Asymptotic analysis of the 1-step recursive Chow test (and variants) in time series models

Andrew Whitby

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2013 (as corrected 27 June)

To my parents.

Acknowledgements

My gratitude goes first and foremost to my supervisor Bent Nielsen, who, throughout this process, has shown almost superhuman patience and tolerance. My examiners, Neil Shephard and Anders Rygh Swensen offered invaluable comments and corrections. Then, I must additionally thank (in alphabetical order), Kasper Lund-Jensen, Diaa Noureldin, Cavit Pakel, Laurent Pauwels, Peter Phillips, Jouni Sohkanen, Rob Taylor, three anonymous referees of Econometric Theory and the two anonymous examiners of my M.Phil thesis, all of whom provided useful feedback on my work. Anna Alekseyeva proofread my thesis, and was a constant support during the final, occasionally frantic, year of this work.

Needless to say, any substantive or typographical errors that remain are mine alone.

Some institutional thanks are due. The simulations were coded in Ox and executed in OxMetrics, which was also to generate all the charts I have included. The LaTeX system was used for typesetting, using K. A. Gillow's `ociamthesis` class. My financial support was provided largely by the Department of Economics; the Oxford Australia Scholarship Fund and James Fairfax; the George Webb Medley Fund; and Balliol College. Without their generous support this work would not have been possible. I have also frequently intruded upon the facilities of Nuffield College, for which I am grateful.

On a personal level, my friends at Balliol and throughout Oxford—too numerous to mention individually—have kept me sane during my time here. Holywell Manor will forever be a second home.

Finally, my family deserves credit for quietly tolerating my extended life as a student, and only rarely asking when I expected to finish. More seriously, I owe to my parents the desire to learn that I hope is reflected, if only imperfectly, in what follows.

Abstract

Title	Asymptotic analysis of the 1-step recursive Chow test (and variants) in time series models
Author	Andrew Whitby, Balliol College
Degree	Doctor of Philosophy
Submitted	Hilary Term 2013

This thesis concerns the asymptotic behaviour of the sequence of 1-step recursive Chow statistics and various tests derived therefrom. The 1-step statistics are produced as diagnostic output in standard econometrics software, and are expected to reflect model misspecification. Such misspecification testing is important in validating the assumptions of a model and so ensuring that subsequent inference is correct. Original contributions to the theory of misspecification testing include (i) a result on the pointwise convergence of the 1-step statistics; (ii) a result on the extreme-value convergence of the maximum of the statistics; and (iii) a result on the weak convergence of an empirical process formed by the statistics.

In Chapter 2, we describe the almost sure pointwise convergence of the 1-step statistic for a broad class of time series models and processes, including unit root and explosive processes. We develop an asymptotic equivalence result, and use this to establish the asymptotic distribution of the maximum of a sequence of 1-step statistics with normal errors. This allows joint consideration of the sequence of 1-step tests via its maximum: the sup-Chow test. In Chapter 3, we use simulation to investigate the power properties of this test and compare it with benchmark tests of structural stability. We find that the sup-Chow test may have advantages when the nature of instability is unknown. In Chapter 4, we consider how the test may be adapted to situations in which the errors cannot be assumed normal. We evaluate several promising approaches, but also note a trade-off between robustness and power. In Chapter 5 we analyse an empirical process formed from the 1-step statistics, and prove a weak convergence result. Under the assumption of normal errors, the limiting distribution reduces to that of a Brownian bridge. The asymptotic approximation appears to works well even in small samples.

Word Count

58,900 words (58.9% of the maximum word limit). Calculated by multiplying the number of words on a representative page (390, p. 24) by the total number of pages, excluding references (151).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis concerns the asymptotic behaviour of the sequence of 1-step recursive Chow statistics and various tests derived from these statistics. Over the following four chapters, we examine the asymptotic behaviour of the pointwise 1-step statistics and the maximum of the 1-step statistics; and an empirical process formed from these statistics. In addition, we investigate some of the finite sample properties of a test based on the maximum. In doing so, we intend to contribute to the understanding of a test that is in active use, offer a new test, and contribute more broadly to the literature of misspecification testing in econometrics.

Within the context of misspecification testing, the 1-step test lies within a family of tests proposed to uncover structural change, and so this is the primary applied context. Various tests have been developed to investigate structural change, including most famously Chow (1960). The 1-step recursive Chow test has its origin here, but in actual use it has outgrown the classical framework, being routinely deployed in autoregressive models. The 1-step test is not unique in this evolution: for instance, see the comments regarding the RESET test in Hendry and Nielsen 2007, s. 13.2.3. The question regarding all such tests is: how well do they work in the settings in which they are used? What results can we derive that may inform our understanding, and where these results hold asymptotically (as they almost inevitably do), how well do they translate to relevant sample sizes?

This thesis endeavours to answer these questions as they pertain to the 1-step recursive Chow test, and the variants we construct.

In Chapter 2, we are motivated initially by the test as it is currently implemented in software (e.g. PcGive). We derive a pointwise result that shows that even in a general autoregressive framework, the classical distributional results still hold, in the limit. This gives us confidence that current tests based on the 1-step statistics continue to provide useful information for investigators, even when the context is a time series one. We then go further, to develop a result by which the sequence of tests can be considered jointly, by considering the the maximum of the sequence (we call this the sup-Chow test).

Our approach to deriving the distribution of the maximum is of some independent interest. We prove the first result almost surely, and use Egorov's theorem to make a uniformity statement about the maximum. We then apply results from extreme value theory to derive the limiting distribution of the maximum, under an assumption about the error distribution.

In Chapter 3 we turn our attention to practical concerns, using Monte Carlo techniques. Firstly, we evaluate the performance of the asymptotic results of Chapter 2 in finite samples. We find these can be improved with some simple corrections. Secondly, we evaluate the power of the sup-Chow test, under a variety of misspecification scenarios. We find the test has useful power to detect mean and variance breaks, and single outliers.

Chapter 4 considers the sup-Chow test when the errors do not follow the normal distribution. This is not a matter of generalising the Chapter 2 results; the 1-step test and the sup-Chow test inherently require the error distribution to be specified (whether that is normal or something else). Instead, we consider two approaches, grounded in different modelling philosophies. In the first approach, we assume the investigator will make an assumption of normal errors, and so run the test in conjunction with other misspecification tests, one of which is designed to reject non-normal residuals. The issue that arises in this case is whether the sup-Chow test has power, conditional on this other test. In the second approach, we assume the investigator prefers to make weaker assumptions about the distribution of the errors. To allow the test to work in such an instance, we consider a variety of simple transformations that allow us to relax the normality assumption (though still requiring that the distribution be within a certain restricted class—the domain of attraction of the Gumbel maximal distribution). Both approaches are partially successful, but each diminishes the power of the test against the alternatives previously considered.

In Chapter 5 we analyse an empirical process formed from the square root of the 1-step statistics. We prove weak convergence to a process involving a pair of correlated Brownian bridges. Under the assumption of normal errors, this limiting distribution reduces to that of a Brownian bridge. This is a particularly convenient result, as in this case tests based on functionals of the empirical process will have their usual distributions. We demonstrate this with a simple Kolmogorov-Smirnov-type statistic, and find that the asymptotic approximation works well in small samples. These results clear the way for a range of tests, based on the empirical process, to be developed and evaluated.

The purpose of the remainder of this chapter is to provide context for the main results that follow in Chapters 2 through 5.

1.1 Background

Economics has long been an empirical discipline. Jevons (1871) states that

> The deductive science of Economics must be verified and rendered useful by the purely empirical science of Statistics. Theory must be invested with the reality and life of fact.

Econometrics as a distinct subdiscipline can be dated to the founding of Econometric Society in 1933. The Society defined econometrics as 'economic theory in its relation to statistics and mathematics' (Frisch 1933, p. 1). As Hoover (2006) notes, in modern use the term emphasises the application of statistical, rather than merely mathematical, techniques to economics (econometrics having separated from 'mathematical' economics). The seminal work of what might be considered 'modern' econometrics is Haavelmo (1944), which firmly established the applicability of probability models to economic data.

The early postwar period, following Haavelmo, saw a growing methodological consensus, but this is not what one finds in econometrics today. Pagan (1987) identifies three dominant methodologies, while Hoover (2006) enumerates no less than five separate methodological approaches. Pagan argues that the present schools of econometric thought arose after 1975, with the onset of 'stagflation' and the inability of the then-dominant large multi-equation models to explain or predict it. This caused a period of methodological introspection, which continues to the present. To an extent, however, methodological differences seem a natural outgrowth of the inherent complexity of the econometrician's task; and, in particular, the necessary but difficult task of drawing causal inference from non-experimental data.

1.1.1 Econometric methodology

Econometric modelling as practised today can be divided between those approaches that are theory-led and those that are data-led (Hoover characterises these, in historical context, as 'theory-of-errors' and 'probabilistic reduction' approaches, respectively). As this thesis largely concerns asymptotic theory, we need not take a strong position in this debate. But where motivation for the work and application of the theory results are concerned, we position ourselves within the London School of Economics (LSE) tradition, from which was born the 1-step recursive Chow test. The LSE approach was originated by Denis Sargan but has come to be associated with David Hendry (Hoover 2006, s. 2.4.3 p. 75). It is firmly in the data-led camp.

The LSE methodology involves a multi-step procedure, starting with the true, unknowable data generating process, and eventually arriving at the econometric model of the process, using a theory of 'reduction'. Reduction involves such considerations as aggregation, data transformation, data partition, marginalizing, sequential factorization, parameter constancy, lag truncation, integrated data, functional form specification, conditional factorization and simultaneity (see Hendry 2009, 1995, for thorough descriptions of these concepts). At each reduction stage, an emphasis is placed on testing a variety of null hypotheses about the model's correct specification. From this approach comes a strong emphasis on model misspecification testing.

1.1.2 Misspecification tests

Misspecification test are statistical tests, congruent with the formulation of Cox and Hinkley (1974, p. 66). They are, however, less easily characterised than the more familiar Neyman-Pearson (N-P) tests (Mizon 1977; Spanos 2006, s. 1.8.1.4, p. 40). Whereas N-P tests are typically used to evaluate restrictions suggested by economic theory, misspecification tests are designed purely to refute the model under consideration, on statistical grounds. Whereas N-P tests are constructed with a clear (though not necessarily point) alternative in mind, misspecification tests are not; when a misspecification test fails, it is considered evidence against the statistical adequacy of the model as a whole. Such tests were used as early as Fisher (1925), according to Spanos (2006, p. 21), and today they are commonly available as standard output from econometrics software.

Misspecification testing is central in the LSE approach. Hendry (1980, p. 403) states that '[t]he three golden rules of econometrics are test, test and test'. They are not solely connected with this methodology, however, and are emphasised, to varying degrees, in most approaches to econometric modelling. Nevertheless, Spanos (2006, p. 7) argues that

> The primary problem of current econometric modelling is unreliable evidence built upon unwarranted inductive inferences. One of the main contributing factors is that the premises of induction, the probabilistic assumptions comprising the underlying statistical model, are often incompletely specified and are rarely probed thoroughly for departures.

The misspecification test statistics we consider meet the two criteria specified by Cox and Hinkley (1974, p. 66). First, that the distribution of the statistic is known under the null hypothesis; and furthermore that this distribution be at least approximately the same in the case of a composite hypothesis. Second, that the larger the value of the statistic, the greater the evidence against the null hypothesis. The first of these criteria seems particularly relevant in cases where the process specified contains nuisance parameters; and we will thus emphasise, amongst other considerations, similarity of tests (in particular, invariance to the dynamic structure of the process).

1.1.3 Structural change

Structural change has been a topic of interest to econometricians since the field was established. For instance, Haavelmo (1944, p. 26) notes that:

> When we try to apply relations established by economic theory to actually observed series for the variables involved, we frequently find that the theoretical relations are "unnecessarily complicated"; we can do well with fewer variables than assumed a priori. But we also know that, when we try to make predictions by such simplified relations for a new set of data, the relations often break down, i.e., there appears to be a *break in the structure* of the data. For the new set of data we might also find a simple relation, but a *different* one. [emphasis in original]

Hendry (1995, s. 2.3) describes structure as 'the set of basic, permanent feature of the economic mechanism'. Causes of structural change may include 'technical progress, R&D, new legislation, institutional changes, regime shifts, financial innovation, shifting demography, evolving social and political mores, as well as conflicts and other major catastrophes'. (Hendry 2009, s.1.4.1.1, p. 17)

The question of structural stability of a model is central to the econometrician's task. Hendry (1995, s. 2.1) notes that 'parameters are of interest if they are uniquely defined, constant over historically relevant time periods and across regimes, and interpretable in the light of subject matter theory.' Structural change is of particular importance in forecasting. Hendry (2009, s.1.4.1.1, p. 17) notes that forecast failure is almost invariably due to structural breaks.

Although it is possible to specify and estimate models in which parameters are not constant over time, Hendry notes that such models 'will be useless for forecasting the future, analysing economic policy, or testing economic theories, since they lack entities on which to base those activities.' This is not to say that a model cannot have entities that do vary, and which we might sometimes call parameters; merely that some underlying parameters must be fixed. As further noted by Hendry, 'models with "varying coefficients" still have an underlying set of constant parameters which characterise the probability mechanism.'

We consider that these activities—forecasting, analysing policy, testing theory—represent most of what is useful about applied econometrics, and so the question of parameter constancy over time is essential to the task of econometric modelling. We consider models without constant parameters (defined as Hendry does) to be incompletely or incorrectly specified, so that parameter *non*constancy is a particular case of model misspecification.

1.2 The 1-step recursive Chow test

The popular econometrics package *PcGive* (Hendry and Doornik 2001, p. 254) includes three variants of the original Chow (1960) test, calculated recursively over time series data and aimed at testing parameter constancy over time, against a break at an *a priori* unknown time. The tests are presented as graphics, calculated per-observation from a chosen recursion start point M (which must be at least k, the number of regressors) up to the sample size S, and hence relax the known-breakpoint requirement. With RSS_s being the sum of squares of residuals for an OLS regression up to s, we have the following tests.

1-step Chow test in which a sequence of single period-s forecasts is made based on data from 1 to $s-1$, with s rolling from M to S.

$$\mathrm{Chow}_{1,s} = \frac{(\mathrm{RSS}_s - \mathrm{RSS}_{s-1})(s-k-1)}{\mathrm{RSS}_{s-1}}, \quad s = M \ldots S. \tag{1.1}$$

Break-point Chow test (or $N{\downarrow}$ Chow test) in which a sequence of multi-period forecasts is made with the break-point between the estimation sub-sample and the forecast sub-sample rolling from $(M-1, M)$ to $(S-1, S)$.

$$\mathrm{Chow}_{N\downarrow,s} = \frac{(\mathrm{RSS}_S - \mathrm{RSS}_{s-1})(s-k-1)}{\mathrm{RSS}_{s-1}(S-t+1)}, \quad s = M \ldots S. \tag{1.2}$$

Forecast Chow test (or $N{\uparrow}$ Chow test) in which the estimation sub-sample is fixed at $1 \ldots M-1$ while the forecast sub-sample grows from a single period M forecast to the entire sub-sample from M to S.

$$\mathrm{Chow}_{N\uparrow,s} = \frac{(\mathrm{RSS}_s - \mathrm{RSS}_{M-1})(M-k-1)}{\mathrm{RSS}_{M-1}(s-M+1)}, \quad s = M \ldots S. \tag{1.3}$$

Figure 1.1 shows the different estimation-forecast structures of the three variants.

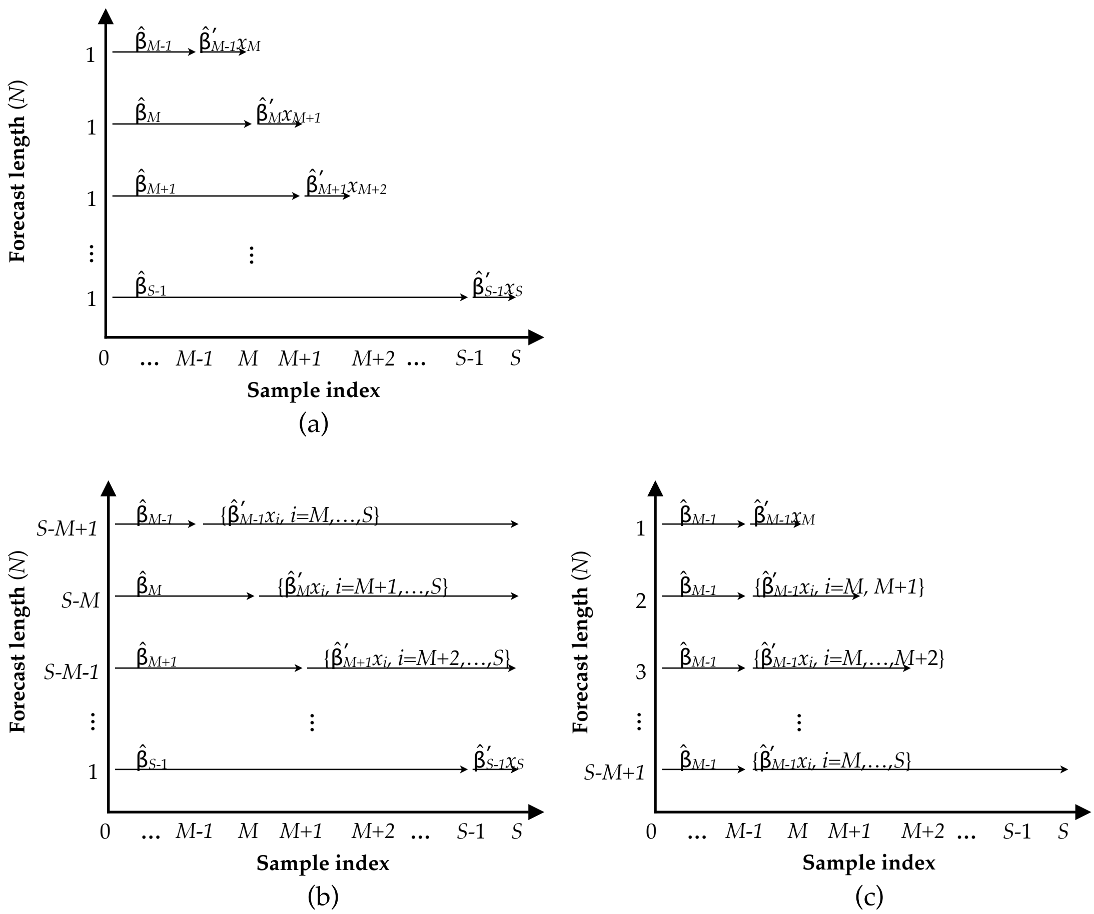

Figure 1.1: The three recursive Chow test variants: (a) 1-step; (b) break-point ($N\downarrow$); and (c) forecast ($N\uparrow$). Reproduced from Ericsson et al. (2005, 181-182).

This thesis focuses on the 1-step test. The 1-step test is essentially the 'prediction interval' test described by Chow, but computed recursively over the sample. This test and the other recursive variants were introduced to *PcGive* during the mid-1980s. (They are not present in 1983 when GIVE was part of the AUTOREG library (Hendry et al. 1983; Hendry 1983), but exist in 1986's PcGive version 4.0 in essentially the same form as in recent versions (Hendry 1986). Prior to this time the Chow test was present primarily in its traditional fixed split-sample form (with data partitioned into an estimation subsample and a forecast subsample). The incorporation of these tests into a popular software package appears to have led to their widespread use, particularly in central banking: see for example Kimura (2001); Celasun and Goswami (2002); Assarsson et al. (2004). Further, since the test is easy to generate and provides graphical reassurance of the model's correct specification, it is likely that far more investigators refer to the test's output (perhaps quite casually) than actually include it in published work.

Theory, meanwhile, has lagged implementation and application. Andrews (1993, p. 825) criticises the tests on the grounds that they lack a well-developed distributional theory. A review of the literature discussing the recursive Chow tests suggests that this criticism has merit. *PcGive*'s manual (Hendry and Doornik 2001, p. 254) cites only Chow's original paper, and implies that the statistic should be F distributed, stating, in addition, that normality is required for the 1-step test (this requirement proves important and is discussed in Chapter 4). Hendry (1995, s. 3.2.2) reviews the classical distributional theory and reports some Monte Carlo results for the break-point variant of the recursive Chow test under an autoregressive distributed lags (ADL) (1,1) data generating process (DGP), noting that 'the standard regression-theory results. . . hold reasonably well despite the presence of dynamics.' There is, however, no theoretical analysis of the test under such processes. The most complete description of the tests is given in Ericsson et al. (2005, pp. 177–183), which provides motivation for the recursive approach and details the differences between the three tests. However even here it is stated that '[u]nder the null hypothesis of correct specification, the statistic. . . is exactly distributed as $F(N, t-k)$ for normal independent e_t with fixed regressors x_t, approximately so for dynamic regressors.' No further information is provided on how 'approximately' this result holds.

Hence there appears to be a general belief that these tests perform relatively well on time series data, even in the absence of a complete theory. This thesis provides support for this belief, using asymptotic theory in Chapter 2 and finite sample simulations in Chapter 3.

A second issue with the tests as currently used is correct interpretation. The recursive testing procedure produces almost as many test statistics as there are observations. While this is useful in identifying the possible timing of a break, it makes the adoption of a strict decision rule difficult. Discussing a time series example of the 1-step test, Hendry and Nielsen (2007) note that as 'about 100 tests are done [with a 1% critical value], which are only slightly dependent, at least two or three rejections would be needed to reject the model based on this type of test.' Clearly, an element of subjective interpretation is required.

It must be said that these tests, with their primarily graphical output, are not intended to be used in a Neyman-Pearson framework, but rather in the iterative, data-driven approach associated with the LSE school. Nevertheless, this does not obviate the joint testing criticism, and this motivates our construction of a supremum version of the 1-step test in Chapter 2, which could be more easily used in a strict setting if desired.

Finally, in concentrating on this test and related constructions, we inevitably maintain a focus, at least implicitly, on short time-horizon forecast performance. This is not to suggest that longer horizons may not otherwise be of interest; but the specific issues relevant to this fall beyond the scope of this thesis.

1.2.1 An example

We introduce an example to demonstrate the issues. This is not intended to make any empirical contribution, but merely to demonstrate the issues facing investigators using the 1-step recursive Chow test, as it is currently implemented, and hence motivate the theoretical work that follows.

The data set we consider comes originally from Hendry and Ericsson (1991) and is provided with Hendry and Nielsen (2007). Amongst other variables, it includes quarterly, seasonally-adjusted M1 (denoted *ma*) and a price level (denoted *pa*), both in logs. This quarterly series spans 1963(1) through 1989(2).

A plot of the data suggests a possible regime change in the early 1980s, with a shift from stationary (or perhaps unit-root) behaviour to mildly explosive behaviour.

Although Hendry and Ericsson consider relatively elaborate dynamic models, we use a simple autoregressive model to demonstrate the test under investigation. In particular, we fit an AR(4) model to $ma - pa = \log(\frac{m}{p})$, the log of the real money stock. The assumed model is

$$\log\left(\frac{m}{p}\right)_s = \beta_1 \log\left(\frac{m}{p}\right)_{s-1} + \beta_2 \log\left(\frac{m}{p}\right)_{s-2} + \beta_3 \log\left(\frac{m}{p}\right)_{s-3} + \beta_4 \log\left(\frac{m}{p}\right)_{s-4} + \varepsilon_t \ ,$$

where $\varepsilon_t \sim \mathsf{N}(0, \sigma^2)$.

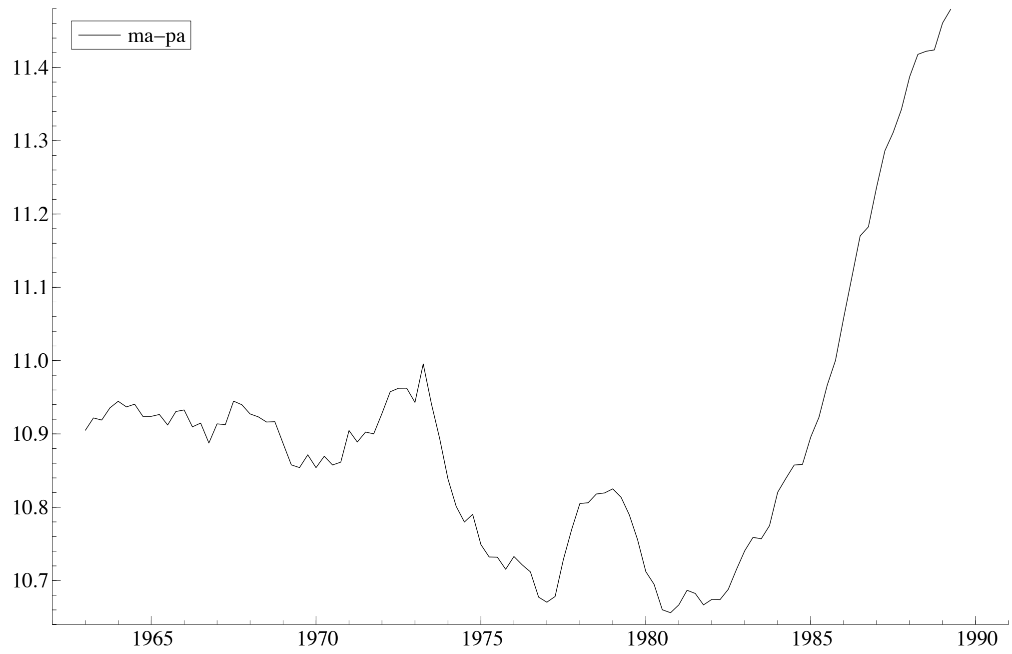

Figure 1.2: Example data series: log of the real quarterly, seasonally adjusted UK money stock (M1).

OLS estimation of this model produces the estimates shown in Table 1.1.

Variable	**Coefficient**	**(Std. Err.)**	***Diagnostics***	
β_1	1.333$**$	(0.1009)	T	102
β_2	-0.027	(0.1656)	R^2	0.987
β_3	-0.290	(0.1665)	$F_{3,98}$	2485$**$
β_4	-0.016	(0.1017)		

Table 1.1: Output from OLS estimation of AR(4) model. $**$ significant at 1% level.

Following model estimation, the 1-Step Chow test statistics can be calculated recursively using *PcGive*'s Recursive Graphics feature. The output is shown in Figure 1.3. The raw statistics are scaled by '1-off critical values from the F-distribution at any selected probability level [here 1%] so that the significant critical values become a straight line at unity' (Hendry and Doornik 2001, p. 254).

In practice, we interpret the test's exceedance of the 1% line around 1973 as evidence of a possible structural break. Given that nearly 100 statistics are plotted at a nominal 1% critical value, a single exceedance is highly likely even with parameter constancy. Nevertheless, the degree to which the test exceeds the 1% line, coupled with our *a priori* knowledge of the 1973 Oil Crisis, invites us to at least consider the possibility of model

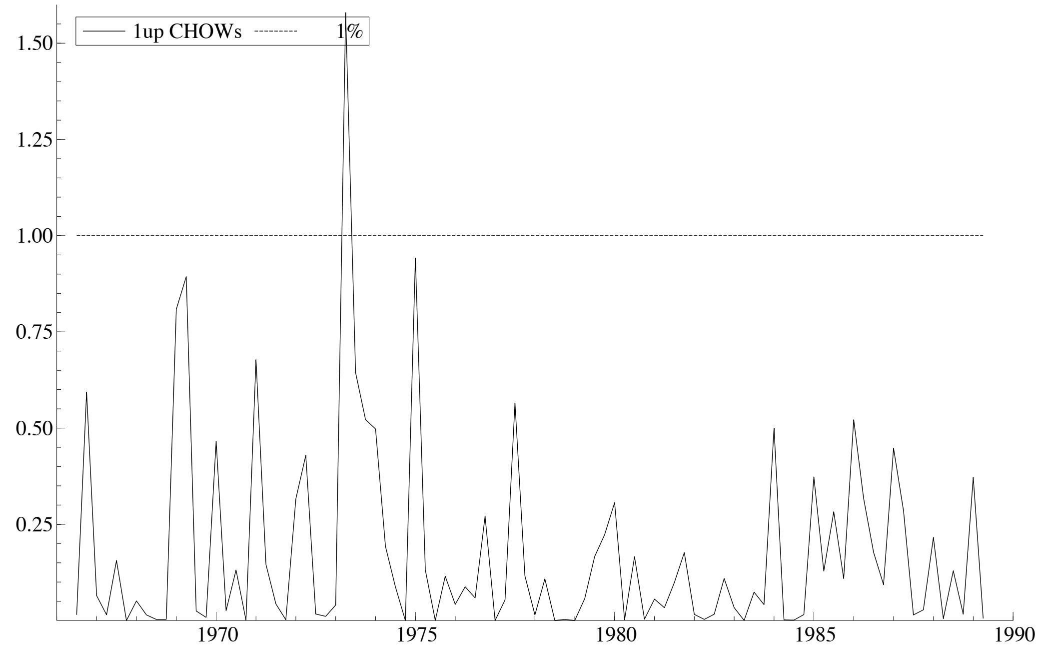

Figure 1.3: 1-Step Chow test statistic calculated for the AR(4) model; standard *PcGive* output.

misspecification, and whether a dummy variable might improve the model's fit.

It is important to recognise what the 1% normalising factor represents. It is the level beyond which 1% of probability mass lies, for a random variable with an F-distribution with $(1, T - k - 1)$ degrees of freedom. The use of this normalisation creates the two issues of interpretation noted above, namely uncertainty about the test's distribution with dynamic regressors, and difficulty interpreting the '1%' critical value in light of the obvious simultaneous testing across the sample.

1.3 Related procedures

The literature on testing for structural breaks is extensive—indeed in his handbook chapter Perron (2006, p. 279) describes it as 'truly voluminous'. Perron divides tests into those that do not explicitly model a break; those that model one break; and those that model multiple breaks. The recursive Chow tests, including the 1-step test, fit into the the first category, which includes the well-known CUSUM and CUSUMSQ tests. The modelled-break category includes, most prominently, the Quandt (1960)–Andrews

(1993) supremum tests. The classical Chow test itself is also a modelled-break test, albeit one in which the breakpoint must be known *a priori*.

In this section, we review the intellectual heritage of the 1-step test. The test can be framed in two ways: either as a time-ordered extension of the 'prediction interval' test of Chow (1960), or as a sequence of studentized recursive residuals. We examine the antecedents of both approaches. We consider the competing approach identified by Perron, modelled-break testing, in Section 1.4, with a particular focus on the Andrews (1993) test.

1.3.1 The classical Chow test

The classical Chow (1960) test—actually comprising several tests—was developed to detect a change in the slope coefficient at a known breakpoint in the classical linear regression model. Although the test's introduction pre-dates modern time series methods, it is clear that Chow had time series applications in mind (Chow 1960, p. 591):

> When a linear regression is used to represent an economic relationship, the question often arises as to whether the relationship remains stable in two periods of time, or whether the same relationship holds for two different groups of economic units. Is the consumption pattern of the American people today the same as it was before World War II?

Chow presented a sequence of tests, which can be shown to be closely related to a standard F test of the hypothesis that some coefficients are equal to zero (see Fisher 1970 for a particularly clear exposition). Under the null hypothesis of no change, with classical assumptions, any one of Chow's statistics follows an F distribution.

Chow considers a classical regression model, with p regressors and n observations. His problem is to determine whether m additional observations come from the same regression. This is essentially a problem of parameter non-constancy within a series of $n + m$ observations, with a known break-point at $n + 1$.

He notes that two relevant tests already existed, at the time of writing: the 'well-known' prediction interval test, useable when $m = 1$ (his equation (8)), and the analysis of covariance test, useable when $m > p$. Chow's original contributions are to offer tests for the intermediate situation, when $2 \leq m \leq p$, and to situate all three cases within the framework of testing general linear hypotheses.

Apart from the reliance on classical assumptions, Chow's tests suffer from another problem for practical time series modelling: the need to specify, *a priori*, the point in time at which the parameter change occurs. In some cases the existence and timing of structural breaks may be known or suspected: for instance in the case of a known policy change (central bank independence, for instance) or a well-studied macroeconomic shock (the oil shocks of the 1970s, for instance). But often this will not be the case, and it is useful—particularly for the purposes of model evaluation—to be able to determine whether some break occurred at an unknown time. One common alternative is to make an informed guess as to the time of the break (or indeed an uninformed guess, by assuming the break occurs at the halfway point). This can be moderately powerful against breaks that occur relatively close to the assumed breakpoint.

1.3.2 The use of recursive residuals

Recursively (or sequentially) calculated regression residuals have been proposed for a wide variety of testing purposes. Consider a linear regression

$$y_t = \beta' x_t + \varepsilon_t \qquad t = 1, \dots, T, \tag{1.4}$$

with y_t scalar, x_t a k-dimensional vector of regressors. 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 RSS_t and recursive residual $\tilde{\varepsilon}_t$, by

$$\hat{\beta}_t = \left(\sum_{s=1}^{t} x_s x_s' \right)^{-1} \left(\sum_{s=1}^{t} x_s y_s' \right) \qquad t = k, \dots, T, \tag{1.5}$$

$$\mathrm{RSS}_t = \sum_{s=1}^{t} (\hat{\beta}_t' x_s - y_s)^2 \qquad t = k, \dots, T, \tag{1.6}$$

$$\tilde{\varepsilon}_t = \left[1 + x_t' \left(\sum_{s=1}^{t-1} x_s x_s' \right)^{-1} x_t \right]^{-1/2} (y_t - \hat{\beta}_{t-1}' x_t) \qquad t = (k+1), \dots, T. \tag{1.7}$$

As Brown et al. (1975) put it, 'it is natural to look at residuals to investigate departures from model specification', and this approach has a relatively long history. The recursive residuals have two advantages over the OLS residuals in this respect: 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. the BLUS—best linear unbiased scalar covariance—residuals of Theil 1965), they have a natural interpretation—in a time series setting—as forecast errors.

The former property was developed in O'Reilly and Quesenberry (1973), which extended to the regression case the literature on goodness-of-fit testing using the probability integral transform (PIT) (Pearson 1938). Then, with exogenous regressors, it is shown that the the recursive residuals are sequentially independent. Such tests share the character of the test we consider in also being, typically, tests of pure significance. Hawkins (1991) develops the approach more completely, drawing parallels with various concepts (leverage, etc.) in the outlier literature.

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 partly motivates the results of Chapter 2.

The recursive residuals form a sequence from $k + 1 \ldots T$ where k is the number of regressors in the model. In principle, many tests that are constructed from the OLS residuals could be adapted to instead use the recursive residuals. Many different functions of the sequence can be contemplated in order to achieve power against various alternatives. As noted by Hawkins (1991), the sequence may, for example, be squared, cumulated, or squared and cumulated to form another sequence of statistics; then various aggregations—mean, maximum, etc.—may be considered to form a scalar summary of the whole sample. A further choice is how to scale the residuals: in particular, whether a full sample or sequential variance estimator is used. Many of these combinations appear, in some form, in the literature.

Surprisingly, the use of recursive residuals to detect outliers in time series data has been little-explored, although there is little doubt they are used for this purpose in practice. 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 labelling of the observations is usually done at random, or in relation to some concomitant variable…'. This difficulty does not exist with time series data, where a natural unique chronological labelling of observations pertains. The section in the same book (p. 396) on detecting outliers in time series is, nevertheless, notably brief, and recursive methods are not considered.

Kianifard and Swallow (1996) outline the development and use of recursive residuals in some detail. Below, we highlight the CUSUM and CUSUMSQ test, which popularised the use of recursive residuals for misspecification testing.

1.3.2.1 CUSUM and CUSUMSQ

The recursive residuals first appear as a regression diagnostic tool with the introduction of the CUSUM and CUSUMSQ tests in Brown et al. (1975). These tests are designed to visually inspect model adequacy (in particular stability over time). The CUSUM is defined

$$W_r = \frac{1}{\hat{\sigma}^2} \sum_{j=k+1}^{r} \tilde{\varepsilon}_j \qquad r = k+1 \ldots T,$$

where

$$\hat{\sigma}^2 = \frac{\sum_{j=k+1}^{T} (\tilde{\varepsilon}_j - \bar{\tilde{\varepsilon}})^2}{n-k-1} \qquad \text{and} \qquad \bar{\tilde{\varepsilon}} = \frac{\sum_{j=k+1}^{T} \tilde{\varepsilon}_j}{T-k}.$$

W_r is plotted against r. Brown et al. (1975) show that the recursive residuals are independent $N(0, \sigma^2)$ distributed under the classical normal model with non-stochastic regressors. Therefore $E(W_r) = 0$, $V(W_r) \approx r - k$ and $C(W_r, W_s) \approx \min(r, s) - k$ for each r. They then approximate W_r using a continuous normal process, and establish, for any size α, bounds within which W_r should vary (under the null hypothesis of no structural change).

The CUSUMSQ is defined

$$s_r = \left(\sum_{j=k+1}^{r} \tilde{\varepsilon}_j^2 \right) \Big/ \left(\sum_{j=k+1}^{T} \tilde{\varepsilon}_j^2 \right).$$

Brown et al. (1975) calculate the joint distribution of the statistics under the same assumptions. Again, bounds can be plotted for s_r, giving a visual indication of any violation of the null hypothesis.

As Brown et al. (1975) work within a framework of fixed regressors, they specifically preclude extension to autoregressive-type processes. Ploberger and Krämer (1986) extends the analysis to stationary autoregressive processes, while Lee et al. (2003) consider a unit root autoregressive process. Recently, Nielsen and Sohkanen (2011) further generalised the test to ADL models with deterministic terms and potentially stationary, unit-root or explosive behaviour. In all cases the asymptotic distributions are the same as in the fixed regressor case.

1.4 Procedures used as benchmarks

In this section we review two tests of structural instability which we consider to be benchmarks. We will use these in Chapters 3 and 4, where we analyse the relative performance of the test developed in Chapter 2 on a number of dimensions.

The first benchmark test is the Quandt-Andrews mid-sample test, first proposed by Quandt (1960) with distribution theory worked out in Andrews (1993). It appears to be the most popular of the modelled-breakpoint tests, being commonly implemented in econometrics software and therefore well-known and readily applicable. It is found in in EViews 7 (QMS 2009) as the command `ubreak` or in R (R Core Team 2013) using the package `strucchange` (Zeileis et al. 2013). Its one significant drawback is low power at the beginning and end of the sample, resulting from a 'trimming' procedure required for the asymptotic results to hold.

