Properties of Tests for Mis-Specification in Non-Stationary Autoregressions

D.Phil thesis

by

Jouni S. Sohkanen
Christ Church and University of Oxford



Trinity Term 2011
Hilary Term 2012
(revised)

to my grandparents

The barrenness of methodological conclusions is often a fitting complement to the weariness entailed by the process of reaching them.

Harrod (1938, p.383)

Acknowledgements

Thanks first to Bent Nielsen who has supervised this work with endless patience. I alone remain responsible for any mistakes in the work. Thanks also to the examiners Jurgen Doornik (internal) and James Davidson (external) for their careful and critical examination of the work.

Regressions are run and graphics produced in OxMetrics (PcGive), Monte Carlo simulations are coded in Ox, and LaTeXis used for typesetting. This work was supported by the Economic and Social Research Council [PTA-031-2006-00174]. I thank Cambridge University Press for their permission to include a version of an article published by them as one of the chapters (automatically granted under their transfer of copyright agreement).

In addition to the gratitude I owe to my supervisor and the D.Phil examiners, I am grateful (in alphabetical order) to Rafaella Giacomini, Andrew Harvey, Søren Johansen, Kasper Lund-Jensen, Diaa Noureldin, Cavit Pakel, Gary Phillips, Anders Rahbek, Andrew Whitby, two anonymous referees of Econometric Theory, and the two anonymous examiners of my M.Phil thesis for helpful comments that have led to improvements in the work. Thanks to the Department of Economics, Nuffield College, and Christ Church for support that has taken many forms, and in particular thanks to my college advisors John Thanassoulis and Jacinta Paula de Campos Vieira Pires for keeping an eye on my progress.

Thanks also to my non-econometrician friends in Oxford, particularly those based at Christ Church, for contributing to a friendly social environment, and to the brilliant executive head chef for feeding us all.

My family and extended family in Finland deserve special thanks for their exceedingly generous moral and financial support, and for remaining interested in my project for such a long time.

Word Count

I calculate the total number of words by multiplying the number of words on a representative page (312) by the total number of pages (164). The total (51168) represents c.51% of the maximum allowed number of words.

Use of language

English is not my first language and I apologise for the inevitable clumsiness of communication. I wish to declare three choices consciously made from the outset for the sake of consistency. First, as the use of English necessitates, against my wishes, a choice between third person singular feminine and masculine pronouns, the use of these pronouns is avoided as far as possible. When unavoidable, however, the masculine forms are used and are to be understood as generic. Second, words suffixed '-ize' are suffixed '-ize' instead of '-ise', regardless of whether the body of the word originates in Latin or Greek. This is in accordance with the policy adopted by the Oxford English Dictionary, and I refer the reader to the justification they use - one that I personally find compelling. Third, the editorial 'we' is used from now on.

Contents

1 Introduction **1**

1 Conceptual background 3

1.1 The econometric model 3

1.2 Model mis-specification 6

1.3 Structural change 8

2 Outline and literature review 11

2.1 The CUSUM of squares tests 12

2.2 CUSQ in autoregressions 15

2.3 Finite sample correction 21

2.4 Joint testing of mis-specification and economic hypotheses in autoregressions 24

3 Recursive graphics 26

3.1 Example: The Great Moderation 29

3.2 Example: The Dotcom Bubble 32

4 Mathematical framework 34

2 Asymptotic behaviour of CUSQ **37**

1 Introduction 37

2 The test statistics 38

3 Model and assumptions 40
4 Asymptotic analysis of $CUSQ^{OLS}$ 43
5 Asymptotic behaviour of the $CUSQ^{REC}$-test 45
6 Simulation study 46
7 Conclusion 47
A Appendix 48
A.1 The case of least squares residuals 48
A.2 Consistency of $\hat{\varphi}_t$ 50
A.3 The case of recursive residuals 52

3 CUSQ in explosive processes **59**
1 Introduction 59
2 Model and Assumptions 60
3 Nuisance terms 62
A Appendix 64

4 Implementation of CUSQ **71**
1 Introduction 71
2 The problem and some existing solutions 73
3 Monte Carlo evidence on the size distortion 79
4 Suggested finite sample correction 82
4.1 Response surface regressions for the AR(1) 83
5 Finite sample correction 84
5.1 Implementation in autoregressions 84
5.2 Implementation in ADLs 85
6 Conclusion 87
A Appendix 88

5 Joint testing **97**

1 Introduction 97

2 Model and assumptions 99

3 Test statistics 103

3.1 Inference on the autoregressive parameters 103

3.2 Unit root test 107

3.3 Test for lag length 108

3.4 Test for variance constancy 109

3.5 Test for normality 110

4 Main results 110

4.1 Behaviour of the tests under the joint null hypothesis . . . 113

5 Simulation study 115

6 Conclusion 117

A Appendix 119

A.1 Auxiliary results on the mis-specification tests 120

A.2 Joint behaviour of the mis-specification tests 120

6 Conclusion **135**

Chapter 1

Introductory remarks

This thesis is a contribution to the theory of statistical mis-specification testing in econometrics. A collection of four research papers written during course of the author's doctoral studies, a version of the first one has been published (Nielsen and Sohkanen 2011) and a version of the third has been submitted to a peer reviewed journal.

We are interested in the stochastic properties, individual and joint, of mis-specification testing when the data are generated by an autoregressive process. Good mis-specification tests are invariant to the dynamic properties of the process, summarized by the location of the roots of its characteristic polynomial. This makes it unnecessary to locate the roots, in other words, do inference, prior to mis-specification testing, which is important as mis-specification testing is used to validate inference. If the individual tests are mutually independent, then the overall significance level of the statistical procedure can be controlled. The ultimate goal is therefore to produce a result that gives the joint distribution of an entire battery of mis-specification tests and inference procedures. Achieving this requires work on individual test procedures and on the stochastic interactions

between them.

In existing literature, several tests are found to be valid to different degrees, most in the stationary and integrated cases, some in regular explosive, some even in singular explosive cases. White's test for heteroscedasticity, for example, is valid in stationary autoregressions and in some unit root specifications (depending on the properties of the innovations), while the Jarque-Bera test for normality and the 1-step Chow test may be applied even when the process is explosive; see e.g. Caceres (2007) and Nielsen and Whitby (2011). Lag length tests based on the likelihood ratio principle have been shown to be valid even in the singular explosive case (Nielsen 2008). In Chapter 2 of this thesis, we derive the asymptotic distribution for two types of CUSUM of squares test, one implemented with standardized 1-step-ahead ordinary least squares prediction errors as used in the original version of this test introduced by Brown, Durbin, and Evans (1975) and another implemented with ordinary least squares residuals. We find that the latter is valid in all but the singular explosive case, but that the former is only valid in the regular explosive case when all the roots are located in the explosive region of the parameter space. That is, if one root is in the non-explosive region, we require all the rest of the roots to be either on the unit circle or in the stationary region. In Chapter 3, we show that a nuisance term arises in the mixed case. In Chapter 4, we derive numerically a finite sample correction to render the tests applicable in modelling. §3 of this chapter contains two examples of applications.

Existing literature on the joint testing problem is not very extensive and relies largely on the classical framework of fixed regressors; see for example Savin and White (1978) and Bera and Jarque (1982), and for simulation results highlighting the problem, McGuirk, Driscoll, and Alwang (1993). In the autoregressive framework of Chapter 5, we consider inference on the parameters associated with

the stationary part of the process, together with tests for a unit root, lag length, residual variance constancy, and normality of the regression innovations. In characterizing the joint distribution of these tests, we rely on asymptotic theory, and show independence in the limit. A simulation experiment suggests that finite sample correlations between some of the tests are statistically significant but small. Asymptotically, then, calibration of the overall significance level of the test procedure is feasible.

In this introductory chapter, we outline the main results of the thesis, and provide some context for them. It is necessary to start with some definitions to avoid ambiguity later on. These are given in a purposefully non-technical manner. For a measure theoretic treatment of the modelling exercise, please see White (1994), and for a comprehensive coverage of the linear regression model, Davidson (2000). Readers familiar with standard econometric terminology can start reading from §2, where we provide a summary of the work.

1 Conceptual background

1.1 The econometric model

> *. . .progress [in macroeconomics] is unlikely as long as macroeconomists require the armor of a stochastic pseudo-world before doing battle with evidence from the real one.*
>
> Summers (1991, p.146)

> *. . .the purpose of constructing a statistical model is not to replace serious economic models with arbitrary statistical descriptions, but rather to provide a framework in which one can compare the economic theories with reality, as measured by the data series.*
>
> Johansen (1995, p.5)

This work, if not premised on the assumption that the statement of Prof. Summers' quoted above is questionable, at least implicitly relies on that assumption for its relevance. This thesis touches on elements of the 'stochastic pseudo-world', or 'statistical model', viewed here less in the sense of Prof. Summers as armour against, and more in the sense of Prof. Johansen as tool for testing theories about, the reality. Arguably, the statistical model constitutes the only well understood formal connection between the observed reality, as measured by data series, and economic theory, as written about in textbooks. It is necessary to appreciate why such a connection is required, how it works when it does, and what happens when it fails.

The *many and varied* problems of econometrics, to use Hendry (1995, §1.2)'s words, can almost without fail be traced back to the lack of causal control of the observed phenomena by the researcher, an unfortunate disadvantage shared by all purely observational disciplines. The central question facing the econometrician is how to say anything meaningful, other than purely descriptive, about the measured reality (data series) without being able to invoke causal interference, and isolating phenomena of interest by making use of such interference. Far from being under 'a scientific illusion', to borrow the title of Summers (1991), the econometrician recognizes that the extent to which the methodology of his subject can mirror that of the successful experimental sciences is limited.

In the absence of causal control, it is never going to be unproblematic to assume some variable fixed or even predetermined in a vastly complex and continually evolving system. Recognition of this fundamental problem leads naturally to a methodology that starts with a model characterizing the joint behaviour of all variables in the system. In an ideal situation, the modeller would succeed in capturing the entire extent of stochastic variation in the data. This would

reveal the probability law governing the observed variation in it. No assumptions involving external causal control would have to be made for valid conclusions to be reached. If sub-entities stochastically invariant to the rest were found, these could then be analyzed in isolation.

This *probability approach to econometrics* was pioneered by the Norwegian statistician Trygve Haavelmo (Haavelmo 1944). Arguably it was the first contribution to the econometric literature to explicitly acknowledge the importance of interaction between theory and data in the modelling process (although there are different interpretations about the views of the Cowles Commission on the subject, see e.g. Hoover (2006)), giving the search for stochastic patterns in the data a central role. Today, this data-driven approach to modelling is most evident in the general-to-specific approach of the LSE tradition particularly strongly associated with the work of Prof. Sir David Hendry and co-authors (for a recent survey, see Hoover (2006) and Hendry (2009)), and the probabilistic reduction approach of Spanos (1986); see also Spanos (2006a), Spanos (2006b).

Rather than attacking the data with the intention of quantifying a theoretical relationship chosen on a priori grounds, the data-driven econometrician, viewing the data as a realization of a stochastic process, sets about searching for an approximation to those features of the data generation process that are of interest to him. Context determines which features are viewed as important.

The apparatus supporting the modelling exercise, the econometric model, consists of an economic model embedded within a statistical model. The embedding is done in a way that ensures relevant statements result from inputs of raw data. In particular, the econometric model can be described as the estimable form of the data-admissibly and theory-conformably restricted statistical model. The statistical model is itself a combination of a probability model and a sampling

rule: a collection of probability measures describing the set of possible data generation processes, in the parametric case indexed by a real-valued vector, and an assumption about the generation of the sample. The latter can be thought of in terms of the extent of departure from independent sampling.

A caricature of the modelling exercise then proceeds by putting this apparatus into use. (1) Data are chosen. (2) Statistical assumptions are made. (3) A stochastic structure within the class defined by the statistical assumptions is chosen of which the data are a representative realization. This step is often known as estimation. (4) Mis-specification testing is conducted to ensure data-admissibility of assumptions. (5) Economic hypotheses are tested as restrictions defined within the chosen model. (6) The final output, the estimated form of the econometric model, sometimes known as the empirical model, is reported and interpreted in light of the economic model.

Crucial in this exercise is the use of a correctly specified statistical model. This allows the economist to determine whether his theory is consistent with the observed stochastic variation in the data series. The hypothesis of adequate specification is testable, and it is this step (4) with which this thesis is mainly concerned.

1.2 Model mis-specification

> *In applying mathematics to subjects such as physics or statistics we make tentative assumptions about the real world which we know are false but which we believe may be useful nonetheless.*
>
> Box (1976)

> *[mis-specification] testing is ... of central importance to all model selection approaches because the presence of mis-specifications jeopardizes the ground for model selection, whichever criterion one uses.*

Mayo and Spanos (2004)

The 'all models are false but some are useful' -argument against mis-specification testing feeds off a confusion of statistical and substantive questions. A substantively adequate econometric model explains the systematic aspects of reality such that there is no systematic departure between theory and the data generating process (Spanos 2009). In mis-specification testing, we are concerned with ensuring statistical adequacy, consistency of model assumptions with the observed data. While establishing statistical adequacy alone does not get us very far (economic theory, algorithms of model discovery, e.g. Doornik (2009), and sheer ingenuity are needed as well), use of a statistically inadequate model as a starting point to model selection does not constitute a reliable method, as statistical adequacy is required to validate inference. A `t`-test, to take an example, has the Student's `t` distribution under certain statistical assumptions. If these are violated, the econometrician loses control of the Type I error probability, and the test loses its meaning.

For terminological clarity, we distinguish between the concepts of mis-specification testing and inference. In doing inference, we are concerned with ascertaining the plausibility of a particular hypothesis restricting the statistical model against a known alternative; the statistical model we are testing within is correctly specified both under the null and the alternative. In mis-specification testing, we use a statistical model (usually the residuals thereof) fitted under a given set of assumptions to test whether the assumptions and the data are coherent; see Mizon (1977, p. 1228). One may put this in terms of a difference between testing 'within' and testing 'without' a model, or between the Neyman-Pearson type hypothesis testing and the Fisher type procedures; see e.g. Spanos (2000). Crucially, the alternative hypothesis is not necessarily a specification of a unique

statistical model. The purpose of mis-specification testing is simply to try to reject the current model. The extent to which this model is placed in jeopardy as a result is referred to as the severity of the test.

To summarize, (a) correct specification is itself a hypothesis that has to be tested using statistical tools, and (b) these tests often have to be based on the same data as the final econometric model. This brings to the fore issues of joint testing that we address in Chapter 5 of this thesis.

1.3 Structural change

> *. . .suppose that we have a set of observations that all confirm the statements that are permissible within our model. Then these statements become facts interpreted in the light of our theoretical model. . .But will the model also hold for future observations? We cannot give any a priori reason for such a supposition. We can only say that, according to a vast record of actual experiences, it seems to have been fruitful to believe in the possibility of such empirical deductions.*
>
> Haavelmo (1944, p.10)

> *Like Bellerophon, the econometrician can ride the Pegasus of skillful and careful modeling to overcome the challenge provided by the chimera of structural change, and in the process greatly enhance his reputation and rewards.*
>
> Anderson and Mizon (1989)

Definitions of structure can be elusive. We adopt the standard Cowles Commission version (see e.g. Haavelmo (1944), and also Christ (1994)). A stochastic equation describing a relation, say an economic one, in a system of such equations is called a structural relation if the equation is known, and if the equation is unaffected by changes in the parameters or functional form of another equation in the system. When all equations in the system have this property, the system defines a structure. A structural relation is autonomous in the sense that it is

invariant to a class of changes. Structural change refers to the breaking down of such a relation. This characterization of structure may be related to the concept of a parametric probability model: a collection of probability measures indexed by a parameter vector. Indeed, Sargan and Desai (1988, p. 27) define structure as the probability model which has been assigned parameter values. This is similar to the definition of empirical model adopted above.

Therefore, in our framework, structural change and model mis-specification are closely related concepts. A parametric probability model is correctly specified if there is a point in the parameter space such that when the probability measure corresponding to that parameter vector is chosen, the resulting structure is the one that generated the data (the local data generation process, or LDGP - see Hendry (2009)). The vector of parameters is said to be identifiable if for each different point in the parameter space there exists a different equivalence class of probability measures, i.e. one that does not agree with the others except on null sets. For a parametric probability model uniquely characterized by its parameters, i.e. one where identification does not fail, structural change therefore implies varying parameters. In the time series context, this entails non-stationarity. We shall use Hendry's terminology (see e.g. Hendry (1995, p.31) or Hendry (1996)) in distinguishing between parameters and coefficients. The latter may vary over time in a constant model. For example, a regression model with a time varying slope can be given a constant parametrization when the slope itself is modelled as following a random process. Then, the model has a constant underlying parametrization but the varying coefficient is retained; see e.g. Hendry (1996).

Sometimes, it may be helpful to distinguish between mis-specification and structural change when talking about parameter constancy. Structural change,

in the time series context, implies the existence of more than one structure in the time period under scrutiny. Take for example some exogenously given regime shifts, introduced into the model to be dummied out in estimation. It is clear that, as Hackl (1980, p. 2) notes, some theoretical knowledge is needed to identify such structural change, and we do not address this problem here. This serves merely as a contrast to the wider mis-specification issues: parameter non-constancy from failure of statistical assumptions is not dependent upon the existence and exogenous break-down of a structure; for simulation evidence highlighting this problem, see Alston and Chalfant (1991). In this thesis, however, we only consider models where parameters associated with the stochastic components of the processes under study are constant ex hypothesi. Therefore non-constancy, whether arising from structural change in the sense defined here or from invalid statistical assumptions, becomes a mis-specification issue in its own right, and tests designed to detect it may be referred to as mis-specification/diagnostic tests. For a comprehensive review of testing for structural breaks, see e.g. Perron (2006).

When applying such tests, it must be borne in mind that the constant parameters hypothesis is set in the context of a model where other assumptions are made. As noted e.g. by Hackl and Westlund (1991, p.3), for the purposes of applying diagnostic and structural change tests and interpreting test outcomes, it is crucial to appreciate the robustness of the relevant test statistics to the choice of the assumptions. The asymptotic theory developed for some test statistics may simply be inappropriate for the type of model required to cope with the modelling challenge at hand. This has been the case, for example, for the CUSUM of squares tests, and many others. One of the contributions of this thesis is to provide results about the robustness of the CUSUM of squares test, and thus validate its application under macroeconometrically relevant assumptions. The finite sample

behaviour of the tests is another question of interest. Kiviet (1986) shows with a simulation experiment that various tests designed to detect serial correlation and predictive failure are defective in small samples if the model contains a lagged dependent variable. Another contribution of this thesis is to provide guidance for the use of the CUSUM of squares test in such models.

Sufficient terminological background has now been provided, and we will proceed to give an outline of what follows.

2 Literature review and outline of the work

Most of this thesis concerns the implementation of the CUSUM of squares test. Owing to its origins in the classical regression setting, it is sometimes implemented in a form that has little or no motivation in the time series context we are considering, but that gives rise to complications. While Chapter 2 is devoted to derivation of asymptotic theory for the test in quite a general linear dynamic model class, Chapter 3 deals with the aforementioned complications. The practical implementation of the test is the subject of Chapter 4, where a finite sample correction is derived, using numerical methods, to render the test implementable into software; see §3 for examples. A brief literature review is given in §2.1. In §2.2 we discuss the contributions of Chapters 2 and 3, and §2.3 contains an outline of Chapter 4.

Understanding the behaviour of the CUSUM of squares test, or indeed of any mis-specification test, in isolation is not sufficient in practice. In modelling, such tests are usually implemented as part of a battery of diagnostics prior to inference. In Chapter 5, we consider the joint behaviour of a battery of some mis-specification tests and the joint behaviour of this battery with the inference

procedure used to test hypotheses within the chosen econometric model. In §2.4, we review the literature relating to this problem, and discuss the contribution made in Chapter 5.

The scope of application of the main results of Chapter 5 is, arguably, confined to the motivating problem. By contrast, the results of Chapters 2 and 4 are readily applicable in empirical modelling. In order to emphasize this, we look at two simple empirical illustrations in §3. We do not intend to make, or indeed make, any substantive contributions to the empirical literature, but hope that the discussion makes the work seem more relevant to practitioners.

A note on the mathematical techniques used in the thesis follows in §4.

2.1 The CUSUM of squares tests

The CUSUM of squares (CUSQ) tests are mis-specification tests, in the sense of Mizon (1977), and have non-trivial asymptotic local power against changes in the variance of the residuals they are implemented with (Ploberger 1989, Ploberger and Krämer 1990, Hansen 1991). They belong to the category of *general* (Ramsey 1974) or *exploratory* (Dufour 1982) tests, tests against a set of rather diffuse alternatives. They use various types of information in the model residuals (other than sign), such as zero mean, homoscedascitic, and autocorrelation, to detect failure of the modelling assumptions.

Exact distribution theory

In the regression context, the CUSUM of squares procedure was suggested by Brown, Durbin, and Evans (1975), who considered the linear regression with

exogeneous regressors

$$y_t = x_t'\beta + \varepsilon_t, \quad \varepsilon_t \sim \mathsf{N}\left(0, \sigma^2\right), \quad t = 1, \ldots, T,$$

and after recursively estimating the model by ordinary least squares, computed standardized one-step-ahead OLS prediction errors

$$\tilde{\varepsilon}_t = \left(1 + x_t' \left(X_{t-1}'X_{t-1}\right)^{-1} x_t\right)^{-1/2} \left(y_t - x_t'\hat{\beta}_{t-1}\right), \quad t = M+1, \ldots T$$

where M is the number of regressors, $\hat{\beta}_t = \left(X_t'X_t\right)^{-1} X_t'Y_t$ with $X_t = \left(x_1', \ldots, x_t'\right)'$ and $Y_t = \left(y_1, \ldots, y_t\right)'$ with $X_t'X_t$ non-singular. With fixed regressors, $\mathsf{Var}(y_t - x_t'\hat{\beta}_{t-1}) = 1 + x_t' \left(X_{t-1}'X_{t-1}\right)^{-1} x_t$, and $\mathsf{Cov}(\tilde{\varepsilon}_t, \tilde{\varepsilon}_{t-s}) = 0$ for all $s \neq t$, so the standardization produces random variables with the same probabilistic properties as the fundamental innovations. With these ingredients, they defined the CUSQ plot by

$$s_{M,t,T} \overset{\mathsf{def}}{=} (\sum_{j=M+1}^{t} \tilde{\varepsilon}_j^2 + \sum_{j=t+1}^{T} \tilde{\varepsilon}_j^2)^{-1} (\sum_{j=M+1}^{t} \tilde{\varepsilon}_j^2)$$

where $\sigma^{-1} \sum_{j=M+1}^{t} \tilde{\varepsilon}_j^2 \sim \chi^2\left(t - M\right)$ and $\sigma^{-1} \sum_{j=t+1}^{T} \tilde{\varepsilon}_j^2 \sim \chi^2\left(T - t\right)$ such that (see e.g. Rao (1973, p. 167, eq. 3a.2.14)) $s_{M,t,T}$ follows the beta distribution with mean and variance respectively of $\left(t - M\right) / \left(T - M\right)$ and $\left(t - M\right)\left(T - t\right) / 2\left(T - M\right)^2 \left(T - M - 1\right)$, i.e. with density

$$\frac{\Gamma\left(\frac{T-M}{2}\right)}{\Gamma\left(\frac{t-M}{2}\right)\Gamma\left(\frac{T-t}{2}\right)} \left((\sum_{s=M+1}^{T} \tilde{\varepsilon}_s^2)^{-1} (\sum_{s=M+1}^{t} \tilde{\varepsilon}_s^2)\right)^{\frac{t-M}{2}-1} \times \left(1 - (\sum_{s=M+1}^{T} \tilde{\varepsilon}_s^2)^{-1} (\sum_{s=M+1}^{t} \tilde{\varepsilon}_s^2)\right)^{\frac{T-t}{2}-1}$$

In deriving an operational test procedure, the authors resorted to approximation of the maximal null-allowable fluctuations of the CUSQ plot around its mean by Pyke (1959)'s modified Kolmogorov-Smirnov statistic, with critical values tablulated by Durbin (1969). It is also possible to draw the CUSQ plot with non-standardized ordinary least squares residuals, although an exact distribution is not available in such a case. Using theory developed by Durbin and Watson (1950), McCabe and Harrison (1980) derived beta-distributed upper and lower bounds for the plot to generate an operational test. While the Brown, Durbin, and Evans (1975) version measures the prediction performance of the model, the OLS version can be interpreted as measuring in-sample residual variance fluctuations, and is as such directly related to the fluctuations tests considered by Ploberger, Krämer, and Kontrus (1989) and Kuan and Hornik (1995), albeit pertaining to fluctuation in the error variance parameter, rather than in the regression coefficients. Rodrigues and Rubia (2011) note that the test may be viewed as an approximate likelihood ratio.

Asymptotic distribution theory

The CUSQ plot, embedded in the appropriate function space, can be shown to have a well defined limit distribution under various sets of assumptions. For the purposes of asymptotic theory, the CUSUM of squares test is taken to mean a normalized version of the supremum of the demeaned plot's absolute value. Previous work on the CUSQ test asymptotics can be delineated along various dimensions, one being the persistence allowed in the data generating process, and another the strength of moment assumptions imposed on the innovations. As usual, a tradeoff seems to be operating here.

The earliest asymptotic result concerning a CUSUM of squares type quan-

tity is presented in Bartlett (1955); see also Durbin (1967). Using an orthogonal transformation on $X_t \sim \mathsf{N}(\mu, \sigma^2)$ to normal zero mean constant variance variables, he derives what is essentially the CUSUM of squares test, and computes the asymptotic probability of its crossing a boundary of the form $c_K\sqrt{2/T}$, where c_K is a critical value, the appropriate quantile of the so called Kolmogorov distribution, the supremum of a standard Brownian bridge. A more recent proof in the independent identically distributed but not necessarily normal case is given in Inclán and Tiao (1994) under an assumption guaranteeing the existence of fourth moments.

Loretan and Phillips (1994) work with financial data (see also Pagan and Schwert (1990)) and consider the asymptotic behaviour of the test when fourth moments do not exist. Under their assumptions, the asymptotic distribution is not a tied down Brownian motion, but a function of a stable Lévy bridge. The usual Brownian bridge asymptotics are also derived by Ploberger and Krämer (1986) who base the test on residuals from a stationary autoregression with martingale difference sequence innovations, Lee, Na, and Na (2003) who allow for the possibility of a unit root, and Deng and Perron (2008a) who work under mixing conditions allowing weak dependence in the regressors with the possibility of ARCH effects in the errors . A version based on transformed (first differenced) residuals in a model with a unit root is considered by Lu, Maekawa, and Lee (2008).

2.2 Behaviour of CUSQ in autoregressions

Judged against the standards of the existing literature, the set-up of Chapter 2 is arguably quite relevant from the empirical macroeconometric perspective. As usual, the cost of generality in the parameter space comes in the form of a

restriction on the innovation process. In contrast to Lee and Park (2001) and Deng and Perron (2008a), who allow for conditional heteroscedasticity, we rule this out with a fourth order conditional moment existence condition, to be able to deal with the case of explosive roots in the data generation process. We note that the mixing assumptions made by Deng and Perron (2008a) do imply assumptions about the existence of unconditional moments as high as the eighth. Our innovation specification resembles that of Ploberger and Krämer (1986), but we are also able to avoid imposing strict restrictions on the location of the characteristic roots of the autoregression that is assumed to generate the data, in sharp contrast to existing literature. Ruling out ARCH innovations necessitates the application of ARCH tests in conjunction with the CUSQ tests. This applies to most diagnostic tests, however, which are normally implemented in a battery; see §2.4 for some discussion of this practice and its implications.

The class of data generating processes is restricted to the possibly unstable vector autoregressive linear class. The model is formed either by marginalization, justified by Granger non-causality of the relevant variables, or conditioning, justified by the appropriate exogeneity concepts (see Engle, Hendry, and Richard (1983)). In either case, the properties inherited from the DGP imply the possibility of stochastic and deterministic trends (i.e. mean non-stationarity), while it is still possible to maintain the hypothesis that the model residuals are well behaved. To illustrate the allowable model class, consider the class of allowable data generating processes for X_t, partitioned into $(Y_t, Z_t)'$. These are p-dimensional

vector autoregressions of order k

$$\begin{pmatrix} Y_t \\ Z_t \end{pmatrix} = \sum_{j=1}^{k} \begin{pmatrix} A_{y,j} \\ A_{z,j} \end{pmatrix} \begin{pmatrix} Y_{t-j} \\ Z_{t-j} \end{pmatrix} + \begin{pmatrix} \xi_{y,t} \\ \xi_{z,t} \end{pmatrix} \tag{2.1}$$

$$\begin{pmatrix} \xi_{y,t} \\ \xi_{z,t} \end{pmatrix} \sim f\left(0, \begin{pmatrix} \Omega_{yy} & \Omega_{yz} \\ \Omega_{zy} & \Omega_{zz} \end{pmatrix}\right) \tag{2.2}$$

where f is a distribution and the variance of the innovation process is positive definite. Apart from the imposed linearity, this class has considerable generality. A vector autoregression can be viewed simply as a summary of covariances in the data (see e.g. Juselius (2006, Ch. 3)), and its order can be selected using well established statistical procedures that are robust to unit roots as well as both regular and singular explosive roots (Nielsen 2006, Nielsen 2008). The autoregressive distributed lag model is obtained by pre-multiplying the VAR(p, k) equation with $(1, -\rho')$, to obtain

$$(1, -\rho') \begin{pmatrix} Y_t \\ Z_t \end{pmatrix} = \sum_{j=1}^{k} (1, -\rho') \begin{pmatrix} A_{y,j} \\ A_{z,j} \end{pmatrix} \begin{pmatrix} Y_{t-j} \\ Z_{t-j} \end{pmatrix} + (1, -\rho') \begin{pmatrix} \xi_{y,t} \\ \xi_{z,t} \end{pmatrix}$$

rewritten as

$$Y_t = \rho' Z_t + \sum_{j=1}^{k} \alpha_j Y_{t-j} + \sum_{j=1}^{k} \beta_j' Z_{t-j} + \varepsilon_t, \qquad \varepsilon_t \sim f\left(0, \tilde{\Omega}\right) \tag{2.3}$$

Since $\rho' = \Omega_{yz}\Omega_{zz}^{-1}$, the residual variance $\tilde{\Omega}$ in the last expression may be written as $\Omega_{yy} - \Omega_{yz}\Omega_{zz}^{-1}\Omega_{zy}$. The other parameters of the autoregressive distributed lag (2.3) and the vector autoregression (2.1) can also be linked via limits of the least squares estimators computed from (2.3); see Chapter 2. Hendry (1995,

Ch. 7) characterizes nine special cases of (2.3) arising out of restrictions on its parameters, showing that the model is rich enough to capture interesting economic phenomena. In particular, this choice of model class allows testing for stability in the popular error correction models, as considered e.g. by Hansen (1992a).

One advantage gained by concentrating on second moments is that questions about the location of characteristic roots can be put to one side while checking diagnostics. The reason is that the asymptotic distribution of the CUSQ test does not vary with the roots (see Chapter 2), and there is evidence that its finite sample distribution exhibits near invariance to the roots (see Chapter 4). This finding is in contrast to existing theory on the so called Quandt-Andrews procedures, where stochastic trends in the data generation process result in different asymptotics from those arising in the stationary case (Hansen 1992b). Separation of the questions of model stability and location of characteristic roots is the most important contribution of this work. To the best knowledge of the author, such a generalization of the CUSQ test has not previously been undertaken.

Another necessary choice for the practitioner is the quantity of which cumulative sums of squares are taken for the purposes of constructing the CUSQ test. For example, in detecting a decline in United States output volatility, McConnell and Perez-Quiros (2000) use deviations from means; see §3. These can of course be interpreted as the OLS residuals from a regression on a constant. In our theory, the test may be based on OLS residuals, where regressors can be non-stationary, or standardized one-step-ahead forecast errors, where again regressors in the model from which predictions are computed may be non-stationary. We do not allow for multiple-steps-ahead forecast errors in the style of Dufour (1982).

The one-step-ahead standardized prediction errors are called recursive residuals by Brown, Durbin, and Evans (1975), who implemented the CUSQ test with

these residuals. They have also been used by Hedayat and Robson (1970) in a test for homoscedasticity against heteroscedasticity that is monotonic in the mean, and in the context of more general functional mis-specification testing by Harvey and Collier (1977). There is also a link between the Chow-type predictive failure tests and the recursive residuals (see e.g. Harvey (1976)): the CUSQ test implemented with recursive residuals essentially tests the model's performance as a prediction instrument, exploiting the fact that predictive failure and model non-constancy are inextricably linked. A model with non-constant parameters is unlikely to forecast well. However, when the CUSQ test is implemented with OLS residuals, the even clearer parameter constancy test interpretation referred to above arises: the test statistic is algebraically equivalent to a statistic giving the largest deviation of the recursively estimated residual variances from the full sample equivalent.

The motivation for implementing the test with the recursive residuals has to do with the relatively simple and convenient probabilistic properties in the classical fixed regressor normal innovations case, reviewed in §2.1. The studentized residuals are independent standard normal under the null hypothesis of constant coefficients, the squares therefore having the χ^2 distribution and the ratios of sums the Beta-distribution. Only a one-sided exact test may be constructed with the available distribution theory, so it is not entirely clear what the advantage of using the forecast errors is even in the classical case, given the work of McCabe and Harrison (1980).

The stochastic regressor case is one where the use of recursive residuals certainly cannot be advocated on the basis of simple probabilistic properties. Although the asymptotic argument works for a purely non-explosive or purely explosive process, as seen in Chapter 2, the interaction of stationary and explosive

components in the data generation process causes non-vanishing nuisance terms to appear when the CUSQ test is implemented with recursive residuals. This is in contrast to the almost complete generality in the parameter space achieved with the use of an OLS residuals-based test (complete apart from the singular explosive case - see Nielsen (2008)). In Chapter 3, we derive an explicit expression for the nuisance term and show that it is non-vanishing. We show further in Chapter 4 that the observed sizes of the two procedures do not differ very much in autoregressive models. The literature suggests that the only important difference has to do with the behaviour of the tests under various alternative hypotheses. For example, Deng and Perron (2008b) find that when a large break occurs in the regression coefficients, the test based on the standardized one-step-ahead forecast errors outperforms the test based on ordinary least squares residuals.

