQ_n$.

Since model (4.4) is non-linear in parameters, in this section, we consider the non-linear least squares estimator $\hat{\theta}_n = \arg\min_{\theta\in\mathbb{R}} Q_n(\theta)$ where $Q_n(\theta) = \sum_{t=1}^n \{y_t - (x_t + \theta)^2\}^2$. In the case of a stationary or deterministic power function regressor we can show strong consistency of the global minimizer, $\hat{\theta}_n$, and hence we can analyze the recursive test.

Assumption 4.15 *Suppose* $x_t = t^\tau$ *for* $\tau > 0$ *or* x_t *is stationary, autoregressive and* $\mathcal{F}_{t-1}$*-adapted with* $\mathsf{E}x_t = \mu_x$ *and* $\mathsf{V}x_t = \sigma_x^2$ *and* $\mathsf{E}x_t^4 < \infty$.

Theorem 4.16 *Consider model (4.4) with* $\theta \in \Theta = \mathbb{R}$ *and Assumptions 3.1, 4.15. Then the global minimizer satisfies*

$$\{\textstyle\sum_{t=1}^n (x_t + \theta_0)^2\}^{1/2}(\hat{\theta}_n - \theta_0) \overset{a.s.}{=} \frac{\sum_{t=1}^n (x_t + \theta_0)\varepsilon_t}{2\{\sum_{t=1}^n (x_t + \theta_0)^2\}^{1/2}} + \mathrm{o}(1) \overset{a.s.}{=} \mathrm{o}\{(\log n)^2\}.$$

Theorem 4.17 *Consider model (4.4) with* $\theta \in \Theta = \mathbb{R}$ *and Assumptions 3.1, 4.15. Then* $CUSQ_n$ *and* $RCUSQ_n$ *converge in distribution to* $\sup_{0\le u\le 1} |\mathcal{B}_u^0|$.

For a random walk regressor we can only show weak consistency of the global minimizer, $\hat{\theta}_n$, and hence we only analyze the non-recursive test.

Theorem 4.18 *Consider model (4.4) with $\theta \in \Theta = \mathbb{R}$ and Assumptions 3.1, 4.3. Then the global minimizer satisfies*

$$\{\sum_{t=1}^{n}(x_t+\theta_0)^2\}^{1/2}(\hat{\theta}_n-\theta_0) = \frac{\sum_{t=1}^{n}(x_t+\theta_0)\varepsilon_t}{2\{\sum_{t=1}^{n}(x_t+\theta_0)^2\}^{1/2}} + \mathrm{o_P}(1) = \mathrm{o_P}\{(\log n)^2\}.$$

Theorem 4.19 *Consider model (4.4) with $\theta \in \Theta = \mathbb{R}$ and Assumptions 3.1, 4.3. Then $CUSQ_n$ converges in distribution to* $\sup_{0\le u\le 1}|\mathcal{B}_u^0|$.

5 Local power analysis

We analyze the local power of the $CUSQ_n$ statistic under the type of departure from the null described in equation (2.2) where $y_t = g(x_t,\theta) + v_t$ and $v_t = \varepsilon_t + \gamma h(z_t)$ with $\gamma = 0$ under the null. This alternative is linear in the parameter γ and we match that with a non-linear function $g(x_t,\theta) = \theta g(x_t)$ that is separable as in (4.3). We choose both g and h to be homogenous and consider a regressor x_t and an unmodelled variable z_t that are both of random walk type. For the local power analysis we choose a drifting sequence for γ approaching zero. We will also argue that the test is consistent, that is, the power approaches unity in large samples for fixed γ alternatives. Simulations in the next section will show that the $CUSQ_n$ test also has power in more general settings regarding the regression function.

In the local power analysis we need to normalize the parameter γ to give a non-trivial drift. Since h is homogenous we have $h(z_t/n^{1/2}) = h(z_t)/h(n^{1/2})$. In addition we need an $n^{1/4}$ factor so that the local parameter is defined through $\gamma = \lambda/\{n^{1/4}h(n^{1/2})\}$. That is, we analyze

$$y_t = \theta g(x_t) + v_t \qquad \text{where} \qquad v_t = \varepsilon_t + \frac{\lambda}{n^{1/4}h(n^{1/2})}h(z_t). \tag{5.1}$$

Such processes satisfy model (4.3) when $\lambda = 0$. For the local analysis we need the following assumption.

Assumption 5.1 (a) *The functions g and h are continuous and homogenous so that $g(x_t/\sqrt{n}) = g(x_t)/g(\sqrt{n})$ and $h(z_t/\sqrt{n}) = h(z_t)/h(\sqrt{n})$.*
(b) *The regressors x_t, z_t satisfy random walk assumptions as outlined in Assumption 4.3.*

Consider the partial sum process $n^{-1/2}\{\sum_{t=1}^{[nu]}(\varepsilon_t^2-\sigma^2), x_{[nu]}, z_{[nu]}\}'$ for $0 \le u \le 1$, where $[nu]$ denotes the integer part of nu. This process converges weakly to the Brownian motion $(\mathcal{B},\mathcal{W}_x,\mathcal{W}_z)'$. Define the process

$$\mathcal{L}_u = \int_0^u \{h(\mathcal{W}_{z,s}) - \frac{\int_0^1 h(\mathcal{W}_{z,r})g(\mathcal{W}_{x,r})dr}{\int_0^1 g^2(\mathcal{W}_{x,r})dr}g(\mathcal{W}_{x,s})\}^2 ds$$

and the bridge process $\mathcal{L}_u^\circ = \mathcal{L}_u - u\mathcal{L}_1$. We then have the following local power result.

Theorem 5.2 *Consider data generating processes (5.1) with $\lambda \in \mathbb{R}$ and Assumptions 3.1, 5.1. Then*

$$CUSQ_n \overset{\mathsf{D}}{\to} \varphi^{-1}\sup_{0\le u\le 1}|\mathcal{B}_u^\circ + \lambda^2\mathcal{L}_u^\circ|.$$

Theorem 5.2 allows us to trace the local power of the test as a function of λ. For $\lambda = 0$ we recognize the null limiting distribution. The parameter λ measures departures from that null. For large λ the term $\lambda^2\mathcal{L}_u^\circ$ dominates which gives the local power.

The test will also be consistent. For this argument we consider data generating processes satisfying

$$y_t = \theta g(x_t) + v_t \qquad \text{where} \qquad v_t = \varepsilon_t + \gamma h(z_t), \tag{5.2}$$

as in model (2.2) with fixed alternatives determined by γ. If λ in (5.1) is replaced by $\lambda_n = \gamma n^{1/4} h(n^{1/2})$ then we get processes of the form (5.2). For a general non-decreasing sequence $\lambda_n > 0$ then $CUSQ_n$ diverges at a rate of $\min(\lambda_n^2, n^{1/2})$, see Remark A.1 in the appendix. In particular in the fixed alternative case where $\lambda_n = \gamma n^{1/4} h(n^{1/2})$, with non-decreasing $h(n^{1/2})$, then $CUSQ_n$ diverges at a rate of $\min(\lambda_n^2, n^{1/2}) = n^{1/2}$ in line with the results found in previous studies. These include the Xiao and Phillips (2002) and Kasparis (2008) analyses of the cumulated sum statistic with non-linearity and non-stationarity and the Deng and Perron (2008a) analysis of the $CUSQ_n$ statistic for stationary variables with structural breaks in the conditional mean.

The $CUSQ_n$ statistic will have non-trivial power to detect misspecifications that are of a different nature than the ones considered in Theorem 5.2. This point is analyzed by simulation in the next section for a wide range of situations. Of particular interest is the case in which the misspecified component is an omitted stationary autoregressive regressor, so that the power of the test is driven by its long run variance –see for instance Deng and Perron (2008b) or Pitarakis (2017) for a theoretical argument on this issue.

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. When studying the size of the test a correctly specified model with iid innovations and a highly persistent regressor is considered. For the power, and given the emphasis in previous sections, we focus here on functional form misspecifications. Specifically, two sets of results are presented. 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 the 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. Table 3 contains a second set of DGPs. Four correctly specified (CS) DGPs and four misspecified (M) DGPs are considered. Specifically, in DGPs 1-4 of Table 3, where the model is correctly specified, the regressor is a nearly integrated process $x_t = (1 + c/n)x_{t-1} + u_t$, with u_t iid $\mathsf{N}(0,1)$ and for which we consider $c = -5$, -20, -50 following Kasparis, Andreou and Phillips (2015). In DGPs 5-8 of Table 3, we consider misspecifications where an autoregressive regressor $z_t = \alpha z_{t-1} + \nu_t$ with ν_t iid $\mathsf{N}(0,1)$ and $\alpha = 0.7$, 0.8, 0.9 has been omitted. The results are as follows.

Table 2, corresponding to DGPs 1-4 in Table 1, 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 (2.7) due to Edgerton and Wells (1994). Notice from Table 4 that the same features are obtained when the regressor is nearly integrated. All these results regarding the size of the test provide clear evidence on the robustness of the procedure with respect to the degree of persistence of the regressors as well as the form of the regression function.

Next, we turn to the power of the test. Table 2, corresponding to DGPs 5-9 in Table 1, reports the power of the $CUSQ$ test for a range of asymptotically homogenous functions. The power increases with the sample size in all cases. This is in line with the power analysis for parameter instabilities conducted by McCabe and Harrison (1980), Ploberger and Krämer (1990), Deng and Perron (2008a), or Turner (2010). It is worth emphasizing that given the goodness-of-fit 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. From Table 5, corresponding to the misspecified models 5-8 in Table 3, it can be seen that the test has non-trivial power to detect autocorrelated residuals and, as expected, the stronger the autocorrelation, the higher the power of the test. As for linear models, the test does not appear to be consistent in this case so it should be combined with a test for temporal dependence of the residuals.

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 6 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 7. 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 7 shows that no test dominates in all cases, but in large samples the $CUSQ_n$ seems to dominate the CUSUM test. 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.

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 $CUSUM_n$ tests based on the cumulated sum of residuals in a non-stationary context. In other words, the cumulated sum of squares statistics is a specification test robust to the non-stationarity and/or persistence properties of the regressors. Hence, 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.

In terms of size, the $CUSQ_n$ test using asymptotic critical values has a size that is nearly uniform over a variety of models in small samples. The asymptotic critical values give a slightly undersized test for small samples, but the size can be controlled nearly perfectly using the response surface of Edgerton and Wells (1994) reported in (2.7). These results are in line with the size control found previously for linear, non-stationary models by Nielsen and Sohkanen (2011).

The test has good power against a variety of non-linear misspecifications within a non-stationary environment. In the formal power analysis we considered an alternative hypothesis with an unmodelled non-linear function of a random walk type variable. We derived the local power function and showed that the test is consistent for fixed alternatives. In the simulation study we compared with the alternatives studied by Kasparis (2008) for his $CUSUM_n$ test. The alternatives include logarithmic, threshold and polynomial alternatives involving random walks while the fitted model is linear cointegration regression. We find a comparable power when (infeasibly) using the best of his bandwidth choices. In large samples the $CUSQ_n$ test appears to have more power than the $CUSUM_n$ test.

The asymptotic properties of the $CUSQ_n$ test are derived under the assumption of conditionally homoskedastic martingale difference errors. This assumes implicitly that any dynamics have been modelled. Indeed, this is consistent with the recommendation of Deng and Perron (2008a) for linear, stationary models. Thus, in practice we recommend to combine the $CUSQ_n$ test with a test for no residual autocorrelation. We are currently studying such a test in a non-linear, non-stationary context. Once such a test is available it will be possible to investigate which alternatives are best picked up by which test and to study the performance of a combined test.

8 Acknowledgements

Comments from the Giuseppe Cavaliere and 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.

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. Recall* $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$ *and* $E(\varepsilon_t^4|\mathcal{F}_{t-1}) = \varphi^2 + \sigma^4$. *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,*

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

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 in terms of auxillary assumptions.

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

Assumption A.3 $n^{-1}\sum_{t=1}^{n}(\hat\varepsilon_t^4 - \varepsilon_t^4) = \mathrm{o}_{\mathsf{P}}(1)$.

Assumption A.4 $\max_{n_0\le t\le n} \left|n^{-1/2}\sum_{s=1}^{t}(\hat\varepsilon_{s,t}^2 - \varepsilon_s^2)\right| = \mathrm{o}_{\mathsf{P}}(1)$.

Lemma A.5 *If Assumptions 3.1, A.2, A.3 are satisfied then* $CUSQ_n \overset{\mathsf{D}}{\to} \sup_{0\le u\le 1}\left|\mathcal{B}_u^0\right|$.

Lemma A.6 *If Assumptions 3.1, A.3, A.4 are satisfied then* $RCUSQ_n \overset{\mathsf{D}}{\to} \sup_{0\le u\le 1}\left|\mathcal{B}_u^0\right|$.

Proof of Lemma A.5: 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 and Assumption A.2 then

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

Introduce the total norm $||f_t||_\infty = \max_{1\le t\le n}|f_t|$. The triangle inequality shows $||\mathcal{A}_{nt}||_\infty \le ||\mathcal{B}_{nt}||_\infty + ||\mathcal{C}_{nt}||_\infty$ as well as $||\mathcal{B}_{nt}||_\infty = ||\mathcal{A}_{nt} - \mathcal{C}_{nt}||_\infty \le ||\mathcal{A}_{nt}||_\infty + ||\mathcal{C}_{nt}||_\infty$ so that $||\mathcal{A}_{nt}||_\infty \ge ||\mathcal{B}_{nt}||_\infty - ||\mathcal{C}_{nt}||_\infty$. In combination $|(||\mathcal{A}_{nt}||_\infty - ||\mathcal{B}_{nt}||_\infty)| \le ||\mathcal{C}_{nt}||_\infty$. Using (A.1) we get $\max_{1\le t\le n}|\mathcal{A}_{nt}| = \max_{1\le t\le n}|\mathcal{B}_{nt}| + \mathrm{o}_{\mathsf{P}}(1)$. Thus, by Lemma A.1 and the Continuous Mapping Theorem applied to the maximum, we have

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

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, $n^{-1}\sum_{t=1}^n (\hat{\varepsilon}_{t,n}^k - \varepsilon_t^k) = \mathrm{o_P}(1)$ for $k = 2, 4$ by Assumptions A.2, A.3. Therefore,

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

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

Proof of Lemma A.6: The proof is largely the same as that of Lemma A.5 but where the maximum is taken over $n_0 \le t \le n$ and the sets $\mathcal{A}_{nt}, \mathcal{C}_{nt}$ are replaced by $\mathcal{A}_{nt}^* = n^{-1/2}\sum_{s=1}^t (\hat{\varepsilon}_{s,t}^2 - n^{-1}\sum_{r=1}^n \hat{\varepsilon}_{r,n}^2)$ and $\mathcal{C}_{nt}^* = n^{-1/2}\sum_{s=1}^t \{(\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2) - n^{-1}\sum_{r=1}^n (\hat{\varepsilon}_{r,n}^2 - \varepsilon_r^2)\}$. □

A.2 Some martingale results

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

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

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

$$||m|| = ||m'm||^{1/2}. \tag{A.2}$$

To see this note $||m||^2 = \max \operatorname{eigen}(m'm)$. Here $m'm$ is symmetric, positive semi-definite so $m'm = V\Lambda V'$ for a diagonal, semi-definite Λ and $VV' = I$. Thus $\max \operatorname{eigen}(m'm) = \max \operatorname{eigen}(\Lambda) = \{\max \operatorname{eigen}(\Lambda^2)\}^{1/2}$. Now $\{\max \operatorname{eigen}(\Lambda^2)\}^{1/2} = \{\max \operatorname{eigen}(m'mm'm)\}^{1/2} = ||m'm||$.

We quote a version of Lemma 1 of Lai and Wei (1982) and provide a triangular array modification.

