A Joint Chow Test for Structural Instability

Bent Nielsen and Andrew Whitby
Department of Economics, University of Oxford

This version: June 25, 2012

Abstract

The classical Chow (1960) test for structural instability requires strictly exogenous regressors and a break-point specified in advance. In this paper we consider two generalisations, the 1-step recursive Chow test (based on the sequence of studentized recursive residuals) and its supremum counterpart, which relax these requirements. We use results on strong consistency of regression estimators to show that the 1-step test is appropriate for stationary, unit root or explosive processes modelled in the autoregressive distributed lags (ADL) framework. We then use results in extreme value theory to develop a new supremum version of the test, suitable for formal testing of structural instability with an unknown break-point. The test assumes normality of errors, and is intended to be used in situations where this can either be assumed or established empirically.

1 Introduction

Identifying structural instability in models is of major concern to econometric practitioners. The Chow (1960) tests are perhaps the most widely used for this purpose, but require strictly exogenous regressors and a break-point specified in advance. As such, a plethora of variants have been developed to meet different requirements. In this paper we consider two generalisations: the 1-step recursive Chow test, based on the sequence of studentized recursive forecast residuals; and its supremum counterpart. The pointwise test is frequently used and reported in applied work, while the supremum test is new. Whereas Chow assumes a classical regression framework, practitioners typically use the one-step test to evaluate dynamic models (e.g. Kimura, 2001; Celasun and Goswami, 2002; Assarsson et al., 2004). Further, since a series of such tests is usually presented graphically to the modeller, multiple testing issues arise, making it difficult to determine how many point failures may be tolerated. These two issues motivate the analysis that follows. First, in Theorem 4.1 we show that the pointwise statistic has the correct asymptotic distribution under fairly general assumptions about the generating process, including lagged dependent variables and deterministic terms. Second, we take advantage of the almost sure convergence earlier proven to construct a supremum version of the 1-step test, applicable to detecting parameter change or at outlier at an unknown point in the sample.

The pointwise 1-step Chow test is essentially the 'prediction interval' test described by Chow, but computed recursively, and over the sample (rather than at an a priori hypothesised change point). It first appears in PcGive version 4.0 (Hendry, 1986) as part of a suite of model misspecification diagnostics. The idea of using residuals calculated recursively to test model misspecification dates from the landmark CUSUM and CUSUMSQ tests (Brown and Durbin,

We thank the participants of the 22nd NZESG Meeting for comments. Bent Nielsen gratefully acknowledges financial support from the Institute for Economic Modelling at the Oxford Martin School. Numerical computations used Ox Professional 6.10 (Doornik, 2007). Correspondence e-mail: andrew.whitby@economics.ox.ac.uk

1968; Brown et al., 1975), which are based on partial sums of (squared) recursive residuals, and have since been generalised to models including lagged dependent variables (Ploberger and Krämer, 1986; Krämer et al., 1988; Nielsen and Sohkanen, 2011). Unlike these tests, the 1-step Chow test does not consider partial sums, but the sequence of recursive residuals itself; in effect testing one-step-ahead forecast failure at each time step. As the following analysis shows, this approach leads to a different type of asymptotics, with a residual sequence behaving like i.i.d. random variables, rather than a partial sum of residuals behaving like a Brownian motion.

Examining the residual sequence to check model specification is, of course, well established. As Brown et al. (1975) put it, 'it is natural to look at residuals to investigate departures from model specification', although this has generally meant the OLS residuals. Other authors (e.g. Galpin and Hawkins, 1984) have suggested plotting the recursive residuals, but in a different manner. The recursive residuals have two advantages over the OLS residuals in many applications: first, under the normal linear model with fixed regressors, they are identically and independently normal; second (and distinguishing them from other i.i.d. normal transformed residuals, e.g. Theil's (1965) BLUS residuals), they have a natural interpretation—in a time series setting—as forecast errors. Ironically, in typical time series settings where the forecast error interpretation is most useful, independence of the residuals does not hold due to the presence of lagged dependent variables, a problem noted by Dufour (1982). This may lead to difficulties drawing firm conclusions from plotted pointwise test sequences, and thus motivates the second part of this paper, which considers a supremum test.

The supremum test considers the maximum of the pointwise 1-step tests, appropriately normalised. It is intended to reflect structural instability anywhere in the sample (with the early part excluded to allow consistent estimation). It relates to work on parameter change at an unknown time, and more particularly with work on tests for outliers at an unknown time. Examples of the former include the already mentioned CUSUM and CUSUMSQ tests, and the Quandt (1960) and Andrews (1993) supremum tests. The test is distinguished from Andrews' test in not imposing any restrictions on the end-of-sample, so that end-of-sample instability may be detected. Additionally, because the 1-step tests behave like an i.i.d. process, the asymptotics differ from these cases, requiring the application of extreme value theory of independent and weakly dependent sequences, rather than the suprema of random-walks.

Seen specifically as an outlier test, the supremum Chow test falls squarely within the tradition of Srikantan (1961), which, however, considers an unknown outlier in a classical setting. Even recently, the majority of work on outliers has taken place outside the time series settings, so for instance, Barnett and Lewis (1994, p. 330) comment that '[recursive residuals] would seem to have potential for the study of outliers, although no major progress on this front is evident. There is a major difficulty in that the labeling of the observations is usually done at random, or in relation to some concomitant variable…'. This difficulty does not exist with time series, where there is a natural chronological labelling of observations. The section in the same book (at p. 396) on detecting outliers in time series is, nevertheless, notably brief, and recursive methods are not considered.

2 The test statistics

The 1-step test applies generally to a linear regression

$$y_t = \beta' x_t + \varepsilon_t \qquad t = 1, \dots, T, \tag{2.1}$$

with y_t scalar, x_t a k-dimensional vector of regressors, and the errors independently, identically Gaussian. For such a regression we can define the sequence of least squares estimators calculated over progressively larger subsamples, along with the corresponding residual sums of squares and recursive residual (or standardized 1-step forecast error), that is

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

$$RSS_t = \sum_{s=1}^{t} (\hat{\beta}_t' x_t - y_t)^2 \qquad t = k, \ldots, T, \tag{2.3}$$

$$\tilde{\varepsilon}_t = \left[1 + x_t' \left(\sum_{s=1}^{t-1} x_s x_s'\right)^{-1} x_t\right]^{-1/2} (y_t - \hat{\beta}_{t-1}' x_t) \qquad t = (k+1), \ldots, T. \tag{2.4}$$

The 1-step Chow test statistic, $\mathsf{C}^2_{1,t}$ is then defined as

$$\mathsf{C}^2_{1,t} = \frac{(RSS_t - RSS_{t-1})(t-k-1)}{RSS_{t-1}} \qquad t = (k+1), \ldots, T, \tag{2.5}$$

and can be expressed as

$$\mathsf{C}^2_{1,t} = \frac{\tilde{\varepsilon}_t^2 (t-k-1)}{RSS_{t-1}}. \tag{2.6}$$

Chow showed that in a classical Gaussian regression model, this statistic would have an exact $\mathsf{F}(1, t-k-1)$ distribution. We first extend this result to show that, for a general class of autoregressive distributed lag (ADL) processes, $\mathsf{C}^2_{1,t} \xrightarrow{\text{d}} \chi^2_{(1)}$, so that asymptotically, the additional dependence does not matter. This result means that comparing the pointwise statistic against an $\mathsf{F}(1, \cdot)$ or $\chi^2_{(1)}$ distribution (as is typically done) is reasonable in large samples. However it still leaves unresolved the difficulty that this test is generally reported graphically, to detect parameter change with an unknown changepoint. To formally treat the problem of multiple testing that occurs in evaluating many pointwise statistics over the entire sample, we introduce a new supremum test.

3 Model and assumptions

We consider the behaviour of the test statistic for ADL models with arbitrary deterministic terms, a class which includes by restriction many commonly posited economic relationships (see Hendry (1995, Chapter 7)). For the purpose of analysis we assume the true data generating model can be represented as a vector autoregression.

