Asymptotic behaviour of the CUSUM of squares test under stochastic and deterministic time trends

Bent Nielsen and Jouni Sohkanen

August 31, 2009

Abstract

We undertake a generalization of the cumulative sum of squares (CUSQ) test to the case of non-stationary autoregressive distributed lag models with quite general deterministic time trends. The test may be validly implemented with either ordinary least squares residuals or standardized forecast errors. Simulations suggest that there is little at stake in the choice between the two in the unit root case under Gaussian innovations, and that there is only very modest variation in the finite sample distribution across the parameter space.

1 Introduction

Brown, Durbin, and Evans (1975) suggested looking at a cumulative sum of squared recursive residuals ($CUSQ$). In the context of fixed regressors and normal innovations they could derive finite sample distributional results. Ploberger and Krämer (1986) derived asymptotic results for time series situations with stationary regressors and martingale difference innovations. This was recently generalized to time series regressions with correlated errors by Deng and Perron (2008a). When it comes to trending data Lu, Maekawa, and Lee (2008) have shown how the $CUSQ$-statistics could be applied to differenced data. Here we investigate the behaviour of the $CUSQ$-statistics when applied directly to the levels of the trending data. We show that usual asymptotic distributions also apply in the context of autoregressive distributed lag models with trending regressors, including the possibility of unit roots, explosive roots and deterministic terms. This shows that the CUSQ test and variants thereof can be applied in autoregressive modelling without prejudice to subsequent inferences. This in turn supports the usage of the $CUSQ$ test in explorative analysis, to use the terminology of Dufour (1982).

Brown, Durbin, and Evans (1975) considered the linear regression

$$y_t = \beta' x_t + \varepsilon_t \qquad \text{for } t = 1, \ldots, T, \tag{1.1}$$

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

$$\hat{\beta}_t = (\sum_{s=1}^{t} x_s x_s')^{-1} \sum_{s=1}^{t} x_s y_s \qquad \text{for } t = M, \ldots, T, \tag{1.2}$$

along with the recursive forecast residuals

$$\tilde{\varepsilon}_t = \{1 + x_t'(\sum_{s=1}^{t-1} x_s x_s')^{-1} x_t\}^{-1/2}(y_t - \hat{\beta}_{t-1}' x_t) \qquad \text{for } t > M, \tag{1.3}$$

the cumulative sums of squares plot with recursive residuals is defined as

$$\mathrm{CUSQ}_{t,T}^{REC} = \sqrt{T}\left(\frac{\sum_{s=M}^{t}\tilde{\varepsilon}_s^2}{\sum_{s=M}^{T}\tilde{\varepsilon}_s^2} - \frac{t-M}{T-M}\right) \qquad \text{for } t \geq M. \tag{1.4}$$

Assuming fixed regressors and Gaussian innovations Brown, Durbin, and Evans (1975) derived the finite sample distribution of $\mathrm{CUSQ}_{t,T}^{REC}$.

In passing Brown, Durbin, and Evans (1975) mentioned an alternative statistic, which was analysed by McCabe and Harrison (1980). Computing recursive residual variances

$$\hat{\sigma}_t = t^{-1}\sum_{s=1}^{t}\hat{\varepsilon}_{s,t}^2 \qquad \text{for } M \leq t. \tag{1.5}$$

based on the least squares residuals

$$\hat{\varepsilon}_{s,t} = y_s - \hat{\beta}_t' x_s \qquad \text{for } M \leq t. \tag{1.6}$$

the cumulative sums of squares plot with least squares residuals is defined as

$$\mathrm{CUSQ}_{t,T}^{OLS} = \sqrt{T}\left(\frac{\hat{\sigma}_t}{\hat{\sigma}_T} - \frac{t}{T}\right) = \sqrt{T}\left(\frac{\sum_{s=1}^{t}\hat{\varepsilon}_{s,t}^2}{\sum_{s=1}^{T}\hat{\varepsilon}_{s,T}^2} - \frac{t}{T}\right) \qquad \text{for } t > M. \tag{1.7}$$

The two $CUSQ$-statistics have the same asymptotic distribution in a range of situations: Viewed as processes in t they converge in distribution to a Brownian bridge. Deng and Perron (2008a) prove this for stationary autoregressions. Here, this result is generalized to autoregressions with trending behaviour.

In the presented analysis the regression (1.1) is generalized to an autoregressive distributed lags regression including deterministic terms. The variables involves are assumed to satisfy a vector autoregression which can have stationary roots, unit roots, and explosive roots, while the deterministic terms include constants, linear trends and seasonal dummies. The analysis can then be based on the results of

Lai and Wei (1985) and Nielsen (2005). The asymptotic analysis is easiest for the $CUSQ^{OLS}$-statistic, where it is not necessary to pay much attention to the different types of components of the process involved.

The asymptotic results are broadly the same the $CUSQ^{REC}$-statistic and the $CUSQ^{OLS}$-statistic, although proven in slightly more generality for the latter. Interestingly, a small scale Monte Carlo study indicates that there is not very much difference in terms of finite sample behaviour for the two 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.

A variant of the $CUSQ^{OLS}$-statistic is the recursively computed residual sum of squared innovations used without confidence bands as an graphical exploratory device in for instance the software PcGive, see Hendry (1986), Doornik and Hendry (2007). Confidence bands are derived from the results for the $CUSQ^{OLS}$-statistic.

The paper is organized so that the time series regressions and the model assumptions are presented in §2. The asymptotic results for the $CUSQ^{OLS}$- and $CUSQ^{REC}$-statistics are presented in §3 and §4, respectively. §5 contains a simulation study involving first order autoregressions. The proofs are given in an appendix.

