

Nielsen, Bent; Whitby, Andrew

Article

A joint chow test for structural instability

Econometrics

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Nielsen, Bent; Whitby, Andrew (2015) : A joint chow test for structural instability, *Econometrics*, ISSN 2225-1146, MDPI, Basel, Vol. 3, Iss. 1, pp. 156-186, <https://doi.org/10.3390/econometrics3010156>

This Version is available at:

<https://hdl.handle.net/10419/171821>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

Article

A Joint Chow Test for Structural Instability

Bent Nielsen ^{1,*} and Andrew Whitby ²

¹ Department of Economics, University of Oxford & Institute of Economic Modelling & Nuffield College, Oxford OX1 1NF, UK

² The World Bank, 1818 H Street NW, Washington DC 20433, USA

* Author to whom correspondence should be addressed; E-Mail: bent.nielsen@economics.ox.ac.uk; Tel: +44-1865-278500.

Academic Editor: Kerry Patterson

Received: 19 December 2014 / Accepted: 26 February 2015 / Published: 12 March 2015

Abstract: The classical Chow test for structural instability requires strictly exogenous regressors and a break-point specified in advance. In this paper, we consider two generalisations, the one-step recursive Chow test (based on the sequence of studentised recursive residuals) and its supremum counterpart, which relaxes these requirements. We use results on the strong consistency of regression estimators to show that the one-step test is appropriate for stationary, unit root or explosive processes modelled in the autoregressive distributed lags (ADL) framework. We then use the 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 the normality of errors and is intended to be used in situations where this can be either assumed nor established empirically. Simulations show that the supremum test has desirable power properties, in particular against level shifts late in the sample and against outliers. An application to U.K. GDP data is given.

Keywords: Chow test; autoregressive distributed lag model; non-stationarity

JEL classifications: C22

1. Introduction

Identifying structural instability in models is of major concern to econometric practitioners. The Chow [1] 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 one-step recursive Chow test, based on the sequence of studentised recursive forecast residuals; and its supremum counterpart. The pointwise test is frequently used and reported in the 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., [2]. 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 6, 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 proven earlier to construct a supremum version of the one-step test, applicable to detecting parameter change or the outlier at an unknown point in the sample. The supremum test offers several advantages useful to modellers: it is simple to compute and has a standard distribution under the null, which does not depend on the autoregressive parameter (even in the unit-root or explosive cases); it focuses attention on end-sample instability; and it is agnostic about the number of breaks, giving power against more complex forms of misspecification. These advantages incur certain costs: the test is not invariant to the distribution of errors (even asymptotically); and other tests are more powerful against particular alternatives.

The pointwise one-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 [3] as part of a suite of model misspecification diagnostics; a similar diagnostic graphic, the “one-step forecast test”, is provided in EViews ([4] p. 180). The idea of using residuals calculated recursively to test model misspecification dates to the landmark cumulated sum (CUSUM) and cumulated sum of squares (CUSUMSQ) tests [5,6], which are based on partial sums of (squared) recursive residuals and have since been generalised to models, including lagged dependent variables [7–9]. Unlike these tests, the one-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 the model specification is, of course, well established. The residuals can be either OLS residuals or recursive residuals, see [6,10]. The recursive residuals have two advantages over the OLS residuals: first, under the normal linear model with fixed regressors, they are identically and independently normal; second, 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, the independence of the residuals does not hold due to the presence of lagged dependent variables, see [11]. 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 one-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 tests for either structural breaks or outliers, both at a possibly unknown time.

Perron [12] divides structural breaks tests into those that do not explicitly model a break, those that model one break and those that model multiple breaks. Our test joins the first category, which includes the already mentioned CUSUM and CUSUMSQ tests. (As an aside, Perron notes that these tests can suffer from the non-monotonicity of power against some alternatives. The risk of this is much reduced by using the one-step Chow statistics, since all parameters are estimated on a growing sub-sample.) The modelled break category includes, most prominently, the Quandt-Andrews (respectively, [13,14] supremum tests. These tests are complicated by a non-standard distribution (tabulated in [15]), but are nevertheless popular in practice, being implemented in several software packages. Our test is distinguished from these by not imposing any restrictions on the end-of-sample, so that end-of-sample instability may be detected. This feature is similar to [15], but a key distinction is that our test is agnostic about the number of breaks in the sample, a useful property in practice. It is also substantially simpler in implementation. Additionally, because the one-step tests behave like an i.i.d. process, the asymptotics differ from full-sample tests, like Quandt-Andrews, requiring the application of the 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 [16], which, however, considers an unknown outlier in a classical setting. Outliers in the Box-Jenkins paradigm have attracted substantial interest, see [17–19]. These authors take a full-sample approach with stepwise elimination of outliers. Although effective in many cases, there is a risk of smearing/masking effects when multiple outliers are present, which is reduced with the recursive the test we present.

Surprisingly, the use of recursive residuals to detect outliers in time series data is relatively unexplored, although there is little doubt that they are used for this purpose in practice. Barnett and Lewis ([20] p. 330) comment that “[recursive residuals] would seem to have the potential for the study of outliers, although no major progress on this front is evident. There is a major difficulty in that the labelling 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 one-step test applies to a linear regression:

$$y_t = \beta'x_t + \varepsilon_t \quad t = 1, \dots, T, \quad (1)$$

with y_t scalar, x_t a k -dimensional vector of regressors and the errors independently and identically distributed. For such a regression, we can define the sequence of least squares estimators calculated over progressively larger subsamples,

$$\hat{\beta}_t = \left(\sum_{s=1}^t x_s x_s' \right)^{-1} \left(\sum_{s=1}^t x_s y_s' \right) \quad t = k, \dots, T, \quad (2)$$

along with the corresponding residual sums of squares

$$RSS_t = \sum_{s=1}^t (\hat{\beta}'_t x_s - y_s)^2 \quad t = k, \dots, T, \tag{3}$$

and recursive residual (or standardised one-step forecast error)

$$\tilde{\varepsilon}_t = \left[1 + x'_t \left(\sum_{s=1}^{t-1} x_s x'_s \right)^{-1} x_t \right]^{-1/2} (y_t - \hat{\beta}'_{t-1} x_t) \quad t = (k + 1), \dots, T. \tag{4}$$

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

$$C^2_{1,t} = \frac{(RSS_t - RSS_{t-1})(t - k - 1)}{RSS_{t-1}} \quad t = (k + 1), \dots, T, \tag{5}$$

and can be expressed as:

$$C^2_{1,t} = \frac{\tilde{\varepsilon}_t^2(t - k - 1)}{RSS_{t-1}}. \tag{6}$$

Chow showed that in a classical Gaussian regression model, this statistic would have an exact $F(1, t - k - 1)$ distribution. We first extend this result to show that, for a general class of Gaussian autoregressive distributed lag (ADL) processes, $C^2_{1,t}$ converges in distribution to a χ^2_1 random variable, so that, asymptotically, the additional dependence does not matter. This result means that comparing the pointwise statistic against an $F(1, \cdot)$ or χ^2_1 distribution (as is typically done) is appropriate in large samples. However, it still leaves unresolved the difficulty that this test is generally reported graphically to detect parameter change with an unknown change point. 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 based on the test statistic:

$$\max_{g(T) \leq t \leq T} C^2_{1,t}, \tag{7}$$

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

To put the test statistics and the asymptotic analysis into perspective, it is useful to review some related, but also somewhat different statistics. The literature on adaptive control, also called tracking, comes to mind, although it does not appear to be applied much in econometrics. It is concerned with tracking the sum of squares innovations. The aim is then to show that:

$$\frac{1}{T} \sum_{t=1}^T (y_t - \hat{\beta}' x_{t-1})^2 - \frac{1}{T} \sum_{t=1}^T \varepsilon_t^2 \tag{8}$$

vanishes. A discussion for a non-explosive autoregressive setup without deterministic terms is given in [21,22]. Asymptotic distribution theory does not seem to be discussed. The reason that the tracking result does not extend to the explosive case is that the residuals are not normalised by the “hat” matrix, in contrast to the residuals in Equation (4). Normalisation by the “hat” matrix also gives excellent finite-sample properties; see Section 5.

The Chow statistic $C_{1,t}^2$ in Equation (5) also involves scaling by a residual variance estimator, so that the asymptotic distribution is free of nuisance parameter; see Theorem 6 below. More fundamentally, the present analysis is concerned with the Chow statistic for individual observations rather than sums, and it will therefore involve extreme value theory.

Another related test that is used in econometrics is the CUSUMSQ test based on the statistic:

$$CUSQ_{t,T} = \sqrt{T} \left(\frac{\sum_{s=k}^t \tilde{\varepsilon}_s^2}{\sum_{s=k}^T \tilde{\varepsilon}_s^2} - \frac{t-k}{T-k} \right), \tag{9}$$

or a similar statistic based on least squares residual variance estimates instead of sums of squared recursive residuals. This test statistic is aimed at detecting non-constancy in the innovation variance rather than detecting individual outliers. Again, it includes normalisation by the “hat” matrix. Distribution theory is discussed in [7,8] for the stationary case and in [9] for general autoregressions, which are possibly explosive and with deterministic terms.

A third related test is Andrews’ sup-F test, see [13]. This is a test for structural breaks in the mean parameters for which the asymptotic theory only applies in the stationary case. The finite sample performances of the Chow test and the sup-F test are compared in Section 6.

3. Model and Assumptions

We consider the behaviour of the test statistic for ADL models with arbitrary deterministic terms, a class that includes by restriction many commonly-positied economic relationships, see ([23] Chapter 7). For the purpose of analysis, we assume that the true data generating model can be represented as a vector autoregression (VAR).