We observe a p-dimensional time series $X_{1-k}, \ldots, X_0, X_1, \ldots X_T$. We will model the series by partitioning X_t as $(Y_t, Z_t)'$ where Y_t is univariate and Z_t is of dimension $p-1$, and then consider the regression of Y_t on the contemporaneous Z_t, lags of both Y_t and Z_t, and a deterministic term D_t. That is,

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

In order to specify the joint distribution of $X_t = (Y_t, Z_t)'$, we assume that X_t follows the vector autoregression

$$X_t = \sum_{j=1}^{k} A_j X_{t-j} + \mu D_{t-1} + \xi_t \qquad t = 1, \ldots, T, \tag{3.2}$$

with the deterministic term D_t given by

$$D_t = \mathbf{D} D_{t-1}. \tag{3.3}$$

We assume the VAR innovations form a martingale difference sequence satisfying the assumption below. The requirement that the innovations have finite moments just beyond 16 stems from a problem with controlling unit root processes (see Nielsen, 2005, Remark 9.3). In the present analysis this constraint emerges in Lemma A.1(i) and is transmitted via Lemma A.2(iv) to Lemma A.5. If $\dim \mathbf{D} = 0$ and the geometric multiplicity of the unit roots equals their algebraic multiplicity (including $I(1)$ but excluding $I(2)$ processes), this could be improved to finite moments greater than 4 using a result of Bauer (2009).

Assumption 3.1. *ξ_t is a martingale difference sequence with respect to the natural filtration $\mathcal{F}_t$, so $\mathsf{E}(\xi_t|\mathcal{F}_{t-1}) = 0$. The initial values $X_0, \dots X_{1-k}$ are $\mathcal{F}_0$-measurable and*

$$\sup_t \mathsf{E}(\|\xi_t\|^\alpha|\mathcal{F}_{t-1}) \overset{a.s.}{<} \infty \qquad \text{for some } \alpha > 16, \tag{3.4}$$

$$\mathsf{E}(\xi_t'\xi_t|\mathcal{F}_{t-1}) \overset{a.s.}{=} \Omega \qquad \text{where } \Omega \text{ is positive definite.} \tag{3.5}$$

The deterministic term D_t follows the approach of Johansen (2000) and Nielsen (2005) and may include, for example, a constant, a linear trend, or periodic functions such as seasonal dummies. The matrix $\mathbf{D}$ has characteristic roots on the unit circle. For example,

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

will generate a constant and three quarterly dummies. The term D_t is assumed to have linearly independent coordinates, formalised as follows.

Assumption 3.2. $|\text{eigen}(\mathbf{D})| = 1$ *and* $\text{rank}(D_1, \dots, D_{\dim \mathbf{D}}) = \dim \mathbf{D}$.

We permit nearly all possible values of the autoregressive parameters A_j in (3.2), excluding only the case of singular explosive roots, which can only arise for a VAR with $p \geq 2$ and multiple explosive roots. See Nielsen (2008) for discussion. Defining the companion matrix

$$\mathbf{B} = \begin{pmatrix} (A_1, \dots, A_{k-1}) & A_k \\ \mathbf{I}_{p(k-1)} & 0 \end{pmatrix},$$

we can express the restriction as follows.

Assumption 3.3. *The explosive roots of $\mathbf{B}$ have geometric multiplicity unity. That is, for all complex λ with $|\lambda| > 1$, $\text{rank}(\mathbf{B} - \lambda \mathbf{I}_{pk}) \geq pk - 1$.*

Additionally, we require that the innovations in the ADL regression equation satisfy a further martingale assumption.

Assumption 3.4. *Let $\mathcal{G}_t$ be the sigma field over $\mathcal{F}_t$ and Z_t. Then $(\varepsilon_t, \mathcal{G}_t)$ is a martingale difference sequence, i.e. $\mathsf{E}(\varepsilon_t|\mathcal{G}_{t-1}) = 0$.*

Finally, the 1-step statistic is such that a distributional assumption must be made in order to derive the limiting distribution of the statistic (since the statistic is an estimate of a single error term, we cannot take advantage of a central limit theorem). Similarly, since the analysis of supremum statistic will rely on extreme value theory, we must impose distributional and independence assumptions on the ADL errors ε_t, in order to uniquely determine the norming sequences applied in Lemma 4.4. We assume normality, which may result from joint normality in the underlying VAR process, and is tested, in practice, under the above assumptions (see Engler and Nielsen, 2009).

Assumption 3.5. $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$.

4 Main results

We must briefly examine the decomposition of the process used in the proofs in order to elucidate the first main result in the explosive case (in the non-explosive case this decomposition becomes trivial). Group the regressors by defining $S'_{t-1} = (Y_{t-1}, Z'_{t-1} \dots, Y_{t-k}, Z'_{t-k}, D'_{t-1})$, and then write (3.2) in companion form, so that

$$S_t = \mathbf{S}S_{t-1} + (\xi'_t, 0')'.$$

Then there exists a regular real matrix $\mathbf{M}$ to block diagonalize $\mathbf{S}$ (see the elaboration in §3 of Nielsen, 2005), so that the process can be decomposed into stationary, unit-root and explosive components:

$$\mathbf{M}S_t = (\mathbf{M}\mathbf{S}\mathbf{M}^{-1})\mathbf{M}S_{t-1} + \mathbf{M}(\xi'_t, 0')',$$

$$\begin{pmatrix} \tilde{U}_t \\ Q_t \\ W_t \end{pmatrix} = \begin{pmatrix} \mathbf{U} & 0 & 0 \\ 0 & \mathbf{Q} & 0 \\ 0 & 0 & \mathbf{W} \end{pmatrix} \begin{pmatrix} \tilde{U}_{t-1} \\ Q_{t-1} \\ W_{t-1} \end{pmatrix} + \begin{pmatrix} e_{\tilde{U},t} \\ e_{Q,t} \\ e_{W,t} \end{pmatrix}, \tag{4.1}$$

where $\tilde{\mathbf{U}}$, $\mathbf{Q}$ and $\mathbf{W}$ have eigenvalues inside, on and outside the unit circle, respectively. For convenience, we group the non-explosive components, so that

$$R_t = \begin{pmatrix} \tilde{U}_t \\ Q_t \end{pmatrix} \qquad \text{and} \qquad \mathbf{R} = \begin{pmatrix} \mathbf{U} & 0 \\ 0 & \mathbf{Q} \end{pmatrix}. \tag{4.2}$$

The first theorem states that the test statistic is almost surely close to a related process in the innovations, q_t^2, under multiple assumptions. This result, paired with a distributional assumption such as 3.5, is sufficient to establish confidence intervals for a single application of the Chow test. It also forms the basis of the supremum test developed below.

Theorem 4.1. *Under Assumptions 3.1, 3.2, 3.3 and 3.4,*

$$\mathsf{C}_{1,t}^2 - (q_t/\sigma)^2 \overset{\text{as}}{\to} 0 \quad \text{as } t \to \infty \tag{4.3}$$

where

$$q_t = \frac{\varepsilon_t - \sum_{s=1}^{\infty} \varepsilon_{t-s} W'(\mathbf{W}^{-s})' F_W^{-1} W}{(1 + W' F_W^{-1} W)^{1/2}} \tag{4.4}$$

$\mathbf{W}$ *is as in (4.1), and as in Nielsen (2005, Corollaries 5.3 and 7.2),* $W = W_0 + \sum_{t=1}^{\infty} \mathbf{W}^{-t} e_{W,t}$ *and* $F_W = \sum_{t=1}^{\infty} \mathbf{W}^{-t} W W' (\mathbf{W}^{-t})'$ *with* F_W *almost surely positive definite.*

Having established pointwise convergence almost surely, we use an argument based on Egorov's Theorem to establish convergence of the supremum of a subsequence. Both the subsequence itself and the lead-in period must grow without bound, to allow the regression estimates to converge.

Lemma 4.2. *Suppose* $\mathsf{C}^2_{1,t} - (q_t/\sigma)^2 \overset{\text{as}}{\to} 0$ *as* $t \to \infty$*. Then*

$$\sup_{g(T)<t\leq T} \left|\mathsf{C}^2_{1,t} - (q_t/\sigma)^2\right| \overset{\text{p}}{\to} 0 \qquad \text{as } T, g(T) \to \infty. \tag{4.5}$$

where $g(T)$ *is an arbitrary function of* T *such that* $g(T) \to \infty$*.*

Now, if an appropriately normalised expression in the maximum over q_t can be shown to converge in distribution, then so will the supremum statistic, with the same normalisation, by asymptotic equivalence. We show that, under the assumption of independent and identical Gaussian innovations, $\max_{1\leq s\leq t} q_s$ does indeed converge to the Gumbel extremal distribution (as $t \to \infty$), which has distribution function:

$$\Pr(\Lambda < x) = \exp[-\exp(-x)] \qquad \text{where } x \in \mathbb{R}. \tag{4.6}$$

A useful property of the Gumbel distribution is the following simple monotonically decreasing transformation to a χ^2 variable, allowing standard distributions to be used:

$$\Lambda \sim \text{Gumbel} \qquad \text{iff} \qquad 2e^{-\Lambda} \sim \chi^2_{(2)}. \tag{4.7}$$

In showing the above convergence we rely on Theorem 1 of Deo (1972), and its corollary, showing that the extremal distribution of the absolute values of a Gaussian sequence is the same in the stationary dependent and independent cases. However, Deo's Lemma 1 gives an incorrect statement of the norming sequences. The incorrect sequences are also quoted without correction in Pakshirajan and Hebbar (1977). Here we state the correct sequences, adopting the notation of Deo (proof in section A.5).