2 Model and assumptions

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

Suppose a p-dimensional time series $X_{1-k}, \dots, X_0, \dots, X_T$ is observed and that X_t is partitioned as $(Y_t, Z_t')'$ 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

$$Y_t = \rho Z_t + \sum_{j=1}^{k} \alpha_j Y_{t-j} + \sum_{j=1}^{k} \beta_j' Z_{t-j} + \nu D_{t-1} + \varepsilon_t, \qquad t = 1, \dots T. \tag{2.1}$$

Here D_t is a deterministic term such as a constant, a linear trend or a seasonal dummy. A frequently used variant of the regression omits the contemporaneous regressor Z_t giving the regression

$$Y_t = \sum_{j=1}^{k} \alpha_j Y_{t-j} + \sum_{j=1}^{k} \beta_j' Z_{t-j} + \nu D_{t-1} + \varepsilon_t, \qquad t = 1, \dots T. \tag{2.2}$$

When the observed time series is univariate so $p = 1$ and $X_t = Y_t$ the regression (2.2) reduces to a univariate autoregression of order k.

In order to characterize the asymptotic distribution of our test statistics, the joint distribution of the time series $X_t = (Y_t, Z_t')'$ needs to be specified. We will assume that X_t satisfies the vector autoregression

$$X_t = \sum_{j=1}^{k} A_j X_{t-j} + \mu D_{t-1} + \xi_t \qquad t = 1, \ldots T, \tag{2.3}$$

$$D_t = \mathbf{D} D_{t-1}. \tag{2.4}$$

where ξ_t is a martingale difference sequence with respect to an increasing sequence of σ-fields $\mathcal{F}_t$. The innovations have to satisfy the following assumption.

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

$$\sup_t \mathsf{E}\{(\xi_t'\xi_t)^{\lambda/2}|\mathcal{F}_{t-1}\} \overset{a.s.}{<} \infty \qquad \textit{for some } \lambda > 4, \tag{2.5}$$

$$\mathsf{E}(\xi_t\xi_t'|\mathcal{F}_{t-1}) \overset{a.s.}{=} \Omega \qquad \textit{where } \Omega \textit{ is positive definite.} \tag{2.6}$$

The formulation for the deterministic term D_t allows a joint autoregressive companion representation of X_t, D_t. The matrix $\mathbf{D}$ has characteristic roots on the complex unit circle, so D_t is a vector of terms such as a constant, a linear trend, or periodic functions like seasonal dummies. For example,

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

will generate a linear trend and a constant, whereas

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

will generate a constant and a dummy for a bi-annual frequency. The deterministic term D_t is assumed to be of polynomial order with linearly independent coordinates.

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 (2.3). This includes stationary roots, roots on the unit circle and a range of explosive roots. The only restriction on the parameter space relates to the explosive roots of the companion matrix

$$\mathbf{B} = \left\{ \begin{array}{cc} (A_1, \ldots, A_{k-1}) & A_k \\ I_{p(k-1)} & 0 \end{array} \right\}. \tag{2.7}$$

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

Assumption C is always satisfied for univariate autoregressions, where $p = 1$, and for vector autoregressions with at most one explosive root. For multivariate autoregressions, Anderson (1959) and Duflo, Senoussi, and Touati (1991) pointed out that this assumption is needed for consistency of the least squares estimators as it ensures positive definiteness of the normalized information matrix associated with the explosive roots; see also Nielsen (2008) for a discussion.

The parameters and innovations of the regressions (2.1) and (2.2) can be linked to the vector autoregression (2.3) through the limits of the least squares estimators arising from (2.1) and (2.2). This also leads to a definition of the innovation terms ε_t appearing in (2.1) and (2.2). For this purpose define

$$\xi_t = \begin{pmatrix} \xi_{y,t} \\ \xi_{z,t} \end{pmatrix}, \qquad \Omega = \begin{pmatrix} \Omega_{yy} & \Omega_{yz} \\ \Omega_{zy} & \Omega_{zz} \end{pmatrix},$$

conformably with $X_t = (Y_t, Z_t')'$. It then holds that for equation (2.1) that

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

where σ^2 is the variance of the innovation ε_t. Similarly, for equation (2.2) it holds

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

For the asymptotic results the above assumptions suffice to establish that the sum of squared residuals is close to the sum of squared innovations. In addition a condition is needed to ensure that the normalized partial sums of squared innovations converge to a Brownian motion. Various conditions could be used here. In line with Assumption A the following assumption suffices.

Assumption D. *For the regression* (2.1) *let* $\mathcal{G}_{t-1}$ *be the* σ*-field over* Z_t *and* $\mathcal{F}_{t-1}$*, while* $\mathcal{G}_t = \mathcal{F}_t$ *for the regression* (2.2)*. Let* $(\varepsilon_t^2 - \sigma^2, \mathcal{G}_t)$ *be a martingale difference sequence satisfying* $\mathsf{Var}(\varepsilon_t^2 - \sigma^2 | \mathcal{G}_{t-1}) = \varphi^2$ *a.s. for some* $\varphi > 0$ *and* $\sup_t \mathsf{E}(|\varepsilon_t|^\lambda | \mathcal{G}_{t-1}) < \infty$ *a.s. for some* $\lambda > 4$.

In the case of independent normally distributed innovations then $\varphi^2 = 2\sigma^4$. For the estimation of φ in non-normal situations one further assumption is needed.

Assumption E. *Let* $(\varepsilon_t^3, \mathcal{G}_t)$ *be a martingale difference sequence, so* $\mathsf{E}(\varepsilon_t^3 | \mathcal{G}_{t-1}) = 0$ *and* $\sup_t \mathsf{E}(|\varepsilon_t|^\lambda | \mathcal{G}_{t-1}) < \infty$ *a.s. for some* $\lambda > 6$.

