Econometric Theory, **27**, 2011, 913–927.
doi:10.1017/S0266466610000563

NOTES AND PROBLEMS

ASYMPTOTIC BEHAVIOR OF THE CUSUM OF SQUARES TEST UNDER STOCHASTIC AND DETERMINISTIC TIME TRENDS

BENT NIELSEN AND JOUNI S. SOHKANEN
University of Oxford

We generalize the cumulative sum of squares (CUSQ) test to the case of nonstationary autoregressive distributed lag models with deterministic time trends. The test may be implemented with either ordinary least squares residuals or standardized forecast errors. In explosive cases the asymptotic theory applies more generally for the least squares residuals-based test. Preliminary simulations of the tests suggest a very modest difference between the tests and a very modest variation with nuisance parameters. This supports the use of the tests in explorative analysis.

1. INTRODUCTION

Cumulative sum of squares (CUSQ) tests are used for testing constancy of the variance of regression errors. The tests were proposed for the fixed regressor case by Brown, Durbin, and Evans (1975). The CUSQ test may be implemented with either least squares residuals or forecast residuals. Here we investigate the behavior of both CUSQ tests when applied to autoregressive distributed lag models, possibly with deterministic trends and unit root and explosive stochastic trends.

We show that the usual asymptotic distribution applies quite generally for the least squares test. This is important in applications as the question of variance constancy can be addressed without having to locate the characteristic roots. For the forecast test the usual asymptotic distribution applies in nonexplosive and purely explosive cases, but nuisance terms may arise in explosive cases. The results generalize work for stationary cases by Ploberger and Krämer (1986) and Deng and Perron (2008a). Lee et al. (2003) considered the least squares test for a unit root autoregression without deterministic terms. The present analysis is based on the results of Lai and Wei (1985) and Nielsen (2005), and is easiest for the least squares test.

The authors received support from the ESRC (RES-000-27-0179 and PTA-031-2006-00174), the Open Society Institute and the Oxford Martin School. Comments from Andrew Whitby are gratefully acknowledged. Address correspondence to Bent Nielsen, Nuffield College, Oxford OX1 1NF, UK; e-mail: bent.nielsen@nuffield.ox.ac.uk.

A small-scale simulation study indicates that there is not much diffence in the finite sample distribution of the two test statistics. This adheres to the findings of Deng and Perron (2008a) that, in the context of stationary models, there is not much difference in size or power when applying the statistics to test for changes in the residual variance. Moreover, the finite sample distributions vary very little with nuisance parameters, indicating that the tests are approximately similar.

The paper is organized so that the two test statistics are presented in §2 while the model assumptions are presented in Section 3. The asymptotic results for the least squares test and the forecast test are presented in Sections 4 and 5, respectively. Section 6 contains a simulation study involving first-order autoregressions. The proofs are given in the Appendix. We refer to Nielsen and Sohkanen (2009) for an empirical illustration.

2. THE TEST STATISTICS

The forecast residual-based test statistic was suggested along with exact distribution results by Brown et al. (1975) for the classical linear regression

$$y_t = \beta' x_t + \varepsilon_t \quad \text{for } t = 1, \ldots, T, \tag{2.1}$$

where y_t is a scalar, x_t is an M-dimensional regressor, and the errors are independently normal, $\mathsf{N}(0, \sigma^2)$-distributed. Computing recursive least squares estimators as

$$\hat{\beta}_t = \left(\sum_{s=1}^{t} x_s x_s' \right)^{-1} \sum_{s=1}^{t} x_s y_s \quad \text{for } t = M, \ldots, T, \tag{2.2}$$

along with the recursive forecast residuals

$$\tilde{\varepsilon}_t = \left\{ 1 + x_t' \left(\sum_{s=1}^{t-1} x_s x_s' \right)^{-1} x_t \right\}^{-1/2} (y_t - \hat{\beta}_{t-1}' x_t) \quad \text{for } t > M, \tag{2.3}$$

the CUSQ plot with recursive residuals is defined as

$$\text{CUSQ}_{t,T}^{REC} = \sqrt{T} \left(\frac{\sum_{s=M}^{t} \tilde{\varepsilon}_s^2}{\sum_{s=M}^{T} \tilde{\varepsilon}_s^2} - \frac{t - M}{T - M} \right) \quad \text{for } t \geq M. \tag{2.4}$$

The alternative least squares residual-based test statistic was mentioned in passing by Brown et al. (1975) and analyzed in detail by McCabe and Harrison (1980). Computing recursive residual variances

$$\hat{\sigma}_t^2 = t^{-1} \sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2 \quad \text{for } M \leq t, \tag{2.5}$$

based on the least squares residuals

$$\hat{\varepsilon}_{s,t} = y_s - \hat{\beta}_t' x_s \quad \text{for } M \leq t, \tag{2.6}$$

the CUSQ plot with least squares residuals is defined as

$$\text{CUSQ}_{t,T}^{OLS} = t/\sqrt{T}\left(\frac{\hat{\sigma}_t^2}{\hat{\sigma}_T^2} - 1\right) = \sqrt{T}\left(\frac{\sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2}{\sum_{s=1}^{T} \hat{\varepsilon}_{s,T}^2} - \frac{t}{T}\right) \quad \text{for } t > M. \tag{2.7}$$

3. MODEL AND ASSUMPTIONS

To facilitate an analysis of trending time series we focus on autoregressive distributed lag regressions and assume vector autoregressive behavior for the variables involved.

Suppose a p-dimensional time series $X_{1-k}, \ldots, X_0, \ldots, X_T$ is observed and that X_t is partitioned as $\left(Y_t, Z_t'\right)'$ where Y_t is univariate and Z_t is of dimension $p-1 \geq 0$. The autoregressive distributed lag regression of order k is given by

$$Y_t = \rho Z_t + \sum_{j=1}^{k} \alpha_j Y_{t-j} + \sum_{j=1}^{k} \beta_j' Z_{t-j} + \nu D_{t-1} + \varepsilon_t, \qquad t = 1, \ldots T, \tag{3.1}$$

where D_t is a deterministic term. When the time series is univariate so $p = 1$ and $X_t = Y_t$, the regression reduces to a univariate autoregression. A variant of the regression omits the contemporaneous regressor Z_t, giving the regression

$$Y_t = \sum_{j=1}^{k} \alpha_j Y_{t-j} + \sum_{j=1}^{k} \beta_j' Z_{t-j} + \nu D_{t-1} + \varepsilon_t, \qquad t = 1, \ldots T. \tag{3.2}$$

In order to characterize the asymptotic distribution of our test statistics, the joint distribution of the time series $X_t = \left(Y_t, Z_t'\right)'$ needs to be specified. We will assume that X_t and D_t satisfy the vector autoregressions

$$X_t = \sum_{j=1}^{k} A_j X_{t-j} + \mu D_{t-1} + \xi_t, \qquad t = 1, \ldots T, \tag{3.3}$$

$$D_t = \mathbf{D} D_{t-1}, \tag{3.4}$$

where $\mathbf{D}$ is a deterministic matrix with properties to be given below. The innovations ξ_t satisfy a martingale difference assumption.