Two further comments are in order. First, the theory developed in this work is applicable when the CUSUM of squares tests are used as one-time retrospective diagnostic tests. Monitoring procedures, used repeatedly as new data arrives, require their own asymptotic theory, where critical values are continually adjusted to avoid a situation where rejection probabilities approach unity; see Chu, Stinchcombe, and White (1996), Carsoule and Franses (2003), and specifically for a repeated CUSUM of squares type test, Lee and Park (2009). Second, while the scope of this work is strictly limited to the behaviour of the CUSUM of squares tests under the null hypothesis, some rudimentary understanding of their power properties can be invaluable as this informs the practitioner's interpretation of the results. In particular, we wish to emphasize that from a local asymptotic power perspective, the CUSUM of squares tests should be thought of as tests of the residual variance constancy only, and not as tests against changes in the autoregressive parameters; see Ploberger (1989), Ploberger and Krämer (1990),

Hansen (1991), and Xu (2008). This conclusion must, however, be qualified if the test is not implemented with residuals from a model with covariates. For instance, (Leybourne, Taylor, and Kim 2006) implement the test with deviations from means and test for a change from trend stationarity to difference stationarity. Another way to qualify the conclusion is to take a non-local view of power. It is of some interest to note that the CUSUM of squares tests does not suffer from non-monotonicity of power issues, unlike many related structural breaks tests; see Deng and Perron (2008b).

2.3 Correcting the finite sample bias of the CUSUM of squares test in autoregressions

Reliance on asymptotic theory of the type developed in Chapter 2 inevitably grows as assumptions defining the statistical model become less restrictive. In fact, moving away from the fixed regressor framework, the hope of finding exact finite sample distributions is all but abandoned in favour of asymptotic procedures. Nevertheless, it is the finite sample behaviour of the test that is of interest to the practitioner. Use of asymptotically valid critical values in sample sizes typical of macroeconomic applications may be undesirable. Edgerton and Wells (1994) note that the asymptotic representation of the test yields a significant error even for sample sizes bigger than 200. These authors address the finite sample discrepancy by extending Durbin (1969)'s tables up a sample size of 10000. As their results are only valid in the classical framework, this is not a satisfactory solution in our setting.

An attempt can be made to rectify this problem in a restrictive class of dynamic models (see e.g. Dufour and Kiviet (1995)). In the more general model class defined in Chapter 2, numerical methods are called for. Bootstrapping,

while in some sense a standard solution to problems of this kind (see for example Tanizaki (1995)), would be impractical given that in typical applications the test is just one among many sharing the same problem. Interestingly, the pilot Monte Carlo study reported in Chapter 2 indicates near invariance to the parameter space of the finite sample distribution of the test. This suggests that a simple correction, a function of T alone, may be available that could be applied directly to the asymptotically valid critical values. In Chapter 4, using a simulation study, we derive this correction for the AR(1) case, and then show that the same rule works to a good approximation in more general models. Importantly, the correction is implementable into software. Its use is illustrated in §3.2.

The methodology used in deriving the finite sample correction is parametric response surface analysis. For a clear review of this technique as well as of the Monte Carlo method in general, please see e.g. Hendry (1984), Davidson and MacKinnon (1993, Ch. 21), and Doornik (2006). Simply put, a Monte Carlo experiment is a computer program where one part generates the data, a random sample of length T, for some parameter values, and another part computes the statistics of interest from the random sample generated by the chosen DGP. This is done a number of times. We call this number the number of replications, denoted M. Sometimes, it is also known as the Monte Carlo sample size, to distinguish it from the econometric sample size, T. Output can take one of many forms. A naïve approach is to simply calculate what is known as the Monte Carlo mean, the simple average across the M computations of the statistic of interest. This is sufficient for our purposes, and with a sufficiently large M, gives us adequate Monte Carlo precision.

The Monte Carlo averages for a range of observed quantiles for both types of CUSUM of squares test are interpreted as a data set of the approximate finite

sample quantiles, each associated with an econometric sample size, T. A central limit argument suggests that the finite sample quantiles approach the asymptotic quantile in inverse square root powers of T. This allows us to relate the sample size T to a measure of divergence between the asymptotic and finite sample quantiles: we simply estimate by OLS a regression model with inverse square root powers of T placed in the design matrix. The measure of divergence is designed such that a constant term may be omitted. The results for the OLS and recursive residuals based CUSUM of squares tests are very similar, with the difference between them growing slightly as we move to higher and higher quantiles (we concentrate of course on the right tail, i.e. the tail relevant for the use of CUSUM of squares test in practice).

As the response surfaces are based on simulated finite sample quantiles from CUSUM of squares tests computed on residuals from the AR(1), it is necessary to assess the performance of these corrections in more general models. Idiosyncrasy of the results would diminish the practical value of the correction. Monte Carlo simulations suggest that the finite sample correction indeed applies approximately in higher order autoregressions with deterministic terms, as well as autoregressive distributed lag models, with no restrictions on the characteristic roots in the DGP, other than non-singularity in the explosive case; see Chapter 2 and Nielsen (2008). Some size distortions are observed in cases where the DGP has a positive unit root and a linear trend is being estimated. We do not venture far enough into the explosive region to cause numerical issues, but if necessary, these could be addressed with a procedure similar to that used by Nielsen and Reade (2007).

2.4 Joint testing of mis-specification and economic hypotheses in autoregressions

One or more of the statistical assumptions may fail to hold for the observed data. When only one does, we follow Bera and Jarque (1982) in calling the failure one-directional, otherwise multi-directional. Whether one-directional or multi-directional, the source of the failure may be difficult to locate. For example, normality tests may pick up functional form problems, and the CUSUM of squares tests fat tailed error distribution problems. In philosophy of science jargon, Duhemian problems stand in the way of any clear rules for re-specification. This is why mis-specification tests, such as the CUSUM of squares tests, are normally implemented as part of a battery.

Using an inappropriate battery may have consequences for the associated error probabilities and may lead to incorrect inference, as well as sub-optimal power properties. To borrow Bera and Jarque (1982)'s terminology, using a test designed as sensitive to one failure only, a one-directional test, when failure is multi-directional, is referred to as under-testing, with the opposite phenomenon called over-testing.

Existing literature dealing with this problem is not very extensive; some examples include Savin and White (1978) and Jarque and Bera (1980). The most comprehensive contribution, however, is made by Bera and Jarque (1982). These authors work with a linear regression with fixed regressors and normal innovations and then define a more general model with a larger parameter vector, allowing for autoregressive dependence as well as heteroscedasticity and non-normality in the innovations, in addition to more regressors and non-linearities in parameters. They then formulate the joint mis-specification test as an LM test on the larger parameter vector $\theta = (\theta_1, \theta_2)$, where θ_1 are nuisance parameters, and the restric-

tions imposed on θ_2 imply correct specification, i.e. the normal model with fixed regressors. This LM test turns out to have an additivity property, whereby the joint test can be written as the sum of the LM tests for each hypothesis separately, that is, the tests have no mutual dependence. Unfortunately, this result does not hold in the time series context.

Even assuming this problem away, i.e. once the appropriate battery is defined, we face a joint testing problem on two levels. First, we must decide how, if at all, to discount the individual tests for the possible covariance between them, and thus control the overall Type I error probability. Second, we must decide how, if at all, to discount subsequent hypotheses tests within the chosen model. If mis-specification tests and inference procedures are stochastically related, the decision of one may affect the decision of the other. If the individual significance levels are not adjusted, we may obtain significant results by default. The term 'data mining' is often associated with such practices (see e.g. Lovell (1983), Hoover and Perez (1999), Hoover and Perez (2000)). In classical statistics, Basu's theorem (Basu 1955) is invoked to show that so called ancillary statistics (such as functions of the model residuals) are independent of so called complete sufficient statistics (such as estimators). This theorem is applied in econometric models by Phillips and McCabe (1983), Phillips and McCabe (1988), and Phillips and McCabe (1989) or Spanos (2007). Again, such results do not hold when temporal dependence is allowed in the process being modelled, and questions about the nature of interaction between the tests remain open.

In Chapter 5, we provide a partial answer to these questions by showing that some commonly used mis-specification tests, such as tests for normality, variance constancy, unit roots, and lag length, are asymptotically stochastically independent of one other and independent of hypothesis tests on the parameters within

the chosen model. Thus, adjustments or discounts are not required asymptotically; we know the significance level of the joint procedure and can choose the appropriate significance level for the individual tests. This knowledge becomes particularly valuable when the number of tests conducted is large; see for example Krolzig and Hendry (2001). We illustrate this in Figure 2.1, where the number of tests, each conducted at the 5% significance level, is plotted on the horizontal axis against the significance level of the joint procedure on the vertical axis. The latter must be bounded above by the sum of the individual significance levels and from below by the individual significance levels. The solid line curve shows the implied level when the tests are independent; see Godfrey (1988, Ch. 1).

Our answer is only a partial one for two reasons. First, we do not even aim for a sufficient level of abstraction to be able to make general statements about misspecification testing and inference. Our results hold for the battery we consider, arguably one that is quite useful for practitioners. Second, the analytical results obtained are asymptotic. This is necessary for making any progress at all in the model class we consider. It is possible, however, to apply numerical techniques to assess the adequacy of the approximations, and we report on a pilot Monte Carlo study showing that the tests are approximately uncorrelated in finite samples also.

3 Implementation: recursive graphics

The asymptotic results developed in Chapter 2, together with the finite sample corrections suggested in Chapter 4, are readily applicable by the practitioner using graphical techniques as well as formal test procedures. Corollary 4.7 of Chapter 2 implies that with approximately $\beta\%$ confidence, the quantities of interest are

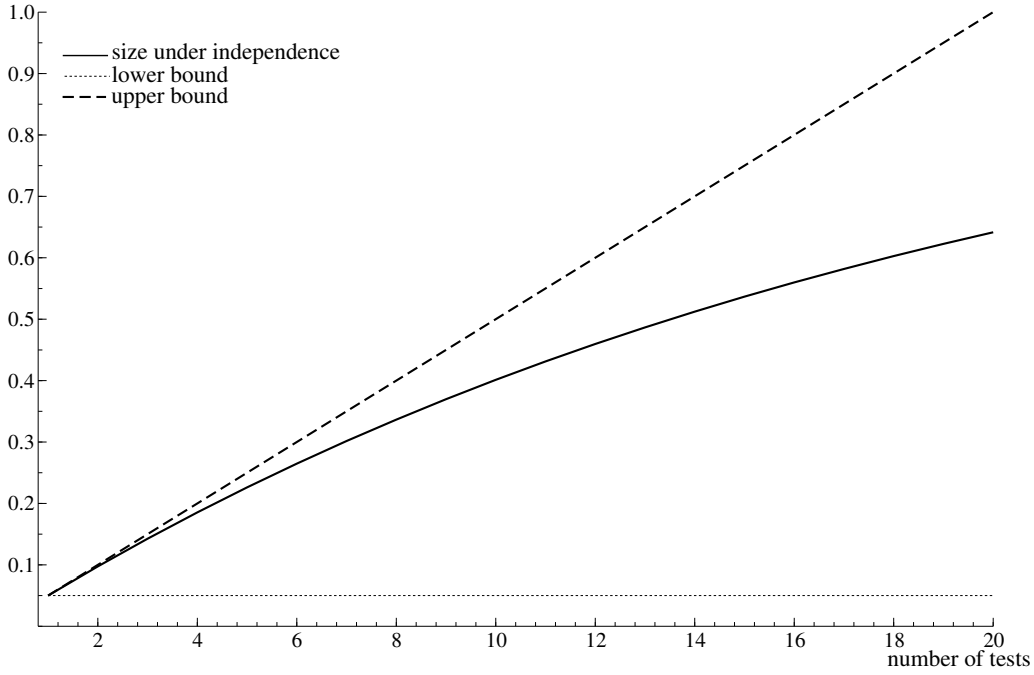


Figure 2.1: Implied joint significance level of the test procedure as a function of the number of tests, each conducted at 5%. The lower bound is given by the significance level of the individual test, and the upper bound by the simple sum of the significance levels. Under independence, the joint significance level is given by $1-(1-0.05)^n$, where n is the number of tests.

bounded as follows.

$$RSS_t \in \left(\hat{RSS}_t - c_\beta \sqrt{2t\hat{\sigma}_T^4}, \hat{RSS}_t + c_\beta \sqrt{2t\hat{\sigma}_T^4} \right) \text{ and}$$
$$\sigma^2 \in \left(\hat{\sigma}_t^2 - c_\beta \sqrt{\frac{2\hat{\sigma}_T^4}{t}}, \hat{\sigma}_t^2 + c_\beta \sqrt{\frac{2\hat{\sigma}_T^4}{t}} \right),$$

where $RSS_t = \sum_{s=1}^{t} \varepsilon_s^2$, $\hat{RSS}_t = \sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2$, and c_β is the appropriate quantile of the standard normal distribution. This allows us to draw point wise confidence bands for the recursive sums of squared residuals graph, currently implemented without such bands in PcGive; see Doornik and Hendry (2007a). The null hypothesis of variance constancy is rejected if a straight line cannot be drawn inside the bands in the case of the RSS graph, or equivalently, if a horizontal line cannot be drawn inside the bands in the case of the residual variance graph.

The validity of the asymptotic theory in the presence of stochastic and deterministic trends allows the practitioner to apply the graphical techniques for integrated, or even explosive data. We illustrate both cases using the examples below. It is not our intention to promote the use of such graphs in place of formal test procedures, to replace systematic approaches with *eyeball econometrics* (Hansen 1991) or to rely on *Sherlock Holmes inference* (Dhyrmes, Howrey, Hymans, Kmenta, Leamer, Quanot, Ramsey, Shapiro, and Zarnowitz 1972). The recursive graphics are likely to be more useful at the early stages of modelling, at the stages where different ideas are being tried out and the modeller has to decide whether to use inference procedures robust to volatility breaks. The graphs are to be viewed as entirely complementary to formal procedures, with one exception: the CUSUM squares plot is drawn with simultaneous bands, and is therefore equivalent to the formal procedure, as a decision is made depending on whether, at any point, the graph crosses one of the bands. With OLS residuals for example,

the α level test rejects if

$$\sup_t |(\sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2 / \sum_{s=1}^{T} \hat{\varepsilon}_{s,T}^2 - t/T)| > c_{K_\alpha}$$

where c_{K_α} is the appropriate quantile of the Kolmogorov distribution, 1.36 in case of a 95% test, for instance; see Schumacher (1984, Tab. 9). One could argue that looking at such a graph is unhelpful, as all the information contained in it can be represented by a binary variable. As long as the practitioner understands this, no harm will result.

Unlike Rodrigues and Rubia (2011), we do not comment on whether timing of the change can be established from the latter graph. It seems to us that finding a good estimator of the break date would require knowledge of the behaviour of the graph under different alternative hypotheses, something we are not concerned with. We refer to Westlund and Törnkvist (1989) for a discussion. Tests that are perhaps more commonly used in the literature for this purpose are the Chow (1960) tests, and the supremum version considered by Quandt (1960) for which asymptotic theory was derived by Andrews (1993); see also Andrews (2003), Hidalgo and Seo (2008). The 1-step Chow test (see Doornik and Hendry (2007a, Ch. 18)) has been shown to be applicable in the same model class we are considering; see Nielsen and Whitby (2011). We refer to Hansen (2001) for a discussion on the estimation of the break point using Quandt-Andrews type procedures.

3.1 Example: The Great Moderation

As an example of an integrated process, take a time series of US GDP from 1947:Q1 to 2006:Q1, obtained from the Bureau of Economic Analysis, given in Figure 3.1 (a) (see also Hendry and Nielsen (2007, Exercise 16.1), Nielsen and

Sohkanen (2009), and Sohkanen (2008)), and fit a fourth order autoregressive model in PcGive (see Doornik and Hendry (2007a)). The recursively estimated model is given in equation (3.1). Cumulative sums of squares of the model's residuals are plotted in panel (b), with point-wise confidence bands. The recursive residual variance estimator is given, with point-wise confidence bands, in panel (c), and the $CUSQ^{OLS}$ test in panel (d), with simultaneous confidence bands. All bands are drawn for the 95% confidence level, but we also add 99% simultaenous bands in panel (d).

Failure to fit any straight line within the bands in panel (b), or a horizontal line within the bands in panel (c) indicates rejection of variance constancy at the significance level associated with the bands. This procedure is, of course, an approximation as the bands are not simultaneous. Nevertheless, the evidence against variance constancy recoverable from panels (b) and (c) is consistent with that reported in panel (d), where the CUSUM of squares test is given with simultaneous bands (In fact, rejection would also have followed at the 99% confidence level, as the appropriate critical value in that case is 1.63 instead of 1.36; see Schumacher (1984, Tab. 9)) This finding is consistent with that of McConnell and Perez-Quiros (2000), who apply the CUSUM of squares test to levels of demeaned post-war US GDP data, that is, to residuals from a regression on a constant. Implementation with residuals from a regression with covariates is to be preferred, however, as Deng and Perron (2008a) note, since the serial correlation in the levels of the series will be modelled parametrically and thus eliminated, resulting in better finite sample performance of the test.

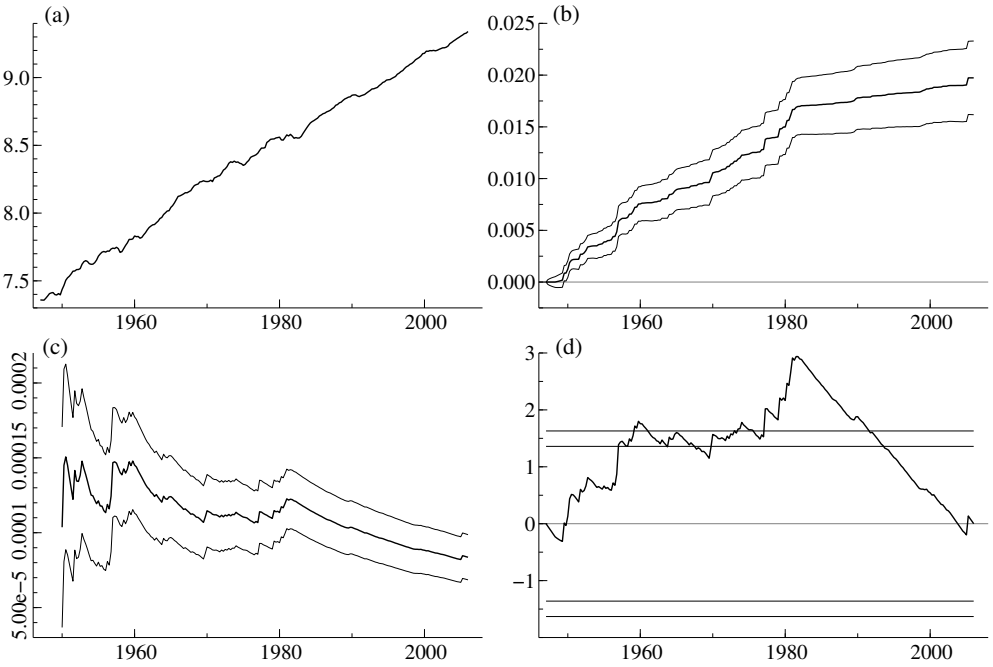


Figure 3.1: (a) US log GDP data and recursive statistics: (b) $R\hat{S}S_t$ with 95% point-wise bands, (c) $\hat{\sigma}_t^2$ with 95% point-wise bands, and (d) $CUSQ^{OLS}$-plot with 95% and 99% simultaneous bands.

$$\begin{aligned}\log(\text{GDP}) = {} & \underset{(0.113)}{0.3285} + \underset{(0.000128)}{0.0003537}\ \text{t} \\ & + \underset{(0.0661)}{1.272}\ \log(\text{GDP})_{t-1} - \underset{(0.106)}{0.1737}\ \log(\text{GDP})_{t-2} \\ & - \underset{(0.106)}{0.2294}\ \log(\text{GDP})_{t-3} + \underset{(0.0656)}{0.08804}\ \log(\text{GDP})_{t-4} \end{aligned} \tag{3.1}$$

σ (residual s.d.)	0.009159	RSS	0.019043
R^2	0.999738	F(5,227)	173000
mean(log (GDP))	8.41443	DW	2.01
var(log (GDP))	0.311547	no. of observations	233
CUSQ^{OLS}	2.918		

3.2 Example: The Dotcom Bubble

Modelling of stock market data by autoregressions may, in the presence of bubbles, result in estimated roots in the explosive region of the parameter space. The assumptions of Chapter 2 do not rule out such a situation, and the CUSUM of squares test and the recursive graphical techniques of Corollary 4.7 of Chapter 2 continue to be valid. Since the use of robust test procedures entails some efficiency loss if in fact the phenomenon the robustification addresses is absent, this is quite beneficial. As an example, we consider the Campbell and Shiller (1988) stock price and dividend data using PcGive (Doornik and Hendry 2007b); see also Engsted and Nielsen (2010). The estimated VAR(2) model of annual prices and dividends from 1874 to 2000, is

$$\begin{aligned} \mathrm{P}_t &= \underset{(0.072)}{0.057} + \underset{(0.091)}{1.236}\,\mathrm{P}_{t-1} - \underset{(0.127)}{0.025}\,\mathrm{P}_{t-2} - \underset{(4.77)}{11.35}\,\mathrm{D}_{t-1} + \underset{(4.47)}{5.998}\,\mathrm{D}_{t-2} \\ \mathrm{D}_t &= \underset{(0.001)}{0.002} + \underset{(0.002)}{0.008}\,\mathrm{P}_{t-1} - \underset{(0.002)}{0.008}\,\mathrm{P}_{t-2} + \underset{(0.087)}{1.107}\,\mathrm{D}_{t-1} - \underset{(0.082)}{0.143}\,\mathrm{D}_{t-2}. \end{aligned}$$

The companion matrix has three eigenvalues corresponding to stationary roots (two complex and one real) and one eigenvalue (real at 1.183) corresponding to an explosive root of the characteristic polynomial. Arguably, this is due to the asset price bubble of the dotcom boom in the late 1990s. Basing the CUSUM of squares test on the OLS residuals, we reject constancy of the residual variance in both equations. The CUSQ test statistics with simultaneous confidence bands, one pair using the asymptotically valid values and another applying the correction derived in Chapter 4, are reported in Figures 3.2 (a) and 3.2 (b). The same model fitted to a sub-sample, 1976-2000 is

$$\mathrm{P}_t = \underset{(1.1)}{-0.339} + \underset{(0.206)}{0.983}\,\mathrm{P}_{t-1} + \underset{(0.306)}{0.351}\,\mathrm{P}_{t-2} + \underset{(24)}{1.904}\,\mathrm{D}_{t-1} - \underset{(23.1)}{4.443}\,\mathrm{D}_{t-2}$$

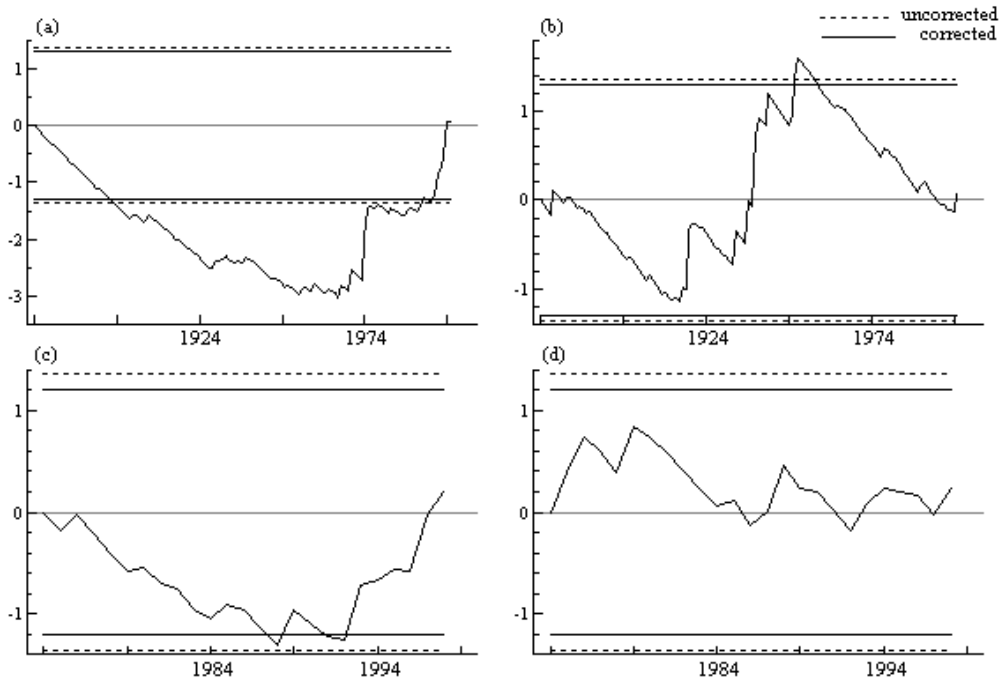


Figure 3.2: The finite sample uncorrected and corrected sup $CUSQ^{OLS}$ test at 5% level for the (a) price equation, (b) dividend equation, (c) price equation in the sub-sample, and (d) dividend equation in the sub-sample

$$\underset{}{\mathrm{D}_t} = \underset{(0.008)}{0.019} + \underset{(0.002)}{0.004}\ \mathrm{P}_{t-1} - \underset{(0.002)}{0.004}\ \mathrm{P}_{t-2} + \underset{(0.181)}{1.329}\ \mathrm{D}_{t-1} - \underset{(0.175)}{0.561}\ \mathrm{D}_{t-2}.$$

Again, the companion matrix has three eigenvalues corresponding to stationary roots (two complex and one real) and one eigenvalue (real at 1.258) corresponding to an explosive root of the characteristic polynomial. In this sub-sample, the test statistic is not significant at the 5% level for the dividend equation, but we do reject constancy in the price equation, using the finite sample corrected test. These tests are reported in Figures 3.2 (c) and 3.2 (d). There appears to be a positive unit root in the model. However, as a linear trend is not estimated, this should not cause the size distortion suggested by the results reported in Chapter 4.

4 Mathematical framework

The scope of this thesis calls for the use of a variety of mathematical methods.

The asymptotic theory in Chapter 2 is developed in stages. First, we use strong consistency results applying in non-stationary autoregressions (Lai and Wei 1982, Lai and Wei 1983, Lai and Wei 1985) developed further to allow for quite a general class of deterministic terms by Nielsen (2005). This may seem complicated relative to alternative approaches such as using results on normed vector spaces, but the martingale theory of Hall and Heyde (1980) employed in the argument essentially establishes strong convergence, and it is therefore natural that the rest of the argument applies in similar generality. The quantities of interest are then re-defined as random functions on a function space endowed with a metric to enable the use functional central limit theory (see Chan and Wei (1988) and Billingsley (1999)) for martingale difference sequences to characterize the asymptotic distribution. Many problems in econometrics have been solved by the application of this powerful tool, e.g. asymptotics of unit root tests pioneered by Phillips (1987); see also Davidson (2006). This is another reason for relying on the strong results in the first stage. The initial argument is effectively needed to show that nuisance quantities vanish, and the almost sure sense in which this is shown is sufficient for neglecting those nuisance quantities when applying the limit theory. In Chapter 3, we also apply martingale limit theory (Hall and Heyde 1980) and the strong consistency results of Nielsen (2005).

In Chapter 4, Monte Carlo methods are utilized and no sophisticated mathematical arguments are necessary. The choice of regressors for response surfaces is informed by central limit theory.

Chapter 5 calls for use of an appropriate generalization of the functional central limit theorem for martingale differences applying to multivariate processes.

We apply Helland (1982, Thm. 3.3). Owing to the discontinuities in some of the mappings applied to the limits, Skorokhod's representation theorem is applied; see Billingsley (1999, Thm. 6.7) and also Chan and Wei (1988).

Throughout this thesis, heavy use is made of the The Bachmann-Landau O-notation (see e.g. von Landau (1927, Ch. 7) or deBruijn (1961, Ch. 1)).

Chapter 2

Asymptotic behaviour of the CUSUM of squares test under stochastic and deterministic time trends[1]

1 Introduction

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

[1] A version of this chapter has been published in *Econometric Theory*, 27, 2011, 913-927 (with Bent Nielsen). The author is grateful to Cambridge University Press for their permission to reproduce a version in this thesis. The version reproduced here adds Corollary 4.7, preceded by a paragraph explaining its contents.

trends. We show that the usual asymptotic distribution applies quite generally for the least squares test. This is important in applications as the question of variance constancy can be addressed without having to locate the characteristic roots. The results generalize work for stationary cases by Ploberger and Krämer (1986) and Deng and Perron (2008a). Lee, Na, and Na (2003) considered the least squares test for a unit root autoregression without deterministic terms. The presented analysis is based on the results of Lai and Wei (1985) and Nielsen (2005), and is easiest for the least squares test. For the forecast test the usual asymptotic distribution applies in purely non-explosive and purely explosive cases but nuisance terms can arise if the characteristic roots of the process are located both in the non-explosive and the explosive region of the parameter space. Analysis of this case is postponed until the next chapter.

The two test statistics are presented in §2 while the model and assumptions are presented in §3. The asymptotic results for the least squares test and the forecast test are presented in §4 and §5, respectively. §6 contains a simulation study involving first order autoregressions. §7 concludes, and proofs are relegated into an appendix.

2 The test statistics

The forecast residual based test statistic was suggested along with exact distribution results by Brown, Durbin, and Evans (1975) for the classical linear regression

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

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

as

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

along with the recursive forecast residuals

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

the CUSQ plot with recursive residuals is defined as

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

The econometric software PcGive (Doornik and Hendry 2007a) features a graphical diagnostic method ('Residual sum of squares') based on the recursively computed least squares residuals

$$\hat{\varepsilon}_{s,t} = y_s - \hat{\beta}_t' x_s \qquad \text{for } M \leq t. \tag{2.5}$$

Normalizing sums of these residuals by t we obtain the recursive residual variances

$$\hat{\sigma}_t^2 = t^{-1} \sum_{s=1}^{t} \hat{\varepsilon}_{s,t}^2 \qquad \text{for } M \leq t. \tag{2.6}$$

and define the CUSQ plot with least squares residuals as

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

We note that McCabe and Harrison (1980) and Deng and Perron (2008a) consider behaviour of the plot defined instead with the non-recursively computed OLS

residuals

$$\hat{\varepsilon}_{s,T} = y_s - \hat{\beta}_T' x_s \tag{2.8}$$

Asymptotically, the use of recursive as opposed to non-recursive OLS residuals in the construction of the plot does not make a difference. The proofs are slightly more complicated in the recursive case we consider.

3 Model and assumptions

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

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

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

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

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

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

assume that X_t and D_t satisfy the vector autoregressions

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

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

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

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

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

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

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

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

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

Assumption B. *The eigenvalues of $\mathbf{D}$ are on the unit circle and in addition* $\operatorname{rank}\left(D_1, \ldots, D_{\dim(\mathbf{D})}\right) = \dim(\mathbf{D})$.

Nearly all values of autoregressive parameters A_j are allowed in the vector autoregression (3.3), including stationary roots, roots on the unit circle and a range of explosive roots. The only exception is the case of singular explosive roots which can arise for vector autoregressions where $p \geq 2$ with more than one

explosive root, see Anderson (1959), Duflo, Senoussi, and Touati (1991), Phillips and Magdalinos (2008) and Nielsen (2008) for further discussion. Thus, define the companion matrices

$$\mathbf{B} = \left\{ \begin{array}{cc} (A_1, \ldots, A_{k-1}) & A_k \\ I_{p(k-1)} & 0 \end{array} \right\}, \ \mu = \left(\begin{array}{c} \mu \\ 0 \end{array} \right), \ \mathbf{S} = \left(\begin{array}{cc} \mathbf{B} & \mu \\ 0 & \mathbf{D} \end{array} \right) \tag{3.7}$$

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

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

$$\xi_t = \left(\begin{array}{c} \xi_t^{(1)} \\ \xi_t^{(2)} \end{array} \right), \qquad \Omega = \left(\begin{array}{cc} \Omega_{yy} & \Omega_{yz} \\ \Omega_{zy} & \Omega_{zz} \end{array} \right),$$

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

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

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

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

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

Assumption D. *For the regression* (3.1), $\mathcal{G}_{t-1}$ *is the* σ*-field generated by* Z_t *and*

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

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

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

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

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

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

The normalized partial sums of squared innovations are asymptotically Brownian. This follows through a direct application of Chan and Wei (1988, Thm. 2.2). A shorthand notation is used to indicate the convergence to a Brownian motion $\mathcal{B}$ of a random function on $[0,1]$. When we say that for $u \in [0,1]$, $G_{\mathrm{int}(Tu)} \to \mathcal{B}_u$ in distribution on $D[0,1]$, we mean that for some g_s, G_T is the random element of $D[0,1]$ (equipped with the uniform metric) defined by $G_T(u) = T^{-1/2}\sum_{s=1}^{\mathrm{int}(Tu)} g_s$, and $G_T \to \mathcal{B}$ in distribution on $D[0,1]$.

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

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

Theorem 4.4. *Assume A, B, C, D. Let $\mathcal{B}^\circ$ be a standard Brownian bridge. Then*
(i) $CUSQ^{OLS}_{int(Tu),T} \to \sigma^{-2}\varphi\mathcal{B}^\circ_u$ *in distribution on* $D[0,1]$,
(ii) $\sup_{t\leq T}|CUSQ^{OLS}_{t,T}| \to \sigma^{-2}\varphi \sup_{u\leq 1}|\mathcal{B}^\circ_u|$ *in distribution on* $\mathbb{R}$.

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

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

Remark 4.6. *The convergence in Theorem 4.5 could be strengthened to almost sure convergence if it were assumed that ε^3_t is a martingale difference. A convergence result for φ_t as a process on $D[0,1]$ could then be deduced. The proof would follow by combining the presented proof with Theorem 2.4 of Nielsen (2005).*

An alternative graphical variance constancy diagnostic is considered without confidence bands by Hendry (1986) and Doornik and Hendry (2007a). The idea is to plot a normalized residual sum of squared residuals, $t^{-1}RSS_t$ where $RSS_t = \sum_{s=1}^t \hat{\varepsilon}^2_{s,t}$. The asymptotic behaviour of $T^{-1/2}RSS_{\text{int}(Tu)}$ follows readily from Lemmas 4.1 and 4.3. Since the function $u^{-1}\mathbf{1}_{(u\geq\nu)}$ is in $D[0,1]$ for all $\nu > 0$, although not for $\nu = 0$, it can be multiplied with $T^{-1/2}RSS_{\text{int}(Tu)}$ and the asymptotic distribution follows from the continuous mapping theorem.