Lemma A.7 (Lai and Wei,1982, Lemma1) *Let $\mathcal{F}_t$ be a filtration so that the $q \times 1$ vector w_t is $\mathcal{F}_{t-1}$ adapted and the scalar m_t is $\mathcal{F}_t$ adapted with $E(m_t|\mathcal{F}_{t-1}) = 0$ and $\sup_t E(m_t^2|\mathcal{F}_{t-1}) < \infty$ a.s. Suppose $n_0 = \inf\{n : \sum_{t=1}^n w_t w_t'$ is invertible$\} < \infty$ a.s. Then, for all $\varsigma > 0$ and $n > n_0$,*

$$\sum_{t=1}^n m_t w_t' (\sum_{t=1}^n w_t w_t')^{-1} \sum_{t=1}^n w_t m_t \overset{a.s.}{=} \mathrm{o}\{(\log ||\sum_{t=1}^n w_t w_t'||)^{1+\varsigma}\} + \mathrm{O}(1).$$

Lemma A.8 *Let $\mathcal{F}_t$ be a filtration so that the $q\times 1$ vector w_t is $\mathcal{F}_{t-1}$ adapted and the scalar m_t is $\mathcal{F}_t$ adapted with $E(m_t|\mathcal{F}_{t-1}) = 0$ and $\sup_t E(m_t^2|\mathcal{F}_{t-1}) < \infty$ 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 $\varsigma > 0$,*

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

Proof of Lemma A.8: Introduce the notation

$$S_{wm,s} = \sum_{t=1}^s w_t m_t \qquad \textit{and} \qquad S_{ww,s} = \sum_{t=1}^s w_t w_t'.$$

We want to prove that $\max_{n_0\le s\le n} \|N_n' S_{wm,s}\| = \mathrm{o}(n^{\varsigma})(1 + \|N_n' S_{ww,n} N_n\|^{1/2+\varsigma})$ *a.s.*

By construction $S_{ww,s}$ is positive semi-definite, while by assumption, $S_{ww,s}$ is invertible and hence positive definite for $s > n_0$. Because $S_{ww,s}$ is a positive definite and symmetric matrix, it can be decomposed as $S_{ww,s} = RR'$ so that $S_{ww,s}^{1/2} = R$ and $S_{ww,s}^{-1/2} = R^{-1}$. Therefore, using these definitions, since $N_n' S_{wm,s} = N_n' S_{ww,s}^{1/2} S_{ww,s}^{-1/2} S_{wm,s}$ we can write

$$N_n' S_{wm,s} = (N_n' R) S_{ww,s}^{-1/2} S_{wm,s} = (N_n' R R' N_n)^{1/2} S_{ww,s}^{-1/2} S_{wm,s} = (N_n' S_{ww,s} N_n)^{1/2} S_{ww,s}^{-1/2} S_{wm,s}.$$

Then, by the norm inequality,

$$||N_n' S_{wm,s}|| = ||(N_n' S_{ww,s} N_n)^{1/2} S_{ww,s}^{-1/2} S_{wm,s}|| \leq ||(N_n' S_{ww,s} N_n)^{1/2}|| ||S_{ww,s}^{-1/2} S_{wm,s}||.$$

Using the identity (A.2) so that $||m|| = ||m'm||^{1/2}$ and $||m^{1/2}|| = ||(m^{1/2})'(m^{1/2})||^{1/2} = ||m||^{1/2}$ we get

$$||N_n' S_{wm,s}|| \leq ||N_n' S_{ww,s} N_n||^{1/2} ||S_{ww,s}^{-1/2} S_{wm,s}|| = ||N_n' S_{ww,s} N_n||^{1/2} ||S_{mw,s} S_{ww,s}^{-1} S_{wm,s}||^{1/2}. \quad \text{(A.3)}$$

Lemma A.7 with Assumption 3.1(a) shows that, for all $\varsigma_1 > 0$ and $n \to \infty$,

$$||S_{mw,n} S_{ww,n}^{-1} S_{wm,n}|| \overset{a.s.}{=} \mathrm{o}\{\left(\log ||S_{ww,n}||^2\right)^{1+\varsigma_1}\} + \mathrm{O}(1). \quad \text{(A.4)}$$

We now argue that this implies that, for all $\varsigma_2 > 0$ and $n \to \infty$,

$$\max_{n_0 \leq s \leq n} ||S_{mw,s} S_{ww,s}^{-1} S_{wm,s}|| \overset{a.s.}{=} \mathrm{o}\{\left(\log ||S_{ww,n}||^2\right)^{1+\varsigma_2}\} + \mathrm{O}(1). \quad \text{(A.5)}$$

We prove (A.4) implies (A.5). Recall that if a sequence x_n on $\mathbb{R}$ is so that $|x_n| = \mathrm{o}(1)$, then $\max_{1 \leq s \leq n} |x_s| = \mathrm{O}(1)$. Now, let y_n and z_n be real sequences so that y_n is positive and non-decreasing and suppose that, for all $\varsigma_1 > 0$,

$$z_n = \mathrm{o}(y_n^{1+\varsigma_1}) + \mathrm{O}(1). \quad \text{(A.6)}$$

We prove that, for all $\varsigma_2 > 0$,

$$\max_{1 \leq s \leq n} |z_s| = \mathrm{o}(y_n^{1+\varsigma_2}) + \mathrm{O}(1). \quad \text{(A.7)}$$

We distinguish between bounded and diverging sequences y_n. If y_n is bounded then (A.6) amounts to $z_n = \mathrm{O}(1)$ so that $\max_{1 \leq s \leq n} |z_s| = \mathrm{O}(1)$ as required in (A.7). If y_n diverges then, for a given ς_2, choose $\varsigma_1 < \varsigma_2$. Then (A.6) amounts to $z_n = \mathrm{o}(y_n^{1+\varsigma_1})$, that is $x_n = z_n / y_n^{1+\varsigma_1} = \mathrm{o}(1)$ so that $\max_{1 \leq s \leq n} |x_s| = \mathrm{O}(1)$. We also have that $y_n^{1+\varsigma_1} = \mathrm{o}(y_n^{1+\varsigma_2})$. Combining these findings and recalling that y_s is positive and non-decreasing, we get

$$\begin{aligned}\max_{1 \leq s \leq n} |z_s| = \max_{1 \leq s \leq n} \frac{|z_s|}{y_s^{1+\varsigma_1}} y_s^{1+\varsigma_1} &\leq \max_{1 \leq s \leq n} \frac{|z_s|}{y_s^{1+\varsigma_1}} \max_{1 \leq s \leq n} y_s^{1+\varsigma_1} \\ &= (\max_{1 \leq s \leq n} |x_s|) y_n^{1+\varsigma_1} = \mathrm{O}(1)\mathrm{o}(y_n^{1+\varsigma_2}) = \mathrm{o}(y_n^{1+\varsigma_2})\end{aligned}$$

as required in (A.7). Thus to prove (A.4) implies (A.5) let $z_s = S_{mw,s} S_{ww,s}^{-1} S_{wm,s}$ and $y_n = \log ||S_{ww,n}||^2$ for each outcome in a set with probability one.

We now argue that (A.5) implies that, for all $\varsigma_3 > 0$,

$$\max_{n_0 \leq s \leq n} ||S_{mw,s} S_{ww,s}^{-1} S_{wm,s}||^{1/2} \overset{a.s.}{=} \mathrm{o}\left(||S_{ww,n}||^{\varsigma_3}\right) + \mathrm{O}(1). \quad \text{(A.8)}$$

Indeed, if $||S_{ww,n}||$ is bounded then the O(1) remainder term dominates. If $||S_{ww,n}||$ diverges we exploit that polynomials dominate logarithms so that $\left(\log ||S_{ww,n}||^2\right)^{1+\varsigma_2} = \mathrm{o}\left(||S_{ww,n}||^{\varsigma_3}\right)$ for any $\varsigma_3 > 0$.

We analyze the remainder term $||S_{ww,n}||^{\varsigma_3}$. Pre- and post-multiplying by $N_n N_n^{-1}$, using the norm inequality and the assumption $N_n^{-1} = \mathrm{O}(n^{\ell})$ $a.s.$ shows

$$||S_{ww,n}||^{\varsigma_3} = ||(N_n')^{-1} N_n' S_{ww,n} N_n N_n^{-1}||^{\varsigma_3} \overset{a.s.}{=} \mathrm{O}(n^{2\ell\varsigma_3}) ||N_n' S_{ww,n} N_n||^{\varsigma_3}.$$

Let $\varsigma_4 = 2\ell\varsigma_3 + \varsigma_3$. Then $n^{2\ell\varsigma_3} \leq n^{\varsigma_4}$ while $||N_n' S_{ww,n} N_n||^{\varsigma_3} \leq 1 + ||N_n' S_{ww,n} N_n||^{\varsigma_4}$ so that

$$||S_{ww,n}||^{\varsigma_3} \overset{a.s.}{=} \mathrm{O}(n^{\varsigma_4})(1 + ||N_n' S_{ww,n} N_n||^{\varsigma_4}).$$

Insert this in (A.8). Since $\mathrm{O}(1) = \mathrm{o}(n^{\varsigma_4})$ we get, for all $\varsigma_4 > 0$,

$$\max_{n_0 \leq s \leq n} ||S_{mw,s} S_{ww,s}^{-1} S_{wm,s}||^{1/2} \overset{a.s.}{=} \mathrm{o}(n^{\varsigma_4})(1 + ||N_n' S_{ww,n} N_n||^{\varsigma_4}).$$

Inserting this in (A.3) and noting $||S_{ww,s}||$ is non-decreasing in s we get

$$\max_{n_0\le s\le n} ||N_n' S_{wm,s}|| \le \max_{n_0\le s\le n} ||N_n' S_{ww,s} N_n||^{1/2} \max_{n_0\le s\le n} ||S_{mw,s} S_{ww,s}^{-1} S_{wm,s}||^{1/2}$$
$$\overset{a.s.}{=} ||N_n' S_{ww,n} N_n||^{1/2} \mathrm{o}(n^{\varsigma_4})(1 + ||N_n' S_{ww,n} N_n||^{\varsigma_4}).$$

Noting that for $x > 0$ then $x^{1/2}(1 + x^{\varsigma_4}) \le 2(1 + x^{1/2+\varsigma_4})$ we get that

$$\max_{n_0\le s\le n} ||N_n' S_{wm,s}|| \overset{a.s.}{=} \mathrm{o}(n^{\varsigma_4})(1 + ||N_n' S_{ww,n} N_n||^{1/2+\varsigma_4})$$

as desired with $\varsigma = \varsigma_4$. □

A.3 Proof of main results

The *CUSQ* statistic is a function of the estimators. In order to separate the randomness coming from the error terms, ε_t, and from the estimators, $\hat{\theta}_n$, we will apply the following result.

Lemma A.9 *Let $\epsilon, \phi > 0$. Suppose a compact set Θ and an $n_\epsilon > 0$ exist so that* $\mathsf{P}(|\hat{\theta}_n| \in \Theta) > 1 - \epsilon$ *for $n > n_\epsilon$. Then, for any measurable, real sequence of functions $G_n(.)$ and $n > n_\epsilon$,*

$$\mathsf{P}\{|G_n(\hat{\theta}_n)| > \phi\} \le \mathsf{P}\{\sup_{\theta\in\Theta} |G_n(\theta)| > \phi\} + \epsilon.$$

Proof of Lemma A.9: For two events $\mathcal{A}, \mathcal{B}$ we have $\mathsf{P}(\mathcal{A}) \le \mathsf{P}(\mathcal{A} \cap \mathcal{B}) + \mathsf{P}(\mathcal{B}^c)$. Thus,

$$\mathsf{P}\{|G_n(\hat{\theta}_n)| > \phi\} \le \mathsf{P}\{|G_n(\hat{\theta}_n)| > \phi, |\hat{\theta}_n| \in \Theta\} + \mathsf{P}\{|\hat{\theta}_n| \notin \Theta\}.$$

The first term is bounded by $\mathsf{P}\{\sup_{\theta\in\Theta} |G_n(\theta)| > \phi\}$ and the second term is small by assumption. Hence, the desired statement follows. □

Proof of Theorem 3.6: We use Lemma A.5 and verify Assumptions A.2 and A.3.

Part I: Assumption A.2.

1. *The problem.* Let $\mathcal{S}_{t,\theta} = n^{-1/2}\{Q_t(\theta) - Q_t(\theta_0)\}$ so that $\mathcal{S}_{t,\hat{\theta}_n} = n^{-1/2}\sum_{s=1}^t (\hat{\varepsilon}_{s,n}^2 - \varepsilon_s^2)$. We show that $\mathcal{S}_{t,\hat{\theta}_n} = \mathrm{o}_{\mathsf{P}}(1)$ uniformly in $1 \le t \le n$. 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 for the normalised parameter and estimator

$$\vartheta_n = N_{n,\theta_0}^{-1}(\theta - \theta_0), \qquad \hat{\vartheta}_n = N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0),$$

noting that under Assumption 3.3 we have $\hat{\vartheta}_n = \mathrm{o}_{\mathsf{P}}(n^\delta)$. We also let, for $1 \le s \le n$,

$$h_{s,n}(\vartheta_n) = g(x_s, \theta) = g(x_s, \theta_0 + N_{n,\theta_0}\vartheta_n),$$

so that $h_{s,n}(0) = g(x_s, \theta_0)$. The derivatives of $h_{s,n}$ with respect to ϑ_n can be expressed in terms of derivates of g with respect to θ as follows

$$\dot{h}_{s,n}(\vartheta_n) = N_{n,\theta_0}' \dot{g}(x_s, \theta), \qquad \ddot{h}_{s,n}(\vartheta_n) = N_{n,\theta_0}' \ddot{g}(x_s, \theta) N_{n,\theta_0}.$$

Let

$$\nabla\ddot{h}_{s,n}(\vartheta_n) = N_{n,\theta_0}' \{\ddot{g}(x_s, \theta) - \ddot{g}(x_s, \theta_0)\} N_{n,\theta_0}.$$

With this notation we get, for instance, that

$$(\theta - \theta_0)'\dot{g}(x_s, \theta_0) = \{N_{n,\theta_0}^{-1}(\theta - \theta_0)\}' N_{n,\theta_0}' \dot{g}(x_s, \theta_0) = \vartheta_n' \dot{h}_{s,n}(0).$$

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

$$\widetilde{\mathcal{S}}_{t,\hat{\theta}_n} = n^{-1/2}\sum_{s=1}^{t}\varepsilon_s \hat{\vartheta}_n' \dot{h}_{s,n}(0) + \frac{1}{2} n^{-1/2}\sum_{s=1}^{t}\varepsilon_s \hat{\vartheta}_n' \ddot{h}_{s,n}(\vartheta_{t,n}^*)\hat{\vartheta}_n, \tag{A.9}$$

for an intermediate point $\vartheta_{t,n}^*$ depending on the summation limit t and $\hat{\vartheta}_n$ so $||\vartheta_{t,n}^*|| \leq ||\hat{\vartheta}_n|| = \mathrm{o_P}(n^{\delta})$. For simplicity we write (A.9) as $\widetilde{\mathcal{S}}_{t,\hat{\theta}_n} = \widetilde{\mathcal{S}}_{t,n,1} + \widetilde{\mathcal{S}}_{t,n,2}/2$.

3. *The martingale term* $\widetilde{\mathcal{S}}_{t,n,1}$. The norm inequality gives

$$|\widetilde{\mathcal{S}}_{t,n,1}| \leq n^{-1/2}||\hat{\vartheta}_n|| \, ||\sum_{s=1}^{t}\varepsilon_s \dot{h}_{s,n}(0)||.$$

Apply Lemma A.8 using Assumptions 3.1, 3.5 to get, for any $\varsigma > 0$,

$$\max_{1\leq t\leq n} ||\sum_{s=1}^{t}\varepsilon_s \dot{h}_{s,n}(0)|| \overset{a.s.}{=} \mathrm{o}(n^{\varsigma})[1 + ||\sum_{s=1}^{n}\{\dot{h}_{s,n}(0)\}\{\dot{h}_{s,n}(0)\}'||^{1/2+\varsigma}].$$

Apply the triangle and norm inequalities to get, for any $\varsigma > 0$,

$$||\sum_{s=1}^{n}\{\dot{h}_{s,n}(0)\}\{\dot{h}_{s,n}(0)\}'||^{1/2+\varsigma} \leq \{\sum_{s=1}^{n}||\dot{h}_{s,n}(0)||^2\}^{1/2+\varsigma},$$

and in combination

$$\max_{1\leq t\leq n} |\widetilde{\mathcal{S}}_{t,n,1}| \overset{a.s.}{=} n^{-1/2}||\hat{\vartheta}_n||\mathrm{o}(n^{\varsigma})[1 + \{\sum_{s=1}^{n}||\dot{h}_{s,n}(0)||^2\}^{1/2+\varsigma}]. \tag{A.10}$$

Now apply that $||\hat{\vartheta}_n|| = \mathrm{o_P}(n^{\delta})$ by Assumption 3.3 while $\sum_{s=1}^{n}||\dot{h}_{s,n}(0)||^2 = \mathrm{O_P}\left(n^{1-2\delta-\eta}\right)$ for some $\eta > 0$ by Assumption 3.4(c) where $0 < \delta < 1/4$ to get

$$\max_{1\leq t\leq n} |\widetilde{\mathcal{S}}_{t,n,1}| = \mathrm{o_P}(n^{\varsigma} n^{\delta-1/2})[1 + \mathrm{O_P}\{n^{(1-2\delta-\eta)(1/2+\varsigma)}\}] = \mathrm{o_P}(1)$$

when $\varsigma > 0$ is chosen so small that $\varsigma \leq 1/2 - \delta$ and $\varsigma + \delta - 1/2 + (1 - 2\delta - \eta)(1/2 + \varsigma) < 0$. The latter is equivalent to $\varsigma(1 - \delta - \eta/2) < \eta/4$.

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

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

Apply the Hölder inequality to get

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

Assumption 3.3 shows $\hat{\vartheta}_n = N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o_P}(n^{\delta})$. The martingale Law of Large Numbers (Chow, 1965, Theorem 5) using Assumption 3.1 shows $n^{-1}\sum_{s=1}^{t}\varepsilon_s^2 \leq n^{-1}\sum_{s=1}^{n}\varepsilon_s^2 \overset{a.s.}{=} \mathrm{O}(1)$. Finally, the summands $||\ddot{h}||$ are non-negative so that

$$\max_{1\leq t\leq n} |\widetilde{\mathcal{S}}_{t,n,2}| = \mathrm{o_P}(n^{2\delta})\{\max_{1\leq t\leq n} \sum_{s=1}^{n}||\ddot{h}_{s,n}(\vartheta_{t,n}^*)||^2\}^{1/2}.$$

To show that $\max_{1\leq t\leq n} |\widetilde{\mathcal{S}}_{t,n,2}| = \mathrm{o_P}(1)$, it suffices to show that

$$\mathcal{D}_n(\hat{\vartheta}_n) = n^{4\delta} \max_{1\leq t\leq n} \sum_{s=1}^{n}||\ddot{h}_{s,n}(\vartheta_{t,n}^*)||^2 = \mathrm{O_P}(1).$$

The intermediate points $\vartheta_{t,n}^*$ all lie on the line from the origin to $\hat{\vartheta}_n$, hence $\mathcal{D}_n(\hat{\vartheta}_n)$ is a function of $\hat{\vartheta}_n$.