We observe a p -dimensional time series $X_{1-k}, \dots, X_0, X_1, \dots, X_T$. We 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,

$$Y_t = \rho Z_t + \sum_{j=1}^k \alpha_j Y_{t-j} + \sum_{j=1}^k \beta'_j Z_{t-j} + \nu D_{t-1} + \varepsilon_t \quad t = 1, \dots, T. \tag{10}$$

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

$$X_t = \sum_{j=1}^k A_j X_{t-j} + \mu D_{t-1} + \xi_t \quad t = 1, \dots, T, \tag{11}$$

with the deterministic term D_t given by:

$$D_t = \mathbf{D}D_{t-1}. \tag{12}$$

The deterministic term D_t follows the approach of [24,25] 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,

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

will generate a constant and a biannually dummy. The term D_t is assumed to have linearly-independent coordinates, formalised as follows.

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

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 ([25] Remark 9.3). In the present analysis, this constraint emerges in Lemma 12 (i) and is transmitted via Lemma 13 (iv) to Lemma 16. If $\dim \mathbf{D} = 0$ and the geometric multiplicity of roots in unity equals their algebraic multiplicity (including $I(1)$, but excluding $I(2)$ processes), this could be improved to finite moments greater than four using the result of [26].

Assumption 2. ξ_t is a martingale difference sequence with respect to the natural filtration \mathcal{F}_t , so $E(\xi_t | \mathcal{F}_{t-1}) = 0$. The initial values X_0, \dots, X_{1-k} are \mathcal{F}_0 -measurable and:

$$\sup_t E(\|\xi_t\|^\alpha | \mathcal{F}_{t-1}) \stackrel{a.s.}{<} \infty \quad \text{for some } \alpha > 16, \tag{13}$$

$$E(\xi'_t \xi_t | \mathcal{F}_{t-1}) \stackrel{a.s.}{=} \Omega \quad \text{where } \Omega \text{ is positive definite.} \tag{14}$$

This assumption also excludes the possibility that the innovations could be heteroscedastic, a common assumption in financial modelling (e.g., autoregressive conditional heteroscedastic, ARCH), but also an increasingly relevant property in macroeconomic work, particularly in light of the ‘‘Great Moderation’’ period [27] and subsequent period of the ‘‘Global Financial Crisis’’. The assumption indicates that such heteroscedasticity should be modelled.

We permit nearly all possible values of the autoregressive parameters A_j in Equation (11), excluding only the case of singular explosive roots, which can only arise for a VAR with $p \geq 2$ and multiple explosive roots, see [28] for a discussion. We can express the restriction in terms of the companion matrix:

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

Assumption 3. The explosive roots of \mathbf{B} have geometric multiplicity of 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 are martingale differences.

Assumption 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., $E(\varepsilon_t | \mathcal{G}_{t-1}) = 0$.

Finally, the one-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 the supremum statistic will rely on extreme value theory, we must impose distributional and independence assumptions on the ADL innovations ε_t , in order to uniquely determine the norming sequences applied in Lemma 9. We assume

normality, which may result from joint normality in the underlying VAR process and is tested, in practice, under the above assumptions, see [29].

Assumption 5. $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$.

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). A two-way decomposition allows us to express separately certain terms that arise in connection with the explosive component of the process. 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 Equation (11) in companion form, so that:

$$S_t = \mathbf{S}S_{t-1} + (\xi'_t, 0)'. \tag{15}$$

Then, there exists a regular real matrix \mathbf{M} to block diagonalize \mathbf{S} (see the elaboration in Section 3 of [25]), so that the process can be decomposed into non-explosive and explosive components, R_t and S_t , respectively. We have:

$$\begin{aligned} \mathbf{M}S_t &= (\mathbf{M}\mathbf{S}\mathbf{M}^{-1})\mathbf{M}S_{t-1} + \mathbf{M}(\xi'_t, 0)', \\ \begin{pmatrix} R_t \\ W_t \end{pmatrix} &= \begin{pmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{W} \end{pmatrix} \begin{pmatrix} R_{t-1} \\ W_{t-1} \end{pmatrix} + \begin{pmatrix} e_{R,t} \\ e_{W,t} \end{pmatrix}, \end{aligned} \tag{16}$$

with \mathbf{R} and \mathbf{W} having eigenvalues inside or on, and outside, the unit circle, respectively.

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 does not require the normality of Assumption 5.

Theorem 6. Under Assumptions 1, 2, 3 and 4,

$$C_{1,t}^2 - (q_t/\sigma)^2 \xrightarrow{as} 0 \quad as \ t \rightarrow \infty, \tag{17}$$

where:

$$q_t = \frac{\varepsilon_t - \sum_{s=1}^{\infty} \varepsilon_{t-s} W'(\mathbf{W}^{-s})' F_W^{-1} W}{(1 + W' F_W^{-1} W)^{1/2}}, \tag{18}$$

and \mathbf{W} is as in Equation (16), and as in [25] (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 the 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 7. Suppose $C_{1,t}^2 - (q_t/\sigma)^2 \xrightarrow{as} 0$ as $t \rightarrow \infty$. Then:

$$\sup_{g(T) < t \leq T} |C_{1,t}^2 - (q_t/\sigma)^2| \xrightarrow{P} 0 \quad \text{as } T, g(T) \rightarrow \infty. \tag{19}$$

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

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, $\max_{1 \leq s \leq t} q_s$, appropriately normalised, does indeed converge to the Gumbel extremal distribution (as $t \rightarrow \infty$), which has distribution function:

$$\Pr(\Lambda < x) = \exp[-\exp(-x)] \quad \text{where: } x \in \mathbf{R}. \tag{20}$$

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:

$$\Lambda \sim \text{Gumbel} \quad \text{iff} \quad 2e^{-\Lambda} \sim \chi_2^2. \tag{21}$$

In showing the above convergence, we rely on Theorem 1 of Deo [30], 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. Here, we state the correct sequences, adopting the notation of Deo (proof in Section A.6).

Lemma 8. 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 $a_n = (2 \log n)^{1/2}$ 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 Equation (18).

Lemma 9. Under Assumption 5,

$$q_t/\sigma \sim N(0, 1), \text{ and} \tag{22}$$

$$a_t(\max_{1 \leq s \leq t} q_s^2 - b_t) \xrightarrow{d} \Lambda, \tag{23}$$

where:

$$a_t = 1/2 \quad \text{and} \quad b_t = \log t^2 - \log \log t - \log \pi, \tag{24}$$

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 one-step Chow test converges in distribution to the Gumbel extremal distribution.

Theorem 10. Under Assumptions 1, 2, 3, 4 and 5, with some $g(T) \rightarrow \infty$,

$$SC_T^2 = \frac{1}{2} \left(\max_{g(T) < t \leq T} C_{1,t}^2 - d_{T-g(T)} \right) \xrightarrow{d} \Lambda \quad \text{as } T \rightarrow \infty, \tag{25}$$

where $C_{1,t}^2$ is the one-step Chow statistic defined in Equation (5) and:

$$d_{T-g(T)} = 2 \left\{ \log[T - g(T)] - \frac{1}{2} \log \log[T - g(T)] - \log \pi \right\}, \tag{26}$$

and Λ is a random variable distributed according to the Gumbel distribution Equation (20).

As a simple corollary, we can transform the test using Equation (21), so that it may be compared against a more readily-available distribution.

Corollary 11. Under the same assumptions, $2 \cdot \exp(-SC_T^2) \sim \chi_2^2$. A test based on this result should reject for small values of the statistic.

5. Finite-Sample Corrections

In practice, we find by simulation 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 one-step statistics appear to be distributed close to $F(1, t - k - 1)$ (as indeed, they are exactly in the classical case) and so use the following transformation to bring the statistics closer to the asymptotic chi-squared distribution:

$$C_{1,t}^{2*} = G^{-1}[F(C_{1,t}^2)] \tag{27}$$

where $F(\cdot)$ and $G(\cdot)$ are the $F(1, t - k - 1)$ and χ_1^2 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. We find that the test performs better if simply compared with the finite maximal distribution, assuming the 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). That is, we approximate the distribution of the maximum, $\max_{g(T) < t \leq T} C_{1,t}^{2*}$, by:

$$\Pr \left\{ \max_{g(T) < t \leq T} C_{1,t}^{2*} \leq x \right\} \approx \Pr \left\{ \max_{g(T) < t \leq T} \varepsilon_{1,t}^2 \leq x\sigma^2 \right\} = [G(x)]^{T-g(T)}. \tag{28}$$

This forms the basis of the finite adjusted sup-Chow test (SC^{2*}), with rejection in the right tail. Note that in this case, no centring or scaling is required, because the null distribution itself depends on T .

6. Simulation Study

We present the results of four simulation experiments done in Ox [31]: size; distributional sensitivity; power against mean shifts; and power against outliers. All of the simulations are done first for a first order autoregression and then for an autoregressive distributed lag model.

6.1. Autoregressive Data-Generating Process

Consider the following data-generating process:

$$\begin{aligned}x_t &= \alpha x_{t-1} + \varepsilon_t, \quad t = 1, \dots, T, \\x_0 &= 0.\end{aligned}\tag{29}$$

Where not otherwise stated, we set $\alpha = 0$ and $\varepsilon_t \stackrel{iid}{\sim} N(0, 1)$. The number of Monte Carlo repetitions and the implied Monte Carlo standard error, MCSE, are indicated in table captions.

The regression model is that of an first order autoregression. It includes an intercept unless otherwise stated. The five tests computed and presented in the tables are: the asymptotic sup-Chow test (SC^2); the corrected sup-Chow test (SC^{*2}); the [13] sup-F test (sup F); an outlier test based on the OLS residuals (sup t^2); and the [32] E_p test for normality (Φ). The nominal size is 5% unless otherwise stated.