Assumption A. Assume $(\xi_t, \mathcal{F}_t)$ is a martingale difference sequence, so $\mathsf{E}(\xi_t | \mathcal{F}_{t-1}) = 0$. The initial values $X_0, \ldots, X_{1-k}$ are $\mathcal{F}_0$-measurable and

$$\sup_t \mathsf{E}\{(\xi_t' \xi_t)^{\lambda/2} | \mathcal{F}_{t-1}\} \overset{a.s.}{<} \infty \quad \text{for some } \lambda > 4, \tag{3.5}$$

$$\mathsf{E}(\xi_t \xi_t' | \mathcal{F}_{t-1}) \overset{a.s.}{=} \Omega \quad \text{where } \Omega \text{ is positive definite.} \tag{3.6}$$

The deterministic term D_t is a vector of polynomial, periodic terms. For example,

$$\mathbf{D} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad \text{with} \quad D_0 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

generates a linear trend, a constant, and a biannual dummy. Specifically, the deterministic term satisfies the following assumption.

Assumption B. $|\text{eigen}(\mathbf{D})| = 1$ and $\text{rank}(D_1, \ldots, D_{\dim \mathbf{D}}) = \dim \mathbf{D}$.

Nearly all values of autoregressive parameters A_j are allowed in the vector autoregression (3.3), including stationary roots, roots on the unit circle, and a range of explosive roots. The only exception is the case of singular explosive roots that can arise for vector autoregressions where $p \geq 2$ with more than one explosive root; see Anderson (1959), Duflo et al. (1991), Phillips and Magdalinos (2008), and Nielsen (2008) for further discussion. Thus, define the companion matrices

$$\mathbf{B} = \begin{Bmatrix} (A_1, \ldots, A_{k-1}) & A_k \\ I_{p(k-1)} & 0 \end{Bmatrix}, \qquad \mu = \begin{Bmatrix} \mu \\ 0 \end{Bmatrix}, \qquad \mathbf{S} = \begin{Bmatrix} \mathrm{B} & \mu \\ 0 & \mathrm{D} \end{Bmatrix}. \tag{3.7}$$

Assumption C. All explosive roots of $\mathbf{B}$ have geometric multiplicity of unity. That is, for all complex λ so $|\lambda| > 1$ then $\text{rank}(\mathbf{B} - \lambda I_{pk}) \geq pk - 1$.

The parameters and innovations of the regressions (3.1) and (3.2) can be linked to the vector autoregression (3.3) through the limits of the least squares estimators arising from (3.1) and (3.2). For this purpose define

$$\xi_t = \begin{pmatrix} \xi_t^{(1)} \\ \xi_t^{(2)} \end{pmatrix}, \qquad \Omega = \begin{pmatrix} \Omega_{yy} & \Omega_{yz} \\ \Omega_{zy} & \Omega_{zz} \end{pmatrix},$$

conformably with $X_t = (Y_t, Z_t')'$. It then holds for equation (3.1) that

$$\rho = \Omega_{yz}\Omega_{zz}^{-1}, \qquad \varepsilon_t = (1, -\rho)\xi_t, \qquad (\alpha_j, \beta_j') = (1, -\rho)A_j,$$
$$\sigma^2 = \Omega_{yy} - \Omega_{yz}\Omega_{zz}^{-1}\Omega_{zy},$$

where σ^2 is the variance of the innovation ε_t. Similarly, for equation (3.2) it holds that

$$(\alpha_j, \beta_j') = (1, 0)A_j, \qquad \varepsilon_t = (1, 0)\xi_t, \qquad \sigma^2 = \Omega_{yy}.$$

In addition an invariance principle for the partial sums of squared innovations $\sum_{s=1}^t \varepsilon_s^2$ and a law of large numbers for $\sum_{s=1}^t \varepsilon_s^4$ are needed. Such results could be assumed. To be more explicit we assume a martingale structure for ε_t^2.

Assumption D. For the regression (3.1), $\mathcal{G}_{t-1}$ is the σ-field generated by Z_t and $\mathcal{F}_{t-1}$, while $\mathcal{G}_t = \mathcal{F}_t$ for the regression (3.2). Suppose $(\varepsilon_t^2 - \sigma^2, \mathcal{G}_t)$ is a martingale difference sequence satisfying $\text{Var}(\varepsilon_t^2 - \sigma^2 | \mathcal{G}_{t-1}) = \varphi^2$ a.s. for some $\varphi > 0$ and $\sup_t \mathsf{E}(|\varepsilon_t|^\lambda | \mathcal{G}_{t-1}) < \infty$ a.s. for some $\lambda > 4$.

4. ASYMPTOTIC ANALYSIS OF THE CUSQ^{OLS}-STATISTIC

Consider the CUSQ^{OLS}-statistic (2.7) based on the autoregressive distributed lags residuals of (3.1) or (3.2). The key to the asymptotic analysis is to generalize Deng and Perron (2008a, Lem. 2) showing that the sum of squared residuals is close to the sum of squared innovations. A first step is the following lemma.

LEMMA 4.1. *Assume A, B, C. Then* $t^{-1/2} \sum_{s=1}^t (\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2) \to 0$ *a.s.*

The next step is to turn this into a result about $x_t = T^{-1/2} \sum_{s=1}^t (\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)$. Due to the next lemma then $\sup_{t \le T}(|x_t|) \to 0$ a.s., so $x_{\text{int}(Tu)}$ vanishes on $D[0,1]$, the space of right-continuous functions on $[0,1]$ with left limits.

LEMMA 4.2. *Let x_t be a sequence so $t^{-1/2} x_t \to 0$. Then* $\sup_{t \le T} T^{-1/2} |x_t| \to 0$.

The normalized partial sums of squared innovations are asymptotically Brownian. This follows through a direct application of Chan and Wei (1988, Thm. 2.2).

LEMMA 4.3. *Assume D. Let $\mathcal{B}$ be a standard Brownian motion. Then, for $u \in [0,1]$, it holds that* $T^{-1/2} \sum_{s=1}^{int(Tu)} (\varepsilon_s^2 - \sigma^2) \to \varphi \mathcal{B}_u$ *in distribution on* $D[0,1]$.

The main result concerning the CUSQ^{OLS}-statistic now follows.

THEOREM 4.4. *Assume A, B, C, D. Let $\mathcal{B}^\circ$ be a standard Brownian bridge. Then*

(i) $CUSQ^{OLS}_{int(Tu),T} \to \sigma^{-2} \varphi \mathcal{B}_u^\circ$ *in distribution on* $D[0,1]$.

(ii) $\sup_{t \le T} |CUSQ^{OLS}_{t,T}| \to \sigma^{-2} \varphi \sup_{u \le 1} |\mathcal{B}_u^\circ|$ *in distribution on* $\mathbb{R}$.

The above result involves a nuisance parameter φ. For normal innovations it holds that $\sigma^{-2}\varphi = \sqrt{2}$. In general, σ^2 is estimated consistently by the sample variance due to Lemma 4.1, whereas φ is estimated consistently by a fourth-moment estimator as shown next.