Corollary 4.7. *Assume A, B, C, D. Let $\mathcal{B}$ be a standard Brownian motion. Then, for $\nu > 0$,* $(Tu)^{-1/2}\{RSS_{int(Tu)} - \sigma^2 Tu\}\mathbf{1}_{(u\geq\nu)} \to \mathbf{1}_{(u\geq\nu)}\varphi u^{-1/2}\mathcal{B}_u$ *in distribution on* $D[0,1]$.

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

Consider now the $CUSQ^{REC}$-statistic (2.4) applied to the regressions (3.1) and (3.2). This statistic is more complicated to describe than the $CUSQ^{OLS}$-statistic. Nuisance terms arise in special cases; see Chapter 3.

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

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

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

The difference between the sum of squared forecast residuals and the sum of squared innovations will in general involve nuisance terms; see Chapter 3.

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

Remark 5.2. *If the process is mixed so that* $\dim(\mathbf{R}) > 0$ *and* $\dim(\mathbf{W}) > 0$ *then non-negligible nuisance terms will appear in Lemma 5.1; see Chapter 3. It is not immediately clear if these nuisance terms will cancel each other.*

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

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

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

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

6 Simulation study

Theorems 4.4 and 5.3 show that the two types of $CUSQ$-statistics have the usual limit distribution in many situations. This leaves the questions of whether the finite sample distributions are different for the two statistics and whether they depend on the autoregressive parameters. These questions are addressed through a small-scale Monte Carlo study. For the important question of the power of these tests we refer to the studies by Deng and Perron (2008a), Deng and Perron (2008b).

The data generating process is a univariate autoregression, $X_t = \alpha X_{t-1} + \varepsilon_t$ for $t = 1, \ldots, T = 100$ with initial value $X_0 = 0$, standard normal innovations, and a range of autoregressive parameters α. The number of repetitions is 10^6. Due to the normality $\sigma^{-2}\varphi = \sqrt{2}$, so the statistics of interest are $S^{OLS} = \max_{M\leq t\leq T} |CUSQ^{OLS}_{t,T}|/\sqrt{2}$ and $S^{REC} = \max_{M\leq t\leq T} |CUSQ^{REC}_{t,T}|/\sqrt{2}$; see (2.6). Theorems 4.4 and 5.3 show that their limit distribution is the supremum of a Brownian bridge. Billingsley (1999, pp. 101-104) gives an analytic expression for the distribution function. In particular, the 95% quantile is 1.36; see Schumacher (1984, Tab. 9).

Table 2.1 reports mean and median of the two statistics as a function of α. The variation of the distribution for the two supremum statistics is very small but significant when taking the Monte Carlo precision into account. This impression

Table 2.1: Simulated means and medians of the CUSQ tests for different values of α. The Monte Carlo standard error is 2.5×10^{-4}. The slight variation in the reported figures is therefore significant.

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

was confirmed when looking at other descriptives such as standard deviation, 95% quantile and p-value of the asymptotic 95% quantile when $\alpha = 0$.

Two conclusions emerge from this small-scale Monte Carlo study. First, there is not much difference in finite sample distribution for the two statistics. Second, there is very little variation in the finite sample distribution with the unknown parameter. This suggests that very simple finite sample corrections could be used; see Chapter 4.

7 Conclusion

In this chapter, we have derived the asymptotic distribution on the CUSUM of squares test implemented with either OLS residuals or standardized one-step-ahead forecast errors. We have also shown how the functional limit result could be used to construct point-wise confidence bands for the recursive sums of squared OLS residuals. The result for the OLS based test is general in the sense that, apart from a restriction to rule out singularity which would cause least squares estimation to be inconsistent, it does not depend upon the location of the roots

in of the vector autoregressive data generating process. In the case of the forecast error based test, nuisance terms appear in the limit unless the the data generation process is restricted to be either purely explosive or purely non-explosive. Expressions are derived for the nuisance terms in the next chapter. These results are asymptotic, of course, and the finite sample behaviour of the tests should be investigated. This task is undertaken in Chapter 4.

A Appendix

Proofs

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

A.1 The case of least squares residuals

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

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

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

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

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

By Nielsen (2005, Cor. 2.6) then

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

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

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

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

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

Proof of Theorem 4.4. (*i*) The Lemmas 4.1, 4.2 and 4.3 imply

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

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

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

and insert the above convergence result.

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

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

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

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

Since $Z_t = \theta S_{t-1} + \xi_t^{(2)}$ for some θ, see (3.3), regression on regressors $Z_t,\ S_{t-1}$ and on regressors $x_t = (\xi_t^{(2)\prime}, S_{t-1}')'$ is equivalent. Define

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

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

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

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

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

Secondly, consider $\mathcal{I}_1 = P_t(\sum_{s=1}^{t}x_s x_s')^{-1/2}\sum_{s=1}^{t}x_s\varepsilon_s^3$. As in (5.1) we can decompose S_{t-1} into autoregressions $U_{t-1}, V_{t-1}, W_{t-1}$ with stationary, unit, and explosive roots. The components $\xi_t^{(2)}, U_{t-1}, V_{t-1}, W_{t-1}$ are asymptotically uncor-

related due to Nielsen (2005, Thms. 2.4, 9.1, 9.2, 9.4) given Assumptions A, B, and C. Thus, as above

$$\mathcal{I}_1 \overset{a.s.}{=} P_t[\{\sum_{s=1}^{t} \xi_s^{(2)}\xi_s^{(2)\prime}\}^{-1/2} \sum_{s=1}^{t} \xi_s^{(2)}\varepsilon_s^3 + \{\sum_{s=1}^{t} U_{s-1}U_{s-1}'\}^{-1/2} \sum_{s=1}^{t} U_{s-1}\varepsilon_s^3 + \{\sum_{s=1}^{t} V_{s-1}V_{s-1}'\}^{-1/2} \sum_{s=1}^{t} V_{s-1}\varepsilon_s^3 + \{\sum_{s=1}^{t} W_{s-1}W_{s-1}'\}^{-1/2} \sum_{s=1}^{t} W_{s-1}\varepsilon_s^3]\{1 + \mathrm{o}(1)\}. \tag{A.2}$$

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

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

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

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

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

$$|P_t(\sum_{s=1}^{t} V_sV_s')^{-1/2}(\sum_{s=1}^{t} V_s)\varepsilon_s^3| \leq |P_t|\{\max_{s\leq t} |(\sum_{s=1}^{t} V_sV_s')^{-1/2}V_s|\} \sum_{s=1}^{t} |\varepsilon_s^3|$$

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

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

$$\|\sum_{s=1}^{t} \mathbf{W}^{-t} W_{s-1} W'_{s-1} (\mathbf{W}')^{-t}\|^{-1/2} (\sum_{s=1}^{t} \|\mathbf{W}^{-t} W_{s-1}\|) \max_{s\leq t} \|\varepsilon_s\|^3$$

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

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

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

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

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

so the last term is bounded. □

A.3 The case of recursive residuals

Lemma 5.1 is proved in three steps. Only the regression (3.1) including Z_t as a regressor is considered. As in the proof of Lemma 4.5 the regressor can be taken as $x_t = (\xi_t^{(2)\prime}, R'_{t-1}, W'_{t-1})'$ where $\xi_t^{(2)}$ is the Z_t-innovation while R_t and W_t are the

non-explosive and explosive components. Define:

$$a_t = \varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s W'_{s-1}(\sum_{s=1}^{t-1} W_{s-1}W'_{s-1})^{-1}W_{t-1}, \quad A_t = W'_{t-1}(\sum_{s=1}^{t-1} W_{s-1}W'_{s-1})^{-1}W_{t-1},$$
$$b_t = \sum_{s=1}^{t-1} \varepsilon_s \xi_s^{(2)\prime}(\sum_{s=1}^{t-1} \xi_s^{(2)}\xi_s^{(2)\prime})^{-1}\xi_t^{(2)}, \quad B_t = \xi_t^{(2)\prime}(\sum_{s=1}^{t-1} \xi_s^{(2)}\xi_s^{(2)\prime})^{-1}\xi_t^{(2)},$$
$$c_t = \sum_{s=1}^{t-1} \varepsilon_s R'_{s-1}(\sum_{s=1}^{t-1} R_{s-1}R'_{s-1})^{-1}R_{t-1}, \quad C_t = R'_{t-1}(\sum_{s=1}^{t-1} R_{s-1}R'_{s-1})^{-1}R_{t-1},$$
$$d_t = \sum_{s=1}^{t-1} \varepsilon_s x'_s(\sum_{s=1}^{t-1} x_s x'_s)^{-1}x_t, \quad D_t = x'_t(\textstyle\sum_{s=1}^{t-1} x_s x'_s)^{-1}x_t,$$
$$f_t^2 = 1 + D_t. \quad \mathcal{I}_{yz} = \sum_{s=1}^{t} \frac{y_s z_s}{f_s^2},\ y_s, z_s \in (a_s, b_s, c_s, d_s, \varepsilon_s).$$

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

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

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

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

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

Lemma A.2. *Assume A. Then* $\mathcal{I}_{\varepsilon b} = \mathrm{o}(t^{1/2})$ *a.s.*

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

a.s. on the set where

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

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

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

a.s. by Nielsen (2005, Thms. 2.4, 5.1, 6.1) given Assumption A. □

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

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

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

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

Due to the partitioned inversion formula

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

it holds $\mathcal{I}_{bb} \leq t^{\eta}\epsilon \sum_{s=1}^{t} \xi_s^{(2)\prime}(\sum_{v=1}^{s} \xi_v^{(2)}\xi_v^{(2)\prime})^{-1}\xi_s^{(2)}$. The sum is of order $\mathrm{O}(\log t)$ due

to Nielsen (2005, Lem. 8.6) assuming A, implying that $\mathcal{I}_{bb}$ is $\mathrm{o}(t^{\eta}) = \mathrm{o}(t^{1/2})$ *a.s.*

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

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

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

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

(i) $\sum_{t=T_0}^{T} h'_t H_t^{-1} h_t = \mathrm{O}(\log \lambda_T^*)$.

(ii) $h'_t H_{t-1}^{-1} h_t = h'_t H_t^{-1} h_t / (1 - h'_t H_t^{-1} h_t)$.

(iii) $\sum_{t=T_0+1}^{T} h'_t H_{t-1}^{-1} h_t = \mathrm{O}(\log \lambda_T^*)$.

Proof of Lemma A.4. *(i)* is the statement of Lai and Wei (1982, Lem. 2.*ii*).

(ii) This follows by (A.4).

(iii) By Lai and Wei (1982, Lem. 2.*i*) then $h'_t H_t^{-1} h_t = 1 - \det H_t / \det H_{t-1}$. Combine this and *(ii)* to get $\sum_{t=T_0+1}^{T} h'_t H_{t-1}^{-1} h_t = \sum_{t=T_0+1}^{T} (\det H_t - \det H_{t-1}) / \det H_{t-1}$. Then complete the argument as in the proof of Lai and Wei (1982, Lemma 2.*ii*). □

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

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

inside and on the complex unit circle, respectively; see Nielsen (2005, §3). These cases can be studied separately.

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

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

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

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

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

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

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

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

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

With this definition and a change of summation order it holds

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

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

$$Z_h = \mathbf{W}^{1-h} W_{h-1} = W_0 + \sum_{s=1}^{h-1} \mathbf{W}^{-s} e_{W,s}$$
$$F_s = \sum_{u=1}^{s-1} \mathbf{W}^{1-s} W_{u-1} W_{u-1}' (\mathbf{W}')^{1-s} = \sum_{u=1}^{s-1} \mathbf{W}^{u-s} Z_u Z_u' (\mathbf{W}')^{u-s}$$

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

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

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

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

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

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

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

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

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

since:

$$\sum_{s-h=0}^{\infty} \tilde{G}_{s-h}^2 = (1 + Z'F^{-1}Z)^{-1}\{1 + Z'F^{-1} \sum_{s-h=0}^{\infty} \mathbf{W}^{h-s} Z Z' (\mathbf{W}')^{h-s} F^{-1} Z\}$$
$$= (1 + Z'F^{-1}Z)^{-1}(1 + Z'F^{-1}FF^{-1}Z) = 1,$$

whereas the sum of cross products satisfies

$$\sum_{s-h=0}^{\infty} \tilde{G}_{s-h}\tilde{G}_{s-h+\ell}.$$
$$=(1 + Z'F^{-1}Z)^{-1}\{-Z'F^{-1}\mathbf{W}^{\ell}Z + Z'F^{-1} \sum_{s-h=0}^{\infty} \mathbf{W}^{h-s} Z Z' (\mathbf{W}')^{h-s} (\mathbf{W}')^{\ell} F^{-1} Z\}$$
$$=(1 + Z'F^{-1}Z)^{-1}\{-Z'F^{-1}\mathbf{W}^{\ell}Z + Z'F^{-1}F(\mathbf{W}')^{\ell}F^{-1}Z\} = 0,$$

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

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

Proof of Lemma 5.1. If $\dim(\mathbf{W}) = 0$ combine Lemmas A.1-A.6. If $\dim(\mathbf{R}) = 0$ apply Lemma A.7. □

Chapter 3

Behaviour of the CUSUM of squares of recursive residuals in processes with both explosive and non-explosive roots

1 Introduction

Lemma 5.1. of Chapter 2 states that, under the assumption that the data generation process is either purely explosive or purely non-explosive, the partial sums of squares of the recursive residuals are asymptotically close to sums of squares of the fundamental innovations. This result is relied upon in the derivation of the asymptotic distribution of the $CUSQ^{REC}$-test; see Theorem 5.3 of Chapter 2. In this chapter, we argue that when the data generation process is mixed, that is, neither purely non-explosive nor purely explosive, a non-vanishing nuisance term arises.

The simulations in Chapter 3 indicate no large distortions in the mixed explosive/non-explosive case, and this suggests that the problem may have limited practical relevance. If the lack of asymptotic similarity in the mixed case causes worries, the practitioner may wish to use the test based on OLS residuals instead. In fact, when restricting attention to behaviour of the CUSUM of squares test under the null hypothesis, there is little reason to implement it with the recursive residuals in the time series context. This conclusion does not necessarily hold once the behaviour under some alternative hypotheses is taken into account. In particular, it has been suggested that the relative power performance of the tests depends upon where in the sample the break lies (Deng and Perron 2008b). Understanding of the behaviour of the recursive residuals based test is therefore important. We stress that if the practitioner wishes, a priori, to rule out explosive behaviour of the variable being modelled, the recursive residuals based test exhibits the same asymptotic similarity as the OLS based test, and this is also the case when the process generating the data is purely explosive. It is only the rather special mixed case where a nuisance term arises.

In §2, we state the necessary assumptions. The main result is stated and discussed in §3 and proofs are in the appendix.

2 Model and Assumptions

The setting is identical to that of Chapter 2, §3. We briefly repeat the assumptions.

The observed p-dimensional time series X_t partitioned as $(Y_t, Z_t')'$, where Y_t

is univariate and Z_t is of dimension $p-1 \geq 0$, is generated by

$$X_t = \sum_{j=1}^{k} A_j X_{t-j} + \mu D_{t-1} + \xi_t \qquad t = 1, \ldots T, \tag{2.1}$$

$$D_t = \mathbf{D} D_{t-1}, \tag{2.2}$$

where $\mathbf{D}$ is a deterministic matrix. The innovations ξ_t satisfy a martingale difference assumption.

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

$$\sup_t \mathsf{E}\{(\xi_t'\xi_t)^{\lambda/2}|\mathcal{F}_{t-1}\} \overset{a.s.}{<} \infty \qquad \text{for some } \lambda > 4, \tag{2.3}$$

$$\mathsf{E}(\xi_t\xi_t'|\mathcal{F}_{t-1}) \overset{a.s.}{=} \Omega \qquad \text{where } \Omega \text{ is positive definite.} \tag{2.4}$$

The deterministic term satisfies the following assumption.

Assumption B. *The eigenvalues of $\mathbf{D}$ are on the unit circle and in addition* $\operatorname{rank}\left(D_1, \ldots, D_{\dim(\mathbf{D})}\right) = \dim(\mathbf{D})$.

We define

$$\mathbf{B} = \begin{Bmatrix} (A_1, \ldots, A_{k-1}) & A_k \\ I_{p(k-1)} & 0 \end{Bmatrix}, \ \mu = \begin{pmatrix} \mu \\ 0 \end{pmatrix}, \ \mathbf{S} = \begin{pmatrix} \mathbf{B} & \mu \\ 0 & \mathbf{D} \end{pmatrix} \tag{2.5}$$

and the companion matrix associated with the stochastic part of the process is taken to satisfy the following assumption.

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

We fit the autoregressive distributed lag model of order k

$$Y_t = \rho' Z_t + \sum_{j=1}^{k} \alpha_j Y_{t-j} + \sum_{j=1}^{k} \beta_j' Z_{t-j} + \nu D_{t-1} + \varepsilon_t, \qquad t = 1, \dots T. \tag{2.6}$$

We define

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

conformably with $X_t = (Y_t, Z_t')'$. A martingale structure is assumed for ε_t^2.

Assumption D. *For the regression* (2.6)*, $\mathcal{G}_{t-1}$ is the σ-field generated by Z_t and $\mathcal{F}_{t-1}$, while $\mathcal{G}_t = \mathcal{F}_t$ when $\rho' = 0$. Suppose $(\varepsilon_t^2 - \sigma^2, \mathcal{G}_t)$ is a martingale difference sequence satisfying $\mathsf{Var}(\varepsilon_t^2 - \sigma^2 | \mathcal{G}_{t-1}) = \varphi^2$ a.s. for some $\varphi > 0$ and $\sup_t \mathsf{E}(|\varepsilon_t|^\lambda | \mathcal{G}_{t-1}) < \infty$ a.s. for some $\lambda > 4$.*

3 Asymptotic analysis of the cumulated sums of squared recursive residuals

The difference between the sum of squared forecast residuals and the sum of squared innovations will in general involve nuisance terms. In order to show this, it is necessary to decompose the data generation process into non-explosive and explosive processes. We define the companion vector $S_{t-1} = (X_{t-1}', \dots X_{t-k}', D_{t-1}')$ and the selection matrix $\iota = (I_p, 0_{(pk-p+\dim \mathbf{D}) \times p})'$. We may then write the process generating the data as the first order vector autoregression $S_t = \mathbf{S} S_{t-1} + \iota \xi_t$, recalling the companion matrix $\mathbf{S}$ defined in (2.5). Then, there is a real matrix

M such that $M\mathbf{S}M^{-1}$ is block diagonal and

$$MS_t = \begin{pmatrix} R_t \\ W_t \end{pmatrix} = \begin{pmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{W} \end{pmatrix} \begin{pmatrix} R_{t-1} \\ W_{t-1} \end{pmatrix} + \begin{pmatrix} e_{R,t} \\ e_{W,t} \end{pmatrix} \tag{3.1}$$

where the moduli of the eigenvalues of $\mathbf{R}$ and $\mathbf{W}$ are at most one and strictly greater than one, respectively, and deterministic components are subsumed into the R_t-process. It can be shown that one of the nuisance terms involves interactions between these different parts of the process. While the processes R_t and W_t themselves can be shown to be asymptotically uncorrelated (see Nielsen (2005, Thm 2.4), the interactions, involving these terms, that arise in the argument turn out not to vanish. In particular, the term involving the aforementioned interactions is not a martingale, and results concerning martingale convergence can therefore not be applied. The other nuisance term will be shown to depend entirely on the explosive component of the data generation process.

Theorem 3.1. *Assume A, B, and C. Then, defining*

$$\mathcal{J}_t = \sum_{s=1}^{t} \left\{ \frac{\sum_{j=1}^{s-1}[\varepsilon_j W'_{j-1}(\sum_{j=1}^{s-1} W_{j-1}W'_{j-1})^{-1}W_{s-1}] \sum_{j=1}^{s-1}[\varepsilon_j R'_{j-1}(\sum_{j=1}^{s-1} R_{j-1}R'_{j-1})^{-1}R_{s-1}]}{1 + W'_{s-1}(\sum_{j=1}^{s-1} W_{j-1}W'_{j-1})^{-1}W_{s-1}} \right\}$$

we have $\sum_{s=1}^{t}(\tilde{\varepsilon}_s^2 - \varepsilon_s^2) = 2\mathcal{J}_t \mathbf{1}_{(dim(\mathbf{W})>0)} \mathbf{1}_{(dim(\mathbf{R})>0)} + \mathrm{o}(t^{1/2})$ *a.s.*

Remark 3.2. *If the process X_t is either purely non-explosive or purely explosive then the term $\mathcal{J}_t$ is absent.*

Remark 3.3. *It can be shown that the term $t^{-1/2}\mathcal{J}_t$ is non-vanishing, although its convergence rate has not been established. In the appendix, we show that $t^{-1/2}\mathcal{J}_t$ is non-vanishing for the special case where the process Y_t is an explosive (univariate)*

first order autoregression, but is modelled as a second order autoregression, so that $R_t = \varepsilon_t$ *up to a constant.*

A Appendix

Proofs

The regressor can be taken as $x_t = (\xi_t^{(2)\prime}, R'_{t-1}, W'_{t-1})'$ where $\xi_t^{(2)}$ is the Z_t-innovation while R_t and W_t are the non-explosive and explosive components. The recursive residual is given by

$$\tilde{\varepsilon}_t = \{1 + x'_t(\sum_{s=1}^{t-1} x_s x'_s)^{-1} x_t\}^{-1/2}\{\varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s x'_s(\sum_{s=1}^{t-1} x_s x'_s)^{-1} x_t\}$$

Define

$$a_t = \varepsilon_t - \sum_{s=1}^{t-1} \varepsilon_s W'_{s-1}(\sum_{s=1}^{t-1} W_{s-1}W'_{s-1})^{-1}W_{t-1}, \quad A_t = W'_{t-1}(\sum_{s=1}^{t-1} W_{s-1}W'_{s-1})^{-1}W_{t-1},$$

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

$$c_t = \sum_{s=1}^{t-1} \varepsilon_s R'_{s-1}(\sum_{s=1}^{t-1} R_{s-1}R'_{s-1})^{-1}R_{t-1}, \quad C_t = R'_{t-1}(\sum_{s=1}^{t-1} R_{s-1}R'_{s-1})^{-1}R_{t-1},$$

$$d_t = \sum_{s=1}^{t-1} \varepsilon_s x'_s(\sum_{s=1}^{t-1} x_s x'_s)^{-1}x_t, \quad D_t = x'_t(\sum^{t-1} x_s x'_s)^{-1}x_t,$$

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

We shall prove Theorem 3.1 by combining two lemmas. We state and prove each in turn, starting with the simpler one. We then consider the behaviour of the nuisance term in a special univariate setting.

Lemma A.1. *Assume A, B, and C. Then*

$$\sum_{s=1}^{t}\left(\varepsilon_s^2-\frac{\varepsilon_s^2}{f_s^2}\right)=\sum_{s=1}^{t}\frac{\varepsilon_s^2 A_s}{1+A_s}\mathbf{1}_{(dim(\mathbf{W})>0)}+\mathrm{o}(t^{1/2})\quad a.s., \tag{A.1}$$

Proof of Lemma A.1. The quantity of interest can be split into two terms as follows

$$\sum_{s=1}^{t}\left(\varepsilon_s^2-\frac{\varepsilon_s^2}{f_s^2}\right)=\sum_{s=1}^{t}\frac{\varepsilon_s^2 A_s}{1+A_s}+\sum_{s=1}^{t}\frac{\varepsilon_s^2(B_s+C_s)}{(1+A_s)(1+A_s+B_s+C_s)}$$

The second term on the right hand side is bounded by $\sum_{s=1}^{t}\varepsilon_s^2(B_s+C_s)\leq(\max_{s\leq t}\varepsilon_s)\sum_{s=1}^{t}(B_s+C_s)$. Since $\max_{s\leq t}\varepsilon_s=\mathrm{o}(t^{\frac{1}{2}-\eta})$ and $\sum_{s=1}^{t}(B_s+C_s)=\mathrm{o}\{\log(t)\}$ by Lemma A.4. (iii) of Chapter 2 and Nielsen (2005, Thm. 7.1), the numerator of the second term vanishes. □

Lemma A.2. *Assume A, B, and C. Then*

$$\sum_{s=1}^{t}\left(\tilde{\varepsilon}_s^2-\frac{\varepsilon_s^2}{f_s^2}\right)=\{\sum_{s=1}^{t}\frac{\varepsilon_s^2 A_s}{1+A_s}+2\sum_{s=1}^{t}\frac{(\varepsilon_s-a_s)c_s}{1+A_s}\mathbf{1}_{(dim(\mathbf{R})>0)}\}\mathbf{1}_{(dim(\mathbf{W})>0)}$$
$$+\mathrm{o}(t^{1/2})\quad a.s, \tag{A.2}$$

and we let $\sum_{s=1}^{t}\{(\varepsilon_s-a_s)c_s/(1+A_s)\}=\mathcal{J}_t$ *to coincide with the definition in Theorem 3.1 .*

The proof of Lemma A.2 requires some auxiliary results, and we will start by establishing these. The first lemma we prove is a generalization of Lemma A.1 of Chapter 2 to the case where we have $\dim(\mathbf{W})\neq 0$.

Lemma A.3. *Assume A, B, and C. Then*

$$\sum_{s=1}^{t}\left(\tilde{\varepsilon}_s^2-\frac{\varepsilon_s^2}{f_s^2}\right)=\{\mathcal{I}_{aa}-\mathcal{I}_{\varepsilon\varepsilon}+\mathcal{I}_{bb}+\mathcal{I}_{cc}+2(\mathcal{I}_{bc}-\mathcal{I}_{ab}-\mathcal{I}_{ac})\}\{1+\mathrm{o}(1)\}, \quad \text{(A.3)}$$

$$f_t^2=(1+A_t+B_t+C_t)\{1+\mathrm{o}(1)\}, \quad \text{(A.4)}$$

both holding a.s.

Proof of Lemma A.3. Notice that $\tilde{\varepsilon}_t=f_t^{-1}(\varepsilon_t-d_t)$, so $\sum_{s=1}^t\tilde{\varepsilon}_s^2=\sum_{s=1}^t f_s^{-2}(\varepsilon_s^2-2\varepsilon_s d_s+d_s^2)$ and therefore

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

Regression on $(Z_t',S_{t-1}')'$ is equivalent to regression on $(\xi_t^{(2)\prime},R_{t-1}',W_{t-1}')'$ since in the autoregression $S_t=\mathbf{S}S_{t-1}+e_{S,t}$, for some θ, $Z_t=\theta S_{t-1}+\xi_t^{(2)}$. The result follows because the components of S_{t-1} separated by the transformation given in (3.1) are asymptotically uncorrelated due to Nielsen (2005, Thm. 2.4). □

Lemma A.4. *Assume A. Then* $\mathcal{I}_{ab}=\mathrm{o}(t^{1/2})$ *a.s.*

Proof of Lemma A.4. Write $\mathcal{I}_{ab}=\mathcal{I}_{\varepsilon b}+(\mathcal{I}_{ab}-\mathcal{I}_{\varepsilon b})$. The term $\mathcal{I}_{\varepsilon b}$ is shown to be $\mathrm{o}(t^{1/2})$ in Lemma A.2 of Chapter 2. (In the published version (Nielsen and Sohkanen 2011), the corresponding statement contains a typo, and it is not necessary to assume $\dim(\mathbf{W})=0$). The term $\mathcal{I}_{ab}-\mathcal{I}_{\varepsilon b}$ is conveniently split in two terms $\mathcal{T}_1$, $\mathcal{T}_2$ say noting that

$$\frac{1}{f_s^2}=\frac{1}{1+A_s+C_s}-\frac{B_s}{(1+A_s+C_s)(1+A_s+B_s+C_s)}. \quad \text{(A.5)}$$

Now, $\mathcal{T}_1=\sum_{s=1}^t g_s b_s$, where $g_s=(a_s-\varepsilon_s)/(1+A_s+C_s)$ is $\mathcal{F}_{s-1}$-measurable and b_s is an $\mathcal{F}_s$-martingale difference, so that $\sum_{s=1}^t g_s b_s$ is an $\mathcal{F}_s$-martingale.

Therefore, by Hall and Heyde (1980, Thm 2.18), we have $\mathcal{T}_1 = \mathrm{o}(t^{1/2})$ on the set where

$$\mathcal{S} = \sum_{s=1}^{\infty} \mathsf{E}(s^{-1} g_s^2 b_s^2 | \mathcal{F}_{s-1}) \leq \sum_{s=1}^{\infty} s^{-1}(a_s - \varepsilon_s)^2 \mathsf{E}(b_s^2 | \mathcal{F}_{s-1}) < \infty \tag{A.6}$$

because $(1 + A_s + B_s) \geq 1$. To see that the last inequality holds, note first that

$$\sum_{u=1}^{s-1} \varepsilon_u \xi_u^{(2)\prime} = \mathrm{O}\{(s \log\log s)^{1/2}\}, \quad (\sum_{u=1}^{s-1} \xi_u^{(2)} \xi_u^{(2)\prime})^{-1} = \mathrm{O}(s^{-1}), \quad \xi_s^{(2)} = \mathrm{o}(s^{1/4}), \tag{A.7}$$

$a.s$ by Nielsen (2005, Thms. 2.4, 5.1, 6.1) Assuming A. Therefore, $\mathsf{E}(b_s^2 | \mathcal{F}_{s-1}) = \mathrm{O}(s^{-1} \log\log s)$. Next,

$$\sum_{s=1}^{t-1} \varepsilon_s W_{s-1}' (\sum_{s=1}^{t-1} W_{s-1} W_{s-1}')^{-1/2} = \mathrm{o}(t^{1/4-2\eta}), \quad (\sum_{s=1}^{t-1} W_{s-1} W_{s-1}')^{-1/2} W_{t-1} = \mathrm{O}(1), \tag{A.8}$$

$a.s.$ for some $\eta > 0$ by Nielsen (2005, Thm. 2.4, Cors. 5.3, 7.2). Therefore, $(a_s - \varepsilon_s)^2 = \mathrm{o}(s^{\frac{1}{2}-4\eta})$ $a.s.$

Now, $\mathcal{T}_2 = \sum_{s=1}^{t} h_s k_s$, where $h_s = (a_s - \varepsilon_s) b_s$ and k_s is the second component of (A.5), that is, $B_s / \{(1 + A_s + C_s)(1 + A_s + B_s + C_s)\}$. Note that k_s is bounded by $B_s = \mathrm{o}(s^{-1/2})$, given (A.7), and the fact that the denominator is larger than one. Thus $\mathcal{T}_2 = \sum_{s=1}^{t} \mathrm{o}(s^{-1/2}) h_s$. But by (A.8), $a_s - \varepsilon_s \overset{a.s.}{=} \mathrm{o}(s^{1/4-2\eta})$ and further by (A.7), $b_s = \mathrm{o}(s^{\frac{1}{4}+\eta})$, so $(a_s - \varepsilon_s) b_s \overset{a.s.}{=} \mathrm{o}(s^{-\eta})$ and $\mathcal{T}_2 = \mathrm{o}(t^{1/2-\eta})$. □

Lemma A.5. *Assume A and B. Then, $\mathcal{I}_{ac} = -\mathcal{J}_t + \mathrm{o}(t^{1/2})$ a.s.*

Proof of Lemma A.5. Write $\mathcal{I}_{ac} = \mathcal{I}_{\varepsilon c} + (\mathcal{I}_{ac} - \mathcal{I}_{\varepsilon c})$ and note $-\mathcal{J}_t = (\mathcal{I}_{ac} - \mathcal{I}_{\varepsilon c})\{1 + \mathrm{o}(1)\}$. We have from Lemma A.5. of Chapter 2 that $\mathcal{I}_{\varepsilon c} = \mathrm{o}(t^{1/2})$. □

Lemma A.6. *Assume A, B, and C. Then $\mathcal{I}_{aa} - \mathcal{I}_{\varepsilon\varepsilon} = \sum_{s=1}^{t} \frac{\varepsilon_s^2 A_s}{1 + A_s} \mathbf{1}_{(dim(\mathbf{W})>0)} + \mathrm{o}(t^{1/2})$ a.s.*

Proof of Lemma A.6. We first show that $\sum_{s=1}^{t}\{a_s^2/(1+A_s) - a_s^2/f_s^2\} = \mathrm{o}(t^{1/2})$ $a.s.$ The expression of interest satisfies

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

where the inequality follows since $A_s, B_s, C_s \geq 0$. By Lemma A.4 (iii) of Chapter 2 and Nielsen (2005, Thm. 7.1), then $\sum_{s=1}^{t}(B_s+C_s) = \mathrm{O}(\log t)$ $a.s.$ Moreover, $a_t = \mathrm{o}(t^{1/4-\eta})$ $a.s$ for some $\eta > 0$ since $\sum_{s=1}^{t-1}\varepsilon_s W_{s-1}'(\sum_{s=1}^{t-1} W_{s-1}W_{s-1}')^{-1/2}$ and ε_t are $\mathrm{o}(t^{1/4-\eta})$ while $(\sum_{s=1}^{t-1} W_{s-1}W_{s-1}')^{-1/2}W_{t-1}$ converges by Nielsen (2005, Thms. 2.4, 5.1, Cors. 5.3, 7.2) given Assumptions A, B, and C. Combining this with Lemma A.1, we may write

$$\begin{aligned}\mathcal{I}_{aa} - \mathcal{I}_{\varepsilon\varepsilon} &= \sum_{s=1}^{t}(\frac{a_s^2-\varepsilon_s^2}{1+A_s}) = \sum_{s=1}^{t}(\frac{a_s^2}{1+A_s} - \varepsilon_s^2) + \sum_{s=1}^{t}\varepsilon_s^2(1 - \frac{1}{1+A_s}) \\ &= \sum_{s=1}^{t}\frac{\varepsilon_s^2 A_s}{1+A_s} + \mathrm{o}(t^{1/2})\ a.s.\end{aligned}$$

where the last equality follows from Lemma A.7. of Chapter 2, and as A_s is absent in the case where $\dim(\mathbf{W}) = 0$, the leading term is absent when $\dim(\mathbf{W}) = 0$. □

Proof of Lemma A.2. First, apply Lemma A.3. The terms $\mathcal{I}_{bb}$, $\mathcal{I}_{cc}$, and $\mathcal{I}_{bc}$ are of order $\mathrm{o}(t^{1/2})$ as the presence of the extra term A_t in f_t^2 in the mixed case does not affect the validity of Lemma A.3 of Chapter 2. The result then follows by combining this with Lemmas A.4, A.5, and A.6. □

Proof of Theorem 3.1. Combine Lemmas A.1 and A.2. □

The nuisance term $\mathcal{J}_t$

It is argued that $t^{-1/2}\mathcal{J}_t$ can be non-vanishing; see Remark 3.3. For a second order autoregression with one root at zero and one root at $\mathbf{W} > 1$, in the notation of (3.1) we can choose $R_t = e_{R,t} = e_{W,t} = \varepsilon_t$ as $\mathcal{J}_t$ is invariant to scaling of R_t and W_t.