Lemma 4.3. *Let* $\{X_n\}$ *be independent Gaussian random variables with mean zero and variance one. Let* $Z_n = \max_{1\leq j\leq n} |X_j|$*. Then* $a_n(Z_n - b_n)$ *converges in distribution to* Λ *where* $a_n = 2\log n$ *and* $b_n = (2\log n)^{1/2} - (8\log n)^{-1/2}(\log\log n + \log\pi)$*.*

The original gives $b_n = (2\log n)^{1/2} - (8\log n)^{-1/2}(\log\log n + 4\pi - 4)$.

Deo's result can then be applied to q_t defined in (4.4).

Lemma 4.4. *Under assumption 3.5,*

$$q_t/\sigma \sim N(0,1), \text{ and} \tag{4.8}$$

$$a_t(\max_{1\leq s\leq t} q_s^2 - b_t) \overset{\text{d}}{\to} \Lambda \tag{4.9}$$

where

$$a_t = 1/2 \qquad \text{and} \qquad b_t = \log t^2 - \log\log t - \log\pi \tag{4.10}$$

and Λ *is a random variable distributed according to the Gumbel (Type 1) law.*

Combining these lemmas gives our main result, that with independent and identically Gaussian innovations, an appropriate normalisation of the supremum 1-step Chow test converges in distribution to the Gumbel extremal distribution.

Theorem 4.5. *Under assumptions 3.1, 3.2, 3.3 and 3.5, and with some* $g(T) \to \infty$*,*

$$\mathsf{SC}^2_T = \frac{1}{2}\left(\max_{g(T)<t\leq T} \mathsf{C}^2_{1,t} - d_{T,g(T)}\right) \overset{d}{\to} \Lambda \qquad \text{as} \qquad T \to \infty \tag{4.11}$$

where $\mathsf{C}^2_{1,t}$ *is the 1-step Chow statistic defined in (2.5) and*

$$d_{T,g(T)} = 2\left\{\log[T - g(T)] - \frac{1}{2}\log\log[T - g(T)] - \log\pi\right\} \tag{4.12}$$

and Λ *is a random variable distributed according to the Gumbel distribution (4.6).*

As a simple corollary, we can transform the test using (4.7) so that it may be compared against a more readily-available distribution.

Corollary 4.6. *Under the same assumptions,* $\mathsf{expSC} = 2 \cdot \exp(-\mathsf{SC}_T^2) \sim \chi^2_{(2)}$. *The test based on this result rejects for small values of* expSC.

5 Simulation study

We present the results of two simulations, the first demonstrating similarity of the test, and the second illustrating how the test may be used in conjunction with a test for normality.

In practice we find in simulations that the test as specified above is over-sized in small samples. To minimise this, we suggest two corrections. For the first correction, we observe that the 1-step statistics appear to be distributed close to $\mathsf{F}(1, t-k-1)$ (as indeed they are exactly in the classical case), and so use the following transform to bring the statistics closer to the asymptotic chi-squared distribution:

$$\mathsf{C}^{2*}_{1,t} = G^{-1}[F(\mathsf{C}^2_{1,t})] \tag{5.1}$$

where $F(\cdot)$ and $G(\cdot)$ are the $\mathsf{F}(1, t-k-1)$ and $\chi^2_{(1)}$ distribution functions, respectively. This first correction results in a test that tends to under-correct, largely a result of relatively slow convergence to the limiting Gumbel distribution. In practice we find that for samples of less than 1000, the test performs better if simply compared with the finite maximal distribution assuming independence and identical distribution of the test statistics (the first assumption holding only in the limit and in the absence of an explosive component, and the second holding only in the limit). Then the maximum, $\max_{g(T)<t\le T} \mathsf{C}^{2*}_{1,t}$, would be distributed exactly as

$$\Pr\left\{\max_{g(T)<t\le T} \mathsf{C}^{2*}_{1,t} \le x\right\} = [G(x)]^{T-g(T)}. \tag{5.2}$$

This forms the basis of the finite adjusted sup-Chow test, with rejection in the right tail. Note that in this case no centring or scaling is required, because the distribution itself depends on T.

In the first experiment, an AR(1) process was simulated with the autoregressive parameter varying in the stationary, unit-root and explosive regions.

$$\begin{aligned} x_t &= \alpha x_{t-1} + \varepsilon_t, \quad \epsilon_t \sim \mathsf{N}(0,1), \quad t = 1, \ldots, T \\ x_0 &= 0 \end{aligned}$$

The finite adjusted sup-Chow was calculated as in (5.2), with $g(T) = T^{1/2}$ and nominal size of 5%. Results are presented in Table 1, and show that the size of the tests does not vary according to the autoregressive parameter. As a consequence it is not necessary to know *a priori* where the autoregressive parameter lies to use this test, avoiding a potential circularity in model construction. The test is uniformly undersized, however for a misspecification test (used to reject potential models) this seems preferable to the oversize of the uncorrected asymptotic form. Further, since the test is approximately similar, it should be possible to apply very simple finite sample corrections to eliminate this size discrepancy.

The second experiment uses a similar data generating process and testing procedure as the first, but in addition to applying the finite adjusted sup-Chow test, the E_p test for normality (Doornik and Hansen, 2008) is applied, and the size of the sup-Chow is calculated conditional on not rejecting normality at the 5% level. This simulates the process a model builder may follow, in using the sup-Chow test as part of a suite of misspecification tests including a test

	Autoregressive coefficient						
T	**-1.03**	**-1.00**	**-0.50**	**0.00**	**0.50**	**1.00**	**1.03**
50	2.62	2.61	2.54	2.54	2.50	2.61	2.61
100	3.47	3.49	3.37	3.39	3.42	3.52	3.55
200	3.99	4.00	3.94	3.94	3.93	4.00	4.11

Table 1: Simulated rejection frequency for finite adjusted sup-Chow under a Gaussian AR(1) process, with nominal size 5%. Number of MC repetitions = 200,000 (all MC standard errors are less than 0.05).

		Error distribution					
	T	Φ	$\chi^2_{(1,\text{centred})}$	t_2	t_5	t_{10}	t_{50}
(a) Rejection rate	**50**	2.45	49.04	49.62	19.56	9.10	3.38
	100	3.37	74.82	77.05	37.70	17.55	5.13
	200	3.89	92.72	93.87	60.06	28.70	7.13
(b) Normality acceptance rate	**50**	85.37	2.37	15.62	57.10	74.32	83.99
	100	85.16	0.01	1.85	36.13	64.94	82.81
	200	85.75	0.00	0.02	13.96	50.84	81.92
(c) Rejection rate given normality acceptance	**50**	1.6	*6.3	8.5	6.1	4.0	2.1
	100	2.2		*8.0	8.2	5.9	2.8
	200	2.6			8.8	7.6	4.0

Table 2: (a) Simulated rejection frequency for adjusted sup-Chow under AR(1) processes with various innovation distributions and nominal size 5%. (b) Simulated non-rejection frequency for normality test. (c) Simulated rejection frequency for adjusted sup-Chow given non-rejection by normality test. Number of MC repetitions = 50,000 (all MC standard errors are less than half the second-least significant digit, except those starred).

for normality. In addition, the distribution of the AR(1) innovations is varied to evaluate the sensitivity to the test, and the conditional test, to non-normality. These results are presented in Table 2

As the table illustrates, the unconditional test is quite sensitive to departures from normality, but used conditional upon non-rejection of a normality test, the size is closer to the nominal size of 5%.

A Proofs

A.1 Notation

Define for any $a_s, b_s \in \{x_s, R_{s-1}, W_{s-1}, \xi_{2,s}, Q_{s-1}, \tilde{U}_{s-1}\}$, the sum $S_{ab} = \sum_{s=1}^{t-1} a_s b_s'$, the correlation $C_{ab} = S_{aa}^{-1/2} S_{ab} S_{bb}^{-1/2}$, and the partial regressions quantities $(a|b)_t = a_t - S_{ab} S_{bb}^{-1} b_t$ and $S_{aa.b} = S_{aa} - S_{ab} S_{bb}^{-1} S_{ba}$.