As will be seen, the test developed in Chapter 2 requires a similar trimming at the beginning of the sample, but not at the end of the sample. So to allow for a complete comparison, we use a second benchmark test, the Andrews (2003*a*) end-of-sample test, a sequel designed to complement the 1993 test, and address exactly the issue of low power towards the end of the sample.

1.4.1 Quandt (1960)–Andrews (1993) mid-sample test

Quandt (1960) addresses the question of a Chow-type test when the break date is unknown. Quandt proposed using the likelihood ratio (LR) test statistic, evaluated at the likelihood-maximising break date. The asymptotic distribution of such a statistic is non-standard, since the timing parameter appears only under the alternative hypothesis and not under the null. This complicates the analysis (see Davies 1977), and prevented Quandt from solving the distribution problem. The general method proposed by Davies for such tests is to consider the maximum of the LR statistic over all possible break dates, within a given set.

Andrews (1993) studies LR, LM and Wald tests of this type, and finds that the asymptotic null distributions are given by the supremum of the square of a standardised tied-down Bessel process of order $p > 1$ (see Hawkins 1977). The original paper provided tabulated critical values, although these are superceded by Andrews (2003*b*). In addition, Hansen (1997) provides a numerical method of approximating this distribution, allowing asymptotic p-values to be computed. (This latter method is used, for instance, in EViews.)

An interesting issue arises in the asymptotic theory: as the sample size grows, the set of break dates cannot extend to either edge of the sample, or else the test diverges to infinity under the null hypothesis. The practical result is that the full set of pointwise LR statistics must be 'trimmed' at either end. Andrews recommends the convention of $(0.15, 0.85)T$ when there is no prior knowledge about the break timing.

In the linear regression setting, the tests have a straightforward expression described in Andrews (1989). We can construct the the ordinary F test statistics $F_T(\tau)$ for a break in the regression parameters at time τT with $\tau \in (0, 1)$. Then, applying the conventional trimming, we calculate $\sup F = \sup_{\tau \in (0.15, 0.85)} F_T(\tau)$. Then it holds that $F_T(\tau) = [(T - 2p)/T] W_T(\tau)/p$, where $W_T(\tau)$ is the Wald test statistic. This allows the $\sup F$ test statistic to be compared with Table 1 of Andrews (1989) (with those critical values divided by p), or the more recent tabulations.

This is the form of the Quandt-Andrews test we use in simulation experiments in Chapters 3 and 4. It is numerically equivalent to the test performed by the `ubreak` command in EViews, which allows us to use, as a benchmark, a test which is actually implemented in software.

The supremum tests of Andrews (1993), while very useful in practice, suffer from several complications. Firstly, while the test statistic is relatively straightforward (particular in the linear regression setting), the asymptotic distributions are non-standard. This is, of course, mitigated by the availability of Hansen (1997) method of approximation. Secondly, the trimming procedure means the test has reduced power against breaks that occur early or late in the sample—and for forecasting, the latter is of particular concern. This motivates the next test we discuss.

1.4.2 Andrews (2003*a*) end-of-sample test

The Andrews (2003*a*) test of end-of-sample instability, the S test, is designed to answer the above criticism of the Andrews (1993) test. It generalises the Chow (1960) F test to situations with (a) errors which are not iid normal, (b) regressors that are not strictly exogenous and (c) estimation methods other than least squares.

The model has d regressors, n observations before the potential changepoint, and m observations after it. The model equation is

$$y_i = \begin{cases} x_i'\beta_0 + \epsilon_i, & \text{for } i = 1, \ldots, n, \\ x_i'\beta_{1i} + \epsilon_i, & \text{for } i = n+1, \ldots, n+m. \end{cases}$$

with assumptions that $\mathsf{E}\varepsilon_i x_i = 0$, $\mathsf{E}x_i x_i'$ is positive definite and $\{(y_i, x_i) : i \geq 1\}$ are stationary and ergodic under the null hypothesis.

The null and alternative hypotheses proposed are

$$H_0 : \begin{cases} \beta_{1i} = \beta_0 \text{ for all } i = n+1, \dots n+m, \text{ and} \\ \{(y_i, x_i) : i \geq 1\} \text{ are stationary and ergodic,} \end{cases}$$

$$H_1 : \begin{cases} \beta_{1i} \neq \beta_0 \text{ for some } i = n+1, \dots n+m, \text{ and/or} \\ \text{the distribution of } (\varepsilon_{n+1}, \dots, \varepsilon_{n+m}) \text{ differs from} \\ \text{that of } (\varepsilon_i, \dots, \varepsilon_{i+m-1}) \text{ for } i = 1, \dots, n-m+1. \end{cases}$$

Define the following vector notation:

$$\begin{aligned} \mathbf{y}_{r,s} &= (y_r, \dots, y_s)', \\ \mathbf{x}_{r,s} &= (x_r, \dots, x_s)', \text{ and} \\ \boldsymbol{\varepsilon}_{r,s} &= (\varepsilon_r, \dots, \varepsilon_s)'. \end{aligned}$$

Then the S test statistic is then constructed as follows. Let

$$\hat{\beta}_{n+m} = \text{LS estimator of } \beta \text{ using observations indexed by } i = 1, \dots, n+m.$$

Then an estimator of the $m \times m$ error covariance matrix $\Sigma_0 = \mathsf{E}\boldsymbol{\varepsilon}_{1,m}\boldsymbol{\varepsilon}_{1,m}'$ (stationary by assumption) is

$$\hat{\Sigma}_{n+m} = \frac{1}{n+1} \sum_{j=1}^{n+1} \hat{\boldsymbol{\varepsilon}}_{j,j+m-1} \hat{\boldsymbol{\varepsilon}}_{j,j+m-1}' \quad \text{where} \quad \hat{\boldsymbol{\varepsilon}}_{j,j+m-1} = \mathbf{y}_{j,j+m-1} - \mathbf{x}_{j,j+m-1} \hat{\beta}_{n+m}$$

Finally, the statistic S is defined as

$$S = \begin{cases} S_{n+1}(\hat{\beta}_{n+m}, \hat{\Sigma}_{n+m}), & \text{for } m > d, \\ P_{n+1}(\hat{\beta}_{n+m}, \hat{\Sigma}_{n+m}), & \text{for } m \leq d, \end{cases}$$

where

$$\begin{aligned} S_j(\beta,\Sigma) &= A_j(\beta,\Sigma)' V_j^{-1}(\Sigma) A_j(\beta,\Sigma), \\ A_j(\beta,\Sigma) &= \mathbf{x}'_{j,j+m-1}\Sigma^{-1}(\mathbf{y}_{j,j+m-1} - \mathbf{x}_{j,j+m-1}\beta), \\ V_j(\Sigma) &= \mathbf{x}'_{j,j+m-1}\Sigma^{-1}\mathbf{x}_{j,j+m-1}, \\ P_j(\beta,\Sigma) &= (\mathbf{y}_{j,j+m-1} - \mathbf{x}_{j,j+m-1}\beta)'\Sigma^{-1}(\mathbf{y}_{j,j+m-1} - \mathbf{x}_{j,j+m-1}\beta), \end{aligned}$$

with $\beta \in \mathbb{R}^d$ and Σ a nonsingular $m \times m$ matrix, for $j = 1, \ldots, n+1$.

In either case ($m > d$ or $m \leq d$), large values of S are suggestive of structural instability and provide evidence against the null hypothesis.

The key difficulty is in obtaining asymptotic results without a parametric assumption on the form of the error distribution. Since the end-of-sample consists of a fixed m periods, central limit theorems are of no use here. Instead, to calculate critical values, Andrews suggests a subsampling technique. Under the assumptions stated, the $S_j(\beta,\Sigma)$ are stationary and ergodic, and the distribution of $S_1(\beta_0,\Sigma_0)$ can be estimated consistently using the empirical distribution of $\{S_j(\beta,\sigma) : j = 1, \ldots, n-m+1\}$ evaluated at consistent estimators of β_0 and Σ_0. Andrews suggests

$$\begin{aligned} \hat{\beta}_{2(j)} = {} & \text{LS estimator of } \beta \text{ using observations indexed by } i = 1, \ldots, n, \\ & \text{with } i \neq j, \ldots, j + \lceil m/2 \rceil - 1, \end{aligned}$$

(where $\lceil \cdot \rceil$ denotes ceiling), for $j = 1, \ldots, n-m+1$, as an estimator of β_0, and $\hat{\Sigma}_{n+m}$ as an estimator of Σ_0. Then define

$$S_j = \begin{cases} S_j(\hat{\beta}_{2(j)}, \hat{\Sigma}_{n+m}), & \text{for } m > d, \\ P_j(\hat{\beta}_{2(j)}, \hat{\Sigma}_{n+m}), & \text{for } m \leq d, \end{cases}$$

for $j = 1 \ldots n-m+1$. The critical values are based on the $1-\alpha$ quantile of the empirical distribution function of these statistics, so that a test p-value may be constructed as

$$p_S = \frac{1}{n-m+1}\sum_{j=1}^{n-m+1} 1(S \le S_j).$$

Andrews finds the S test generally has better size than the F test whenever the assumption of uncorrelated normal errors is violated. Size-corrected power is higher for the F test when the degree of autocorrelation is mild, and higher for the S test when the autocorrelation is more severe. For all the tests, power increases rapidly in m (the length of the end-sample period) and only slowly in n. This is because the test ultimately depends upon only m residuals.

Two issues are not fully addressed in the paper.

Firstly, the test requires stationary regressors, and while Andrews claims it has little (undesired) power against non-stationarity involving a one-time change in error distribution, it is not clear that this would extend to unit-root type non-stationarity. The solution proposed for this case is to use differenced data; however that requires the investigator to simultaneously consider structural instability and unit root behaviour. This may cause difficulties, as structural breaks can cause false-positives in tests for unit roots, conflating two separate specification issues (see e.g. Perron 2006, p. 279).

Secondly, given a sample, it is not clear how an investigator should choose n and m. As far as we are aware, this test is not implemented in any common econometrics software, so we cannot refer to typical use. Given that Andrews recommends excluding the first and last 15% of observations for the (complementary) Andrews (1993) tests, a logical choice would be to set m equal to 15% of the sample size. However, the simulations presented in Andrews (2003*a*) assume that parameters change and remain changed for the entire end-sample period from $n+1$ to $n+m$. If instability occurred in only a small part of that, this would likely diminish power.

In the simulations of Chapters 3 and 4, we compare two versions of this test, one with $m = 0.15T$, and one with $m = 1$.

1.5 Mathematical techniques

The subject of this thesis is the sequence of 1-step Chow statistics, described in Section 1.2 above. In Chapters 2 and 5 we analyse this sequence from three quite different perspectives, with each requiring particular mathematical methods. First, we prove a result concerning pointwise asymptotic equivalence of the 1-step sequence and a simpler sequence, the limiting distribution of which can be readily derived. Second, we consider the distribution of the maximum over this sequence. Finally, we analyse an empirical process constructed from the 1-step statistics.

Throughout, we rely on results of Nielsen (2005), which concerns the strong convergence of autoregressive estimators and related quantities. To take advantage of these results, we adopt the flexible model framework used in this paper (autoregressive distributed lags regression with a DGP modelled as a vector autoregression). This framework supports a great variety of econometric models (see Hendry 1995, Chapter 7). Nielsen's results are a generalisation of the strong convergence results of Lai and Wei (1985).

The use of almost sure convergence, rather than the more common convergence in probability, serves a number of purposes. It provides a relatively straightforward method of dealing with maxima and sums over sequences, as compared with the more elaborate arguments that must be used to analyse the same using convergence in probability (see, for example, Deng and Perron 2008). Intuitively, complicated sums and maxima arise in relation to the 1-step recursive Chow statistic exactly because of the recursive estimation: where a quantity involving OLS residuals might involve a single sum, the same quantity involving recursive residuals will have a double sum, and it is in such situations that almost sure results prove convenient.

Of course, these results do not come for free. In general, the Nielsen (2005) results require the existence of higher moments. In the case of the Chapter 2 results, this assumption is amplified by the need to bound various covariances of process components, with the result that 16+ moments must be finite. For the Chapter 5 results, only 4+ moments are required. If heavy tails are a concern, these assumptions might prove problematic. Of course an assumption of normal errors will trivially satisfy such requirements.

The almost sure results also allow the investigation of the sup-Chow statistic, via an application of Egorov's theorem that is of some technical interest. For the first part, we take advantage of the pointwise almost sure results to make a convergence-in-probability statement on the supremum statistic. The cost of this approach is that an asymptotically growing subsample of the sequence must be excluded from calculation of the maximum. This reflects early instability in the recursive parameter estimates due to small estimation subsamples, and consequently seems unavoidable for a statistic of this nature. (The practical implication is that the test thus derived will have no power against breaks or outliers early in the sample; but the sup-Chow is not alone in this characteristic. It will have power at the end-of-sample, which is of particular relevance in forecast scenarios.) From there, we use the techniques of extreme value theory to find the asymptotic distribution of the maximum (appropriately scaled and centred). Since the sequence under consideration, though simpler than 1-step sequence itself, remains correlated, we cannot use the results of the classical theory and instead rely on further developments applicable to dependent sequences; the key reference here is Leadbetter et al. (1982).

In Chapter 5 we turn to the empirical process approach. This necessitates the use of weak convergence of probability measures on function spaces, a topic well described by, for example, Billingsley (1968). The majority of arguments serve to show asymptotic equivalence, uniformly on the unit interval, of the empirical process to standard processes, so that the weak convergence results follow with relatively little additional work (essentially direct application of theorems on weighted empirical processes, by Shorack and Wellner 1986). The key theorem here follows Theorem 2.2.5 of Koul (2002) in outline, although with very substantial modifications. Koul's theorem applies to a regression scenario with scale and location error, in common with our result, however recursive estimation creates additional complications. Once again, the almost sure results of Nielsen (2005) allow us to resolve many of these complications.

Where not specifically highlighted above, or cited in the text, the reader can safely assume that any mathematical techniques used are well-described in Davidson (1994). Short but lucid summaries of the relevant asymptotic theory are provided in a handbook chapter by the same author (Davidson 2006). A good explanation of the Landau O and

o notation, used throughout, is found in de Bruijn (1961).

1.6 Thesis structure

This thesis consists of this introduction, four content chapters, a conclusion and an appendix.

Chapter 2 presents asymptotic analysis of the 1-step test. An almost sure pointwise convergence result is developed, applicable to a broad class of time series models and processes. This result can be used directly to derive the distribution of the pointwise statistics for a given error distribution. In addition, we analyse the behaviour of the maximum of the 1-step sequence, using extreme value theory to derive an asymptotic result on the distribution of the maximum, given normal errors. These provide the basis for a new test, the sup-Chow test, which allows joint consideration of the 1-step sequence. A version of this chapter has been circulated as a discussion paper (Nielsen and Whitby 2012).

Chapter 3 applies Monte Carlo simulation techniques to investigate the finite sample size and power of the sup-Chow test, with normal errors. We propose a pragmatic correction that substantially improves size performance of the test in small samples. We investigate power under various plausible alternatives of structural change and the presence of outliers. Comparison is made with the benchmark tests described above.

Chapter 4 considers how the sup-Chow test may be adapted to situations in which the errors cannot be assumed normal. We investigate two main approaches: first, using the test in combination with a separate test for normality of the residuals; second, transforming the test to make it more robust against non-normal errors.

Chapter 5 presents asymptotic analysis of an empirical process formed by the 1-step statistics. We prove weak convergence to a process involving two correlated Brownian bridges, and show that this process reduces to another Brownian bridge when the reference distribution is normal. A potential application is demonstrated using a Kolmogorov-Smirnov-type statistic, with the asymptotic approximation proving accurate in small samples.

Chapter 6 concludes, and enumerates some directions for future work.

Appendix A provides detail of the common framework used for simulations in the preceding chapters.

Chapter 2

The 1-step Chow test and the sup-Chow test

2.1 Introduction

Identifying structural instability in models is of major concern to econometricians. 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. In this paper we consider two variants that relax those requirements: 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 2.4.1 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

A version of this chapter has been circulated as the discussion paper Nielsen and Whitby (2012).

supremum version of the 1-step test, applicable to detecting parameter change or at 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 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; a similar diagnostic graphic, the 'one-step forecast test' is provided in EViews (QMS 2009, p. 180). The idea of using residuals calculated recursively to test model misspecification dates from the landmark CUSUM and CUSUMSQ tests (Brown and Durbin 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 it tests 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. 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 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 is based on 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 tests for either structural breaks or outliers, both at a possibly unknown time.

Regarding structural breaks, Perron (2006) divides 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 non-monotonicity of power against some alternatives. The risk of this seems much reduced by using the 1-step Chow statistics, since all parameters are estimated on a growing subsample.) The modelled-break category includes, most prominently, the Quandt (1960)–Andrews (1993) supremum tests. These tests are complicated by a non-standard distribution (tabulated in Andrews 2003*a*), but are nevertheless popular in practice, being implemented in several software packages. Our test is distinguished from these in not imposing any restrictions on the end-of-sample, so that end-of-sample instability may be detected. This feature is similar to Andrews (2003*a*), 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 1-step tests behave like an i.i.d. process, the asymptotics differ from full-sample tests like Quandt–Andrews, requiring the application of extreme value theory of independent and weakly dependent sequences, rather than the suprema of random-walks.

Seen, alternatively, as an outlier test, the supremum Chow test falls squarely within the tradition of Srikantan (1961), which considers an unknown outlier in a classical setting. Outliers in the Box-Jenkins paradigm have attracted substantial interest, in particular the sequence of Chang and Tiao (1983); Chang et al. (1988); Chen and Liu (1993) which build on the classification of Fox (1972). 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 by the use of recursive estimation in the test we present.

Surprisingly, the use of recursive residuals to detect outliers in time series data has been little-explored, although they are likely used for this purpose in practice. 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 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.

2.2 The test statistics

The 1-step test applies to a linear regression

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

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, along with the corresponding residual sums of squares and recursive residual (or standardised 1-step forecast error) $\tilde{\varepsilon}_t$. We have

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

$$\mathrm{RSS}_t = \sum_{s=1}^{t} (\hat{\beta}_t' x_s - y_s)^2 \qquad t = k, \ldots, T, \tag{2.3}$$

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

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

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

and can be expressed as

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

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, $\mathsf{C}^2_{1,t}$ 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 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.

2.3 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 (VAR).

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,

$$Y_t = \rho Z_t + \sum_{j=1}^{k} \alpha_j Y_{t-j} + \sum_{j=1}^{k} \beta_j' Z_{t-j} + \nu D_{t-1} + \varepsilon_t \qquad t = 1, \ldots, T. \tag{2.7}$$

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

$$X_t = \sum_{j=1}^{k} A_j X_{t-j} + \mu D_{t-1} + \xi_t \qquad t = 1, \ldots, T, \tag{2.8}$$

with the deterministic term D_t given by

$$D_t = \mathbf{D} D_{t-1}. \tag{2.9}$$

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

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 Nielsen 2005, Remark 9.3). In the present analysis this constraint emerges in Lemma 2.A.1(i) and is transmitted via Lemma 2.A.2(iv) to Lemma 2.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 2.3.2. *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*

$$\sup_t \mathsf{E}(\|\xi_t\|^\alpha|\mathcal{F}_{t-1}) \overset{a.s.}{<} \infty \qquad \text{for some } \alpha > 16, \tag{2.10}$$

$$\mathsf{E}(\xi_t'\xi_t|\mathcal{F}_{t-1}) \overset{a.s.}{=} \Omega \qquad \text{where } \Omega \text{ is positive definite.} \tag{2.11}$$

This assumption also excludes the possibility that the innovations could 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 (Stock and Watson 2002) and subsequent period of 'global financial crisis'.

We permit nearly all possible values of the autoregressive parameters A_j in (2.8), 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 2.3.3. *The explosive roots of $\mathbf{B}$ have geometric multiplicity unity. That is, for all complex λ with $|\lambda| > 1$, $\operatorname{rank}(\mathbf{B} - \lambda \mathbf{I}_{pk}) \geq pk - 1$.*

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

Assumption 2.3.4. *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 the 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 2.4.4. 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 2.3.5. $\varepsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$.

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

$$S'_{t-1} = (Y_{t-1}, Z'_{t-1} \ldots, Y_{t-k}, Z'_{t-k}, D'_{t-1}),$$

and then write (2.8) in companion form, so that

$$S_t = \mathbf{S} S_{t-1} + (\xi'_t, 0')'. \tag{2.12}$$

Then there exists a regular real matrix $\mathbf{M}$ to block diagonalize $\mathbf{S}$ (see the elaboration in section 3 of Nielsen 2005), so that the process can be decomposed into non-explosive and explosive components, R_t and S_t respectively. We have

$$\mathbf{M}S_t = (\mathbf{MSM}^{-1})\mathbf{M}S_{t-1} + \mathbf{M}(\xi_t', 0')',$$
$$\begin{pmatrix} R_t \\ W_t \end{pmatrix} = \begin{pmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{W} \end{pmatrix} \begin{pmatrix} R_{t-1} \\ W_{t-1} \end{pmatrix} + \begin{pmatrix} e_{R,t} \\ e_{W,t} \end{pmatrix}, \tag{2.13}$$

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, paired with a distributional assumption such as 2.3.5, 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 2.4.1. *Under Assumptions 2.3.1, 2.3.2, 2.3.3 and 2.3.4,*

$$\mathsf{C}_{1,t}^2 - (q_t/\sigma)^2 \overset{\text{as}}{\to} 0 \quad \textit{as } t \to \infty, \tag{2.14}$$

where

$$q_t = \frac{\varepsilon_t - \sum_{s=1}^{\infty} \varepsilon_{t-s} W'(\mathbf{W}^{-s})' F_W^{-1} W}{(1 + W' F_W^{-1} W)^{1/2}}, \tag{2.15}$$

and $\mathbf{W}$ *is as in (2.13), 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 2.4.2. *Suppose* $\mathsf{C}_{1,t}^2 - (q_t/\sigma)^2 \overset{\text{as}}{\to} 0$ *as* $t \to \infty$*. Then*

$$\sup_{g(T) < t \leq T} \left| \mathsf{C}_{1,t}^2 - (q_t/\sigma)^2 \right| \overset{\text{p}}{\to} 0 \quad \textit{as } T, g(T) \to \infty. \tag{2.16}$$

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

$$\Pr(\Lambda < x) = \exp[-\exp(-x)] \qquad \text{where } x \in \mathbb{R}. \tag{2.17}$$

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:

$$\Lambda \sim \text{Gumbel} \qquad \text{iff} \qquad 2e^{-\Lambda} \sim \chi^2_2. \tag{2.18}$$

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 2.A.6).

Lemma 2.4.3. *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 $a_n = (2\log n)^{1/2}$ 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 (2.15).

Lemma 2.4.4. *Under assumption 2.3.5,*

$$q_t/\sigma \sim N(0,1), \text{ and} \tag{2.19}$$

$$a_t(\max_{1\leq s\leq t} q_s^2 - b_t) \xrightarrow{\text{d}} \Lambda, \tag{2.20}$$

where

$$a_t = 1/2 \qquad \textit{and} \qquad b_t = \log t^2 - \log\log t - \log \pi, \tag{2.21}$$

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 2.4.5. *Under assumptions 2.3.1, 2.3.2, 2.3.3 and 2.3.5, with some* $g(T) \to \infty$*,*

$$\mathsf{SC}_T^2 = \frac{1}{2}\left(\max_{g(T)<t\leq T} \mathsf{C}_{1,t}^2 - d_{T-g(T)}\right) \xrightarrow{d} \Lambda \qquad \textit{as} \qquad T \to \infty, \tag{2.22}$$

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

$$d_{T-g(T)} = 2\left\{\log[T-g(T)] - \frac{1}{2}\log\log[T-g(T)] - \log\pi\right\}, \tag{2.23}$$

and Λ *is a random variable distributed according to the Gumbel distribution (2.17).*

As a simple corollary, we can transform the test using (2.18) so that it may be compared against a more readily-available distribution.

Corollary 2.4.6. *Under the same assumptions,* $2 \cdot \exp(-\mathsf{SC}_T^2) \sim \chi_2^2$*. A test based on this result should reject for* small *values of the statistic.*

2.5 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 1-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:

$$\mathsf{C}_{1,t}^{2*} = G^{-1}[F(\mathsf{C}_{1,t}^2)] \tag{2.24}$$

where $F(\cdot)$ and $G(\cdot)$ are the $F(1, t-k-1)$ and χ^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. We find that 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). That is, we approximate the distribution of the maximum, $\max_{g(T)<t\leq T} \mathsf{C}^{2*}_{1,t}$, by

$$\Pr\left\{\max_{g(T)<t\leq T} \mathsf{C}^{2*}_{1,t} \leq x\right\} \approx \Pr\left\{\max_{g(T)<t\leq T} \varepsilon^2_{1,t} \leq x\right\} = [G(x)]^{T-g(T)}. \tag{2.25}$$

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.

2.6 Simulation study

We present the results a single simulation experiment regarding size, to test the asymptotic distribution derived above in small samples. The simulation uses the following data-generating process:

$$\begin{aligned} x_t &= \beta x_{t-1} + \varepsilon_t \qquad\qquad t = 1, \ldots, T, \qquad\qquad \text{(DGP1)} \\ \varepsilon_t &\overset{\text{iid}}{\sim} \mathsf{N}(0, 1), \\ x_0 &= 0, \end{aligned}$$

with β taking values in a range spanning stationary, unit-root and explosive processes, and with observation sequences of varying lengths T. Two model equations are evaluated, one containing an intercept

$$x_t = b_0 + b_1 x_{t-1} + u_t \qquad\qquad t = 1, \ldots, T, \qquad\qquad \text{(M1)}$$

and one excluding an intercept

$$x_t = b_1 x_{t-1} + u_t \qquad t = 1, \ldots, T. \qquad \text{(M2)}$$

The two sup-Chow tests evaluated are described in Theorem 2.4.5 and at (2.25) respectively; for the function $g(T)$ we use $T^{1/2}$. All tests are presented at a nominal size of 5% and 1%, reflecting the sizes most likely to be use for misspecification testing.

		Autoregressive coefficient (β)							
T		**-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

Table 2.1: Simulated rejection frequency (%) for SC^2, SC^{2*} and $\sup F$ under a Gaussian AR(1) process. Nominal size 5% and 1%. Number of MC repetitions, $M = 200000$. MCSE < 0.1.

As noted above, the SC^2 test is uniformly oversized, especially at the 1% size. The

SC^{2*} test is correctly sized and approximately similar, with simulated size varying 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.

2.7 Discussion

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 1-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 normality of residuals); rejection would draw the investigator'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: we anticipate the test will be less powerful against any particular alternative than tests which explicitly model that alternative (e.g. a single mean break). This motivates the use of multiple different tests, with failure of any one signalling misspecification and triggering further investigation. In real data sets, breaks may not be of the single mean-shift variety, and the parallel popularity of CUSUM-type and Andrews-type tests suggests both approaches have value.

The test is not invariant to the error distribution, even asymptotically—a feature it shares with many outlier tests. This issue is investigated in some detail in Chapter 4.

2.A Proofs

2.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 $S_{aa.b} = S_{aa} - S_{ab}S_{bb}^{-1}S_{ba}$. We use the spectral matrix norm, so $\|A\| = [\lambda_{\max}(AA')]^{1/2}$, where $\lambda_{max}(AA')$ indicates the largest eigenvalue of AA'.

2.A.2 Three-way process decomposition

We elaborate on the decomposition of the companion form (2.12) given in (2.13). 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 $\mathbf{M_3}$ to block diagonalize $\mathbf{S}$ into stationary, unit-root and explosive components:

$$\mathbf{M_3} S_t = (\mathbf{M_3}\mathbf{S}\mathbf{M_3}^{-1})\mathbf{M_3} S_{t-1} + \mathbf{M_3}(\xi_t', 0')',$$
$$\begin{pmatrix} \tilde{U}_t \\ Q_t \\ W_t \end{pmatrix} = \begin{pmatrix} \mathbf{U} & 0 & 0 \\ 0 & \mathbf{Q} & 0 \\ 0 & 0 & \mathbf{W} \end{pmatrix} \begin{pmatrix} \tilde{U}_{t-1} \\ Q_{t-1} \\ W_{t-1} \end{pmatrix} + \begin{pmatrix} e_{\tilde{U},t} \\ e_{Q,t} \\ e_{W,t} \end{pmatrix}, \tag{2.26}$$

where $\tilde{\mathbf{U}}$, $\mathbf{Q}$ and $\mathbf{W}$ have eigenvalues inside, on and outside the unit circle, respectively. We can now express the two-way decomposition presented at (2.13) as follows:

$$R_t = \begin{pmatrix} \tilde{U}_t \\ Q_t \end{pmatrix} \qquad \text{and} \qquad \mathbf{R} = \begin{pmatrix} \mathbf{U} & 0 \\ 0 & \mathbf{Q} \end{pmatrix}. \tag{2.27}$$

2.A.3 Preliminary asymptotic results

The ADL model (2.7) 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 (2.8) 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 (2.13), 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 2.A.1. *Suppose Assumptions 2.3.2, 2.3.1 and 2.3.3 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 $C_{\tilde{U}W} \overset{\text{a.s.}}{=} \text{o}(t^{\beta-1/2})$ and $C_{QW} \overset{\text{a.s.}}{=} \text{o}(t^{-\zeta/2})$ by N05

Theorems 9.1 and 9.2 respectively. The latter term will dominate since $\alpha > 16/7$ under Assumption 2.3.2.

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 analogous 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 2.3.2.

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 2.A.2. *Under Assumptions 2.3.2, 2.3.1 and 2.3.3 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) $\sum_{s=1}^{t-1} \varepsilon_s (R|W)'_s S_{RR}^{-1/2} \overset{\text{a.s.}}{=} \mathrm{O}[(\log t)^{1/2}] + \mathrm{o}(t^{\beta-1/16})$.

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 2.A.1(i). □

Lemma 2.A.3. *Under Assumptions 2.3.2, 2.3.1, 2.3.3 and 2.3.4,*

(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 2.A.2(i) and Lemma 2.A.1(ii). □

2.A.4 Proof of Theorem 2.4.1

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

$$\tilde{\varepsilon}_t = \frac{\varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s x'_s \left(\sum_{s=1}^{t-1} x_s x'_s\right)^{-1} x_t}{\left[1 + x'_t \left(\sum_{s=1}^{t-1} x_s x'_s\right)^{-1} x_t\right]^{1/2}}. \tag{2.28}$$

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

Lemma 2.A.4. *Under Assumptions 2.3.2, 2.3.1 and 2.3.3,*

$$x_t' S_{xx}^{-1} x_t - W' F_W^{-1} W = \mathrm{o}(t^{-\zeta}) \text{ a.s.} \tag{2.29}$$

for all $\zeta < 1/8$ *with* W *and* F_W *as in Theorem 2.4.1.*

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

We next state a lemma concerning the main numerator term in (2.28).

Lemma 2.A.5. *Under Assumptions 2.3.2, 2.3.1, 2.3.3 and 2.3.4*

$$\sum_{s=1}^{t-1} \varepsilon_s x_s' S_{xx}^{-1} x_t - G_t F_W^{-1} W = \mathrm{o}(t^{\beta - 1/8}) \text{ a.s.} \tag{2.30}$$

for all $\beta > 1/\alpha$*, where* W *and* F_W *are defined as in Theorem 2.4.1, and*

$$G_t = \sum_{s=1}^{t-1} \varepsilon_{t-s} W' (\mathbf{W}^{-s})' = \mathrm{o}(t^{\beta}) \text{ a.s. }. \tag{2.31}$$

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

$$\left\| \sum_{s=1}^{t-1} \varepsilon_s x_s' S_{xx}^{-1} x_t - \sum_{s=1}^{t-1} \varepsilon_s W_{s-1}' S_{WW}^{-1} W_{t-1} \right\| \leq \left\| \sum_{s=1}^{t-1} \varepsilon_s (\xi_2 | R, W)_s' S_{\xi\xi \cdot RW}^{-1} (\xi_2 | R, W)_t' \right\| \\ + \left\| \sum_{s=1}^{t-1} \varepsilon_s (R|W)_s' S_{RR \cdot W}^{-1} (R|W)_t \right\|, \tag{2.32}$$

and we consider each term on the right separately.

For the first term in (2.32) we use Lemma 2.A.1(iv) to write

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

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

For the second term in (2.32) we use Lemma 2.A.1(iii) to write

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

and then apply Lemma 2.A.2(iv) and Lemma 2.A.1(vii) to arrive at $\mathrm{o}(t^{\beta-1/8})$ 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

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

Many of these terms are familiar from the proof of Lemma 2.A.4, 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 $\mathrm{o}(t^\beta)$ since the latter two terms are bounded, while $\varepsilon_s = \mathrm{o}(s^\beta)$ by Lai and Wei (1985, Theorem 1). For the former term we have

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

where at the second last line we use that $\sum_{u=0}^{\infty} \mathbf{W}^{-u} e_{W,u+s} = \mathrm{o}(s^{\beta})$ 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 2.4.1. We aim to show that

$$\mathsf{C}_{1,t}^2 - (q_t/\sigma)^2 \overset{\text{a.s.}}{=} \mathrm{o}(1). \tag{2.33}$$

Using (2.6) we can rewrite this expression as

$$\frac{\tilde{\varepsilon}_t^2}{(t-k-1)^{-1}\mathrm{RSS}_{t-1}} - \left(\frac{q_t}{\sigma}\right)^2 = \tilde{\varepsilon}_t^2 \left[\frac{(t-k-1)}{\mathrm{RSS}_{t-1}} - \frac{1}{\sigma^2}\right] + \frac{\tilde{\varepsilon}_t^2 - q_t^2}{\sigma^2}. \tag{2.34}$$

We first consider the difference $\tilde{\varepsilon}_t^2 - q_t^2$. We have from (2.28),

$$\tilde{\varepsilon}_t^2 - q_t^2 = \frac{(\varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s x_s' S_{xx}^{-1} x_t)^2}{1 + x_t' S_{xx}^{-1} x_t} - \frac{(\varepsilon_t - G_t F_W^{-1} W)^2}{1 + W' F_W^{-1} W} = \frac{(\varepsilon_t - A_3)^2}{1 + A_1} - \frac{(\varepsilon_t - A_4)^2}{1 + A_2} \tag{2.35}$$

$$= \frac{A_2 - A_1}{(1 + A_1)(1 + A_2)} (\varepsilon_t - A_3)^2 + \frac{1}{1 + A_2} (A_4 - A_3)(2\varepsilon_t - A_4 - A_3), \tag{2.36}$$

where

$$A_1 = x_t' S_{xx}^{-1} x_t \qquad A_2 = W' F_W^{-1} W \qquad A_3 = \sum_{s=1}^{t-1} \varepsilon_s x_s' S_{xx}^{-1} x_t \qquad A_4 = G_t F_W^{-1} W. \tag{2.37}$$

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 2.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 2.A.5 respectively, while $A_4 - A_3$ is $\mathrm{O}(t^{\beta-1/8})$ by Lemma 2.A.5. So the first term of the sum is $\mathrm{o}(t^{2\beta-1/8})$ almost surely.

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

Thus the second term in (2.34) will vanish as long as $2\beta < 1/8$ or $\alpha > 16$ in Assumption 2.3.2, 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 proved, while

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

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

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

for $\gamma < 1/2$. So the first term in (2.34) will vanish as long as $2\beta < 1/2$, which is satisfied by Assumption 2.3.2.

2.A.5 Proof of Lemma 2.4.2

Proof. Theorem 2.4.1 shows that $\mathsf{C}_{1,t}^2 - q_t^2$ vanishes almost surely. Egorov's theorem (Davidson 1994, Theorem 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

$$\sup_{g(T) < t \le T} |\mathsf{C}_{1,t}^2 - q_t^2| \xrightarrow{\mathrm{P}} 0 \text{ as } T \to \infty.$$

□

2.A.6 Proof of Lemma 2.4.3 (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:

$$\Pr\{|X_1| < x\} = F(x) = \sqrt{2/\pi} \int_0^x e^{-t^2/2}\, dt = 2\Phi(x), \quad x \geq 0 \tag{2.38}$$

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

$$n(1 - F(u_n)) \to e^{-x} \qquad \text{as } n \to \infty. \tag{2.39}$$

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

$$1 - F(u) \sim f(u)/u \quad \text{as } u \to \infty, \tag{2.40}$$

so that combining (2.39) and (2.40) we have that $(1/n)e^{-x}u_n/f(u_n) \to 1$. Taking logs and substituting the density f, we have

$$-\log n - x + \log u_n - \tfrac{1}{2}\log(\pi/2) + u_n^2/2 \to 0. \tag{2.41}$$

Dividing through by $\log n$,

$$-1 - \frac{x}{\log n} + \frac{\log u_n}{\log n} - \frac{\log(\pi/2)}{\log n} + \frac{u_n^2}{2\log n} \to 0, \tag{2.42}$$

then for any fixed x, the second and fourth terms vanish trivially. The third term vanishes by substituting (2.39) 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),

$$2\log u_n - \log 2 - \log\log n \to 0. \tag{2.43}$$

Substituting this result into (2.41), we have that

$$-\log n - x + \tfrac{1}{2}\log 2 + \tfrac{1}{2}\log\log n - \tfrac{1}{2}\log(\pi/2) + u_n^2/2 \to 0. \tag{2.44}$$

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. □

2.A.7 Proof of Lemma 2.4.4

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 2.3.5 q_t/σ is an independent standard normal sequence, and q_t^2/σ^2 is an independent χ^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 2.3.5 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:

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

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 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

$$\max_t q_t^2/\sigma^2 < u_t \qquad \text{iff} \qquad \max_t |q_t/\sigma| < \sqrt{u_t} \tag{2.45}$$

where $|q_t/\sigma|$ has the half-normal distribution. Lemma 1 and Theorem 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

$$c_t(\max_{1\le s\le t} |q_s/\sigma| - d_t) \xrightarrow{\mathrm{d}} \Lambda \tag{2.46}$$

with

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

(Note that the centring sequence—here d_t, originally b_n—is incorrect in the original. A correction is provided as Lemma 2.4.3) Taking $\sqrt{u_t(z)} = c_t z + d_t$ and using (2.45) and (2.46), we have

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

giving norming sequences

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

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

2.A.8 Proof of Theorem 2.4.5

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

$$\begin{aligned} &\frac{1}{2}\left[\max_{g(T)\le t\le T}(\mathsf{C}_{1,t}^2) - d_{T-g(T)}\right] \\ &\quad\le \frac{1}{2}\max_{g(T)\le t\le T}\left(\mathsf{C}_{1,t}^2 - (q_t/\sigma)^2\right) + \frac{1}{2}\left[\max_{g(T)\le t\le T}(q_t/\sigma)^2 - d_{T-g(T)}\right] \end{aligned} \tag{2.47}$$

where the left term vanishes in probability by Lemma 2.4.2 and the right term converges in distribution by Lemma 2.4.4. 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.