3 Asymptotic analysis of the $CUSQ^{OLS}$-statistic

We now consider the $CUSQ^{OLS}$-statistic (1.7) based on the autoregressive distributed lags residuals of (2.1) or (2.2). The key to the asymptotic analysis is to generalize

Lemma 2 of Deng and Perron (2008a) showing that the sum of squared residuals is close to the sum of squared innovations. The following Lemma is proved in Appendix §A.1.

Lemma 3.1. *Assume A, B, C. Then* $\sup_{t \leq T} T^{-1/2} |\sum_{s=1}^{t} (\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)| \to 0$ *a.s.*

The normalized partial sums of squared innovation are asymptotically Brownian. This follows through a direct application of Chan and Wei (1988, Theorem 2.2), as stated in the next result.

Lemma 3.2. *Assume D. Let $\mathcal{B}$ be a standard Brownian motion and let $D[0,1]$ denote the space of right-continuous functions on $[0,1]$ with left limits. Then, for $0 \leq u \leq 1$, it holds $T^{-1/2} \sum_{s=1}^{Tu} (\varepsilon_u^2 - \sigma^2) \to \varphi \mathcal{B}_u$ in distribution on $D[0,1]$.*

The above result involves a nuisance parameter φ which needs to be estimated. In the case of normal innovations $\varphi^2 = 2\sigma^4$ so φ can be estimated from the sample variance of the residuals. For non-normal innovations a more natural estimator involves the fourth moment of the residuals. The consistency of such an estimator is given in the next result which is proved in Appendix §A.2.

Theorem 3.3. *A, B, C, D, E. Then* $\hat{\varphi}_t^2 = t^{-1} \sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^4 - (t^{-1} \sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2)^2 \to \varphi^2$ *a.s.*

The main result concerning the $CUSQ^{OLS}$-statistic now follows, with a proof given in Appendix §A.1.

Theorem 3.4. *Assume A, B, C, D. Let $\mathcal{B}^\circ$ be a standard Brownian bridge. Then*
(i) $CUSQ_{int(Tu),T}^{OLS} \to \varphi \mathcal{B}_u^\circ$ *in distribution on $D[0,1]$.*
(ii) $\sup_{t \leq T} |CUSQ_{t,T}^{OLS}| \to \sup_{u \leq 1} |\varphi \mathcal{B}_u^\circ|$ *in distribution on $\mathbb{R}$.*
Note that Theorem 3.3 provides consistent estimators for φ.

An alternative graphical variance constancy diagnostic is considered without confidence bands by Hendry (1986), Doornik and Hendry (2007). The idea is to plot a normalized residual sum of squared residuals, $t^{-1} RSS_t$ where $RSS_t = \sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2$. The asymptotic behaviour of $T^{-1/2} RSS_{\text{int}(Tu)}$ follows readily from Lemmas 3.1 and 3.2. Since the function $u^{-1} 1_{(u \geq \nu)}$ is in $D[0,1]$ for all $\nu > 0$, although not for $\nu = 0$, it can be multiplied with $T^{-1/2} RSS_{\text{int}(Tu)}$ and the asymptotic distribution follows from the continuous mapping theorem.

Corollary 3.5. *Assume A, B, C, D. Let $\mathcal{B}$ be standard Brownian. Then, for $\nu > 0$,* $(Tu)^{-1/2} \{RSS_{int(Tu)} - \sigma^2 Tu\} 1_{(u \geq \nu)} \to 1_{(u \geq \nu)} \varphi u^{-1/2} \mathcal{B}_u$ *in distribution on $D[0,1]$.*

4 Asymptotic behaviour of the $CUSQ^{REC}$-test

We now turn to the asymptotic behaviour of the $CUSQ^{REC}$-statistic (1.4) applied to the regressions (2.1) and (2.2). This statistic is more complicated to describe than the $CUSQ^{OLS}$-statistic. The asymptotic results are not quite as general in that the vector autoregression is assumed to be either purely non-explosive or purely explosive.

In order to formulate a generalization of Deng and Perron (2008a, Lemma 2) it is necessary to decompose the vector autoregression into its non-explosive 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 $\mathbf{B}$ defined in (2.7) the vector autoregression satisfies a first order vector autoregression $S_t = \mathbf{B}S_{t-1} + \iota\xi_t$. As noted in for instance Nielsen (2005, §3), there exists a real matrix M so $M\mathbf{B}M^{-1}$ is block diagonal and

$$MS_t = \begin{pmatrix} R_t \\ W_t \end{pmatrix} = \begin{pmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{W} \end{pmatrix} \begin{pmatrix} R_{t-1} \\ W_{t-1} \end{pmatrix} + \begin{pmatrix} e_{R,t} \\ e_{W,t} \end{pmatrix}, \tag{4.1}$$

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

Lemma 2 of Deng and Perron (2008a) is now generalized, albeit to a lesser extent than seen for the situation of OLS residuals in Lemma 3.1. The issue is that cross terms between explosive and non-explosive terms are not easy to deal with as explained in Remark A.2 in the Appendix §A.3 and are therefore ruled out. Even with that restriction the proof of the following Lemma given in Appendix §A.3 is more involving than that of Lemma 3.1 and requires the additional Assumption D.

Lemma 4.1. *Assume A, B, C, D. Assume the process is* ***either*** *purely non-explosive* ***or*** *purely explosive. Then* $\sup_{t\le T} T^{-1/2}|\sum_{s=1}^{t}(\tilde\varepsilon_s^2 - \varepsilon_s^2)| \to 0$ *a.s.*

A limiting result for the $CUSQ^{RES}$ then follows by exactly the same argument as that of Theorem 3.4.