As in the proof of Lemma A.7 of Chapter 2, we define

$$Z_t = \mathbf{W}^{1-t}W_{t-1} = W_0 + \sum_{s=1}^{t-1}\mathbf{W}^{-s}e_{W,s}$$
$$F_t = \sum_{s=1}^{t-1}\mathbf{W}^{1-t}W_{s-1}W'_{s-1}(\mathbf{W}')^{1-t} = \sum_{s=1}^{t-1}\mathbf{W}^{s-t}Z_sZ'_s(\mathbf{W}')^{s-t},$$

with almost sure limits, respectively of $Z = W_0 + \sum_{s=1}^{\infty}\mathbf{W}^{-s}e_{W,s}$ and $F = \sum_{s=1}^{\infty}\mathbf{W}^{-s}ZZ'(\mathbf{W}')^{-u}$ by Lai and Wei (1985, Lem. 2, Cor. 2). Then, we can rewrite

$$A_t = Z'_tF_t^{-1}Z_t \overset{a.s.}{=} Z'F^{-1}Z\{1+\mathrm{o}(1)\} \tag{A.9}$$
$$-(a_t-\varepsilon_t) = \sum_{s=1}^{t-1}\varepsilon_s\mathbf{W}^{s-t}Z'_sF_t^{-1}Z_t \overset{a.s.}{=} \sum_{s=1}^{t-1}\varepsilon_s\mathbf{W}^{s-t}Z'F^{-1}Z\{1+\mathrm{o}(1)\}, \tag{A.10}$$

with exponentially small remainder terms. We next claim that $\mathcal{J}_T = \mathcal{J}_{1,T}\mathcal{J}_{2,T}\{1+\mathrm{o}(1)\}$ $a.s$, where

$$\mathcal{J}_{1,T} = \frac{Z'F^{-1}Z}{1+Z'F^{-1}Z}, \qquad \mathcal{J}_{2,T} = \sum_{t=1}^{T}\frac{1}{t\sigma^2}(\sum_{s=1}^{t-1}\varepsilon_s\mathbf{W}^{s-t})(\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_{s-1})\varepsilon_{t-1}.$$

To verify this, we first use $f_t^2 = 1 + A_t + \mathrm{o}(1)$ from the proof of Lemma A.6 to write $\mathcal{J}_T = -\sum_{t=1}^{T}\{(a_t-\varepsilon_t)c_t/(1+A_t)\}\{1+\mathrm{o}(1)\}$, where

$$c_t = \sum_{s=1}^{t-1}\varepsilon_s\varepsilon_{s-1}(\sum_{s=1}^{t-1}\varepsilon_{s-1}^2)^{-1}\varepsilon_{t-1}$$

since $R_t = \varepsilon_t$. Then, we apply the limits given in (A.9) and (A.10) and the Law of Large Numbers that gives $\sum_{s=1}^{t-1} \varepsilon_{s-1}^2 = t\sigma^2\{1 + \mathrm{o}(1)\}$ $a.s.$ We proceed to analyze the terms $\mathcal{J}_{1,T}$ and $\mathcal{J}_{2,T}$.

The term $\mathcal{J}_{1,T}$ has a continuous distribution by Nielsen (2005, Cor. 5.3); see also Lai and Wei (1983). Therefore, if we show that when the innovations are independently and identically normally distributed and $\mathbf{W} > 1$, $\mathsf{E}(\mathcal{J}_{2,T}^2)$ is the sum of non-negative terms, then for $\mathcal{J}_{2,T}$ to be non-vanishing, it suffices to show that one of those non-negative terms is non-vanishing. From the definition of $\mathcal{J}_{2,T}$, we see that if the innovations are indeed normal and $\mathbf{W} > 1$, then $\mathcal{J}_{2,T}$ it is a triple sum where each element is a product of the form $\varepsilon_{s_1}\varepsilon_{s_2}\varepsilon_{s_2-1}\varepsilon_{t_1}$, that is, a product of four i.i.d. normals and a strictly positive coefficient $\mathbf{W}^{s_1-t_1}/(t_1\sigma^2)$. Then, $\mathsf{E}(\mathcal{J}_{2,T}^2)$ is a six-fold sum of the expectation of eight normals $\mathsf{E}(\varepsilon_{s_1}\varepsilon_{s_2}\varepsilon_{s_2-1}\varepsilon_{t_1}\varepsilon_{s_3}\varepsilon_{s_4}\varepsilon_{s_4-1}\varepsilon_{t_2})$ with strictly positive coefficients $\mathbf{W}^{s_1-t_1}/(t_1\sigma^2)\mathbf{W}^{s_3-t_2}/(t_2\sigma^2)$. By i.i.d. normality, these expectations are either zero or positive, and not all are zero. Therefore, $\mathsf{E}(\mathcal{J}_{2,T}^2) \geq 0$. It then suffices to look at one term and to show that it does not vanish. In particular, our quantity of interest is bounded below by

$$\begin{aligned}
\mathsf{E}(\mathcal{J}_{2,T}^2) &\geq \sum_{t_1<t_2} \frac{\mathbf{W}^{-2}}{t_1 t_2 \sigma^4} \mathsf{E}\{\varepsilon_{t_1-1}^2 \varepsilon_{t_2-1}^2 (\sum_{s=1}^{t_1-2} \varepsilon_s \varepsilon_{s-1})(\sum_{s=1}^{t_2-2} \varepsilon_s \varepsilon_{s-1})\} \\
&= \sum_{t_1<t_2} \frac{\mathbf{W}^{-2}}{t_1 t_2 \sigma^4} \mathsf{E}(\varepsilon_{t_1-1}^2) \mathsf{E}\{(\sum_{s=1}^{t_1-2} \varepsilon_s \varepsilon_{s-1})^2\} \mathsf{E}(\varepsilon_{t_2-1}^2) \\
&= \sum_{t_1<t_2} \frac{\mathbf{W}^{-2}\sigma^4(t_1-2)}{t_1 t_2}
\end{aligned}$$

Note that $\mathbf{W}$ is a scalar in this example. The last sum has leading term $\mathbf{W}^{-2}T\sigma^4$. This shows that $T^{-1/2}\mathcal{J}_T$ is non-vanishing, albeit the convergence rate has not been established.

Chapter 4

Practical implementation of the CUSUM of squares test in non-stationary time series[1]

1 Introduction

The asymptotic behaviour of the CUSUM of squares tests was investigated in Chapter 2, where the problem was solved for a large class of models. However, with the exception of the one-sided CUSUM of squares test under restrictive assumptions (see Brown, Durbin, and Evans (1975)), the exact behaviour of the procedure is not well understood. Reliance on asymptotic theory, as Edgerton and Wells (1994) note, yields significant errors even for quite large sample sizes. These authors compute approximate critical values for larger sample sizes, but only deal with the one-sided test under classical assumptions.

There does not exist a finite sample correction for the CUSUM of squares test

[1] A version of this chapter has been submitted to a peer reviewed journal. At the time of writing, decision by the journal is pending.

that would be valid for a practically interesting model class, such as dynamic linear econometric models with stationary, integrated, and or explosive regressors. Bootstrapping (see e.g. the procedure advocated by Tanizaki (1995)), while a standard numerical solution to finite sample problems, becomes impractical once several tests are conducted. This is often the case in explorative analysis. In this chapter, we derive a simple finite sample correction for the AR(1) case through response surface analysis using a data generation process with normal innovations. Simulations suggest that this will also apply approximately in higher order autoregressions with deterministic terms, as well as in autoregressive distributed lag models, whether the location of the characteristic roots implies stationary, integrated, or explosive behaviour for the process.

The correction we suggest is a simple function of the sample size, and thus readily implementable into software. The relevance of this is underlined by the recent interest in robustifying specification test procedures to quite general forms of global non-homoscedasticity (see e.g. Phillips and Xu (2006), Cavaliere, Rahbek, and Taylor (2010), Cavaliere, Harvey, Leybourne, and Taylor (2011)). Reliable diagnostic procedures to test for the presence of this phenomenon are called for, and the CUSUM of squares test is strong candidate. Furthermore, as will be shown in Chapter 5, use of the CUSUM of squares test preceding inference is asymptotically costless in the sense that it is stochastically independent of hypothesis tests concerning the autoregressive parameters.

Existing finite sample distribution theory and approximations are reviewed in §2 to give the reader an overview of the problem. A simulation experiment giving the empirical quantiles of the tests as a function of the sample size is described and summarized in §3. In §4 ,we explain the finite sample correction methodology and summarize the response surface regressions that are used to

derive the suggestion for a correction. We then illustrate the use of the correction in autoregressions and autoregressive distributed lag (ADL) models in §5. §6 concludes and additional Monte Carlo evidence is presented in the appendix.

2 The problem and some existing solutions

Recall from Chapters 2 and 3 the distinction between the CUSUM of squares plot and the CUSUM of squares test. The latter can be thought of in terms of a binary variable indicating whether the former crosses simultaneous confidence bands. Exact distribution of the plot is available under strong assumptions, but for the construction of a two-sided test, an approximation is necessary. Only asymptotic results are available for the CUSUM squares test statistic, the supremum of the plot over the sample for which variance constancy is being retrospectively tested. Use of the asymptotic approximation results in significant errors, and a correction is therefore required.

Consider the scalar classical regression setting, where, for $t = 1, \ldots, T$, a dependent variable y_t is regressed on a vector of strongly exogenous regressors x_t

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

The parameters β are estimated recursively by OLS which gives

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

where M is a positive integer growing with $\dim(\beta)$, along with the recursive

standardized one-step-ahead forecast residuals

$$\tilde{\varepsilon}_t = \{1 + x_t'(\sum_{s=1}^{t-1} x_s x_s')^{-1} x_t\}^{-1/2}(y_t - \hat{\beta}_{t-1}' x_t) \qquad \text{for } t > M > 0$$

When $\varepsilon_t \sim NI(0, \sigma^2)$, an immediate consequence of this setting is that $(\tilde{\varepsilon}_t/\sigma)^2 \sim \chi^2(1)$. The CUSUM of squares plot is defined as

$$\begin{aligned} s_t &= \left(\sum_{j=M+1}^{T} \tilde{\varepsilon}_j^2\right)^{-1} \left(\sum_{j=M+1}^{t} \tilde{\varepsilon}_j^2\right) \\ &= \left(\sum_{j=M+1}^{t} \tilde{\varepsilon}_j^2 + \sum_{j=t+1}^{T} \tilde{\varepsilon}_j^2\right)^{-1} \left(\sum_{j=M+1}^{t} \tilde{\varepsilon}_j^2\right) \end{aligned}$$

where $\sum_{j=M+1}^{t}(\tilde{\varepsilon}_j/\sigma)^2 \sim \chi^2(t-M)$ and $\sum_{j=t+1}^{T}(\tilde{\varepsilon}_j/\sigma)^2 \sim \chi^2(T-t)$ such that s_t is beta-distributed with mean given by $(t-M)/(T-M)$, and with variance $(t-M)(T-t)/2(T-M)^2(T-M-1)$. This information is sufficient to construct a one-sided test, by choice of a critical value c_α giving the rejection region

$$s_t - \mathsf{E}(s_t) = \left(\sum_{j=M+1}^{T} \tilde{\varepsilon}_j^2\right)^{-1} \left(\sum_{j=M+1}^{t} \tilde{\varepsilon}_j^2\right) - \frac{t-M}{T-M} > c_\alpha$$

for $t = M+1, \ldots, T$ and a significance level α.

In practice, a two-sided procedure is often required, and it is this problem that Brown, Durbin, and Evans (1975) set out to solve. No exact results are available, however, and these authors resort to an approximation. The idea is to find a pair of lines $s_t = \pm c_0 + (t-M)/(T-M)$ such that the probability that the sample path of s_t crosses one or both lines is α, the required significance level. The maximum positive and negative deviations of the s_ts from the mean value

line can be written as

$$c^{+} = \max_{i=1,\ldots,T-M-1} \left(s_{M+i} - \frac{i}{T-M} \right) \quad \text{and} \quad c^{-} = \max_{i=1,\ldots,T-M-1} \left(\frac{i}{T-M} - s_{M+i} \right)$$

and Brown, Durbin, and Evans (1975) suggest approximating these by Pyke (1959)'s modified Kolmogorov-Smirnov statistic

$$C^{+} = \max_{t} \left(U_{\tilde{t}} - \frac{t-M}{T-M} \right) \quad \text{and} \quad C^{-} = \max_{t} \left(\frac{t-M}{T-M} - U_{\tilde{t}} \right)$$

where $U_{\tilde{t}}$ is the $\frac{1}{2}(t-M)^{th}$ order statistic from a draw of $\frac{1}{2}(T-M)-1$ independent random variables X indexed by t, so $\mathsf{P}(X_t \leq x) = x$ for $0 \leq x \leq 1$. Durbin (1969) tabulates values of C^{+} for $\alpha = \{0.10, 0.05, 0.025, 0.01, 0.005\}$ for sample sizes up to 200. This approximation works since if we draw $1/2(T-M)-1$ observations denoted $x_t, t = 1, \ldots, 1/2(T-M)-1$ from the standard uniform distribution, then the $1/2(t-M)^{th}$ order statistic u has density

$$\begin{aligned} f_U(u) &= \frac{\left(\frac{1}{2}(T-M)-1\right)!}{\left(\frac{1}{2}(T-M)-1\right)!\left(\frac{1}{2}(T-t)-1\right)!} u^{\frac{1}{2}(t-M)-1}(1-u)^{\frac{1}{2}(T-t)-1} \\ &= \frac{\Gamma\left(\frac{1}{2}(T-M)\right)}{\Gamma\left(\frac{1}{2}(t-M)\right)\Gamma\left(\frac{1}{2}(T-t)\right)} u^{\frac{1}{2}(t-M)-1}(1-u)^{\frac{1}{2}(T-t)-1} \end{aligned}$$

i.e. the density of a beta$\left(\frac{1}{2}(T-M), \frac{1}{2}(T-t)\right)$-distributed random variable.

The use of standardized forecast errors is motivated by their simple probabilistic properties in the classical setting. In such a setting, it also possible to implement the test with non-standardized ordinary least squares residuals, although an exact distribution is not available even for the one-sided test. Using theory developed by Durbin and Watson (1950), McCabe and Harrison (1980) derived the finite sample distribution of upper and lower bounding variables to the CUSQ test constructed from ordinary least squares residuals, and found that

these bounding variables are beta-distributed.

We are interested in retrospective variance constancy testing. In other words, we want to know whether the null of variance constancy is true in the entire sample under scrutiny. In particular, we are not interested in how many times a plot crosses boundaries, but whether or not it does. Therefore, simultaneous confidence bands serve our purpose, and we want to look at the supremum of the plot and compare that with the supremum of its distribution under the null. Once the fixed regressors assumption is dropped, the finite sample distribution theory becomes to a large extent intractable. Some restrictive results are available, for instance in Dufour and Kiviet (1995). However, in the quest for a procedure that is valid in an economically interesting class of linear models, one has to resort to asymptotic theory. In the literature, generality has been sought in different directions, and the most general available results on the CUSQ test are those of Deng and Perron (2008a) who impose strong mixing assumptions, require the existence of the eight unconditional moments, and do not allow for integrated data generation processes, and those of Chapter 2, where we work in the context of an unstable vector autoregression, imposing conditional moment assumptions which rule out autoregressive conditional heteroscedasticity in the disturbances, but only require the existence of fourth conditional moments. Under both sets of assumptions, the test viewed as a process in the appropriate function space exhibits scaled Brownian bridge asymptotics, such that a supremum transformation of the studentized version results in a statistic with the Kolmogorov distribution with analytically computable quantiles. The normalization necessary to achieve this behaviour is

$$\sqrt{T}\frac{\sigma^2}{\varphi}\left\{\left(\sum_{j=M+1}^{T} w_j^2\right)^{-1}\left(\sum_{j=M+1}^{t} w_j^2\right)-\frac{t}{T}\right\},$$

where the w_j are either the recursive residuals or the OLS residuals. In our simulations, the innovation process will be generated from the normal distribution which implies $\varphi/\sigma^2 = \sqrt{2}$. Note that the limit of the supremum of this CUSUM of squares statistic is the same as that first derived by Bartlett (1955) for the i.i.d. case and again by Inclán and Tiao (1994). Furthermore, and unlike in the case of the CUSUM test, the distribution does not depend on whether OLS or recursive residuals are used. While the OLS version is computationally simpler and has a nice interpretation in terms of in-sample variance fluctuations, the recursive residuals based procedure has a nice interpretation also, namely as a simulation of the forecasting performance of the model. We therefore look at both tests, defining

$$\sup_{M\leq t\leq T} |\mathrm{CUSQ}_{t,T}^{OLS}| = \sup_{M\leq t\leq T} \left| \sqrt{T/2} \left(\frac{\sum_{s=M}^{t} \hat{\varepsilon}_{s,t}^2}{\sum_{s=M}^{T} \hat{\varepsilon}_{s,t}^2} - \frac{t}{T} \right) \right| \tag{2.1}$$

$$\sup_{M\leq t\leq T} |\mathrm{CUSQ}_{t,T}^{REC}| = \sup_{M\leq t\leq T} \left| \sqrt{T/2} \left(\frac{\sum_{s=M}^{t} \tilde{\varepsilon}_{s}^2}{\sum_{s=M}^{T} \tilde{\varepsilon}_{s}^2} - \frac{t}{T} \right) \right| \tag{2.2}$$

where the degrees of freedom correction in the recursive residuals based test has been omitted as it is asymptotically irrelevant.

Reliance on asymptotic theory, as Edgerton and Wells (1994) note, yields significant errors even for quite large sample sizes. These authors compute critical values for sample sizes larger than those considered by Durbin (1969), but only deal with the one-sided test under classical assumptions. In this chapter, we set out to correct the asymptotically valid test directly. To illustrate the idea underlying the finite sample correction, consider the simplest possible dynamic setting with an endogenous regressor, the AR(1) model with zero mean, and, using Ox

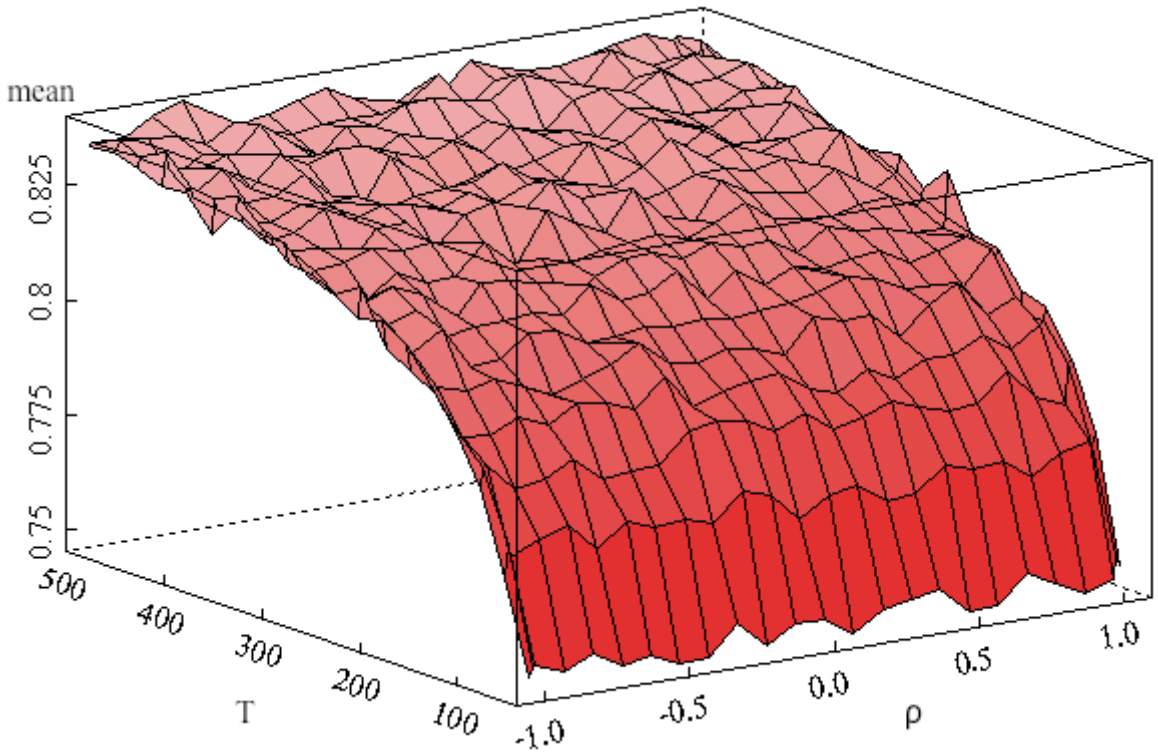


Figure 2.1: Monte Carlo response surface for means of $CUSQ^{OLS}$ test with 10^5 replications for each sample size/persistence combination

(Doornik 2007), compute the mean of the CUSQ test with OLS residuals from a Monte Carlo experiment for 19 sample sizes and 25 values of the autoregressive parameter ranging from the negative explosive, negative unit root, through stationary to positive unit root and the positive explosive region. Ten thousand repetitions are taken for each combination of sample size and persistence parameter, and the results are plotted in Figures 2.1 and 2.2. We observe that most of the variation takes place along the time dimension, and this is promising from the viewpoint of finite sample correction. The nuisance parameter can be ignored in the design of the correction. The reason why we shall talk about near invariance is that, judged against the simulation error in experiments with a higher number of replications, such as the pilot study reported in Chapter 2 using $M = 10^6$, the variation in the mean and median of the statistic are significant.

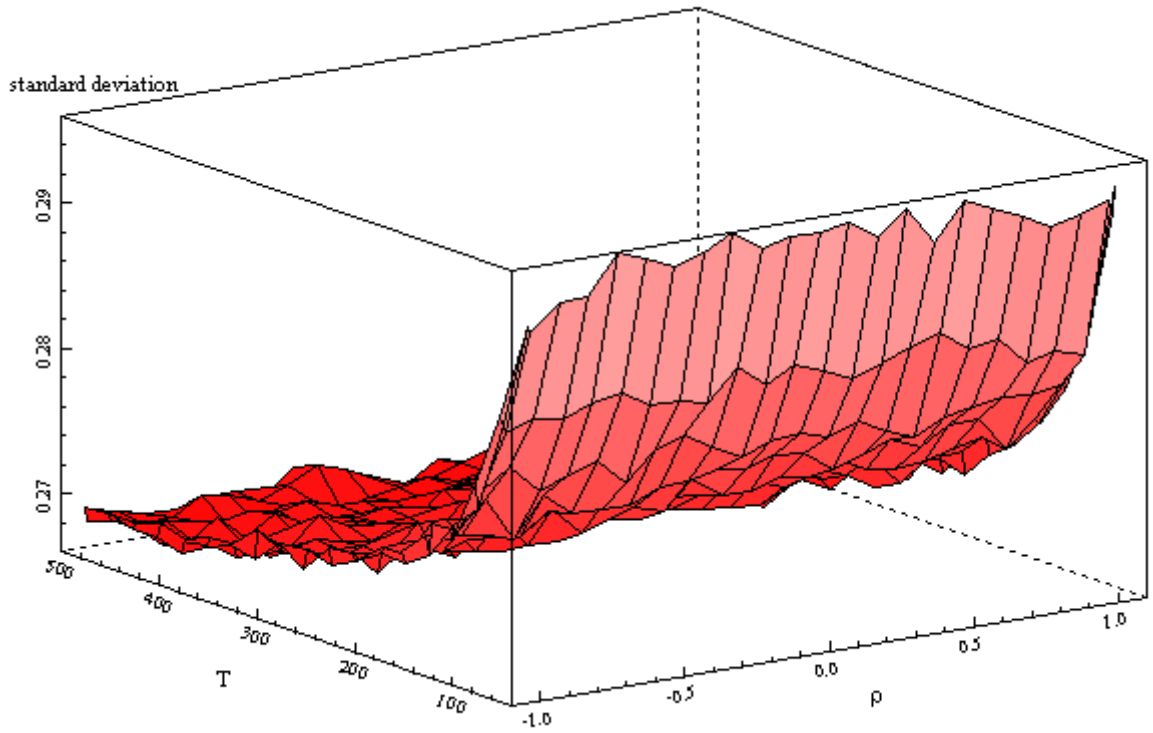


Figure 2.2: Monte Carlo response surface for standard deviations $CUSQ^{OLS}$ test with 10^5 replications for each sample size/persistence combination

3 Monte Carlo evidence on the size distortion

There are two separate questions of interest. The primary question is what happens as the sample size grows. That is, how quickly the empirical quantiles of the CUSUM of squares tests defined in (2.1) and (2.2) approach the asymptotic quantiles. A secondary question is whether there is any difference between the test (2.1), implemented with OLS residuals, and the test (2.2) based on recursive residuals. Some light is thrown on these questions by Figures 3.1 and 3.2, where we report results of a Monte Carlo experiment where the residuals on which the tests are based are derived from an AR(1) with zero persistence and normally distributed innovations. That is, the model from which residuals are computed is

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathsf{N}(0, \sigma^2), \quad \text{for } t = 1, \ldots, T,$$

and in the data generation process we set $x_0 = 0$ and $\rho = 0$. The number of replications is 10^5 and the Ox default seed is used for the random number generator; see Doornik (2007, p.191). In Figure 3.1, we report the variation in the empirical quantiles of the CUSUM of squares test (2.1) based on OLS residuals as a function of the sample size for a fixed point in the parameter space. The observed quantiles are plotted against T^{-1} in order to better illustrate convergence to asymptotic quantiles. The plots are therefore downward sloping (the finite sample distortion is negative). While locally approximately linear, globally they look somewhat more quadratic. We see that even for rather large sample sizes, there is a considerable distance between the asymptotic and the empirical quantiles. Note that more noise is observed for higher quantiles. More Monte Carlo repetitions could be required when generating the empirical quantiles. However, if the number of repetitions were increased, we would still run into the same problem of a difference in relative precision between quantiles when moving to higher and higher quantiles. As the purpose of this correction is a very practical one, and since subsequent investigations in this chapter show that the suggested correction works satisfactorily, we regard this as a minor problem. The relatively high noise in higher quantiles will, however, be reflected in relatively low R^2 in the response surface regressions. The R^2 does not rise appreciably when more powers of T are added to the design matrix.

In Figure 3.2, the differences in the observed quantiles of the OLS based test (2.1) and those of the recursive residuals based test (2.2) are reported. Noise is again observed in the case of the higher quantiles, but this appears to die out rather quickly as T grows, and there appears not to be much difference between the empirical quantiles of the two tests. This observation will also be reflected in the results of the response surface regressions reported in §4.1.

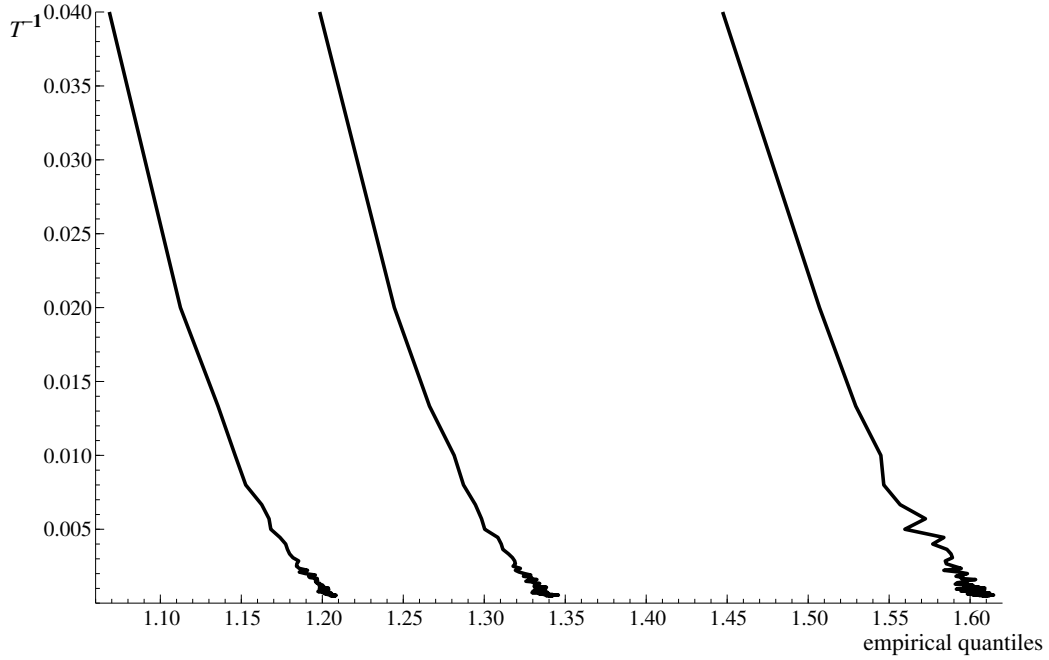


Figure 3.1: Empirical 90%, 95%, and 99% quantiles of the uncorrected $\sup CUSQ^{OLS}$ test on residuals from an AR(1) with $\alpha = 0$ and $T = \{25, 50, \ldots, 1975, 2000\}$

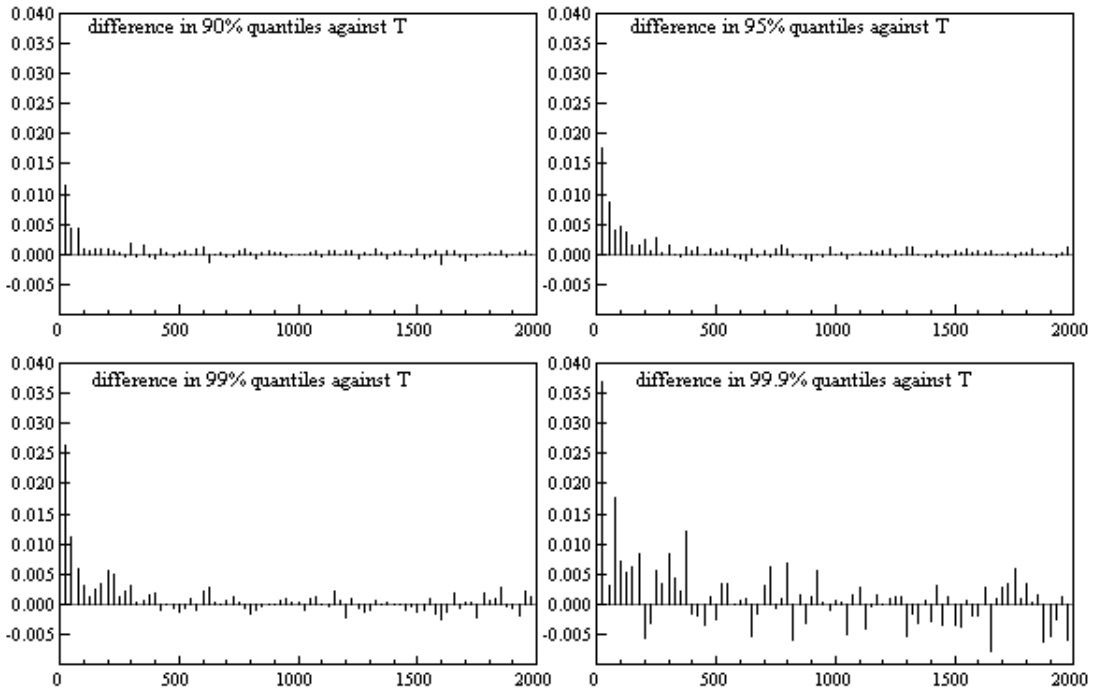


Figure 3.2: Differences in 90%, 95%, 99%, and 99.9% quantiles of the uncorrected $\sup CUSQ^{OLS}$ and $\sup CUSQ^{REC}$ tests on residuals from an AR(1) with $\alpha = 0$

4 Suggested finite sample correction

The evidence presented in the preceding section indicates under-rejection of the null, and rather slow convergence of the quantiles. In this section, we derive a finite sample correction based on the empirical quantiles reported in the preceding section. This is done using a response surface that relates a measure of departure of the observed quantiles from the asymptotic quantiles and a simple function of the sample size. The response surface is fitted to the test constructed with an AR(1) model with the true autoregressive parameter zero. Later we show that this correction, derived on the basis of i.i.d. data, works satisfactorily when the test is constructed with residuals from second order autoregressions, or autoregressive distributed lag models, whether stationary or unstable.

A possible explanation for why the correction works well across the parameter space for a fixed sample size is that either any bias in the autoregressive coefficient estimate does not vary wildly with the value of the true coefficient or any large distortion from true innovations in the numerator (sum of squared residuals up to $1, \ldots t$) of the test is offset partly by a similarly large distortion in the denominator (sum of squared residuals up to T). An exception to the good properties of the correction appears in the unit root case with deterministic components, where over-rejection of the null of constancy occurs. It is not clear what causes this, as the asymptotic theory gives no guidance here and it has not been possible to suggest a finite sample theory. This is in fact why the simulations are conducted. A modeller who suspects a unit root and includes a deterministic trend into his model should be cautious if using the suggested correction, or should not use it at all. Note that this distortion does not arise in the ADL simulations as deterministic terms are not included there.