The two sup-Chow tests are described in Theorem 10 and Equation (28), respectively; for the function $g(T)$, we use $T^{1/2}$. The sup-F test is a linear regression form of Andrews' sup-W (Wald) test with 15% trimming, as used for the simulations in [33] and implemented in EViews 7 by the command `ubreak`. It is the maximum of the Chow F-tests calculated over break points $0.15T \leq \lambda \leq 0.85T$, such that for each break point λ , under the alternative, the model is estimated separately for each subsample ($1 \dots \lambda$ and $\lambda + 1 \dots T$), whereas under the hypothesis, the model is estimated for the full sample. The null distribution is given by asymptotic approximation to a non-standard distribution with simulated critical values given in [15]. The sup- t^2 test examines the maximum of the squared full-sample OLS residuals, externally studentised as in ([34] s. 2.2.9); that is, residual t is normalised using an estimate of the error variance that excludes residual t itself. These squared statistics are $F_{1,t-k-1}$ distributed under normality of the errors, but not independent; hence, the use of the Bonferroni inequality to find significance levels is recommended by [34]. We find in simulation that, despite dependence, the exact maximal distribution of T independent $F_{1,t-k-1}$ random variables is a reasonable choice for the sample sizes and processes we consider, except those that are near-unit-root. Finally, in experiments where we wish to evaluate the performance of the sup-Chow test conditional on residuals having satisfied a normality test, we use the [32] E_p test of the OLS residuals.

In the first experiment (Table 1), we vary the autoregressive parameter through the stationary, unit-root and explosive regions and consider the effect of either including or excluding an intercept from the model. As noted above, the SC^2 test is uniformly oversized. The SC^{*2} test is correctly sized and approximately similar, with simulated size varying very little across the parameter space. There is some tendency towards inflated sizes under near-unit-root processes when an intercept is included in the model, but the extent of this is quite limited (7% simulated size). The key consequence of this result is that it is not necessary to know *a priori* where the autoregressive parameter lies to effectively apply the SC^{*2} test, avoiding a potential circularity in model construction.

In simulations that are not reported here, we also investigated the sup F test. This test is not valid in the non-stationary case. Thus, the same patterns were seen, albeit with a larger effect. The simulated size is as high as 44% in the unit root case.

Table 1. Simulated rejection frequency for SC^2 and SC^{*2} under the Gaussian autoregression in Equation (29). 200,000 repetitions, $MCSE \leq 0.1$.

T		Autoregressive Coefficient (α)							
		-1.03	-1.00	-0.50	0.00	0.50	0.90	1.00	1.03
5% Nominal Size									
Intercept included in model (M1)									
25	SC^2	14.52	14.44	13.92	14.40	15.82	19.28	19.86	20.21
	SC^{*2}	5.30	5.26	5.01	5.24	5.78	7.28	7.59	7.75
50	SC^2	12.80	12.72	12.32	12.60	13.50	16.13	16.97	17.43
	SC^{*2}	5.17	5.15	4.92	5.05	5.38	6.52	7.00	7.27
100	SC^2	10.43	10.41	10.15	10.36	10.73	12.34	13.27	13.85
	SC^{*2}	5.09	5.05	4.95	5.00	5.08	5.82	6.38	6.74
No intercept included in model (M2)									
25	SC^2	15.28	15.23	14.51	14.49	14.62	15.11	15.33	15.39
	SC^{*2}	5.49	5.46	5.20	5.19	5.23	5.44	5.49	5.53
50	SC^2	13.17	13.12	12.72	12.72	12.71	13.01	13.20	13.21
	SC^{*2}	5.33	5.29	5.06	5.03	5.09	5.20	5.26	5.27
100	SC^2	10.64	10.60	10.27	10.31	10.34	10.42	10.62	10.63
	SC^{*2}	5.17	5.12	5.02	4.99	5.01	5.03	5.13	5.16
1% Nominal Size									
No intercept included in model (M1)									
25	SC^2	6.30	6.29	5.98	6.24	7.06	8.87	9.16	9.33
	SC^{*2}	1.09	1.10	1.04	1.07	1.24	1.64	1.75	1.79
50	SC^2	4.77	4.77	4.56	4.70	5.16	6.44	6.83	7.04
	SC^{*2}	1.08	1.08	1.02	1.06	1.15	1.45	1.57	1.60
100	SC^2	3.36	3.36	3.24	3.31	3.47	4.22	4.59	4.74
	SC^{*2}	1.05	1.05	1.02	1.02	1.04	1.21	1.36	1.43
No intercept included in model (M2)									
25	SC^2	6.71	6.69	6.25	6.22	6.34	6.59	6.71	6.73
	SC^{*2}	1.16	1.15	1.06	1.04	1.06	1.12	1.18	1.19
50	SC^2	4.98	4.97	4.70	4.69	4.78	4.90	4.95	4.95
	SC^{*2}	1.12	1.11	1.04	1.05	1.05	1.07	1.10	1.09
100	SC^2	3.41	3.37	3.24	3.24	3.25	3.33	3.39	3.40
	SC^{*2}	1.06	1.05	1.02	1.01	1.00	1.00	1.03	1.04

The second experiment evaluates the sensitivity of the SC^{*2} test to failures of Assumption 5, in particular the non-normality of the errors. Table 2 presents simulated sizes for both the SC^{*2} and sup F tests under a range of error distributions. The former is very sensitive to departures from normality, while the latter is not. In the second part of the table we consider a further scenario, in which a model builder runs the structural instability tests only if a test for normal residuals is not rejected. This yields

three additional tests, the normality test Φ and the $SC^{*2}|\Phi$ and $\sup F|\Phi$ tests, each conditional on the normality hypothesis having not been rejected. We also consider joint tests $SC^{*2} + \Phi$ and $\sup F + \Phi$ that first tests normality and then tests for break if normality cannot be rejected. As the table illustrates, some size distortion remains, but the inflation of the unconditional test is largely controlled in the conditional case. As noted in Section 8 below, we recommend using the test in this way if the normality of the errors cannot be safely assumed.

Table 2. Simulated rejection frequency for SC^{*2} and $\sup F$, possibly combined with normality test Φ , under autoregression in Equation (29) with various error distributions. 50,000 repetitions, $MCSE \leq 0.25$.

<i>T</i>		Error Distribution				
		Φ	t_{50}	t_{10}	t_5	$\chi_{3,cent.}^2$
Unconditional tests						
50	SC^{*2}	5.0	6.6	15.0	28.8	40.9
	$\sup F$	6.0	6.0	6.0	5.9	6.3
100	SC^{*2}	5.0	7.4	22.2	45.0	59.9
	$\sup F$	4.9	5.0	4.9	4.7	4.9
Joint tests						
50	Φ	4.9	6.7	19.0	41.0	95.1
	$SC^{*2} \Phi$	3.4	3.9	6.2	8.0	* 7.4
	$\sup F \Phi$	6.0	6.0	6.1	6.0	* 6.7
	$SC^{*2} + \Phi$	8.1	10.4	24.0	45.8	* 95.5
	$\sup F + \Phi$	10.6	12.3	23.9	44.6	* 95.4
100	Φ	4.8	7.6	28.7	63.0	100.0
	$SC^{*2} \Phi$	3.3	4.2	7.2	8.6	
	$\sup F \Phi$	5.0	5.0	5.0	4.9	
	$SC^{*2} + \Phi$	8.0	11.5	33.9	66.2	* 100.0
	$\sup F + \Phi$	9.5	12.2	32.3	64.8	* 100.0

The third experiment considers the power of the tests against a single shift in the mean level of the process. The data generating process is:

$$x_t = \gamma I_{t>\tau} + \varepsilon_t, \quad t = 1, \dots, T, \tag{30}$$

$$x_0 = 0.$$

and we allow γ and τ to vary as presented. The regression model remains a first order autoregression with an intercept. The level shift is therefore not modelled. Table 3 shows simulated sizes for the unconditional tests as in the previous experiment. We note that the $\sup F$ test performs well for a break at mid-sample, but is outperformed by the SC^{*2} test for breaks occurring near the end of the sample. We also consider conditional tests as in the previous experiment. There are two main observations: firstly,

the normality test is increasingly likely to reject as the break magnitude becomes large; but secondly, the SC*2 test still has power (attenuated by around one-half) to detect the break in this case.

Table 3. Simulated rejection frequency for SC*2 and sup F, possibly combined with normality test Φ, under the process in Equation (30) with a break of magnitude γ at time τ. 50,000 repetitions, MCSE ≤ 0.5.

T		Break Timing (τ)						
		0.5T		T-2		T-1		
		Post-Break Constant (γ)						
		0.0	2.0	4.0	2.0	4.0	2.0	4.0
Unconditional tests								
25	SC*2	5.5	13.7	51.2	21.0	78.7	16.0	70.3
	sup F	10.3	90.4	99.9	19.9	44.8	14.1	28.1
50	SC*2	5.2	17.1	67.6	18.6	83.2	13.3	69.9
	sup F	6.0	99.8	100.0	10.2	34.5	7.1	11.5
100	SC*2	5.1	20.0	77.9	16.1	84.2	11.5	67.6
	sup F	4.9	100.0	100.0	7.2	31.2	5.4	7.4
Joint tests								
25	Φ	5.1	4.3	15.6	9.6	37.9	9.9	53.9
	SC*2 Φ	3.8	12.5	45.3	14.8	66.7	9.4	38.3
	sup F Φ	10.2	90.4	99.9	19.6	52.1	13.1	22.6
	SC*2 + Φ	8.7	16.2	53.8	22.9	79.4	18.4	71.5
	sup F + Φ	14.8	90.8	99.9	27.3	70.3	21.7	64.3
50	Φ	4.8	4.4	17.0	11.1	61.9	9.5	58.6
	SC*2 Φ	3.5	15.7	62.6	11.3	58.6	7.2	31.1
	sup F Φ	6.0	99.8	100.0	9.8	33.9	6.9	9.0
	SC*2 + Φ	8.2	19.4	69.0	21.1	84.2	16.1	71.5
	sup F + Φ	10.6	99.9	100.0	19.8	74.8	15.8	62.3
100	Φ	4.7	4.2	15.6	11.2	74.7	8.5	57.4
	SC*2 Φ	3.4	18.7	74.7	8.9	43.5	6.4	28.4
	sup F Φ	4.9	100.0	100.0	6.8	19.3	5.4	6.3
	SC*2 + Φ	8.0	22.1	78.7	19.1	85.7	14.4	69.5
	sup F + Φ	9.4	100.0	100.0	17.2	79.6	13.5	60.0

The fourth experiment (Table 4) considers the power of the tests against a single innovation outlier at the process mid-point. The data generating process is:

$$\begin{aligned}
 x_t &= \alpha x_{t-1} + \delta I_{t=T/2+1} + \varepsilon_t, \quad t = 1, \dots, T, \\
 x_0 &= 0.
 \end{aligned}
 \tag{31}$$

and we allow α and δ to vary. Both tests presented have similar power in most circumstances, with an outlier larger than three-times the error standard deviation being detected with useful frequency. The OLS-based sup t² test has slightly better power than the Chow test. The conditional evaluations show that both tests retain power in situations where the normality test is not rejected.