Theorem 4.2. *Assume A, B, C, D. Assume the process is* ***either*** *purely non-explosive* ***or*** *purely explosive. Let* $\mathcal{B}^\circ$ *be a standard Brownian bridge. Then*
(i) $CUSQ^{REC}_{int(Tu),T} \to \varphi\mathcal{B}^\circ_u$ *in distribution on* $D[0,1]$.
(ii) $\sup_{t\le T}|CUSQ^{REC}_{t,T}| \to \sup_{u\le 1}|\varphi\mathcal{B}^\circ_u|$ *in distribution on* $\mathbb{R}$.
Note that Theorem 3.3 provides consistent estimators for φ.

Table 1: Statistics for which there is no significant variation accross a range of of values of autoregressive parameter α

	S^{OLS}	S^{REC}
standard deviation	0.256	0.256
95% quantile	1.27	1.28
p-value of asymptotic 95% quantile	0.032	0.033
MCSE for above p-value	0.0002	0.0002

5 Simulation study

Theorems 3.4 and 4.2 show that the two types of $CUSQ$-statistics have the same limit distribution in many situations. For the $CUSQ^{OLS}$-statistic, in particular, this does not depend on the autoregressive parameters apart from the regularity assumption C for the explosive roots. This leaves the questions whether the 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 the studies by Deng and Perron (2008a), Deng and Perron (2008b).

The data generating process is a univariate first order autoregression, $X_t = \alpha X_{t-1} + \epsilon_t$ for $t = 1, \ldots, T = 100$ with initial value $X_0 = 0$, innovation variance of unity and a range for the autoregressive parameters α. The statistics of interest are the supremum statistics $S^{OLS} = \max_{M \leq t \leq T} |CUSQ_{t,T}^{OLS}|/\hat{\varphi}_T$ and $S^{REC} = \max_{M \leq t \leq T} |CUSQ_{t,T}^{REC}|/\hat{\varphi}_T$, where $\hat{\varphi}_T^2 = 2\hat{\sigma}_T^4$, see (1.5). The theorems 3.4 and 4.2 show that these statistics converge in distribution to 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, Table 9).

The results for the simulation study are reported in Tables 1, 2. The variation of the distribution for the two supremum statistics is very small and could hardly be picked up with 10^6 repetitions. Table 1 reports statistics like standard deviation, 95%-quantile and p-value of the asymptotic 95%-quantile for which there was no significant variation for different values of α, whereas Table 2 reports mean and median for which one can just about discern a slight variation in α.

Two conclusions emerge from this small scale Monte Carlo study. First, there is not much difference in finite sample distribution for the two statistics. Secondly, 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.

Table 2: Statistics for which there is a slight significant variation accross a range of of values of autoregressive parameter α

	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

6 Example: United States GDP 1947-2006

To illustrate the results log quarterly, seasonally adjusted GDP data for the US for 1947:1 to 2006:1 are considered. The data originate from the Bureau of Economic Analysis; see also Hendry and Nielsen (2007). In the figure, panel (a) show the time series in levels. A fourth order autoregression with a constant and a linear trend was fitted recursively. Panel (b) shows cumulative sums of the model residuals with point-wise confidence bands. Panel (c) shows the recursive residual variance estimator with point-wise confidence bands. Panel (d) shows the $CUSQ^{OLS}$-statistic in panel (d), with simultaneous confidence bands. All bands are draw for the 5% significance level. We note that there is evidence against constancy of the residual variance. This finding is consistent with that of McConnell and Perez-Quiros (2000), who apply the CUSUM of Squares test to levels of demeaned post-war US GDP data. Our finding is supported by our theory that validates the use of the CUSUM of Squares test in this context. Furthermore, it is based on residuals from a regression model, which is often preferred to the approach using demeaned data as pointed out by Deng and Perron (2008b).

Acknowledgements

The first author received financial support from ESRC grant RES-000-27-0179.

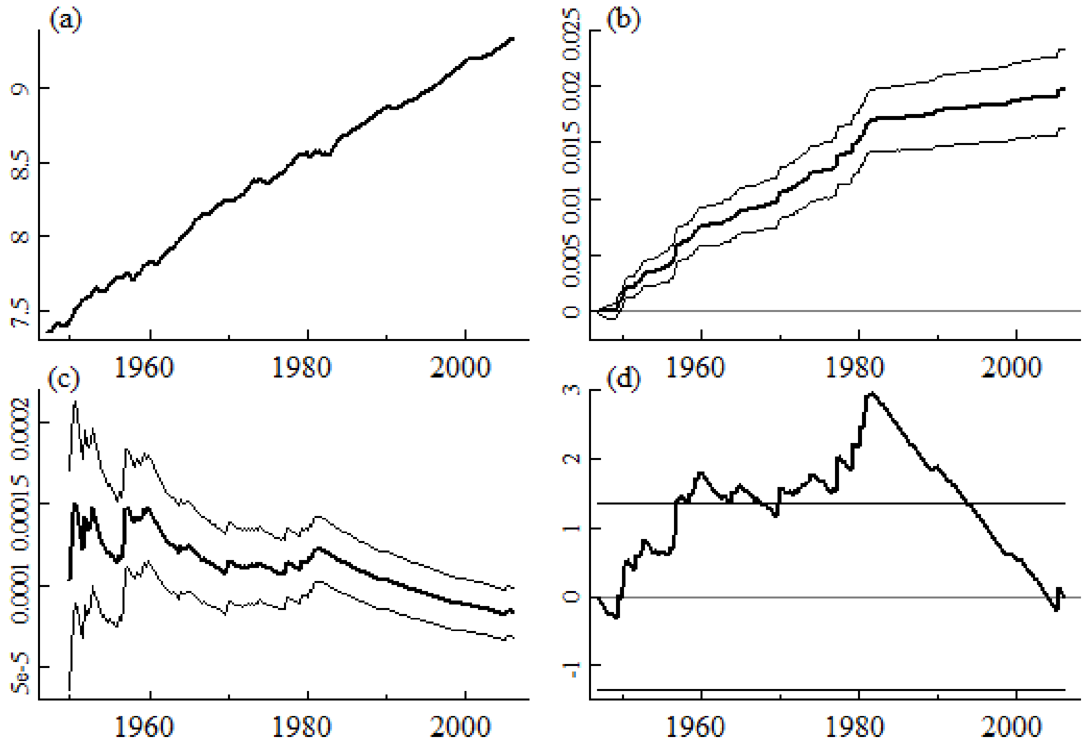

Figure 1: US log GDP data and recursive statistics.