4.1 Response surface regressions for the AR(1)

Let $q_T(p)$ represent the true unknown quantile function of the test for sample size t, with $q_\infty(p)$ the quantile function of the limit distribution. We have 80 observations on $q_T(p)$ for each $p \in \{0.90, 0.95, 0.99, 0.999\}$ from the Monte Carlo simulations, for T ranging from 20 to 2000. A response surface could be fitted by regressing $y_T(p) = q_T(p)/q_\infty(p) - 1$ on some powers of T. We follow the strategy adopted in previous literature and choose $\{T^{-1/2}, T^{-1}, T^{-3/2}, \ldots, T^{-N/2}\}$ for a fixed $N < \infty$; see e.g. Edgerton and Wells (1994) . The finite sample quantiles could then be found by computing $\hat{q}_T(p)$. For the recursive forecast residuals based test, fixing $N = 6$, and estimating by ordinary least squares, it appears that the regressors T^{-3} and $T^{-5/2}$ may be excluded in all cases.

$$\hat{y}_T^{\mathrm{REC}}(0.90) = \underset{(0.0125)}{-\ 0.458}\ T^{-1/2} \underset{(0.279)}{-\ 2.132}\ T^{-1} + \underset{(1.19)}{5.435}\ T^{-3/2}$$

$$\hat{y}_T^{\mathrm{REC}}(0.95) = \underset{(0.0308)}{-\ 0.6829}\ T^{-1/2} + \underset{(1.19)}{3.137}\ T^{-1} \underset{(12.4)}{-\ 33.85}\ T^{-3/2} + \underset{(36.5)}{93.93}\ T^{-2}$$

$$\hat{y}_T^{\mathrm{REC}}(0.99) = \underset{(0.048)}{-\ 0.6712}\ T^{-1/2} + \underset{(1.85)}{4.881}\ T^{-1} \underset{(19.3)}{-\ 48.31}\ T^{-3/2} + \underset{(56.8)}{123.5}\ T^{-2}$$

$$\hat{y}_T^{\mathrm{REC}}(0.999) = \underset{(0.125)}{-\ 0.8566}\ T^{-1/2} + \underset{(4.83)}{12.91}\ T^{-1} \underset{(50.3)}{-\ 123.2}\ T^{-3/2} + \underset{(148)}{322}\ T^{-2}$$

It seems that the further we are in the tails of the distribution, the slower is convergence in T. The R^2s are 0.99, 0.99, 0.97, and 0.82, respectively. The results for the OLS based test are as follows.

$$\hat{y}_T^{\mathrm{OLS}}(0.90) = \underset{(0.0126)}{-\ 0.4581}\ T^{-1/2} \underset{(0.281)}{-\ 2.072}\ T^{-1} + \underset{(1.2)}{6.309}\ T^{-3/2}$$

$$\hat{y}_T^{\mathrm{OLS}}(0.95) = \underset{(0.0298)}{-\ 0.6795}\ T^{-1/2} + \underset{(1.15)}{3.075}\ T^{-1} \underset{(12)}{-\ 29.79}\ T^{-3/2} + \underset{(35.3)}{82.84}\ T^{-2}$$

$$\hat{y}_T^{\mathrm{OLS}}(0.99) = \underset{(0.0475)}{-\ 0.6873}\ T^{-1/2} + \underset{(1.83)}{5.623}\ T^{-1} \underset{(19)}{-\ 52.91}\ T^{-3/2} + \underset{(56.2)}{140}\ T^{-2}$$

$$\hat{y}_T^{\mathrm{OLS}}(0.999) = \underset{(0.119)}{-\,0.9394}\; T^{-1/2} + \underset{(4.59)}{16.39}\; T^{-1} - \underset{(47.8)}{156.1}\; T^{-3/2} + \underset{(141)}{421.6}\; T^{-2}$$

The R^2s are 0.99, 0.99, 0.97, and 0.78, respectively.

We have described the divergence of the finite sample and asymptotic quantiles by a function of the sample size. Note the similarity of the correction for the two different tests. The approximate finite sample critical value for a $100p\%$ confidence level test is given by $\hat{q}_T(p) = q_\infty(p)(\hat{y}_T(p) + 1)$, where e.g. $q_\infty(0.10) \approx 1.22$, $q_\infty(0.05) \approx 1.36$, and $q_\infty(0.01) \approx 1.63$; see e.g. Schumacher (1984, Tab. 9). We now check the performance of the correction in autoregressions as well as ADLs.

5 Implementation of the suggested correction

5.1 Implementation in autoregressions

Using the suggested finite sample correction we report in Figures 5.1 and 5.2 the empirical rejection frequencies, implied by use of 90% 95% and 99% critical values respectively, of the uncorrected and the corrected CUSUM of squares tests implemented with OLS residuals from AR(1) and AR(2) specifications, with normal errors, zero initial values, and a range of values for the autoregressive parameters. We consider models without any deterministic terms, as well as models with a constant and models with both a constant and a linear trend. In the data generation process, however, the coefficients on the deterministic terms are set to zero in all these cases to ensure that deterministic trends do not dominate stochastic trends, so that we can observe any effects of those stochastic trends. For a fixed sample size of 100 we make 10^5 replications, again using Ox (Doornik 2007).

When varying the autoregressive coefficients in the data generation process,

we move in the parameter space such that points in the stationary, integrated, and explosive regions are covered. Each tick on the horizontal axes of Figures 5.1 and 5.2, corresponds to a point in the parameter space. In the AR(2) case, there is no immediate way to order the points, and we have chosen to order them according to the maximal eigenvalue of the companion matrix of their companion form. We note that there is little variation in most of the parameter space and the correction works well, but some size distortion occurs in the unit root I(1) case when a linear trend is also estimated in the model. This results in over-rejection of the mis-specification null, consistent with a conservative modelling strategy. While the size-performance of the test based on the one-step ahead forecast errors is slightly inferior to that based on OLS residuals, the difference is very small and we only report results for the OLS based test graphically. We refer to the appendix for full tables of results.

5.2 Implementation in ADLs

We also consider an ADL(1,1) specification under weak exogeneity, understood in the sense of Engle, Hendry, and Richard (1983). The data generation process is

$$\begin{pmatrix} Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} (\alpha+\rho\omega) & (\beta+\rho\lambda) \\ \omega & \lambda \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Z_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & \rho \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{Y,t} \\ \varepsilon_{Z,t} \end{pmatrix}$$
$$\begin{pmatrix} \varepsilon_{Y,t} \\ \varepsilon_{Z,t} \end{pmatrix} \sim \mathsf{N}_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{YY} & \sigma_{YZ} \\ \sigma_{YZ} & \sigma_{ZZ} \end{pmatrix} \right\},$$

where the covariance matrix has zero off-diagonal elements. This is equivalent to writing an equation for the first process, $Y_t = \rho Z_t + \alpha Y_{t-1} + \beta Z_{t-1} + \varepsilon_{Y,t}$, and for the second process $Z_t = \lambda Z_{t-1} + \omega Y_{t-1} + \varepsilon_{Z,t}$, and specifying that $\varepsilon_{Y,t}$ and

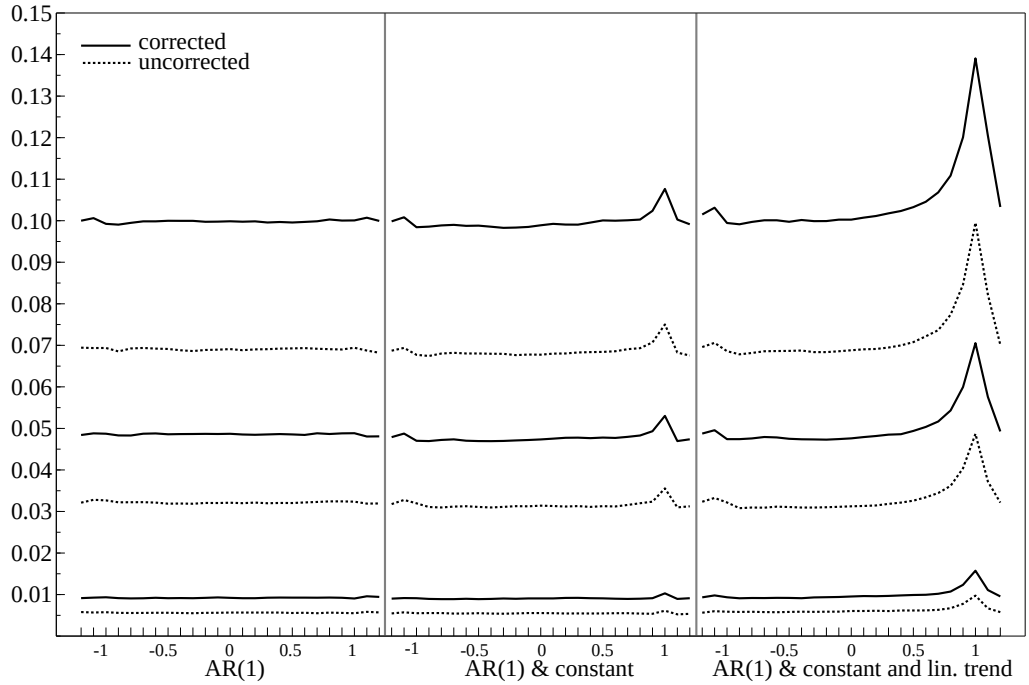


Figure 5.1: Rejection frequencies of the corrected and uncorrected $\sup CUSQ^{OLS}$ test with residuals from AR(1), $T = 100$, $\alpha = \{-1.2, -1.1, \ldots, 0, \ldots 1.1, 1.2\}$

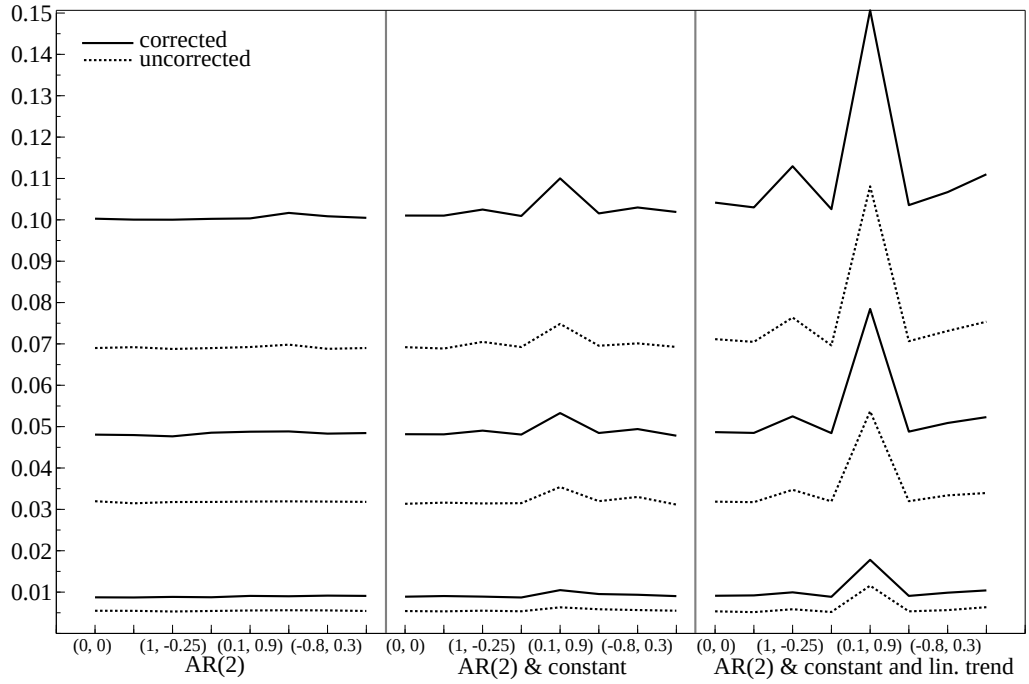


Figure 5.2: Rejection frequencies of the corrected and uncorrected $\sup CUSQ^{OLS}$ test with residuals from AR(2), $T = 100$ and parameters $\{(0,0)\,,\ (0.1,-0.2)\,,\ (1,-0.25)\,,\ (-1,-0.5)\,,\ (0.1,0.9)\,,\ (0,-1)\,,\ (-0.8,0.3)\,,\ (0.25,1)\}$; 1^{st} through 4^{th} ticks are stationary, 5^{th} is is I(1), 6^{th} I(2), 7^{th} and 8^{th} are explosive.

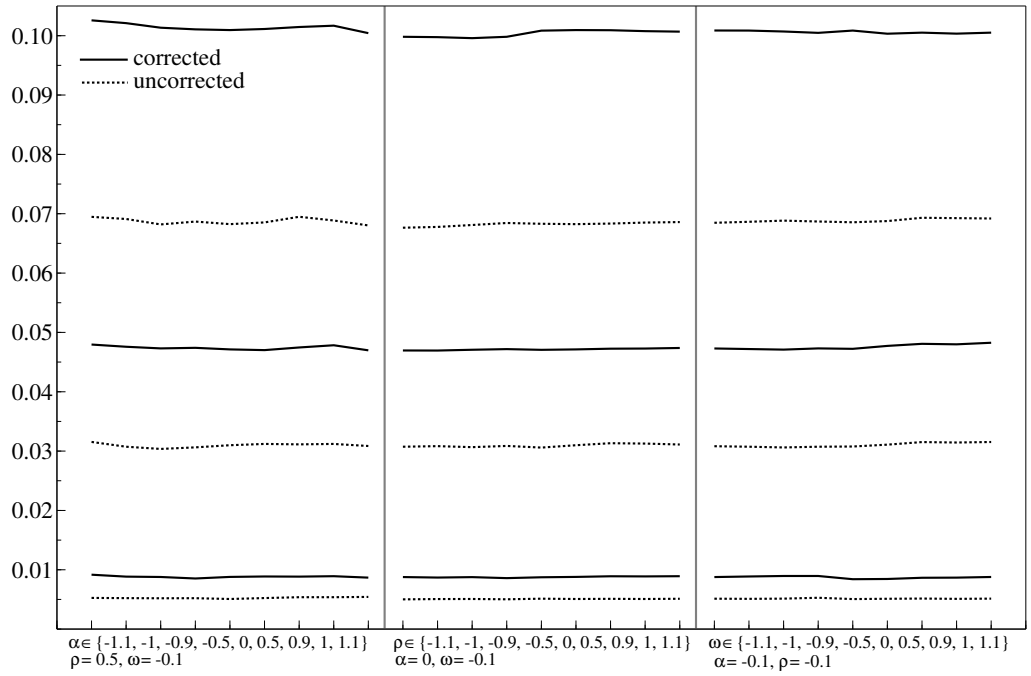


Figure 5.3: Rejection frequencies of the corrected and uncorrected sup $CUSQ^{OLS}$ test with residuals from ADL(1,1), $T = 100$; $\lambda = 1$, $\beta = 0.1$

$\varepsilon_{Z,t}$ are jointly normal and uncorrelated. We generate Y_ts, estimate the ADL, and perform the CUSUM of squares tests, all of this 10^5 times, altering the parameters. The results are presented graphically in Figure 5.3. The correction appears reasonable. The size distortion observed in the pure autoregressive case does not arise because the model does not contain deterministic terms. As for the pure autoregressions, we do not graph results for the forecast error-based test as it differs very little from the OLS based test; see the appendix for tables of these results.

6 Conclusion

There are no exact finite sample results that would enable the use of the CUSUM of squares tests in a model class that would be of interest to macroeconometri-

cians. The asymptotic representation of the test is inadequate because of slow convergence of the quantiles, in other words, the possibility of large errors in inference even in reasonably large samples. In this chapter we address this problem by numerically deriving for the AR(1) model a finite sample correction for the asymptotically valid CUSUM of squares tests implemented either with OLS residuals or standardized OLS prediction errors. We then show, by simulation, that the correction is valid in quite a general class of linear dynamic models. Some size distortion occurs in the positive unit root case when a linear trend is estimated. Future research should look at the empirical size of a non-normality corrected version of the test under fat tailed error distributions. In addition, the behaviour of the tests under residual variance breaks should be investigated in order that a judgment could be made about their size-corrected power. Existing work on the latter problem includes Ploberger (1989); for results on power against changes in regression coefficients, see Deng and Perron (2008b). This would be of relevance in detecting volatility breaks in macroeconomic time series, a well documented phenomenon (see for example Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Blanchard and Simon (2001), Stock and Watson (2002), Sensier and van Dijk (2004)).

A Appendix

Additional Monte Carlo evidence

We report in tabular form some simulation results reported graphically in figures elsewhere in this chapter, as well as additional evidence that may be of interest to practitioners.

Observed size in autoregressions under normality

Tables 4.1 to 4.3 show results from an AR(1) model without deterministics, with a constant, and with a constant and a linear trend, respectively, whereas tables 4.4 to 4.6 show corresponding results for the AR(2) data generating process. Each simulation uses 10^5 repetitions, and the nominal size is set to 5% by choosing as the critical value the 95^{th} percentile of the Kolmogorov distribution, which is 1.36; see Schumacher (1984, Tab. 9).

The results for the OLS test were already given in graphically in Figures 5.1 and 5.2. As we see in Tables 4.1 to 4.6, the behaviour of the CUSUM of squares tests in first and second order autoregressions under normality is quite stable across different designs. A size distortion is observed when a linear trend is estimated in the model and the data generation process includes one stochastic trend but no deterministic trend. A similar distortion is present, albeit a smaller one, in the case where a constant is estimated in the model.

Table 4.1: Simulated statistics for $\sup CUSQ^{OLS}$ (left) and $\sup CUSQ^{REC}$ (right), no constant/trend estimated

	AR(1) parameter value in DGP							AR(1) parameter value in DGP						
T = 100 M = 10^5	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.796	0.796	0.795	0.795	0.795	0.795	0.796	0.797	0.796	0.795	0.795	0.795	0.796	0.797
median	0.756	0.755	0.755	0.756	0.754	0.754	0.756	0.758	0.756	0.756	0.756	0.755	0.755	0.758
95%	1.281	1.282	1.281	1.281	1.282	1.283	1.279	1.279	1.280	1.277	1.277	1.278	1.279	1.277
obs. size	0.032	0.033	0.033	0.033	0.033	0.033	0.032	0.032	0.032	0.032	0.031	0.032	0.032	0.032
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.2: Simulated statistics for $\sup CUSQ^{OLS}$ (left) and $\sup CUSQ^{REC}$ (right), constant, no trend estimated

	AR(1) parameter value in DGP							AR(1) parameter value in DGP						
T = 100 M = 10^5	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.797	0.796	0.795	0.795	0.801	0.808	0.797	0.804	0.802	0.801	0.801	0.808	0.816	0.801
median	0.757	0.756	0.755	0.755	0.762	0.768	0.758	0.766	0.763	0.762	0.762	0.769	0.776	0.762
95%	1.279	1.276	1.275	1.277	1.284	1.297	1.276	1.286	1.284	1.280	1.281	1.291	1.309	1.279
obs. size	0.032	0.032	0.031	0.032	0.033	0.036	0.032	0.033	0.033	0.032	0.032	0.034	0.038	0.032
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.3: Simulated statistics for $\sup CUSQ^{OLS}$ (left) and $\sup CUSQ^{REC}$ (right), constant and trend estimated

	AR(1) parameter value in DGP							AR(1) parameter value in DGP						
T = 100 M = 10^5	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.802	0.799	0.798	0.800	0.830	0.851	0.812	0.816	0.813	0.811	0.809	0.840	0.861	0.817
median	0.762	0.760	0.759	0.760	0.791	0.810	0.772	0.777	0.773	0.771	0.770	0.801	0.821	0.777
95%	1.282	1.277	1.276	1.279	1.322	1.355	1.284	1.299	1.294	1.292	1.291	1.333	1.370	1.288
obs. size	0.033	0.032	0.031	0.032	0.041	0.049	0.033	0.036	0.035	0.034	0.034	0.044	0.053	0.033
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.4: Simulated statistics for $\sup CUSQ^{OLS}$ (left) and $\sup CUSQ^{REC}$ (right), no constant/trend estimated

	AR(2) parameter values in DGP (inverse roots of characteristic polynomial)							AR(2) parameter values in DGP (inverse roots of characteristic polynomial)						
T = 100 M = 10^5	0, 0 (0,0)	1, −0.25 (0.5,0.5)	0.1, 0.9 (1,−0.9)	0.25, 1 (1.13,−0.88)	−0.8, 0.3 (0.28,−1.08)	0.1, −0.2 (c. stat.)	−1, −0.5 (c. expl.)	0, 0 (0,0)	1, −0.25 (0.5,0.5)	0.1, 0.9 (1,−0.9)	0.25, 1 (1.13,−0.88)	−0.8, 0.3 (0.28,−1.08)	0.1, −0.2 (c. stat.)	−1, −0.5 (c. expl.)
mean	0.798	0.798	0.799	0.800	0.801	0.798	0.798	0.801	0.801	0.802	0.804	0.805	0.801	0.801
median	0.759	0.759	0.760	0.761	0.762	0.759	0.759	0.762	0.763	0.764	0.766	0.767	0.762	0.762
95%	1.279	1.278	1.282	1.281	1.280	1.279	1.281	1.279	1.278	1.281	1.280	1.283	1.279	1.280
obs. size	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.033	0.032	0.033	0.032	0.032
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.5: Simulated statistics for $\sup CUSQ^{OLS}$ (left) and $\sup CUSQ^{REC}$ (right), constant, no trend estimated

	AR(2) parameter values in DGP (inverse roots of characteristic polynomial)							AR(2) parameter values in DGP (inverse roots of characteristic polynomial)						
T = 100 M = 10^5	0, 0 (0,0)	1, −0.25 (0.5,0.5)	0.1, 0.9 (1,−0.9)	0.25, 1 (1.13,−0.88)	−0.8, 0.3 (0.28,−1.08)	0.1, −0.2 (c. stat.)	−1, −0.5 (c. expl.)	0, 0 (0,0)	1, −0.25 (0.5,0.5)	0.1, 0.9 (1,−0.9)	0.25, 1 (1.13,−0.88)	−0.8, 0.3 (0.28,−1.08)	0.1, −0.2 (c. stat.)	−1, −0.5 (c. expl.)
mean	0.802	0.805	0.815	0.808	0.805	0.801	0.801	0.809	0.810	0.827	0.817	0.814	0.808	0.809
median	0.763	0.767	0.776	0.771	0.765	0.763	0.762	0.771	0.773	0.788	0.780	0.775	0.770	0.770
95%	1.280	1.283	1.298	1.279	1.285	1.280	1.280	1.285	1.287	1.313	1.284	1.295	1.285	1.286
obs. size	0.032	0.032	0.036	0.032	0.033	0.032	0.032	0.033	0.033	0.039	0.033	0.035	0.033	0.033
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.6: Simulated statistics for $\sup CUSQ^{OLS}$ (left) and $\sup CUSQ^{REC}$ (right), constant and trend estimated

	AR(2) parameter values in DGP (inverse roots of characteristic polynomial)							AR(2) parameter values in DGP (inverse roots of characteristic polynomial)						
T = 100 M = 10^5	0, 0 (0,0)	1, −0.25 (0.5,0.5)	0.1, 0.9 (1,−0.9)	0.25, 1 (1.13,−0.88)	−0.8, 0.3 (0.28,−1.08)	0.1, −0.2 (c. stat.)	−1, −0.5 (c. expl.)	0, 0 (0,0)	1, −0.25 (0.5,0.5)	0.1, 0.9 (1,−0.9)	0.25, 1 (1.13,−0.88)	−0.8, 0.3 (0.28,−1.08)	0.1, −0.2 (c. stat.)	−1, −0.5 (c. expl.)
mean	0.811	0.825	0.866	0.837	0.814	0.810	0.806	0.824	0.835	0.885	0.849	0.831	0.822	0.822
median	0.773	0.788	0.825	0.800	0.776	0.772	0.767	0.787	0.798	0.845	0.813	0.793	0.785	0.784
95%	1.282	1.294	1.373	1.293	1.290	1.281	1.281	1.297	1.306	1.394	1.301	1.308	1.295	1.297
obs. size	0.032	0.035	0.054	0.034	0.034	0.032	0.032	0.035	0.037	0.061	0.036	0.038	0.035	0.036
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

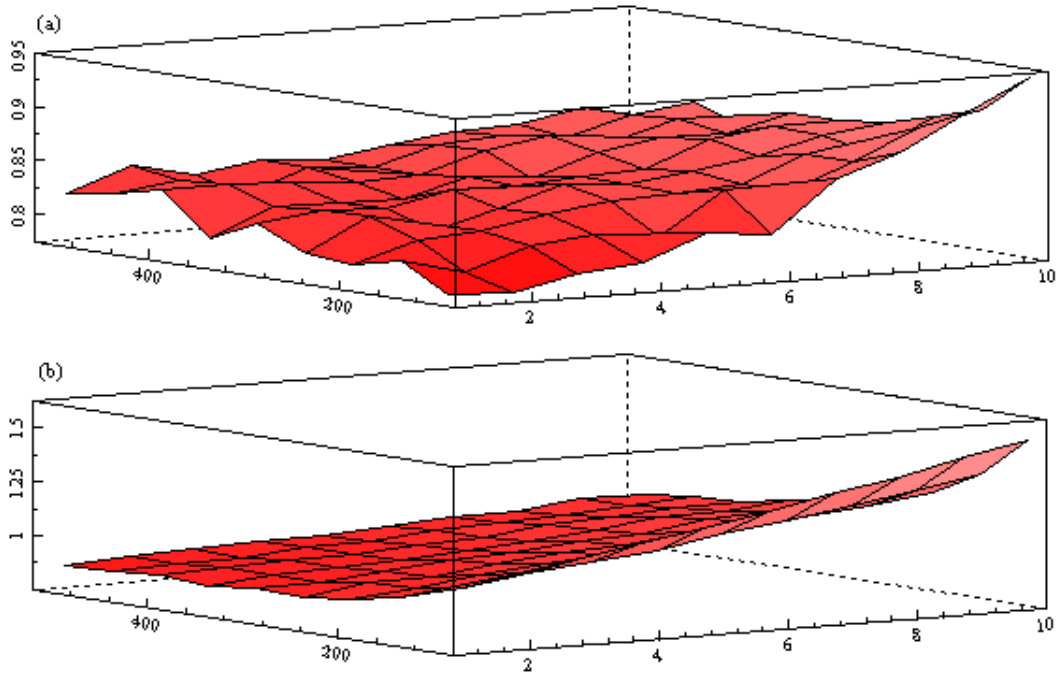


Figure 7.1: Monte Carlo response surface for mean of (a) $\sup CUSQ^{OLS}$ and (b) $\sup CUSQ^{REC}$ with 10^3 replications for each sample size/lag order combination

Higher order autoregressions

When considering higher order autoregressions, the estimation sample must be appropriately extended together with the lag length. We note that, owing to further lags, the estimation uncertainty results in greater finite sample bias in the test. Figure 7.1 illustrates this, with all autoregressive coefficients equal to zero in the data generation process.

Autoregressive distributed lag models

Imposing strong exogeneity, we generate Y_ts, estimate the autoregressive distributed lag model, and perform the CUSUM of squares tests, all of this 10^5 times, altering the regression parameters. The results for three different values of ρ are summarized in Tables 4.7-4.9, those for λ in Tables 4.10-4.12, and for β in Tables 4.13-4.15. The results for the OLS based test were already reported in

graphical form in Figure 5.3. The mechanism that generates the data is given in §5.2, and the parameter values used in the simulations are given in the tables.

Near invariance of the size of the procedure is observed, just as in the pure autoregressive case. Here the models do not include deterministic terms, and the size distortion observed in the purely autogressive case when a stochastic trend was present in the data generation process, and deterministic terms were included in the model, does not arise.

Table 4.7: Simulated statistics for $\sup CUSQ^{OLS}$ and $\sup CUSQ^{REC}$, ADL(1,1): $\beta = 0.5$, $\rho = 0.5$, $\lambda = 0.5$

T = 100 M = 10^5	AR(1) parameter value in ADL(1,1)							AR(1) parameter value in ADL(1,1)						
	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.805	0.803	0.802	0.803	0.803	0.804	0.804	0.814	0.812	0.811	0.811	0.811	0.812	0.812
median	0.766	0.764	0.764	0.765	0.765	0.764	0.765	0.776	0.774	0.773	0.773	0.774	0.773	0.774
95%	1.285	1.286	1.282	1.284	1.282	1.281	1.282	1.291	1.292	1.289	1.291	1.288	1.288	1.289
obs. size	0.033	0.033	0.032	0.033	0.032	0.032	0.032	0.035	0.035	0.033	0.034	0.034	0.034	0.033
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.8: Simulated statistics for $\sup CUSQ^{OLS}$ and $\sup CUSQ^{REC}$, ADL(1,1): $\beta = 0.5$, $\rho = 1$, $\lambda = 0.5$

T = 100 M = 10^5	AR(1) parameter value in ADL(1,1)							AR(1) parameter value in ADL(1,1)						
	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.805	0.803	0.802	0.803	0.803	0.803	0.804	0.814	0.812	0.811	0.811	0.811	0.811	0.812
median	0.765	0.764	0.764	0.765	0.765	0.764	0.765	0.775	0.773	0.773	0.773	0.773	0.773	0.774
95%	1.285	1.282	1.282	1.284	1.283	1.280	1.282	1.291	1.291	1.289	1.291	1.289	1.287	1.290
obs. size	0.033	0.033	0.032	0.033	0.033	0.032	0.032	0.034	0.035	0.033	0.034	0.034	0.033	0.033
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.9: Simulated statistics for $\sup CUSQ^{OLS}$ and $\sup CUSQ^{REC}$, ADL(1,1): $\beta = 0.5$, $\rho = 1.1$, $\lambda = 0.5$

T = 100 M = 10^5	AR(1) parameter value in ADL(1,1)							AR(1) parameter value in ADL(1,1)						
	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.805	0.803	0.803	0.803	0.803	0.803	0.804	0.813	0.812	0.811	0.811	0.811	0.811	0.812
median	0.765	0.764	0.764	0.765	0.765	0.764	0.765	0.775	0.773	0.772	0.773	0.773	0.773	0.774
95%	1.284	1.283	1.282	1.284	1.283	1.279	1.282	1.291	1.291	1.290	1.291	1.290	1.288	1.290
obs. size	0.033	0.033	0.032	0.033	0.033	0.032	0.032	0.034	0.035	0.034	0.034	0.034	0.034	0.033
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.10: Simulated statistics for $\sup CUSQ^{OLS}$ and $\sup CUSQ^{REC}$, ADL(1,1): $\beta = 0.5$, $\rho = 0.5$, $\lambda = 0.5$

T = 100 M = 10^5	AR(1) parameter value in ADL(1,1)							AR(1) parameter value in ADL(1,1)						
	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.805	0.803	0.802	0.803	0.803	0.804	0.804	0.814	0.812	0.811	0.811	0.811	0.812	0.812
median	0.766	0.764	0.764	0.765	0.765	0.764	0.765	0.776	0.774	0.773	0.773	0.774	0.773	0.774
95%	1.285	1.286	1.282	1.284	1.282	1.281	1.282	1.291	1.292	1.289	1.291	1.288	1.288	1.289
obs. size	0.033	0.033	0.032	0.033	0.032	0.032	0.032	0.035	0.035	0.033	0.034	0.034	0.034	0.033
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.11: Simulated statistics for $\sup CUSQ^{OLS}$ and $\sup CUSQ^{REC}$, ADL(1,1): $\beta = 0.5$, $\rho = 0.5$, $\lambda = 1$

T = 100 M = 10^5	AR(1) parameter value in ADL(1,1)							AR(1) parameter value in ADL(1,1)						
	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.805	0.803	0.802	0.803	0.804	0.804	0.803	0.814	0.812	0.811	0.811	0.811	0.811	0.811
median	0.767	0.765	0.763	0.764	0.764	0.764	0.765	0.776	0.774	0.773	0.773	0.772	0.773	0.772
95%	1.283	1.283	1.281	1.284	1.282	1.282	1.284	1.292	1.293	1.290	1.291	1.290	1.290	1.289
obs. size	0.033	0.033	0.032	0.033	0.033	0.033	0.033	0.035	0.035	0.034	0.034	0.034	0.034	0.034
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.12: Simulated statistics for $\sup CUSQ^{OLS}$ and $\sup CUSQ^{REC}$, ADL(1,1): $\beta = 0.5$, $\rho = 0.5$, $\lambda = 1.1$

T = 100 M = 10^5	AR(1) parameter value in ADL(1,1)							AR(1) parameter value in ADL(1,1)						
	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.808	0.803	0.802	0.803	0.803	0.803	33.030	0.815	0.812	0.811	0.811	0.811	0.811	2.016
median	0.768	0.764	0.764	0.764	0.765	0.765	2.020	0.777	0.774	0.773	0.773	0.773	0.772	1.013
95%	1.292	1.284	1.281	1.283	1.284	1.284	74.868	1.292	1.293	1.289	1.290	1.289	1.289	6.196
obs. size	0.035	0.033	0.032	0.033	0.033	0.033	0.551	0.035	0.035	0.033	0.034	0.034	0.034	0.369
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)

Table 4.13: Simulated statistics for $\sup CUSQ^{OLS}$ and $\sup CUSQ^{REC}$, ADL(1,1): $\beta = 0.5$, $\rho = 0.5$, $\lambda = 0.5$

T = 100 M = 10^5	AR(1) parameter value in ADL(1,1)							AR(1) parameter value in ADL(1,1)						
	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.805	0.803	0.802	0.803	0.803	0.804	0.804	0.814	0.812	0.811	0.811	0.811	0.812	0.812
median	0.766	0.764	0.764	0.765	0.765	0.764	0.765	0.776	0.774	0.773	0.773	0.774	0.773	0.774
95%	1.285	1.286	1.282	1.284	1.282	1.281	1.282	1.291	1.292	1.289	1.291	1.288	1.288	1.289
obs. size	0.033	0.033	0.032	0.033	0.032	0.032	0.032	0.035	0.035	0.033	0.034	0.034	0.034	0.033
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.14: Simulated statistics for $\sup CUSQ^{OLS}$ and $\sup CUSQ^{REC}$, ADL(1,1): $\beta = 1$, $\rho = 0.5$, $\lambda = 0.5$

T = 100 M = 10^5	AR(1) parameter value in ADL(1,1)							AR(1) parameter value in ADL(1,1)						
	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.805	0.803	0.802	0.803	0.803	0.803	0.804	0.814	0.812	0.811	0.811	0.811	0.811	0.811
median	0.766	0.764	0.764	0.765	0.765	0.764	0.765	0.776	0.773	0.773	0.773	0.773	0.773	0.774
95%	1.285	1.283	1.282	1.284	1.283	1.280	1.282	1.291	1.291	1.289	1.292	1.290	1.287	1.290
obs. size	0.033	0.033	0.032	0.033	0.033	0.032	0.032	0.035	0.034	0.034	0.034	0.034	0.033	0.033
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.15: Simulated statistics for $\sup CUSQ^{OLS}$ and $\sup CUSQ^{REC}$, ADL(1,1): $\beta = 1.1$, $\rho = 0.5$, $\lambda = 0.5$

T = 100 M = 10^5	AR(1) parameter value in ADL(1,1)							AR(1) parameter value in ADL(1,1)						
	-1.2	-1	-0.9	0	0.9	1	1.2	-1.2	-1	-0.9	0	0.9	1	1.2
mean	0.805	0.803	0.802	0.803	0.803	0.803	0.804	0.814	0.812	0.811	0.811	0.811	0.811	0.812
median	0.766	0.764	0.764	0.765	0.765	0.764	0.765	0.766	0.773	0.773	0.773	0.773	0.773	0.774
95%	1.285	1.283	1.282	1.284	1.283	1.279	1.282	1.291	1.291	1.289	1.292	1.289	1.288	1.290
obs. size	0.033	0.033	0.032	0.033	0.033	0.032	0.032	0.035	0.034	0.034	0.034	0.034	0.034	0.033
(MCSE)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Chapter 5

Joint mis-specification testing and inference in autoregressions

1 Introduction

Mis-specification tests, as reported for example in software packages like OxMetrics (Doornik and Hendry 2007a, Doornik and Hendry 2007b), are typically given in a battery where the values of individual test statistics are reported together with indications of significance of the individual tests at particular levels (usually 5% or 1%). Two distinct worries arise. First, in the absence of information about the probabilistic interaction between the mis-specification tests it is far from clear how the battery should be interpreted. Generally, it is accepted as good practice to only use a model when all tests are passed individually at some sufficiently low levels of statistical significance; see e.g. Krolzig and Hendry (2001) and Doornik (2009). However, this does not allow the modeller to calibrate the overall significance level of the mis-specification test battery. Secondly, the extent to which mis-specification affects the results of subsequent inference is not clear when the

model does not fit into the classical regression framework.