Chapter 3

Simulations of the sup-Chow test

3.1 Introduction

In Chapter 2 we developed asymptotic theory for the maximum of the 1-step Chow statistics, providing an alternative to the usual graphical summaries of this sequence. This allowed the framing of a misspecification test, the sup-Chow test, which would provide summary evidence of misspecification of the type to which the 1-step statistics are sensitive, anywhere in the sample.

As a misspecification test in the sense of Section 1.1, this test is not conducted with a specific model as the alternative hypothesis. Instead, it is expected (or perhaps hoped) that the test will have power against a diffuse range of misspecification errors. In this chapter, we use Monte Carlo simulations to test that expectation.

Although the assumptions of Chapter 2 permit general vector autoregressive (VAR) processes with multiple lags and deterministic terms, our simulation experiments involve only AR(1) processes (occasionally reducing to simple white noise) and corresponding AR(1) regressions. It is not possible to explore the universe of permissible models by simulation, and a principle of parsimony suggests that the simplest dynamic models are an appropriate starting point, and may illuminate general properties of the test, simplicity notwithstanding.

The alternatives we explore are necessarily arbitrary. Such exploration is nonethe-

Technical details of the simulations, and an index of the DGPs (DGPx) and models (Mx) used is proved in Appendix A less valuable. As Cox and Hinkley (1974, p. 65) note in the context of test of pure significance, 'without some idea of what are meaningful departures from H_0, the problem of testing consistency with it is meaningless.' In generating alternatives, we first take inspiration from the history of the recursive Chow statistics (described in 1.2), which like the Chow (1960) tests are generally seen as assessing parameter constancy. The processes we consider are parameterized by at most a constant term, an autoregressive coefficient, and an error variance, and so it is these parameters we vary. We take additional inspiration from the structure of the 1-step statistic, an approximation of a single unobserved error, and consequently also investigate alternatives which directly influence a single error—in other words, outliers.

As indicated in the Chapter 1, we compare, where appropriate, the sup-Chow test with two existing tests for structural instability at an unknown time (Andrews 1993, 2003*a*). This provides some sense of the comparative performance of the sup-Chow test, although we caution that the limited selection of processes and alternatives means this comparison is necessarily incomplete.

The chapter unfolds as follows. Section 3.2 considers, as a preliminary matter, the finite sample size of the test, and various corrections to improve size. (After this section, all tests are corrected for true size.) Section 3.3 evaluates the power of the test against changes in the model parameters. Section 3.4 investigates power against a single outlier. Section 3.5 considers the performance of the test in comparison with Andrews (1993) and Andrews (2003*a*). The chapter concludes with a brief discussion in Section 3.6.

Finally, in this chapter we consider the test only under the maintained assumption of normally-distributed errors (Assumption 2.3.5). We consider departures from this assumption, and various methods of addressing such departures, in Chapter 4.

3.2 Size of the test

In determining critical values and rejection regions for tests, we work with the asymptotic distribution as derived in Theorem 2.4.5 of Chapter 2. Such distributions arise as limits when the sample size grows without bound; hence they are only approximate for finite sample sizes, as are critical values calculated from them. Using an exact distribution, we would expect a test to produce Type I (rejecting under the null) errors in the proportion of the nominal size of the test (typically 1%, 2.5%, 5% or 10%). With the asymptotic distribution, this will not be the case, and examining the true rejection frequency corresponding to a particular nominal size provides insight into the accuracy of the approximation and speed of convergence.

We can further use size simulations to confirm that properties of the test established asymptotically also hold in finite samples (for example, invariance to regression parameters).

3.2.1 The asymptotic test

We first validate by simulation the results obtained analytically in Chapter 2. There we showed that, under a well-specified ADL model with normal errors, the sup-Chow statistic converges in distribution to the Gumbel extreme value distribution. This convergence did not depend upon the roots of the process, and neither did the limit distribution, and so the same distribution should apply to stationary, unit-root and explosive processes.

To test this result, we consider a simple AR(1) process

$$\begin{aligned} x_t &= \beta x_{t-1} + \varepsilon_t \qquad\qquad t = 1, \dots, T, \qquad\qquad \text{(DGP1)} \\ \varepsilon_t &\overset{\text{iid}}{\sim} N(0,1), \\ x_0 &= 0, \end{aligned}$$

with β taking values in a range spanning stationary, unit-root and explosive processes, and with observation sequences of varying lengths T. The model equation is

$$x_t = b_0 + b_1 x_{t-1} + u_t \qquad\qquad t = 1, \dots, T. \qquad\qquad \text{(M1)}$$

The nominal size is set at 5%. The results are given in Table 3.1. (As for all the simulation experiments reported, the number of Monte Carlo repetitions, M, and Monte Carlo standard error, MCSE, is given in the table.)

It is immediately apparent that the sup-Chow test is over-sized (that is, it will reject

	$\beta =$								
T	**-1.03**	**-1.00**	**-0.90**	**-0.50**	**0.00**	**0.50**	**0.90**	**1.00**	**1.03**
50	12.8	12.7	12.5	12.3	12.6	13.5	16.1	17.0	17.4
100	10.4	10.4	10.2	10.2	10.4	10.7	12.3	13.3	13.8
200	8.7	8.5	8.4	8.4	8.4	8.7	9.6	10.2	10.6
500	6.8	6.7	6.7	6.6	6.7	6.8	7.1	7.6	7.6
1000	6.8	5.8	5.8	5.8	5.8	5.8	6.0	6.3	7.7

Table 3.1: Simulated rejection frequency (%) for SC^2 statistic under (DGP1)/(M1). Nominal size 5%. $M = 200000$. MCSE < 0.1.

too often) up until the length of the series $T = 1000$. Moreover, the disparity between the nominal and true size depends on T, suggesting that a further finite sample correction may be of use. We consider a number of variants designed to pragmatically address this issue in the following section.

On the other dimension considered, the test is reasonably well-behaved: true size does not vary greatly with β, except in the positive near-unit-root through explosive range (and this discrepancy diminishes with sample size as the theory implies it must). This near-similarity is a useful feature: were it not the case, and size depended on the autoregressive parameter, then a circularity would ensue between model estimation and model specification testing, which would prove difficult to overcome.

3.2.2 Finite sample variants

In this section we consider three finite-sample corrections, motivated by the over-size of the sup-Chow test in small and medium finite samples. Although it is possible to address this precisely by simulating true critical values for tests (as we do when considering power), this requires considerable computation, as it must be done for each model and significance level. In practice it is more useful to have simple size-corrections, even if these are not perfect, leaving some residual size discrepancy.

The variants we propose and test are inspired by combining the asymptotic results obtained for general processes, with finite sample results that would hold under stricter assumptions, and which might be expected to hold approximately without these assumptions. Figure 3.1 gives an overview of how these variants are constructed from the basic sup-Chow test.

To construct the first variant ($\mathsf{SC}^{2\dagger}$), we note that asymptotically, the studentized squared forecast residuals approximate the normalised squared errors, which are independent χ^2 distributed (by assumption). However in practice for finite samples, the 1-step statistics appear to be distributed close to $\mathsf{F}(1, t-k-1)$ (as indeed they are exactly absent dependence, with fixed regressors). We cannot construct an extreme-value test directly on this basis, as that theory requires identically distributed random variables; however, we can use an inverse-probability transform to map the F-distributed 1-step statistic into χ^2 random variables, and then apply the maximum, centring and scaling as before. As the sequence of $F(1, t)$ distributions converges to χ^2_1, this correction has no asymptotic effect, but should provide a finite sample improvement.

Specifically, with F the CDF for a random variable with an (Snedecor) F distribution with $(1, t-k-1)$ degrees of freedom and G the distribution function for a random variable with a χ^2_1 distribution, we define the adjusted 1-step statistic

$$\mathsf{C}^{2*}_{1,t} = G^{-1}[F(\mathsf{C}^2_{1,t})], \tag{3.1}$$

and the corresponding supremum statistic

$$\mathsf{SC}^{2\dagger}_T = \frac{1}{2}\left(\max_{g(T)<t\leq T} \mathsf{C}^{2*}_{1,t} - d_{T,g(T)}\right), \tag{3.2}$$

(with other quantities as defined in Theorem 2.4.5).

For the second variant ($\mathsf{SC}^{2\ddagger}$), we replace the extreme value results used with finite sample equivalents. As a distribution under the null, we use 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). That is, we approximate the distribution of

	Not F-corrected	F-corrected
Extreme value distribution	SC^2	$\mathsf{SC}^{2\dagger}$
Finite maximal distribution	$\mathsf{SC}^{2\ddagger}$	SC^{2*}

Figure 3.1: Construction of the three variant tests.

		$\beta =$				
	T	**0.00**	**0.50**	**0.90**	**1.00**	**1.03**
SC^2	**50**	12.6	13.5	16.1	17.0	17.4
	100	10.4	10.7	12.3	13.3	13.8
	200	8.4	8.7	9.6	10.2	10.6
	500	6.7	6.8	7.1	7.6	7.6
	1000	5.8	5.8	6.0	6.3	7.7
$\mathsf{SC}^{2\dagger}$	**50**	3.9	4.1	5.1	5.5	5.7
	100	4.0	4.1	4.7	5.2	5.4
	200	4.2	4.2	4.6	4.9	5.1
	500	4.3	4.3	4.4	4.8	4.7
	1000	4.3	4.4	4.4	4.6	6.0
$\mathsf{SC}^{2\ddagger}$	**50**	14.8	15.8	18.7	19.7	20.2
	100	12.0	12.5	14.3	15.4	16.0
	200	9.7	10.0	11.0	11.8	12.2
	500	7.7	7.8	8.1	8.7	8.7
	1000	6.6	6.7	6.8	7.1	8.6
SC^{2*}	**50**	5.0	5.4	6.5	7.0	7.3
	100	5.0	5.1	5.8	6.4	6.7
	200	5.0	5.1	5.5	5.9	6.1
	500	5.1	5.1	5.2	5.6	5.5
	1000	5.0	5.0	5.1	5.3	6.7

Table 3.2: Simulated rejection frequency (%) for sup-Chow variants under (DGP1)/(M1). Nominal size 5%. $M = 200000$. MCSE < 0.1.

the maximum, $\max_{g(T)<t\leq T} \mathsf{C}^2_{1,t}$, by

$$\Pr\left\{\max_{g(T)<t\leq T} \mathsf{C}^2_{1,t} \leq x\right\} \approx \Pr\left\{\max_{g(T)<t\leq T} \varepsilon^2_{1,t} \leq x\right\} = [G(x)]^{T-g(T)}, \tag{3.3}$$

where again G is the distribution function for a random variable with a χ^2 distribution with one degree of freedom. Note that in this case no centring or scaling is required, because the distribution itself depends on T.

The third and final variant (SC^{2*}) considered in this section simply combines the two separate corrections used previously. We use the approximation

$$\Pr\left\{\max_{g(T)<t\leq T} \mathsf{C}^{2*}_{1,t} \leq x\right\} \approx \Pr\left\{\max_{g(T)<t\leq T} \varepsilon^2_{1,t} \leq x\right\} = [G(x)]^{T-g(T)}. \tag{3.4}$$

The experiment is conducted with (DGP1) and (M1), as before, and results are reported in Table 3.2, for a nominal size of 5%.

The results demonstrate that the two corrections have offsetting effects. The F-correction of $\mathsf{SC}^{2\dagger}$ has a dramatic effect, but somewhat over-corrects leading to an under-sized test which slowly approaches the nominal 5% size from below. Moving from the extreme value to the finite maximal distribution in $\mathsf{SC}^{2\ddagger}$ has the opposite effect, slightly increasing the rejection frequency for all T. The combination in SC^{2*} results in a test that is well-sized at all sample sizes for stationary processes, and somewhat oversized with non-stationary processes at small sample sizes (but considerably improved from the asymptotic test). Of the four variants, this is our preferred test.

3.3 Power against structural breaks

We consider three specific types of structural break that are applicable to the stationary AR(1) case. First, we consider a change in the process mean created by a change in the constant term. Second, we consider a change in the autoregressive coefficient that does not affect the long run process mean. Third, we consider a change in the variance of the error term.

We evaluate only the preferred test SC^{2*} under these experiments.

3.3.1 Process mean

Clements and Hendry (1998, p. 49) observe that for forecasting, it is breaks in the process mean that matter most. We model a break in the mean of an AR(1) process directly as follows:

$$\begin{aligned} x_t &= \gamma I(t > \tau T) + \varepsilon_t \qquad\qquad t = 1, \ldots, T, \qquad\qquad \text{(DGP2)} \\ \varepsilon_t &\overset{\text{iid}}{\sim} N(0, 1), \\ x_0 &= 0, \end{aligned}$$

where x_t is the process observed, τT is the break time and I is an indicator function. We use model (M1) once again. Since there is no autoregressive term in the DGP, the change in the constant term translates directly into a shift in the process mean. As the error term is standard normal, the mean shift γ can be viewed in units of the pre-break error (or process) standard deviation.

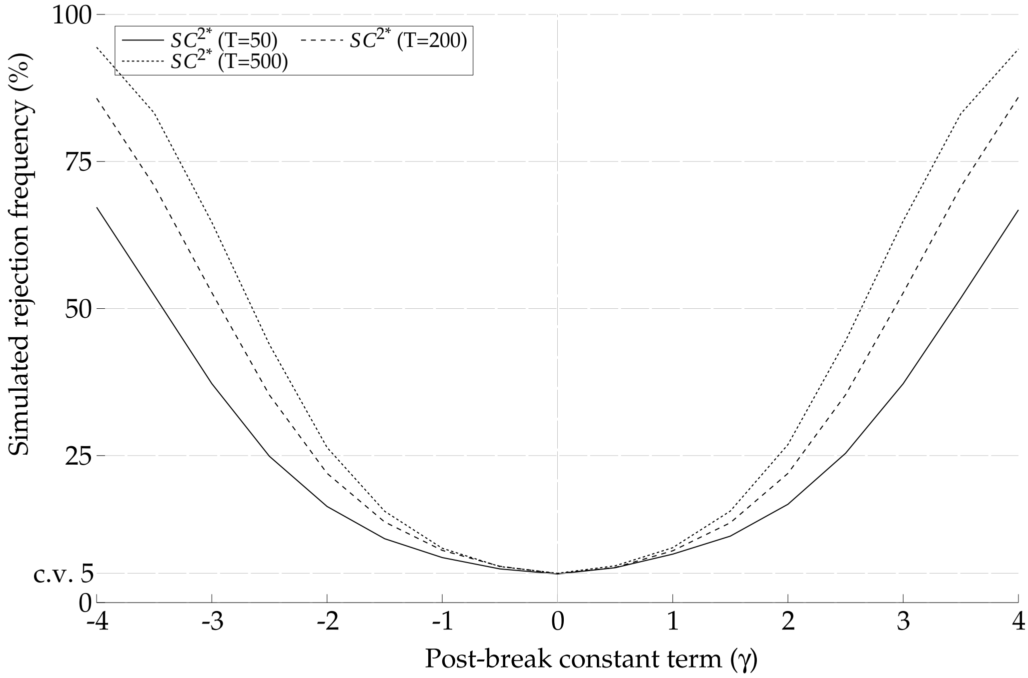

Figure 3.2: Simulated rejection frequency (%) with a single break in the constant term (and process mean) at $\tau = 0.5$ (sample mid-point), under (DGP2)/(M1). Nominal size 5%. $M = 20000$.

We model data generating processes of this type in the experiment reported in Figures 3.2 and 3.3.

Figure 3.2 shows that SC^{2*} does have power against a break in the process mean. A relatively large break is required before the rejection rate exceeds 50%.

Figure 3.3 shows how the test's power varies depending on the timing of the mean break. The shape of the curve indicates that the test has peak power when the break occurs somewhere beyond 90% of the way through the sample. The very lower power against early breaks in the smallest sample ($T = 25$) reflects the left-trimming required for the convergence argument in Theorem 2.4.5. With the left-trimming point set at $g(T) = T^{1/2}$, $\sqrt{50} \approx 7$ 1-step statistics will be ignored at the start of the sample; the effect is to mask breaks occurring at 10% ($t = 5$) or 20% ($t = 10$) of the sample size.

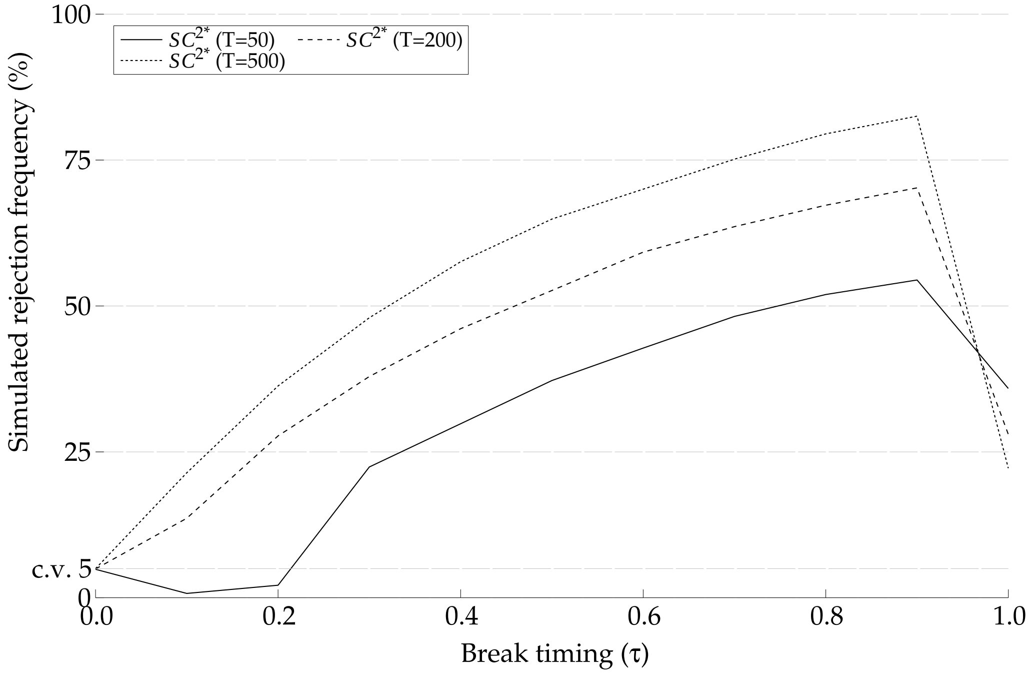

Figure 3.3: Simulated rejection frequency (%) with a single break in the constant term of $\gamma = 3$, occurring at $t = \tau T$, under (DGP2)/(M1). Nominal size 5%. $M = 20000$.

3.3.2 Autoregressive coefficient

We model a break in the autoregressive coefficient of an AR(1) process as follows:

$$x_t = \Delta\beta I(t \geq \tau T)x_{t-1} + \varepsilon_t \qquad t = 1, \ldots, T, \qquad \text{(DGP3)}$$
$$\varepsilon_t \overset{\text{iid}}{\sim} N(0, 1),$$
$$x_0 = 0,$$

where x_t is the process observed, τT is the break time and I is an indicator function. We use model (M1). In this experiment, we set the autoregressive coefficient before the break to zero, so that the pre-break process is simply the error process, whereas the process becomes AR(1) after time τT, with a coefficient of $\Delta\beta$.

We model data generating processes of this type in the experiment reported in Figures 3.4 and 3.5.

Figure 3.4 shows that the SC^{2*} does have some power against a break in the autoregressive coefficient, but perhaps not useful power. In the experiment, rejection rates do

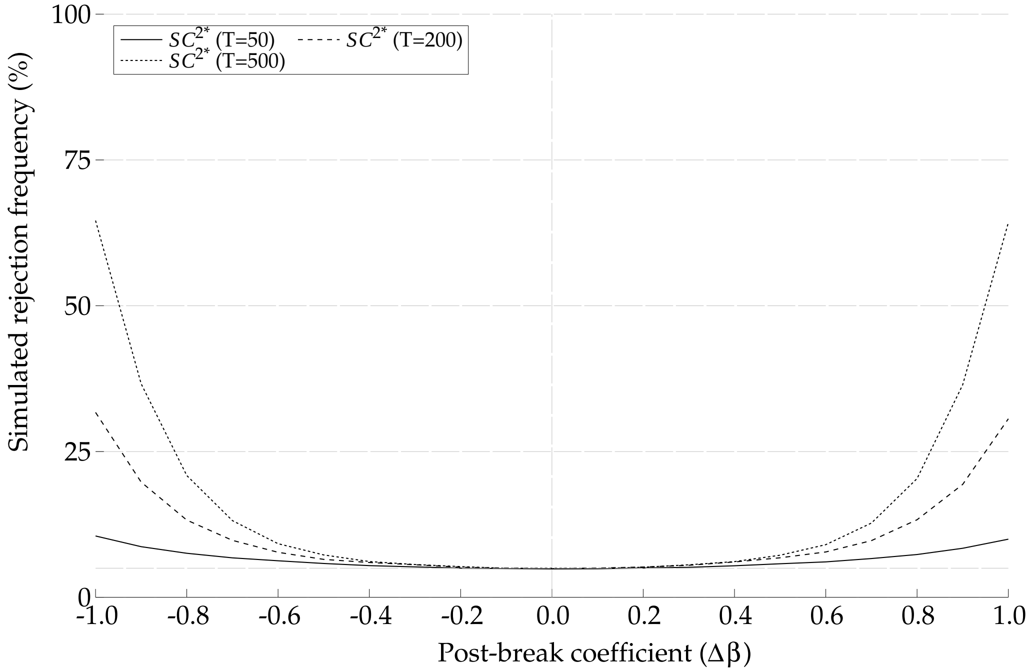

Figure 3.4: Simulated rejection frequency (%) with a break in the AR(1) coefficient of $\Delta\beta$ at $\tau = 0.5$ (sample mid-point), under (DGP3)/(M1). Nominal size 5%. $M = 20000$.

not substantially exceed the size of the test until the magnitude of the change $|\Delta\beta|$ exceeds 0.6. This is consistent with findings by Clements and Hendry (1999) and Hendry (2000), that changes of this type are difficult to detect in general, unless they have an impact on long-run mean (in this experiment they do not).

Figure 3.5 shows the effect of break timing on the power of the test, choosing a relatively large break magnitude ($\Delta\beta = 0.8$). It is apparent that power peaks when the break is around 80% of the way through the sample, regardless of sample size, with no power to detect breaks at either end of the sample.

In order to confirm the hypothesis that the low power results from the inability of the autoregressive coefficient to influence the process mean, we run an additional experiment. The DGP is as before, but with the addition of a unit constant term.

$$x_t = 1 + \Delta\beta I(t \geq \tau T)x_{t-1} + \varepsilon_t \qquad t = 1, \ldots, T, \qquad \text{(DGP4)}$$

$$\varepsilon_t \overset{\text{iid}}{\sim} N(0, 1),$$

$$x_0 = 0,$$

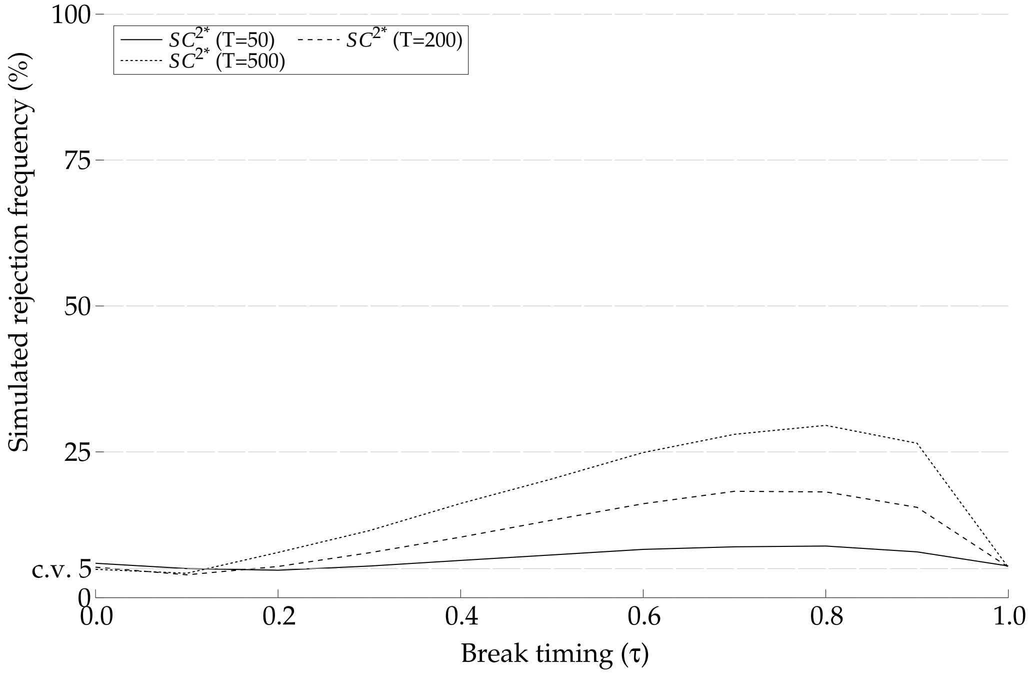

Figure 3.5: Simulated rejection frequency (%) with a break in the AR(1) coefficient of $\Delta\beta = 0.8$ at time τT, under (DGP3)/(M1). Nominal size 5%. $M = 20000$.

where x_t is the process observed, τT is the break time and I is an indicator function. We use model (M1).

Under this DGP we can express the process mean as

$$\mathsf{E}x_t = \frac{1}{1 - \Delta\beta I(t \geq \tau T)} = \begin{cases} 1, & \text{for } t < \tau T, \\ (1-\Delta\beta)^{-1}, & \text{for } t \geq \tau T, \end{cases}$$

so in effect this experiment tests both the change in the parameter, and also the resulting change in the process mean.

The effect of including a constant term is substantial, and can be seen by comparing Figure 3.4 with Figure 3.6.

3.3.3 Error variance

A change in the variance of the error term in a stochastic process may occur, for instance, if variability of unmodelled inputs increases. We model a break in the variance of the

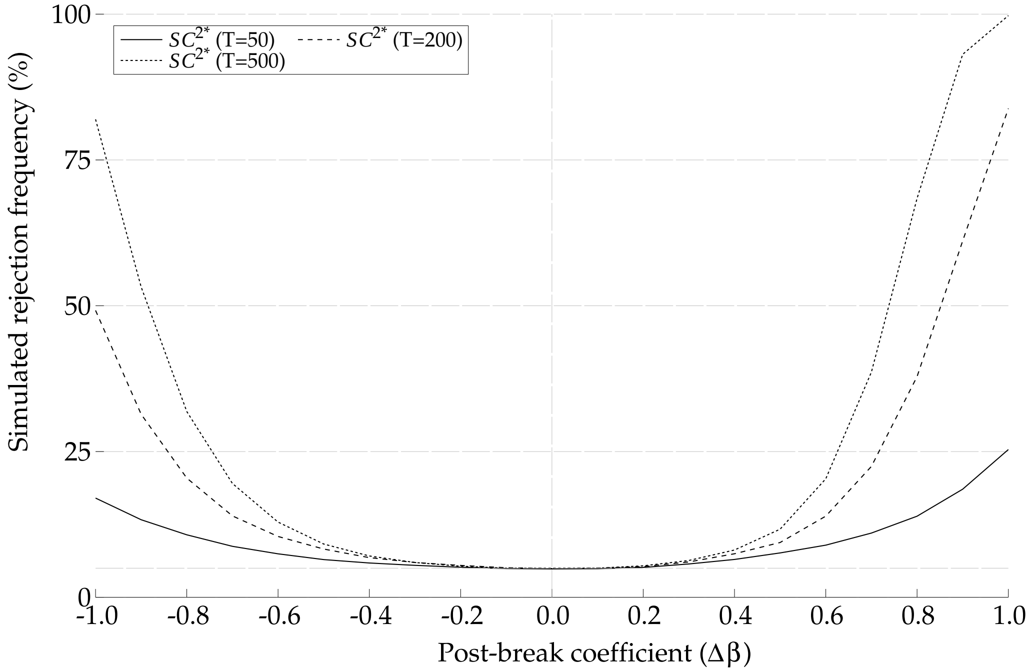

Figure 3.6: Simulated rejection frequency (%) with a break in the AR(1) coefficient of $\Delta\beta$ at $\tau = 0.5$ (sample mid-point), under (DGP4)/(M1). Nominal size 5%. $M = 20000$.

error term of an AR(1) process directly, as follows:

$$x_t = \{1 + I(t \geq \tau T)(\sigma - 1)\}\varepsilon_t \qquad t = 1, \ldots, T, \qquad \text{(DGP5)}$$
$$\varepsilon_t \overset{\text{iid}}{\sim} N(0, 1),$$
$$x_0 = 0,$$

where x_t is the process observed, τT is the break time and I is an indicator function. We include no constant or autoregressive terms; the process is simply white noise. We use model (M1). In this experiment, the error variance before the break is unity, whereas after the break it changes to σ^2. It remains normally distributed with mean zero throughout.

Figures 3.7 and 3.8 show the results of the experiment modelling data generating processes of this type. We see that the SC^{2*} test has power against a step change in the error variance, with relatively large gains in power for all three sample sizes as the magnitude of the changes rises to 3 or 4. In contrast with the other structural break experiments, power of the SC^{2*} test as a function of break time (Figure 3.8) shows a

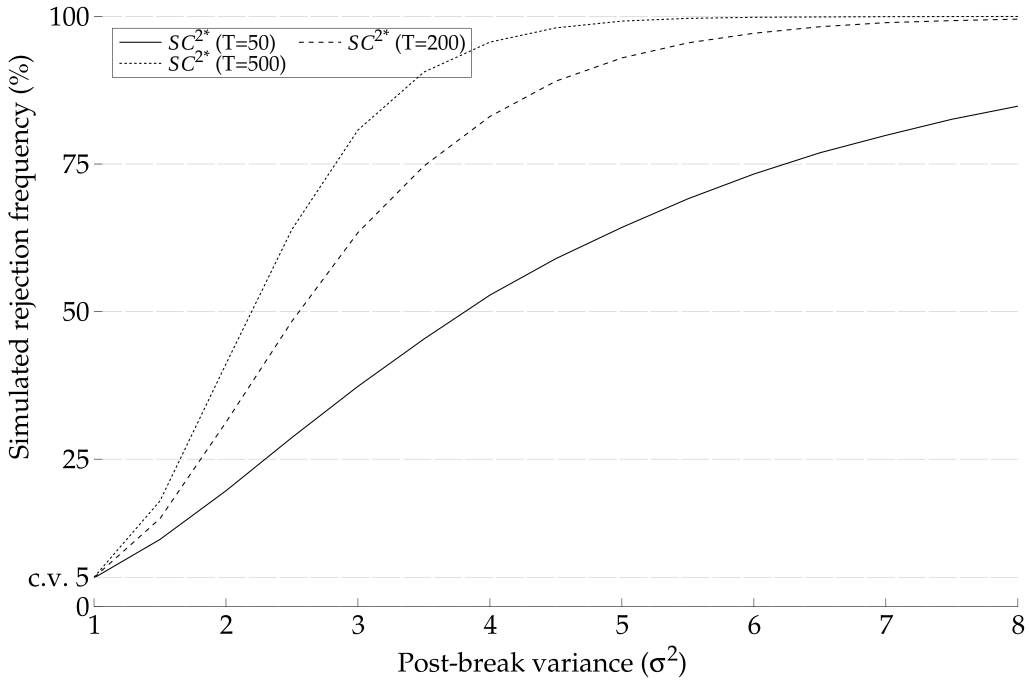

Figure 3.7: Simulated rejection frequency (%) with a break in the error variance from 1 to σ^2 at $\tau = 0.5$ (mid-sample), under (DGP5)/(M1). Nominal size 5%. $M = 20000$.

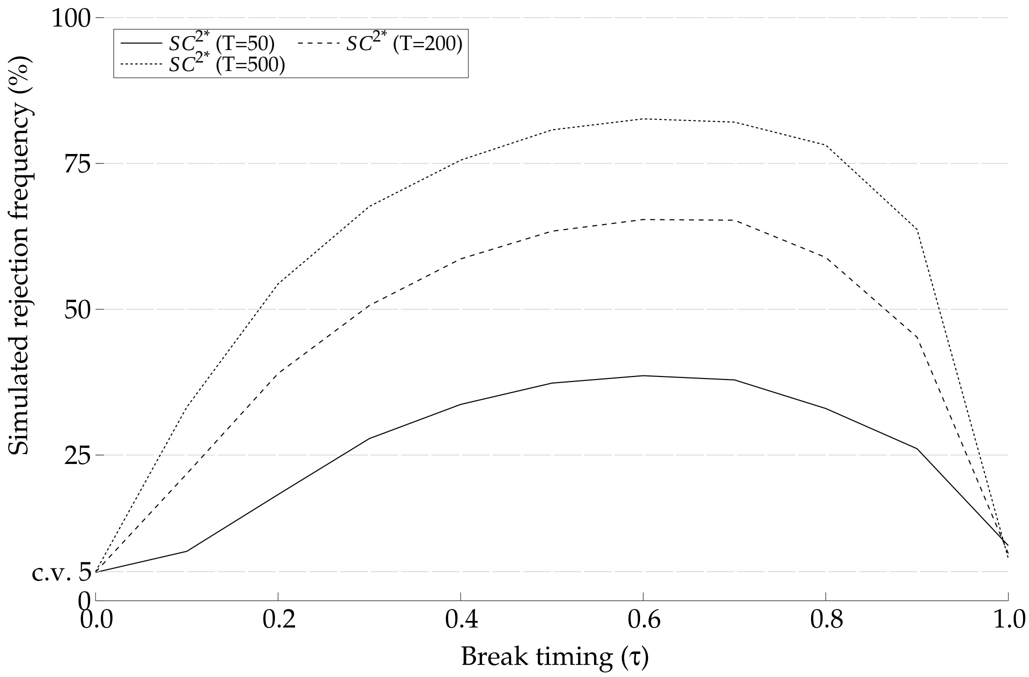

Figure 3.8: Simulated rejection frequency (%) with a break in the error variance from 1 to $\sigma^2 = 3.0$, at time τT, under (DGP5)/(M1). Nominal size 5%. $M = 20000$.

much more curved shape for a change in error variance. The curve also peaks somewhat earlier—a change in error variance is most likely to be detected if it appears around 60–70% through the sample, for the sample sizes considered.

3.4 Power against outliers

Fox (1972) considers outliers in time series, dividing these into two categories. Additive (Fox's 'Type I') outliers lie outside the time series structure of a process, affect only a single observation, and correspond to genuine measurement errors in an experimental framework. Innovation (Fox's 'Type II') outliers lie within the time series structure of a process, and so affect not only an initial observation, but also subsequent observations. In order to understand the properties of the sup-Chow test, we examine power against both types of outlier.

3.4.1 Additive outliers

Additive outliers (Fox's type I) affect only a single observation. In the stationary AR(1) framework, we model this as follows:

$$\begin{aligned} x_t^* &= \beta x_{t-1}^* + \varepsilon_t \qquad\qquad t = 1, \ldots, T, \qquad \text{(DGP6)} \\ \varepsilon_t &\overset{\text{iid}}{\sim} N(0,1), \\ x_0^* &= 0, \\ x_t &= x_t^* + \delta I(t = \lfloor \tau T \rfloor + 1), \end{aligned}$$

where x_t is the process observed (with error), x_t^* is the true (outlier-free) process, $\lfloor \tau T \rfloor + 1$ is the outlier time and I is an indicator function. We use $\beta = 0.5$, to ensure that a distinction exists between an additive outlier and an innovation outlier. We use model (M1).

We model data generating processes of this type in the experiment reported in Figures 3.9 and 3.10. The first of these shows rejection frequency as a function of outlier magnitude (δ), with the outlier time fixed at mid-sample ($\tau = 0.5$), for a range of sample sizes. The second of these shows rejection frequency as a function of outlier time (τ),

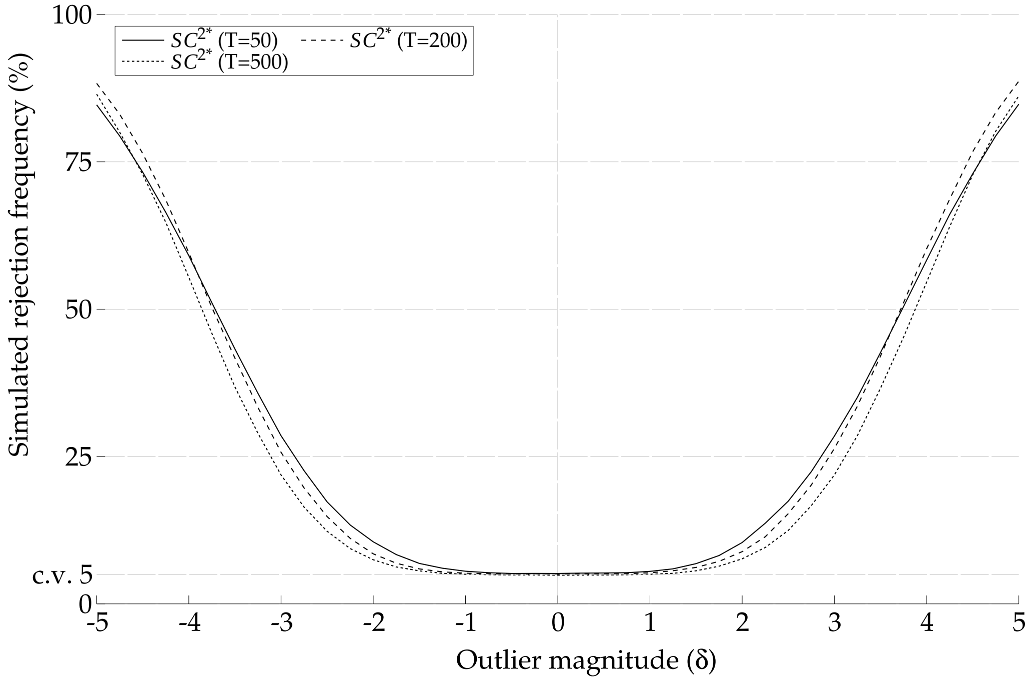

Figure 3.9: Simulated rejection frequency (%) with an additive outlier at $\tau = 0.5$ (mid-sample), under (DGP6)/(M1). Nominal size 5%. Magnitude relative to error variance of 1.0. $M = 20000$.

with the outlier magnitude fixed at a magnitude of $\delta = 4$. As the error standard deviation is set at unity, the outlier magnitudes can be interpreted in standard-deviation units.