5. *The term* $\mathcal{D}_n(\hat{\vartheta}_n)$. Since $||n^{-\delta}\hat{\vartheta}_n|| = \mathrm{o_P}(1)$ by Assumption 3.3 we get from Lemma A.9 that it suffices to replace $n^{-\delta}\hat{\vartheta}_n$ by deterministic $n^{-\delta}\vartheta$ varying in a compact region with high probability. That is, we have that $\mathcal{D}_n(\hat{\vartheta}_n) = \mathrm{O_P}(1)$ if, for some $\epsilon > 0$, it is shown that $\sup_{||n^{-\delta}\vartheta||\leq\epsilon} \mathcal{D}_n(\vartheta) = \mathrm{O_P}(1)$. The effect of replacing $n^{-\delta}\hat{\vartheta}_n$ by $n^{-\delta}\vartheta$ on the intermediate points is captured as follows. For

$t = 1, \dots, n$, the original intermediate points $\vartheta^*_{t,n}$ are on the line from the origin to $\hat{\vartheta}_n$ and satisfy $||\vartheta^*_{t,n}|| \le ||\hat{\vartheta}_n||$. They are now replaced by intermediate points $\vartheta^*_{t,\vartheta}$ on the line from the origin to ϑ, where the index n in $\vartheta^*_{t,n}$ has been replaced by ϑ in $\vartheta^*_{t,\vartheta}$ to emphasize the dependence on the point ϑ replacing $\hat{\vartheta}_n$. The intermediate points $\vartheta^*_{t,\vartheta}$ remain stochastic as they stem from the mean value expansion in (A.9).

We now bound the function $\mathcal{D}_n(\vartheta) = n^{4\delta} \max_{1\le t\le n} \sum_{s=1}^n ||\ddot{h}_{s,n}(\vartheta^*_{t,\vartheta})||^2$ as follows. Since $\vartheta^*_{t,\vartheta}$ lies between zero and some point ϑ so that $||\vartheta|| \le \epsilon n^\delta$, then, for each t, we can bound

$$\sum_{s=1}^n ||\ddot{h}_{s,n}(\vartheta^*_{t,\vartheta})||^2 \le \sup_{||\vartheta|| \le \epsilon n^\delta} \sum_{s=1}^n ||\ddot{h}_{s,n}(\vartheta)||^2.$$

The bound is uniform in t so that

$$\mathcal{D}_n(\vartheta) \le n^{4\delta} \sup_{||\vartheta|| \le \epsilon n^\delta} \sum_{s=1}^n ||\ddot{h}_{s,n}(\vartheta)||^2.$$

It is also uniform in $||\vartheta|| \le \epsilon n^\delta$ so that

$$\sup_{||n^{-\delta}\vartheta|| \le \epsilon} \mathcal{D}_n(\vartheta) \le n^{4\delta} \sup_{||\vartheta|| \le \epsilon n^\delta} \sum_{s=1}^n ||\ddot{h}_{s,n}(\vartheta)||^2,$$

which is $\mathrm{o}_\mathsf{P}(1)$ by Assumption 3.4(d).

6. *The compensator.* The compensator term can be written as

$$\overline{\mathcal{S}}_{t,\theta} = n^{-1/2} \sum_{s=1}^t \{\nabla g_s(\theta)\}^2 = n^{-1/2} \sum_{s=1}^t \{g(x_s, \theta_0 + N_{n,\theta_0}\vartheta) - g(x_s, \theta_0)\}^2 = \overline{\mathcal{S}}_{t,\theta_0 + N_{n,\theta_0}\vartheta}. \quad \text{(A.11)}$$

Since $\hat{\vartheta}_n = N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o}_\mathsf{P}(n^\delta)$ then, by Lemma A.9 we get $\overline{\mathcal{S}}_{t,\hat{\theta}_n} = \overline{\mathcal{S}}_{t,\theta_0 + N_{n,\theta_0}\hat{\vartheta}_n} = \mathrm{o}_\mathsf{P}(1)$ uniformly in t if $\mathcal{A}_n = \sup_{||\vartheta|| \le \epsilon n^\delta} \max_{1\le t\le n} \overline{\mathcal{S}}_{t,\theta_0 + N_{n,\theta_0}\vartheta}$ vanishes. Now, $\overline{\mathcal{S}}_{t,\theta_0 + N_{n,\theta_0}\vartheta} \le \overline{\mathcal{S}}_{n,\theta_0 + N_{n,\theta_0}\vartheta}$ together with Assumption 3.4(a) ensures that $\mathcal{A}_n \le \sup_{||\vartheta|| \le \epsilon n^\delta} \overline{\mathcal{S}}_{n,\theta_0 + N_{n,\theta_0}\vartheta} = \mathrm{o}_\mathsf{P}(1)$.

Part II: Assumption A.3.

1. *The problem.* Let $\mathcal{V}_{n,\theta} = n^{-1} \sum_{t=1}^n [\{\varepsilon_t - \nabla g_s(\theta)\}^4 - \varepsilon_t^4]$ where $\nabla g_s(\theta) = g(x_s, \theta) - g(x_s, \theta_0)$ as before, so that $\mathcal{V}_{n,\hat{\theta}_n} = n^{-1} \sum_{t=1}^n (\hat{\varepsilon}_t^4 - \varepsilon_t^4)$. By Lemma A.9, using Assumption 3.3, it suffices to show that $\mathcal{V}_{n,\theta} = \mathcal{V}_{n,\theta_0 + N_{n,\theta_0}\vartheta} = \mathrm{o}_\mathsf{P}(1)$ uniformly over $||\vartheta|| \le \epsilon n^\delta$ or equivalently $||N_{n,\theta_0}^{-1}(\theta - \theta_0)|| \le \epsilon n^\delta$.

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, the triangle and Hölder inequalities give

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

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(b) while $n^{-1} \sum_{t=1}^n \{\nabla g_s(\theta)\}^4 = \mathrm{o}_\mathsf{P}(1)$ by Assumption 3.4(b). □

Proof of Theorem 3.8. We use Lemma A.6 and verify Assumptions A.3, A.4.

Part I: Assumption A.4: We show $B_n = \max_{n_0 \le t \le n} |n^{-1/2} \sum_{s=1}^t (\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)| = \mathrm{o}_\mathsf{P}(1)$.

First, we turn the strong convergence properties in Assumption 3.7 into uniform properties using Egorov's theorem (Davidson 1994, Theorem 18.4). Choose $\nu, \tau, \varphi > 0$ small. The strong convergence assumptions are that $n^{-\delta}\hat{\vartheta}_n = n^{-\delta} N_{n,\theta_0}^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{o}(1)$ *a.s.* and

$$\mathcal{A}_n^* = n^{-1/2} \sup_{\theta: ||N_{n,\theta_0}^{-1}(\theta - \theta_0)|| \le n^\delta \epsilon} \sum_{s=1}^n \{\nabla g(x_s, \theta)\}^2 \overset{a.s.}{=} \mathrm{o}(1), \quad \text{(A.12)}$$

$$\mathcal{C}_n^* = n^{2\delta + \eta - 1} \sum_{s=1}^n ||N_{n,\theta_0}' \dot{g}(x_s, \theta_0)||^2 \overset{a.s.}{=} \mathrm{o}(1), \quad \text{(A.13)}$$

$$\mathcal{D}_n^* = n^{4\delta} \sup_{\theta: ||N_{n,\theta_0}^{-1}(\theta - \theta_0)|| \le n^\delta \epsilon} \sum_{s=1}^n ||\ddot{h}_{s,n}(x_s, \theta)||^2 \overset{a.s.}{=} \mathrm{o}(1). \quad \text{(A.14)}$$

Egorov's theorem implies that for all $\nu, \tau > 0$ there exists a $t_0 \geq n_0$ and a set Ω_ν with probability $\mathsf{P}(\Omega_\nu) \geq 1-\nu$ so that for $t \geq t_0$ then $t^{-\delta}||\hat{\vartheta}_t||$, $\mathcal{A}^*_t$, $\mathcal{C}^*_t$ and $\mathcal{D}^*_t$ are all bounded by τ on Ω_ν. We will consider small $\tau > 0$ so we can assume $\tau < \epsilon$.

Second, we want to show $B_n = \mathrm{o}_{\mathsf{P}}(1)$, that is for all $\epsilon > 0$ there exists n_ϵ so that for $n \geq n_\epsilon$ then $\mathsf{P}(|B_n| > \epsilon) < 2\epsilon$. Now, bound the probability so that $\mathsf{P}(|B_n| > \epsilon) \leq \mathsf{P}(|B_n| > \epsilon, \Omega_\nu) + \mathsf{P}(\Omega_\nu^c)$. The latter probability is small if ν is small. Thus it suffices to show $B_n = \mathrm{o}_{\mathsf{P}}(1)$ on Ω_ν. We bound

$$B_n \leq n^{-1/2} \max_{n_0 \leq t < t_0} |\textstyle\sum_{s=1}^t (\hat{\varepsilon}^2_{s,t} - \varepsilon^2_s)| + \max_{t_0 \leq t \leq n} |n^{-1/2} \textstyle\sum_{s=1}^t (\hat{\varepsilon}^2_{s,t} - \varepsilon^2_s)|.$$

Since t_0 is finite and the summands of the first sum only depend on t_0 and not n, then the normalization by $n^{-1/2}$ ensures that first term is $\mathrm{O}_{\mathsf{P}}(n^{-1/2})$ and vanishes. We show that the second term vanishes. For this we follow part I in the proof of Theorem 3.6 with a series of smaller adjustments.

1. *The problem.* We now want to show $\mathcal{S}_{t,\hat{\theta}_t} = n^{-1/2} \sum_{s=1}^t (\hat{\varepsilon}^2_{s,t} - \varepsilon^2_s) = \mathrm{o}_{\mathsf{P}}(1)$ uniformly in $t_0 \leq t \leq n$.

2. *Expand the martingale* $\widetilde{\mathcal{S}}_{t,\hat{\theta}_t}$. With the notation in the proof of Theorem 3.6 we have

$$\vartheta_t = N^{-1}_{t,\theta_0}(\theta - \theta_0), \qquad \dot{h}_{s,t}(\vartheta_t) = N'_{t,\theta_0}\dot{g}(x_s, \theta), \qquad \ddot{h}_{s,t}(\vartheta_t) = N'_{t,\theta_0}\ddot{g}(x_s, \theta)N_{t,\theta_0}. \tag{A.15}$$

Thus we can expand $\widetilde{\mathcal{S}}_{t,\hat{\theta}_t} = \widetilde{\mathcal{S}}_{t,t,1} + \widetilde{\mathcal{S}}_{t,t,2}/2$ as in (A.9) where

$$\widetilde{\mathcal{S}}_{t,t,1} = n^{-1/2} \textstyle\sum_{s=1}^t \varepsilon_s \hat{\vartheta}'_t \dot{h}_{s,t}(0), \qquad \widetilde{\mathcal{S}}_{t,t,2} = n^{-1/2} \textstyle\sum_{s=1}^t \varepsilon_s \hat{\vartheta}'_t \ddot{h}_{s,t}(\vartheta^*_{t,t})\hat{\vartheta}_t,$$

for an intermediate point $\vartheta^*_{t,t}$ depending on the summation limit t and $\hat{\vartheta}_t$ so $||\vartheta^*_{t,t}|| \leq ||\hat{\vartheta}_t||$.

3. *The martingale term* $\widetilde{\mathcal{S}}_{t,t,1}$. Here we will use the $\dot{g}$ notation initially. Thus, let $\dot{g}_s = \dot{g}(x_s, \theta_0)$ and write

$$\widetilde{\mathcal{S}}_{t,t,1} = n^{-1/2} \hat{\vartheta}'_t N'_{t,\theta_0} \textstyle\sum_{s=1}^t \dot{g}_s \varepsilon_s. \tag{A.16}$$

Apply Lemma A.8 with $n = t$ to get, for all $\varsigma > 0$,

$$N'_{t,\theta_0} \textstyle\sum_{s=1}^t \dot{g}_s \varepsilon_s \overset{a.s.}{=} \mathrm{o}(t^\varsigma)(1 + ||\textstyle\sum_{s=1}^t N'_{t,\theta_0}\dot{g}_s\dot{g}'_s N_{t,\theta_0}||^{1/2+\varsigma}). \tag{A.17}$$

Recall that $\dot{h}_s(0) = N'_{t,\theta_0}\dot{g}(x_s, \theta_0)$. Use the triangle and norm inequalities to get

$$\mathcal{T}_t = ||\textstyle\sum_{s=1}^t N'_{t,\theta_0}\dot{g}_s\dot{g}'_s N_{t,\theta_0}|| = ||\textstyle\sum_{s=1}^t \{\dot{h}_{s,t}(0)\}\{\dot{h}_{s,t}(0)\}'|| \leq \textstyle\sum_{s=1}^t ||\dot{h}_{s,t}(0)||^2.$$

Using the definitions of $\mathcal{C}^*$ and $\dot{h}$ in (A.13), (A.15), respectively, gives the further bound

$$\mathcal{T}_t \leq \textstyle\sum_{s=1}^t ||\dot{h}_{s,t}(0)||^2 = t^{1-2\delta-\eta}\mathcal{C}^*_t \leq t^{1-2\delta-\eta}\tau.$$

Insert this in the expression (A.17) while noting that if a sequence x_t satisfies $x_t = \mathrm{o}(t^\varsigma)$ then $\max_{t_0 \leq t \leq n} |x_t| = \mathrm{O}(n^\varsigma)$ so that

$$\max_{t_0 \leq t \leq n} |N'_{t,\theta_0} \textstyle\sum_{s=1}^t \dot{g}_s \varepsilon_s| \overset{a.s.}{=} \mathrm{O}(n^\varsigma)\{1 + (n^{1-2\delta-\eta}\tau)^{1/2+\varsigma}\}.$$

Insert this in the expression for $\widetilde{\mathcal{S}}_{t,t,1}$ in (A.16) along with the bound $||\hat{\vartheta}_t|| \leq t^\delta\tau \leq n^\delta\tau$ on Ω_ν to get, for all $\varsigma > 0$,

$$\max_{t_0 \leq t \leq n} |\widetilde{\mathcal{S}}_{t,t,1}| \overset{a.s.}{=} \tau\mathrm{O}(n^{-1/2+\delta+\varsigma})\{1 + (n^{1-2\delta-\eta}\tau)^{1/2+\varsigma}\} = \mathrm{o}(1),$$

since $\varsigma > 0$ can be chosen so small that $-1/2 + \delta + \varsigma < 0$ and $\delta + \varsigma + (1 - 2\delta - \eta)(1/2 + \varsigma) < 1/2$ or equivalently $2\varsigma(1 - \delta - \eta/2) < \eta/2$.