A.2 Preliminary Asymptotic Results

The ADL model (3.1) becomes

$$Y_t = \rho Z_t + \theta' S_{t-1} + \varepsilon_t \qquad t = 1, \ldots, T.$$

where θ is the vector of coefficients. Then from (3.2) we have $Z_t = \Pi S_{t-1} + \xi_{2,t}$, where ξ_t has been partitioned conformably with X_t. Then, the residuals from regressing Y_t on $(Z_t', S_{t-1}')'$ could also be obtained by regressing Y_t on $(\xi_{2,t}', S_{t-1}')'$, or as result of the decomposition above at (4.1), on $x_t = (\xi_{2,t}', R_{t-1}', W_{t-1}')'$—so we can analyse the test statistic (2.6) as if these were the actual regressors.

Many results refer to Nielsen (2005), hereafter abbreviated N05.

Lemma A.1. *Suppose Assumptions 3.1, 3.2 and 3.3 hold with $\alpha > 4$ only. Then for all $\beta > 1/\alpha$ and $\zeta < 1/8$,*

(i) $C_{RW} \overset{\text{a.s.}}{=} \text{o}(t^{-\zeta/2})$,

(ii) $C_{\xi S} \overset{\text{a.s.}}{=} \text{o}(t^{\beta-1/2})$,

(iii) $S_{RR\cdot W}^{-1} \overset{\text{a.s.}}{=} S_{RR}^{-1/2} \cdot \{1 + \text{o}(1)\} \cdot S_{RR}^{-1/2}$,

(iv) $S_{\xi\xi\cdot S}^{-1} \overset{\text{a.s.}}{=} S_{\xi\xi}^{-1/2} \cdot \{1 + \text{o}(1)\} \cdot S_{\xi\xi}^{-1/2}$,

(v) $S_{RR}^{-1/2} R_{t-1} \overset{\text{a.s.}}{=} \text{o}(t^{-\zeta/2})$,

(vi) $S_{WW}^{-1/2} W_{t-1} \overset{\text{a.s.}}{=} \text{O}(1)$,

(vii) $S_{RR}^{-1/2} (R|W)_t \overset{\text{a.s.}}{=} \text{o}(t^{-\zeta/2})$, *and*

(viii) $S_{\xi\xi}^{-1/2} (\xi_2|S)_t \overset{\text{a.s.}}{=} \text{o}(t^{\beta-1/2})$.

Proof. Result (i) is proven by decomposing the correlation to apply results from N05, so that

$$\begin{aligned}
\|C_{RW}\| &= \|S_{RR}^{-1/2} S_{RW} S_{WW}^{-1/2}\| \\
&\leq \left\| \begin{pmatrix} 1 & C_{\tilde{U}Q} \\ C_{Q\tilde{U}} & 1 \end{pmatrix}^{-1/2} \right\| \left\| \begin{pmatrix} S_{\tilde{U}\tilde{U}}^{-1/2} & 0 \\ 0 & S_{QQ}^{-1/2} \end{pmatrix} \begin{pmatrix} S_{\tilde{U}W} \\ S_{QW} \end{pmatrix} S_{WW}^{-1/2} \right\| \\
&\overset{\text{a.s.}}{=} \text{O}(1) \begin{pmatrix} C_{\tilde{U}W} \\ C_{QW} \end{pmatrix}
\end{aligned}$$

where the last line follows because with $C_{\tilde{U}Q}$ is vanishing almost surely by N05 Theorem 9.4. Then the result follows since $C_{\tilde{U}W} \overset{\text{a.s.}}{=} \mathrm{o}(t^{\beta-1/2})$ and $C_{QW} \overset{\text{a.s.}}{=} \mathrm{o}(t^{-\zeta/2})$ by N05 Theorems 9.1 and 9.2 respectively. The latter term will dominate since $\alpha > 16/7$ under Assumption 3.1.

Result (ii) is proved by noting that $\|C_{\xi S}\| \leq \|S_{\xi\xi}^{-1/2}\| \|S_{\xi S} S_{SS}^{-1/2}\|$, with the first normed term $\mathrm{O}(t^{-1/2})$ by N05 Theorem 2.8 and the second $\mathrm{o}(t^{\beta})$ by N05 Theorem 2.4.

Result (iii) follows by writing

$$\begin{aligned} S_{RR.W}^{-1} &= (S_{RR} - S_{RW} S_{WW}^{-1} S_{WR})^{-1} \\ &= S_{RR}^{-1/2} (I - C_{RW} C_{WR})^{-1} S_{RR}^{-1/2} \end{aligned}$$

and applying (i) to show that C_{RW} is vanishing.

Result (iv) is exactly analagous but substitute (ii) for (i).

Result (v) follows by again decomposing R_t. Namely,

$$S_{RR} = \begin{pmatrix} S_{\tilde{U}\tilde{U}}^{1/2} & 0 \\ 0 & S_{QQ}^{1/2} \end{pmatrix} \begin{pmatrix} 1 & C_{\tilde{U}Q} \\ C_{Q\tilde{U}} & 1 \end{pmatrix} \begin{pmatrix} S_{\tilde{U}\tilde{U}}^{1/2} & 0 \\ 0 & S_{QQ}^{1/2} \end{pmatrix}$$

so that

$$\|S_{RR}^{-1/2} R_{t-1}\| \leq \left\| \begin{pmatrix} 1 & C_{\tilde{U}Q} \\ C_{Q\tilde{U}} & 1 \end{pmatrix}^{-1/2} \right\| \left\| \begin{pmatrix} S_{\tilde{U}\tilde{U}}^{-1/2} & 0 \\ 0 & S_{QQ}^{-1/2} \end{pmatrix} \begin{pmatrix} \tilde{U}_{t-1} \\ Q_{t-1} \end{pmatrix} \right\|$$

Then the first normed quantity on the right hand side is bounded since $C_{\tilde{U}Q}$ is vanishing by N05 Theorem 9.4. The second normed quantity comprises $S_{\tilde{U}\tilde{U}}^{-1/2} \tilde{U}_{t-1}$ stacked with $S_{QQ}^{-1/2} Q_{t-1}$. By N05 Theorem 8.3 we have $S_{\tilde{U}\tilde{U}}^{-1/2} = \mathrm{O}(t^{-1/2})$ and by Lai and Wei (1985, Theorem 1(i)) we have that $\tilde{U}_{t-1} = \mathrm{o}(t^{\beta})$, so $S_{\tilde{U}\tilde{U}}^{-1/2} \tilde{U}_{t-1} = \mathrm{o}(t^{\beta-1/2})$.

We cannot bound $S_{QQ}^{-1/2}$ independently in the same way, but since Q_t contains only the unit-root components (with eigenvalues on the unit circle), we can apply N05 Theorem 8.4, which states that for some η, $\max_{t^{\eta} \leq s < t} Q_s' \left(\sum_{s=1}^{t} Q_{s-1} Q_{s-1}'\right)^{-1} Q_s = \mathrm{o}(t^{-\zeta})$ for all $\zeta < 1/8$ and so *a fortiori* $Q_{t-1}' \left(\sum_{s=1}^{t} Q_{s-1} Q_{s-1}'\right)^{-1} Q_{t-1} = \mathrm{o}(t^{-\zeta})$. But then $\|S_{QQ}^{-1/2} Q_{t-1}\|^2 = Q_{t-1}' S_{QQ}^{-1} Q_{t-1}$, and we can then use the matrix identity $b'(\mathbf{A} + bb')^{-1} b = b' \mathbf{A}^{-1} b (1 + b' \mathbf{A}^{-1} b)^{-1}$ (Searle, 1982, p. 151) to write:

$$Q_{t-1}' S_{QQ}^{-1} Q_{t-1} = \frac{Q_{t-1}' \left(\sum_{s=1}^{t} Q_{s-1} Q_{s-1}'\right)^{-1} Q_{t-1}}{1 - Q_{t-1}' \left(\sum_{s=1}^{t} Q_{s-1} Q_{s-1}'\right)^{-1} Q_{t-1}}$$

which is $\mathrm{o}(t^{-\zeta})$, so that $S_{QQ}^{-1/2} Q_{t-1} = \mathrm{o}(t^{-\zeta/2})$.