Several features are apparent from Figure 3.9. First, the test power *decreases* slowly with sample size. This is expected, since even in the absence of outliers, errors of a given magnitude will become more likely the longer the series. For this reason, outlier tests in general do not have power asymptotically. Second, the test has no appreciable power for outliers less than 2 in magnitude. Again, this is not surprising, given that the test is set at a level of 5% and the errors are normal. Outliers below this magnitude are simply indistinguishable from non-outlier errors at this level.

The results of the timing experiment (Figure 3.10) are quite similar to the mean break case (Figure 3.3). Again, the low early power with small samples largely reflects the left-trimming parameter (using a rule of $g(T) = T^{1/2}$).

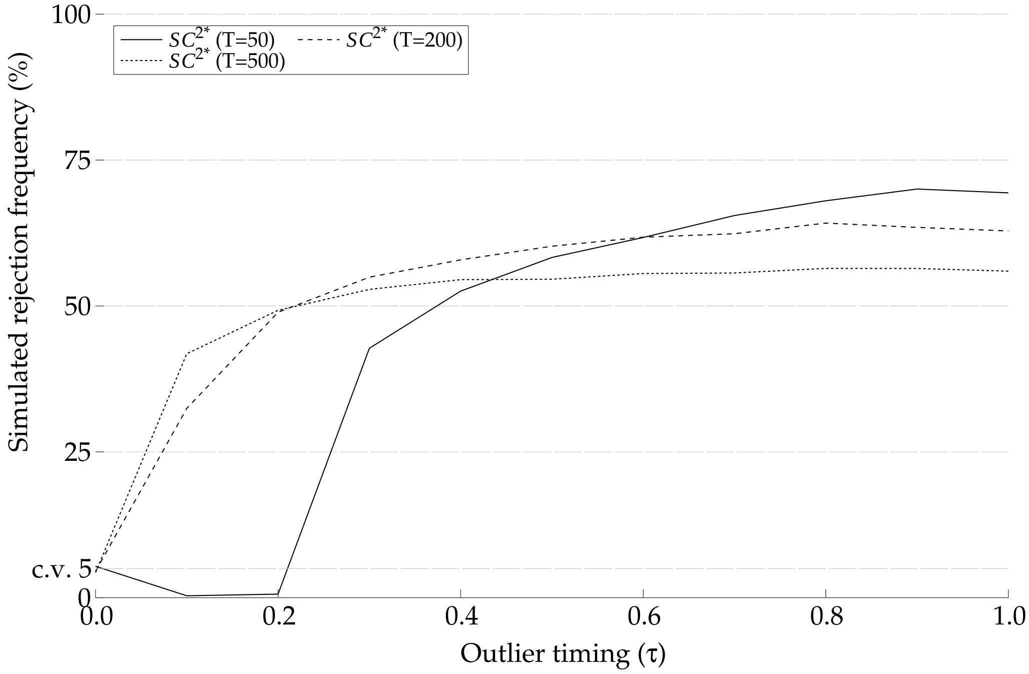

Figure 3.10: Simulated rejection frequency (%) with an additive outlier of magnitude $\delta = 4.0$, at time $t = \lfloor \tau T \rfloor + 1$, under (DGP6)/(M1). Nominal size 5%. $M = 20000$.

3.4.2 Innovation outliers

Innovation outliers (Fox's type II) occur at one observation, but are then integrated into subsequent observations through the autocorrelation of the process. In the stationary AR(1) framework, we model this as follows:

$$x_t = \beta x_{t-1} + \delta I(t = \lfloor \tau T \rfloor + 1) + \varepsilon_t \qquad t = 1, \ldots, T, \tag{DGP7}$$
$$\varepsilon_t \overset{\text{iid}}{\sim} N(0,1),$$
$$x_0 = 0,$$

where x_t is the process observed, $\lfloor \tau T \rfloor + 1$ is the outlier time and I is an indicator function. This differs from the additive outlier scenario in the effect of the outlier on

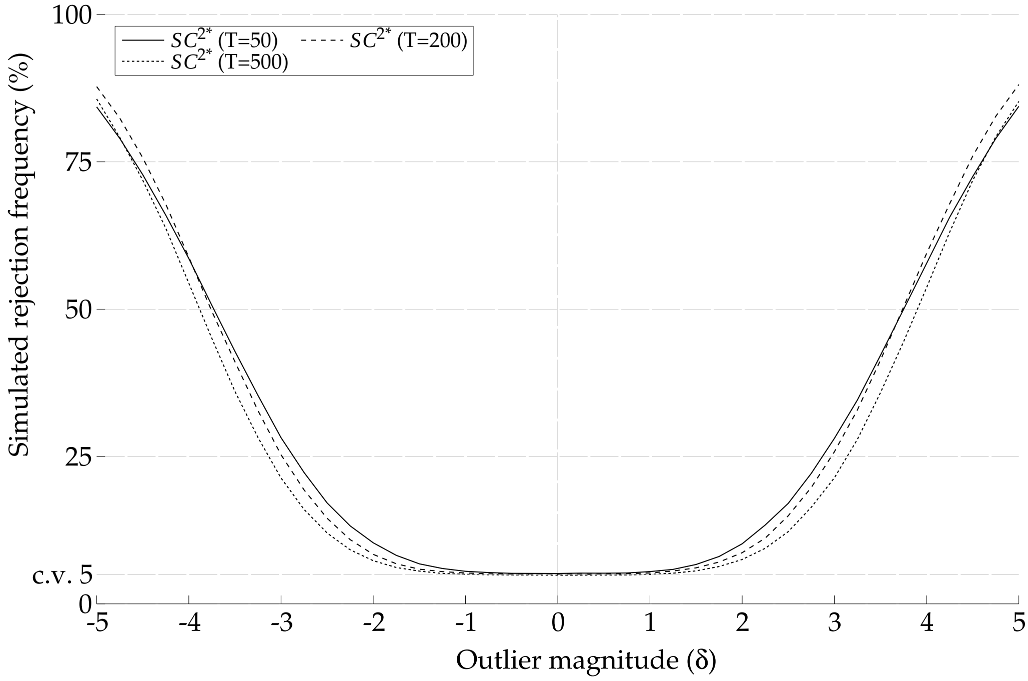

Figure 3.11: Simulated rejection frequency (%) with an innovation outlier of magnitude δ at $\tau = 0.5$ (mid-sample), under (DGP7)/(M1). Nominal size 5%. Magnitude relative to error variance of 1.0. $M = 20000$.

observations for $t > \lfloor \tau T \rfloor + 1$. In particular, we have

$$
\begin{aligned}
x_{\lfloor \tau T \rfloor + 1} &= \beta x_{\lfloor \tau T \rfloor} + \delta + \epsilon_{\tau T} \\
x_{\lfloor \tau T \rfloor + 2} &= \beta^2 x_{\lfloor \tau T \rfloor} + \beta\delta + \beta\epsilon_{\lfloor \tau T \rfloor + 1} + \epsilon_{\tau T + 1} \\
&\vdots
\end{aligned}
$$

so that the outlier, η is incorporated into subsequent observations with geometrically declining—but non-zero—effect.

We model data generating processes of this type in the experiment reported in Figures 3.11 and 3.12. These correspond to Figures 3.9 and 3.10 in the previous section and bear similar interpretation.

The general features of Figure 3.11 are similar to Figure 3.9, suggesting that the tests are similarly powerful against either additive or innovation outliers. The results of the timing experiment for innovation outliers (Figure 3.12) are very similar to those for additive outliers.

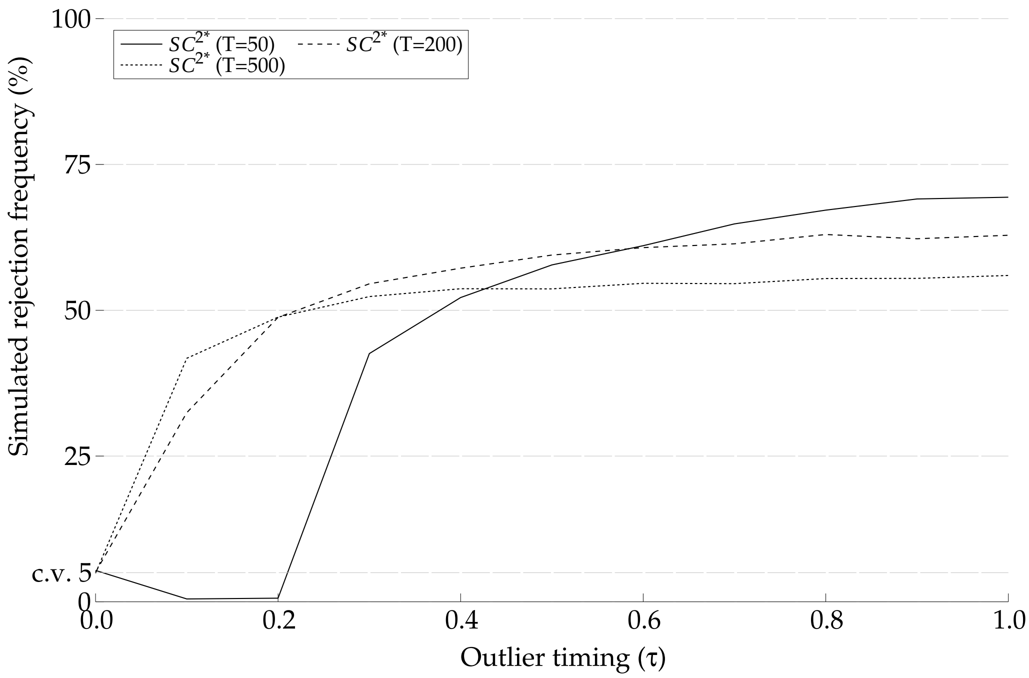

Figure 3.12: Simulated rejection frequency (%) with an innovation outlier of magnitude $\delta = 4.0$, at time $t = \lfloor \tau T \rfloor + 1$, under (DGP7)/(M1). Nominal size 5%. $M = 20000$.

3.5 Comparisons

We consider comparisons of size-adjusted power against the test of Andrews (1993) and Andrews (2003*a*), described in Sections 1.4.1 and 1.4.2 respectively. The former, the $\sup F$ test, is a commonly-used test of structural change at an unknown changepoint, and so is an obvious candidate against which to compare the SC^{2*} test. The latter, the S test, is a complementary test, designed to address the weak power of the $\sup F$ test near the sample endpoints.

Both tests have free parameters, which can be specified by the investigator to reflect either a prior suspicion about the timing of structural instability, or to reflect ignorance of the same. In the case of the $\sup F$ test, the parameters are the end points of a trimming region $(a, b) \subset [0, 1]$, over which the supremum of pointwise statistics is taken. We assume no prior information about the timing of structural change (in line with the motivation of general misspecification testing), and so adopt $(a, b) = (0.15, 0.85)$ as recommended in Andrews (1993). In the case of the S test, the starting point of the end-of-sample must be specified. Andrews (2003*a*) provides no guidance on this, and

		Break timing (τ)								
		0.5T			T-2			T-1		
		Post-break constant (γ)								
T		**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**
25	SC^{2*}	5.1	13.2	50.4	5.1	20.4	78.0	5.1	15.3	69.9
	$F_{0.5}$	4.9	95.4	99.9	4.9	5.8	9.0	4.9	4.8	4.4
	$\sup F$	5.0	78.5	99.4	5.0	11.0	29.3	5.0	7.2	18.6
	$S_{0.15}$	3.6	1.6	0.3	3.6	22.6	43.9	3.6	14.1	36.6
	S_1	6.7	8.6	17.6	6.7	41.1	88.7	6.7	50.5	96.7
50	SC^{2*}	4.9	16.5	67.1	4.9	18.2	82.9	4.9	12.9	69.7
	$F_{0.5}$	5.0	100.0	100.0	5.0	6.3	14.7	5.0	5.0	5.0
	$\sup F$	5.0	99.8	100.0	5.0	8.7	31.8	5.0	5.9	10.1
	$S_{0.15}$	4.0	0.7	0.0	4.0	12.6	32.7	4.0	7.6	15.4
	S_1	5.4	8.1	21.3	5.4	44.7	94.5	5.4	49.7	97.1
100	SC^{2*}	4.9	19.4	77.4	4.9	15.8	84.1	4.9	11.1	67.4
	$F_{0.5}$	4.9	100.0	100.0	4.9	5.8	14.3	4.9	4.9	4.8
	$\sup F$	4.9	100.0	100.0	4.9	7.2	31.2	4.9	5.4	7.3
	$S_{0.15}$	4.5	0.1	0.0	4.5	9.5	34.0	4.5	6.1	10.1
	S_1	4.7	7.5	22.0	4.7	46.5	96.2	4.7	49.0	97.3

Table 3.3: Simulated rejection frequency (%) with a single break in the constant term (and process mean) of γ at τ, under (DGP2)/(M1). Nominal size 5%. $M = 200000$. MCSE < 0.2.

we consider two extremes: $S_{0.15}$, for which the final 15% of the sample is tested, and S_1, for which only the final observation is tested. The first of these is a logical choice given the 15% trimming of the $\sup F$ statistic, while the latter should have the greatest power against change at the very end of the sample (and corresponds most closely to SC^{2*} statistic which is built on the 1-step Chow statistics).

In addition to these statistics, we include the classical Chow (1960) F test for a break at mid-sample, labelled $F_{0.5}$.

We again use the following DGP, and the model (M1)

$$x_t = \gamma I(t > \tau T) + \varepsilon_t \qquad t = 1, \ldots, T, \qquad \text{(DGP2)}$$
$$\varepsilon_t \overset{\text{iid}}{\sim} N(0,1),$$
$$x_0 = 0.$$

We consider three different break timings: one at mid-sample, and two at the end-of-sample, in the last and second-last observations respectively. We also evaluate three different break magnitudes. The results are presented in Table 3.3.

The results are relatively unsurprising. The mid-sample $F_{0.5}$ test is consistently the most powerful test, of those evaluated, for a break at mid-sample, with almost certain rejection for even the smaller break at the smallest sample size. The $\sup F$ test also performs extremely well in this scenario. The end-of-sample S tests have little power against this kind of break (nor are they expected to). The SC^{2*} test is not as powerful against this alternative as the $\sup F$ test, but still has useful power for sufficiently large breaks and sample sizes.

Against the end-of-sample breaks, the mid-sample $F_{0.5}$ and $\sup F$ tests have little power, although the $\sup F$ appears to improve rapidly as the break recedes from the end of the sample. The $S_{0.15}$ has poor power against changes at the very end of the sample, while the S_1 test performs much better. The SC^{2*} once again occupies a middle ground, not as powerful as the S_1 test (designed specifically for this narrow alternative) but more powerful than the $S_{0.15}$ test, and much more powerful than the full-sample tests.

3.6 Discussion

The sup-Chow test, like the 1-step test from which it is constructed, is intended as a misspecification test. It is therefore not conducted with a specific model as the alternative hypothesis. Nevertheless it is useful to develop a sense of the types of misspecification against which it may have power. While the alternatives examined are necessarily arbitrary and simplistic, in the absence of other evidence the results in this chapter provide good rules-of-thumb for when the test may be useful.

For the case of structural breaks, we find that the test has power against shifts in the process mean or changes in the error variance, but less so against changes in the autoregressive parameter. This is broadly in line with the findings of Clements and Hendry (1999) and Hendry (2000), that changes of the latter type are difficult to detect in general. Compared with our benchmark tests (Andrews 1993, 2003*a*), the results suggest that the sup-Chow's main value is in detecting structural change near the end of the sample, but when the end-sample period cannot be precisely identified (which is, of course, the likely case for a general misspecification test).

We also find that the test has power against outliers, although the recursive estimation combined with the left trimming procedure has a dramatic effect, masking the effect of outliers early in the sample. We expect that this issue could be overcome by re-running the test on the sample with the time-ordering reversed, although then the interpretation as a test of forecast failure is sacrificed.

Finally, it bears reiterating once more that the scope for model misspecification is essentially limitless, while this chapter, in common with most such analyses, examines only a limited range of alternatives. While the results should provide some confidence about what the test may or may not reveal, that confidence must be checked in the face of the unknowable generating processes of real data.

Chapter 4

Distributional robustness of the sup-Chow test

4.1 Introduction

Much of the theory of Chapter 2 is developed under only relatively weak assumptions on the distribution of the error term, but to operationalise this theory for a misspecification test, we necessarily adopt a specific distributional form in Theorem 2.4.5. In a modelling methodology that emphasises strong assumptions (see, e.g. Spanos 2006, p. 7) this is not a major issue: an error distribution is specified (ordinarily, the normal), and then misspecification testing is used to accept or reject this assumption. For investigators who prefer to make weaker assumptions, however, the test's dependence on the error distribution is problematic. And while, in the former case, a strong distributional assumption may not itself be a problem, there is a question of dependence between the sup-Chow test and another misspecification test that might be used, in practice, to test the assumption. In this chapter, we confront this issue of error distribution under both approaches.

Before proceeding, it is useful to review why this distribution sensitivity arises. While the asymptotic behaviour of maxima is somewhat like that of averages, it is different in an important respect. By the central limit theorem, averages of independent

Technical details of the simulations, and an index of the DGPs (DGPx) and models (Mx) used is proved in Appendix A random variables converge in distribution to the normal, for a wide variety of underlying distributions, with a scaling transform, $\sqrt{T}$, that does not depend on the underlying distribution, and a simple and consistently-estimable centring transform, the mean. In a similar way, maxima of independent random variables converge in distribution to the Gumbel for a wide variety of underlying distributions. However, crucially, the scaling and centring transforms required to effect this convergence depend on the underlying distribution, even asymptotically. So either an underlying distribution must be assumed, as is done in previous chapters, or these transforms must be estimated from the data, which amounts to estimating the tail behaviour of the underlying distribution. For this reason, statistical tests based on extreme value theory are necessarily either sensitive to the error distribution, or rather complicated.

The chapter continues as follows. In Section 4.2 we illustrate the issue by simulating the size of the test under various non-normal error distributions. In Section 4.3, we consider the size and power of the test in conjunction with a separate test of normal residuals. In Section 4.4, we consider several simple transformations which improve the distributional robustness of the test, moving closer to a test that is distribution-free. In Section 4.5 we review some of the comparisons presented in Chapter 3. Finally, Section 4.6 concludes the chapter with a discussion of the results.

4.2 Sensitivity to distribution assumption

In Chapters 2 and 3 two main variants of the sup-Chow test are introduced, under an assumption of normal errors. We have the asymptotic sup-Chow test, based on an extreme value distribution, SC^2,

$$\mathsf{SC}^2_T = c_{T-g(T)} \left(\max_{g(T)<t\leq T} \mathsf{C}^2_{1,t} - d_{T-g(T)} \right) \xrightarrow{d} \Lambda, \tag{4.1}$$

where $g(T) = T^{1/2}$, Λ is a random variable with the Gumbel distribution, and

$$\begin{aligned} c_{T-g(T)} &= \frac{1}{2}, \\ d_{T-g(T)} &= 2\left\{ \log[T-g(T)] - \frac{1}{2}\log\log[T-g(T)] - \log\pi \right\}. \end{aligned}$$

We reject if the test statistic falls in the α right tail of the distribution.

We also introduce the finite adjusted sup-Chow test, SC^{2*},

$$\mathsf{SC}^{2*}_{\mathsf{T}} = \max_{g(T)<t\leq T} \mathsf{C}^{2*}_{1,t}$$

distributed approximately as follows

$$\Pr\left\{\mathsf{SC}^{2*}_{1,t} \leq x\right\} \approx \Pr\left\{\max_{g(T)<t\leq T} \varepsilon^2_{1,t} \leq x\right\} = [G(x)]^{T-g(T)},$$

where G is the χ^2_1 distribution function. Rejection is again in the right tail. This test was generally preferred for its minimal size distortion in small samples.

In both cases the test depends on the assumption of normal errors. In the first case, the norming sequences c_T and d_T depend on the particular distributional form. Those specified are derived under the assumption of normal errors (and, hence, χ^2_1 pointwise statistics). Given correct norming sequences, the $\mathsf{SC}^2_{\mathsf{T}}$ test statistic would converge to the Gumbel extreme value distribution for a range of error distributions; but the sequences depend on the distribution. In the second case the error distribution directly enters the approximating distribution for the statistic. Therefore we should expect that the tests will be of incorrect size when the errors have a non-normal distribution.

We demonstrate this initial point by simulating the following process.

$$\begin{aligned} x_t &= \varepsilon_t & t &= 1, \dots, T, \\ \varepsilon_t &\overset{\text{iid}}{\sim} F, \end{aligned} \tag{DGP8}$$

with F representing a variety of tested error distributions with varying skewness and kurtosis. The centred χ^2, t and uniform distributions test robustness to asymmetry, thick tails and thin tails, respectively. The model equation is as used in Chapter 3; we repeat it here for convenience.

$$x_t = b_0 + b_1 x_{t-1} + u_t \qquad t = 1, \dots, T. \tag{M1}$$

The tests are performed at a nominal size of 5%. The results are presented in Table 4.1.

		Error distribution					
T		Φ	$\chi^2_{3,\text{cent.}}$	$U[0,1]$	t_5	t_{10}	t_{50}
50	SC^2	12.6	52.8	1.3	40.8	25.9	15.1
	SC^{2*}	5.1	41.0	0.2	28.7	15.1	6.7
100	SC^2	10.4	68.9	0.3	54.7	31.3	13.5
	SC^{2*}	5.0	59.9	0.0	45.0	22.2	7.3
200	SC^2	8.3	83.6	0.0	70.8	39.0	12.8
	SC^{2*}	5.0	78.8	0.0	64.6	32.1	8.6

Table 4.1: Simulated rejection frequency (%) for SC^2 statistic under (DGP8)/(M1) for a variety of error distributions. Nominal size 5%. $M = 100000$. MCSE < 0.2. $\chi^2_{3,\text{cent.}}$ indicates a χ^2 distribution centred about its mean.

As expected, the simulated size is near-nominal with normal errors, but other distributions cause large size distortion. Only for the Student's t distribution with 50 degrees of freedom does the test have close to nominal size (and such a distribution is so close to the normal distribution as to be visually indistinguishable from it at any reasonable scale). As would be expected, with thick tails the statistic rejects too often, while with the thin tails of the uniform distribution it almost never rejects. Finally, the rejection frequencies are increasing in T, so even in the moderate t cases, the test is asymptotically mis-sized.

This poses a problem for the test in practical use: either the assumption of normal errors must itself by tested, or a distribution-robust transformation of the test must be constructed. In the following sections, we investigate both approaches.

4.3 Pre-testing for the error distribution

As discussed in Section 1.1, the setting in which we anticipate use of the sup-Chow is as part of a suite of misspecification tests. It is common in such a setting to make a strong distributional assumption about the residuals—in particular, normal distribution—and then test the fit of the residuals with this assumption. Such a test can take the form of either a graphical plot of the residuals (e.g. a QQ plot), or a formal test of the distribution of the residuals. In this setting, the sup-Chow test may be run only subsequent to a test for normal residuals, and conditional on non-rejection of that assumption.

This use case raises two questions, which we aim to answer in this section. First, does pre-testing for normal residual distribution resolve the distorted test size that occurs under other error distributions? Second, if it does, does the sup-Chow test have useful residual conditional power?

Before addressing these questions, we briefly introduce the normality test that we use.

4.3.1 Doornik and Hansen (2008) test for normality

We use the Doornik and Hansen (2008) univariate E_p statistic to pre-test for normality of the residuals. This test, derived from Shenton and Bowman (1977), belongs to a family of tests based on skewness and kurtosis. The statistic is

$$E_p = z_1^2 + z_2^2 \tag{4.2}$$

where z_1 and z_2 are measures of the sample skewness and kurtosis respectively, transformed to improve small sample performance. The transformations are quite complicated and we reserve them to Section 4.B. The statistic is approximately χ^2_2-distributed under the null hypothesis of normality, so we reject normality of a residual sequence if E_p is the 5% right-tail of this distribution. Note that we test normality of the OLS residuals (i.e. using the full-sample estimator) rather than the RLS residuals, as this would be the default behaviour in software.

In the tables that follow we denote the E_p test as Φ and indicate that another test is conducted given acceptance of normality under the E_p test using $|\Phi$. For example, $\mathsf{SC}^{2*}|\Phi$ indicates that only those samples passing the E_p test have been tested using the SC^{2*} test, and the rejection frequencies reported are relative to the total of such samples (which will also be reported, under the line Φ, as an acceptance frequency rather than a rejection frequency). We describe a sample against which the null of normality is not rejected as being 'normal-like'.

4.3.2 Simulation of the conditional sup-Chow test

tested being normal-like. We repeat the experiment of Section 4.2, that is, (DGP8) with (M1). The results appear in Table 4.2.

The results are mixed. The symmetric but thicker-tailed t distributions produce the

		Main tests (rejection)						**Normality test, Φ (acceptance)**					
		Error distribution $\rightarrow$											
T		Φ	χ^2_3	$U_{[0,1]}$	t_5	t_{10}	t_{50}	Φ	χ^2_3	$U_{[0,1]}$	t_5	t_{10}	t_{50}
50	SC^{2*}	5.1	41.0	0.2	28.7	15.1	6.7	95.2	3.3	56.3	58.2	80.6	93.3
	SC$^{2*}\vert\Phi$	3.5	*6.7	0.2	8.1	6.2	4.2						
100	SC^{2*}	5.0	59.9	0.0	45.0	22.2	7.3	95.2	0.0	5.2	36.3	71.1	92.5
	SC$^{2*}\vert\Phi$	3.4	na	*0.0	8.5	7.3	4.2						
200	SC^{2*}	5.0	78.8	0.0	64.6	32.1	8.6	95.1	0.0	0.0	13.6	55.6	91.2
	SC$^{2*}\vert\Phi$	3.4	na	na	8.3	8.5	4.8						

Table 4.2: Simulated rejection frequency (%) for SC^2 statistic, unconditional and conditional on accepting hypothesis of normality (Φ), under (DGP8)/(M1), for a variety of error distributions. Nominal size 5%. $M = 100000$. MCSE < 0.3 except where starred.

best results: a substantial proportion of samples pass the normality test, and the conditional size of the SC^{2*} test is relatively close to the nominal 5% level. The results for the χ^2 and uniform errors are less convincing: few samples pass the normality test. As a result the Monte Carlo standard errors are very high for the SC$^{2*}|\Phi$ test in these columns, providing little information. Clearly, however, this approach is ineffective against such error distributions.

Note, also, the effect of pre-testing on the SC^{2*} test when the errors do, in fact, follow the normal distribution. Rejection frequency is reduced following pre-testing. As a result of this effect, we anticipate that pre-testing for normality will reduce the power of the test as well. In many cases a structural break will cause the residual distribution to have non-normal characteristics (e.g. bimodal, or having heavier tails), which will cause the Doornik-Hansen test to reject. The question is whether the SC^{2*} has any residual power to detect structural instability in samples that have already been accepted as normal-like. To test this, we simulate the following process with an innovation outlier at mid-sample,

$$x_t = \beta x_{t-1} + \delta I(t = \lfloor \tau T \rfloor + 1) + \varepsilon_t \qquad t = 1, \dots, T, \qquad \text{(DGP7)}$$
$$\varepsilon_t \overset{\text{iid}}{\sim} N(0,1),$$
$$x_0 = 0,$$

where x_t is the process observed, $\lfloor \tau T \rfloor + 1$ is the outlier time (here, with $\tau = 0.5$) and I is an indicator function. The errors are distributed normally. We use model (M1). The

		Main tests (rejection)						Normality test, Φ (acceptance)					
		Outlier magnitude (δ) $\rightarrow$											
T		**0.0**	**1.0**	**2.0**	**3.0**	**4.0**	**5.0**	**0.0**	**1.0**	**2.0**	**3.0**	**4.0**	**5.0**
							$\beta = 0.5$						
50	SC^{2*}	5.3	5.6	10.3	28.2	58.2	84.6	95.2	95.0	92.5	82.2	60.1	32.8
	$SC^{2*}\|\Phi$	4.3	4.5	7.7	19.0	38.9	61.6						
100	SC^{2*}	5.1	5.4	9.6	27.9	60.9	87.8	94.2	94.0	92.1	82.7	60.9	32.6
	$SC^{2*}\|\Phi$	4.2	4.5	7.3	19.2	42.3	68.0						
200	SC^{2*}	5.0	5.2	8.7	25.7	59.4	88.0	93.5	93.4	91.9	84.6	64.8	35.9
	$SC^{2*}\|\Phi$	4.3	4.4	6.9	18.7	43.5	70.6						
							$\beta = 0.9$						
50	SC^{2*}	6.4	6.7	11.7	29.7	59.2	84.9	80.7	80.9	80.3	78.1	73.5	67.1
	$SC^{2*}\|\Phi$	5.4	5.7	10.2	26.8	55.5	82.1						
100	SC^{2*}	5.8	6.2	10.6	28.6	61.2	87.8	69.2	69.1	68.9	67.9	65.2	60.5
	$SC^{2*}\|\Phi$	5.2	5.6	9.8	26.9	58.8	86.1						
200	SC^{2*}	5.4	5.6	9.0	26.0	59.5	88.2	59.7	59.7	59.8	59.4	58.2	55.2
	$SC^{2*}\|\Phi$	5.2	5.3	8.7	25.0	58.0	87.3						

Table 4.3: Simulated rejection frequency (%) with an innovation outlier of magnitude δ at $\tau = 0.5$, under (DGP7)/(M1). Nominal size 5%. Magnitude relative to error variance of 1.0. $M = 50000$. MCSE < 0.4

results are presented in Table 4.3.

First, the Doornik-Hansen test of the residuals (Φ) is sensitive to the autoregressive parameter in the model, so we present both mild $\beta = 0.5$ and severe $\beta = 0.9$ cases. In both cases, however, the results are quite encouraging. The sup-Chow test appears to have power which is orthogonal to the normality test, so that even in the worst cases, $SC^{2*}|\Phi$ rejects around two-thirds as often as SC^{2*}. Furthermore, the combined power of both stages of testing is quite high (although the two-stage procedure will not be size correct). This suggests that the test could feasibly be used in combination with a normality test and still provide useful evidence of misspecification.

4.4 Transformations to improve robustness

The above results provide some reassurance that the test may be used in conjunction with a test of the error distribution, and still retain some power (albeit attenuated). There may be circumstances in which it is not, however, appropriate to make a specific parametric assumption about the error distribution. In such a circumstances, it would be desirable to have a test which has a known asymptotic distribution that does not depend on the error distribution (or at least that works for a wide class of distributions).

The following sections investigate various transformations to achieve this end.

4.4.1 Difference (spacing) transformations

One interesting and straightforward method of increasing the robustness of tests based on extreme values is offered by Burridge and Taylor (2006), in response to a similar problem of distribution-sensitivity in a test proposed by Perron and Rodríguez (2003). A combination of transformation and subsampling is used to create an asymptotically robust statistic, although finite sample performance is still limited.

The underlying principle is developed in Theorems 2 and 3 of Weissman (1978), which Burridge and Taylor then apply. It relies on a convenient property of the Gumbel extreme value distribution, which we now describe.

4.4.1.1 Theory of spacing transformation

We restate Weissman's result here for convenience. Consider a independent random sample $X_1, X_2, \ldots, X_T$ from a distribution function F. Denote the order statistics of the sample $X_{(1),t} \geq X_{(2):T} \geq \ldots \geq X_{(T):T}$, where we also index by the sample size T.

Proposition 4.4.1 (Theorems 2 and 3, Weissman (1978)). *Suppose F is in the maximum domain of attraction of the Gumbel distribution, so that there exist sequences a_T and b_T such that for fixed i, $m_{i,T} = a_T^{-1}(X_{(i):T} - b_T)$ converges in distribution to a Gumbel random variable as $T \to \infty$. Then for each fixed L, the random vector $(m_{1,T}, m_{2,T}, \ldots, m_{L,T})$ converges in distribution to some $(m_1, m_2, \ldots, m_L)$, and moreover, $\{m_1 - m_2, m_2 - m_3, \ldots, m_{L-1} - m_L, m_L\}$ are independent and each $m_i - m_{i+1}$ follows the exponential distribution with mean i^{-1}.*

Burridge and Taylor apply this to avoid directly estimating the tail behaviour of the underlying error distribution. In the naive extreme value approach, we need to know (or estimate) both the scaling sequence a_T and the centring sequence b_T. But by focusing instead on differences, we immediately eliminate the centring sequences b_T. The scaling sequences remain embedded, since they are required to transform $X_{(i):T}$ into $m_{i,T}$ and are not eliminated by taking differences of adjacent statistics. We have

$$m_{i,T} - m_{i+1,T} = a_T^{-1}(X_{(i):T} - X_{(i+1):T}) \xrightarrow{d} \mathsf{Exp}(i^{-1}).$$

But then we can normalise the spacings so that

$$ia_T^{-1}(X_{(i):T} - X_{(i+1):T}) \xrightarrow{d} \mathsf{Exp}(1),$$

and then

$$i(X_{(i):T} - X_{(i+1):T}) \overset{a}{\sim} \mathsf{Exp}(a_T),$$

with these weighted spacings asymptotically independent over i by Proposition 4.4.1. Then the first k spacings, for some finite k and with large T, are approximately independent and identical in distribution; and this suggests the very simple testing scheme adopted by Burridge and Taylor, which we adapt in the next section.

4.4.1.2 Application to the sup-Chow test

For a test at a nominal size $\alpha \times 100\%$, calculate the first $L = 1/\alpha$ spacings of the individual pointwise statistics (in our case, the finite adjusted 1-step Chow statistics). That is, calculate

$$i\left(\mathsf{C}^{2*}_{1,(i):T} - \mathsf{C}^{2*}_{1,(i+1):T}\right) \quad \text{for } 1 \le i \le L. \tag{4.3}$$

These form a sample which is approximately independent and identical, and so the probability that any particular weighted spacing is the largest must be $1/L = \alpha$. However, if the maximum of the pointwise statistics is large (as we would expect it to be in the case of an outlier, and perhaps certain structural breaks), then the first weighted spacing will be the largest of k with a probability greater than α (assuming the subsequent spacings are not also contaminated by the outlier/break). Hence we adopt the criterion to reject when the first weighted spacing is the largest, that is, reject if

$$1 \cdot \left(\mathsf{C}^{2*}_{1,(1):T} - \mathsf{C}^{2*}_{1,(2):T}\right) > \max_{1<i\le L}\left\{i\left(\mathsf{C}^{2*}_{1,(i):T} - \mathsf{C}^{2*}_{1,(i+1):T}\right)\right\}. \tag{4.4}$$

This gives a test with asymptotically correct size, and potentially power against some interesting alternatives. We label this test $\mathsf{SC}^{2*}_{\mathsf{D}}$ in tables.

Note that the discrete nature of this test means only certain nominal sizes can be achieved, and the smaller the nominal size, the larger the sample required to perform the test. The first problem is minor, especially since standard test sizes of 1%, 5% and 10% have integral reciprocals (100, 20, 10). The second problem is more serious: a 1% test—commonly used when multiple misspecification test are being run—requires examination of 100 order statistics, which in practice means a sample larger than 100 (since the left trimming must be applied).

4.4.1.3 A finite sample correction

Burridge and Taylor find a size issue for finite samples for the test they are considering, with simulation results that are 'either too liberal or too conservative in finite samples according to the distribution from which the sample is drawn'. They suggest several solutions to address this, based on an expansion of the quantile function.

They first consider a non-parametric correction based on Cheng and Peng (2002), but reject this as impractical; for this reason we do not investigate this approach. Similarly, they find a simple estimator fails to perform well for Gaussian data, and reject this. The approach they recommend is simulation of weights appropriate for a Gaussian underlying distribution with a particular, reasonable sample size, in the hope that these weights—though imperfect—will also improve size for other underlying distributions of interest. The sample size for which they optimise is 3000 (for reasons that are not entirely clear).

We apply a similar correction, using weights simulated from chi-squared data (since the 1-step statistics are asymptotically chi-squared), and for a sample size in our target range, 200. These weights are reported in Section 4.A. We label this test $\mathsf{SC}^{2*}_{\mathsf{DW}}$ in tables.

4.4.1.4 Simulation of the spacing-transformed sup-Chow

We conduct a size simulation using the same experiment as in Section 4.2. Recall that the process was:

$$\begin{aligned} x_t &= \varepsilon_t & t &= 1, \dots, T, \\ \varepsilon_t &\overset{\text{iid}}{\sim} F, \end{aligned} \tag{DGP8}$$

with F representing a variety of tested error distributions with varying skewness and kurtosis, with model (M1). The results are presented in Table 4.4.

Note that the uniform and t distributions are not in the maximum domain of attraction of the Gumbel distribution (being in the domain of attraction of the Weibull and Fréchet distributions, respectively—see Embrechts et al. 1997, pp. 153–160), so the arguments of previous sections do not in fact apply to these error distributions; we retain these for comparative purposes, and in fact find that for the t distributions, the transformation is nevertheless an improvement.

While the test is clearly mis-sized for these finite samples (reflecting the asymptotic arguments used), it can be seen that the size improves as T increases, even for non-Gaussian errors, in contrast to the behaviour of the sup-Chow test. Therefore, this test is at least asymptotically robust to distributional misspecification (or simply distributional uncertainty) within the class of distributions in the domain of attraction of the Gumbel. The finite-sample weightings of SC^{2*}_{DW} improve the normal case, but make little difference when the distribution is non-normal.

To investigate the power of the spacings test, we conduct a simulation using (DGP7), with an innovation outlier at mid-sample, and model (M1). The results are presented in Table 4.5.