Our aim in this chapter is to simultaneously treat the problems of joint mis-specification testing and of the possible interaction between mis-specification testing and classical hypothesis testing within the chosen model. Existing work on the first problem, joint mis-specification testing, includes Savin and White (1978) and Bera and Jarque (1982). The latter authors work with a linear regression with fixed regressors and normal innovations and then define a more general model with a larger parameter vector, allowing for autoregressive dependence as well as heteroscedasticity and non-normality in the innovations, in addition to more regressors and non-linearities in parameters. They then formulate the joint mis-specification test as an LM test on the larger parameter vector $\theta = (\theta_1, \theta_2)$, where θ_1 are nuisance parameters, and the restrictions imposed on θ_2 imply correct specification, i.e. the normal model with fixed regressors. Under the null hypothesis, this LM test turns out to have an additivity property, whereby the joint test can be written as the sum of the LM tests for each hypothesis separately, that is, the tests have no mutual dependence. Relating to the second problem, a few independence results exist for some mis-specification tests and inference procedures in contexts where Basu's theorem (Basu 1955) applies and functions of the model's residuals are independent of sufficient statistics for its parameters; see Phillips and McCabe (1983), Phillips and McCabe (1988), and Phillips and McCabe (1989), or Spanos (2007) for a slightly more abstract treatment of the problem. To the best knowledge of this author, existing work does not treat either problem in a time series context.

While Bera and Jarque (1982) test residual normality, homoscedasticity, and independence, as well as linearity of the regression equation in a model with fixed regressors, we adopt the more general model class of a finite order autoregression

and test residual normality and covariance stationarity, as well as lag length, the presence of a unit root, and hypotheses concerning the autoregressive parameters in the stationary part of the process. We find that the mis-specification tests are not only mutually independent asymptotically, but also each independent asymptotically of subsequent inference within the chosen model, under the null hypothesis of correct specification. The implications for finite sample behaviour are not very clear, but a small scale simulation study suggests that independence may be regarded as a good approximation for typical sample sizes.

The model class is defined in §2, the test statistics included in the analysis are treated separately in §3, and the main results of the chapter, along with some auxiliary results are presented in §4. §5 presents results from a small scale simulation study, and §6 concludes. Proofs are given in the appendix.

2 Model and assumptions

We will consider a number of mis-specification tests. The strength of assumptions under which asymptotic theory exists for each of these tests varies. While a very general result is available for the lag length test (Nielsen 2006), making it valid even in singular explosive vector autoregressions (see Nielsen (2008)), and the CUSUM of squares tests are valid under the assumptions of Chapter 2, the normality test has not been shown to be valid in such a general model class; Kilian and Demiroglu (2000) consider a vector autoregession allowing for a constant and one stochastic trend, while Caceres (2007, Ch. 5) shows that the test is valid in an explosive first order autoregression. We are unaware of existing results for higher order explosive autoregressions or conditional models such as the autoregressive distributed lag used in Chapters 2 and 3. We therefore choose to consider an

autoregressive model of order k, with characteristic roots restricted to the non-explosive region. These choices also simplify inference on the integrated part of the process, as we do not have to make assumptions about weak exogeneity that would be necessary in the case of a vector autoregression. We proceed to state our assumptions formally.

The model class in which inference is conducted is a univariate autoregression with a constant and a linear trend, although when considering the normality test, we shall make an assumption ruling out the linear trend. The scalar process x_t satisfies the autoregression

$$x_t = \mu_1 + \mu_2 t + \alpha_1 x_{t-1} + \ldots + \alpha_k x_{t-k} + \varepsilon_t, \quad t = 1, \ldots, T. \tag{2.1}$$

We let $X_{k,t-1} = (x_{t-1}, \ldots x_{t-k})'$ denote the $(k \times 1)$-vector of lagged values, $\alpha_k = (\alpha_1, \ldots \alpha_k)'$ the parameters associated with the stochastic part of the process, and with $\mu = (\mu_1, \mu_2)'$, $d_t = (1, t)'$, define the $(k+2)$-dimensional vector of regressors $S_{k,t-1} = (X'_{k,t-1}, d'_t)'$ and the parameter vector $\beta_k = (\alpha'_k, \mu')'$ so that we can rewrite (2.1) as $x_t = \beta'_k S_{k,t-1} + \varepsilon_t$. Let $\Lambda(z)$ be the characteristic polynomial of the process x_t, as defined in (2.1), with k (complex) roots denoted $(\lambda_1, \ldots \lambda_k)$ and let $|\lambda_j|$ denote the absolute value (modulus) of the j^{th} (complex) root. We restrict attention to the non-explosive region of the parameter space in (2.1) with the following assumption.

Assumption A. $|\lambda_j| \geq 1$ *for* $j = 1, \ldots k$

In the sequel, $(\mathcal{F}_t,\ t = 0, \ldots T)$ denotes an increasing sequence of σ-fields. The innovations of the autoregression (2.1) are taken to satisfy the following martingale difference assumption.

Assumption B. *The adapted sequence* $(\varepsilon_t, \mathcal{F}_t)$ *is a martingale difference sequence so that* $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ *and its higher moments are governed by the following two conditions.*

$$\sup_t \mathsf{E}(|\varepsilon_t|^\lambda|\mathcal{F}_{t-1}) < \infty \ a.s \quad \textit{for some } \lambda > 4, \tag{2.2}$$

$$\mathsf{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2 > 0 \ a.s. \tag{2.3}$$

We also take the the initial values $x_0, \ldots, x_{1-k}$ to be $\mathcal{F}_0$-measurable. The variance constancy test considered later requires us to make an assumption about the behaviour of the squared innovations. The difference between this and Assumption D of Chapter 2 is the σ-field we condition on. This is because, in the model under consideration here, we do do not include a contemporaneous stochastic regressor.

Assumption C. *The adapted sequence* $(\varepsilon_t^2 - \sigma^2, \mathcal{F}_t)$ *is a martingale difference sequence, such that* $\mathsf{E}\{(\varepsilon_t^2 - \sigma^2)|\mathcal{F}_{t-1}\} = 0$*, and satisfies, for some* $\varphi > 0$*,*

$$\mathsf{Var}(\varepsilon_t^2 - \sigma^2|\mathcal{F}_{t-1}) = \varphi^2 \ a.s. \tag{2.4}$$

The variance constancy test will depend asymptotically on the nuisance parameter φ^2, the fourth conditional moment of the innovations. In order to ensure that it is possible to estimate this quantity consistently in strong sense, a martingale assumption must be made concerning the third conditional moments; see Nielsen and Sohkanen (2009, Thm 3.3). While we are not really concerned with consistent estimation of φ^2 here, this assumption is also required in showing independence in the limit between the variance constancy test and the other tests.

Assumption D. *The adapted process* $(\varepsilon_t^3, \mathcal{F}_t)$ *is a martingale difference sequence*

so that $\mathsf{E}(\varepsilon_t^3|\mathcal{F}_{t-1}) = 0$ *and for some* $\lambda > 6$*,*

$$\sup_t \mathsf{E}(|\varepsilon_t|^\lambda|\mathcal{F}_{t-1}) < \infty \ a.s. \tag{2.5}$$

For the purposes of dealing with normality tests, even higher powers of the innovations need to be controlled. The asymptotic distribution of the test is the same under normality and under a martingale difference assumption on $\{(\varepsilon_t^4 - 3\sigma^4), \mathcal{F}_t\}$ if $\mathsf{E}(|\varepsilon_t|^\lambda|\mathcal{F}_{t-1})$ for some $\lambda > 8$ is bounded uniformly in t. Therefore, it can be thought of as a test for approximate normality. However, in making an asymptotic independence argument involving the normality test, we shall require that the population moments of the innovations match the normal ones up to eighth order. This is very close to assuming normality, and we therefore choose to do just that when considering the normality test. Furthermore, asymptotic distribution theory for the normality test we consider only exists in the presence of a stochastic trend when there is no linear trend in the model. Since neither normality nor the restriction on the linear trend are required for the convergence or independence arguments involving the other tests, for the sake of conciseness we formulate both requirements specific to the normality test in a single assumption as follows.

Assumption E. *The sequence* (ε_t) *is independently and identically normally distributed as* $\mathsf{N}(0, \sigma^2)$*. Furthermore, in the autoregressions defined by* (2.1) *and* (3.1)*,* $\mu_2 = 0$ *such that there is a constant but not a linear trend in the data generation process.*

Finally, some notation is required in what follows. For any two processes y_t and z_t we write 'y_t corrected for z_t' as $(y_t|z_t) = y_t - \sum_{j=1}^{T} y_j z_j (\sum_{j=1}^{T} z_j z_j')^{-1} z_t$.

3 Test statistics

The model given in (2.1) is fitted, that is, x_t is regressed on $(x_{t-1}, \ldots, x_{t-k})$, and $(1, t)$, in

$$x_t = \mu_1 + \mu_2 t + \alpha_1 x_{t-1} + \ldots + \alpha_k x_{t-k} + \varepsilon_t, \quad t = 1, \ldots, T.$$

Recall the shorthand notation $S_{k,t-1} = (X'_{k,t-1}, d'_t)'$ with $X_{k,t-1} = (x_{t-1}, \ldots x_{t-k})'$ and $d_t = (1, t)'$ for the regressors of this model. We shall make use of this in the arguments that follow. Once the initial model is fitted, mis-specification tests, such as tests for lag length, residual variance constancy, and normality, are reported in order to ascertain whether the residuals exhibit, to a good approximation, the properties assumed about the innovations of (2.1). When this is the case, inference is conducted concerning the parameters of interest $\beta_k = (\alpha'_k, \mu')'$, where $\alpha_k = (\alpha_1, \ldots \alpha_k)'$ and $\mu = (\mu_1, \mu_2)'$ and the subscript k on β_k is included to emphasize the lag length of autoregression. This exercise is complicated by the fact that the distributional properties of the relevant estimators depend upon the characteristic roots of the data generating process. For simplicity of presentation, in formulating the main results, we choose not to consider inference on the parameters associated with the constant term that is left unrestricted under the unit root hypothesis.

3.1 Inference on the autoregressive parameters

While the mis-specification tests considered in this chapter are valid in the presence of unit roots, a mapping to a (trend) stationary process is needed for valid inference on the parameters of interest. In the context of vector autoregressions, Doornik and Hendry (2007b, Ch. 6) refer to this as *reduction to I(0)*, a proce-

dure that relies on exploiting cointegrating relationships and differencing. For the sake of simplicity, we consider the case of one unit root, with all remaining roots stationary. While Nielsen (2001) shows that this latter assumption is redundant when testing for a unit root, it is consistent with Assumption A. In this univariate case, we make use of first differencing only. In particular, we shall consider the transformed model

$$\begin{aligned}\Delta x_t &= \mu_1 + \mu_2 t + \{(\sum_{j=1}^{k} \alpha_j) - 1\} x_{t-1} - \sum_{l=1}^{k-1} \sum_{j=l}^{k} \alpha_j \Delta x_{t-l} + \varepsilon_t \\ &= \mu_1 + \mu_2 t + \gamma_1 x_{t-1} + \sum_{l=1}^{k-1} \gamma_{2,l} \Delta x_{t-l} + \varepsilon_t, \end{aligned} \tag{3.1}$$

with the characteristic polynomial $\tilde{\Lambda}(z) = (1-z)^{-1}\Lambda(z)$, where $\Lambda(z)$ is the characteristic polynomial of (2.1). The presence of deterministic terms introduces some additional complications when trying to find out whether to use (2.1) or (3.1) as the model. When testing for the presence of a unit root, a joint hypothesis of the form $\gamma_1 = 0$ and $\mu = \tilde{\mu}$ is required for reasons to be explained below. We introduce the following shorthand notation for the regressors and parameters of (3.1) under this hypothesis:

$$\begin{aligned}\tilde{S}_{k-1,t-1} &= \{(\Delta x_{t-1}, \ldots, \Delta x_{t-k+1}), (1, t)\}' \\ \delta_{k-1} &= \{(\gamma_{2,1}, \ldots, \gamma_{2,k-1}), \tilde{\mu}'\}'. \end{aligned}$$

In oder to illustrate the problem, we choose to consider separately the case of no deterministic terms, that is, the case of $\mu = 0$, before proceeding with the more general case.

The case of no deterministic terms

Setting $\mu = 0$, the model given in (2.1) has a stochastic trend if $(\sum_{j=1}^{k} \alpha_j) - 1 = 0$ and $\sum_{l=1}^{k-1} \sum_{j=l}^{k} \alpha_j \neq -1$ (where the latter restriction rules out the case of two stochastic trends). Equivalently, this is the case in (3.1) when $\gamma_1 = 0$ and $\sum_{l=1}^{k-1} \gamma_{2,l} \neq 1$. To see this, note that under the restriction, the model in levels is $x_t = x_{t-1} + \{(\gamma_{2,1}, \ldots, \gamma_{2,k-1}), 0'\}\tilde{S}_{k-1,t-1} + \varepsilon_t$ and the presence of the stochastic trend may be demonstrated using recursive substitution. The regression is balanced only when first differences of x_t are used as the dependent variable.

The case of deterministic terms

Given $\sum_{l=1}^{k-1} \gamma_{2,l} \neq 1$, under the hypothesis $\gamma_1 = 0$ the behaviour of the x_t-process defined in (2.1) depends on what is assumed about the deterministic d_t-process. This is a well known problem in unit root testing. Normally, the unit root hypothesis involves restricting the highest order deterministic term to zero; see e.g. Nielsen (2001). That is, if the deterministic term in the model in levels, (2.1), is a constant, then in addition to the restriction $\gamma_1 = 0$, the unit root hypothesis involves a restriction of this constant to zero. In our setting, this would be the case if $\mu = (\mu_1, 0)$ in (2.1), so that $\tilde{\mu} = (0, 0)$. Then, the model in levels only has a stochastic trend under the null hypothesis. If, however, the deterministic term in (2.1) includes a constant and a linear trend, then it is the coefficient on this trend that is restricted to zero together with $\gamma_1 = 0$ in the formulation of the unit root hypothesis, and the model in levels under the null has a constant and a linear trend. Again, in our setting this would be the case if $\mu = (\mu_1, \mu_2)$ in (2.1), so that $\tilde{\mu} = (\mu_1, 0)$. If this restriction on the linear trend were omitted, the model in levels under the null would have a quadratic trend, which arguably would be unsatisfactory in most applications.

In the analysis that follows it is helpful to think of the data generation process as a composite of an integrated and a stationary stochastic process, mirroring the arguments of Chapters 2 and 3. As noted e.g. in Nielsen (2005, §3) for a vector autoregression S_t with characteristic roots on and outside the circle, there exists a real invertible matrix M with the property that

$$MS_t = \begin{pmatrix} U_t \\ V_t \end{pmatrix} = \begin{pmatrix} \mathbf{U} & 0 \\ 0 & \mathbf{V} \end{pmatrix} \begin{pmatrix} U_{t-1} \\ V_{t-1} \end{pmatrix} + \begin{pmatrix} e_{U,t} \\ e_{V,t} \end{pmatrix} \tag{3.2}$$

where $\max|\text{eigen}(\mathbf{U})| < 1$ and the deterministic terms are subsumed into the V_t-process. If $\dim(\mathbf{V}) \neq 0$, then the process is non-stationary. Under the restriction $\mathbf{V} = 1$, its order of integration is unity. This decomposition allows us to separate the problem of inference on the stationary stochastic part of the process from unit root testing with its complications regarding deterministic terms. In particular, inference on any of the parameters associated with the stationary stochastic part of the process can be conducted by examining the vector

$$\mathcal{S}^{\mathsf{inf}} = (\sum_{t=1}^{T} U_{t-1}U_{t-1}')^{-1/2} \sum_{t=1}^{T} U_{t-1}\varepsilon_t. \tag{3.3}$$

It remains to relate this rather abstract framework to the models given in (2.1) and (3.1). Under the $I(1)$ hypothesis, $\sum_{l=1}^{k-1}\gamma_{2,l} \neq 1$, $\gamma_1 = 0$ and $\mu = \tilde{\mu}$ (the latter translating as the exclusion of the highest order deterministic term from the model), we shall set $U_{t-1} = (\iota_{k-1}\tilde{S}_{k-1,t-1}|d_t)$, where ι_{k-1} is the selection matrix $(I_{k-1}, 0_{(k-1)\times\dim(\mu)})$. When the I(1) hypothesis is not maintained, we set $U_{t-1} = (\iota_k S_{k,t-1}|d_t)$, where $\iota_k = (I_k, 0_{(k)\times\dim(\mu)})$. More concretely, in the former case, U_{t-1} is the vector of residuals from the regression of the first differences of the stochastic regressors on the deterministic terms, in other words the vector

$(\Delta x_{t-1}, \ldots, \Delta x_{t-k+1})'$ corrected for d_t. In the latter case, U_{t-1} is the vector $(x_{t-1}, \ldots, x_{t-k})'$ corrected for d_t. Under Assumptions A and B, the quantity $\mathcal{S}^{\mathsf{inf}}$ is asymptotically normally distributed by the central limit theorem for martingale differences; see e.g. Chan and Wei (1988, Thm. 2.2). Having described inference on the stationary part of the process, we proceed to discuss, in turn, each of the four other tests considered in this chapter. We will leave it up to the reader to decide whether to view the unit root test as a mis-specification test or as part of inference on the parameters of the model.

3.2 Unit root test

We shall consider the likelihood ratio test presented in Nielsen (2001). As was mentioned above, the relevant test statistic of the hypothesis $\gamma_1 = 0$ and $\mu = \tilde{\mu}$ varies with $\tilde{\mu}$, as does its asymptotic distribution. The test statistic is $\mathsf{LR}^{\mathsf{ur}} = -T\log(1-\hat{\lambda}^2)$, where, in terms of the notation introduced in the previous section, with ι_{k-1} the selection matrix $(I_{k-1}, 0_{(k-1)\times\dim(\mu)})$,

$$\hat{\lambda} = \frac{\sum_{t=1}^{T}(\Delta x_t|\iota_{k-1}\tilde{S}_{k-1,t-1})(x_{t-1}|\iota_{k-1}\tilde{S}_{k-1,t-1})}{\{\sum_{t=1}^{T}(\Delta x_t|\iota_{k-1}\tilde{S}_{k-1,t-1})^2 \sum_{t=1}^{T}(x_{t-1}|\iota_{k-1}\tilde{S}_{k-1,t-1})^2\}^{1/2}}, \tag{3.4}$$

when $\mu = 0$ in the data generation process so no deterministic terms are included in (2.1), and therefore (3.1). If $\mu_2 = 0$ in the data generation process, but a constant is included in (2.1), it is restricted to zero in (3.1) so we jointly test $\gamma_1 = 0$ and $\mu_1 = 0$ in

$$\Delta x_t = (\gamma_1, \mu_1)\begin{pmatrix} x_{t-1} \\ 1 \end{pmatrix} + \sum_{l=1}^{k-1}\gamma_{2,l}\Delta x_{t-l} + \varepsilon_t \tag{3.5}$$

and form $\hat{\lambda}$ with the residuals $(\Delta x_t|\iota_{k-1}\tilde{S}_{k-1,t-1})$ and $((x_{t-1},1)|\iota_{k-1}\tilde{S}_{k-1,t-1})$. Finally, if a constant and trend are both included in (2.1), then unrestricted constant appears under the null hypothesis in (3.1). That is, we jointly test $\gamma_1 = 0$ and $\mu_2 = 0$ in

$$\Delta x_t = (\gamma_1, \mu_2) \begin{pmatrix} x_{t-1} \\ t \end{pmatrix} + \mu_1 + \sum_{l=1}^{k-1} \gamma_{2,l} \Delta x_{t-l} + \varepsilon_t \tag{3.6}$$

and the relevant sample correlation is formed with the residuals $(\Delta x_t|\tilde{\iota}_k\tilde{S}_{k-1,t-1})$ and $((x_{t-1},t)|\tilde{\iota}_k\tilde{S}_{k-1,t-1})$, where we define $\tilde{\iota}_k = (I_k, 0_{k\times 1})$, such that $\tilde{\iota}_k\tilde{S}_{k-1,t-1} = (\Delta x_{t-1}, \dots \Delta x_{t-k+1}, 1)'$.

3.3 Test for lag length

The test for lag length is the likelihood ratio test of $\mathsf{H}_m : (\alpha_{k+1}, \dots, \alpha_{k+m}) = 0$ for a positive integer m against its negation in the model

$$x_t = \sum_{j=1}^{k+m} \alpha_j x_{t-j} + \mu' d_t + \varepsilon_t, \quad t = 1, \dots, T. \tag{3.7}$$

Again, we use shorthand for the regressors, defining $S_{k+m,t-1} = (X'_{k+m,t-1}, d'_t)'$ where $X_{k+m,t-1} = (x_{t-1}, \dots, x_{t-k-m})'$, and let

$$\hat{\sigma}^2_{k+m} = T^{-1} \sum_{t=1}^{T} (\varepsilon_t|S_{k+m,t-1})^2 \quad \text{and} \quad \hat{\sigma}^2_k = T^{-1} \sum_{t=1}^{T} (\varepsilon_t|S_{k,t-1})^2, \tag{3.8}$$

and the likelihood ratio test of $\mathsf{H}_m : (\alpha_{k+1}, \dots, \alpha_{k+m}) = 0$ can be written as

$$\mathsf{LR}_{\mathsf{H}_m} = -T \log(\hat{\sigma}^2_{k+m}/\hat{\sigma}^2_k) = -T \log\left(1 - \frac{\hat{\sigma}^2_k - \hat{\sigma}^2_{k+m}}{\hat{\sigma}^2_k}\right). \tag{3.9}$$

Sometimes, the test for lag length is implemented through an auxiliary regression of the fitted residuals on their own lags and the lags of the process included in the original regression; see Breusch (1978), Godfrey (1978), and Pagan (1984). This is the approach taken, for example, in the econometric software PcGive (Hendry 1986, Doornik and Hendry 2007a).

3.4 Test for variance constancy

The test of variance constancy (against non-stationary volatility) is the CUSUM of squares test using either the recursive residuals of Brown, Durbin, and Evans (1975) or ordinary least squares residuals. Its implementation calls for recursive estimation of the model. That is, we compute

$$\hat{\beta}_{k,t} = (\sum_{s=1}^{t} S_{k,s-1} S'_{k,s-1})^{-1} \sum_{s=1}^{t} S_{k,s-1} x_s, \quad \text{for } t = M, \ldots, T. \tag{3.10}$$

For example if the recursive residuals are used, the test is based on the statistic

$$\mathsf{CUSQ}^{\mathsf{REC}} = T^{1/2} \left(\sum_{s=1}^{t} \tilde{\varepsilon}_s^2 / \sum_{s=1}^{T} \tilde{\varepsilon}_s^2 - (t-M)/(T-M) \right), \tag{3.11}$$

where the standardized one-step-ahead prediction errors $\tilde{\varepsilon}_t$ are defined, for $t > M$ as

$$\tilde{\varepsilon}_t = \{1 + S'_{k,t-1} (\sum_{s=1}^{t-1} S_{k,s-1} S'_{k,s-1})^{-1} S_{k,t-1}\}^{-1/2} (x_t - \hat{\beta}'_{k,t-1} S_{k,t-1}) \tag{3.12}$$

3.5 Test for normality

The Jarque-Bera omnibus test for normality is based on the sum of sample skewness and sample excess kurtosis, that is on the sum of

$$\hat{S} = \frac{T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_t - \bar{\hat{\varepsilon}})^3}{(T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_t - \bar{\hat{\varepsilon}})^2)^{3/2}} \quad \text{and} \quad \hat{K} = \frac{T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_t - \bar{\hat{\varepsilon}})^4}{(T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_t - \bar{\hat{\varepsilon}})^2)^2} - 3, \tag{3.13}$$

where $\bar{\hat{\varepsilon}} = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t$; see Bowman and Shenton (1975) and Jarque and Bera (1987). For the model given in (2.1), the test statistic is $\mathsf{JB}^{\mathsf{omnibus}} = T(\hat{S}^2/6 + \hat{K}^2/24)$. De-meaning of the residuals is important in case the data generation process contains a stochastic trend. If neglected, a non-vanishing nuisance term appears in the limit; see Lee and Wei (1999). We note that the use of the third and fourth sample moments to test for normality is not a recent innovation. This approach is taken by Thiele (1889).

4 Main results

We start by establishing some more notation. Throughout,

$$s_{\gamma_2,T} = (T\sigma^2)^{-1}\sum_{t=1}^{T}(\iota_{k-1}\tilde{S}_{k-1,t-1}|d_t)(\iota_{k-1}\tilde{S}_{k-1,t-1}|d_t)'$$

is a $(k-1)\times(k-1)$ matrix and we let $\mathcal{B}_{U,1}$ and $\mathcal{B}_{\mathcal{Y},1}$ be standard normal distributions of dimensions $k-1$ and m, respectively when the $I(1)$ hypothesis is maintained so $U_{t-1} = (\iota_{k-1}\tilde{S}_{k-1,t-1}|d_t)$, and of dimensions k and m, respectively, when the $I(1)$ hypothesis is not maintained so $U_{t-1} = (\iota_k S_{k,t-1}|d_t)$, while $\mathcal{B}_{1,u}$ is a standard scalar Brownian motion.

The marginal asymptotic distributions of the test statistics for the presence of a stochastic trend, and of the tests for lag length, variance breaks, and normality introduced in the previous section are characterized before the their joint behaviour is considered. The lemmas below give the necessary results. We emphasize that these are existing results (references are given in the proofs), and our contribution in this chapter relates to the asymptotic interactions between the tests.

First, we consider the test for a unit root.

Lemma 4.1. *Assume A and B. Then, if* $\mathbf{V} = 1$*, the likelihood ratio test* $LR^{ur} = -T\log(1-\hat{\lambda}^2)$*, with* $\hat{\lambda}$ *given in* §3.2*, is distributed asymptotically as the quantity* $\int_0^1 d\mathcal{B}_{1,u}F_u'\{\int_0^1 F_uF_u'\}^{-1}\int_0^1 F_u d\mathcal{B}_{1,u}$*, where the process* F_u *is defined in Nielsen (2001, Thm. 3). In particular,* $F_u = \mathcal{B}_{1,u}$ *if there are no deterministic terms in* (2.1)*,* $F_u = (\mathcal{B}_{1,u}, 1)$ *if a constant is included in* (2.1)*, and* $F_u = (\mathcal{B}_{1,u} - \int_0^1 \mathcal{B}_{1,v}dv, u - 1/2)$ *if both a constant and a linear trend are included in* (2.1)*.*

Remark 4.2. *The restriction* $\mathbf{V} = 1$ *ensures that there is at most one stochastic trend in the data generation process. In terms of the model* (3.1)*, we can instead write* $\sum_{l=1}^{k-1}\gamma_{2,l} \neq 1$*.*

Next, the likelihood ratio test for lag length is considered.

Lemma 4.3. *Assuming A and B, the LR test of* $H_m : (\alpha_{k+1}, \ldots, \alpha_{k+m})' = 0$*, considered by Nielsen (2006), depends asymptotically only on the* $\mathcal{F}_T$*-measurable quantity* $(\mathcal{S}_m^{lag})^2$*, where*

$$\mathcal{S}_m^{lag} = \left(\sum_{t=1}^{T}\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'\right)^{-1/2}\sum_{t=1}^{T}\mathcal{Y}_{t-1}\varepsilon_t \tag{4.1}$$

with $\mathcal{Y}_t$ the process $C'(U_{t-1}, \varepsilon_t)'$ a linear combination of ε_t and U_{t-1} independent of U_t. The test given in (3.9) *is asymptotically distributed as $\chi^2(m)$ under Assumptions A and B; see Nielsen (2006, Thm. 2.3).*

The CUSUM of squares test was considered in Chapter 2. Regardless of implementation by forecast errors or ordinary least squares residuals, it can be shown to depend asymptotically on the fundamental innovations under our assumptions, and to exhibit Brownian asymptotics.

Lemma 4.4. *Assume A, B, and C. Then, the CUSUM of squares tests defined in* (3.11) *behaves asymptotically as the quantity*

$$\mathcal{S}^{\mathsf{CUSQ}} = T^{1/2} \left(\frac{\sum_{s=1}^{t} \varepsilon_s^2}{\sum_{s=1}^{T} \varepsilon_s^2} - \frac{t}{T} \right) \tag{4.2}$$

distributed asymptotically as a standard Brownian bridge scaled by a nuisance parameter. This result continues to hold if ordinary least squares residuals, instead of the recursive residuals, are used in the implementation.

The Jarque-Bera normality test was generalized for vector autoregressions with unit roots by Kilian and Demiroglu (2000); for a treatment of the test in stationary VARs, see Lütkepohl (2007, Ch. 4). The following lemma suffices for our purposes in the univariate autoregressive setting.

Lemma 4.5. *Assume A, B, C, D, and E. Then, the test implemented with residuals from the autoregression* (2.1) *as $\mathsf{JB}^{\mathsf{omnibus}} = T(\hat{S}^2/6 + \hat{K}^2/24)$, where $\hat{S}$ and $\hat{K}$ are defined in* (3.13)*, and the test implemented with the true errors, have the same asymptotic distribution. In particular, $\mathsf{JB}^{\mathsf{omnibus}} = \mathcal{S}^{\mathsf{JB}} + \mathrm{o}_P(1)$, where*

$\mathcal{S}^{JB} = T(S/6 + K/24)$ *with*

$$S = \frac{T^{-1}\sum_{t=1}^{T}(\varepsilon_t - \bar{\varepsilon})^3}{(T^{-1}\sum_{t=1}^{T}(\varepsilon_t - \bar{\varepsilon})^2)^{3/2}} \quad \textit{and} \quad K = \frac{T^{-1}\sum_{t=1}^{T}(\varepsilon_t - \bar{\varepsilon})^4}{(T^{-1}\sum_{t=1}^{T}(\varepsilon_t - \bar{\varepsilon})^2)^2} - 3, \tag{4.3}$$

is asymptotically distributed as $\chi^2(2)$.

Finite sample corrections for the normality tests have been discussed ever since Fisher (1930), the main issue being the small sample correlation between the third and fourth sample moments. Doornik and Hansen (2008) show how to implement these corrections for a test in the multivariate case.

4.1 Behaviour of the tests under the joint null hypothesis

As the tests require assumptions of varying strength, we split the main result into three theorems. We believe this adds clarity to the argument as the way the assumptions are being used becomes apparent. Proofs are in the appendix. In formulating the theorems, we do not consider inference on the unrestricted constant under the unit root hypothesis. The first theorem concerns inference and lag length testing.

Theorem 4.6. *If* $\mathbf{V} = 1$*, then under Assumptions A and B, it holds*

$$\begin{pmatrix} s_{\gamma_2,T}^{1/2}\iota_{k-1}(\hat{\delta}_{k-1} - \delta_{k-1}) \\ LR^{ur} \\ LR_{H_m} \end{pmatrix} \to \begin{pmatrix} \mathcal{B}_{U,1} \\ \int_0^1 d\mathcal{B}_{1,u}F_u'\{\int_0^1 F_uF_u'\}^{-1}\int_0^1 F_u d\mathcal{B}_{1,u} \\ \mathcal{B}_{y,1}'\mathcal{B}_{y,1} \end{pmatrix} \tag{4.4}$$

in distribution on $\mathbb{R}^3$ *as* $T \to \infty$*, where* F_u *is defined in Lemma 4.1 and the three elements of the asymptotic distribution are mutually independent.*

The second theorem concerns inference, lag length and variance constancy testing. We let $\mathcal{B}^0_{2,u}$ be a standard Brownian bridge.