Table 4. Simulated rejection frequency for SC^{*2} and $\sup F$, possibly combined with normality test Φ , under the process in Equation (31) with a break of magnitude δ . 50,000 repetitions, $MCSE \leq 0.5$.

α	T		Outlier Magnitude (δ)					
			0.0	1.0	2.0	3.0	4.0	5.0
Unconditional tests								
0.5	50	SC^{*2}	5.5	5.8	10.6	28.7	58.7	84.9
		$\sup t^2$	4.7	5.0	10.5	31.8	66.2	90.7
	100	SC^{*2}	5.2	5.5	9.8	28.1	61.1	87.9
		$\sup t^2$	4.8	5.1	9.8	29.9	65.1	90.8
0.9	50	SC^{*2}	6.6	7.0	12.0	30.2	59.7	85.3
		$\sup t^2$	4.7	4.9	10.3	31.3	65.5	90.3
	100	SC^{*2}	5.9	6.3	10.8	28.9	61.4	87.9
		$\sup t^2$	4.9	5.1	9.8	29.7	64.9	90.6
Joint tests								
0.5	50	Φ	4.9	5.0	9.6	27.7	59.3	86.6
		$SC^{*2} \Phi$	3.8	4.0	5.8	10.8	18.0	25.6
		$\sup t^2 \Phi$	2.2	2.4	4.0	10.3	22.7	36.8
		$SC^{*2} + \Phi$	8.5	8.8	14.8	35.5	66.6	90.0
		$\sup t^2 + \Phi$	7.0	7.3	13.2	35.2	68.5	91.5
	100	Φ	4.8	5.0	8.7	25.0	57.1	86.0
		$SC^{*2} \Phi$	3.5	3.7	5.2	11.0	20.9	32.0
		$\sup t^2 \Phi$	2.7	2.8	4.4	10.9	24.3	40.0
		$SC^{*2} + \Phi$	8.1	8.5	13.4	33.2	66.1	90.5
		$\sup t^2 + \Phi$	7.3	7.7	12.7	33.2	67.5	91.6
0.9	50	Φ	4.9	5.1	9.4	27.1	58.6	85.9
		$SC^{*2} \Phi$	4.9	5.2	7.3	13.1	21.6	30.6
		$\sup t^2 \Phi$	2.2	2.3	4.0	10.3	22.6	37.5
		$SC^{*2} + \Phi$	9.5	10.0	16.0	36.6	67.5	90.2
		$\sup t^2 + \Phi$	6.9	7.2	13.0	34.6	68.0	91.2
100	100	Φ	4.7	5.0	8.6	24.7	56.8	85.8
		$SC^{*2} \Phi$	4.2	4.4	6.2	12.2	22.2	33.6
		$\sup t^2 \Phi$	2.7	2.8	4.4	10.9	24.2	39.5
		$SC^{*2} + \Phi$	8.7	9.2	14.3	33.8	66.4	90.6
		$\sup t^2 + \Phi$	7.3	7.6	12.6	32.9	67.2	91.4

6.2. Autoregressive Distributed Lag Data-Generating Process

We consider a bivariate data-generating process, written in triangular equilibrium correction form as

$$\Delta y_t = \frac{1}{4} \Delta z_t - \frac{1}{4} (y_{t-1} - z_{t-1}) + \nu + \varepsilon_t, \tag{32}$$

$$\Delta z_t = (\psi - 1)z_{t-1} + \eta_t \tag{33}$$

$$y_0 = z_0 = 0,$$

where $\varepsilon_t, \eta_t \stackrel{iid}{\sim} N_2(0, I_2)$. The characteristic roots of the system are ψ and $3/4$. When $\psi = 1$, the model is cointegrated. When $\psi = 1/4$, the model is stationary. In both cases, $y_t - z_t$ is stationary.

We then fit the univariate autoregressive distributed lag model

$$y_t = \rho z_t + \alpha y_{t-1} + \beta z_{t-1} + \nu + \varepsilon_t, \tag{34}$$

and investigate the residuals $\hat{\varepsilon}_t$ using the Chow statistics.

The first experiment (Table 5) evaluates the size of the Chow tests. Here, ψ varies, while $\nu = 0$ in the data generating process. The results are in line with those seen for the autoregressive situation in Table 1. The second experiment is not done in this situation.

Table 5. Simulated rejection frequency for SC^2, SC^{*2} under an autoregressive distributed lag process in Equations (32), (33). 50,000 repetitions, $MCSE \leq 0.2$.

<i>T</i>		Autoregressive Coefficient (ψ)	
		0.25	1.00
50	SC^2	15.0	16.3
	SC^{*2}	6.1	6.7
100	SC^2	11.6	12.2
	SC^{*2}	5.4	5.8

The third experiment (Table 6) evaluates the power of the Chow tests against a single shift in the mean level. This is done by replacing ν by $\nu I_{t>\tau}$ in the data generating process Equation (32). The results are in line with those seen for the autoregressive situation in Table 3: There is good size control. The power is nearly uniform in ψ and comparable to the power reported in Table 3.

The fourth experiment (Table 7) evaluates the power of the Chow tests against a single innovation outlier at the process mid-point. This is done by replacing ν by $\nu I_{t=T/2+1}$ in the data generating process Equation (32). The results are in line with those seen for the autoregressive situation in Table 4.

Table 6. Simulated rejection frequency for SC*² under process in Equations (32), (33) with a break of magnitude ν at time τ . $T = 50$. 50,000 repetitions, MCSE ≤ 0.5 .

ψ		Break Timing(τ)						
		0.5T		T-2		T-1		
		Post-Break Constant (ν)						
		0.0	2.0	4.0	2.0	4.0	2.0	4.0
0.25	SC* ²	6.1	16.3	64.6	19.7	83.2	13.8	68.2
	Φ	4.9	6.8	34.1	10.9	61.8	9.1	54.9
	SC* ² Φ	4.5	13.4	51.2	12.7	58.8	8.4	33.1
	SC* ² + Φ	9.2	19.3	67.8	22.2	84.3	16.7	69.8
1.00	SC* ²	6.7	16.0	62.0	19.4	81.2	14.2	67.2
	Φ	5.0	6.0	25.1	9.7	54.5	8.7	51.8
	SC* ² Φ	5.3	13.7	52.6	13.5	60.8	9.1	35.2
	SC* ² + Φ	10.0	18.9	64.5	21.9	82.2	17.0	68.8

Table 7. Simulated rejection frequency for SC*² under process in Equations (32), (33) with an outlier of magnitude ν at mid-sample. $T = 50$. 50,000 repetitions, MCSE ≤ 0.5 .

ψ		Outlier Magnitude (ν)					
		0.0	1.0	2.0	3.0	4.0	5.0
0.25	SC* ²	6.1	6.3	10.4	25.5	53.0	79.1
	Φ	4.9	5.2	9.1	25.4	55.4	83.1
	SC* ² Φ	4.5	4.6	6.3	10.8	17.5	24.8
	SC* ² + Φ	9.2	9.6	14.8	33.5	63.2	87.3
1.00	SC* ²	6.7	7.1	11.0	25.4	51.5	77.4
	Φ	5.0	5.3	9.3	25.9	55.9	83.5
	SC* ² Φ	5.3	5.5	7.1	11.7	18.5	26.0
	SC* ² + Φ	10.0	10.5	15.7	34.6	64.1	87.8

7. Empirical Illustration

As an empirical illustration, consider log quarterly U.K. gross domestic product¹, y , say, for the period 1991:1 to 2013:3. This gives a total sample length of 91, of which two observations are held back as initial values. The data is provided as supplementary material. Figure 1a shows the series.

An autoregression with two lags, an intercept and a linear trend is fitted to the data:

$$\hat{y}_t = 1.65y_{t-1} - 0.66y_{t-2} + 0.17 + 0.000045t \tag{35}$$

(se) (0.08) (0.08) (0.14) (0.000075)

$$\hat{\sigma} = 0.0050, \quad \text{log-likelihood} = 347.439,$$

$$\chi^2_{norm}[2] = 4.3 (p = 0.12), \quad F_{ar(1-5)}[5, 80] = 1.2 (p = 0.32), \tag{36}$$

$$\max C^{2*} = 12.9 (p = 0.026) \{ \arg \max = 2008 : 2 \} \tag{37}$$

¹ ABMI series from Office of National Statistics, seasonally adjusted, 2010 prices, release 20 December 2013.