The results are encouraging. Power against this alternative is comparable to the SC^{2*} test; hence this would seem to be a worthwhile transformation of the test. Once again the finite sample weights of SC^{2*}_{DW} make only a minor difference.

Finally, since this test relies on a correct sampling distribution for the largest $(1/\alpha)$ of the 1-step statistics, we anticipate that it will have reduced power against more than

T		Error distribution Φ	$\chi^2_{3,\text{cent.}}$	$U[0,1]$	t_5	t_{10}	t_{50}
50	SC^{2*}	5.1	41.0	0.2	28.7	15.1	6.7
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	8.9	32.7	0.9	24.4	16.2	10.2
	$\mathsf{SC}^{2*}_{\mathsf{DW}}$	7.6	31.2	0.7	22.9	14.7	8.8
100	SC^{2*}	5.0	59.9	0.0	45.0	22.2	7.3
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.8	24.5	1.4	24.0	14.8	8.2
	$\mathsf{SC}^{2*}_{\mathsf{DW}}$	5.6	23.3	1.0	22.5	13.2	6.9
200	SC^{2*}	5.0	78.8	0.0	64.6	32.1	8.6
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.1	18.1	2.6	24.0	14.6	7.5
	$\mathsf{SC}^{2*}_{\mathsf{DW}}$	4.9	16.5	2.1	22.4	12.9	6.2

Table 4.4: Simulated rejection frequency (%) for spacings tests under (DGP8)/(M1) for a variety of error distributions. Nominal size 5%. $M = 100000$. MCSE < 0.2.

β	T		Outlier magnitude (δ) **0.0**	**1.0**	**2.0**	**3.0**	**4.0**	**5.0**
0.5	**50**	SC^{2*}	5.3	5.6	10.3	28.2	58.2	84.6
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	8.9	9.1	13.7	30.2	56.8	79.4
		$\mathsf{SC}^{2*}_{\mathsf{DW}}$	7.7	7.7	12.2	28.3	54.9	78.1
	100	SC^{2*}	5.1	5.4	9.6	27.9	60.9	87.8
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.8	6.9	10.0	24.4	52.6	79.3
		$\mathsf{SC}^{2*}_{\mathsf{DW}}$	5.6	5.7	8.7	22.4	50.3	77.8
	200	SC^{2*}	5.0	5.2	8.7	25.7	59.4	88.0
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.1	6.2	8.3	20.4	48.1	77.6
		$\mathsf{SC}^{2*}_{\mathsf{DW}}$	4.9	5.1	7.1	18.5	45.8	75.9
0.9	**50**	SC^{2*}	6.4	6.7	11.7	29.7	59.2	84.9
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	8.8	9.1	13.4	29.1	54.7	77.1
		$\mathsf{SC}^{2*}_{\mathsf{DW}}$	7.6	7.8	11.9	27.2	52.8	75.8
	100	SC^{2*}	5.8	6.2	10.6	28.6	61.2	87.8
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.9	7.0	10.0	23.7	51.1	77.7
		$\mathsf{SC}^{2*}_{\mathsf{DW}}$	5.7	5.8	8.5	21.8	48.6	76.0
	200	SC^{2*}	5.4	5.6	9.0	26.0	59.5	88.2
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.0	6.1	8.1	20.1	47.2	77.0
		$\mathsf{SC}^{2*}_{\mathsf{DW}}$	4.9	5.0	6.8	18.2	44.8	75.2

Table 4.5: Simulated rejection frequency (%) for spacings tests with an innovation outlier of magnitude δ at $\tau = 0.5$, under (DGP7)/(M1). Nominal size 5%. Magnitude relative to error variance of 1.0. $M = 50000$. MCSE < 0.3.

a single outlier, and limited power against parameter changes (since these will typically cause a sequence of high 1-step statistics, that is, a run of forecast failures).

4.4.2 Ratio transformations

The Burridge and Taylor (2006) spacings transformation is effective in removing dependency on the centring term, but dependency on the scaling term remains. Taking a ratio of scale-dependent statistics is a standard approach to eliminate an unknown scale, and a logical next step from the differencing of the spacings approach. We are not, however, aware of a previous application of this technique to an extreme-value-based test such as we or Burridge and Taylor (2006) consider. We investigate two transformations based on this idea: one simple and one more complicated.

4.4.2.1 Ratio of first two spacings

The simplest ratio statistic we can construct, and the most direct extension of the Burridge and Taylor approach, is the ratio of the first two spacings. We have from above that

$$ia_t^{-1}(X_{(i),t} - X_{(i+1),t}) \xrightarrow{d} \mathsf{Exp}(1) \tag{4.5}$$

independent over i. Then if we take the ratio of two spacings, we have, with $i \neq j$,

$$\frac{(1/2)i(X_{(i),t} - X_{(i+1),t})}{(1/2)j(X_{(j),t} - X_{(j+1),t})} \xrightarrow{d} \frac{Y/2}{Z/2} \tag{4.6}$$

where Y and Z are exponential random variables with parameter $\frac{1}{2}$, independent of each other. Since the $\mathsf{Exp}(\frac{1}{2})$ distribution is the χ^2_2 distribution, the ratio will have an $F(2,2)$ distribution (the numerator and denominator degrees of freedom cancel).

By taking the ratio we have eliminated the a_t scaling transform. This avoids the need to use the sampling approach of Burridge and Taylor; instead we can simply calculate the first four orders statistics, and compare the ratio of the the differences with a critical value drawn from the $F(2,2)$ distribution, with rejection in the right tail. We label this test $\mathsf{SC}_\mathsf{R}^{2*}$ in tables.

		Error distribution					
T		Φ	$\chi^2_{3,\text{cent.}}$	$U[0,1]$	t_5	t_{10}	t_{50}
50	SC^{2*}	5.1	41.0	0.2	28.7	15.1	6.7
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	8.9	32.7	0.9	24.4	16.2	10.2
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.2	8.0	4.6	7.3	6.1	5.3
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	6.6	25.6	2.5	17.3	11.4	7.2
100	SC^{2*}	5.0	59.9	0.0	45.0	22.2	7.3
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.8	24.5	1.4	24.0	14.8	8.2
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.1	6.3	4.8	7.5	5.9	5.4
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	6.3	20.3	2.8	20.6	12.6	7.4
200	SC^{2*}	5.0	78.8	0.0	64.6	32.1	8.6
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.1	18.1	2.6	24.0	14.6	7.5
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.1	6.2	5.0	7.9	6.0	5.4
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	6.1	17.1	3.6	23.2	14.2	7.6

Table 4.6: Simulated rejection frequency (%) for ratio tests under (DGP8)/(M1) for a variety of error distributions. Nominal size 5%. $M = 100000$. MCSE < 0.2.

4.4.2.2 Ratio of first spacing to a sum of spacings

A second ratio transformation of spacings that might be expected to have better power properties is the following.

$$\frac{1 \cdot (X_{(1),t} - X_{(2),t})}{(L-1)^{-1} \sum_{j=2}^{L} j(X_{(j),t} - X_{(j+1),t})} \xrightarrow{d} \frac{Y/2}{W/(2L-2)} \tag{4.7}$$

for some suitable fixed L. Here, Y is again a χ^2_2 random variable, while W has the distribution of a sum of $L-1$ independent χ^2_2 variables, which is $\chi^2_{2(L-1)}$. Then the ratio of Y and W, each divided by their degrees of freedom, will have a $F(2, 2L-2)$ distribution. Again, we reject in the right tail of the distribution.

Although the asymptotics of Proposition 4.4.1 take L as fixed in T, we find better results over a variety of sample sizes if we allow L to grow with increasing sample size as $L = \sqrt{T - g(T)}$. We label this test $\mathsf{SC}^{2*}_{\mathsf{RS}}$ in tables.

4.4.2.3 Simulation of the ratio-transformed sup-Chow

We again conduct a size simulation using (DGP8)—white noise—and model (M1), with the same error distributions as in Section 4.2. The results are presented in Table 4.6.

The simple ratio test SC^{2*}_{R} appears to have much better robustness against deviations from normal errors than either the spacings test (included for comparison) or the sum-

			Outlier magnitude (δ)					
α	T		**0.0**	**1.0**	**2.0**	**3.0**	**4.0**	**5.0**
0.5	**50**	SC^{2*}	5.3	5.6	10.3	28.2	58.2	84.6
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	8.9	9.1	13.7	30.2	56.8	79.4
		$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.0	4.9	5.7	8.2	13.4	20.0
		$\mathsf{SC}^{2*}_{\mathsf{RS}}$	5.0	5.1	7.7	19.4	42.0	65.7
	100	SC^{2*}	5.1	5.4	9.6	27.9	60.9	87.8
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.8	6.9	10.0	24.4	52.6	79.3
		$\mathsf{SC}^{2*}_{\mathsf{R}}$	4.9	5.0	5.5	7.4	12.9	20.5
		$\mathsf{SC}^{2*}_{\mathsf{RS}}$	4.9	5.0	7.3	19.5	45.6	74.0
	200	SC^{2*}	5.0	5.2	8.7	25.7	59.4	88.0
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.1	6.2	8.3	20.4	48.1	77.6
		$\mathsf{SC}^{2*}_{\mathsf{R}}$	4.8	4.8	5.1	7.2	12.6	20.5
		$\mathsf{SC}^{2*}_{\mathsf{RS}}$	5.0	5.2	7.0	18.7	45.9	76.9
0.9	**50**	SC^{2*}	6.4	6.7	11.7	29.7	59.2	84.9
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	8.8	9.1	13.4	29.1	54.7	77.1
		$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.1	5.1	5.5	8.2	13.3	19.3
		$\mathsf{SC}^{2*}_{\mathsf{RS}}$	5.1	5.1	7.8	18.5	39.6	62.9
	100	SC^{2*}	5.8	6.2	10.6	28.6	61.2	87.8
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.9	7.0	10.0	23.7	51.1	77.7
		$\mathsf{SC}^{2*}_{\mathsf{R}}$	4.9	5.1	5.4	7.5	12.6	20.2
		$\mathsf{SC}^{2*}_{\mathsf{RS}}$	5.0	5.1	7.4	18.7	44.1	72.1
	200	SC^{2*}	5.4	5.6	9.0	26.0	59.5	88.2
		$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.0	6.1	8.1	20.1	47.2	77.0
		$\mathsf{SC}^{2*}_{\mathsf{R}}$	4.8	4.8	5.1	7.0	12.2	20.5
		$\mathsf{SC}^{2*}_{\mathsf{RS}}$	5.0	5.0	6.9	18.0	45.2	76.1

Table 4.7: Simulated rejection frequency (%) for ratio tests with an innovation outlier of magnitude δ at $\tau = 0.5$ (mid-sample), under (DGP7)/(M1). Nominal size 5%. Magnitude relative to error variance of 1.0. $M = 50000$. MCSE < 0.3.

ratio test SC^{2*}_{RS}. The intuition for this is that these transformation are based on asymptotic behaviour implied by Proposition 4.4.1. The SC^{2*}_{R} test uses only the three largest order statistics, so this asymptotic approximation holds quite well. The SC^{2*}_{D} and SC^{2*}_{RS} tests both rely on a larger number of order statistics, and further from the extrema, the approximation breaks down.

To investigate the power of the ratio tests against a single outlier, we conduct the same experiment as in the previous section, with (DGP7) (an innovation outlier) and (M1). Results are presented in Table 4.7.

The power results once again show the very different behaviour of the two ratio tests. The SC^{2*}_{RS} test is again very similar to the spacings test, showing power comparable to both it and the original sup-Chow test. The simple ratio test, SC^{2*}_{R}, is relatively less powerful.

From the transformations considered, there appears to be a trade-off, in finite samples, between distributional robustness and power against a single outlier. This make intuitive sense, since an outlier can only be defined relative to a given reference distribution, so that what might be considered an outlier for a distribution with lighter tails would not be for a distribution with heavier tails. Even within the class of distributions imposed by Proposition 4.4.1 (Gumbel domain of attraction), this trade-off remains.

Finally, given the similar performance of the the spacings ($\mathsf{SC}^{2*}_{\mathsf{D}}$ and $\mathsf{SC}^{2*}_{\mathsf{DW}}$) tests and the ratio-sum test ($\mathsf{SC}^{2*}_{\mathsf{RS}}$), there seems nothing to recommend the former, given the non-standard subsampling method and the limitations discussed in Section 4.4.1.

4.4.3 Power of transformed tests against a structural break

The context for the spacings approach of Burridge and Taylor (2006) is outlier testing, and it seems likely that the various transformations described above would have a detrimental effect on power against other more complex alternatives (for instance, parametric change). All the tests rely on the spacing between the largest and second-largest pointwise statistics, so any misspecification which leads to both of these statistics being large may go undetected.

To examine this possibility, we simulate power against a single shift in the mean level of the process, using (M1) with

$$
\begin{aligned}
x_t &= \gamma I(t > \tau T) + \varepsilon_t \qquad\qquad t = 1, \ldots, T, \qquad\qquad \text{(DGP2)}\\
\varepsilon_t &\overset{\text{iid}}{\sim} N(0, 1),\\
x_0 &= 0,
\end{aligned}
$$

where x_t is the process observed, τT is the break time and I is an indicator function. The results are presented in Table 4.8.

As expected, the transformed tests have relatively little power against structural breaks occurring anywhere except the very end of the sample. The end-of-sample break performance is comparable to the original sup-Chow test, which is unsurprising given

		Break timing (τ)								
		0.5T			T-2			T-1		
		Post-break constant (γ)								
T		**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**
25	SC^{2*}	5.3	13.4	50.6	5.3	20.6	78.2	5.3	15.7	69.9
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	4.9	5.3	10.0	4.9	5.4	8.1	4.9	5.7	12.4
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	4.9	5.9	19.8	4.9	6.7	15.2	4.9	7.1	27.0
50	SC^{2*}	5.0	16.7	67.0	5.0	18.2	82.7	5.0	13.0	69.5
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	9.0	10.9	38.3	9.0	14.3	37.2	9.0	13.5	56.3
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.0	5.5	8.6	5.0	5.7	7.3	5.0	5.5	13.5
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	5.0	6.2	23.6	5.0	8.0	22.0	5.0	7.8	40.8
100	SC^{2*}	5.0	19.7	77.6	5.0	15.9	84.0	5.0	11.3	67.4
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.8	8.6	29.1	6.8	10.9	33.0	6.8	10.0	52.9
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.1	5.2	7.1	5.1	5.5	6.5	5.1	5.4	13.0
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	4.9	6.3	21.7	4.9	7.9	26.0	4.9	7.2	45.8

Table 4.8: Simulated rejection frequency (%) of transformed tests, with a single break in the constant term (and process mean) of γ at τ, under (DGP2)/(M1). Nominal size 5%. $M = 50000$. MCSE < 0.3.

the similarity between a structural break and an outlier at the end of the sample.

4.5 Comparisons

The issue of distribution-dependence is not limited to the sup-Chow tests. For instance, Burridge and Taylor (2006) propose their method to deal with exactly such a dependence in an outlier test proposed by Perron and Rodríguez (2003). The tests of Andrews (1993) and Andrews (2003*a*), however, which we used as benchmark in Chapter 3, do not suffer from this limitation. It is therefore of some interest to now compare the performance of the conditional and transformed sup-Chow tests against these tests.

We first confirm the distributional robustness of the Andrews tests. We use (DGP8) and (M1) as in previous simulations of size in this chapter. Recall from Chapter 3 that $\sup F$ is the Andrews (1993) test of structural instability anywhere in the sample, while $S_{0.15}$ and S_1 are two variants of the Andrews (2003*a*) end-of-sample test (see Section 3.5 for more details). The results are presented in Table 4.9.

Indeed, as expected, both the $\sup F$ and S tests are comparatively well-sized under all the distributions tested. In this respect, these tests outperform any of the transformed variants of the sup-Chow that we have considered.

		Error distribution					
T		Φ	$\chi^2_{3,\text{cent.}}$	$U[0,1]$	t_5	t_{10}	t_{50}
50	$\sup F$	6.1	6.3	6.3	5.9	6.0	6.0
	$S_{0.15}$	4.0	3.9	3.7	4.1	3.9	3.9
	S_1	5.4	5.4	3.6	5.6	5.6	5.7
100	$\sup F$	4.9	4.8	5.3	4.7	4.9	5.0
	$S_{0.15}$	4.5	4.2	4.6	4.6	4.4	4.5
	S_1	4.7	4.8	3.5	4.9	4.8	4.8
200	$\sup F$	4.6	4.4	5.0	4.3	4.6	4.8
	$S_{0.15}$	4.5	4.2	4.5	4.4	4.6	4.6
	S_1	4.9	4.8	4.2	4.9	4.9	4.9

Table 4.9: Simulated rejection frequency (%) for Andrews (1993) and Andrews (2003*a*) tests under (DGP8)/(M1) for a variety of error distributions. Nominal size 5%. $M = 100000$. MCSE < 0.1.

We also compare the power of the conditional and transformed tests against the $\sup F$ and S tests. We simulate (DGP2)—a break in the process mean—with model (M1), and present the results in two tables. Table 4.10 includes the conditional sup-Chow tests of Section 4.3, while Table 4.11 includes the transformed sup-Chow tests of Section 4.4.

We know from previous sections that both robustness techniques—conditioning and transformations—tend to attenuate the power of the test. This is evident in both tables. Whereas the SC^{2*} test has a clear power advantage over the long end-of-sample $S_{0.15}$ test for breaks at the very end of the sample, this is not true of either the conditional ($\mathsf{SC}^{2*}|\Phi$) or transformed (SC^{2*}_{RS}) versions of the test; the number of cases in which the sup-Chow tests are preferred is fewer if either of these robustness procedures are used.

		Main tests (rejection)									Normality test, Φ (acceptance)								
		Break timing (τ) $\rightarrow$																	
		0.5T			T-2			T-1			0.5T			T-2			T-1		
		Post-break constant (γ)																	
T		**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**
25	SC^{2*}	5.1	13.2	50.4	5.1	20.4	78.0	5.1	15.3	69.9	95.1	98.1	73.7	95.1	89.0	38.0	95.1	89.8	41.2
	$\mathsf{SC}^{2*}\vert\Phi$	3.7	12.9	43.9	3.7	13.4	51.4	3.7	8.7	32.2									
	$\sup F$	5.0	78.5	99.4	5.0	11.0	29.3	5.0	7.2	18.6									
	$S_{0.15}$	3.6	1.6	0.3	3.6	22.6	43.9	3.6	14.1	36.6									
	S_1	6.7	8.6	17.6	6.7	41.1	88.7	6.7	50.5	96.7									
50	SC^{2*}	4.9	16.5	67.1	4.9	18.2	82.9	4.9	12.9	69.7	95.2	97.9	32.7	95.2	87.2	21.2	95.2	90.2	39.0
	$\mathsf{SC}^{2*}\vert\Phi$	3.3	16.2	54.8	3.3	9.9	37.9	3.3	6.7	27.7									
	$\sup F$	5.0	99.8	100.0	5.0	8.7	31.8	5.0	5.9	10.1									
	$S_{0.15}$	4.0	0.7	0.0	4.0	12.6	32.7	4.0	7.6	15.4									
	S_1	5.4	8.1	21.3	5.4	44.7	94.5	5.4	49.7	97.1									
100	SC^{2*}	4.9	19.4	77.4	4.9	15.8	84.1	4.9	11.1	67.4	95.2	95.2	1.5	95.2	87.9	18.1	95.2	91.3	42.0
	$\mathsf{SC}^{2*}\vert\Phi$	3.3	19.2	*62.4	3.3	8.2	32.0	3.3	6.0	27.4									
	$\sup F$	4.9	100.0	100.0	4.9	7.2	31.2	4.9	5.4	7.3									
	$S_{0.15}$	4.5	0.1	0.0	4.5	9.5	34.0	4.5	6.1	10.1									
	S_1	4.7	7.5	22.0	4.7	46.5	96.2	4.7	49.0	97.3									

Table 4.10: Simulated rejection frequency (%) of unconditional and conditional tests compared with Andrews (1993) and Andrews (2003*a*) tests, with a single break in the constant term (and process mean) of γ at τ, under (DGP2)/(M1). Nominal size 5%. $M = 200000$. MCSE < 0.3 except where starred.

4.6 Discussion

Sensitivity to the underlying error distribution is as an issue with the sup-Chow test. Depending on the modelling approach taken, this manifests either as inadvertent power against a misspecified error distribution, or simply as size distortion (if no particular error distribution is specified).

In addressing the former case, we demonstrate that although the test is sensitive to error distribution, it has power against misspecification above and beyond that of a test for normality. This means that, in practice, the sup-Chow test could be usefully included as one of a battery of misspecification tests. Of course, this naturally raises questions relating to the size of the joint misspecification procedure; this is a general problem, and goes beyond the scope of this thesis (see Sohkanen 2011 for a treatment of this problem).

In addressing the latter case, we evaluate the spacings approach of Burridge and Taylor (2006), and find this partially effective in reducing distribution sensitivity (although in our case their proposed finite-sample correction makes only a minor improvement). We extend this approach by considering ratios of spacings, and find that this can provide comparable robustness, size and power, and a simpler method of test construction than the spacings procedure. We see no reason that Burridge and Taylor (2006) could not also be extended in this manner.

In either case, however, we discover a trade-off between increasing the robustness of the test against non-normal errors, and power against the alternatives we consider. The result is that sup-Chow tests perform worse in comparison with our benchmark tests. Consequently, the approaches suggested in this chapter offer only a partial response to the distribution sensitivity issue.

Finally, it is important to note that this sensitivity to distribution arises specifically because in Chapter 2 we choose to investigate the maximum as a summary of the statistics, and avail ourselves of extreme value theory. We know from the theory of partial sums (e.g. as used by the CUSUM test, see Section 1.3.2.1) that not every such summary suffers from this drawback. Quantiles of the sequence, for example, might prove more robust. This partly motivates our investigation of the empirical process in Chapter 5.

		Break timing (τ)								
		0.5T			T-2			T-1		
		Post-break constant (γ)								
T		**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**	**0.0**	**2.0**	**4.0**
25	SC^{2*}	5.1	13.2	50.4	5.1	20.4	78.0	5.1	15.3	69.9
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	4.9	5.8	20.0	4.9	6.5	15.2	4.9	7.0	27.0
	$\sup F$	5.0	78.5	99.4	5.0	11.0	29.3	5.0	7.2	18.6
	$S_{0.15}$	3.6	1.6	0.3	3.6	22.6	43.9	3.6	14.1	36.6
	S_1	6.7	8.6	17.6	6.7	41.1	88.7	6.7	50.5	96.7
50	SC^{2*}	4.9	16.5	67.1	4.9	18.2	82.9	4.9	12.9	69.7
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	4.9	6.0	23.7	4.9	7.9	22.0	4.9	7.7	40.9
	$\sup F$	5.0	99.8	100.0	5.0	8.7	31.8	5.0	5.9	10.1
	$S_{0.15}$	4.0	0.7	0.0	4.0	12.6	32.7	4.0	7.6	15.4
	S_1	5.4	8.1	21.3	5.4	44.7	94.5	5.4	49.7	97.1
100	SC^{2*}	4.9	19.4	77.4	4.9	15.8	84.1	4.9	11.1	67.4
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	5.0	6.1	21.5	5.0	8.1	26.2	5.0	7.3	45.7
	$\sup F$	4.9	100.0	100.0	4.9	7.2	31.2	4.9	5.4	7.3
	$S_{0.15}$	4.5	0.1	0.0	4.5	9.5	34.0	4.5	6.1	10.1
	S_1	4.7	7.5	22.0	4.7	46.5	96.2	4.7	49.0	97.3

Table 4.11: Simulated rejection frequency (%) of transformed tests compared with Andrews (1993) and Andrews (2003*a*) tests, with a single break in the constant term (and process mean) of γ at τ, under (DGP2)/(M1). Nominal size 5%. $M = 200000$. MCSE < 0.1.

4.A Burridge and Taylor-style finite-sample weight vector

Weights calculated in the same fashion as Burridge and Taylor (2006) for 100 chi-squared(1) independent draws (with 10^6 repetitions).

$$
\begin{aligned}
W(100) = \langle & 1.000, 0.488, 0.320, 0.237, 0.187, 0.154, 0.131, 0.113, 0.100, 0.089,\\
& 0.080, 0.073, 0.067, 0.061, 0.057, 0.053, 0.049, 0.046, 0.043, 0.041,\\
& 0.038, 0.036, 0.035, 0.033, 0.031, 0.030, 0.028, 0.027, 0.026, 0.025,\\
& 0.024, 0.023, 0.022, 0.021, 0.020, 0.020, 0.019, 0.018, 0.017, 0.017,\\
& 0.016, 0.016, 0.015, 0.015, 0.014, 0.014, 0.013, 0.013, 0.012, 0.012,\\
& 0.012, 0.011, 0.011, 0.010, 0.010, 0.010, 0.009, 0.009, 0.009, 0.009,\\
& 0.008, 0.008, 0.008, 0.007, 0.007, 0.007, 0.007, 0.006, 0.006, 0.006,\\
& 0.006, 0.006, 0.005, 0.005, 0.005, 0.005, 0.004, 0.004, 0.004, 0.004,\\
& 0.004, 0.003, 0.003, 0.003, 0.003, 0.003, 0.003, 0.002, 0.002, 0.002,\\
& 0.002, 0.002, 0.002, 0.001, 0.001, 0.001, 0.001, 0.001, 0.000, 0.000\rangle
\end{aligned}
$$

4.B Doornik and Hansen (2008) E_p statistic computation

This section briefly summarises the computation of the E_p statistic. We refer the reader to Doornik and Hansen (2008) for an explanation of these transformations.

Let $(x_1, \ldots, x_T)$ be a sample of T observations. Define the sample central moments, skewness and kurtosis as

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t, \qquad m_i = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^i, \qquad \sqrt{b_1} = \frac{m_3}{m_2^{3/2}}, \qquad b_2 = \frac{m_4}{m_2^2}.$$

Then the transformations are as follows. For z_2, first define various polynomial terms,

$$\begin{aligned}
\delta &= (T-3)(T+1)(T^2+15T-4),\\
a &= \frac{(T-2)(T+5)(T+7)(T^2+27T-70)}{6\delta},\\
c &= \frac{(T-7)(T+5)(T+7)(T^2+2T-5)}{6\delta},\\
k &= \frac{(T+5)(T+7)(T^3+37T^2+11T-313)}{12\delta},
\end{aligned}$$

and then define

$$\alpha = a + b_1 c \qquad\qquad \chi = (b_2 - 1 - b_1)2k$$

so that finally

$$z_2 = \left[\left(\frac{\chi}{2\alpha}\right)^{1/3} - 1 + \frac{1}{9\alpha}\right](9\alpha)^{1/2}.$$

For z_1, first define several preliminary terms,

$$\begin{aligned}
\beta &= \frac{3(T^2+27T-70)(T+1)(T+3)}{(T-2)(T+5)(T+7)(T+9)},\\
\omega^2 &= -1 + [2(\beta-1)]^{1/2},\\
\delta &= \frac{1}{[\log(\sqrt{\omega^2})]^{1/2}},
\end{aligned}$$

and then define

$$y = \sqrt{b_1}\left[\frac{\omega^2 - 1}{2}\frac{((T+1)(T+3))}{6(T-2)}\right]^{1/2},$$

so that finally

$$z_1 = \delta \log\left[y + (y^2+1)^{1/2}\right].$$

The the statistic E_p is then computed as $E_p = z_1^2 + z_2^2$.

Chapter 5

Empirical process of the 1-step Chow statistics

5.1 Introduction

Empirical process results have proved useful in understanding distribution theory for a range of statistical tests based on the empirical distribution function. This class includes the well known Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests (for other examples see Stephens 1974). The classical tests of this type apply to sequences of independent, identically distributed random variables with a completely specified reference distribution, with extensions to the case in which the mean and variance are estimated. In the typical econometric context of regression, a natural extension is to consider the empirical distribution of regression residuals (rather than unobserved errors) and build tests corresponding to KS and AD from this instead.

Since the 1-step recursive Chow statistics can be viewed as squared studentized recursive residuals, it is of some interest to investigate their empirical process. This might have particular advantages where outliers or structural breaks are suspected to have contaminated the regression estimates, for reasons examined in Chapter 3. Furthermore, an understanding of the empirical process helps provided a complete and rounded perspective on the 1-step statistics. It builds on the pointwise and extreme-value results presented in Chapter 2, and the partial-sum results presented elsewhere (see Nielsen and Sohkanen 2011) to provide a more complete picture. Finally, the empirical process we study, based on the 1-step statistics, converges weakly to a standard Brownian bridge (in the case of a normal reference distribution). This makes the distributional theory of statistics derived from the process, such as the KS statistics, relatively convenient. The distinguishes our result from many other results on (full sample) least squares autoregression residuals, in which the limiting process has a more complex structure.

The empirical process of the autoregression residuals has been studied by a number of authors, including Boldin (1983), Koul and Levental (1989), Ling (1998), Lee and Wei (1999), Pierce (1985) and Engler and Nielsen (2009). As Engler and Nielsen discuss, in the case of a first order autoregression without an intercept, the limiting distribution depends on the presence of a unit root, being non-Gaussian in that case (but otherwise Gaussian). In the case of a first order autoregression with an intercept, the non-Gaussian component vanishes and the same Gaussian distribution arises irrespective of any possible unit root. This is an encouraging result from the point of view of misspecification testing, as it separates the concern of testing the innovation distribution from knowledge of the roots of the process.

It is thus similarly desirable to derive a weak convergence result for the empirical process of 1-step Chow statistics which is invariant to the location of the process roots. We achieve this for processes with stationary and unit roots, under a general framework that includes deterministic terms. The difficulty in the explosive case relates to a term described in Nielsen and Sohkanen (2011) and analysed in Sohkanen (2011). The key differences from Engler and Nielsen's paper arise because of the recursive estimation inherent in the 1-step statistics. Terms resulting from recursive estimation error are time-varying (within a fixed sample), which complicates the analysis. They are, however, predictable, and we take advantage of this property, and additional strong consistency results, to overcome this complication.

The chapter proceeds as follows. In Section 5.2 we recall the definition of the 1-step recursive Chow statistic. In Section 5.3 we present the model and assumptions under which our main results hold (the model will be familiar from Chapter 2, though a different set of assumptions is required). We present the theory results in two sections.

Section 5.4 presents the main results on weak convergence of the empirical process of 1-step residuals. These results depend heavily on more general results on empirical processes with scale and location error. Since these latter results are of some independent interest, they are presented separately in Section 5.5, with assumptions that hold under the specific assumptions made in Section 5.4. Section 5.6 describes a simple simulation, while Section 5.7 briefly discusses some issues that arise from the results. Proofs follow in the chapter appendices.

5.2 The 1-step Chow statistic

Consider a linear regression

$$Y_t = \beta' S_t + \varepsilon_t \qquad t = 1, \dots, T,$$

with S_t a K-dimensional vector of regressors, and the errors ε_t independently, identically distributed. For such a regression we can define the sequence of least squares estimators calculated over progressively larger subsamples, along with the corresponding variance estimator,

$$\hat{\beta}_t = \left(\sum_{s=1}^{t} S_s S_s'\right)^{-1} \left(\sum_{s=1}^{t} S_s Y_s'\right) \qquad t = K, \dots, T,$$

$$\hat{\sigma}_t^2 = (t-K)^{-1} \sum_{s=K}^{t} (Y_s - \hat{\beta}_t' S_s)^2 \qquad t = K, \dots, T$$

The (unscaled) recursive residuals are defined as

$$\tilde{\varepsilon}_t := (Y_t - \hat{\beta}_{t-1}' S_t)/f_t \qquad t = (K+1), \dots, T,$$

where $f_t^2 := 1 + S_t' \left(\sum_{s=1}^{t-1} S_s S_s'\right)^{-1} S_t$. We can then define the 1-step Chow test statistic as

$$c_t^2 := \hat{\sigma}_{t-1}^{-2} \tilde{\varepsilon}_t^2 \qquad t = (K+1), \dots, T,$$

and with σ^2 the variance of ε_t, we can write

$$c_t = \frac{\varepsilon_t - (\hat{\beta}_{t-1} - \beta)' S_t}{\hat{\sigma}_{t-1} f_t} = \frac{\varepsilon_t - \sigma \hat{b}_t}{\sigma(1 + \hat{a}_t)} = \frac{\eta_t - \hat{b}_t}{1 + \hat{a}_t}$$

where

$$\eta_t := \varepsilon_t / \sigma \qquad \hat{a}_t := \hat{\sigma}_{t-1} f_t / \sigma - 1 \qquad \hat{b}_t := (\hat{\beta}_{t-1} - \beta)' S_t / \sigma, \tag{5.1}$$

noting that $\hat{a}_t$ and $\hat{b}_t$ are then predictable processes (measurable at time $t-1$).

5.3 Model and assumptions

We adopt the general framework of Engler and Nielsen (2009) and Chapter 2. The model equation is an ADL model with arbitrary deterministic terms. 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 $X_{1-K-k}, \ldots, X_0, X_1, \ldots X_T$ (this indexing ensures that the first 1-step statistic can be calculated for $t = 1$, simplifying later expressions). 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 of dimension d; with the total number of regressors $K = 1 + pk + d$. That is,

$$Y_t = \rho Z_t + \sum_{j=1}^{k} \alpha_j Y_{t-j} + \sum_{j=1}^{k} \beta_j' Z_{t-j} + \nu D_{t-1} + \sigma \eta_t \qquad t = 1-K, \ldots, T, \tag{5.2}$$

conditional on $(X_{-K}, \ldots, X_{-K-k})$.

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

$$X_t = \sum_{j=1}^{k} A_j X_{t-j} + \mu D_{t-1} + \xi_t \qquad t = 1-K, \ldots, T, \tag{5.3}$$

with the deterministic term D_t given by

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

This allows us to apply the results of Nielsen (2005) on strong convergence of various regression quantities.

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,

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

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

Assumption 5.3.1. $|\text{eigen}(\mathbf{D})| = 1$ *and* $\text{rank}(D_1, \ldots, D_{\dim \mathbf{D}}) = \dim \mathbf{D}$.

We assume the VAR innovations form a martingale difference sequence satisfying the assumption below.

Assumption 5.3.2. *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 $X_0, \ldots X_{1-k}$ are $\mathcal{F}_0$-measurable and*

$$\sup_t \mathsf{E}(\|\xi_t\|^\alpha|\mathcal{F}_{t-1}) \overset{a.s.}{<} \infty \qquad \text{for some } \alpha > 4,$$
$$\mathsf{E}(\xi_t'\xi_t|\mathcal{F}_{t-1}) \overset{a.s.}{=} \Omega \qquad \text{where } \Omega \text{ is positive definite.}$$

Unlike in Chapter 2, we restrict the autoregressive parameters A_j in (5.3) to exclude explosive roots. 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 5.3.3. *The eigenvalues of* $\mathbf{B}$ *lie inside or on the unit circle. That is, the process contains no explosive component.*

Additionally, we require that the innovations in the ADL regression equation satisfy three further assumptions.

The first specifies the joint distribution of the innovations.

Assumption 5.3.4. *The innovations* η_t *of the regression equation (5.2) are independent and identically distributed with distribution function* $F(x) = \Pr(\eta_t \leq x)$ *having* $\mathsf{E}(\eta_t) = 0$ *and* $\mathsf{Var}(\eta_t) = 1$.

It follows that $\mathsf{E}(\varepsilon_t) = 0$ and $\mathsf{Var}(\varepsilon_t) = \sigma^2$.

The second assumption is needed to ensure that the regression equation (5.2) is an ADL equation when ρ is unrestricted.

Assumption 5.3.5. *Let* $\mathcal{G}_t$ *be the sigma field over* $\mathcal{F}_t$ *and* Z_t *(when* ρ *is unrestricted) or let* $\mathcal{G}_t = \mathcal{F}_t$ *(when* ρ *is restricted). Then* $(\varepsilon_t, \mathcal{G}_t)$ *is a martingale difference sequence, i.e.* $\mathsf{E}(\varepsilon_t|\mathcal{G}_{t-1}) = 0$.

The third assumption places further restrictions on the distribution function F. These are mostly technical, with the exception of (ii), which imposes symmetry.

Assumption 5.3.6 (The distribution F)**.** *For the distribution of the i.i.d. error process* η_t*, we have*

(i) F *has Lebesgue density* f *satisfying* $f > 0$ *on the set* $\{x : 0 < F(x) < 1\}$*; and second derivative* f'*,*

(ii) f *is symmetric about 0 ,*

(iii) $f[F^{-1}(u)]$ *and* $F^{-1}(u)f[F^{-1}(u)]$ *are uniformly continuous in* $0 \leq u \leq 1$*,*

(iv) $\sup_{x\in\mathbb{R}}(1+|x|)f(x) < \infty$ *and* $\lim_{x\to\pm\infty}[xf(x)] = 0$*, and*

(v) $\sup_{x\in\mathbb{R}}(1+x^2)f'(x) < \infty$.

Remark 5.3.7. *Part (iii) is standard (see, e.g. (21) of Shorack and Wellner 1986, p. 196). It will, for example, be satisfied by any distribution that has a continuous density f and support on the entire real line and $\lim_{x\to\infty} xf(x) = 0$. Then the quantile function $F^{-1}(u)$ is continuous on $(0,1)$ by the continuity of F and the everywhere-positivity of f on its support. Since in addition f is continuous, then the composition $F^{-1}(u)f[F^{-1}(u)]$ must be continuous on $(0,1)$. Then we have $\lim_{x\to\infty} xf(x) = 0$ and since F has support on the entire real line, $\lim_{u\to 1} F^{-1}(u) = \infty$. Together we then have that $\lim_{u\to 1} F^{-1}(u)f[F^{-1}(u)] = 0$; the limit to 0 follows similarly by symmetry. Now we have defined $F^{-1}(u)f[F^{-1}(u)]$ on the closed interval $[0,1]$ and uniform continuity follows from continuity and the Heine-Cantor theorem (Rudin 1976, Theorem 4.19, p.91). It therefore follows that part (iii) can be deduced from parts (ii) and (iv); we retain it as it is a familiar condition in this literature.*

Note that under this assumption certain one-to-one relationships hold between the error distribution, $F(x)$ and the absolute error distribution $G(x)$, their derivatives and their inverses. With

$$F(x) = \Pr(\eta_1 \le x) \quad \text{for} \quad x \in \mathbb{R} \qquad \text{and} \qquad G(x) = \Pr(|\eta_1| \le x) \quad \text{for} \quad x \in \mathbb{R}_0^+,$$

we have

$$G(y) = 2[F(y) - 1/2], \qquad G^{-1}(u) = F^{-1}\left[u/2 + 1/2\right], \qquad g(y) = 2f(y). \tag{5.4}$$

Also, G shares the assumed properties of F, apart from symmetry. We will use both distributions as convenient in the sequel. In general x will be a number on either the real line or the non-negative half-line, depending on the case, while we will reserve u and v for numbers in the unit interval (the range of either F or G).