Theorem 4.7. *Under assumptions A, B, C, D, and the restriction* $\mathbf{V} = 1$*, it holds*

$$\begin{pmatrix} s^{1/2}_{\gamma_2,T}\iota_{k-1}(\hat{\delta}_{k-1} - \delta_{k-1}) \\ LR^{ur} \\ LR_{H_m} \\ CUSQ^{REC} \end{pmatrix} \rightarrow \begin{pmatrix} \mathcal{B}_{U,1} \\ \int_0^1 d\mathcal{B}_{1,u}F_u'\{\int_0^1 F_uF_u'\}^{-1}\int_0^1 F_u d\mathcal{B}_{1,u} \\ \mathcal{B}'_{\mathcal{Y},1}\mathcal{B}_{\mathcal{Y},1} \\ \sigma^{-2}\varphi\mathcal{B}^0_{2,u} \end{pmatrix} \tag{4.5}$$

in distribution on the product space $\mathbb{R}^3 \times D[0,1]$ *as* $T \rightarrow \infty$*, where* F_u *is defined in Lemma 4.1 and the four elements of the asymptotic distribution are mutually independent.*

We now present a theorem concerning the entire set of tests considered in this chapter, letting $\mathcal{B}_{3,1}$ and $\mathcal{B}_{4,1}$ be standard normal distributions.

Theorem 4.8. *Under Assumptions A, B, C, D, E, and the restriction* $\mathbf{V} = 1$*, it holds*

$$\begin{pmatrix} s^{1/2}_{\gamma_2,T}\iota_{k-1}(\hat{\delta}_{k-1} - \delta_{k-1}) \\ LR^{ur} \\ LR_{H_m} \\ CUSQ^{REC} \\ JB^{omnibus} \end{pmatrix} \rightarrow \begin{pmatrix} \mathcal{B}_{U,1} \\ \int_0^1 d\mathcal{B}_{1,u}F_u'\{\int_0^1 F_uF_u'\}^{-1}\int_0^1 F_u d\mathcal{B}_{1,u} \\ \mathcal{B}'_{\mathcal{Y},1}\mathcal{B}_{\mathcal{Y},1} \\ \sigma^{-2}\varphi\mathcal{B}^0_{2,u} \\ \mathcal{B}^2_{3,1} + \mathcal{B}^2_{4,1} \end{pmatrix} \tag{4.6}$$

in distribution on the product space $\mathbb{R}^4 \times D[0,1]$ *as* $T \rightarrow \infty$*, where* F_u *is either* $\mathcal{B}_{1,u}$ *or* $(\mathcal{B}_{1,u}, 1)$*, and the five elements of the asymptotic distribution are mutually independent.*

Remark 4.9. *If dim*$(\mathbf{V}) = 0$*, then the results stated in Theorems 4.6, 4.7, and 4.8 hold with the second elements deleted and* $\iota_{k-1}(\hat{\delta}_{k-1} - \delta_{k-1})$ *replaced by the* k *vector* $\iota_k(\hat{\beta}_k - \beta_k)$ *and* $s_{\gamma_2,T}^{1/2}$ *replaced by* $s_{\alpha,T}^{1/2}$*, with* $s_{\alpha,T}$ *the* $k \times k$ *matrix* $(T\sigma^2)^{-1}\sum_{t=1}^{T}(\iota_k S_{k,t-1}|d_t)(\iota_k S_{k,t-1}|d_t)'$*. We do not consider inference on the parameters associated with the constant and linear trend.*

The independence result included in each theorem implies that, asymptotically, calculation of the significance level of the joint test is straightforward. As a simple example, suppose that the individual significance level of each test is c and n tests are conducted. In the absence of information about the mutual dependence between the tests, all we can say is that the the joint significance level is bounded above by nc. Under asymptotic independence, the significance level of the joint test is $1-(1-c)^n$ asymptotically; see Godfrey (1988) and Figure 2.1 of Chapter 1.

5 Simulation study

Using Ox (Doornik 2007), we perform a small scale Monte Carlo experiment to investigate the pairwise sample correlations of the tests in a second order autoregression. There is one (real) unit root in the DGP and the other (real) root is in the stationary region. Specifically, the process generating the data is

$$Y_t = 0.4Y_{t-1} + 0.6Y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim \mathsf{N}(0,1) \tag{5.1}$$

which can be rewritten as

$$\Delta Y_t = -0.6\Delta Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathsf{N}(0,1) \tag{5.2}$$

The test statistics considered are the t-test on the coefficient of the first lagged difference in the AR(1) in differences, the likelihood ratio unit root test, the likelihood ratio test on the significance of the third lag in levels, the CUSUM of squares test implemented with recursive residuals from the model in levels, and the Jarque-Bera normality test on ordinary least squares residuals from the model in levels.

For each repetition $m = 1, \ldots, M$, values of the test statistics $t_{i,m}$ for $i = 1, \ldots, 5$ are recorded. Then, the pairwise correlations given in Tables 5.1, 5.2, and 5.3 are calculated as $c_{i,j} = \{\sum_{m=1}^{M}(t_{i,m}-\bar{t}_i)(t_{j,m}-\bar{t}_j)\}/\{\sum_{m=1}^{M}(t_{i,m}-\bar{t}_i)^2 \sum_{m=1}^{M}(t_{j,m}-\bar{t}_j)^2\}^{1/2}$, where $\bar{t}_i = M^{-1}\sum_{m=1}^{M} t_{i,m}$ and $\bar{t}_j = M^{-1}\sum_{m=1}^{M} t_{j,m}$. Across the M independent repetitions, the shape of the distribution of the $c_{i,j}$s will depend on how far the population value is from zero, and on the value of M. In particular, if the absolute value of the correlation were close to one, the distribution would be highly skewed. Although, in this application, we expect the correlations to be close to zero, we apply the Fisher transformation to compute the confidence intervals just in case; see for example Johnson (1978). That is, we first apply the transformation $\tilde{c}_{i,j} = \log\{(1 + c_{i,j})/(1 - c_{i,j})\}/2$ and then compute a confidence interval for this transformed variable as $\tilde{c}_{i,j} \pm 1.96/\sqrt{M}$. The confidence intervals reported in Tables 5.1, 5.2, and 5.3 may be recovered by application of the mapping $f(x) = (2\exp(x) - 1)/(2\exp(x) + 1)$.

An asterisk indicates that the correlation is significantly different from zero at the 5% level, and a double asterisk at the 1% level. As we see in Table 5.1, in half of the cases, the 95% confidence intervals on the pairwise correlations do not contain zero when $T = 100$. That is, for $T = 100$, these correlations are not significantly different from zero at the 5% significance level. What is interesting is that there is no statistically significant correlation between the inference

Table 5.1: Pairwise sample correlations (with 95% confidence intervals) of some specification and mis-specification tests in the AR(2) defined in (5.1), $T = 100$, $M = 10^5$

	$\mathsf{t}_{\mathsf{H}_0:\alpha=0.4}$	$\mathsf{LR}_{\mathsf{H}_m:\alpha_3=0}$	$\mathsf{LR}^{\mathsf{ur}}$	$\mathsf{CUSQ}^{\mathsf{REC}}$	$\mathsf{JB}^{\mathsf{omnibus}}$
$\mathsf{t}_{\mathsf{H}_0:\alpha=0.4}$	1	−0.004 (−0.010,0.002)	−0.001 (−0.007,0.005)	−0.001 (−0.007,0.006)	−0.005 (−0.011,0.001)
$\mathsf{LR}_{\mathsf{H}_{\mathsf{mis}}:\alpha_3=0}$		1	0.042 (0.036,0.048)**	0.028 (0.022,0.034)**	0.083 (0.078,0.089**)
$\mathsf{LR}_{\mathsf{ur}}$			1	0.000 (−0.007,0.006)	0.007 (0.001,0.014)**
$\mathsf{CUSQ}^{\mathsf{REC}}$				1	0.169 (0.163,0.175)**
$\mathsf{JB}^{\mathsf{omnibus}}$					1

procedure used within the chosen model and any of the mis-specification tests. Similarly, we cannot reject zero correlation between the unit root test and the variance constancy test. When the sample size increases, to $T = 500$, only three out of the ten correlations remain significant, all involving the Jarque-Bera normality test. As the sample size is further increased to $T = 1000$ only two out of the ten correlations are significant at the 1% level, again both involving the Jarque-Bera test.

6 Conclusion

We make two contributions in this chapter. First, we show precisely how the diagnostic testing information is, as Hendry (1995, p.545) puts it, *effectively independent of the sufficient statistics*. There is no need to discount the results. Second, we show that the diagnostic tests themselves are mutually independent. The investigator can therefore control the overall size of the statistical proce-

Table 5.2: Pairwise sample correlations (with 95% confidence intervals) of some specification and mis-specification tests in the AR(2) defined in (5.1), $T = 500$, $M = 10^5$

	$\mathsf{t}_{\mathsf{H}_0:\alpha=0.4}$	$\mathsf{LR}_{\mathsf{H}_m:\alpha_3=0}$	$\mathsf{LR}^{\mathsf{ur}}$	$\mathsf{CUSQ}^{\mathsf{REC}}$	$\mathsf{JB}^{\mathsf{omnibus}}$
$\mathsf{t}_{\mathsf{H}_0:\alpha=0.4}$	1	−0.002 (−0.009,0.004)	−0.005 (−0.011,0.001)	0.002 (−0.005,0.008)	−0.009 (−0.015,−0.003)**
$\mathsf{LR}_{\mathsf{H}_{\mathsf{mis}}:\alpha_3=0}$		1	0.004 (−0.002,0.011)	0.005 (−0.001,0.011)	0.018 (0.012,0.025)**
$\mathsf{LR}_{\mathsf{ur}}$			1	−0.003 (−0.009,0.003)	0.003 (−0.003,0.009)
$\mathsf{CUSQ}^{\mathsf{REC}}$				1	0.071 (0.064,0.077)**
$\mathsf{JB}^{\mathsf{omnibus}}$					1

Table 5.3: Pairwise sample correlations (with 95% confidence intervals) of some specification and mis-specification tests in the AR(2) defined in (5.1), $T = 1000$, $M = 10^5$

	$\mathsf{t}_{\mathsf{H}_0:\alpha=0.4}$	$\mathsf{LR}_{\mathsf{H}_m:\alpha_3=0}$	$\mathsf{LR}^{\mathsf{ur}}$	$\mathsf{CUSQ}^{\mathsf{REC}}$	$\mathsf{JB}^{\mathsf{omnibus}}$
$\mathsf{t}_{\mathsf{H}_0:\alpha=0.4}$	1	−0.006 (−0.012,0.000)	−0.001 (−0.007,0.005)	0.001 (−0.005,0.008)	−0.003 (−0.009,0.004)
$\mathsf{LR}_{\mathsf{H}_{\mathsf{mis}}:\alpha_3=0}$		1	0.003 (−0.003,0.009)	0.000 (−0.006,0.006)	0.010 (0.004,0.016)**
$\mathsf{LR}_{\mathsf{ur}}$			1	−0.004 (−0.010,0.002)	0.001 (−0.005,0.007)
$\mathsf{CUSQ}^{\mathsf{REC}}$				1	0.041 (0.035,0.047)**
$\mathsf{JB}^{\mathsf{omnibus}}$					1

dure with ease. While more work is needed to cover further test procedures and more general model classes (an interesting extension would be the model class of Chapter 2), this work is an improvement on previous literature in the sense that more tests are covered, and covered in a more general model class. Due to the temporal dependence allowed in the model class, Basu's theorem ((Basu 1955) is not applicable to functions of the residuals and estimators. This is why exact results have not been obtained, and further work is needed to assess how good the asymptotic approximations are.

A Appendix

Proofs

Notation: For a vector m of length k, we let Euclidean norm be denoted $\|m\| = (m_1^2 + \ldots + m_k^2)^{1/2}$, and the supremum norm $\|m\|_\infty = \sup_j |m_j|,\ j = 1, \ldots, k$. For any real number r, int(r) denotes its integer part. D[0,1] denotes the cádlág space of right continuous functions with left limits on the interval [0,1].

First, we provide proofs of the auxiliary results used to reduce the tests considered to analyzable entities, functionals of quantities that we show to converge jointly and to be asymptotically independent in the limit. These quantities involve martingale transforms with the stationary part of the process, U_t, as well as of the $\mathcal{Y}_t$-process (appearing in the limit of the likelihood ratio lag length test), random functions defined by the partial sums of the first two powers of the innovations, and sums of the third and fourth powers of the standardized innovations. Then, having proved a result concerning joint convergence involving a stochastic integral, we give proofs of the theorems.

A.1 Auxiliary results on the mis-specification tests

Proof of Lemma 4.1. This is Theorem 3 in Nielsen (2001). Strictly speaking, our Assumption A is unnecessary for the theorem to hold, but not making it would complicate the subsequent arguments. □

Proof of Lemma 4.3. Given Assumption B, the form (4.1) follows by Lemmas 3.2, 3.3, 3.4, 3.5, 3.6, 3.8, 3.9, and 3.10 in Nielsen (2006), and the asymptotic distribution by Nielsen (2006, Thm. 2.3). As in the case of the unit root test, our Assumption A is unnecessary. Not making it would require us to separately characterize inference on the stationary and explosive parts of the process, however, and would therefore complicate matters. □

Proof of Lemma 4.4. When the test is constructed with ordinary least squares (recursive forecast) residuals, the first result follows by Lemma 4.1 (Lemma 5.1) of Chapter 2 and the asymptotic distribution by Theorem 4.4 (i) (Theorem 5.3 (i)) of Chapter 2 by Assumptions B and C. (A, B, and C). □

Proof of Lemma 4.5. The result follows by Proposition 2 in Kilian and Demiroglu (2000). While their argument is made for the case of a vector autoregression, the result we state follows as a special case when the dimension of the VAR is set to one and its lag length to k. □

A.2 Joint behaviour of the mis-specification tests

Recall that U_t denotes the stationary stochastic part of the data generation process and $\mathcal{Y}_t$, arising as the leading term in an asymptotic analysis of the lag length test, a linear combination of U_{t-1} and ε_t that is independent of U_t.

As in Chapter 2, shorthand notation is used to indicate weak convergence of a random function to a Brownian motion $\mathcal{B}$ on the unit interval. When we say that

for $u \in [0,1]$, $G_{\text{int}(Tu)} \to \mathcal{B}_u$ in distribution on $D[0,1]$, we mean that for some g_s, G_T is the random element of $D[0,1]$ (equipped with the uniform metric) defined by $G_T(u) = T^{-1/2}\sum_{s=1}^{\text{int}(Tu)} g_s$, and $G_T \to \mathcal{B}$ in distribution on $D[0,1]$.

Lemma A.1. *Assume A and B. Then, for $u \in [0,1]$*

$$\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} U_{t-1}\varepsilon_t/\sigma,\ \sum_{t=1}^{T}\mathcal{Y}_{t-1}\varepsilon_t/\sigma,\ \sum_{s=1}^{int(Tu)}\varepsilon_s/\sigma\right) \to (\mathcal{B}_{U,1},\ \mathcal{B}_{\mathcal{Y},1},\ \mathcal{B}_{1,u}) \tag{A.1}$$

in distribution on $\mathbb{R}^2 \times D[0,1]$ as $T \to \infty$, where $\mathcal{B}_{U,1}$ and $\mathcal{B}_{\mathcal{Y},1}$ are independent standard normal distributions, and both are independent of the standard Brownian motion $\mathcal{B}_{1,u}$.

Proof of Lemma A.1. *Convergence of marginal distributions:* Write $\mathcal{M}_s = T^{-1/2}\zeta_{s-1}\varepsilon_s$, where $\zeta_{s-1} = (U'_{s-1}, \mathcal{Y}'_{s-1}, 1)'$ is $\mathcal{F}_{s-1}$-measurable. Therefore, under Assumption B, convergence of each element follows by the functional central limit theorem for martingale difference sequences by Helland (1982, Thm 3.2) (see also Chan and Wei (1988, Thm. 2.2)) under the conditions

$$\sum_{s=1}^{\text{int}(Tu)} \mathsf{E}(\|\mathcal{M}_s\|^2|\mathcal{F}_{s-1}) = \mathsf{O}_\mathsf{P}(1) \tag{A.2}$$

$$\sum_{s=1}^{\text{int}(Tu)} \mathsf{E}\{\|\mathcal{M}_s\|^2 I(\|\mathcal{M}_s\| > \epsilon)|\mathcal{F}_{s-1}\} \xrightarrow{\mathsf{P}} 0 \quad \text{for } \epsilon > 0. \tag{A.3}$$

Using Assumption B, in order to show (A.2) it is sufficient to show that

$$T^{-1}\sigma^2\Big(\sum_{s=1}^{\text{int}(Tu)} U_{s-1}U'_{s-1},\ \sum_{s=1}^{\text{int}(Tu)}\mathcal{Y}_{s-1}\mathcal{Y}'_{s-1}, \text{int}(Tu)\Big)' = \mathsf{O}(1)\ a.s. \tag{A.4}$$

where the order follows by Nielsen (2005, Thm. 7.1), and this implies boundedness in probability. To verify the Lindeberg condition, apply the conditional Markov

inequality (see Davidson (1994, 10.17, p.152)) for $\kappa, \epsilon > 0$ and rewrite (A.3) as

$$\begin{aligned} T^{-1} \sum_{s=1}^{\text{int}(Tu)} & \mathsf{E}\{\|\zeta_{s-1}\varepsilon_s\|^2 I(\|\zeta_{s-1}\varepsilon_s\| > T^{1/2}\epsilon)|\mathcal{F}_{s-1}\} \\ & \leq T^{-1} \sum_{s=1}^{\text{int}(Tu)} (T^{1/2}\epsilon)^{-\kappa} \mathsf{E}\{\|\zeta_{s-1}\varepsilon_s\|^{2+\kappa}|\mathcal{F}_{s-1}\} \\ & = T^{-1} \sum_{s=1}^{\text{int}(Tu)} (T^{1/2}\epsilon)^{-\kappa} \|\zeta_{s-1}\|^{2+\kappa} \mathsf{E}\{|\varepsilon_s|^{2+\kappa}|\mathcal{F}_{s-1}\} \\ & \leq T^{-1} \sum_{s=1}^{\text{int}(Tu)} K\epsilon^{-\kappa} \|\zeta_{s-1}\|^2 (T^{-1} \max_{0\leq s\leq \text{int}(Tu)} \|\zeta_s\|^2)^{\kappa/2} = \mathsf{o_P}(1), \end{aligned} \tag{A.5}$$

where we have replaced $\mathsf{E}\{|\varepsilon_s|^{2+\kappa}|\mathcal{F}_{s-1}\}$ with a constant $K > 0$ using Assumption B and the asymptotic order follows by (A.4).

Joint convergence and independence: We apply Helland (1982, Thm 3.3) (see also Chan and Wei (1988, Thm. 2.2)) under the condition

$$\sum_{s=1}^{\text{int}(Tu)} \mathsf{E}(\mathcal{M}_s^i \mathcal{M}_s^{j\prime}|\mathcal{F}_{s-1}) \xrightarrow{\mathsf{P}} 0 \quad \text{for } i \neq j \tag{A.6}$$

where $\mathcal{M}_s^i$ is the i^{th} element of $\mathcal{M}_s$. Using the definition of $\mathcal{M}_s$ and Assumption B, we may rewrite this as

$$T^{-1} \sum_{s=1}^{\text{int}(Tu)} \sigma^2(U_{s-1}\mathcal{Y}_{s-1}', \ U_{s-1}, \ \mathcal{Y}_{s-1}) \xrightarrow{\mathsf{P}} 0, \tag{A.7}$$

since we are effectively showing that the off-diagonal elements of the matrix $\sigma^2 T^{-1} \sum_{s=1}^{\text{int}(Tu)} (U_{s-1}, \mathcal{Y}_{s-1}, 1)'(U_{s-1}, \mathcal{Y}_{s-1}, 1)$ vanish in probability. The first statement in (A.7) holds by Lemma 4.3 which allows us to write $\mathcal{Y}_{s-1} = C'(U_{s-2}, \varepsilon_{s-1})'$

and $U_{s-1} = (0,1)(\varepsilon_s, U_{s-1})' = C'_\perp(U_{s-2}, \varepsilon_{s-1})'$. It follows that

$$\mathsf{Cov}(Y_{s-1}, U_{s-1}) = C' \begin{pmatrix} F & 0 \\ 0 & \sigma^2 \end{pmatrix} C_\perp = 0, \tag{A.8}$$

where we follow Nielsen (2006), in letting the variance of the stationary part be denoted $F = \lim_{T\to\infty} \sum_{t=1}^{T} U_{t-1}U'_{t-1}$. The second and third statements of (A.7) hold by Nielsen (2005, Th. 5.1). □

We now extend the result to include squares of the innovations.

Lemma A.2. *Assume A, B, C, and D. Then,*

$$\frac{1}{\sqrt{T}} \left\{ \sum_{t=1}^{T} \begin{pmatrix} U_{t-1} \\ \mathcal{Y}_{t-1} \end{pmatrix}' \frac{\varepsilon_t}{\sigma}, \sum_{s=1}^{int(Tu)} \begin{pmatrix} \varepsilon_s/\sigma \\ \varepsilon_s^2 - \sigma^2 \end{pmatrix}' \right\} \to (\mathcal{B}_{U,1},\ \mathcal{B}_{\mathcal{Y},1},\ \mathcal{B}_{1,u},\ \varphi\mathcal{B}_{2,u}) \tag{A.9}$$

in distribution on $\mathbb{R}^2 \times D[0,1]^2$ as $T \to \infty$, where $\mathcal{B}_{U,1}$ and $\mathcal{B}_{\mathcal{Y},1}$ are independent standard normal distributions, each independent of the standard Brownian motion $\mathcal{B}_{1,u}$ and of the scaled standard Brownian motion $\varphi\mathcal{B}_{2,u}$. The latter two are also mutually independent.

Proof of Lemma A.2. *Convergence of marginal distributions:* As in the proof of Lemma A.1, we apply the central limit theorem for martingale difference sequences (Helland (1982, Thm 3.2)) to show convergence. We can do this because by Assumption C, ε_s^2 has a martingale structure. The conditions (A.2) and (A.3) must be shown to apply with $\mathcal{M}_s$ extended with $T^{-1/2}(\varepsilon_s^2 - \sigma^2)$. The former is guaranteed by Assumption C since it ensures that $E\{(\varepsilon_s^2 - \sigma^2)^2 | \mathcal{F}_{s-1}\}$ is bounded almost surely and therefore in probability also. To show the Lindeberg condition,

write for some $\kappa, \epsilon > 0$,

$$\begin{aligned} T^{-1} \sum_{s=1}^{\text{int}(Tu)} \mathsf{E}\{(\varepsilon_s^2 - \sigma^2)^2 I(|\varepsilon_s^2 - \sigma^2| > T^{1/2}\epsilon)|\mathcal{F}_{s-1}\} \\ \leq T^{-(1+\frac{\kappa}{2})} \epsilon^{-\kappa} \sum_{s=1}^{\text{int}(Tu)} \mathsf{E}\{|\varepsilon_s^2 - \sigma^2|^{2+\kappa}|\mathcal{F}_{s-1}\} \\ = \mathsf{o_P}(1), \end{aligned} \tag{A.10}$$

by the conditional Markov inequality and Assumption B.

Joint convergence and independence: The condition (A.6) must be shown to apply with the extended $\mathcal{M}_t$-vector. In particular, we have

$$\mathcal{M}_s = T^{-1/2}(U_{s-1}'\varepsilon_s,\ \mathcal{Y}_{s-1}'\varepsilon_s,\ \varepsilon_s,\ \varepsilon_s^2 - \sigma^2).$$

Therefore, given the proof of Lemma A.1, to show (A.6), it remains to show

$$\begin{aligned} T^{-1} \sum_{s=1}^{\text{int}(Tu)} \mathsf{E}[(U_{s-1}'\varepsilon_s(\varepsilon_s^2 - \sigma^2)|\mathcal{F}_{s-1}] \xrightarrow{\mathsf{P}} 0 \\ T^{-1} \sum_{s=1}^{\text{int}(Tu)} \mathsf{E}[\mathcal{Y}_{s-1}'\varepsilon_s(\varepsilon_s^2 - \sigma^2)|\mathcal{F}_{s-1}] \xrightarrow{\mathsf{P}} 0 \\ T^{-1} \sum_{s=1}^{\text{int}(Tu)} \mathsf{E}[(\varepsilon_s(\varepsilon_s^2 - \sigma^2)|\mathcal{F}_{s-1}] \xrightarrow{\mathsf{P}} 0 \end{aligned}$$

Since ε_s and ε_s^3 are martingale difference sequences and U_{s-1} and $\mathcal{Y}_{s-1}$ are $\mathcal{F}_{s-1}$-measurable processes, the above conditions hold trivially (the quantities of interest are equal to zero). □

We now extend the result to include higher powers of the innovations.

Lemma A.3. *Assume A, B, C, D, and E. Then,*

$$\frac{1}{\sqrt{T}}\left\{\sum_{t=1}^{T}\begin{pmatrix} U_{t-1} \\ \mathcal{Y}_{t-1}\end{pmatrix}'\frac{\varepsilon_t}{\sigma},\ \sum_{s=1}^{int(Tu)}\begin{pmatrix} \varepsilon_s/\sigma \\ \varepsilon_s^2-\sigma^2\end{pmatrix}',\begin{pmatrix} TS/\sqrt{6} \\ TK/\sqrt{24}\end{pmatrix}'\right\}$$
$$\to (\mathcal{B}_{U,1},\ \mathcal{B}_{\mathcal{Y},1},\ \mathcal{B}_{1,u},\ \varphi\mathcal{B}_{2,u},\ \mathcal{B}_{3,1},\ \mathcal{B}_{4,1}) \tag{A.11}$$

in distribution on $\mathbb{R}^4 \times D[0,1]^2$ *as* $T \to \infty$*, where* S *and* K *are defined by* (4.3) *and* $\mathcal{B}_{U,1}$*,* $\mathcal{B}_{\mathcal{Y},1}$*,* $\mathcal{B}_{3,1}$*, and* $\mathcal{B}_{4,1}$ *are independent standard normal distributions, each independent of the standard Brownian motion* $\mathcal{B}_{1,u}$ *and the scaled standard Brownian motion* $\varphi\mathcal{B}_{2,u}$*. The latter two are also mutually independent.*

Proof of Lemma A.3. We first expand

$$T^{-1}\textstyle\sum_{t=1}^{T}(\varepsilon_t-\bar{\varepsilon})^2 = T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 + \mathrm{o_P}(1), \tag{A.12}$$

so that in dealing with TS and TK it suffices to consider the numerators

$$\sum_{t=1}^{T}(\varepsilon_t-\bar{\varepsilon})^3 = \sum_{t=1}^{T}\varepsilon_t^3 - 3\sigma^2\sum_{t=1}^{T}\varepsilon_t + \mathrm{o_P}(T^{1/2}) \tag{A.13}$$

$$\sum_{t=1}^{T}(\varepsilon_t-\bar{\varepsilon})^4 - 3(T^{-1}\sum_{t=1}^{T}(\varepsilon_t-\bar{\varepsilon})^2)^2$$
$$= \sum_{t=1}^{T}(\varepsilon_t^4-3\sigma^4) - 6\sigma^2\sum_{t=1}^{T}(\varepsilon_t^2-\sigma^2) + \mathrm{o_P}(T^{1/2}) \tag{A.14}$$

These expansions imply that, since normalization is by $T^{1/2}$, the remainder terms are small in probability. Since the leading terms are linear functions of martingale difference sequences we again apply Helland (1982, Thm 3.2). First, (A.2) and (A.3) must be shown to apply with $\mathcal{M}_s$ extended with $T^{-1/2}\{(\varepsilon_s^2-\sigma^2),\ (\varepsilon_s^3-3\sigma^2\varepsilon_s),\ (\varepsilon_s^4-3\sigma^4-6\sigma^2(\varepsilon_s^2-\sigma^2))\}'$ Independence follows if (A.6) applies with this extended $\mathcal{M}_s$-vector. We start by verifying the convergence of marginal

distributions.

Convergence of marginal distributions: Given the expansions (A.13) and (A.14), note that (A.2) follows if $T^{-1}\sum_{s=1}^{\text{int}(Tu)}\mathsf{E}(|\varepsilon_s|^\lambda|\mathcal{F}_{s-1})$ is bounded for some $\lambda > 8$ as the highest order moment we require to be bounded in verifying (A.2) is the eighth. This holds by Assumption E, which is in fact stronger an assumption than it is necessary to make in order to establish the convergence of marginal distributions. (In the independence argument something more than mere boundedness of eighth order moments will be used.) To show (A.3), consider first the term $T^{-1/2}(\varepsilon_s^3 - 3\sigma^2\varepsilon_s)$. Write for $\kappa, \epsilon > 0$,

$$\begin{aligned} T^{-1}\sum_{s=1}^{\text{int}(Tu)}&\mathsf{E}\{(\varepsilon_s^3 - 3\sigma^2\varepsilon_s)^2 I(|\varepsilon_s^3 - 3\sigma^2\varepsilon_s| > T^{1/2}\epsilon)|\mathcal{F}_{s-1}\} \\ &\leq T^{-(1+\frac{\kappa}{2})}\epsilon^{-\kappa}\sum_{s=1}^{\text{int}(Tu)}\mathsf{E}\{|(\varepsilon_s^3 - 3\sigma^2\varepsilon_s)|^{2+\kappa}|\mathcal{F}_{s-1}\} \\ &= \mathrm{o}_\mathsf{P}(1), \end{aligned} \tag{A.15}$$

again using the conditional Markov inequality. The order follows from Assumption E. A similar argument can made with $T^{-1/2}\{\varepsilon_s^4 - 3\sigma^4 - 6\sigma^2(\varepsilon_s^2 - \sigma^2)\}$, by application of (A.14), and boundedness of the relevant conditional moments is guaranteed by Assumption E.

Joint convergence and independence: Since U_{s-1}, $\mathcal{Y}_{s-1}$ are $\mathcal{F}_{s-1}$-measurable, whenever $\mathcal{M}_s^i$ is $T^{-1/2}U_{s-1}\varepsilon_s$ or $T^{-1/2}\mathcal{Y}_{s-1}\varepsilon_s$, we can make use of the expansions (A.13), and (A.14), such that (A.6) is verified with $\mathcal{M}_s^j = T^{-1/2}(\varepsilon_s^3 - 3\sigma^2\varepsilon_s) + \mathrm{o}_\mathsf{P}(1)$ or $\mathcal{M}_s^j = T^{-1/2}\{\varepsilon_s^4 - 3\sigma^4 - 6\sigma^2(\varepsilon_s^2 - \sigma^2)\} + \mathrm{o}_\mathsf{P}(1)$ as the sum of conditional expectations of the cross products $\mathcal{M}_s^i\mathcal{M}_s^{j\prime}$ then vanishes in probability by Assumption E and by Nielsen (2005, Thm. 5.1) because the stationary processes U_{s-1} and $\mathcal{Y}_{s-1}$ are $\mathcal{F}_{s-1}$-measurable.

In order to verify concisely the the asymptotic orders of the remaining sums of conditional expectations, we write the entire quantity of interest as

$$\begin{aligned}
&\sum_{s=1}^{\text{int}(Tu)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3\sigma^2 & 0 & 1 & 0 \\ 0 & -6\sigma^2 & 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_s \\ \varepsilon_s^2 - \sigma^2 \\ \varepsilon_s^3 \\ \varepsilon_s^4 - 3\sigma^4 \end{pmatrix} + o_{\mathsf{P}}(T^{1/2}) \\
&= \sum_{s=1}^{\text{int}(Tu)} \mathcal{C}\xi_s + o_{\mathsf{P}}(T^{1/2}) \\
&= T^{1/2} \sum_{s=1}^{\text{int}(Tu)} \begin{pmatrix} \mathcal{M}_s^1 \\ \mathcal{M}_s^2 \\ \mathcal{M}_s^3 \\ \mathcal{M}_s^4 \end{pmatrix} + o_{\mathsf{P}}(T^{1/2})
\end{aligned}$$

For asymptotic independence, it must be shown that (A.6) holds. That is, $\sum_{s=1}^{\text{int}(Tu)} \mathsf{E}(\mathcal{M}_s^i \mathcal{M}_s^{j\prime} | \mathcal{F}_{s-1}) \xrightarrow{\mathsf{P}} 0$ for $i \neq j$. Given the expansion above, it is sufficient to show that $T^{-1} \sum_{s=1}^{\text{int}(Tu)} \mathcal{C}\mathsf{E}(\xi_s \xi_s' | \mathcal{F}_{s-1}) \mathcal{C}'$ is a diagonal matrix. We first compute the conditional expectation under the assumption of normality.

$$\mathsf{E}(\xi_s \xi_s' | \mathcal{F}_{s-1}) = \begin{pmatrix} \sigma^2 & 0 & 3\sigma^4 & 0 \\ 0 & 2\sigma^4 & 0 & 12\sigma^6 \\ 3\sigma^4 & 0 & 15\sigma^6 & 0 \\ 0 & 12\sigma^6 & 0 & 96\sigma^8 \end{pmatrix}$$

Then, we carry out the multiplication $\mathcal{C}\mathsf{E}(\xi_s\xi_s'|\mathcal{F}_{s-1})\mathcal{C}'$, that is,

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3\sigma^2 & 0 & 1 & 0 \\ 0 & -6\sigma^2 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma^2 & 0 & 3\sigma^4 & 0 \\ 0 & 2\sigma^4 & 0 & 12\sigma^6 \\ 3\sigma^4 & 0 & 15\sigma^6 & 0 \\ 0 & 12\sigma^6 & 0 & 96\sigma^8 \end{pmatrix} \begin{pmatrix} 1 & 0 & -3\sigma^2 & 0 \\ 0 & 1 & 0 & -6\sigma^2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and obtain the diagonal matrix

$$\begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & 2\sigma^4 & 0 & 0 \\ 0 & 0 & 6\sigma^6 & 0 \\ 0 & 0 & 0 & 24\sigma^8 \end{pmatrix}.$$

This implies that $T^{-1}\sum_{s=1}^{\mathrm{int}(Tu)} \mathcal{C}\mathsf{E}(\xi_s\xi_s'|\mathcal{F}_{s-1})\mathcal{C}'$ is a diagonal matrix, which in turn implies (A.6), that is, $\sum_{s=1}^{\mathrm{int}(Tu)} \mathsf{E}(\mathcal{M}_s^i\mathcal{M}_s^{j\prime}|\mathcal{F}_{s-1}) \xrightarrow{\mathsf{P}} 0$ for $i \neq j$.