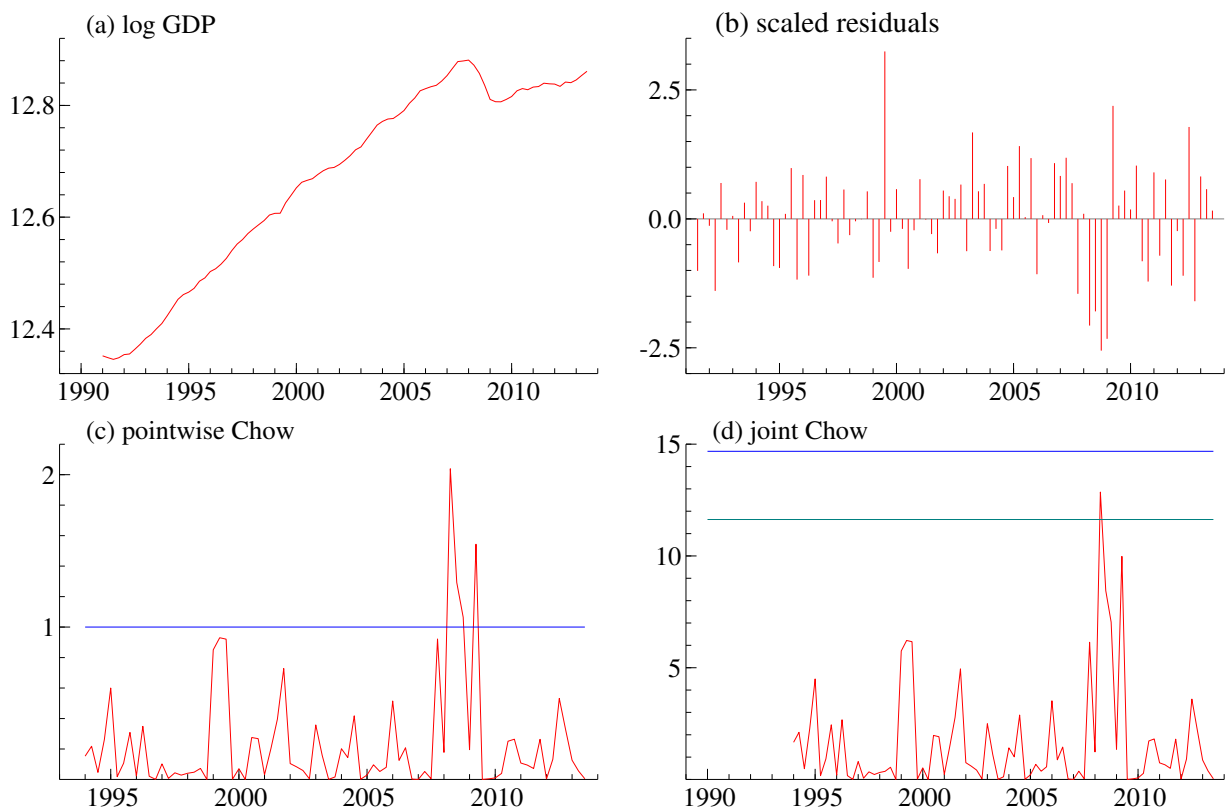


Figure 1. U.K. GDP series. **(a)** Series in logs; **(b)** scaled residuals; **(c)** pointwise one-step Chow tests; the horizontal line is the 1% critical value; **(d)** simultaneous Chow test; the horizontal lines are the 1% (top) and 5% (bottom) critical values.

The specification tests in Equation (36) are a cumulant-based test for normality and a test for autocorrelated residuals. They are valid for non-stationary autoregressive distributed lag models; see [29,35,36]. These specification tests do not indicate particular problems with the model. The specification test in Equation (37) is the supremum Chow test $C_{1,t}^{2*}$ from Equation (27) evaluated according to the finite sample approximation Equation (28). This indicates a specification problem that is worst for 2008:2. A graphical version of this test is discussed below.

Figure 1b shows the scaled residuals. There appears to be clustering of first positive and then negative residuals around the start of the financial crisis in 2006–2009. This tendency is, however, not sufficient to trigger the specification tests for autocorrelation and non-normality in Equation (36).

Figure 1c shows pointwise one-step Chow tests. This is the standard output from PcGive. The scale is a probability scale, where the horizontal line indicates the critical value for pointwise tests at a 1% level. The exact nature of the probability scale is not documented in [37]. This plot represents pointwise tests based on 79 statistics with an unclear correlation structure. Thus, researchers have traditionally interpreted this plot with a grain of salt; see ([38] p.197). In this plot, there is a cluster of pointwise rejections, so it would seem prudent to question the stability of the model.

Figure 1d shows the Chow statistics $C_{1,t}^{2*}$ (see Equation (27)), along with horizontal lines indicating the critical values of simultaneous 5% and 1% tests. These are computed according to the finite sample approximation Equation (28). The maximum of the test statistics is 12.9; see Equation (37). This is

between the 5% and the 1% critical levels of 11.6 and 14.7, respectively. Thus, with a 5% level, it would seem prudent to question the stability of the model.

What could be done to remedy the situation? The usual answer is to inspect the data, think about the economic context and the economic question of interest and update the empirical model accordingly. While the economic question is vague in this illustration, the context of the specification problem is the financial crisis. In 2008, there is downward shift both in productivity and in growth rates. How big these effects were remains unclear, and so, the Office of National Statistics keeps revising the data for this period; see [39,40]. Now, one way to capture the downward shift in productive and growth rates is to use an impulse indicator and a step indicator. This gives the model:

$$\widehat{y}_t = \underset{(se)}{1.33}y_{t-1} - \underset{(0.08)}{0.42}y_{t-2} + \underset{(0.24)}{1.17} + \underset{(0.00016)}{0.00077}t - \underset{(0.0045)}{0.015}1_{t=2008:2} - \underset{(0.0038)}{0.019}1_{t \geq 2008:3} \quad (38)$$

$$\widehat{\sigma} = 0.0043, \quad \log\text{-likelihood} = 361.397,$$

$$\chi_{norm}^2[2] = 2.7 (p = 0.25), \quad F_{ar(1-5)}[5, 78] = 0.5 (p = 0.79), \quad (39)$$

$$\max C^{2*} = 8.9 (p = 0.198) \{\arg \max = 2009 : 2\}. \quad (40)$$

The formal asymptotic theory does not cover this situation with dummies. We proceed with these results nonetheless. The specification tests (39), (40) for the new model (38) are clearly better than the tests (37), (37) for the original model (35). An asymptotic theory for this type of iterative specification of a regression model is given in [41].

It is worth noting how the plot Figure 1d is computed. First, $C_{1,t}^2$ is computed using Equation (5) for $t = 11, \dots, 89$, so that $g(T) = 10$ and $T = 89$. This can be done easily using standard regression techniques. For instance, to compute $C_{1,11}^2$, run a regression including over the sample $1, \dots, 11$, including a dummy for Observation 11. Then, $C_{1,11}^2$ is the F -statistic for testing the absence of the dummy variable. Secondly, $C_{1,11}^{2*}$ is computed from $C_{1,11}^2$ by first transforming using the $F(1, t - k - 1)$ distribution function, with $t = 11$ and $k = 4$ in this case, and then by the inverse of G , the χ_1^2 distribution function. The critical values are computed according to the finite sample approximation (28). For instance, for $p = 0.05$, then $x = G^{-1}\{(1 - p)^{1/[T-g(T)]}\} = 11.72$, so that $[G(x)]^{T-g(T)} = 1 - p$.

8. Conclusions

We advocate the sup-Chow test as a general misspecification test to be used as part of an iterative modelling strategy. It is a relatively simple transformation of the existing one-step Chow test (or, similarly, the EViews one-step forecast test), with a standard and easily calculated null distribution, which does not vary substantially in the AR(1) parameter space. We anticipate that it would be used as one of a battery of tests (including the normality of residuals); rejection would draw the modeller's attention to the pointwise plot, which would help identify the cause and timing of the failure.

By construction, the test is sensitive to parameter changes and outliers and is somewhat agnostic about the timing and number of these breaks. This makes it useful against a variety of simple and complex misspecification types. However, there is a clear trade-off, and as the first columns of Table 3 show, the test is less powerful against a particular alternative than the [13] sup-F test, which explicitly models a single break. This motivates the use of multiple different tests, with the failure of any one signalling misspecification and triggering further investigation. In real datasets, breaks may not be of the single

mean-shift variety, and the parallel popularity of CUSUM-type and Andrews-type tests suggests that both approaches have value.

The test is not invariant to the error distribution, even asymptotically; a feature it shares with most outlier tests and end-sample tests of parameter change. There are two different solutions to this, depending on the modelling approach being used.

If normality is assumed and tested, there is no problem. As Table 2 shows, the pre-test for normality affects the size and power of the test, but substantial power remains following the pre-test: that is, the sup-Chow test has power somewhat orthogonal to a common normality test.

If normality is not maintained, the sup-Chow test as presented cannot be used. The solution to this problem would involve a subsampling technique to recover the distribution of the errors from the known good part of the sample. This is the approach taken in [15], but it necessarily complicates the test. An appealing alternative is offered in [42] in the context of another test applying extreme value theory to outlier detection. The authors resolve the distribution dependency in three stages. First, extreme value theory itself means that a wide class of error distributions converge in the maximum to the same extreme value distribution: the Gumbel, discussed in Equation (20) above. Second, the centring factor required for extreme value convergence is eliminated by examining differences between order statistics, applying a convenient theorem of [43]. Third, the scaling factor is implicitly sampled by examining the largest $(1/\alpha)$ pointwise statistics for a level α test. Preliminary experiments suggest that this approach can be applied effectively to the sup-Chow test, but as in [42], a further correction is needed to control the size of the test; hence, this remains future work.