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

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

On Ω_ν then $||\hat{\vartheta}_t|| \le \tau t^\delta$. The martingale Law of Large Numbers by Chow (1965, Theorem 5) using Assumption 3.1 shows $n^{-1}\sum_{s=1}^t \varepsilon_s^2 \le n^{-1}\sum_{s=1}^n \varepsilon_s^2 = \mathrm{O}(1)$ on a set $\Omega_{a.s.}$ with probability one. Note $\mathsf{P}(\Omega_\nu \cap \Omega_{a.s.}^c) \le \mathsf{P}(\Omega_{a.s.}^c) = 0$. Thus, on $\Omega_\nu \cap \Omega_{a.s.}$, where $\mathsf{P}(\Omega_\nu \cap \Omega_{a.s.}) = \mathsf{P}(\Omega_\nu)$, we have

$$\max_{t_0 \le t \le n} |\widetilde{\mathcal{S}}_{t,t,2}| = \mathrm{O}(\tau^2)\mathcal{D}_n^{1/2} \qquad \text{where} \qquad \mathcal{D}_n = \max_{t_0 \le t \le n} t^{4\delta} \textstyle\sum_{s=1}^t ||\ddot{h}_{s,t}(\vartheta_{t,t}^*)||^2.$$

Since τ can be chosen arbitrarily small then we get $\max_{t_0 \le t \le n} |\widetilde{\mathcal{S}}_{t,t,2}| = \mathrm{o}_{\mathsf{P}}(1)$ on the general sample space Ω if $\mathcal{D}_n = \mathrm{O}(1)$ on $\Omega_\nu \cap \Omega_{a.s.} \subset \Omega_\nu$.
5. *The term* $\mathcal{D}_n$. Since $\vartheta_{t,t}^*$ is on the line from 0 to $\hat{\vartheta}_t$, we get $||\vartheta_{t,t}^*||^2 \le ||\hat{\vartheta}_t||^2$. On Ω_ν we have $t^{-\delta}||\hat{\vartheta}_t|| \le \tau < \epsilon$ so we get, see (A.14),

$$\mathcal{D}_n \le \max_{t_0 \le t \le n} \sup_{||\vartheta|| \le \epsilon t^\delta} t^{4\delta} \textstyle\sum_{s=1}^t ||\ddot{h}_{s,t}(\vartheta)||^2 = \max_{t_0 \le t \le n} \mathcal{D}_t^* \le \tau = \mathrm{O}(1).$$

6. *The compensator.* We show $\mathcal{A}_n = \max_{n_0 \le t \le n} \overline{\mathcal{S}}_{t,\hat{\theta}_t} = \mathrm{o}_{\mathsf{P}}(1)$ on Ω_ν, where $\overline{\mathcal{S}}_{t,\theta} = n^{-1/2}\sum_{s=1}^t \{\nabla g_s(\theta)\}^2$. Recall from (A.11) that $\overline{\mathcal{S}}_{t,\theta} = \overline{\mathcal{S}}_{t,\theta_0 + N_{n,\theta_0}\vartheta}$. Since $\hat{\vartheta}_t = N_{t,\theta_0}^{-1}(\hat{\theta}_t - \theta)$, where $t^{-\delta}||\hat{\vartheta}_t|| \le \tau < \epsilon$ on Ω_ν, see (A.12), then on Ω_ν,

$$\mathcal{A}_n \le \max_{n_0 \le t \le n} \sup_{||\vartheta|| \le \epsilon t^\delta} \overline{\mathcal{S}}_{t,\theta_0 + N_{t,\theta_0}\vartheta} = \max_{n_0 \le t \le n} \sup_{\theta: ||N_{t,\theta_0}^{-1}(\theta - \theta_0)|| \le \epsilon t^\delta} \overline{\mathcal{S}}_{t,\theta} = \max_{n_0 \le t \le n} \mathcal{A}_t^* \le \tau.$$

Since τ can be chosen arbitrarily small then $\mathcal{A}_n = \mathrm{o}_{\mathsf{P}}(1)$.

Part II: Assumption A.3: Since the Assumptions of Theorem 3.8 imply those of Theorem 3.6 then Part II of the proof of Theorem 3.6 applies. □

Proof of Theorem 3.10. Assumption 3.4(c) follows from Assumption 3.9(a) with $k = 2$.

Assumption 3.4(d) is the same as Assumption 3.9(b) with $k = 2$.

Assumption 3.4(a, b). Recall the notation in item 3 in the proof of Theorem 3.6 and expand

$$g(x_t, \theta) - g(x_t, \theta_0) = \vartheta_n' \dot{h}_{t,n}(0) + \frac{1}{2}\vartheta_n' \ddot{h}_{t,n}(\vartheta_{t,n}^*)\vartheta_n,$$

where $\vartheta_{t,n}^*$ is an intermediate point depending on x_t so $||\vartheta_{t,n}^*|| \le ||\vartheta_n||$. Raise this to the power $k = 2$ or $k = 4$ and apply the inequality $|x + y|^k \le C(|x|^k + |y|^k)$ to see that

$$|g(x_t, \theta) - g(x_t, \theta_0)|^k \le C||\vartheta_n||^k ||\dot{h}_{t,n}(0)||^k + C||\vartheta_n||^{2k} ||\ddot{h}_{t,n}(\vartheta_{t,n}^*)||^k.$$

We want to show Assumption 3.4(a, b), hence, we only have to consider $||\vartheta_n|| \le \epsilon n^\delta$. Also, since ϑ_t is in the line between 0 and ϑ_n we have that $||\ddot{h}_{t,n}(\vartheta_{t,n}^*)||^k \le \sup_{||\vartheta|| \le \epsilon n^\delta} ||\ddot{h}_{t,n}(\vartheta)||^k$. Then cumulate to get

$$|\textstyle\sum_{t=1}^n \{g(x_t, \theta) - g(x_t, \theta_0)\}^k| \le C\epsilon^k n^{\delta k} \sum_{t=1}^n ||\dot{h}_{t,n}(0)||^k + C\epsilon^{2k} n^{2\delta k} \sum_{t=1}^n \sup_{||\vartheta|| \le \epsilon n^\delta} ||\ddot{h}_{t,n}(\vartheta)||^k,$$

which is $\mathrm{o}_{\mathsf{P}}(n^{1/2})$ for $k = 2$ and $\mathrm{o}_{\mathsf{P}}(n)$ for $k = 4$ uniformly in θ due to Assumption 3.9. □

A.4 Proof for autoregressive distributed lag model

We first appeal to the following result, that in turn builds on Lai and Wei (1982, 1985) recalling that the normalization is chosen as $N_n{}^{-1} = (\sum_{t=1}^n x_t x_t')^{1/2}$.

Lemma A.10 (Nielsen 2005, Theorems 2.4, 7.1, Lemma 8.2) *Consider model (4.1) with* $\theta \in \Theta = \mathbb{R}^p$ *and Assumptions 3.1, 4.1. Then*
(i) $N_n^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{O}\{(\log n)^{1/2}\}$ *a.s.;*
(ii) $(N_n N_n')^{-1} = \sum_{t=1}^n x_t x_t' = \mathrm{O}(n^{2\ell})$ *a.s. for some finite* $\ell > 0$*;*
(iii) $n_0 = \inf(n : \sum_{t=1}^n x_t x_t'$ *is invertible*$) < \infty$ *a.s.*

The assumptions for the $CUSQ_n$ and $RCUSQ_n$ statistics are now easily demonstrated.

Proof of Theorem 4.2. We apply Theorems 3.6, 3.8 and check Assumptions 3.1-3.5 and 3.7. *Assumption 3.1* follows by Assumption 4.1(a). *Assumption 3.2* holds since N_n^{-1} grows at most polynomially by Lemma A.10(ii). *Assumption 3.3* holds $a.s.$ for any $\delta > 0$ due to Lemma A.10(i). *Assumption 3.4* holds $a.s.$ as follows. For (a, b) use that $g(x_t, \theta) - g(x_t, \theta_0) = x_t' N_n N_n^{-1}(\theta - \theta_0)$ so that

$$\mathcal{G} = \sum_{t=1}^n |g(x_t, \theta) - g(x_t, \theta_0)|^k \leq ||N_n^{-1}(\theta - \theta_0)||^k \sum_{t=1}^n ||N_n' x_t||^k,$$

for $k = 2, 4$. Since $\sum_{t=1}^n N_n' x_t x_t' N_n = I$ then $||N_n' x_t||^2 \leq 1$ so that $\sum_{t=1}^n ||N_n' x_t||^k \leq \sum_{t=1}^n ||N_n' x_t||^2$ for $k = 2, 4$. For a vector v the spectral norm satisfies $||v||^2 = v'v = \mathrm{tr}(vv')$. Thus, $\sum_{t=1}^n ||N_n' x_t||^2 \leq \mathrm{tr} \sum_{t=1}^n N_n' x_t x_t' N_n = \mathrm{tr}(I_p) = p$. Noting that $N_n^{-1}(\theta - \theta_0) = \mathrm{o}(n^\delta)$ $a.s.$ we get $\mathcal{G} \overset{a.s.}{=} \mathrm{o}(n^{k\delta})$. For ($c$) use that $\dot{g}(x_t, \theta_0) = x_t$ so that $\sum_{t=1}^n ||N' \dot{g}||^2 = \sum_{t=1}^n ||N_n' x_t||^2 \leq p$. For ($d$) note that $\ddot{g} = 0$. *Assumption 3.5* holds by Lemma A.10(iii). *Assumption 3.7* also holds because Assumptions 3.3, 3.4 were demonstrated $a.s.$ □

A.5 Proofs for separable models with homogeneous functions

We need a Law of Iterated Logarithm for the regressor x_t satisfying Assumption 4.3. If the intercept μ in the vector autoregressive component is absent we can appeal to Lai and Wei (1985, Theorem 1). For general μ we need the following generalization.

Lemma A.11 ***(Nielsen 2005, Theorem 5.1,i,ii)*** *Suppose x_t satisfies Assumption 4.3. Then, for some $\omega > 0$, we have $x_n = \sum_{s=1}^{n-1} \eta_s + \mathrm{o}(n^{1/2-\omega}) = \mathrm{O}\{(n \log\log n)^{1/2}\}$ a.s.*

For almost sure arguments we use a stochastic normalization matrix $N_n = [\sum_{t=1}^n \{g(x_t)\}^2]^{-1/2}$. This has the following property.

Lemma A.12 *Let Assumptions 4.3, 4.4(i) hold. Let $N_n^{-2} = \sum_{t=1}^n \{g(x_t)\}^2$. Then $N_n^{-2} = \mathrm{O}(n^\ell)$ a.s. for some $\ell > 0$.*

Proof of Lemma A.12. The homogeneity of g in Assumption 4.4(i) yields

$$N_n^{-2} = n\{g(n^{1/2})\}^2 \frac{1}{n} \sum_{t=1}^n \{g(x_t)/g(n^{1/2})\}^2 = n\{g(n^{1/2})\}^2 \frac{1}{n} \sum_{t=1}^n \{g(x_t/n^{1/2})\}^2.$$

The bound $g(x) \leq C(1 + |x|^{\ell_1})$ for some finite $C, \ell_1 > 0$ in Assumption 4.4(i) gives that $\{g(n^{1/2})\}^2 = \mathrm{O}(n^{\ell_1})$. Moreover we get $|g(x_t/n^{1/2})| \leq C(1 + |x_t/n^{1/2}|^{\ell_1})$. Note that $x_t/n^{1/2}$ can be written as $(x_t/\sqrt{t \log\log t})(\sqrt{t \log\log t}/n^{1/2})$. By Lemma A.11, using Assumption 4.3, we have that for almost every realization, there exists a C_1 so that $x_t/\sqrt{t \log\log t} \leq C_1$. Since $\sqrt{t \log\log t}/n^{1/2} \leq \sqrt{\log\log n}$, then $\max_{1 \leq t \leq n} |x_t/n^{1/2}| = \mathrm{O}\{(\log\log n)^{1/2}\}$ $a.s.$ and therefore

$$n^{-1} \sum_{t=1}^n \{g(x_t/n^{1/2})\}^2 \leq n^{-1} \sum_{t=1}^n C^2 (1 + |x_t/n^{1/2}|^{\ell_1})^2.$$

Since, $(x + y)^2 \leq 2(x^2 + y^2)$ we get

$$n^{-1} \sum_{t=1}^n \{g(x_t/n^{1/2})\}^2 \leq 2C^2 (1 + \max_{1 \leq t \leq n} |x_t/n^{1/2}|^{2\ell_1}) \overset{a.s.}{=} \mathrm{O}\{(\log\log n)^{\ell_1}\}.$$

In combination, we get $N_n^{-2} = n\mathrm{O}(n^{\ell_1})\mathrm{O}\{(\log\log n)^{\ell_1}\}$ so that $N_n^{-2} = \mathrm{O}(n^{\ell_2})$ $a.s.$ for some finite $\ell_2 \geq 2 + \ell_1$. □

Proof of Theorem 4.5. By Lemma A.7 using Assumptions 3.1, 4.4(ii), we get

$$\{N_n^{-1}(\hat{\theta}_n - \theta)\}^2 = \frac{\{\sum_{t=1}^n g(x_t)\varepsilon_t\}^2}{\sum_{t=1}^n \{g(x_t)\}^2} \overset{a.s.}{=} \mathrm{o}\{(\log[\sum_{t=1}^n \{g(x_t)\}^2])^{1+\varsigma}\} + \mathrm{O}(1) \qquad \text{for all } \varsigma > 0.$$

Since $\sum_{t=1}^{n}\{g(x_t)\}^2 = N_n^{-2} = \mathrm{O}(n^{\ell})$ *a.s.* for some finite ℓ by Lemma A.12 using Assumptions 4.3, 4.4(i) then $N_n^{-1}(\hat{\theta}_n - \theta) = \mathrm{o}\{(\log n)^{1+\varsigma}\} = \mathrm{o}(n^{\delta})$ *a.s.* for all $\delta > 0$. □

Proof of Theorem 4.6. We apply Theorems 3.6, 3.8 and check Assumptions 3.1-3.5 and 3.7. Note $g(x_t, \theta) = \theta g(x_t)$ so that $\dot{g}(x_t, \theta) = g(x_t)$ and $\ddot{g}(x_t, \theta) = 0$. *Assumption 3.1* is assumed.

Assumption 3.2: $N_n^{-2} = \sum_{t=1}^{n}\{g(x_t)\}^2$ is of polynomial order by Lemma A.12 using Assumptions 4.3, 4.4. We also have that N_n^{-2} is invertible for $n > n_0$ by Assumption 4.4(ii).

Assumption 3.3 holds *a.s.* for any $\delta > 0$ by Theorem 4.5 using Assumptions 3.1, 4.3, 4.4.

Assumption 3.4: For (a, b) use that $g(x_t, \theta) - g(x_t, \theta_0) = (\theta - \theta_0)g(x_t)$ so that for $k = 2, 4$,

$$\mathcal{G} = \sum_{t=1}^{n}|g(x_t, \theta) - g(x_t, \theta_0)|^k \leq |N_n^{-1}(\theta - \theta_0)|^k \sum_{t=1}^{n}|N_n g(x_t)|^k.$$

Since $N_n^2 \sum_{t=1}^{n}\{g(x_t)\}^2 = 1$ then $N_n^2\{g(x_t)\}^2 \leq 1$ and $\sum_{t=1}^{n}|N_n g(x_t)|^k \leq \sum_{t=1}^{n}|N_n g(x_t)|^2 = 1$ while $N_n^{-1}(\theta - \theta_0) = \mathrm{o}(n^{\delta})$ *a.s.* and therefore $\mathcal{G} = \mathrm{o}(n^{k\delta})$ *a.s.* For (c) we get $\sum_{t=1}^{n}|N_n \dot{g}(x_t, \theta_0)|^2 = \sum_{t=1}^{n}|N_n g(x_t)|^2 = 1$. For (d) note that $\ddot{g} = 0$.

Assumption 3.5 is assumed in Assumption 4.4(ii).

Finally, *Assumption 3.7* holds because Assumptions 3.3, 3.4 were demonstrated *a.s.* □

A.6 Proofs for separable models with asymptotically homogeneous functions

Proof of Thorem 4.10. *Part I:* By Assumption 4.3, $x_t = \sum_{s=1}^{t-1}\eta_s + \alpha'\mathbf{x}_{t-1}$. Hence, by Lemma A.11 using Assumption 4.3,

$$n^{-1/2}x_{[un]} = n^{-1/2}\sum_{s=1}^{[un]-1}\eta_s + n^{-1/2}\alpha'\mathbf{x}_{[un]-1} \overset{a.s.}{=} n^{-1/2}\sum_{s=1}^{[un]-1}\eta_s + \mathrm{o}(1).$$

The random walk term satisfies the functional central limit theorem by Brown (1971) using Assumption 4.3. Hence, the result follows.

Part II: Since T is asymptotically homogeneous we can write

$$T(x_t) = T(n^{1/2}n^{-1/2}x_t) = v(n^{1/2})H(n^{-1/2}x_t) + R(n^{-1/2}x_t, n^{1/2}),$$

where $H(x)$ is locally integrable and $|R(n^{-1/2}x_t, n^{1/2})|$ is bounded as described in Definition 4.8. Hence,

$$\{nv(n^{1/2})\}^{-1}\sum_{t=1}^{n}T(x_t) = n^{-1}\sum_{t=1}^{n}H(n^{-1/2}x_t) + \{nv(n^{1/2})\}^{-1}\sum_{t=1}^{n}R(n^{-1/2}x_t, n^{1/2}),$$

where the first term satisfies

$$n^{-1}\sum_{t=1}^{n}H(n^{-1/2}x_t) \overset{\mathsf{D}}{\to} \int_0^1 H(\sigma_\eta W_u)du,$$

by Theorem 2.1 in Pötscher (2004) and the second term vanishes as we next show. By Definition 4.8, the remainder term $R(n^{-1/2}x_t, n^{1/2})$ can be bounded by

$$|R(n^{-1/2}x_t, n^{1/2})| \leq a(n^{1/2})P(n^{-1/2}x_t),$$

where $\sup_{n\to\infty} a(n^{1/2})/v(n^{1/2}) = 0$ and P is locally integrable. Thus, by the triangle inequality,

$$|\{nv(n^{1/2})\}^{-1}\sum_{t=1}^{n}R(n^{-1/2}x_t, n^{1/2})| \leq \{a(n^{1/2})/v(n^{1/2})\}n^{-1}\sum_{t=1}^{n}P(n^{-1/2}x_t) = \mathrm{o}_{\mathsf{P}}(1),$$

applying again Theorem 2.1. in Pötscher (2004) to the term $n^{-1}\sum_{t=1}^{n}P(n^{-1/2}x_t)$. □

Proof of Thorem 4.13. Let $N_n^{-1} = [\sum_{t=1}^{n}\{g(x_t)\}^2]^{1/2}$. By Lemma A.7 using Assumptions 3.1, 4.12, we get, for all $\varsigma > 0$,

$$\{N_n^{-1}(\hat{\theta}_n - \theta)\}^2 = \frac{\{\sum_{t=1}^{n} g(x_t)\varepsilon_t\}^2}{\sum_{t=1}^{n}\{g(x_t)\}^2} \overset{a.s.}{=} \mathrm{o}\{(\log[\sum_{t=1}^{n}\{g(x_t)\}^2])^{1+\varsigma}\} + \mathrm{O}(1). \quad \text{(A.18)}$$

By Theorem 4.10 using Assumption 4.11, we have

$$N_n^{-2} = \sum_{t=1}^n \{g(x_t)\}^2 = \mathrm{O}_{\mathsf{P}}\{nv(n^{1/2})\} = \mathrm{O}_{\mathsf{P}}(n^{1+\ell}), \tag{A.19}$$

where the last equality holds since $v(n^{1/2}) = \mathrm{o}(n^\ell)$ for some $\ell > 0$ by Definition 4.8. As a consequence

$$(\log[\sum_{t=1}^n \{g(x_t)\}^2])^{1+\varsigma} = \mathrm{o}_{\mathsf{P}}(n^\omega),$$

for any $\omega > 0$. Inserting this into (A.18) shows that

$$\{N_n^{-1}(\hat{\theta}_n - \theta)\}^2 = \frac{\{\sum_{t=1}^n g(x_t)\varepsilon_t\}^2}{\sum_{t=1}^n \{g(x_t)\}^2} = \mathrm{o}_{\mathsf{P}}(n^\omega),$$

which gives $N_n^{-1}(\hat{\theta}_n - \theta) = \mathrm{o}_{\mathsf{P}}(n^\delta)$ for $\delta = \omega/2$ as desired. □

Proof of Thorem 4.14. We apply Theorem 3.10 so that Assumption 3.9 implies Assumption 3.4 and Theorem 3.6. Hence we need to check Assumptions 3.1, 3.2, 3.3, 3.5 and 3.9. Note $g(x_t, \theta) = \theta g(x_t)$ so that $\dot{g}(x_t, \theta) = g(x_t)$ and $\ddot{g}(x_t, \theta) = 0$. *Assumption 3.1* is assumed.