Appendix: Proofs

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

A.1 The case of least squares residuals

Proof of Lemma 3.1. Partition ξ_t as $(\xi_t^{(1)}, \xi_t^{(2)})'$ and partition the least squares residuals, $\hat{\xi}_{s,t}$, of X_t on $X_{t-1}, \ldots X_{t-k}$ 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 (2.2) 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, Corollary 2.6) to see that $t^{-1}\sum_{s=1}^{t}(\hat{\xi}_{s,t}^{\otimes 2} - \xi_s^2) = \mathrm{o}(t^{-1/2})$ *a.s.*, assuming A, B, C. The result (A.1) then follows.

Secondly, if Z_t is included as regressor as in (2.1) then

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

By Nielsen (2005, Corollary 2.6) then

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

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

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

Using that $(1, -\rho)\Omega(0, I)' = 0$ along with Nielsen (2005, Theorem 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.

Now, (A.1) implies that for almost every outcome and for any $\epsilon > 0$, there exists a finite t_0 such that

$$\frac{1}{t}\sum_{s=1}^{t}(\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2) < \frac{\epsilon}{\sqrt{t}} \qquad \forall\ t > t_0.$$

Moreover, since $t \leq T$, then $T^{-1/2}t \leq t^{1/2}$ for all $t > t_0$, which implies

$$\frac{1}{\sqrt{T}}\sum_{s=1}^{t}(\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2) < \epsilon \quad \forall\ t > t_0 \tag{A.2}$$

Since t_0 is finite, we also have $\max_{t \leq t_0}\{t^{-1}\sum_{s=1}^{t}(\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)\}$ is finite, whereas $T^{-1/2}t$ vanishes for $t < t_0$ and $T \to \infty$. In combination we have that for T sufficiently large

$$\frac{1}{\sqrt{T}} \sum_{s=1}^{t} (\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2) < \epsilon. \tag{A.3}$$

Combining (A.2) and (A.3) gives the desired result. □

Proof of Theorem 3.4. (i) Note that

$$\sum_{s=1}^{\text{int}(Tu)} (\hat{\varepsilon}_{s,\text{int}(Tu)}^2 - \sigma^2) = \sum_{s=1}^{\text{int}(Tu)} (\hat{\varepsilon}_{s,\text{int}(Tu)}^2 - \varepsilon_s^2) + \sum_{s=1}^{\text{int}(Tu)} (\varepsilon_s^2 - \sigma^2).$$

The Lemmas 3.1 and 3.2 imply $T^{-1/2}\sum_{s=1}^{\text{int}(Tu)}(\hat{\varepsilon}_{s,\text{int}(Tu)}^2 - \sigma^2) \to \varphi \mathcal{B}_u$ in distribution.

Next, rewrite the $CUSQ$-statistic as

$$\begin{aligned} CUSQ_{\text{int}(Tu),T}^{OLS} &= T^{1/2}\left(\frac{\sum_{s=1}^{\text{int}(Tu)} \hat{\varepsilon}_{s,\text{int}(Tu)}^2}{\sum_{s=1}^{T} \hat{\varepsilon}_{s,T}^2} - \frac{t}{T}\right) \\ &= \frac{T^{-1/2}\{\sum_{s=1}^{\text{int}(Tu)}(\hat{\varepsilon}_{s,\text{int}(Tu)}^2 - \sigma^2) - \frac{t}{T}\sum_{s=1}^{T}(\hat{\varepsilon}_{s,T}^2 - \sigma^2)\}}{T^{-1}\sum_{s=1}^{T} \hat{\varepsilon}_{s,T}^2}. \end{aligned}$$

Then insert the above convergence result for the partial sums.

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

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

Proof of Theorem 3.3. The result is proved for the regression (2.1) included Z_t as regressor. The argument for the regression (2.2) can be deduced in a similar way.

Due to Lemma 3.1 the second moments of the residuals and of the innovations have the same limit. If the same is shown for the fourth moments $t^{-1}\sum_{s=1}^{t}\hat{\varepsilon}_{s,t}^4$ then the desired result follows from a Law of Large Numbers applied to the squares of the martingale differences $\varepsilon_s = (1, -\rho)\xi_s$ and $\varepsilon_s^2 - \sigma^2$, assuming A, D.

Recall the companion vector $S_{t-1} = (X_{t-1}', \ldots, X_{t-k}', D_{t-1})'$. Since the residuals from regressing Z_t on S_t equal those of regressing $\xi_t^{(2)}$ on S_t the regressor is then $x_t = (\xi_t^{(2)\prime}, S_{t-1}')'$. Define

$$P_t = \sum_{s=1}^{t} \varepsilon_s x_s' (\sum_{s=1}^{t} x_s^{\otimes 2})^{-1/2}, \qquad Q_{s,t} = (\sum_{s=1}^{t} x_s^{\otimes 2})^{-1/2} x_s.$$

The residuals satisfy $\hat{\varepsilon}_{s,t} = \varepsilon_t - P_t Q_{s,t}$. A binomial expansion of $\hat{\varepsilon}_{s,t}^4$ shows that it suffices to prove that $\mathcal{I}_m = \sum_{s=1}^{t}(P_t Q_{s,t})^m \varepsilon_s^{4-m} = \text{o}(t)$ $a.s.$ for $m = 1, \ldots, 4$.