Acknowledgements

The comments from the referees are gratefully acknowledged.

Author Contributions

The authors made equal contributions.

A. Proofs

A.1. Notation

We use the spectral matrix norm, so $\|M\| = [\lambda_{\max}(MM')]^{1/2}$, where $\lambda_{\max}(MM')$ indicates the largest eigenvalue of MM' . For a symmetric, non-negative definite matrix M , then $M^{1/2}$ refers to a matrix that is multiplied with its transpose to give M . The choice of the matrix square root does not matter, due to the choice of the matrix norm and the properties of the eigenvalues of the products. 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. Three-Way Process Decomposition

We elaborate on the decomposition of the companion form Equation (15) given in Equation (16). Whereas, there, it was decomposed into non-explosive and explosive components, we now further decompose the non-explosive components into stationary and unit-root components. As before, there exists a regular real matrix M_3 to block diagonalize S into stationary, unit-root and explosive components:

$$M_3 S_t = (M_3 S M_3^{-1}) M_3 S_{t-1} + M_3 (\xi'_t, 0')',$$

$$\begin{pmatrix} \tilde{U}_t \\ Q_t \\ W_t \end{pmatrix} = \begin{pmatrix} U & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & W \end{pmatrix} \begin{pmatrix} \tilde{U}_{t-1} \\ Q_{t-1} \\ W_{t-1} \end{pmatrix} + \begin{pmatrix} e_{\tilde{U},t} \\ e_{Q,t} \\ e_{W,t} \end{pmatrix}, \tag{41}$$

where \tilde{U} , Q and W have eigenvalues inside, on and outside the unit circle, respectively. We can now express the two-way decomposition presented in Equation (16) as follows:

$$R_t = \begin{pmatrix} \tilde{U}_t \\ Q_t \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} U & 0 \\ 0 & Q \end{pmatrix}. \tag{42}$$

A.3. Preliminary Asymptotic Results

The ADL model Equation (10) becomes:

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

where θ is the vector of coefficients. Then, from Equation (11), 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 a result of the decomposition above in Equation (16), on $x_t = (\xi'_{2,t}, R'_{t-1}, W'_{t-1})'$; so, we can analyse the test statistic Equation (6) as if these were the actual regressors.

Lemma 12. *Suppose Assumptions 1, 2 and 3 hold with $\alpha > 4$ only. Then, for all $\beta > 1/\alpha$ and $\zeta < 1/8$,*

- (i) $C_{RW} \stackrel{\text{a.s.}}{=} o(t^{-\zeta/2})$,
- (ii) $C_{\xi S} \stackrel{\text{a.s.}}{=} o(t^{\beta-1/2})$,
- (iii) $S_{RR \cdot W}^{-1} \stackrel{\text{a.s.}}{=} S_{RR}^{-1/2} \cdot \{1 + o(1)\} \cdot S_{RR}^{-1/2}$,
- (iv) $S_{\xi_2 \xi_2 \cdot S}^{-1} \stackrel{\text{a.s.}}{=} S_{\xi_2 \xi_2}^{-1/2} \cdot \{1 + o(1)\} \cdot S_{\xi_2 \xi_2}^{-1/2}$,
- (v) $S_{RR}^{-1/2} R_{t-1} \stackrel{\text{a.s.}}{=} o(t^{-\zeta/2})$,
- (vi) $S_{WW}^{-1/2} W_{t-1} \stackrel{\text{a.s.}}{=} O(1)$,
- (vii) $S_{RR}^{-1/2} (R|W)_t \stackrel{\text{a.s.}}{=} o(t^{-\zeta/2})$, and
- (viii) $S_{\xi_2 \xi_2}^{-1/2} (\xi_2|S)_t \stackrel{\text{a.s.}}{=} o(t^{\beta-1/2})$.

Proof. Result (i) is proven by decomposing the correlation to apply results from [25], so that:

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

where the last line follows, because $C_{\tilde{U}Q}$ is vanishing almost surely by [25], Theorem 9.4. Then the result follows, since $C_{\tilde{U}W} \stackrel{\text{a.s.}}{=} o(t^{\beta-1/2})$ and $C_{QW} \stackrel{\text{a.s.}}{=} o(t^{-\zeta/2})$ by [25], Theorems 9.1 and 9.2. The latter term will dominate since $\alpha > 16/7$ under Assumption 2.

Result (ii) is proven by noting that $\|C_{\xi S}\| \leq \|S_{\xi\xi}^{-1/2}\| \|S_{\xi S} S_{SS}^{-1/2}\|$, with the first normed term $O(t^{-1/2})$ by [25], Theorem 2.8, and the second $o(t^\beta)$ by [25], 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 analogous, but substitutes (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 [25], 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 [25], Theorem 8.3, we have $S_{\tilde{U}\tilde{U}}^{-1/2} = O(t^{-1/2})$ and by [44], Theorem 1(i), we have that $\tilde{U}_{t-1} = o(t^\beta)$, so $S_{\tilde{U}\tilde{U}}^{-1/2} \tilde{U}_{t-1} = o(t^{\beta-1/2})$.

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 [25], Theorem 8.4, which states that for some η , $\max_{t^\eta \leq s < t} Q'_s (\sum_{s=1}^t Q_{s-1} Q'_{s-1})^{-1} Q_s = o(t^{-\zeta})$ for all $\zeta < 1/8$, and so, *a fortiori*, $Q'_{t-1} (\sum_{s=1}^t Q_{s-1} Q'_{s-1})^{-1} Q_{t-1} = o(t^{-\zeta})$. However, 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 $b'(\mathbf{A} + bb')^{-1}b = b'\mathbf{A}^{-1}b(1 + b'\mathbf{A}^{-1}b)^{-1}$ ([45] p. 151) to write:

$$Q'_{t-1} S_{QQ}^{-1} Q_{t-1} = \frac{Q'_{t-1} (\sum_{s=1}^t Q_{s-1} Q'_{s-1})^{-1} Q_{t-1}}{1 - Q'_{t-1} (\sum_{s=1}^t Q_{s-1} Q'_{s-1})^{-1} Q_{t-1}},$$

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

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

Result (vi) follows directly from [44], 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 $o(t^{-\zeta/2})$ by (v), as is the second by (i), while the third is bounded by (vi).

Result (viii) is proven in a similar fashion. Write:

$$\begin{aligned} \|S_{\xi_2\xi_2}^{-1/2}(\xi_2|S)_t\| &= \|S_{\xi_2\xi_2}^{-1/2}\xi_{2,t} - S_{\xi_2\xi_2}^{-1/2}S_{\xi_2S}S_{SS}^{-1}S_{t-1}\| \\ &= \|S_{\xi_2\xi_2}^{-1/2}\xi_{2,t}\| - \|C_{\xi_2S}\| \|S_{SS}^{-1/2}S_{t-1}\|. \end{aligned}$$

Then, the first of the normed quantities is $o(t^{\beta-1/2})$ by [25], Theorem 2.8, and the result is $\xi_t = o(t^\beta)$ by [44], Theorem 1; the second is $O(t^{\beta-1/2})$ by (ii); and the third is $O(1)$, since we use a partial regression transformation to write:

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

and then apply (iii) and (vii) and (vi), respectively. \square

Lemma 13. Under Assumptions 1, 2 and 3 with $\alpha > 4$ and with $\beta > 1/\alpha$,

- (i) $\sum_{s=1}^{t-1} \varepsilon_s S'_{s-1} S_{SS}^{-1/2} \stackrel{\text{a.s.}}{=} o(t^\beta)$,
- (ii) $\sum_{s=1}^{t-1} \varepsilon_s R'_{s-1} S_{RR}^{-1/2} \stackrel{\text{a.s.}}{=} O[(\log t)^{1/2}]$,
- (iii) $\sum_{s=1}^{t-1} \varepsilon_s W'_{s-1} S_{WW}^{-1/2} \stackrel{\text{a.s.}}{=} o(t^\beta)$,
- (iv) $\sum_{s=1}^{t-1} \varepsilon_s (R|W)'_s S_{RR}^{-1/2} \stackrel{\text{a.s.}}{=} O[(\log t)^{1/2}] + o(t^{\beta-1/16})$.

Proof. For (i), (ii) and (iii), use [25], Theorem 2.4. For (iv), write:

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

and then apply (ii), (iii) and Lemma 12 (i). \square

Lemma 14. Under Assumptions 1, 2, 3 and 4,

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

Proof. For (i), use [46], Lemma 1(iii), and [44], Corollary 1(iii). For (ii), write:

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

and then apply (i), Lemma 13 (i) and Lemma 12 (ii). \square

A.4. Proof of Theorem 6

We proceed to examine the behaviour of $\tilde{\varepsilon}_t$, the one-step forecast residuals. From Equation (6), we can write:

$$\tilde{\varepsilon}_t = \frac{\varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s x'_s \left(\sum_{s=1}^{t-1} x_s x'_s\right)^{-1} x_t}{\left[1 + x'_t \left(\sum_{s=1}^{t-1} x_s x'_s\right)^{-1} x_t\right]^{1/2}}. \tag{43}$$

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

Lemma 15. Under Assumptions 1, 2 and 3,

$$x'_t S_{xx}^{-1} x_t - W' F_W^{-1} W = o(t^{-\zeta}) \text{ a.s.} \tag{44}$$

for all $\zeta < 1/8$ with W and F_W as in Theorem 6.

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:

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

The first normed term on the right-hand side is $o(t^{2\beta-1})$, and the second is $o(t^{-\zeta})$ by Lemma 12, Parts (iv) and (viii); and (iii) and (vii), respectively. The second term will dominate, since $\alpha > 16/7$, so $\|x'_t S_{xx}^{-1} x_t - W'_{t-1} S_{WW}^{-1} W_{t-1}\| = o(t^{-\zeta})$.