Assumption 3.2: By Theorem 4.10 using Assumption 4.11 , we have $N_n^{-2} = \sum_{t=1}^n \{g(x_t)\}^2 = \mathrm{O}_{\mathsf{P}}(n^{1+\ell})$ as shown in the proof of Theorem 4.13 –see (A.19). We also have that N_n^{-2} is invertible for $n > n_0$ by Assumption 4.12.

Assumption 3.3 holds by Theorem 4.13 using Assumptions 3.1, 4.3, 4.9, 4.11, 4.12.

Assumption 3.5 is assumed in Assumption 4.12.

Assumption 3.9: For (a) let $k = 2, 4$ and consider $\sum_{t=1}^n |N_n g(x_t)|^k$. Since $N_n^2 \sum_{t=1}^n \{g(x_t)\}^2 = 1$ then $N_n^2 \{g(x_t)\}^2 \leq 1$ and $\sum_{t=1}^n |N_n g(x_t)|^k \leq \sum_{t=1}^n |N_n g(x_t)|^2 = 1 = \mathrm{o}_{\mathsf{P}}(n^{k/4 - k\delta})$ for $0 < \delta < 1/4$. For (b) note that $\ddot{g} = 0$. □

A.7 Proofs for power curve model

We start with some algebraic manipulations of the criterion function. Since the data generating process satisfies $y_t = (x_t + \theta_0)^2 + \varepsilon_t$, the criterion function is

$$Q_n(\theta) = \sum_{t=1}^n \{y_t - (x_t + \theta)^2\}^2 = \sum_{t=1}^n \{\varepsilon_t + (x_t + \theta_0)^2 - (x_t + \theta)^2\}^2. \tag{A.20}$$

This is a fourth order random polynomial and as such can have one or two local minimizers. Indeed, the fourth order polynomial has (a) one minimizer when its derivative, a third order polynomial, has only one real root; and (b) two local minimizers when its derivative has three distinct real roots. We will analyze these minimizers, first, for general regressors in Theorem A.17 below and, then, for the three specific regressors in the proof of Theorems 4.16 and 4.18. Finally we check the conditions for the cumulated sum of squares statistics results in the proofs of Theorems 4.17 and 4.19.

Add and subtract θ_0 to θ in the criterion function (A.20) and let $\vartheta = \theta - \theta_0$ and $\tilde{x}_t = x_t + \theta_0$ to get

$$Q_n(\theta) = \sum_{t=1}^n \{\varepsilon_t + \tilde{x}_t^2 - (\tilde{x}_t + \vartheta)^2\}^2 = \sum_{t=1}^n (\varepsilon_t - 2\vartheta\tilde{x}_t - \vartheta^2)^2.$$

Normalize $Q_n(\theta)$ as

$$D_n(\vartheta) = Q_n(\theta) - Q_n(\theta_0) = \sum_{t=1}^n \{\vartheta^4 + 4\tilde{x}_t\vartheta^3 + (4\tilde{x}_t^2 - 2\varepsilon_t)\vartheta^2 - 4\tilde{x}_t\varepsilon_t\vartheta\}. \tag{A.21}$$

These are random polynomials so we can exploit the following continuity result.

Lemma A.13 (Hammersley, 1956, Theorem 4.2) *The q-value function consisting of the zeros of the polynomial $\sum_{j=0}^q c_j z^j = 0$ is continuous at any point $(c_0, \ldots, c_q)$ such that $c_q \neq 0$ and $(c_0, \ldots, c_{q-1})$ are finite.*

In the analysis of the criterion function D_n in (A.21) we will balance the coeffcents to $\vartheta, \vartheta^2, \vartheta^3$ and ϑ^4 in various ways. When doing that it is convenient to work with the correlation-type coefficient

$$c_{x,n} = \frac{\sum_{t=1}^n \tilde{x}_t}{(n\sum_{t=1}^n \tilde{x}_t^2)^{1/2}}. \tag{A.22}$$

We analyze the criterion function under either of the following assumptions. The first covers the random walk case while the second covers the stationary and power function cases.

Assumption A.14 (weak) *Suppose x_t is $\mathcal{F}_{t-1}$-adapted and satisfies*
(a) $\{n^{-1}\sum_{t=1}^n (x_t+\theta_0)^2\}^{-1} = \mathrm{O}_{\mathsf{P}}(1)$;
(b) $c_{x,n} \overset{\mathsf{D}}{\to} c_x$ *in distribution for some some random variable c_x satisfying* $\mathsf{P}(c_x^2 < 1) = 1$;
(c) $\sum_{t=1}^n (x_t+\theta_0)^2 = \mathrm{O}_{\mathsf{P}}(n^\ell)$ *for some* $\ell > 0$.

Assumption A.15 (strong) *Suppose x_t is $\mathcal{F}_{t-1}$-adapted and satisfies*
(a) $\{n^{-1}\sum_{t=1}^n (x_t+\theta_0)^2\}^{-1} = \mathrm{O}(1)$ *a.s.;*
(b) $c_{x,n} \to c_x$ *a.s. for some deterministic c_x satisfying* $c_x^2 < 1$;
(c) $\sum_{t=1}^n (x_t+\theta_0)^2 = \mathrm{O}(n^\ell)$ *a.s. for some* $\ell > 0$.

Notation A.16 *If either of Assumption A.14 or A.15 is satisfied we write* $(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-1} = \mathrm{o}_v(1)$ *and* $c_{x,n}^2 \to_v c_x^2$ *while* $c_x^2 <_v 1$.

We can then establish the consistency properties of the model.

Theorem A.17 *Consider model (4.4) with $\theta \in \Theta = \mathbb{R}$. Suppose ε_t satisfies Assumption 3.1 while x_t satisfies Assumption A.14 or A.15. Recall the Notation A.16. Then the global minimizer satisfies*

$$\{\textstyle\sum_{t=1}^n (x_t+\theta_0)^2\}^{1/2}(\hat{\theta}_n - \theta_0) = \frac{\sum_{t=1}^n (x_t+\theta_0)\varepsilon_t}{2\{\sum_{t=1}^n (x_t+\theta_0)^2\}^{1/2}} + \mathrm{o}_v(1) = \mathrm{o}_v\{(\log n)^2\}.$$

If $c_x^2 > 8/9$ then the criterion function has a local minimizer satisfying

$$\{n^{-1}\textstyle\sum_{t=1}^n (x_t+\theta_0)^2\}^{-1/2}(\hat{\theta}_{local} - \theta_0) = -\frac{3}{2}c_x\{1 + (1-\frac{8}{9c_x^2})^{1/2}\} + \mathrm{o}_v(1).$$

Proof of Thorem A.17. We derive the properties of the global and the local minimizers by analyzing the criterion function D_n defined in (A.21) in detail. We start by finding the roots of the first derivative of D_n, which correspond to the extrema points in D_n.

1. Finding roots of the first derivative of D_n. Divide the polynomial D_n in (A.21) by $4n(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^2$ and normalize the parameter so that $\rho = \vartheta/(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{1/2}$ and

$$\frac{D_n(\vartheta)}{4n(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^2} = \underline{D}_n(\rho) = \frac{1}{4}\rho^4 + \underline{c}_{x,n}\rho^3 + \underline{b}_{x,n}\rho^2 - \underline{a}_{x,n}\rho,$$

with coefficients

$$\underline{c}_{x,n} = c_{x,n} = \frac{\sum_{t=1}^n \tilde{x}_t}{(n\sum_{t=1}^n \tilde{x}_t^2)^{1/2}}, \qquad \underline{b}_{x,n} = 1 - \frac{\sum_{t=1}^n \varepsilon_t}{2\sum_{t=1}^n \tilde{x}_t^2}, \qquad \underline{a}_{x,n} = n^{1/2}\frac{\sum_{t=1}^n \tilde{x}_t\varepsilon_t}{(\sum_{t=1}^n \tilde{x}_t^2)^{3/2}}.$$

To find the extrema of D_n, or equivalently of $\underline{D}_n$, we find the derivative

$$\underline{\dot{D}}_n(\rho) = \frac{\partial}{\partial\rho}\underline{D}_n(\rho) = \rho^3 + 3\underline{c}_{x,n}\rho^2 + 2\underline{b}_{x,n}\rho - \underline{a}_{x,n}. \tag{A.23}$$

We start by checking that the coefficients satisfy the conditions of the Hammersley Lemma A.13. If that is the case, then by continuity, we can learn about the roots $\underline{\dot{D}}_n$ by studying the roots of the limit of $\dot{\underline{D}}_n$. In particular, we will be interested in the number of real roots as this informs about the number of extrema of $\underline{D}_n$.

We argue that the coefficients of $\dot{\underline{D}}_n(\rho)$ in (A.23) satisfy

$$c_{x,n}^2 \leq 1, \qquad c_{x,n} \to_v c_x, \qquad \underline{b}_{x,n} \to_v 1, \qquad \underline{a}_{x,n} \to_v 0.$$

The Cauchy-Schwarz inequality shows $c_{x,n}^2 \leq 1$, while $c_{x,n} \to_v c_x$ by part (b) of Assumption A.14 or A.15. Rewrite

$$1 - \underline{b}_{x,n} = \{\frac{(n \log\log n)^{1/2}}{n}\}\frac{(n\log\log n)^{-1/2}\sum_{t=1}^n \varepsilon_t}{2n^{-1}\sum_{t=1}^n \tilde{x}_t^2}.$$

Apply the Law of Iterated Logarithm in Lemma A.11 to $\sum_{t=1}^n \varepsilon_t$ and part (a) of either Assumption A.14 or A.15 to $\sum_{t=1}^n \tilde{x}_t^2$. In this way, we can establish a rate for $1 - \underline{b}_{x,n}$, which will be used later in the proof, and show that $\underline{b}_{x,n} \to_v 1$. That is, for all $\omega > 0$,

$$1 - \underline{b}_{x,n} = \{\frac{(n\log\log n)^{1/2}}{n}\}\mathrm{O}_{a.s.}(1)\mathrm{o}_v(1) = \mathrm{o}_v(n^{\omega-1/2}) = \mathrm{o}_v(1), \tag{A.24}$$

by the dominance of powers over logarithms. Finally, using first Lemma A.7 and then that power functions dominate logarithms we get

$$\underline{a}_{x,n} \overset{a.s.}{=} n^{1/2}\frac{\mathrm{o}[\{\log(\sum_{t=1}^n \tilde{x}_t^2)\}^2]}{\sum_{t=1}^n \tilde{x}_t^2}.$$

Now, apply dominance of powers over logarithms and part (a) of Assumption A.14 or A.15 to show that, for all $0 < \varsigma < 1/2$,

$$\underline{a}_{x,n} \overset{a.s.}{=} \mathrm{o}(n^{1/2})(\textstyle\sum_{t=1}^n \tilde{x}_t^2)^{\varsigma-1} = \mathrm{o}(n^{1/2})n^{\varsigma-1}(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{\varsigma-1} = \mathrm{o}_v(1).$$

We can now apply the Hammersley Lemma A.13 since the highest order coefficient, to ρ^3 in polynomial (A.23), is unity and hence non-zero, while the other coefficients are finite. Thus, the roots of $\dot{\underline{D}}_n(\rho)$ are continuous functions of the coefficients. Applying continuity to the $a.s.$ sequence or through the Continuous Mapping Theorem shows that the roots converge to the roots of the limiting polynomial

$$\dot{\underline{D}}_n(\rho) \to_v \dot{\underline{D}}(\rho) = \rho^3 + 3c_x\rho^2 + 2\rho = \rho(\rho^2 + 3c_x\rho + 2),$$

which has roots

$$\rho_1 = 0, \qquad \rho_\pm = -\frac{3}{2}\{c_x \pm (c_x^2 - \frac{8}{9})^{1/2}\}.$$

We have $\underline{D}_n(\rho) \to \underline{D}(\rho) = \rho^4/4 + c_x\rho^3 + \rho^2$. The location of the roots of $\dot{\underline{D}}(\rho)$ indicates the location of the local minimizers of $\underline{D}(\rho)$. Since $\underline{D}(\rho)$ has a positive coefficient to ρ^4 then it must have either one local minimizer or two local minimizers with a local maximum in between.

If $c_x^2 < 8/9$ then $\rho_\pm$ are complex so that $\dot{\underline{D}}(\rho)$ has one real root and $\underline{D}(\rho)$ has one local minimizer at zero. If $c_x^2 = 8/9$ then $\rho_+ = \rho_-$ and $|\rho_+| = (3/2)(8/9)^{1/2} = \sqrt{2}$ so that $\dot{\underline{D}}(\rho)$ has one single real root and one double real root. In this case, $\underline{D}(\rho)$ has one local minimizer at zero and a point of inflection at $\rho_+ = \rho_-$. If $8/9 < c_x^2 \leq 1$ then $\rho_\pm$ are real, distinct and non-zero so that $\dot{\underline{D}}(\rho)$ has three distinct real roots and $\underline{D}(\rho)$ has two local minimizers. Thus, $\underline{D}(\rho)$ has one local minimizer at zero if $c_x^2 \leq 8/9$ and two local minimizers if $8/9 < c_x^2 \leq 1$.

2. Ordering the limiting roots when $8/9 < c_x^2 \leq 1$. Since $c_x^2 > 0$ we can write

$$\rho_\pm = -\frac{3}{2}c_x\{1 \pm (1 - \frac{8}{9c_x^2})^{1/2}\}.$$

We have $0 < c_x^2 - 8/9 \leq 1/9$ and $8/9 < c_x^2$ so that $0 < 1 - 8/(9c_x^2) \leq 1/(9c_x^2) < 1/8$. We then get that the roots have the same sign and satisfy

$$|\rho_-| < |\rho_+| < \frac{3}{2}|c_x|(1 + \frac{1}{\sqrt{8}}) \leq \frac{3}{2}(1 + \frac{1}{\sqrt{8}}) \approx 2.03,$$
$$|\rho_+| > |\rho_-| > \frac{3}{2}|c_x|(1 - \frac{1}{\sqrt{8}}) > \frac{3}{2}(\frac{\sqrt{8}}{3})(1 - \frac{1}{\sqrt{8}}) \approx 0.91.$$

In particular, we find $\rho_1 = 0 < |\rho_-| < |\rho_+|$ and therefore the roots are either ordered as $\rho_1 = 0 < \rho_- < \rho_+$ or $\rho_1 = 0 > \rho_- > \rho_+$. Therefore, in both cases, ρ_1 and ρ_+ are local minimizers for $\underline{D}$ while ρ_- is a local maximum.

Recall that $\rho = \vartheta/(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{1/2}$ so that $\vartheta = \rho(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{1/2}$. Since $n^{-1}\sum_{t=1}^n \tilde{x}_t^2$ is bounded away from zero by part (a) of AssumptionA.14 or A.15 we get that $|\vartheta_+| > |\vartheta_-| > 0$. When $n^{-1}\sum_{t=1}^n \tilde{x}_t^2$ diverges we get that $|\vartheta_+| > |\vartheta_-|$ diverge.