THEOREM 4.5. *Assume A, B, C, D. Then* $\hat{\varphi}_t^2 = t^{-1} \sum_{s=1}^t \hat{\varepsilon}_{s,t}^4 - (t^{-1} \sum_{s=1}^t \hat{\varepsilon}_{s,t}^2)^2 \xrightarrow{P} \varphi^2$.

Remark 4.6. The convergence in Theorem 4.5 could be strengthened to almost sure convergence if it were assumed that ε_t^3 is a martingale difference.

A convergence result for φ_t as a process on $D[0,1]$ could then be deduced. The proof would follow by combining the presented proof with Theorem 2.4 of Nielsen (2005).

5. ASYMPTOTIC BEHAVIOR OF THE CUSQ^{REC}-TEST

Consider now the CUSQ^{REC}-statistic (2.4) applied to regressions (3.1) and (3.2). This statistic is more complicated to describe than the CUSQ^{OLS}-statistic. A nuisance term arises in special cases.

In order to generalize Deng and Perron (2008a, Lem. 2) decompose the vector autoregression into its nonexplosive and explosive parts. Thus, define the companion vector $S_{t-1} = (X'_{t-1}, \dots X'_{t-k}, D'_{t-1})$ and the selection matrix $\iota = (I_p, 0_{(pk-p+\dim \mathbf{D})\times p})'$. Recalling the companion matrix $\mathbf{S}$ defined in (3.7), the vector autoregression satisfies a first-order vector autoregression $S_t = \mathbf{S}S_{t-1} + \iota\xi_t$. As noted in, for instance, Nielsen (2005, Sect. 3), there exists a real matrix M so $M\mathbf{S}M^{-1}$ is block diagonal and

$$MS_t = \begin{pmatrix} R_t \\ W_t \end{pmatrix} = \begin{pmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{W} \end{pmatrix} \begin{pmatrix} R_{t-1} \\ W_{t-1} \end{pmatrix} + \begin{pmatrix} e_{R,t} \\ e_{W,t} \end{pmatrix}, \tag{5.1}$$

where the absolute values of the eigenvalues of $\mathbf{R}$ and $\mathbf{W}$ are at most one and greater than one, respectively. Deterministic components are subsumed into the R_t -process.

The difference between the sum of squared forecast residuals and the sum of squared innovations will in general involve a nuisance term.

LEMMA 5.1. *Assume A, B, C and that either* $\dim \mathbf{R} = 0$ *or* $\dim \mathbf{W} = 0$. *Then* $\sum_{s=1}^{t}(\tilde{\varepsilon}_s^2 - \varepsilon_s^2) = \mathrm{o}(t^{1/2})$ *a.s.*

Remark 5.2. If the process is mixed so that $\dim \mathbf{R} > 0$ and $\dim \mathbf{W} > 0$ then several nonnegligible nuisance terms will appear in Lemma 5.1. It is not immediately clear if these nuisance terms will cancel each other.

A limiting result for the CUSQ^{REC} then follows by exactly the same argument as that of Theorem 4.4, replacing Lemma 4.1 by Lemma 5.1.

THEOREM 5.3. *Assume A, B, C, D and that either* $\dim \mathbf{R} = 0$ *or* $\dim \mathbf{W} = 0$. *Then*

(i) $CUSQ^{REC}_{int(Tu),T} \to \sigma^{-2}\varphi\mathcal{B}^{\circ}_u$ *in distribution on* $D[0,1]$,

(ii) $\sup_{t\le T} |CUSQ^{REC}_{t,T}| \to \sigma^{-2}\varphi \sup_{u\le 1} |\mathcal{B}^{\circ}_u|$ *in distribution on* $\mathbb{R}$.

6. SIMULATION STUDY

Theorems 4.4 and 5.3 show that the two types of CUSQ-statistics have the usual limit distribution in many situations. This leaves the questions of whether the

TABLE 1. Simulated means and medians of the CUSQ tests for different values of α. The Monte Carlo standard error is 2.5×10^{-4}. The slight variation in the reported figures is therefore significant.

	S^{OLS}		S^{REC}	
α	Mean	Median	Mean	Median
-1.2	0.790	0.750	0.797	0.758
-1.0	0.789	0.749	0.796	0.757
-0.9	0.789	0.748	0.795	0.756
0.0	0.788	0.748	0.795	0.756
0.9	0.789	0.748	0.795	0.756
1.0	0.789	0.749	0.796	0.757
1.2	0.790	0.749	0.797	0.758

finite sample distributions are different for the two statistics and whether they depend on the autoregressive parameters. These questions are addressed through a small-scale Monte Carlo study. For the important question of the power of these tests, we refer to Deng and Perron (2008a, 2008b).

The data generating process is a univariate autoregression, $X_t = \alpha X_{t-1} + \varepsilon_t$ for $t = 1, \ldots, T = 100$ with initial value $X_0 = 0$, standard normal innovations, and a range of autoregressive parameters α. The number of repetitions was 10^6. Due to the normality $\sigma^{-2}\varphi = \sqrt{2}$, so the statistics of interest are $S^{OLS} = \max_{M \le t \le T} |\text{CUSQ}_{t,T}^{OLS}|/\sqrt{2}$ and $S^{REC} = \max_{M \le t \le T} |\text{CUSQ}_{t,T}^{REC}|/\sqrt{2}$; see (2.5). Theorems 4.4 and 5.3 show that their limit distribution is the supremum of a Brownian bridge. Billingsley (1999, pp. 101–104) gives an analytic expression for the distribution function. In particular, the 95% quantile is 1.36; see Schumacher (1984, Tab. 9)

Table 1 reports the mean and median of the two statistics as a function of α. The variation of the distribution for the two supremum statistics is very small but significant when taking the Monte Carlo precision into account. This impression was confirmed when looking at other descriptives such as standard deviation, 95% quantile, and p-value of the asymptotic 95% quantile when $\alpha = 0$.

Two conclusions emerge from this small-scale Monte Carlo study. First, there is not much difference in finite sample distribution for the two statistics. Second, there is very little variation in the finite sample distribution with the unknown parameter. This suggests that very simple finite sample corrections could be used.

REFERENCES

Anderson, T.W. (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676–687.

Billingsley, P. (1999) *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley.

Brown, R., J. Durbin, & J. Evans (1975) Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society*. Series B (Methodological) 37, 149–192.

Chan, N.H. & C.Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregessive processes. *Annals of Statistics* 16, 367–401.

Deng, A. & P. Perron (2008a) The limit distribution of the CUSUM of squares test under general mixing conditions. *Econometric Theory* 24, 809–822.

Deng, A. & P. Perron (2008b) 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.

Duflo, M., R. Senoussi, & R. Touati (1991) Propriétés asymptotiques presque sûre de l'estimateur des moindres carrés d'un modèle autorégressif vectoriel. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* 27, 1–25.

Hall, P. & C. Heyde (1980) *Martingale Limit Theory and Its Application* (*Probability and Mathematical Statistics*). Academic Press.