5.4 Main result

Define the empirical distribution function of 1-step statistics as

$$\hat{G}_T(x) := T^{-1} \sum_{t=1}^{T} I(|c_t| \le x) \quad \text{for } x \in \mathbb{R}_0^+.$$

With $u = G(x)$ define the 1-step empirical process as

$$\mathbb{G}_T^*(u) := T^{1/2} \left\{ \hat{G}_T(x) - G(x) \right\} = T^{1/2} \left\{ \hat{G}_T(G^{-1}(u)) - u \right\} \tag{5.5}$$

Then we can state our main result as follows.

Theorem 5.4.1 (Main result)**.** *Under Assumptions 5.3.1, 5.3.2, 5.3.3, 5.3.4, 5.3.5, 5.3.6,*

$$\mathbb{G}_T^* \Rightarrow \mathbb{X}_G \quad \textit{as} \quad T \to \infty,$$

where

$$\mathbb{X}_G(u) := \mathbb{U}(u) + \frac{1}{\sqrt{2}} G^{-1}(u) g[G^{-1}(u)] \int_0^1 [G^{-1}(s)]^2 \, \mathrm{d}\mathbb{W}(s),$$

and $\mathbb{U}$, $\mathbb{W}$ *are jointly normal Brownian bridges, with* $\mathsf{Cov}[\mathbb{U}(u), \mathbb{W}(v)] = \frac{1}{\sqrt{2}}(u \wedge v - uv)$.

The limiting distribution is expressed in terms of functionals of a pair of Brownian bridges. Shorack and Wellner (1986, pp. 91–95) provide details on how to calculate the variances and covariances required; alternatively these are worked out in Sections 5.B.4 and 5.C. An elegant corollary relating to the covariance between the two terms follows.

Corollary 5.4.2 (Gaussian reference distribution)**.** *If the reference distribution F is Gaussian, the process $\mathbb{X}_G$ is itself a Brownian bridge.*

This corollary is particularly valuable for constructing tests around this process. It provides a simple limiting process under null hypotheses that involve normal errors. The Brownian bridge is a standard process and distributions of the functionals of the Brownian bridge are well-documented. For example, the Kolmogorov-Smirnov statistic constructed around the process $\mathbb{G}_T^*$ will have the usual distribution (the Kolmogorov distribution). This is not the case, for instance, with the residuals process studied by Engler and Nielsen (2009).

When the reference distribution is not normal, the limiting process is non-standard. In that case, it is relatively simple to simulate the process $\mathbb{X}_F$. The most direct method involves constructing the correlated Brownian bridges $\mathbb{U}$ and $\mathbb{W}$ on a grid over $[0,1]$ then computing the integral directly.

First, simulate the vector of Brownian motions $(\mathbb{B},\mathbb{B}_W)'$ by taking partial sums $\sum x_t$ and $\sum x_{W,t}$ of generated independent normal increments distributed respectively as

$$\begin{pmatrix} x_t \\ x_{W,t} \end{pmatrix} \overset{\text{iid}}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]. \tag{5.6}$$

The resulting Brownian motions $\mathbb{B}$ and $\mathbb{B}_W$ satisfy $\mathsf{Cov}[\mathbb{B}(u),\mathbb{B}_W(v)] = \rho(u\wedge v)$. Then transform these into Brownian bridges using the usual method,

$$\mathbb{U}(u) = \mathbb{B}(u) - u\mathbb{B}(1) \qquad \text{and} \qquad \mathbb{W}(u) = \mathbb{B}_W(u) - u\mathbb{B}_W(1). \tag{5.7}$$

The resulting Brownian bridges $\mathbb{U}$ and $\mathbb{W}$ satisfy $\mathsf{Cov}[\mathbb{U}(u),\mathbb{W}(v)] = \rho(u\wedge v - uv)$ as required.

5.5 Auxiliary results

In this section we introduce some auxiliary results on empirical processes with time-varying scale and location errors. We use these results to prove the main result in Section 5.4, but set them apart as they are of independent interest. For this section, we continue to have

$$c_t = \frac{\eta_t - \hat{b}_t}{1+\hat{a}_t},$$

so that

$$I\left(c_t \le x\right) = I\left(\eta_t \le x + \hat{a}_t x + \hat{b}_t\right),$$

but now we treat $\hat{a}_t$ and $\hat{b}_t$ as abstract sequences satisfying the following assumption.

Assumption 5.5.1 (The processes $\hat{a}_t$ and $\hat{b}_t$)**.** *For the scale and location processes $\hat{a}_t$ and $\hat{b}_t$ with $t = 1 \dots T$, we have*

(i) $\hat{a}_t$ *and* $\hat{b}_t$ *are predictable (that is, measurable time* $t-1$*),*

(ii) $\hat{a}_t > -1$ *almost surely,*

(iii) $\hat{a}_t = o(1)$ *in t almost surely,*

(iv) $T^{-1/2}\sum_{t=1}^{T} \hat{a}_t = O_p(1)$*,*

(v) $T^{-1/2}\sum_{t=1}^{T} \hat{a}_t^2 = o_p(1)$*, and*

(vi) $T^{-1/2}\sum_{t=1}^{T} \hat{b}_t^2 = o_p(1)$*.*

Now, define the empirical distribution function of $|c_t|$, and its average conditional expectation as (noting that henceforth when we write $\mathsf{E}_{t-1}(\cdot)$ we mean $\mathsf{E}(\cdot|\mathcal{G}_{t-1})$)

$$\hat{G}_T(x) := T^{-1}\sum_{t=1}^{T} I(|c_t| \leq x) \quad \text{and} \quad \bar{G}_T(x) := T^{-1}\sum_{t=1}^{T} \mathsf{E}_{t-1} I(|c_t| \leq x), \tag{5.8}$$

for $x \in \mathbb{R}_0^+$. Define the corresponding empirical distribution without estimation error, and its average conditional expectation

$$\hat{G}_T^0(x) := T^{-1}\sum_{t=1}^{T} I(|\eta_t| \leq x) \quad \text{and} \quad \bar{G}_T^0(x) := T^{-1}\sum_{t=1}^{T} \mathsf{E}_{t-1} I(|\eta_t| \leq x) = G(x).$$

for $x \in \mathbb{R}_0^+$.

Define the process

$$\mathbb{G}_T(u) := T^{1/2}\left\{\hat{G}_T(x) - \bar{G}(x)\right\} = T^{1/2}\left\{\hat{G}_T(G^{-1}(u)) - \bar{G}_T(G^{-1}(u))\right\}$$

and the standard empirical process as

$$\mathbb{G}_T^0(u) := T^{1/2}\left\{\hat{G}_T^0(x) - G(x)\right\} = T^{1/2}\left\{\hat{G}_T^0(G^{-1}(u)) - u\right\}. \tag{5.9}$$

Then the two auxiliary theorems are as follows.

Theorem 5.5.2 (Uniform convergence of the empirical process)**.** *Under Assumptions 5.3.6, 5.5.1,*

$$\sup_{u\in[0,1]} \left|\mathbb{G}_T(u) - \mathbb{G}_T^0(u)\right| \xrightarrow{p} 0.$$

Theorem 5.5.3 (Uniform convergence of the average d.f.s)**.** *Under Assumptions 5.3.6 and 5.5.1,*

$$\sup_{x\in\mathbb{R}_0^+} \left|T^{1/2}\left\{\bar{G}_T(x) - \bar{G}_T^0(x)\right\} - 2A_T(x)\right| \xrightarrow{p} 0.$$

where

$$A_T(x) = xf(x)T^{-1/2}\sum_{t=1}^{T}\hat{a}_t.$$

5.6 Simulations

In this section, we present one simulation to demonstrate the result. This also serves to reveal the finite sample behaviour of the empirical process. We focus on the Kolmogorov-Smirnoff D statistic of the square root of the 1-step Chow statistics, which is simply the maximum of the empirical process (5.5), and which therefore, by Corollary 5.4.2 and the continuous mapping theorem, has the distribution $\sup_{0\leq u\leq 1}|\mathbb{B}(u)|$ where $\mathbb{B}$ is a Brownian bridge. This distribution is standard, and has a 5% critical value of 1.36 (see, for example Stephens 1974, Table 1A).

We simulate the following process,

$$\begin{aligned} x_t &= \beta x_{t-1} + \varepsilon_t \qquad\qquad t = 1,\ldots,T, \qquad\qquad \text{(DGP1)}\\ \varepsilon_t &\overset{\text{iid}}{\sim} N(0,1),\\ x_0 &= 0, \end{aligned}$$

with β taking values in a range spanning stationary and unit-root processes, and with observation sequences of varying lengths T. The model equation is

$$x_t = b_0 + b_1 x_{t-1} + u_t \qquad t = 1, \ldots, T. \tag{M1}$$

The nominal size is set at 5%. The results are shown in Table 5.1.

	$\beta =$						
T	**-1.00**	**-0.90**	**-0.50**	**0.00**	**0.50**	**0.90**	**1.00**
50	5.1	5.0	4.9	5.1	5.2	6.2	6.5
100	5.0	4.8	4.8	4.9	5.2	5.8	6.4
200	4.8	4.9	4.9	4.9	5.0	5.6	6.0
500	4.8	4.9	4.8	4.8	4.9	5.0	5.5

Table 5.1: Simulated rejection frequency (%) for KS (Kolmogorov-Smirnov D) statistic under (DGP1)/(M1). Nominal size 5%. $M = 50000$. MCSE < 0.2.

It is clear from the table that Kolmogorov-Smirnov D statistic is well-sized across the parameter range and for all sample sizes.

This simple result demonstrates the utility of the weak convergence result for the empirical process of 1-step Chow statistics. The general result applies to any empirical distribution function-type test involving the 1-step Chow statistic. We leave the task of investigating which such statistics might prove useful for future work.

5.7 Discussion

The result of Theorem 5.4.1 parallels the result on the empirical process of studentized autoregressive OLS residuals in Engler and Nielsen (2009), and many of the applications given there would apply here also.

One interesting difference from that paper, is that the limiting $\mathbb{X}_G$ process studied here reduces to a Brownian bridge when a Gaussian reference distribution is used. This not only makes many standard results applicable, but potentially makes asymptotic power calculations, considering a different error distribution against a null hypothesis of normality, relatively straightforward.

It is possible that a similar result would follow for the *absolute value* of the OLS residuals; since it is by taking the absolute value (in combination with symmetry assumption) that we eliminate the bias due to location estimation (the term B_T, in the proofs). However, this elegant result also derives from the specific structure of the A_T term, and the weighting scheme implied by recursive estimation; so it may be that this result is unique to statistics based on the recursive residuals.

It would be of interest to study the process of the signed, un-squared 1-step Chow statistics. One difficulty we anticipate is that the location estimation bias term would then not vanish. This corresponds to the issue revealed in Ling (1998) and Lee and Wei (1999), and analysed by Engler and Nielsen. But whereas those authors find the inclusion of an intercept is sufficient for the corresponding term to be zero, in the case of the recursive residuals this is unlikely to be the case, as the sum of recursive residuals is not constrained to zero, even when an intercept is included. This raises interesting questions, which are however beyond the scope of this paper, and so are reserved for future work.

The present result does not cover processes with explosive components, unlike the results of Chapter 2; these are ruled out by Assumption 5.3.3. It would be be desirable to have such a result. This would require a substantially different proof technique, however, as the proof of Assumption 5.5.1(vi) depends on results of Nielsen and Sohkanen (2011), which appear not to hold for explosive processes (see Sohkanen 2011). Further, Lemma 5.B.1 does not hold for explosive processes, so it is possible that the explosive roots may enter the limiting distribution.

Another interesting difference from the supremum result in Chapter 2 is that, for the empirical process, the early-sample instability is not an issue. Whereas the supremum result of Theorem 2.4.5 required a left-trimming, or 'burn-in' period, to ensure consistency of the early-sample estimators, the result of Theorem 5.4.1 does not. The reason is that the empirical distribution involves sums, not suprema. Intuitively, while the distinction between a 99th percentile and a maximum may be insignificant for a small sample, asymptotically the two are completely different. While this difference is easily explicable asymptotically, it is interesting that this early sample instability does not appear to cause problems for the smallest sample size ($T = 50$) in Table 5.1.

5.A Proofs of auxiliary results

For the proofs, define $\hat{F}_T, \bar{F}_T, \hat{F}_T^0, \bar{F}_T^0, \mathbb{F}_T, \mathbb{F}_T^0$ corresponding to their G-counterparts defined at (5.8)–(5.9), except where $|c_t|$ or $|\eta_t|$ appear, write c_t or η_t respectively (i.e. without modulos).

5.A.1 A lemma on Taylor series remainders

Lemma 5.A.1 (Lagrange remainders of Taylor expansions)**.** *With F any distribution function with density f and second-derivative f'.*

$$\left|F(x + \hat{a}_t x + \hat{b}_t) - F(x)\right| \leq \sup_{x\in\mathbb{R}} [(1 + |x|)f(x)] \cdot C(|\hat{a}_t| + |\hat{b}_t|),$$

and

$$\left|F(x + \hat{a}_t x + \hat{b}_t) - F(x) - f(x)(\hat{a}_t x + \hat{b}_t)\right|$$
$$\leq \sup_{x\in\mathbb{R}} \left[(1 + x^2)f'(x) + (1 + |x|)f(x)\right] \cdot D\left(\hat{a}_t^2 + \hat{b}_t^2\right),$$

for some constants C and D not dependent on x or t.

Proof. We prove the statements in two parts, for $|\hat{a}_t| < \frac{1}{2}$ and $|\hat{a}_t| \geq \frac{1}{2}$.

First consider the case when $|\hat{a}_t| < \frac{1}{2}$. For any x^* in $(x, \hat{a}_t x + \hat{b}_t)$ (or $(\hat{a}_t x + \hat{b}_t, x)$) , we have $-|\hat{a}_t x + \hat{b}_t| \leq x^* - x \leq |\hat{a}_t x + \hat{b}_t|$, which implies

$$-|\hat{a}_t||x| - |\hat{b}_t| \leq x^* - x \leq |\hat{a}_t||x| + |\hat{b}_t|.$$

Then we have

$$x \leq \frac{x^* + |\hat{b}_t|}{1 - |\hat{a}_t|} \quad \text{and} \quad |x| \leq \frac{|x^*| + |\hat{b}_t|}{1 - |\hat{a}_t|} \qquad \text{if } x \geq 0,$$
$$x \geq \frac{x^* - |\hat{b}_t|}{1 - |\hat{a}_t|} \quad \text{and} \quad |x| \leq \left|\frac{x^* - |\hat{b}_t|}{1 - |\hat{a}_t|}\right| \leq \frac{|x^*| + |\hat{b}_t|}{1 - |\hat{a}_t|} \qquad \text{if } x < 0,$$

and for any x,

$$x^2 \leq \frac{x^{*2} + 2|\hat{b}_t||x^*| + \hat{b}_t^2}{(1-|\hat{a}_t|)^2},$$

Then since $1 - |a_t| > 1/2$, we have

$$|x| \leq 2\left(|x^*| + |\hat{b}_t|\right) \qquad \text{and} \qquad x^2 \leq 4\left(x^{*2} + 2|\hat{b}_t||x^*| + \hat{b}_t^2\right) \leq 8\left(x^{*2} + \hat{b}_t^2\right). \tag{5.10}$$

We use these results to control the Taylor expansions in the statement of the lemma.

For the linear (mean value) expansion we have

$$\begin{aligned}\left|F(x + \hat{a}_t x + \hat{b}_t) - F(x)\right| &= \left|f(x^*)(\hat{a}_t x + \hat{b}_t)\right| \\ &\leq 2f(x^*)|x^*||\hat{a}_t| + 2f(x^*)|\hat{a}_t||\hat{b}_t| + f(x^*)|\hat{b}_t| \\ &\leq \sup_{x\in\mathbb{R}}\left[(1+|x|)f(x)\right] \cdot C'(|\hat{a}_t| + |\hat{b}_t|),\end{aligned} \tag{5.11}$$

for some constant C' not dependent on x or t, since $|\hat{a}_t||\hat{b}_t| < \frac{1}{2}|\hat{b}_t|$.

Similarly, for the quadratic expansion we have

$$\begin{aligned}&\left|F(x + \hat{a}_t x + \hat{b}_t) - F(x) - f(x)(\hat{a}_t x + \hat{b}_t)\right| \\ &= \frac{1}{2}\left|f'(x^*)(\hat{a}_t x + \hat{b}_t)^2\right| \\ &\leq \frac{1}{2}\left|f'(x^*)x^2\hat{a}_t^2\right| + \left|f'(x^*)x\hat{a}_t\hat{b}_t\right| + \frac{1}{2}\left|f'(x^*)\hat{b}_t^2\right|.\end{aligned}$$

Taking this term by term and applying (5.10) we have

$$\begin{aligned}\frac{1}{2}\left|f'(x^*)x^2\hat{a}_t^2\right| &\leq 4f'(x^*)x^{*2}\hat{a}_t^2 + 4f'(x^*)\hat{a}_t^2\hat{b}_t^2 \\ &\leq \sup_{x\in\mathbb{R}}\left[(1+x^2)f'(x)\right] \cdot D_1'\left[\hat{a}_t^2 + \hat{a}_t^2|\hat{b}_t| + \hat{a}_t^2\hat{b}_t^2\right], \\ \left|f'(x^*)x\hat{a}_t\hat{b}_t\right| &\leq 2f'(x^*)|x^*||\hat{a}_t||\hat{b}_t| + 2f'(x^*)|\hat{a}_t|\hat{b}_t^2 \\ &\leq \sup_{x\in\mathbb{R}}\left[(1+x^2)f'(x)\right] \cdot D_2'\left[|\hat{a}_t||\hat{b}_t| + |\hat{a}_t|\hat{b}_t^2\right], \text{ and} \\ \frac{1}{2}\left|f'(x^*)\hat{b}_t^2\right| &\leq \sup_{x\in\mathbb{R}} f'(x) \cdot D_3'\hat{b}_t^2,\end{aligned}$$

for constants D_1', D_2', D_3' not dependent on x or t. Then for some other constant D, the entire quadratic term is bounded, with

$$\left|F(x + \hat{a}_t x + \hat{b}_t) - F(x) - f(x)(\hat{a}_t x + \hat{b}_t)\right| \leq \sup_{x\in\mathbb{R}} \left[(1 + x^2)f'(x)\right] \cdot D' \left(\hat{a}_t^2 + \hat{b}_t^2\right), \quad (5.12)$$

since $|\hat{a}_t||\hat{b}_t| \leq \frac{1}{2}\left(|\hat{a}_t|^2 + |\hat{b}_t|^2\right)$, $\hat{a}_t^2|\hat{b}_t| \leq \frac{1}{2}\hat{a}^2\left(1 + \hat{b}^2\right)$, $|\hat{a}_t|\hat{b}_t^2 \leq \frac{1}{2}\hat{b}_t^2\left(1 + \hat{a}_t^2\right) \leq \frac{1}{2}\hat{b}_t^2(1 + \frac{1}{4})$ and $\hat{a}_t^2\hat{b}_t^2 < \hat{b}_t^2$ (with the last two of these inequalities following from $|\hat{a}_t| < \frac{1}{2}$).

Now consider the case when $|\hat{a}_t| > \frac{1}{2}$. Then in the linear case we have

$$\left|F(x + \hat{a}_t x + \hat{b}_t) - F(x)\right| \leq 1 < 2|\hat{a}_t| < C''\left(|\hat{a}_t| + |\hat{b}_t|\right). \quad (5.13)$$

In the quadratic case we have

$$\begin{aligned}\left|F(x + \hat{a}_t x + \hat{b}_t) - F(x) - f(x)(\hat{a}_t x + \hat{b}_t)\right| &\leq 1 + \sup_{x\in\mathbb{R}} |x| f(x)|\hat{a}_t| + \sup_{x\in\mathbb{R}} f(x)|\hat{b}_t| \\ &< 4\hat{a}_t^2 + \sup_{x\in\mathbb{R}} |x| f(x)\hat{a}_t^2 + 2\sup_{x\in\mathbb{R}} f(x)|\hat{b}_t||\hat{a}_t| \\ &\leq \sup_{x\in\mathbb{R}} \left[(1 + |x|)f(x)\right] \cdot D''(\hat{a}_t^2 + \hat{b}_t^2), \quad (5.14)\end{aligned}$$

where in the last step we have used that $|\hat{b}_t||\hat{a}_t| \leq \frac{1}{2}(a_t^2 + b_t^2)$.

Then the main result follows by combining (5.11), (5.12), (5.13) and (5.14). □

5.A.2 Three lemmas on martingale inequalities

For the proof of Theorem 5.5.2, we take advantage of a theorem due to Bercu and Touati (2008). We first state the theorem, as a lemma, and then present a sequence of lemmas which are used in the application.

Lemma 5.A.2 (Theorem 2.1 of Bercu and Touati 2008)**.** *Let (M_n) be a locally square integrable martingale adapted to a filtration $(\mathcal{F}_n)$ with $M_0 = 0$. Then, for all $w, z > 0$,*

$$\Pr\left(|M_n| \geq w, [M]_n + \langle M\rangle_n \leq z\right) \leq 2\exp\left(-\frac{w^2}{2z}\right),$$

where $\langle M\rangle_n$ and $[M]_n$ are the predictable quadratic variation and total quadratic variation, *respectively, given by*

$$\langle M\rangle_n = \sum_{k=1}^{n} \mathsf{E}\left[\Delta M_k^2 | \mathcal{F}_{k-1}\right] \quad \textit{and} \quad [M]_n = \sum_{k=1}^{n} \Delta M_k^2.$$

Proof. See Bercu and Touati (2008, Theorem 2.1) □

Lemma 5.A.3 (Martingale iteration: termination condition)**.** *Let $I_{j,Tt}$ be an indicator function for some event which depends on j and t, which is measurable time t. Define $M_{j,T} = \sum_{t=1}^{T} I_{j,Tt} - \mathsf{E}_{t-1} I_{j,Tt}$, and note that $M_{j,T}$ is a martingale. Then if $\#\mathbb{J}$, the number of elements in $\mathbb{J}$, is $O(T^{1/2})$, we have*

$$\Pr\left(\max_{j\in\mathbb{J}} |T^{-2} M_{j,T}| \geq \frac{T^{-1/2}}{2}\right) \to 0 \textit{ as } T \to \infty.$$

Proof. We first use Boole's inequality to bound the maximum

$$\Pr\left(\max_{j\in\mathbb{J}} \left\{T^{-2} M_{j,T}\right\} > \frac{T^{-1/2}}{2}\right) \leq \sum_{j\in\mathbb{J}} \Pr\left(T^{-2} M_{j,T} > \frac{T^{-1/2}}{2}\right),$$

then use Chebyshev's inequality and the martingale property of $M_{j,T}$ on the summand, so that, for any j,

$$\begin{aligned}\Pr\left(T^{-2} M_{j,T} > \frac{T^{-1/2}}{2}\right) &\leq \frac{4}{T^{-1}} T^{-4} \mathsf{E}\left\{\sum_{t=1}^{T}\left[\sum_{t=1}^{T} I_{j,Tt} - \mathsf{E}_{t-1} I_{j,Tt}\right]\right\}^2 \\ &= 4T^{-3}\mathsf{E}\left\{T\sum_{t=1}^{T}\left[\sum_{t=1}^{T} I_{j,Tt} - \mathsf{E}_{t-1} I_{j,Tt}\right]^2\right\} \\ &= 4T^{-2}\sum_{t=1}^{T}\mathsf{E}\left\{\sum_{t=1}^{T} I_{j,Tt} - \mathsf{E}_{t-1} I_{j,Tt}\right\}^2 \\ &= O(T^{-1}).\end{aligned}$$

Then the sum of $\#\mathbb{J} = O(T^{1/2})$ terms will be $O(T^{-1/2})$ and the Lemma is proven.

□

Lemma 5.A.4 (Martingale iteration)**.** *Let $I_{j,Tt}$ be an indicator function for some event which* *depends on j and t, which is measurable time t. Define $M_{j,T} = \sum_{t=1}^{T} I_{j,Tt} - \mathsf{E}_{t-1} I_{j,Tt}$, and note that $M_{j,T}$ is a martingale. Let $\#\mathbb{J}$, the number of elements in $\mathbb{J}$, be $O(T^{1/2})$. Then we have, for any $d \geq 0$, if*

$$\max_{j \in \mathbb{J}} \sum_{t=1}^{T} \mathsf{E}_{t-1} I_{j,Tt} = O_p(T^{3/4}),$$

and

$$\Pr\left(\max_j |T^{-(d+1)} M_{j,T}| \geq \frac{y_T}{2}\right) \to 0 \text{ as } T \to \infty,$$

for some sequences ϵ_T, y_T such that

$$\epsilon_T^2/(y_T \log T) \to \infty \qquad \text{and} \qquad T^{d+1/4} y_T \to \infty,$$

then

$$\Pr\left(\max_j |T^{-(d+1)/2} M_{j,T}| \geq \epsilon_T\right) \to 0 \text{ as } T \to \infty.$$

Proof. We have that

$$T^{-(d+1)/2} M_{j,T} = T^{-(d+1)/2} \sum_{t=1}^{T} \{I_{j,Tt} - \mathsf{E}_{t-1} I_{j,Tt}\}.$$

Then $\tilde{M}_{j,T} := T^{-(d+1)/2} M_{j,T}$ is a martingale with respect to $\mathcal{F}_{T-1}$ with increments $\Delta \tilde{M}_t = T^{-(d+1)/2}\{I_{j,Tt} - \mathsf{E}_{t-1} I_{j,Tt}\}$. We additionally define

$$\begin{aligned}
[\tilde{M}]_{j,T} &:= \sum_{t=1}^{T} (\Delta \tilde{M}_t)^2 \\
&= T^{-(d+1)} \sum_{t=1}^{T} \{I_{j,Tt} - \mathsf{E}_{t-1} I_{j,Tt}\}^2 \\
&= T^{-(d+1)} \sum_{t=1}^{T} \left\{ I_{j,Tt} - 2 I_{j,Tt} \mathsf{E}_{t-1}(I_{j,Tt}) + [\mathsf{E}_{t-1}(I_{j,Tt})]^2 \right\},
\end{aligned}$$

and

$$\begin{aligned}\langle\tilde{M}\rangle_{j,T} &:= \sum_{t=1}^{T}\mathsf{E}_{t-1}\left[(\Delta\tilde{M}_t)^2\right]\\ &= T^{-(d+1)}\sum_{t=1}^{T}\mathsf{E}_{t-1}\left(\left\{I_{j,Tt}-2I_{j,Tt}\mathsf{E}_{t-1}(I_{j,Tt})+[\mathsf{E}_{t-1}(I_{j,Tt})]^2\right\}\right)\\ &= T^{-(d+1)}\sum_{t=1}^{T}\left\{\mathsf{E}_{t-1}(I_{j,Tt})-[\mathsf{E}_{t-1}(I_{j,Tt})]^2\right\},\end{aligned}$$

so that

$$\begin{aligned}[\tilde{M}]_{j,T}+\langle\tilde{M}\rangle_{j,T} &= T^{-(d+1)}\sum_{t=1}^{T}\{I_{j,Tt}-2I_{j,Tt}\mathsf{E}_{t-1}(I_{j,Tt})+\mathsf{E}_{t-1}(I_{j,Tt})\}\\ &\leq T^{-(d+1)}\sum_{t=1}^{T}\{I_{j,Tt}+\mathsf{E}_{t-1}(I_{j,Tt})\}\\ &= T^{-(d+1)}\sum_{t=1}^{T}\{[I_{j,Tt}-\mathsf{E}_{t-1}(I_{j,Tt})]+2\mathsf{E}_{t-1}(I_{j,Tt})\}.\end{aligned}$$

Then, for the lemma, we are required to prove for a deterministic sequence ϵ_T,

$$\Pr\left(\max_j|\tilde{M}_{j,T}|\geq\epsilon_T\right)\to 0 \text{ as } T\to\infty.$$

Write, for each T and some sequence y_T,

$$\begin{aligned}\Pr\left(\max_j|\tilde{M}_{j,T}|\geq\epsilon_T\right) &\leq \Pr\left(\max_j|\tilde{M}_{j,T}|\geq\epsilon_T,\max_j\left\{[\tilde{M}]_{j,T}+\langle\tilde{M}\rangle_{j,T}\right\}\leq y_T\right)\\ &\quad+\Pr\left(\max_j\left\{[\tilde{M}]_{j,T}+\langle\tilde{M}\rangle_{j,T}\right\}>y_T\right).\end{aligned}$$

Then for the first term we can use Boole's inequality to bound the probability

$$\begin{aligned}&\Pr\left(\max_j|\tilde{M}_{j,T}|\geq\epsilon_T,\max_j\left\{[\tilde{M}]_{j,T}+\langle\tilde{M}\rangle_{j,T}\right\}\leq y_T\right)\\ &\leq\sum_{i\in\mathbb{J}}\Pr\left(|\tilde{M}_{j,T}|\geq\epsilon_T,\max_j\left\{[\tilde{M}]_{j,T}+\langle\tilde{M}\rangle_{j,T}\right\}\leq y_T\right)\\ &\leq\sum_{j\in\mathbb{J}}\Pr\left(|\tilde{M}_{j,T}|\geq\epsilon_T,[\tilde{M}]_{j,T}+\langle\tilde{M}\rangle_{j,T}\leq y_T\right),\end{aligned}$$

and we can apply Lemma 5.A.2 to the summand to get, for each j and T,

$$\Pr\left(|\tilde{M}_{j,T}| \geq \epsilon_T, [\tilde{M}]_{j,T} + \langle\tilde{M}\rangle_{j,T} \leq y_T\right) \leq 2\exp\left(-\frac{\epsilon_T^2}{2y_T}\right),$$

which vanishes exponentially since by assumption $\epsilon_T^2/(y_T \log T) \to \infty$. Since this bound for the summand does not depend on j, the sum of $\#\mathbb{J} = O(T^{1/2})$ terms will then be $o(1)$.

We divide the second term into a martingale and a sum of conditional expectations.

$$\begin{aligned}
&Pr\left(\max_j\left\{[\tilde{M}]_{j,T} + \langle\tilde{M}\rangle_{j,T}\right\} > y_T\right) \\
&\quad\leq \Pr\left(\max_j\left[T^{-(d+1)}\sum_{t=1}^{T}\{[I_{j,Tt} - \mathsf{E}_{t-1}(I_{j,Tt})] + 2\mathsf{E}_{t-1}(I_{j,Tt})\}\right] > y_T\right) \\
&\quad\leq \Pr\left(\max_j\left\{T^{-(d+1)}\sum_{t=1}^{T}[I_{j,Tt} - \mathsf{E}_{t-1}(I_{j,Tt})]\right\} > \frac{y_T}{2}\right) \\
&\qquad + \Pr\left(2\max_j\left\{T^{-(d+1)}\sum_{t=1}^{T}\mathsf{E}_{t-1}(I_{j,Tt})\right\} > \frac{y_T}{2}\right). \qquad (5.15)
\end{aligned}$$

Now observe that the first subterm is equal to

$$\Pr\left(\max_j\left\{T^{-(d+1)}M_{j,T}\right\} > \frac{y_T}{2}\right),$$

which is vanishing by assumption. For the second subterm, we also have by assumption that

$$\max_{j\in\mathbb{J}}\sum_{t=1}^{T}\mathsf{E}_{t-1}I_{j,Tt} = O_p(T^{3/4}).$$

So then

$$\max_j\left\{T^{-(d+1)}\sum_{t=1}^{T}\mathsf{E}_{t-1}(I_{j,Tt})\right\} = O_p(T^{-(d+1/4)}),$$

which means that the second subterm will vanish, since $T^{d+1/4}y_T \to \infty$ by assumption.

□

5.A.3 A lemma on the average distribution functions

Recall that for $x \in \mathbb{R}$,

$$A_T(x) = xf(x)T^{-1/2}\sum_{t=1}^{T}\hat{a}_t,$$

and define similarly

$$B_T(x) = f(x)T^{-1/2}\sum_{t=1}^{T}\hat{b}_t.$$

Lemma 5.A.5. *Under Assumptions 5.3.6 and 5.5.1*

$$\sup_{x\in\mathbb{R}}\left|T^{1/2}\left\{\bar{F}_T(x) - \bar{F}_T^0(x)\right\} - A_T(x) - B_T(x)\right| \xrightarrow{p} 0.$$

Proof. Expand the quantity of interest,

$$\begin{aligned}&T^{1/2}\left\{\bar{F}_T(x) - \bar{F}_T^0(x)\right\} - A_T(x) - B_T(x)\\ &\quad\leq T^{-1/2}\sum_{t=1}^{T}\left|F(x+\hat{a}_t x+\hat{b}_t) - F(x) - xf(x)\hat{a}_t - f(x)\hat{b}_t\right|.\end{aligned}$$

Then, as a preliminary matter, note that since $a_t = o(1)$ almost surely by Assumption 5.5.1(iii), there exists some random t_0 such that for all $t > t_0$, $|a_t| < 1/2$ with probability one. We break the sum at this point, with the initial t_0 terms made negligible by the $T^{-1/2}$ weight, since the summand is bounded (F is bounded and Assumption 5.3.6(iv)).

For the tail sum, use a Taylor expansion of F in x, and Lemma 5.A.1 with Assumption 5.3.6(iv) and (v) to see that,

$$\begin{aligned}&T^{-1/2}\sum_{t=t_0}^{T}\left|F(x+\hat{a}_t x+\hat{b}_t) - F(x) - f(x)(\hat{a}_t x+\hat{b}_t)\right|\\ &\quad\leq CT^{-1/2}\sum_{t=1}^{T}\left(\hat{a}_t^2+\hat{b}_t^2\right).\end{aligned}$$

These sums vanish by Assumption 5.5.1(v) and (vi). □

5.A.4 Proof of Theorem 5.5.2

We are required to prove

$$\sup_{x\in\mathbb{R}_0^+} \left|\mathbb{G}_T(x) - \mathbb{G}_T^0(x)\right| \xrightarrow{p} 0.$$

The proof is based loosely on Theorem 2.2.5 of Koul (2002), but with several substantial modifications to suit our problem. In particular, we adopt a different (symmetric) partition in the chaining argument, we substitute a stronger martingale inequality in the analysis of $\mathcal{B}_{Tj,1}$ and $\mathcal{B}_{Tj,2}$, and we rely on the almost sure convergence of $\hat{a}_t$ at various points.

Observe that for some a and non-negative x

$$I(|a| \leq x) = I(-x \leq a \leq x) = I(a \leq x) - I(a \leq -x),$$

so that we can write, for $x \geq 0$,

$$\begin{aligned}\mathbb{G}_T(x) - \mathbb{G}_T^0(x) &= \mathbb{F}_T(x) - \mathbb{F}_T(-x) - [\mathbb{F}_T^0(x) - \mathbb{F}_T^0(-x)] \\ &= [\mathbb{F}_T(x) - \mathbb{F}_T^0(x)] - [\mathbb{F}_T(-x) - \mathbb{F}_T^0(-x)].\end{aligned}$$

Written out in full, each of these two terms is (now for any $x \in \mathbb{R}$)

$$\begin{aligned}&\mathbb{F}_T(x) - \mathbb{F}_T^0(x) \\ &\quad = T^{1/2}\left\{\hat{F}_T(x) - \bar{F}(x) - \left[\hat{F}_T^0(x) - \bar{F}^0(x)\right]\right\} \\ &\quad = T^{-1/2}\sum_{t=1}^{T}\left\{I(c_t \leq x) - \mathsf{E}_{t-1}I(c_t \leq x) - [I(\eta_t \leq x) - \mathsf{E}_{t-1}I(\eta_t \leq x)]\right\}.\end{aligned}$$

The partition Fix a $\delta > 0$ and partition $\mathbb{R}$ symmetrically into

$$-\infty = x_{-r_T} < -x_{-(r_T-1)} < \cdots < x_{-1} < x_0 = 0 < x_1 < \cdots < x_{r_T-1} < x_{r_T} = \infty,$$

such that $x_j = F^{-1}[1/2 + j\delta/T^{1/2}], -(r_T - 1) \leq j \leq r_T - 1$ and $r_T = \lfloor T^{1/2}/(2\delta)\rfloor + 1$. Then by construction,

$$T^{1/2}[F(x_j) - F(x_{j-1})] \leq \delta \text{ for all } 1 \leq j \leq r_T, \tag{5.16}$$

with equality holding for all but $j = r_T$ and $j = r_{-(r_T-1)}$. And by symmetry,

$$\begin{aligned} x_{-j} &= F^{-1}[1/2 - j\delta/(2T^{1/2})] \\ &= F^{-1}[1 - (1/2 + j\delta/(2T^{1/2})] \\ &= -F^{-1}[1/2 + j\delta/(2T^{1/2})] \\ &= -x_j. \end{aligned}$$

The main decomposition For brevity we denote for any $2 \leq 1 \leq r_T - 1$,

$$\begin{aligned} I &:= I(\eta_t \leq x + \hat{a}_t x + \hat{b}_t) & I_j &:= I(\eta_t \leq x_j + \hat{a}_t x_j + \hat{b}_t) \\ F &:= \mathsf{E}_{t-1} I = \mathsf{E}_{t-1} I(\eta_t \leq x + \hat{a}_t x + \hat{b}_t) & F_j &:= \mathsf{E}_{t-1} I_j = \mathsf{E}_{t-1} I(\eta_t \leq x_j + \hat{a}_t x_j + \hat{b}_t) \\ I^0 &:= I(\eta_t \leq x) & I_j^0 &:= I(\eta_t \leq x_j) \\ F^0 &:= \mathsf{E}_{t-1} I^0 = \mathsf{E}_{t-1} I(\eta_t \leq x) & F_j^0 &:= \mathsf{E}_{t-1} I_j^0 = \mathsf{E}_{t-1} I(\eta_t \leq x_j). \end{aligned}$$

Further denote negative-x quantities as

$$I_- := I(\eta_t \leq -x - \hat{a}_t x + \hat{b}_t) \qquad I_{-j} := I(\eta_t \leq x_{-j} + \hat{a}_t x_{-j} + \hat{b}_t),$$

and correspondingly for F, I^0 and F^0. In the tails (that is for $j = -r_T$ or $j = r_T$) instead use limits, such as

$$I_{-r_T} := \lim_{x \to -\infty} I = 0 \qquad I_{r_T} := \lim_{x \to \infty} I = 1$$

and correspondingly for F, I^0 and F^0.