Considering the maximum of these components, we have again that the latter dominates and $\|S_{RR}^{-1/2} R_{t-1}\| = \mathrm{O}(t^{-\zeta/2})$ since $\alpha > 16/7$ under Assumption 3.1.

Result (vi) follows directly from Lai and Wei (1985, Lemma 4(i)).

Result (vii) follows from (i), (v) and (vi). Write

$$\begin{aligned} \|S_{RR}^{-1/2} (R|W)_t\| &= \|S_{RR}^{-1/2} R_{t-1} - S_{RR}^{-1/2} S_{RW} S_{WW}^{-1} W_{t-1}\| \\ &\leq \|S_{RR}^{-1/2} R_{t-1}\| + \|C_{RW}\| \|S_{WW}^{-1/2} W_{t-1}\| \end{aligned}$$

giving three normed quantities to bound. The first is $\mathrm{o}(t^{-\zeta/2})$ by (v), as is the second by (i), while the third is bounded by (vi).

Result (viii) is proved in a similar fashion. Write

$$\begin{aligned}\|S_{\xi\xi}^{-1/2}(\xi_2|S)_t\| &= \|S_{\xi\xi}^{-1/2}\xi_{2,t} - S_{\xi\xi}^{-1/2}S_{\xi S}S_{SS}^{-1}S_{t-1}\| \\ &= \|S_{\xi\xi}^{-1/2}\xi_{2,t}\| - \|C_{\xi S}\|\|S_{SS}^{-1/2}S_{t-1}\|\end{aligned}$$

Then the first of the normed quantities is $\mathrm{o}(t^{\beta-1/2})$ by N05 Theorem 2.8 and the result that $\xi_t = \mathrm{o}(t^\beta)$ (Lai and Wei, 1985, Theorem 1); the second is $\mathrm{O}(t^{\beta-1/2})$ by (ii); and the third is $\mathrm{O}(1)$ since we use a partial regression transformation to write

$$\begin{aligned}\|S_{SS}^{-1/2}S_{t-1}\|^2 &= S_{t-1}'S_{SS}^{-1}S_{t-1} \\ &= (R|W)_t'S_{RR\cdot W}^{-1}(R|W)_t + W_{t-1}'S_{WW}^{-1}W_{t-1},\end{aligned}$$

and then apply (iii) and (vii), and (vi), respectively. □

Lemma A.2. *Under Assumptions 3.1, 3.2 and 3.3 with $\alpha > 4$; and with $\beta > 1/\alpha$,*

(i) $\sum_{s=1}^{t-1}\varepsilon_s S_{s-1}'S_{SS}^{-1/2} \overset{\text{a.s.}}{=} \mathrm{o}(t^\beta)$,

(ii) $\sum_{s=1}^{t-1}\varepsilon_s R_{s-1}'S_{RR}^{-1/2} \overset{\text{a.s.}}{=} \mathrm{O}[(\log t)^{1/2}]$,

(iii) $\sum_{s=1}^{t-1}\varepsilon_s W_{s-1}'S_{WW}^{-1/2} \overset{\text{a.s.}}{=} \mathrm{o}(t^\beta)$,

(iv) $\sum_{s=1}^{t-1}\varepsilon_s (R|W)_s'S_{RR}^{-1/2} \overset{\text{a.s.}}{=} \mathrm{o}(t^{\beta-1/16})$*, which vanishes if $\alpha > 16$.*

Proof. Results (i), (ii) and (iii) by N05 Theorem 2.4.

Result (iv) follows by writing

$$\sum_{s=1}^{t-1}\varepsilon_s(R|W)_s'S_{RR}^{-1/2} = \sum_{s=1}^{t-1}\varepsilon_s R_{s-1}'S_{RR}^{-1/2} - \sum_{s=1}^{t-1}\varepsilon_s W_{s-1}'S_{WW}^{-1/2}C_{WR}$$

and then applying (ii), (iii) and Lemma A.1(i). □

Lemma A.3. *Under Assumptions 3.1, 3.2, 3.3 and 3.4,*

(i) $\sum_{s=1}^{t-1}\varepsilon_s\xi_{2,s}'S_{\xi\xi}^{-1/2} \overset{\text{a.s.}}{=} \mathrm{O}[(\log t)^{1/2}]$,

(ii) $\sum_{s=1}^{t-1}\varepsilon_s(\xi_2|S)_s'S_{\xi\xi}^{-1/2} \overset{\text{a.s.}}{=} \mathrm{o}(t^{2\beta-1/2}) + \mathrm{O}[(\log t)^{1/2}]$*, the latter term dominating when $\alpha > 4$.*

Proof. Result (i) by Lai and Wei (1982, Lemma 1(iii)) and Lai and Wei (1985, Corollary 1(iii)).

Result (ii) follows by writing

$$\sum_{s=1}^{t-1}\varepsilon_s(\xi_2|S)_s'S_{\xi\xi}^{-1/2} = \sum_{s=1}^{t-1}\varepsilon_s\xi_{2,s}'S_{\xi\xi}^{-1/2} - \sum_{s=1}^{t-1}\varepsilon_s S_{s-1}'S_{SS}^{-1/2}C_{S\xi}$$

and then applying (i), Lemma A.2(i) and Lemma A.1(ii). □

A.3 Proof of Theorem 4.1

We proceed by examining the behaviour of $\tilde{\epsilon}_t$, the one-step forecast residuals. From (2.6), we can write these

$$\tilde{\varepsilon}_t = \frac{\varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s x_s' \left(\sum_{s=1}^{t-1} x_s x_s'\right)^{-1} x_t}{\left[1 + x_t' \left(\sum_{s=1}^{t-1} x_s x_s'\right)^{-1} x_t\right]^{1/2}} \tag{A.1}$$

We break the result into two lemmas, one describing denominator and one the numerator, with similar reasoning in each case.

Lemma A.4. *Under Assumptions 3.1, 3.2 and 3.3,*

$$x_t' S_{xx}^{-1} x_t - W' F_W^{-1} W = \mathrm{o}(t^{-\zeta}) \ \textit{a.s.} \tag{A.2}$$

for all $\zeta < 1/8$ with W and F_W as in Theorem 4.1.

Proof. Divide the statistic into two parts using that

$$\|x_t' S_{xx}^{-1} x_t - W' F_W^{-1} W\| \leq \|x_t' S_{xx}^{-1} x_t - W_{t-1}' S_{WW}^{-1} W_{t-1}\| + \|W_{t-1}' S_{WW}^{-1} W_{t-1} - W' F_W^{-1} W\|.$$

We use a partial regression transformation to divide the first part into two partial components

$$\|x_t' S_{xx}^{-1} x_t - W_{t-1}' S_{WW}^{-1} W_{t-1}\| \leq \|(\xi_2|R,W)_t' S_{\xi\xi.RW}^{-1} (\xi_2|R,W)_t\| + \|(R|W)_t' S_{RR.W}^{-1} (R|W)_t\|$$

The first normed term on the right hand side is $\mathrm{o}(t^{2\beta-1})$ and the second is $\mathrm{o}(t^{-\zeta})$ by Lemma A.1 parts (iv) and (viii); and (iii) and (vii), respectively. The second term will dominate since $\alpha > 16/7$ so $\|x_t' S_{xx}^{-1} x_t - W_{t-1}' S_{WW}^{-1} W_{t-1}\| = \mathrm{o}(t^{-\zeta})$.

The lemma is then proven by rewriting the second step

$$\begin{aligned} W_{t-1}' S_{WW}^{-1} W_{t-1} - W' F_W^{-1} W =& (\mathbf{W}^{-(t-1)} W_{t-1})' [(\mathbf{W}^{t-1})' S_{WW}^{-1} \mathbf{W}^{t-1} - F_W^{-1}] (\mathbf{W}^{-(t-1)} W_{t-1}) \\ &+ (\mathbf{W}^{-(t-1)} W_{t-1} - W)' F_W^{-1} (\mathbf{W}^{-(t-1)} W_{t-1}) \\ &+ W' F_W^{-1} (\mathbf{W}^{-(t-1)} W_{t-1} - W) \end{aligned}$$

and noting that $\mathbf{W}^{t-1} W_{t-1} - W = \mathrm{O}(\lambda_{\min}(\mathbf{W})^{-t})$ by N05 (Corollary 5.3(i)) and $(\mathbf{W}^{t-1})' S_{WW}^{-1} \mathbf{W}^{t-1} - F_W^{-1} = \mathrm{O}(\lambda_{\min}(\mathbf{W})^{-2t})$ by N05 (Corollary 7.2), while all the other terms are bounded by the same corollaries. □

We next state a lemma concerning the main numerator term in (A.1).