Lai, T.L. & 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.

Lai, T.L. & C.Z. Wei (1985) Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In P. Krishnaiah (ed.), *Multivariate Analysis VI*, pp. 375–393. Elsevier Science.

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

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

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

Nielsen, B. (2008) Singular Vector Autoregressions with Deterministic Terms: Strong Consistency and Lag Order Determination. Discussion paper, Nuffield College.

Nielsen, B. & J.S. Sohkanen (2009) Asymptotic Behaviour of the CUSUM of Squares Test under Stochastic and Deterministic Time Trends. Discussion paper, Nuffield College.

Phillips, P. & T. Magdalinos (2008) Limit theory for explosively cointegrated systems. *Econometric Theory* 24, 865–887.

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

Schumacher, M. (1984) Two-sample tests of Cramer-von Mises and Kolmogorov-Smirnov type for randomly censored data. *International Statistical Review* 52, 263–281.

APPENDIX: Proofs

Notation: For a matrix m, let $\|m\|^2 = \lambda_{\max}\left(mm'\right)$, where $\lambda_{\max}$ gives the greatest eigenvalue of the matrix.

A.1. The Case of Least Squares Residuals

Proof of Lemma 4.1. Partition ξ_t as $(\xi_t^{(1)}, \xi_t^{(2)\prime})'$ and partition the least squares residuals, $\hat{\xi}_{s,t}$, of X_t on $X_{t-1}, \dots X_{t-k}$ and D_{t-1} conformably. We start by arguing

$$\frac{1}{t}\sum_{s=1}^{t}(\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2) \overset{a.s.}{=} \mathrm{o}(t^{-1/2}). \tag{A.1}$$

First, if Z_t is excluded as regressor as in (3.2) then $\sum_{s=1}^{t}\hat{\varepsilon}_{s,t}^2 = \sum_{s=1}^{t}(\hat{\xi}_{s,t}^{(1)})^2$. Combine this with Nielsen (2005, Cor. 2.6) to see that $t^{-1}\sum_{s=1}^{t}(\hat{\xi}_{s,t}\hat{\xi}_{s,t}' - \xi_s\xi_s') = \mathrm{o}(t^{-1/2})$ a.s., assuming Assumptions A, B, C. The result (A.1) then follows.

Second, if Z_t is included as regressor as in (3.1), then

$$\sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2 = \sum_{s=1}^{t} (\hat{\xi}_{s,t}^{(1)})^2 - \sum_{s=1}^{t} \hat{\xi}_{s,t}^{(1)} \hat{\xi}_{s,t}^{(2)\prime} \left\{ \sum_{s=1}^{t} \hat{\xi}_{s,t}^{(2)} \hat{\xi}_{s,t}^{(2)\prime} \right\}^{-1} \sum_{s=1}^{t} \hat{\xi}_{s,t}^{(2)} \hat{\xi}_{s,t}^{(1)}.$$

By Nielsen (2005, Cor. 2.6) then

$$\sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2 \overset{a.s.}{=} \left[\sum_{s=1}^{t} (\xi_s^{(1)})^2 - \sum_{s=1}^{t} \xi_s^{(1)} \xi_s^{(2)\prime} \left\{ \sum_{s=1}^{t} \xi_s^{(2)} \xi_s^{(2)\prime} \right\}^{-1} \sum_{s=1}^{t} \xi_s^{(2)} \xi_s^{(1)} \right] \{1 + \mathrm{o}(t^{-1/2})\}.$$

Since $\xi_s^{(1)} = \varepsilon_s + \rho \xi_s^{(2)}$ then

$$\sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2 \overset{a.s.}{=} \left[\sum_{s=1}^{t} \varepsilon_s^2 - \sum_{s=1}^{t} \varepsilon_s \xi_s^{(2)\prime} \left\{ \sum_{s=1}^{t} \xi_s^{(2)} \xi_s^{(2)\prime} \right\}^{-1} \sum_{s=1}^{t} \xi_s^{(2)} \varepsilon_s \right] \{1 + \mathrm{o}(t^{-1/2})\}.$$

Using that $\mathsf{E}(\varepsilon_s \xi_s^{(2)\prime}) = (1, -\rho)\Omega(0, I)' = 0$ along with Nielsen (2005, Thm. 2.8) shows that $t^{-1}\sum_{s=1}^{t} \varepsilon_s \xi_s^{(2)\prime} = \mathrm{o}(t^{-1/4})$ a.s. so that (A.1) follows. ■

Proof of Lemma 4.2. Since $t^{-1/2}x_t \to 0$ then a finite t_0 exists such that $t^{-1/2}|x_t| < \epsilon$ for all $t > t_0$. Since $t \le T$, then $T^{-1/2} \le t^{-1/2}$ so $T^{-1/2}|x_t| < \epsilon$ for all $t > t_0$. It follows that $\sup_{t>t_0} T^{-1/2}|x_t| < \epsilon$. Moreover, since t_0 is finite, then $\max_{t \le t_0} |x_t|$ is finite and we also have $\sup_{t \le t_0} T^{-1/2}|x_t| < \epsilon$. In combination we have that for T sufficiently large then $\sup_{t \le T} T^{-1/2}|x_t| < \epsilon$. The desired result follows since ϵ was arbitrary. ■

Proof of Theorem 4.4.

(i) Lemmas 4.1, 4.2, and 4.3 imply

$$T^{-1/2} \sum_{s=1}^{\mathrm{int}(Tu)} (\hat{\varepsilon}_{s,\mathrm{int}(Tu)}^2 - \sigma^2)$$
$$= T^{-1/2} \sum_{s=1}^{\mathrm{int}(Tu)} (\hat{\varepsilon}_{s,\mathrm{int}(Tu)}^2 - \varepsilon_s^2) + T^{-1/2} \sum_{s=1}^{\mathrm{int}(Tu)} (\varepsilon_s^2 - \sigma^2) \overset{\mathsf{D}}{\to} \varphi \mathcal{B}_u$$

on $D[0, 1]$. Next, rewrite the CUSQ-statistic as

$$\mathrm{CUSQ}_{\mathrm{int}(Tu),T}^{OLS} = \frac{T^{-1/2}\{\sum_{s=1}^{\mathrm{int}(Tu)} (\hat{\varepsilon}_{s,\mathrm{int}(Tu)}^2 - \sigma^2) - T^{-1} t \sum_{s=1}^{T} (\hat{\varepsilon}_{s,T}^2 - \sigma^2)\}}{T^{-1}\sum_{s=1}^{T} \hat{\varepsilon}_{s,T}^2},$$

and insert the above convergence result.

(ii) Taking supremum entails taking a continuous mapping on $D[0, 1]$. ■

A.2. Consistency of $\hat{\varphi}_t$

Proof of Theorem 4.5. The result is proved for the regression (3.1) including Z_t as regressor. The argument for the regression (3.2) can be made in a similar way.