3. Evaluating $\underline{D}(\rho_+)$ when $8/9 < c_x^2 \leq 1$. We know that $\dot{\underline{D}}(\rho_+) = 0$. Thus, we can eliminate the fourth power term in $\underline{D}$ as follows

$$\underline{D}(\rho_+) = \underline{D}(\rho_+) - \frac{1}{4}\rho_+\dot{\underline{D}}(\rho_+) = \frac{1}{4}c_x\rho_+^3 + \frac{1}{2}\rho_+^2 = \frac{1}{4}\rho_+^2(c_x\rho_+ + 2).$$

Using first that $0 < 1 - 8/(9c_x^2) \leq 1/(9c_x^2)$, as argued in item 2, and then that $c_x^2 <_v 1$ by part (b) of Assumption A.14 or A.15 we get

$$c_x\rho_+ + 2 = 2 - \frac{3}{2}c_x^2\{1 + (1 - \frac{8}{9c_x^2})^{1/2}\} \geq 2 - \frac{3}{2}c_x^2(1 + \frac{1}{3|c_x|}) = -\frac{3}{2}(|c_x| - 1)(|c_x| + \frac{4}{3}) >_v 0.$$

The second order polynomial, with a negative coefficient on c_x^2, is positive between its roots $c_x = -4/3$ and $c_x = 1$. Hence, $c_x\rho_+ + 2 > 0$ whenever $|c_x| < 1$ or equivalently $c_x^2 < 1$. Thus, $c_x\rho_+ + 2 >_v 0$. Furthermore, since $|\rho_+| > 0.9$ we then get $\rho_+^2/4 > 0.2$ and therefore $\underline{D}(\rho_+) >_v 0$. As $\underline{D}(0) = 0$ then ρ_+ cannot be a global minimizer for the limiting polynomial $\underline{D}(\rho)$.

The analysis so far shows that the limiting polynomial $\underline{D}$ has a unique global minimum at zero. Since $\underline{D}_n(\rho)$ approaches $\underline{D}(\rho)$ continuously by the Hammersley Lemma A.13 then, to show that $\rho_1 = 0$ is the global minimizer of $D_n(\vartheta)$ for large n, it suffices to show that the local minimum of $\underline{D}_n(\rho)$ in the vicinity of ρ_+ is larger than the local minimum of $\underline{D}_n(\rho)$ in the vicinity of $\rho_1 = 0$. We start by showing, in item 4, that the local minimum of $\underline{D}_n(\rho)$ in the vicinity of ρ_+ is positive. In items 5 and 6, we will argue that the local minimum of $\underline{D}_n(\rho)$ in the vicinity of $\rho_1 = 0$ is lower.

4. Showing that the minimum of $D_n(\rho)$ in the vicinity of ρ_+ is positive for large n. The argument is slightly different for the weak and strong convergence cases under Assumption A.14 and A.15, respectively. Hence, we outline these arguments separately.

The strong consistency case under Assumption A.15: We have that $c_{x,n} \to c_x$ *a.s.* where the constant c_x satisfies $c_x^2 < 1$. Then item 3 shows not only that $\underline{D}(\rho_+) > 0$, but also that a small $\epsilon > 0$ exists so that $\underline{D}(\rho_+) > \epsilon$. Since $\underline{D}_n(\rho)$ approaches $\underline{D}(\rho)$ continuously by the Hammersley Lemma A.13 then for almost every outcome the local minimum of $\underline{D}_n(\rho)$ in the vicinity of ρ_+ will be larger than $\epsilon/2$ for large n. Thus, if the local minimum in the vicinity of 0 is less than $\epsilon/2$, this must be the global minimum. We will show that this is the case in item 6.

The weak convergence case under Assumption A.14: We have that $c_{x,n} \to c_x$ in distribution where the random variable c_x satisfies $\mathsf{P}(|c_x| < 1) = 1$. This will only result in a second local minimum when $c_x^2 > 8/9$. Item 3 shows that on the set $c_x^2 > 8/9$ we have $\mathsf{P}\{\underline{D}(\rho_+) > 0\} = 1$. This implies that for all $\epsilon > 0$ there exists a continuity point $\delta > 0$ of $\underline{D}(\rho_+)$ so that $\mathsf{P}[\{c_x^2 > 8/9\} \cap \{\underline{D}(\rho_+) \leq \delta\}] \leq \epsilon/4$. Since $\underline{D}_n(\rho)$ approaches $\underline{D}(\rho)$ continuously by the Hammersley Lemma A.13 then the probability that $\underline{D}_n(\rho)$ has two local minimizers converges to $\mathsf{P}(c_x^2 > 8/9)$. Moreover, by the same continuity argument, the probability for the joint event that $\underline{D}_n(\rho)$ has two local minimizers and the local minimum in the vicinity of ρ_+ is smaller than $\delta/2$ will be at most $\epsilon/2$ for large n. Thus, if it is shown that the local minimum in the vicinity of 0 exceeds $\delta/4$ with probability of at most $\epsilon/2$ for large n, then the local minimum at 0 is not the global minimum with probability of at most ϵ for large n. Equivalently, if the local minumum in the vicinity of 0 is less than $\delta/4$ with probability of at least $1-\epsilon/2$ for large n, then the local minimum at 0 is the global minimum with probability of at least $1-\epsilon$ for large n. We will show that this is the case in item 6.

5. Analyzing the minimum of D_n in the vicinity of 0 for large n. We now study the root $\rho_1 = 0$. Recall that in item 1 we normalized the coefficients ϑ so that $\rho = \vartheta(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-1/2}$. We show that the root $\rho_1 = 0$ corresponds to the consistent root, $(\sum_{t=1}^n \tilde{x}_t^2)^{1/2}\vartheta$, on the ϑ scale. Moreover, we want to find the asymptotic properties of its corresponding minimizer. To do all this, we study $D_n(\vartheta)$ in a shrinking region around $\vartheta = 0$. Moreover, to find the asymptotic properties of the consistent root normalize ϑ by $(\sum_{t=1}^n \tilde{x}_t^2)^{1/2}$ and define the parameter $\lambda = \vartheta(\sum_{t=1}^n \tilde{x}_t^2)^{1/2}$ so that $\vartheta = \lambda(\sum_{t=1}^n \tilde{x}_t^2)^{-1/2}$. Note that $\lambda = 0$ corresponds to $\vartheta = 0$. This gives the following relationship $\lambda = \rho(n^{-1/2}\sum_{t=1}^n \tilde{x}_t^2)$.

The shrinking region is defined by $|\lambda| \leq n^{1/8}$. On the ϑ scale this region is indeed shrinking because

$$|\vartheta| = |\lambda|(\textstyle\sum_{t=1}^n \tilde{x}_t^2)^{-1/2} = n^{-1/2}|\lambda|(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-1/2} \leq n^{-1/2+1/8}(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-1/2} = \mathrm{O}_v(n^{-3/8}),$$

since $(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-1} = \mathrm{O}_v(1)$ by part (a) of Assumption A.14 or A.15, see Notation A.16. The region is also shrinking on the ρ scale because

$$|\rho| = |\lambda|(n^{-1/2}\textstyle\sum_{t=1}^n \tilde{x}_t^2)^{-1} \leq n^{1/8}(n^{-1/2}\sum_{t=1}^n \tilde{x}_t^2)^{-1} = n^{1/8-1/2}(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-1} = \mathrm{O}_v(n^{-3/8}).$$

We check that the asymptotic local minimimizer ρ_+ is outside this shrinking region. From item 2 we know that either $\rho_+ \in [-2.04, -0.90]$ or $\rho_+ \in [0.90, 2.04]$. Therefore ρ_+ is outside the shrinking region where $|\rho| = \mathrm{O}_v(n^{-3/8})$ for large n.

We study the properties of D_n uniformly in $|\lambda| \leq n^{1/8}$. The polynomial D_n in (A.21) can be rewritten as

$$\frac{1}{4}D_n(\vartheta) = \underline{\underline{D}}_n(\lambda) = \underline{\underline{d}}_{x,n}\lambda^4 + \underline{\underline{c}}_{x,n}\lambda^3 + \underline{\underline{b}}_{x,n}\lambda^2 - \underline{\underline{a}}_{x,n}\lambda$$

with coefficients

$$\underline{\underline{d}}_{x,n} = \frac{n}{4(\sum_{t=1}^n \tilde{x}_t^2)^2}, \quad \underline{\underline{c}}_{x,n} = \frac{\sum_{t=1}^n \tilde{x}_t}{(\sum_{t=1}^n \tilde{x}_t^2)^{3/2}}, \quad \underline{\underline{b}}_{x,n} = 1 - \frac{\sum_{t=1}^n \varepsilon_t}{2\sum_{t=1}^n \tilde{x}_t^2}, \quad \underline{\underline{a}}_{x,n} = \frac{\sum_{t=1}^n \tilde{x}_t\varepsilon_t}{(\sum_{t=1}^n \tilde{x}_t^2)^{1/2}}.$$

We will argue that

$$\sup_{|\lambda|\leq n^{1/8}} \underline{\underline{d}}_{x,n}\lambda^4 = \mathrm{O}_v(n^{-1/2}), \qquad \sup_{|\lambda|\leq n^{1/8}} \underline{\underline{c}}_{x,n}\lambda^3 = \mathrm{O}_v(n^{-1/8}), \qquad \sup_{|\lambda|\leq n^{1/8}} (\underline{\underline{b}}_{x,n} - 1)\lambda^2 = \mathrm{O}_v(n^{-1/8}).$$

The bound for $\underline{\underline{d}}_{x,n}\lambda^4$ follows from part (a) of Assumption A.14 or A.15 showing that

$$\underline{\underline{d}}_{x,n}\lambda^4 = (1/4)nn^{-2}(n^{-1}\textstyle\sum_{t=1}^n \tilde{x}_t^2)^2\lambda^4 = n^{-1}\mathrm{O}_v(1)n^{4/8} = \mathrm{O}_v(n^{-1/2}).$$

The bound for $\underline{\underline{c}}_{x,n}\lambda^3$ follows similarly using that $c_x^2 \leq 1$ by the Cauchy-Schwarz inequality, that is

$$\underline{\underline{c}}_{x,n}\lambda^3 = \frac{\sum_{t=1}^n \tilde{x}_t}{(n\sum_{t=1}^n \tilde{x}_t^2)^{1/2}}(n^{-1}\textstyle\sum_{t=1}^n \tilde{x}_t^2)^{-1}n^{-1/2}\lambda^3 = \mathrm{O}(1)\mathrm{O}_v(1)n^{-1/2}n^{3/8} = \mathrm{O}_v(n^{-1/8}).$$

The bound for $(\underline{\underline{b}}_{x,n} - 1)\lambda^2$ follows since $\underline{\underline{b}}_{x,n} - 1 = \underline{b}_{x,n} - 1 = \mathrm{o}_v(n^{\omega-1/2})$ for any $\omega > 0$ as derived in (A.24) while $\lambda^2 \leq n^{2/8}$ so that $(\underline{\underline{b}}_{x,n} - 1)\lambda^2 = \mathrm{o}_v(n^{\omega-1/2})n^{1/4} = \mathrm{O}_v(n^{-1/8})$. In summary, we get, uniformly in $|\lambda| \leq n^{1/8}$, that

$$\underline{\underline{D}}_n(\lambda) = \lambda^2 - \underline{\underline{a}}_{x,n}\lambda + \mathrm{O}_v(n^{-1/8}). \tag{A.25}$$

This quadratic criterion function has minimum at

$$\lambda_1 = \underline{\underline{a}}_{x,n}/2 + \mathrm{O}_v(n^{-1/16}) = \frac{\sum_{t=1}^n \tilde{x}_t\varepsilon_t}{2(\sum_{t=1}^n \tilde{x}_t^2)^{1/2}} + \mathrm{O}_v(n^{-1/16}).$$

We note that $\underline{\underline{a}}_{x,n} = \mathrm{o}_v(\log n)$, see Lemma A.7 with part (c) of Assumption A.14 or A.15. In particular $|\lambda_1| \leq n^{1/8}$ for large n. Since $\lambda = \vartheta(\sum_{t=1}^n \tilde{x}_t^2)^{1/2}$ we get that the corresponding minimizer on the ϑ scale satisfies

$$(\textstyle\sum_{t=1}^n \tilde{x}_t^2)^{1/2}\vartheta_1 = \frac{\sum_{t=1}^n \tilde{x}_t \varepsilon_t}{2(\sum_{t=1}^n \tilde{x}_t^2)^{1/2}} + \mathrm{o}_v(1),$$

which is the desired solution, noting that $\mathrm{O}_v(n^{-1/16}) = \mathrm{o}_v(1)$.

6. Evaluating $D_n(\vartheta_1)$. Inserting $\vartheta_1 = \lambda_1(\sum_{t=1}^n \tilde{x}_t^2)^{-1/2}$ with $\lambda_1 = \underline{\underline{a}}_{x,n}/2 + \mathrm{O}_v(n^{-1/16})$ and$|\lambda_1| \leq n^{1/8}$ in (A.25) we get

$$D_n(\vartheta_1) = 4\underline{\underline{D}}_n(\lambda_1) = -\underline{\underline{a}}_{x,n}^2 + \mathrm{O}_v(n^{-1/8}).$$

The first term, $-\underline{\underline{a}}_{x,n}^2$, is non-positive. The second term may be positive, but vanishes. Thus, by the argument in item 4 we get that ϑ_1 is the global minimizer in the limit. Indeed, in the strong consistency case we get that for any $\epsilon > 0$ then for each outcome $D_n(\vartheta_1) \leq \epsilon/2$ for large n. In the weak consistency case we get that for any $\epsilon, \delta > 0$ then $\mathsf{P}\{D_n(\vartheta_1) \leq \delta/4\} > 1 - \epsilon/2$ for large n. □

We need some preliminary results for stationary and power function regressors.

Lemma A.18 *Let x_t be stationary and autoregressive with $Ex_t = \mu_x$ and $\mathsf{V}x_t = \sigma_x^2$ and $Ex_t^4 < \infty$. Then $n^{-1}\sum_{t=1}^n \tilde{x}_t \to \tilde{\mu}_x = \mu_x + \theta_0$ a.s. and $n^{-1}\sum_{t=1}^n \tilde{x}_t^2 \to \sigma_x^2 + \tilde{\mu}_x^2$ a.s.*

Proof of Lemma A.18. Apply Phillips and Solo (1992, Theorems 3.13, 3.16). □

Lemma A.19 *Let $x_t = t^\tau$ for some $\tau > 0$. Then $n^{-(1+k\tau)}\sum_{t=1}^n \tilde{x}_t^k \to (1 + k\tau)^{-1}$ for $k > 0$.*

Proof of Lemma A.19. Approximate $n^{-(1+k\tau)}\sum_{t=1}^n (t^\tau + \theta_0)^k$ by $\int_0^1 u^{k\tau} du = (1 + k\tau)^{-1}$. □

We can now prove the desired results for stationary and power function regressors.

Proof of Theorem 4.16. Since *Assumption 3.1* is assumed we can apply Theorem A.17 if Assumption A.15 holds.

Let x_t be stationary and autoregressive. We check Assumption A.15 using Lemma A.18. (a), (c) follow since $n^{-1}\sum_{t=1}^n \tilde{x}_t^2 \to \sigma_x^2 + \tilde{\mu}_x^2$ *a.s.* For (b) note $c_{x,n}^2 \to c_x^2 = \tilde{\mu}_x^2/(\sigma_x^2 + \tilde{\mu}_x^2)$ *a.s.* where $c_x^2 \leq 8/9$ when $\tilde{\mu}_x^2 \leq 8\sigma_x^2$. Thus, the objective function has one minimizer for processes with small to medium signal-to-noise ratio and two minimizers otherwise.

Let $x_t = t^\tau$ for $\tau > 0$. We check Assumption A.15 using Lemma A.19. (a), (c) follow since $n^{-(1+2\tau)}\sum_{t=1}^n \tilde{x}_t^2 \to (1 + 2\tau)^{-1}$. For (b) note $c_{x,n}^2 \to c_x^2 = (1 + 2\tau)/(1 + \tau)^2$ where $c_{x,n}^2 \leq 8/9$ when $\tau \geq 1/2$. □

Proof of Theorem 4.17. We apply Theorems 3.6, 3.8 with $N_{n,\theta_0}^{-1} = \{\sum_{t=1}^n (x_t + \theta_0)^2\}^{1/2}$ and check Assumptions 3.1-3.5 and 3.7. *Assumption 3.1* is assumed. Note that $g(x_t, \theta) = (x_t + \theta)^2$ so that $\dot{g}(x_t, \theta) = 2(x_t + \theta)$ and $\ddot{g}(x_t, \theta) = 2$. Let $\vartheta = \theta - \theta_0$ and $\tilde{x}_t = x_t + \theta_0$.

Assumption 3.2: In the stationary, autoregressive case $n^{-1}N_{n,\theta_0}^{-1} = n^{-1}\sum_{t=1}^n \tilde{x}_t^2$ converges *a.s.* by Lemma A.18. In the power case $n^{-1-2\tau}N_{n,\theta_0}^{-1}$ converges by Lemma A.19. In both cases the normalization is of polynomial order. These results also show $(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-1} = \mathrm{O}(1)$ *a.s.*

Assumption 3.3: This is satisfied *a.s.* for any $\delta > 0$. Indeed, Theorem 4.16 shows $N_{n,\theta_0}^{-1}(\theta - \theta_0) = \mathrm{o}\{(\log n)^2\} = \mathrm{o}(n^\delta)$ *a.s.*

Assumption 3.4. We need part (b) weakly and the parts (a, c, d) strongly. The set $\mathcal{S}_n = \{\theta : |N_{n,\theta_0}^{-1}(\theta - \theta_0)| \leq n^\delta \epsilon\}$ is equivalent to the set where $\vartheta^2 \sum_{t=1}^n \tilde{x}_t^2 \leq n^{2\delta}\epsilon^2$. We have that $(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-1} = \mathrm{O}(1)$ *a.s.* using Lemmas A.18, A.19 in the autoregressive and in the trend case, respectively. Therefore for large n we have that $\vartheta = \mathrm{O}(n^{2\delta-1})$ *a.s.* on the set $\mathcal{S}_n$.