Observe that all the above expressions are monotonic in the suppressed x or x_j, as long as $\hat{a}_t > -1$, assured by Assumption 5.5.1(ii), with the positive-x expressions increasing in x and the negative-x expressions decreasing in x. Using this fact, for a given $0 \leq x_{j-1} \leq x < x_j$, we have

$$\begin{aligned}
&\mathbb{G}_T(x) - \mathbb{G}_T^0(x) \\
&\quad = [\mathbb{F}_T(x) - \mathbb{F}_T^0(x)] - [\mathbb{F}_T(-x) - \mathbb{F}_T^0(-x)] \\
&\quad = \left| T^{-1/2} \sum_{t=1}^{T} \left([I - F - I^0 + F^0] - [I_- - F_- - I_-^0 + F_-^0] \right) \right| \\
&\quad \leq \left| T^{-1/2} \sum_{t=1}^{T} ([I_j - F_{j-1} - I_{j-1}^0 + F_j^0] - [I_{-j} - F_{-(j-1)} - I_{-(j-1)}^0 + F_{-j}^0]) \right| \\
&\qquad + \left| T^{-1/2} \sum_{t=1}^{T} ([I_{j-1} - F_j - I_j^0 + F_{j-1}^0] - [I_{-(j-1)} - F_{-j} - I_{-j}^0 + F_{-(j-1)}^0]) \right| \\
&\quad \leq |\mathcal{B}_{Tj,1}| + |\mathcal{B}_{Tj,3}| + |\mathcal{B}_{Tj,4}| + |\mathcal{B}_{T,-j,1}| + |\mathcal{B}_{T,-j,3}| + |\mathcal{B}_{T,-j,4}| \\
&\qquad + |\mathcal{B}_{Tj,2}| + |\mathcal{B}_{Tj,3}| + |\mathcal{B}_{Tj,4}| + |\mathcal{B}_{T,-j,2}| + |\mathcal{B}_{T,-j,3}| + |\mathcal{B}_{T,-j,4}|
\end{aligned}$$

with

$$\begin{aligned}
\mathcal{B}_{Tj,1} &:= T^{-1/2} \sum_{t=1}^{T} \left\{ (I_j - F_j) - (I_{j-1}^0 - F_{j-1}^0) \right\} \\
\mathcal{B}_{Tj,2} &:= T^{-1/2} \sum_{t=1}^{T} \left\{ (I_{j-1} - F_{j-1}) - (I_j^0 - F_j^0) \right\} \\
\mathcal{B}_{Tj,3} &:= T^{-1/2} \sum_{t=1}^{T} (F_j - F_{j-1}) - (B_{T,j} - B_{T,j-1}) \\
\mathcal{B}_{Tj,4} &:= T^{-1/2} \sum_{t=1}^{T} (F_j^0 - F_{j-1}^0)
\end{aligned}$$

where $B_{T,j} := B_T(x_j)$ for $-(r_T - 1) < j < (r_T - 1)$, and $B_{T,-r_T} := 0$ and $B_{T,r_T} := 0$, recalling that

$$B_T(x) = f(x) T^{-1/2} \sum_{t=1}^{T} \hat{b}_t.$$

Then to prove, as required, that

$$\sup_{x \in \mathbb{R}_0^+} |\mathbb{G}_T(x) - \mathbb{G}_T^0(x)| = o_p(1),$$

it suffices to show that $\max_{-(r_T-1)\leq j\leq r_T} |\mathcal{B}_{j,k}| = o_p(1)$ for $k \in \{1,2,3,4\}$. We will write $\max_{-(r_T-1)\leq j\leq r_T}$ as $\max_j$, and $\sum_{j=-(r_T-1)}^{r_T}$ as $\sum_j$.

The term $\mathcal{B}_{Tj,1}$ Convergence of this term follows a double iteration of Lemma 5.A.4, with $I_{j,Tt} = (I_j - I_{j-1}^0)$, so that $\mathsf{E}_{t-1} I_{j,Tt} = F_j - F_{j-1}^0$; and $\mathbb{J} = -(r_T - 1)\dots r_T$, which has $O(T^{1/2})$ elements as required. This lemma requires two premises to be proven.

The first premise is

$$\max_{j\in\mathbb{J}} \sum_{t=1}^{T} \mathsf{E}_{t-1} I_{j,Tt} = \max_j \sum_{t=1}^{T} I_j - I_{j-1}^0 = O_p(T^{3/4}).$$

Consider first the conditional expectation, for which we have that

$$\mathsf{E}_{t-1}(I_j - I_{j-1}^0)$$
$$= \begin{cases} [F(x_j) - 0] + [F(x_j + \hat{a}_t x_j + \hat{b}_t) - F(x_j)], & \text{if } j = (-r_T - 1), \\ [F(x_j) - F(x_{j-1})] + [F(x_j + \hat{a}_t x_j + \hat{b}_t) - F(x_j)], & \text{if } -(r_T - 2) \leq j \leq r_T - 1, \\ 1 - F(x_{j-1}), & \text{if } j = r_T. \end{cases} \tag{5.17}$$

In all three cases the first term, involving only the grid points x_j and not $\hat{a}_t$ or $\hat{b}_t$, is less than or equal to $\delta T^{-1/2}$, by construction. Then for the second term in the first and second cases, we have,

$$F(x_j + \hat{a}_t x_j + \hat{b}_t) - F(x_j) \leq C(|\hat{a}_t| + |\hat{b}_t|),$$

for some constant C not dependent on j or t (by Lemma 5.A.1, and Assumption 5.3.6(iv)). Then combining these results we have that, for all $1 \leq j \leq r_T$,

$$\mathsf{E}_{t-1}(I_{j,Tt}) \leq C[\delta T^{-1/2} + |\hat{a}_t| + |\hat{b}_t|].$$

We then have, for any j,

$$\begin{aligned}
&\sum_{t=t_0}^{T} \mathsf{E}_{t-1}(I_{j,Tt}) \\
&\quad \leq C\left\{T^{1/2}\delta + \sum_{t=1}^{T}|\hat{a}_t| + \sum_{t=1}^{T}|\hat{b}_t|\right\} \\
&\quad = O_p(T^{1/2}\delta) + o_p(T^{3/4}) + o_p(T^{3/4}) \\
&\quad = o_p(T^{3/4}),
\end{aligned}$$

since $\sum|\hat{a}_t|$ and $\sum|\hat{b}_t|$ are both $o_p(T^{3/4})$ by Assumption 5.5.1(v) and (vi) and the Cauchy-Schwarz inequality.

We show the second premise by applying Lemma 5.A.3 with $I_{j,Tt} = I_j - F_j$ as above, so that

$$\Pr\left(\max_j |T^{-3/2}\mathcal{B}_{Tj,1}| \geq \frac{T^{-1/2}}{2}\right) \to 0 \text{ as } T \to \infty.$$

Then apply Lemma 5.A.4 with $d = 1, y_T = T^{-1/2}, \epsilon_T = T^{-1/8}$, to get

$$\Pr\left(\max_j |T^{-1/2}\mathcal{B}_{Tj,1}| \geq T^{-1/8}\right) \to 0 \text{ as } T \to \infty.$$

Then choose an arbitrary ϵ and apply Lemma 5.A.4 once more with $d = 0, y_T = 2T^{-1/8}, \epsilon_T = \epsilon$ to get

$$\Pr\left(\max_j |\mathcal{B}_{Tj,1}| \geq \epsilon\right) \to 0 \text{ as } T \to \infty,$$

as required.

The term $\mathcal{B}_{Tj,2}$ Convergence of this term follows as for $\mathcal{B}_{Tj,1}$, except with $I_{j,Tt} := I_{j-1} - I_j^0$. Then at (5.17) we have

$$\mathsf{E}_{t-1}(I_{j,Tt})$$
$$= \begin{cases} 0 - F(x_j), & \text{if } j = -(r_T - 1), \\ [F(x_{j-1}) - F(x_j)] + [F(x_{j-1} + \hat{a}_t x_{j-1} + \hat{b}_t) - F(x_{j-1})], & \text{if } -(r_T - 2) \leq j \leq r_t - 1, \\ [F(x_{j-1}) - 1] + [F(x_{j-1} + \hat{a}_t x_{j-1} + \hat{b}_t) - F(x_{j-1})], & \text{if } j = r_T. \end{cases}$$

Everything else follows unchanged.

The term $\mathcal{B}_{Tj,4}$ We can apply (5.16) so that

$$\begin{aligned} \max_j \left| T^{-1/2} \sum_{t=1}^{T} (F_j^0 - F_{j-1}^0) \right| &\leq T^{1/2} \max_j |F(x_j) - F(x_{j-1})| \\ &\leq T^{1/2} \cdot T^{-1/2} \delta \\ &= \delta, \end{aligned}$$

and this vanishes, since δ is chosen arbitrarily.

The term $\mathcal{B}_{Tj,3}$ We have $\mathcal{B}_{Tj,3} = T^{-1/2} \sum_{t=1}^{T} (F_j - F_{j-1}) - (B_{T,j} - B_{T,j-1})$. In the right tail case (when $j = r_T$ and $x \geq x_{r_{T-1}-1}$) we can write this

$$\begin{aligned} \mathcal{B}_{Tj,3} = T^{-1/2} \sum_{t=1}^{T} [F_{r_T} - F_{r_T}^0] - \left\{ T^{-1/2} \sum_{t=1}^{T} [F_{r_T-1} - F_{r_T-1}^0] - A_{T,r_T-1} - B_{T,r_T-1} \right\} \\ + T^{-1/2} \sum_{t=1}^{T} [F_{r_T}^0 - F_{r_T-1}^0] - A_{T,r_T-1} - B_{T,r_T}. \end{aligned}$$

The first term is zero, the second term is uniformly vanishing in probability by Lemma 5.A.5, and the third term vanishes in the same way as $\mathcal{B}_{Tj,4}$ above. The fifth term is zero by definition. The fourth term we can analyse as follows.

$$|A_T(x_{r_T-1})| \leq 2 \, |f(x_{r_T-1}) x_{r_T-1}| \left| T^{-1/2} \sum_{t=1}^{T} \hat{a}_t \right|.$$

Then the weighted sum is $O_p(1)$ by Assumption 5.5.1(iv), so we must consider $x_{r_T-1}f(x_{rT-1})$. We can then use that, by construction, $|\lim_{x\to\infty} F(x) - F(x_{r_T-1})| \leq T^{-1}\delta$, to write

$$\begin{aligned}
&x_{r_T-1}f(x_{r_T-1})\\
&\quad= x_{r_T-1}f(x_{r_T-1}) - \lim_{x\to\infty}[xf(x)]\\
&\quad= F^{-1}[F(x_{r_T-1})]f\{F^{-1}[F(x_{r_T-1})]\} - \lim_{x\to\infty}\left(F^{-1}[F(x)]f\{F^{-1}[F(x)]\}\right)\\
&\quad\leq \max_{|y-z|\leq T^{-1/2}\delta}\left|F^{-1}[y]f\{F^{-1}[y]\} - F^{-1}[z]f\{F^{-1}[z]\}\right|,
\end{aligned}$$

and this vanishes by the uniform continuity of $F^{-1}(\cdot)f[F^{-1}(\cdot)]$ (Assumption 5.3.6(iii)).

The left-tail case ($j = 1$ and $x < x_1$) follows similarly.

In the non-tail cases, we have

$$\begin{aligned}
\mathcal{B}_{Tj,3} &= T^{-1/2}\sum_{t=t_0}^{T}(F_j - F_{j-1}) - (B_{T,j} - B_{T,j-1})\\
&= \left\{T^{-1/2}\sum_{t=t_0}^{T}[F_j - F_j^0] - A_{T,j} - B_{T,j}\right\}\\
&\quad - \left\{T^{-1/2}\sum_{t=t_0}^{T}[F_{j-1} - F_{j-1}^0] - A_{T,j-1} - B_{T,j-1}\right\}\\
&\quad + T^{-1/2}\sum_{t=t_0}^{T}[F_j^0 - F_{j-1}^0] + [A_{T,j} - A_{T,j-1}].
\end{aligned}$$

The first two terms are uniformly vanishing in probability by Lemma 5.A.5. The third term vanishes in the same way as $\mathcal{B}_{Tj,4}$. Then it remains to consider the fourth term. We have

$$\begin{aligned}
\max_j |A_{T,j} - A_{T,j-1}| &= \max_j \left|T^{-1/2}\sum_{t=1}^{T}[f(x_j)x_j\hat{a}_t - f(x_{j-1})x_{j-1}\hat{a}_t]\right|\\
&\leq \left|T^{-1/2}\sum_{t=1}^{T}\hat{a}_t\right| \max_j |f(x_j)x_j - f(x_{j-1})x_{j-1}|.
\end{aligned}$$

The first normed term is $O_p(1)$ by Assumption 5.5.1(iv). Then we have

$$\begin{aligned}
&\max_j |f(x_j)x_j - f(x_{j-1})x_{j-1}| \\
&\quad = \max_j \left| f\{F^{-1}[F(x_j)]\}F^{-1}[F(x_j)] - f\{F^{-1}[F(x_{j-1})]\}F^{-1}[F(x_{j-1})] \right| \\
&\quad \leq \max_{|y-z| \leq T^{-1/2}\delta} \left| f\{F^{-1}[y]\}F^{-1}[y] - f\{F^{-1}[z]\}F^{-1}[z] \right|,
\end{aligned}$$

and this vanishes by the uniform continuity of $F^{-1}(\cdot)f[F^{-1}(\cdot)]$ (Assumption 5.3.6(iii)).

5.A.5 Proof of Theorem 5.5.3

We are required to prove

$$\sup_{x \in \mathbb{R}_0^+} \left| T^{1/2}\left\{\bar{G}_T(x) - \bar{G}_T^0(x)\right\} - 2A_T(x) \right| \xrightarrow{p} 0$$

We expand the supremand using that $A_T(x) = -A_T(-x)$ and $B_T(x) = B_T(-x)$ by symmetry, writing

$$\begin{aligned}
&\left| T^{1/2}\left\{\bar{G}_T(x) - \bar{G}_T^0(x)\right\} - 2A_T(x) \right| \\
&\quad \leq \left| T^{1/2}\left\{\bar{F}_T(x) - \bar{F}_T^0(x)\right\} - A_T(x) - \left[T^{1/2}\left\{\bar{F}_T(-x) - \bar{F}_T^0(-x)\right\} - A_T(-x) \right] \right| \\
&\quad \leq \left| T^{1/2}\left\{\bar{F}_T(x) - \bar{F}_T^0(x)\right\} - A_T(x) - B_T(x) \right| \\
&\quad\quad - \left| T^{1/2}\left\{\bar{F}_T(-x) - \bar{F}_T^0(-x)\right\} - A_T(-x) - B_T(-x) \right|.
\end{aligned}$$

Then the theorem follows with two applications of Lemma 5.A.5.

5.B Proof of main result (Theorem 5.4.1)

We prove Theorem 5.4.1 directly by showing that the assumptions of Theorems 5.5.2 and 5.5.3 are satisfied by the process (5.2), then deriving the limiting distribution that results using a pair of theorems of Shorack and Wellner (1986). We first present several intermediate lemmas.

5.B.1 A lemma on f_t

Lemma 5.B.1. *Under Assumptions 5.3.1, 5.3.2, 5.3.3*

$$f_t^2 - 1 = o(1) \qquad \textit{a.s.},$$
$$\sum_{t=1}^{T}(f_t^2 - 1) = O(\log T) \qquad \textit{a.s.}\ .$$

Proof. We have by 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) that

$$f_t^2 - 1 = S_t' \left(\sum_{s=1}^{t-1} S_s S_s'\right)^{-1} S_t = \frac{S_t' \left(\sum_{s=1}^{t} S_s S_s'\right)^{-1} S_t}{1 - S_t' \left(\sum_{s=1}^{t} S_s S_s'\right)^{-1} S_t} \tag{5.18}$$

Then we have by Nielsen (2005, Lemma 8.6) that

$$S_t' \left(\sum_{s=1}^{t} S_s S_s'\right)^{-1} S_t = o(1) \text{ in } t \qquad \text{a.s.},$$
$$\sum_{t=1}^{T} S_t' \left(\sum_{s=1}^{t} S_s S_s'\right)^{-1} S_t = O(\log T) \qquad \text{a.s.}\ .$$

where Nielsen's Assumptions 2.1 and 2.2 are satisfied by Assumption 5.3.2; Nielsen's Assumption 2.3 is satisfied by Assumption 5.3.1; and explosive roots are ruled out by Assumption 5.3.3. Then we can apply these directly by noting that the denominator in (5.18) must eventually be greater than $1/2$, so that

$$f_t^2 - 1 = o(1) \qquad \text{a.s.},$$
$$\sum_{t=1}^{T}(f_t^2 - 1) = O(\log T) \qquad \text{a.s.}\ .$$

□

5.B.2 A lemma on sums of $\hat{a}_t$

Lemma 5.B.2. *Under Assumptions 5.3.1, 5.3.2, 5.3.3,*

(i) $T^{-1/2} \sum_{t=1}^{T} \hat{a}_t^2 = o_p(1)$*, and*

(ii) $T^{-1/2}\sum_{t=1}^{T}\hat{a}_t = \tilde{A}_T + o_p(1)$,

where

$$\hat{a}_t = \frac{\hat{\sigma}_{t-1}f_t}{\sigma} - 1 \qquad \textit{and} \qquad \tilde{A}_T := T^{-1/2}\sum_{t=1}^{T} w_{1,T}(\eta_t^2 - 1).$$

Proof of (i) Write $\hat{a}_t^2 = \left(\sqrt{\hat{\sigma}_{t-1}^2 f_t^2/\sigma^2} - 1\right)^2$, then apply the Taylor expansion

$$\left(\sqrt{x} - 1\right)^2 = \frac{1}{4}(x-1)^2 + O_p\left[(x-1)^3\right],$$

resulting in

$$T^{-1/2}\sum_{t=1}^{T}\hat{a}_t^2 = \frac{1}{4}T^{-1/2}\sum_{t=1}^{T}(\hat{\sigma}_{t-1}^2 f_t^2/\sigma^2 - 1)^2 + O\left[T^{-1/2}\sum_{t=1}^{T}(\hat{\sigma}_{t-1}^2 f_t^2/\sigma^2 - 1)^3\right] \quad (5.19)$$

We expand $\hat{\sigma}_{t-1}^2 f_t^2/\sigma^2 - 1$ as

$$\frac{\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2}{\sigma^2} = \left(\frac{\hat{\sigma}_{t-1}^2}{\sigma^2} - 1\right) + \left(\frac{\hat{\sigma}_{t-1}^2}{\sigma^2} - 1\right)(f_t^2 - 1) + \left(f_t^2 - 1\right).$$

By Nielsen (2005, Corollary 2.9) we have that $\hat{\sigma}_{t-1}^2 - \sigma^2 = o(t^{\epsilon-1/2})$ a.s. for any $\epsilon > 0$, with with Nielsen's Assumptions 2.1, 2.3 and 2.7 satisfied by Assumptions 5.3.2 (in particular, with 4+ moments finite), 5.3.1 and 5.3.2 respectively; $f_t^2 - 1$ is $o(1)$ almost surely from above, so the entire term is $o(t^{\epsilon-1/2})$. Since this property holds almost surely, the maxima of the summands in each of the terms of (5.19) are $o(T^{\epsilon-1})$ a.s. and $o(T^{\epsilon-3/2})$ a.s. respectively; and the sums themselves are $o(T^{\epsilon})$ a.s. and $o(T^{\epsilon-1/2})$ a.s., so both terms vanish due to the $T^{-1/2}$ factor.

Proof of (ii) First apply the Taylor expansion

$$\frac{\hat{\sigma}_{t-1}f_t}{\sigma} - 1 = \sqrt{1 + \frac{\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2}{\sigma^2}} - 1 = \frac{\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2}{2\sigma^2} + O_p\left[\left(\frac{\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2}{\sigma^2}\right)^2\right].$$

Then write

$$T^{-1/2}\sum_{t=1}^{T}\hat{a}_t = T^{-1/2}\sum_{t=1}^{T}\frac{\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2}{2\sigma^2} + O_p\left[T^{-1/2}\sum_{t=1}^{T}\left(\frac{\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2}{\sigma^2}\right)^2\right]. \tag{5.20}$$

The remainder term vanishes by (i). Write the the leading term as

$$\begin{aligned}T^{-1/2}\sum_{t=1}^{T}\frac{\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2}{2\sigma^2} &= \frac{1}{2}T^{-1/2}\sum_{t=1}^{T}\left(\frac{\hat{\sigma}_{t-1}^2}{\sigma^2} - 1\right) \\ &\quad + \frac{1}{2}T^{-1/2}\sum_{t=1}^{T}\left(\frac{\hat{\sigma}_{t-1}^2}{\sigma^2} - 1\right)(f_t^2 - 1) + \frac{1}{2}T^{-1/2}\sum_{t=1}^{T}\left(f_t^2 - 1\right).\end{aligned} \tag{5.21}$$

The third of these terms is $o_p(1)$ from above, and since $f_t^2 - 1 = o(1)$ almost surely by Lemma 5.B.1, the second will be $o_p(1)$ as long as the first is $O_p(1)$, which we proceed to show.

We have that

$$\frac{1}{2}T^{-1/2}\sum_{t=1}^{T}\left(\frac{\hat{\sigma}_{t-1}^2}{\sigma^2} - 1\right) = \frac{1}{2}T^{-1/2}\sum_{t=1}^{T}\left[t^{-1}\sum_{s=1}^{t}\frac{\hat{\varepsilon}_{s,t}^2}{\sigma^2} - 1\right]$$

First we observe that the estimation error is negligible, with

$$\begin{aligned}T^{-1/2}\sum_{t=1}^{T}\left[t^{-1}\sum_{s=1}^{t}\left(\frac{\hat{\varepsilon}_{s,t}^2}{\sigma^2} - \frac{\varepsilon_s^2}{\sigma^2}\right)\right] &= \sigma^{-2}T^{-1/2}\max_{1\leq t\leq T}\left|\sum_{s=1}^{t}(\hat{\varepsilon}_{s,t}^2 - \varepsilon_s^2)\right|\left|\sum_{t=1}^{T}t^{-1}\right| \\ &= T^{-1/2}o(T^{1/2-\epsilon})O(\log T) \qquad \text{a.s.} \\ &= o(1) \quad \text{a.s.,}\end{aligned} \tag{5.22}$$

which follows since the term in the maximum is $o(T^{1/2-\epsilon})$ almost surely by Nielsen (2005, Corollary 2.6) as long as ε_1 has finite moments greater than 4, which is satisfied by Assumption 5.3.2. So we can focus on

$$\frac{1}{2}T^{-1/2}\sum_{t=1}^{T}\left[t^{-1}\sum_{s=1}^{t}\frac{\varepsilon_s^2}{\sigma^2} - 1\right] = \frac{1}{2}T^{-1/2}\sum_{t=1}^{T}\left[\eta_t^2 w_{t,T-1} - 1\right],$$

with $w_{t,T}$ defined as in (5.23) and the the right hand side following from the substitution

$\eta_t = \varepsilon_t/\sigma_t$ and resummation. Since $\sum_{t=1}^{T} w_{t,T-1} = T = \sum_{t=1}^{T} 1$ by (5.26), we can rewrite this sum as

$$\frac{1}{2}T^{-1/2}\sum_{t=1}^{T}\left[\eta_t^2 w_{t,T} - 1\right] = \frac{1}{2}T^{-1/2}\sum_{t=1}^{T} w_{t,T}\left[\eta_t^2 - 1\right]$$
$$= \frac{1}{2}T^{-1/2}\sum_{t=1}^{T} w_{t,T}\left[\eta_t^2 - \mathsf{E}\eta_t^2\right].$$

This term is $O_p(1)$, as required, by the Lindeberg-Feller CLT (details appear in Section 5.C.2).

5.B.3 Some results on weights relating to the harmonic numbers

We will require some simple summation results to derive the limiting distribution.

Define a triangular array of deterministic weights $\{w_{1,T} \ldots w_{T,T}; T \geq 1\}$, by

$$w_{t,T} := \sum_{s=t}^{T}(1/s). \tag{5.23}$$

We can then calculate the sum of the weights

$$\sum_{t=1}^{T} w_{t,T} = \sum_{t=1}^{T}\sum_{s=t}^{T}\frac{1}{s} = \sum_{s=1}^{T}\frac{1}{s}\sum_{t=1}^{s} 1 = T, \tag{5.24}$$

and the sum of squared weights

$$\begin{aligned}
\sum_{t=1}^{T} w_{t,T}^2 &= \sum_{t=1}^{T}\sum_{s=t}^{T}\frac{1}{s}\sum_{r=t}^{T}\frac{1}{r} \\
&= \sum_{s=1}^{T}\frac{1}{s}\sum_{t=1}^{s}\sum_{r=t}^{T}\frac{1}{r} \\
&= \sum_{s=1}^{T}\frac{1}{s}\left(s\sum_{r=s+1}^{T}\frac{1}{r}+\sum_{r=1}^{s}\frac{1}{r}\sum_{t=1}^{r}1\right) \\
&= \sum_{s=1}^{T}\sum_{r=s+1}^{T}\frac{1}{r}+\sum_{s=1}^{T}\frac{1}{s}\cdot s \\
&= \sum_{s=1}^{T}\sum_{r=s}^{T}\frac{1}{r}-\sum_{s=1}^{T}\frac{1}{s}+T \\
&= 2T-\sum_{s=1}^{T}\frac{1}{s}. \qquad (5.25)
\end{aligned}$$

Then $\sum_{s=1}^{T}\frac{1}{s}=H_T$, is the Tth harmonic number, with $H_T=O(\log T)$, so we have

$$T^{-1}\sum_{t=1}^{T} w_{t,T}^2 = 2+o(1). \qquad (5.26)$$

5.B.4 Joint convergence of $\mathbb{G}_T^0$ and A_T

We follow Shorack and Wellner (1986, p. 93). In particular, we invoke two theorems which we reproduce here as lemmas.

Lemma 5.B.3 (The special construction, Theorem 3.1.1 of Shorack and Wellner 1986)**.** *Suppose that* $\{c_{1,T},\dots,c_{T,T};T\geq 1\}$ *and*

$$\max_{1\leq t\leq T}\frac{c_{1,T}^2}{\sum_{t=1}^{T}c_{1,T}^2}\to 0 \qquad \textit{and} \qquad \rho_T=\frac{T^{-1}\sum_{t=1}^{T}c_{1,T}}{\sqrt{T^{-1}\sum_{t=1}^{T}c_{1,T}^2}}\to\rho. \qquad (5.27)$$

Then there exists a triangular array of row independent uniform (0,1) random variables $\{\xi_{1,T},\dots,\xi_{T,T};T\geq 1\}$ *and Brownian bridges* $\mathbb{U}$ *and* $\mathbb{W}$ *having*

$$\mathsf{Cov}[\mathbb{U}(u),\mathbb{W}(v)]=\rho[u\wedge v-uv] \qquad \textit{for } 0\leq u,v\leq 1,$$

that are all defined on a common probability space $(\Omega, \mathcal{A}, P)$, for which

$$\sup_{0\leq u\leq 1} |\mathbb{U}_T(u) - \mathbb{U}(u)| \to 0, \text{ and}$$

$$\sup_{0\leq u\leq 1} |\mathbb{W}_T(u) - \mathbb{W}(u)| \to 0,$$

on every $\omega \in \Omega$, where

$$\mathbb{U}_T(u) = T^{-1/2} \sum_{t=1}^{T} \left(1(\xi_t \leq u) - u\right), \text{ and}$$

$$\mathbb{W}_T(u) = \frac{1}{\sqrt{\sum_{t=1}^{T} c_{1,T}^2}} \sum_{t=1}^{T} c_{1,T} \left(1(\xi_t \leq u) - u\right).$$

For the next lemma we need some additional notation. Let $\mathcal{L}_2$ be the space of square integrable deterministic functions on $[0,1]$. For $h, \tilde{h} \in \mathcal{L}_2$, denote

$$\bar{h} := \int_0^1 h(u)\,\mathrm{d}u, \qquad |[h]| := \left\{\int_0^1 h^2(u)\right\}^{1/2}, \qquad \sigma_h^2 := |[h]|^2 - \bar{h}^2.$$

and

$$\sigma_{h,\tilde{h}} := \int_0^1 \left(h(u) - \bar{h}\right)\left(\tilde{h}(u) - \bar{\tilde{h}}\right)\mathrm{d}u.$$

Lemma 5.B.4 (Theorem 3.1.2 of Shorack and Wellner 1986)**.** *Let $h \in \mathcal{L}_2$. Under the hypotheses of Lemma 5.B.3, we can claim the existence of a random variable on $(\Omega, \mathcal{A}, P)$, to be denoted by $\int_0^1 h\,\mathrm{d}\mathbb{W}$ for which*

$$\int_0^1 h\,\mathrm{d}\mathbb{W}_T - \int_0^1 h\,\mathrm{d}\mathbb{W} \xrightarrow{p} 0 \qquad \text{with} \qquad \int_0^1 h\,\mathrm{d}\mathbb{W} \sim N(0, \sigma_h^2).$$

Moreover, $\mathbb{U}$, $\mathbb{W}$ and $\int_0^1 h\,\mathrm{d}\mathbb{W}$ are jointly normal, with

$$\mathsf{Cov}\left[\mathbb{W}(u), \int_0^1 h\,\mathrm{d}\mathbb{W}\right] = \int_0^u h(s)\,\mathrm{d}s - u\bar{h},$$

and

$$\mathsf{Cov}\left[\mathbb{U}(u), \int_0^1 h \, \mathrm{d}\mathbb{W}\right] = \rho\sigma_{1[0,u],h}.$$

Then we can apply these lemmas to the present problem as follows. We have $\xi_{t,T} = G^{-1}(|\eta_t|)$ for all $t \leq T$ and $T \geq 1$, and these satisfy the requirements of uniform distribution and row independence. We have $c_{t,T} = w_{t,T}$ and these satisfy the requirements of (5.27),

$$\max_{1\leq t\leq T} \frac{w_{1,T}^2}{\sum_{t=1}^T w_{1,T}^2} \to 0 \qquad \text{and} \qquad \rho_T = \frac{T^{-1}\sum_{t=1}^T w_{1,T}}{\sqrt{T^{-1}\sum_{t=1}^T w_{1,T}^2}} = \frac{1}{\sqrt{2+o(1)}} \to \frac{1}{\sqrt{2}} = \rho,$$

by (5.24) and (5.26).

We then have that $\mathbb{U}_T = \mathbb{G}_T^0$, and furthermore we use a Stieltjes integral to write

$$\tilde{A}_T^W = \frac{1}{\sqrt{\sum_{t=1}^T w_{1,T}^2}} \sum_{t=1}^T w_{1,T}\left(\eta_t^2 - 1\right) \stackrel{\mathrm{d}}{=} \int_0^1 h(u) \, \mathrm{d}\mathbb{W}_T,$$

with $h(u) = [G^{-1}(u)]^2$. This function is in $\mathcal{L}_2$ since G has finite fourth moments and

$$\int_0^1 h^2 = \int_0^1 [G^{-1}(u)]^4 \, \mathrm{d}u = \int_0^\infty x^4 \, \mathrm{d}G(x) < \infty,$$

by the change-of-variables $u = G(x)$. Similarly we have

$$\bar{h} := \int_0^1 h(u) \, \mathrm{d}u = \int_0^\infty x^2 \, \mathrm{d}G(x) = \mathsf{E}(|\eta|^2) = 1,$$

$$|[h]|^2 := \int_0^1 h^2(u) \, \mathrm{d}u = \int_0^\infty x^4 \, \mathrm{d}G(x) = \mathsf{E}(|\eta|^4), \text{ and}$$

$$\sigma_h^2 := |[h]|^2 - \bar{h}^2 = \mathsf{E}(\eta^4) - 1, \tag{5.28}$$

and with $\tilde{h}_u(v) = I(0 \leq v \leq u)$, we have

$$\bar{\tilde{h}}_u := \int_0^1 \tilde{h}_u(v) \, \mathrm{d}v = \int_0^\infty I(x \leq G^{-1}(u)) \, \mathrm{d}G(x) = u.$$

Then

$$\begin{aligned}\sigma_{I[0,u],h} &:= \int_0^1 \left(\tilde{h}_u(v) - \bar{\tilde{h}}_u\right)\left(h(v) - \bar{h}\right)\,\mathrm{d}v \\ &= \int_0^1 \tilde{h}_u(v)h(v)\,\mathrm{d}v - \int_0^1 \bar{\tilde{h}}_u(v)h(v)\,\mathrm{d}v - \int_0^1 \tilde{h}_u(v)\bar{h}\,\mathrm{d}v + \int_0^1 \bar{\tilde{h}}_u\bar{h}\,\mathrm{d}v \\ &= \int_0^{G^{-1}(u)} x^2\,\mathrm{d}G(x) - u.\end{aligned}$$

It follows by Lemma 5.B.3 that

$$\sup_{0\le u\le 1}\left|\mathbb{G}_T^0(u) - \mathbb{U}(u)\right| \to 0 \quad \text{surely,} \tag{5.29}$$

and by Lemma 5.B.4 that

$$\tilde{A}_T^W \xrightarrow{p} \int_0^1 h\,\mathrm{d}\mathbb{W}, \tag{5.30}$$

and these are jointly normal.

Then we can write, using Lemma 5.B.2,

$$\begin{aligned}A_T(x) &= \frac{1}{4}xg(x)T^{-1/2}\sum_{t=1}^{T}\hat{a}_t \\ &= \frac{1}{4}xg(x)T^{-1/2}\sum_{t=1}^{T} w_{1,T}\left(\eta_t^2 - 1\right) + o_p(1) \\ &= \frac{1}{4}xg(x)\left(T^{-1}\sum_{t=1}^{T} w_{1,T}^2\right)^{1/2}\tilde{A}_T^W + o_p(1),\end{aligned}$$

and we have that $T^{-1}\sum_{t=1}^{T} w_{1,T}^2 \to 2$, so by (5.30) and the continuous mapping theorem,

$$A_T(G^{-1}(u)) \xrightarrow{p} A(G^{-1}(u)) = \frac{1}{2\sqrt{2}}G^{-1}(u)g[G^{-1}(u)]\int_0^1 h\,\mathrm{d}\mathbb{W}.$$

Then $\int_0^1 h\,\mathrm{d}\mathbb{W} \sim N(0, \mathsf{E}(\eta_1^4) - 1)$ so

$$A_T(G^{-1}(u)) \xrightarrow{d} N\left\{0, \frac{1}{8}[G^{-1}(u)]^2[g\circ G^{-1}(u)]^2[\mathsf{E}(\eta_1^4) - 1]\right\}, \tag{5.31}$$

and

$$\mathsf{Cov}\left[\mathbb{U}(u), \int_0^1 h\,\mathrm{d}\mathbb{W}\right] = \rho\sigma_{1[0,u],h} = \frac{1}{\sqrt{2}}\left(\int_0^{G^{-1}(u)} x^2\,\mathrm{d}G(x) - u\right), \tag{5.32}$$

so

$$\mathsf{Cov}\left[\mathbb{U}(u), A[F^{-1}(v)]\right] = \frac{1}{4}G^{-1}(u)g[G^{-1}(u)]\left(\int_0^{G^{-1}(u)} x^2\,\mathrm{d}G(x) - u\right). \tag{5.33}$$

Weak convergence follows trivially from the uniform continuity of $G^{-1}g \circ G^{-1}$.

5.B.5 Proof that Assumption 5.5.1 holds

Part (i) Follows from (5.1).

Part (ii) Follows directly from non-negativity of $\hat{\sigma}_{t-1}$, f_t and σ.

Part (iii) First apply the Taylor expansion

$$\hat{a}_t = \frac{\hat{\sigma}_{t-1}f_t}{\sigma} - 1 = \sqrt{1 + \frac{\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2}{\sigma^2}} - 1 = \frac{\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2}{2\sigma^2} + O_p\left[\left(\frac{\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2}{\sigma^2}\right)^2\right],$$

then write $\hat{\sigma}_{t-1}^2 f_t^2 - \sigma^2 = (\hat{\sigma}_{t-1}^2 - \sigma^2)f_t^2 + \sigma^2(f_t^2 - 1)$. Then we have that $\hat{\sigma}_{t-1}^2 - \sigma^2 = o(1)$ almost surely by Nielsen (2005, Corollary 2.9), with Nielsen's Assumptions 2.1, 2.3 and 2.7 satisfied by Assumptions 5.3.2, 5.3.1 and 5.3.2 respectively; while $f_t^2 - 1 = o(1)$ almost surely as above.

Part (iv) Follows directly from the convergence in distribution of (5.31).

Part (v) Follows directly from Lemma 5.B.2.

Part (vi) We have that

$$T^{-1/2}\sum_{t=1}^{T}\hat{b}_t^2 = \sigma^{-2}T^{-1/2}\sum_{t=1}^{T}[(\hat{\beta}_{t-1} - \beta)' S_t]^2$$

Define the grouped lagged regressors

$$R'_{t-1} = (Y_{t-1}, Z'_{t-1} \dots, Y_{t-k}, Z'_{t-k}, D'_{t-1}),$$

so that the regressors in the ADL model can be written $(Z_t, R_{t-1})'$. Then we can equivalently regress on $(\xi_{2,t}, R_{t-1})'$. This allows us to use the argument of Nielsen and Sohkanen (2011, Lemma A.1) to write

$$\begin{aligned}\hat{b}_t &= (\hat{\beta}_{t-1} - \beta)' S_t \\ &= \sum_{s=1}^{t} \varepsilon_s S'_s \left(\sum_{s=1}^{t-1} S_s S'_s \right)^{-1} S_t \\ &= (k_t + r_t)\{1 + o(1)\},\end{aligned}$$

almost surely, where

$$k_t = \sum_{s=1}^{t} \varepsilon_s \xi'_{2,s} \left(\sum_{s=1}^{t-1} \xi_{2,s} \xi'_{2,s} \right)^{-1} \xi_{2,t} \qquad r_t = \sum_{s=1}^{t} \varepsilon_s R'_{s-1} \left(\sum_{s=1}^{t-1} R_{s-1} R'_{s-1} \right)^{-1} R_{t-1}.$$

So then

$$T^{-1/2} \sum_{t=1}^{T} \hat{b}_t^2 = T^{-1/2} \sum_{t=1}^{T} \left(k_t^2 + 2 k_t r_t + r_t^2 \right) \{1 + o(1)\} \quad \text{a.s.} \, .$$

Then each of the sums $\sum k_t^2, \sum k_t r_t, \sum r_t^2$ is $o(T^{1/2})$ almost surely by Nielsen and Sohkanen (2011, Lemma A.3) (noting that although the summands there differ by f_t^2 in the denominator, the arguments used do not depend on that).