Lemma A.5. *Under Assumptions 3.1, 3.2, 3.3 and 3.4*

$$\sum_{s=1}^{t-1} \varepsilon_s x_s' S_{xx}^{-1} x_t - G_t F_W^{-1} W = \mathrm{o}(t^{\beta - 1/8}) \ \textit{a.s.} \tag{A.3}$$

for all $\beta > 1/\alpha$, where W and F_W are defined as in Theorem 4.1, and

$$G_t = \sum_{s=1}^{t-1} \varepsilon_{t-s} W' (\mathbf{W}^{-s})' = \mathrm{o}(t^{\beta}). \tag{A.4}$$

Proof. Once again we take the proof in two steps, using that

$$\left\|\sum_{s=1}^{t-1} \varepsilon_s x_s' S_{xx}^{-1} x_t - G_t F_W^{-1} W\right\|$$
$$\leq \left\|\sum_{s=1}^{t-1} \varepsilon_s x_s' S_{xx}^{-1} x_t - \sum_{s=1}^{t-1} \varepsilon_s W_{s-1}' S_{WW}^{-1} W_{t-1}\right\| + \left\|\sum_{s=1}^{t-1} \varepsilon_s W_{s-1}' S_{WW}^{-1} W_{t-1} - G_t F_W^{-1} W\right\|.$$

For the first step, we again decompose using a partial regression transformation, so that

$$\left\|\sum_{s=1}^{t-1} \varepsilon_s x_s' S_{xx}^{-1} x_t - \sum_{s=1}^{t-1} \varepsilon_s W_{s-1} S_{WW}^{-1} W_{t-1}\right\| \leq \left\|\sum_{s=1}^{t-1} \varepsilon_s (\xi_2|R,W)_s' S_{\xi\xi\cdot RW}^{-1} (\xi_2|R,W)_t'\right\| + \left\|\sum_{s=1}^{t-1} \varepsilon_s (R|W)_s' S_{RR\cdot W}^{-1} (R|W)_t\right\| \tag{A.5}$$

and we consider each term on the right separately.

For the first term in (A.5) we use Lemma A.1(iv) to write

$$\left\|\sum_{s=1}^{t-1} \varepsilon_s (\xi_2|R,W)_s' S_{\xi\xi\cdot RW}^{-1} (\xi_2|R,W)_t'\right\| \leq \left\|\sum_{s=1}^{t-1} \varepsilon_s (\xi_2|R,W)_s' S_{\xi\xi}^{-1/2}\right\| \mathrm{O}(1) \left\|S_{\xi\xi}^{-1/2} (\xi_2|R,W)_t'\right\|$$

and then apply Lemma A.3(ii) and Lemma A.1(viii) to arrive at $\mathrm{o}(t^{\beta-1/2})$.

For the second term in (A.5) we use Lemma A.1(iii) to write

$$\left\|\sum_{s=1}^{t-1} \varepsilon_s (R|W)_s' S_{RR\cdot W}^{-1} (R|W)_t\right\| \leq \left\|\sum_{s=1}^{t-1} \varepsilon_s (R|W)_s' S_{RR}^{-1/2}\right\| \mathrm{O}(1) \left\|S_{RR}^{-1/2} (R|W)_t\right\|$$

and then apply Lemma A.2(iv) and Lemma A.1(vii) to arrive at $\mathrm{o}(t^{\beta-1/8})$. Overall then, the first step is dominated by this second term.

For the second step we have to show the bounding rate for

$$\begin{aligned}
&\sum_{s=1}^{t-1} \varepsilon_s W_{s-1}' S_{WW}^{-1} W_{t-1} - G_t F_W^{-1} W \\
&= \left[\sum_{s=1}^{t-1} \varepsilon_s W_{s-1}' (\mathbf{W}^{-(t-1)})'\right] \left[(\mathbf{W}^{t-1})' S_{WW}^{-1} \mathbf{W}^{t-1}\right] \mathbf{W}^{-(t-1)} W_{t-1} - G_t F_W^{-1} W \\
&= \left[\sum_{s=1}^{t-1} \varepsilon_s W_{s-1}' (\mathbf{W}^{-(t-1)})' - G_t\right] \left[(\mathbf{W}^{t-1})' S_{WW}^{-1} \mathbf{W}^{t-1}\right] \mathbf{W}^{-(t-1)} W_{t-1} \\
&\quad + G_t \left[(\mathbf{W}^{t-1})' S_{WW}^{-1} \mathbf{W}^{t-1} - F_W^{-1}\right] \mathbf{W}^{-(t-1)} W_{t-1} \\
&\quad + G_t F_W^{-1} \left[\mathbf{W}^{-(t-1)} W_{t-1} - W\right]
\end{aligned}$$

Many of these terms are familiar from the analysis of (A.3), and the only new terms to bound are $\sum_{s=1}^{t-1} \varepsilon_s W_{s-1}' (\mathbf{W}^{-(t-1)})' - G_t$ and G_t. For the latter we have

$$\|G_t\| = \left\|\sum_{s=1}^{t-1} \varepsilon_{t-s} W' (\mathbf{W}^{-s})'\right\| \leq \left\|\max_{1\leq s<t} \varepsilon_s\right\| \|W'\| \left\|\sum_{s=1}^{t-1} (\mathbf{W}^{-s})'\right\|$$

which is $o(t^\beta)$ since the latter two terms are bounded, while $\varepsilon_s = o(s^\beta)$ by Lai and Wei (1985, Theorem 1). For the former term we have

$$\begin{aligned}
&\left\| \sum_{s=1}^{t-1} \varepsilon_s W'_{s-1} (\mathbf{W}^{-(t-1)})' - G_t \right\| \\
=&\left\| \sum_{s=1}^{t-1} \varepsilon_s \left[\mathbf{W}^{-(t-1)} W_{s-1} - \mathbf{W}^{-(t-s)} W \right]' \right\| \\
=&\left\| \sum_{s=1}^{t-1} \varepsilon_s \left[\mathbf{W}^{s-t)} \sum_{p=s}^{\infty} \mathbf{W}^{-p} e_{W,p} \right]' \right\| \\
\leq&\left\| \max_{1\leq s<t} \varepsilon_s \right\| \|\mathbf{W}^{-t}\| \sum_{s=1}^{t-1} \left\| \sum_{u=0}^{\infty} \mathbf{W}^{-u} e_{W,u+s} \right\| \\
=&O(t^\beta) O(\lambda_{\min}(\mathbf{W})^{-t}) o(t^{1+\beta}) \\
=&o(t^{2\beta+1} \lambda_{\min}(\mathbf{W})^{-t})
\end{aligned}$$

where at the second last line we use that $\sum_{u=0}^{\infty} \mathbf{W}^{-u} e_{W,u+s} = o(s^\beta)$ by Nielsen (2008, Corollary 4.3). Combining these results, we see that this second step vanishes exponentially fast, and the first step dominates the expression of interest, giving the result.

The order of G_t follows by writing

$$\begin{aligned}
G_t &= \sum_{s=1}^{t-1} \varepsilon_{t-s} W' (\mathbf{W}^{-s})' \\
&\leq \left\| \max_{1\leq s<t} \varepsilon_s \right\| \|W\| \left\| \sum_{s=1}^{t-1} (\mathbf{W}^{-s}) \right\|,
\end{aligned}$$

and applying Lai and Wei (1985, Theorem 1). □

Proof of Theorem 4.1. We aim to show that

$$\mathsf{C}_{1,t}^2 - (q_t/\sigma)^2 \overset{\text{a.s.}}{=} o(1). \tag{A.6}$$

Using (2.6) we can rewrite this expression as

$$\frac{\tilde{\varepsilon}_t^2}{(t-k-1)^{-1} RSS_{t-1}} - \left(\frac{q_t}{\sigma}\right)^2 = \tilde{\varepsilon}_t^2 \left[\frac{(t-k-1)}{RSS_{t-1}} - \frac{1}{\sigma^2} \right] + \frac{\tilde{\varepsilon}_t^2 - q_t^2}{\sigma^2}. \tag{A.7}$$