First, argue that $P_t = \mathrm{o}(t^{1/4})$ *a.s.* The components of $x_s = (\xi_s^{(2)\prime}, S_{s-1}')'$ are asymptotically uncorrelated due to Nielsen (2005, Theorem 2.4) assuming A, B, C. Thus

$$P_t \overset{a.s.}{=} \{\sum_{s=1}^{t} \varepsilon_s \xi_s^{(2)\prime} (\sum_{s=1}^{t} \xi_s^{(2)\otimes 2})^{-1/2} + \sum_{s=1}^{t} \varepsilon_s S_{s-1}' (\sum_{s=1}^{t} S_{s-1}^{\otimes 2})^{-1/2}\}\{1 + \mathrm{o}(1)\}.$$

This is of the desired order due to Nielsen (2005, Theorems 2.4, 2.8, Corollary 2.6) assuming A, B, C, and the construction $(1, -\rho)\Omega(0, I)' = 0$.

Secondly, consider $\mathcal{I}_1 = P_t(\sum_{s=1}^{t} x_s^{\otimes 2})^{-1/2} \sum_{s=1}^{t} x_s \varepsilon_t^3$. Recall the decomposition $MS_s = (R_s', W_s')'$ in (4.1). The components are asymptotically uncorrelated due to Nielsen (2005, Theorems 9.1, 9.2) assuming A, B, C. Thus

$$\mathcal{I}_1 \overset{a.s.}{=} P_t\{\sum_{s=1}^{t} \begin{pmatrix} \xi_s^{(2)\otimes 2} & 0 & 0 \\ 0 & R_{s-1}^{\otimes 2} & 0 \\ 0 & 0 & W_{s-1}^{\otimes 2} \end{pmatrix}\}^{-1/2} \sum_{s=1}^{t} \begin{pmatrix} \xi_s^{(2)} \\ R_{s-1} \\ W_{s-1} \end{pmatrix} \varepsilon_t^3 \{1 + \mathrm{o}(1)\}$$

The term P_t was dealt with above.

The terms involving $\xi_s^{(2)}$ are $\mathrm{O}\{(\log\log t)^{1/2}\}$ due to Nielsen (2005, Theorem 2.4) assuming E.

The terms involving R_s are $\mathrm{O}\{(\log t)^{1/2}\}$ due to Lai and Wei (1982, Lemma 1), Nielsen (2005, Theorems 7.1) assuming A, B, E.

The terms involving W_s can be bounded by

$$\| \sum_{s=1}^{t} (\mathbf{W}^{-t} W_{s-1})^{\otimes 2} \|^{-1/2} (\sum_{s=1}^{t} \|\mathbf{W}^{-t} W_{s-1}\|) \max_{s \le t} \|\varepsilon_s\|^3$$

The first two terms are convergent due to Nielsen (2005, Corollaries 5.3, 7.2) assuming A, C. The latter term is $\mathrm{o}(t^{3/4})$ since $\varepsilon_t = \mathrm{o}(t^{1/4})$ by Nielsen (2005, Theorem 5.1) assuming A.

Secondly, consider $\mathcal{I}_m$ for $m \ge 2$. The following bound holds

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

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

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

so the last term is bounded. □

A.3 The case of recursive residuals

Lemma 4.1 is proved in three steps. As in the proof of Lemma 3.3 only the regression (2.1) including Z_t as a regressor is considered, whereas the argument for the regression (2.2) is slightly simpler. The regressor is written as $x_t = (\xi_t^{(2)\prime}, R_{t-1}', W_{t-1}')'$, where $\xi_t^{(2)}$ is the innovation term for the contemporaneous regressor Z_t while R_t and W_t are the non-explosive and explosive components. The residuals $\tilde{\varepsilon}_t$ will be decomposed in a similar way. Thus, define:

$$
\begin{aligned}
a_t &= \varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s W_{s-1}' (\sum_{s=1}^{t-1} W_{s-1}^{\otimes 2})^{-1} W_{t-1}, & A_t &= W_{t-1}' (\sum_{s=1}^{t-1} W_{s-1}^{\otimes 2})^{-1} W_{t-1}, \\
b_t &= \sum_{s=1}^{t-1} \varepsilon_s \xi_s^{(2)\prime} (\sum_{s=1}^{t-1} \xi_s^{(2)\otimes 2})^{-1} \xi_t^{(2)}, & B_t &= \xi_t^{(2)\prime} (\sum_{s=1}^{t-1} \xi_s^{(2)\otimes 2})^{-1} \xi_t^{(2)}, \\
c_t &= \sum_{s=1}^{t-1} \varepsilon_s R_{s-1}' (\sum_{s=1}^{t-1} R_{s-1}^{\otimes 2})^{-1} R_{t-1}, & C_t &= R_{t-1}' (\sum_{s=1}^{t-1} R_{s-1}^{\otimes 2})^{-1} R_{t-1}, \\
f_t^2 &= 1 + A_t + B_t + C_t. & \mathcal{I}_{aa} &= \sum_{s=1}^{t} \frac{a_s^2}{f_s^2}
\end{aligned}
$$

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

Lemma A.1. *Assume A, B, C. Then* $\sum_{s=1}^t \tilde{\varepsilon}_s^2 = \mathcal{I}_{aa} + \mathrm{o}(t^{1/2})$ *a.s.*

Proof of Lemma A.1. *Purely explosive case*: Result trivial.
Purely non-explosive case: Note $a_t = \varepsilon_t$, $A_t = 0$, and the regressor decomposes as $x_t = (\xi_t^{(2)\prime}, R_{t-1}')'$.

The components of x_t are asymptotically uncorrelated due to Nielsen (2005, Theorem 2.4) assuming A, B, C. The recursive forecast residual (1.3) then satisfies $\tilde{\varepsilon}_t = \{(a_t - b_t - c_t)/f_t\}\{1 + \mathrm{o}(1)\}$ *a.s.* It has to be argued that terms involving b_s or c_s are $\mathrm{o}(t^{1/2})$.