Due to Lemma 4.1, $t^{-1}\sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2$ and $t^{-1}\sum_{s=1}^{t} \varepsilon_s^2$ have the same limit. If the same is shown for $t^{-1}\sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^4$ and $t^{-1}\sum_{s=1}^{t} \varepsilon_s^4$, then the desired result follows from a law of large numbers applied to $t^{-1}\sum_{s=1}^{t} \varepsilon_s^2$ and $t^{-1}\sum_{s=1}^{t} \varepsilon_s^4$, assuming A and D.

Since $Z_t = \theta S_{t-1} + \xi_t^{(2)}$ for some θ (see (3.3)), regression on regressors Z_t, S_{t-1} and on regressors $x_t = (\xi_t^{(2)\prime}, S_{t-1}^{\prime})^{\prime}$ is equivalent. Define

$$P_t = \sum_{s=1}^{t} \varepsilon_s x_s^{\prime} \left(\sum_{s=1}^{t} x_s x_s^{\prime} \right)^{-1/2}, \qquad Q_{s,t} = \left(\sum_{s=1}^{t} x_s x_s^{\prime} \right)^{-1/2} x_s,$$

so $\hat{\varepsilon}_{s,t} = \varepsilon_t - P_t Q_{s,t}$. To prove $\sum_{s=1}^{t} (\hat{\varepsilon}_{s,t}^4 - \varepsilon_s^4) = \mathrm{o}(t)$, apply a binomial expansion to $\hat{\varepsilon}_{s,t}^4$ so it suffices to prove $\mathcal{I}_m = \sum_{s=1}^{t} (P_t Q_{s,t})^m \varepsilon_s^{4-m} = \mathrm{o}_P(t)$ for $m = 1, \ldots, 4$.

First, argue that $P_t = \mathrm{o}(t^{1/4})$ a.s. The series $\xi_s^{(2)}$ and S_{s-1} are asymptotically uncorrelated due to Nielsen (2005, Thm. 2.4) given Assumptions A, B, C, and D. Thus

$$P_t \overset{a.s.}{=} \left\{ \sum_{s=1}^{t} \varepsilon_s \xi_s^{(2)\prime} \left(\sum_{s=1}^{t} \xi_s^{(2)} \xi_s^{(2)\prime} \right)^{-1/2} + \sum_{s=1}^{t} \varepsilon_s S_{s-1}^{\prime} \left(\sum_{s=1}^{t} S_{s-1} S_{s-1}^{\prime} \right)^{-1/2} \right\} \{1 + \mathrm{o}(1)\}.$$

This is of the desired order due to Nielsen (2005, Thms. 2.4, 2.8, Cor. 2.6) given Assumptions A, B, C, and the construction $\mathsf{E}(\varepsilon_s \xi_s^{(2)\prime}) = 0$.

Second, consider $\mathcal{I}_1 = P_t (\sum_{s=1}^{t} x_s x_s^{\prime})^{-1/2} \sum_{s=1}^{t} x_s \varepsilon_t^3$. As in (5.1) we can decompose S_{t-1} into autoregressions $U_{t-1}, V_{t-1}, W_{t-1}$ with stationary, unit and explosive roots. The components $\xi_t^{(2)}, U_{t-1}, V_{t-1}, W_{t-1}$ are asymptotically uncorrelated due to Nielsen (2005, Thms. 2.4, 9.1, 9.2, 9.4) given Assumptions A, B, C. Thus, as above,

$$\begin{aligned} \mathcal{I}_1 \overset{a.s.}{=} P_t \Bigg[& \left\{ \sum_{s=1}^{t} \xi_s^{(2)} \xi_s^{(2)\prime} \right\}^{-1/2} \sum_{s=1}^{t} \xi_s^{(2)} \varepsilon_t^3 + \left\{ \sum_{s=1}^{t} U_{s-1} U_{s-1}^{\prime} \right\}^{-1/2} \sum_{s=1}^{t} U_{s-1} \varepsilon_t^3 \\ & + \left\{ \sum_{s=1}^{t} V_{s-1} V_{s-1}^{\prime} \right\}^{-1/2} \sum_{s=1}^{t} V_{s-1} \varepsilon_t^3 + \left\{ \sum_{s=1}^{t} W_{s-1} W_{s-1}^{\prime} \right\}^{-1/2} \sum_{s=1}^{t} W_{s-1} \varepsilon_t^3 \Bigg] \\ & \{1 + \mathrm{o}(1)\}. \end{aligned} \tag{A.2}$$

The first term of (A.2) involving $\xi_s^{(2)}$ is bounded by

$$\left| P_t \left(\sum_{s=1}^{t} \xi_s^{(2)} \xi_s^{(2)\prime} \right)^{-1/2} \sum_{s=1}^{t} \xi_s^{(2)} \varepsilon_t^3 \right| \leq |P_t| \left\{ \max_{s \leq t} \left| \left(\sum_{s=1}^{t} \xi_s^{(2)} \xi_s^{(2)\prime} \right)^{-1/2} \xi_s^{(2)} \right| \right\} \textstyle\sum_{s=1}^{t} |\varepsilon_s^3|.$$

Here $P_t = \mathrm{o}(t^{1/4})$ *a.s* given Assumption A. The second term is $\mathrm{o}(t^{-1/4-\eta})$ a.s. for some $\eta > 0$ since $t^{-1} \sum_{s=1}^{t} \xi_s^{(2)} \xi_s^{(2)\prime}$ is convergent and $\xi_s^{(2)} = \mathrm{o}(t^{-1/4-\eta})$ a.s. for all $\eta > 0$; see Nielsen (2005, Thms. 5.1, 6.1)). The third term is $\sum_{s=1}^{t} |\varepsilon_s^3| = \mathrm{o}(t^{1+\eta})$ a.s. for all $\eta > 0$; see Nielsen (2005 Thm. 7.3). Overall, the first term of (A.2) is $\mathrm{o}(t)$.

The second term of (A.2) involving U_{s-1} is analyzed the same way.

The third term of (A.2) involving V_{s-1} is bounded by

$$\left| P_t \left(\sum_{s=1}^{t} V_s V_s^{\prime} \right)^{-1/2} \left(\sum_{s=1}^{t} V_s \right) \varepsilon_t^3 \right| \leq |P_t| \left\{ \max_{s \leq t} \left| \left(\sum_{s=1}^{t} V_s V_s^{\prime} \right)^{-1/2} V_s \right| \right\} \sum_{s=1}^{t} |\varepsilon_s^3|.$$

Introduce normalizations for the unit root process V_{s-1} as in Chan and Wei (1988) to see that the second term is $\mathrm{O}_p(t^{-1/2})$ assuming A, B. The last term is $\mathrm{o}(t^{1+\eta})$ for all $\eta > 0$; see Nielsen (2005, Thm. 7.3). Overall, the bound is $\mathrm{o}(t^{3/4+\eta})$.