Assumption 3.4(a) holds *a.s.* Use that $g(x_t, \theta) - g(x_t, \theta_0) = \vartheta^2 + 2\vartheta\tilde{x}_t$. By the inequality $(y + z)^2 \leq 2(y^2 + z^2)$ we find

$$\mathcal{G}_{n2} = \textstyle\sum_{t=1}^n \{g(x_t, \theta) - g(x_t, \theta_0)\}^2 \leq 2n\vartheta^4 + 8\vartheta^2\sum_{t=1}^n \tilde{x}_t^2$$

On the set $\mathcal{S}_n$ we have $\vartheta = \mathrm{O}(n^{2\delta-1})$ and $\vartheta^2 \sum_{t=1}^n \tilde{x}_t^2 \le n^{2\delta}\epsilon^2$ so that $\mathcal{G}_{n2} = \mathrm{O}(n^{8\delta-3}) + \mathrm{O}(n^{2\delta}\epsilon^2) = \mathrm{o}(n^{1/2})$ *a.s.* when $\delta > 0$ is chosen sufficiently small.

Assumption 3.4(b) holds *a.s.* Proceed as before using $(y+z)^4 \le C(y^4+z^4)$ and $\sum_{t=1}^n \tilde{x}_t^4 \le (\sum_{t=1}^n \tilde{x}_t^2)^2$ to get, on $\mathcal{S}_n$,

$$\mathcal{G}_{n4} = \textstyle\sum_{t=1}^n \{g(x_t,\theta) - g(x_t,\theta_0)\}^4 \le C\{n\vartheta^8 + (\vartheta^2 \sum_{t=1}^n \tilde{x}_t^2)^2\} \overset{a.s.}{=} \mathrm{O}(n^{16\delta-7}) + \mathrm{O}(n^{4\delta}\epsilon^4) = \mathrm{o}(n)$$

when $\delta > 0$ is chosen sufficiently small.

Assumption 3.4(c) also holds *a.s.* since $\sum_{t=1}^n ||N_{n,\theta_0}' \dot{g}(x_t,\theta_0)||^2 = 1$.

Assumption 3.4(d) also holds *a.s.* We have $\ddot{g} = 2$ so that, noting $(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-1} = \mathrm{O}(1)$ *a.s.*,

$$\sup_{\theta:||N_{n,\theta_0}^{-1}(\theta-\theta_0)|| \le n^{\delta}\epsilon} \textstyle\sum_{t=1}^n ||N_{n,\theta_0}^2 2||^2 = 2^2 n N_{n,\theta_0}^4 = 2^2 n^{-1}(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-2} \overset{a.s.}{=} \mathrm{O}(n^{-1}).$$

Assumption 3.5: $\sum_{t=1}^n \{\dot{g}(x_t,\theta_0)\}^2 = 4\sum_{t=1}^n \tilde{x}_t^2$ which diverges since $(n^{-1}\sum_{t=1}^n \tilde{x}_t^2)^{-1} = \mathrm{O}(1)$ *a.s.* Finally, *Assumption 3.7* holds because Assumptions 3.3, 3.4(a,c,d) were demonstrated *a.s.* □

Lemma A.20 *Let x_t satisfy the random walk type Assumption 4.3. Then*
(a) $n^{-1/2}x_{[nu]} \overset{\mathsf{D}}{\to} \mathcal{W}_{x,u}$ *on* $D[0,1]$ *where* $\mathcal{W}_{x,u}$ *is a Brownian motion;*
(b) $(n^{-3/2}\sum_{t=1}^n x_t, n^{-2}\sum_{t=1}^n x_t^2) \overset{\mathsf{D}}{\to} (\int_0^1 \mathcal{W}_{x,u}du, \int_0^1 \mathcal{W}_{x,u}^2 du)$;
(c) $\int_0^1 \mathcal{W}_{x,u}^2 du - (\int_0^1 \mathcal{W}_{x,u}du)^2 \overset{\mathsf{D}}{=} \int_0^1 \mathcal{B}_u^2 du$ *where* $\mathcal{B}_u$ *is a Brownian bridge;*
(d) $(\int_0^1 \mathcal{W}_{x,u}du)^2$, $\int_0^1 \mathcal{W}_{x,u}^2 du$, $\int_0^1 \mathcal{B}_u^2 du$, $(\int_0^1 \mathcal{W}_{x,u}d\mathcal{W}_{x,u})^2/\int_0^1 \mathcal{W}_{x,u}^2 du$ *are positive with probability one;*
(e) $\sum_{t=1}^n \tilde{x}_t^2 = \mathrm{O}(n^2 \log\log n)$ *a.s.;*
(f) $\liminf_{n\to\infty} n^{-2}\log\log n \sum_{t=1}^n \tilde{x}_t^2 > 0$ *a.s.*

Proof of Lemma A.20. Recall that $x_t = \sum_{s=1}^t \eta_s + \alpha'\mathbf{x}_{t-1}$ and $\tilde{x}_t = x_t + \theta_0$.

(a) Apply Chan and Wei (1988, Theorem 2.2), noting that $\mathbf{x}_t = \mathrm{o}(n^{1/2})$ *a.s.*, see Lemma A.11.

(b) Use the Continuous Mapping Theorem from $D[0,1]$ to $\mathbb{R}^2$ for integrals.

(c) Apply, for instance, Donati-Martin and Yor (1991, Proposition 1).

(d) The variable $\int_0^1 \mathcal{W}_{x,u}du$ is normally distributed, see Dickey and Fuller (1979). Hence its square is positive *a.s.* Chan and Wei (1988, Lemma 3.1.1) show $\int_0^1 \mathcal{W}_{x,u}^2 du > 0$ *a.s.* Their proof extends to $\int_0^1 \mathcal{B}_u^2 du$. The Ito formula shows that $\int_0^1 \mathcal{W}_{x,u}d\mathcal{W}_{x,u} = (\mathcal{W}_{x,1}^2 - 1)/2$, see also Dickey and Fuller (1979). This has a continuous distribution so that $(\int_0^1 \mathcal{W}_{x,u}d\mathcal{W}_{x,u})^2/\int_0^1 \mathcal{W}_{x,u}^2 du > 0$ *a.s.*

(e) Note $\sum_{t=1}^n \tilde{x}_t^2 \le n^2 \max_{1\le t\le n}(n^{-1/2}\tilde{x}_t)^2$ and apply the Law of Iterated Logarithm in Lemma A.11.

(f) Nielsen (2005, equation 3.1) gives the decomposition $\mathbf{x}_t = \tilde{\mathbf{x}}_t + \tilde{\mu}$, where $\tilde{\mathbf{x}}_t$ follows the vector autoregressive equation (4.2) with $\mu = 0$ and some vector $\tilde{\mu}$. Thus, we can write $\tilde{x}_t = v_t + u_t$ where $v_t = \sum_{s=1}^t \eta_s + \alpha'\tilde{\mu} + \theta_0$ and $u_t = \alpha'\tilde{\mathbf{x}}_t$. Thus, we can expand

$$\textstyle\sum_{t=1}^n \tilde{x}_t^2 = \sum_{t=1}^n v_t^2 + 2\sum_{t=1}^n u_t v_t + \sum_{t=1}^n u_t^2. \tag{A.26}$$

We get $\liminf_{n\to\infty} n^{-2}\log\log n \sum_{t=1}^n v_t^2 > 0$ *a.s.* from the Donsker-Varadhan Law of Iterated Logarithm, see Lai and Wei (1982, equation 3.23). We bound $|\sum_{t=1}^n u_t v_t| \le n \max_{1\le t\le n}|u_t| \max_{1\le t\le n}|v_t|$ and show this is $\mathrm{o}(n^2/\log\log n)$. Indeed $\max_{1\le t\le n}|v_t| = \mathrm{o}\{(n\log\log n)^{1/2}\}$ *a.s.* by Lemma A.11, while $\max_{1\le t\le n}|u_t| = \mathrm{o}(n^{1/2-\omega})$ *a.s.* for some $\omega > 0$ see A.11. In combination $|\sum_{t=1}^n u_t v_t| = \mathrm{o}\{n(n\log\log n)^{1/2}n^{1/2-\omega}\} = \mathrm{o}(n^2/\log\log n)$ *a.s.* Finally, $\sum_{t=1}^n u_t^2 = \mathrm{O}(n) = \mathrm{o}(n^2/\log\log n)$ *a.s.* by Lai and Wei (2005, Corollary 1,iii). Thus, the first term in the expansion (A.26) dominates and gives the desired rate. □

Proof of Theorem 4.18. Since Assumption 3.1 is satisfied we apply Theorem A.17 if Assumption A.14 holds, which we will check. Apply Lemma A.20(b,c,d). Now, (a), (c) in Assumption

A.14 follows since $n^{-2}\sum_{t=1}^{n}x_t^2 \overset{\mathsf{D}}{\to} \int_0^1 \mathcal{W}_{x,u}^2 du$ which is positive $a.s.$ For (b) note $c_{x,n}^2 \overset{\mathsf{D}}{\to} c_x^2 = (\int_0^1 \mathcal{W}_{x,u}^2 du)^{-1}(\int_0^1 \mathcal{W}_{x,u} du)^2$. This has a distribution on $(0,1)$. □

Proof of Theorem 4.19. We apply Theorem 3.6 with $N_{n,\theta_0}^{-1} = \{\sum_{t=1}^{n}(x_t+\theta_0)^2\}^{1/2}$ and show Assumptions 3.1-3.5. *Assumption 3.1* is assumed. Note that $g(x_t,\theta) = (x_t+\theta)^2$ so that $\dot{g}(x_t,\theta) = 2(x_t+\theta)$ and $\ddot{g}(x_t,\theta) = 2$. Let $\vartheta = \theta - \theta_0$ and $\tilde{x}_t = x_t + \theta_0$.

Assumption 3.2: Lemma A.20(e,f) shows $N_{n,\theta_0}^{-2} = \mathrm{O}(n^2 \log\log n) = \mathrm{O}(n^3)$ $a.s.$ and also that $\liminf_{n\to\infty} n^{-2}(\log\log n)N_{n,\theta_0}^2 > 0$ $a.s.$

Assumption 3.3 is satisfied weakly for any $\delta > 0$ by Theorem 4.18 using Assumptions 3.1, 4.3.

Assumption 3.4, 3.5: same as in the proof of Theorem 4.17 due to the stochastic normalization while $(n^{-1}\sum_{t=1}^{n}\tilde{x}_t^2)^{-1} = \mathrm{O}(1)$ $a.s$ also in this case by Lemma A.20(e,f). □

A.8 Proof of the local power result

We start by deriving limiting results for the least squares product sums appearing in the analysis of model (5.1).

Lemma A.21 *Suppose Assumption 5.1 holds. Then, we have jointly on $D^{27}[0,1]$, endowed with the Skorokhod metric with common distortion across the coordinates so that addition is continuous,*
(a) $n^{-1/2}\{\sum_{t=1}^{[nu]}\varepsilon_t, x_{[nu]}, z_{[nu]}\}' \overset{\mathsf{D}}{\to} (\mathcal{W}_{\varepsilon,u}, \mathcal{W}_{x,u}, \mathcal{W}_{z,u})'$;
(b) $n^{-1/2}\sum_{t=1}^{[nu]}(\varepsilon_t^2 - \sigma^2) \overset{\mathsf{D}}{\to} \mathcal{B}_u$;
(c) $\{g(x_{[nu]}/n^{1/2}), h(z_{[nu]}/n^{1/2})\}' \overset{\mathsf{D}}{\to} (\mathcal{G}_u, \mathcal{H}_u)' = \{g(\mathcal{W}_{x,u}), h(\mathcal{W}_{z,u})\}'$;
(d) $S_{gg,u} = n^{-1}\sum_{t=1}^{[nu]}g^2(x_t/n^{1/2}) \overset{\mathsf{D}}{\to} \mathcal{I}_{gg,u} = \int_0^u \mathcal{G}_s^2 ds$;
(e) $S_{gh,u} = n^{-1}\sum_{t=1}^{[nu]}g(x_t/n^{1/2})h(z_t/n^{1/2}) \overset{\mathsf{D}}{\to} \mathcal{I}_{gh,u} = \int_0^u \mathcal{G}_s\mathcal{H}_s ds$;
(f) $S_{g\varepsilon,u} = n^{-1/2}\sum_{t=1}^{[nu]}g(x_t/n^{1/2})\varepsilon_t \overset{\mathsf{D}}{\to} \mathcal{I}_{g\varepsilon,u} = \int_0^u \mathcal{G}_s d\mathcal{W}_{\varepsilon,s}$;
Further if $m_{t,n} = m(z_t/n^{1/2}, x_t/n^{1/2}) = h(z_t/n^{1/2}) - S_{gh,1}S_{gg,1}^{-1}g(x_t/n^{1/2})$ *then*
(g) $m_{[nu],n} \overset{\mathsf{D}}{\to} \mathcal{M}_u = m(\mathcal{W}_{z,u}, \mathcal{W}_{x,u}) = \mathcal{H}_u - \mathcal{I}_{gh,1}\mathcal{I}_{gg,1}^{-1}\mathcal{G}_u$;
(h) $n^{-1}\sum_{t=1}^{[nu]}m_{t,n}^j \overset{\mathsf{D}}{\to} \int_0^u \mathcal{M}_s^j ds$ *for* $j = 1,2,3,4,6$;
(i) $n^{-1/2}\sum_{t=1}^{[nu]}m_{t,n}\varepsilon_t \overset{\mathsf{D}}{\to} \int_0^u \mathcal{M}_s d\mathcal{W}_{\varepsilon,s}$;
(j) $n^{-1}\sum_{t=1}^{[nu]}m_{t,n}^j\varepsilon_t^k = \mathrm{O}_{\mathsf{P}}(1)$ *for* $k = 1,2,3$, $j = 0,\ldots,4-k$;
(k) $n^{-1}\sum_{t=1}^{[nu]}m_{t,n}g(x_t/n^{1/2}) \overset{\mathsf{D}}{\to} \int_0^u \mathcal{M}_s\mathcal{G}_s ds$.

Proof of Lemma A.21. $(a),(b)$ Under Assumption 5.1 the listed partial sum process converges weakly to the Brownian motion $(\mathcal{W}_\varepsilon, \mathcal{W}_x, \mathcal{W}_z)'$ due to Chan and Wei (1988, Theorem 2.2) and the Cramér-Wold device (Billingsley, 1968).

(c) Use the Continuous Mapping Theorem (Billingsley, 1968) for the mapping from $D^2[0,1]$ by g,h to $D^2[0,1]$, where g,h are assumed continuous in Assumption 5.1(a).

$(d),(e),(h),(k)$ Use the Continuous Mapping Theorem for partial product sums mapping from $D^8[0,1]$ to $D^8[0,1]$.

$(f),(i)$ For linear g,h and $u = 1$ apply Chan and Wei (1988, Theorem 2.4). For the general case apply Jakubowski, Ménin and Pages (1989), see also Kurtz and Protter (1991).

(g) Same as (c) noting that addition is continuous on the chosen space so that m is a continuous mapping from $D^2[0,1]$ to $D[0,1]$.

(j) The triangle inequality gives

$$\mathcal{M}_{jk} = |n^{-1}\sum_{t=1}^{[nu]}m_{t,n}^j\varepsilon_t^k| \leq n^{-1}\sum_{t=1}^{[nu]}|m_{t,n}|^j|\varepsilon_t|^k.$$

Further, the Hölder inequality gives the further bound, for $k = 1,2,3$ and $j = 0,\ldots,4-k$,

$$\mathcal{M}_{jk} \leq \{n^{-1}\sum_{t=1}^{[nu]}|m_{t,n}|^{4j/(4-k)}\}^{(4-k)/4}\{n^{-1}\sum_{t=1}^{[nu]}|\varepsilon_t|^4\}^{k/4}.$$

Here $4j/(4-k) \leq 6$. The Jensen inequality gives the further bound

$$\mathcal{M}_{jk} \leq \{n^{-1}\textstyle\sum_{t=1}^{[nu]}|m_{t,n}|^6\}^{j/6}\{n^{-1}\sum_{t=1}^{[nu]}|\varepsilon_t|^4\}^{k/4}.$$

The first average is $\mathrm{O}_{\mathsf{P}}(1)$ by (h). The second average converges to $\varphi^2 - \sigma^4$ as in Lemma A.1. $\square$

Proof of Theorem 5.2: The model is $y_t = \theta g(x_t) + v_t$ where $v_t = \varepsilon_t + \lambda n^{-1/4}h(z_t/n^{1/2})$. Let $N_n^{-1} = n^{1/4}g(n^{1/2})$. Recall $g(x_t/n^{1/2}) = g(x_t)/g(n^{1/2})$ and $h(z_t/n^{1/2}) = h(z_t)/h(n^{1/2})$ by Assumption 5.1.