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

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

Due to the partitioned inversion formula

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

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

Consider $\mathcal{I}_{cc} = \sum_{s=1}^{t} c_s^2/f_s^2$. A similar argument shows $\mathcal{I}_{cc} = \mathrm{o}(t^{1/2})$ *a.s.* The only slight difference is the bound for $\mathcal{S}_2 = \sum_{s=1}^{t-1} \varepsilon_s R_{s-1}^{\prime}(\sum_{s=1}^{t-1} R_{s-1}^{\otimes 2})^{-1/2}$. By Nielsen (2005, Theorem 2.4), assuming A, B, C, this bound is $\mathcal{S}_2^2 = \mathrm{O}(\log t)$, which is of course still $\mathrm{o}(t^{\eta})$ for all $\eta > 0$.

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

Consider $\mathcal{I}_{ab} = \sum_{s=1}^{t} a_s b_s/f_s^2$. Since $a_s = \varepsilon_s$ and b_s/f_s^2 is $\mathcal{G}_{t-1}$-measurable then $\mathcal{I}_{ab}$ is a martingale. Applying Hall and Heyde (1980, Theorem 2.18) shows $t^{-1/2}\mathcal{I}_{ab}$ vanishes on the set where $\sum_{s=1}^{\infty} \mathsf{E}(s^{-1}\varepsilon_s^2 b_s^2/f_s^4|\mathcal{G}_{s-1}) < \infty$. Assumption D shows that $\mathsf{E}(\varepsilon_s^2|\mathcal{G}_{t-1}) = \sigma^2$, so it suffices to consider the set where $\sum_{s=1}^{\infty} s^{-1}b_s^2/f_s^4$ is finite. Since $\sum_{u=1}^{s-1} \varepsilon_u \xi_u^{(2)\prime}(\sum_{u=1}^{s-1} \xi_u^{(2)\otimes 2})^{-1/2} = \mathrm{O}\{(\log\log s)^{1/2}\}$, $\xi_s^{(2)} = \mathrm{o}(s^{1/4})$, and $(\sum_{u=1}^{s-1} \xi_u^{(2)\otimes 2})^{-1/2} = \mathrm{O}(s^{-1})$ by Nielsen (2005, Theorems 2.4, 5.1, 6.1) while $1 \leq f_s^2$ then this set has probability one.

Consider $\mathcal{I}_{ac} = \sum_{s=1}^{t} a_s c_s/f_s^2$. The term c_s is based on the component R_s which has both stationary and unit roots. Thus, decompose R_s into stationary and unit root components. These are uncorrelated by Nielsen (2005, Theorem 9.4), so can be treated separately. The stationary case matches the analysis for $\mathcal{I}_{ab}$. Thus, assume R_s only has unit roots. A martingale argument is also made here with the difference that $\sum_{u=1}^{s-1} \varepsilon_u R_{u-1}^{\prime}(\sum_{u=1}^{s-1} R_{u-1}^{\otimes 2})^{-1/2} = \mathrm{O}\{(\log s)^{1/2}\}$ and $R_{s-1}^{\prime}(\sum_{u=1}^{s-1} R_{u-1}^{\otimes 2})^{-1}R_{s-1} = \mathrm{o}(s^{-\eta})$ for some $\eta > 0$ by Nielsen (2005, Theorems 2.4, 8.4). □

Remark A.2. *The difficulty in considering the case with both explosive and non-explosive terms arises in connection with the cross terms $\mathcal{I}_{ab}$, $\mathcal{I}_{ac}$. In general $a_s \neq \varepsilon_s$. Hence these terms are not martingales.*

Lemma A.3. *Let $h_1, h_2, \ldots$ be p-dimensional vectors and let $H_T = \sum_{t=1}^{T} h_t^{\otimes 2}$. Assume H_T is non-singular for some T_0. Let λ_T^* be the maximal eigenvalue of H_T Then*
(i) $\sum_{t=T_0}^{T} h_t^{\prime} H_t^{-1} h_t = \mathrm{O}(\log \lambda_T^*)$.
(ii) $\sum_{t=T_0+1}^{T} h_t^{\prime} H_{t-1}^{-1} h_t = \mathrm{O}(\log \lambda_T^*)$.

Proof of Lemma A.3. (*i*) is the statement of Lai and Wei (1982, Lemma 2,*ii*).

(*ii*) is proved in a similar way. By (A.4) then $h_t^{\prime} H_{t-1}^{-1} h_t = h_t^{\prime} H_t^{-1} h_t/(1 - h_t^{\prime} H_t^{-1} h_t)$, whereas by Lai and Wei (1982, Lemma 2(i)) then $h_t^{\prime} H_t^{-1} h_t = 1 - \det H_t/\det H_{t-1}$. In combination this shows $\sum_{t=T_0+1}^{T} h_t^{\prime} H_{t-1}^{-1} h_t = \sum_{t=T_0+1}^{T} (\det H_t - \det H_{t-1})/\det H_{t-1}$. Then complete the argument as in the proof of Lai and Wei (1982, Lemma 2,*ii*). □

Lemma A.4. *Assume A, B, C. Then* $\sum_{s=1}^{t}\{a_s^2/f_s^2 - a_s^2/(1 + A_s)\} = \mathrm{o}(t^{1/2})$ *a.s.*