The fourth term of (A.2) involving W_{s-1} is bounded by

$$\left\|\sum_{s=1}^{t} \mathbf{W}^{-t} W_{s-1} W_{s-1}' (\mathbf{W}')^{-t}\right\|^{-1/2} \left(\sum_{s=1}^{t} \|\mathbf{W}^{-t} W_{s-1}\|\right) \max_{s\le t} \|\varepsilon_s\|^3.$$

The first two terms are convergent, while the last term is $\mathrm{o}(t^{3/4})$ since $\varepsilon_t = \mathrm{o}(t^{1/4})$; see Nielsen (2005, Cors. 5.3, 7.2, Thm. 5.1) assuming A, C.

Third, consider $\mathcal{I}_m$ for $m \geq 2$. The following bound holds:

$$\mathcal{I}_m \leq \|P_t\|^m \max_{s\le t} \|\varepsilon_s\|^{4-m} \sum_{s=1}^{t} (P_{s,t}' P_{s,t})^{m/2}.$$

The first two terms are $\mathrm{o}(t)$ by the arguments above. For the latter term, note that $P_{s,t}' P_{s,t} \leq 1$. Thus, for $m/2 \geq 1$,

$$\sum_{s=1}^{t} (P_{s,t}' P_{s,t})^{m/2} \leq \sum_{s=1}^{t} P_{s,t}' P_{s,t} = \sum_{s=1}^{t} \mathrm{tr}(P_{s,t} P_{s,t}') = \mathrm{tr}(I_{pk}) = pk,$$

so the last term is bounded. ■

A.3. The Case of Recursive Residuals. Lemma 5.1 is proved in three steps. Only the regression (3.1) including Z_t as a regressor is considered. As in the proof of Theorem 4.5, the regressor can be taken as $x_t = (\xi_t^{(2)\prime}, R_{t-1}', W_{t-1}')'$ where $\xi_t^{(2)}$ is the Z_t-innovation while R_t and W_t are the nonexplosive and explosive components. Define

$$a_t = \varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s W_{s-1}' \left(\sum_{s=1}^{t-1} W_{s-1} W_{s-1}'\right)^{-1} W_{t-1},$$

$$A_t = W_{t-1}' \left(\sum_{s=1}^{t-1} W_{s-1} W_{s-1}'\right)^{-1} W_{t-1},$$

$$b_t = \sum_{s=1}^{t-1} \varepsilon_s \xi_s^{(2)\prime} \left(\sum_{s=1}^{t-1} \xi_s^{(2)} \xi_s^{(2)\prime}\right)^{-1} \xi_t^{(2)}, \qquad B_t = \xi_t^{(2)\prime} \left(\sum_{s=1}^{t-1} \xi_s^{(2)} \xi_s^{(2)\prime}\right)^{-1} \xi_t^{(2)},$$

$$c_t = \sum_{s=1}^{t-1} \varepsilon_s R_{s-1}' \left(\sum_{s=1}^{t-1} R_{s-1} R_{s-1}'\right)^{-1} R_{t-1}, \qquad C_t = R_{t-1}' \left(\sum_{s=1}^{t-1} R_{s-1} R_{s-1}'\right)^{-1} R_{t-1},$$

$$d_t = \sum_{s=1}^{t-1} \varepsilon_s x_s' \left(\sum_{s=1}^{t-1} x_s x_s'\right)^{-1} x_t, \qquad D_t = x_t' \left(\sum_{s=1}^{t-1} x_s x_s'\right)^{-1} x_t$$

$$f_t^2 = 1 + D_t, \qquad \mathcal{I}_{yz} = \sum_{s=1}^{t} \frac{y_s z_s}{f_s^2}, \ y_s, z_s \in (a_s, b_s, c_s, d_s, \varepsilon_s).$$

LEMMA A.1. *Assume A, B, C and that* $\dim \mathbf{W} = 0$. *Then*

$$\sum_{s=1}^{t}\left(\tilde{\varepsilon}_s^2 - \frac{\varepsilon_s^2}{f_s^2}\right) = \left\{\mathcal{I}_{bb} + 2\mathcal{I}_{bc} + \mathcal{I}_{cc} - 2(\mathcal{I}_{\varepsilon b} + \mathcal{I}_{\varepsilon c})\right\}\left\{1 + \mathrm{o}(1)\right\} \qquad a.s.$$

$$f_t^2 = (1 + B_t + C_t)\{1 + \mathrm{o}(1)\} \qquad a.s.$$

Proof of Lemma A.1. Since $\tilde{\varepsilon}_t f_t = \varepsilon_t - d_t$ then

$$\sum_{s=1}^{t}\left(\tilde{\varepsilon}_s^2 - \frac{\varepsilon_s^2}{f_s^2}\right) = \sum_{s=1}^{t}\frac{1}{f_s^2}(d_s^2 - 2\varepsilon_s d_s) = \mathcal{I}_{dd} - 2\mathcal{I}_{\varepsilon d}.$$

The components of x_t are asymptotically uncorrelated due to Nielsen (2005, Thm. 2.4) given Assumptions A, B, C and $\dim \mathbf{W} = 0$. It then holds that $d_t = (b_t + c_t)\{1 + \mathrm{o}(1)\}$ *a.s.* and the first result follows. The second result follows by a similar argument. ■

LEMMA A.2. *Assume Assumption A and that* $\dim \mathbf{W} = 0$. *Then* $\mathcal{I}_{\varepsilon b} = \mathrm{o}(t^{1/2})$ *a.s.*

Proof of Lemma A.2. The term $\mathcal{I}_{\varepsilon b}$ is a $\mathcal{G}_t$-martingale since b_s/f_s^2 is $\mathcal{G}_{s-1}$-measurable. Therefore, by Hall and Heyde (1980, Thm. 2.18), $\mathcal{I}_{\varepsilon b} = \mathrm{o}(t^{1/2})$ a.s. on the set where

$$\mathcal{S} = \sum_{s=1}^{\infty}\mathsf{E}(s^{-1}\varepsilon_s^2 b_s^2/f_s^4|\mathcal{G}_{s-1}) = \sum_{s=1}^{\infty}s^{-1}b_s^2 f_s^{-4}\mathsf{E}(\varepsilon_s^2|\mathcal{G}_{s-1}) < \infty.$$

It suffices to show that $\mathcal{S} = \mathrm{O}(\sum_{s=1}^{\infty}s^{-3/2}\log\log s) = \mathrm{O}(1)$ a.s. Note that $f_s^2 \geq 1$, and $\sup_s \mathsf{E}(\varepsilon_s^2|\mathcal{G}_{s-1}) < \infty$ given Assumption A. Further, $b_s = \mathrm{o}\{(s^{-1/2}\log\log s)^{1/2}\}$ since

$$\sum_{u=1}^{s-1}\varepsilon_u \xi_u^{(2)\prime} = \mathrm{O}\{(s\log\log s)^{1/2}\}, \qquad \left(\sum_{u=1}^{s-1}\xi_u^{(2)}\xi_u^{(2)\prime}\right)^{-1} = \mathrm{O}(s^{-1}), \qquad \xi_s^{(2)} = \mathrm{o}(s^{1/4}), \tag{A.3}$$

a.s. by Nielsen (2005, Thms. 2.4, 5.1, 6.1) assuming A. ■

LEMMA A.3. *Assume A, B, C. Then* $\mathcal{I}_{bb}, \mathcal{I}_{cc}, \mathcal{I}_{bc} = o(t^{1/2})$ *a.s.*

Proof of Lemma A.3. Apply the expansion of f_t^2 in Lemma A.1 as $1 + B_t + C_t$ while ignoring the $\mathrm{o}(1)$ remainder term for notational simplicity.