We first consider the difference $\tilde{\varepsilon}_t^2 - q_t^2$. We have from (A.1),

$$\tilde{\varepsilon}_t^2 - q_t^2 = \frac{(\varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s x'_s S_{xx}^{-1} x_t)^2}{1 + x'_t S_{xx}^{-1} x_t} - \frac{(\varepsilon_t - G_t F_W^{-1} W)^2}{1 + W' F_W^{-1} W} = \frac{(\varepsilon_t - A_3)^2}{1 + A_1} - \frac{(\varepsilon_t - A_4)^2}{1 + A_2} \tag{A.8}$$

$$= \frac{A_2 - A_1}{(1 + A_1)(1 + A_2)} (\varepsilon_t - A_3)^2 + \frac{1}{1 + A_2} (A_4 - A_3)(2\varepsilon_t - A_4 - A_3), \tag{A.9}$$

where

$$A_1 = x'_t S_{xx}^{-1} x_t \qquad A_2 = W' F_W^{-1} W \qquad A_3 = \sum_{s=1}^{t-1} \varepsilon_s x'_s S_{xx}^{-1} x_t \qquad A_4 = G_t F_W^{-1} W. \tag{A.10}$$

Both denominators are bounded from below by unity, since A_1 and A_2 are non-negative. In the first numerator, $A_1 - A_2$ is $\mathrm{o}(t^{-\zeta})$ by Lemma A.4. The factor $\varepsilon_t - A_3 = \varepsilon_t - A_4 + A_4 - A_3$ is $\mathrm{o}(t^{\beta})$ since ε_t and A_4 are both $\mathrm{o}(t^{\beta})$ by Lai and Wei (1985, Theorem 1) and Lemma A.5 respectively, while $A_4 - A_3$ is $\mathrm{O}(t^{\beta-1/8})$ by Lemma A.5. So the first term of the sum is $\mathrm{o}(t^{2\beta-1/8})$.

In the second numerator, $A_4 - A_3$ is $\mathrm{O}(t^{\beta-1/8})$ by Lemma A.5, while ε_t and A_4 are each $\mathrm{o}(t^{\beta})$ as above, so that the whole second term is also $\mathrm{o}(t^{2\beta-1/8})$.

Thus the second term in (A.7) will vanish as long as $2\beta < 1/8$ or $\alpha > 16$ in Assumption 3.1, as required. To show the same for the first term, note that $\tilde{\varepsilon}_t^2 = q_t^2 + (\tilde{\varepsilon}_t^2 - q_t^2)$, where the difference vanishes as just proved, while

$$q_t^2 = \frac{(\varepsilon_t - A_4)^2}{1 + A_2} = \mathrm{o}(t^{2\beta})$$

since, as above, ε_t and A_4 are both $\mathrm{o}(t^{\beta})$ as above, while A_2 is nonnegative. Then N05 (Corollary 2.9) implies that

$$\frac{(t-k-1)}{RSS_{t-1}} - \frac{1}{\sigma^2} = \mathrm{o}(t^{\gamma})$$

for $\gamma < 1/2$. So the first term in (A.7) will vanish as long as $2\beta < 1/2$, which is satisfied by Assumption 3.1.

A.4 Proof of Lemma 4.2

Proof. Theorem 4.1 shows that $\mathsf{C}_{1,t}^2 - q_t^2$ vanishes almost surely. Egorov's theorem (Davidson, 1994, 18.4) then shows that $\mathsf{C}_{1,t}^2 - q_t^2$ vanishes uniformly on a set with large probability. That is,

$$\forall \epsilon > 0 \exists T_0 : \Pr(\sup_{t > T_0} |\mathsf{C}_{1,t}^2 - q_t^2| < \epsilon) > 1 - \epsilon.$$

This implies that for any sequence $g(T)$ which increases to infinity, then $\sup_{g(T) < t \leq T} |\mathsf{C}_{1,t}^2 - q_t^2| \xrightarrow{\mathrm{p}} 0$ as $T \to \infty$. □

A.5 Proof of Lemma 4.3 (correction to Lemma 1 of Deo (1972))

Proof. The first part of Deo's lemma, determining the domain of attraction as Λ, is correct. The second part, determining the norming sequences, is in error. Deo cites Cramér (1946, p. 374) for this calculation. There Cramér calculates the norming sequences for a sequence of independent standard normal random variables (with a right tail differing from the density of interest in only a constant factor). We follow the slightly more direct approach of Leadbetter et al. (1982, Theorem 1.5.3).

Since $\{X_n\}$ are independent standard normal random variables, $\{|X_n|\}$ are independent random variables identically distributed with the half-normal density, that is, the normal density folded around zero:

$$\Pr\{|X_1| < x\} = F(x) = \sqrt{2/\pi} \int_0^x e^{-t^2/2}\, dt = 2\Phi(x), \quad x \geq 0 \tag{A.11}$$

We are interested in probabilities of the form $\Pr\{a_n(Z_n - b_n) < x\}$, which may be rewritten $\Pr\{Z_n \leq u_n\}$, where $u_n(x) = x/a_n + b_n$. We seek a_n, b_n such that the sequence u_n satisfies (1.5.1) in Leadbetter et al. (1982, Theorem 1.5.1), namely

$$n(1 - F(u_n)) \to e^{-x} \qquad \text{as } n \to \infty. \tag{A.12}$$

Apply a modified version of the well-known normal tail relation,

$$1 - F(u) \sim f(u)/u \quad \text{as } u \to \infty, \tag{A.13}$$

so that combining (A.12) and (A.13) we have that $(1/n)e^{-x}u_n/f(u_n) \to 1$. Taking logs and substituting the density f, we have

$$-\log n - x + \log u_n - \tfrac{1}{2}\log(\pi/2) + u_n^2/2 \to 0. \tag{A.14}$$

Dividing through by $\log n$,

$$-1 - \frac{x}{\log n} + \frac{\log u_n}{\log n} - \frac{\log(\pi/2)}{\log n} + \frac{u_n^2}{2\log n} \to 0, \tag{A.15}$$

then for any fixed x, the second and fourth terms vanish trivially. The third term vanishes by substituting (A.12) for n and twice applying L'Hôpital's rule. It then follows that $\frac{u_n^2}{2\log n} \to 1$, or (taking logarithms again),

$$2\log u_n - \log 2 - \log\log n \to 0. \tag{A.16}$$

Substituting this result into (A.14), we have that

$$-\log n - x + \tfrac{1}{2}\log 2 + \tfrac{1}{2}\log\log n - \tfrac{1}{2}\log(\pi/2) + u_n^2/2 \to 0. \tag{A.17}$$

so that rearranging,

$$u_n^2 = 2\log n\left\{1 + \frac{x - \frac{1}{2}\log\pi - \frac{1}{2}\log\log n}{\log n} + \mathrm{o}\left(\frac{1}{\log n}\right)\right\},$$

and hence the maximum of n half-normal random variables has the form

$$u_n = (2\log n)^{1/2}\left\{1 + \frac{x - \frac{1}{2}\log\pi - \frac{1}{2}\log\log n}{2\log n} + \mathrm{o}\left(\frac{1}{\log n}\right)\right\}.$$

It then follows from Leadbetter et al. (1982, Theorem 1.5.3) that $\Pr\{Z_n \leq u_n\} \to \exp(-e^{-x})$, and rearranging gives the norming sequences. □

A.6 Proof of Lemma 4.4

Proof. Consider the normalised linear process

$$q_t/\sigma = (\varepsilon_t/\sigma)(1 + W'F_W^{-1}W)^{-1/2} - \sum_{s=1}^{\infty}(\varepsilon_{t-s}/\sigma)W'(\mathbf{W}^{-s})'F_W^{-1}W(1 + W'F_W^{-1}W)^{-1/2}$$

In the case without explosive components, this reduces to

$$q_t/\sigma = (\varepsilon_t/\sigma)$$

so that under Assumption 3.5 q_t/σ is an independent standard normal sequence, and q_t^2/σ^2 is an independent $\chi^2_{(1)}$ sequence. Then classical extreme value theory gives the lemma with the norming sequences a_t and b_t as stated (see, for instance p. 56 of Embrechts et al., 1997, noting that the χ^2 distribution is a special case of the gamma distribution).

When an explosive component is present, q_t/σ under Assumption 3.5 is still marginally standard normal. However dependence between members of the sequence means that classical extreme value theory cannot be applied. In particular, we have:

$$\begin{aligned}\mathsf{E}(q_t/\sigma) &= 0\\ \mathsf{Var}(q_t/\sigma) &= 1\\ \mathsf{Covar}(q_s/\sigma, q_t/\sigma) = r(s,t) = r_{|t-s|} &= 2W'(F_W^{-1})\mathbf{W}^{-|t-s|}W(1+W'F_W^{-1}W)^{-1}\end{aligned}$$

The general approach to dealing with dependent sequences is outlined in Leadbetter and Rootzen (1988); as long as the dependence is not too great, the same limiting results hold.