1. The estimator satisfies

$$\hat{\theta}_n - \theta = \frac{\sum_{i=1}^n g(x_t)v_t}{\sum_{i=1}^n g^2(x_t)} = n^{-1/2}\frac{n^{-1/2}\sum_{i=1}^n g(x_t)\varepsilon_t}{n^{-1}\sum_{i=1}^n g^2(x_t)} + \lambda n^{-1/4}\frac{n^{-1}\sum_{i=1}^n g(x_t)h(z_t/n^{1/2})}{n^{-1}\sum_{i=1}^n g^2(x_t)}.$$

Using first the homogeneity of g and then Lemma A.21(d,e,f) we get

$$N_n^{-1}(\hat{\theta}_n - \theta) = n^{-1/4}\frac{S_{g\varepsilon,1}}{S_{gg,1}} + \lambda\frac{S_{gh,1}}{S_{gg,1}} \overset{\mathsf{D}}{\to} \lambda\frac{\mathcal{I}_{gh,1}}{\mathcal{I}_{gg,1}}.$$

2. The residuals are $\hat{v}_{t,n} = v_t - (\hat{\theta}_n - \theta)g(x_t)$. Recalling $v_t = \varepsilon_t + \lambda n^{-1/4}h(z_t/n^{1/2})$ we get

$$\hat{v}_{t,n} - \varepsilon_t = \lambda n^{-1/4}h(z_t/n^{1/2}) - \{N_n^{-1}(\hat{\theta}_n - \theta)\}N_n g(x_t).$$

Using the homogeneity of g, item 1 and the definition of the function m in Lemma A.21 we get

$$n^{1/4}(\hat{v}_{t,n} - \varepsilon_t) = \lambda h(z_t/n^{1/2}) - \{\lambda\frac{S_{hg,1}}{S_{gg,1}} + n^{-1/4}\frac{S_{g\varepsilon,1}}{S_{gg,1}}\}g(x_t/n^{1/2}) = \lambda m_{t,n} + n^{-1/4}\frac{S_{g\varepsilon,1}}{S_{gg,1}}g_{t,n}, \quad \text{(A.27)}$$

where $m_{t,n} = m(z_t/n^{1/2}, x_t/n^{1/2}) = h(z_t/n^{1/2}) - S_{hg,1}S_{gg,1}^{-1}g(x_t/n^{1/2})$, see Lemma A.21$(g)$, and $g_{t,n} = g(x_t/n^{1/2})$. The second term does not depend on λ and it is $\mathrm{O}_{\mathsf{P}}(n^{-1/4})$ uniformly in t. Lemma A.21(g) then shows $n^{1/4}(\hat{v}_{[nu],n} - \varepsilon_{[nu]}) \overset{\mathsf{D}}{\to} \lambda\mathcal{M}_u$.

3. The test statistic is $CUSQ_n = \max_{1\leq t\leq n}|\mathcal{V}_{t,n}|/\hat{\varphi}_n$, where

$$\mathcal{V}_{t,n} = n^{-1/2}\{\textstyle\sum_{s=1}^t \hat{v}_{s,n}^2 - (t/n)\sum_{s=1}^n \hat{v}_{s,n}^2\}, \qquad \hat{\varphi}_n^2 = n^{-1}\sum_{t=1}^n \hat{v}_{t,n}^4 - (n^{-1}\sum_{t=1}^n \hat{v}_{t,n}^2)^2.$$

3.1. The numerator is location invariant. Thus replace $\hat{v}_{s,n}^2$ by $\hat{v}_{s,n}^2 - \sigma^2$ and note $\mathcal{V}_{t,n} = \tilde{\mathcal{V}}_{t,n} - (t/n)\tilde{\mathcal{V}}_{n,n}$ where $\tilde{\mathcal{V}}_{t,n} = n^{-1/2}\sum_{s=1}^t(\hat{v}_{s,n}^2 - \sigma^2)$. Write $\hat{v}_{s,n} = (\hat{v}_{s,n} - \varepsilon_s) + \varepsilon_s$ and expand

$$\tilde{\mathcal{V}}_{t,n} = n^{-1/2}\textstyle\sum_{s=1}^t\{(\varepsilon_s^2 - \sigma^2) + 2(\hat{v}_{s,n} - \varepsilon_s)\varepsilon_s + (\hat{v}_{s,n} - \varepsilon_s)^2\}.$$

Apply expansion (A.27) for $n^{1/4}(\hat{v}_{t,n} - \varepsilon_t)$ to get

$$\begin{aligned}\tilde{\mathcal{V}}_{t,n} &= n^{-1/2}\textstyle\sum_{s=1}^t\{(\varepsilon_s^2 - \sigma^2) + 2(n^{-1/4}\lambda)m_{s,n}\varepsilon_s + (n^{-1/4}\lambda)^2 m_{s,n}^2\} \\ &\quad + n^{-1/2}\textstyle\sum_{s=1}^t\{2n^{-1/2}g_{s,n}\varepsilon_s\frac{S_{g\varepsilon,1}}{S_{gg,1}} + 2n^{-3/4}\lambda m_{s,n}g_{s,n}\frac{S_{g\varepsilon,1}}{S_{gg,1}} + n^{-1}g_{s,n}^2\frac{S_{g\varepsilon,1}^2}{S_{gg,1}^2}\}. \quad \text{(A.28)}\end{aligned}$$

Apply Lemma A.21(d,e,i,k) to get

$$\tilde{\mathcal{V}}_{t,n} = n^{-1/2}\textstyle\sum_{s=1}^t(\varepsilon_s^2 - \sigma^2) + \lambda^2 n^{-1}\sum_{s=1}^t m_{s,n}^2 + \mathrm{O}_{\mathsf{P}}(n^{-1/2}) + \mathrm{O}_{\mathsf{P}}(n^{-1/4}\lambda).$$

Applying Lemma A.1, A.21(h) then shows, on $D[0,1]$, $\tilde{\mathcal{V}}_{t,n} \overset{\mathsf{D}}{\to} \mathcal{B}_u + \lambda^2\mathcal{L}_u$, where $\mathcal{B}_u$ and $\mathcal{L}_u = \int_0^u \mathcal{M}_s^2 ds$. In particular we can form the bridge processes $\mathcal{B}_u^\circ = \mathcal{B}_u - u\mathcal{B}_1$ and $\mathcal{L}_u^\circ = \mathcal{L}_u - u\mathcal{L}_1$ so that $\max_{1\leq t\leq n}|\mathcal{V}_{t,n}| \overset{\mathsf{D}}{\to} \sup_{0\leq u\leq 1}|\mathcal{B}_u^\circ + \lambda^2\mathcal{L}_u^\circ|$.

3.2. The denominator. First, item 3.1 shows $n^{-1}\sum_{t=1}^{n}\hat{v}_{t,n}^{2} = \sigma^2 + n^{-1}\sum_{t=1}^{n}(\hat{v}_{t,n}^{2} - \sigma^2) = \sigma^2 + \mathrm{o_P}(1)$. Second, noting $\hat{v}_{t,n} = \varepsilon_t + (\hat{v}_{t,n} - \varepsilon_t)$ a binomial expansion gives

$$n^{-1}\sum_{t=1}^{n}\hat{v}_{t,n}^{4} = n^{-1}\sum_{t=1}^{n}\varepsilon_t^4 + \sum_{j=1}^{4}\binom{4}{j}n^{-j/4}[n^{-1}\sum_{t=1}^{n}\{n^{1/4}(\hat{v}_{t,n} - \varepsilon_t)\}^j\varepsilon_t^{4-j}].$$

Expansion (A.27) shows $n^{1/4}(\hat{v}_{t,n} - \varepsilon_t) = \lambda m_{t,n} + \mathrm{O_P}(n^{-1/4})$ so that a binomial expansion gives

$$\{n^{1/4}(\hat{v}_{t,n} - \varepsilon_t)\}^j = (\lambda m_{t,n})^j + \sum_{h=1}^{j}\binom{j}{h}\mathrm{O_P}(n^{-h/4})(\lambda m_{t,n})^{j-h}.$$

Insert this to get

$$n^{-1}\sum_{t=1}^{n}\hat{v}_{t,n}^{4} = n^{-1}\sum_{t=1}^{n}\varepsilon_t^4 + \sum_{j=1}^{4}\binom{4}{j}(n^{-1/4}\lambda)^j(n^{-1}\sum_{t=1}^{n}m_{t,n}^{j}\varepsilon_t^{4-j})$$
$$+\sum_{j=1}^{4}\binom{4}{j}\sum_{h=1}^{j}\binom{j}{h}(n^{-1/4}\lambda)^{j-h}\mathrm{O_P}(n^{-h/2})n^{-1}\sum_{t=1}^{n}m_{t,n}^{j-h}\varepsilon_t^{4-j}. \quad \text{(A.29)}$$

Lemma A.21(h,j) shows $n^{-1}\sum_{t=1}^{n}m_{t,n}^{j}\varepsilon_t^{4-j} = \mathrm{O_P}(1)$ so that $n^{-1}\sum_{t=1}^{n}\hat{v}_{t,n}^{4} = n^{-1}\sum_{t=1}^{n}\varepsilon_t^4 + \mathrm{o_P}(1) = \varphi^2 + \sigma^4 + \mathrm{o_P}(1)$, see also Lemma A.1. In combination we have $\hat{\varphi}_n^2 = \varphi^2 + \mathrm{o_P}(1)$.

4. Combine items to get the desired result. □

Remark A.1 *The consistency of the $CUSQ_n$ test can be explored by inspecting the expansions (A.28), (A.29) for numerator and denominator of the test statistic. It can be seen that if λ in (5.1) is replaced by a non-decreasing sequence λ_n then $CUSQ_n$ diverges at the rate of* $\min(\lambda_n^2, n^{1/2})$. *This result arises because the remainder terms are of order $n^{-1/4}\lambda_n$. The argument goes as follows.*
If $\lambda_n/n^{1/4} \to 0$ but $\lambda_n \to \infty$ then the term $\lambda_n^2\mathcal{L}_u^\circ$ dominates in the numerator (A.28) and φ^2 dominates in the denominator (A.29) so that

$$CUSQ_n \sim \lambda_n^2 \sup_{0\le u\le 1}\frac{|\mathcal{L}_u^\circ|}{\varphi}.$$

If $\lambda_n/n^{1/4} = c \in \mathbb{R}_+$ then $\lambda_n^2\mathcal{L}_u^\circ = n^{1/2}c^2\mathcal{L}_u^\circ$ dominates in the numerator (A.28) as before but all terms in the first j sum contribute in the denominator (A.29) so that, for $E(\varepsilon_t^3|\mathcal{F}_{t-1}) = \mu_3$,

$$CUSQ_n \sim \sup_{0\le u\le 1}\frac{n^{1/2}c^2|\mathcal{L}_u^\circ|}{(\varphi^2 + 4c\mu_3\int_0^1\mathcal{M}_s ds + 6\sigma^2\int_0^1\mathcal{M}_s^2 ds + \int_0^1\mathcal{M}_s^4 ds)^{1/2}}.$$

If $\lambda_n/n^{1/4} \to \infty$ the term $\lambda_n^2\mathcal{L}_u^\circ$ dominates in the numerator (A.28) as before and the $j = 4$ term in the first j sum dominates in the denominator (A.29). Thus $\hat{\varphi}_n^2 \sim (n^{-1/4}\lambda_n)^4\int_0^1\mathcal{M}_s^4 ds$ so that

$$CUSQ_n \sim \sup_{0\le u\le 1}\frac{\lambda_n^2|\mathcal{L}_u^\circ|}{\{(n^{-1/4}\lambda_n)^4\int_0^1\mathcal{M}_s^4 ds\}^{1/2}} = n^{1/2}\sup_{0\le u\le 1}\frac{|\mathcal{L}_u^\circ|}{(\int_0^1\mathcal{M}_s^4)^{1/2}}.$$

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, respectively. x_t is $I(\tau)$ with $\tau = 0.7, 1, 2$. ε_t, v_t are $i.i.d.N(0,1)$. 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: 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	$1+0.9x_t 1(v_t\le 0)+0.5x_t 1(v_t>0)+\varepsilon_t$	$\theta_1+\theta_2 x_t 1(v_t\le 0)+\theta_3 x_t 1(v_t>0)$
CS	4	$1+0.3\|x_t\|^{1.5}+\varepsilon_t$	$\theta_1+\theta_2\|x_t\|^{\theta_3}$
M	5	$1+0.5x_t+z_t+\varepsilon_t$	$\theta_1+\theta_2 x_t$
M	6	$1+0.5x_t^2+z_t+\varepsilon_t$	$\theta_1+\theta_2 x_t^2$
M	7	$1+0.9x_t 1(v_t\le 0)+0.5x_t 1(v_t>0)+z_t+\varepsilon_t$	$\theta_1+\theta_2 x_t 1(v_t\le 0)+\theta_3 x_t 1(v_t>0)$
M	8	$1+0.3\|x_t\|^{1.5}+z_t+\varepsilon_t$	$\theta_1+\theta_2\|x_t\|^{\theta_3}$

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. $x_t=(1+c/n)x_{t-1}+u_t$ with $c=-2,-10,-50$ for CS and $c=0$ for M. $z_t=\alpha z_{t-1}+\nu_t$ with $\alpha=0.8,0.9,0.99$. ε_t, ν_t,u_t,v_t are $i.i.d.N(0,1)$ and independent of each other.

Table 4: Size: Finite Sample Performance

$x_t=(1+c/n)x_{t-1}+u_t$		$c=-5$			$c=-20$			$c=-50$	
$CUSQ_n$		n			n			n	
*	DGP	100	500	1000	100	500	1000	100	500	1000
CS	1	0.031	0.041	0.043	0.030	0.040	0.043	0.031	0.040	0.044
CS	2	0.032	0.041	0.043	0.032	0.040	0.043	0.028	0.040	0.044
CS	3	0.031	0.039	0.042	0.029	0.040	0.042	0.030	0.040	0.042
CS	4	0.031	0.040	0.042	0.030	0.040	0.043	0.030	0.039	0.044

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

Table 5: Power: Finite Sample Performance

$x_t=x_{t-1}+u_t$	$z_t=\alpha z_{t-1}+\nu_t$	$\alpha=0.7$			$\alpha=0.8$			$\alpha=0.9$		
$CUSQ_n$		n			n			n		
*	DGP	100	500	1000	100	500	1000	100	500	1000
M	5	0.104	0.189	0.215	0.185	0.387	0.443	0.371	0.771	0.838
M	6	0.107	0.190	0.217	0.189	0.387	0.444	0.374	0.772	0.841
M	7	0.103	0.193	0.222	0.186	0.390	0.444	0.368	0.765	0.846
M	8	0.096	0.187	0.212	0.174	0.381	0.436	0.347	0.760	0.835

M denotes misspecification; hence, power is considered in those cases. 10000 replications are conducted.

Table 6: Power performance comparison with Kasparis (2008)

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

$z_t = z_{t-1} + w_t$ 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 are $i.i.d.N(0,1)$ and $D = [1\ .2\ .1,\ .3\ 2\ 0,\ 0\ .1\ 1.2]$

Table 7: Power performance comparison with Kasparis (2008)

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

References

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.

Chan, N.H. and C.Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. Annals of Satistics 16, 367-401.

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.

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 (2008a) 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.

Deng, A. and P. Perron (2008b) The limit distribution of the CUSUM of squares test under general mixing conditions. Econometric Theory 24, 809-822.

Dickey, D.A. and Fuller, W.A. (1979) Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427-431.

Donati-Martin, C. and Yor, M. (1991) Fubini's theorem for double Wiener integrals and the variance of the Brownian path. Annales de l'Institut Henri Poincaré 27, 181-200.

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.

Hammersley, J.M. (1956) The zeros of a random polynomial. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability 2, 89-111.

Jakubowski, A., Ménin, J. and Pages, G. (1989) Convergence en loi des suites d'intégrales stochastiques sur l'espace D^1de Skorokhod. Probability Theory and Related Fields 81, 111-137.

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.

Kasparis, I., Andreou, E. and P.C.B. Phillips (2015) Nonparametric predictive regression. Journal of Econometrics 185, 468-494.

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

Kurtz, T.G. and P. Protter (1991) Weak limit theorems for stochastic integrals and stochastic differential equations. Annals of Probability 19, 1035-1070.

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.

Liang, H., P.C.B. Phillips and Q. Wang (2016) Weak convergence to stochastic integrals for econometric applications. Econometric Theory 32, 1349-1375.

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. 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.

Phillips, P.C.B. (2007) Regression with slowly varying regressors and nonlinear trends. Econometric Theory 23, 557-614.

Phillips. P.C.B. and Solo, V. (1992) Asymptotics for linear processes. Annals of Statistics 20, 971–1001.

Pitarakis, JY (2017) A Simple Approach for Diagnosing Instabilities in Predictive Regressions, Oxford Bulletin of Economics and Statistics, forthcoming.

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.

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

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

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 (2012) A specification test for nonlinear nonstationary models. The Annals of Statistics 40, 727–758.

Wu, C.F. (1981) Asymptotic theory of nonlinear least squares estimation. The Annals of Statistics 9, 501–513.

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