Proof of Lemma A.4. *Purely explosive case*: Result trivial.
Purely non-explosive case: The expression of interest, $\mathcal{D}$ say, satisfies

$$\mathcal{D} = \sum_{s=1}^{t} \frac{a_s^2}{1+A_s}\left(1 - \frac{1+A_s}{f_s^2}\right) = \sum_{s=1}^{t} \frac{a_s^2(B_s+C_s)}{(1+A_s)f_s} \leq (\max_{s\leq t} a_s^2) \sum_{s=1}^{t}(B_s+C_s),$$

where the inequalty follows since $1 \leq f_s$, $1 \leq 1+A_s$, $0 \leq B_s$, and $0 \leq C_s$. Lemma A.3 together with Nielsen (2005, Theorem 7.1) shows $\sum_{s=1}^{t}(B_s+C_s) = \mathrm{O}(\log t)$ *a.s.* Moreover, $a_t = \mathrm{o}(t^{1/4-\eta})$ *a.s.* for all $\eta > 0$ since ε_t and $\sum_{s=1}^{t-1}\varepsilon_s W_{s-1}'(\sum_{s=1}^{t-1} W_{s-1}^{\otimes 2})^{-1/2}$ are $\mathrm{o}(T^{1/4-\eta})$ by Nielsen (2005, Theorems 2.4, 5.1) while $(\sum_{s=1}^{t-1} W_{s-1}^{\otimes 2})^{-1/2} W_{t-1}$ is convergent by Nielsen (2005, Corrolaries 5.3, 7.2) assuming A, B, C. □

Lemma A.5. *Assume A, B, C. Then* $\sum_{s=1}^{t}\{\varepsilon_s^2 - a_2^2/(1+A_s)\} = \mathrm{o}(t^{1/2})$ *a.s.*

Proof of Lemma A.5. *Purely non-explosive case*: Result trivial.
Purely explosive case: Define $a_s = \sum_{h=1}^{s} C_{s-h,s}\varepsilon_h$, where

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

With this definition and a change of summation order it holds

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

It has to be argued that the sums in s of the coefficients $C_{s-h,s}$ are close to zero.

Defining the quantities

$$\begin{aligned} Z_h &= \mathbf{W}^{1-h} W_{h-1} = W_0 + \sum_{s=1}^{h-1} \mathbf{W}^{-s} e_{W,s} \\ F_s &= \sum_{u=1}^{s-1} (\mathbf{W}^{1-s} W_{u-1})^{\otimes 2} = \sum_{u=1}^{s-1} (\mathbf{W}^{u-s} Z_u)^{\otimes 2} \end{aligned}$$

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

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

Lai and Wei (1985, Lemma 2, Corollary 2) give the convergence results

$$Z_h \overset{a.s.}{\to} Z = W_0 + \sum_{s=1}^{\infty} \mathbf{W}^{-s} e_{W,s}, \qquad F_h \overset{a.s.}{\to} F = \sum_{u=1}^{\infty} (\mathbf{W}^{-u} Z)^{\otimes 2}. \tag{A.5}$$

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

$$\tilde{C}_{s-h} = \begin{cases} -Z'F^{-1}\mathbf{W}^{h-s}Z/(1+Z'F^{-1}Z) & \text{for } s-h>0, \\ 1/(1+Z'F^{-1}Z) & \text{for } s-h=0. \end{cases}$$

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

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

The approximating sums involving $\tilde{C}_{s-h}$-coefficients are identical to one and zero, respectively, since:

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

whereas the sum of cross products satisfies

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

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 justify the approximation (A.6). First, the tail sums $\sum_{s-h=t+1}^{\infty} \tilde{C}_{s-h}^2$ and $\sum_{s-h=t+1}^{\infty} \tilde{C}_{s-h}\tilde{C}_{s-h+\ell}$ vanish exponentially with $\mathbf{W}^{s-h}$. Secondly, the convergence results in (A.5) also have an exponential rate. This means that if $h > H$ where $H \to \infty$ at at $\log T$-rate then the difference $C_{s-h,s} - \tilde{C}_{s-h} = \mathrm{o}(T^{-n})$ for any integer n. These observations can then be applied in argument as that of the last paragraphs of the proof of Lemma 3.1. □

Proof of Lemma 4.1. Combining this with Lemmas A.1, A.4, A.5 shows that $\sum_{s=1}^{t}(\tilde{\varepsilon_s}^2 - \varepsilon_s^2) = \mathrm{o}(t^{1/2})$. The argument is then finished as in the last paragraph of the proof of Lemma 3.1. □

References

Anderson, T. W. (1959): "On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations," *The Annals of Mathematical Statistics*, 30, 676–687.

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

Brown, R., J. Durbin, and 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., and C. Z. Wei (1988): "Limiting distributions of least squares estimates of unstable autoregressive processes," *The Annals of Statistics*, 16, 367–401.

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

——— (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.

Doornik, J. A., and D. F. Hendry (2007): *Empirical econometric modelling using PcGive 12, vol. 1 and 2.* London: Timberlake Consultants Press.

Duflo, M., R. Senoussi, and 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.

Dufour, J.-M. (1982): "Recursive stability analysis of regression relationships: an exploratory methotology," *Journal of Econometrics*, 19, 31–76.

Hall, P., and C. Heyde (1980): *Martingale limit theory and its application Probability and mathematical statistics (New York, NY).* Academic Press.

Hendry, D., and B. Nielsen (2007): *Econometric modeling.* Princeton: Princeton University Press.

Hendry, D. F. (1986): *PC-GIVE 4.0.* Mimeo, Nuffield College, University of Oxford.

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

——— (1985): "Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems," *Multivariate Analysis VI*, pp. 375–393.

Lu, X., K. Maekawa, and S. Lee (2008): "The CUSUM of squares test for the stability of regression models with non-stationary regressors," *Economics Letters*, 100, 234–237.

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.

McConnell, M. M., and G. Perez-Quiros (2000): "Output fluctuations in the United States: what has changed since the early 1980's?," *American Economic Review*, 90, 1464–1476.

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

——— (2008): "Singular vector autoregressions with deterministic terms: Strong consistency and lag order determination," Nuffield college working paper.

Ploberger, W., and 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.