The lemma is then proven by rewriting the second step:

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

and noting that, by [25], Corollary 5.3(i), $\mathbf{W}^{t-1} W_{t-1} - W = O(\lambda_{\min}(\mathbf{W})^{-t})$ and, by [25], Corollary 7.2, $(\mathbf{W}^{t-1})' S_{WW}^{-1} \mathbf{W}^{t-1} - F_W^{-1} = O(\lambda_{\min}(\mathbf{W})^{-2t})$, almost surely, while all of the other terms are bounded by the same corollaries. \square

We next state a lemma concerning the main numerator term in Equation (43).

Lemma 16. Under Assumptions 1, 2, 3 and 4:

$$\sum_{s=1}^{t-1} \varepsilon_s x'_s S_{xx}^{-1} x_t - G_t F_W^{-1} W = o(t^{\beta-1/8}) \text{ a.s.} \tag{45}$$

for all $\beta > 1/\alpha$, where W and F_W are defined as in Theorem 6, and:

$$G_t = \sum_{s=1}^{t-1} \varepsilon_{t-s} W' (\mathbf{W}^{-s})' = o(t^\beta) \text{ a.s.} \tag{46}$$

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

$$\begin{aligned} & \left\| \sum_{s=1}^{t-1} \varepsilon_s x'_s S_{xx}^{-1} x_t - G_t F_W^{-1} W \right\| \\ & \leq \left\| \sum_{s=1}^{t-1} \varepsilon_s x'_s S_{xx}^{-1} x_t - \sum_{s=1}^{t-1} \varepsilon_s W'_{s-1} S_{WW}^{-1} W_{t-1} \right\| + \left\| \sum_{s=1}^{t-1} \varepsilon_s W'_{s-1} S_{WW}^{-1} W_{t-1} - G_t F_W^{-1} W \right\|. \end{aligned}$$

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

$$\begin{aligned} \left\| \sum_{s=1}^{t-1} \varepsilon_s x'_s S_{xx}^{-1} x_t - \sum_{s=1}^{t-1} \varepsilon_s W'_{s-1} S_{WW}^{-1} W_{t-1} \right\| & \leq \left\| \sum_{s=1}^{t-1} \varepsilon_s (\xi_2 | R, W)'_s S_{\xi_2 \xi_2 \cdot RW}^{-1} (\xi_2 | R, W)'_t \right\| \\ & + \left\| \sum_{s=1}^{t-1} \varepsilon_s (R|W)'_s S_{RR \cdot W}^{-1} (R|W)_t \right\|, \end{aligned} \tag{47}$$

and we consider each term on the right separately. For the first term in Equation (47), use Lemma 12 (iv) to write:

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

and then apply Lemma 14 (ii) and Lemma 12 (viii) to arrive at $o(t^{3\beta-1}) + o[t^{\beta-1/2}(\log t)^{1/2}]$ almost surely. For the second term in Equation (47), we use Lemma 12 (iii) to write:

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

and then apply Lemma 13 (iv) and Lemma 12 (vii) to arrive at $o(t^{\beta-1/8})$ almost surely. Overall then, as long as $\alpha > 4$, the first step is dominated by this second term.

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

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

Many of these terms are familiar from the proof of Lemma 15, 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 [44], Theorem 1. For the former term, we have:

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

where at the second to last line, we use that $\sum_{u=0}^{\infty} \mathbf{W}^{-u} e_{W,u+s} = o(s^\beta)$ by [28], 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 [44], Theorem 1. \square

Proof of Theorem 6. We aim to show that:

$$C_{1,t}^2 - (q_t/\sigma)^2 \stackrel{\text{a.s.}}{=} o(1). \tag{48}$$

Using Equation (6), we can rewrite this expression as:

$$\frac{\tilde{\varepsilon}_t^2}{(t-k-1)^{-1} \text{RSS}_{t-1}} - \left(\frac{q_t}{\sigma}\right)^2 = \tilde{\varepsilon}_t^2 \left[\frac{(t-k-1)}{\text{RSS}_{t-1}} - \frac{1}{\sigma^2} \right] + \frac{\tilde{\varepsilon}_t^2 - q_t^2}{\sigma^2}. \tag{49}$$

We first consider the difference $\tilde{\varepsilon}_t^2 - q_t^2$. We have from Equation (43),

$$\tilde{\varepsilon}_t^2 - q_t^2 = \frac{(\varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s x'_s S_{xx}^{-1} x_t)^2}{1 + x'_t S_{xx}^{-1} x_t} - \frac{(\varepsilon_t - G_t F_W^{-1} W)^2}{1 + W' F_W^{-1} W} = \frac{(\varepsilon_t - A_3)^2}{1 + A_1} - \frac{(\varepsilon_t - A_4)^2}{1 + A_2} \tag{50}$$

$$= \frac{A_2 - A_1}{(1 + A_1)(1 + A_2)} (\varepsilon_t - A_3)^2 + \frac{1}{1 + A_2} (A_4 - A_3) (2\varepsilon_t - A_4 - A_3), \tag{51}$$

where:

$$A_1 = x'_t S_{xx}^{-1} x_t \quad A_2 = W' F_W^{-1} W \quad A_3 = \sum_{s=1}^{t-1} \varepsilon_s x'_s S_{xx}^{-1} x_t \quad A_4 = G_t F_W^{-1} W. \tag{52}$$

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 $o(t^{-\zeta})$ by Lemma 15. The factor $\varepsilon_t - A_3 = \varepsilon_t - A_4 + A_4 - A_3$ is $o(t^\beta)$, since ε_t

and A_4 are both $o(t^\beta)$ by [44], Theorem 1, and Lemma 16, respectively, while $A_4 - A_3$ is $O(t^{\beta-1/8})$ by Lemma 16. Therefore, the first term of the sum is $o(t^{2\beta-1/8})$ almost surely.

In the second numerator, $A_4 - A_3$ is $O(t^{\beta-1/8})$ by Lemma 16, while ε_t and A_4 are each $o(t^\beta)$ as above, so that the whole second term is also $o(t^{2\beta-1/8})$ almost surely.

Thus, the second term in Equation (49) will vanish as long as $2\beta < 1/8$ or $\alpha > 16$ in Assumption 2, 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 proven, while:

$$q_t^2 = \frac{(\varepsilon_t - A_4)^2}{1 + A_2} = o(t^{2\beta}) \text{ a.s.},$$

since, as above, ε_t and A_4 are both $o(t^\beta)$, while A_2 is nonnegative. Then, [25], Corollary 2.9, implies that:

$$\frac{(t - k - 1)}{\text{RSS}_{t-1}} - \frac{1}{\sigma^2} = o(t^\gamma) \text{ a.s.},$$

for $\gamma < 1/2$. Therefore, the first term in Equation (49) vanishes as long as $2\beta < 1/2$, which is satisfied by Assumption 2. \square

A.5. Proof of Lemma 7

Proof. Theorem 6 shows that $C_{1,t}^2 - q_t^2$ vanishes almost surely. Egorov’s theorem ([47], Theorem 18.4) then shows that $C_{1,t}^2 - q_t^2$ vanishes uniformly on a set with large probability. That is,

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

This implies that for any sequence $g(T)$ which increases to infinity, then $\sup_{g(T) < t \leq T} |C_{1,t}^2 - q_t^2|$ vanishes in probability as T increases. \square

A.6. Proof of Lemma 8 (Correction to Lemma 1 of [30])

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 ([48] 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 [49], 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:

$$\Pr\{|X_1| < x\} = F(x) = \sqrt{2/\pi} \int_0^x e^{-t^2/2} dt = 2\Phi(x), \quad x \geq 0 \tag{53}$$

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 [49], Theorem 1.5.1, namely:

$$n(1 - F(u_n)) \rightarrow e^{-x} \quad \text{as } n \rightarrow \infty. \tag{54}$$

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

$$1 - F(u) \sim f(u)/u \quad \text{as } u \rightarrow \infty, \tag{55}$$

so that combining Equations (54) and (55), we have that $(1/n)e^{-x}u_n/f(u_n) \rightarrow 1$. Taking logs and substituting the density f , we have:

$$-\log n - x + \log u_n - \frac{1}{2} \log(\pi/2) + u_n^2/2 \rightarrow 0. \tag{56}$$

Dividing through by $\log n$,

$$-1 - \frac{x}{\log n} + \frac{\log u_n}{\log n} - \frac{\log(\pi/2)}{\log n} + \frac{u_n^2}{2\log n} \rightarrow 0, \tag{57}$$

then, for any fixed x , the second and fourth terms vanish trivially. The third term vanishes by substituting Equation (54) for n and twice applying L'Hôpital's rule. It then follows that $\frac{u_n^2}{2\log n} \rightarrow 1$, or (taking logarithms again),

$$2 \log u_n - \log 2 - \log \log n \rightarrow 0. \tag{58}$$

Substituting this result into Equation (56), we have that:

$$-\log n - x + \frac{1}{2} \log 2 + \frac{1}{2} \log \log n - \frac{1}{2} \log(\pi/2) + u_n^2/2 \rightarrow 0. \tag{59}$$

so that rearranging,

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

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

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

It then follows from [49], Theorem 1.5.3, that $\Pr\{Z_n \leq u_n\} \rightarrow \exp(-e^{-x})$, and rearranging gives the norming sequences. \square

A.7. Proof of Lemma 9

Proof. Consider the normalised linear process:

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

In the case without explosive components, this reduces to:

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

so that under Assumption 5, q_t/σ is an independent standard normal sequence and q_t^2/σ^2 is an independent χ_1^2 sequence. Then, classical extreme value theory gives the lemma with the norming

sequences a_t and b_t as stated (see, for instance, [50] p. 56), 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 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:

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

The general approach to dealing with dependent sequences is outlined in [51]; as long as the dependence is not too great, the same limiting results hold.