5.B.6 Proof of Corollary 5.4.2

We have from Theorem 5.4.1 and Section 5.B.4 that the process $\mathbb{X}_G$ is Gaussian, so it remains to show that when the reference distribution F is Gaussian, the covariances of $\mathbb{X}_G$ match those of Brownian bridge.

Recall that we have

$$\mathbb{X}_G(u) := \mathbb{U}(u) + \frac{1}{\sqrt{2}} G^{-1}(u) g[G^{-1}(u)] \int_0^1 [G^{-1}(s)]^2 \, \mathrm{d}\mathbb{W}(s).$$

When F is Gaussian, we have from (5.4) that

$$G(y) = 2[\Phi(y) - 1/2], \qquad G^{-1}(u) = \Phi^{-1}\left[u/2 + 1/2\right], \qquad g(y) = 2\phi(y).$$

Then we are interested in the covariance

$$\begin{aligned}
&\mathsf{Cov}\left[\mathbb{X}_G(u), \mathbb{X}_G(v)\right] \\
&\quad = \mathsf{Cov}[\mathbb{U}(u), \mathbb{U}(v)] \qquad (5.34)\\
&\quad + \frac{1}{\sqrt{2}} G^{-1}(v) g[G^{-1}(v)] \cdot \mathsf{Cov}\left\{\mathbb{U}(u), \int_0^1 [G^{-1}(s)]^2 \, \mathrm{d}\mathbb{W}(s)\right\} \\
&\quad + \frac{1}{\sqrt{2}} G^{-1}(u) g[G^{-1}(u)] \cdot \mathsf{Cov}\left\{\int_0^1 [G^{-1}(s)]^2 \, \mathrm{d}\mathbb{W}(s), \mathbb{U}(v)\right\} \\
&\quad + \frac{1}{\sqrt{2}} G^{-1}(u) g[G^{-1}(u)] \frac{1}{\sqrt{2}} G^{-1}(v) g[G^{-1}(v)] \cdot \mathsf{Var}\left\{\int_0^1 [G^{-1}(s)]^2 \, \mathrm{d}\mathbb{W}(s)\right\}.
\end{aligned}$$

We have from (5.28) that

$$\mathsf{Var}\left\{\int_0^1 [G^{-1}(s)]^2 \, \mathrm{d}\mathbb{W}(s)\right\} = \mathsf{E}(\eta_1^4) - 1 = 3 - 1 = 2,$$

where the result follows because we have assumed $\eta_1 \sim N(0,1)$.

And we have from (5.32) that

$$\begin{aligned}
\mathsf{Cov}\left\{\mathbb{U}(u), \int_0^1 [G^{-1}(s)]^2 \, \mathrm{d}\mathbb{W}(s)\right\} &= \frac{1}{\sqrt{2}} \int_0^{G^{-1}(u)} x^2 \, \mathrm{d}G(x) - u \\
&= \frac{1}{\sqrt{2}} \int_0^{G^{-1}(u)} (x^2 - 1) \, \mathrm{d}G(x) - u \\
&= \frac{1}{\sqrt{2}} \Big[-x g(x)\Big]_0^{G^{-1}(u)} \\
&= -\frac{1}{\sqrt{2}} G^{-1}(u) g[G^{-1}(u)],
\end{aligned}$$

where we have used that $\phi'(x) = -x\phi(x)$ and $\phi''(x) = (x^2 - 1)\phi(x)$.

Combining these results back in to (5.34) we find that the second, third and fourth terms cancel, leaving

$$\mathsf{Cov}\left[\mathbb{X}_G(u), \mathbb{X}_G(v)\right] = \mathsf{Cov}[\mathbb{U}(u), \mathbb{U}(v)].$$

Then since $\mathbb{U}$ is a Brownian bridge, we have shown that $\mathbb{X}_G$ has the covariance structure of a Brownian bridge, as required.

5.C Alternative derivations

In this section we present, for interest only, direct derivations of the variance and covariances calculated in Section 5.B.4. We begin by presenting some additional results on the weights $w_{t,T}$ which arise in the covariance expressions.

5.C.1 Some further results on weights

For the following sections we also require some double sums. We have

$$\begin{aligned}\sum_{t=1}^{T}\sum_{s=t}^{T} w_{s,T} &= \sum_{s=1}^{T} s w_{s,T} \\ &= \sum_{s=1}^{T} s \sum_{t=s}^{T} \frac{1}{t} \\ &= \sum_{t=1}^{T} \frac{1}{t} \sum_{s=1}^{t} s \\ &= \frac{T(T+3)}{4},\end{aligned}$$

and

$$\begin{aligned}\sum_{t=1}^{T}\sum_{s=1}^{t} w_{s,T} &= \sum_{s=1}^{T} w_{s,T}(T-s+1) \\ &= T\sum_{s=1}^{T} w_{s,T} - \sum_{s=1}^{T} s w_{s,T} + \sum_{s=1}^{T} w_{s,T} \\ &= \frac{T(3T+1)}{4},\end{aligned}$$

so then

$$\sum_{t=1}^{T}\sum_{s=t+1}^{T} w_{s,T} = \sum_{t=1}^{T}\sum_{s=t}^{T} w_{s,T} - \sum_{t=1}^{T} w_{t,T} = \frac{T(T-1)}{4}, \text{ and}$$

$$\sum_{t=1}^{T}\sum_{s=1}^{t-1} w_{s,T} = \sum_{t=1}^{T}\sum_{s=1}^{t} w_{s,T} - \sum_{t=1}^{T} w_{t,T} = \frac{3T(T-1)}{4}.$$

5.C.2 Marginal distribution

We first derive the marginal asymptotic distribution of $A_T(x)$. We have

$$A_T(x) = xf(x)T^{-1/2}\sum_{t=1}^{T}\hat{a}_t = \frac{1}{2}xg(x)T^{-1/2}\sum_{t=1}^{T}\hat{a}_t.$$

Then from Lemma 5.B.2 we have

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{T}\hat{a}_t &= T^{-1/2}\sum_{t=1}^{T} w_{1,T}(\eta_t^2 - 1) + o_p(1) \\ &= T^{-1/2}\sum_{t=1}^{T} w_{1,T}\left[\eta_t^2 - \mathsf{E}\eta_t^2\right] + o_p(1). \end{aligned}$$

Now we can show convergence in distribution of the leading term, writing

$$S_T := \frac{1}{2}T^{-1/2}\sum_{t=1}^{T} w_{t,T}\left[\eta_t^2 - \mathsf{E}\eta_t^2\right] = \sum_{t=1}^{T-1} X_t,$$

where $X_t = \frac{1}{2}T^{-1/2}w_{t,T}(\eta_t^2 - 1)$, with $\mathsf{E}X_t = 0$ and $\mathsf{V}X_t = \frac{1}{4}T^{-1}w_{t,T-1}^2(\mathsf{E}\eta_1^4 - 1)$. Define the sum of variances

$$\begin{aligned} s_T^2 &:= \sum_{t=1}^{T}\mathsf{V}X_t \\ &= \frac{1}{4}(\mathsf{E}\eta_1^4 - 1)T^{-1}\sum_{t=1}^{T} w_{t,T-1}^2 \\ &\to \frac{1}{2}(\mathsf{E}\eta_1^4 - 1), \end{aligned}$$

using (5.26). Then by the Lindeberg-Feller central limit theorem, we have

$$\frac{S_{T-1}}{\sqrt{s_{T-1}^2}} = \frac{\frac{1}{2}T^{-1/2}\sum_{t=1}^{T} w_{t,T}\left[\eta_t^2 - 1\right]}{\sqrt{(1/2)(\mathsf{E}\eta_t^4 - 1)}} \xrightarrow{d} N(0,1), \tag{5.35}$$

as long as a Lindeberg condition is satisfied. We show the stronger Lyupanov condition, which requires that for some $\delta > 0$,

$$\frac{1}{s_T^{2+\delta}}\sum_{t=1}^{T}\mathsf{E}\left(|X_t|^{2+\delta}\right) \to 0.$$

Since s_T converges, this reduces to showing $\sum_{t=1}^{T}\mathsf{E}\left(|X_t|^{2+\delta}\right) \to 0$. We have that for any $\delta > 0$,

$$\mathsf{E}\left(|X_t|^{2+\delta}\right) = \frac{1}{2^{2+\delta}}T^{-1-\delta/2}w_{t,T}^{2+\delta}\mathsf{E}\left[(\eta_t^2 - 1)^{2+\delta}\right].$$

On the right hand side, the expectation is finite by Assumption 5.3.2. We have that $w_{t,T} \leq w_{1,T} = O(\log T)$ for any t, so $w_{t,T}^{2+\delta}$ is increasing logarithmically. Hence with the factor $T^{-1-\delta/2}$, the entire expression will vanish as required.

So the conditions of the Lindeberg-Feller CLT are satisfied and combining (5.35) with (5.22), (5.21) and (5.20), we have

$$T^{-1/2}\sum_{t=1}^{T}\hat{a}_t \xrightarrow{d} N\left(0, \frac{1}{2}(\mathsf{E}\eta_t^4 - 1)\right),$$

and

$$A_T(x) = \frac{1}{2}xg(x)T^{-1/2}\sum_{t=1}^{T}\hat{a}_t \xrightarrow{d} N\left(0, \frac{1}{8}x^2[g(x)]^2(\mathsf{E}\eta_t^4 - 1)\right).$$

5.C.3 Covariances

Next we derive the asymptotic covariance between $\mathbb{G}_T^0(G(x))$ and $A_T(y)$. We have

$$\mathbb{G}_T^0(G(x)) = T^{-1/2}\sum_{t=1}^{T}\left[I(|\eta_t| \le x) - G(x)\right], \text{ and}$$
$$A_T(y) = \frac{1}{4}yg(y)T^{-1/2}\sum_{t=1}^{T} w_{t,T}\left[\eta_t^2 - 1\right] + o_p(1).$$

We ignore the $o_p(1)$ term which does not enter in the limit. Then both terms are mean zero and we are interested in

$$\begin{aligned}
&\mathsf{E}\left[\left(T^{-1/2}\sum_{t=1}^{T}\left[I(|\eta_t| \le x) - G(x)\right]\right)\left(\frac{1}{4}yg(y)T^{-1/2}\sum_{t=1}^{T} w_{t,T}\left[\eta_t^2 - 1\right]\right)\right] \\
&\quad = \frac{1}{4}yg(y)T^{-1}\mathsf{E}\left[\left(\sum_{t=1}^{T} I(|\eta_t| \le x) - TG(x)\right)\left(\sum_{t=1}^{T} w_{t,T}\eta_t^2 - T\right)\right] \\
&\quad = \frac{1}{4}yg(y)T^{-1}\mathsf{E}\left[\sum_{t=1}^{T} I(|\eta_t| \le x)\sum_{s=1}^{T} w_{s,T}\eta_s^2\right] - \frac{1}{4}yf(y)TG(x). \qquad (5.36)
\end{aligned}$$

Then consider the expectation term, breaking the double sum into diagonal, upper and lower triangle terms.

$$\begin{aligned}
&\mathsf{E}\left[\sum_{t=1}^{T} I(|\eta_t| \le x)\sum_{s=1}^{T} w_{s,T}\eta_s^2\right] \\
&\quad = \mathsf{E}\left[\sum_{t=1}^{T} I(|\eta_t| \le x)w_{t,T}\eta_t^2 + \sum_{t=1}^{T}\sum_{s=t+1}^{T} I(|\eta_t| \le x)w_{s,T}\eta_s^2 + \sum_{t=1}^{T}\sum_{s=1}^{t-1} I(|\eta_t| \le x)w_{s,T}\eta_s^2\right]. \qquad (5.37)
\end{aligned}$$

The diagonal sum is easily analysed

$$\mathsf{E}\left[\sum_{t=1}^{T} I(|\eta_t| \le x)w_{t,T}\eta_t^2\right] = \sum_{t=1}^{T} w_{t,T}\mathsf{E}\left[I(|\eta_t| \le x)\eta_t^2\right] = T\mathsf{E}\left[I(|\eta_1| \le x)\eta_1^2\right].$$

The upper triangle sum is analysed using the independence of η_t and η_s for $s \neq t$.

$$\begin{aligned}\mathsf{E}\left[\sum_{t=1}^{T}\sum_{s=t+1}^{T} I(|\eta_t| \leq x) w_{s,T}\eta_s^2\right] &= \sum_{t=1}^{T}\sum_{s=t+1}^{T} w_{s,T}\mathsf{E}\left[I(|\eta_t| \leq x)\right]\mathsf{E}\left[\eta_s^2\right] \\ &= G(x)\sum_{t=1}^{T}\sum_{s=t+1}^{T} w_{s,T} \\ &= \frac{1}{4}T(T-1)G(x).\end{aligned}$$

And the same approach is used on the lower triangle sum.

$$\begin{aligned}\mathsf{E}\left[\sum_{t=1}^{T}\sum_{s=1}^{t-1} I(|\eta_t| \leq x) w_{s,T}\eta_s^2\right] &= \sum_{t=1}^{T}\sum_{s=1}^{t-1} w_{s,T}\mathsf{E}\left[I(|\eta_t| \leq x)\right]\mathsf{E}\left[\eta_s^2\right] \\ &= G(x)\sum_{t=1}^{T}\sum_{s=1}^{t-1} w_{s,T} \\ &= \frac{1}{4}3T(T-1)G(x).\end{aligned}$$

Then recombining the parts of (5.37) we have

$$\mathsf{E}\left[\sum_{t=1}^{T} I(|\eta_t| \leq x)\sum_{s=1}^{T} w_{s,T}\eta_s^2\right] = T\mathsf{E}\left[I(|\eta_1| \leq x)\eta_1^2\right] + \left(T^2 - T\right)G(x),$$

and substituting this back into (5.36) the asymptotic covariance between $\mathbb{G}_T^0(G(x))$ and $A_T(y)$ is equal to

$$\frac{1}{4}yg(y)\left(\mathsf{E}\left[I(|\eta_1| \leq x)\eta_1^2\right] - \frac{1}{2}G(x)\right) = \frac{1}{4}yg(y)\left(\int_0^x r^2\,\mathrm{d}G(r) - G(x)\right).$$

Chapter 6

Conclusion

This thesis is intended to contribute to the literature on misspecification tests in time series econometrics. Such misspecification testing is important in validating the assumptions of a model and so ensuring that subsequent inference is correct. We have focused in particular on the 1-step recursive Chow statistic, which, though available as diagnostic output in common econometrics software, has not previously attracted significant theoretical attention.

In Chapter 2, we directly addressed the 1-step statistic, proving an almost sure pointwise convergence result under a broad class of time series models and processes, including unit root and explosive processes. From this result, one can derive distributional results for any particular assumption about the error distribution. We then used this result to establish the asymptotic distribution of the maximum of a sequence of 1-step statistics with normal errors, which allows joint consideration of the sequence of 1-step tests. We construct a further misspecification test from this: the sup-Chow test.

As this test is novel, questions naturally arise about its power against various plausible forms misspecification. In Chapter 3, we endeavoured to answer these questions, using simulation to investigate the power properties of this test, both alone and in comparison with benchmark tests of structural instability. We found that the sup-Chow test may have advantages when the nature of structural instability is unknown, though the benchmarks dominate when the alternative can be more precisely specified.

In Chapter 4, we considered how the test may be adapted to a situation in which the errors cannot be assumed normal. We considered two main approaches, each appropriate to a different modelling philosophy. When the error distribution is fully specified, this assumption should first be tested, and our results on the conditional size and power of the sup-Chow test are relevant. When the error distribution is not specified, the proposed transformations of the test may be of more interest. In either approach, we observed a trade-off between robustness and power.

In Chapter 5 we analysed an empirical process formed from the square root of the 1-step statistics. We proved weak convergence to a process involving a pair of correlated Brownian bridges. Under the assumption of normal errors, this limiting distribution reduces to that of a Brownian bridge. We found that the asymptotic approximation works well in small samples, at least for the case of a Kolmogorov-Smirnov-type statistic. These results clear the way for a range of tests, based on the empirical process, to be developed and evaluated.

Several problems remain open. First, neither of the Chapter 4 approaches to distribution sensitivity is completely satisfactory, with power decreasing as robustness increases. It may be that this is inherent to the structure of the 1-step statistics. Further investigation would be worthwhile. Second, it would be appealing to generalise the empirical process results of Chapter 5 to explosive processes, as in Chapter 2. Although a relatively uncommon concern for empirical investigators, such processes do arise in, for example, the modelling of financial bubbles (see, e.g. Engsted and Nielsen 2012). If explosive roots are permitted, however, several key results relied upon would not be available, and so a substantially different proof may be required. Third, having established the weak convergence result of Chapter 5, it would be of great practical interest to investigate the properties of tests based on the empirical process of 1-step statistics, as Chapter 3 does in respect of the theory of Chapter 2. In a similar vein, with the simple limiting process, it should be relatively straightforward to investigate asymptotic power of tests constructed from the empirical process. Chapter 5 is a quite general result, however, and the number of tests which could be adapted to take advantage of it is large; hence this remains a project for future work.

Appendix A

Common simulation framework

A.1 Computational environment

A common computational environment was used for all simulations reported in this thesis. OxMetrics 6.10 Professional for OS X (Doornik 2007), incorporating the Ox matrix programming interpreter, was used as the programming and execution environment.

Standard normal random numbers were generated by the `rann` function using the ziggurat method, using Ox 6.10's default uniform random number generator (Marsaglia multiply-with-carry, `MWC8222_52`, with a period of approximately 2^{8222}) and seed. We list the number of Monte Carlo replications in tables and figures as M (e.g. $M = 5000$). In general, identical sequences of random numbers were used for identical experiments, even when they appear in different sections, to ensure consistency of reported results.

The code for recursive estimation and calculation of the 1-step Chow statistic was written according to the description given in Section 1.2; testing confirms that these routines are numerically equivalent to those implemented in *PcGive*, to the accuracy reported by that package.

The simulations comprise around 7000 lines of Ox code ($\sim$1000 lines core probability and statistics; $\sim$1500 lines testing of core probability and statistics; $\sim$2000 lines descriptions of experiments; $\sim$1000 lines output and presentation; $\sim$1500 lines utility procedures).

A.2 Experiments

We use eight different DGPs in simulations. The page reference refers to the place in the text where the DGP is defined. WN means (independent) white noise.

Label	Short description	Ref.
DGP1	Gaussian AR(1) no constant	p. 38
DGP2	Gaussian WN; with mean break	p. 61
DGP3	Gaussian AR(1) no constant; with AR coefficient break	p. 63
DGP4	Gaussian AR(1) with constant; with AR coefficient break	p. 64
DGP5	Gaussian WN; with variance break	p. 66
DGP6	Gaussian AR(1); with additive outlier	p. 68
DGP7	Gaussian AR(1); with innovation outlier	p. 70
DGP8	Non-Gaussian WN	p. 78

We fit only two models in experiments.

Label	Short description	Ref.
M1	AR(1) autoregression with intercept	p. 38
M2	AR(1) autoregression without intercept	p. 39

A.3 Tests

Below, we list the various tests which are reported for simulations. The page reference refers to the place in text where the test is described.

A.3.1 Sup-Chow tests

Symbol	Test	Ref.	Parameterization	Validated against
SC^2	asymptotic	p. 37	$g(T) = \sqrt{T}$	PcGive (1-step test)
$\mathsf{SC}^{2\dagger}$	variant	p. 59	$g(T) = \sqrt{T}$	PcGive (1-step test)
$\mathsf{SC}^{2\ddagger}$	variant	p. 60	$g(T) = \sqrt{T}$	PcGive (1-step test)
SC^{2*}	preferred variant	p. 60	$g(T) = \sqrt{T}$	PcGive (1-step test)
SC^{2*}_D	spacings	p. 84	$g(T) = \sqrt{T}$	PcGive (1-step test)
SC^{2*}_{DW}	finite-weighted spacings	p. 85	$g(T) = \sqrt{T}$	PcGive (1-step test)
SC^{2*}_R	ratio	p. 88	$g(T) = \sqrt{T}$	PcGive (1-step test)
SC^{2*}_{RS}	ratio-sum	p. 89	$g(T) = \sqrt{T}$	PcGive (1-step test)

A.3.2 Other tests

Symbol	Test	Ref.	Parameterization	Validated against
$F_{0.5}$	mid-point F			
$\sup F$	Andrews (1993)	p. 19	$(a, b) = (0.15, 0.85)$	EViews `ubreak`
$S_{0.15}$	Andrews (2003*a*)	p. 19	$m = 0.15T$	
S_1	Andrews (2003*a*)	p. 19	$m = 1$	
Φ	Doornik and Hansen (2008) E_p	p. 80		

A.3.3 Size-corrected power

Size-corrected power results are generated to separate the issues of size distortion and power. For (DGP1) and (M1), for a test of nominal size α, we simulate the empirical distribution of the test statistic of interest, under the null hypothesis, and select the $1 - \alpha$ quantile. This is then used as the critical value in the experiment.

The exception is for the subsampled statistics (SC^{2*}_D, SC^{2*}_{DW}, $S_{0.15}$ and S_1), for which size is discrete; in these cases we generate p-values, p, and report the $1 - \alpha$ quantile of $1 - p$ (which should be $1 - \alpha$ for a correctly-sized test). In these cases the size-correction is at best partially effective.

The results of this simulation are provided in Table A.1 below.

Right-tail quantile (α)	Statistic	T **25**	**50**	**100**	**200**	**500**
0.01	Φ	9.538	9.842	10.000	9.961	9.649
	SC^2	9.449	7.697	6.505	5.805	5.089
	$\mathsf{SC}^{2\dagger}$	4.263	4.259	4.297	4.380	4.356
	$\mathsf{SC}^{2\ddagger}$	22.279	20.278	19.281	19.225	19.537
	SC^{2*}	11.906	13.402	14.864	16.374	18.072
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	0.933	0.974	0.988	0.994	0.998
	$\mathsf{SC}^{2*}_{\mathsf{DW}}$	0.933	0.974	0.988	0.992	0.986
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	107.771	104.531	105.209	101.319	98.088
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	0.815	0.575	0.427	0.318	0.209
	$\mathsf{SC}^{2*}_{\mathsf{RSE}}$	13.191	8.105	6.699	6.049	5.535
	$\sup F$	21.472	17.271	16.031	15.512	15.477
	$F_{0.5}$	5.101	4.791	4.679	4.586	4.615
	$S_{0.15}$	1.000	1.000	1.000	1.000	1.000
	S_1	1.000	1.000	0.990	0.990	0.990
0.05	Φ	5.920	5.921	5.984	5.971	6.013
	SC^2	5.073	4.533	3.994	3.621	3.295
	$\mathsf{SC}^{2\dagger}$	2.680	2.753	2.776	2.795	2.820
	$\mathsf{SC}^{2\ddagger}$	13.527	13.949	14.258	14.856	15.949
	SC^{2*}	8.741	10.390	11.822	13.204	15.000
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	0.933	0.974	0.975	0.963	0.958
	$\mathsf{SC}^{2*}_{\mathsf{DW}}$	0.933	0.974	0.958	0.950	0.944
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	20.941	19.656	19.420	19.736	19.174
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	0.672	0.423	0.303	0.219	0.140
	$\mathsf{SC}^{2*}_{\mathsf{RSE}}$	6.133	4.406	3.910	3.645	3.432
	$\sup F$	14.632	12.257	11.732	11.541	11.510
	$F_{0.5}$	3.071	3.001	3.009	2.999	2.991
	$S_{0.15}$	0.947	0.944	0.943	0.950	0.946
	S_1	0.957	0.958	0.949	0.949	0.950

Table A.1: Simulated quantiles for statistics used in this thesis, under (DGP1)/(M1). Note for SC^{2*}_{D}, SC^{2*}_{DW}, $S_{0.15}$ and S_1 the values reported are $1-p$ where p represents a p-value. $M = 200000$.

Table A.2 illustrates the effect of size correction on the SC family of statistics.

		$\alpha =$							
		0.00		0.50		1.00		1.03	
		Corrected							
T		**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
25	SC^2	14.4	5.1	15.7	5.7	19.8	7.5	20.3	7.7
	$\mathsf{SC}^{2\dagger}$	3.8	5.2	4.3	5.7	5.5	7.4	5.7	7.5
	$\mathsf{SC}^{2\ddagger}$	17.1	5.1	18.6	5.7	23.2	7.5	23.6	7.7
	SC^{2*}	5.4	5.2	5.8	5.7	7.6	7.4	7.7	7.5
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.5	5.0	5.4	4.9	5.4	4.9	5.4	5.0
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	6.7	5.0	6.8	4.9	6.6	4.9	6.7	4.9
50	SC^2	12.6	4.9	13.5	5.3	16.9	7.1	17.4	7.3
	$\mathsf{SC}^{2\dagger}$	3.9	4.9	4.1	5.2	5.5	6.7	5.7	7.0
	$\mathsf{SC}^{2\ddagger}$	14.8	4.9	15.8	5.3	19.6	7.1	20.2	7.3
	SC^{2*}	5.1	4.9	5.3	5.2	6.9	6.7	7.2	7.0
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	8.9	8.9	9.0	9.0	8.8	8.8	8.8	8.8
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.2	5.0	5.2	5.0	5.4	5.2	5.2	5.0
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	6.6	5.0	6.6	5.0	6.5	4.9	6.6	5.0
100	SC^2	10.4	5.0	10.7	5.2	13.3	6.7	13.9	7.1
	$\mathsf{SC}^{2\dagger}$	4.0	4.9	4.1	5.0	5.2	6.3	5.5	6.7
	$\mathsf{SC}^{2\ddagger}$	12.0	5.0	12.5	5.2	15.4	6.7	16.1	7.1
	SC^{2*}	5.0	4.9	5.1	5.0	6.4	6.3	6.8	6.7
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.8	6.8	6.8	6.8	6.9	6.9	6.8	6.8
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.1	5.0	5.0	4.9	5.2	5.1	5.0	4.9
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	6.3	5.0	6.2	5.0	6.3	5.0	6.2	5.0
200	SC^2	8.3	4.9	8.5	5.1	10.1	6.1	10.5	6.4
	$\mathsf{SC}^{2\dagger}$	4.2	4.9	4.2	5.0	4.9	5.8	5.0	6.0
	$\mathsf{SC}^{2\ddagger}$	9.6	4.9	9.9	5.1	11.7	6.1	12.1	6.4
	SC^{2*}	5.0	4.9	5.1	5.0	5.9	5.8	6.0	6.0
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	6.1	6.1	6.1	6.1	6.1	6.1	5.9	5.9
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.1	4.9	5.1	4.9	5.1	4.9	5.1	5.0
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	6.1	5.0	6.2	5.0	6.1	5.0	5.9	4.9
500	SC^2	6.7	5.0	6.7	5.0	7.6	5.7	7.6	5.7
	$\mathsf{SC}^{2\dagger}$	4.3	5.1	4.3	5.1	4.7	5.5	4.7	5.4
	$\mathsf{SC}^{2\ddagger}$	7.7	5.0	7.7	5.0	8.6	5.7	8.6	5.7
	SC^{2*}	5.1	5.1	5.1	5.1	5.6	5.5	5.5	5.4
	$\mathsf{SC}^{2*}_{\mathsf{D}}$	5.7	5.7	5.7	5.7	5.6	5.6	5.7	5.7
	$\mathsf{SC}^{2*}_{\mathsf{R}}$	5.1	5.0	5.1	5.0	5.0	5.0	5.1	5.1
	$\mathsf{SC}^{2*}_{\mathsf{RS}}$	6.0	5.1	6.1	5.2	6.1	5.2	6.1	5.2

Table A.2: Illustration of size-correction. Simulated rejection frequency (%) for the SC^2 statistics under (DGP1)/(M1). Nominal size 5%. Note that SC^{2*}_D cannot be calculated when $T = 25$. $M = 100000$. MCSE < 0.2.

References

Andrews, Donald W. K. (1989), Tests for parameter instability and structural change with unknown change point, Cowles Foundation Discussion Papers 943, Cowles Foundation for Research in Economics, Yale University.

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

Andrews, Donald W. K. (2003*a*), 'End-of-sample instability tests', *Econometrica* **71**(6), 1661–1694.

Andrews, Donald W. K. (2003*b*), 'Tests for parameter instability and structural change with unknown change point: A corrigendum', *Econometrica* **71**(1), 395–397.

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.

Bercu, Bernard and Abderrahmen Touati (2008), 'Exponential inequalities for self-normalized martingales with applications', *Annals of Applied Probability* **18**(5), 1848–1869.

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

Boldin, M. (1983), 'Estimation of the distribution of noise in an autoregression scheme', *Theory of Probability & Its Applications* **27**(4), 866–871.

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.

Burridge, Peter and A. M. Robert Taylor (2006), 'Additive outlier detection via extreme-value theory', *Journal of Time Series Analysis* **27**(5), 685–701.

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.

Chang, I., G. C. Tiao and C. Chen (1988), 'Estimation of time series parameters in the presence of outliers', *Technometrics* **31**, 193–204.

Chang, I. and G.C. Tiao (1983), Estimation of time series parameters in the presence of outliers, Technical Report 8, University of Chicago, Statistics Research Center.

Chen, C. and L. M. Liu (1993), 'Forecasting time series with outliers', *Journal of Forecasting* **12**, 13–35.

Cheng, M. Y. and L. Peng (2002), 'Regression modeling for nonparametric estimation of distribution and quantile functions', *Statistica Sinica* **12**, 1043–60.

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

Clements, Michael P. and David F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge University Press, Cambridge, Mass.

Clements, M.P. and D.F. Hendry (1999), *Forecasting Non-stationary Economic Time Series*, MIT Press, Cambridge, Mass.

Cox, D. R and D. V Hinkley (1974), *Theoretical Statistics*, Chapman and Hall, London.

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

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

Davidson, James (2006), Asymptotic methods and functional central limit theorems, *in* T. C.Mills and K.Patterson, eds, 'Palgrave Handbook of Econometrics', Vol. 1, Palgrave Macmillan.

Davies, R. B. (1977), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika* **64**(2), 247–254.

de Bruijn, N.G. (1961), *Asymptotic methods in analysis*, North-Holland, Amsterdam.

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

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.

Doornik, J.A. (2007), *Object-Oriented Matrix Programming Using Ox*, 3rd edn, Timberlake Consultants Press and Oxford, London.

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.

Engsted, Tom and Bent Nielsen (2012), 'Testing for rational bubbles in a coexplosive vector autoregression', *The Econometrics Journal* **15**(2), 226–254.

Ericsson, Neil R., Julia Campos and Hong-Anh Tran (2005), Pc-Give and David Hendry's econometric methodology, *in* J.Campos, N. R.Ericsson and D. F.Hendry, eds, 'General-to-Specific Modelling', Vol. 1 of *The International Library of Critical Writings in Econometrics*, Edward Elgar, Cheltenham, UK, pp. 140–250.

Fisher, Franklin M (1970), 'Tests of equality between sets of coefficients in two linear regressions: An expository note', *Econometrica* **35**(2), 361–366.

Fisher, R.A. (1925), *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.

Fox, A. J. (1972), 'Outliers in time series', *Journal of the Royal Statistical Society. Series B (Methodological)* **34**(3), pp. 350–363.

Frisch, R. (1933), 'Editor's note', *Econometrica* **1**(1), 1–4.

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.

Haavelmo, Trygve (1944), 'The probability approach in econometrics', *Econometrica* **12**, (iii)–(vi), 1–115.

Hansen, Bruce E. (1997), 'Approximate asymptotic p values for structural-change tests', *Journal of Business & Economic Statistics* **15**(1), 60–67.

Hawkins, Douglas M. (1977), 'Testing a sequence of observations for a shift in location', *Journal of the American Statistical Association* **72**(357), 180–186.

Hawkins, Douglas M. (1991), 'Diagnostics for use with regression recursive residuals', *Technometrics* **33**(2), 221–234.

Hendry, D. F. (2009), The methodology of empirical econometric modelling: Applied econometrics through the looking-glass, *in* T. C.Mills and K.Patterson, eds, 'Palgrave Handbook of Econometrics', Vol. 2, Palgrave Macmillan.

Hendry, David F. (1980), 'Econometrics—alchemy or science?', *Economica* **47**(188), 387–406.

Hendry, David F. (1983), *Technical Manual for GIVE*, Nuffield College, Oxford.

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.

Hendry, David F. (2000), 'On detectable and non-detectable structural change', *Structural Change and Economic Dynamics* **11**(1-2), 45–65.

Hendry, David F. and Bent Nielsen (2007), *Econometric Modeling: A Likelihood Approach*, Princeton University Press, Princeton.

Hendry, David F. and Jurgen A. Doornik (2001), *Empirical Econometric Modelling Using PcGive*, Vol. 1, Timberlake Consultants Ltd, London.

Hendry, David F., Mary Morgan and Frank Srba (1983), *GIVE Guide*.

Hendry, David F. and Neil R. Ericsson (1991), 'Modeling the demand for narrow money in the United Kingdom and the United States', *European Economic Review* **35**(4), 833–881.

Hoover, Kevin D. (2006), The methodology of econometrics, *in* T. C.Mills and K.Patterson, eds, 'Palgrave Handbook of Econometrics', Vol. 1, Palgrave Macmillan.

Jevons, William Stanley (1871), *The Theory of Political Economy*, Macmillan, London.

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

Kianifard, Farid and William H. Swallow (1996), 'A review of the development and application of recursive residuals in linear models', *Journal of the American Statistical Association* **91**(433), 391–400.

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.

Koul, Hira L. and Shlomo Levental (1989), 'Weak convergence of the residual empirical process in explosive autoregression', *The Annals of Statistics* **17**(4), 1784–1794.

Koul, H.L. (2002), *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd edn, Springer, New York.

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

Lai, T. L. and C.-Z. Wei (1982), 'Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems', *Annals of Statistics* **10**(1), 154–166.

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, North Holland, Amsterdam, pp. 375–393.

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.

Lee, Sangyeol and Ching-Zong Wei (1999), 'On residual empirical processes of stochastic regression models with applications to time series', *The Annals of Statistics* **27**(1), 237–261.

Lee, Sangyeol, Okyoung Na and Seongryong Na (2003), 'On the CUSUM of squares test for variance change in nonstationary and nonparametric time series models', *Annals of the Institute of Statistical Mathematics* **55**(3), 467–485.

Ling, Shiqing (1998), 'Weak convergence of the sequential empirical processes of residuals in nonstationary autoregressive models', *The Annals of Statistics* **26**(2), 741–754.

Mizon, Grayham E. (1977), 'Inferential procedures in nonlinear models: An application in a UK industrial cross section study of factor substitution and returns to scale', *Econometrica* **45**(5), 1221–1242.

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 Andrew Whitby (2012), A joint Chow test for structural instability, Discussion Paper 2012-W07, Nuffield College.

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.

O'Reilly, Federico J. and C. P. Quesenberry (1973), 'The conditional probability integral transformation and applications to obtain composite chi-square goodness-of-fit tests', *The Annals of Statistics* **1**(1), 74–83.

Pagan, Adrian Rodney (1987), 'Three econometric methodologies: A critical appraisal', *Journal of Economic Surveys* **1**(1), 3–24.

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.

Pearson, E. S. (1938), 'The probability integral transformation for testing goodness of fit and combining independent tests of significance', *Biometrika* **30**(1/2), 134–148.

Perron, Pierre (2006), Dealing with structural breaks, *in* T. C.Mills and K.Patterson, eds, 'Palgrave Handbook of Econometrics', Palgrave Macmillan, pp. 278–352.

Perron, Pierre and Gabriel Rodríguez (2003), 'Searching for additive outliers in nonstationary time series', *Journal of Time Series Analysis* **24**(2), 193–220.

Pierce, Donald A. (1985), 'Testing normality in autoregressive models', *Biometrika* **72**(2), 293–297.

Ploberger, Werner and Walter Krämer (1986), 'On studentizing a test for structural change', *Economics Letters* **20**(4), 341–344.

QMS (2009), *EViews 7 User's Guide II*.

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.

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org*

Rudin, Walter (1976), *Principles of Mathematical Analysis*, 3rd edn, McGraw-Hill, New York.

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

Shenton, L. R. and K. O. Bowman (1977), 'A bivariate model for the distribution of $\sqrt{b_1}$ and b_2', *Journal of the American Statistical Association* **72**(357), 206–211.

Shorack, Galen R. and Jon A. Wellner (1986), *Empirical Processes with Applications to Statistics*, John Wiley & Son, New York.

Sohkanen, Jouni S. (2011), Properties of tests for mis-specification in non-stationary autoregressions, PhD thesis, University of Oxford.

Spanos, Aris (2006), Econometrics in retrospect and prospect, *in* T. C.Mills and K.Patterson, eds, 'Palgrave Handbook of Econometrics', Vol. 1, Palgrave Macmillan.

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.

Stephens, M. A. (1974), 'EDF statistics for goodness of fit and some comparisons', *Journal of the American Statistical Association* **69**(347), 730–737.

Stock, James H. and Mark W. Watson (2002), Has the business cycle changed and why?, *in* 'NBER Macroeconomics Annual 2002', Vol. 17, MIT Press.

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

Weissman, Ishay (1978), 'Estimation of parameters and larger quantiles based on the k largest observations', *Journal of the American Statistical Association* **73**(364), 812–815.

Zeileis, Achim, Friedrich Leisch, Kurt Hornik, Christian Kleiber and Bruce Hansen (2013), *Package strucchange*.
URL: *http://cran.r-project.org/web/packages/strucchange/strucchange.pdf*