We take advantage of the relationship between the $\chi^2_{(1)}$ and normal distributions to use existing results on dependent normal sequences to analyse the limiting behaviour of q_t^2/σ^2. In particular, we have

$$\max_t q_t^2/\sigma^2 < u_t \qquad \text{iff} \qquad \max_t |q_t/\sigma| < \sqrt{u_t} \tag{A.18}$$

where $|q_t/\sigma|$ has the half-normal distribution. Lemma 1 of Deo (1972) and its Corollary consider just such processes, under a square-summability condition that holds here: $\sum r_s^2 = 4 < \infty$. Then Deo's result is

$$c_t(\max_{1\le s\le t} |q_s/\sigma| - d_t) \xrightarrow{\text{d}} \Lambda$$

with

$$\begin{aligned}c_t &= (2\log t)^{1/2}\\ d_t &= (2\log t)^{1/2} - (8\log t)^{-1/2}(\log\log t + \log\pi).\end{aligned}$$

(Note that the centring sequence—here d_t, originally b_n—is incorrect in the original. A correction is provided as Lemma 4.3) Taking $\sqrt{u_t(z)} = c_t z + d_t$ and using (A.18) and (A.19), we have

$$\Pr\left\{\frac{c_t}{2d_t}\left(\max_{1\le s\le t} q_t^2/\sigma^2 - d_t^2\right)\right\} \xrightarrow{\text{d}} \Lambda$$

giving norming sequences

$$\begin{aligned}a_t' &= \frac{c_t}{2d_t} \quad \text{(scaling)}\\ b_t' &= d_t^2 \quad \text{(centring)}.\end{aligned}$$

The equivalence between a_t', b_t' and a_t, b_t is proved by showing that $a_t/a_t' \to 1$ and $a_t(b_t' - b_t) \to 0$. □

A.7 Proof of Theorem 4.5

By a property of inequalities we can establish a lower bound on the supremum statistic,

$$\frac{1}{2}\left[\max_{g(T)\le t\le T}(\mathsf{C}^2_{1,t}) - d_{T-g(T)}\right] \le \frac{1}{2}\max_{g(T)\le t\le T}\left(\mathsf{C}^2_{1,t} - (q_t/\sigma)^2\right) + \frac{1}{2}\left[\max_{g(T)\le t\le T}(q_t/\sigma)^2 - d_{T-g(T)}\right] \tag{A.19}$$

where the left term vanishes in probability by Lemma 4.2 and the right term converges in distribution to by Lemma 4.4. We can establish a similar upper bound, so that the normalised supremum statistic is bounded above and below by quantities that converge in distribution, and the theorem is proved.

References

Andrews, Donald W. K (1993), 'Tests for parameter instability and structural change with unknown change point', *Econometrica* **61**(4), 821–856.

Assarsson, Bengt, Claes Berg and Per Jansson (2004), Investment in Swedish manufacturing: Analysis and forecasts, Technical Report 2, Sveriges Riksbank.

Barnett, V. and T. Lewis (1994), *Outliers in statistical data*, 3rd edn, Wiley.

Bauer, Dietmar (2009), 'Almost sure bounds on the estimation error for OLS estimators when the regressors include certain MFI(1) processes', *Econometric Theory* **25**(02), 571–582.
URL: *http://dx.doi.org/10.1017/S0266466608090579*

Brown, R. L. and J. Durbin (1968), Methods of investigating whether a regression relationship is constant over time, *in* 'Selected Statistical Papers, European Meeting', Mathematical Centre Tracts No. 26, Amsterdam.

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. Series B (Methodological)* **37**(2), 149–192.

Celasun, Oya and Mangal Goswami (2002), An analysis of money demand and inflation in the Islamic Republic of Iran, IMF Working Paper WP/02/205, International Monetary Fund.

Chow, Gregory C. (1960), 'Tests of equality between sets of coefficients in two linear regressions', *Econometrica* **28**(3), 591–605.

Cramér, Harald (1946), *Mathematical Methods in Statistics*, Princeton University Press, Princeton.

Davidson, James (1994), *Stochastic Limit Theory*, Oxford University Press, Oxford.

Deo, Chandrakant M. (1972), 'Some limit theorems for maxima of absolute values of gaussian sequences', *Sankhyā: The Indian Journal of Statistics, Series A* **34**(3), 289–292.
URL: *http://www.jstor.org/stable/25049808*

Doornik, J.A. (2007), *Object-Oriented Matrix Programming Using Ox*, 3rd edn, Timberlake Consultants Press and Oxford, London.
URL: *http://www.doornik.com*

Doornik, Jurgen A. and Henrik Hansen (2008), 'An omnibus test for univariate and multivariate normality', *Oxford Bulletin of Economics and Statistics* **70**(s1), 927–939.

Dufour, Jean-Marie (1982), 'Recursive stability analysis of linear regression relationships: An exploratory methodology', *Journal of Econometrics* **19**(1), 31–76.

Embrechts, Paul, Claudia Klüppelberg and Thomas Mikosch (1997), *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

Engler, Eric and Bent Nielsen (2009), 'The Empirical Process of Autoregressive Residuals', *Econometrics Journal* **12**, 367–381.

Galpin, Jacqueline S. and Douglas M. Hawkins (1984), 'The use of recursive residuals in checking model fit in linear regression', *The American Statistician* **38**(2), 94–105.

Hendry, David F. (1986), 'Using PC-GIVE in econometrics teaching.', *Oxford Bulletin of Economics and Statistics* **48**(1), 87–98.

Hendry, David F. (1995), *Dynamic Econometrics*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

Johansen, Søren (2000), 'A Bartlett correction factor for tests on the cointegrating relations', *Econometric Theory* (16), 740–778.

Kimura, Takeshi (2001), The impact of financial uncertainties on money demand in Europe, *in* H.-J.Klöckers and C.Willeke, eds, 'Monetary Analysis: Tools and Applications', European Central Bank, Frankfurt am Main, pp. 97–116.

Krämer, Walter, Werner Ploberger and Raimund Alt (1988), 'Testing for structural change in dynamic models', *Econometrica* **56**(6), pp. 1355–1369.

Lai, T. L. and C. Z. Wei (1985), Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems, *in* P. R.Krishnaiah, ed., 'Multivariate Analysis', Vol. VI, pp. 375–393.

Lai, Tze Leung and Ching Zong Wei (1982), 'Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems', *Annals of Statistics* **10**(1), 154–166.

Leadbetter, M. R., Georg Lindgren and Holger Rootzén (1982), *Extremes and related properties of random sequences and processes*, Springer Series in Statistics, Springer-Verlag, New York.

Leadbetter, M. R. and Holger Rootzen (1988), 'Extremal theory for stochastic processes', *The Annals of Probability* **16**(2), 431–478.
URL: *http://www.jstor.org/stable/2243819*

Nielsen, Bent (2005), 'Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms', *Econometric Theory* **21**(03), 534–561.

Nielsen, Bent (2008), Singular vector autoregressions with deterministic terms: Strong consistency and lag order determination, Oxford Economics Papers 2008-W14, Department of Economics, University of Oxford.

Nielsen, Bent and Jouni S. Sohkanen (2011), 'Asymptotic behavior of the cusum of squares test under stochastic and deterministic time trends', *Econometric Theory* **27**(04), 913–927.
URL: *http://dx.doi.org/10.1017/S0266466610000563*

Pakshirajan, R. P. and H. Vishnu Hebbar (1977), 'Limit theorems for extremes of absolute values of gaussian sequences', *Sankhyā: The Indian Journal of Statistics, Series A* **39**(2), 191–195.
URL: *http://www.jstor.org/stable/25050092*

Ploberger, Werner and Walter Krämer (1986), 'On studentizing a test for structural change', *Economics Letters* **20**(4), 341–344.
URL: *http://www.sciencedirect.com/science/article/pii/0165176586900091*

Quandt, Richard E. (1960), 'Tests of the hypothesis that a linear regression system obeys two separate regimes', *Journal of the American Statistical Association* **55**(290), 324–330.

Searle, Shayle R. (1982), *Matrix Algebra Useful for Statistics*, John Wiley and Sons.

Srikantan, K. S. (1961), 'Testing for the single outlier in a regression model', *Sankhya: The Indian Journal of Statistics, Series A* **23**(3), pp. 251–260.

Theil, H. (1965), 'The analysis of disturbances in regression analysis', *Journal of the American Statistical Association* **60**(312), 1067–1079.