We take advantage of the relationship between the χ^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:

$$\max_t q_t^2/\sigma^2 < u_t \quad \text{iff} \quad \max_t |q_t/\sigma| < \sqrt{u_t} \tag{60}$$

where $|q_t/\sigma|$ has the half-normal distribution. Lemma 1 and Theorem 1 of [30] (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:

$$c_t(\max_{1 \leq s \leq t} |q_s/\sigma| - d_t) \xrightarrow{d} \Lambda \tag{61}$$

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 8.) Taking $\sqrt{u_t(z)} = c_t z + d_t$ and using Equations (60) and (61), we have:

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

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 proven by showing that $a_t/a'_t \rightarrow 1$ and $a_t(b'_t - b_t) \rightarrow 0$. \square

A.8. Proof of Theorem 10

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

$$\begin{aligned} & \frac{1}{2} \left[\max_{g(T) \leq t \leq T} (C_{1,t}^2) - d_{T-g(T)} \right] \\ & \leq \frac{1}{2} \max_{g(T) \leq t \leq T} (C_{1,t}^2 - (q_t/\sigma)^2) + \frac{1}{2} \left[\max_{g(T) \leq t \leq T} (q_t/\sigma)^2 - d_{T-g(T)} \right] \end{aligned} \quad (62)$$

where the left term vanishes in probability by Lemma 7 and the right term converges in distribution by Lemma 9. 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 proven. \square

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Chow, G.C. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **1960**, *28*, 591–605.
2. Kimura, T. The impact of financial uncertainties on money demand in Europe. In *Monetary Analysis: Tools and Applications*; Klöckers, H.-J., Willeke, C., Eds.; European Central Bank: Frankfurt am Main, Germany, 2001; pp. 97–116.
3. Hendry, D.F. Using PC-GIVE in econometrics teaching. *Oxf. Bull. Econ. Stat.* **1986**, *48*, 87–98.
4. Quantitative Micro Software (QMS). *EViews 7 User's Guide II*; QMS: Irvine, CA, US, 2009.
5. Brown, R.L.; Durbin, J. Methods of investigating whether a regression relationship is constant over time. In *Selected Statistical Papers, European Meeting*; Mathematical Centre Tracts, 26; Mathematisch Centrum: Amsterdam, the Netherlands, 1968.
6. Brown, R.L.; Durbin, J.; Evans, J.M. Techniques for testing the constancy of regression relationships over time. *J. R. Stat. Soc. Ser. B* **1975**, *37*, 149–192.
7. Krämer, W.; Ploberger, W.; Alt, R. Testing for structural change in dynamic models. *Econometrica* **1988**, *56*, 1355–1369.
8. Ploberger, W.; Krämer, W. On studentizing a test for structural change. *Econ. Lett.* **1986**, *20*, 341–344.
9. Nielsen, B.; Sohkanen, J.S. Asymptotic behavior of the CUSUM of squares test under stochastic and deterministic time trends. *Econom. Theory* **2011**, *27*, 913–927.
10. Galpin, J.S.; Hawkins, D.M. The use of recursive residuals in checking model fit in linear regression. *Am. Stat.* **1984**, *38*, 94–105.
11. Dufour, J.-M. Recursive stability analysis of linear regression relationships: An exploratory methodology. *J. Econom.* **1982**, *19*, 31–76.
12. Perron, P. Dealing with structural breaks. In *Palgrave Handbook of Econometrics*; Mills, T.C., Patterson, K., Eds.; Palgrave Macmillan: Basingstoke, UK, 2006; pp. 278–352.

13. Andrews, D.W.K. Tests for parameter instability and structural change with unknown change point. *Econometrica* **1993**, *61*, 821–856.
14. Quandt, R.E. Tests of the hypothesis that a linear regression system obeys two separate regimes. *J. Am. Stat. Assoc.* **1960**, *55*, 324–330.
15. Andrews, D.W.K. End-of-sample instability tests. *Econometrica* **2003**, *71*, 1661–1694.
16. Srikantan, K.S. Testing for the single outlier in a regression model. *Sankhya Indian J. Stat. Ser. A* **1961**, *23*, 251–260.
17. Chang, I.; Tiao, G.C.; Chen, C. Estimation of time series parameters in the presence of outliers. *Technometrics* **1988**, *31*, 193–204.
18. Chen, C.; Liu, L.M. Forecasting time series with outliers. *J. Forecast.* **1993**, *12*, 13–35.
19. Fox, A.J. Outliers in time series. *J. R. Stat. Soc. Ser. B* **1972**, *34*, 350–363.
20. Barnett, V.; Lewis, T. *Outliers in Statistical Data*, 3rd ed.; Wiley: New York, NY, USA, 1994.
21. Lai, T.L.; Wei, C.-Z. Extended least squares and their applications to adaptive control and prediction in linear systems. *Autom. Control* **1986**, *31*, 898–906.
22. Dufflo, M. *Random Iterative Models*; Springer-Verlag: Berlin, Germany, 1997.
23. Hendry, D.F. *Dynamic Econometrics*; Oxford University Press: Oxford, UK, 1995.
24. Johansen, S. A Bartlett correction factor for tests on the cointegrating relations. *Econom. Theory* **2000**, *16*, 740–778.
25. Nielsen, B. Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms. *Econom. Theory* **2005**, *21*, 534–561.
26. Bauer, D. Almost sure bounds on the estimation error for OLS estimators when the regressors include certain MFI(1) processes. *Econom. Theory* **2009**, *25*, 571–582.
27. Stock, J.H.; Watson, M.W. Has the business cycle changed and why? *NBER Macroecon. Annu.* **2002**, *17*, 159–230.
28. Nielsen, B. *Singular Vector Autoregressions with Deterministic Terms: Strong Consistency and Lag Order Determination*; Nuffield College Discussion Paper; Nuffield College: Oxford, UK, 2008.
29. Engler, E.; Nielsen, B. The empirical process of autoregressive residuals. *Econom. J.* **2009**, *12*, 367–381.
30. Deo, C.M. Some limit theorems for maxima of absolute values of Gaussian sequences. *Sankhyā Indian J. Stat. Ser. A* **1972**, *34*, 289–292.
31. Doornik, J.A. *Object-Oriented Matrix Programming Using Ox*, 3rd ed.; Timberlake Consultants Press: London, UK, 2007.
32. Doornik, J.A.; Hansen, H. An omnibus test for univariate and multivariate normality. *Oxf. Bull. Econ. Stat.* **2008**, *70*, 927–939.
33. Andrews, D.W.K. *Tests for Parameter Instability and Structural Change with Unknown Change Point*; Cowles Foundation Discussion Papers; Cowles Foundation for Research in Economics, Yale University: New Haven, CT, USA, 1990; Volume 943.
34. Cook, R.D.; Weisberg, S. *Residuals and Influence in Regression*; Chapman and Hall: New York, NY, USA, 1982.

35. Kilian, L.; Demiroglu, U. Residual based tests for normality in autoregressions: Asymptotic theory and simulations. *J. Econ. Bus. Control* **2000**, *18*, 40–50.
36. Nielsen, B. Order determination in general vector autoregressions. In *Time Series and Related Topics: In Memory of Ching-Zong Wei*; IMS Lecture Notes and Monograph Series; Ho, H.-C., Ing, C.-K., Lai, T.L., Eds.; Institute of Mathematical Statistics: Beachwood, Ohio, USA, 2006; Volume 52, pp. 93–112.
37. Doornik, J.A.; Hendry, D.F. *Empirical Econometric Modelling—PcGive 14*; Timberlake Consultants: London, UK, 2013; Volume 1.
38. Hendry, D.F.; Nielsen, B. *Econometric Modelling*; Princeton University Press: Princeton, NJ, USA, 2007.
39. Lynch, R.; Richardson, C. Discussion. *J. Off. Stat.* **2004**, *20*, 623–629.
40. Patterson, K.D.; Heravi, S.M. Revisions to official data on U.S. GNP: A multivariate assessment of different vintages (with discussion). *J. Off. Stat.* **2004**, *20*, 573–602.
41. Johansen, S.; Nielsen, B. *Outlier Detection Algorithms for Least Squares Time Series*; Nuffield College Discussion Paper 2014-W04; Nuffield College: Oxford, UK, 2014.
42. Burridge, P.; Taylor, A.M.R. Additive outlier detection via extreme-value theory. *J. Time Ser. Anal.* **2006**, *27*, 685–701.
43. Weissman, I. Estimation of parameters and larger quantiles based on the k largest observations. *J. Am. Stat. Assoc.* **1978**, *73*, 812–815.
44. Lai, T.L.; Wei, C.-Z. Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In *Multivariate Analysis VI*; Krishnaiah, P.R., Ed.; North Holland: Amsterdam, the Netherlands, 1985; pp. 375–393.
45. Searle, S.R. *Matrix Algebra Useful for Statistics*; John Wiley and Sons: New York, NY, USA, 1982.
46. Lai, T.L.; Wei, C.-Z. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Stat.* **1982**, *10*, 154–166.
47. Davidson, J. *Stochastic Limit Theory*; Oxford University Press: Oxford, UK, 1994.
48. Cramér, H. *Mathematical Methods in Statistics*; Princeton University Press: Princeton, NJ, USA, 1946.
49. Leadbetter, M.R.; Lindgren, G.; Rootzén, H. *Extremes and Related Properties of Random Sequences and Processes*; Springer-Verlag: New York, NY, USA, 1982.
50. Embrechts, P.; Klüppelberg, C.; Mikosch, T. *Modelling Extremal Events for Insurance and Finance*; Springer: Berlin, Germany, 1997.
51. Leadbetter, M.R.; Rootzen, H. Extremal theory for stochastic processes. *Ann. Probab.* **1988**, *16*, 431–478.