Consider $\mathcal{I}_{bb}$. The denominator satisfies $f_s^2 \geq 1 + B_s$. Further, $\mathcal{S}_1 = \sum_{s=1}^{t-1}\varepsilon_s\xi_s^{(2)\prime}(\sum_{s=1}^{t-1}\xi_s^{(2)}\xi_s^{(2)\prime})^{-1/2} = \mathrm{O}\{(\log\log t)^{1/2}\}$ by Nielsen (2005, Thm. 2.4). Thus, for almost every outcome and $\epsilon > 0$ then for large t and $s \leq t$, it holds that $\mathcal{S}_1^2 \leq t^{\eta}\epsilon$ for all $\eta > 0$. This implies that for large t

$$\mathcal{I}_{bb} \leq t^{\eta}\epsilon\sum_{s=1}^{t}\left\{\xi_s^{(2)\prime}\left(\sum_{v=1}^{s-1}\xi_v^{(2)}\xi_v^{(2)\prime}\right)^{-1}\xi_s^{(2)}\right\}\Bigg/\left\{1 + \xi_s^{(2)\prime}\left(\sum_{v=1}^{s-1}\xi_v^{(2)}\xi_v^{(2)\prime}\right)^{-1}\xi_s^{(2)}\right\}.$$

Due to the partitioned inversion formula

$$\begin{aligned} A_{12}A_{22}^{-1}A_{21}(1 + A_{12}A_{22}^{-1}A_{21})^{-1} &= 1 - (1,0)\begin{pmatrix} 1 & A_{12} \\ A_{21} & A_{22}\end{pmatrix}^{-1}\begin{pmatrix}0\\1\end{pmatrix} \\ &= A_{12}(A_{22} + A_{21}A_{12})^{-1}A_{21}, \end{aligned} \tag{A.4}$$

it holds that $\mathcal{I}_{bb} \leq t^{\eta}\epsilon \sum_{s=1}^{t} \xi_s^{(2)\prime}(\sum_{v=1}^{s} \xi_v^{(2)}\xi_v^{(2)\prime})^{-1}\xi_s^{(2)}$. The sum is of order $\mathrm{O}(\log t)$ due to Nielsen (2005, Lem. 8.6) assuming A, implying that $\mathcal{I}_{bb}$ is $\mathrm{o}(t^{\eta}) = \mathrm{o}(t^{1/2})$ a.s.

Consider $\mathcal{I}_{cc}$. A similar argument shows $\mathcal{I}_{cc} = \mathrm{o}(t^{1/2})$ a.s. The only slight difference is the bound for $\mathcal{S}_2 = \sum_{s=1}^{t-1} \varepsilon_s R_{s-1}'(\sum_{s=1}^{t-1} R_{s-1}R_{s-1}')^{-1/2}$. By Nielsen (2005, Thm. 2.4), assuming A, B, C, this bound is $\mathcal{S}_2^2 = \mathrm{O}(\log t)$, which is still $\mathrm{o}(t^{\eta})$ for all $\eta > 0$.

Consider $\mathcal{I}_{bc}$. The Hölder inequality implies $\mathcal{I}_{bc} = \mathrm{o}(t^{1/2})$ a.s. ■

A modified version of Lemma 2 of Lai and Wei (1982) is needed.

LEMMA A.4. *Let $h_1, h_2, \ldots$ be p-dimensional vectors and let $H_T = \sum_{t=1}^{T} h_t h_t'$. Assume H_T is nonsingular for some T_0. Let λ_T^* be the maximal eigenvalue of H_T. Then*

(i) $\sum_{t=T_0}^{T} h_t' H_t^{-1} h_t = O(\log \lambda_T^*)$,
(ii) $h_t' H_{t-1}^{-1} h_t = h_t' H_t^{-1} h_t / (1 - h_t' H_t^{-1} h_t)$,
(iii) $\sum_{t=T_0+1}^{T} h_t' H_{t-1}^{-1} h_t = O(\log \lambda_T^*)$.

Proof of Lemma A.4.

(i) This is the statement of Lai and Wei (1982, Lem. 2.ii).
(ii) This follows by (A.4).
(iii) By Lai and Wei (1982, Lem. 2.i) then $h_t' H_t^{-1} h_t = 1 - \det H_t / \det H_{t-1}$. Combine this and (ii) to get $\sum_{t=T_0+1}^{T} h_t' H_{t-1}^{-1} h_t = \sum_{t=T_0+1}^{T} (\det H_t - \det H_{t-1}) / \det H_{t-1}$. Then complete the argument as in the proof of Lai and Wei (1982, Lem. 2.ii). ■

LEMMA A.5. *Assume A, B. Then, $\mathcal{I}_{\varepsilon c} = o(t^{1/2})$ a.s.*

Proof of Lemma A.5. Note that $\mathcal{I}_{\varepsilon c}$ is a $\mathcal{G}_t$-martingale. As in the proof of Lemma A.2 argue that $\sum_{t=1}^{\infty} t^{-1} c_t^2 / f_t^4 < \infty$. Since $f_t \geq 1$ it suffices that $c_t = \mathrm{o}(t^{-\eta})$ for some $\eta > 0$. The similarity transformation M in (5.1) can be chosen so that $\mathbf{R}$ is block diagonal with elements $\mathbf{U}$ and $\mathbf{V}$ with eigenvalues inside and on the complex unit circle, respectively; see Nielsen (2005, Sect. 3). These cases can be studied separately.

If $\mathbf{R} = \mathbf{U}$, apply Nielsen (2005, Thms. 2.4, 5.1, 6.2) to see that $c_t = \mathrm{o}(t^{-1/4})$.