□

Each of the tests given in (4.6) may be constructed from the random vectors defined in (A.11) by the application of an appropriate functional. The limit distribution of the unit root test involves the stochastic integral, $\int_0^1 \mathcal{B}_{u,1}d\mathcal{B}_{u,1}$. This causes complications, as the set of discontinuities associated with this functional is not small in Lebesgue-measure and the application of the mapping theorem in Billingsley (1999, p. 20) is ruled out. Consequently, establishing joint convergence and independence of the tests becomes more complicated. To solve this problem, we adopt the strategy followed by Chan and Wei (1988).

Heuristically speaking, this necessitates the construction of a so called Skorokhod embedding, essentially establishing copies, equal in distribution to the

quantities of interest. The existence of these quantities on the same probability space as the original ones is guaranteed by Skorokhod's representation theorem (see Billingsley (1999, Thm 6.7)). The theorem also guarantees almost sure convergence of the copies to the common stochastic limits. An application of Egoroff's theorem (see Davidson (1994, 18.4, p. 283)) shows that the functionals of the desired quantities converge to functionals of the stochastic limits in probability. Equality in distribution of the original quantities and the copies then ensures that the same functionals of the original quantities converge to the same limits in distribution. Once this has been established, the quantities of interest have been defined on the same probability space and an independence argument can be formulated. This is made precise in the proof of the lemma below. We let $\mathcal{B}_{U,1}$, $\mathcal{B}_{\mathcal{Y},1}$, $\mathcal{B}_{1,u}$, $\mathcal{B}_{2,u}$, $\mathcal{B}_{3,1}$, and $\mathcal{B}_{4,1}$ be defined as in Lemmas A.1, A.2, and A.3.

Lemma A.4. *Let* $N_T = \sum_{t=1}^{T}(U_{t-1}, \mathcal{Y}_{t-1})\varepsilon_t/\sigma$ *and* $V_t = \sum_{s=1}^{t}\varepsilon_s/\sigma$. *Then, under Assumptions A and B, it holds*

$$T^{-1/2}(N_T,\ V_t,\ T^{-1/2}\sum_{t=1}^{T} V_{t-1}\varepsilon_t/\sigma\ ) \to \{(\mathcal{B}_{U,1},\ \mathcal{B}_{\mathcal{Y},1}),\ \mathcal{B}_{1,u},\ \textstyle\int_0^1 \mathcal{B}_{1,u}d\mathcal{B}_{1,u}\} \quad \text{(A.16)}$$

in distribution on $\mathbb{R}^2 \times D[0,1] \times \mathbb{R}$ *as* $T \to \infty$ *and the first two elements of the asymptotic distribution are mutually independent and also independent of* $(\mathcal{B}_{1,u}, \int_0^1 \mathcal{B}_{1,u}dB_{1,u})$.

Proof of Lemma A.4. This is an extension and modification of the proof of Theorem 2.4 by Chan and Wei (1988). Let $G_T = T^{-1}\sum_{t=1}^{T} V_{t-1}\varepsilon_t/\sigma$.

Convergence: By Lemma A.1 and Skorokhod's representation theorem, on $(\Omega, \mathcal{F}, \mathsf{P})$, there exist $\tilde{N}_T$ and $\tilde{V}_t$ in $D[0,1]$ such that $(\tilde{N}_T, \tilde{V}_t) = (N_T, V_t)$ holds in distribution

and

$$\sup_t |(T^{-1/2}\tilde{N}_T,\ T^{-1/2}\tilde{V}_t) - (\mathcal{B}_{U,1},\ \mathcal{B}_{\mathcal{Y},1},\ \mathcal{B}_{1,u})| \to 0 \quad a.s. \tag{A.17}$$

as $T \to \infty$. Note that, given Lemma A.1, $\tilde{N}_T$ and $\tilde{V}_t$ are asymptotically mutually independent. We then construct $\tilde{G}_T = T^{-1}\sum_{t=1}^{T}\tilde{V}_{t-1}(\tilde{V}_t - \tilde{V}_{t-1})$ and it follows that, in distribution,

$$(T^{-1/2}\tilde{N}_T,\ T^{-1/2}\tilde{V}_t,\ \tilde{G}_T\ ) = (T^{-1/2}N_T,\ T^{-1/2}V_t,\ G_T) \tag{A.18}$$

The convergence $\tilde{G}_T \xrightarrow{\mathsf{P}} \int_0^1 \mathcal{B}_{1,u}d\mathcal{B}_{1,u}$ can be shown by the argument on pp. 376-377 of Chan and Wei (1988) as the set on which the distance defined by the left hand side of (A.17) vanishes has high probability. That is, by Egoroff's theorem, for $\epsilon > 0$ and for each $\omega \in \Omega_\epsilon \subset \Omega$ such that $\mathsf{P}(\Omega_\epsilon) \geq 1 - \epsilon$

$$\sup_{\omega\in\Omega_\epsilon}\{\sup_t |(T^{-1/2}\tilde{N}_T,\ T^{-1/2}\tilde{V}_t) - (\mathcal{B}_{U,1},\ \mathcal{B}_{\mathcal{Y},1},\ \mathcal{B}_{1,u})|\} \to 0 \tag{A.19}$$

Therefore, combining statements (2.25), (2.27), and (2.28) of Chan and Wei (1988), we can write

$$\mathbf{1}_{\Omega_\epsilon}\tilde{G}_T = \mathbf{1}_{\Omega_\epsilon}\int_0^1 \mathcal{B}_{1,u}d\mathcal{B}_{1,u} + o_{\mathsf{P}}(1) \tag{A.20}$$

Independence: The application of Skorokhod's representation theorem ensures that the random functions of interest are defined on the same probability space $(\Omega, \mathcal{F}, \mathsf{P})$. The above convergence argument allows us to write, for an $\mathcal{F}$-measurable function n_T, such that $n_T(\mathcal{B}_1) = \sum_{t=1}^{T}\mathcal{B}_{1,\frac{t-1}{T}}(\mathcal{B}_{1,\frac{t}{T}} - \mathcal{B}_{1,\frac{t-1}{T}})$, and for any three

events $\{A_1, A_2, A_3\} \in \mathcal{F}$, and a large enough T,

$$\begin{aligned}
&\mathsf{P}(T^{-1/2}\tilde{N}_T \in A_1,\ T^{-1/2}\tilde{V}_{T,t} \in A_2,\ \tilde{G}_T \in A_3) \\
&\qquad \approx \mathsf{P}\{(\mathcal{B}_U, \mathcal{B}_\mathcal{Y}) \in A_1,\ \mathcal{B}_1 \in A_2,\ n_T(\mathcal{B}_1) \in A_3\} \\
&\qquad = \mathsf{P}\{(\mathcal{B}_U, \mathcal{B}_\mathcal{Y}) \in A_1\}\mathsf{P}\{\mathcal{B}_1 \in A_2,\ n_T(\mathcal{B}_1) \in A_3\} \\
&\qquad \approx \mathsf{P}\{(\mathcal{B}_U, \mathcal{B}_\mathcal{Y}) \in A_1\}\mathsf{P}\{\mathcal{B}_1 \in A_2, \int_0^1 \mathcal{B}_{1,u} d\mathcal{B}_{1,u} \in A_3\} \qquad \text{(A.21)}
\end{aligned}$$

where the first approximation follows by (A.19) together with the definition of $\tilde{G}_T$, and the equality by Lemma A.1 and because n_T is an $\mathcal{F}$-measurable function. The second approximation follows by (A.20) together with $\mathsf{P}(\Omega_\epsilon) \geq 1 - \epsilon$. □

Lemma A.5. *Let N_T, $V_{T,t}$, and G_T be as defined in Lemma A.4 and define $S_{T,t} = \sum_{s=1}^{t}(\varepsilon_s^2 - \sigma^2)$. Then, under Assumptions A, B, C, and D, both* (A.16) *and the mutual independence of the elements of the asymptotic distribution hold as described, with the left hand side extended with $T^{-1/2}S_{T,t}$ and the right hand side with $\varphi\mathcal{B}_{2,u}$, so convergence is takes place on $\mathbb{R}^2 \times D[0,1] \times \mathbb{R} \times D[0,1]$. Furthermore, $\varphi\mathcal{B}_{2,u}$ is independent of the other elements of the asymptotic distribution.*

Proof of Lemma A.5. Apply Lemma A.2 and the proof of Lemma A.4, mutatis mutandis. □

Lemma A.6. *Let N_T, $V_{T,t}$, $S_{T,t}$, and G_T be as defined in Lemmas A.4 and A.5, and define $J_T = \{\sum_{t=1}^{T}(\varepsilon_t - \bar{\varepsilon})^3/\sqrt{6\sigma^6},\ (\sum_{t=1}^{T}(\varepsilon_t - \bar{\varepsilon})^4 - 3\sigma^4)/\sqrt{24\sigma^8}\}$. Then, under Assumptions A, B, C, D, and E, both* (A.16) *and the mutual independence of the elements of the asymptotic distribution hold as described, with the left hand side extended with $T^{-1/2}(S_{T,t}, J_T)$ and the right hand side with $(\varphi\mathcal{B}_{2,u},\ \mathcal{B}_{3,1},\ \mathcal{B}_{4,1})$, so convergence takes place on $\mathbb{R}^2 \times D[0,1] \times \mathbb{R} \times D[0,1] \times \mathbb{R}^2$. Furthermore, $\varphi\mathcal{B}_{2,u}$*

and $(\mathcal{B}_{3,1},\ \mathcal{B}_{4,1})$ *are mutually independent and independent of the other elements of the asymptotic distribution.*

Proof of Lemma A.6. Apply Lemma A.3 and the proof of Lemma A.4, mutatis mutandis. □

Proof of Theorem 4.6. By equation (3.3), and Lemma and 4.3, we may instead of $s_{\gamma_2,T}^{1/2}\iota_{k-1}(\hat{\delta}_{k-1}-\delta_{k-1})$ and $\mathsf{LR}_{\mathsf{H}_m}$ consider, respectively, $\mathcal{S}^{\mathsf{inf}}$ and $\mathcal{S}^{\mathsf{lag}}$ written for a continuous mapping $f:(\mathbb{R}\times\mathbb{R}_+)\mapsto\mathbb{R}$, as

$$\mathcal{S}^{\mathsf{inf}}=f\left(\begin{array}{c}\sum_{t=1}^{T}U_{t-1}\varepsilon_t\\ \sum_{t=1}^{T}U_{t-1}U_{t-1}'\end{array}\right),\ \mathcal{S}^{\mathsf{lag}}=f\left(\begin{array}{c}\sum_{t=1}^{T}\mathcal{Y}_{t-1}\varepsilon_t\\ \sum_{t=1}^{T}\mathcal{Y}_{t-1}\mathcal{Y}_{t-1}'\end{array}\right) \tag{A.22}$$

Furthermore, using Lemma 4.1 and the argument of Lemma A.4, we may instead of $\mathsf{LR}^{\mathsf{ur}}$, for continuous mappings $g_1:(D[0,1]\times\mathbb{R})\mapsto(\mathbb{R}_+\times\mathbb{R})$ with $u\in[0,1]$, and $g_1:(\mathbb{R}_+\times\mathbb{R})\mapsto\mathbb{R}$, write $g_1\circ g_2(\tilde{G}_T)$. Setting $U_{t-1}=(\iota_{k-1}\tilde{S}_{k-1,t-1}|d_t)$, the result then follows from Lemmas A.2 and A.4 and the mapping theorem of Billingsley (1999, p.20). □

Proof of Theorem 4.7. We use the proof of Theorem 4.6 and further note that by Lemma 4.4 we may instead of $\mathsf{CUSQ}^{\mathsf{REC}}$ consider $\mathcal{S}^{\mathsf{CUSQ}}$, written for a continous mapping $h:D[0,1]\mapsto D[0,1]$ and $u\in[0,1]$, as $\mathcal{S}^{\mathsf{CUSQ}}=h\{\sum_{s=1}^{\mathrm{int}(Tu)}(\varepsilon_s^2-\sigma^2)\}$. Then, applying Lemmas A.2 and A.5 and the mapping theorem, the result follows by setting $U_{t-1}=(\iota_{k-1}\tilde{S}_{k-1,t-1}|d_t)$. □

Proof of Theorem 4.8. We use the proofs of Theorems 4.6 and 4.7. To deal with the normality test, note that by Lemma 4.5, instead of $\mathsf{JB}^{\mathsf{omnibus}}$ we may consider $\mathcal{S}^{\mathsf{JB}}=k(T^{-1/2}\{\sum_{t=1}^{T}(\varepsilon_t-\bar{\varepsilon})^3/\sqrt{6\sigma^6},\ (\sum_{t=1}^{T}(\varepsilon_t-\bar{\varepsilon})^4-3\sigma^4)/\sqrt{24\sigma^8}\})$, for the purposes of the asymptotic arguments, where $k:\mathbb{R}^2\mapsto\mathbb{R}_+$, is a continous mapping. The result follows by letting $U_{t-1}=(\iota_{k-1}\tilde{S}_{k-1,t-1}|d_t)$ and applying

Lemmas A.3 and A.6 together with the mapping theorem. □

Proof of Remark 4.9. Let $U_{t-1} = (\iota_k S_{k,t-1} | d_t)$ and proceed as in proofs of Theorems 4.6, 4.7, and 4.8, omitting the unit root test and therefore application of Lemmas A.4, A.5, and A.6. □

Chapter 6

Concluding Remarks

This thesis was written as a contribution to a research programme aiming to improve the reliability of statistical methods used in empirical econometric modelling by reducing context dependence of the standard toolbox.

The need to account for a variety of patterns in temporal dependence, not only integration of different orders, but also the explosive behaviour of macroeconomic series observed under exceptional circumstances, or in the presence of so called bubbles, can mean that the practitioner sometimes finds himself without sufficient theoretical backup. One instance of this is the practice of mis-specification testing. Integral to the modelling exercise, this largely automated practice is carried out prior to conducting inference, and effectively serves as a quality check on the results. In the absence of substantive adequacy, a model's passing of a battery of diagnostics has limited relevance, of course; verification of statistical adequacy should be viewed as a necessary but not sufficient pre-condition for proceeding with inference. Given the automatic reporting of diagnostics by software, the modeller is required to make a judgement as to whether this verification is itself reliable in the model under scrutiny. One of the characteristics of good

mis-specification tests is the obviation of such judgements, and this requires distribution theory that applies with some sufficient degree of generality. Similarity or asymptotic similarity with near invariance in finite samples and robustness to irrelevant mis-specifications ensure that a mis-specification test procedure truly checks statistical adequacy without introducing prejudice to subsequent inference, or to the outputs of other mis-specification test procedures.

Understanding of the stochastic properties of the mis-specification test battery requires both piecemeal work on the distribution theory of each test and some understanding of the stochastic interactions between the tests and between the tests and subsequent inference procedures. Our contribution on the first point is to generalize the CUSUM of squares tests to a model class where a variety of temporal dependence patterns in the data is allowed and to show asymptotic similarity (Chapter 2) with some caveats (Chapter 3), and near invariance in finite samples (Chapter 4). On the second point, our contribution is to characterize the joint distribution of some mis-specification testing and inference procedures and to show asymptotic independence and near uncorrelatedness in finite samples (Chapter 5).

A number of problems remain unsolved. Concerning the joint testing problem, the main results of Chapter 5 could be developed to apply at the same level of generality as those of Chapter 2. In particular, an extension to the case where the data are generated by a vector autoregression, possibly with explosive roots, and with the possible inclusion of a contemporaneous stochastic regressor at the level of the model, is an objective for future research. While the current argument is general enough to cover a range of autoregressive models, these extensions could make it more persuasive. The addition of more mis-specification tests into the battery considered could also be desirable. Concerning the asymp-

totic theory of the CUSUM of squares tests, further research on the application of non-parametric procedures to the problem, such as the use of kernels (see e.g. Hansen (2008)), as well as on the repeated application of the test for monitoring purposes would be welcome. Concerning the nuisance terms arising in the mixed explosive/non-explosive case, more work on the behaviour of these terms, numerical or analytical, could help determine exactly how important a role they play. The simulation experiments reported in Chapter 4 suggest that in the case of the second order autoregression, the mixed case does not necessarily cause problems in practice. Instead, there is evidence of a size distortion in single unit root case, when deterministic terms, and in particular a linear trend, are included in the model. There is scope for further simulation experiments to add to the evidence on this case, to be able of determine how much of a knife edge phenomenon it is.

Bibliography

Alston, J. M., and J. A. Chalfant (1991): "Unstable models from incorrect forms," *American Journal of Agricultural Economics*, 73, 1171–1181.

Anderson, G. J., and G. E. Mizon (1989): "What can statistics contribute to the analysis of economic structural change?," in *Statistical analysis and forecasting of economic structural change*, ed. by P. Hackl, pp. 3–21. Springer-Verlag.

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

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

——— (2003): "End-of-sample instability tests," *Econometrica*, 71, 1661–1694.

Bartlett, M. S. (1955): *An introduction to stochastic processes*. Cambridge University Press.

Basu, D. (1955): "On statistics independent of a complete sufficient statistic," *The Indian Journal of Statistics*, 15, 377.

Bera, A. K., and C. M. Jarque (1982): "Model specification tests: a simultaneous approach," *Journal of Econometrics*, 20, 59–82.

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

Blanchard, O., and J. Simon (2001): "The long and large decline in U.S. output volatility," *Brookings Papers on Economic Activity*, 2001, 135–164.

Bowman, K. O., and L. R. Shenton (1975): "Omnibus test contours for departures from normality based on $\sqrt{b_1}$ b_2," *Biometrika*, 62, 243–250.

Breusch, T. S. (1978): "Testing for autocorrelation in dynamic linear models," *Australian Economic Papers*, 17, 334–355.

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

Caceres, C. (2007): "Asymptotic properties of tests for mis-specification," D.Phil thesis, University of Oxford.

Campbell, J. Y., and R. J. Shiller (1988): "Stock prices, earnings, and expected dividends," *The Journal of Finance*, 43, 661–676.

Carsoule, F., and P. H. Franses (2003): "A note on monitoring time-varying parameters in an autoregression," *Metrika*, 57, 51–62.

Cavaliere, G., D. I. Harvey, S. Leybourne, and A. R. Taylor (2011): "Testing for unit roots in the presence of a possible break in trend and non-stationary volatility," *Econometric Theory*, 27, 957–991.

Cavaliere, G., A. Rahbek, and R. Taylor (2010): "Testing for co-integration in vector autoregressions with non-stationary volatility," *Journal of Econometrics*, 158, 7–24.

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

Chow, G. C. (1960): "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28, 591–605.

Christ, C. F. (1994): "The Cowles Commission's contributions to econometrics at Chicago, 1939-1955," *Journal of Economic Literature*, 38, 30–59.

Chu, C.-S. J., M. Stinchcombe, and H. White (1996): "Monitoring structural change," *Econometrica*, 64, 1045–1065.

Davidson, J. (1994): *Stochastic limit theory (Advanced Texts in Econometrics)*. Oxford University Press.

——— (2000): *Econometric theory*. Blackwell Publishers.

——— (2006): "Asymptotic methods and functional central limit theorems," in *Palgrave handbook of econometrics, vol. 1*, ed. by K. Patterson, and T. C. Mills, pp. 159–214. Palgrave Macmillan.

Davidson, R., and J. G. MacKinnon (1993): *Estimation and inference in econometrics*. Oxford University Press.

deBruijn, N. G. (1961): *Asymptotic methods in analysis*. North-Holland Publishing Co.

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

——— (2008b): "A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change," *Journal of Econometrics*, 142, C26–C52.

Dhyrmes, P. J., E. Howrey, S. H. Hymans, J. Kmenta, E. E. Leamer, R. E. Quanot, J. B. Ramsey, H. T. Shapiro, and V. Zarnowitz (1972): "Critieria for evaluation of econometric models," *Annals of Economic and Social Measurement*, 1, 291–325.

Doornik, J. A. (2006): "The role of simulation in econometrics," in *Palgrave handbook of econometrics, vol. 1*, ed. by K. Patterson, and T. C. Mills, pp. 787–811. Palgrave Macmillan.

——— (2007): *An Object-oriented Matrix Programming Language Ox^{TM} 5*. Timberlake Consultants Ltd.

——— (2009): "Autometrics," in *The methodology and practice of econometrics: a Festschrift in honour of David F. Hendry*, ed. by J. Castle, and N. Shephard, pp. 88–122. Oxford University Press.

Doornik, J. A., and H. Hansen (2008): "An omnibus test for univariate and multivariate normality," *Oxford Bulletin of Economics and Statistics*, 70, 927–939.

Doornik, J. A., and D. F. Hendry (2007a): *Empirical econometric modeling: PcGive^{TM}12, Volume I*. Timberlake Consultants Ltd.

——— (2007b): *Empirical econometric modeling: PcGive^{TM}12, Volume II.* Timberlake Consultants Ltd.

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

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

Dufour, J.-M., and J. F. Kiviet (1995): "Exact tests for structural change in first-order dynamic models," Université de Montréal working paper.

Durbin, J. (1967): "Tests of serial independence based on the cumulated periodogram," *Bulletin of the International Statistical Institute*, 42, 1039–1049.

——— (1969): "Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals," *Biometrika*, 56, 1–15.

Durbin, J., and G. S. Watson (1950): "Testing for serial correlation in least squares regression: I," *Biometrika*, 37, 409–428.

Edgerton, D., and C. Wells (1994): "Critical values for the CUSUMSQ statistic in medium and large sized samples," *Oxford Bulletin of Economics and Statistics*, 56, 355–365.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983): "Exogeneity," *Econometrica*, 51, 277–304.

Engsted, T., and B. Nielsen (2010): "Testing for rational bubbles in a coexplosive vector autoregression," CREATES Research Paper 2010-25.

Fisher, R. A. (1930): "The moments of the distribution for normal samples of measures of departure from normality," *Proceedings of the Royal Society of London A*, 130, 16–28.

Godfrey, L. (1988): *Misspecification tests in econometrics (Econometric Society Monographs No. 16)*. Cambridge University Press.

Godfrey, L. G. (1978): "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables," *Econometrica*, 46, 1293–1301.

Haavelmo, T. (1944): "The probability approach in econometrics," *Econometrica*, 12, iii–vi + 1–115.

Hackl, P. (1980): *Testing the constancy of regression models over time (Angewandte Statistik Und Ökonometrie)*. Vandehoeck & Ruprecht.

Hackl, P., and A. H. Westlund (1991): *Economic structural change: analysis and forecasting*. Springer-Verlag.

Hall, P., and C. C. Heyde (1980): *Martingale limit theory and its application*. Academic Press.

Hansen, B. E. (1991): "A comparison of tests for parameter instability: an examination of asymptotic local power," University of Rochester working paper.

——— (1992a): "Testing for parameter instability in linear models," *Journal of Policy Modeling*, 14, 517–533.

——— (1992b): "Tests for parameter instability in regressions with I(1) processes," *Journal of Business & Economic Statistics*, 10, 321–335.

——— (2001): "The new econometrics of structural change: dating breaks in U.S. labor productivity," *Journal of Economic Perspectives*, 15, 117–128.

——— (2008): "Uniform convergence rates for kernel estimation with dependent data," *Econometric Theory*, 24, 726–748.

Harrod, R. F. (1938): "Scope and method of economics," *The Economic Journal*, 48, 383–412.

Harvey, A. C. (1976): "An alternative proof and generalization of a test for structural change," *The American Statistician*, 30, 122–123.

Harvey, A. C., and P. Collier (1977): "Testing for functional misspecification in regression analysis," *Journal of Econometrics*, 6, 103–119.

Hedayat, A., and D. S. Robson (1970): "Independent stepwise residuals for testing homoscedasticity," *Journal of the American Statistical Association*, 65, 1573–1581.

Helland, I. S. (1982): "Central limit theorems for martingales with discrete or continuous time," *Scandinavian Journal of Statistics*, 9, 79–94.

Hendry, D. F. (1984): "Monte Carlo experimentation in econometrics," in *Handbook of econometrics, vol. 2*, ed. by Z. Griliches, and M. D. Intriligator, pp. 937–976. Elsevier Science Publishers.

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

——— (1995): *Dynamic econometrics (Advanced Texts in Econometrics)*. Oxford University Press.

——— (1996): "On the constancy of time-series econometric equations," Invited lecture, Irish Economic Association.

——— (2009): "The methodology of empirical econometric modeling: applied econometrics through the looking glass," in *Palgrave handbook of econometrics, vol. 2*, ed. by K. Patterson, and T. C. Mills, pp. 3–67. Palgrave Macmillan.

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

Hidalgo, J., and M. H. Seo (2008): "Testing for structural stability in the whole sample," Working paper.

Hoover, K. D. (2006): "The methodology of Econometrics," in *Palgrave handbook of econometrics, vol. 1*, ed. by K. Patterson, and T. C. Mills, pp. 61–87. Palgrave Macmillan.

Hoover, K. D., and S. J. Perez (1999): "Data mining reconsidered: encompassing and the general-to-specific approach to specification search," *Econometrics Journal*, 2, 167–191.

——— (2000): "Three attitudes towards data mining," *Journal of Economic Methodology*, 7, 195–210.

Incl\'an, C., and G. C. Tiao (1994): "Use of cumulative sums of squares for retrospective detection of changes of variance," *Journal of the American Statistical Association*, 89, 913–923.

Jarque, C. M., and A. K. Bera (1980): "Efficient tests for normality, homoscedasticity and serial dependence of regression residuals," *Economics Letters*, 6, 255–259.

——— (1987): "A test for normality of observations and regression residuals," *International Statistical Review*, 55, 163–172.

Johansen, S. (1995): *Likelihood-based inference in cointegrated vector autoregressive models (Advanced Texts in Econometrics)*. Oxford University Press.

Johnson, N. J. (1978): "Modified t tests and confidence intervals for asymmetrical populations," *Journal of the American Statistical Association*, 73, 536–544.

Juselius, K. (2006): *The cointegrated VAR model (Advanced Texts in Econometrics)*. Oxford University Press.

Kilian, L., and U. Demiroglu (2000): "Residual-based tests for normality in autoregressions: asymptotic theory and simulation evidence," *Journal of Business & Economic Statistics*, 18, 40–50.

Kim, C.-J., and C. R. Nelson (1999): "Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle," *Review of Economic Studies*, 81, 608–616.

Kiviet, J. F. (1986): "On the rigour of some misspecification tests for modelling dynamic relationships," *Review of Economic Studies*, 53, 241–261.

Krolzig, H.-M., and D. F. Hendry (2001): "Computer automation of general-to-specific model selection procedures," *Journal of Economic Dynamics and Control*, 25, 831–866.

Kuan, C.-M., and K. Hornik (1995): "Generalized fluctuation test: a unifying view," *Econometric Reviews*, 14, 135–161.

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

——— (1983): "A note on martingale difference sequences satisfying the local Marcinkiewicz-Zygmund condition," *Bulletin of the Institute of Mathematics, Academia Sinica*, 11, 1–13.

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

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

Lee, S., and S. Park (2001): "The CUSUM of squares test for scale changes in infinite order moving average processes," *Scandinavian Journal of Statistics*, 28, 625–644.

——— (2009): "The monitoring test for the stability of regression models with nonstationary regressors," *Economics Letters*, 105, 250–252.

Lee, S., and C.-Z. Wei (1999): "On residual empirical processes of stochastic regression models with applications to time series," *The Annals of Statistics*, 27, 237–261.

Leybourne, S., R. Taylor, and T.-H. Kim (2006): "CUSUM of squares-based tests for a change in persistence," *Journal of Time Series Analysis*, 28, 408–433.

Loretan, M., and P. C. B. Phillips (1994): "Testing the covariance stationarity of heavy-tailed time series: an overview of the theory with applications to several financial datasets," *Journal of Empirical Finance*, 1, 211–248.

Lovell, M. C. (1983): "Data mining," *The Review of Economics and Statistics*, 65, 1–11.

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

Lütkepohl, H. (2007): *New introduction to multiple time series analysis*. Springer-Verlag.

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

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

McGuirk, A. M., P. Driscoll, and J. Alwang (1993): "Misspecification testing: a comprehensive approach," *American Journal of Agricultural Economics*, 75, 1044–1055.

Mizon, G. 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, 1221–1242.

Nielsen, B. (2001): "The asymptotic distribution of unit root tests of unstable autoregressive processes," *Econometrica*, 69, 211–219.

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

——— (2006): "Order determination in general vector autoregressions," *IMS Lecture Notes - Monograph Series*, 52, 93–112.

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

Nielsen, B., and J. J. Reade (2007): "Simulating properties of the likelihood ratio test for a unit root in an explosive second-order autoregression," *Econometric Reviews*, 26, 487–501.

Nielsen, B., and J. S. Sohkanen (2009): "Asymptotic Behaviour of the CUSUM of Squares Test under Stochastic and Deterministic Time Trends," Nuffield college working paper.

Nielsen, B., and J. S. Sohkanen (2011): "Asymptotic behaviour of the CUSUM of squares test under stochastic and deterministic time trends," *Econometric Theory*, 27, 913–927.

Nielsen, B., and A. Whitby (2011): "Asymptotic properties of the 1-step Chow test," forthcoming as a Nuffield college working paper.

Pagan, A. R. (1984): "Model evaluation by variable addition," in *Econometrics and quantitative economics*, ed. by D. F. Hendry, and K. F. Wallis. Basil Blackwell, Oxford.

Pagan, A. R., and G. W. Schwert (1990): "Testing for covariance stationarity in stock market data," *Economics Letters*, 33, 165–170.

Perron, P. (2006): "Dealing with structural breaks," in *Palgrave handbook of econometrics, vol. 1*, ed. by K. Patterson, and T. C. Mills, pp. 278–352. Palgrave Macmillan.

Phillips, G. D. A., and B. P. M. McCabe (1983): "The independence of tests for structural change in regression models," *Economics Letters*, 12, 283–287.

——— (1988): "Some applications for Basu's independence theorem in testing econometric models," *Statistica Neederlandica*, 42, 37–46.

——— (1989): "A sequential approach to testing for structural change in econometric models," *Empirical Economics*, 14, 151–165.

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

Phillips, P. C. B. (1987): "Time series regression with a unit root," *Econometrica*, 55, 277–301.

Phillips, P. C. B., and K.-L. Xu (2006): "Inference in autoregression under heteroscedasticity," *Journal of Time Series Analysis*, 27, 289–308.

Ploberger, W. (1989): "The local power of the CUSUM-SQ test against heteroscedasticity," in *Statistical analysis and forecasting of economic structural change*, ed. by P. Hackl, pp. 127–133. Springer-Verlag.

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

——— (1990): "The local power of the CUSUM and CUSUM of squares tests," *Econometric Theory*, 6, 335–347.

Ploberger, W., W. Krämer, and K. Kontrus (1989): "A new test for structural stability in the linear regression model," *Journal of Econometrics*, 40, 307–318.

Pyke, R. (1959): "The supremum and infimum of the Poisson process," *The Annals of Mathematical Statistics*, 30, 568 576.

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

Ramsey, J. B. (1974): "Classical model selection through specification error tests," in *Frontiers in econometrics*, ed. by P. Zarembka, pp. 13–46. Academic Press.

Rao, C. R. (1973): *Linear statistical inference and its applications (2^{nd} ed.)*. New York: Wiley.

Rodrigues, P. M. M., and A. Rubia (2011): "The effects of additive outliers and measurement errors when testing for structural breaks in variance," *Oxford Bulletin of Economics and Statistics*, 73, 449–468.

Sargan, D. J., and M. Desai (1988): *Lectures on advanced econometric theory*. Basil Blackwell Oxford.

Savin, N. E., and K. J. White (1978): "Estimation and testing for functional form and autocorrelation," *Journal of Econometrics*, 8, 1–12.

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

Sensier, M., and D. van Dijk (2004): "Testing for volatility changes in U.S. macroeconomic time series," *Review of Economic Studies*, 86, 833–839.

Sohkanen, J. S. (2008): "Properties of tests for constancy of residual variance in autoregressions," M.Phil thesis, University of Oxford.

Spanos, A. (1986): *Statistical foundations of econometric modeling*. Cambridge University Press.

——— (2000): "Revisiting data mining: 'hunting' with or without a license," *Journal of Economic Methodology*, 7.

——— (2006a): "Econometrics in retrospect and prospect," in *Palgrave handbook of econometrics, vol. 1*, ed. by K. Patterson, and T. C. Mills, pp. 3–58. Palgrave Macmillan.

——— (2006b): "Where do statistical models come from? Revisiting the problem of specification," *IMS Lecture Notes - Monograph Series*, 49, 98–119.

——— (2007): "Using the residuals to test the validity of a statistical model: revisiting sufficiency and ancillarity," Virginia Tech working paper.

——— (2009): "Philosophy of Econometrics," Virginia Tech working paper.

Stock, J. H., and M. W. Watson (2002): "Has the business cycle changed and why?," *NBER Macroeconomics Annual*, 17, 159–218.

Summers, L. H. (1991): "The scientific illusion in empirical macroeconomics," *The Scandinavian Journal of Economics*, 93, 129–148.

Tanizaki, H. (1995): "Asymptotically exact confidence intervals of CUSUM and CUSUMSQ tests," *Communications in Statistics - Simulation and Computation*, 24, 1019–1036.

Thiele, T. N. (1889): *Almindelig Iagttagelseslre: sandsynlighedsregning of mindste kvadraters methode*. Copenhagen: C.A. Reitzel.

von Landau, E. (1927): *Vorlesungen über Zahlentheorie (Aus der analytischen und geometrischen Zahlentheorie)*. Leipzig, S. Hirzel.

Westlund, A. H., and B. Törnkvist (1989): "On the identification of time for structural changes by MOSUM-SQ and CUSUM-SQ procedures," in *Statistical analysis and forecasting of economic structural change*, ed. by P. Hackl, pp. 97–126. Springer-Verlag.

White, H. (1994): *Estimation, inference and specification analysis (Econometric Society Monographs No. 22)*. Cambridge University Press.

Xu, K.-L. (2008): "Testing against nonstationary volatility in time series," *Economics Letters*, 101, 288–292.