If $\mathbf{R} = \mathbf{V}$ then apply Lemma A.4(ii) in combination with Nielsen (2005, Thms. 2.4, 8.4) to see that $c_t = \mathrm{o}(t^{-\eta})$ for some $\eta > 0$. ■

LEMMA A.6. *Assume A, B, C and that* $\dim \mathbf{W} = 0$. *Then* $\sum_{s=1}^{t} (\varepsilon_s^2 - \varepsilon_s^2 / f_s^2) = \mathrm{o}(t^{1/2})$ *a.s.*

Proof of Lemma A.6. The expression of interest satisfies

$$\sum_{s=1}^{t} \varepsilon_s^2 \left(1 - \frac{1}{f_s^2}\right) = \sum_{s=1}^{t} \frac{\varepsilon_s^2 D_s}{f_s^2} \leq \left(\max_{s \leq t} \varepsilon_s^2\right) \sum_{s=1}^{t} D_s,$$

where the inequality follows since $D_s \geq 0$. By Lemmas A.1, A.4(iii), and Nielsen (2005, Thm. 7.1) then $\sum_{s=1}^{t} D_s = \mathrm{O}(\log t)$ a.s. Moreover, $\varepsilon_t = \mathrm{o}(t^{1/2-\eta})$ *a.s.* for some $\eta > 0$ by Nielsen (2005, Thm. 5.1) assuming A, B, C. ■

LEMMA A.7. *Assume A, B, C. Then $\sum_{s=1}^{t} \{\varepsilon_s^2 - a_s^2 / (1 + A_s)\} = o(t^{1/2})$ a.s.*

Proof of Lemma A.7. Define $K_s = \sum_{u=1}^{s} W_{u-1} W_{u-1}'$, $g_s = \sum_{h=1}^{s} G_{s-h,s}\varepsilon_h$, where

$$G_{s-h,s} = \begin{cases} -W_{s-1}' K_{s-1}^{-1} W_{h-1} (1 + W_{s-1}' K_{s-1}^{-1} W_{s-1})^{-1/2} & \text{for } h < s, \\ (1 + W_{s-1}' K_{s-1}^{-1} W_{s-1})^{-1/2} & \text{for } h = s. \end{cases}$$

With this definition and a change of summation order it holds that

$$\begin{aligned} \sum_{s=1}^{t} g_s^2 &= \sum_{s=1}^{t}\sum_{h=1}^{s} G_{s-h,s}^2 \varepsilon_h^2 + 2\sum_{s=1}^{t}\sum_{h=1}^{s} G_{s-h,s}\varepsilon_h \sum_{\ell=1}^{h-1} G_{s-h+\ell,s}\varepsilon_{h-\ell} \\ &= \sum_{h=1}^{t}\varepsilon_h^2 + \sum_{h=1}^{t}\left\{\left(\sum_{s=h}^{t} G_{s-h,s}^2\right) - 1\right\}\varepsilon_h^2 + 2\sum_{h=1}^{t}\sum_{\ell=1}^{h-1}\left(\sum_{s=h}^{t} G_{s-h,s} G_{s-h+\ell,s}\right)\varepsilon_h\varepsilon_{h-\ell}. \end{aligned}$$

It has to be argued that the sums in s are close to zero. Define

$$Z_h = \mathbf{W}^{1-h} W_{h-1} = W_0 + \sum_{s=1}^{h-1} \mathbf{W}^{-s} e_{W,s},$$

$$F_s = \sum_{u=1}^{s-1} \mathbf{W}^{1-s} W_{u-1} W_{u-1}' (\mathbf{W}')^{1-s} = \sum_{u=1}^{s-1} \mathbf{W}^{u-s} Z_u Z_u' (\mathbf{W}')^{u-s};$$

the coefficients $G_{s-h,s}$ can be rewritten as

$$G_{s-h,s} = \begin{cases} -Z_s' F_s^{-1} \mathbf{W}^{h-s} Z_h \{1 + Z_s' F_s^{-1} Z_s\}^{-1/2} & \text{for } h < s, \\ \{1 + Z_s' F_s^{-1} Z_s\}^{-1/2} & \text{for } h = s. \end{cases}$$

Lai and Wei (1985, Lem. 2, Cor. 2) give the convergence results

$$Z_h \overset{a.s.}{\to} Z = W_0 + \sum_{s=1}^{\infty} \mathbf{W}^{-s} e_{W,s}, \qquad F_h \overset{a.s.}{\to} F = \sum_{u=1}^{\infty} \mathbf{W}^{-u} Z Z' (\mathbf{W}')^{-u}. \tag{A.5}$$

The limiting matrix F is positive definite a.s. under Assumption C, see Lai and Wei (1985, Cor. 2), Nielsen (2008, Rem. 2.3). Thus introduce the coefficients

$$\tilde{G}_{s-h} = \begin{cases} -Z' F^{-1} \mathbf{W}^{h-s} Z (1 + Z' F^{-1} Z)^{-1/2} & \text{for } s > h, \\ (1 + Z' F^{-1} Z)^{-1/2} & \text{for } s = h, \end{cases}$$

and approximate the sums of the coefficients $G_{s-h,s}$ by

$$\sum_{s=h}^{t} G_{s-h,s}^2 \approx \sum_{s-h=0}^{\infty} \tilde{G}_{s-h}^2, \qquad \sum_{s=h}^{t} G_{s-h,s} G_{s-h+\ell,s} \approx \sum_{s-h=0}^{\infty} \tilde{G}_{s-h}\tilde{G}_{s-h+\ell}. \tag{A.6}$$

The approximating sums with $\tilde{G}_{s-h}$ are identical to one and zero, respectively, since

$$\begin{aligned} \sum_{s-h=0}^{\infty} \tilde{G}_{s-h}^2 &= (1 + Z' F^{-1} Z)^{-1} \left\{ 1 + Z' F^{-1} \sum_{s-h=0}^{\infty} \mathbf{W}^{h-s} Z Z' (\mathbf{W}')^{h-s} F^{-1} Z \right\} \\ &= (1 + Z' F^{-1} Z)^{-1} (1 + Z' F^{-1} F F^{-1} Z) = 1, \end{aligned}$$

whereas the sum of cross products satisfies

$$\begin{aligned}
&\sum_{s-h=0}^{\infty} \tilde{G}_{s-h}\tilde{G}_{s-h+\ell}. \\
&\quad = (1+Z'F^{-1}Z)^{-1}\Bigg\{-Z'F^{-1}\mathbf{W}^{\ell}Z + Z'F^{-1} \\
&\qquad\qquad \times \sum_{s-h=0}^{\infty} \mathbf{W}^{h-s}ZZ'(\mathbf{W}')^{h-s}(\mathbf{W}')^{\ell}F^{-1}Z\Bigg\} \\
&\quad = (1+Z'F^{-1}Z)^{-1}\Big\{-Z'F^{-1}\mathbf{W}^{\ell}Z + Z'F^{-1}F(\mathbf{W}')^{\ell}F^{-1}Z\Big\} = 0,
\end{aligned}$$

where the last identity follows since the scalar $Z'F^{-1}\mathbf{W}^{\ell}Z$ is equal to $Z'(\mathbf{W}')^{\ell}F^{-1}Z$.

Two observations are needed to justify the approximation (A.6). First, the tail sums $\sum_{s-h=t+1}^{\infty}\tilde{C}_{s-h}^{2}$ and $\sum_{s-h=t+1}^{\infty}\tilde{G}_{s-h}\tilde{G}_{s-h+\ell}$ vanish exponentially with $\mathbf{W}^{s-h}$. Second, the convergence results in (A.5) also have an exponential rate. This means that if $h > H$ where $H \to \infty$ at $\log T$-rate then the difference $G_{s-h,s} - \tilde{G}_{s-h} = \mathrm{o}(T^{-n})$ for any integer n. Then, apply Lemma 4.2. ∎

Proof of Lemma 5.1. If $\dim \mathbf{W} = 0$ combine Lemmas A.1–A.6. If $\dim \mathbf{R} = 0$ apply Lemma A.7. ∎